Assessing the Gains from E-Commerce[*]

Paul Dolfen

Stanford

Liran Einav

Stanford and NBER

Peter J. Klenow

Stanford and NBER

Benjamin Klopack

Stanford

Jonathan D. Levin

Stanford and NBER

Larry Levin

Visa, Inc.

Wayne Best

Visa, Inc.

October 31, 2018

Abstract

E-commerce represents a rapidly growing share of U.S. retail spending. We use transactions-level data on credit and debit cards from Visa, Inc. between 2007 and 2017 to quantify the resulting consumer surplus. We estimate that e-commerce spending reached 8% of consumption by 2017, yielding consumers the equivalent of a 1% permanent boost to consumption, or over $1k per household. The gains arose from saving the travel costs of buying from local merchants, but more important from accessing merchants online who do not have local brick-and-mortar stores. Cardholders with incomes above $50k gained more than twice as much (as a % of consumption) than households below $50k. Perhaps surprisingly, consumers in densely populated counties gained more as a % of consumption.

1.   Introduction

According to the U.S. Census Bureau, e-commerce spending doubled as a share of retail sales from 2007 to 2017, reaching 10% of overall sales. In addition to large online-only megastores, many traditional brick-and-mortar retailers have launched online entities that sell the same products available in the retailer's physical stores.

For consumers, shopping online differs in important ways from visiting a brick-and-mortar store. Because online retailers are less constrained by physical space, they can offer a wider variety of products.[1] E-commerce also enables consumers to access stores that do not have a physical location near them. Finally, consumers can purchase a product online that they may have previously purchased at a brick-and-mortar store without making a physical trip. We refer to these as variety gains and convenience gains, respectively.

In this paper we attempt to quantify the benefits for consumers from the rise of online shopping by leveraging a large and dataset of consumer purchases: the universe of Visa credit and debit card transactions between 2007 and 2017. In 2017, roughly 22% of consumption flowed through Visa. Our data include detailed information on each transaction, but no personally identifiable information about individual cardholders. We begin by describing the features of this unique dataset and presenting some descriptive facts on the growth of e-commerce.

To quantify the convenience gains from e-commerce, we posit a simple binary choice model of consumer behavior in which consumers decide whether to make a purchase at a given merchant's online or offline sales channel. Each consumer is defined by her location in geographical space relative to the location of a retailer. We show that a consumer located farther away from a given merchant's brick-and-mortar store is more likely to buy online. We use this

---

[1]Brynjolfsson et al. (2003) found that the number of book titles available at Amazon was 23 times larger than those available at a typical Barnes & Noble. Quan and Williams (2016) document a related pattern in the context of shoes.

distance gradient, estimates of the cost of travel, and information on the distribution of distances of each merchant's customers to estimate the convenience value of shopping online. Using this within-merchant substitution, we estimate that gains from convenience reached no more than 0.4% of consumer spending by 2017.

To quantify the variety gains from e-commerce, we write down a richer model in which variety-loving consumers can adjust the number of merchants they visit online and offline. The gains here are increasing in the share of spending online, and decreasing in the degree of substitutability between online and offline spending. We estimate substitutability by exploiting how spending at online vs. offline merchants varies as a function of consumer distance to each offline merchant. We again convert travel distance into dollars. We also use cross-sectional variation across cards to estimate how much consumers are willing to trade off shopping at a greater variety of merchants vs. spending more at each merchant. Within this framework, we estimate consumer gains from increased spending online to be about 1.1% of all consumption by 2017. This is tantamount to $1,150 per household in 2017. The gains are twice as large – even as a % of consumption – for richer households (income above $50k per year) than poorer households (below $50k per year), and are higher in more densely populated counties.

Our work is related to several papers that attempt to quantify the benefit to consumers from the internet. Goolsbee and Klenow (2006) develop an approach based on the time spent using the internet at home. Using estimates of the opportunity cost of time, they estimate surplus for the median consumer of 2-3% of consumption. Brynjolfsson and Oh (2012) use a similar approach that also considers data on internet speed and the share of time spent on different websites. They estimate the value from free digital services alone to be roughly 1% of consumption. Varian (2013) estimates the value of time savings from internet search engines. Syverson (2017) looks at the question of whether the observed slowdown in labor productivity can be explained by mismeasurement

of digital goods and ICT more generally. He concludes that surplus from ICT is not large enough to explain much of the productivity slowdown, which exceeds 1% *per year* for over a decade.

Our paper also contributes to a broader literature that tackles the question of how to measure consumer surplus from new products. Broda and Weinstein (2010) and Redding and Weinstein (2016) estimate the value of variety using AC Nielsen scanner data. Broda and Weinstein (2006) quantify the value of rising import variety. Brynjolfsson, Hu, and Smith (2003) look at the gains to consumers from accessing additional book titles at online booksellers.[2]

The rest of the paper is organized as follows. Section 2 introduces the data and how we construct some of the key variables. Section 3 presents summary statistics and initial facts. Sections 4 and 5 estimate the convenience and variety gains, respectively, from e-commerce. Section 6 briefly concludes.

2.   Data and Variable Construction

Our primary dataset is the universe of all credit and debit card transactions in the United States that were cleared through the Visa network between January 2007 and December 2017.[3] We complement the Visa data with data from a major credit reporting bureau, as well as information at the county level from both the U.S. Census and Internal Revenue Service.

The unit of observation in the raw data is a signature-based (not PIN-based) transaction between a cardholder and a merchant. We observe the transaction amount, the date of the transaction, a unique card identifier, the type of card (credit or debit), and a merchant identifier and ZIP code (but not street address). The merchant identifier is linked by Visa to the merchant's name and

---

[2]Quan and Williams (2016) make and illustrate the important point that, if demand is location-specific, then representative consumer frameworks can overstate variety gains..

[3]The Visa network is the largest network in the market. It accounted for 40 to 50% of credit card transaction volume and over 70% of debit card volume over this period, with Mastercard, American Express, and Discover sharing the rest of the volume; see, e.g., https://wallethub.com/edu/market-share-by-credit-card-network/25531.

industry classification (NAICS). In contrast, cards used by the same person or household are not linked to each other, and information about the cardholder is limited to what one could infer from the card's transactions. That is, our sample is completely anonymized, and we do not observe the name, address, demographics, or any other personally identifiable information about the cardholder.

The 2007–2017 Visa data contain an annual average of 380 million cards, 35.9 billion transactions, and $1.93 trillion in sales. Of these sales, 55% were credit transactions and 45% were debit transactions. Figure 1 presents Visa spending as a share of U.S. consumption and nominal GDP, respectively. Visa volume has been steadily increasing over time, from approximately 14% of consumption in 2007 to almost 22% of consumption in 2017.[4] In Section 4 below, where we focus on substitution between online and offline channels within a merchant, we further limit the analysis to the five retail NAICS categories where the online transaction share was between 10% and 90%.[5]

Key variables. Each transaction indicates whether it occurred in person ("CP" for Card Present, meaning that the card was physically swiped) or not ("CNP" for Card Not Present). Most CP transactions are broken further into e-commerce, mail order, phone order, and recurring transactions. We treat phone, mail, and recurring transactions as offline. For some CNP transactions, this further breakdown is missing. We assume the e-commerce fraction of such missing values that is the same as the fraction of non-missing CNP values that is classified as e-commerce. Denoting ECI as the E-Commerce Indicator within CNP transactions, we infer e-commerce spending within 3-digit NAICS years as

---

[4]Our analysis sample uses all transactions between 2007 and 2017 that pass standard filters used by the Visa analytics team. We exclude transactions at merchants not located in the U.S., those not classified as sales drafts, and those that did not occur on the Visa credit/signature debit network. (Transactions not involving sales drafts include chargebacks, credit voucher fees, and other miscellaneous charges.) We also drop cards that transact with fewer than 5 merchants over the card's lifetime, as many of the dropped cards are specialized gift cards.

[5]The Census Bureau NAICS 44 and 45 are Retail Trade. Based on their online transaction share in the Visa data, we use merchants in the following five categories to estimate convenience gains: furniture and home furnishings stores; electronics and appliance stores; clothing and clothing accessories stores; sporting goods, hobby, musical instruments and book stores; miscellaneous store retailers.

Figure 1: Visa spending as a share of consumption and GDP
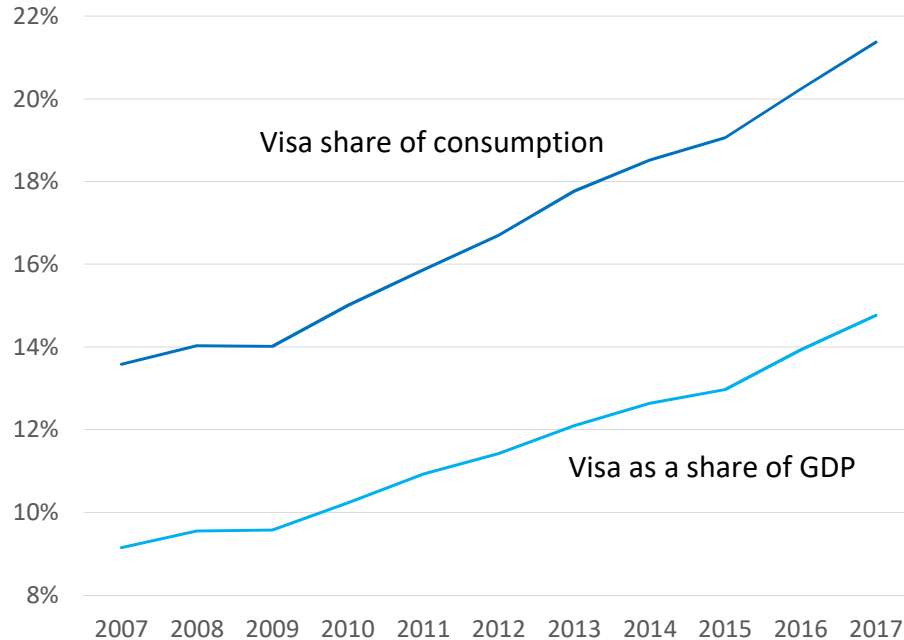
$$\text{e-commerce} = \frac{\text{ECI}}{\text{ECI} + \text{phone/mail/recurring}} \times \text{CNP}$$

Figure 1 lists the NAICS that contain a nontrivial share of spending with the ECI flag. This includes some retail and non-retail NAICS. It excludes NAICS such as Communication, which contains ample CNP spending on cell phone bills but which occurs predominantly through recurring payments. The non-retail NAICS with a significant ECI presence are all related to travel and transportation. We include these NAICS on the grounds that they provide some of the same convenience and variety as online options in retail NAICS (e.g. booking travel online rather than visiting or calling travel agent). That said, we will show the robustness of our results to concentrating on the retail NAICS only.

Two other important variables in our analysis are card location and income. For card location, we can infer a card's preferred shopping location from its transaction history. Recall that we observe the 5-digit ZIP code of the merchant for each offline transaction. We use this to define a card's location as a longitude-latitude pair given by the transaction-weighted average ZIP centroid.[6] Using this location variable, we then construct a distance variable for each offline transaction, which is given by the straight-line distance between the longitude-latitude pair of the card and the ZIP centroid of the merchant (recall that we do not observe the merchant's street address).

For about 50% of the credit cards in 2016 and 2017, we have more precise information about the cardholder address as well as income from a large credit rating agency. We use this location as a robustness check on our estimates with shopping centroid. We use income to break down online spending shares and the gains from e-commerce by affluence.

3. Summary Statistics and Initial Facts

The growth of online spending. We start by documenting the increasing importance of online spending during our sample. Table 2 documents the rising share of online spending within Visa in selected NAICS. The online share was already quite high in 2007 in some categories, such as air transport. And in some categories, such as food, the online share remained low in 2017.[7]

To estimate the share of online spending in *all* U.S. consumption, we first scale up Visa online spending by the inverse of Visa's share in national credit and debit card spending. This assumes Visa spending is representative of all card spending in terms of its online share, and that all spending online occurs through debit and credit cards. Finally, we divide by overall U.S. consumption

---

[6]In doing so, we limit attention to ZIP codes in which the card transacted often enough (we use 20 transaction per ZIP over the card's lifetime) in order to omit transactions that were not part of the card's primary purchasing area. This means that less active cards also are excluded from our analysis that uses card location.

[7]NAICS such as gasoline had essentially no online spending in either year.

of goods and services (including the service flow from housing):[8]

$$\text{U.S. online share} = \frac{\text{Total U.S. card spending}}{\text{U.S. Consumption}} \times \text{Visa online share}$$

Figure 2 shows our estimates of the share of online spending in all consumption from 2007 and 2017, growing from about 5% of spending in 2007 to almost 8% in 2017. Defined more narrowly using retail NAICS, the online share rose from about 3.5% in 2007 to 5% in 2017.

Figure 2: Share of Visa spending online
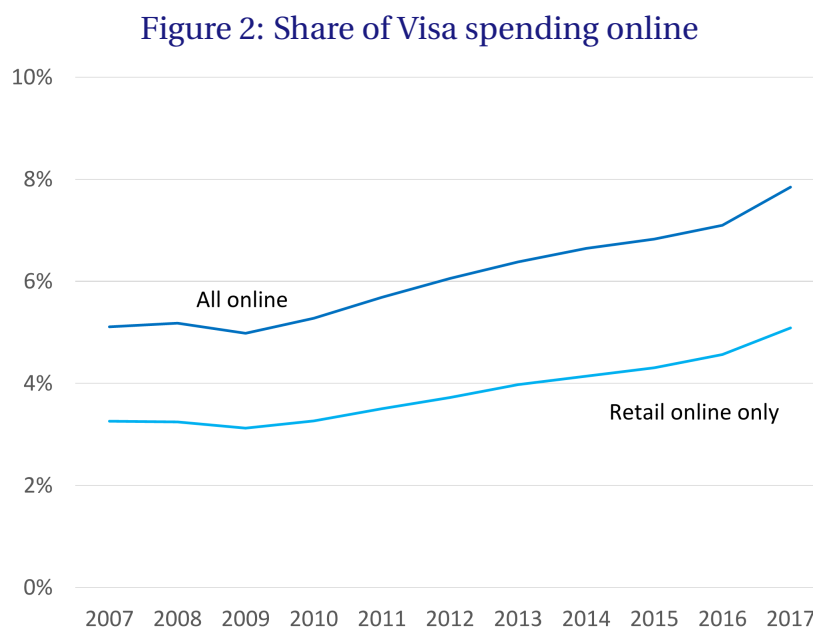
Heterogeneity by income and population density. There are two primary channels by which consumers likely benefit from the increased availability of the online channel: convenience and availability. From a convenience perspective, e-commerce allows consumers to avoid the trip to the offline store, and the potential time and hassle costs associated with parking, transacting, and carrying

---

[8]The Federal Reserve Bank of Boston estimates the fraction of all consumer spending transacted using debit and credit cards.

Table 1: E-Commerce NAICS

Retail NAICS	Example
Nonstore Retailers	Amazon
Clothing	Nordstrom
Miscellaneous Retail	Staples
General Merchandise	Walmart
Electronics	Best Buy
Building Material, Garden Supplies	Home Depot
Furniture	Bed Bath & Beyond
Sporting Goods, Hobby	Nike
Health, Personal Care	CVS
Food	Safeway
Ground Transportation	Uber

Non-Retail NAICS	Examples
Admin. Support Services	Expedia Travel
Air Transportation	American Airlines
Accommodation	Marriott
Car Parts	AutoZone
Rental Services	Hertz Rent-a-Car

Table 2: Visa online shares (% of spending in each NAICS)

	2007	2017
Nonstore Retailers	90	96
Air Transport	87	97
Electronics	42	51
Furniture	35	43
Clothing	22	37
General Merchandise	8	15
Food	5	6

home the purchased items. It seems plausible that these convenience benefits are largest for more affluent consumers.

The availability benefits might be particularly important for consumers who live in more rural areas, where there are fewer offline merchants. E-commerce is essentially available to everyone everywhere, thus making many more merchants available to consumers.

Even though we observe (estimated) income for about one-half of Visa credit cards in 2016 and 2017 through a credit bureau, not all households have credit or debit cards. To adjust for the card-less, we scale down the Visa online spending share in a given county-income pair by the ratio of Visa cards to the number of IRS tax return filers and dependents in that county-income pair:

$$s_{cy} = \frac{\text{Visa online spending}_{cy}}{\text{Total Visa spending}_{cy}} \cdot \alpha_{cy}$$

where s_{cy} is our estimate of the online share of all consumption for income group y in county c, and α_{cy} is our estimate of the share of households with cards in that group:

$$\alpha_{cy} = \frac{\text{\# of Visa Cards}_{cy}}{\text{Tax Filers}_{cy}}$$

Again, we are assuming online spending only occurs through credit and debit cards, so that the cardless are not online at all.

Figure 3 plots the online share of consumption for four income groups and ten county population density deciles in 2016. The share is 3-4% for households with incomes below \$50k, and 6-14% for households with incomes above \$50k. Within each income group, the online share is increasing in population density, particularly for high incomes. This is perhaps surprising because the density of brick-and-mortar retailers is increasing in population density.

Figure 4 displays our online share estimates for all U.S. counties. Online penetration is distinctly higher in the Northeast and in the West and Mountain regions than in the South or Midwest.

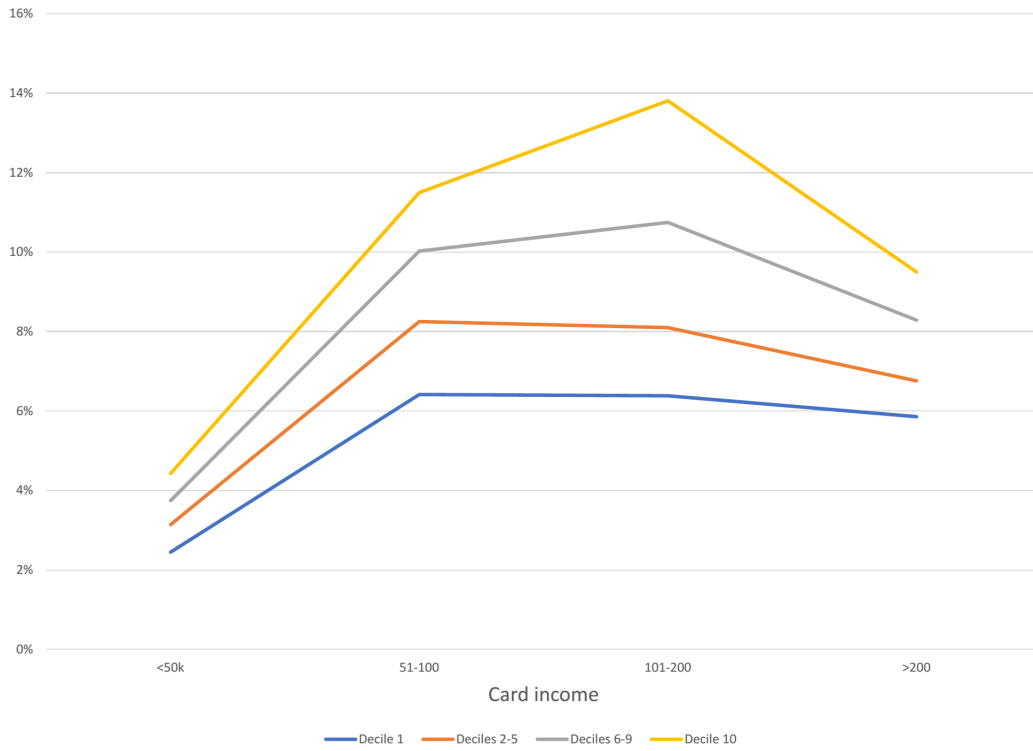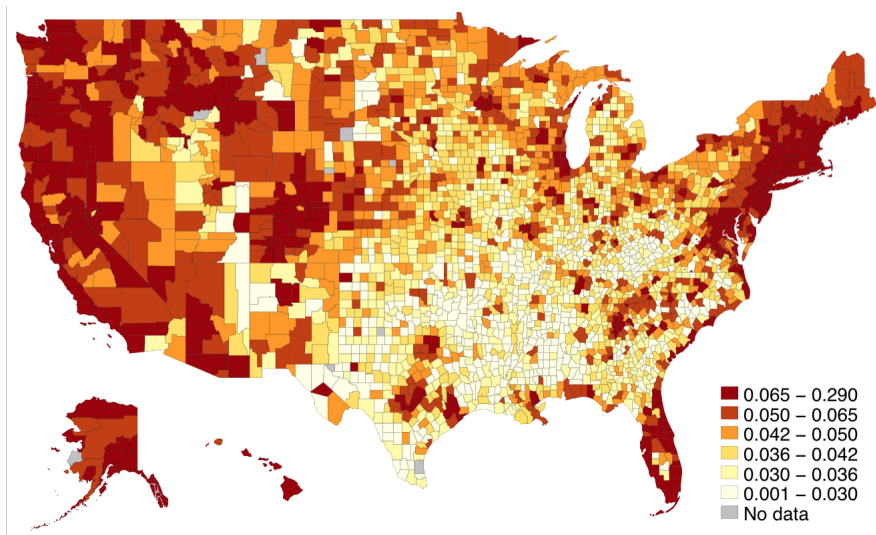## Figure 3: Online shares by card income and density

Figure 4: Online shares by county

4. Estimates of Convenience Surplus

In this section we focus on a specific gain from e-commerce: the ability it provides to avoid the physical shopping trip to a brick and mortar store, and instead buy the same basket of goods from the same merchant via its e-commerce channel. Given that e-commerce provides a wider choice set of merchants than what would otherwise be available to consumers, this direct convenience gain would surely be smaller than the overall gain, which accounts for merchant substitution. Yet, it seems natural to begin assessing the gain from convenience given that doing so is simpler and requires fewer modeling assumptions.

Specification. To quantify these convenience gains, we estimate a simple binary choice between online and offline transaction. We assume that consumer i has full information of the items she plans to buy, the merchant she plans to buy the items from, and the associated prices. We make the strong assumption that prices are the same online and offline for a given merchant, consistent with evidence in Cavallo (2017). The only remaining choice is thus whether to transact online or offline.

   We assume the utility for consumer i of making an online purchase at merchant j is given by

$$u_{ij}^o = \gamma_j^o + \epsilon_{ij}^o, \tag{1}$$

where γ_j^o is the average merchant-specific utility from the online channel and ϵ_{ij}^o is an online consumer-merchant component, which we assume is drawn from a type I extreme value distribution, iid across merchants and consumers.

   Meanwhile, we assume the utility for consumer i of making an offline purchase at merchant j is given by

$$u_{ij}^b = \gamma_j^b - \beta \cdot dist_{ij} + \epsilon_{ij}^b, \tag{2}$$

where γ_j^b is the average merchant-specific utility from the offline channel, and $dist_{ij}$ is the straight-line distance between the location of consumer i and the

nearest store of merchant j.[9] ϵ_{ij}^{b} is an offline consumer-merchant component, which we assume is similarly drawn from a type I extreme value distribution, iid across merchants and consumers.

Equations (1) and (2) give rise to a simple logit regression of an indicator variable that is equal to 1 for an online purchase (and 0 for an offline purchase) on distance $dist_{ij}$ and merchant fixed effects.

Estimation and results. We estimate this logit specification on a random sample of 1% of all cards in 2017 for which we observe the home zipcode. To capture merchants where the choice of online and offline is meaningful, we use transactions in the five mixed-channel retail categories (described in the previous section) where the consumer was within 50 miles of the offline store.

Table 3 presents summary statistics for this sample. Online transactions account for 15-30% of the overall number of transactions and for 25-40% of the total dollar amount, except for electronics where the online share of transactions is much greater (47% of transactions). The most robust pattern in Table 3 is the distance of the consumer to the nearest offline store, which is systematically shorter for offline transactions than for online ones. This is the key variation which we rely on in the analysis below.

Figure 5 pools across the five retail categories, and relates the online share to distance in the raw data, as well as the estimated relationship using the logit specification. As expected, the online share increases with distance. That is, as the nearest brick-and-mortar store is further away, the online channel becomes relatively more attractive, and the online share increases. Comparing cases where the offline store is nearby to cases where the offline store is 30-50 miles away, the online share roughly triples, from approximately 14% to 45%.

Using our logit specification, we estimate a β coefficient of -0.013 (with a standard error less than 0.0001), which implies that moving a consumer from 10 to 20 miles away from a physical store increases the share of purchases made

---

[9]The store location is recorded by Visa as a latitude-longitude pair, while the location of the consumer is based on the centroid of the zip+4 billing zipcode.

Table 3: Summary statistics by NAICS

NAICS	Furniture	Electronics	Clothing	Sport, Music, and Books	Misc. stores
NAICS code	442	443	448	451	453
Transactions	711,178	932,867	3,570,316	1,780,257	1,391,438
Online share					
Transactions	0.168	0.474	0.265	0.253	0.220
Spending	0.244	0.295	0.302	0.238	0.393
Ticket size (dollars):					
Offline	130.9	235.2	79.7	61.8	49.2
	(10.7 - 237.0)	(11.3 - 617.0)	(13.7 - 158.2)	(6.8 - 134.9)	(6.6 - 100.4)
Online	209.3	90.7	96.4	57.1	114.6
	(23.2 - 419.3)	(5.0 - 158.9)	(15.6 - 194.9)	(4.9 - 133.6)	(17.6 - 196.3)
Distance to nearest offline store (miles):					
Offline	7.0	6.0	6.3	7.2	4.7
	(1.1 - 16.5)	(1.1 - 13.4)	(1.0 - 14.2)	(1.3 - 16.9)	(0.7 - 10.6)
Online	8.7	12.8	9.1	10.9	10.5
	(1.4 - 21.4)	(31.8 - 0.0)	(23.8 - 0.0)	(28.4 - 0.0)	(27.9 - 0.0)

The table shows summary statistics for the transactions used in the convenience analysis. The ticket size panel gives the average dollars per transaction for each NAICS and channel (online or offline). Distance to the nearest store is calculated as the as-the-crow-flies distance between a consumer's location and the nearest offline branch of the merchant where the transaction was made. The first row in each of the bottom two panels contains the average ticket size or distance. The numbers below, in parentheses, are the 10th and 90th percentiles.

online by approximately 0.9 percentage points (evaluated at the average distance from the offline store).

Estimates of convenience gains. This simple model allows us to estimate the value of e-commerce in a straightforward way. We can evaluate the consumer surplus from e-commerce using the difference between consumer surplus when both online and offline options are available and when only the offline option is available. Applying the well-known properties of the extreme value distribution, the convenience gain of each transaction by consumer i at merchant j is

$$\Delta CS_{ij} = \frac{ln[exp(\gamma_j^b - \beta \cdot dist_{ij}) + exp(\gamma_j^o)] - (\gamma_j^b - \beta \cdot dist_{ij})}{\beta}. \qquad (3)$$

It is important to note that we do not observe and therefore do not use prices in our analysis. The consumer surplus expression uses travel distance as a determinant of the full price. To monetize miles, we multiply the number

Figure 5: Online share vs. distance to merchant store
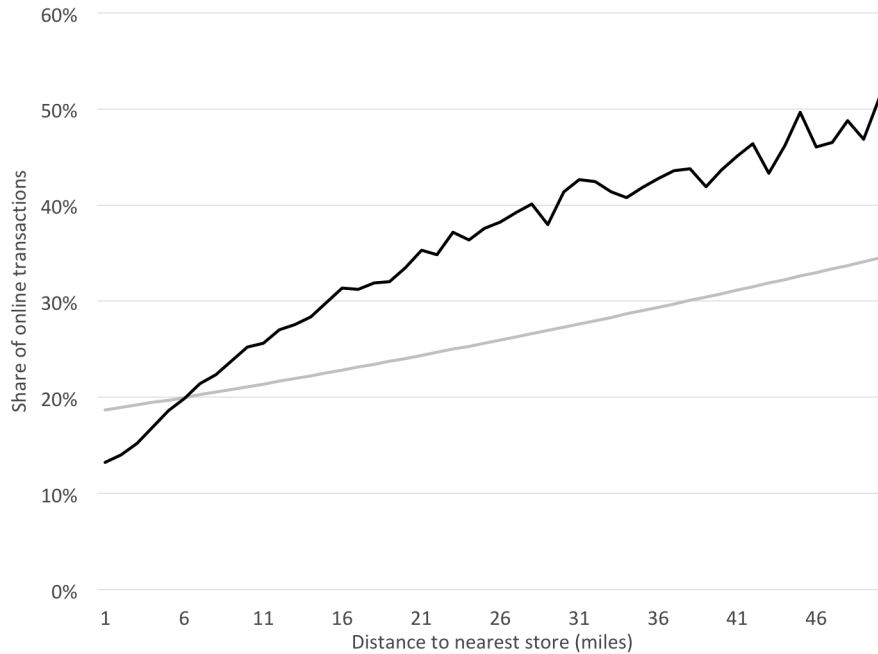
of miles by two to get the roundtrip distance, and assume that each mile costs $0.80 in time costs and $0.79 in direct costs, for a total of $1.59 for each one-way mile and $3.18 for each round-trip mile between the consumer and the store.[10]

Applying equation (3) to all the transactions in our data, we obtain an average convenience gain (across all transactions in the sample) of 11.3 mile equivalents. Using the conversion factor above ($3.18 per round trip mile), the convenience gain per transaction comes to $36 dollars.

The average ticket size in our sample is $96 and the average distance between consumer and store is 10 miles. Thus convenience gains from the online option are on the order of 28% for purchases in the five NAICS categories used in the estimation. Together, transactions in these five categories made by con-

---

[10]To obtain the monetary cost of a mile, we use estimates from Einav et al. (2016), who report summary statistics for a large number of short-distance trips of breast cancer patients. They report that an average trip takes 10.9 minutes to travel 5.3 straight-line miles, with an actual driving distance of 7.9 miles. The BLS reports that the average hourly wage from 2007-2017 was $23 per hour after tax. As an estimate for the driving cost, we use the average of the IRS reimbursement rate from 2007-2017 of $0.535 per mile, which considers the cost of fuel and depreciation of the car. Thus, the time cost of driving one mile is given by $23/60 \cdot 10.9/5.3 = \$0.80$ and the driving cost of of one mile is $0.535 \cdot 7.9/5.3 = \$0.79$.

sumers who were closer than 50 miles to an offline outlet of the same merchant make up about 7% of all dollars, implying that the total convenience gains as a share of Visa spending of about 2%, or roughly 0.4% of all consumption.

5. Estimates of variety surplus

While the model in the previous section allows us to place some quantitative bounds on an important benefit from e-commerce, it does not allow for substitution across merchants, thereby ignoring potential consumer gains from access to a wider variety of shopping options.

This channel may be first order. The set of merchants that consumers visit online and offline are largely different. To illustrate this, in Table 4 we show the proportion of online spending that occurred at merchants where a given card also shopped offline. Each entry in the table gives the share of online spending by the amount the same card spent offline at that merchant. For example, the entry in the first row, third column shows that 10% of total online sales were made at merchants for which cards spent $0 offline and between $10 and $100 online. The table shows that 88% of online spending occurred at merchants that were not visited offline, suggesting that cross-merchant substitution may be a predominant source of consumer surplus.

5.1. Model Setup

To capture these gains from variety, we write down a stylized model that allows substitution across merchants and calibrate it using moments calculated from the Visa data.

Consumer problem. Consumers allocate spending across a set of M merchants in both online and offline channels, and must pay fixed costs that are increasing in the number of merchants visited. Consumers maximize:

Table 4: Within-card merchant overlap between online and offline spending

		Online spending for card-merchant in…					
		0	($0,$10)	[$10,$100)	[$100,$500)	>$500	Total
Offline spending for card-merchant in…	0	0.00	0.00	0.10	0.31	0.47	0.88
	($0,$10)	0.00	0.00	0.00	0.00	0.00	0.00
	[$10,$100)	0.00	0.00	0.01	0.01	0.02	0.04
	[$100,$500)	0.00	0.00	0.00	0.01	0.02	0.04
	>$500	0.00	0.00	0.00	0.00	0.04	0.04
	Total	0.00	0.00	0.11	0.34	0.54	1.00

Each cell in the table gives the share of total online spending in 2014 by the amount of offline and online dollars spent at a given merchant by a card. Each observation in the underlying data is a card-merchant combination with an entry for offline and online spending. For example, the cell in the first row and third column contains the share of online dollars corresponding to card-merchant combinations where a card spent $0 offline at a merchant and between $10 and $100 online at that same merchant. The "total" row (column) gives the sum of the cells across all columns (rows) in that row (column). All cells (excluding the total row and column) sum to 1.

$$\max U = \left[\sum_{m=1}^{M} (q_m \cdot x_m)^{1-\frac{1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \tag{4}$$

subject to

$$M_b^{\phi} F_b + M_o^{\phi} F_o + \sum_{m=1}^{M} p_m \cdot x_m \le w$$

and

$$M = M_b + M_o$$

where q_m is the "quality" of merchant m, x_m is the quantity purchased from them, M_b (M_o) is the number of merchants shopped at in-store (online), $F_b(F_o)$ are the fixed costs of shopping in-store (online), and w is the consumer's wage income (the same as the nominal wage given a fixed unit of labor supply per consumers).

The parameter σ is the elasticity of substitution across merchants. Values of $\sigma < \infty$ imply a "love of variety." The parameter ϕ governs how fast fixed costs to visiting merchants increase with the number of merchants visited. We assume

that $\phi > 1$ to get an interior solution.[11]

Merchant problem. Merchants choose prices to maximize their flow profits

$$\max_{p_m} \ \pi_m \ = \ p_m y_m - wL_m - wK_j$$

subject to

$$y_m = \frac{M_j}{M_{j,market}} L \, x_m \ \text{ and } \ y_m = Z_m L_m$$

where y_m denotes the total units sold across all consumers, L_m is the labor employed by the merchant, K_j is overhead labor, L is the total number of consumers, and Z_m is productivity for merchant m. Here $j = o$ or b, so overhead labor is allowed to differ for online versus offline merchants. $M_j \leq M_{j,market}$ – we make the simplifying assumption that brick-and-mortar (online) sellers are entertained by a random subset of the L_m consumers. Consumers decide how many merchants to buy from (and how much to buy) within the merchants they visit based on their CES demand above. Merchants are monopolistic competitors.

Shopping technology. Firms in transportation/internet sectors hire labor L_b to produce transportation services to help consumers access brick-and-mortar retailers, and hire labor L_o to provide internet/computer services to help consumers access online retailers:

$$L \cdot M_b^\phi \ = \ Y_b \ = \ A_b L_b$$

$$L \cdot M_o^\phi = \ Y_o \ = \ A_o L_o$$

This sector is perfectly competitive so that its firms price at marginal cost:

$$F_b \ = \ \frac{w}{A_b} \ \text{ and } \ F_o \ = \ \frac{w}{A_o}$$

---

[11]This convex cost specification can be thought of as a reduced-form for a menu of merchants with rising fixed costs of shopping at them.

The transportation/internet technologies therefore pin down the "intercept" of the convex costs of accessing merchants offline (picture driving longer distances to access more) and online (imagine some retailers provide more convenient account sign-up) given above. The share of consumer spending online may have risen, in part, because it has become easier to access online merchants due to rising A_o and therefore falling F_o.

Free entry and market clearing. We allow free entry because we want to capture the possibility that the rise of online spending has come at the expense of offline merchants. This could take the form of a shrinking number of brick-and-mortar merchants, *ceteris paribus*, cutting into the gains consumers enjoy from online spending.

For each market j, we assume that expected profits across merchants offline (online) are zero:

$$E_j[\pi_m] = 0$$

Thus the number of online and offline merchants is determined endogenously so that any variable profits just offset the cost of overhead labor. This follows the well-known Hopenhayn (1992) structure wherein firms pay the overhead cost before observing their productivity draw Z_m. They enter to the point where expected profits is zero.

Meanwhile, labor market clearing requires

$$L = \sum_m L_m + M_{b,market} K_b + M_{o,market} K_o + L_b + L_o$$

as economywide labor is allocated to merchant production of consumer goods, merchant overhead, and transportation and internet services.

5.2. Model Solution

Symmetric technologies and outcomes. To focus on the online versus offline dimension, we now assume symmetry in many places. In particular, we assume

all merchants have the same process efficiency:

$$Z_m = Z$$

We assume all offline (online) merchants have the same quality, though we do allow quality to differ offline and online:

$$q_m = q_b \ \textbf{for} \ m \in M_{b,market}$$

$$q_m = q_o \ \textbf{for} \ m \in M_{o,market}$$

Because all merchants face the same wage, have the same process efficiency, and are monopolistic competitors facing the common elasticity of demand σ, they price at a common markup over their common marginal cost:

$$p_m = p = \frac{\sigma}{\sigma - 1} \cdot \frac{w}{Z}$$

With prices the same, consumers will spend the same amount ($p_m x_m$) at each offline and online merchant. Denote these spending levels o online and b offline. Due to differences in quality, spending per merchant online versus offline satisfies

$$\frac{o}{b} = \left(\frac{q_o}{q_b}\right)^{\sigma-1}.$$

The higher the quality online relative to offline, the higher the spending per merchant online relative to offline. Consumers may be shifting their spending online because of rising "quality" in the form of a growing variety of offerings online and/or growing ease of finding the right product online.

In turn, merchant profits online and offline are

$$\pi_o = \frac{M_o}{M_{o,market}} L \cdot \frac{o}{\sigma} - wK_o$$

$$\pi_b = \frac{M_b}{M_{b,market}} L \cdot \frac{b}{\sigma} - wK_b$$

In equilibrium, the number of merchants in the market and visited are

$$M_{b,market} = \frac{1}{1+k} \cdot \frac{1}{\sigma} \cdot \frac{(\sigma-1)\phi}{1+(\sigma-1)\phi} \cdot \frac{L}{K_b}$$

$$M_{o,market} = \frac{k}{1+k} \cdot \frac{1}{\sigma} \cdot \frac{(\sigma-1)\phi}{1+(\sigma-1)\phi} \cdot \frac{L}{K_o}$$

$$M_b = \left[\frac{1}{1+(\sigma-1)\phi} \cdot \frac{1}{1+k} \cdot A_b \right]^{\frac{1}{\phi}}$$

$$M_o = \left[\frac{1}{1+(\sigma-1)\phi} \cdot \frac{k}{1+k} \cdot A_o \right]^{\frac{1}{\phi}}$$

where $k \equiv \left(\frac{q_o}{q_b} \right)^{\frac{\phi}{\phi-1}(\sigma-1)} \left(\frac{A_o}{A_b} \right)^{\frac{1}{\phi-1}}$. The number of online merchants relative to offline merchants – both available and visited – increases in their relative quality (q_o/q_b) and ease of access (A_o/A_b).

The utility-maximizing share of spending online is

$$s_o \equiv \frac{oM_o}{oM_o + bM_b} = \frac{k}{k+1} \tag{5}$$

The online share s_o rises with q_o/q_b and A_o/A_b. Consumers gain from rising s_o if it is due to a combination of online options becoming better (rising q_o) and easier access to online merchants (rising A_o).

Consumption-equivalent welfare is proportional to

$$Z \cdot M^{1/(\sigma-1)} \cdot \bar{q}$$

where average quality is defined as

$$\bar{q} \equiv \left[\frac{q_b^{\sigma-1} \cdot M_b + q_o^{\sigma-1} \cdot M_o}{M} \right]^{1/(\sigma-1)}$$

welfare is increasing in process efficiency (Z) and the variety (M) and quality $(\bar{q})$ of merchants visited. In terms of exogenous driving forces, consumption-

equivalent welfare is proportional to

$$Z \cdot \left(q_b^{\frac{\phi}{\phi-1}(\sigma-1)} A_b^{\frac{1}{\phi-1}} + q_o^{\frac{\phi}{\phi-1}(\sigma-1)} A_o^{\frac{1}{\phi-1}} \right)^{\frac{\phi-1}{\phi} \frac{1}{\sigma-1}} \tag{6}$$

Consumers are better off if process efficiency rises (higher A), the quality of products available improves (higher q_b and q_o), and if shopping becomes easier offline (higher A_b) and/or online (higher A_o).

For given Z, q_b, A_b, consumer gains from rising q_o and A_o can be quantified from s_o, the share of spending online, and the values for parameters σ and ϕ. The gains are increasing in s_o, for given parameter values, as s_o is an increasing function of the quality and accessibility of online options. For given s_o, the gains are falling with σ. Consumers can more easily substitute from offline to online options when σ is higher, so online offerings do not need to improve as much (in quality or accessibility) to explain a given rise in online share. Also for given s_o, the gains are increasing in ϕ. The harder it is to add merchants visited online or offline, the bigger the improvement in the online option needed to explain a given rise in online share.

5.3.   Calibration of ϕ and σ

We first estimate ϕ, the parameter that governs the convexity of fixed costs with respect to the number of merchants visited. To do this, we use the observation that the level of ϕ affects the relationship between total expenditure ($oM_o + bM_b$), spending per merchant (o and b), and the number of merchants visited (M_o and M_b) across consumers. A higher value of ϕ gives rise to a steeper Engel curve on the intensive margin, with an elasticity of $1 - 1/\phi$ for spending per merchant, and a flatter Engel curve on the extensive margin, with an elasticity of $1/\phi$ for the number of merchants visited. We obtain an estimate for ϕ using empirical Engel curves.

Specifically, we exploit the following decomposition of spending into the extensive and intensive margins:

Table 5: Estimates of fixed shopping cost convexity

	2007	2017
$\widehat{\phi}$	1.73	1.75
# of cards	283M	462M
R^2	0.67	0.67

$$\ln M = \alpha + \frac{1}{\phi} \cdot \ln(oM_o + bM_b) \tag{7}$$

$$\ln\left(\frac{oM_o + bM_b}{M}\right) = \eta + \frac{\phi - 1}{\phi} \cdot \ln(oM_o + bM_b) \tag{8}$$

where $M = M_o + M_b$. To consistently estimate the parameter ϕ from (7) and (8) via OLS, we must assume that any idiosyncratic fixed shopping costs are uncorrelated with total spending across consumers.[12]

In Table 5, we present our estimates for ϕ. We perform the estimation separately for 2007 and 2017. Across the two years the average point estimate is 1.74. The standard errors are too small to mention given the hundreds of millions of cards in each regression. A ϕ of 2 would imply that 50% of additional card spending is on the extensive margin and 50% is on the intensive margin. Our estimate is modestly below 2, implying the extensive margin accounts for 57% of and the intensive margin accounts for 43% of variation across high and low spending cards.

---

[12]Since the decomposition is exact, the estimate of ϕ will be identical regardless of which of the two equations is used.

To calibrate σ, the elasticity of substitution across merchants, we use variation induced by differences in physical distance between each card i and each merchant j. Assuming that variation in distance is uncorrelated with individual preferences (conditional on chain fixed effects), we can use substitution patterns across merchants as a function of relative distance to identify σ.

We estimate across-merchant substitution using purchases for the 1% sample of cards in 2017, as described in Section 4. For each card i, we look at online purchases as well as offline purchases made within 20 miles of i's location. We construct, for each individual i and NAICS, all pairs of physical stores j and k such that i buys in one of these stores and compute $|dist_{ij} - dist_{ik}|$. We also construct all pairs of physical stores j and online merchants k such that i buys from one of these. We then calculate the share of combined trips for each pair that were made to the farther (or physical) store, and average across cards for each NAICS. In Figure 6, we show the fraction of trips to the farther store as a function of the relative distance between the two stores.

Figure 6: Relative trips as a function of distance
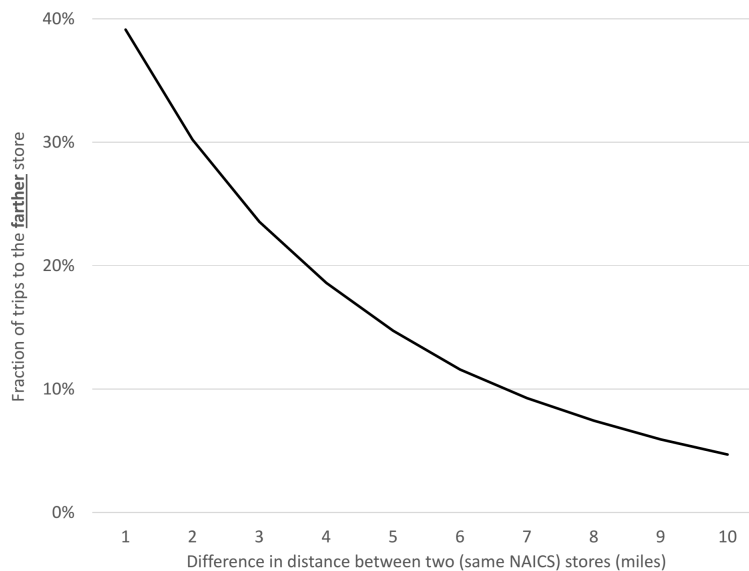
As in the convenience analysis, we convert distance into effective price variation. We estimate a roundtrip mile involves \$3.18 in direct and indirect travel costs. We add these travel costs to the average ticket size of Visa transactions in the pair of merchants. This gives us the relative price of the total bundle — Visa ticket size plus travel costs — for going to the closer store (or shopping online) vs. the farther store (or the brick-and-mortar store). We then regress the log relative number of trips on log relative prices inclusive of travel costs, controlling for merchant fixed effects:[13]

$$\ln\left(\frac{Trips_j}{Trips_k}\right) = \ln\left(\frac{q_j}{q_k}\right) - \sigma \cdot \ln\left(\frac{p_{jk} + \tau_{ij}}{p_{jk} + \tau_{ik}}\right)$$

Here p_{jk} is average ticket size at merchants j, k; τ = transportation costs for i to j or k; and the fixed effects capture relative merchant quality. Again, we run regressions for both online-offline and offline-offline samples. The implicit residual in this regression is idiosyncratic preferences for merchants.

As shown in Table 6, we estimate an elasticity of substitution between competing online and offline merchants of $\widehat{\sigma} = 4.3$.[14] As this regression involves 3.6 million merchant pair observations, the standard errors are tiny. The high R^2 of 0.97 indicates that merchant fixed effects plus distance account for almost all variation in relative trips to merchants. Still, there could be endogeneity bias if people locate closer to merchants they prefer. This would actually bias our estimate of σ upward.

For comparison, Table 6 also reports our estimate of the elasticity of substitution across competing offline merchants. This is higher at $\widehat{\sigma} = 6.1$. Although our model preferences feature a common σ, we think the σ for online-offline competition is the relevant one for evaluating the gains to consumers from switching from offline to online spending. Still, we will report robustness of

---

[13]The number of trips corresponds to the quantities x_m in our model if we assume a fixed basket of items bought at the same prices across competing outlets.

[14]We do not compare online and offline arms of the same merchant, such as Walmart stores vs. Walmart.com. This elasticity is presumably higher than the cross-merchant elasticity. Recall that the vast majority of card online spending is at merchants that the card visits online only.

Table 6: Estimates of substitutability

	online-offline	offline-offline
$\widehat{\sigma}$	4.3	6.1
# of obs	3.6M	14.0M
R^2	0.97	0.94

our welfare calculation to using the higher σ across physical stores.

5.4. Consumer surplus

Using our estimates of ϕ and σ and the online share s_o calculated from the Visa data, we can calculate consumption-equivalent changes in consumer welfare from the rise of e-commerce. We present our estimates for these welfare gains in Table 7. Using our baseline estimates of ϕ and σ, we calculate an increase in consumer surplus equivalent to 0.4% between 2007 and 2017. Relative to a counterfactual where the online channel is completely unavailable, e-commerce in 2017 resulted in gains for consumers of 1.1% overall. These counterfactuals assume fixed levels of quality and accessibility offline (q_b and A_b) and fixed efficiency in producing goods (Z). Thus, they involve increasing quality and accessibility of online merchants (q_o and A_o) to account for the rising in the spending share of online merchants (s_o).

Table 7 also illustrates how the gains change with the parameter values. If we round ϕ up to 2 (from 1.74) the welfare gains increase from 1.1% to 1.2% of consumption. If we use the higher, offline σ of 6.1 (rather than 4.3) the gains fall to 0.7% of consumption. These sensitivity checks go in the expected direction.

Table 7: Consumption-equivalent welfare gains from e-commerce

	ϕ	σ	s_o^{2017} vs. s_o^{2007}	s_o^{2017} vs. $s_o = 0$
Baseline	1.74	4.3	0.4%	**1.1%**
High ϕ	*2*	4.3	0.4%	1.2%
High σ	1.74	*6.1*	0.2%	0.7%

As we highlighted in Section 3, the online share is not uniform across the U.S. population. Households with incomes above $50k and in more densely populated counties exhibited higher online shares. In Tables 8 and 9, we show consumption-equivalent welfare gains broken out by income groups and quartiles of county population density.[15]

Cardholders with income of $50k$ or less enjoyed gains equivalent to 0.5% of their consumption from online shopping. Richer households enjoyed twice the gains, in the range of $1.1 - 1.5\%$ of their consumption. The gains were also increasing in population density, rising from 0.8% for the sparest counties to 1.3% for the most densely populated counties.

We have framed these gains as a percentage of all consumption, but at least two other benchmarks are useful. One is expressing consumption-equivalent surplus as a share of online spending. Since e-commerce ends up at around 8% of consumption, by our estimate, surplus is equivalent to about 14% of e-commerce spending.[16] Another benchmark is consumer surplus from online

---

[15]We use the same ϕ and σ values of 1.74 and 4.3 for every group, but use group-specific online spending shares s_o. Each quartile contains approximately 25% of the population.

[16]This is modest compared to the Cohen, Hahn, Hall, Levitt and Metcalfe (2016) estimate of consumer surplus equal to 160% of spending on Uber.

Table 8: Welfare gains by cardholder income

Income in $	Gains from s_o^{2017} vs. $s_o = 0$
0-50k	0.5%
50k-100k	1.3%
100k-200k	1.5%
200k+	1.1%

Table 9: Welfare gains by county population density

	s_o^{2017} vs. $s_o = 0$
Quartile 1 (sparse)	0.8%
Quartile 2	1.0%
Quartile 3	1.2%
Quartile 4 (dense)	1.3%

spending as a share of Visa spending, which we can observe. (Recall that we assumed Visa was representative of all credit and debit card spending.) Online spending reached 21% of overall Visa spending in 2017. We calculate that surplus from online shopping equaled 3.1% of Visa spending in 2017.

Focusing on Visa alone enables us to entertain a nested CES structure as a robustness check.[17] One reason to do this is that substitutability is surely higher within NAICS than across NAICS, whereas our CES utility in equation (4) assumes the same elasticity within and across NAICS. We did estimate σ across NAICS within industries, keeping in mind that such substitutability was sure to be higher. By moving to a nested CES structure, we can also allow σ to vary by NAICS. We implement this for the 16 3-digit NAICS plus a grouping of all other consumer NAICS. For 10 of the 17 categories, there is a physical store component which allows us to estimate σ. For the other 7 categories, such as nonstore retailers, there is no or very limited brick and mortar presence; for them we simply assume the overall estimate of $\sigma = 4.3$. Table 10 provides the estimates for each category, ranked from most to least substitutability. It ranges from a high of 7.7 for building maerials and garden supplies to a low of 3.4 for electronics and appliance stores. We assume the upper nest, which aggregate these lower nests, is simply Cobb-Douglas.

In Table 11 we present our estimated welfare gains under nested CES. The first row is the baseline of a single nest for comparison, where the gains equal 3.1% of Visa spending. A key ambiguity that arises with the nests is how to treat nonstore retailers, which contains online-only retailers such as Amazon. When we treat it as its own nest, the welfare gain amounts of 5.6% of Visa spending. When we allocate nonstore retail spending based on estimates of Amazon's sales by NAICS, the gains come to 4.7%.[18] We prefer this latter estimate, as one would expect Amazon sales in electronics and appliances, clothing, etc. to fit into the nests for those respective NAICS. The welfare gains are larger

---

[17]This would be difficult to do for the whole economy, as we would need to extrapolate Visa spending to the rest of the economy in each NAICS.

[18]Source: eMarketer estimates for 2017.

Table 10: Estimates of substitutability by NAICS

	$\widehat{\sigma}$
Building Material, Garden Supplies	7.7
Motor Vehicle and Parts Dealers	7.5
Furniture and Home Furnishings Stores	7.4
General Merchandise Stores	5.8
Health and Personal Care Stores	5.5
Clothing and Clothing Accessories Stores	5.2
Miscellaneous Store Retailers	5.2
Nonstore Retailers	4.3
Air Transportation	4.3
Ground Transportation	4.3
Rental and Leasing Services	4.3
Administrative and Support Services	4.3
Accommodation	4.3
All others	4.3
Sporting Goods, Hobby, Music, Book Stores	4.2
Food and Beverage Stores	3.6
Electronics and Appliance Stores	3.4

Table 11: Nested CES Welfare Gains (% of Visa spending)

	2017
Single nest (baseline)	3.1%
17 Nests (nonstore retail its own nest)	5.6%
16 Nests (nonstore retail allocated)	4.7%

with this nested approach because it assumes (we think realistically) that there is more limited substitutability across NAICS than assumed by our baseline single CES formula. We hesitate to make this nested approach our baseline, however, because of the uncertainty in allocating nonstore retail spending to other NAICS, and in extrapolating Visa spending to all card spending by NAICS.

5.5. Producer surplus?

Our stylized model features free entry conditions for both offline and online merchants. As a direct result, the shift in consumer spending has no impact on producer surplus. Still, within the model we can ask what the rise of e-commerce did to brick-and-mortar merchants. Table 12 indicates the effect of rising q_o and A_o, holding fixed Z, L, q_b and A_b. Interestingly, the effects are rather modest: a 3.7% decline in spending at brick and mortar stores, with a 1.6% decline in spending per surviving physical store and 2.1% decline in the number of physical stores. The effect on the profits of brick-and-mortar retailers is zero by construction.

Table 12: Retail Apocalypse?

2007–2017 Change

b	–1.6%
M_b	–2.1%
$M_{b,market}$	–3.7%
Profits	0%

6. Conclusions

The advent of the internet, the rise of Amazon, and the increased popularity of e-commerce have dramatically changed the retail landscape over the last two decades. These changes had a huge impact on almost all retail sectors, in terms of both consumer surplus and producer surplus. In this paper, we focus on the consumer side. We take advantage of a unique data source — about half of all credit and debit transactions in the U.S. running through the Visa network – and attempt to quantify the consumer gains associated with the rise of e-commerce.

We report two estimates, capturing two types of gains likely associated with the increased availability of e-commerce. The first is the pure convenience gain, which we think of as the ability to purchase online instead of offline exactly the same set of items from the same merchant at the same prices. We estimate a binary consumer choice of online vs. offline transactions, and estimate that the convenience gains are equivalent to at most 0.4% of consumptiong. We then write down a stylized representative consumer model which allows for substitution across merchants and variety gains. Our main estimate using this model is a welfare gain equivalent to 1.1% of consumption in 2017. Given consumption

per household of roughly $100k per household in 2017, this amounts to about $1,100 per household.

Obviously, any single number that attempts to summarize such a dramatic change in purchasing behavior should be taken with great caution. First, surplus is likely to be even more heterogeneous than we have characterized – e.g., across product categories and consumer locations. Second, it relies on highly stylized modeling assumptions. Decomposing this estimate across products and consumers is a promising agenda for future work, as would be assessing the sensitivity of these estimates to alternative assumptions.

The Visa data is unique in its granularity and coverage, and as such allows us to obtain an estimate that covers multiple consumer sectors. At the same time, a primary limitation of the Visa data is that we only observe spending, not prices, and our primary strategy in this paper is to use variation in travel distance and monetize it. This type of analysis is complementary to existing work that uses more detailed data on transactions, albeit in a narrower context of data, such as books, shoes, or airlines. Finding ways to combine these narrower estimates from specific contexts and our more aggregate estimate from a broader set of data is yet another fruitful agenda for additional work.

References

Broda, Christian and David E Weinstein, "Product creation and destruction: Evidence and price implications," *The American Economic Review*, 2010, *100* (3), 691–723.

Brynjolfsson, Erik and JooHee Oh, "The attention economy: measuring the value of free digital services on the Internet," 2012.

_ , Yu Hu, and Michael D Smith, "Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers," *Management Science*, 2003, *49* (11), 1580–1596.

Cavallo, Alberto, "Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers," *The American Economic Review*, 2017, *107* (1), 283–303.

Cohen, Peter, Robert Hahn, Jonathan Hall, Steven Levitt, and Robert Metcalfe, "Using big data to estimate consumer surplus: The case of Uber," Technical Report, National Bureau of Economic Research 2016.

Einav, Liran, Amy Finkelstein, and Heidi Williams, "Paying on the margin for medical care: Evidence from breast cancer treatments," *American Economic Journal: Economic Policy*, 2016, *8* (1), 52–79.

Goolsbee, Austan and Peter J Klenow, "Valuing Consumer Products by the Time Spent Using Them: An Application to the Internet," *The American Economic Review*, 2006, *96* (2), 108–113.

Hopenhayn, Hugo A, "Entry, exit, and firm dynamics in long run equilibrium," *Econometrica*, 1992, pp. 1127–1150.

Quan, Thomas W and Kevin R Williams, "Product Variety, Across-Market Demand Heterogeneity, and the Value of Online Retail," 2016.

Redding, Stephen J and David E Weinstein, "A unified approach to estimating demand and welfare," Technical Report, National Bureau of Economic Research 2016.

Syverson, Chad, "Challenges to mismeasurement explanations for the US productivity slowdown," *Journal of Economic Perspectives*, 2017, *31* (2), 165–86.

Varian, Hal, "The value of the internet now and in the future," *The Economist*, 2013.