

# Assessing the Gains from E-Commerce\*

Preliminary – please don't quote or circulate without explicit permission

Liran Einav	Peter J. Klenow	Benjamin Klopck
Stanford and NBER	Stanford and NBER	Stanford
Jonathan D. Levin	Larry Levin	Wayne Best
Stanford and NBER	Visa, Inc.	Visa, Inc.

September 18, 2017

## 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 2014 to quantify the resulting consumer surplus. We estimate that the gains from e-commerce reached the equivalent of a 1.3% permanent boost to consumption by 2014, or about \$1,250 per household. The gains arose mostly from accessing a wider variety of merchants online, but also from saving the travel costs of buying items in brick-and-mortar stores. The richest counties gained roughly twice as much as the poorest counties (top vs. bottom quartiles), and densely populated counties gained more than sparsely populated counties.

---

\*We are grateful to Raviv Murciano-Goroff and Paul Dolfen for terrific research assistance, and to Sam Kortum for comments on an earlier draft. All results have been reviewed to ensure that no confidential information about Visa merchants or cardholders is disclosed.

## 1. Introduction

Over the last twenty years, e-commerce has grown swiftly in prominence. The U.S. Census Bureau reports that nominal e-commerce spending in the retail sector increased by 216% between 2002 and 2008, while offline retail spending increased by 24% during the same period (Lieber and Syverson (2012)). 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.<sup>1</sup> 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 detailed dataset of consumer purchases: the universe of Visa credit and debit card transactions between 2007 and 2014. 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. For example, the share of online spending in all Visa spending rose from 12.5% in 2007 to 22.5% in 2014.

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 loca-

---

<sup>1</sup>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.

tion 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 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, our preliminary estimates suggest that gains from convenience are about 1.0% of total spending on the Visa network.

To quantify the variety gains from e-commerce, we write down a richer model in which variety-loving consumers can substitute across merchants both online and offline. To pin down how much consumers are willing to substitute across merchants, we exploit the extent to which consumer spending at competing offline merchants varies as a function of consumer distance to each merchant. To do so, we again convert distance into dollars to relate the choice of merchant to the relative price of buying a given bundle of goods at competing merchants. 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 merchant variety to be about 3.6% of Visa spending and 1.3% of all consumption by 2014. This is equivalent to \$1,250 per household in 2014. The estimated gains are twice as large in richer counties (top vs. bottom quartile), and notably 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 find that surplus for the median consumer was around \$3,000 per year, or around 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 about \$100 billion per year, or roughly 1% of consumption. Varian (2013) estimates the value of time savings from internet

search engines. Syverson (2016) 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.

Our paper also contributes to a broader literature that tackles the question of how to measure consumer surplus from new products. Redding and Weinstein (2016) and Broda and Weinstein (2010) estimate the value of variety using AC Nielsen scanner data. Broda and Weinstein (2006) quantify the value of the increased availability of new goods via globalization. Brynjolfsson, Hu, and Smith (2003) look at the gains for consumers from accessing additional book titles at online booksellers.<sup>2</sup>

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. Section 4 estimates the convenience gains and Section 5 the variety gains from e-commerce. Section 6 briefly concludes.

## 2. Data and Variable Construction

The primary data for the analysis 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 2014.<sup>3</sup> We complement the Visa data with county-level information from the census.

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 ad-

---

<sup>2</sup>Quan and Williams (2016) make and illustrate the important point that if demand is location-specific these representative consumer frameworks, which we adopt as well, over estimate the variety gains.

<sup>3</sup>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>.

dress). The merchant identifier is linked by Visa to the merchant's name and 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 2007–2014 Visa data contain an annual average of 373 million cards, 31.9 billion transactions, and \$1.7 trillion in sales, split almost evenly between credit and debit transactions. Figure 1 presents the volume of transactions in the Visa data as a share of U.S. nominal GDP and consumption. Visa volume has been steadily increasing over time, from approximately 10% of GDP and 14% of consumption in 2007 to 13% and 20%, respectively, in 2014.<sup>4</sup> 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%.<sup>5</sup>

**Key variables.** Each transaction indicates whether it occurred in person (“card present”, meaning that the card was physically swiped) or not (“card not present”). “Card not present” transactions are broken into e-commerce, mail order, phone order, and recurring transactions. Throughout our analysis, we treat only the e-commerce transactions as online, and all other transactions (including phone, mail, and recurring) as offline or brick-and-mortar interchangeably.

Two other important variables in our analysis are card affluence and loca-

---

<sup>4</sup>Our analysis sample uses all transactions between 2007 and 2014 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.

<sup>5</sup>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

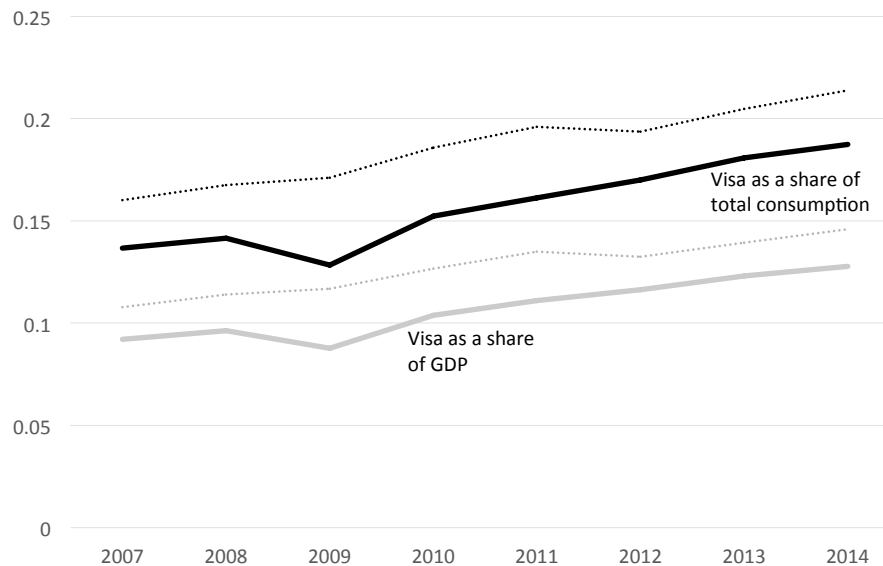


Figure presents, by year, the ratio of the total dollar amount transacted in the Visa data to consumption (black line) and GDP (gray line), with the latter two provided by the census. Dashed thin lines reflect the raw Visa data. Thick solid lines reflect the Visa data, after applying internal (to Visa) standard filters.

tion. As mentioned earlier, we infer a card's location from its transaction history. For card affluence, we use the average monthly spending on the card by year. That is, in each year, we use the ratio of the total spending by the card to the number of months over which the card was active (defined as the last month minus the first month with at least one transaction). Our measure results in a separate affluence for each card-year combination for cards that had at least 20 transactions in a year.

For card location, 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.<sup>6</sup> 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

<sup>6</sup>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.

(recall that we do not observe the merchant's street address).

### 3. Summary Statistics and Initial Facts

**The growth of online spending.** We start by documenting the increasing importance of online spending during our sample. Figure 2 shows that the share of online spending across all merchant categories nearly doubled between 2007 and 2014, growing from 12.5% of spending in 2007 to 22.5% in 2014.

Figure 2: Share of Visa spending online

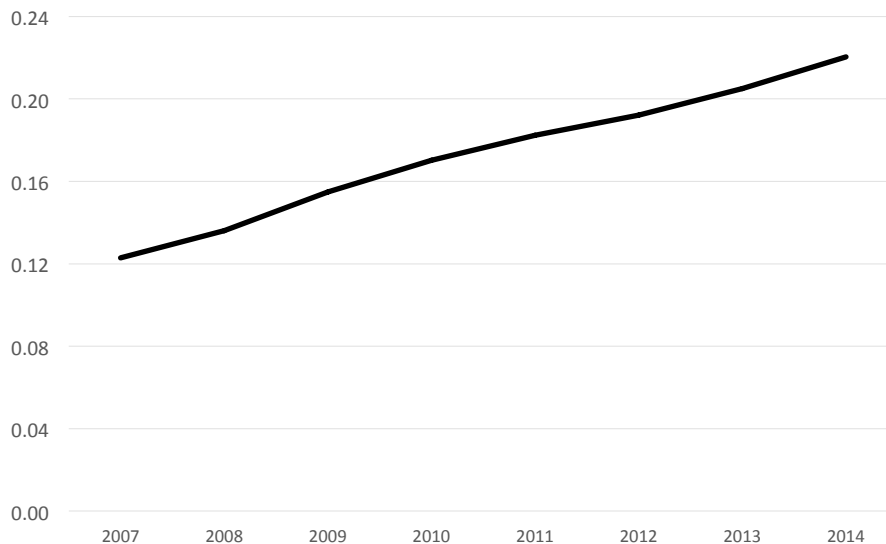


Figure presents, by year, the ratio of the total dollar amount transacted online to the total dollar amount transacted overall. Figure uses the Visa data, after applying internal (to Visa) standard filters. Online includes only transactions categorized as "e-commerce" by Visa. Offline includes in-person transactions (where a card was physically swiped) as well as transactions coded as mail order, phone order, and recurring transactions.

One way to decompose total spending in a channel (online or offline) is to write it as the product of three components: (i) the number of cards that transact; (ii) the average number of unique merchants a given cards transact at; and (iii) the average dollars spent per unique merchant. Table 1 presents this decomposition separately for online and offline transactions. The bottom two rows display each component online *relative* to offline. Interestingly, as indicated in the final row, each of the three components contributes about one-third to the overall doubling of the online share between 2007 and 2014.

Table 1: Decomposition of the growth in online share

	Share of cards transacting	Spending per unique merchant	Number of unique merchants
Online transactions			
In 2007	0.605	213.8	4.3
In 2014	0.740	264.3	5.5
Ratio	1.224	1.236	1.270
Offline transactions			
In 2007	1.000	174.4	23.3
In 2014	0.978	168.4	23.2
Ratio	0.978	0.966	0.996
Relative online growth	1.25	1.28	1.28
Contribution to overall online share	0.31	0.35	0.34

The first column in the table shows the share of total cards that transacted at least once online (in top panel) or offline (in bottom panel) in 2007 and 2014. The second column shows the total spending per merchant for cards that transacted online or offline. The third column shows the number of unique merchants visited per card for cards that transacted online or offline. In both of the top two panels, the Ratio row is calculated as the given variable in 2014 divided by its value in 2007. In the bottom panel, relative online growth is calculated as the online 2014/2007 ratio (top panel) divided by the offline 2014/2007 ratio (bottom panel). The contribution to overall online share is calculated as the relative online growth for the variable corresponding to the column divided by the sum of the relative online growth for all three variables.

As shown in Figure 3, the level and growth of online spending varies considerably across retail categories. While the share of online spending across all categories (with a single exception) is higher in 2014 relative to 2007, there is large heterogeneity. Some spending categories, such as food and general merchandise stores, remain almost entirely offline. Other categories, such as clothing and electronics, already had non-trivial online segment in 2007, and these are the categories whose online share increased the most.

**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



Figure 3: Growth in online share of spending by category, 2007-2014

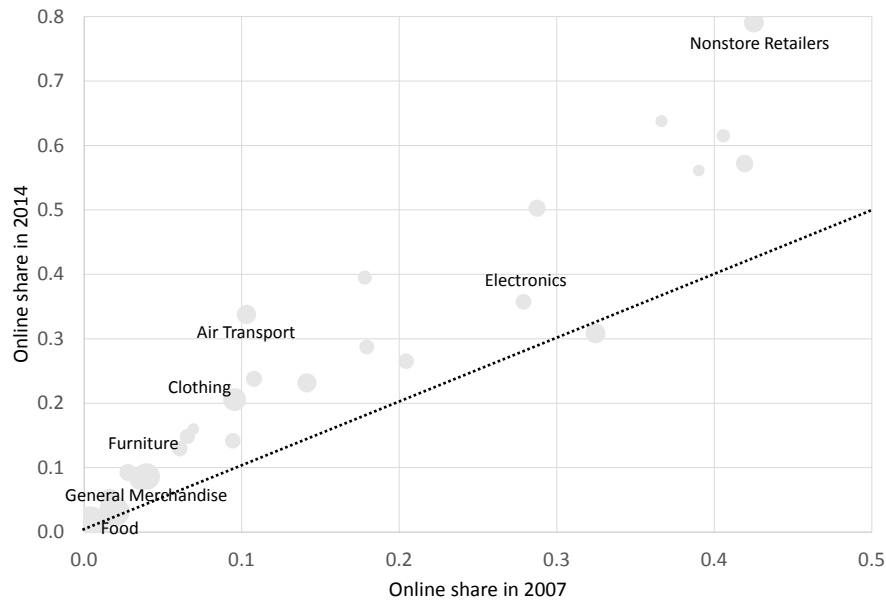


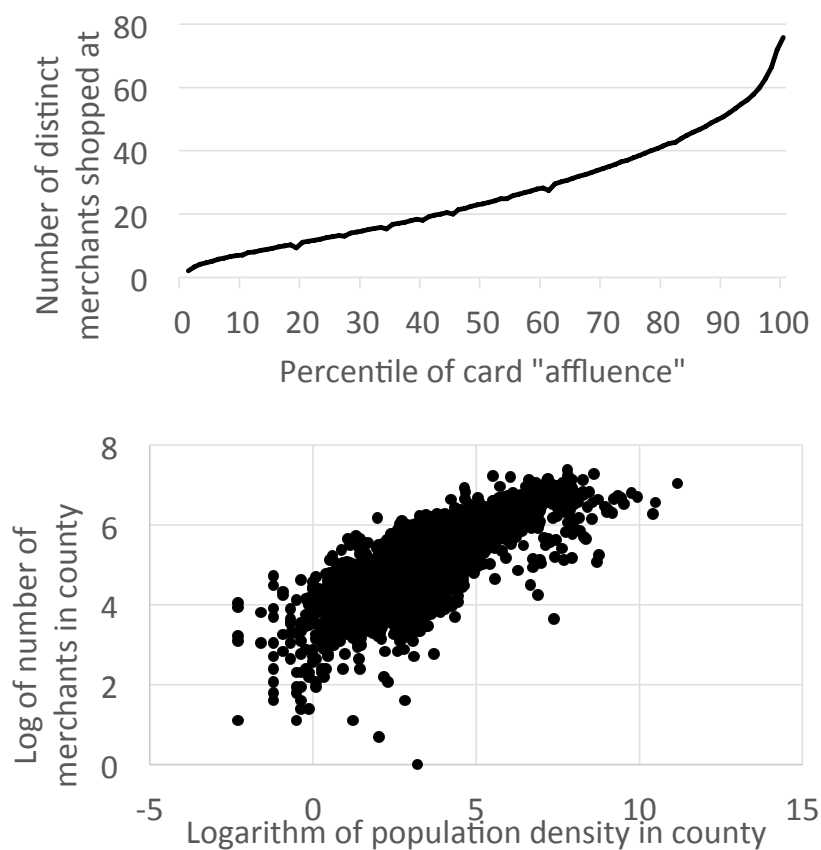
Figure presents the online share in 2014 against the online share in 2007, by industry category, for the largest 30 categories. Dashed line is the 45-degree line. Size of each bubble is proportional to its 2014 share of overall spending. Selected labels are displayed.

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.

To motivate this analysis, the top panel of Figure 4 shows how steeply the number of unique merchants ascends with the affluence of a card. The bottom panel illustrates that the number of available merchants in a county is roughly proportional to the population density in the county. Cards at the 90th percentile of affluence shop at approximately five times the number of merchants as cards at the 10th percentile of the affluence distribution — and thus may benefit more from the e-commerce channel via convenience. And the densest urban counties have 500-1000 distinct merchants while the least dense counties have less than 20 merchants — making the potential surplus from e-commerce high in such counties.

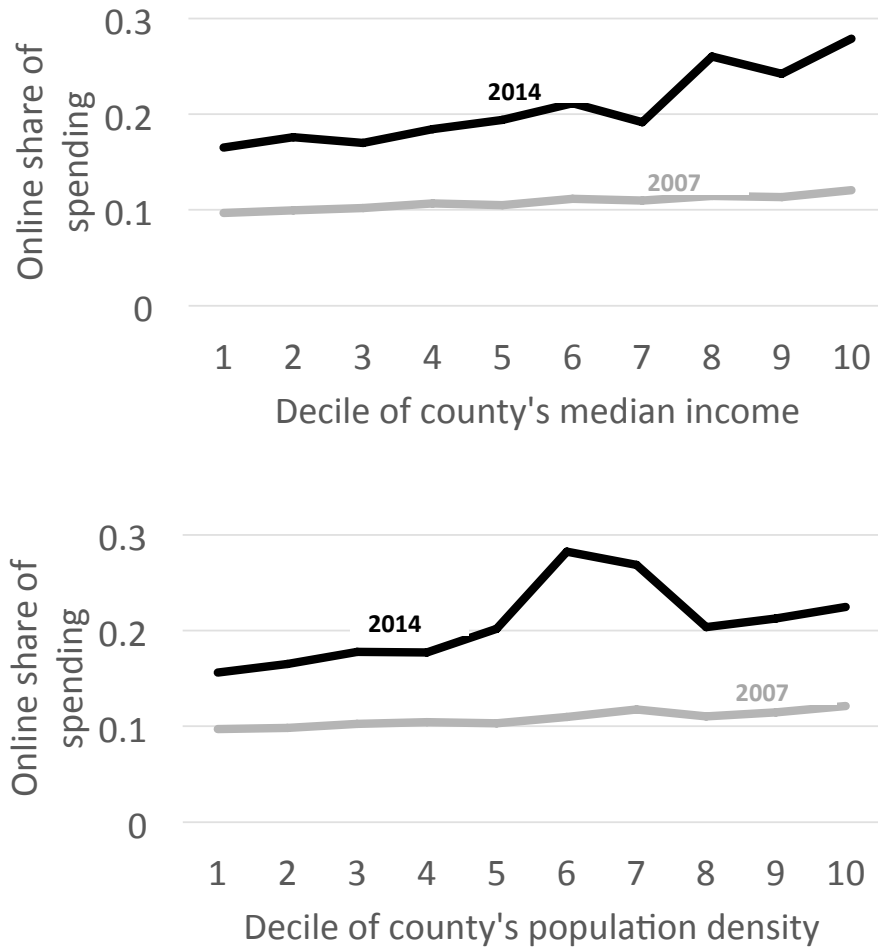
Figure 4: Unique merchants by card affluence and population density



Top panel plots the number of unique merchants shopped at for each card against the percentile of the card's affluence (average monthly spending). The bottom panel plots the log-log relationship between the number of available merchants in each county and the population density (per squared mile) in each county.

With these facts in mind, Figure 5 presents the growth in online share separately by decile of median county income (top panel) and decile of county population density (bottom panel). As one can see, the online spending share as of 2007 was quite uniform across these deciles. But from 2007 to 2014 the growth of e-commerce was more pronounced for the highest income counties and, to a lesser extent, for counties with greater population density. This suggests that merchant availability might be less important than convenience. The next two sections explore these channels more formally.

Figure 5: Online shares by county income and density



Top panel plots the share of online spending by county decile, where counties are sorted by their median income. Bottom panel repeats the same exercise, but counties are sorted by their population density. Cards are associated to counties based on the modal zip code in which they transact. Cards are placed in deciles so that each decile contains about 10% of the cards.

#### 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 (which are assumed to be the same online and offline, as is often the case according to 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

$$u_{ij}^o = \gamma_j^o + \epsilon_{ij}^o, \quad (1)$$

where  $\gamma_j^o$  is the average merchant-specific utility from the online channel and  $\epsilon_{ij}^o$  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

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

where  $\gamma_j^b$  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  $j$ .<sup>7</sup>  $\epsilon_{ij}^b$  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)

---

<sup>7</sup>Recall that, as described in section 2, the store location is given by the centroid of the store's ZIP code, while the location of the consumer is based on the transaction-weighted average location of the card's transactions.

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 2014. To capture merchants where the choice of online and offline is meaningful, we use transactions in the five mixed-channel retail categories described in Section 3, where the consumer was within 50 miles of the offline store.

Table 2 presents summary statistics for this sample. Online transactions account for 15-25% of the overall number of transactions and for 20-30% of the total dollar amount, except for electronics where the online share is much greater (63% of transactions and 43% of dollar volume). The typical online transaction is for a lower amount than an offline one, with electronics again being a major exception. The most robust pattern in Table 2 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 6 pools across the five retail categories, and presents the online share as a function of 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 more than doubles, from approximately 8% to 16-18%.

Using our logit specification, we estimate a  $\beta$  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 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

Table 2: Summary statistics by NAICS

NAICS	Furniture	Electronics	Clothing	Sport, Music, and Books	Misc. stores
NAICS code	442	443	448	451	453
Transactions	1,290,271	2,532,749	6,662,967	4,565,437	4,001,782
Online share					
Transactions	0.163	0.633	0.200	0.255	0.147
Spending	0.206	0.433	0.231	0.241	0.287
Ticket size (dollars):					
Offline	117.8 (9.8 - 214.2)	127.8 (5.0 - 292.3)	69.2 (11.9 - 139.1)	51.8 (5.9 - 113.0)	44.3 (5.3 - 86.4)
Online	157.9 (21.1 - 296.1)	56.5 (3.0 - 93.9)	83.0 (15.9 - 166.4)	48.1 (2.2 - 111.7)	103.3 (13.0 - 190.0)
Distance to nearest offline store (miles):					
Offline	14.4 (1.1 - 40.2)	11.2 (0.9 - 29.2)	12.7 (0.9 - 35.2)	13.7 (1.0 - 37.7)	10.9 (0.7 - 28.6)
Online	28.1 (1.6 - 85.7)	26.9 (1.3 - 83.2)	19.5 (1.1 - 59.3)	21.0 (1.3 - 64.8)	22.2 (1.1 - 69.4)

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.

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

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

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 of miles by two to get the roundtrip distance, and assume that each mile costs \$1.05 in time costs and \$0.80 in direct costs, for a total of \$3.71 for each one-way

Figure 6: Online share vs. distance to merchant store

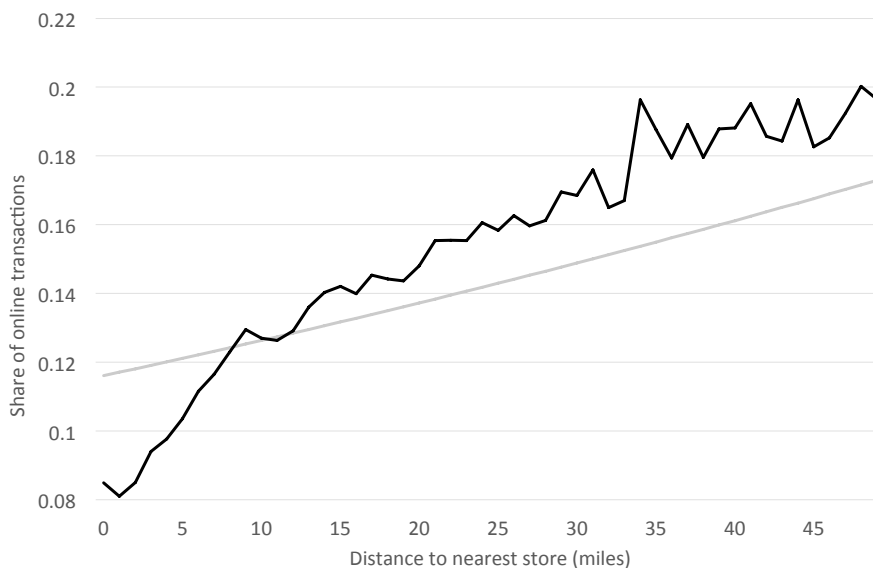


Figure presents the impact of distance to the nearest store on the propensity to shop online. Sample covers all transactions at merchants for which an online and an offline option (within 50 miles from the card holder) are available. The black line presents the share of online transactions as a function of distance to the nearest offline store of that merchant, where each point is the average online share within a one-mile distance bin. Gray line present the fitted logit regression, accounting for merchant fixed effects.

mile between the consumer and the store.<sup>8</sup>

Applying equation (3) to all the transactions in our data, we obtain an average convenience gain (across all transactions in the sample) of 16.6 mile equivalents. Using the conversion factor above (\$3.71 per round trip mile), the convenience gain per transaction comes to \$62 dollars.

A well known concern with such an estimate is that it is driven by the relatively fat tails of the logit distribution, which could substantially amplify the value we attribute to e-commerce. We thus decompose this estimate into two separate parts. First, we consider the direct value of e-commerce which is captured by removing the need to travel to the offline store. This gain is fully cap-

<sup>8</sup>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 2007-2014 was \$30.78 per hour. As an estimate for the driving cost, we use the average of the IRS reimbursement rate from 2007-2014 of \$0.537 per mile, which considers the cost of fuel and depreciation of the car. Thus, the time cost of driving one mile is given by  $\$30.78/60 \cdot 10.9/5.3 = \$1.05$  and the driving cost of of one mile is  $\$0.537 \cdot 7.9/5.3 = \$0.80$ .

tured by our estimate of  $\beta$ . The second is the gain from e-commerce attributable to the unobserved term, which is likely sensitive to the distributional assumptions. To do so, for each transaction, we integrate separately over the values of the error term, which would imply an online choice even if the offline choice was associated with  $dist_{ij} = 0$ , and then over the values which imply an online choice only because  $dist_{ij} > 0$ . This decomposition results in 4.2 miles and \$16. We view the latter as a lower bound for the convenience gains from e-commerce. This lower bound is likely to be less sensitive to distributional assumptions.

The average ticket size in our sample is \$62 and the average distance between consumer and store is 14 miles. Using the miles to dollars conversion described above, our lower bound on consumer surplus of 4.2 miles from the presence of the online option is worth about \$15.5. This implies that the gains from convenience from the online option are on the order of 14% for purchases in the five NAICS categories used in the estimation. Together, transactions in these categories made by consumers 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 are about 1%.<sup>9</sup>

## 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 3 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

---

<sup>9</sup>We obtain our 14% number as the ratio of the consumer surplus (\$15.5) to the total cost of a transaction ( $\$62 + \$3.71 \cdot 14 = 114\$$ ).



Table 3: Within-card merchant overlap between online and offline spending

		Online spending for card-merchant in...					Total
		0	(\$0,\$10)	[\$10,\$100)	[\$100,\$500)	>\$500	
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.

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 demand model that allows substitution across merchants and calibrate it using moments calculated from the Visa data. In our framework, 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:

$$\max U = \left[ \sum_{m=1}^M (b_m + q \cdot o_m)^{1-\frac{1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \quad (4)$$

subject to

$$M_b^\phi F_b + M_o^\phi F_o + \sum_{m=1}^M (b_m + o_m) \leq y$$

where  $b_m$  ( $o_m$ ) is the brick-and-mortar (online) spending at merchant  $m$ ,  $q$  is the relative “quality” of the online channel,  $M_b$  ( $M_o$ ) is the number of merchants shopped at in-store (online),  $F_b$  ( $F_o$ ) are the fixed costs of shopping in-store (online), and  $y$  is the consumer’s total expenditure.

This formulation implicitly assumes that all merchants charge the same prices, and that a given merchant charges the same prices online and offline.<sup>10</sup> In turn, the consumer’s utility-maximizing level of spending is equal across merchants within online and offline channels:  $o_m = o$  and  $b_m = b$ . In addition, online and offline spending are perfect substitutes, so a consumer will never shop both offline and online within a given merchant. This feature of the model is roughly consistent with the low levels of within-card merchant overlap we document in Table 3.

The parameter  $\sigma$  is the elasticity of substitution across merchants. Values of  $\sigma < \infty$  imply a “love of variety.” The parameter  $\phi$  governs the shape of the fixed cost function, or how fast fixed costs from visiting merchants increase with the number of merchants visited. We assume that  $\phi > 1$  to get an interior solution.<sup>11</sup>

## 5.2. Model Solution

Maximizing utility in (4) yields the optimal levels of  $o$ ,  $b$ ,  $M_o$  and  $M_b$ :

$$o = (\sigma - 1)\phi F_o^{\frac{1}{\phi}} \left[ \left( \frac{k}{k+1} \right) \left( \frac{1}{1 + (\sigma - 1)\phi} \right) y \right]^{\frac{\phi-1}{\phi}} \quad (5)$$

$$b = (\sigma - 1)\phi F_b^{\frac{1}{\phi}} \left[ \left( \frac{1}{k+1} \right) \left( \frac{1}{1 + (\sigma - 1)\phi} \right) y \right]^{\frac{\phi-1}{\phi}} \quad (6)$$

<sup>10</sup>See Cavallo (2017) for evidence that most merchants in his sample do charge identical prices in their online and offline channels.

<sup>11</sup>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.

$$M_o = \left[ \left( \frac{k}{k+1} \right) \left( \frac{1}{1 + (\sigma - 1)\phi} \right) \frac{1}{F_o} y \right]^{\frac{1}{\phi}} \quad (7)$$

$$M_b = \left[ \left( \frac{1}{k+1} \right) \left( \frac{1}{1 + (\sigma - 1)\phi} \right) \frac{1}{F_b} y \right]^{\frac{1}{\phi}} \quad (8)$$

where

$$k \equiv q^{\frac{\phi}{\phi-1}(\sigma-1)} \left( \frac{F_b}{F_o} \right)^{\frac{1}{\phi-1}}. \quad (9)$$

Meanwhile,  $s_o$ , the online share of total spending, can be written as:

$$s_o \equiv \frac{oM_o}{oM_o + bM_b} = \frac{k}{k+1}. \quad (10)$$

In this model, the observed increase in  $s_o$  between 2007 and 2014 can be driven by an increase in  $q$  (the relative quality of online shopping relative to offline) or an increase in  $F_b/F_o$  (the ratio of fixed costs in the offline channel relative to online). Finally, maximized utility can be written as:

$$W = \left( \frac{1}{1 - s_o} \right)^{\frac{\phi-1}{\phi} \frac{1}{\sigma-1}} \left( \frac{1}{F_b} \right)^{\frac{1}{\phi} \frac{1}{\sigma-1}} \frac{(\sigma - 1)\phi y^{\frac{1}{\sigma-1}} \left( \sigma - \frac{\phi-1}{\phi} \right)}{[1 + (\sigma - 1)\phi]^{\frac{1}{\sigma-1} \left( \sigma - \frac{\phi-1}{\phi} \right)}}. \quad (11)$$

For a given level of expenditures  $y$  and fixed costs for online stores  $F_b$ , consumer welfare is increasing in the share of online spending,  $s_o$ . As  $s_o$  increases, consumers benefit because online options become better (through higher relative quality  $q$ ) and/or because online merchants become easier to access (lower fixed costs  $F_o$ ). In our welfare analysis, we assume that  $F_b$  remains constant and attribute movement in the ratio  $F_o/F_b$  to changes in  $F_o$ .

### 5.3. Calibration of $\phi$ and $\sigma$

We first estimate  $\phi$ , 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  $\phi$  affects the relationship between total expenditure ( $y$ ), spending per merchant ( $o$  and  $b$ ), and number of merchants visited ( $M_o$  and  $M_b$ ). A higher value of  $\phi$  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  $\phi$  using empirical Engel curves.

Rearranging equations (5)-(8), we obtain the following decomposition of spending into the extensive and intensive margins:

$$\ln M = \alpha + \frac{1}{\phi} \cdot \ln(oM_o + bM_b) \quad (12)$$

$$\ln \left( \frac{oM_o + bM_b}{M} \right) = \eta + \frac{\phi - 1}{\phi} \cdot \ln(oM_o + bM_b) \quad (13)$$

where  $M = M_o + M_b$ . To consistently estimate the parameter  $\phi$  from (12) and (13) via OLS, we must assume that there are no idiosyncratic fixed costs or online/offline preferences that are correlated with a cardholder's total expenditures.<sup>12</sup>

In Table 4, we present our estimates for  $\phi$ . We perform the estimation separately for 2007 and 2014. Across the two years the average point estimate is 1.8. The standard errors are too small to mention given the hundreds of millions of cards in each regression. A  $\phi$  of 2 would imply that 50% of additional card spending is on the extensive margin; our estimate is modestly below that, so that the extensive margin accounts for 56% of marginal spending, and the intensive margin accounts for 44%.

To calibrate  $\sigma$ , the elasticity of substitution across merchants, we use varia-

---

<sup>12</sup>Since the decomposition is exact, the estimate of  $\phi$  will be identical regardless of which of the two equations is used.

Table 4: Estimates of  $\phi$ 

	2007	2014
$\phi$	1.78	1.82
# of cards	287M	453M
$R^2$	0.66	0.67

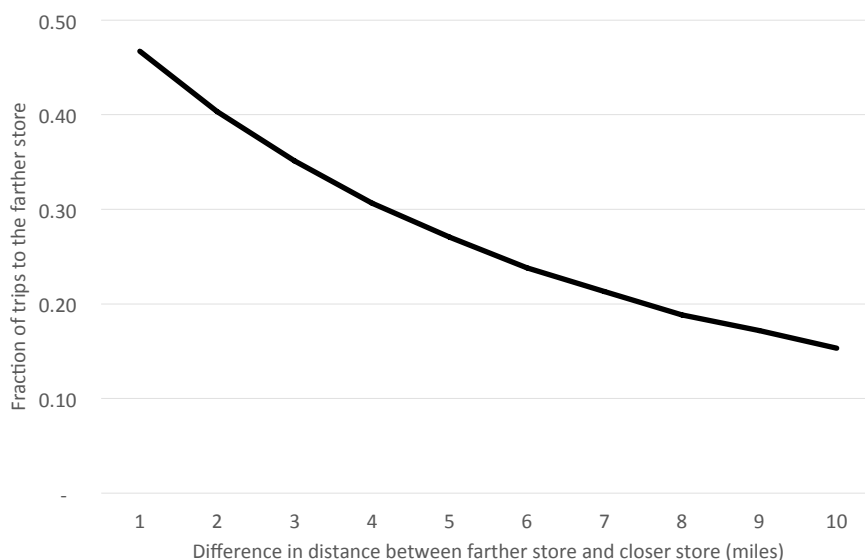
$\phi$  is estimated as the inverse of the OLS coefficient from a regression of  $\log(\#$  of unique merchants) on  $\log(\text{spending})$ , where each observation is a cardholder. This relationship follows from the solution to our CES model. The table below gives our point estimates for this regression performed separately in 2007 and 2014.

tion 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  $\sigma$ .

We estimate across-merchant substitution using purchases for the 1% sample of cards described in Section 4. For each card  $i$ , we look at offline purchases made within 10 miles of  $i$ 's location. We construct all pairs of stores  $j$  and  $k$  within each NAICS where  $i$  made a purchase in one of the stores and compute  $|dist_{ij} - dist_{ik}|$ . We then calculate the share of combined trips for each pair that were made to the farther store, and average across cards for each NAICS. In Figure 7, we show the fraction of trips to the farther store as a function of the relative distance between the two stores.

Consider a store that is 11 miles away from consumers instead of 1 mile away. This will cause the fraction of trips to the farther store to fall from 47% to 15% of combined trips. Conversely, the share of trips to the closer store rise from 53% to 85% of combined trips. This gives us the change in relative consumption. We need to convert distance into a price equivalent in order to arrive at an elasticity of substitution. As we did above for estimating convenience gains, we convert a mile of distance into \$3.71 in direct and indirect roundtrip travel costs. Assuming the consumer is 6 miles from the closer store and 16 miles from the farther store, the consumer's total travel costs are \$22.26 to the closer store and \$59.36 to the further store. We add these travel costs to the

Figure 7: Relative trips as a function of distance



The figure shows substitution between stores as a function of the relative distance between the two stores and the consumer. The x-axis gives the difference in distances between a pair of stores in the same NAICS. Estimation is performed using all purchases made by a 1% sample of cards in 2014. We construct all pairs of stores within the NAICS where a consumer made a purchase in at least one of the stores. The black line above gives the share of transactions that occurred at the further away store, averaged across all NAICS categories.

average ticket size of Visa transactions across all NAICS, which is \$42.25. This gives us the relative price of the total bundle (Visa ticket size plus travel costs) for going to the closer store vs. the farther store. The resulting ratio of relative trips to relative price gives us a point estimate of  $\sigma = 3.6$ .<sup>13</sup>

## 5.4. Results

Using our estimates of  $\phi$  and  $\sigma$  and the online share calculated from the Visa data, we can calculate consumption-equivalent changes in consumer welfare from the rise of e-commerce using equation (11). We present our estimates for these welfare gains in Table 5. Using our baseline estimates of  $\phi$  and  $\sigma$ , we calculate a change in consumer surplus of 1.7% between 2007 and 2014. Relative to a counterfactual where the online channel is completely unavailable,

<sup>13</sup>If we restrict our attention to only e-commerce NAICS, the distance gradient is similar but the average ticket size is larger, \$62 vs. \$42, leading to a smaller percentage change in price and a higher elasticity of  $\sigma = 3.8$ .

Table 5: Consumption-equivalent welfare gains from e-commerce

	$\phi$	$\sigma$	$s_o^{2014}$ vs. $s_o^{2007}$	$s_o^{2014}$ vs. $s_o = 0$
Baseline	1.8	3.6	1.7%	3.6%
Sensitivity:				
High $\phi$	2	3.6	1.9%	4.1%
High $\sigma$	1.8	4	1.5%	3.2%

The table above gives the consumption-equivalent gains in consumer welfare for two scenarios: the increase in online share between 2007 and 2014 and the level of online shopping in 2014 versus a counterfactual where online shopping is not available. Consumer welfare is calculated from the solution to the CES model detailed above. In our baseline estimation, we use the values of  $\phi$  and  $\sigma$  we estimate, using the approach described above. We perform additional sensitivity analysis to illustrate how our estimates for the welfare gains from e-commerce vary with alternate values of the parameters.

e-commerce in 2014 resulted in gains for consumers of 3.6% overall. Using alternate parameters for  $\phi$  and  $\sigma$  (rounding up  $\phi$  to 2 and  $\sigma$  to 4) shows that these parameter values clearly matter.

As we highlighted in Section 3, the level and growth in online share was not uniform across the U.S. population. Cards located in counties with higher median income and higher population density generally spent online at higher rates and saw their online share of spending grow faster. Motivated by this heterogeneity, we perform welfare analysis separately for these different segments of cards. In Tables 6 and 7, we show consumption-equivalent welfare growth broken out by quartiles of county income and county population density.<sup>14</sup>

The results of this exercise show that growth in welfare was generally increasing in county income and population density. Between 2007 and 2014, cardholders living in higher income counties experienced e-commerce welfare gains that were 130% larger than those living in the poorest counties. Relative to a counterfactual where no online option exists, consumers living in the most

<sup>14</sup>We use the same  $\phi$  and  $\sigma$  values of 1.8 and 3.6 for every group, but use group-specific online spending shares  $s_o$ . Each quartile contains approximately 25% of the population. The average of the quartile gains are smaller than the aggregate gains in Table 5 because the quartiles are population weighted, whereas the aggregate is purely dollar weighted.

Table 6: Welfare gains by county income

	$s_o^{2014}$ vs. $s_o^{2007}$	$s_o^{2014}$ vs. $s_o = 0$
Quartile 1 (poorest)	1.2%	2.7%
Quartile 2	1.3%	2.9%
Quartile 3	1.5%	3.2%
Quartile 4 (richest)	2.8%	4.6%

The table above gives the consumption-equivalent gains in consumer welfare for two scenarios: the increase in online share between 2007 and 2014 and the level of online shopping in 2014 versus a counterfactual where online shopping is not available. Consumer welfare is calculated from the solution to the CES model detailed above. Welfare gains are calculated separately for each quartile of counties. Cards are placed into the county in which it had the most transactions and counties are split into quartiles so that each quartile contains the same number of cards. The welfare analysis for each quartile is calculated using our baseline sigma and phi estimated above.

Table 7: Welfare gains by county population density

	$s_o^{2014}$ vs. $s_o = 0$	$s_o^{2014}$ vs. $s_o = 0$
Quartile 1 (least dense)	1.1%	2.6%
Quartile 2	1.4%	3.0%
Quartile 3	2.6%	4.3%
Quartile 4 (most dense)	1.7%	3.5%

The table above gives the consumption-equivalent gains in consumer welfare for two scenarios: the increase in online share between 2007 and 2014 and the level of online shopping in 2014 versus a counterfactual where online shopping is not available. Consumer welfare is calculated from the solution to the CES model detailed above. Welfare gains are calculated separately for each quartile of counties. Cards are placed into the county in which it had the most transactions and counties are split into quartiles so that each quartile contains the same number of cards. The welfare analysis for each quartile is calculated using our baseline sigma and phi estimated above.



densely populated counties experienced welfare gains 70% larger than their counterparts in the more sparsely populated counties. Higher density counties generally gained more, but not monotonically and by smaller amounts.

In Table 8 we perform several additional exercises to assess the sensitivity of our main estimate of the welfare gains to the specific values of the parameters  $\phi$  and  $\sigma$ . To give a sense of how our top-line numbers might vary, we re-estimate  $\phi$  and  $\sigma$  per the approach described above separately for quartiles of counties along two dimensions of heterogeneity: county density and income. We calculate the welfare gains using the average online shares in 2014 taken across the whole sample. Overall, our estimate ranges from 3.5% to 7% for the overall gains from e-commerce.

Table 8: Additional sensitivity analysis of the main estimate

		Sigma estimated by quartile of county density				
			1st (lowest)	2nd	3rd	4th (highest)
		Phi/sigma	2.2	2.9	3.3	3.8
Phi estimated by quartile of county density	1st (lowest)	1.9	7.0%	4.9%	4.2%	3.5%
	2nd	1.8	6.7%	4.7%	4.1%	3.4%
	3rd	1.8	6.8%	4.7%	4.1%	3.4%
	4th (highest)	1.8	6.7%	4.7%	4.1%	3.4%
		Sigma estimated by quartile of county income				
			1st (lowest)	2nd	3rd	4th (highest)
		Phi/sigma	2.8	3.2	3.4	3.5
Phi estimated by quartile of county income	1st (lowest)	1.9	5.2%	4.3%	4.1%	3.9%
	2nd	1.8	5.0%	4.2%	3.9%	3.8%
	3rd	1.8	4.9%	4.1%	3.9%	3.7%
	4th (highest)	1.8	4.9%	4.1%	3.8%	3.7%

Each cell in the table gives the welfare gains from e-commerce, computed with different values of our estimated  $\phi$  and  $\sigma$  parameters. We compute a separate parameter value for each quartile of counties, ordered by density (top panels) or income (bottom panels) so that each quartile contains about 25% of the cards. We compute welfare gains using the aggregate share of online sales in 2014.

We have framed these gains as a percentage of Visa spending, 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 22.5% of Visa spending, surplus is equivalent to about 16% of e-commerce spending.<sup>15</sup> In the other direction, it is important to recognize that Visa spending is only 20% of all consumption by 2014. Visa might be representative, however, of all credit and debit card spending. Using an industry estimate from WalletHub.com, all such spending was a little over 36% of consumption in 2014.<sup>16</sup> Thus, if consumer surplus from e-commerce was 3.6% of Visa spending in 2014, this would be equivalent to a more modest 1.3% of all consumption. Given consumption per household of \$96,270 in 2014 according to the BEA, this amounts to \$1,254 per household.

Our estimated gains from e-commerce are equal to about 0.9% of GDP. Although this is arguably large economically, it took something like 20 years to reach this level. Even if none of these gains showed up in measured growth and they all occurred in the last ten years, they would understate growth by only 9 basis points per year. This would not overturn the conclusion of Syverson (2016) that rising mismeasurement was not a major contributor to the growth slowdown in the U.S.<sup>17</sup>

## 6. Conclusions

The advent of the internet, the rise of Amazon, and the increased popularity of e-commerce more generally 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,

---

<sup>15</sup>This is modest compared to the Cohen et al. (2016) estimate of consumer surplus equal to 160% of spending on Uber.

<sup>16</sup><https://wallethub.com/edu/market-share-by-credit-card-network/25531>

<sup>17</sup>The U.S. Bureau of Labor Statistics reports that growth in multifactor productivity fell from 2.7 percentage points per year from 1996–2005 to 0.9 percentage point per year from 2006–2016, or 180 basis points.

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 least 1.0% of total Visa spending. 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 about 3.6% of total Visa spending, which maps into approximately 1.3% of consumption.

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 heterogeneous across product categories, income groups, 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 retail 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 the transaction, 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.
- Lieber, Ethan and Chad Syverson, “Online versus offline competition,” *Oxford Handbook of the Digital Economy*, 2012, pp. 189–223.
- 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,” Technical Report, National Bureau of Economic Research 2016.

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