

# Assessing the Gains from E-Commerce

By PAUL DOLFEN, LIRAN EINAV, PETER J. KLENOW, BENJAMIN KLOPACK,  
JONATHAN D. LEVIN, LARRY LEVIN, WAYNE BEST\*

*E-commerce represents a rapidly growing share of consumer spending in the U.S. 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 e-commerce reached 8% of consumption by 2017, yielding the equivalent of a 1% boost to their consumption, or over \$1,000 per household per year. While some of the gains arose from avoiding travel costs to local merchants, most of the gains stemmed from substituting to merchants available online but not locally. Higher income consumers gained more, as did consumers in more densely populated counties.*

*JEL: L81, L86, O33, O47*

*Keywords: E-commerce*

According to the U.S. Census Bureau, e-commerce spending doubled as a share of retail sales from 5% in 2007 to 10% by 2017. In addition to large online-only megastores, many traditional brick-and-mortar retailers have launched online entities that sell the same products available in their physical stores.

For consumers, shopping online differs in important ways from visiting a brick-and-mortar store. Less constrained by physical space, online retailers 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 to consumers from the rise of online shopping by leveraging a large and detailed transaction-level dataset

\* Dolfen: Uber, 1455 Market St #400, San Francisco, CA 94103, jpdolfen@gmail.com. Einav: Department of Economics, Stanford University, 579 Jane Stanford Way, Stanford, CA 94305, and NBER, leinav@stanford.edu. Klenow: Department of Economics, Stanford University, 579 Jane Stanford Way, Stanford, CA 94305, and NBER, klenow@stanford.edu. Klopac: Department of Economics, Texas A&M University, 2935 Research Pkwy, College Station, TX 77845, bklopac@tamu.edu. Levin: Graduate School of Business, Stanford University, 655 Knight Way, Stanford, CA 94305, and NBER, jdlevin1@stanford.edu. Levin: Visa, Inc., 900 Metro Center Blvd, Foster City, CA 94404, llewin@visa.com. Best: Visa, Inc., 900 Metro Center Blvd, Foster City, CA 94404, wbest@visa.com. We are grateful to Yue Cao and Raviv Murciano-Goroff for terrific research assistance, and to Yu Jeffrey Hu, Jessie Handbury, Sam Kortum, two anonymous referees, Simon Gilchrist (the Editor), and many seminar participants for comments on earlier drafts. Financial support from the National Science Foundation (Grant # SES-1729395) is gratefully acknowledged. All results have been reviewed to ensure that no confidential information about Visa merchants or cardholders is disclosed.

<sup>1</sup>Erik Brynjolfsson, Yu Hu and Michael D Smith (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 (2018) document a related pattern in the context of shoes.

of consumer purchases: the universe of Visa credit and debit card transactions between 2007 and 2017. We begin by describing the features of this dataset and presenting some descriptive facts on the growth of e-commerce. By the end of our sample, in 2017, roughly 22% of consumption flowed through Visa.

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 channel.<sup>2</sup> We show that a consumer located farther away from a merchant's brick-and-mortar store is more likely to purchase in their online channel. 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, we estimate that gains from convenience reached no more than 0.4% of consumer spending by 2017.

To quantify the variety and quality gains from e-commerce, we write down a model in which 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 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, again converting travel distance into dollars. We also use variation across cardholders 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 percent of consumption — for richer households (annual income above \$50,000) than poorer households (below \$50,000), 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. With one exception, this line of work focuses on the broader surplus from the internet, rather than e-commerce more narrowly. Goolsbee and Klenow (2006) propose an approach based on time spent on 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.

More recently, Syverson (2017) looks at the question of whether the observed slowdown in labor productivity growth can be explained by mismeasurement of digital goods and Information and Communications Technology (ICT) more gen-

<sup>2</sup>Throughout the paper, we use “convenience” to refer to within-merchant gains (e.g., through the reduced time and travel cost), and “variety” gains to refer to substitution across merchants. This is naturally an approximation given that within-merchant benefits may also account for some variety gains through increased product availability online.

erally. He concludes that surplus from ICT is not large enough to explain much of the growth slowdown, which exceeds 1% *per year* for over a decade. A closer paper to ours is Yoon J. Jo, Misaki Matsumura and David E. Weinstein (2019), which uses data from Rakuten (the Japanese version of Amazon) to estimate the gains from online sales in Japan, including those through price convergence across regions. Couture et al. (2020) study a program that increased internet access in Chinese villages and find more modest gains.

Perhaps the closest paper to ours is the concurrent and independent work of Huang and Bronnenberg (2020). They use a detailed consumer survey from the Netherlands in the narrower context of apparel purchases to measure the consumer gains from e-commerce. Like us, they use variation in distance to the brick-and-mortar store to measure the convenience gain of e-commerce, and an economic model to measure the variety gains. Their findings are broadly similar to ours. Moreover, their narrower context and more detailed data on prices also allow them to measure a third, indirect component of consumer gains, derived from more intense price competition.

Our paper also contributes to the literature examining the broader impact of the internet, and to the literature that measures consumer surplus from new products.<sup>3</sup> Broda and Weinstein (2010) and Redding and Weinstein (2020) estimate the value of growing variety using scanner data, Broda and Weinstein (2006) quantify the value of rising import variety, and Brynjolfsson, Hu and Smith (2003) look at the gains to consumers from accessing additional book titles at online bookstores.<sup>4</sup>

The rest of the paper is organized as follows. Section I introduces the data and how we construct some of the key variables. Section II presents summary statistics and initial facts. Sections III and IV estimate, respectively, the convenience and variety gains from e-commerce. Section V briefly concludes. The online Appendix provides much more information about the data, the construction of the samples, the way we measure e-commerce, and the construction of figures and tables, as well as additional derivations associated with the solution of the variety model in Section IV.

## I. 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.<sup>5</sup> We complement the Visa data with data from a major

<sup>3</sup>For example, Akerman, Gaarder and Mogstad (2015) study the effect of firms' broadband adoption on labor productivity and wages, and Hjort and Poulsen (2019) assess the employment effects of the arrival of fast internet in Africa.

<sup>4</sup>Quan and Williams (2018) make and illustrate the important point that, if demand is location-specific, then representative consumer frameworks can overstate variety gains.

<sup>5</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->

credit reporting bureau, as well as publicly available information at the county level from the U.S. Census and the Internal Revenue Service.

Online Appendix A provides a detailed description of the data source and sample construction, and we attempt to summarize it here. 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 (as well as the street address in the most recent years). The merchant identifier is linked by Visa to the merchant's name and industry classification (NAICS). In contrast, different cards used by the same person or household are not linked to each other, and information about the cardholder is generally limited to what one could infer from the card's transactions (with the exception of approximately half of the cards active in 2016 and 2017, which are matched to cardholder data from a credit bureau).

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.<sup>6</sup> As we will see, Visa's rising share of overall consumption reflects in part the rising importance of e-commerce. But Visa has also captured a rising share of credit card spending, though not of debit card spending or the number of credit cards.<sup>7</sup>

In Section III below, where we focus on substitution between online and offline channels within a merchant, we further limit the analysis to the nine retail NAICS where the online transaction share was between 10% and 90%, and to merchants within those categories that had online shares between 1% and 99%.<sup>8</sup> These include the following categories, which are used to estimate convenience gains: automotive stores; furniture and home furnishings stores; electronics and appliance stores; building material stores; health and personal care stores; clothing and clothing accessories stores; sporting goods, hobby, musical instruments and book stores; general merchandise stores; and miscellaneous store retailers.

**Key variables.** Each transaction has an indicator for whether it occurred in person (“CP” for Card Present, meaning that the card was physically swiped) or not (“CNP” for Card Not Present). Roughly half of CNP transactions are

share-by-credit-card-network/25531.

<sup>6</sup>As described in online Appendix A, 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.

<sup>7</sup><https://wallethub.com/edu/cc/market-share-by-credit-card-network/25531/>.

<sup>8</sup>Census Bureau NAICS codes 44 and 45 cover Retail Trade.

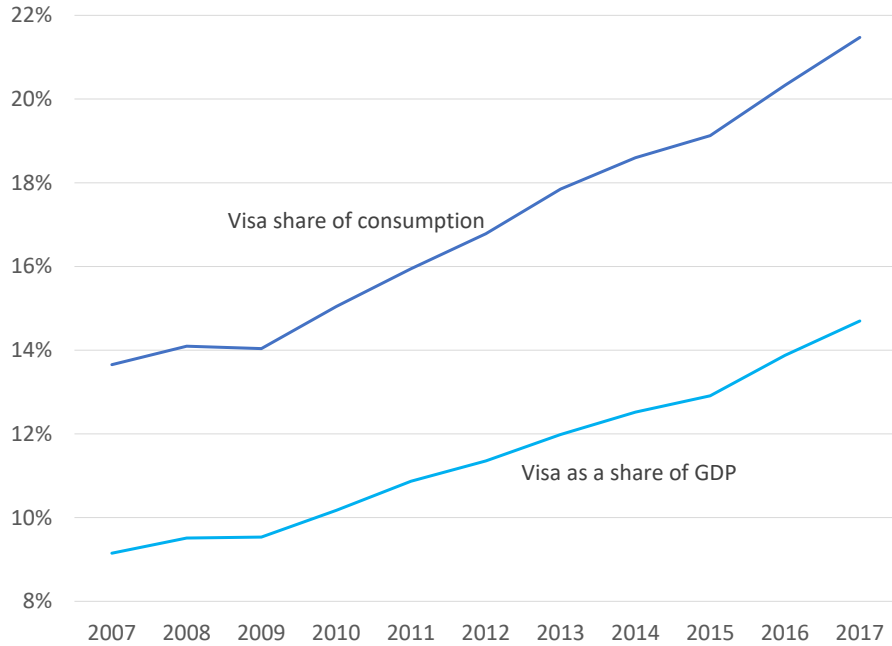


FIGURE 1. VISA SPENDING AS A SHARE OF CONSUMPTION AND GDP

*Note:* Visa credit and debit card spending; GDP and Consumption from the BEA.

broken down further into e-commerce, mail order, phone order, and recurring transactions; in the other half, this additional breakdown is missing. We treat phone, mail, and recurring transactions as offline. In the convenience analysis in Section III, we exclude CNP transactions that are not classified as e-commerce from the sample to more sharply focus on the substitution between offline and online purchases.

For other analyses in the paper, we infer the fraction of e-commerce in transactions with missing breakdowns based on industry-level averages. In online Appendix B we report additional analysis, which suggests that the breakdown between e-commerce and other CNP categories (when available) does not vary systematically within industries, thus allowing us to rely on a NAICS-level measure of whether a transaction is e-commerce or an “offline CNP” (phone, mail, or recurring transaction). Denoting ECI as the e-commerce indicator within CNP transactions,  $i$  as the 3-digit NAICS category, and  $t$  as the year, we thus infer e-commerce spending as

$$\text{e-commerce}_{it} = \frac{\text{ECI}_{it}}{\text{ECI}_{it} + \text{phone/mail/recurring}_{it}} \times \text{CNP}_{it}.$$

Table 1 lists the NAICS that contain a nontrivial share of spending with the e-commerce indicator. This includes many retail and some non-retail NAICS categories. It excludes NAICS categories such as Communication, which contains ample CNP spending on cell phone bills but which occurs predominantly through recurring payments. The non-retail NAICS categories with a significant e-commerce presence are all related to travel and transportation. We include these NAICS categories in our analysis on the grounds that they provide convenience and variety benefits akin to online options in retail NAICS categories (e.g. booking travel online rather than visiting or calling a travel agent).

TABLE 1—E-COMMERCE CATEGORIES

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

*Note:* NAICS categories that we classify as containing e-commerce spending.

Two other important variables in our analysis are card location and income. For those cards that are matched with credit bureau data in 2016 and 2017, we have a direct measure of the 9-digit ZIP code of the cardholder's residential address, as well as credit bureau information about her (categorical) annual income.<sup>9</sup> We

<sup>9</sup>The 9-digit ZIP code classification (or "ZIP+4") is used by the U.S. Postal Service to indicate a particular segment of a delivery route, and is significantly smaller than the 5-digit ZIP code. A typical 9-digit ZIP code contains a group of houses on a particular block.

complement these data with proxies for location that we can obtain from each card's transactions, which can be constructed for the entire sample. Specifically, we infer a card's preferred shopping location from its transaction history by defining a card's location as a longitude-latitude pair given by the transaction-weighted average ZIP centroid of its transactions.<sup>10</sup>

**Data limitations.** While the Visa data capture a substantial amount of overall consumer spending, they also have several important limitations. We briefly describe these issues here and provide additional detail in online Appendices A and B.

First, we do not observe specific items purchased, nor their prices or quantities. The data contain only the total dollar amount of the transaction. Second, for much of our sample period, we are unable to accurately attribute transactions to small merchants. Visa assigns a unique merchant identifier to each large chain, but aggregates transactions from some smaller chains and single establishment firms together into a single merchant identifier, by NAICS. As a result, we use only transactions from these large firms to obtain those parameter estimates in Section III and Section IV, which require knowledge of the merchant identity.<sup>11</sup> Our aggregate measures, including our estimates of the U.S. online share and the convenience and variety gains from e-commerce, are based on the transactions of all merchants.

Third, as mentioned, for about half of transactions where a card was not physically present, we cannot distinguish between transactions that were processed online from those that were processed by phone, mail, or as recurring charges. For CNP transactions with missing breakdowns, we assume the e-commerce fraction is the same as the NAICS-year fraction of non-missing CNP values that is classified as e-commerce. As shown in online Appendix Figure A2, when the CNP breakdown is available, e-commerce accounts for the vast majority of CNP spending in the e-commerce sectors, which makes it less likely that the analysis would be sensitive to this missing information. online Appendix B provides more detail.

Fourth, as described in online Appendix A, we exclude PIN-debit transactions from all analysis. Due to regulatory changes during our sample period, the volume of PIN-debit transactions processed by Visa fluctuates significantly over time, as some merchants began to reroute these transactions to other card companies. Finally, we cannot link cards across households. If some households use multiple cards, we treat each card separately.

<sup>10</sup>We limit attention to ZIP codes in which the card transacted 20 or more times 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 are excluded from our analysis that uses card location.

<sup>11</sup>Transactions from these large "named" merchants accounted for about 59% of dollars and 69% of transactions.

## II. 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 categories. The online share was already quite high in 2007 in some categories, such as air transport, while in some categories, such as food and gasoline, the online share remained very low in 2017.

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.<sup>12</sup> 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.<sup>13</sup> Finally, we divide by overall U.S. consumption of goods and services (including the service flow from housing):

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

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

**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 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

<sup>12</sup>We divide Visa’s credit and debit card spending by the estimate of national credit and debit card spending at WalletHub.com (<https://WalletHub.com/edu/market-share-by-credit-card-network/25531>). These estimates are based on the SEC filings of the major card companies. We calculate this share year by year to account for the growing card spending share of consumption. Visa had a market share of 70.9% in the debit card market and 52.9% in the credit card market (in terms of transaction volume in 2017). The second largest company was Mastercard with a market shares of 29.1% and 22.0% in the debit and credit card markets, respectively. American Express and Discover had market shares of 21.1% and 3.8% respectively in the U.S. credit market.

<sup>13</sup>This implies that we assume away e-commerce spending through PayPal. While we lack sufficient data to quantify PayPal’s share of US online spending, we can assess PayPal’s size relative to the major credit card networks in terms of global transaction volume. In 2017, the transaction volume on the major payment card networks was as follows: \$8.4 trillion on Visa, \$4.3 trillion on Mastercard, \$1.2 trillion on American Express and \$0.2 trillion on Discover (Source: Visa 10-k 2018). This implies a total payment card transaction volume of \$14.2 trillion on these four networks. PayPal’s total transaction volume in 2017 was \$0.5T (Source: PayPal 10-k 2017). Furthermore, 20% of PayPal’s transaction volume reflects P2P transactions. Source: PayPal (<https://www.paypal.com/stories/us/paypal-reports-fourth-quarter-and-full-year-2017-results>).



TABLE 2—VISA ONLINE SHARES IN SELECT NAICS CATEGORIES

	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

*Note:* % of Visa credit and debit card spending in select NAICS categories.

in more rural areas and smaller cities, where there are fewer offline merchants.<sup>14</sup> 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  $\alpha_{cy}$  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 that online spending occurs only through credit and debit cards, so the cardless are not online at all. See online Appendix C for details. In the final step we scale down all  $s_{cy}$  values to match our estimated aggregate U.S. e-commerce share. The implied fraction of households with credit cards is markedly lower in poorer county-income pairs, consistent with evidence that higher income households are more likely to have cards.<sup>15</sup>

We estimate an online share of 3.4% of consumption for households with incomes

<sup>14</sup>See Handbury and Weinstein (2014) for evidence that variety is greater in larger cities.

<sup>15</sup>According to the Federal Reserve Board's Report on the Economic Well-Being of US Households in 2018, 61% of households with annual income of \$40,000 or less had one or more cards in 2018, versus 90% of households with income between \$40,000 and \$100,000 and 98% of households with income higher than \$100,000.

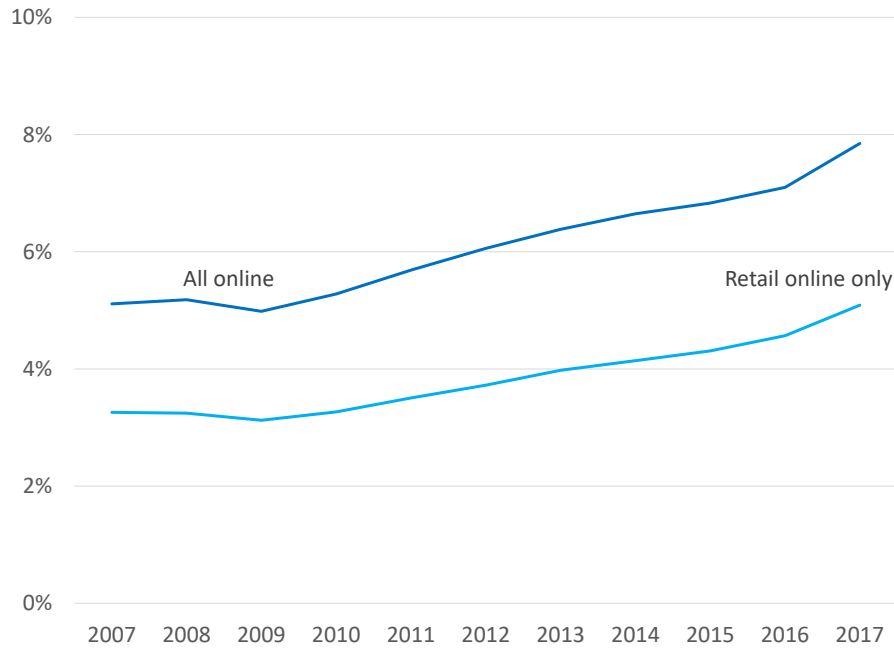


FIGURE 2. ESTIMATED SHARE OF ONLINE SPENDING IN THE U.S.

*Note:* We estimate e-commerce spending on the Visa network and extrapolate it to the the U.S. economy assuming: 1) that Visa is representative of all card spending in terms of online share, and 2) all online spending is done using credit or debit cards. ‘All online’ refers to our baseline estimate of e-commerce spending in all consumer categories. ‘Retail online only’ refers to our alternative estimate which only counts online spending in retail industries as e-commerce. Total consumption (the denominator for each series) is from the BEA. Online Appendix C provides further details on the extrapolation from Visa to the US economy.

of \$50k and below in 2017, and 9.7% of consumption for households with incomes above \$50k. Using population per square mile from the 2010 Census, counties with above-median population density have a population-weighted average online share of 7.9% of consumption, whereas below-median density counties have an average share of 5.1%. This is perhaps surprising because the density of brick-and-mortar retailers is increasing in population density.

Figure 3 displays our online share estimates for all U.S. counties in 2017. This provides a finer geographic breakdown than simply rural versus urban populations. Online penetration is distinctly higher in the Northeast and in the West and Mountain regions than in the South or Midwest (excepting Florida and some areas of Texas).

Figure 4 plots merchants visited per card versus dollars spent per card across cards in 2017—for online and offline spending separately. It shows that higher spending cards access more merchants, but more so offline than online. That is,

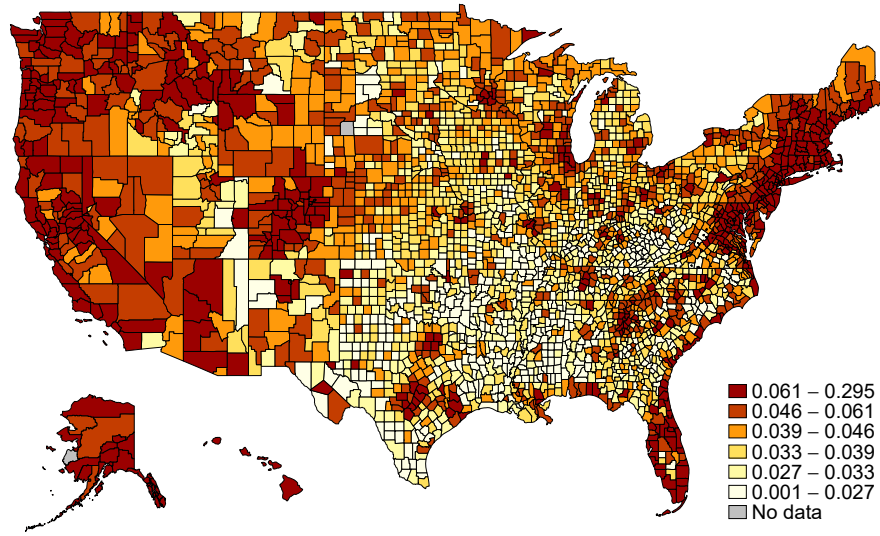


FIGURE 3. ONLINE SHARES BY COUNTY IN 2017

*Note:* This figure displays the online share in each county calculated from the Visa data and adjusted by the propensity of county residents to use a credit card. Each card is placed in a county-income bin according to their home billing ZIP code and estimated household income. We compute the online share for each county-income bin from their Visa credit card spending and multiply it by the ratio of credit card accounts to population in that county income bin, normalized to match our estimate of the aggregate online share of spending. As a measure of population in each county-income bin, we use IRS data on the number of tax filers. The plot shows the online share (aggregated across cardholders of different incomes) within each county. See online Appendix C for more details.

online merchant variety is less unequal than offline merchant variety. Figure 5 plots average merchants visited per card in a county against population density in the county. It indicates that offline merchant variety rises modestly with density, but online variety is essentially flat with respect to density.

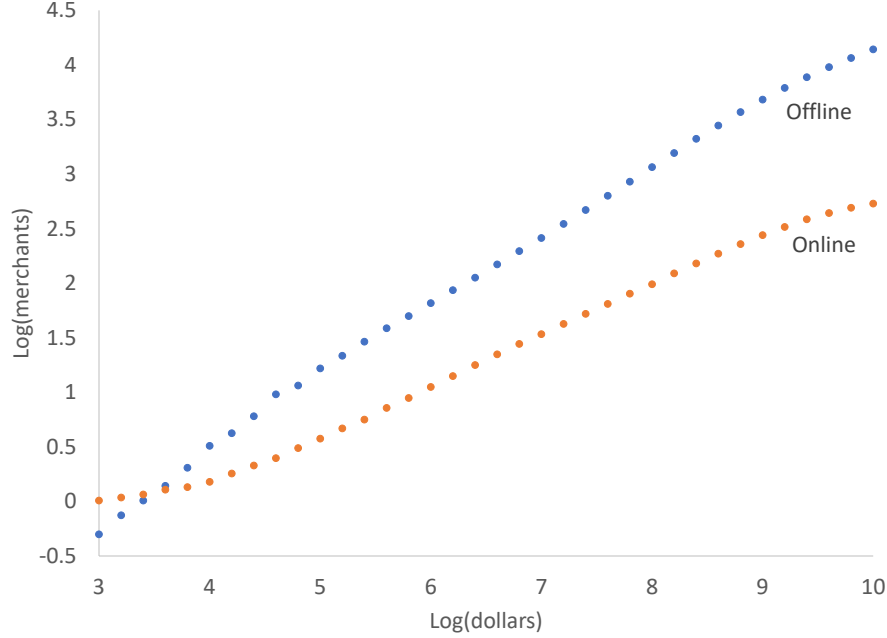


FIGURE 4. MERCHANTS VISITED VS. DOLLARS SPENT PER CARD, ONLINE AND OFFLINE

*Note:* The figure shows a bin scatter of the natural logarithm of the number of merchants visited per card against the natural logarithm of the total dollar amount spent per card in 2017, separately for online and offline purchases. Cards are grouped into bins based on their total online and offline 2017 spending. Each point show the average number of online or offline merchants per card for accounts of a given level of online or offline spending.

### III. Estimates of Convenience Surplus

In this section we focus on a specific gain from e-commerce: shopping via a merchant's e-commerce channel saves costs associated with traveling to that merchant's physical store.<sup>16</sup> Given that e-commerce allows a consumer to access a wider set of merchants than would otherwise be available, this direct convenience gain is surely smaller than the overall gain, which accounts for merchant substitution. Yet we begin by assessing the gains from convenience given that doing so is simpler and requires fewer modeling assumptions.

**Specification.** To quantify these convenience gains, we estimate a simple binary

<sup>16</sup>The thought experiment here is that goods are at the merchant's distribution center, and from there they can either get shipped to the merchant's physical store or to the consumer directly (at the same cost), which means that the extra travel cost to the physical store is incremental. Note that the merchant fixed effects could also incorporate differences in online vs. offline prices.

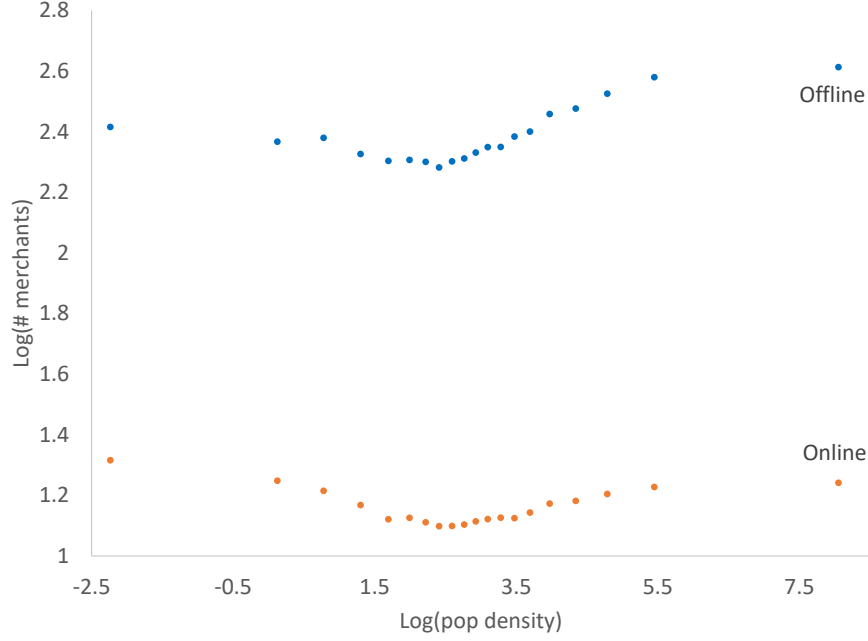


FIGURE 5. MERCHANTS VISITED VS. COUNTY POPULATION DENSITY, ONLINE AND OFFLINE

*Note:* The figure shows a bin scatter of the natural logarithm of the average number of merchants visited per card against the natural logarithm of county population density, separately for online and offline purchases. The sample includes credit cards that are matched to credit bureau data that includes the card's billing zipcode, which we use to locate cards in counties. County population density is calculated by the US Census Bureau as people per square kilometer using ACS 2014-2018 population estimates. We compute the average number of merchants per card for each county, and then place counties into 20 bins by population density so that each bin contains an equal number of counties. Each point in the plot shows the average number of merchants visited for counties of a given population density.

choice between an online and offline transaction. We assume consumers know the prices and items they will buy from each merchant. We make the strong assumption that prices are the same online and offline for a given merchant.<sup>17</sup> The only remaining choice is thus whether to transact online or offline.

Let the utility of consumer  $i$  from buying online at merchant  $j$  be

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

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

<sup>17</sup>Cavallo (2017) reports that, in the United States, in 69% of the cases he examined prices are identical online and offline within merchants.

type I extreme value distribution, iid across merchants and consumers. Similarly, let the utility of consumer  $i$  from buying offline at merchant  $j$  be

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

where  $\gamma_j^b$  is the average merchant-specific utility from the offline channel,  $\epsilon_{ij}^b$  is an offline consumer-merchant component, which we assume is similarly drawn from a type I extreme value distribution, iid across merchants and consumers. Finally,  $dist_{ij}$  is the straight-line distance between the location of consumer  $i$  and the nearest store of merchant  $j$ . The store location is recorded by Visa as a latitude-longitude pair, while in our baseline the location of the consumer is based on the centroid of the 9-digit billing ZIP code of the cardholder.

One may be concerned that many shopping trips cover more than a single merchant, so the distance measure we use is measured with error. To address this concern, we re-estimate the model but instead of billing address we use the cardholder's average shopping location (see Section I) from which to compute travel distances (we can do this for about half the sample). The estimated coefficient on distance is, in fact, lower, implying consumer surplus that is approximately 30% higher than the estimate reported below.

Equations (1) and (2) give rise to a simple logit regression, in which we regress 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, which capture the merchant-specific relative attractiveness of its online store,  $\gamma_j^o - \gamma_j^b$ .

**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 ZIP code. To capture merchants where the choice of online and offline is meaningful, we use transactions in the nine mixed-channel retail categories at 'named' merchants (described in the previous section) that had both an online and offline presence where the consumer was within 50 miles of the offline store.<sup>18</sup>

Table 3 presents summary statistics for this sample. The categories with the highest share of online transactions are clothing stores, health and personal care stores, and electronics stores. In these three categories, online purchases made up between 22% and 32% of total transactions and sales. 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 6 pools across the nine retail categories, and relates the online share to distance in the raw data (which does not account for any merchant-specific online attractiveness), as well as the estimated relationship using the logit specification (which is a bit flatter, as it includes merchant fixed effects, and therefore accounts

<sup>18</sup>We define a merchant within these NAICS as having offline and online available when its online share is between 1% and 99% of its total transactions.

TABLE 3—SUMMARY STATISTICS BY NAICS CATEGORY

NAICS NAICS code	Clothing 448	Health and Personal Care 446	Electronics 443	Other -
Transactions	1,757,549	432,415	412,371	12,159,454
Online Share				
Transactions	0.299	0.259	0.218	0.087
Spending	0.317	0.291	0.257	0.151
Ticket size (dollars):				
Offline	84.9 (15.1 - 161.0)	53.6 (8.5 - 103.6)	237.5 (14.1 - 628.1)	65.1 (7.0 - 135.2)
Online	92.6 (18.7 - 182.9)	63.0 (2.9 - 130.0)	295.2 (16.1 - 855.9)	121.9 (9.6 - 227.0)
Distance to nearest offline store (miles):				
Offline	7.1 (1.2 - 16.4)	5.6 (0.9 - 12.5)	6.0 (1.1 - 13.3)	4.9 (0.8 - 10.8)
Online	9.5 (1.4 - 24.1)	8.3 (1.1 - 22.3)	8.0 (1.3 - 21.1)	7.4 (1.2 - 18.2)

*Note:* The table shows statistics for the transactions used in the convenience analysis, which include purchases in nine mixed-channel NAICS at merchants with an online share between 1% and 99% where the merchant had an offline store within 50 miles of the consumer's location. The table shows statistics for the three NAICS categories included in the sample with the highest online share of transactions. The other six NAICS categories are included in Other; these include Automotive Stores (441); Furniture Stores (442); Building Material Stores (444); Sporting Goods, Music, and Book Stores (451); General Merchandise Stores (452); and Miscellaneous Stores (453). In this analysis, we include only offline and e-commerce transactions, excluding mail order, recurring transactions and other forms of Card Not Present purchases. 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.

for merchant-specific effects). 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 40-50 miles away, the online share roughly quadruples, from approximately 7% to about 30%.

Using our logit specification, we estimate a  $\beta$  coefficient of 0.026 (with a standard error of about 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 2.6 percentage points.<sup>19</sup>

**Estimates of convenience gains.** This simple model allows us to estimate the

<sup>19</sup>The predicted online share at merchant  $j$  for card  $i$  that lives  $dist_{ij}$  miles away from the nearest location  $j$  is  $\frac{\exp(\gamma_j + \beta dist_{ij})}{1 + \exp(\gamma_j + \beta dist_{ij})}$  (where  $\gamma_j = \gamma_j^o - \gamma_j^b$ ). The change in predicted online share at 20 miles vs. 10 miles is  $\frac{\exp(\gamma_j + 0.026 \cdot 20)}{1 + \exp(\gamma_j + 0.026 \cdot 20)} - \frac{\exp(\gamma_j + 0.026 \cdot 10)}{1 + \exp(\gamma_j + 0.026 \cdot 10)}$ , which is approximately 2.6 percentage points, averaged across merchants.

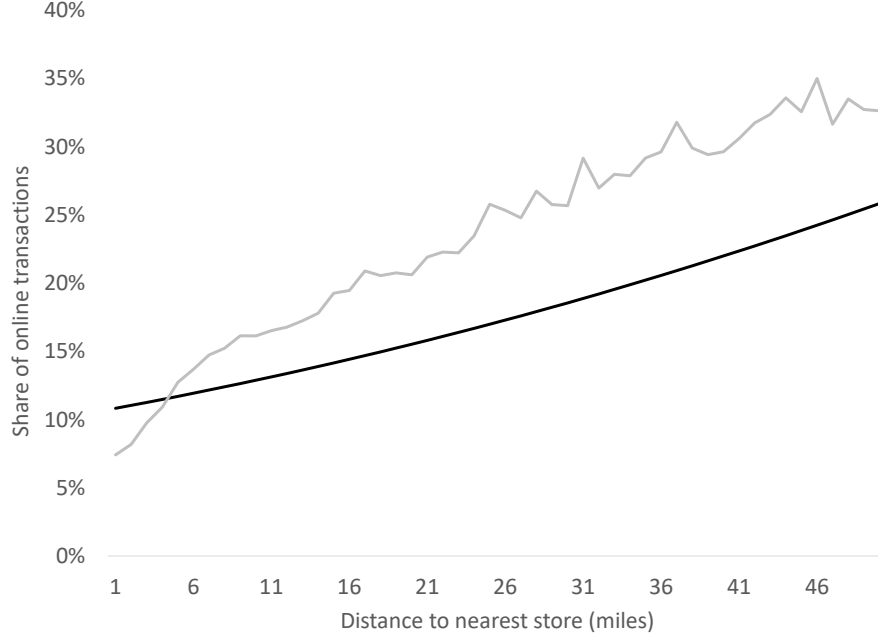


FIGURE 6. ONLINE SHARE VS. DISTANCE TO MERCHANT STORE

*Note:* The figure shows the share of transactions that occur online as a function of the distance between the card and the nearest outlet of the merchant. The sample includes transactions made by 1% of cards in 2017 at merchants in the nine mixed-channel NAICS categories listed in the data section. We include transactions at merchants that had a location within 50 miles of the card's billing ZIP code. The grey line shows a bin scatter of the share of these transactions that occurred online in the raw data. Each point gives the average share of transactions that were online for cards in a bin of size one mile. For example, the leftmost point on the grey line shows that cards that were between zero and one mile away from an outlet of a merchant conducted about 7% of their transactions with that merchant in the online channel. The black line shows the predicted share of online transactions from a logit regression of an indicator for whether the transaction was online on the distance between the card and merchant and a set of merchant fixed effects. We note that the black line is flatter than the grey line because the raw data does not account for merchant effects, while the logit regression does.

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

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



We do not observe and therefore do not use prices in our analysis. Instead, we use travel distance as a determinant of the full price. To monetize miles, we 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.<sup>20</sup>

Applying equation (3) to all the transactions in our data, we obtain an average convenience gain (across all transactions in the sample) of 5.8 mile equivalents.<sup>21</sup> Using the conversion factor above (\$3.18 per round trip mile), the convenience gain per transaction comes to \$18 dollars.

We note that this convenience gain arises from two distinct factors, which we estimate to contribute about equally to the overall convenience gain estimate. One is the reduction in travel cost that arises from the online channel. The second is the merchant-specific fixed effect,  $\gamma_j^o - \gamma_j^b$ , which captures (merchant-specific) increased utility from using its online channel.

The average ticket size in our sample is \$77 and the average distance between consumer and store is 6 miles. Thus convenience gains from the online option are on the order of 19% for purchases in the nine NAICS categories in our estimation. Together, transactions in these nine categories at merchants with both online and offline options made by consumers who were closer than 50 miles to an offline outlet of the same merchant make up about 10% of all Visa transactions, implying that the total convenience gains as a share of Visa spending are about 1.9%, or roughly 0.4% of all consumption (and these gains include an annual reduction of approximately 5 billion miles driven — or 50 miles per household — which may be associated with additional positive externalities).

Our estimates here assume no impact of rising e-commerce spending on the number of physical stores. One advantage of taking a broader view of the gains from e-commerce in the next section is that we can write down a simple model in which offline choices endogenously shrink in response. Our estimates in this section also assume people will be buying from the same merchants online and offline. In the next section we will treat online and offline spending as imperfect substitutes, and estimate this degree of substitutability.

<sup>20</sup>To obtain the monetary cost of a mile, we use estimates from Einav, Finkelstein and Williams (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 one mile is  $\$0.535 \cdot 7.9/5.3 = \$0.79$ .

<sup>21</sup>This estimate is 33% higher (7.7 mile equivalents) if we allow the distance coefficient and the merchant fixed effect to vary with the size of the transaction.

#### IV. Estimates of variety and quality gains

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. For example, 88% of total online sales in the Visa data were associated with card-merchant combinations for which the offline sales associated with this card-merchant combination were zero.<sup>22</sup> That is, as cardholders move online the vast majority of their online spending is associated with merchants they have not used before, suggesting that cross-merchant substitution may be a predominant source of consumer surplus.

##### A. Model Setup

To capture gains from variety and quality, we write down a stylized model with substitution across merchants and calibrate it using moments from 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:<sup>23</sup>

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

subject to

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

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 consumer).

<sup>22</sup>This calculation is based on all active cards in 2014.

<sup>23</sup>Our baseline model assumes a single elasticity of substitution across all merchants. We will later generalize this assumption using nested CES preferences, where each nest corresponds to a 3-digit industry. The resulting welfare gains in the more general setup exceed the welfare gains we find using our baseline model. We prefer our baseline model for two reasons. First, we do not need to take a stance on the ambiguity of how spending at Nonstore Retailers (NAICS 454) can be mapped to other industries. Second, we do not need to extrapolate Visa spending to economy-wide spending for every industry.

The parameter  $\sigma$  is the elasticity of substitution across merchants. Values of  $\sigma < \infty$  imply a “love of variety.” The parameter  $\phi$  governs how fast fixed costs to visiting merchants increase with the number of merchants visited. We assume  $\phi > 1$  to get an interior solution.<sup>24</sup>

In this setup, consumers gain if the costs of accessing online merchants falls over time, allowing consumers better access to some combination of more varieties, higher quality varieties, and cheaper varieties.

**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

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

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.

We make the simplifying assumption that each brick-and-mortar (online) seller is entertained by a random subset of the  $L$  consumers. For example, suppose  $M_j/M_{j,market}$  is 90%. Then each consumer entertains a random 90% of the merchants. The consumer then decides how much to buy from each merchant they visit based on their CES preferences in (4) above. Merchants are monopolistic competitors who face an elasticity of demand  $\sigma$  from the customers who visit them. Merchants price to sell to the customers who visit them, but do not price to entice more customers to visit them because of the random assignment. We make this assumption to simplify the pricing problem and because we cannot see merchant prices in the Visa data.<sup>25</sup>

**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:

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

<sup>24</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.

<sup>25</sup>Cavallo (2018) presents evidence that online competition has changed pricing patterns (e.g. the frequency of price changes) and inflation dynamics (such as exchange rate pass-through). See Goolsbee and Klenow (2018) for evidence that inflation is lower online than offline.

The transportation/internet technologies therefore pin down the “intercept” of the convex costs of accessing merchants offline (picture driving longer distances to access more stores) and online (imagine some retailers provide more convenient account sign-up). 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, reducing 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 are 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 economy-wide labor is allocated to merchant production of consumer goods, merchant overhead, and transportation and internet services.

### B. 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 offline (online) merchants have the same process efficiency:

$$Z_m = Z_b \text{ for } m \in M_{b,market}$$

$$Z_m = Z_o \text{ for } m \in M_{o,market}$$

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

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

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

Because all offline (online) merchants have the same process efficiency, face the

same wage, and are monopolistic competitors facing the common elasticity of demand  $\sigma$ , they price at a common markup over their common marginal cost:

$$p_m = p_b = \frac{\sigma}{\sigma - 1} \cdot \frac{w}{Z_b} \text{ for } m \in M_{b,market}$$

$$p_m = p_o = \frac{\sigma}{\sigma - 1} \cdot \frac{w}{Z_o} \text{ for } m \in M_{o,market}$$

With prices the same, consumers will spend the same amount  $p_b x_b$  ( $p_o x_o$ ) at each offline (online) merchant. Spending per merchant online versus offline satisfies<sup>26</sup>

$$\frac{p_o x_o}{p_b x_b} = \left( \frac{q_o/p_o}{q_b/p_b} \right)^{\sigma-1}.$$

The higher is quality (relative to price) online relative to offline, the higher the spending per merchant online relative to offline.

In turn, merchant profits online and offline are

$$\pi_o = \frac{M_o}{M_{o,market}} L \cdot \frac{p_o x_o}{\sigma} - w K_o$$

$$\pi_b = \frac{M_b}{M_{b,market}} L \cdot \frac{p_b x_b}{\sigma} - w K_b$$

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

<sup>26</sup>Online Appendix D provides further details on the model solution.

The utility-maximizing share of spending online is

$$(5) \quad s_o \equiv \frac{M_o \cdot p_o x_o}{M_o \cdot p_o x_o + M_b \cdot p_b x_b} = \frac{k}{k+1}$$

The online share  $s_o$  rises with  $q_o/q_b$ ,  $Z_o/Z_b$ , and  $A_o/A_b$ . Consumers gain from rising  $s_o$  if it is due to a combination of online options becoming better (rising  $q_o$ ), online options becoming cheaper (rising  $Z_o$ ), and easier access to online merchants (rising  $A_o$ ). Note that higher quality and lower price online are isomorphic in terms of their impact.

Consumption-equivalent welfare is proportional to<sup>27</sup>

$$M^{\frac{1}{\sigma-1}} \cdot \left[ \frac{1}{M} \sum_m (q_m \cdot Z_m)^{\sigma-1} \right]^{\frac{1}{\sigma-1}}$$

Welfare is increasing in process efficiency ( $Z$ ), variety ( $M$ ), and quality ( $q$ ) of merchants visited. In terms of exogenous driving forces, consumption-equivalent welfare is proportional to

$$\left( \left[ A_o^{\frac{1}{\phi}} (Z_o q_o)^{\sigma-1} \right]^{\frac{\phi}{\phi-1}} + \left[ A_b^{\frac{1}{\phi}} (Z_b q_b)^{\sigma-1} \right]^{\frac{\phi}{\phi-1}} \right)^{\frac{1}{\sigma-1} \frac{\phi-1}{\phi}}$$

Consumers are better off if process efficiency rises so that products are cheaper (higher  $Z_o$  and  $Z_b$ ), 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  $Z_b$ ,  $q_b$ ,  $A_b$ , consumer gains from rising  $Z_o$ ,  $q_o$  and  $A_o$  can be quantified from  $s_o$ , the share of spending online, and the values for  $\sigma$  and  $\phi$ :

$$(6) \quad Z_b \cdot q_b \cdot A_b^{\frac{1}{\phi(\sigma-1)}} \left( \frac{1}{1-s_o} \right)^{\frac{\phi-1}{\phi(\sigma-1)}}$$

Welfare gains are increasing in  $s_o$ , which itself is increasing in the quality and accessibility of online options. For given  $s_o$ , the gains are falling with  $\sigma$ . Consumers can more easily substitute from offline to online options when  $\sigma$  is higher, so online offerings do not need to improve as much (in price, quality, or accessibility) to explain a given rise in online share. The gains are increasing in  $\phi$ . 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.

<sup>27</sup>That is, doubling this expression has the same impact on utility as doubling the quantity  $x_m$  of every good bought.

To be clear, by using the spending share online to quantify welfare gains, we are isolating the effects of falling  $Z$ ,  $q$  and  $A$  online *relative* to offline.

### C. 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 exploit how  $\phi$  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  $\phi$  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.

Specifically, we exploit the following decomposition of spending into the extensive and intensive margins:<sup>28</sup>

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

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

where  $M = M_o + M_b$ . To consistently estimate the parameter  $\phi$  from (7) and (8) via OLS, we must assume that any idiosyncratic fixed shopping costs are uncorrelated with total spending across consumers.<sup>29</sup>

Table 4 contains our estimates for  $\phi$ . We estimate  $\phi$  separately for 2007 and 2017, and the estimates are similar in the two years. Our average point estimate across the two years is 1.74. The standard errors are tiny given that each regression involves hundreds of millions of cards. A  $\phi$  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, so that the extensive margin plays a bigger role (access costs are less convex): the extensive margin accounts for 57% of additional spending and the intensive margin accounts for 43% of additional spending across high and low spending cards.

We can also estimate versions of equation (7) or (8) separately for online and offline transactions. The resulting estimate for  $\phi$  turns out to be larger offline (about 2.5) than online (about 1.6). We will discuss below how entertaining a smaller (larger)  $\phi$  would lower (raise) the gains from online access.<sup>30</sup>

<sup>28</sup>We provide the full derivation for the estimating equations below in online Appendix D.

<sup>29</sup>The decomposition is exact, so the  $\phi$  estimate is the same from either of the two equations.

<sup>30</sup>Another source of heterogeneity could be that high income individuals have a high opportunity cost

TABLE 4—ESTIMATES OF FIXED SHOPPING COST CONVEXITY

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

*Note:* Each column represents a separate regression. The estimates of  $\phi$  are from the OLS regression  $\ln M = \alpha + \frac{1}{\phi} \cdot \ln(oM_o + bM_b) + \epsilon$ , where  $M$  denotes distinct merchants visited and  $oM_o + bM_b$  denotes total card spending. One observation is a card-year. We run this regression separately for 2007 and 2017. See online Appendix C for more details.

To calibrate  $\sigma$ , the elasticity of substitution across merchants, we use variation in online spending induced by physical distance between each card  $i$  and each brick-and-mortar merchant  $j$ . We assume a cardholder's distance to physical stores is uncorrelated with individual shopping costs online versus at physical stores (conditional on chain fixed effects).

We estimate the elasticity of substitution using purchases for the 1% sample of cards in 2017 described in Section III. 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 category, all pairs of physical stores  $j$  and online merchants  $k$  such that  $i$  buys from at least one of these. We then calculate the share of combined trips for each pair that were made online, and average across cards for each NAICS category. In Figure 7, we show this fraction of combined purchases made online as a function of card distance to each physical store.

For comparison, we also generate an offline substitution estimate by constructing all pairs of physical stores  $j$  and  $k$  such that  $i$  buys in at least one of these 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 category. See Figure A3 in online Appendix C.

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

of time, and hence high fixed shopping costs. This could bias  $\phi$  upward, leading us to overstate the gains from e-commerce. To gauge how big a problem this might be, we used the credit reporting agency data to control for household income for Visa credit cards in 2017. As expected, for given card spending, richer households purchased from fewer merchants. But the implied  $\phi$  fell very little, from 1.69 to 1.68, once controlling for income. See the notes on Table 4 in online Appendix C.



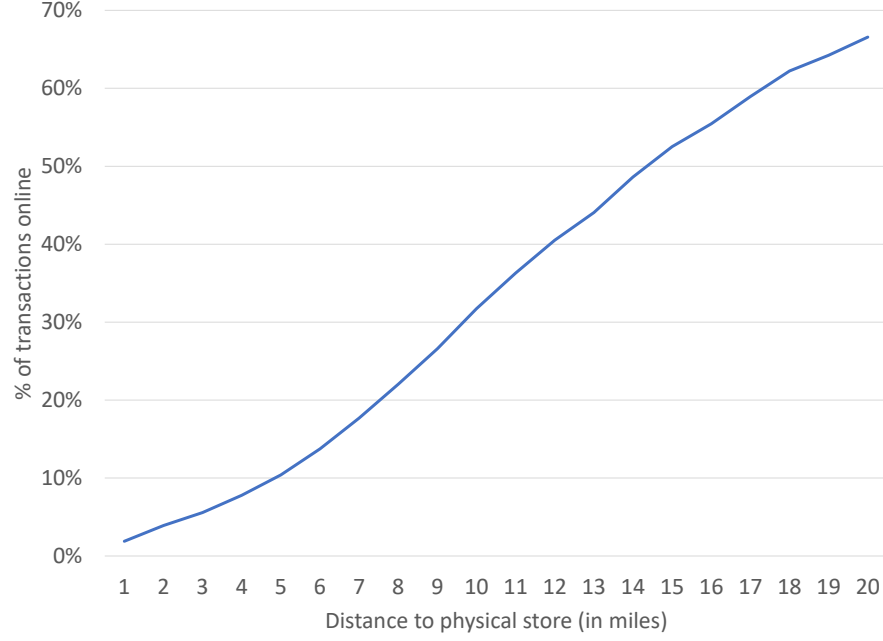


FIGURE 7. % TRANSACTIONS ONLINE VS. DISTANCE TO A PHYSICAL STORE

*Note:* This graph is based on a 1% random sample of cards in 2017. The underlying observations are card-store-merchant triples such that the card transacted either offline at the store or online at the merchant (or both), the store is within 20 miles of the cards, and the store and the merchant are in the same 3-digit retail e-commerce industry. The x-axis is distance of the store from the card (in 1 mile bins). The y-axis is percentage of online transactions out of total transactions. See online Appendix C for more details.

size plus travel costs — for going to the closer store (or shopping online) vs. the farther store (or the brick-and-mortar store). We group consumers based on their distance from two merchants  $j$  and  $k$  (denoted  $d_j$  and  $d_k$ ).<sup>31</sup> We regress the log relative number of trips on log relative prices inclusive of travel costs, controlling for merchant fixed effects:<sup>32</sup>

$$(9) \quad \ln \left( \frac{Trips_{j,d_j}}{Trips_{k,d_k}} \right) = \ln \left( \frac{q_j}{q_k} \right) - \sigma \cdot \ln \left( \frac{p_{jk} + \tau_{d_j}}{p_{jk} + \tau_{d_k}} \right)$$

<sup>31</sup>In the offline-online estimation, the online merchant has  $\tau = 0$  (since the consumer does not travel to purchase online).

<sup>32</sup>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.

Here  $p_{jk}$  is average ticket size at merchants  $j, k$ ;  $\tau$  = transportation costs for traveling  $d_j$  or  $d_k$  miles 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 an idiosyncratic preference for merchants.

As shown in Table 5, we estimate an elasticity of substitution  $\sigma$  between online and offline merchants of 4.3. This regression involves 3.6 million merchant pair observations, so 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 bias our estimate of  $\sigma$  upward.

For comparison, Table 5 also reports our estimate of the elasticity of substitution across offline merchants. This  $\sigma$  is higher at 6.1. Although our model preferences feature a common  $\sigma$ , we think the  $\sigma$  for online-offline competition is the relevant one for evaluating the gains to consumers from switching from offline to online spending. We will report robustness of our welfare calculation to using the higher  $\sigma$  across physical stores.<sup>33</sup>

As a robustness check, we estimated  $\sigma$  using  $p_j$  and  $p_k$  instead of  $p_{jk}$  in (9). That is, instead of assuming a common price across merchants based on their average ticket sizes, we assumed their respective ticket sizes reflect price differences. When we do so, we obtain a modestly lower  $\sigma$  of 4.0, versus our baseline of 4.3 when comparing offline and online options. When we estimate  $\sigma$  across competing offline merchants, we likewise get a smaller  $\sigma$  of 5.4 with price heterogeneity versus 6.0 when we assume common prices.

In online Appendix C we check the robustness of our  $\sigma$  estimates to using card and merchant longitude-latitude, rather than ZIP-centroid. We only have card location for the 50% of credit cards for which we have credit bureau data. In online Appendix Table A.1 we report these  $\sigma$  estimates. Incorporating the more precise location measure modestly increases our estimates of  $\sigma$ .<sup>34</sup> We do not make this the baseline since the sample is restricted. Also, since cardholders can make multi-destination shopping trips, it is not clear whether the card address or shopping ZIP-centroid is a better yardstick for shopping distance. Still, we report robustness below to using a value of  $\sigma$  in excess of 6.

#### D. Consumer surplus

Using our estimates of  $\phi$  and  $\sigma$  and the online share  $s_o$  in the Visa data, we can calculate consumption-equivalent changes in consumer welfare from the rise of e-

<sup>33</sup>We are unable to measure substitution between two online merchants using this approach as there is no variation in travel time.

<sup>34</sup>Using this restricted sample of cards with matched credit bureau data, we estimate  $\sigma$  to be 5.8 when we locate cards and businesses using their zip centroids and 6.3 when we use card and merchant addresses.

TABLE 5—ESTIMATES OF SUBSTITUTABILITY

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

*Note:* Each column represents a separate regression. Coefficients are from the regression  $\ln\left(\frac{Trips_{j,d_j}}{Trips_{k,d_k}}\right) = \ln\left(\frac{q_j}{q_k}\right) - \sigma \cdot \ln\left(\frac{p_{jk} + \tau d_j}{p_{jk} + \tau d_k}\right)$ . Observations are transactions from a 1% random sample of cards in 2017 wherein the card transacted with at least one of stores  $j$  and  $k$  (located at distance  $d_j$  and  $d_k$  miles from the consumer) at competing merchants in the same industry and in a retail e-commerce NAICS category. In ‘online-offline’  $j$  is a merchant with online sales and  $k$  a store within 20 miles of the card. In ‘offline-offline’ both  $j$  and  $k$  are stores within 20 miles of the card.  $p_{jk}$  denotes the average ticket size across merchants  $j$  and  $k$  and  $\tau$  a monetized cost of the return trip to the store. Both regressions are implemented using cross-store fixed effects. See online Appendix C for more details and robustness checks.

commerce. See Table 6. We estimate an increase in consumer surplus from rising e-commerce between 2007 and 2017 equivalent to 0.38% of annual consumption. Relative to a counterfactual where the online channel is completely unavailable, e-commerce in 2017 resulted in gains for consumers of 1.06% 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 rise in the spending share of online merchants ( $s_o$ ).

Table 6 also illustrates how the gains change with the parameter values. If we use the lower  $\phi$  estimated from spending on offline merchants only (1.58 versus the baseline value of 1.74), the welfare gains fall from 1.06% to 0.91% of consumption. If we use the higher, offline  $\sigma$  of 6.1 (rather than 4.3) the gains fall to 0.68% of consumption. These sensitivity checks go in the expected direction.

As we highlighted in Section II, 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 7 we show welfare gains by splits of income and county population density.<sup>35</sup>

Cardholders with income of \$50k or less enjoyed gains equivalent to 0.45% of their consumption from online shopping. Richer households enjoyed more than

<sup>35</sup>We use the same  $\phi$  and  $\sigma$  values of 1.74 and 4.3 for every group, but use group-specific online spending shares  $s_o$ .

TABLE 6—CONSUMPTION-EQUIVALENT WELFARE GAINS FROM E-COMMERCE

	$\phi$	$\sigma$	$s_o^{2017}$ vs. $s_o^{2007}$	$s_o^{2017}$ vs. $s_o = 0$
Baseline	1.74	4.3	0.38%	<b>1.06%</b>
Offline $\phi$	1.58	4.3	0.33%	0.91%
Offline $\sigma$	1.74	6.1	0.24%	0.68%

*Note:* The consumption-equivalent welfare gain is  $\left( \frac{1-s_o^{old}}{1-s_o^{new}} \right)^{\frac{\phi-1}{\phi} \frac{1}{\sigma-1}}$ , where  $s_o$  denotes the online share.  $Z$ ,  $A_b$  and  $q_b$  are held constant across years. The results are obtained by substituting in the respective values of  $s_o$ ,  $\phi$  and  $\sigma$ .

twice the gains, which were equivalent to 1.32% of their consumption. The gains were also increasing in population density, rising from 0.67% for the sparsest counties to 1.07% for the most densely populated counties.

We have framed these gains as a percentage of all consumption, but it is also interesting to express 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.<sup>36</sup>

We have assumed the online share of Visa spending is representative of all credit and debit card spending. If we assume, further, that Visa is representative within each NAICS category, then we can entertain a nested CES structure as a robustness check. Substitutability is surely higher within than across NAICS categories, whereas our CES utility in equation (4) assumes the same elasticity within and across NAICS categories.<sup>37</sup>

By moving to a nested CES structure, we can allow  $\sigma$  to vary by NAICS category. We implement this for ten 3-digit NAICS categories with a physical store component along with online spending. For five 3-digit NAICS categories which are big online but have little offline spending (such as Air Transportation), we use the overall estimate of  $\sigma = 4.3$ . We do the same for a catch-all category containing all NAICS dominated by offline spending (such as Gasoline). Table 8 provides the  $\sigma$  estimates for the 10 overlapping online-offline categories. The elasticity ranges from a high of 7.7 for building materials and garden supplies to a

<sup>36</sup>This is modest compared to the Peter Cohen, Robert Hahn, Jonathan Hall, Steven Levitt and Robert Metcalfe (2016) estimate of consumer surplus equal to 160% of spending on Uber.

<sup>37</sup>We did estimate  $\sigma$  across merchants within NAICS categories, above, keeping in mind that such substitutability was sure to be higher.

TABLE 7—WELFARE GAINS BY CARDHOLDER INCOME AND COUNTY POPULATION DENSITY

	Gains from $s_o^{2017}$ vs. $s_o = 0$
Income $\leq$ \$50k	0.45%
Income $>$ \$50k	1.32%
Below-median density	0.67%
Above-median density	1.07%

*Note:* The consumption-equivalent welfare gain is  $\left(\frac{1}{1-s_o^{2017}}\right)^{\frac{\phi-1}{\phi} \frac{1}{\sigma-1}}$ , where  $s_o^{2017}$  denotes the online spending share in 2017 for an income or density group. We use the same  $\phi$  and  $\sigma$  for each group.

low of 3.4 for electronics and appliance stores. We assume the upper nest, which aggregates our 16 lower CES nests, is simply Cobb-Douglas.

An ambiguity that arises with the nests is how to treat the nonstore retailer NAICS category, which contains online-only retailers such as Amazon. We allocate nonstore retail spending based on estimates of Amazon’s sales by NAICS.<sup>38</sup> That is, we allocate nonstore retailer spending into electronics and appliances, clothing, etc. based on estimates of the distribution of Amazon’s sales.

In Table 9 we report the welfare gain under nested CES. The gain computed with 16 nests is 1.05%, nearly identical to our baseline estimate of 1.06% of consumption with a single nest.

Our stylized model features free entry for both offline and online merchants. As a 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 10 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 3.4% decline in the number of physical stores, with a 1.5% decline in spending per store and a 2% decline in the number of physical stores visited per card. The effect on the profits of brick-and-mortar retailers is zero by construction.<sup>39</sup>

<sup>38</sup>Source: eMarketer estimates for 2017 found in <https://www.cnbc.com/2018/07/12/amazon-to-take-almost-50-percent-of-us-e-commerce-market-by-years-end.html>.

<sup>39</sup>Farrell, Relihan and Ward (2018) document the lackluster growth in offline retail spending amid rapidly rising retail spending. Relihan (2017) estimates that online grocery shopping crowds out offline grocery shopping, but crowds in spending at coffee shops.

TABLE 8—ESTIMATES OF SUBSTITUTABILITY BY NAICS CATEGORY

	$\sigma$
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
Sporting Goods, Hobby, Music, Book Stores	4.2
Food and Beverage Stores	3.6
Electronics and Appliance Stores	3.4

*Note:* Estimates are across offline versus online merchants within each listed NAICS. For other NAICS (Air Transportation, Ground Transportation, Rental and Leasing Services, Administrative and Support Services, Accommodation) the offline component was sufficiently limited that we used the overall offline-online estimate of  $\sigma = 4.3$ . See online Appendix C for more details.

TABLE 9—NESTED CES WELFARE GAIN IN 2017

Single nest (baseline)	1.06%
16 nests (nonstore retail allocated)	1.05%

*Note:* We compare the welfare gains under nested CES preferences to our single nest benchmark. Each nest is a 3-digit NAICS. The consumption equivalent welfare gain with nested CES preferences equals  $\left(\prod_m (1 - s_m)^{-\frac{\alpha_m}{\sigma_m - 1}}\right)^{\frac{\phi - 1}{\phi}}$ . The results are obtained by substituting in the sector specific online shares  $s_m$  and elasticities of substitution  $\sigma_m$ . The outer nest Cobb-Douglas elasticities  $\alpha_m$  are calibrated using spending shares. See online Appendix C for more details.

TABLE 10—RETAIL APOCALYPSE?

		2007–2017 Change
$b$	Per card spending per offline merchant	–1.5%
$M_b$	Per card # of offline merchants bought from	–2.0%
$M_{b,market}$	total # of offline merchants in the market	–3.4%
$\Pi$	Profits of offline merchants	0%

*Note:* The change in the share of spending online is a sufficient statistic for assessing changes in spending per offline merchant, number of offline merchants visited and number of offline merchants in the market in our model (holding  $Z$ ,  $A_b$ ,  $K_b$  and  $q_b$  constant). The corresponding formulae are given by  $b_{2017}/b_{2007} = [(1 - s_{2017}) / (1 - s_{2007})]^{\frac{\phi - 1}{\phi}}$ ,  $M_{b,2017}/M_{b,2007} = [(1 - s_{2017}) / (1 - s_{2007})]^{\frac{1}{\phi}}$ ,  $M_{b,market,2017}/M_{b,market,2007} = (1 - s_{2017}) / (1 - s_{2007})$ . The results are obtained by using our baseline estimate of  $\phi = 1.74$ .

## V. Conclusion

We take advantage of a unique data source — all credit and debit card 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. 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 find convenience gains equivalent to at most 0.4% of consumption. We then write down a representative consumer model which allows for substitution across merchants and both variety and quality gains. Our main estimate using this model is a welfare gain equivalent to over 1% of consumption in 2017, or over \$1,000 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. Future work could usefully analyze the impact of online spending on prices, quality, variety, and the number of stores offline.

## REFERENCES

- Akerman, Anders, Ingvil Gaarder, and Magne Mogstad.** 2015. “The Skill Complementarity of Broadband Internet.” *Quarterly Journal of Economics*, 130(4): 1781–1824.
- Broda, Christian, and David E. Weinstein.** 2006. “Globalization and the Gains from Variety.” *Quarterly Journal of Economics*, 121(2): 541–585.
- Broda, Christian, and David E. Weinstein.** 2010. “Product Creation and Destruction: Evidence and Price Implications.” *American Economic Review*, 100(3): 691–723.
- Brynjolfsson, Erik, and JooHee Oh.** 2012. “The Attention Economy: Measuring the Value of Free Digital Services on the Internet.”

- Brynjolfsson, Erik, Yu Hu, and Michael D Smith.** 2003. "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers." *Management Science*, 49(11): 1580–1596.
- Cavallo, Alberto.** 2017. "Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers." *American Economic Review*, 107(1): 283–303.
- Cavallo, Alberto.** 2018. "More Amazon Effects: Online Competition and Pricing Behaviors." *Jackson Hole Economic Symposium Conference Proceedings (Federal Reserve Bank of Kansas City)*.
- Cohen, Peter, Robert Hahn, Jonathan Hall, Steven Levitt, and Robert Metcalfe.** 2016. "Using Big Data to Estimate Consumer Surplus: The Case of Uber." National Bureau of Economic Research Working Paper 22627.
- Couture, Victor, Benjamin Faber, Yizhen Gu, and Lizhi Liu.** 2020. "Connecting the Countryside via E-Commerce: Evidence from China." *American Economic Review: Insights*.
- Einav, Liran, Amy Finkelstein, and Heidi Williams.** 2016. "Paying on the Margin for Medical Care: Evidence from Breast Cancer Treatments." *American Economic Journal: Economic Policy*, 8(1): 52–79.
- Farrell, Diana, Lindsay Relihan, and Marvin Ward.** 2018. "Shopping, Near and Far: Local Commerce in the Digital Age." *JPMorgan Chase Institute*.
- Goolsbee, Austan, and Peter J. Klenow.** 2006. "Valuing Consumer Products by the Time Spent Using Them: An Application to the Internet." *American Economic Review*, 96(2): 108–113.
- Goolsbee, Austan D., and Peter J. Klenow.** 2018. "Internet Rising, Prices Falling: Measuring Inflation in a World of E-Commerce." *AEA Papers and Proceedings*, 108: 488–92.
- Handbury, Jessie, and David E. Weinstein.** 2014. "Goods Prices and Availability in Cities." *Review of Economic Studies*, 82(1): 258–296.
- Hjort, Jonas, and Jonas Poulsen.** 2019. "The Arrival of Fast Internet and Employment in Africa." *American Economic Review*, 109(3): 1032–79.
- Hopenhayn, Hugo A.** 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica*, 1127–1150.
- Huang, Yufeng, and Bart J Bronnenberg.** 2020. "Gains from Convenience and the Value of E-commerce." *Available at SSRN 3596460*.
- Jo, Yoon J., Misaki Matsumura, and David E. Weinstein.** 2019. "The Impact of E-Commerce on Relative Prices and Consumer Welfare." National Bureau of Economic Research Working Paper 26506.



- Quan, Thomas W., and Kevin R. Williams.** 2018. "Product Variety, Across-market Demand Heterogeneity, and the Value of Online Retail." *RAND Journal of Economics*, 49(4): 877–913.
- Redding, Stephen J., and David E. Weinstein.** 2020. "Measuring Aggregate Price Indices with Taste Shocks: Theory and Evidence for CES Preferences." *Quarterly Journal of Economics*, 135(1): 503–560.
- Relihan, Lindsay.** 2017. "Is Online Retail Killing Coffee Shops? Estimating the Winners and Losers of Online Retail using Customer Transaction Microdata."
- Syverson, Chad.** 2017. "Challenges to Mismeasurement Explanations for the US Productivity Slowdown." *Journal of Economic Perspectives*, 31(2): 165–86.
- Varian, Hal.** 2013. "The Value of the Internet Now and in the Future." *The Economist*.