ONLINE APPENDIX
Assessing the Gains from E-Commerce

by Dolfen, Einav, Klenow, Klopack, Levin, Levin, and Best

October 2021

This online appendix contains the following four items. Appendix A describes the data sources and the data samples we use for our analyses. Appendix B provides further detail on how we measure online spending in our data and the definition of e-commerce we use. Appendix C provides more detailed notes about the construction of the figures and tables in the main text. Finally, Appendix D contains further details about model derivations.

Appendix A. Data and Samples

A.1. The Raw Datasets

Visa Transaction Table

The main source of raw data is a proprietary dataset by Visa Inc. covering the universe of transactions on the Visa network. This is a transaction-level dataset, which contains information on all transactions (both credit and debit cards) that were processed through the Visa network. For the current paper, we use all transactions that occurred between January 1, 2007 and December 31, 2017.

The main variables for our analysis are the card number, a merchant identifier, the transaction ZIP code (which is only meaningful for brick-and-mortar transactions), the transaction amount, and the transaction date (as well as the exact time, which we do not currently use). One limitation of this dataset is that multiple outlets of the same merchant within the same ZIP code are not distinguishable (we address this issue with a different Visa dataset – the GMR table – discussed below). To that end, the transaction table also contains an establishment identifier (available since mid 2015) which can be linked to the GMR table.

We furthermore observe two variables that identify whether or not the card was present for the transaction (the card would be present for a brick and mortar transaction, but not for e.g. an e-commerce transaction). The first indicator is always available and distinguishes between card present (hereafter referred to as CP) and card not present (hereafter referred to as CNP). The second indicator allows a further breakdown of CNP transactions into various categories: e-commerce, mail order, phone order, or recurring transactions (e.g. phone bills). These two indicators will be the basis for our measures of e-commerce on the Visa network.

We do not directly observe any card attributes in this dataset. But we use the card's transactions to create a card-year location variable based on the brick and mortar transactions of the card-year. In particular, we define a card-year's location to be the transaction-weighted average of the longitudes and latitudes of the brick and mortar transaction locations of that card. We only use transaction ZIP codes in which the card transacted 20 or more times in a given year to avoid contamination of the card location by e.g. transactions during holidays. We use the ZIP code centroid to assign longitude and latitude to transaction ZIP codes. We also have access to a different measure of card location based on credit bureau data (discussed below). This is however only available for a subset of cards and only in more recent years. It is worth noting that our measure of card location based on a card's transaction performs very well when compared to this external data.

One important limitation of the transaction data concerns the merchant identifier. While every transaction is assigned a merchant identifier, this identifier does not always allow us to infer the exact merchant. The Visa data distinguishes between two types of merchants, "named" and "unnamed" merchants. Roughly, "named" merchants are large chains for which Visa assigns a unique merchant id, i.e. there is a unique mapping between Visa's merchant id and the merchant. "Unnamed" merchants are typically smaller chains and single establishment merchants, for whom Visa did not assign a separate ID. These "unnamed" merchants have a unique establishment identifier, and are classified (by Visa) to a specific industry (NAICS), but multiple stores of the same chain are not associated with each other. While most of our analysis will be based on transactions at both "named" and "unnamed" merchants, two parts of the paper require linking outlets of the same merchants, and are therefore based on "named" merchants only. First, the estimates of the convenience benefits of e-eommerce. Second, the estimate of the elasticity of substitution across different merchants. We should note, however, that our aggregates of online spending on the Visa network, our estimate of total online spending in the US and our estimate of the welfare benefit of e-commerce are all based on spending at both "named" and "unnamed" merchants. Overall, 58% of dollar volume in our sample are transacted at "named" merchants.

Visa GMR Table

Global Merchant Repository (GMR) is an effort by Visa to create a master file of merchant information from data provided by the merchant's acquiring bank and from external data providers. All Visa transactions were linked to a GMR entity via a unique identifier starting in mid 2015, and are thus available for the last 2.5 years of our sample. Each GMR-stamped transaction is mapped to a merchant ID and a store ID. For each store, GMR contains the mailing address and a corresponding latitude/longitude information.

Credit Bureau Data

We have access to an additional dataset that provides cardholder-level demographics which can be linked to a sample of Visa credit cards. This dataset is provided by a large credit bureau. About 50% of active credit cards (but not debit cards) in 2016 and 2017 were linked to an entity in this dataset. For each cardholder matched to the credit bureau data, we observe the cardholder's age and their 9-digit billing ZIP code, as well as their estimated household income, marital status, number of children, and education level.

A.2. The Analyses Samples

The Transaction Sample

We impose several sample restrictions. We focus on the transaction of Visa credit and debit cards at U.S. merchants (i.e. we discard non-Visa cards that occasionally transact through the Visa network, as well as Visa pre-paid Cards, and we discard transactions that occur outside the U.S.). Furthermore, we focus on Credit and Debit-Signature transactions only. This mainly excludes Debit-PIN transaction. Following the Durbin Amendment in 2010 (part of the Dodd-Frank Bill), Visa was not able to restrict how merchants routed Debit-PIN transactions. Therefore, starting in 2012 (when the law went into effect), the data exhibit significant fluctuations in the volume of Debit-PIN transactions of stores, making it less reliable. We thus focus on Credit and Debit-Signature transactions, for which merchants' network routing appears fairly consistent. The transactions that satisfy our filters account for 91.5% of the total dollar volume on the Visa network.

We further impose the additional restriction that cards in our sample must have transacted with at least five merchants over their lifetime. This filter was chosen to exclude cards that are only used for one merchant and gift cards (there is a large number of cards that only transact with one merchant for a total of USD 50 or USD 100). Overall, transactions that satisfy this filter as well account for 87.6% of the total dollar volume on the Visa network.

The Convenience Sample

In the convenience analysis (Section 4), we use 2017 transactions from a 1% random sample of the cards that were matched to the credit bureau data in nine "mixed" retail NAICS categories – i.e., those that had online share between 10% and 90%. Those include the following 3-digit NAICS codes: 441, 442, 443, 444, 446, 448, 451, 452, and 453. We count offline transactions as those that were marked as occurring face-to-face (i.e. with the card physically swiped) and online as those that were marked with the e-commerce indicator. We exclude phone order, mail order, and recurring transactions from this analysis.

For each transaction, we calculate the distance between the merchant and the card as the distance between the card's ZIP+4 from the credit bureau data and the closest offline branch of that merchant (defined as the latitude and longitude of the store as recorded by Visa). We keep transactions that occurred at merchants that had an offline presence within 50 miles of the consumer's location. We also exclude merchants that had a greater than 99% online share or less than 1% online share within our sample transactions.

The Variety Sample

The variety sample consists of a random 1% sample of cards in 2017. We only consider transactions of these cards at named merchants (because controlling for the exact merchant identity is important) and in the narrower set of e-commerce industries (excluding Nonstore Retail) (3-digit NAICS 441, 442, 443, 444, 445, 446, 448, 451, 452, 453). We choose these industries (as opposed to the baseline set of industries) because our estimation strategy again relies on distance to the merchant which is less relevant in the non-retail e-commerce industries (e.g. Hotels, Car Rental). We also exclude Nonstore Retail because distance to brick and mortar stores is not meaningful in Nonstore Retail. We furthermore restrict our analysis to transactions for which the card is located within 20 miles of the merchant (using the transaction based measure of card location and locating a physical store and the ZIP code centroid).

The first σ we estimate is based on choices between online and offline merchants. For that we use all the transactions in the Variety Sample. We construct all pairs of physical store j and CNP merchant k such that card i buys from one of these. We require that both merchants are in the same 3 digit industry, that the store is within 20 miles of the card location and that merchant k has CNP revenues in that year.

The second estimated σ is based on the comparison of different offline choices. For this estimation, we only use the CP transactions in our Variety Sample. We then construct, for each individual i and 3-digit NAICS, all pairs of physical stores j and k such that i buys in at least one of these stores. We furthermore require that both stores are in the same 3 digit industry and both within a 20 mile radius of the card location.

Appendix B. Measuring E-Commerce in the Visa Data

The E-Commerce Industries

We define e-commerce industries to be industries that are affected through gains in buying/shopping convenience and/or increased variety by the rise of e-commerce. We hence choose to not include industries in which the convenience is only in terms of payment. Examples of industries which we believe to only be affected through convenience in terms of payment are utilities, telecommunication, and broadcasting.

Our baseline set of e-commerce industries is as follows: Motor Vehicle and Parts Dealers (441), Furniture and Home Furnishing Stores (442), Electronics and Appliances Stores (443), Building Material and Garden Equipment and Supplies Dealers (444), Food and Beverage Stores (445), Health and Personal Care Stores (446), Clothing and Clothing Accessories Stores (448), Sporting Goods, Hobby, Musical Instrument, and Book Stores (451), General Merchandise Stores (452), Miscellaneous Store Retailers (453), Nonstore Retailers (454), Air Transportation (481), Transit and Ground Passenger Transportation (485), Rental and Leasing Services (532), Administrative and Support Services (561), Accommodation (721)

We will however also report an alternative estimate of e-commerce which is based on a narrower set of industries, by focusing only on the above 3-digit NAICSs that belong to the 2-digit Retail NAICSs (44 and 45).

E-Commerce Variables in the Visa data

In the following we will distinguish between transactions that are CNP (Card Not Present) and CP (Card Present). CP transactions are brick and mortar transactions in which the card was swiped, whereas CNP transactions refer to transactions for which the card was not swiped, and contain four indicators in the data: E-commerce, Recurring Transactions (e.g. utilities, phone bills), Mail Order, and Telephone Order.

As highlighted in the Data section, we observe two variables at the transaction level that allow us to distinguish whether a card was present or not during the transaction. The first is a CNP indicator that distinguishes between CP and CNP transactions. This variable is automatically created by Visa and is available for every transaction (given how important it is for fraud detection). The second variable is an e-commerce indicator. This variable is filled in by merchants and allows a further breakdown of CNP transactions into e-commerce, Mail Order, Telephone Order and Recurring Transactions. One limitation of this latter indicator is that it is not required by Visa and about half of the values are missing (the share of missing values has been declining over time). This implies that we often cannot identify whether a CNP transaction is e-commerce or not. Since this variable is filled in by merchants, whether or not the e-commerce indicator is missing varies quite a lot from merchant to merchant.

In our final sample, 47% of the dollar volume is classified as CNP using the CNP indicator. Furthermore 20% of dollars can be classified as e-commerce and 4% as either Mail Order, Telephone Order or Recurring Transaction using Visa's e-commerce indicator. This implies that half the CNP transaction dollars on the Visa network cannot be broken down any further using information contained in Visa's datasets.

As a consequence we will not estimate e-commerce in the Visa data by only using Visa's e-commerce indicator. First, the large number of missing values would bias downwards our estimates of the e-commerce share. Second, the declining share of missing values over time would bias upwards our estimate of the rise in e-commerce. We also do not classify e-commerce by only using the CNP indicator either because this measure would overstate the e-commerce share since it also includes other CNP transactions. Below we describe our classification procedure, which combines the information of both these indicators to obtain our estimate of e-commerce in the Visa data.

Measuring e-commerce in the Visa Data

Given the data limitation described above, we combine the information of both the CNP and e-commerce indicators to estimate e-commerce spending. The underlying idea is to create (3-digit industry-year) weights that map CNP spending into e-commerce spending. We choose this strategy as every transaction on the Visa network is either classified as CP or CNP.

To create these weights, we only keep CNP transactions that have a valid e-commerce indicator (i.e. allow us to distinguish whether or not the CNP transaction was an e-commerce transaction). We then calculate the share of CNP dollars that are e-commerce dollars on this clean subsample. We do this exercise by 3-digit industry-year pairs to allow for different mappings across industries and time. To then obtain an estimate of e-commerce spending by industry-year, we multiply CNP spending in that industry-year in our full sample by the weights we calculated in the previous step. We find that 40% of the dollar volume is classified as online dollars using this methodology (Recall that 47% were CNP).

Further Details on Missing E-Commerce Indicators

We choose to create these weights at an industry-year level as opposed to at the firm level for the following reason. While the CP/CNP classification is automatically filled in by Visa for each transaction, the e-commerce indicator is filled in by the merchants. As a consequence, the e-commerce indicator is often present for either all or none of the transactions of a given merchant. In particular, merchant-years where the share of CNP dollars with missing e-commerce indicator is below 5% or above 95% account for 38% of all CNP dollars in the sample. The distribution of the share of CNP dollars with missing e-commerce classification across merchant-years is displayed in Figure A1.

We also examined manually the identity of some of the merchants with no transactions classified as e-commerce. We looked at Air Transportation, Accommodation, and Administrative and Support Services because, as we discuss below, these three industries account for a large share of imputed e-commerce dollars. The three largest merchants (in terms of CNP revenues) that do not populate the e-commerce indicator field in each of these industries are: Southwest, Alaskan Airlines, and PAL Airlines; Holiday Inn, Marriott, and Hyatt; and Funjet Vacations, Southwest Vacations, and Delta Vacations.

Further Details on Imputed E-Commerce Dollars

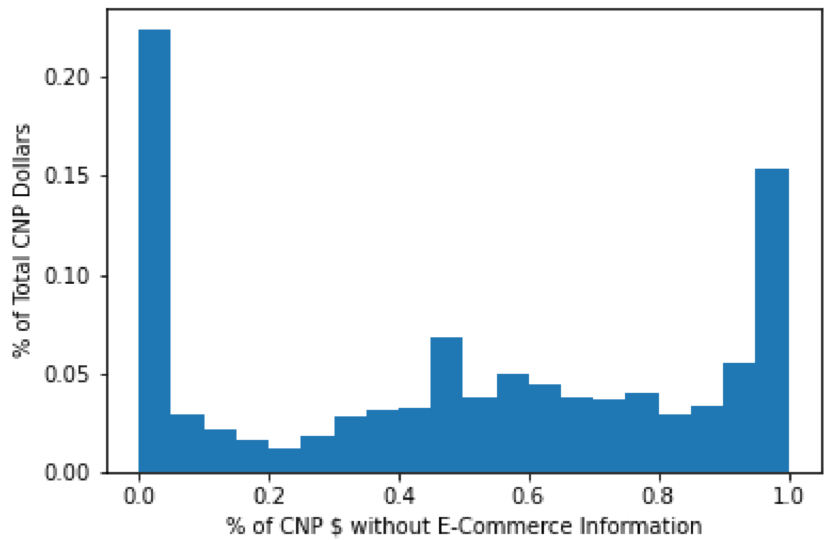
As just mentioned, the three-digit industries that account for most of the imputed e-commerce Dollars are Air Transportation (NAICS 481), Accommodation (NAICS 721) and Administrative and Support Services (NAICS 561). The last industry contains firms such as Expedia, Airbnb, Hotels.com. Out of all the e-commerce dollars that we impute in our sample, Air Transportation accounts for 20%, Accommodation accounts for 14% and Administrative and Support Services account for 9%.

We also checked whether the industries that account for the majority of the imputed e-commerce dollars have a high e-commerce share out of total CNP spending (as opposed to recurring, mail order, and telephone order transactions). Figure A2 plots the e-commerce share of industry CNP dollars (conditional on observing a valid e-commerce indicator) against the industry's share of total generated e-commerce dollars. Note that the y-coordinates sum to one across industries. First, note that all e-commerce industries have a high e-commerce share of CNP dollar. This is precisely because we chose the e-commerce industries such that the CNP industry spending reflects our definition of e-commerce (as opposed to e.g. paying utility bills online, which would be classified as a recurring transaction). Second, even among the e-commerce industries, the industries that account for the most imputed e-commerce dollars are the ones with the highest e-commerce share of CNP dollars.

Our Estimate of E-Commerce Spending on the Visa Network

To obtain our estimate of e-commerce spending on the Visa network in any given year, we add the estimated online dollars across our e-commerce industries and then divide by total spending on the Visa network in that year. As

Figure A1: Distribution of Share of CNP Dollars with Missing E-Commerce Classification Across Merchant-Years



Note: Underlying data is at a merchant-year level. For each (named) merchant-year in an e-commerce industry, calculate the share of all CNP dollars that have a missing e-commerce indicator. Bin merchant-years based on this fraction. Bars denote the share of total CNP dollars that each bin accounts for. Bar heights add up to one.

discussed above, we report two separate estimates, one using our baseline definition of the e-commerce industries and one counting only the online dollars from our e-commerce industries that are part of the Retail NAICS (44 and 45).

We classify 20% of the dollars spent on the Visa network as e-commerce, and 13% when using the narrower set of industries. Recall that 47% of dollars are CNP and 40% are estimated to be spent online (without using our e-commerce industry restrictions).

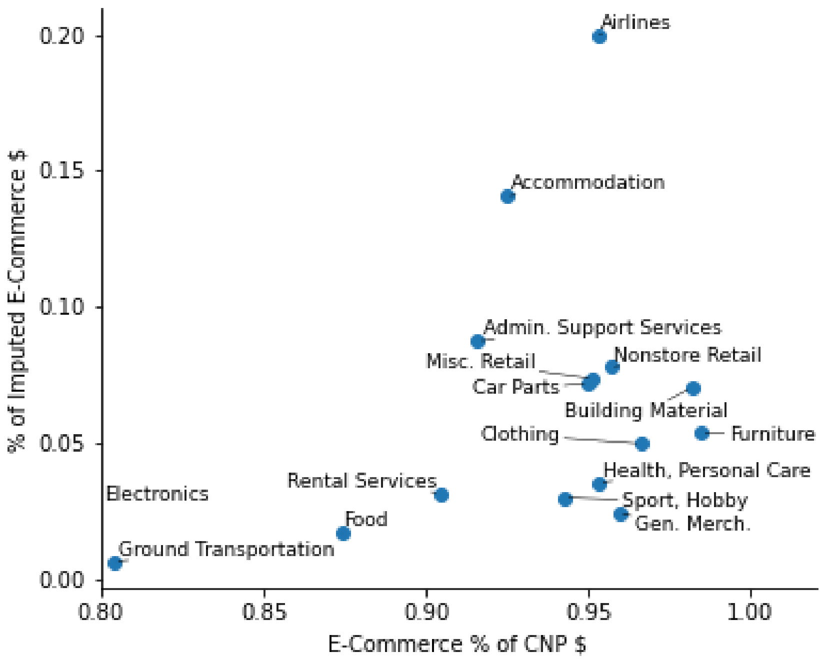
Appendix C. Additional Notes on Figures and Tables

- Figure 1: Visa spending by year is calculated as total sales draft transaction spending on Visa credit and debit cards in our sample. GDP and Consumption estimates are from the U.S. Bureau of Economic Analysis (BEA).
- Table 1: List of 3-digit NAICS that we associate with e-commerce, along with example merchants falling in each of the NAICS.
- Table 2: Estimate of the share of online spending on the Visa network in select 3-digit NAICS categories.
- Figure 2: Estimates of e-commerce spending in the U.S. as a share of all consumption. We estimate e-commerce spending on the Visa network (as discussed in Appendix B) 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. We first calculate, respectively, total amount of credit and debit card spending on the Visa network (by year) based on our transaction data. We then use external information on Visa's share of total credit and debit card spending to calculate the total amount of card spending (by year) in the U.S.[1] Having calculated total card spending by year, we then multiply this by the corresponding online share on the Visa network. Using our two assumptions, this number is our estimate of total online spending in the U.S. by year. We then divide by total consumption to obtain our estimate of the U.S. online share.[2] 'All online' refers to our baseline estimate of e-commerce spending in all consumer

[1]The external information is provided by WalletHub: https://WalletHub.com/edu/market-share-by-credit-card-network/25531/. WalletHub calculates market shares for credit and debit card spending based on the SEC filings of all major card providers.

[2]We also did an analogous aggregation taking into account the different online shares for credit and debit cards. The results are qualitatively and quantitatively very similar.

Figure A2: E-Commerce Share against Share of Imputed Dollars



Note: Underlying data is at an industry level and contains all e-commerce industries. The horizontal axis calculates, for each industry, the share of CNP transactions with a valid e-commerce indicator that are e-commerce (as opposed to recurring transactions, mail orders and telephone orders). The vertical axis displays, for every industry, the share of industry generated e-commerce dollar to total generated e-commerce dollars (i.e. these shares sum up to one across all e-commerce industries).

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.

- Figure 3: This figure displays the 2017 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-card income bin according to their home billing ZIP code and estimated household income. We compute the online share for each county-card income bin from their Visa credit card spending (as discussed in Appendix B) and then adjust for differential propensities to use credit cards. The adjustment is made in the following way: Using the credit bure denotes total spendingau data, we count the different number of Visa cards in the county - card income bins. We then obtain the population equivalent, i.e. the total number of people in the different groups, using 2015 IRS data on the number of tax filers. The population is calculated as the number of single filers + number of married filers × 2 + number of head of household filers + number of dependents. Combining these two numbers, we calculate an adjustment factor that maps the different subgroups in Visa to their population equivalent, namely

$$\alpha_{cy} = \frac{\text{Visa Cards}_{cy}}{\text{Population}_{cy}}$$

The adjusted online share is then calculated using

$$\widetilde{\text{Online Share}}_{cy} = \alpha_{cy} \cdot \text{Visa Online Share}_{cy}$$

In the final step we scale down the different $\widetilde{\text{Online Share}}_{cy}$ such that the aggregate of Online Share$_{cy}$ matches our estimated total U.S. e-commerce share. The plot shows the online share (aggregated across cardholders of different incomes) within each county.

- Figure 4: The figure shows the number of merchants visited per card against the total spending per card for offline and online transactions separately. The sample includes all credit and debit cards, subject to the filters discussed above and in the main text. Cards are binned by their (logged) amount of offline/online spending. Within each bin, the figure shows the average of the log(# of merchants visited) in offline/online channels. CNP transactions in which the e-commerce categorization is missing are treated as probabilistic e-commerce transactions, consistent with the way in which we attribute CNP spending to e-commerce. For example, a CNP transaction with a missing e-commerce indicator in a NAICS category in which 70% of CNP transactions are e-commerce is treated as an e-commerce transaction with 70% probability for a given card. If the card made three transactions at the merchant, the probability that at least one transaction by that card at that merchant was an e-commerce transaction is $1-(1-0.7)^3 = 0.973$, which we would use as the number of visits for that card-merchant combination. Because of these fractional visits, some values of log(# of merchants visited) in the figure are below zero.

- Figure 5: The figure shows the average number of merchants visited per card against the population density of the cardholder's county of residence. The sample includes only cards that can be matched to a billing zipcode using the credit bureau data. To construct the table, we first compute the number of merchants visited and total spending for offline and online transactions separately at the card level. Then we aggregate to the county level based on the cardholder's billing zipcode. Finally, we group counties into twenty bins, so that each bin contains approximately the same number of counties. For each bin, the plot shows the natural log of the average number of merchants visited against the population density of the bin.

- Table 3: The table shows summary statistics for the transactions used in the convenience analysis. The ticket size panel gives the average dollars per transaction by 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.

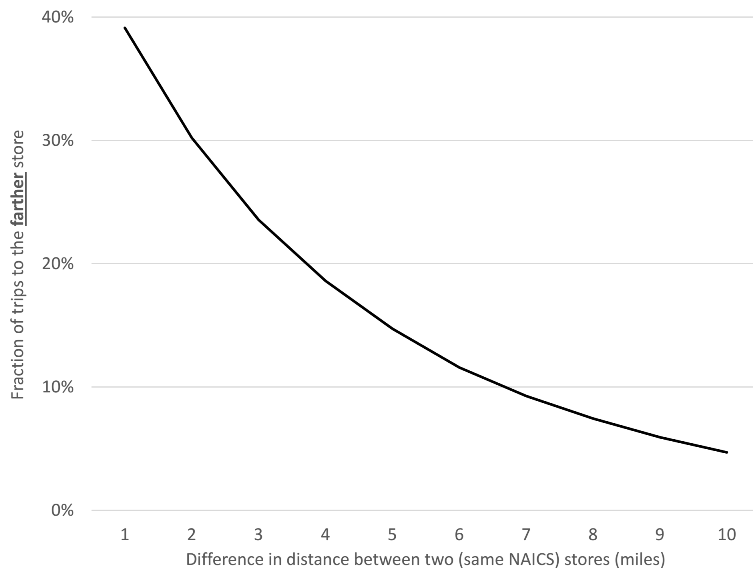
- Figure 6: The figure shows the share of transactions that occur online as a function of the distance between the card and the nearest outlet of the merchant. The sample includes transactions made by 1% of cards in 2017 at merchants in the nine mixed-channel NAICS listed in the data section. We include transactions at merchants that had a location within 50 miles of the card's billing ZIP code. The black 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 black line shows that cards that were between zero and one mile away from an outlet of a merchant conducted about 12% of their transactions with that merchant in the online channel. The grey 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.

- Table 4: Each column represents a separate regression. The estimates of ϕ 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 spending. One observation is a card-year. We run this regression separately for 2007 and 2017. As a robustness check, we ran this regression controlling for household income using credit reporting agency data. The sample is 127 million cards in 2017. For given card spending, richer households purchased from fewer merchants (elasticity -0.05). But the implied ϕ fell very little, from 1.69 to 1.68, once controlling for income.

- Figure 7: The 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 card, 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. We aggregate to the distance level by summing the online and offline transactions across card-store-merchant triples. Finally the share of transactions online is calculated as a function of the distance to the store and the observations are connected with a smoothed curve.

  We also conduct a related analysis of card choices between two (offline) stores as a function of distance to the stores. In particular, for card-store-store triples, we calculate the share of transactions at the farther store as a function of the differential distance between the stores. This relationship is depicted in Figure A3. The graph is based on a 1% random sample of cards in 2017. The underlying observations are card-store-store triples such that the card visited at least one of the two stores, both stores are within 20 miles of the card, and the stores are from merchants in the same 3-digit retail e-commerce industry. The x-axis is differential distance of the two stores from the card (in 1 mile bins). The y-axis is the share of transactions at the farther store. We aggregate to the differential distance level by summing the farther and closer transactions across card-store-store triples. Finally the share of transactions at the farther store is calculated as a function of the differential distance and the observations are connected with a smoothed curve.

Figure A3: Relative trips as a function of distance



- Estimating σ: The estimates of σ are based on the variety sample described in Appendix A.2. We restrict attention to transactions that are either CNP or CP and within 20 miles of (transaction based) card location. Based on this we will create two different datasets, each of which will yield a separate estimate of σ. The first dataset, hereafter referred to as 'offline-offline' dataset, is at the level of observation of card-store-store such that the card visits at least one of the stores. Both stores are required to be within a 20 mile radius of the card location and in the same 3 digit industry. The second dataset, hereafter referred to as 'online-offline' is at the level of observation of card-store-online merchant such that the card transacts with at least one of the entities. The store is again required to be within a 20 mile radius of the card location, the online merchant is a merchant with positive CNP sales that year and both are in the same 3 digit industry. In both datasets there are four additional variables, namely distance between the card and the merchant (set to zero for CNP purchases) (in 1 mile bins) and the number of transactions at each of the merchants. We then aggregate both datasets to the level of merchant j, merchant k, distance to j, distance to k by summing transactions and regress
$$\ln\left(\frac{Trips_j}{Trips_k}\right) = \ln\left(\frac{q_j}{q_k}\right) - \sigma \ln\left(\frac{p_{jk} + \tau_j}{p_{jk} + \tau_k}\right)$$
where p_{jk} is the average ticket size at merchants j and k (dollar weighted) and τ_j is the cost of travelling (a return trip) to j. τ_j consists of several components: First, we convert straight-line miles into driving miles (and driving time): 1 straight line mile requires 1.5 miles of driving on average (Einav et al, 2016), and one mile of driving requires 1.4 minutes of driving (Einav et al, 2016). Second, we calculate the time cost of driving. An average hourly after-tax wage of \$23 (BLS) implies a time cost of $1 \times 1.5 \times \frac{14}{600} \times 23 = \$0.80$. Third, we calculate the monetary cost of driving. An average fuel plus depreciation per mile of \$0.53 (IRS) implies a monetary cost of $1 \times 1.5 \times 0.53 = \$0.79$. Combining these three terms, the (round trip) cost of driving a (straight-line) mile is $2 \times (0.80 + 0.79) = \$3.18$. We implement the above described regression using merchant cross fixed effects to control for $\ln(q_j/q_k)$. We run this regression on both datasets to obtain, respectively, the 'offline-offline' σ and 'offline-online' σ.

- Table 5: Each column represents a separate regression. Coefficients are from the regression $\ln\left(\frac{Trips_j}{Trips_k}\right) = \ln\left(\frac{q_j}{q_k}\right) - \sigma \ln\left(\frac{p_{jk}+\tau_j}{p_{jk}+\tau_k}\right)$. 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 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. The resulting tables at the card-merchant1-merchant2 level are then aggregated to a merchant1-merchant2-distance1-distance2 level where distance denotes store distance from the card (aggregated to 1 mile bins) by summing transactions. p_{jk} denotes the average ticket size across merchants j and k and τ a monetized cost of the return trip to the store. Both regressions are implemented using cross-store fixed effects (i.e., fixed effects for the (j, k) pair).

  We also conduct several robustness checks for our estimate of substitutability. We focus on the 'online-offline' estimate because this is our baseline estimate for the welfare calculations. These baseline estimates (both in aggregate and by industry) are displayed in the first column of Table A1. We then estimate the same regressions using a 2% sample of the cards in 2017 for which we observe credit bureau data. The resulting etsimates are displayed in the second column of Table A1. Lastly, we run the same regressions on this sample using alternative location measures. In particular, we locate cards using the longitude and latitude of their billing address (available from the credit bureau data) and locate stores using their longitude and latitude (available from the Visa GMR Table). The results of this regression are diplayed in the third column of Table A1.

- Table 6: The consumption-equivalent welfare gain is $\left(\frac{1-s_{old}}{1-s_{new}}\right)^{\frac{\phi-1}{\phi}\frac{1}{\sigma-1}}$, where s denotes the U.S. online share in that year (holding Z, A_b and q_b constant). The results are obtained by substituting in the datapoints for s and using the values of ϕ and σ shown in the Table.

- Table 7: The income split is for the subset of households with credit reporting agency data on income. Counties are sorted by population density in 2017, then placed into top or bottom half of the population by density. County population and density are obtained from the 2010 Census.

- Table 8: Estimates are across offline versus online merchants within each listed NAICS category. For other e-commerce NAICS categories (Air Transportation, Ground Transportation, Rental and Leasing Services, Administrative and Support Services, Accommodation) the offline component was sufficiently limited that we

Table A1: Substitutability Robustness

	Baseline	Credit Bureau Sample	Longitude-Latitude
Aggregate	4.3	5.8	6.3
Building Material, Garden Supplies	7.7	7.8	8.8
Motor Vehicle and Parts Dealers	7.5	7.3	8.2
Furniture and Home Furnishings Stores	7.4	7.6	7.5
General Merchandise Stores	5.8	6.2	6.3
Health and Personal Care Stores	5.5	5.3	5.9
Clothing and Clothing Accessories Stores	5.2	6.1	5.8
Miscellaneous Store Retailers	5.2	5.7	5.9
Sporting Goods, Hobby, Music, Book Stores	4.2	4.5	4.4
Food and Beverage Stores	3.6	5.9	6.6
Electronics and Appliance Stores	3.4	3.9	3.8

used the overall offline-online estimate of $\sigma = 4.3$. [By comparison, we tend to estimate higher elasticities of substitution between competing offline merchants. We estimate a pooled elasticity of 5.01 across offline retail merchants within 3-digit categories. Interestingly, for some categories with little or no online option, we estimate lower elasticities (such as 2.92 across restaurants, which is itself a 3-digit non-retail NAICS).]

- Table 9: We compare the welfare gains under nested CES preferences to our single nest benchmark. Each nest is a 3-digit NAICS. We distribute purchases at nonstore retailers (NAICS 454) to the other nests using eMarketer estimates of the composition of nonstore retail spending. 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 σ_m. The outer nest Cobb-Douglas elasticities α_m are calibrated using spending shares. To calculate the sector specific online shares s_m, we take the sectoral online shares on the Visa network and multiply by the share of total consumption that happens on credit and debit cards (0.372). This is analogous to how we calculate overall online share, and assumes that spending on the Visa network is representative of all card spending, and that all online spending happens on credit and debit cards.

- Table 10: Changes in online share are 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 (conditional on ϕ). The corresponding formulae are given by $b_{2017}/b_{2007} = \left[(1-s_{2017})/(1-s_{2007})\right]^{\frac{\phi-1}{\phi}}$, $M_{b,2017}/M_{b,2007} = \left[(1-s_{2017})/(1-s_{2007})\right]^{\frac{1}{\phi}}$, $M_{b,market,2017}/M_{b,market,2017} = (1-s_{2017})/(1-s_{2007})$. The results are obtained by using our baseline estimate of $\phi = 1.74$. For this exercise, we use online shares computed from the set of retail NAICS categories only.

Appendix D. Model Derivations

1. The Consumer Problem: The first order conditions of the consumer problem are

$$p_o x_o = (\sigma - 1)\,\phi M_o^{\phi-1} F_o$$
$$p_b x_b = (\sigma - 1)\,\phi M_b^{\phi-1} F_b$$
$$\frac{p_o x_o}{p_b x_b} = \left(\frac{q_o/p_o}{q_b/p_b}\right)^{\sigma-1}$$

The first order conditions pin down the online share s of the optimal consumption bundle, namely

$$s = \frac{M_o \cdot p_o x_o}{M_o \cdot p_o x_o + M_b \cdot p_b x_b} = \frac{k}{k+1}$$

where $k = \left(\frac{q_o}{q_b}\frac{Z_o}{Z_b}\right)^{\frac{\phi}{\phi-1}(\sigma-1)}\left(\frac{F_b}{F_o}\right)^{\frac{1}{\phi-1}}$. Furthermore it can be shown that

$$M_o \cdot p_o x_o + M_b \cdot p_b x_b = \frac{(\sigma-1)\,\phi}{1+(\sigma-1)\,\phi} \times w$$

Using this, the relation between oM_o and bM_b and the identities $F_b = \frac{w}{A_b}$, $F_o = \frac{w}{A_o}$ we obtain the analytic solution to the consumer problem given in the main text.

The number of brick-and-mortar and online merchants visited, respectively, are given by:

$$\begin{aligned} M_b &= \left[\frac{1}{1+(\sigma-1)\phi} \cdot \frac{1}{1+k} \cdot A_b\right]^{\frac{1}{\phi}} \\ &= \left[\frac{1}{1+(\sigma-1)\phi} \cdot \frac{1}{1+k} \cdot \frac{w}{F_b}\right]^{\frac{1}{\phi}} \end{aligned}$$

$$\begin{aligned} M_o &= \left[\frac{1}{1+(\sigma-1)\phi} \cdot \frac{k}{1+k} \cdot A_o\right]^{\frac{1}{\phi}} \\ &= \left[\frac{1}{1+(\sigma-1)\phi} \cdot \frac{1}{1+k} \cdot \frac{w}{F_o}\right]^{\frac{1}{\phi}} \end{aligned}$$

Combining the expressions above for M_b and M_o yields the total number of merchants visited:

$$\begin{aligned} M &= M_o + M_b \\ &= \left[\frac{1}{1+(\sigma-1)\phi}\right]^{\frac{1}{\phi}} \left[\left(\frac{1}{1+k} \cdot \frac{1}{F_b}\right)^{\frac{1}{\phi}} + \left(\frac{k}{1+k} \cdot \frac{1}{F_o}\right)^{\frac{1}{\phi}}\right] w^{\frac{1}{\phi}} \end{aligned}$$

Combining this with the expression for $M_o \cdot p_o x_o + M_b \cdot p_b x_b$ above yields the estimating equation used in calibrating ϕ in the main text in equations (7) and (8).

$$\begin{aligned} M &= \left(\frac{1}{\sigma-1}\right)^{\frac{1}{\phi}} \times \left[\left(\frac{1}{1+k} \cdot \frac{1}{F_b}\right)^{\frac{1}{\phi}} + \left(\frac{k}{1+k} \cdot \frac{1}{F_o}\right)^{\frac{1}{\phi}}\right] \times [M_o p_o x_o + M_b p_b x_b]^{\frac{1}{\phi}} \\ \log M &= \frac{1}{\phi}\log\left(\frac{1}{\sigma-1}\right) + \log\left[\left(\frac{1}{1+k} \cdot \frac{1}{F_b}\right)^{\frac{1}{\phi}} + \left(\frac{k}{1+k} \cdot \frac{1}{F_o}\right)^{\frac{1}{\phi}}\right] + \frac{1}{\phi}\log[M_o p_o x_o + M_b p_b x_b] \end{aligned}$$

2. Supply side: The optimal price of any online/ offline firm is, respectively, equal to

$$\begin{aligned} p_o &= \frac{\sigma}{\sigma-1}\frac{w}{Z_o} \\ p_b &= \frac{\sigma}{\sigma-1}\frac{w}{Z_b} \end{aligned}$$

This, combined with the free entry condition, pins down production labor for offline and online firms to equal, respectively, $(\sigma-1)\,K_b$ and $(\sigma-1)\,K_o$. We then use the definition of shipping labor and the solution to the consumer problem to find

$$\begin{aligned} L_o &= \left(\frac{k}{k+1}\right)\left(\frac{1}{1+(\sigma-1)\,\phi}\right) L \\ L_b &= \left(\frac{1}{k+1}\right)\left(\frac{1}{1+(\sigma-1)\,\phi}\right) L \end{aligned}$$

Substituting the expression for production labor and shipping labor into the labor market clearing condition

$$M_{o,mkt}K_o + M_{b,mkt}K_b = \frac{1}{\sigma}\frac{(\sigma-1)\,\phi}{1+(\sigma-1)\,\phi}L$$

Lastly, combining the zero profit conditions for online and offline merchants yields

$$\frac{M_o/M_{o,mkt}}{M_b/M_{b,mkt}} = \frac{K_o}{K_b}\left(\frac{Z_o}{Z_b}\frac{q_o}{q_b}\right)^{(1-\sigma)}$$

Using the solution to the consumer problem then yields

$$\frac{M_{o,mkt}}{M_{b,mkt}} = k\frac{K_b}{K_o}$$

Combing this with the above expression of the labor market clearing conditions yields the analytic solution

$$M_{b,mkt} = \frac{1}{1+k}\frac{1}{\sigma}\frac{(\sigma-1)\,\phi}{1+(\sigma-1)\,\phi}\frac{L}{K_b}$$
$$M_{o,mkt} = \frac{k}{1+k}\frac{1}{\sigma}\frac{(\sigma-1)\,\phi}{1+(\sigma-1)\,\phi}\frac{L}{K_o}$$

3. Consumer surplus: Denote by s the estimated U.S. e-commerce share. It can be shown that welfare can be expressed as

$$W \propto A_b^{\frac{1}{\phi}\frac{1}{\sigma-1}} \times \left(\frac{1}{1-s_o}\right)^{\frac{\phi-1}{\phi}\frac{1}{\sigma-1}} \times \frac{w}{p}$$

Conditional on σ, ϕ and A_b, the consumption equivalent welfare gain Δ stemming from the rise in e-commerce can be obtained from

$$W(\frac{w}{p},\ s_{new}) = W\left(\Delta \times \frac{w}{p},\ s_{old}\right)$$
$$\left(\frac{1}{1-s_{new}}\right)^{\frac{\phi-1}{\phi}\frac{1}{\sigma-1}} \times \frac{w}{p} = \left(\frac{1}{1-s_{old}}\right)^{\frac{\phi-1}{\phi}\frac{1}{\sigma-1}} \times \Delta \times \frac{w}{p}$$
$$\left(\frac{1}{1-s_{new}}\right)^{\frac{\phi-1}{\phi}\frac{1}{\sigma-1}} = \left(\frac{1}{1-s_{old}}\right)^{\frac{\phi-1}{\phi}\frac{1}{\sigma-1}} \times \Delta$$
$$\Delta = \left(\frac{1-s_{old}}{1-s_{new}}\right)^{\frac{\phi-1}{\phi}\frac{1}{\sigma-1}}$$

Substituting in the discussed values for s, ϕ, σ will hence deliver the result. The welfare calculations by card income/ county density group are done analogously.[3]

4. Consumer surplus in nested CES case: The welfare gains in the nested CES case can be expressed as

$$\Delta = \left(\prod_m \left(\frac{1-s_{m,old}}{1-s_{m,new}}\right)^{\frac{\alpha_m}{\sigma_m-1}}\right)^{\frac{\phi-1}{\phi}}$$

where m denotes the nests, s_m the online share within nest, α_m the outer nest elasticity (Cobb-Douglas) and σ_m the nest specific elasticity. The α_m are calibrated using spending shares, the σ_m estimated by industry (analogously to the baseline σ) and the s_m observed in the Visa data.

5. Producer surplus/ Retail Apocalypse: Here we examine the impact of changing q_o/q_b, A_o/A_b on b, M_b, $M_{b,mkt}$ through the lens of the model. It can be shown that the online share s is a sufficient statistic for all three counterfactuals and that the predicted changes can be expressed as follows:

$$\frac{b_{2017}}{b_{2007}} = \left[\frac{1-s_{2017}}{1-s_{2007}}\right]^{\frac{\phi-1}{\phi}}$$

[3]The underlying Visa e-commerce shares for the different card groups are adjusted for the card-less as described above before substituting into the above formula.

$$\frac{M_{b,2017}}{M_{b,2007}} = \left[\frac{1 - s_{2017}}{1 - s_{2007}}\right]^{\frac{1}{\phi}}$$

$$\frac{M_{b,market,2017}}{M_{b,market,2017}} = \frac{1 - s_{2017}}{1 - s_{2007}}$$

As we are describing the Retail Apocalypse we will use an estimate of the online share in the retail industries only, rather than in all of the U.S. economy. The online share in U.S. Retail is calculated analogously to the overall U.S. online share. In particular, we calculate, by year, total online revenues for online merchants in the retail NAICS. We then divide retail revenues on the Visa network by Visa's share of total card spending to obtain an estimate of total online spending at retail merchants in that year. In the final step we will divide this estimate by the BEA's Retail Trade Gross Output estimate to obtain an estimate of the online share in U.S. retail. The resulting estimates are 6.0% in 2007 and 9.5% in 2017.