Assessing the Gains from E-Commerce

Paul Dolfen, Stanford
Liran Einav, Stanford and NBER
Pete Klenow, Stanford and NBER
Ben Klopack, Stanford
Jonathan Levin, Stanford and NBER
Larry Levin, Visa
Wayne Best, Visa

October 2018

Berkeley+Stanford Trade Lunch Talk

Motivation

- E-commerce is a rapidly growing share of U.S. retail spending
  - reached 10% in 2017
- How big is the consumer surplus from e-commerce?
- How are the gains distributed by geography and affluence?

What we do and find

- Document the rise of e-commerce using Visa data
- Estimate the *pure convenience* gains from shopping online
  - 0.4% of consumption
- Allow for *variety* gains from e-commerce
  - Consumer surplus from e-commerce 1% of consumption
- Gains are increasing in county population density
- Gains are half as big for incomes below $50k

Related literature

Gains from e-commerce and the internet

- Brynjolffson and collaborators (2003, 2012, 2017)
- Goolsbee and Klenow (2006, 2018)
- Varian (2013)
- Syverson (2016)
- Couture, Faber, Gu and Liu (2018)

Consumer surplus from new products more generally

- Feenstra (1994)
- Hausman (1997, 1999)
- Weinstein and collaborators (2006, 2010, 2018)

Outline

1. **Visa data and basic facts**
2. Estimating the *pure convenience* gains from shopping online
3. Estimating the *variety* gains from e-commerce

Visa data

Raw data is similar to line items in monthly statements:

- Transaction amount and day
- Unique card identifiers (credit and debit)
- Store name, NAICS, ZIP (longitude-latitude in recent years)
- January 2007 through December 2017

Merged with *Experian* data the last few years:

- Card income
- Card location

Sample of raw data

Someone's profligate card activity:

10/06	1 800 CONTACTS 800-266-8228 UT	559.92
10/06	TAMARINE RESTAURANT PALO ALTO CA	147.00
10/07	ISEE PROGRAM 212-672-9800 NY	135.00
10/06	AMAZON MKTPLACE PMTS AMZN.COM/BILL WA	14.00
10/06	AMAZON MKTPLACE PMTS AMZN.COM/BILL WA	6.25
10/06	MOLLIE ST PALO ALTO PALO ALTO CA	66.07
10/06	BON APPETIT 51157154 STANFORD CA	9.09
10/07	AMAZON MKTPLACE PMTS AMZN.COM/BILL WA	83.97
10/08	FEDEXOFFICE 00051011 PALO ALTO CA	344.52
10/08	ORCHARD SUPPLY #690 MOUNTAIN VIEW CA	19.64
10/10	LITTLE DAVID PEST CTRL 408-294-1955 CA	255.00
10/09	HOSTNINE 404-627-7789 NC	3.95
10/11	AMIGOS GRILL PORTOLA VALLE CA	67.59
10/11	TRADER JOE'S #207 QPS PALO ALTO CA	133.61
10/09	BON APPETIT 51157154 STANFORD CA	4.19
10/13	ISING SILICON VALLEY ISINGSV.COM CA	400.00

Visa data confidentiality

All results have been reviewed to ensure that no confidential information about Visa merchants or cardholders is disclosed.

Cards are anonymized, and we report no data on individual cards. Cardholder information is based solely on the card's transactions.

We report no data on specific merchants or from recent months – which is why the analysis sample ends in December 2017.

Visa data caveats

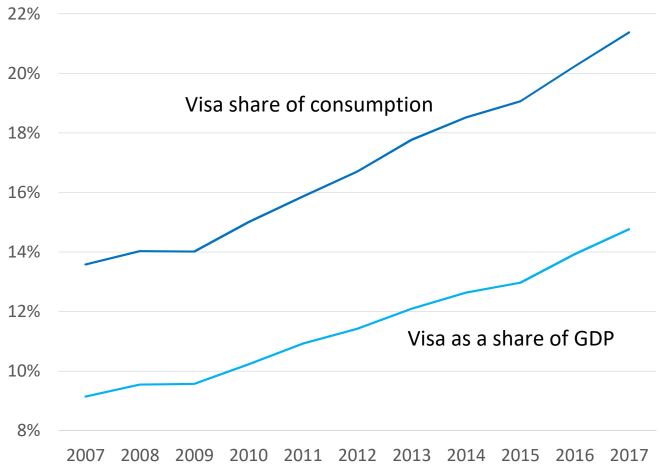
- No details on items bought or prices
- Cannot tie multiple cards to households
- Tremendous card turnover
- Will rely heavily on monetized distance to get at WTP

Visa summary statistics

U.S. annual averages from 2007 through 2017

- 380 million cards
- 35.9 billion transactions
- $1.93 trillion in sales
  - 55% credit, 45% debit

Flowing through Visa



Sources: Visa and BEA

Experian data

- Consumer credit reporting agency
- Merged with Visa cards (only in recent years)
- Match for roughly 50% of Visa credit cards in 2016
- Cardholder demographics (e.g. income and education)

E-commerce in the Visa data

Visa transaction flags:

- CP Card Present (brick-and-mortar)
- CNP Card Not Present
  - phone or mail order
  - recurring bill payments
  - ECI e-commerce indicator
  - missing values

For missing values we allocate within 3-digit NAICS years:

$$\text{e-commerce} = \frac{\text{ECI}}{\text{ECI} + \text{phone/mail/recurring}} \quad \text{CNP}$$

E-Commerce industries

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

Non-Retail	Example
Admin, Support Services	Expedia Travel
Air Transportation	American Airlines
Accommodation	Marriott
Car Parts	AutoZone
Rental Services	Hertz Rent-A-Car

Online Visa spending shares (in %), selected NAICS

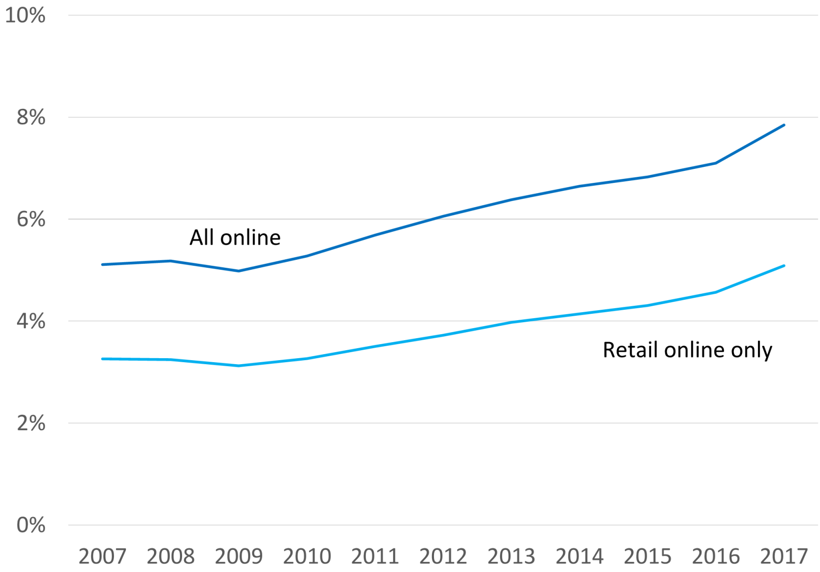
	2007	2017
Nonstore Retailers	90	96
Air Transport	87	97
Electronics	42	51
Furniture	35	43
Clothing	22	37
General Merchandise	8	15
Food	5	6

Estimating e-commerce in the U.S. overall

$$\text{U.S. Online Share} = \frac{\text{Total Card Spending}}{\text{Consumption}} \quad \text{Visa Online Share}$$

- Calculate e-commerce share in Visa as described above
- Assume Visa representative of all card transactions
- Assume non-card transactions are all offline

Share of U.S. consumption online



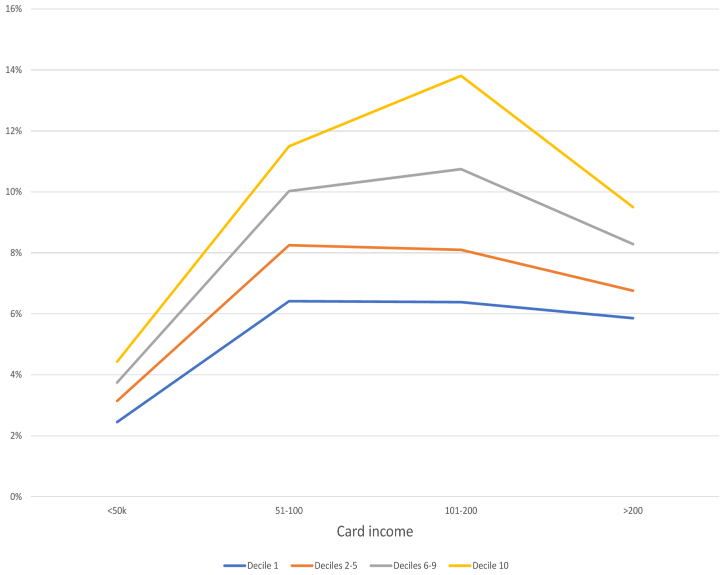
Online shares for county-income groups

To adjust for the card-less:

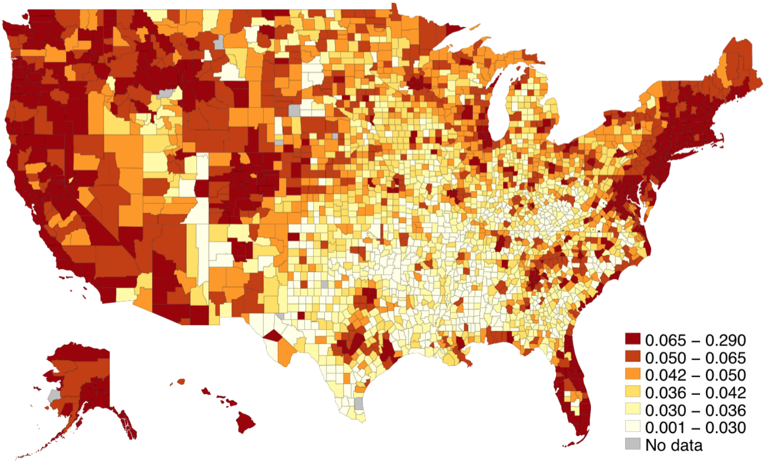
$$\text{OnlineShare}_{cy} = \frac{\text{VisaOnline}_{cy}}{\text{TotalVisa}_{cy}} \quad \alpha_{cy}$$

- county c, income group y
- α_{cy} is the share of Visa in total spending
- assumes online spending only through credit+debit cards
- $\widehat{\alpha}_{cy} = \frac{\text{VisaCards}_{cy}}{\text{Pop}_{cy}}$
  - VisaCards_{cy} based on Experian subsample with y info
  - Pop_{cy} based on IRS tax return filers and dependents

Adjusted online share by card income in 2016



Adjusted online share by county in 2016



Outline

1. Visa data and basic facts
2. **Estimating the *pure convenience* gains from shopping online**
3. Estimating the *variety* gains from e-commerce

Estimates of convenience surplus

- Assume basket is fixed and, for a given merchant, identical prices online and offline (see Cavallo, 2017)
- Focus on the binary decision when both options are available:

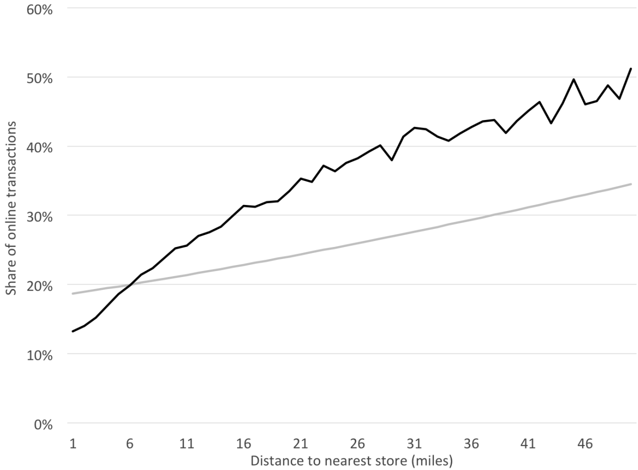
$$u_{ij}^o = \gamma_j^o + \epsilon_{ij}^o$$

$$u_{ij}^b = \gamma_j^b \quad \beta \;\; distance_{ij} + \epsilon_{ij}^b$$

i is the individual, j is the merchant, and the distance is between the individual i and the nearest store of merchant j

If iid Type I extreme value distribution of idiosyncratic preferences, logit online/offline indicator on distance and merchant fixed effects

Pr(shop online) vs. distance to merchant store



- An observation is an individual transaction
- Sample = transactions in 5 “mixed online/offline” retail categories (1% sample of cards) in 2017

Converting distance into WTP (willingness to pay)

- A straight-line mile requires 1.5 miles of driving on average (Einav et al, 2016)
- 1.4 minutes per mile of driving on average (Einav at al, 2016)
- 2017–2017 average hourly after-tax wage of $23 per hour (BLS)
- 2007–2017 average fuel + depreciation per mile = $0.535 (IRS)
- Each mile counts as two miles of round trip travel
- Each mile costs $0.79 in direct costs and $0.80 in time costs, for a total of $3.18 per mile (roundtrip)

Consumer surplus from convenience

- Results imply consumer gain of 11.3 mile-equivalents per transaction
- Convenience gains per online transaction = \$36 (11.3 \$3.18)
- Typical transaction is \$96 and 10.3 miles away, so convenience value is 28% (= \$36/\$129)
- In 2017, share of all Visa spending at "mixed" merchants with distance < 50 miles in our 5 "mixed" NAICSs was 7%, so overall convenience gains are 2% (28% of the 7%) of Visa spending and 0.4% of consumption

Interpreting the pure convenience gains

- Convenience gains are driven by consumer substitution from offline to online within the *same* merchant
- However, 88% of spending online is on merchants at which the same card never transacts offline
- Much of the gains may be due to accessing new merchants rather substitution within merchants' offline/online arms
- To quantify this, we next write down a stylized GE model with offline and online merchants, and calibrate it using the Visa data

Outline

1. Visa data and basic facts
2. Estimating the *pure convenience* gains from shopping online
3. **Estimating the *variety* gains from e-commerce**

Consumer problem

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

subject to

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

- q_m = "quality" of merchant m
- x_m = quantity purchased from merchant m
- p_m = price per unit at merchant m
- $M = M_b + M_o$ = total merchants bought from
- M_b (M_o) = # of merchants shopped at in-store (online)
- F_b (F_o) = scale of fixed costs for shopping in-store (online)

Comments on the consumer problem

- Merchants are either online or offline (not both)
  - Broadly consistent with low merchant overlap within cards
- $\sigma > 1$ is the elasticity of substitution across *merchants*
  - $\sigma <$ 1) "love of variety"
- ϕ governs how fast fixed shopping costs rise with the # of online and brick-and-mortar merchants shopped at
  - $\phi > 1$ so we get an interior solution despite love of variety

Producer problem

$$\max_{p_m} \; \pi_m = p_m y_m \quad wL_m \quad wK_j$$

subject to

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

- $j = o$ or b
- $M_j \quad M_{j,market}$
- Brick-and-mortar (online) sellers split their market evenly
- K_j = overhead labor needed to operate

Free entry and market clearing

For each market j:

$$E_j[\pi_m] = 0$$

Labor market clearing:

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

Shopping technology

$$L \quad M_b^{\phi} = Y_b = A_b L_b$$

$$L \quad M_o^{\phi} = Y_o = A_o L_o$$

Perfectly competitive so marginal cost pricing:

$$F_b = \frac{w}{A_b}$$

$$F_o = \frac{w}{A_o}$$

Symmetric technologies

Process efficiency:

$$Z_m = Z$$

Quality offline:

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

Quality online:

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

Symmetric outcomes

Pricing:

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

Spending per merchant online (o) and offline (b):

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

Profits:

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

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

Merchants in GE

$$\text{Define } k \quad \left(\frac{q_o}{q_b}\right)^{\frac{\phi}{\phi \quad 1}(\sigma \quad 1)} \left(\frac{A_o}{A_b}\right)^{\frac{1}{\phi \quad 1}}$$

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

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

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

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

GE comparative statics

	$\frac{M_{o,market}}{M_{b,market}}$	$\frac{M_o}{M_b}$	$\frac{o}{b}$
$\frac{A_o}{A_b}$	+	+	0
$\frac{q_o}{q_b}$	+	+	+

Online spending share

Let s_o denote the share of card spending online:

$$s_o \quad \frac{oM_o}{oM_o + bM_b} = \frac{k}{k + 1}$$

where $k \quad \left(\frac{q_o}{q_b}\right)^{\frac{\phi}{\phi \quad 1}(\sigma \quad 1)} \left(\frac{A_o}{A_b}\right)^{\frac{1}{\phi \quad 1}}$

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

Quantitative strategy

Calibrate:

- ϕ = convexity of fixed shopping costs
- σ = elasticity of substitution across merchants

Then back out the combination of q_o/q_b and A_o/A_b from s_o

Estimating ϕ (convexity of fixed shopping costs)

According to the model, we can estimate ϕ using one of two regressions that yield the same answer by construction:

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

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

Extensive and intensive margin Engel Curve slopes should reflect ϕ

Caveat: This assumes no idiosyncratic fixed costs or online/offline preferences correlated with a card's total expenditures

Estimates of ϕ (convexity of fixed shopping costs)

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

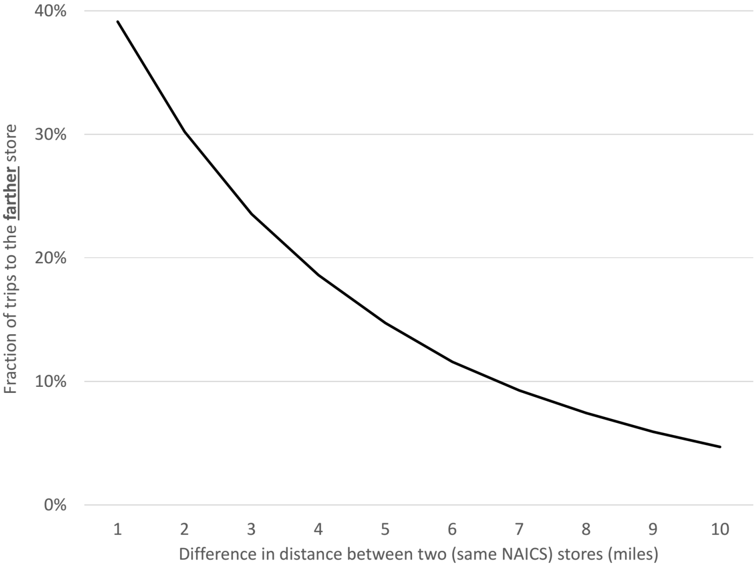
Standard errors are tiny ...

Estimating σ (elasticity of substitution across merchants)

- We use a 1% sample of cards that transacted in 2017
- For each card i, we look at online purchases and offline purchases made within 20 miles of i's location

1. We construct, for each individual i and NAICS, all pairs of physical stores j and k such that i buys in one of these stores
2. We also construct all pairs of physical store j and online merchant k such that i buys from one of these

Relative trips vs. distance



Converting distance into WTP (willingness to pay)

- A straight-line mile requires 1.5 miles of driving on average (Einav et al, 2016)
- 1.4 minutes per mile of driving on average (Einav at al, 2016)
- 2017–2017 average hourly wage = $23 per hour (BLS)
- 2007–2017 average fuel + depreciation per mile = $0.535 (IRS)
- Each mile counts as two miles of round trip travel
- Each mile costs $0.80 in direct costs and $0.79 in time costs, for a total of $3.18 per roundtrip mile

Estimating σ (continued)

- Assuming distance is uncorrelated with preferences (controlling for merchant fixed effects), we can use how visits change with distance to estimate σ

- Aggregating to the merchant j, merchant k, $dist_{ij}$, $dist_{ik}$ level:

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

- p_{jk} = average ticket size at merchants j, k
- τ = transportation costs for i to j or k
- We capture relative quality with cross fixed effects

- Regress on both online-offline and offline-offline samples

Estimates of σ

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

Standard errors are tiny (on the order of 0.001)

Welfare

Consumption-equivalent welfare is proportional to

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

where average quality is

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

In terms of exogenous driving forces, welfare is proportional to

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

Consumption-equivalent welfare gains from e-commerce

Due to rising q_o/q_b and A_o/A_b:

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

Welfare gains by card income in 2017

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

Welfare gains by county density in 2017

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

Quartiles based on population (25% in each quartile).

Conclusions

1. *Convenience* gains 0.4% of consumer spending
2. Allowing for *variety* gains, consumer surplus from online spending 1% of consumption
3. Consumer surplus from e-commerce is:
   - smallest for incomes below $50k (less likely to have cards)
   - bigger in more densely populated counties

Retail Apocalypse

Due to changing q_o/q_b and A_o/A_b (holding A and L fixed):

	2007–2017 Change
b	–1.6%
M_b	–2.1%
$M_{b,market}$	–3.7%
Profits	0%

Our projects underway with Visa data

1. Assessing the Gains from E-Commerce
2. Customers are (Almost) Everything
3. Store entry: Too Little, Too Much, or Just Right?
4. A Heat Map of High-Frequency ZIP Code Spend

Cards account for most sales variation

Independent variable is ln(Sales)

	Dependent Variable		
Fixed Effects	ln(Cards)	ln(Trans/Card)	ln(Sales/Trans)
Year	.653	.038	.309
Year, NAICS	.846	.029	.125
Year, Merchant	.916	.035	.049
Year, Merchant-Zip	.859	.046	.095