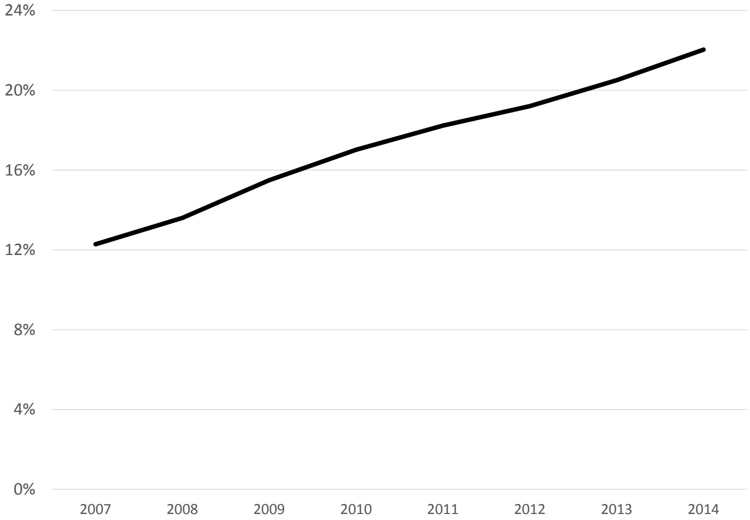
Assessing the Gains from E-Commerce

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

July 20, 2017

NBER IT and Digitization meeting, Cambridge, MA

Share of Visa spending online



Source: Visa credit and debit card data

Motivation

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

What we do and (*preliminarily*) find

- Document the rise of e-commerce in Visa data
- Estimate the *pure convenience* gains from shopping online
  - $\approx 1.0\%$ of Visa spending
- Allow for *variety* gains from e-commerce
  - Consumer surplus from online spending $\approx 3.6\%$ of Visa spending
- How do gains differ by county density and affluence?
  - Higher in richer and denser counties

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 similar to line items in monthly statements:

- Transaction amount and day
- Unique card identifiers (credit and debit)
- Store name, NAICS, ZIP (long-lat down the road)
- January 2007 through December 2014

Two data constructs:

- Card location
- Card affluence

Note:

- No details on items bought or prices
- Will rely heavily on (monetized) distance to get at WTP

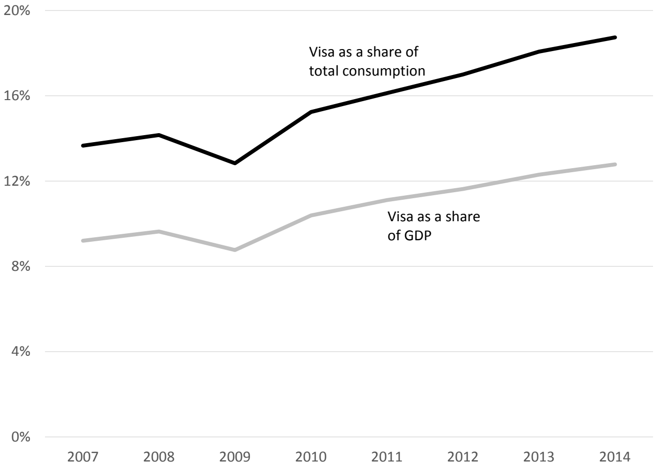
Visa data confidentiality

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, and observe a 2-year blackout – which is why the analysis sample ends in December 2014.

A lot is flowing through Visa



Sources: Visa, BEA, and Census Bureau

Annual averages: 373M cards, 31.9B transactions, $1.7T in sales

Online vs. offline Visa spending

Visa transaction flags

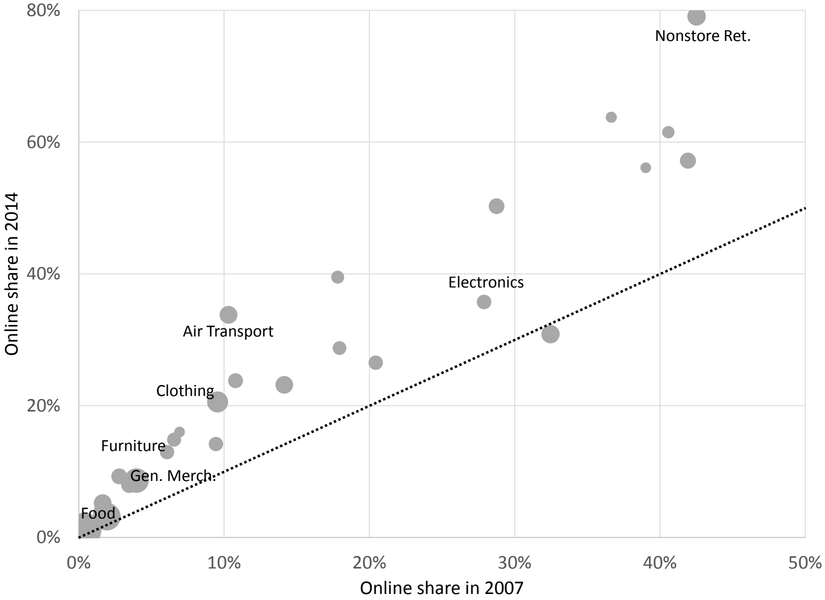
- CP $\equiv$ Card Present
- CNP $\equiv$ Card Not Present

We define offline to include:

- CP (brick-and-mortar)
- CNP phone + mail + recurring

Visa defines the rest of the CNP categories as *e-commerce*

Online spending shares

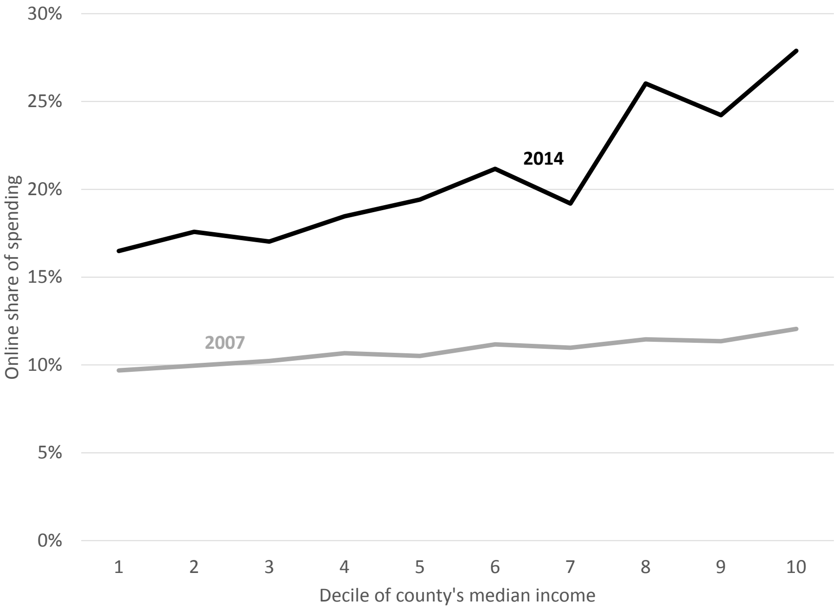


Decomposing the growth of online spending

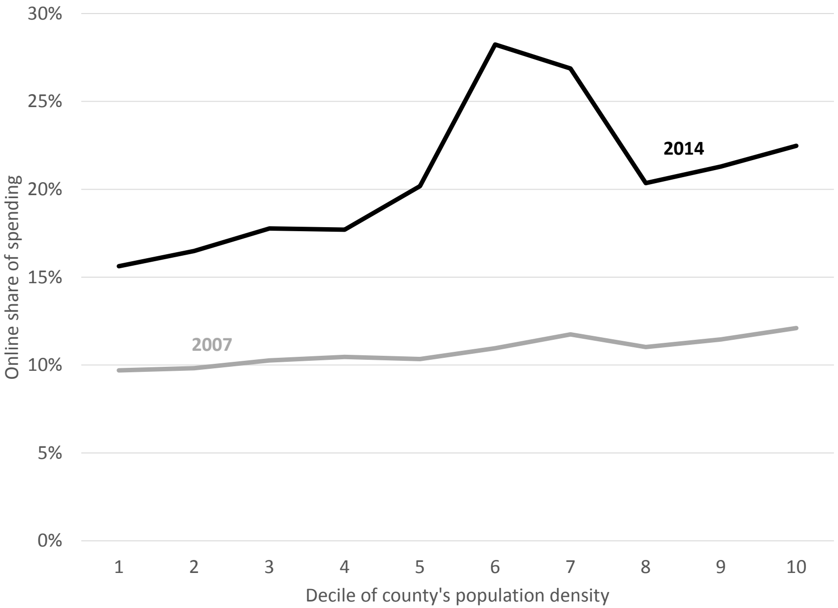
	2007	2014	$\frac{2014}{2007}$	Share*
Online card share	0.61	0.74	1.22	31%
Online spending per merchant ($)	214	264	1.24	35%
Online merchants per card	4.3	5.5	1.27	34%
Offline spending per merchant ($)	174	168	0.97	
Offline merchants per card	23	23	1.00	

* Contribution to growth in (spending online)/(spending offline)

Online share by county income



Online share by county density



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

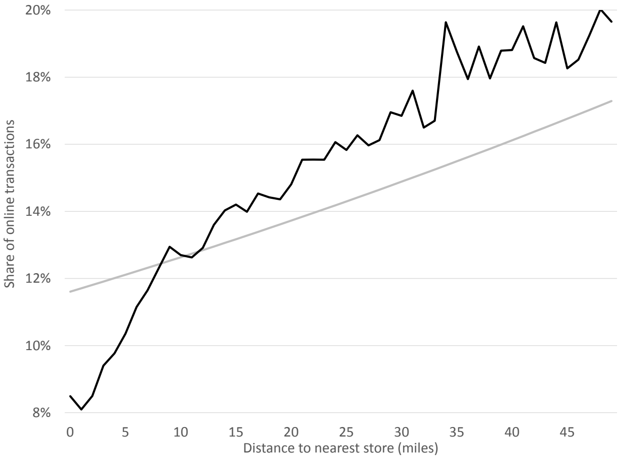
- Assume basket is fixed and identical prices online and offline
- Focus on the binary decision when both options are available

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

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

Gives rise to a simple logit specification of online/offline indicator on distance and merchant fixed effects

Pr(shop online) vs. distance to merchant store



- An observation is an individual transaction
- Sample is transactions in 5 “mixed online/offline” retail categories (by random 1% sample of cards) in 2014

Consumer surplus from convenience

- Results imply consumer gain of 16.6 “mile equivalents” per (any) transaction
- But 4.2 “directly” off the data; rest is driven by the (logit) tail
- Convert 4.2 miles to \$15.5 (see paper for details), which implies convenience gain of $\approx 14\%$ for the average transaction
- In 2014, share of all Visa spending at “mixed” merchants with distance < 50 miles in our 5 “mixed” NAICSs was 7%, so overall convenience gains are $\approx 1.0\%$ (7% of 14%)

Interpreting the pure convenience gains

- Convenience gains are driven by substitution from offline to online for the *same* merchant
- However, 88% of spending online is at merchants at which the same card has never transacted offline
- So much of the gains are captured by cross-merchant substitution rather than within merchants
- To quantify this, we next write down a stylized model which allows such substitution, 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

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

subject to

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

- b_m (o_m) = brick-and-mortar (online) spending on merchant m
- q = relative "quality" of the online channel
- 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

- The implicit assumption is that all merchants charge the same price (across merchants and online vs. offline)
  - as a result, optimal spending is the same across different offline (online) merchants visited: $o_m = o$ and $b_m = b$
- Offline and online are perfect substitutes, so a consumer spends online or offline (or neither) for a given merchant
  - Broadly consistent with low within-card merchant overlap
- $\sigma > 1$ is the elasticity of substitution across *merchants*
  - $\sigma < \infty \Rightarrow$ "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

Online spending share and gains from e-commerce

Let s_o denote the share of card spending online:

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

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

- s_o rises with q and F_b/F_o; trends in o, b, M_o, and M_b tell us how much s_o has risen because of rising q vs. falling F_o (for given F_b)

At the optimum, for given expenditures y and fixed costs for offline stores F_b, welfare rises with the share of spending online s_o

Consumers gain from rising s_o due to online options getting better (rising q) and easier access to online merchants (falling F_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:

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

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

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)

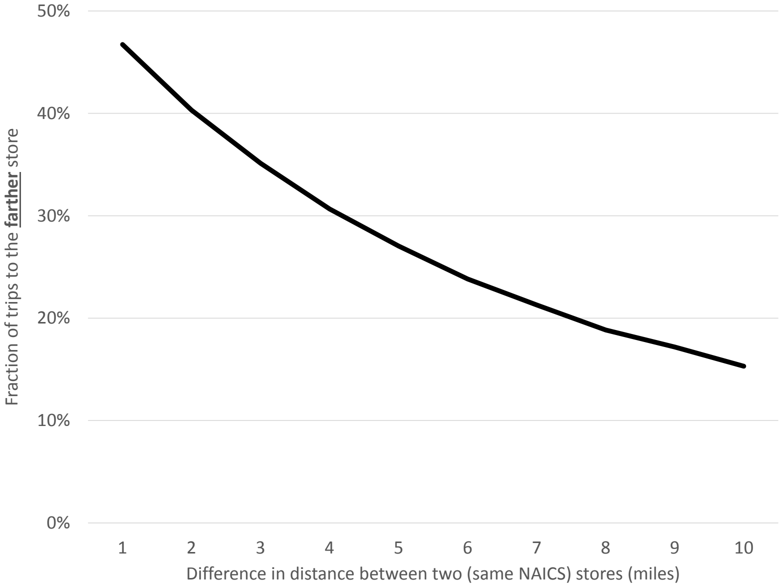
	2007	2014
$\widehat{\phi}$	1.78	1.82
# of cards	287M	453M
R^2	0.66	0.67

Standard errors are tiny ...

Estimating σ (elasticity of substitution across merchants)

- We use a 1% sample of cards that transacted in 2014
- For each card i, we look at offline purchases made within 10 miles of i's location
- construct all pairs of stores j and k for each NAICS and, for each individual i who buys in one of these stores, compute $|dist_{ij} - dist_{ik}|$
- calculate the share of combined trips to each pair made to the farther store at each absolute distance

Relative trips vs. distance



Estimating σ (continued)

- Assuming distance is uncorrelated with preferences (controlling for chain fixed effects), we can use how visits change with distance to estimate σ
- Moving from 6 to 16 miles away implies an effective price change of \$37 (round trip); given an average ticket size of \$42, this is a price increase of 57%
- The trip *share* drops by about 68%, implying a σ of approximately 3.6

Consumption-equivalent welfare gains from e-commerce

	ϕ	σ	s_o^{2014} vs. s_o^{2007}	s_o^{2014} vs. $s_o = 0$
Baseline	1.8	3.6	1.7%	**3.6%**
Sensitivity checks:				
High ϕ	2	3.6	1.9%	4.1%
High σ	1.8	4	1.5%	3.2%

Welfare gains by county affluence

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

Note: Quartiles based on cards; 25% of cards in each quartile

Welfare gains by county density

	s_o^{2014} vs. s_o^{2007}	s_o^{2014} vs. $s_o = 0$
Quartile 1 (sparse)	1.1%	2.6%
Quartile 2	1.4%	3.0%
Quartile 3	2.6%	4.3%
Quartile 4 (dense)	1.7%	3.5%

Note: Quartiles based on cards; 25% of cards in each quartile

Conclusions

1. *Convenience* gains $\approx$ 1.0% of Visa spending
2. Allowing for *variety* gains, consumer surplus from online spending $\approx$ 3.6% of Visa spending
   - Roughly 1.3% of all consumer spending and 0.9% of GDP
3. Consumer surplus from e-commerce is:
   - Larger for consumers in richer counties
   - Larger for consumers in more dense counties