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

June 2019

Technology Diffusion and Productivity Workshop
Federal Reserve Bank of Richmond
What we do

- Document the rise of e-commerce using Visa data

- Estimate resulting consumer surplus > 1% of consumption

- Find gains are increasing in county population density

- Find gains are twice as big for incomes above $50k
Gains from e-commerce and the internet

- Syverson (2016)
- Couture, Faber, Gu and Liu (2018)
- Allcott, Braghieri, Eichmeyer and Gentzkow (2019)

Consumer surplus from new products more generally

- Feenstra (1994)
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
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

Visa share of consumption

Visa as a share of GDP

Sources: Visa and BEA
Experian data

- Consumer credit reporting agency

- Merged with Visa cards (only in recent years)

- Can match roughly 50% of Visa credit cards 2016–2017

- Cardholder demographics (e.g. income and education)
E-commerce in the Visa data

Visa transaction flags:

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

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

$$
e-commerce = \frac{ECI}{ECI + \text{phone/mail/recurring}} \times \text{CNP}
$$
## E-Commerce industries

<table>
<thead>
<tr>
<th>Retail</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonstore Retail</td>
<td>Amazon</td>
</tr>
<tr>
<td>Clothing</td>
<td>Nordstrom</td>
</tr>
<tr>
<td>Misc Retail</td>
<td>Staples</td>
</tr>
<tr>
<td>General Merchandise</td>
<td>Walmart</td>
</tr>
<tr>
<td>Electronics</td>
<td>Best Buy</td>
</tr>
<tr>
<td>Building Material, Garden Supplies</td>
<td>Home Depot</td>
</tr>
<tr>
<td>Furniture</td>
<td>Bed Bath &amp; Beyond</td>
</tr>
<tr>
<td>Sporting Goods, Hobby</td>
<td>Nike</td>
</tr>
<tr>
<td>Health, Personal Care</td>
<td>CVS</td>
</tr>
<tr>
<td>Food</td>
<td>Safeway</td>
</tr>
<tr>
<td>Ground Transportation</td>
<td>Uber</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-Retail</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Admin, Support Services</td>
<td>Expedia Travel</td>
</tr>
<tr>
<td>Air Transportation</td>
<td>American Airlines</td>
</tr>
<tr>
<td>Accommodation</td>
<td>Marriott</td>
</tr>
<tr>
<td>Car Parts</td>
<td>AutoZone</td>
</tr>
<tr>
<td>Rental Services</td>
<td>Hertz Rent-A-Car</td>
</tr>
</tbody>
</table>
Share of visa spending online, select industries

- Nonstore retail (mostly internet-first retailers)
- Air transport
- Electronics
- Furniture
- Clothing
- General merchandise
- Food

2007 vs 2017
U.S. Online Share $= \frac{\text{Total Card Spending}}{\text{Consumption}} \cdot \text{Visa Online Share}$

- 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

- All online
- Retail online only

Fraction of households with cards:

\[
\hat{\alpha}_{cy} \propto \frac{\# \text{ of Visa Cards}_{cy}}{\text{Tax Filers}_{cy}}
\]

Fraction of all consumption on e-commerce for each county-income:

\[
\hat{s}_{cy} \propto \frac{\text{Visa online spending}_{cy}}{\text{Total Visa spending}_{cy}} \cdot \hat{\alpha}_{cy}
\]
Online share of all consumer spending:

- Below-median density counties: 6.4%
- Above-median density counties: 9.1%
- Cardholder income $\leq$ $50k$: 3.4%
- Cardholder income $> $50k: 9.7%
1. Visa data and basic facts

2. Estimating the *pure convenience* gains from shopping online

3. Estimating the *variety* gains from e-commerce
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

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

subject to

\[ M^\phi F_b + M^\phi F_o + \sum_{m=1}^{M} p_m \cdot x_m \leq w \]

- \( q_m \) = “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
  - $\sigma < \infty \implies \text{“love of variety”}$

- $\phi$ 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

\[
\max_{p_m} \pi_m = p_m y_m - wL_m - wK_j
\]

subject to

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

- \( j = o \) or \( b \)
- \( M_j \leq M_{j,market} \)
- 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:

$$L = \sum_{m} L_m + L_b + L_o + M_{b,\text{market}} K_b + M_{o,\text{market}} K_o$$
Shopping technology

\[ L \cdot M_b^\phi = Y_b = A_b L_b \]

\[ L \cdot M_o^\phi = Y_o = A_o L_o \]

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:

\[ q_m = q_b \quad \text{for} \quad m \in M_{b, market} \]

Quality online:

\[ q_m = q_o \quad \text{for} \quad m \in M_{o, market} \]
Symmetric outcomes

Pricing:

\[ p_m = p = \frac{\sigma}{\sigma - 1} \cdot \frac{w}{Z} \]

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

\[ \frac{o}{b} = \left( \frac{q_o}{q_b} \right)^{\sigma - 1} \]

Profits:

\[ \pi_o = \frac{M_o}{M_{o,\text{market}}} L \cdot \frac{o}{\sigma} - wK_o \]

\[ \pi_b = \frac{M_b}{M_{b,\text{market}}} L \cdot \frac{b}{\sigma} - wK_b \]
Define \( k \equiv \left( \frac{q_o}{q_b} \right)^{\frac{\phi}{\phi-1}} \left( \frac{A_o}{A_b} \right)^{\frac{1}{\phi-1}} \)

\[
M_{b,\text{market}} = \frac{1}{1+k} \cdot \frac{1}{\sigma} \cdot \frac{(\sigma - 1)\phi}{1 + (\sigma - 1)\phi} \cdot \frac{L}{K_b}
\]

\[
M_{o,\text{market}} = \frac{k}{1+k} \cdot \frac{1}{\sigma} \cdot \frac{(\sigma - 1)\phi}{1 + (\sigma - 1)\phi} \cdot \frac{L}{K_o}
\]

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

\[
M_o = \left[ \frac{1}{1 + (\sigma - 1)\phi} \cdot \frac{k}{1+k} \cdot A_o \right]^{\frac{1}{\phi}}
\]
GE comparative statics

\[
\begin{array}{ccc}
\frac{M_{o,market}}{M_{b,market}} & \frac{M_o}{M_b} & \frac{o}{b} \\
\hline
\frac{A_o}{A_b} & + & + & 0 \\
\frac{q_o}{q_b} & + & + & + \\
\end{array}
\]
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 $k \equiv \left( \frac{q_o}{q_b} \right)^{\frac{\phi}{\phi - 1}}(\sigma - 1) \left( \frac{A_o}{A_b} \right)^{\frac{1}{\phi - 1}}$

- $s_o$ 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 better (rising $q_o$) and easier to access (rising $A_o$) online options
Consumption-equivalent welfare is proportional to

\[ Z \cdot M^{1/(\sigma - 1)} \cdot \bar{q} \]

where average quality is

\[ \bar{q} \equiv \left[ \frac{q_b^{\sigma-1} \cdot M_b + q_o^{\sigma-1} \cdot M_o}{M} \right]^{1/(\sigma-1)} \]
In terms of exogenous variables, welfare is proportional to

\[ Z \cdot \left( q_b \phi_{\phi-1}^{\phi-1}(\sigma-1) A_b^{\frac{1}{\phi-1}} + q_o \phi_{\phi-1}^{\phi-1}(\sigma-1) A_o^{\frac{1}{\phi-1}} \right)^{\frac{\phi-1}{\phi} \frac{1}{\sigma-1}} \]

For given \( Z, q_b, \) and \( A_b, \) welfare is increasing in \( s_o : \)

\[ Z \cdot q_b \cdot A_b^{\phi(\sigma-1)} \left( \frac{1}{1 - s_o} \right)^{\frac{\phi-1}{\phi(\sigma-1)}} \]
Calibrate:

- $\phi = \text{convexity of fixed shopping costs}$
- $\sigma = \text{elasticity of substitution across merchants}$

Then infer the welfare gain from the path of $s_o$
Estimating $\phi$ (convexity of fixed shopping costs)

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

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

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

Extensive and intensive margin Engel Curve slopes should reflect $\phi$

**Caveat**: This assumes idiosyncratic fixed costs are uncorrelated with a card’s total expenditures
Estimates of $\phi$ (convexity of fixed shopping costs)

<table>
<thead>
<tr>
<th></th>
<th>2007</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\phi}$</td>
<td>1.73</td>
<td>1.75</td>
</tr>
<tr>
<td># of cards</td>
<td>283M</td>
<td>462M</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.67</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Standard errors are tiny ...
Estimating $\sigma$

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

- Aggregating merchant pairs $\{j, k\}$ with the same $\{dist_{ij}, dist_{ik}\}$:

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

- $p_{jk} = \text{average ticket size at merchants } j, k$
- $\tau = \text{transportation costs for } i \text{ to } j \text{ or } k$
- $\tau = 0 \text{ for online transactions}$
- Capture relative quality with cross fixed effects
- Regress on both online-offline and offline-offline samples
Transactions online vs. distance to a physical store

% of transactions online

Distance to physical store (in miles)
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 et 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
Estimates of $\sigma$

<table>
<thead>
<tr>
<th></th>
<th>online-offline</th>
<th>offline-offline</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\sigma}$</td>
<td>4.3</td>
<td>6.1</td>
</tr>
<tr>
<td># of obs</td>
<td>3.6M</td>
<td>14.0M</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.97</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Standard errors are tiny (on the order of 0.001)
### Consumption-equivalent gains from e-commerce

<table>
<thead>
<tr>
<th></th>
<th>( \phi )</th>
<th>( \sigma )</th>
<th>Gains</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong></td>
<td>1.74</td>
<td>4.3</td>
<td><strong>1.06%</strong></td>
</tr>
<tr>
<td><strong>Offline ( \phi )</strong></td>
<td>1.58</td>
<td>4.3</td>
<td><strong>0.91%</strong></td>
</tr>
<tr>
<td><strong>Offline ( \sigma )</strong></td>
<td>1.74</td>
<td>6.1</td>
<td><strong>0.68%</strong></td>
</tr>
</tbody>
</table>
Consumption-equivalent gains by income and density

Card income $\leq 50k$ 0.45%

Card income $> 50k$ 1.32%

Below-median density counties 0.85%

Above-median density counties 1.24%
Substitutability by NAICS

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: These are the 10 mixed offline/online 3-digit NAICS
Consumption-equivalent gains by 2017

1 big CES nest (baseline) 1.06%

16 CES nests (allocating nonstore retail) 1.62%

Note: assumes Cobb-Douglas aggregation of nests
<table>
<thead>
<tr>
<th>Due to rising $q_o$ and $A_o$</th>
<th>2007–2017 Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$ spending per offline merchant</td>
<td>$-1.6%$</td>
</tr>
<tr>
<td>$M_b$ # of offline merchants bought from</td>
<td>$-2.1%$</td>
</tr>
<tr>
<td>$M_{b,\text{market}}$ # of offline merchants in the market</td>
<td>$-3.7%$</td>
</tr>
<tr>
<td>$\Pi$ profits of offline merchants</td>
<td>$0%$</td>
</tr>
</tbody>
</table>
Conclusions

1. Allowing for *variety* gains, surplus ≈ 1% of consumption

2. Consumer surplus from e-commerce is:
   - smaller for incomes below $50k (less likely to have cards)
   - larger in more densely populated counties

3. Modest implications for growth and inequality trends