The effect of Internet distribution on brick-and-mortar sales

Author: Andrea Pozzi

汇报人：侯力铭

2019年3月8日

Source: RAND Journal of Economics Vol. 44, No. 3, Fall 2013 pp. 569–583

CONTENT

01 Introduction

Introduction

This paper contributes new evidence on the effect of online distribution on a retailer's traditional sales by describing the case of a large supermarket chain (henceforth, the Retailer) which added an e-commerce service to its network of brick-and-mortar stores.

(1) Trade off or new business

(2) Mechanisms: travel costs (distance), competition (market structure)

(3)Endogeneity: IV(availability of online shopping, the distribution of coupons)

Contributions: only a limited empirical evidence; spatial differentiation.

Unique data and rigorous analysis.

Introduction

(1) **Household level analysis**

examining household behavior and investigate whether the introduction of Internet shopping leads customers to spend more at the Retailer.

$$Total\_Expenditure_{it} = \alpha_i + \tau_t + \beta Online\_Expenditure_{it} + \varepsilon_{it},$$

β : new business for the chain;

(1 − β) : trade off.

(2) **Store-level analysis**

using aggregated data on store sales to look directly at how this reflects on revenues of the chain.

$$\ln(Total\_sales_{szt}) = \alpha_s + \tau_t + \beta Online\_Available_{zt} + \varepsilon_{szt},$$

$$Incremental\ sales = (Fitted\ sales|_{\beta=\hat{\beta}} - Fitted\ sales|_{\beta=0}).$$

02

Environment and data description

Environment and data description

(1) The Retailer operates over 1500 brick-and-mortar stores across the US, and its online shopping was available in over 1600 zip codes; it is the only grocer offering Internet shopping in roughly 70 percent of these zip codes.

(2) The Retailer commits to offering the same prices and promotions in-store and online.

(3) To shop online, customers must register, providing their loyalty card number. The loyalty card number identifies the household in the data and allows for matching its online and in-store purchases.

(4) Online orders must be worth at least $50 to be processed and payment occurs at checkout by credit or debit card.

(5) The delivery fee is set at $9.95, but the Retailer frequently issues coupons offering discounts. The fee is also waived or reduced for large orders.

Environment and data description

Household-level data:

(1) The Retailer provided scanner data relative to all the shopping trips, online and in-store, made at the chain between June 2004 and June 2006 by a sample of almost 10,000 households.

(2) Households are in the sample if they shopped at least once in a supermarket store and at least once on the Internet in the period.

Store-level data:

(1) UPC(Universal Product Classification Code) provides a weekly panel detailing sales on the Retailer’s revenues for a sample of 118 stores between January 2004 and December 2006.

TABLE 1 Household Shopping Behavior, By Channel of Purchase

			Percentiles				
	Mean	*Standard Deviation*	*10th*	*25th*	*50th*	*75th*	*90th*
Panel A: All Trips ($n = 1{,}492{,}166$)							
Monthly expenditure	426.15	335.38	79.33	182.99	358.75	589.72	845.24
Trips per month	7.61	6.94	2	3	6	10	15
Expenditure per trip	56.01	68.17	4.46	10.97	29.79	76.9	148.32
Basket size	19.14	24.47	1	3	9	27	53
Total trips	160.05	143.53	32	66	125	212	320
Panel B: In-Store Trips ($n = 1{,}372{,}180$)							
Monthly expenditure	326.73	302.98	25.52	99.95	250.48	472.78	722.69
Trips per month	7	7.02	1	2	5	9	15
Expenditure per trip	46.71	58.39	4.08	9.99	25.82	60.22	120.26
Basket size	15.52	20	1	3	7	21	43
Total trips	147.18	144.4	20	52	110	199	309
Panel C: Online Trips ($n = 119{,}986$)							
Monthly expenditure	99.42	200.7	0	0	0	143.13	337.57
Trips per month	.61	1.08	0	0	0	1	2
Expenditure per trip	162.52	80.38	80.47	108.34	149.27	194.19	257.81
Basket size	60.49	31.8	29	40	55	74	97
Total trips	12.87	17.33	1	3	7	16	32
Panel D: Distance to Closest Supermarket Stores ($n = 9{,}323$)							
Distance to Retailer's store	1.43	2.01	0.37	0.64	1.08	1.70	2.56
Distance to competitors' store	2.55	5.27	0.53	0.86	1.44	2.47	4.45

Notes: Total and per trip expenditures are expressed in 2006 dollars. Figures for expenditure per trip and basket size are averages of households averages (i.e., the average expenditure per trip of the average household). Basket size is defined as the number of items (UPCs) purchased in a shopping trip. Distance from the closest supermarket store of the Retailer and from the closest competitor is computed in miles using data provided by the Retailer (for the former) and geodesic coordinates from References USA (for the latter). The sample includes the over 9000 households who shopped *at least* once online and *at least* once in-store at the grocery chain between June 2004 and June 2006.

03

Household level analysis

Household level analysis

Regression form:

$$Total\_Expenditure_{it} = \alpha_i + \tau_t + \beta Online\_Expenditure_{it} + \varepsilon_{it},$$

(1) β : represents incremental revenues (new business for the chain);

(2) (1 − β) : gives an estimate of crowding out (trade off).

(3) αi is household fixed effect (unobserved heterogeneity among households).

(4) τt is time fixed effect (seasonal patterns and aggregate shocks to demand for grocery).

(5) Total expenditure and Online expenditure are expressed in 2006 dollars and computed net of promotional discounts and the fee paid for home delivery.

TABLE 2 The Effect of Introducing Internet Shopping on Households' Consumption at the Retailer

	(1) OLS	(2) OLS	(3) IV	(4) OLS	(5) OLS
Online expenditure	0.665***	0.750***	0.553***	0.620***	0.683***
	(0.012)	(0.024)	(0.028)	(0.010)	(0.011)
Online expenditure*		− 0.008*			
distance from competitors		(0.005)			
Online expenditure*		0.042***			
distance from Retailer		(0.011)			
Lagged total expenditure				0.159***	
				(0.009)	
Household f.e.	Yes	No	Yes	Yes	Yes
Observations	196,148	172,113	167,590	180,725	239,167
R-squared	0.29	0.26	0.22	0.25	0.31
Number of hhid	9,323	7,789	7,789	9,194	11,629

Notes: This table reports estimates of the composition of online expenditure for customers of the Retailer. The model estimated is the one in equation (1): the coefficient on *online expenditure* (β in equation (1)) represents incremental revenues and $(1 - \beta)$ gives an estimate of crowding out. The unit of observation is a household-month; standard errors (in parenthesis) are clustered at the household level. In column 2, I include demographic variables from the Census 2000 matched using the block group of residence of the household. Variables included are: share of males, share of Blacks, share of Hispanics, share of people aged 25-34, 35-44, 45-54, 55-64, and over 65, share of families, share of college graduates, share of employed, median household income, and share of commuters for 60 minutes or longer. I also include the distance in miles between the household residence and the closest store of the chain and the distance in miles between the household residence and closest store of a competitor. The former is computed using data provided by the Retailer, the latter using geodesic coordinates from References USA. These coefficients are not reported for reasons of brevity, but full results are available upon request. The instruments used in column 3 are a dummy for availability of the service in the zip code of residence of the household and an indicator variable signaling the availability of a coupon for a household in a particular month. In column 5, I consider only expenditure in perishable and nonstorable items. All monetary amounts are expressed in 2006 dollars. Significance levels: *10%, **5%, ***1%.

Household level analysis

(1) Why the addition of an Internet distribution channel induces the gain of new business?

- Reducing transportation costs for shoppers.
- **Column 2:** households living further away from stores of the chain indeed generate more additional business.

(2) What are the sources of the incremental sales?

- Market expansion(substituting for the outside good) and business stealing(business diverted from rival grocers).
- But without additional data or strong assumptions, we cannot separately identify.

Household level analysis

(3) Endogeneity of online expenditure: unobserved individual shocks to demand for grocery correlated with the choice of shopping on the web.

- **Instrumental variables:**

① availability of online shopping in the zip code of residence of the household;

(Although the Retailer's decision to introduce online distribution in a market may be based on the **expected demand**, as long as the timing of rollout is uncorrelated with demand considerations, the instrument is valid.)

② the distribution of coupons entitling customers to a discount fee for the Internet service in a given month

(The Retailer follows a "blanket" approach and mails coupons with discounts to all registered customers living in a given zip code. Therefore, coupons availability is by construction orthogonal to individual shocks to demand for grocery)

- **Column 3:** The first stage (not reported) shows that both instruments are positively and significantly correlated with online expenditure.

TABLE A1 Impact of Future E-Commerce Availability on Zip Code Level Sales of the Chain

	(1)	(2)	(3)
Available	262.4***	358.9***	108***
	(37.4)	(119.2)	(40.3)
Available in t + 1		82.9	−89.2
		(110.1)	(60.3)
Available in t + 2		72.4	−77.8
		(86.8)	(64.3)
Available in t + 3		104.1	−75.2
		(89.4)	(55.2)
Available in t + 4		58.2	−121.3*
		(85.2)	(73.1)
Available in t + 5		74.6	−55.2
		(70.9)	(49.5)
N	8,319	8,319	8,319
Zip code f.e.	Yes	No	Yes

Notes: This table assesses the impact of future and current availability of online grocery on the total sales of the chain to the households included in the sample, aggregated at the zip code level. *Available* is a dummy variable that takes value one in each month where the Retailer offers online grocery in the zip code. The set of indicator variables *Available in t + s* denote that the Retailer will start offering online grocery in the zip code in *s* months. Standard errors (in parenthesis) are clustered at the zip code level. Year-month fixed effects are included in all specifications. The sample includes only the zip codes where the Retailer introduced online grocery between June 2004 and June 2006. Significance levels: *10%, **5%, ***1%.

- If introduction of online grocery is decided as a response to increased demand, current expenditure for grocery in a market could be correlated with future availability of the service. Otherwise, the leads should not be significant.

Household level analysis

(4) Inventory motive: (the intertemporal cannibalization of online shopping on brick-and-mortar sale, since e-commerce is well suited for large stock-up purchases which fulfill grocery demand for current and future periods.)

- **Column 4** controls for lagged expenditure in grocery which proxies for household inventory.
- **Column 5** considers only expenditure in perishable grocery products, such as eggs or milk, which cannot be stockpiled.

- The resulting changes in the estimated share of new business gained online are small and do not alter the economic bottom line.

Household level analysis

(5) Implications for the Retailer's revenues.

$$Incremental\ sales = (Fitted\ sales|_{\beta=\hat{\beta}} - Fitted\ sales|_{\beta=0}).$$

The estimated value of the channel ranges between 11.5 and 14 millions of dollars over the two years.

04 Store level analysis

Store level analysis

Regression form:

$$\ln(Total\_sales_{szt}) = \alpha_s + \tau_t + \beta Online\_Available_{zt} + \varepsilon_{szt},$$

(1) where **s** indexes a particular store set in zip code z and **t** indicates a month.

(2) αs is store fixed effect (time-invariant unobserved differences across locations).

(3) τt is time fixed effect (seasonal patterns).

(4) *Online Available* signals that the e-commerce service was provided in the market where the store is located.

Store level analysis

The comparison of household and store level analysis:

(1) Whereas individual data are available only for households using the loyalty card, store revenues also include transactions by customers who do not hold one.

(2) Store level analysis allows for a more direct approach to quantifying the impact of e-commerce on revenues.

(3) Household level analysis does not consider regular customers who use the loyalty card only when shopping online and new customers who started shopping at the chain after the service was introduced. Store data include purchases of both these groups, allowing me to circumvent the problem.

TABLE 3 The Effect of Introducing Internet Shopping on Store Revenues

	(1)	(2)	(3)	(4)	(5)	(6)
Access in the store's	0.13**			0.28**	0.25**	0.25**
zip code	(0.012)			(0.115)	(.118)	(.096)
% zip codes with access		0.06**				
in the store's market		(0.032)				
% zip codes with coupons			0.08*			
in the store's market			(0.046)			
Access in the store's				− 0.01	− 0.17	
zip code * monopoly				(0.125)	(.157)	
Access in the store's				− 0.31*	− 0.08	
zip code * duopoly				(0.176)	(0.163)	
Access in the store's				0.01		
zip code * two competitors				(0.140)		
Access in the store's				− 0.49***		
zip code * three competitors				(0.187)		
Access in the store's						− 0.13*
zip code * online competitors						(0.081)
Store f.e.	Yes	Yes	Yes	No	No	No
Observations	3,041	2,926	2,926	2,963	2,963	2,807
R-squared	0.22	0.02	0.03	0.20	0.15	0.14

Notes: The dependent variable is the logarithm of total monthly store revenues. A store's market is defined as the zip code where it is located in columns 1, 4, 5, and 6; whereas it includes all the zip codes to whose centroid the store is closer than any other store of the chain in columns 2 and 3. The shares of zip codes with access to the online service and of those with coupons for delivery fee are computed using population weights based on information from Census 2000. The number of competitors in a store's zip code is computed using information on store location from Reference USA. Column 4 considers all supermarket stores competing with the Retailer 's chain, whereas column 5 only includes stores of major supermarket chains. Specifications in columns 4 and 5 include market level controls from Census 2000: share of Blacks, share of Hispanics, share of people aged 25-34, 35-44, 45-54, 55-64, and over 65, share of families, share of college graduates, median household income. The point estimates of the market structure dummies in colums 4-6 are not reported for reasons of brevity. All specifications include month-year fixed effects. Standard errors (in parenthesis) are clustered at the store level. Significance levels: *10%, **5%, ***1%.

Store level analysis

(1) Column 1: define the market of a store as the zip code where it is located.

- Store revenues go up by 13 percent after online shopping becomes available in the zip code.

(2) Column 2: first broaden the definition of a store market to include all the zip codes whose centroid is closer to it than to any other outlet of the chain, then calculate the share of the zip codes in its market in which e-commerce is available, weighting each zip code by its population.

- Increases in the penetration of the web service in the market of a store have a positive and sizeable effect on its revenues.

(3) Column 3: keep the same market definition as in column 2 and regress log revenues on the fraction of zip codes in the store's market that have been targeted for coupon distribution.

- Store revenues go up in months when coupons stimulate access to Internet commerce in its area.

Store level analysis

(4) how the impact of e-commerce on store revenues varies with market structure?

- **Column 4:** uses four separate indicator variables denoting different market structures and considers all supermarket stores competing with the Retailer ’s chain.

- The dummies for market structure are all positive: the Retailer enjoys higher revenues in markets where fewer rivals are present.

- The interaction dummies for the case of one and three competitors are negative: As the excluded group is “four or more competitors,” this implies that the revenue surge induced by the introduction of the service is lower for markets with fewer competitors.

Store level analysis

(5) how the impact of e-commerce on store revenues varies with market structure?

- **Column 5:** considers only outlets of "big competitors," that is, multistore chains with number of employees and revenues similar to those of the Retailer. And only defines three dummies: monopoly markets, duopoly, and markets with two or more competitors.

- The interaction coefficients have the expected sign: revenues increase less in markets where there is lower potential for business stealing. However, they are not significant.

- The results for the whole sample may be driven by the effect on small chains and individual stores, who suffer the bulk of the business stealing.

Store level analysis

(6) how the impact of e-commerce on store revenues varies with market structure?

- **Column 6:** considers if rival chains respond to the Retailer's decision to introduce e-commerce by doing the same thing, then what the effect of competition in the supply of the service on the amount of new business gained on the Internet.

- The interaction between online availability and online competition implies that the additional business generated online is split among the grocers providing the service. In particular, the presence of a rival e-grocer halves the revenue growth induced by the Internet channel for the Retailer.

05 Conclusion

Conclusions

This paper presented results on the effect of the introduction of an online shopping service for a large supermarket chain and showed that selling online allows the Retailer to considerably expand its sales with only modest self-cannibalization and document two interesting features of this result.

First, the reduction in transportation cost for customers shopping online at the Retailer is one of its driving forces.

Second, the author relates the magnitude of the revenue enhancement to the strength of the competition faced by the chain, then he finds stronger effects in areas where the chain faces more brick-and-mortar competitors and lower effects if there is an alternative online outlets.


THANK YOU !
2019.3.8