MIS7420

Seminar in Management Information Systems: Paper Replication with R

Prof. Zhiqiang Zheng, Prof. Vijay Mookerjee

Author: Beyza Celik, Luoying Chen, Yihong Liu, Duc Vu

NETID: BXC190000, LXC190027, YXL180111, DDV110020

Contents

List of Figures

List of Tables

List of Codes

1 Data Cleaning Process

In this section, we present our codes for data cleaning and panel data preparation. Package `dplyr` ?, `haven` ?, `sqldf` ?, `zoo` ?, `plm` ?, are used in this process.

Notice that the provided data only contain users' browsing transactions with purchase. When we apply search engine filter on reference domain, besides google.com, yahoo.com, bing.com, some other search engines (msn.com, aol.com, live.com, mywebsearch.com) and five vendors (amazon.com, staples.com, dell.com, walmart.com, bestbuy.com) are also involved. For product categories, we only consider those sold at Circuit City and exclude three other types (Business machines, Office furniture, Office supplies).

For `CCStorePresent`, its value is the same as `Store_Close_Status` in original data. And we set `AfterStoreClosing` to 0 if the time of user transaction is before November, 2008 and 1 otherwise. For `BBStorePresent`, we set it as 1 for user transactions if there's any Best Buy stores around user's location and 0 otherwise, using the `bestbuyzipcodes_sample` dataset. `NoReferringDomain` is the variable we construct if the user directly came to the target website without any referring domain. And `ReferringDomainIsSearchEngine` is the variable we construct if the user was referred by search engine to the target website.

Finally, two panel data are constructed by concatenating `sales_allotherzipcode` with `sales_ccity0mile` and `sales_ccity5mile`. After the concatenation and aggregation, we found that the built panel data are unbalanced, in a sense that, for instance, `zip_code` 75080 only has 2 records, instead of 24 (2 years). It happens because (a) the provided data are a small sample from the whole original one; (b) the original data might not cover the full 2 years period. Unbalanced panel data has been studied by many researchers ?, like unbalanced seemingly unrelated regression ?. Here we adopt a naive solution: we impute the missing values for one `zip_code` by averaging those non-missing values of this `zip_code`.

2 Paper Replication

In this section, we provide our replication for this paper. Names for subsections correspond to the tables in the published paper. Package `stargazer` ? and `estout` ? are used to export estimation into LaTeX format.

2.1 Table 1

Table 1 shows the summary statistics of top five vendors by sales volume.

Table 1: Summary Statistics of Top Five Vendors by Sales Volume

Domain Name	**Total Transactions**	**Total Sales**	**Total Pages Viewed**	**Pages Per Dollar**	**Total Duration**	**Mins Per Dollar**
dell.com	1,620	483,703.300	66,953	0.138	57,225.660	0.118
amazon.com	10,904	354,573.300	464,383	1.310	369,227.900	1.041
staples.com	5,927	236,982.300	247,163	1.043	166,189.900	0.701
walmart.com	1,977	156,606.100	80,397	0.513	68,434.890	0.437
bestbuy.com	1,230	149,950.400	50,627	0.338	36,735.900	0.245

Codes for generating Table 1 are listed below.

2.2 Table 2

Table 2 summarizes the frequency of referral channels for various online retailers.

Table 2: Summary Statistics of Referring Domain Categories

Domain Name	**Total Transactions**	**Referred by Search Engine**	**Direct to Website**	**Referred by Others**
amazon.com	10,904	2,955(27.1%)	7,018(64.4%)	931(8.6%)
bestbuy.com	1,230	258(21.0%)	901(73.3%)	71(5.8%)
All Others	36,794	6,999(19.0%)	25,483(69.3%)	4,312(11.7%)
All Transactions	48,928	10,212(20.9%)	33,402(68.3%)	5,314(10.9%)

Codes for generating Table 2 are listed below.

2.3 Table 3

Table 3 reports the model-free average DID values for some outcome variables.

Table 3: Average Difference-in-Difference (DID) of the Outcome Variables

Outcome Variable	Groups	After Store Closure	Before Store Closure	First Difference (se)	DID
Amazon Sales	Control	3.418	3.303	0.115 (0.031)	-0.167
	Treatment	3.351	3.403	-0.052 (0.212)	
Amazon PagesPerDollar	Control	1.188	1.147	0.041 (0.025)	0.257
	Treatment	1.363	1.065	0.298 (0.153)	
Amazon MinsPerDollar	Control	1.016	0.975	0.041 (0.025)	0.263
	Treatment	1.187	0.882	0.304 (0.137)	
bestbuy.com Sales	Control	3.418	3.303	0.354 (0.031)	0.623
	Treatment	3.351	3.403	0.976 (0.212)	
bestbuy.com PagesPerDollar	Control	1.188	1.147	-0.109 (0.025)	0.074
	Treatment	1.363	1.065	-0.035 (0.153)	
bestbuy.com MinsPerDollar	Control	1.016	0.975	-0.084 (0.025)	-0.012
	Treatment	1.187	0.882	-0.096 (0.137)	

Codes for generating Table 3 are listed below.

2.4 Table 4

In order to examine whether a competing online retailer benefits from the presence of a local showroom, we run the following regressions for Amazon.com and BestBuy.com:

$$
\begin{aligned}
&\log\left(\texttt{TotalMonthlySales}+1\right)_{i,t} \\
&= \mu_i + \tau_t \\
&+ \beta_1\ \texttt{CCStorePresent}_i \times \texttt{AfterStoreClosing}_t \\
&+ \beta_2\ \texttt{CCStorePresent}_i \times \texttt{AfterStoreClosing}_t \times \texttt{BBStorePresent}_i \\
&+ \epsilon_{i,t}
\end{aligned} \tag{1}
$$

Table 4: Results of the Sales Effect (All Product Categories)

	log(TotalMonthlySales + 1)			
	Amazon-0 Mile	Amazon-5 Miles	BestBuy-0 Mile	BestBuy-5 Miles
	(1)	(2)	(3)	(4)
β_1	0.014	−0.005	−0.002	−0.002
	(0.015)	(0.008)	(0.033)	(0.008)
β_2	−0.033	0.003	0.009	0.002
	(0.022)	(0.010)	(0.036)	(0.010)
Observations	68,472	75,096	14,664	16,848
R^2	0.00003	0.00001	0.00002	0.00000
Adjusted R^2	-0.044	-0.044	-0.045	-0.045
F Statistic	1.091 (df = 2; 65594)	0.278 (df = 2; 71942)	0.154 (df = 2; 14028)	0.035 (df = 2; 16121)

Note: *p<0.1; **p<0.05; ***p<0.01

Codes for generating Table 4 are listed below.

2.5 Table 5

To measure the impact of the exit of local showrooms on consumer online search intensity and the moderating effect of Best Buy Stores as an alternative local showroom, we run the following regressions:

$$
\begin{aligned}
&\log(\texttt{PagesPerDollar} + 1,\ \texttt{MinsPerDollar} + 1)_{i,t} \\
&= \mu_i + \tau_t \\
&+ \beta_1\ \texttt{CCStorePresent}_i \times \texttt{AfterStoreClosing}_t \\
&+ \beta_2\ \texttt{CCStorePresent}_i \times \texttt{AfterStoreClosing}_t \times \texttt{BBStorePresent}_i \\
&+ \epsilon_{i,t}
\end{aligned} \tag{2}
$$

Table 5: Results of the Search Effect (All Product Categories)

	log(PagesPerDollar + 1)				log(MinsPerDollar + 1)			
	Amazon-0 Mile	Amazon-5 Miles	BestBuy-0 Mile	BestBuy-5 Miles	Amazon-0 Mile	Amazon-5 Miles	BestBuy-0 Mile	BestBuy-5 Miles
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1	0.003	-0.019^{***}	0.001	0.002	0.004	-0.021^{***}	0.001	0.003
	(0.012)	(0.007)	(0.016)	(0.004)	(0.012)	(0.007)	(0.013)	(0.003)
β_2	-0.068^{***}	0.018^{**}	0.003	-0.001	-0.057^{***}	0.022^{***}	0.0004	-0.002
	(0.018)	(0.009)	(0.018)	(0.005)	(0.017)	(0.008)	(0.014)	(0.004)
Observations	68,472	75,096	14,664	16,848	68,472	75,096	14,664	16,848
R^2	0.0004	0.0001	0.00003	0.00004	0.0003	0.0001	0.00001	0.0001
Adjusted R^2	-0.043	-0.044	-0.045	-0.045	-0.044	-0.044	-0.045	-0.045
F Statistic	12.530^{***} (df = 2; 65594)	3.985^{**} (df = 2; 71942)	0.202 (df = 2; 14028)	0.337 (df = 2; 16121)	8.867^{***} (df = 2; 65594)	5.187^{***} (df = 2; 71942)	0.046 (df = 2; 14028)	0.451 (df = 2; 16121)

Note: *p<0.1; **p<0.05; ***p<0.01

Codes for generating Table 5 are listed below.

2.6 Table 6

To capture the expected change in the odds ratio of the impact of Circuit City store closures and the moderating effect of Best Buy stores as an alternative local showroom, we run the following regressions:

$$
\begin{aligned}
&\text{Logit}\left(\texttt{ReferringDomainIsSearchEngine}, \texttt{NoReferringDomain}\right)_{i,t} \\
&= \mu_i + \tau_t \\
&+ \beta_1 \ \texttt{CCStorePresent}_i \times \texttt{AfterStoreClosing}_t \\
&+ \beta_2 \ \texttt{CCStorePresent}_i \times \texttt{AfterStoreClosing}_t \times \texttt{BBStorePresent}_i \\
&+ \epsilon_{i,t}
\end{aligned} \tag{3}
$$

Table 6: Results of the Sales Effect: Experience and Search Products

	log(TotalMonthlySales + 1)							
	Amazon-0 Mile-Exp	Amazon-5 Miles-Exp	Amazon-0 Mile-Search	Amazon-5 Miles-Search	BestBuy-0 Mile-Exp	BestBuy-5 Miles-Exp	BestBuy-0 Mile-Exp	BestBuy-5 Miles-Search
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1	−0.005	−0.007	0.005	−0.008	−0.011	−0.009	−0.001	−0.010
	(0.017)	(0.010)	(0.013)	(0.006)	(0.009)	(0.007)	(0.023)	(0.008)
β_2	−0.043*	0.009	−0.002	0.009		0.013	0.000	0.009
	(0.024)	(0.012)	(0.018)	(0.008)		(0.008)	(0.028)	(0.010)
Observations	32,112	35,568	52,392	57,648	10,224	11,712	5,664	6,600
R^2	0.0002	0.00002	0.00000	0.00003	0.0001	0.0002	0.00000	0.0002
Adjusted R^2	-0.044	-0.044	-0.044	-0.044	-0.046	-0.045	-0.048	-0.047
F Statistic	2.775* (df = 2; 30749)	0.318 (df = 2; 34061)	0.101 (df = 2; 50184)	0.774 (df = 2; 55221)	1.377 (df = 1; 9774)	1.297 (df = 2; 11199)	0.004 (df = 2; 5403)	0.746 (df = 2; 6300)

Note: *p<0.1; **p<0.05; ***p<0.01

Codes for generating Table 6 are listed below.

2.7 Table 7

We then test whether the show-rooming effect is stronger for experience goods, by grouping products into physical experience goods and search goods. Table 7 and 8 presents the results for sales model on physical experience products and search products separately.

Table 7: Results of the Online Search Effect: Experience Products

	log(PagesPerDollar + 1)				log(MinsPerDollar + 1)			
	Amazon-0 Mile	Amazon-5 Miles	BestBuy-0 Mile	BestBuy-5 Miles	Amazon-0 Mile	Amazon-5 Miles	BestBuy-0 Mile	BestBuy-5 Miles
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1	0.007	-0.037^{***}	0.006^{**}	0.001	0.006	-0.039^{***}	0.003	0.001
	(0.015)	(0.008)	(0.002)	(0.002)	(0.015)	(0.008)	(0.002)	(0.001)
β_2	-0.077^{***}	0.030^{***}		-0.0001	-0.067^{***}	0.034^{***}		-0.001
	(0.020)	(0.010)		(0.002)	(0.020)	(0.010)		(0.002)
Observations	32,112	35,568	10,224	11,712	32,112	35,568	10,224	11,712
R^2	0.001	0.001	0.001	0.00003	0.001	0.001	0.0003	0.0001
Adjusted R^2	-0.043	-0.044	-0.045	-0.046	-0.044	-0.044	-0.046	-0.046
F Statistic	12.857^{***} (df = 2; 30749)	10.009^{***} (df = 2; 34061)	5.763^{**} (df = 1; 9774)	0.143 (df = 2; 11199)	10.349^{***} (df = 2; 30749)	11.626^{***} (df = 2; 34061)	2.508 (df = 1; 9774)	0.438 (df = 2; 11199)

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Codes for generating Table 7 are listed below.

2.8 Table 8

Table 8: Results of the Online Search Effect: Search Products

	log(PagesPerDollar + 1)				log(MinsPerDollar + 1)			
	Amazon-0 Mile	Amazon-5 Miles	BestBuy-0 Mile	BestBuy-5 Miles	Amazon-0 Mile	Amazon-5 Miles	BestBuy-0 Mile	BestBuy-5 Miles
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1	0.001	0.006	0.001	0.009*	0.003	0.004	0.0001	0.009*
	(0.012)	(0.006)	(0.014)	(0.005)	(0.012)	(0.006)	(0.012)	(0.005)
β_2	−0.019	−0.002	−0.000	−0.007	−0.019	0.001	−0.000	−0.008
	(0.017)	(0.008)	(0.017)	(0.006)	(0.017)	(0.008)	(0.015)	(0.006)
Observations	52,392	57,648	5,664	6,600	52,392	57,648	5,664	6,600
R^2	0.00005	0.00002	0.00000	0.001	0.00004	0.00003	0.00000	0.001
Adjusted R^2	-0.044	-0.044	-0.048	-0.047	-0.044	-0.044	-0.048	-0.047
F Statistic	1.138 (df = 2; 50184)	0.553 (df = 2; 55221)	0.011 (df = 2; 5403)	1.590 (df = 2; 6300)	0.935 (df = 2; 50184)	0.696 (df = 2; 55221)	0.0001 (df = 2; 5403)	1.927 (df = 2; 6300)

Note: *p<0.1; **p<0.05; ***p<0.01

Codes for generating Table 8 are listed below.

2.9 Table 9

Table 9 presents the effect of store closure on referring domain.

Table 9: Results of Logistic Regression for Referring Domain

	ReferringDomainIsSearchEngine		NoReferringDomain	
	Amazon-0 Mile	BestBuy-0 Mile	Amazon-0 Mile	BestBuy-0 Mile
	(1)	(2)	(3)	(4)
β_1	-0.817^{*}	-15.12^{***}	0.325	-0.223
	(0.337)	(0.611)	(0.346)	(1.259)
β_2	0.697	14.43^{***}	-0.415	0.916
	(0.564)	(0.944)	(0.544)	(1.615)
Observations	10,791	1,225	10,791	1,225
Note:				*p<0.05; **p<0.01; ***p<0.001

Stata codes for generating Table 9 are listed below.

2.10 Table 10

By applying more traditional online search measures, we perform the same DID analysis for Amazon and bestbuy.com, to further investigate if the increase in search intensity manifests itself independent of sales amount.

Table 10: Results of the Online Sales and Search Effect (All Product Categories)

	log(SalesPerTransaction + 1)		log(PagesPerTransaction + 1)		log(MinsPerTransaction + 1)	
	Amazon-0 Mile	BestBuy-0 Mile	Amazon-0 Mile	BestBuy-0 Mile	Amazon-0 Mile	BestBuy-0 Mile
	(1)	(2)	(3)	(4)	(5)	(6)
β_1	0.012	−0.001	0.004	0.0002	0.006	0.0002
	(0.013)	(0.032)	(0.009)	(0.017)	(0.011)	(0.020)
β_2	−0.018	0.010	−0.021*	0.005	−0.021	−0.003
	(0.019)	(0.034)	(0.013)	(0.018)	(0.016)	(0.021)
Observations	68,472	14,664	68,472	14,664	68,472	14,664
R^2	0.00002	0.00003	0.0001	0.00004	0.00003	0.00001
Adjusted R^2	-0.044	-0.045	-0.044	-0.045	-0.044	-0.045
F Statistic	0.539 (df = 2; 65594)	0.213 (df = 2; 14028)	1.867 (df = 2; 65594)	0.304 (df = 2; 14028)	0.939 (df = 2; 65594)	0.066 (df = 2; 14028)

Note: *p<0.1; **p<0.05; ***p<0.01

Codes for generating Table 10 are listed below.

2.11 Table 11

To further investigate if the increase in search intensity has a causal link to the Circuit City store closures and not due to other endogenous reasons, we adopt coarsened exact matching algorithm to match each zip code from the treatment group with an equivalent zip code from the control group, using zip code level demographics (average household age, average income and average household size). The matching results left us with 56 zip codes in each group. Using the data from the combined 112 zip codes, we ran the models for sales and search.

Table 11: Results of the Online Sales and Search Effect After Matching Zip Codes: TotalMonthlySales, PagesPerDollar, and MinsPerDollar (All Product Categories)

	log(TotalMonthlySales + 1)		log(PagesPerDollar + 1)		log(MinsPerDollar + 1)	
	Amazon-0 Mile	BesyBuy-0 Mile	Amazon-0 Mile	BesyBuy-0 Mile	Amazon-0 Mile	BesyBuy-0 Mile
	(1)	(2)	(3)	(4)	(5)	(6)
β_1	0.019	−0.0002	0.006	−0.001	0.003	−0.0002
	(0.019)	(0.002)	(0.012)	(0.003)	(0.011)	(0.002)
β_2	−0.026		−0.023		−0.024*	
	(0.024)		(0.016)		(0.013)	
Observations	1,776	384	1,776	384	1,776	384
R^2	0.001	0.00002	0.001	0.0001	0.002	0.00003
Adjusted R^2	-0.058	-0.113	-0.057	-0.113	-0.056	-0.113
F Statistic	0.740 (df = 2; 1677)	0.008 (df = 1; 344)	1.183 (df = 2; 1677)	0.030 (df = 1; 344)	1.931 (df = 2; 1677)	0.012 (df = 1; 344)

Note: *p<0.1; **p<0.05; ***p<0.01

Codes for generating Table 11 are listed below.

2.12 Table 12

To examine the possible heterogeneity within the geographic zip code area which may be unaccounted for, we add location specific demographics in the regression equations as interaction terms with DID term.

Table 12: Results of the Online Sales and Search Effect with Zip Code Demographics as Interactions and Time Fixed Effects (All Product Categories)

	log(TotalMonthlySales + 1)		log(PagesPerDollar + 1)		log(MinsPerDollar + 1)	
	Amazon-0 Mile	BestBuy-0 Mile	Amazon-0 Mile	BestBuy-0 Mile	Amazon-0 Mile	BestBuy-0 Mile
	(1)	(2)	(3)	(4)	(5)	(6)
β_1	−0.00001	−0.00001	0.0001	0.00001	0.0001	0.00000
	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
β_2	−0.00001	−0.0001	−0.0002*	0.00002	−0.0002	0.00001
	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Observations	68,472	14,664	68,472	14,664	68,472	14,664
R^2	0.00000	0.00004	0.00005	0.00002	0.00003	0.00001
Adjusted R^2	-0.044	-0.045	-0.044	-0.045	-0.044	-0.045
F Statistic	0.019 (df = 2; 65594)	0.255 (df = 2; 14028)	1.478 (df = 2; 65594)	0.114 (df = 2; 14028)	1.131 (df = 2; 65594)	0.053 (df = 2; 14028)

Note: *p<0.1; **p<0.05; ***p<0.01

Codes for generating Table 12 are listed below.

2.13 Table 13

In order to address the serial correlation issue, the first solution is to ignore the time series and the results is in Table 13.

Table 13: Results of the Online Sales and Search Effect After Matching Zip Codes: TotalMonthlySales, PagesPerDollar, and MinsPerDollar (All Product Categories)

	log(TotalMonthlySales + 1)		log(PagesPerDollar + 1)		log(MinsPerDollar + 1)	
	Amazon-0 Mile	BesyBuy-0 Mile	Amazon-0 Mile	BesyBuy-0 Mile	Amazon-0 Mile	BesyBuy-0 Mile
	(1)	(2)	(3)	(4)	(5)	(6)
β_1	0.023	−0.001	0.007	−0.003	0.009	−0.001
	(0.015)	(0.003)	(0.010)	(0.006)	(0.008)	(0.004)
β_2	−0.026		−0.023		−0.024*	
	(0.024)		(0.015)		(0.013)	
Observations	1,776	208	1,776	208	1,776	208
R^2	0.001	0.0003	0.001	0.001	0.002	0.0004
Adjusted R^2	-0.043	-0.083	-0.043	-0.083	-0.042	-0.083
F Statistic	1.169 (df = 2; 1700)	0.052 (df = 1; 191)	1.182 (df = 2; 1700)	0.197 (df = 1; 191)	1.663 (df = 2; 1700)	0.078 (df = 1; 191)

Note: *p<0.1; **p<0.05; ***p<0.01

Codes for generating Table 13 are listed below.

2.14 Table 14

For the serial correlation issue, another solution is to to use a White-like estimator to calculate the variance-covariance matrix of the error term.The results is in Table 14.

Table 14: Results of the Online Sales and Search Effect with Arbitrary Variance-Covariance Matrix Corrections (All Product Categories)

	log(TotalMonthlySales + 1)		log(PagesPerDollar + 1)		log(MinsPerDollar + 1)	
	Amazon-0 Mile	BestBuy-0 Mile	Amazon-0 Mile	BestBuy-0 Mile	Amazon-0 Mile	BestBuy-0 Mile
	(1)	(2)	(3)	(4)	(5)	(6)
β_1	0.019	−0.001	0.006	−0.003	0.003	−0.001
	(0.019)	(0.003)	(0.006)	(0.006)	(0.011)	(0.004)
β_2	−0.026		−0.023***		−0.024***	
	(0.028)		(0.001)		(0.006)	
Observations	1,776	208	1,776	208	1,776	208
R^2	0.001	0.0002	0.001	0.001	0.002	0.0003
Adjusted R^2	-0.058	-0.156	-0.057	-0.156	-0.056	-0.156
F Statistic	0.740 (df = 2; 1677)	0.036 (df = 1; 179)	1.183 (df = 2; 1677)	0.136 (df = 1; 179)	1.931 (df = 2; 1677)	0.054 (df = 1; 179)

Note: *p<0.1; **p<0.05; ***p<0.01

Codes for generating Table 14 are listed below.

2.15 Table C1

Table 15: Change in Demographics after Circuit City Store Closure

Group	**Before Store Closure**			**After Store Closure**			**First Difference of Mean (p-value)**		
	Mean Age	**Mean Income**	**Mean Education**	**Mean Age**	**Mean Income**	**Mean Education**	**Mean Age**	**Mean Income**	**Mean Education**
Control	7.048	4.479	97.957	6.937	4.498	97.999	-0.111 (<0.0001)	0.019 (0.300)	0.042 (0.639)
Treated	7.68	4.971	98.632	6.645	4.739	96.843	-1.035 (<0.0001)	-0.232 (0.029)	-1.789 (0.004)

Codes for generating Table 15 are listed below.

2.16 Table D1-D3

In this section, we further investigate the relationship between search and sales to understand the underlying conversion.

2.16.1 Table D1

In order to capture how much time a user would spend on a page on average before making a purchase, we define a new search intensity measure called minutes per page. And we run the following model to explore the relationship between search intensity and change in sales:

$$\begin{aligned}
&\log\left(\texttt{Sales}+1\right)_{i,t} \\
&= \mu_i + \tau_t \\
&+ \beta_1\ \texttt{MinsPerPage}_{i,t} \\
&+ \beta_2\ \texttt{ExperienceGood}_{i,t} \\
&+ \beta_3\ \texttt{MinsPerPage}_{i,t} \times \texttt{ExperienceGood}_{i,t} \\
&+ \epsilon_{i,t}
\end{aligned} \tag{4}$$

Table 16: Search Intensity Effects on Sales for Amazon

	(1) Log(Sales + 1)
β_1	2.376***
	(0.0435)
β_2	3.194***
	(0.0675)
β_3	-2.153***
	(0.0744)
Observations	10791

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Stata codes for generating Table 16 are listed below.

2.16.2 Table D2

Next, we explored the correlation between product characteristics and search intensity during a transaction for Amazon sales:

$$\begin{aligned}
&\log\left(\texttt{PagesViewed}, \texttt{MinsSpent} + 1\right)_{i,t} \\
&= \mu_i + \tau_t \\
&+ \beta_1 \ \texttt{ProductPrice}_{i,t} \\
&+ \beta_2 \ \texttt{ExperienceGood}_{i,t} \\
&+ \epsilon_{i,t}
\end{aligned} \tag{5}$$

Table 17: Product Characteristics Effects on Search Intensity for Amazon

	(1) Log(PagesViewed)	(2) Log(MinsSpent + 1)
β_1	0.00465***	0.00450***
	(0.000843)	(0.000785)
β_2	3.156***	2.923***
	(0.0581)	(0.0558)
Observations	10791	10791

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Stata codes for generating Table 17 are listed below.

2.16.3 Table D3

To study the search path used by experienced goods buyers, we run the following regression to assess whether experience goods buyers come directly to Amazon or through a search engine:

$$\begin{aligned}&\text{Logit}\left(\texttt{RefDomainIsAmazon}, \texttt{RefDomainIsSearchEngine}\right)_{i,t}\\&= \tau_t\\&+ \beta_1\ \texttt{ExperienceGood}_{i,t}\\&+ \epsilon_{i,t}\end{aligned} \tag{6}$$

Table 18: Product Characteristics Effects on Search Intensity for Amazon

	(1)	(2)
	RefDomainIsAmazon	ReferringDomainIsSearchEngine
ExperienceGood	-4.274***	-0.828***
	(0.310)	(0.0672)
Observations	10791	10791

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Stata codes for generating Table 18 are listed below.

2.17 Table E1-E3

In this section, we examine the effect of a physical store closure on some other online retailers, to investigate concern that the effect of the Circuit City store may also be felt by the other smaller online consumer electronic stores.

2.17.1 Table E1

In order to check which online seller benefits due to the exit of the offline Circuit City store, we run the following regression model:

$$
\begin{aligned}
&\log(\texttt{TotalMonthlySales}+1)_{i,t} \\
&= \mu_i + \tau_t \\
&+ \beta_1\ \texttt{CCStorePresent}_i \times \texttt{AfterStoreClosing}_t \\
&+ \beta_2\ \texttt{CCStorePresent}_i \times \texttt{AfterStoreClosing}_t \times \texttt{BBStorePresent}_i \\
&+ \epsilon_{i,t}
\end{aligned} \tag{7}
$$

Table 19: Results of the Sales Effect (All Product Categories)

	log(TotalMonthlySales + 1)			
	staples.com-0 Mile	walmart.com-0 Mile	dell.com-0 Mile	circuitcity.com-0 Mile
	(1)	(2)	(3)	(4)
β_1	−0.027	0.026	−0.006	0.003
	(0.064)	(0.018)	(0.018)	(0.036)
β_2	0.082	−0.034	−0.018	0.013
	(0.075)	(0.022)	(0.026)	(0.051)
Observations	8,352	24,912	19,440	3,940
R^2	0.0003	0.0001	0.0001	0.0001
Adjusted R^2	-0.046	-0.044	-0.045	-0.058
F Statistic	1.004 (df = 2; 7979)	1.332 (df = 2; 23849)	0.834 (df = 2; 18605)	0.094 (df = 2; 3722)

Note: *p<0.1; **p<0.05; ***p<0.01

Codes for generating Table 19 are listed below.

2.17.2 Table E2

We run the following regression model to investigate if the change in search intensity as seen for amazon.com is also prevalent in the other major online seller categories.

$$
\begin{aligned}
&\log(\texttt{PagesPerDollar}+1, \texttt{MinsPerDollar}+1)_{i,t} \\
&= \mu_i + \tau_t \\
&+ \beta_1\ \texttt{CCStorePresent}_i \times \texttt{AfterStoreClosing}_t \\
&+ \beta_2\ \texttt{CCStorePresent}_i \times \texttt{AfterStoreClosing}_t \times \texttt{BBStorePresent}_i \\
&+ \epsilon_{i,t}
\end{aligned}
\tag{8}
$$

The results of the search models are presented in Table 20 and 21.

Table 20: Results of the Online Search Effect (All Product Categories)

	log(PagesPerDollar + 1)			
	staples.com-0 Mile	walmart.com-0 Mile	dell.com-0 Mile	circuitcity.com-0 Mile
	(1)	(2)	(3)	(4)
β_1	0.010	0.004	0.001	−0.002
	(0.027)	(0.009)	(0.004)	(0.012)
β_2	−0.017	−0.002	−0.002	0.0002
	(0.031)	(0.011)	(0.005)	(0.016)
Observations	8,352	24,912	19,440	3,940
R^2	0.00004	0.00001	0.00001	0.00002
Adjusted R^2	-0.047	-0.045	-0.045	-0.058
F Statistic	0.171 (df = 2; 7979)	0.123 (df = 2; 23849)	0.083 (df = 2; 18605)	0.030 (df = 2; 3722)
Note:				*p<0.1; **p<0.05; ***p<0.01

Codes for generating Table 20 are listed below.

2.17.3 Table E3

Table 21: Results of the Online Search Effect (All Product Categories)

	log(MinsPerDollar + 1)			
	staples.com-0 Mile	walmart.com-0 Mile	dell.com-0 Mile	circuitcity.com-0 Mile
	(1)	(2)	(3)	(4)
β_1	−0.011	0.002	−0.001	−0.002
	(0.022)	(0.008)	(0.003)	(0.010)
β_2	0.008	−0.001	−0.001	0.00003
	(0.027)	(0.010)	(0.005)	(0.014)
Observations	8,352	24,912	19,440	3,940
R^2	0.00003	0.00000	0.00001	0.00002
Adjusted R^2	-0.047	-0.045	-0.045	-0.058
F Statistic	0.137 (df = 2; 7979)	0.056 (df = 2; 23849)	0.124 (df = 2; 18605)	0.037 (df = 2; 3722)
Note:				*p<0.1; **p<0.05; ***p<0.01

Codes for generating Table 21 are listed below.

2.18 Table G1-G3

2.18.1 Table G1

In order to further rule out other alternate explanations, we extracted the product categories most purchased by returning customers. We accordingly run the following regressions (with the selected product category sales as the outcome variable), for all the focal online competitors to Circuit City.

$$
\begin{aligned}
&\log(\texttt{TotalMonthlySales}+1)_{i,t} \\
&= \mu_i + \tau_t \\
&+ \beta_1 \ \texttt{CCStorePresent}_i \times \texttt{AfterStoreClosing}_t \\
&+ \beta_2 \ \texttt{CCStorePresent}_i \times \texttt{AfterStoreClosing}_t \times \texttt{BBStorePresent}_i \\
&+ \epsilon_{i,t}
\end{aligned} \tag{9}
$$

Table 22: Results of the Sales Effect (Music, Movies and Videos, Console Video Games)

	log(TotalMonthlySales + 1)		
	amazon.com-0 Mile	bestbuy.com-0 Mile	circuitcity.com-0 Mile
	(1)	(2)	(3)
β_1	0.005	−0.001	−0.002
	(0.013)	(0.024)	(0.043)
β_2	0.008	0.000	
	(0.019)	(0.028)	
Observations	52,416	5,808	810
R^2	0.00002	0.00000	0.00000
Adjusted R^2	-0.044	-0.048	-0.092
F Statistic	0.535 (df = 2; 50207)	0.004 (df = 2; 5541)	0.001 (df = 1; 741)
Note:			*p<0.1; **p<0.05; ***p<0.01

Codes for generating Table 22 are listed below.

2.18.2 Table G2

We further, included all the zip codes in our dataset even if they did not have any sale at one of the five top online competitors. The results are given below:

Table 23: Results of the Sales Effect (All Products; All Online Sellers in the Control Group)

	log(TotalMonthlySales + 1)				
	amazon.com-0 Mile	bestbuy.com-0 Mile	staples.com-0 Mile	walmart.com-0 Mile	circuitcity.com-0 Mile
	(1)	(2)	(3)	(4)	(5)
β_1	0.014	−0.002	−0.027	−0.006	0.003
	(0.015)	(0.033)	(0.064)	(0.018)	(0.036)
β_2	−0.033	0.009	0.082	−0.018	0.013
	(0.022)	(0.036)	(0.075)	(0.026)	(0.051)
Observations	68,472	14,664	8,352	19,440	3,940
R^2	0.00003	0.00002	0.0003	0.0001	0.0001
Adjusted R^2	-0.044	-0.045	-0.046	-0.045	-0.058
F Statistic	1.091 (df = 2; 65594)	0.154 (df = 2; 14028)	1.004 (df = 2; 7979)	0.834 (df = 2; 18605)	0.094 (df = 2; 3722)

Note: *p<0.1; **p<0.05; ***p<0.01

Codes for generating Table 23 are listed below.

2.18.3 Table G3

There might also be a concern that we use individual level transactions to tease out how customers landed into Amazon's site after the store closure, not the zip code level aggregated data. Hence, we aggregated all the sales per zip code per month and calculated the ratio of sales navigation originating from a search engine and also those going directly to amazon.com. We then run a regression on both ratios as outcome variables.

$$
\begin{aligned}
&(\texttt{AmazonReferringDomainIsSearchEngine Ratio}, \texttt{NoReferringDomain Ratio})_{i,t} \\
&= \mu_i + \tau_t \\
&+ \beta_1\ \texttt{CCStorePresent}_i \times \texttt{AfterStoreClosing}_t \\
&+ \beta_2\ \texttt{CCStorePresent}_i \times \texttt{AfterStoreClosing}_t \times \texttt{BBStorePresent}_i \\
&+ \epsilon_{i,t}
\end{aligned} \tag{10}
$$

Table 24: Effect Referring Domain on Amazon Sales

	ReferringDomainIsSearchEngineRatio	NoReferringDomainRatio
	Amazon	Amazon
	(1)	(2)
β_1	-0.010^{**}	0.009^{*}
	(0.005)	(0.005)
β_2	0.011	-0.008
	(0.007)	(0.007)
Observations	73,416	73,416
R^2	0.0001	0.00004
Adjusted R^2	-0.044	-0.044
F Statistic (df = 2; 70332)	1.961	1.422
Note:		*p<0.1; **p<0.05; ***p<0.01

Codes for generating Table 24 are listed below.

3 Advanced Method

In this section, we present some advanced methods on the provided data.

3.1 Generalized Synthetic Control Method

In this dataset, the DID model is a good fit because 1) the treatment of all treated units begins at a same time t, which produces the parallel trend, and 2) no treatment heterogeneity exists. However, we still give a test to the GSC method. The GSC method merges the Interactive Fixed Effects (IFE) model and the Synthetic Control (SC) method. It accepts one treatment variable and observable control variables and handles unobserved factors in the model. R package `gsynth` is used in this section.

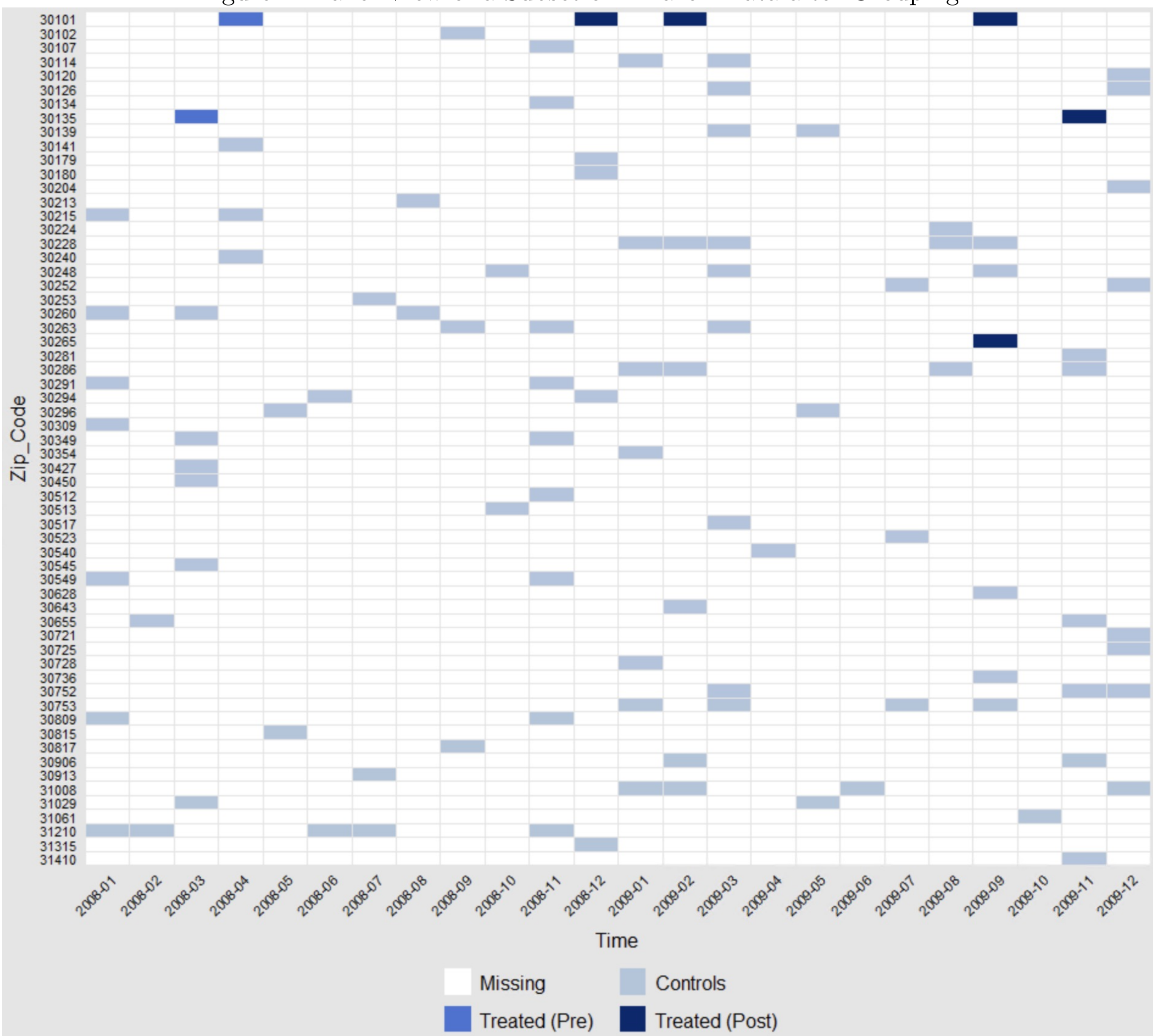
Because the three-way interaction variable `CCStorePresent` × `AfterStoreClosing` × `BBStorePresent` in the original paper is one of the way used to address endogeneity concern and the GSC model is able to handle the omitted variables problem and only accepts one treatment variable, we decide to focus on the interaction variable `CCStorePresent` × `AfterStoreClosing`. Also, because we are lacking BestBuy data regarding the pre-treatment periods and the number of units, the GSC model is not able to fit into Bestbuy data after grouping, which leads us to focus on Amazon data only.

Xu (2017) suggests that data should have at least 10 pre-treatment periods and 40 control units to apply the model. Since our data begins in January 2018 and the treatment begins in November 2018, the condition for the number of pre-treatment periods is satisfied. On the other hand, after grouping the data based on zip code and month, we have 2796 control units in the 0-mile data and 2855 control units in the 5-mile data. Therefore, the condition for the number of control variables is also satisfied.

R package `gsynth` allows us to feed in one outcome variable, one treatment variable and control variables and get the best value of r, the number of latent variables, directly after cross-validation. Besides, the model from the package drops the units which only have data records in less than n time periods, where n can be changes by users but should be at least 3. With larger n, the model involves less bias. However, because our data is imbalance and sparse, we are not able to use n larger than 3 because otherwise too many data will be dropped and the model cannot be applied. Also, we are restricted with r no greater than 1 because of the same reason. Figure 1 shows the panel view of a subset of data as an example. We can observe how sparse the data is and see the limited number of available periods in each units.

With all restrictions stated above, we run the model on sales effect and search breadth and depth on both 0-mile data and 5-mile data, with outcome variable `AmazonTotalMonthlySales`, `AmazonPagesPerDollar` and `AmazonMinsPerDollar`, respectively. Table 25 shows the chosen number of latent variables and the

Figure 1: Panel View of a Subset of Amazon Data after Grouping



Average Treatment effect on the Treated unit (ATT) in each of the models.

Table 25: Number of Latent Variables and the Averaage Treatment Effect for Amazon Data

	AmazonTotalMonthlySales		AmazonPagesPerDollar		AmazonMinsPerDollar	
	Amazon-0 Mile	Amazon-5 Miles	Amazon-0 Mile	Amazon-5 Miles	Amazon-0 Mile	Amazon-5 Miles
	(1)	(2)	(3)	(4)	(5)	(6)
ATT	−0.863	−0.248	−0.210	0.024	−0.293	−0.293
	(1.316)	(0.229)	(18.580)	(0.211)	(1.038)	(1.067)
r*	0	0	1	0	0	0
MSPE of r	1.294	1.312	0.015	1.375	0.342	0.342

Codes for each of the models are listed below.

3.2 PSM and LA-PSM

3.3 Causal Forest

In this section we focus on transaction level data and our treated group is determined by using a a variable called `treatment` which takes value of 1 if `CCStorePresent` and `AfterStoreClosing` are 1 and takes value of 0 otherwise.

Codes for constructing treatment variable are listed below.
We first investigate the treatment effects on Amazon sales using the transactions within the zip code where a Circuit City store was closed. For each transaction $i = 1, \cdots, n$, we observe a binary treatment indicator `treatment` (W_i), a real valued outcome `prod_totprice` (Y_i), as well as 10 categorical covariates which are `hoh_most_education`, `census_region`, `household_size`, `hoh_oldest_age`, `children`, `racial_background`, `connection_speed`, `country_of_origin`, `prod_category_type` and `BBStorePresent`; and 4 real-valued covariates which are `pages_viewed`, `duration`, `prod_qty`, `household_income`. We expanded out categorical random variables via one-hot encoding, thus resulting in covariates $X_i \in \mathbb{R}^p$ with $p = 38$.

We define causal effects via the potential outcomes model (Imbens and Rubin, 2015): For each sample i, the potential outcomes denoted by $Y_i(0)$ and $Y_i(1)$ corresponding to the outcome we would have observed if the i-th sample was in control or treatment group, and assume that we observe $Y_i = Y_i(Wi)$. The average treatment effect is then defined as $\tau = \mathbb{E}[Y_i(1) - Y_i(0)]$, and the conditional average treatment effect function is $\tau(x) = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i] = x$.

Codes for estimating treatment effects on Amazon sales within the zip code where a Circuit City store was closed are listed below.
We use the package `grf` to apply causal forest on our data and also to estimate the average treatment effect. the confidence interval for the average treatment effect is presented in Table 26. Since the confidence interval includes zero we cannot conclude that treatment effect is significant.

Table 26: 95% CI for the ATE on Amazon Sales (Zero Mile Data)

2.5%	Estimate	97.5%
-19.28	-7.68	3.92

Table 27: 95% CI for the ATE on Amazon Pages Per Dollar (Zero Mile Data)

2.5%	Estimate	97.5%
-39.25	-23.62	-7.98

4 Reference

Xu Y (2017) Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis.* 25(1):57-76