SIEPR - Giannini Data Center

Data originating from sharing agreement with a "large retail chain"

Sofia B. Villas-Boas

Department of Agricultural and Resource Economics

Data presentation September 10, 2012

Background

- These data originate from a Data Sharing Agreement with a large U.S. retail chain and U. C. Berkeley, now extended to all U. Cs. & other schools
- I am the main Berkeley contact with the retailer.
- This data sharing agreement allows any University of California researcher, faculty, or graduate student, to submit a project to the Retail chain via the data center and ask for the necessary data to design and perform such project using existing or new data.

Collaboration

The KEY Determinant of successful collaboration so far for research projects and data requests is the constant formal and informal feedback on current process and procedures from

- retailer's perspective
- academia perspective

Collaboration

- The retailer shares retail insights and provides feedback on internal details and procedures relevant to specific research projects
- In addition to sharing results, researchers also share and increase communication regarding research methods and data analysis, including data code.
- Researchers provide preliminary initial results to the retailer within first three month of approved project (this is the most effective way to move on a project)
- Possible participation of researchers at internal retailer meetings relevant to their research topics (this is very useful)

Overview Talk

- SIEPR-GIANNINI Data Center
  - Deposit of all data originating from the sharing agreement
  - Functional web page and communication platform
- Data description and Data requests procedure
- Examples of Projects and Data Used

SIEPR-GIANNINI Data Center

Funding sources:

Stanford Institute for Economic Policy Research SIEPR

Economic research institute that brings together economist from all over Stanford campus (Economics, Business, Political Sciences, Law, Hoover Institute)

GIANNINI · FOUNDATION OF AGRICULTURAL ECONOMICS UNIVERSITY OF CALIFORNIA

Foundation that promotes and supports research activity in economics (funds libraries at UC Berkeley and Davis Campus)

Contacts :

Sofia Villas-Boas (sberto@berkeley.edu) and Kristin Kiesel (kiesel@saclink.csus.edu)

Data presentation September 10, 2012

SIEPR-GIANNINI Data Center

- Pilot of web page and communication platform
  http://are.berkeley.edu/SGDC/

"*" Are password protected, email me if you want to access and be part of mailing list

- ✓Information resource
- ✓Data archive
- ✓Secure data delivery option

- **EXISTING DATA** based requests:
  - Requests for existing data can be processed by center internally and included in quarterly progress report for update and approval by retail chain as long as involved researchers are already covered by existing data sharing agreement, the center uses caution to ensure that proposed projects are not in conflict with other existing projects and also taking into account general retailer's objectives
  - Requests for existing data that involve PI's/authors not currently covered by institutional data sharing agreements will be included in quarterly requests to be granted and to provide individual agreements by the retailer to the PI's institution if the retailer wishes to do so (this is how the schools have been added)
- **NEW DATA** pulls:
  - New data requests will still be submitted quarterly for approval by the retailer

Scanner Data and Auxiliary Data

1. store-level scanner data (*)
2. transaction-level scanner data(*)

(*) These data originate every time an item gets scanned at the cash register.

3. Data on product attributes, store characteristics.

4. Also other retail chain data (personnel, square footage and shelf layout, e.g.) are available that were/can be requested to develop specific projects.

Scanner Data

1. store-level scanner data

Retailer stores 4 year window by product, store week:

upc_id	upc_desc	qty	tot_net_amount	promo_week_id	store_id
00012345678	best_Milk_Ever_GL	45	250	201212	345
00012345678	best_Milk_Ever_GL	49	260	201213	345
00012345678	best_Milk_Ever_GL	12	45	201214	345
00012345678	best_Milk_Ever_GL	4	25	201212	3499
00012345678	best_Milk_Ever_GL	5	26	201213	3449
00012345678	best_Milk_Ever_GL	1	4	201214	3499

2. transaction-level scanner data

Retailer stores 2 years by transaction product, store, time:

upc_id	upc_desc	qty	tot_net_amt	promo_week_id	store_id	hh_id	date
00012345678	best_Milk_Ever_GL	2	12.5	201212	3499	778	3/22
00012345678	best_Milk_Ever_GL	2	12.5	201212	3499	123	3/26
00012345678	best_Milk_Ever_GL	3	18	201213	3449	778	3/29
00012345678	best_Milk_Ever_GL	2	8	201213	3449	123	4/1
00012345678	best_Milk_Ever_GL	1	4	201214	3499	778	4/8

Sample of Data Requests

- Analysis of productivity- 2 years *data (for peak hour time periods) of scanned items per second by cashier id, store, time,* Alexandre Mas, Enrico Moretti (Berkeley)
- Changes in consumer expenditures and commodity prices – four years weekly *product level price per product UPC over time and space,* Sofia B. Villas-Boas, Jeff Perloff and Charles Seguin (Berkeley)
- Demand for over the counter brand name and generic drugs – two years of *consumer level product choices (quantity and revenues) over time and stores for over the counter drugs -* Sofia Villas-Boas, and Mariana Carrera (Berkeley)
- Estimating wine demand – *4 years product store level weekly wine quantity and revenues,* Aviv Nevo and Carlo Prato (Northwestern)

Examples

1. Measure the causal effect of an exogenous shock on consumers' purchases. Combine household/store level scanner data with data on exogenously varying variables, time, or space changing variables. (Example Sideways)

2. Use store level field experimental data (Example OTC pilot this Summer)

Sideways Effects of Wine

With Fred Finan, Econ Berkeley

- We exploit the release of the movie *Sideways* in the U.S. markets to examine the extent to which third-party endorsement affects consumer behavior in wine purchases.

"If anyone orders Merlot, I'm leaving. I am not drinking any f -- Merlot.", - Miles in "Sideways", 2004.

"It's thin-skinned, temperamental, ripens early. It's not a survivor. Only when someone has taken the time to truly understand its potential can pinot be coaxed into its fullest expression (...)"

- Miles in "Sideways", 2004.

"I don't know that anybody is ready to lay this thing at the feet of Sideways, but there is clearly some small version of the French paradox," occurring. It is a classic example of a third-party endorsement, which is what we live and die on in the wine industry."

A spokesman for the pinot noir best-selling brand, 2004.

"Should there be a Sideways: The Musical, Merlot producers might stomp themselves to death en masse." - S.F. Chronicle, 2004.

- Using a data set of purchases for different stores of a large U.S. retail chain, we investigate if sales of different varieties of wine changed in markets where *Sideways* was released.

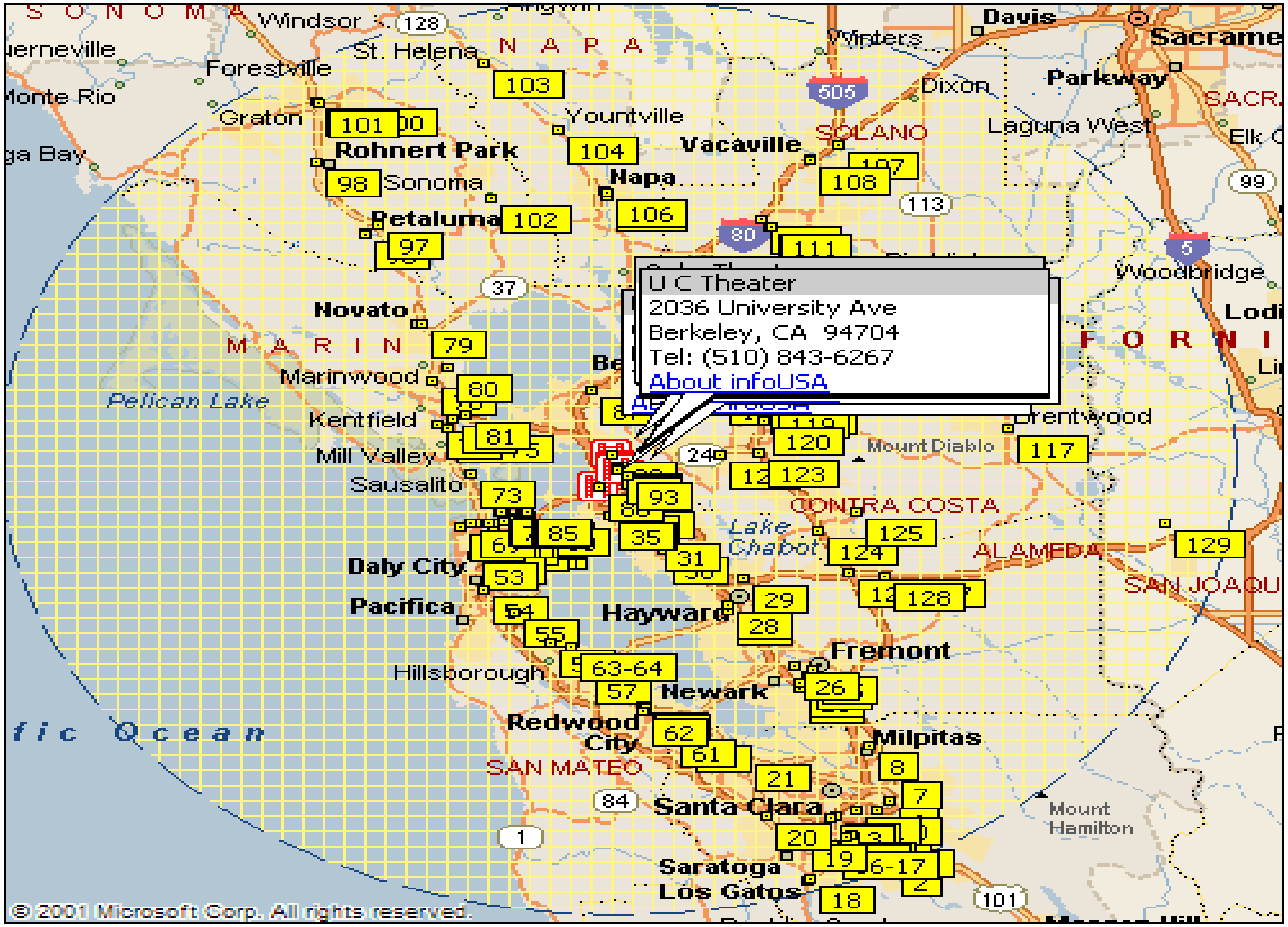
Measuring Sideways **Causal Effects**

Change in stores near movie theater showing sideways relative to changes in stores not near a theater showing sideways

•The way we approach this is to look at the change in wine bottles sold in stores that are really affected by the movie (Treated Stores) and compare to stores that are likely not affected by the movie but are quite similar (Control stores)

•Those are the control stores that have in common with the affected stores that for instance " prices of wine are the same and change the same way as in the affected stores, prices of substitutes like beer are the same and move the same as in the affected stores, have the same religious holiday happening where people may not drink so much wine as the affected stores."

Sideways Effects of Wine

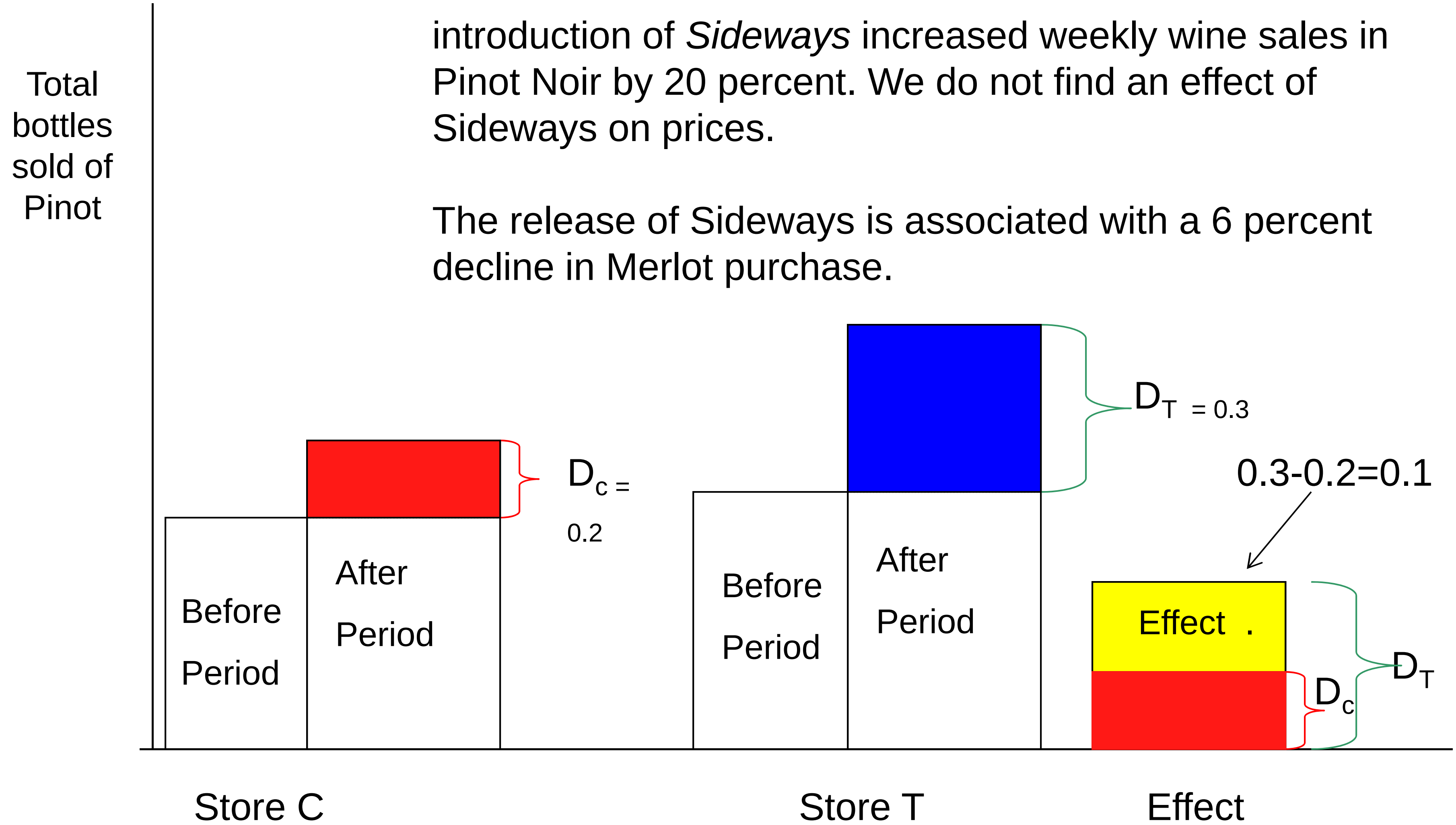


Sample of retail stores (in yellow) and movie theaters (in red) showing *Sideways* in the Bay Area.

Difference in difference

introduction of *Sideways* increased weekly wine sales in Pinot Noir by 20 percent. We do not find an effect of Sideways on prices.

The release of Sideways is associated with a 6 percent decline in Merlot purchase.



What Leads Consumers to Choose Generic Over-the-Counter Drugs (OTC)?

Preliminary Results from Pilot Intervention in 2012

With Mariana Carrera, Case Western Reserve University

Motivating Fact 1.

Based on our analysis of 2010-2011 purchase data when the price difference between generic brand and national brand OTC drugs doubles that only increases generic take-up rate by 2-3%.

Motivating Fact 2.

Generic OTC brand market share is about 20% when national brands take the rest despite generic drugs being much cheaper

Why aren't customers more responsive to price differences?

-Habitually choose the brand without noticing that a generic equivalent costs less?

-Fear that the quality of generics is worse?

-Price difference is difficult to interpret? (e.g. different counts in brand and generic packages)

-Need to know what others choose?

Pilot Test conducted May 16-June 13

Information treatments.

- **Treatment**: Posted tags with product-specific information beneath the price tags of generic products in selected treatment categories.
- Several stores, each given a different information treatment.
- Of eleven OTC drug categories with similar availability of generics: six were selected to receive treatment, and five served as controls.

We measured the effect of each treatment on weekly store sales.

- **Outcome 1: Quantity sold ($q_b + q_g$)**
- **Outcome 2: Generic purchase share $q_g/(q_b + q_g)$**
- Analysis conducted at the level of *active ingredient*, e.g. loratadine/Claritin
- Compare outcomes in:
  - *Treated* categories vs. *untreated* categories, at treated stores.
  - Treated categories at *treated* stores vs. at *untreated* stores.
- Regression analysis, adjusting for seasonal trends, price changes, and differences across stores.

Test 1: Are consumers doubtful of generic drug quality?

Three different statements based on FDA approval information posted.

"Bioequivalent"

Did you know?

The FDA determined this product to be bioequivalent and therapeutically equivalent to Zyrtec tablets.

Ref: ANDA078336 12/27/2007

[redacted]-brand drugs are subject to the same strict standards of good manufacturing practice as all others with the same active ingredients.

"Approved by FDA"

Did you know?

This product contains the same active ingredient as Zantac150 and has been approved by the FDA.

Ref: ANDA078653 11/16/2007

[redacted]-brand drugs are subject to the same strict standards of good manufacturing practice as all others with the same active ingredients.

"Same Active Ingredient"

Did you know?

This product contains the same active ingredient as Bayer regular-strength aspirin.

[redacted]-brand drugs are subject to the same strict standards of good manufacturing practice as all others with the same active ingredients.

Yes, but more information can have mixed results.

The share of generic purchases increased significantly in the group of bioequivalent products, but decreased among products receiving other labels.

Test 2: Are consumers inattentive to the price savings associated with buying generic? Probably yes.

Did you know?

Customers who choose Safeway-brand cetirizine save

60%.*

[redacted]brand drugs are subject to the same strict standards of good manufacturing practice as all others with the same active ingredients.

**Average savings per dose relative to Zyrtec, 2011 sales at the Br[redacted]d store.*

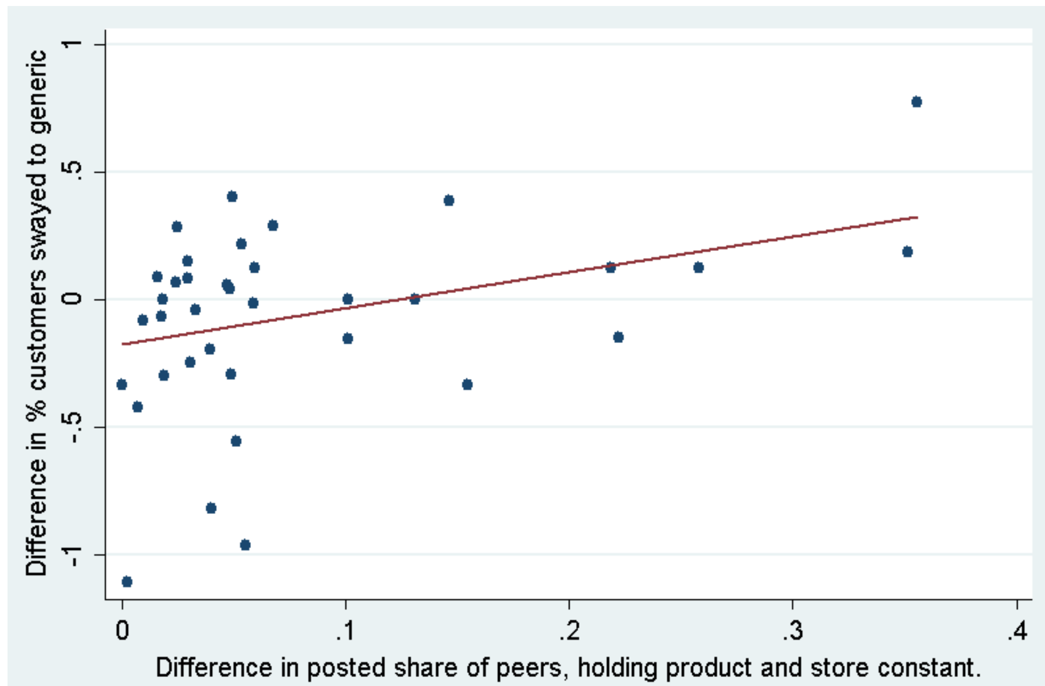
Percentage posted ranged from:
14% (fexofenadine/Allegra) to
68% (aspirin/Bayer).

The average generic purchase share across treatment products increased **10.5** percentage points.

Total quantity purchased also increased.

Test 3. Would more consumers try the generic if they learned how many fellow consumers buy it?

Yes, information on peer purchases is powerful, especially when the share of customers who buy generic is high.



We would need to increase the price difference by $2.43 (a doubling almost) to achieve the 5% average generic share increase due to the peers treatment

Our results suggest that an additional 10 percent in the *posted* share is associated with an increase of 6.2 percentage points in the effect.

Link: http://are.berkeley.edu/SGDC/

SIEPR-Giannini Data Center, a joint project between Stanford University and University of California, Berkeley for archiving and documenting existing data sets. It aims at connecting existing projects, providing commonly used resources, and making data more easily accessible for future projects.

Funding for this project is provided by the Stanford Institute for Economic Policy Research (SIEPR) and the Giannini Foundation.

Contacts :

Sofia Villas-Boas (sberto@berkeley.edu) and Kristin Kiesel (kiesel@saclink.csus.edu)

Thank you for your attention