

Private Sector Data in Economic Research

Jonathan Levin
Stanford and NBER

BEA/CNSTAT Meeting
November 19, 2015

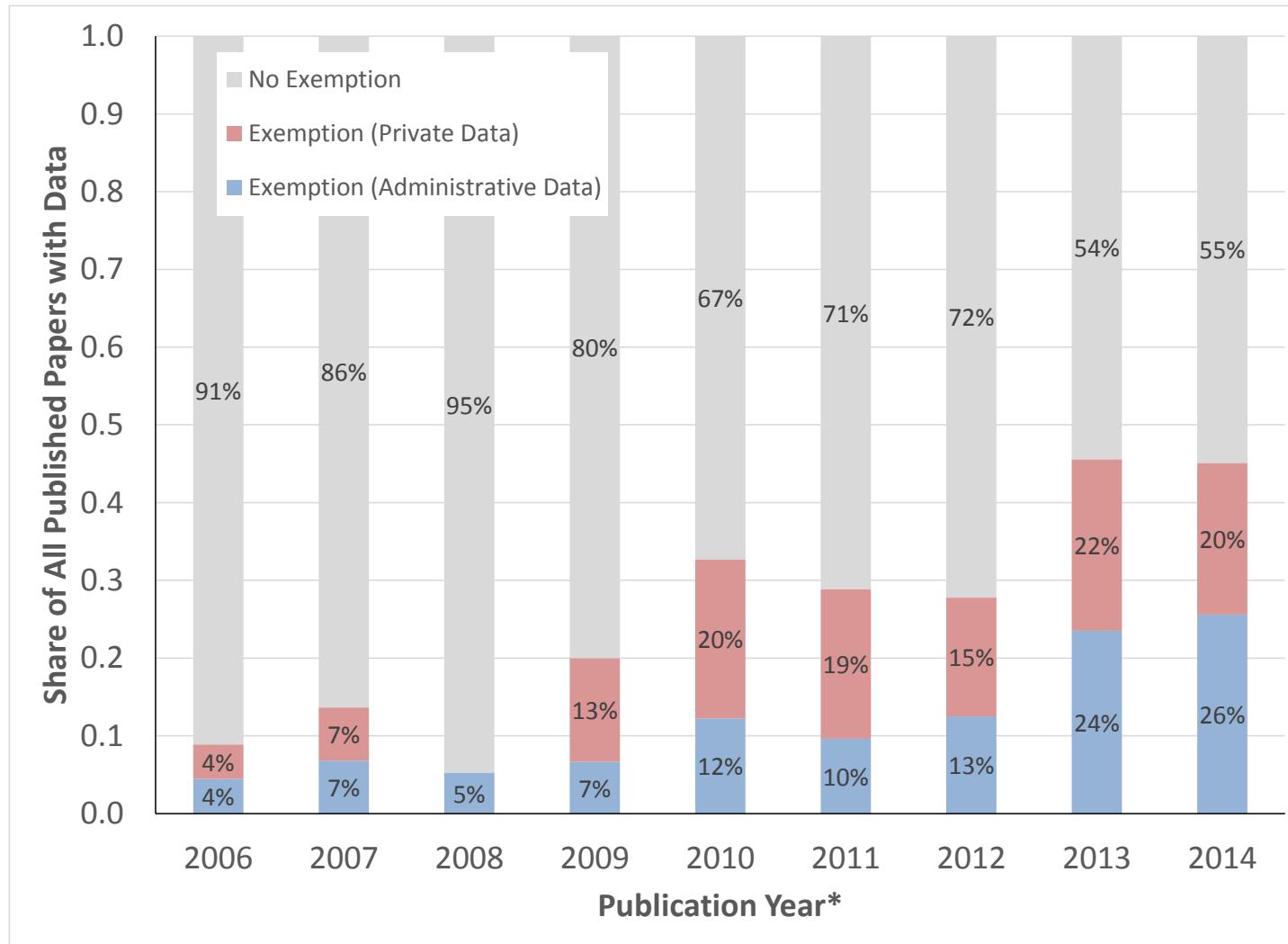
Rise of Empirical Research in Economics

TABLE 4
PERCENT DISTRIBUTIONS OF METHODOLOGY OF PUBLISHED ARTICLES, 1963–2011*

Year	Theory	Theory with simulation	Type of study		
			Empirical: borrowed data	Empirical: own data	Experiment
1963	50.7	1.5	39.1	8.7	0
1973	54.6	4.2	37.0	4.2	0
1983	57.6	4.0	35.2	2.4	0.8
1993	32.4	7.3	47.8	8.8	3.7
2003	28.9	11.1	38.5	17.8	3.7
2011	19.1	8.8	29.9	34.0	8.2

From Hamermesch (2013, *JEL*): “Six Decades of Economics Publishing”

Rise of Proprietary Data Use



From Einav and Levin (2014, *Science*): “Economics in the Age of Big Data”

Plan for Talk

- Discuss economic research using private sector data. Focus on proprietary data, rather than commercially available datasets, or information products. Expect others may comment on latter.
- For concreteness, describe a few of my own experiences collaborating with companies for economic research. (Note: will talk more about the process than the actual research!)
- Try to identify some general challenges and opportunities in using this sort of data, both for research + national statistics.

Example 1: Subprime Lending

- Collaboration with subprime auto sales and financing company
 - With Will Adams, Liran Einav, Mark Jenkins.
- Detailed operational data: hundreds of thousands of loan applications and offers, loans, payment and recovery outcomes.
 - Lens into behavior of poor, less stable households, and into functioning of high-risk consumer credit markets. Started in 2005 (good timing).
 - Many striking data features: 80% markups on cars, 30% interest rates, but still hard to lend profitably ... 65% default rates, low recoveries.
- How did it work?
 - Serendipitous - undergraduate advisee was roommate of the owner's son ... company transferred us their data to us in stages. Review process before posting papers ... became complicated around subprime crisis.

Difficulties in High-Risk Consumer Lending

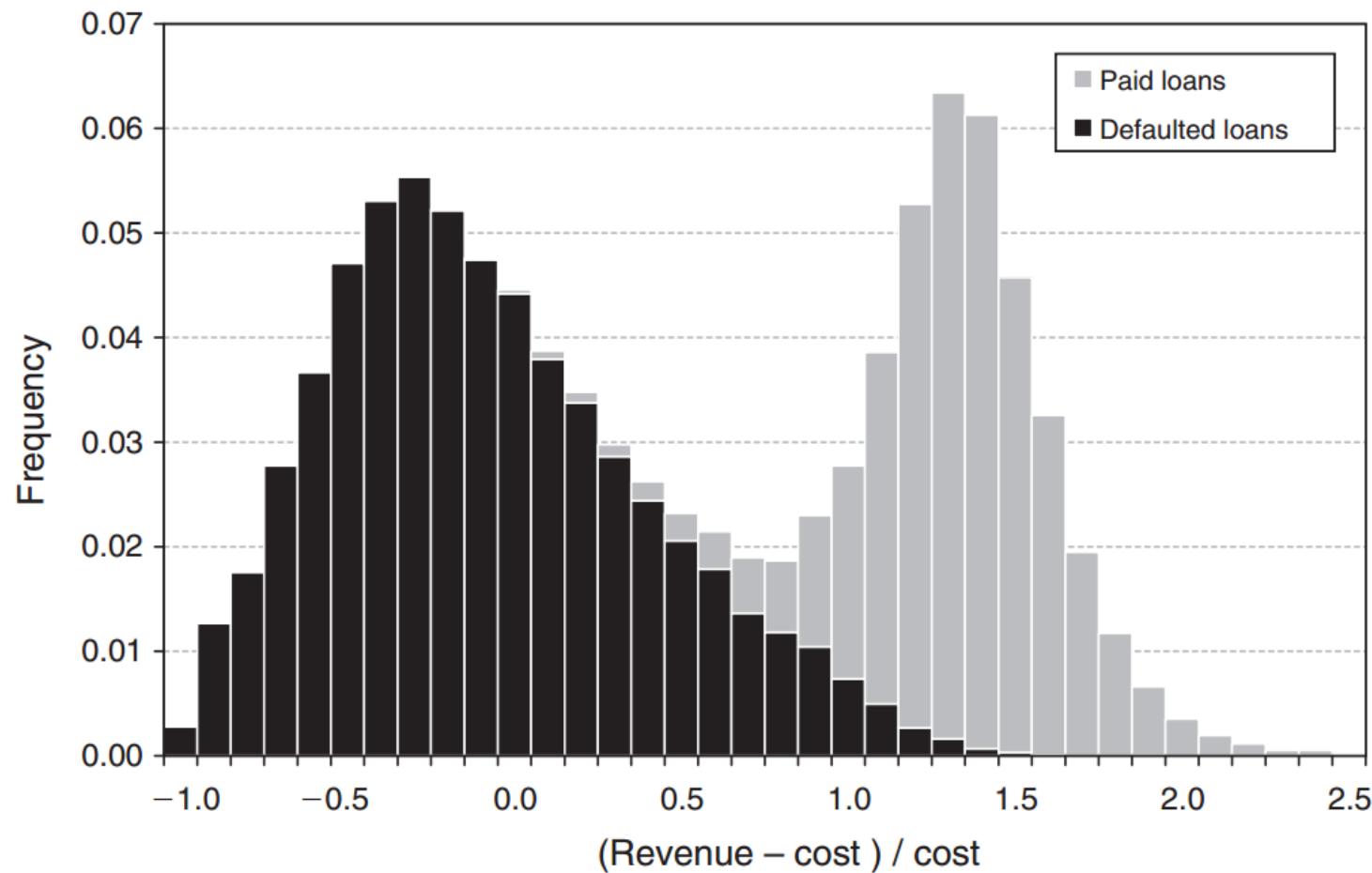


FIGURE 1B. RATE OF RETURN HISTOGRAM

From Adams, Einav, Levin (*American Economic Review*, 2009)

Example 2: eBay

- Collaboration with eBay to study internet commerce
 - With Liran Einav, Neel Sundaresan, multiple graduate students.
- Marketplace data: listings, transactions, browsing data (one billion listings a year ... auctions, posted prices, etc.).
 - Granular, large-scale data on consumer search and purchasing behavior, seller decisions about pricing, competition, cross-border trade, etc.
- How did it work?
 - Also serendipitous - Neel invited me to give a talk at eBay ... Data stayed on eBay computers & graduate students got “employee” access ... eBay reviewed papers to ensure no disclosure of sensitive information.

Estimating Sales Tax Sensitivity

- An eBay search results page:

View as: Sort by: Best Match Page 1 of 11640

	Sinamay Fascinator Mini Top 7" Hat w/Veil/3Flowers, White	1 Bid	\$22.99	2m
Quick look				
	Aeropostale mens embellished baseball cap Expedited shipping available	Top-rated seller	Buy It Now	\$13.99 29d 14h 15m Free shipping
More options				
	Miami Heat Snapback Mitchell and Ness	1 Bid	\$4.99	2m
Quick look				
	Chicago Bulls Snapback Mitchell and Ness	5 Bids	\$14.00	2m
Quick look				
	NWT MENS TOMMY HILFIGER LOGO BASEBALL CAP Expedited shipping available	Top-rated seller	Buy It Now	\$16.50 27d 16h 3m
Quick look				

- No info about sales tax ... only collected in-state ... or seller location.

Tax Sensitivity via “Tax Surprises”

- eBay item page:



Sinamay Fascinator Mini Top 7" Hat w/Veil/3Flowers, White

Like

Item condition: New with tags

Ended: Jul 10, 2011 21:54:35 PDT

Bid history: 1 bid

Winning bid: US \$22.99

Add to list

Shipping: \$6.99 Standard Shipping | See all details

Delivery: Estimated within 4-8 business days

Returns: No Returns Accepted

eBay Buyer Protection
Covers your purchase price plus original shipping.
Learn more

Seller info
dahlia-land (4)
100% Positive feedback

Save this seller
See other items

Other item info
Item number: 120745390657
Item location: Forest Hills, NY, United States
Ships to: United States
Payments: PayPal See details

Print | Report item

Enlarge

Description Shipping and payments

Seller assumes all responsibility for this listing.

Last updated on Jul 03, 2011 21:55:16 PDT [View all revisions](#)

Share: [Email](#) [Facebook](#) [Twitter](#)

- NYC buyer must pay an extra 8.875% in sales taxes --- we can see if he/she is less likely to purchase (or to purchase something else).

Example 3: Visa

- Collaboration with Visa to study consumers and retail sector.
 - With Liran Einav, Pete Klenow, Wayne Best, multiple graduate students.
- Visa network transactions: annual averages ... 197 million cards, 35.7 billion swipes, \$1.9 trillion sales. Sizeable fraction of retail.
 - Information on consumers – spending patterns, reaction to shocks, etc.
 - Broad census of retail outlets – see revenue and linked customers.
- How does it work?
 - Lucky initial meeting ... but multi-year process to get things going
 - Sensitive data. Requires care on legal and security issues. Data stays inside Visa, graduate students get access via Visa computers.

Private Data for Research: The Good

- Private companies collect *huge* amounts of data on economic activity.
- Often different visibility from government or commercial data – e.g. details of interests, search, communication, choice sets, decisions. Or other novelties such as “matched” consumer / retailer data, different populations.
- Often great opportunities for research design – granularity makes it possible to find natural experiments to identify consumer and firm responses.
- Encourages research that is practical and commercially valuable ... some would view this as a negative, but in my view often a positive.

The Bad: Access, Replicability, etc.

- Access and relationship issues
 - Access is uneven – requires a lot of luck and goodwill.
 - Limits on topics - not completely free to choose research projects.
 - Potential for hold-ups in publishing / clearing research.
- Replication and verification is difficult.
 - In many cases not impossible, but costs/time reqts are much higher.
 - Access also a key limitation for follow-on research....
- Not sure these can be overcome for all private company data, but could be improved greatly (with lots of work) for commercial datasets, perhaps through RDC model / expansion.

The Ugly: Data Issues

- Data is a convenience sample ...
 - May not be representative of any particular population other than company's customers, workers or clients.
- Variation in how private companies manage + validate data
 - Data gets stored and checked when it matters for business.
 - Old data sometimes deleted. Nature of data collection, sample, variable can change often. Data is very often not homogeneous over time.
- Significant set up costs and idiosyncrasy
 - Requires huge numbers of decisions – what to include or exclude from the sample, defining of categorizing variables of interest, how to aggregate, how to deal with data quality problems, missing variables.
 - Don't have long history & institutional memory we do with gov't survey.

Private Data and National Statistics

- Better measures of the same things – geo + frequency.
 - Problem is not “does useful data exist” --- often there are multiple sources --- it’s “where to selectively sink the resources to utilize it”.
 - Ground truth problem – BEA / BLS plays role of standard setter – low bar to fit a predictive model (though higher bar to fit a durable model).
 - Gov’t agencies could decide to outsource standard – e.g. use Zillow for housing component of inflation. Better? Cheaper? Surely trade-offs.
- Measurement of different things
 - Measures of well-being – time in traffic, health, physical activity.
 - Measures of social interaction
 - Measures of time use – what are people doing
 - Measures of environmental quality
 - Should these be part of national accounts? I have no idea, but certainly from a research standpoint, they are of considerable interest.