Digitization, access to information, and their impacts on traditional institutions

Imke Reimers

Northeastern University

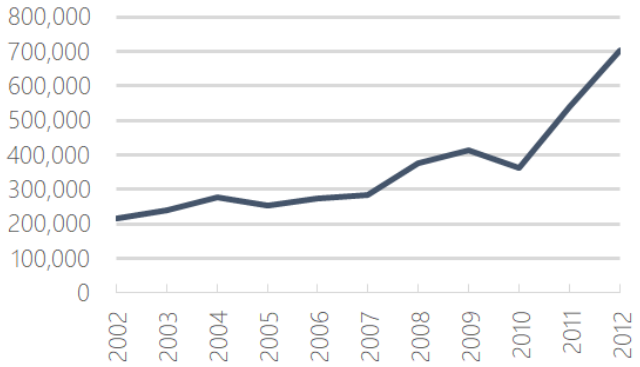
i.reimers@northeastern.edu

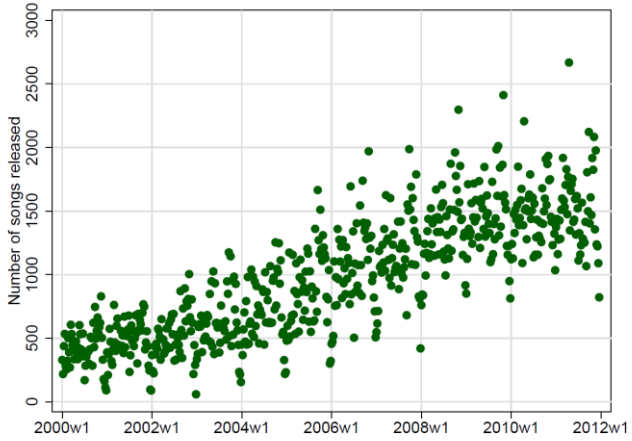
Nov 14, 2019

Giessen Workshop on the Economics of FBP Systems

Digitization…

- … has lowered costs of production and distribution
  - Many more products available
  - Potential welfare benefits, especially to consumers, are substantial



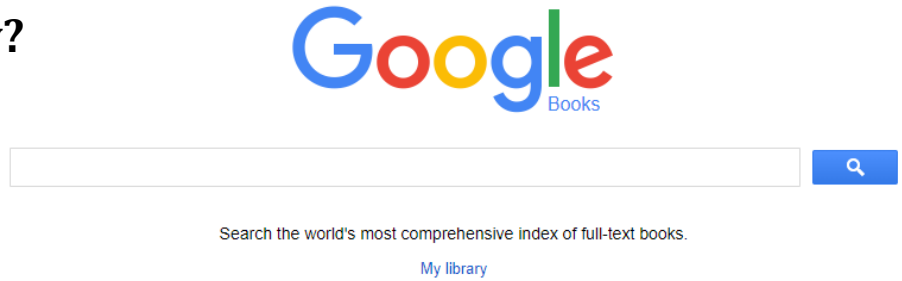


- … but how are traditional institutions affected by the changes?

How does digitization impact existing institutions?

- Changes in how authors can reach consumers (**self-publishing**)
  - How are traditional publishers affected?
  - **Can they benefit from entry?**

- Digitization of existing work (e.g. **Google Books**)
  - How are sales of traditional formats affected?
  - **Can digitization aide discovery?**



Digital Disintermediation and Efficiency in the Market for Ideas

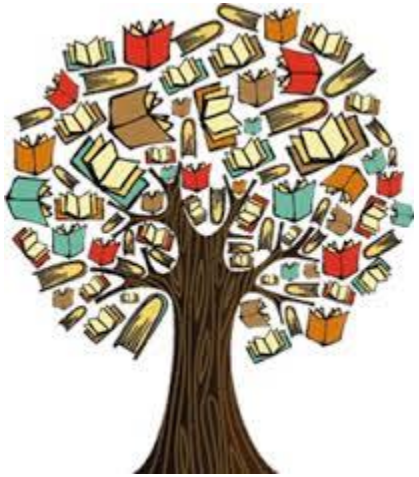
With Christian Peukert (ETH Zurich)

Circumventing the gatekeepers

- Inventors
  - Venture capital vs. Kickstarter

- Academics
  - Academic journals vs. own website, SSRN, etc.

- Book publishing
  - **Traditional publishers vs. self-publishing**



The Martian

Novel by Andy Weir
Self-published in 2011
re-published by Crown Publishing in 2014

Imkes Tagebuch (German Edition) Kindle Edition

by Chunying Xie (Translator)

Be the first to review this item

Look inside

See all formats and editions

Kindle
from $1.00

Read with Our **Free App**

Dieses ist ein fiktiver Eintrag ins Tagebuch. Er spiegelt mein ereignisreiches Leben wider.

Length: 1 pages | Enhanced Typesetting: Enabled | Language: German

Product Details

File Size: 138 KB
Print Length: 1 pages
Simultaneous Device Usage: Unlimited
Publication Date: October 21, 2010
Language: German
ASIN: B0048EL5OO
X-Ray: Not Enabled
Word Wise: Not Enabled
Lending: Not Enabled
Enhanced Typesetting: Enabled
Amazon Best Sellers Rank: #2,939,423 Paid in Kindle Store (See Top 100 Paid in Kindle Store)
#1624 in Kindle Store > Kindle eBooks > Foreign Languages > German > **Biographies, Diaries & True Accounts**
#1965 in Kindle Store > Kindle Short Reads > 15 minutes (1-11 pages) > **Biographies & Memoirs**
#35804 in Kindle Store > Kindle eBooks > Biographies & Memoirs > **Memoirs**

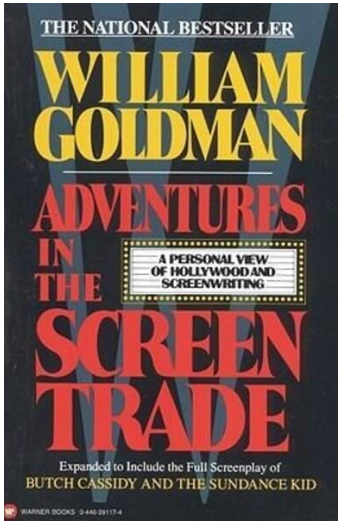
Key issue

Digital self-publishing platforms are challenging traditional gatekeepers

- How does it impact *license payments*?
  - Better outside option for authors – upward pressure on license fees

- Can it improve *efficiency*?
  - More books in the market
  - These may help publishers better predict ex-post appeal

- We test this with data on 90,000 license deals, 2002-2015



Data on expected and realized appeal

Expected appeal:

- **License deals** reported on Publishers Marketplace (2002 – 2015)
  - Author, working title, editor, publisher, genre
  - 5 size categories (<49k to >500k)

Realized appeal:

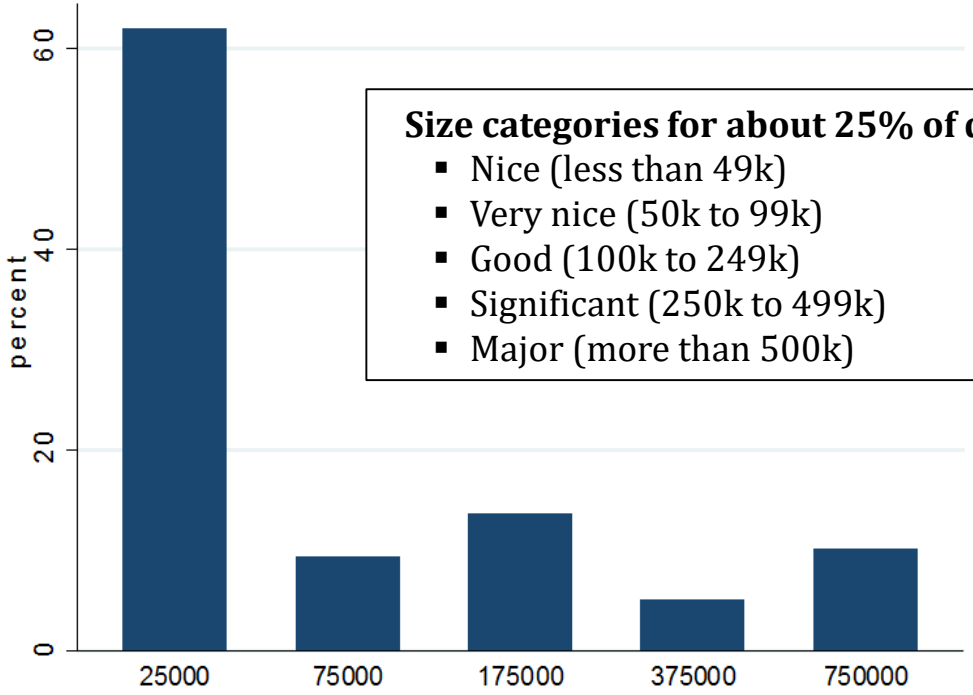
- **Unit sales** inferred from Nielsen Bookscan and USA Today (2002 – 2016)
  - Snapshots of weekly top 100 bestsellers

License deal data

- 52,000 book deals
- 40,000 rights deals

12/04/2006. Fiction: General/Other

Lynn York's second novel, a follow up to her debut The Piano Teacher, to Trena Keating at Plume, in a very nice deal, by Suzanne Gluck at the William Morris Agency.



Size categories for about 25% of deals

- Nice (less than 49k)
- Very nice (50k to 99k)
- Good (100k to 249k)
- Significant (250k to 499k)
- Major (more than 500k)

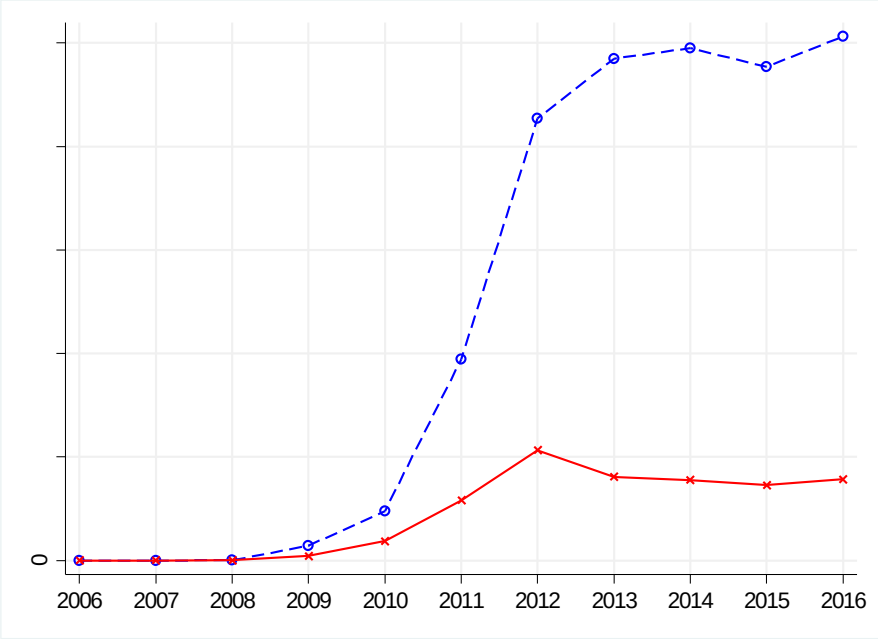
Identification and estimation

Identifying the effect of self-publishing

- **Problem**: digitization happened for all authors at the same time
  - Amazon's Kindle (and KDP): November 2007
  - Apple's iPad (Apr 2010); self-publishing platforms

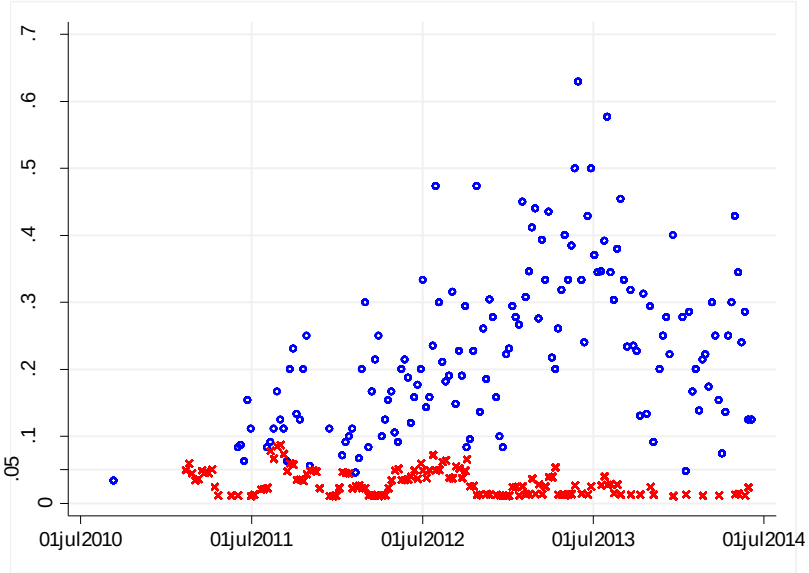
Romance books are more appealing for self-publishing than other genres…

- Low costs of entry into self-publishing
- Relatively high demand for e-book editions



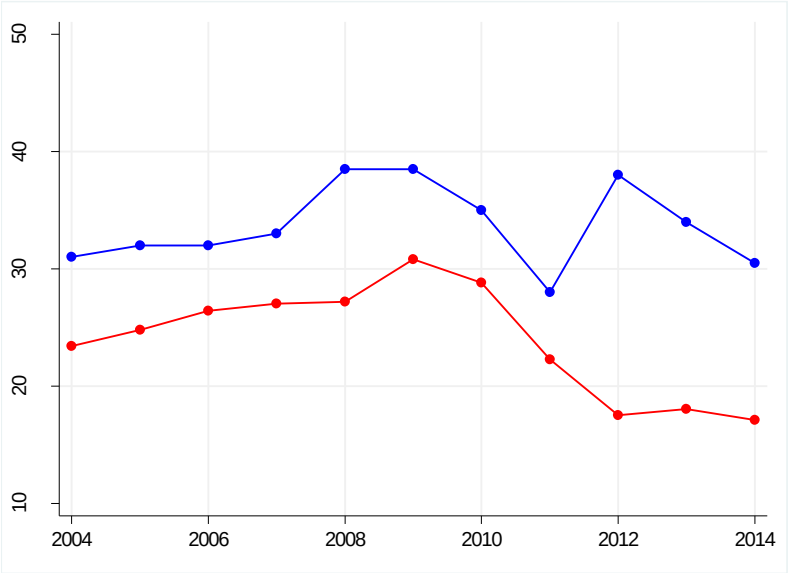
Romance authors as the treatment group

Self-publishing is particularly popular



Share of originally self-published books in Top 100
o Romance, × Non-Romance

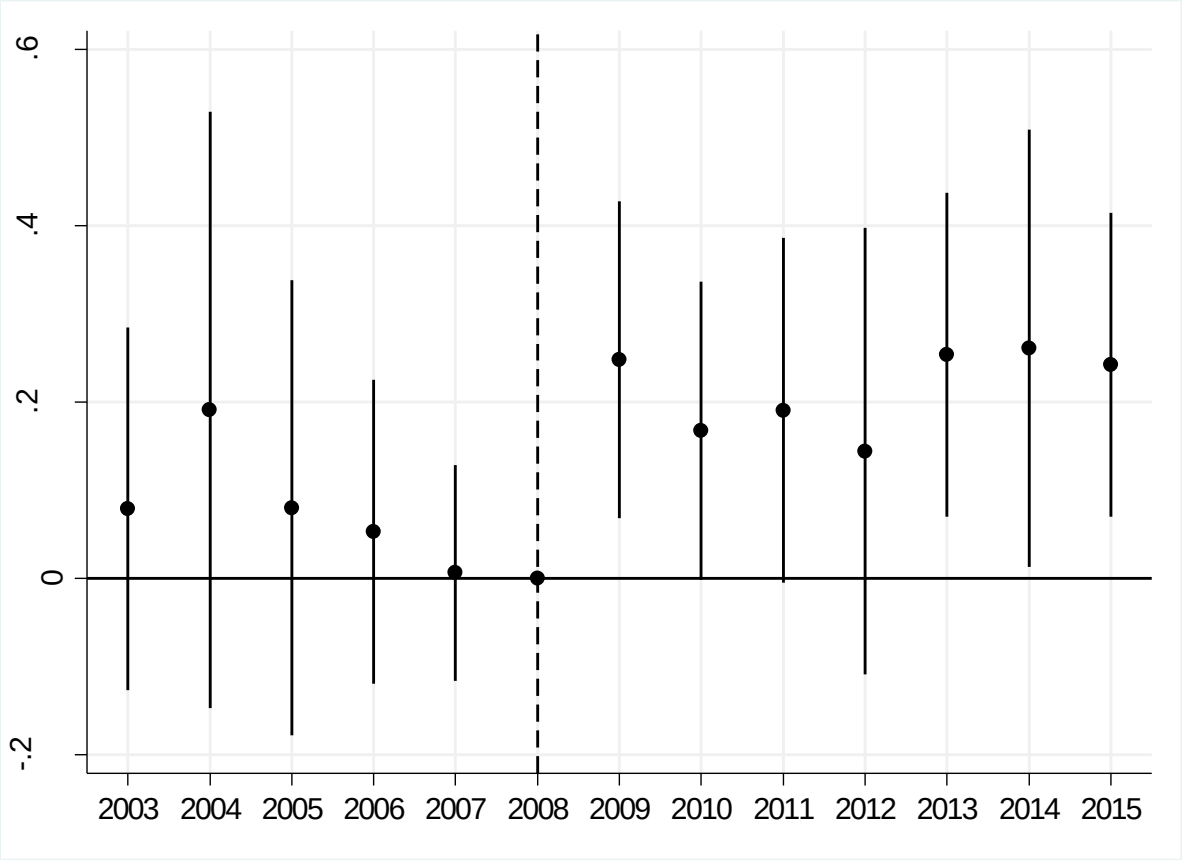
Traditional demand did not increase



Total unit sales per genre
o Romance, o Non-Romance

1) Do license deals increase?

Romance deals increase significantly



Notes: OLS estimates of annual differences in *LogSize* between the treatment group (romance authors) and the control group. The omitted year is 2008 – the first year of recorded e-reader ownership. Standard errors are clustered on the genre-level, and bars indicate 90% confidence bands.

Romance deals increase significantly

Table 2: Results: Changes in license deals

	(1) DV: Log(Size)		(2) DV: Size		(3) DV: Deal category	
Romance	-0.138**	(0.050)	-22.895**	(7.669)	-0.156**	(0.055)
After2008 × Romance	0.140***	(0.037)	31.990***	(8.202)	0.167***	(0.042)
Acclaimed	0.151***	(0.030)	27.112***	(4.486)	0.175***	(0.032)
Prev. bestseller	0.984***	(0.082)	201.510***	(12.368)	1.158***	(0.089)
Contested	0.670***	(0.070)	117.900***	(13.810)	0.766***	(0.081)
Debut	0.046	(0.054)	15.997	(10.790)	0.063	(0.062)
Orig. self-published	0.405*	(0.189)	92.313**	(33.069)	0.481*	(0.217)
Sequel	0.161***	(0.047)	26.972**	(11.614)	0.181**	(0.058)
Observations	14771		14771		14771	
R^2	0.541		0.410		0.526	

Notes: Editor, month-year fixed effects, and constant not reported.
Standard errors in parentheses, clustered on the genre-level. * $p<0.10$, ** $p<0.05$ *** $p<0.01$

TV/movie and international rights

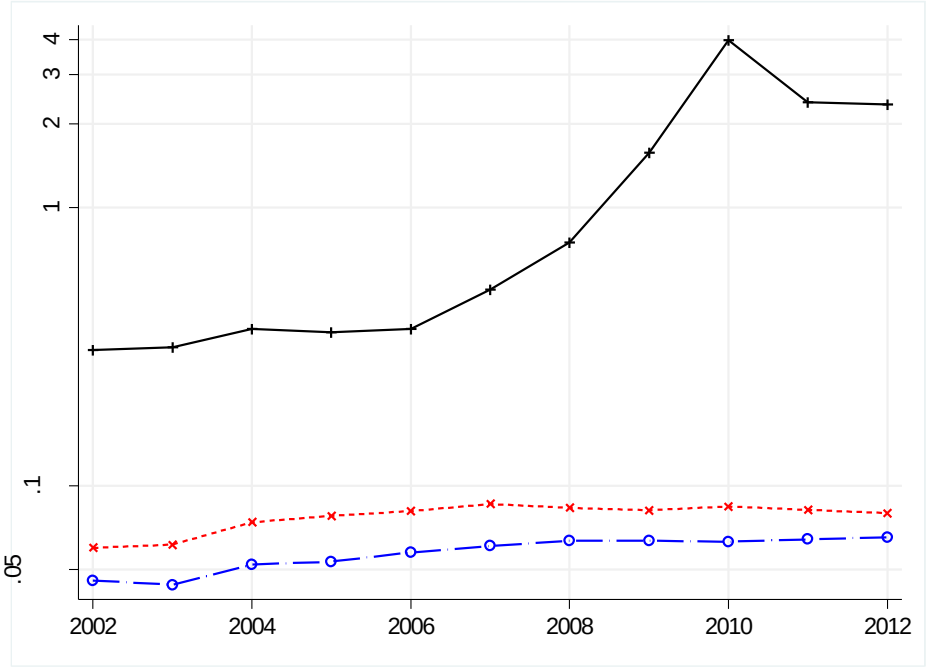
- No competition from self-publishing
- So we *shouldn't* see an effect there

Table 3: Results: Changes in rights deals (placebo exercises)

	(1) DV: Log(Size)	(2) DV: Size	(3) DV: Category
After2008 × Romance	-0.061	-2.487	-0.062
	(0.109)	(19.223)	(0.123)
Observations	8194	8194	8194
R^2	0.527	0.423	0.515

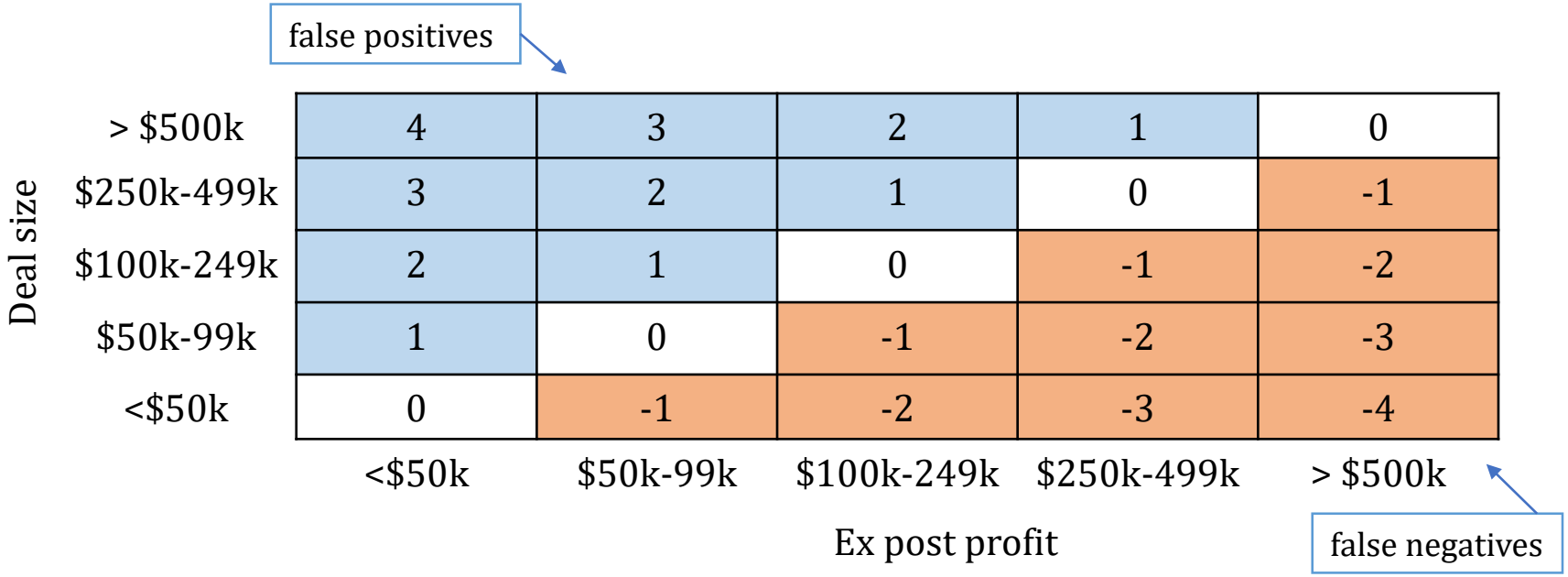
Notes: Editor and month-year fixed effects and coefficients of control variables not reported. Standard errors in parentheses, clustered on the genre-level. * $p<0.10$, ** $p<0.05$ *** $p<0.01$

2) Do predictions become more precise with additional entry?



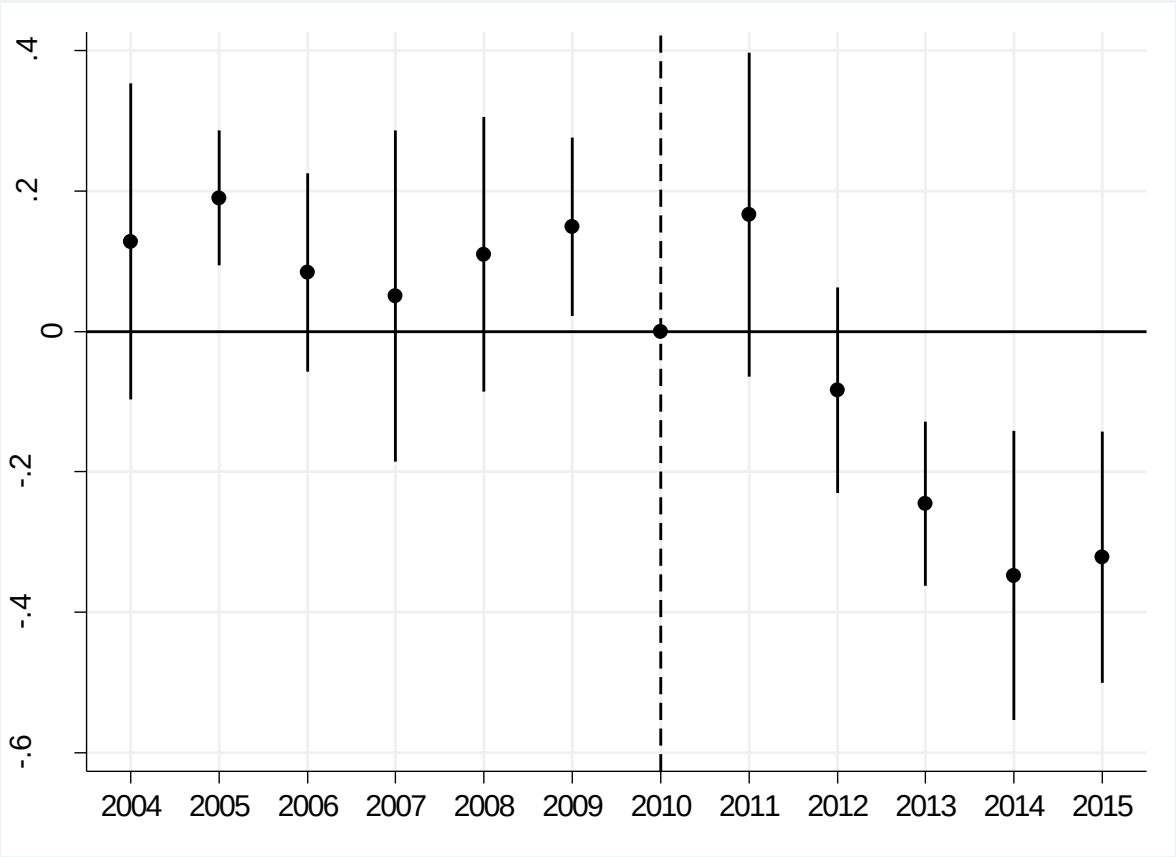
Defining prediction error

- We compare license deals (ex-ante predictions) to future profits (ex-post appeal)
  - (Anecdotal information: publisher profits are ~35% of revenues)



- The absolute value of this gives us the dependent variable in the next regressions

Prediction error decreases significantly



Notes: OLS estimates of annual differences in |*error*| between the treatment group (romance authors) and the control group. The omitted year is 2010 – the first year of significant differences in SP-supply. Standard errors are clustered on the genre-level, and bars indicate 90% confidence bands.

Publishers make fewer errors

Table 3: Results: Changes in predicting ex-post appeal

	(1) Abs(error)	(2) I(error)	(3) error^2	(4) False neg.	(5) False pos.
After 2010 × Romance	-0.236*** (0.070)	-0.091*** (0.011)	-0.643* (0.299)	-0.046** (0.017)	-0.045*** (0.011)
Observations	14771	14771	14771	14771	14771
R^2	0.336	0.380	0.231	0.076	0.396

Notes: Editor, month-year fixed effects, controls and constant not reported.
Standard errors in parentheses, clustered on the genre-level. * $p<0.10$, ** $p<0.05$ *** $p<0.01$

- Fewer "false negatives"
  - 81.9% decrease at the mean
  - could be due to a shift in bargaining power: authors get better deals
- Fewer "false positives"
  - 13.4% decrease at the mean
  - can't be explained by shifts in bargaining power

To summarize: the market for ideas

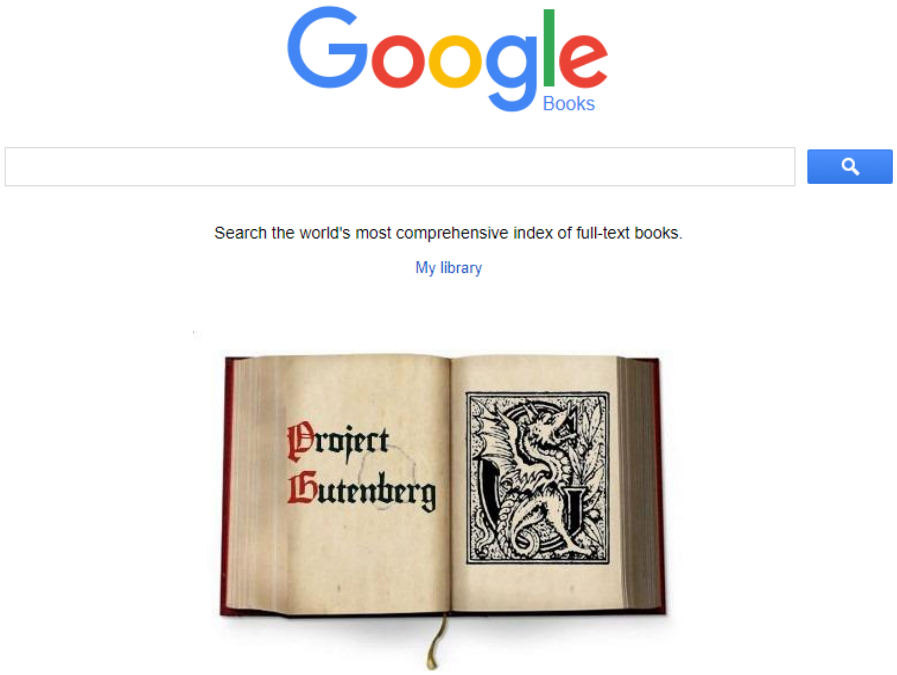
- Greater variety of available titles (at lower prices)
  - Could improve consumer welfare
- Larger license fees for authors
  - Increased incentives to produce?
- Better prediction of ex-post appeal
  - Reallocation of resources could benefit (almost) everyone
- **Self-publishing and traditional publishers as <u>complements</u>**

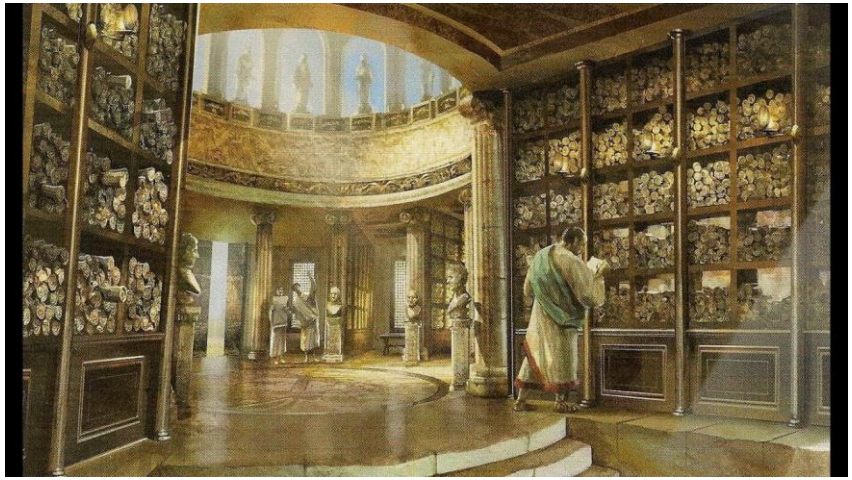
Digitization and the Demand for Physical Works: Evidence from the Google Books Project

With Abhishek Nagaraj (UC Berkeley)

Large-scale digitization efforts

- Digitization has expanded access to existing works
- Possibility of a (searchable) repository of all knowledge in digital form





Large-scale digitization efforts

- Digitization has expanded access to existing works
- Possibility of a (searchable) repository of all knowledge in digital form

No. 15-849

IN THE

Supreme Court of the United States

THE AUTHORS GUILD, *et al.*,
Petitioners,

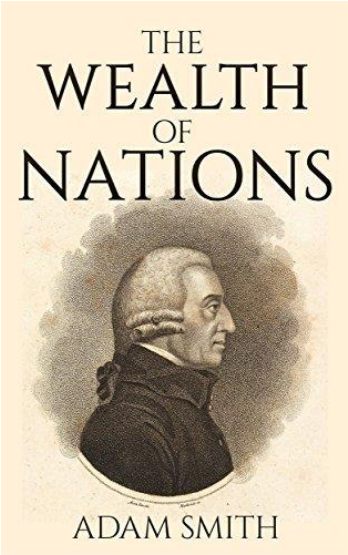
v.

GOOGLE INC.,
Respondent.

ON PETITION FOR A WRIT OF CERTIORARI TO THE
UNITED STATES COURT OF APPEALS
FOR THE SECOND CIRCUIT

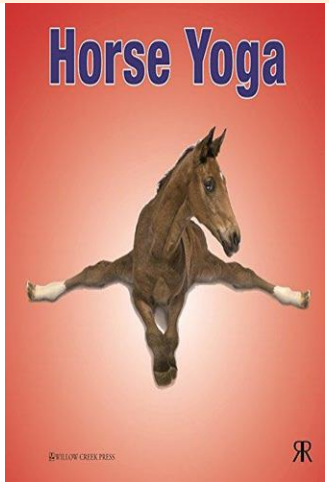
Key question

What is the impact of digitization on demand for physical works?



Digitization as a substitute for physical copies – it decreases demand

- Especially for popular books



Digitization could enable discovery – it increases demand

- Especially for obscure books

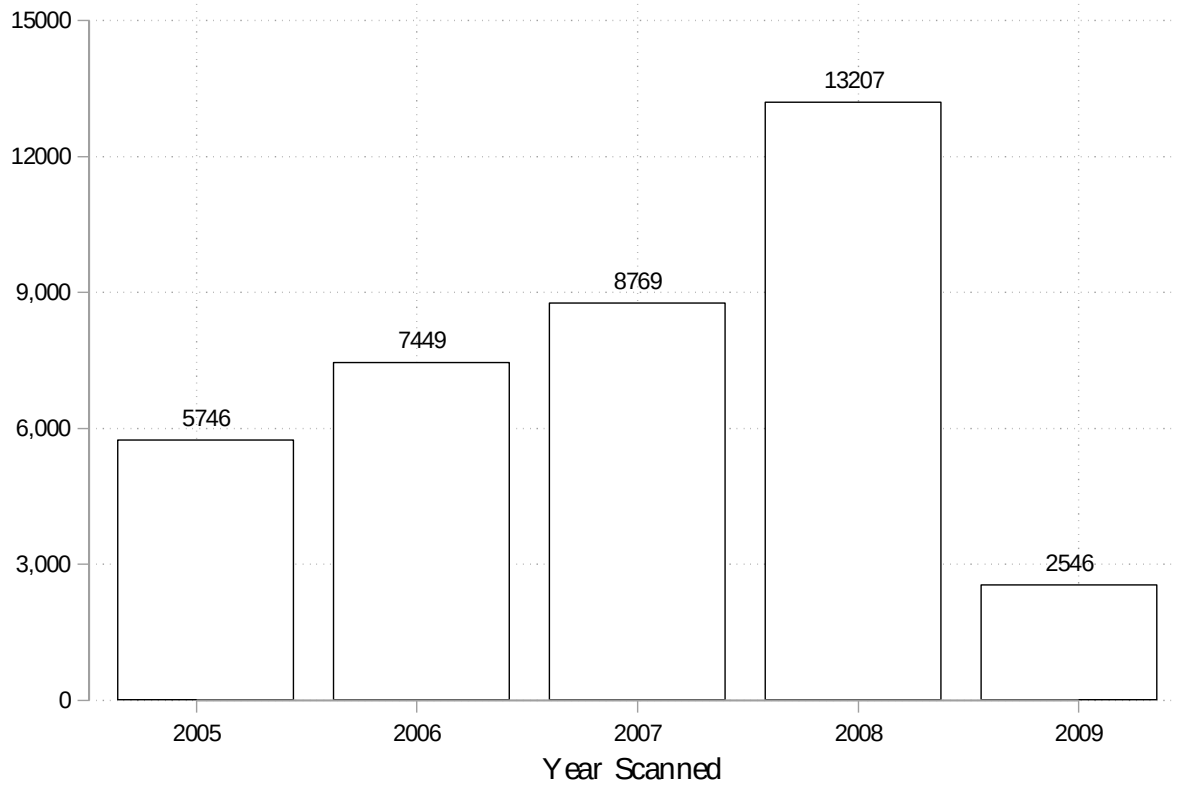
NOTE: most of the digitized books are “old” and “academic”

Empirical setting and data

A natural experiment: Harvard libraries

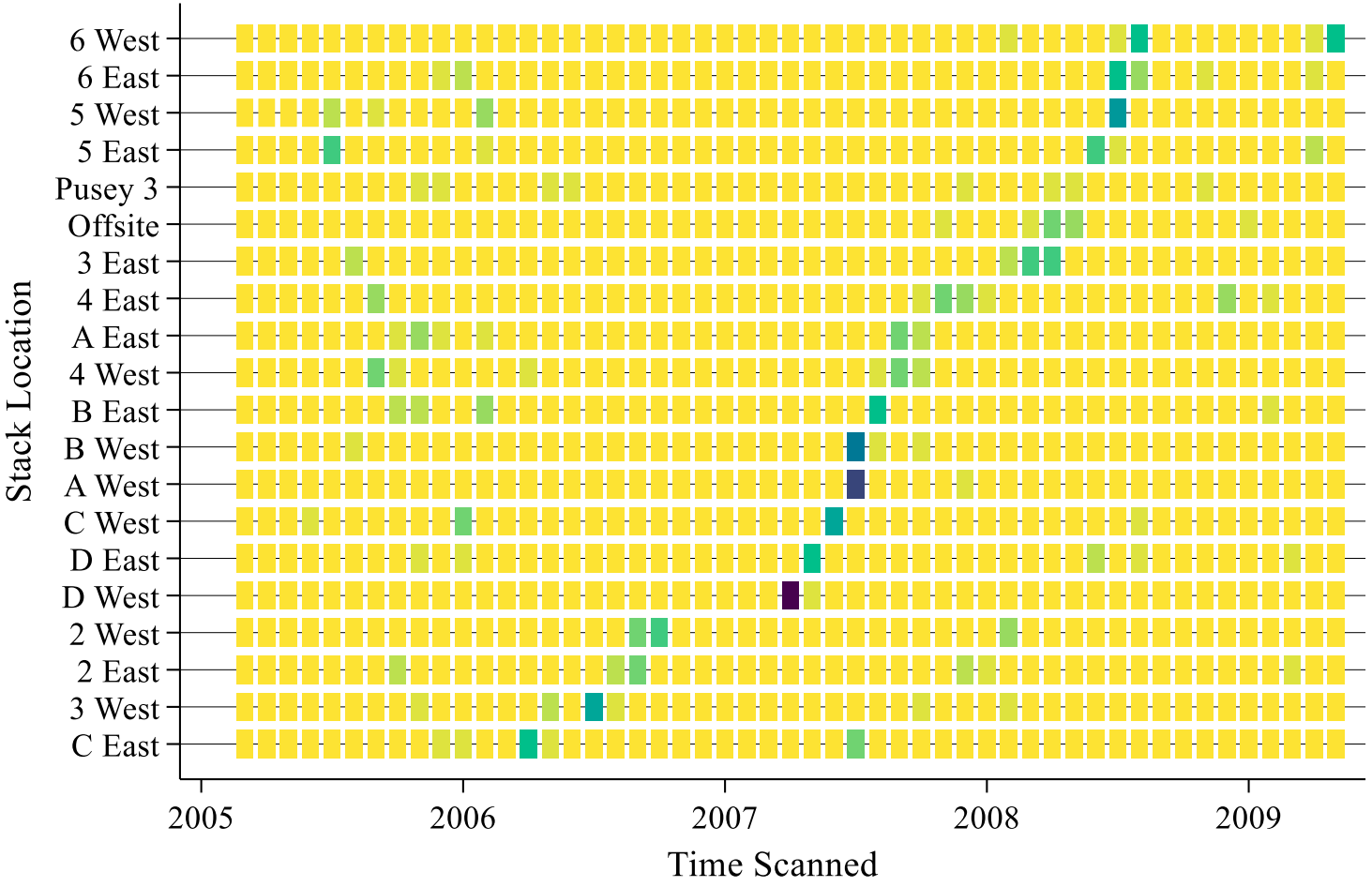
- Harvard was among the first few libraries to join forces with Google Books
  - Google Books digitized all *out-of-copyright* works from Harvard's Widener library
  - (digitization – and searchability – of entire books, rather than snippets)
  - 43% of titles were digitized
- Digitization of Harvard's catalog was labor- and time-intensive
  - It took (at least) five years, from 2005 to 2009
- Books were digitized on a *shelf-by-shelf* basis
  - Convenience, rather than popularity

Digitization was time-intensive



37,743 books were digitized, 50,263 were not

Shelf-by-shelf digitization



Digitization for most locations was very concentrated

Data: sales, loans, and availability

We observe sales for ~9200 books from the Harvard libraries from 2003 to 2011

- **Loans**: Harvard Widener library
  - Digitization date through borrower codes
  - ~88,000 books have at least 1 loan
  - 0.25 loans per year

- **Sales:** NPD (Nielsen) Bookscan
  - All titles English-language titles with >3 loans
  - 802 sales per year
  - (median annual sales are 0)

- **Availability:** Bowker Books-in-Print
  - All ISBNs for each title
  - 1.08 new editions per year

Estimation and results

Research design

- How did demand change for digitized books
- compared to those that are not (yet) digitized

$$sales_{it} = \alpha \times PostScan_{it} + \gamma_i + \mu_t + \epsilon_{it}$$

- $PostScan_{it} = 1$ in years after book i has been digitized
- (Book and year fixed effects)

- Estimated via Poisson and linear probability models
  - (but OLS works too)

Impacts on sales

Estimates for the impact on sales by popularity

	Sales (Poisson)	**Any Sales (LPM)**	**Sales (Poisson)**	**Any Sales (LPM)**
Post-scanned	0.297*	0.078***	0.349**	0.067***
	(0.153)	(0.005)	(0.190)	(0.006)
Post-scanned × popular			-0.201	0.024***
			(0.221)	(0.009)
Book FEs	✓	✓	✓	✓
Year FEs	✓	✓	✓	✓
N	82,836	82,836	82,836	82,836

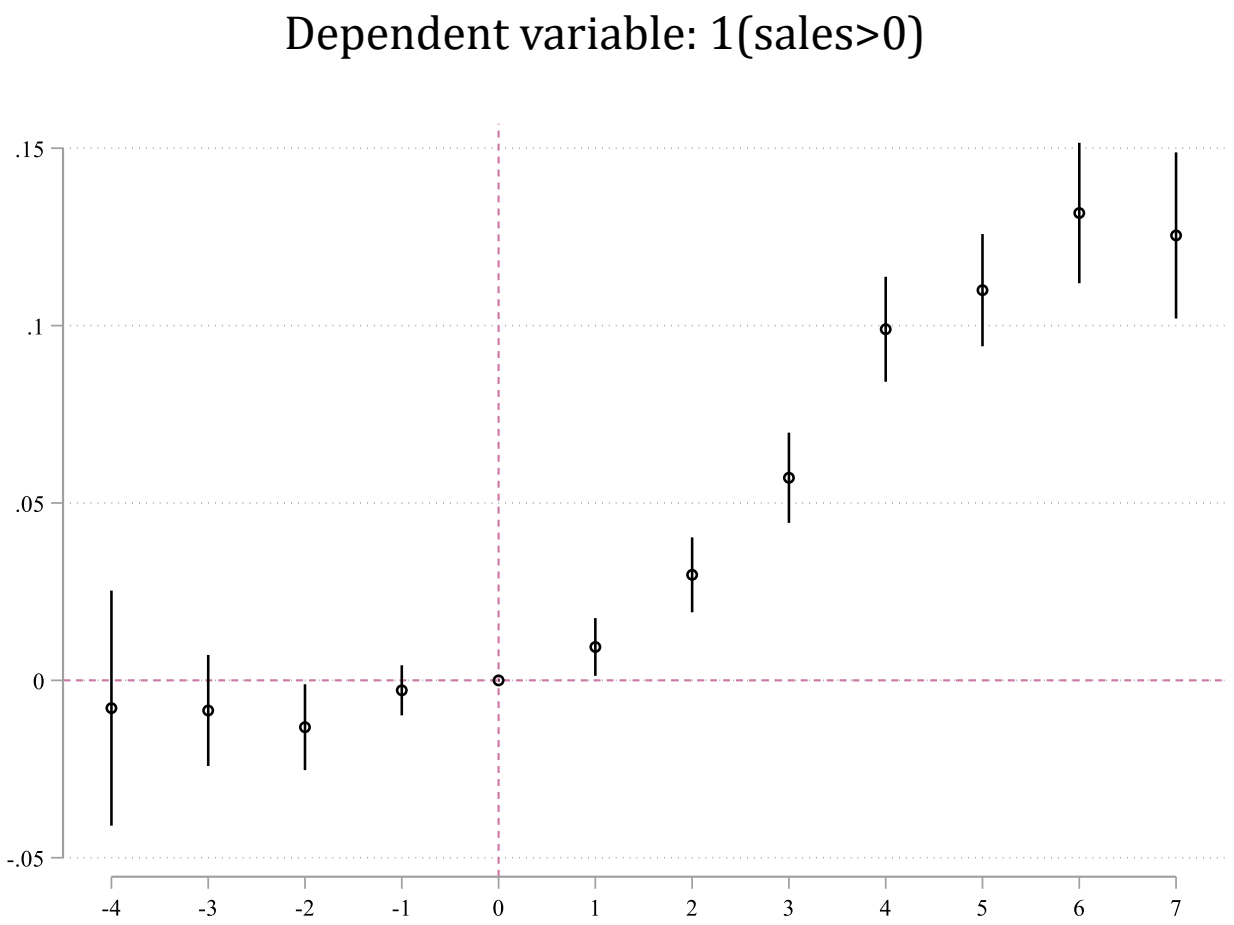
- "Popular" books: checked out at least once in 2003 + 2004

Summarizing the coefficients

Impact on sales:

- Digitization increased sales by about 34%
- It increased the likelihood of a sale by 7.8 percentage points
    - Or 92% at the mean
- 16% increase for popular books; 42% for less popular books

Estimated effects over time: sales

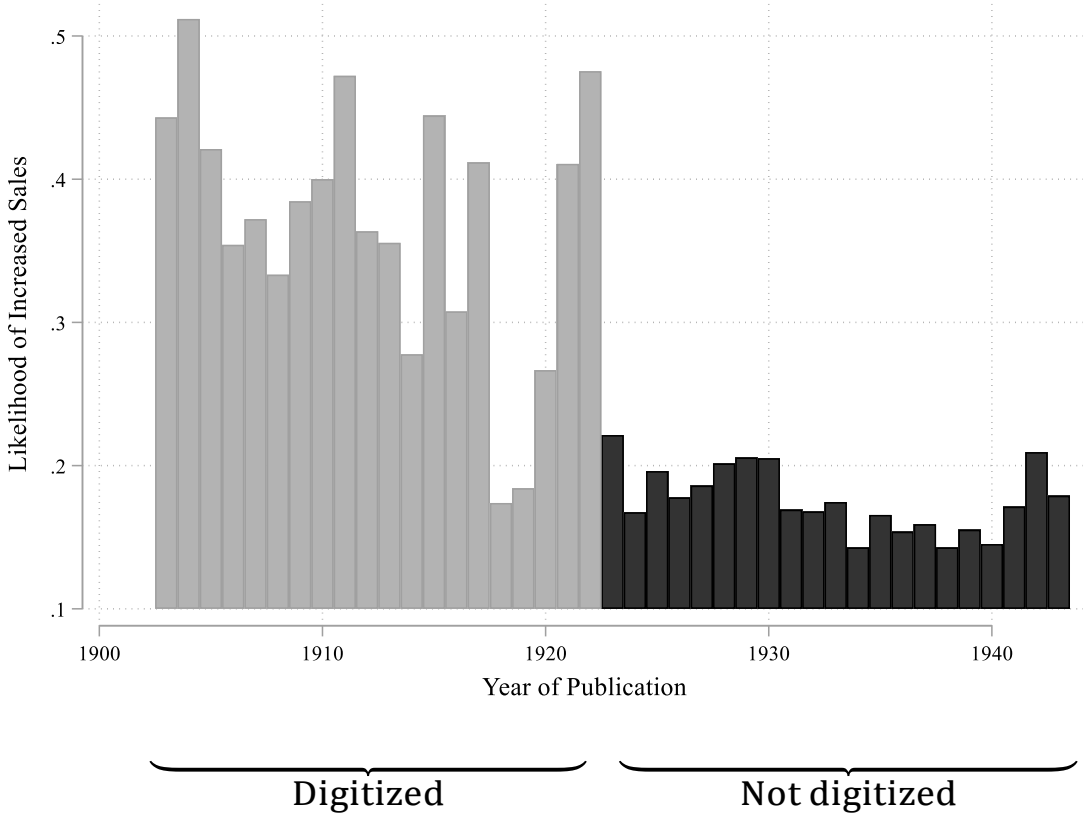


Discovery, or availability?

- **Did digitization lead to an increase in in-print editions?**
  - Yes, the prob. of being available increases by 19 percentage points
- **Did digitization lead to decreases in prices?**
  - No evidence of this
- **Are the impacts driven by these changes?**
  - No. All effects remain strong after controlling for availability

Additional evidence: sales by publication date

% of books from each vintage that had more sales in 2010/11 than in 2003/04



Likelihood of *increased sales* is much <u>larger</u> for digitized books

To summarize: demand for physical editions

- Digitization can increase physical demand under two conditions
  - A book is not otherwise well known
  - Consumers like physical versions
- **This seems to be the case for most works**

- Caveat: we only observe digitization of a specific set of books
  - We can't say much about recent bestsellers

To summarize: effects of digitization

- Digitization brings about information that…
  - Can help allocate resources to the best books
  - Can increase awareness of existing works

It's not all bad for traditional institutions

Thank you!