

Editors Influence Topics of Papers Published in AER

Ali Sina Önder¹ Sergey V. Popov² Sascha Schweitzer¹

¹University of Bayreuth

²Cardiff Business School

Palma, September 15, 2017

Paper Submissions

- Authors write papers and submit them to journals
- Authors want to get published, editors evaluate their publications
- What gets published?
- Much literature re citations, affiliations, etc.

Paper Submissions

- Authors write papers and submit them to journals
 - Authors want to get published, editors evaluate their publications
 - What gets published?
 - Much literature re citations, affiliations, etc.
- Q: Is there a selection based on topics?

Paper Submissions

- Authors write papers and submit them to journals
- Authors want to get published, editors evaluate their publications
- What gets published?
- Much literature re citations, affiliations, etc.

Q: Is there a selection based on topics?

A: Assign an editor, get a bias *towards* editor's topics.

Simple Model of Topic Selection

- Two topics, 1 and 2.
- π_i - prob of a paper of quality $q = 1$, $1 - \pi_i$ get quality $q = 0$.
- Editors see an estimate of quality $\tilde{q} = q + \varepsilon$, noise uniform on $[-b_i, b_i]$, publish if $\tilde{q} > \hat{q} \in (0, 1)$.

Prob to publish a paper of quality q on topic i :

$$P(q + \varepsilon > \bar{q}) = P(\varepsilon > \bar{q} - q) = \frac{b_i - (\bar{q} - q)}{2b_i}$$

Amount of papers of papers on Topic i :

$$(1 - \pi_i) \overbrace{\frac{b_i - (\bar{q} - 0)}{2b_i}}^{\text{bad paper is published}} + \pi_i \overbrace{\frac{b_i - (\bar{q} - 1)}{2b_i}}^{\text{good paper is published}} = \frac{1}{2} - \frac{\pi_i - \bar{q}}{2b_i}.$$

Result

Amount of papers of papers on Topic i :

$$(1 - \pi_i) \overbrace{\frac{b_i - (\bar{q} - 0)}{2b_i}}^{\text{bad paper is published}} + \pi_i \overbrace{\frac{b_i - (\bar{q} - 1)}{2b_i}}^{\text{good paper is published}} = \frac{1}{2} - \frac{\pi_i - \bar{q}}{2b_i}.$$

- Increase in b_i leads to more papers of topic i accepted if $\pi_i > \bar{q}$, decrease otherwise.

=> same with share of topic i .

Change in editor = change in competence = change in b_i .

Result

Amount of papers of papers on Topic i :

$$(1 - \pi_i) \overbrace{\frac{b_i - (\bar{q} - 0)}{2b_i}}^{\text{bad paper is published}} + \pi_i \overbrace{\frac{b_i - (\bar{q} - 1)}{2b_i}}^{\text{good paper is published}} = \frac{1}{2} - \frac{\pi_i - \bar{q}}{2b_i}.$$

- Increase in b_i leads to more papers of topic i accepted if $\pi_i > \bar{q}$, decrease otherwise.

=> same with share of topic i .

Change in editor = change in competence = change in b_i .

There can be biases, too!

Overview

- People use JEL codes, keywords, title/abstract words,..

Step 1 Topic analysis gives us topics

- Count words in papers (look at word combinations too)
- Latent Dirichlet Allocation of words in the document
- MALLET package, prepared for that purpose by computer scientists
- Get topic loadings in documents

Step 2 See if AER's loadings change correlates with editor's loadings

Latent Dirichlet Allocation

- Topic is a bag of words with a probability attached to each word.
- Topics have a probability distribution over them
- Article is a bag of words:
 - use Dirichlet distribution to pick an ex-ante distribution of topics
 - pull words according to distribution over words from topics
 - get the article
- The number of topics is picked by the researcher, allocation of words to topics and probabilities is data-driven
 - Gibbs sampling: start with “stupid” topics, update probabilities iteratively

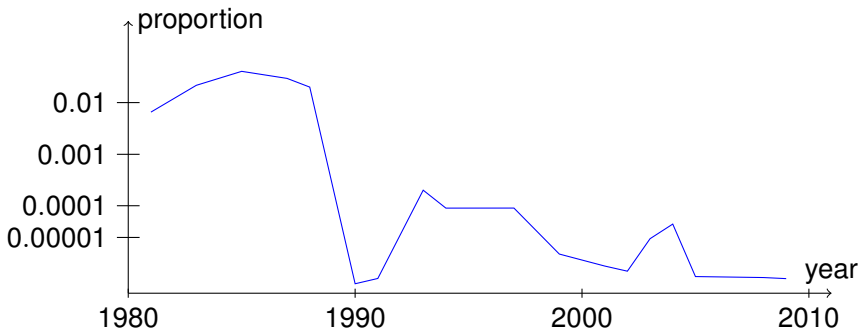
Texts

For each editor assigned in year t : Six documents:

- AER texts from $[t - 2, t + 1]$ “before tenure”
- AER texts from $[t + 2, t + 5]$ “during tenure”
- Rest of top-5 texts from $[t - 2, t + 1]$ “before tenure”
- Rest of top-5 texts from $[t + 2, t + 5]$ “during tenure”
- Editor’s texts from $[t - 2, t + 1]$ “before tenure”
- Editor’s texts from $[t + 2, t + 5]$ “during tenure”

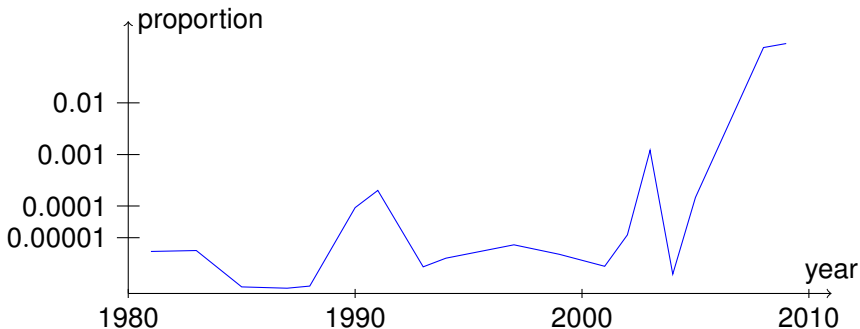
Publication lag of 2 years, window of 4 years.

Topic Development



*result present profit analysi demand assum economi
product factor inflat growth assumpt privat requir competit
includ distribut run standard*

Topic Development



*incom labor capit problem present consid industri analysi
level higher section report research test theori two degre
economist demand*

Topics

*level social model rule rate competit growth impli measur
econom behavior function indic secur account like evid
decad limit*

*model estim time employ new invest innov product
economy approach network show busi crisi jel state
financi vari make*

- Could be a research topic.
- Could be verbiage.
- No way to know.

Topics

*level social model rule rate competit growth impli measur
econom behavior function indic secur account like evid
decad limit*

*model estim time employ new invest innov product
economi approach network show busi crisi jel state
financi vari make*

- Could be a research topic.
- Could be verbiage.
- No way to know. Does not matter.

Extracting Editor's Idiosyncratic Tastes

Multicollinearity between variables. Obtain editor's idiosyncratic preferences before:

$$Edit_{i,before}^C = \alpha_0 + \alpha_1 AER_{i,before}^C + \alpha_2 BM_{i,before}^C + \pi_i^C$$

Then look at how they change:

$$\Delta Edit_i^C = \beta_0 + \beta_1 \Delta AER_i^C + \beta_2 \Delta BM_i^C + \phi_i^C$$

Then try to predict changes in editor's topics:

$$Edit_{i,during}^C = \gamma_0 + \gamma_1 \hat{\pi}_i^C + \gamma_2 \hat{\phi}_i^C + \psi_i^C$$

Keep $\widehat{Edit}_{i,during}^C$: not explained by values or changes in AER or benchmark.

Change in AER Topics During

| | (21) | (22) | (6) | (23) |
|--------------------|-------------------------|-------------------------|----------------------|----------------------|
| | ΔAER | ΔAER | $\Delta Benchmark$ | $\Delta[AER - Bm]$ |
| Editor's Pref | 0.00597*** (0.00152) | 0.00581*** (0.00150) | 0.00290 (0.00205) | 0.00306 (0.00232) |
| <i>Benchmark</i> | | | | |
| $\Delta Benchmark$ | | 0.0552** (0.0174) | | |
| Start.Year.FE | Yes | Yes | Yes | Yes |
| Topic.FE | Yes | Yes | Yes | Yes |
| Observations | 4200 | 4200 | 4200 | 4200 |
| R^2 | 0.137 | 0.141 | 0.0716 | 0.0787 |
| F | 2.153 | 2.220 | 1.718 | 1.777 |

Standard errors in parentheses. ⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

AER Topics During

| | (18) | (19) | (20) |
|---------------------------|------------|------------|------------------|
| | <i>AER</i> | <i>AER</i> | <i>Benchmark</i> |
| Editor's Pref | 0.176*** | 0.137** | 0.0955 |
| | (0.0412) | (0.0447) | (0.0649) |
| <i>Benchmark</i> | | 9.561*** | |
| | | (1.398) | |
| Δ <i>Benchmark</i> | | | |
| Start.Year.FE | Yes | Yes | Yes |
| Topic.FE | Yes | Yes | Yes |
| Observations | 5600 | 5600 | 5600 |
| R^2 | | | |
| F | | | |

Standard errors in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Conclusion

- We find that editors steer words in published papers
 - $\approx +6$ words per 1000 for very narrow editors
 - if papers are all same length, +6 papers per 1000
 - like half paper per year
- Does not have to be good or bad
- What to do:
 - citations of papers affiliated with topics?..
 - use paper texts instead of abstracts
 - better title