Leveraging growth modeling to examine topic usage over time

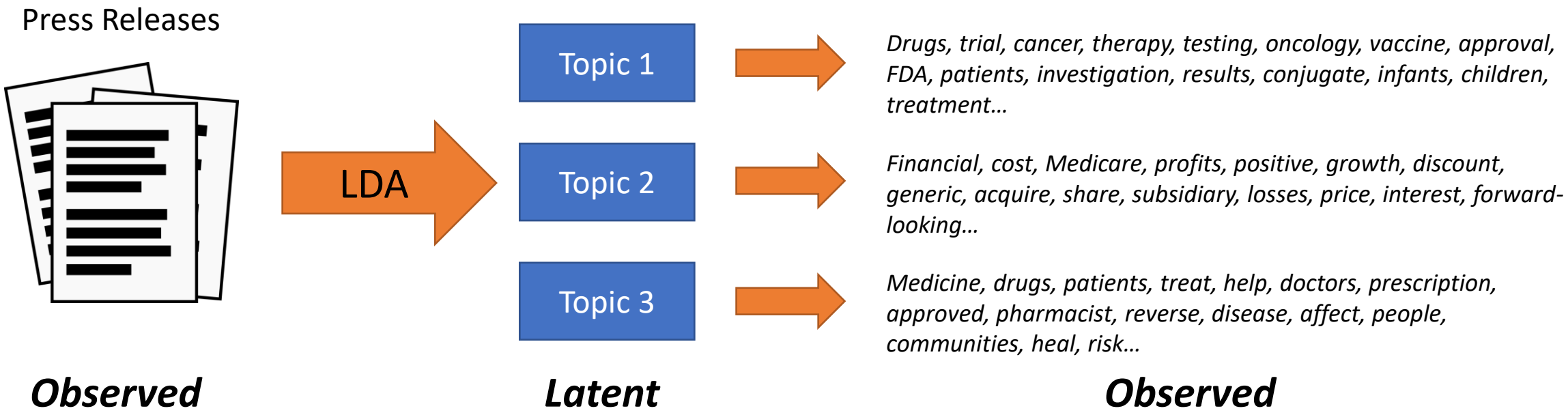
Miles A. Zachary

Auburn University

HARBERT | inspiring BUSINESS

Topic modeling

- **Topic modeling** uses statistical associations of words in a collection of texts (i.e., a corpus) to generate latent topics
  - Uses an algorithm to uncover clusters of co-occurring words that jointly represent higher-order concepts (e.g., novelty or innovation)



Topics over time?

Three Similarities

1. Each is regarding constructs directly or indirectly measured through rhetoric
2. Focus is on general rather than specific rhetoric
3. Interested in examining changes in generalized rhetoric over time

- What if you wanted to exami
- For example:
  - How do perceptions of a company's work environment change following the adoption of new policies and procedures?
  - How does leader rhetoric regarding social justice issues change in response to social movements like BLM or Time's Up?
  - How do corporate values change over time? Are changes to corporate values punctuated by CEO and/or board changes?

Topics over time? – A possible solution

Company values statement for company *j* at founding

LDA

Topic 1

Topic 2

Topic 3

Shareholders, expect, growth, increasing…

Costs, risk, below, careful, expansion…

Employees, people, valued, inclusive…

Topic 1

Topic 2

Topic 3

Dictionaries

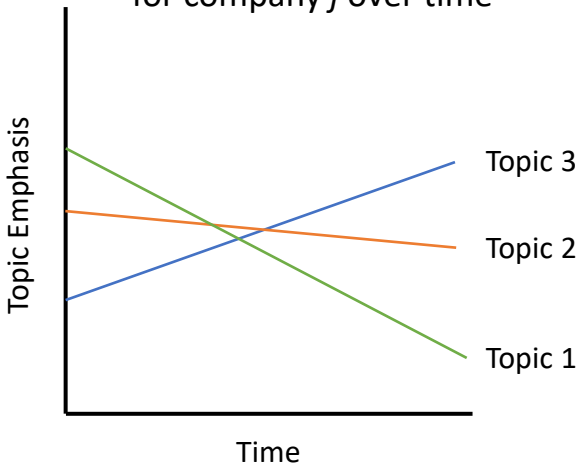
CATA

Company values statements for company *j* from founding to 5 years after founding

$$Topic_{ij} = [\beta_{00} + \beta_{10}(Time_{ij})] + [u_{0j} + u_{1j}(Time_{ij}) + r_{ij}]$$

Growth Modeling using RCM

Change in company values topics for company *j* over time



Topics over time? – An example

- Companies communicate information to stakeholders about themselves to project a desired image (Elsbach, 1994; 2003, Gioia & Thomas, 1996)
- RQs: How does the IPO process change projected image? How do changes in ownership influence such changes?
- Data
  - 168 IPO firms between 2009-2012
  - “About Us” website information -5 years before to +5 years following IPO
  - Founder/VC ownership collected from IPO prospectuses, proxy statements, and annual reports
- Analysis
  - Topic models & CATA in R using *tm* and *topicmodels* packages
  - Discontinuous growth modeling using *mixed* in Stata

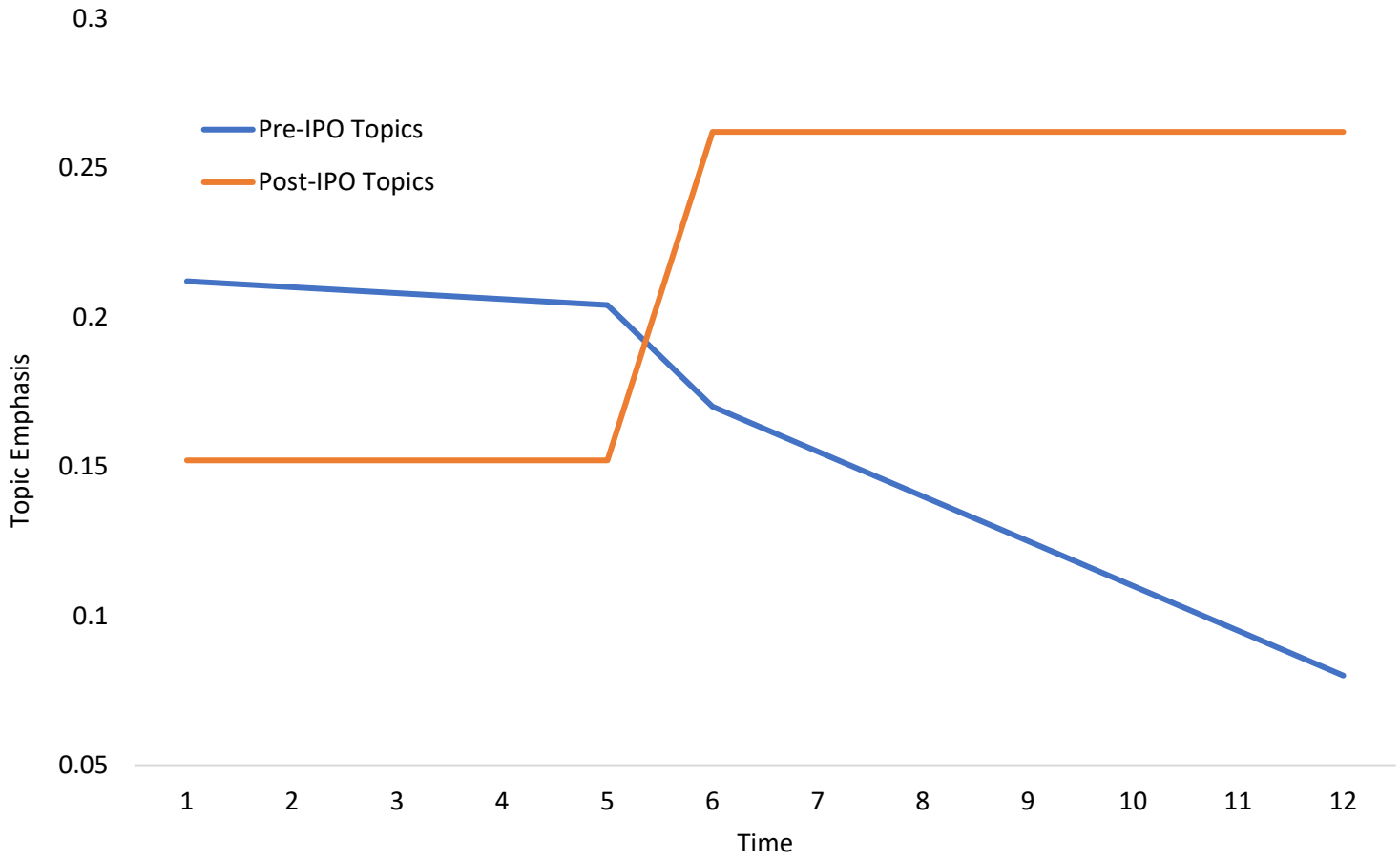
Topics over time? – An example

Content analysis code structure overview

1. Topic model pre-IPO text data for firm *j* (3 topics)
2. Store top 25 words per topic (3 pre-IPO topic word lists)
3. Topic model post-IPO data for firm *j* (3 topics)
4. Store top 25 words per topic (3 post-IPO topic word lists)
5. CATA analyze entire corpus for firm *j* using all 6 word lists
6. Repeated steps 1-5 for the remaining 167 IPO firms
7. Export to .dta using write.dta (*foreign* library)

	ticker	year	ipoyear	topic	time	time_ipo	time_post	pretopic_std	posttopic_~d
1	ACOM	2004	2009	1	0	0	0	.118705	.1151079
2	ACOM	2005	2009	1	1	0	0	.2201835	.1100917
3	ACOM	2006	2009	1	2	0	0	.2230769	.0923077
4	ACOM	2007	2009	1	3	0	0	.2610442	.1084337
5	ACOM	2008	2009	1	4	0	0	.2610442	.1084337
6	ACOM	2009	2009	1	5	1	0	.1153846	.025641
7	ACOM	2010	2009	1	6	1	1	.0642202	.0458716
8	ACOM	2011	2009	1	7	1	2	.0566038	.0314465
9	ACOM	2012	2009	1	8	1	3	.0897436	.4102564
10	ACOM	2013	2009	1	9	1	4	.0503597	.1798561
11	ACOM	2004	2009	2	0	0	0	.1978417	.0539568
12	ACOM	2005	2009	2	1	0	0	.1284404	.0642202
13	ACOM	2006	2009	2	2	0	0	.1384615	.0769231
14	ACOM	2007	2009	2	3	0	0	.1084337	.0803213
15	ACOM	2008	2009	2	4	0	0	.1084337	.0803213
16	ACOM	2009	2009	2	5	1	0	.0641026	.1153846
17	ACOM	2010	2009	2	6	1	1	.0366972	.2293578
18	ACOM	2011	2009	2	7	1	2	.0314465	.2012579
19	ACOM	2012	2009	2	8	1	3	.0384615	.1410256
20	ACOM	2013	2009	2	9	1	4	.028777	.0791367
21	ACOM	2004	2009	3	0	0	0	.0863309	.0647482
22	ACOM	2005	2009	3	1	0	0	.2477064	.146789
23	ACOM	2006	2009	3	2	0	0	.2384615	.1307692
24	ACOM	2007	2009	3	3	0	0	.2570281	.124498
25	ACOM	2008	2009	3	4	0	0	.2570281	.124498
26	ACOM	2009	2009	3	5	1	0	.1410256	.1025641
27	ACOM	2010	2009	3	6	1	1	.1100917	.3119266
28	ACOM	2011	2009	3	7	1	2	.1257862	.2767296
29	ACOM	2012	2009	3	8	1	3	.0769231	.0897436
30	ACOM	2013	2009	3	9	1	4	.057554	.0863309

Topics over time? – An example



Topics over time? – An example

- Other findings
  - Projected image changes more quickly when founders and VCs retain less ownership following IPO
  - Changes to projected image at IPO are positively associated with analyst recommendations
  - Retaining pre-IPO rhetoric is negatively associated with market-to-book ratio

Topics over time?

- Things to consider
  - Why does it matter?
    - Be able to argue why rhetorical change matters; link to important outcomes
  - Text data! A lot of longitudinal text data
    - Sources: Company reports (e.g., annual reports, shareholder letters), social media, Glassdoor, Internet Archive
    - Garbage in, garbage out
  - Some basic R programming skills
    - Familiarize yourself with general procedures and topic modeling libraries
  - Explore different timeframes and topic configurations
    - Topic modeling and growth models are sensitive to cut-offs
    - Try different combinations to better understand rates of change