DPIR SPRING SCHOOL

Scaling models

13 April 2016

Andy Eggers

Suppose we had a **document-term matrix** like this.

	Word 1	Word 2	Word 3	…
Article 1	0	14	2	…
Article 2	1	8	0	…
Article 3	0	7	1	…
Article 4	2	3	0	…
…	…	…	…	…

What can we learn from it?

Data structure

The document-term matrix comes from raw texts.

A lot of information is lost!

Bag of words. (Bag of features.)

	these	that	those
Article 1	0	14	2
Article 2	1	8	0
Article 3	0	7	1
Article 4	2	3	0

Article	Word	Count
1	these	0
2	these	1
3	these	0
4	these	2
1	that	14
2	that	8
3	that	7
4	that	3
1	those	2
2	those	0
3	those	1
4	those	0

Article 1: That that “that” that. That that those; that that that that. That that that that those.

Article 2: That that/these that that! That that that.

Article 3: That that those — that that that that that.

Article 4: These that! These that that.

A dimension-reduction approach

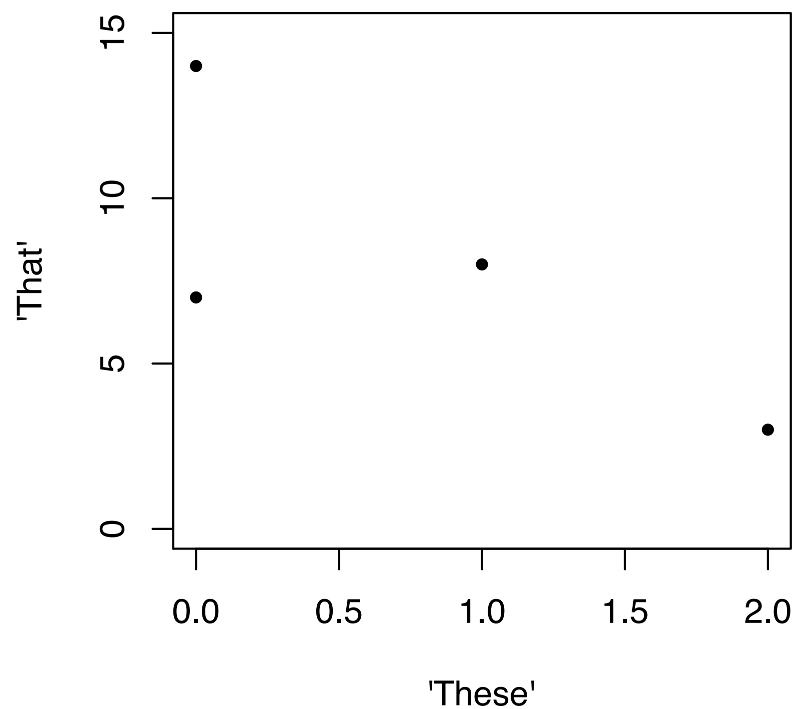
	these	that	those
Article 1	0	14	2
Article 2	1	8	0
Article 3	0	7	1
Article 4	2	3	0

Think about plotting the word frequencies in space, starting with just two words.

Now draw a line that is as close to the points as possible by some metric.

Finally, "project" points onto line: you have reduced two dimensions to one.

Imagine with three dimensions, or 10,000!



Toward a modeling approach: simpler example

Imagine we just had one word and a covariate.

	Word 23	x (Ideology score)
Legislator 1	6	34
Legislator 2	0	67
Legislator 3	3	49
Legislator 4	8	12
…	…	…

Q: How could you relate ideology to word frequency?

What would this tell you?

A: Regress frequency on x (OLS or Poisson/negative binomial regression):

$$\lambda_i = \alpha + \beta x_i$$

or

$$\lambda_i = e^{\alpha + \beta x_i}$$

A slightly less simple example

How could we extend this to more than one word?

	Word	Frequency	x (Ideology score)
Legislator 1	1	3	34
Legislator 2	1	2	67
Legislator 3	1	7	49
Legislator 4	1	20	12
…	…	…	…
Legislator 1	2	2	34
Legislator 2	2	1	67
Legislator 3	2	0	49
Legislator 4	2	1	12
…	…	…	…

Same thing, but allow for

- different intercept for each legislator
- different intercept for each bill
- different slope for each bill

$$\lambda_{ij} = \alpha_i + \psi_j + \beta_j x_i$$

or

$$\lambda_{ij} = e^{\alpha_i + \psi_j + \beta_j x_i}$$

Doing the seemingly impossible

Now suppose the ideology score was missing. What now?

	Word	Frequency	x (Ideology score)
Legislator 1	1	3	?
Legislator 2	1	2	?
Legislator 3	1	7	?
Legislator 4	1	20	?
…	…	…	…
Legislator 1	2	2	?
Legislator 2	2	1	?
Legislator 3	2	0	?
Legislator 4	2	1	?
…	…	…	…

Statistical model is the same as if x was observed, but x becomes an additional parameter to estimate.

Why does this work? Because the same legislator is observed for many words, and x is assumed to affect frequency of many words.

Suppose n legislators and k bills: how many parameters are we estimating?

Scaling text: wordfish

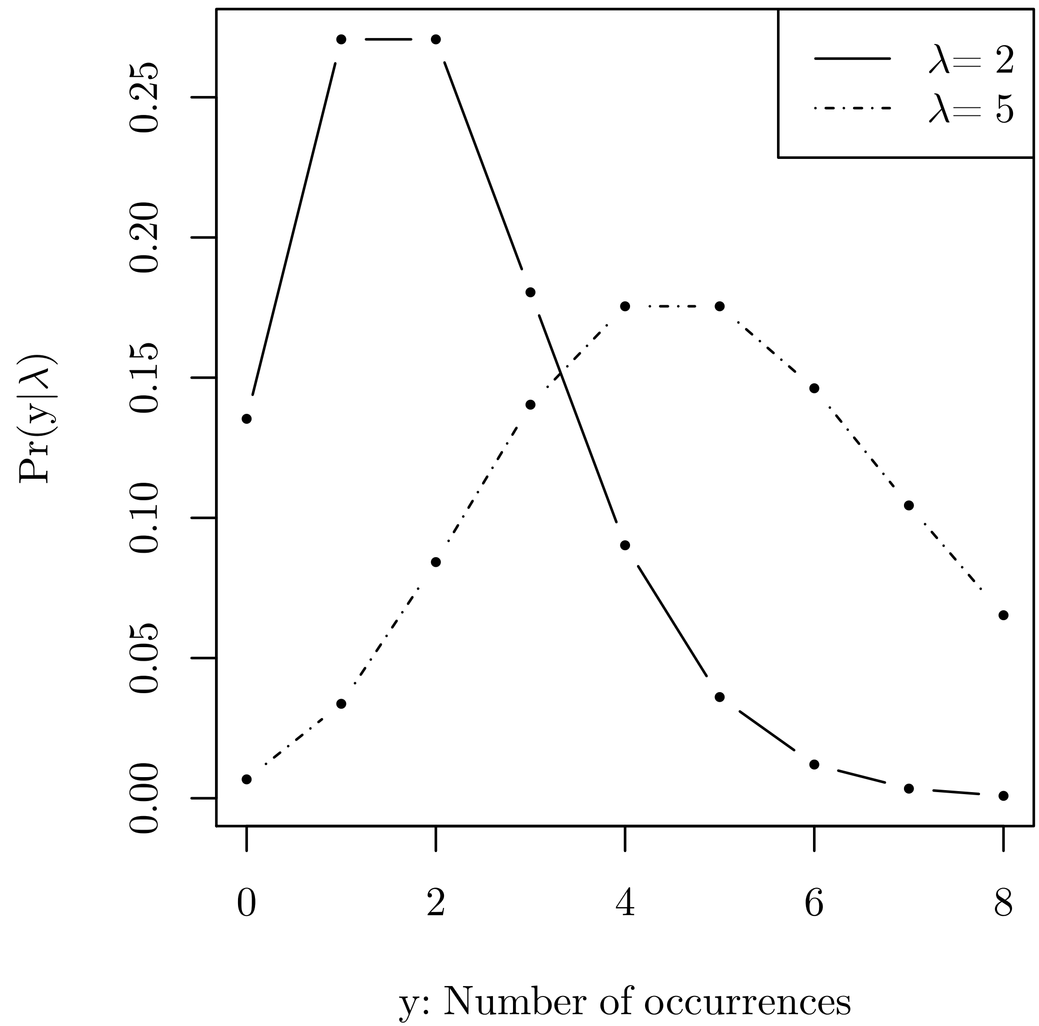
Think about rates at which parties use different words in party manifestos.

Consider this model for the rate λ for party i using word j at time t:

$$\lambda_{ijt} = e^{\alpha_{it} + \psi_j + \beta_j \omega_{it}}$$

where

- α_{it} is party-year fixed effect
- ψ_j is word fixed effect
- β_j is word weight, i.e. discrimination parameter
- ω_{it} is party i's position in year t



Scaling text: wordfish

Consider this model for the rate λ for party i using word j at time t:

$$\lambda_{ijt} = e^{\alpha_{it} + \psi_j + \beta_j \omega_{it}}$$

where

- α_{it} is party-year fixed effect
- ψ_j is word fixed effect
- β_j is word weight, i.e. discrimination parameter
- ω_{it} is party i's position in year t

Consider the words "and" and "deficit".

Q: What values of ψ_j and β_j would you expect for these words?

A: For the word "and":

- high ψ_j, because it is a common word
- small (in magnitude) β_j because its frequency is not likely to differ between parties

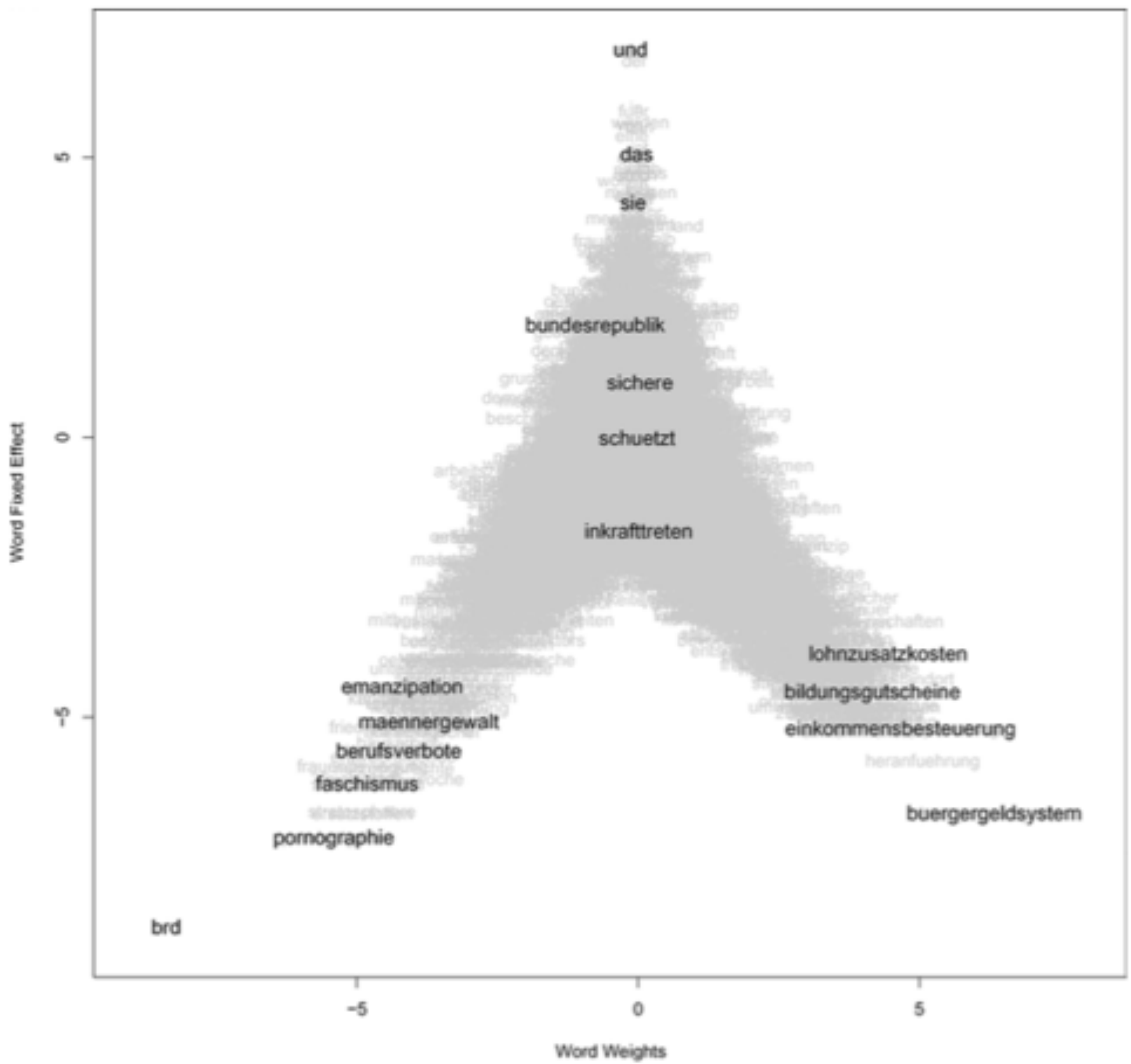
For the word "deficit":

- lower ψ_j
- larger (in magnitude) β_j; for example, if the right talks about "deficits" more frequently and party positions are oriented so that right is positive, β_j should be large and positive.

Eiffel Tower of words

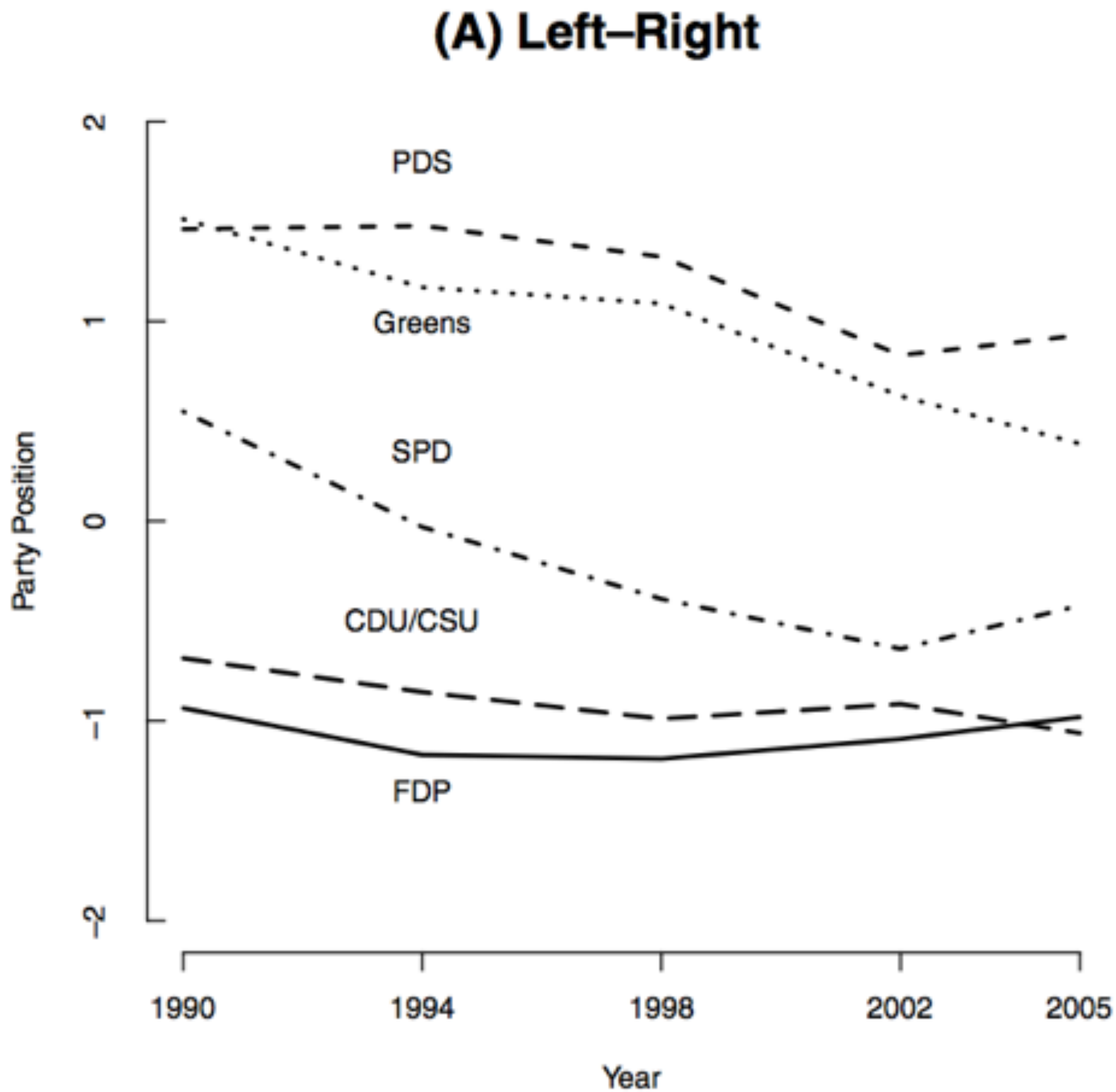
Slapin and Proksch, 2008

FIGURE 2 **Word Weights vs. Word Fixed Effects. Left-Right Dimension, Germany 1990–2005 (Translations given in text)**



Estimated party positions in Germany

Slapin and Proksch, 2008



Use of scaling models beyond text

- Measuring student ability and question difficulty in educational testing (origin of item response theory)
- Measuring ideology of contributors and ideological appeal of candidates using campaign contribution data (Bonica)
- Measuring ideology of groups of citizens (e.g. French women) using responses to survey questions (Caughey & Warshaw, group IRT)
- Measuring judges' ideology and how it changes over time (Martin & Quinn)

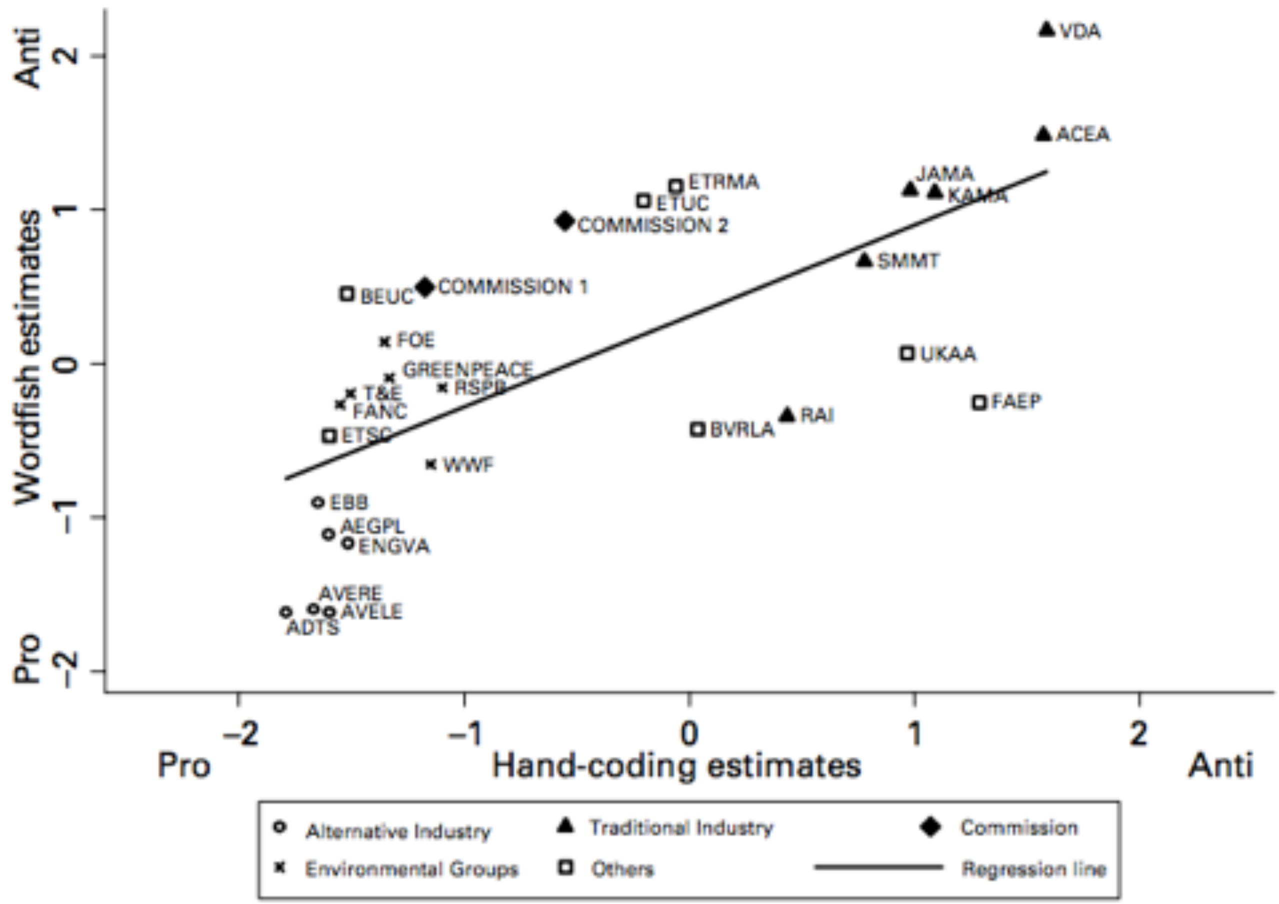
Another example

European Union Politics
DOI: 10.1177/1465116509346782
Volume 10 (4): 535–549

Measuring Interest Group Influence Using Quantitative Text Analysis

Heike Klüver
University of Mannheim, Germany

Kluver uses Wordfish to study policy statements on CO_2 emission regulations submitted by EU interest groups.



Discussion

When will scaling methods be useful for text?