

Modelling research funding in R

How much of REF *research quality* is really *journal reputation*?

David Selby

Department of Statistics, University of Warwick

26th September 2019

Research Excellence Framework

The REF is the system for assessing the quality of research in UK higher education institutions, according to

Quality of **outputs**

Impact outside academia

Research **environment**

Research Excellence Framework

Output profiles for *Economics & Econometrics*, REF2014

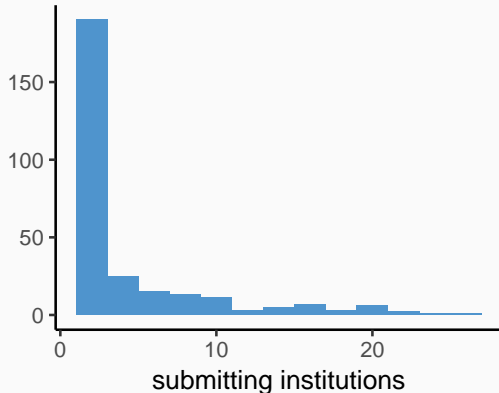
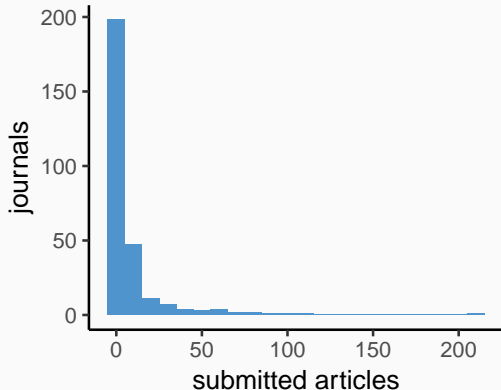
Institution	4*	3*	2*	1*	uc
UCL	70	28	2		
LSE	56	33	5	1	5
Cambridge	54	39	5	1	
Oxford	43	44	11	2	
Warwick	43	51	7		
...					

Economics & Econometrics REF2014 data

28 institutions

2,600 submitted articles

321 unique journals, books or reports



Identifying publications: a graph problem

ID	Volume title	ISSN
1	Journal of Political Economy	1234
2	the journal of political economy	12-3 4
3	Journal of Political Economy	1234
4	The Journal of Political Economy	
5	Journal of Something Else	6789
6	Journal of Political Economy	4321
7	J POL ECON	4321

Identifying publications: a graph problem

ID	Volume title	ISSN
1	Journal of Political Economy	1234
2	the journal of political economy	12-3 4
3	Journal of Political Economy	1234
4	The Journal of Political Economy	
5	Journal of Something Else	6789
6	Journal of Political Economy	4321
7	J POL ECON	4321

```
tidyr::gather(data, 'identifier', 'value', -ID) %>%  
  igraph::graph_from_data_frame() %>%  
  igraph::clusters()
```

Ecological Inference

Inferring **individual** behaviour from **aggregate** data

e.g. US voting rights litigation; voter transition models

	Vote	No vote	
Black	?	?	X_i
White	?	?	$N_i - X_i$
	Y_i	$N_i - Y_i$	N_i

Ecological Inference

Long history:

Goodman (1953)

Ecological regression

Brown and Payne (1986)

Dirichlet-multinomial model

Rosen et al. (2001)

Who voted for the Nazis?

Flaxman et al. (2015)

Who voted for Obama?

Poisson Binomial distribution

Sum of independent Bernoulli trials,
not necessarily identically distributed.

$$Y \sim \text{Poisson-Binomial}(p_1, p_2, \dots, p_N)$$

Poisson Binomial distribution

Sum of independent Bernoulli trials,
not necessarily identically distributed.

$$Y \sim \text{Poisson-Binomial}(p_1, p_2, \dots, p_N)$$

- Y # 4* ratings awarded
- N # articles submitted
- p_j Pr(article j rated 4*)

Articles in the same journal are assumed iid:

e.g. $\mathbf{p} = (p_1, p_1, p_1, p_2, p_2, p_3, p_5)$.

Model Prior

$$p_j \sim \text{Beta}(\gamma\mu, \gamma(1 - \mu))$$

for each journal j

$$\mu \sim \text{Uniform}(0, 1)$$

$$\gamma \sim \text{Gamma}(10, \frac{1}{10})$$

Likelihood

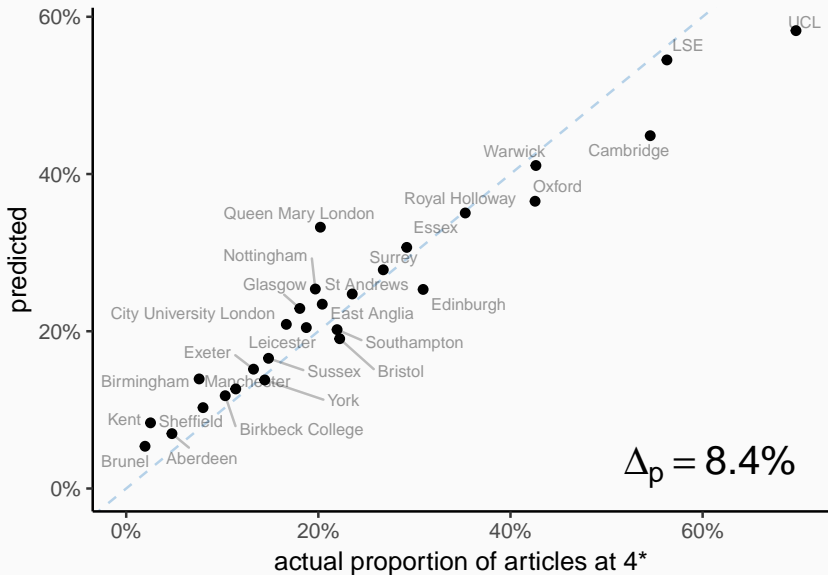
$$Y_i \sim \text{Poisson-Binomial}(p_1, \dots, p_{N_i})$$

for each institution i

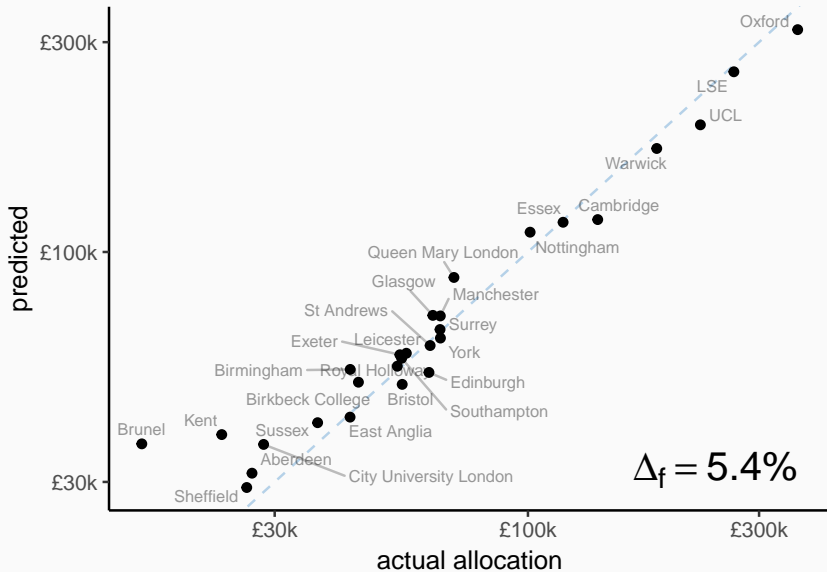
Journal probabilities of 4^* in REF2014



Predicted vs. actual REF2014 profile



Predicted vs. actual funding allocation



Discussion

Can we automate the REF?

outputs only

not all fields

reputations might change by 2021

Future work

Modelling other fields

e.g. maths & stats

Comparison with citation metrics

Thank you

David Selby

selbydavid.com

@TeaStats

D.Selby@warwick.ac.uk

David Firth

warwick.ac.uk/dfirth

@firthcity

D.Firth@warwick.ac.uk