Wordfish & Interest Group Influence

Contributor
Saniya Jain

To what extent do business elites influence labor policy?

Labor interest group vs Business Associations

Wordfish is a computer program written in the R statistical language to extract political positions from text documents. Word frequencies are used to place documents onto a single dimension. Wordfish is a scaling technique and does not need any anchoring documents to perform the analysis. Instead, it relies on a statistical model of word counts. The current implementation assumes a Poisson distribution of word frequencies. Positions are estimated using an expectation-maximization algorithm. Confidence intervals for estimated positions can be generated from a parametric bootstrap.The name Wordfish pays tribute to the French meaning of the word "poisson".

A Poisson distribution is a discrete probability distribution. It gives the probability of an event happening a certain number of times (k) within a given interval of time or space.

Organisation	Text	Labor
AFLCIO	The American Federation of Labor and Congress of Industrial Unions (AFL-CIO) submits these comments in support of the Equal Employment Opportunity Commission's ("EEOC" or "Commission") Notice of Proposed Rulemaking ("NPRM"), RIN 3046–AB30, Regulations To Implement the Pregnant Workers Fairness Act, published in the Federal Register on August 11, 2023.1	1
UFCW	The United Food and Commercial Workers International Union (UFCW) submits these comments in support of the Equal Employment Opportunity Commission's ("EEOC or	1
NEA	The American Federation of Labor and Congress of Industrial Unions (AFL-CIO) submits these comments in support of the Equal Employment Opportunity	1
NFIB	The National Federation of Independent Business (NFIB)1 submits these comments in	0
USCC	The U.S. Chamber of Commerce ("Chamber") presents these comments to the Equal Employment Opportunity Commission ("Commission" or "EEOC") in response to its Notice of	0
PROP_RULE	Federal Register / Vol. 88, No. 154 / Friday, August 11, 2023 / Proposed Rules	2
FINAL_RULE	Federal Register / Vol. 89, No. 77 / Friday, April 19, 2024 / Rules and Regulations	2
NADA	The National Automobile Dealers Association (NADA)1 respectfully submits the following	0
NTFTI	On behalf of the National Taskforce on Tradeswomen Issues, in support of the Equal Employment Opportunity Commission's ("EEOC" or "Commission") Notice of Proposed Rulemaking ("NPRM"), RIN 3046-AB30, Regulations To Implement the Pregnant Workers	1
NPWF	The National Partnership for Women & Families appreciates the opportunity to comment in support of the Equal Employment Opportunity Commission's (EEOC's) proposed regulations and enforcement guidance implementing the Pregnant Workers Fairness Act (PWFA). As an	1

Excel File with Comments

```
# 1. Remove URLs
text_vector_clean <- stri_replace_all_regex(text_vector_clean,
                                            "https?://\\S+|www\\.\\S+",
                                            "")

# 2. Replace smart quotes, dashes, bullets, and weird characters
text_vector_clean <- stri_replace_all_regex(text_vector_clean,
  c("[\u2018\u2019\u201A\u201B\u2039\u203A]",    # curly single quotes
    "[\u201C\u201D\u201E\u201F\u00AB\u00BB]",    # curly double quotes
    "[\u2022●]",                                 # bullets
    "[--]",                                      # en/em dashes
    "[^[:alnum:][:space:]'-])",                  # remove all other non-alphanum
  c("'", "\"", " ", "-", " "),
  vectorize_all = FALSE
)

# 3. Collapse extra whitespace
text_vector_clean <- str_squish(text_vector_clean)

# Reassign names
names(text_vector_clean) <- comments$Organisation
```

Pre-processing

Create a Corpus

```
# Tokenize and clean
tokens_q <- tokens(corpus_q,
                   remove_punct = TRUE,
                   remove_numbers = TRUE) %>%
   tokens_tolower() %>%
   tokens_remove(stopwords("en")) %>%
   tokens_remove(c("nbsp", "said", "also", "one", "will", "may", "must", "upon")) %>%
   tokens_keep(min_nchar = 3) %>%
   tokens_select(pattern = "^[a-z]+$", valuetype = "regex")  # keep only purely lower case words
```

```
Corpus consisting of 10 documents.
AFLCIO :
"The American Federation of Labor and Congress of Industrial ..."

UFCW :
"The United Food and Commercial Workers International Union U..."

NEA :
"The American Federation of Labor and Congress of Industrial ..."

NFIB :
"The National Federation of Independent Business NFIB 1 submi..."

USCC :
"The U S Chamber of Commerce Chamber presents these comments ..."

PROP_RULE :
"54714 Federal Register Vol 88 No 154 Friday August 11 2023 P..."

[ reached max_ndoc ... 4 more documents ]
```

```
Tokens consisting of 10 documents.
AFLCIO :
 [1] "american"   "federation" "labor"       "congress"   "industrial" "unions"     "afl"         "cio"         "submits"    "comments"   "support"
[12] "equal"
[ ... and 2,440 more ]

UFCW :
 [1] "united"      "food"        "commercial"  "workers"    "international" "union"      "ufcw"        "submits"    "comments"
[10] "support"     "equal"       "employment"
[ ... and 2,455 more ]

NEA :
 [1] "american"   "federation" "labor"       "congress"   "industrial" "unions"     "afl"         "cio"         "submits"    "comments"   "support"
[12] "equal"
[ ... and 2,440 more ]

NFIB :
 [1] "national"   "federation"  "independent" "business"   "nfib"       "submits"    "comments"    "response"   "notice"     "proposed"
[11] "rulemaking"  "titled"
[ ... and 1,524 more ]

USCC :
 [1] "chamber"    "commerce"    "chamber"     "presents"   "comments"   "equal"      "employment"  "opportunity" "commission" "commission"
[11] "eeoc"        "response"
[ ... and 2,500 more ]

PROP_RULE :
 [1] "federal"    "register"    "vol"         "friday"     "august"     "proposed"   "rules"       "equal"       "employment" "opportunity"
[11] "commission"  "cfr"
[ ... and 2,591 more ]
```

Document Term Matrix

```
Document-feature matrix of: 10 documents, 3,013 features (70.90% sparse) and 0 docvars.
        features
docs      american federation labor congress industrial unions afl cio submits comments
  AFLCIO        4         2      4      1          1          3    1   1       1        1
  UFCW         4         0      1      0          0          0    0   0       1        1
  NEA          4         2      4      1          1          3    1   1       1        1
  NFIB         0         1      0      3          0          0    0   0       1        1
  USCC         0         0      1     10          0          0    0   0       0        3
  PROP_RULE    1         0      5      2          0          1    0   0       0       11
[ reached max_ndoc ... 4 more documents, reached max_nfeat ... 3,003 more features ]
```

Key Intuition:
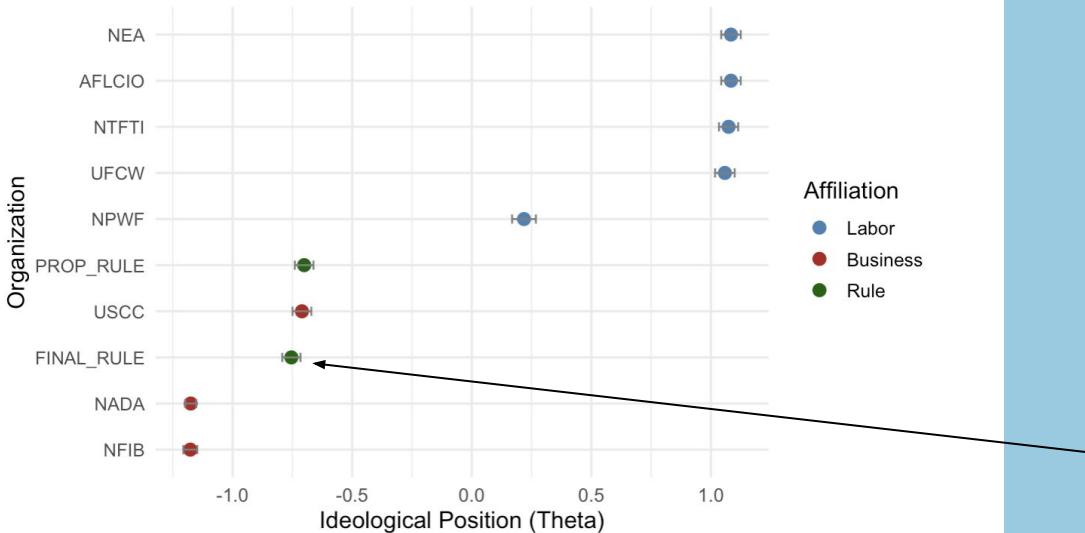
Wordfish treats each word count as a little clue about a document's position.
It uses the Poisson model to statistically link word frequency patterns with latent ideological
positions — all without needing labels.

```
# Run Wordfish model
wf_quanteda <- textmodel_wordfish(dfm_q, dir = c(4, 1))
```
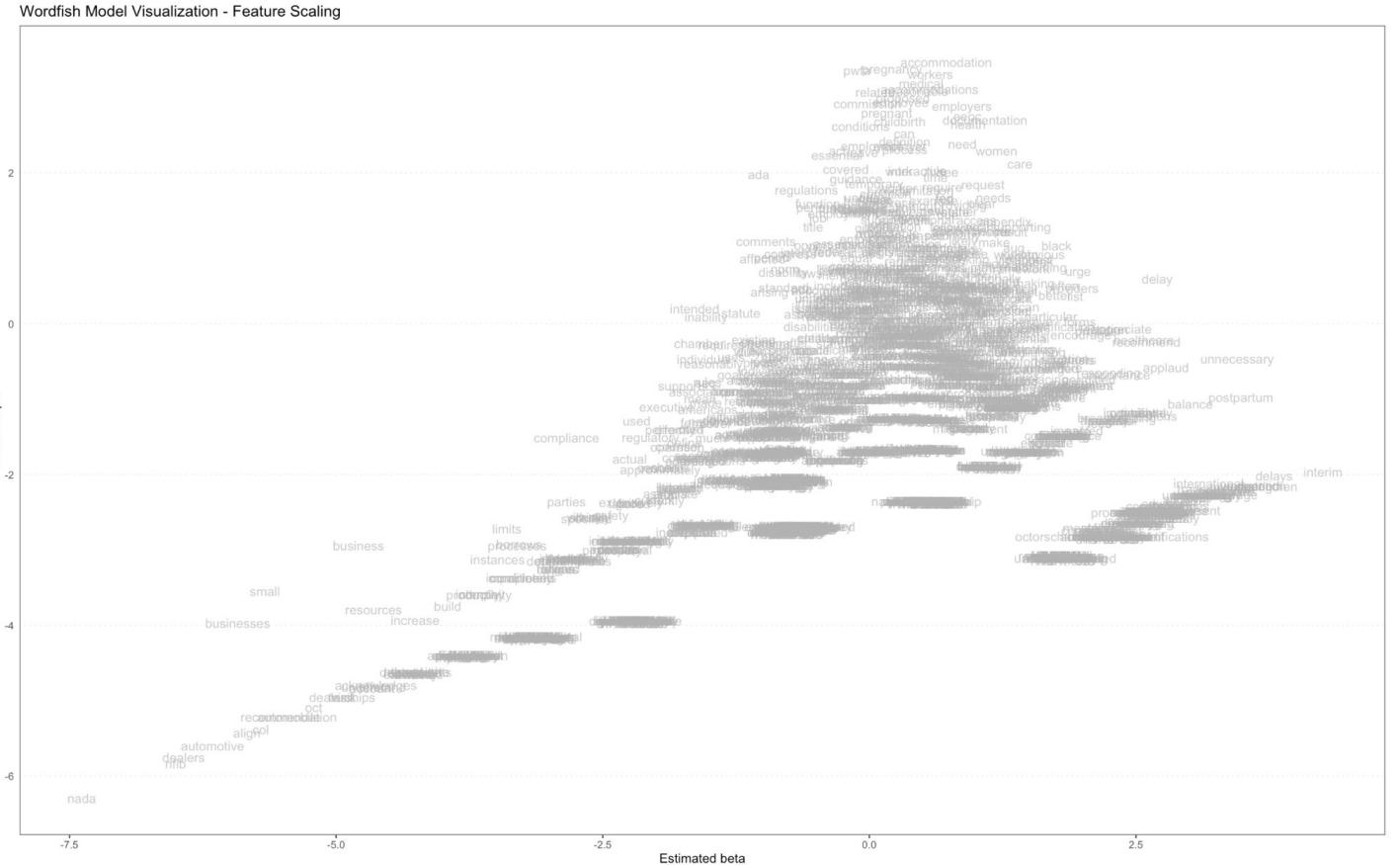
Anchoring dimensions

Regulations To Implement the Pregnant Workers Fairness Act
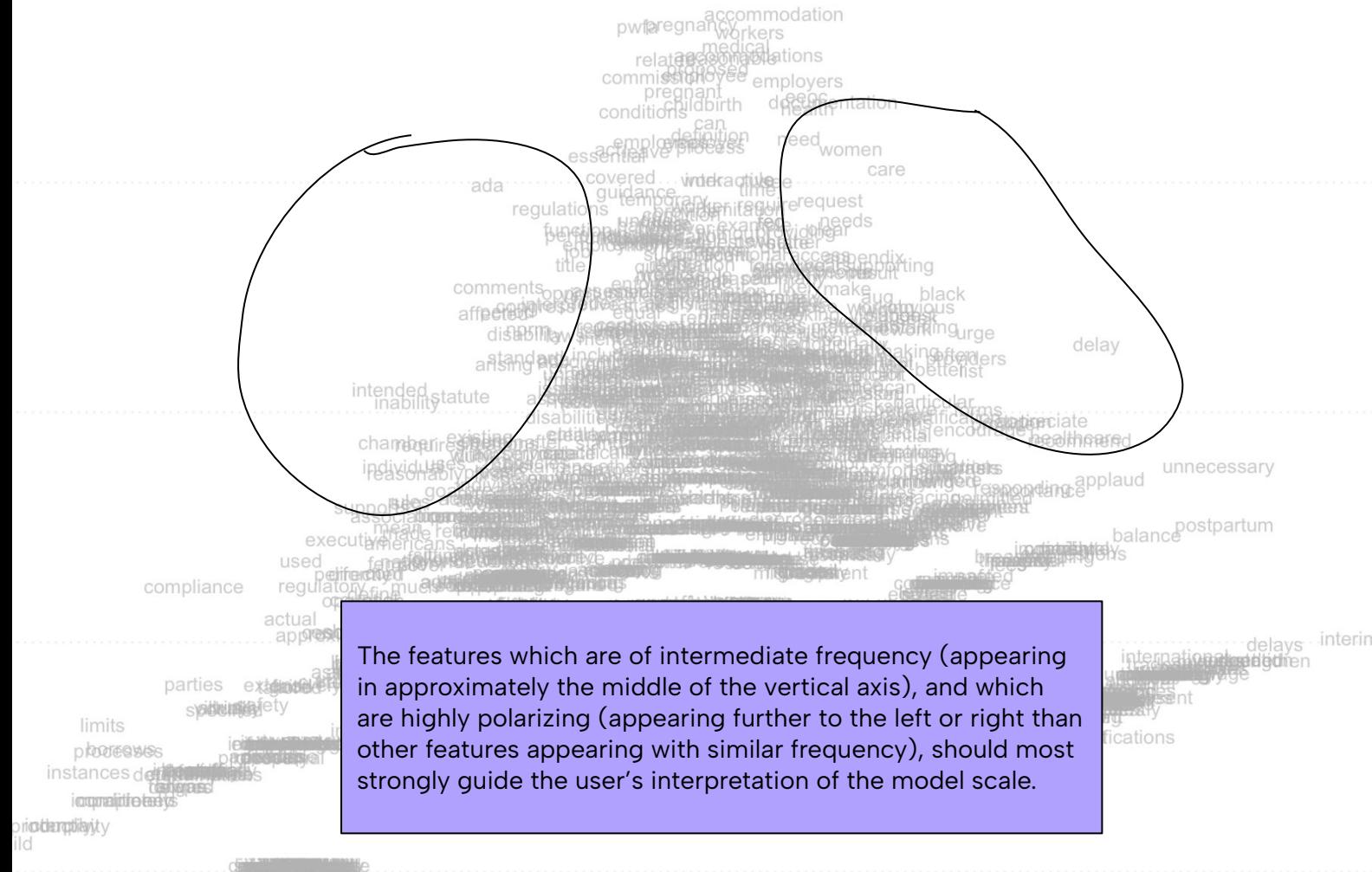
Ideological position of Final rule is slight left leaning

Word level Visualization

Wordfish Model Visualization - Feature Scaling

- The horizontal axis (beta) represents how extreme the scaling of each feature is: features far to the left and far to the right drive the placement of documents on the document–level left–right scale
- Centrally located features have minimal impact.
- The vertical axis (psi) describes how common that feature is: more common features appear at the top of the model, while rarer ones appear at the bottom.

The features which are of intermediate frequency (appearing in approximately the middle of the vertical axis), and which are highly polarizing (appearing further to the left or right than other features appearing with similar frequency), should most strongly guide the user's interpretation of the model scale.

Description: df [60 × 4]

	Side <chr>	word <chr>	beta <dbl>	psi <dbl>
1203	Left	compliance	-2.839069	-1.51278224
1295	Left	actual	-2.247233	-1.79024001
1224	Left	used	-2.181691	-1.28705431
1335	Left	regulatory	-2.051401	-1.50277669
2260	Left	approximately	-1.960472	-1.93035515
1233	Left	resume	-1.937565	-1.90952341
1411	Left	concerns	-1.932051	-1.90451783
1288	Left	shall	-1.924984	-1.89810756
1518	Left	executive	-1.899444	-1.10179779
1227	Left	performed	-1.825503	-1.40301877
1701	Left	directly	-1.821949	-1.39983892
1304	Left	operation	-1.799570	-1.63116471
1523	Left	counsel	-1.795205	-1.62727313
1580	Left	supports	-1.747428	-0.82269945
1229	Left	define	-1.732248	-1.57141879
2205	Left	com	-1.676403	-1.38878293
1581	Left	intended	-1.637495	0.20331739
1756	Left	fertility	-1.630552	-1.81879437
1857	Left	addressed	-1.630552	-1.81879437
1784	Left	calling	-1.618847	-1.80863377
1524	Left	chamber	-1.594018	-0.26109080
1675	Left	female	-1.588401	-1.31229609
1230	Left	mean	-1.586642	-0.99232516
2632	Left	seeing	-1.580313	-1.77532697
1644	Left	contemplated	-1.579254	-1.77441505
1670	Left	removing	-1.579254	-1.77441505
1049	Left	association	-1.576064	-0.89620012
1201	Left	individual	-1.552543	-0.47053129
1925	Left	secretariat	-1.545908	-1.74578177
1996	Left	seq	-1.533929	-0.78002665

394	Right	interim	4.253839	−1.96089944
310	Right	postpartum	3.484290	−0.97165104
279	Right	unnecessary	3.447779	−0.46058998
448	Right	balance	3.009267	−1.06239289
233	Right	applaud	2.782691	−0.57564848
280	Right	delay	2.699896	0.59559330
305	Right	recommend	2.596703	−0.23948644
88	Right	healthcare	2.574038	−0.21695329
282	Right	suggestions	2.563074	−1.22464177
396	Right	immediately	2.517437	−1.17947772
495	Right	contribute	2.517437	−1.17947772
509	Right	timely	2.517437	−1.17947772
522	Right	racism	2.517437	−1.17947772
274	Right	appreciate	2.365204	−0.06909320
426	Right	importance	2.334095	−0.67499132
956	Right	breastfeeding	2.319672	−1.24887269
510	Right	way	2.309414	−1.23899982
357	Right	responding	2.244913	−0.64663894

	Side	word	beta	psi
	<chr>	<chr>	<dbl>	<dbl>
295	Right	thoughtful	2.242420	−1.28025408
514	Right	rural	2.242420	−1.28025408
647	Right	record	2.242420	−1.28025408
691	Right	invasive	2.242420	−1.28025408
886	Right	privacy	2.242420	−1.28025408
312	Right	lactation	2.200430	−0.07035232
540	Right	fees	2.156622	−1.31695728
798	Right	provider	2.147395	−0.09191579
679	Right	permitted	2.057838	−0.73968280
324	Right	encourage	1.984412	−0.14661395
71	Right	forms	1.975391	0.02858987
48	Right	safe	1.959492	−1.42342636

Thank You!