

Cultural Change

Evidence from Three Centuries of U.S. Local Newspapers

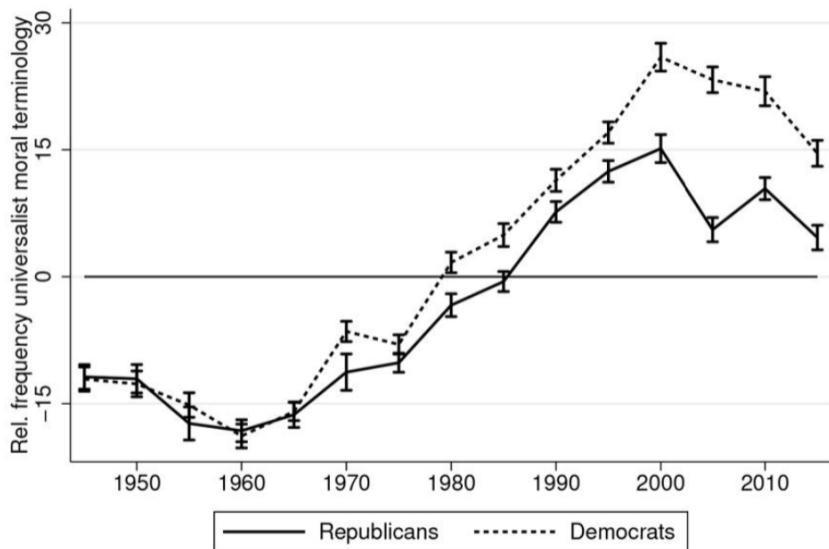
Max Posch
(formerly Max Winkler)
University of Exeter

ASREC Conference 2023
March 18, 2023

Studying cultural change through text analysis of historical corpora

- Historical text documents are an exceptional source to track the evolution of cultural traits across space and over time
- A popular design to study cultural change using historical text comprises:
 1. Drawing on an existing **dictionary** of keywords presumably related to the cultural trait of interest
 2. Counting the prevalence of these keywords in a readily available corpus of historical texts, such as the **US Congressional Record** or **Google Books Ngrams**

Example: evolution of universalism (Enke, JPE, 2020)



Example: evolution of norm tightness (Jackson et al., NHB, 2019)

Table 1 | Words selected to comprise the final tightness and looseness indices

Tight words	Loose words
Restrain	Allow
Prevent	Freedom
Comply	Create
Constrain	Variability
Uniformity	Autonomy
Adhere	Openness
Enforce	Leeway
Proscribe	Flexibility
Abide	Broadmindedness
Dictate	Transformatory
Circumscribe	Customize
Impose	Subjectivities
Uphold	Modify
Discourage	Limitless
Compel	Empower
Forbid	Adaptiveness
Confine	Pluralism
Govern	Personalize
Prohibit	Encourage
Preclude	Diverse

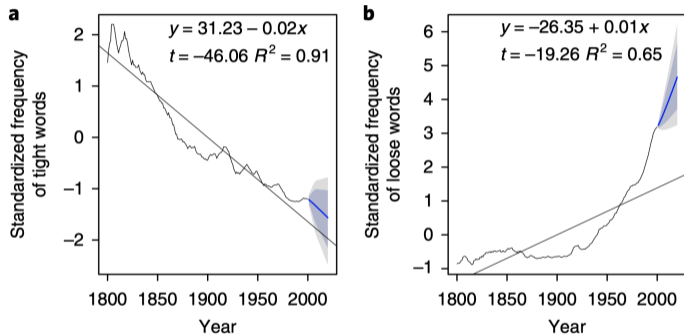


Fig. 1 | Frequencies in tight and loose words in books from 1800 to 2000.

Limitations of this approach

1. Data

- Difficult to disentangle true cultural changes from changes in topics or genre composition over time – e.g., large rise in fiction in Google Books corpus after 2000
 - Hard to examine causal relationships – limited spatial variation along with time variation
- ⇒ **We access text from thousands of local newspapers going back to 1700s**

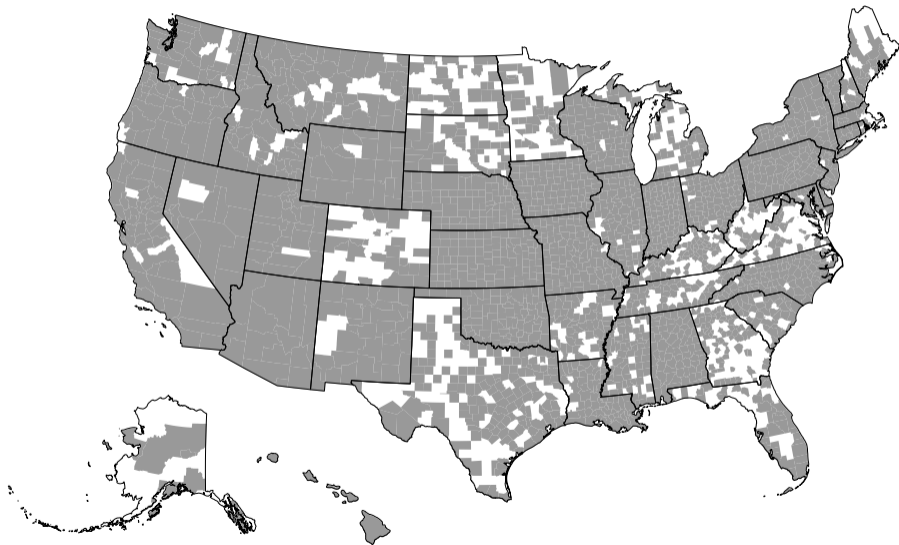
2. Method

- Open questions about best practices for keyword selection; e.g., semantic changes over time, emergence of new words, part-of-speech balance
 - Limited options to validate against conventional psychological data (e.g., from surveys)
- ⇒ **Today: we propose refinements and validation checks for dictionary-based approaches**
- ⇒ **Future: we focus on (contextualized) embeddings from language models**

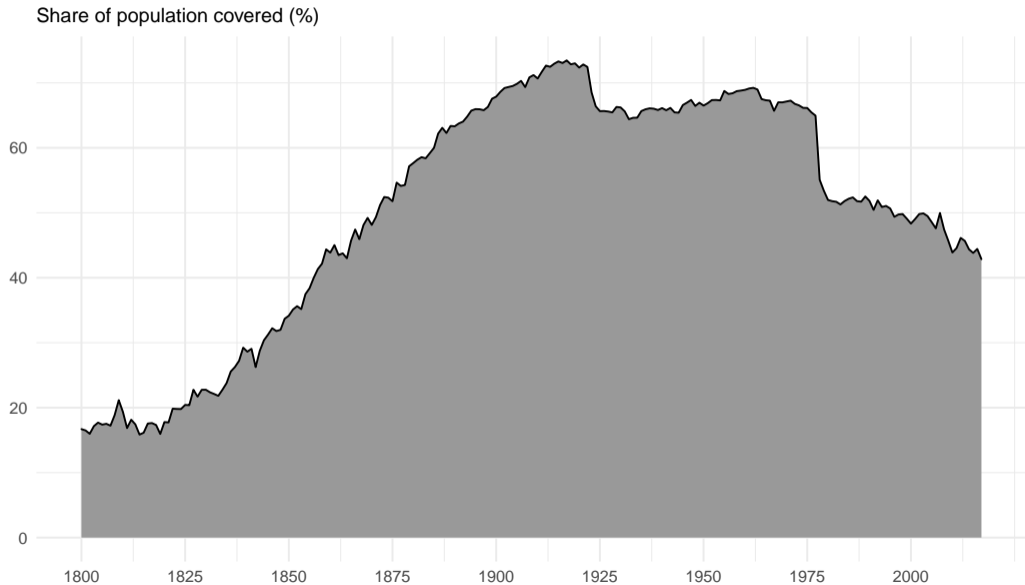
Data: digitized U.S. newspaper archives

- We draw on **newspapers.com** and **newspaperarchive.com** to compile a novel database containing **more than 1B pages** from local newspapers
 - We can link newspaper text to key metadata: newspaper name, publication date, page, city of publication, circulation, political affiliation, and much more
 - The database covers 2,405 U.S. counties from all states and goes back to the 1700s
 - Newspaper markets in the U.S. have been **highly local** – only one (two) daily newspaper in 77% (14%) of counties between 1869 and 2004 – Gentskowitz + 2011
- ⇒ Testable hypothesis: **language in newspapers reflects local culture**

Newspaper database covers 2,405 counties from all states



Newspapers cover large but unbalanced population share over time



Method: dictionary-based approach

- We draw on the **Tight-Loose dictionary** created by Jackson et al. (NHB, 2019)
- The dictionary contains 2x20 keywords whose Google News word2vec coordinates are close to the coordinates of 2x8 seed words related to Tight-Loose theory

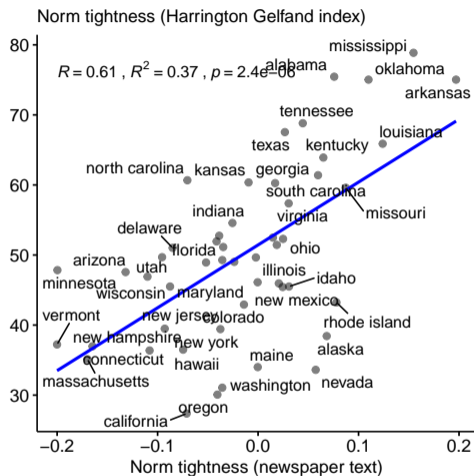
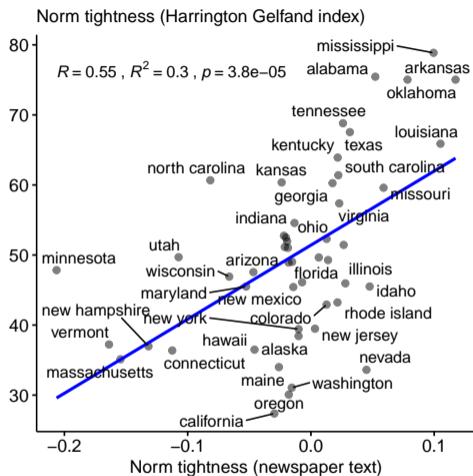
$$\text{Norm Tightness} = \text{Avg}_t \left[\text{Scale} \left[\frac{\# \text{ tight word}_t}{\text{doc length}} \right] \right] - \text{Avg}_i \left[\text{Scale} \left[\frac{\# \text{ loose word}_i}{\text{doc length}} \right] \right]$$

where a document is all newspaper text in newspaper i located in county c in year t

Method: refining and validating the dictionary

- **Semantic substitution:** Substitute each word with semantically closest neighbor, then recompute the construct of interest and report its correlation with the original measure ($\rho \approx 1$)
- **Part-of-speech balancing:** We augment the dictionary by balancing nouns, verbs, adjectives, adverbs, etc.
 - e.g., if dictionary contains *prohibit*, we add *prohibited*, *prohibiting*, *prohibition*, *prohibitive*, and *prohibitively*
 - $\rho = 0.896$
- **Validation against survey-based data and other proxy measures** (next slides)

Does the text-based measure correlate with conventional psych data?

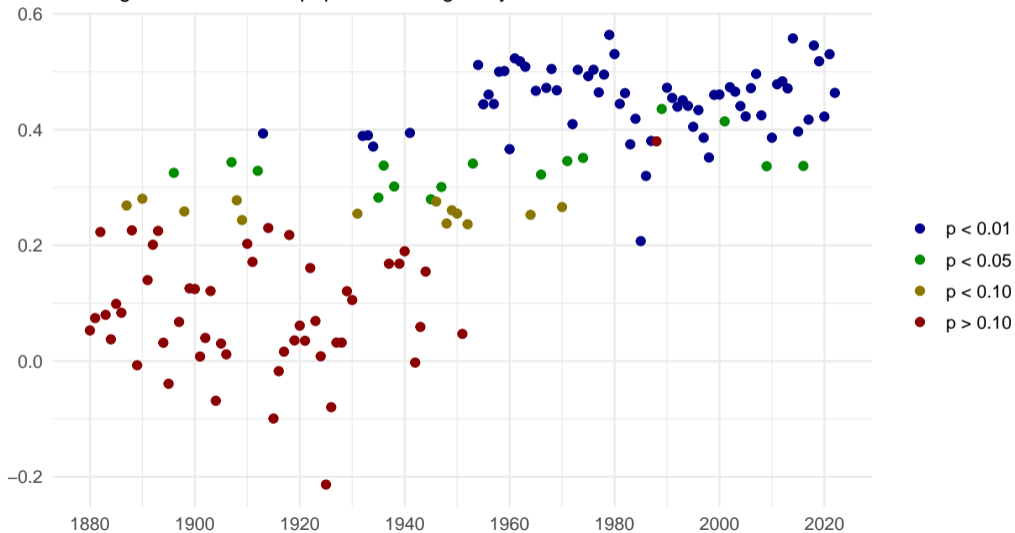


Data shown on y-axis are from Harrington Gelfand (PNAS, 2014). Left: original dictionary; right: refined dictionary.

Refined dictionary performs about **10%** better than the original norm tightness dictionary created by Jackson et al. (NHB, 2019)

Text from before 1950s less correlated with modern psych data

Correlations between Harrington Gelfand index and norm tightness from newspaper text in a given year



Decade-by-decade rank correlations suggest break during 1950s

decade1900	1	0.56	0.32	0.28	0.21	0.11	0.09	0	-0.04	-0.07
decade1910	0.56	1	0.56	0.41	0.35	0.26	0.24	0.11	0.07	0.07
decade1920	0.32	0.56	1	0.72	0.55	0.31	0.24	0.1	0.03	0.02
decade1930	0.28	0.41	0.72	1	0.69	0.42	0.32	0.17	0.09	0.04
decade1940	0.21	0.35	0.55	0.69	1	0.59	0.45	0.28	0.17	0.17
decade1950	0.11	0.26	0.31	0.42	0.59	1	0.79	0.66	0.5	0.44
decade1960	0.09	0.24	0.24	0.32	0.45	0.79	1	0.78	0.55	0.48
decade1970	0	0.11	0.1	0.17	0.28	0.66	0.78	1	0.69	0.59
decade1980	-0.04	0.07	0.03	0.09	0.17	0.5	0.55	0.69	1	0.76
decade1990	-0.07	0.07	0.02	0.04	0.17	0.44	0.48	0.59	0.76	1

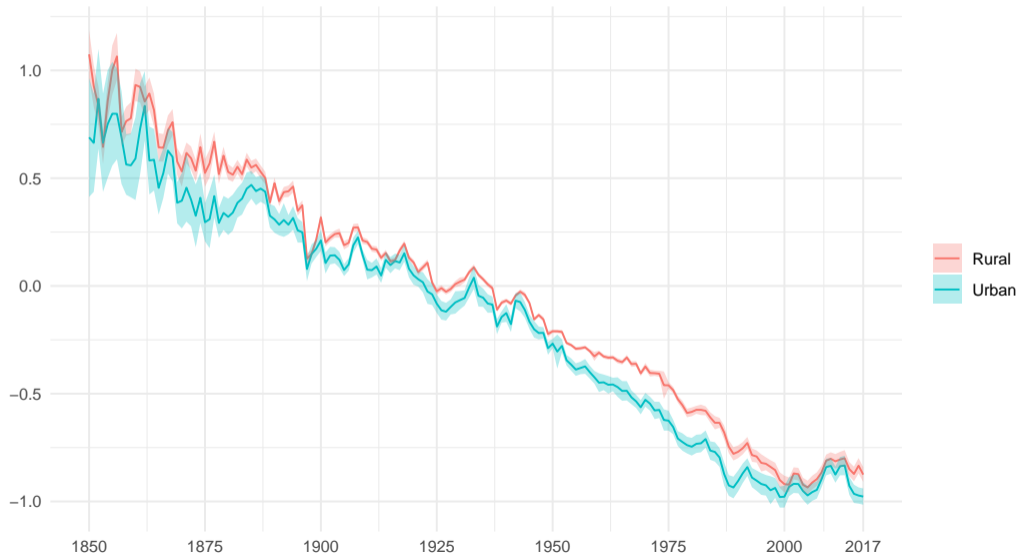
Frequency of tight vs. loose words in newspapers from 1728 to 2022

Norm tightness in newspapers (yearly averages and 95% CI)



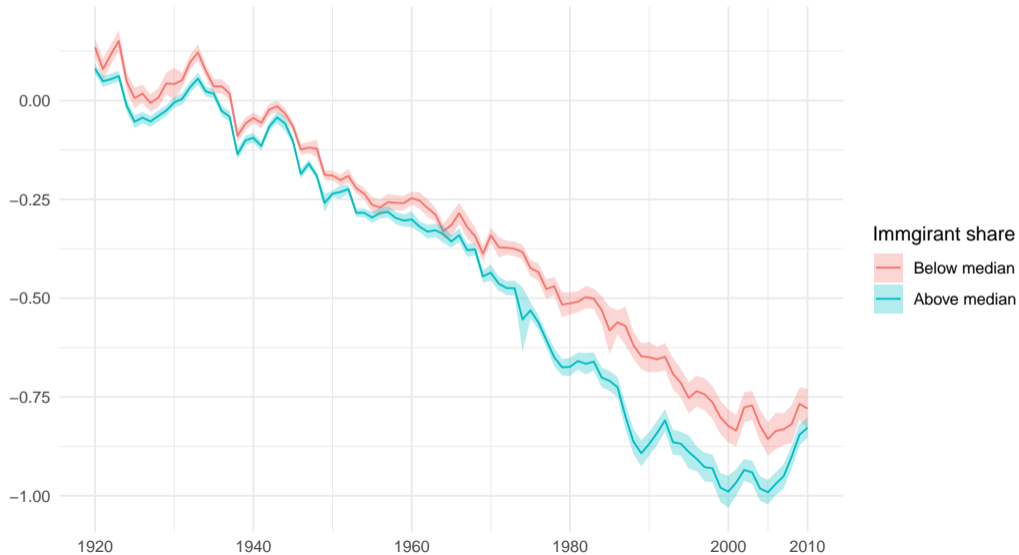
Looser language in newspapers in urban compared to rural locations

Norm tightness in newspapers (yearly averages and 95% CI)



Looser language in newspapers in places with higher immigrant share

Norm tightness in newspapers (yearly averages and 95% CI)



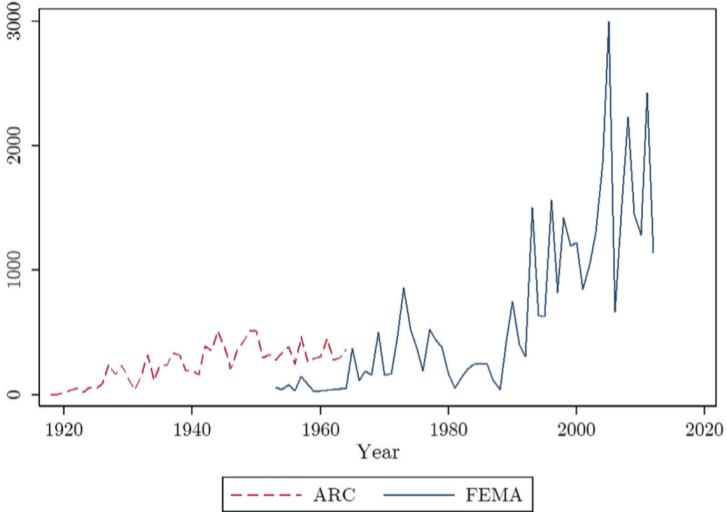
Do natural disasters tighten social norms?

- **Data:** all U.S. federally designated natural disasters from 1918 to 2012, aggregated to county-decade level - Boustan + 2020
- **Empirical strategy:** standard difference-in-differences equation

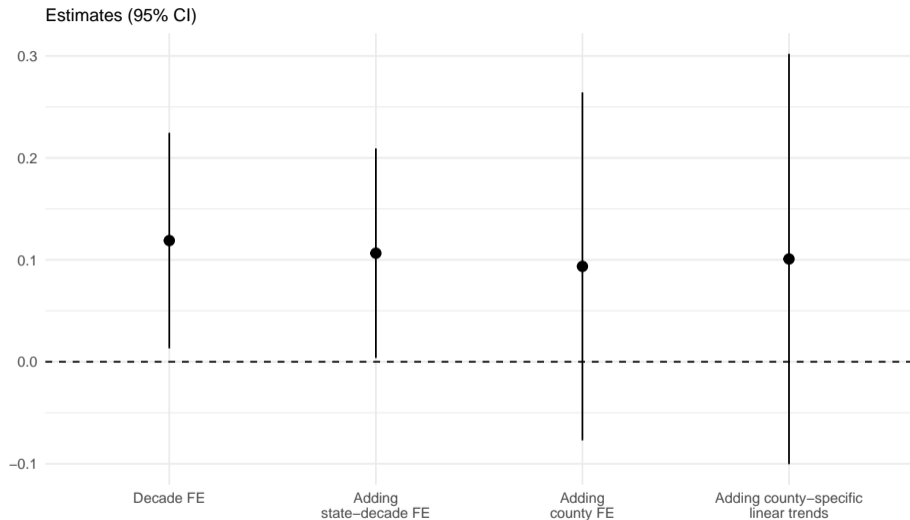
$$\text{Tightness}_{c(s)t} = \beta \text{ Severe disaster indicator}_{ct} + \alpha_c + \alpha_{st} + \alpha_c \times t + \varepsilon_{ct}$$

- c, s, t denote county, state, year
- **Severe disaster indicator** = 1 if a disaster with ≥ 25 fatalities occurs b/w $t - 10$ and t
- α_c : county-fixed effects, absorbing time-invariant factors (e.g., geography)
- α_{st} : decade-fixed effects, absorbing time-variant state-specific factors (e.g., economic, cultural, or environmental trends, state laws)
- $\alpha_c \times t$: county-specific linear trends, removing local factors that smoothly change over time (e.g., local long-term economic progress)
- Standard errors clustered on states

Annual disaster count in U.S. from 1918 to 2012 trends upwards



Natural disasters increase norm tightness in newspapers



⇒ Severe disasters increase norm tightness by ≈ 0.1 s.d. or **10%** of gap b/w VT and AR

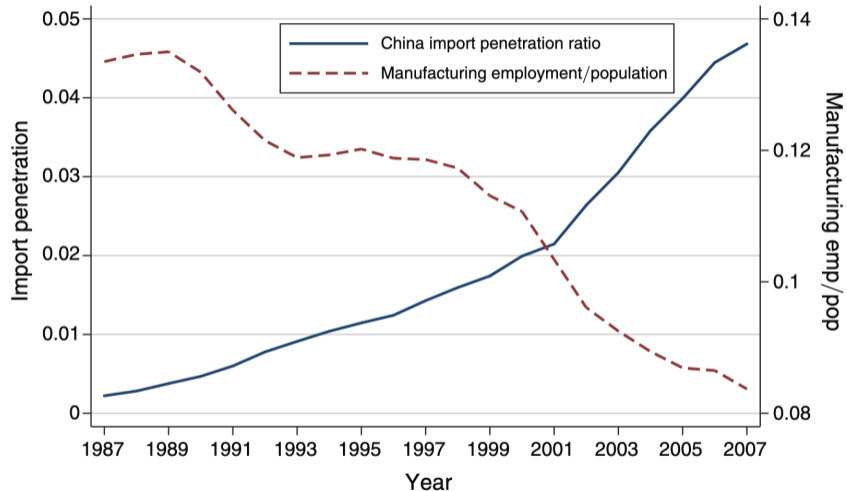
Do adverse economic shocks tighten social norms?

- **Data:** Exogenous variation in local U.S. labor market conditions induced by trade with China b/w 1990 and 2007 – the “China Shock” – Autor + 2013
- **Empirical strategy:** First-difference equation

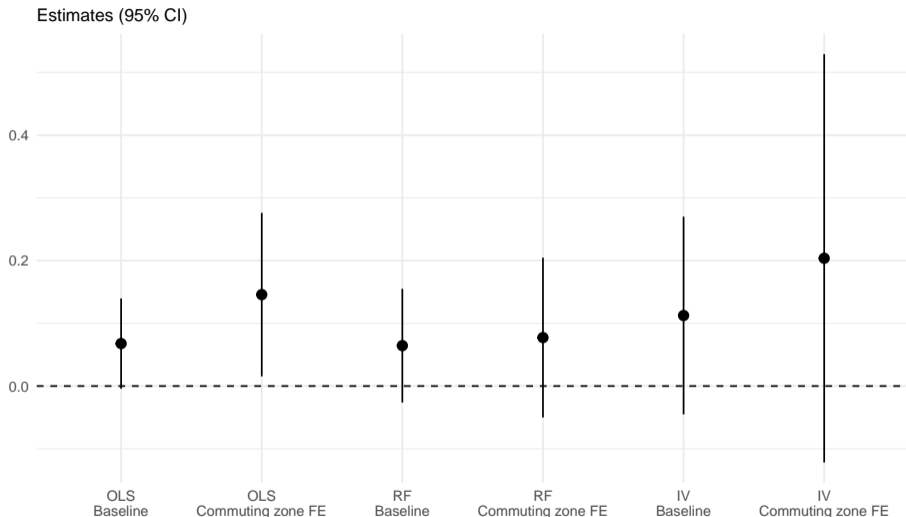
$$\Delta \text{Tightness}_{c(r)t} = \beta \Delta \text{Local trade exposure}_{ct} + X_c \gamma + \alpha_r + \alpha_t + \varepsilon_{c(r)t}$$

- c, r, t denote commuting zone, census-region, period
- Δ **Local trade exposure** is change in local import competition with China from $t - 1$ to t
- X_c : baseline employment share in manufacturing
- α_r : census-region fixed effects, removing regional factors that smoothly change over time
- α_t : period-fixed effects, absorbing time-variant factors affecting all commuting zone
- Standard errors clustered on states

Local trade exposure and manufacturing employment in U.S.



Adverse economic shocks increase norm tightness in newspapers



⇒ 1% ↑ trade exposure increase norm tightness by ≈ 0.1 s.d. or 10% of gap b/w VT - AR

Take-aways

- We study historical psychology through the lens of thousands of local newspapers from all U.S. states and going back to the 1700s
- We propose best practices dictionary based approaches; validation checks are key
- Consistent with previous studies, we find norms in the U.S. loosened over the past 200 years, while local adverse shocks causes local tightening
- Unable to conclude that the geography of norm tightness changed in the 1950s; possible that the use or meaning of keywords included in the dictionary changed
 - ⇒ Will train language models on the newspaper corpus to track embeddings over time

Our interdisciplinary newspaper team



Mohammad Atari
Psychology



Prमित Chaudhuri
Classics and Comparative Literature



Joseph Dexter
Data Science



Max Posch
Economics



Joe Henrich
Anthropology and Psychology



Jonathan Schulz
Economics