Cultural Change
Evidence from Three Centuries of U.S. Local Newspapers

Max Posch
(formerly Max Winkler)
University of Exeter

Workshop on Kinship, Historical Psychology and European Medieval Development
October 6, 2023

Analyzing cultural change via historical texts

- Historical text documents are increasingly used to study cultural change when conventional survey data are not available
  - For example, historical books, newspapers, parliamentary or presidential speech
- Assumption: the text reflects the culture of the people who read, wrote or spoke it
- These studies typically rely on **time series data** and **cannot leverage spatial variation**, making it hard to examine causality

This project: a database of local culture from digitized US newspapers

- We access **more than 1B pages** of local newspapers available through **newspapers.com**, **newspaperarchive.com** or **chroniclingamerica.gov**

- We can link newspaper text to key metadata: newspaper name, publication date, page, city of publication, circulation, political affiliation, and more

- The database covers 2,405 U.S. counties from all states and goes back to the 1700s

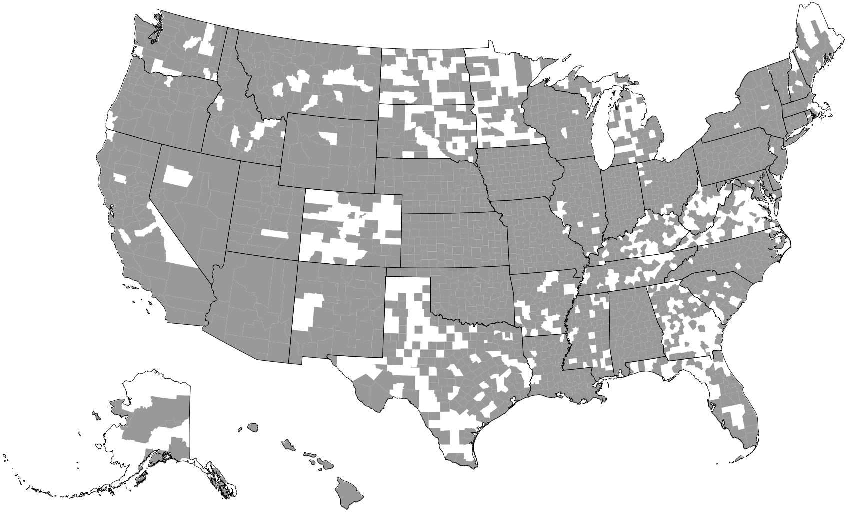
- Newspaper markets in the U.S. have been **highly local** — only one (two) daily newspaper in 77% (14%) of counties between 1869 and 2004 – Gentzkow + 2011

Limitations of our newspaper data

- Full-text access for around 25% of the corpus, otherwise access to frequencies of keywords and keyword combinations at the page level
- Poor OCR quality, especially for older newspapers, and no article segmentation
- We will assess the importance of these issues for our goals by replicating our measurement using the corpus recently published by Dell et al.
- Their corpus has better OCR quality and segmented articles for approx. 20M pages

Newspaper database covers 2,405 counties from all states

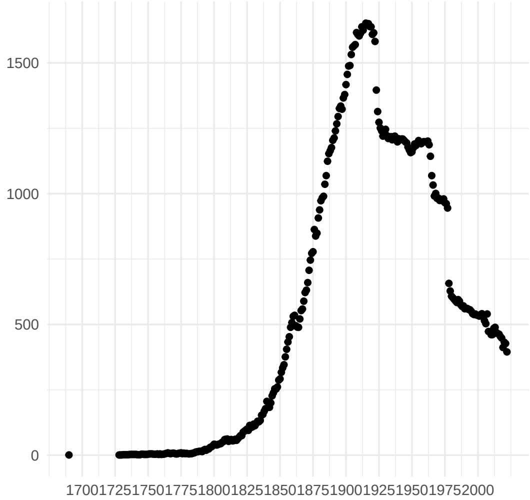
‣ newspapers.com ‣ newspaperarchive.com



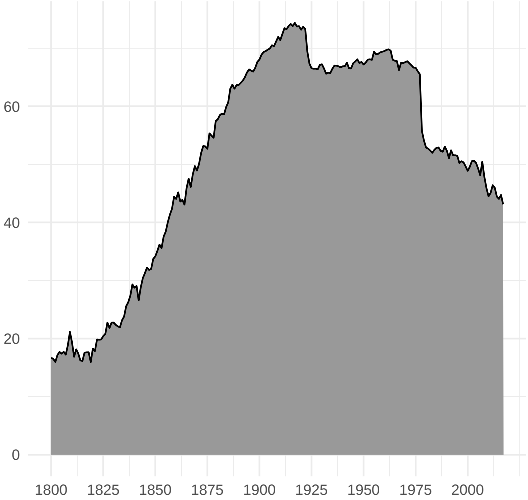
Number of counties covered unbalanced over time

- newspapers.com
- newspaperarchive.com

Counties with at least one newspaper page in a given year



Share of population covered (%)



Today, focus on a specific cultural trait: norm tightness

Sources: Harrington, Gelfand (PNAS, 2014); Jackson et al. (Nature Human Behavior, 2019)

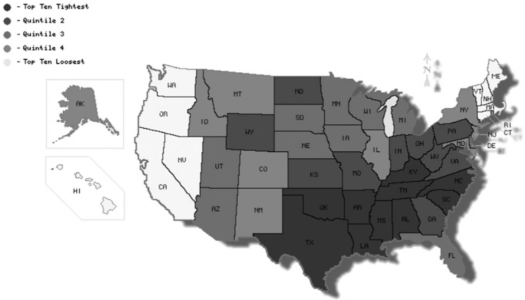


Fig. 1. Patterns of tightness-looseness at the state level in the United States. States are organized into quintiles based upon tightness-looseness index scores. This map was constructed at www.diymaps.net.

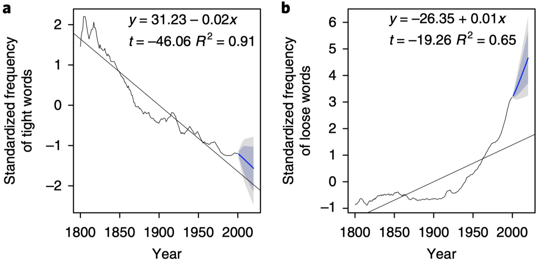


Fig. 1 | Frequencies in tight and loose words in books from 1800 to 2000.

How we measure norm tightness from newspaper text

- The literature knows **two types of approaches** to detect concepts, such as norm tightness, from text:
  1. **Dictionary-based:** Create a dictionary of words that are semantically close to the concept of interest and count their frequency in the text
  2. **Algorithmic-based:** Use a machine learning algorithm to detect the concept of interest from the text

- Both approaches have advantages and disadvantages:
  - **Dictionary-based:** transparent, but coarse
  - **Algorithmic-based:** often higher accuracy, but opaque (and limited to full-text corpora)

- We use two approaches and compare their performances:
  1. Dictionary augmented with machine judgement
  2. Contextualized Construct Representation (CCR), which is based on BERT (a language model that captures contextual semantic information)

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 embeddings* are close to the *embeddings* of 2x8 seed words related to Tight-Loose theory

- *Embeddings* are vectors of numbers that represent the meaning of a word based on co-occurrence across the whole corpus; intuitively, words that appear in similar corpus contexts get similar vectors

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

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

---

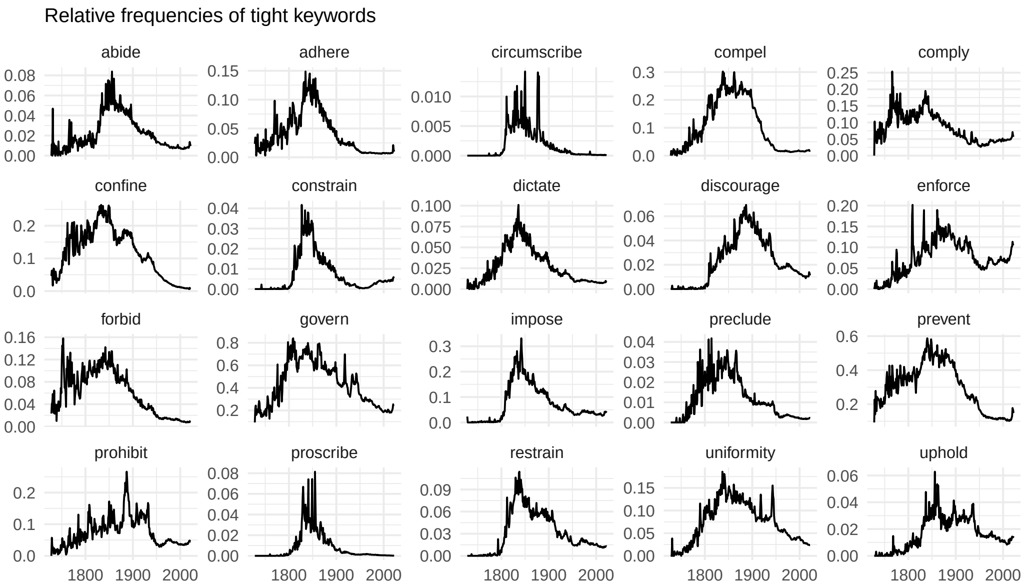
Tight words: restrain, prevent, comply, constrain, uniformity, adhere, enforce, proscribe, abide, dictate, circumscribe, impose, uphold, discourage, compel, forbid, confine, govern, prohibit, preclude. **Loose words:** allow, freedom, create, variability, autonomy, openness, leeway, flexibility, broadmindedness, transformatory, customize, subjectivities, modify, limitless, empower, adaptiveness, pluralism, personalize, encourage, diverse

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 proxy measure from the literature** (in four slides)

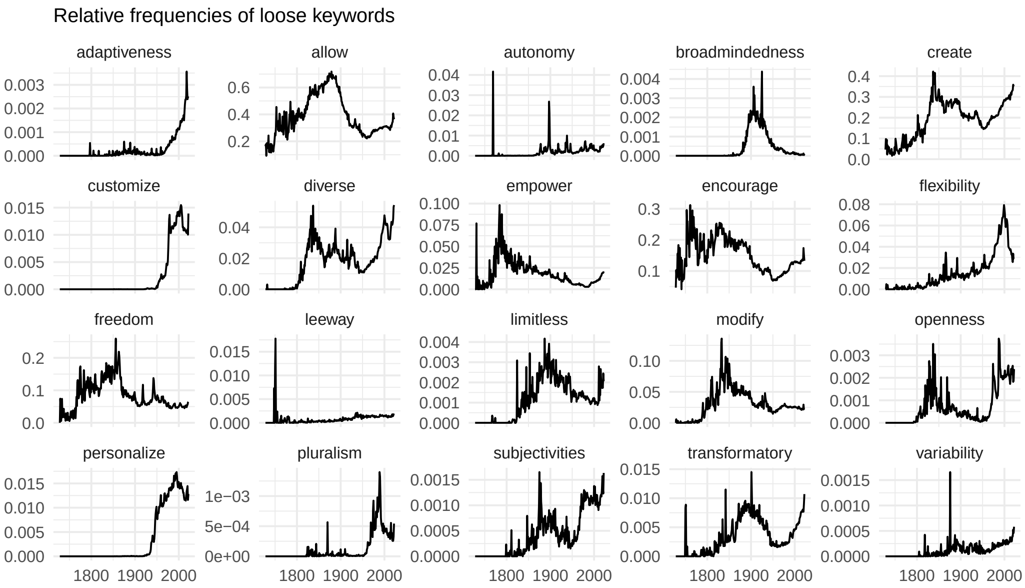
Relative frequency of tight words in newspapers

▸ newspaperarchive.com



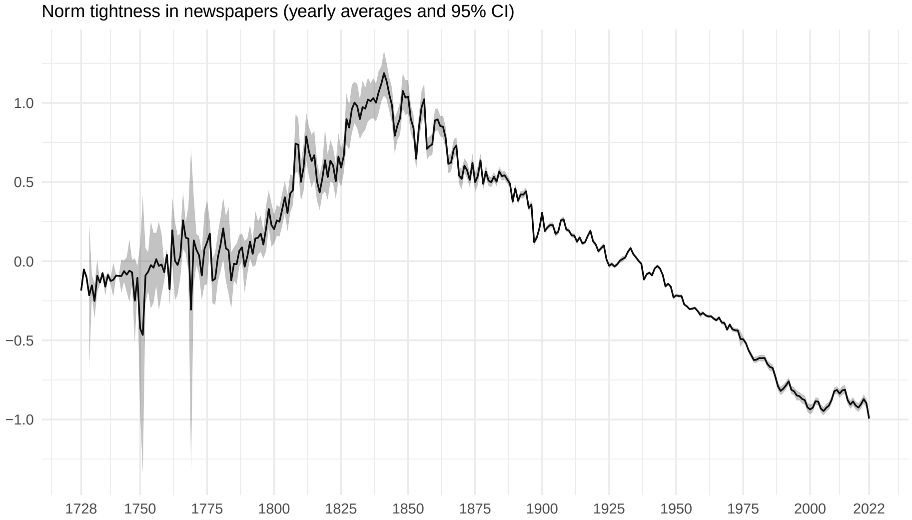
Relative frequency of loose words in newspapers

newspaperarchive.com



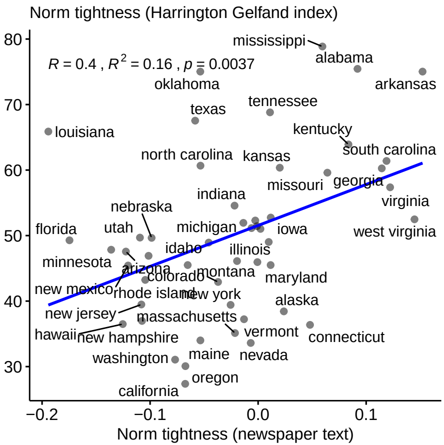
Frequency of tight vs. loose words in newspapers from 1728 to 2022

▸ newspaperarchive.com



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

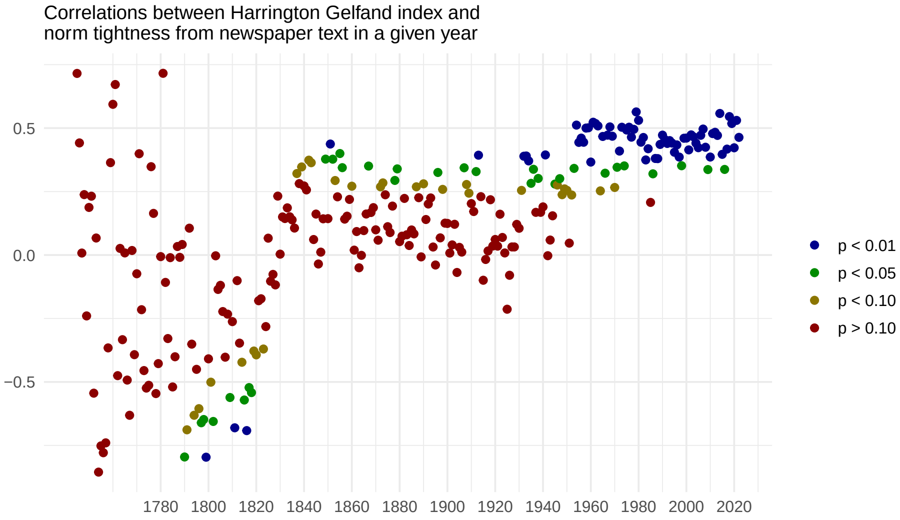
- newspaperarchive.com



Data shown on y-axis are from Harrington Gelfand (PNAS, 2014).

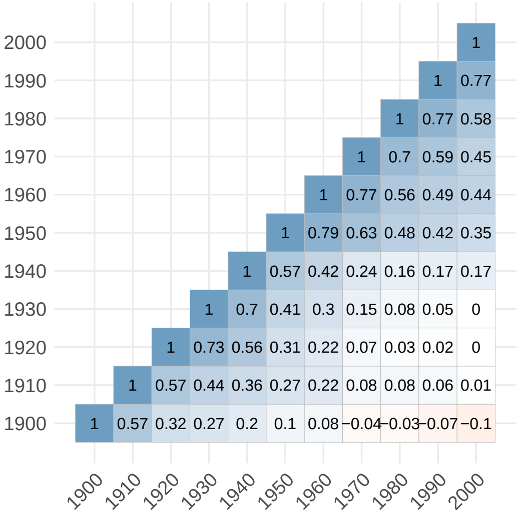
Text after 1950 highly correlated with modern psych data

- newspaperarchive.com



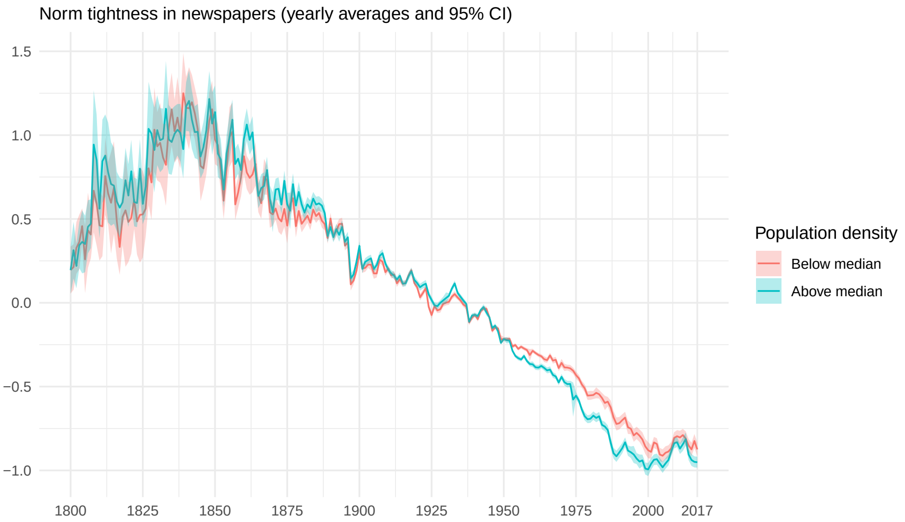
Decade-by-decade rank correlations suggest break during 1940s

▸ newspaperarchive.com



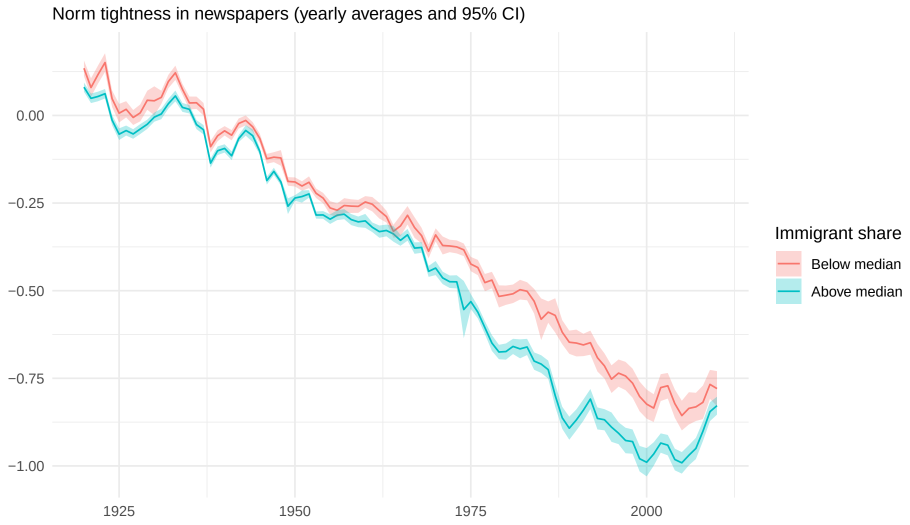
Urban newspapers increasingly looser than rural ones after 1950

- newspaperarchive.com



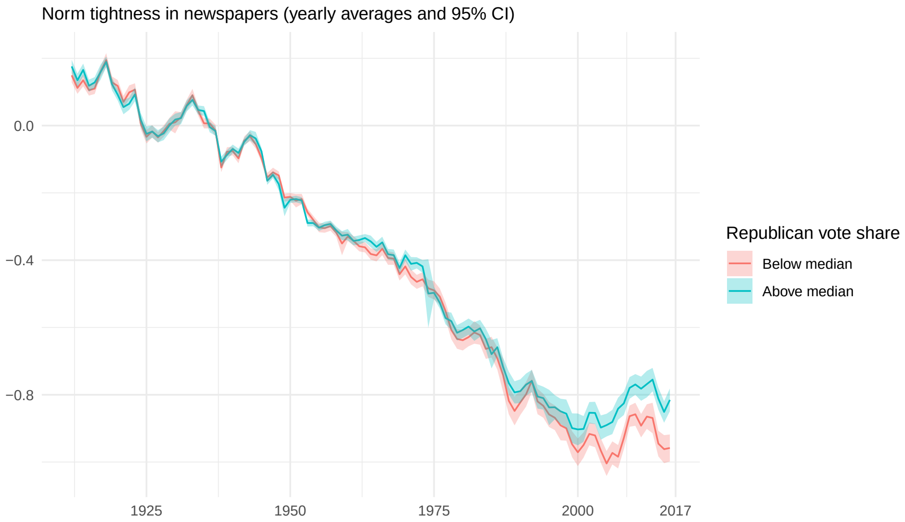
Newspapers in high-immigration places increasingly looser after 1970

▸ newspaperarchive.com



Newspapers in Republican places increasingly tighter after 2000

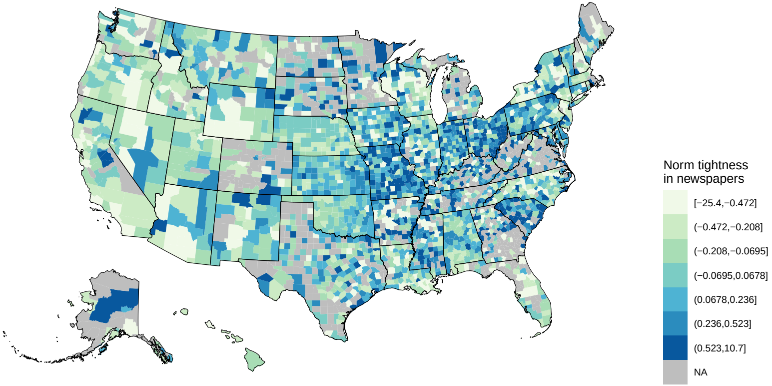
‣ newspaperarchive.com



Norm tightness across US counties

averaged over all years

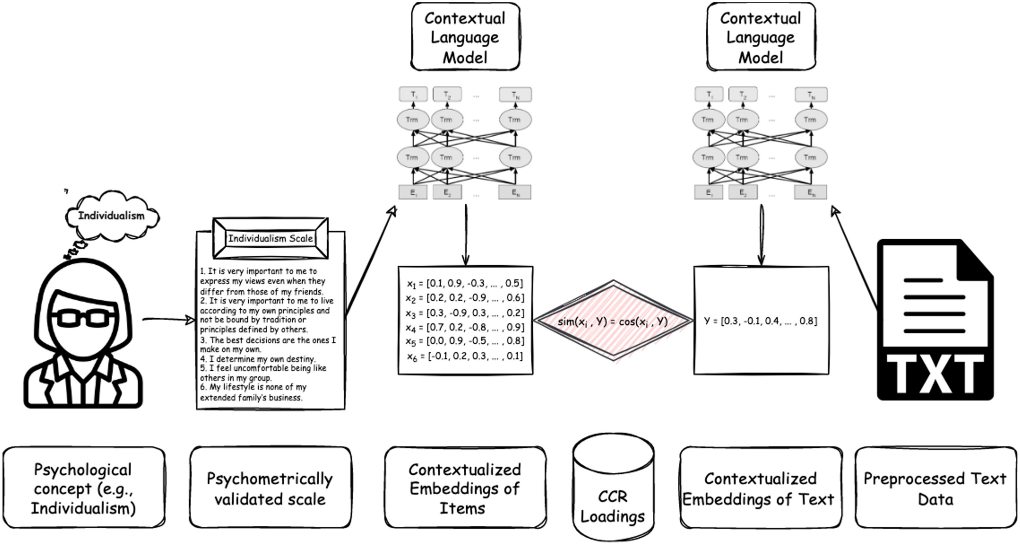
▸ averaged over 1950-2022



$\rho = 0.81$ between norm tightness averaged over all years and averaged over 1950 to 2022

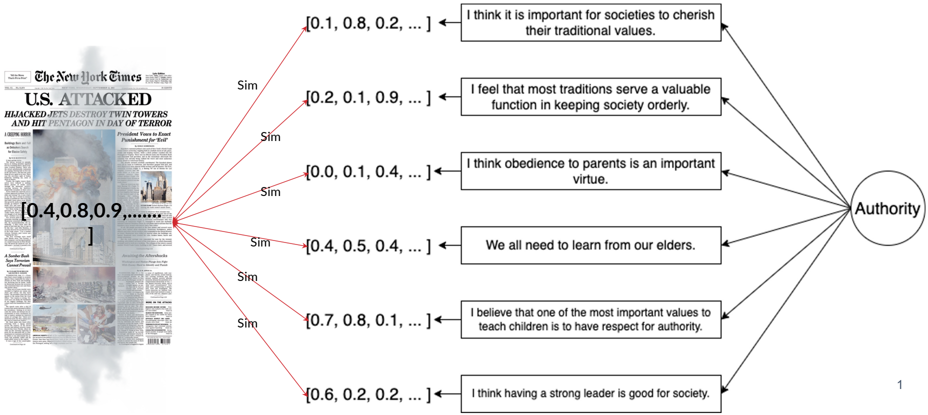
Method: contextualized construct representation (CCR)

Source: Atari, Omrani, Dehghani (2023)



Method: contextualized construct representation (CCR)

Source: Atari, Omrani, Dehghani (2023)



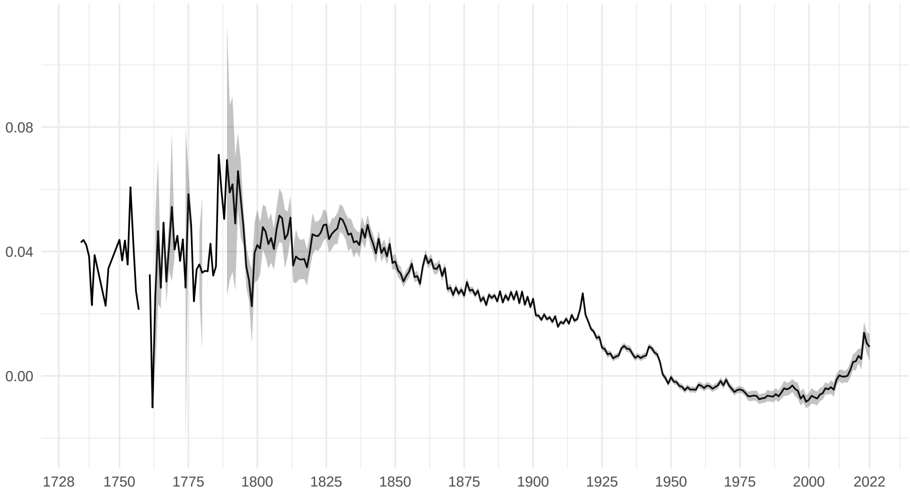
1

Survey questions: 1. There are many social norms that people are supposed to abide by in this country. 2. In this country, there are very clear expectations for how people should act in most situations. 3. People agree upon what behaviors are appropriate versus inappropriate in most situations this country. 4. People in this country have a great deal of freedom in deciding how they want to behave in most situations. 5. In this country, if someone acts in an inappropriate way, others will strongly disapprove. 6. People in this country almost always comply with social norms.

CCR-based measure of norm tightness

▸ Validation

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



- Hope you're convinced that we're capturing some aspects of norm tightness in newspapers

- Let's put this to work by testing core hypotheses of cultural evolution theory:
  - **Do adverse shocks tighten social norms?**
  - **Do looser norms promote innovation?**

Do natural disasters tighten social norms?

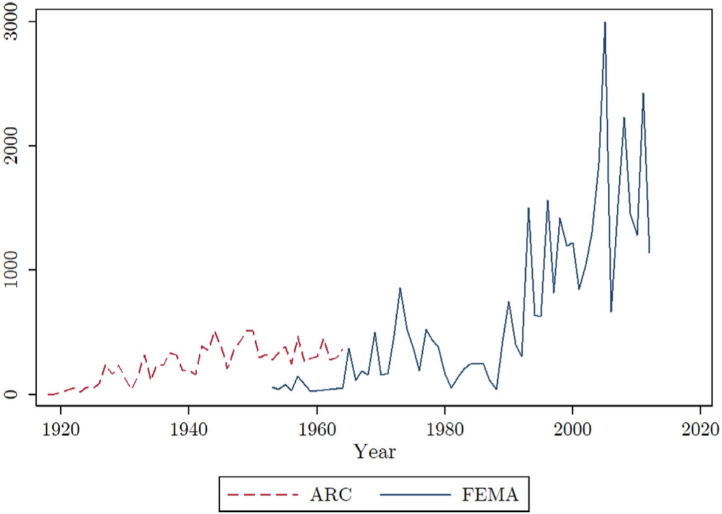
- **Data:** all U.S. federally designated natural disasters from 1918 to 2012, aggregated to county-decade level – Boustan + 2020

- **Empirical strategy:** standard difference-in-differences equation

$$\text{Tightness}_{c(s)t} = \beta \text{ Severe disaster indicator}_{c(s)t} + \alpha_c + \alpha_t + \gamma \times (X_{c(s)} \times t) + \varepsilon_{c(s)t}$$

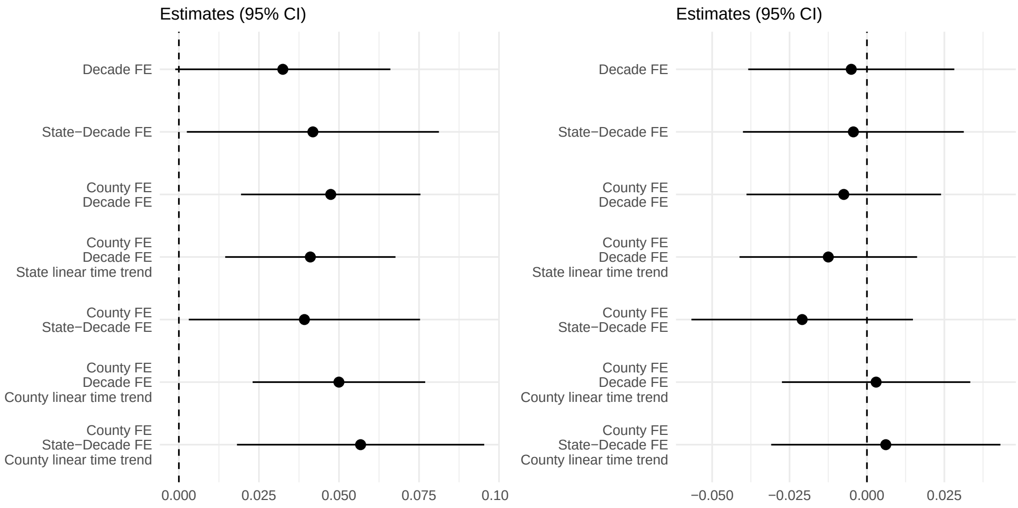
  - $\boldsymbol{c, s, t}$ denote county, state, year
  - **Severe disaster indicator** $= 1$ if a disaster with ≥ 25 fatalities occurs b/w $t - 10$ and t
  - $\boldsymbol{\alpha_c}$ and $\boldsymbol{\alpha_t}$: county and decade fixed effects
  - $\boldsymbol{X_{c(s)} \times t}$: state or county-specific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

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



Boustan, Kahn, Rhode, Yanguas (J Urban Econ, 2020)

Natural disasters increase norm tightness in newspapers



$\Rightarrow$ Severe disasters increase norm tightness by $\approx$ 0.05 s.d. or 5% of gap b/w VT and AR

Do adverse economic shocks tighten social norms?

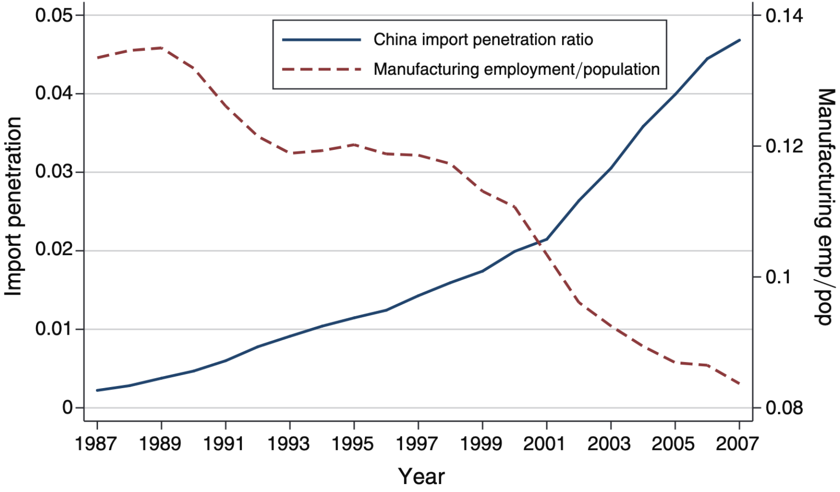
- **Data:** Exogenous variation in local U.S. labor market condtions induced by trade with China b/w 1990 and 2007 — the "China Shock" – Autor + 2013

- **Empirical strategy:** First-difference equation

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

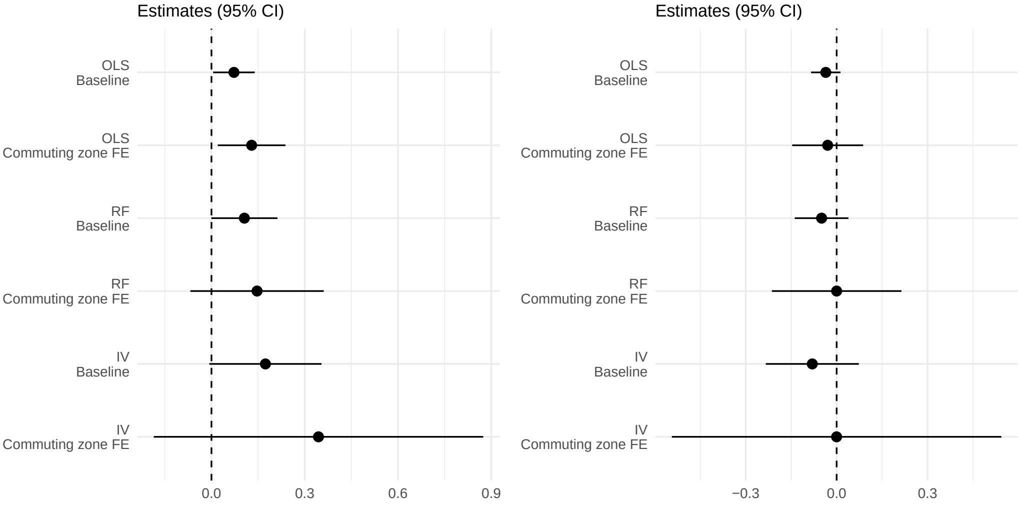
  - $\boldsymbol{c, r, t}$ denote commuting zone, census-region, period
  - **Δ Local trade exposure** is change in local import competition with China from $t-1$ to t
  - $\boldsymbol{X_c}$: baseline employment share in manufacturing
  - $\boldsymbol{\alpha_r}$: census-region fixed effects, removing regional factors that smoothly change over time
  - $\boldsymbol{\alpha_t}$: 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.



Autor, Dorn, Hanson (AER, 2013)

Adverse economic shocks increase norm tightness in newspapers



$\Rightarrow$ Positive estimates of effect of trade exposure (left), while no pre-trend (right)

Does norm tightness affect innovation?

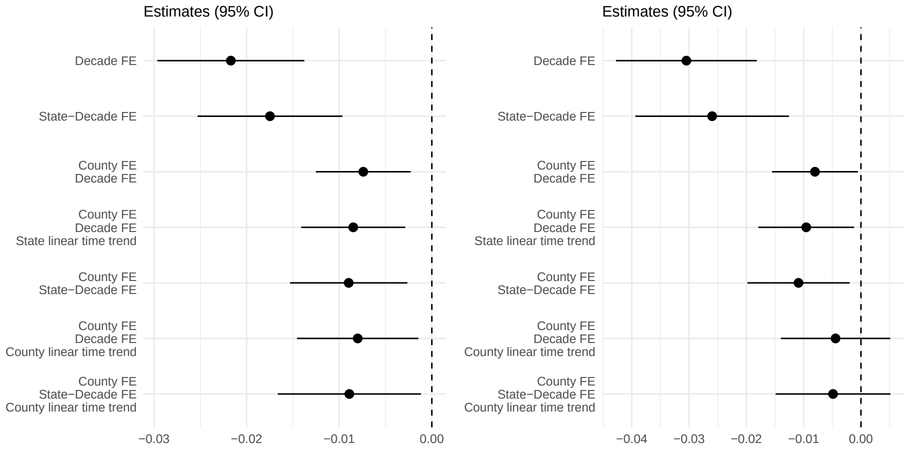
- **Data:** Comprehensive Universe of U.S. Patents (CUSP); all patents issued by the USPTO between 1836 and 2015, aggregated to county-decade level – Berkes 2018

- **Empirical strategy:** standard difference-in-differences equation

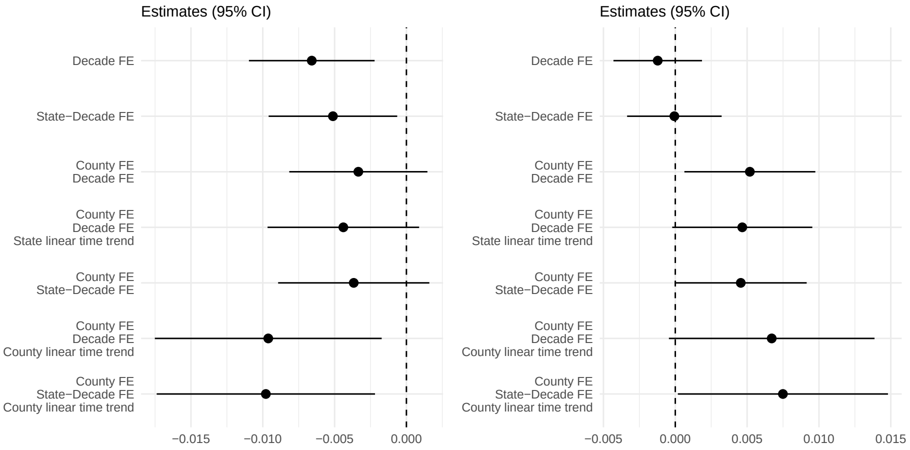
$$\text{Patents p. 10,000 ppl}_{c(s)t} = \beta\,\text{Tightness}_{c(s)t} + \alpha_c + \alpha_t + \gamma \times (X_{c(s)} \times t) + \varepsilon_{c(s)t}$$

  - $\boldsymbol{c, s, t}$ denote county, state, decade
  - **Patents p. 10,000 ppl**: number of patents filed b/w t and $t+10$ normalized by county population in t
  - **Tightness**: avg. tightness among newspapers in i b/w t and $t+10$
  - $\boldsymbol{\alpha_c}$ and $\boldsymbol{\alpha_{st}}$: county and decade fixed effects
  - $\boldsymbol{X_{c(s)} \times t}$: state or county-specific linear time trends, removing local factors that smoothly change over time (e.g., local long-term economic progress)
  - Standard errors clustered on counties

Norm tightness associated with fewer patents p.c., but pretrend



Adding one-period lagged dependent variable



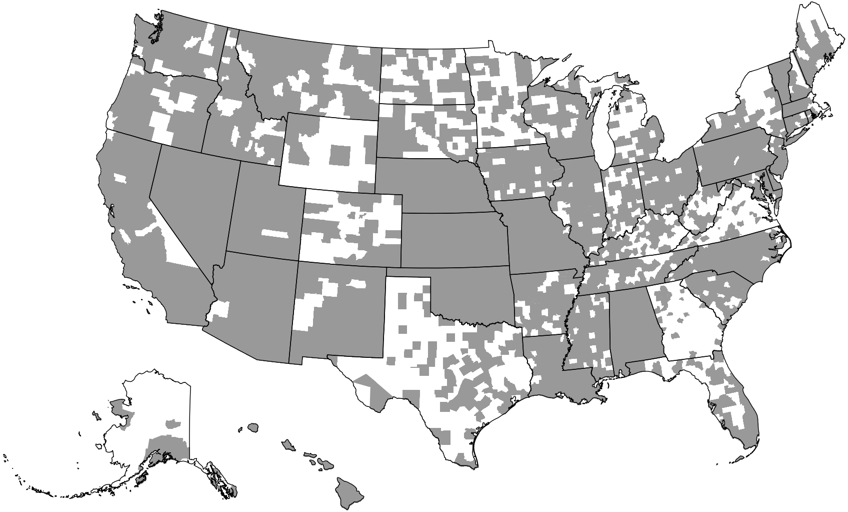
Take-aways

- We study historical psychology through the lens of thousands of local newspapers from all U.S. states and going back to the 1700s
- Consistent with previous studies, we find norms in the U.S. loosened over the past 200 years, while local adverse shocks causes local tightening
- Looser norms are associated with more innovation
- Next steps: measuring more traits (e.g., individualism, moral universalism, religiosity, honor culture, gender norms, etc.)

Appendix

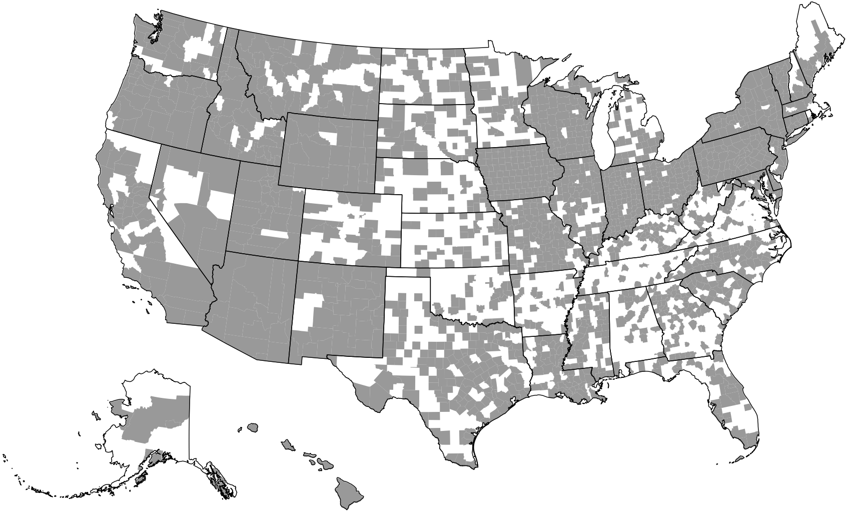
Newspapers.com database covers 2,057 counties from all states

Back



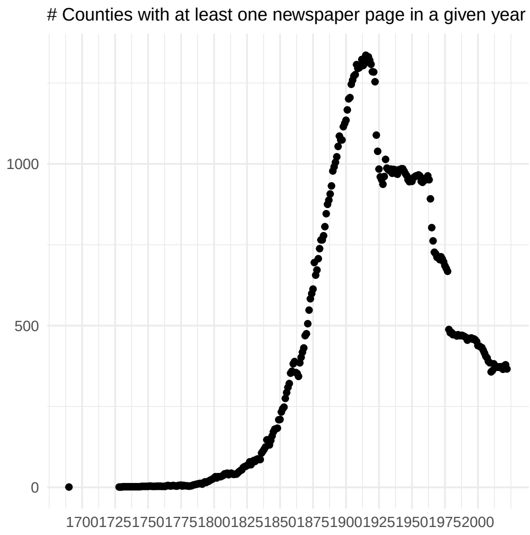
Newspaperarchive database covers 1,866 counties from all states

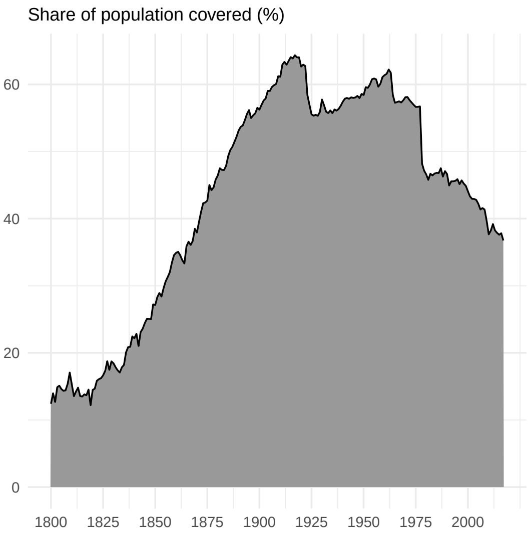
Back



Newspapers.com database

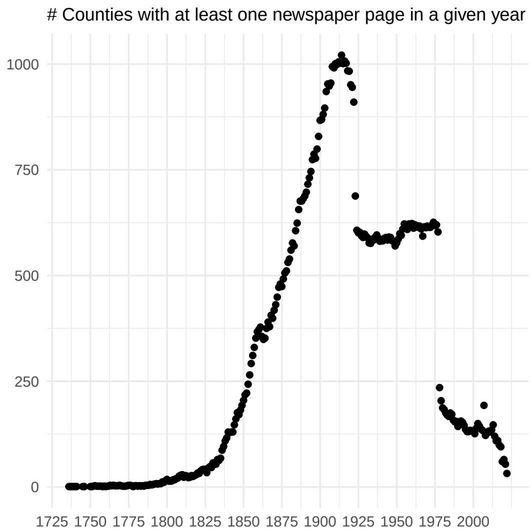
Back

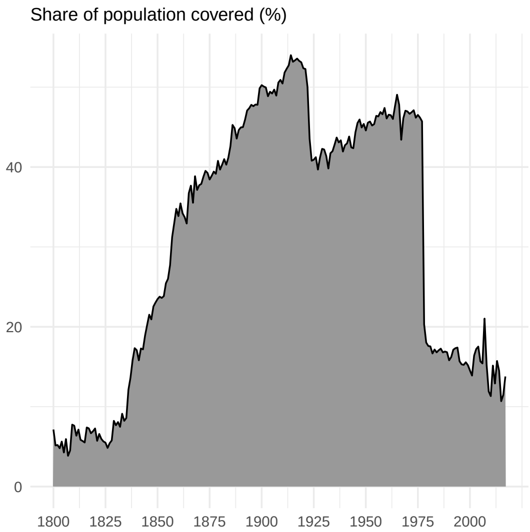




Newspaperarchive database

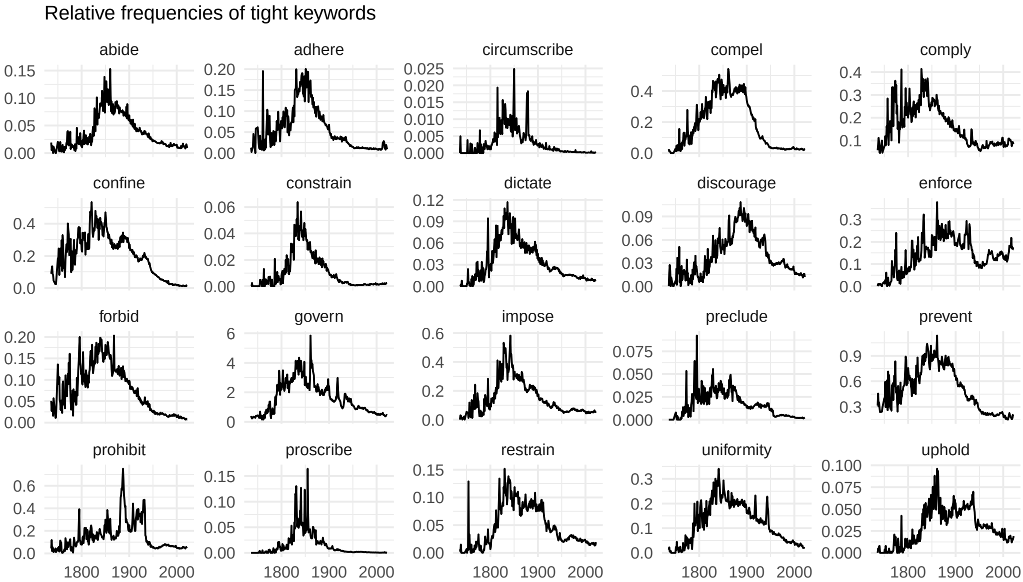
Back





Newspaperarchive sample

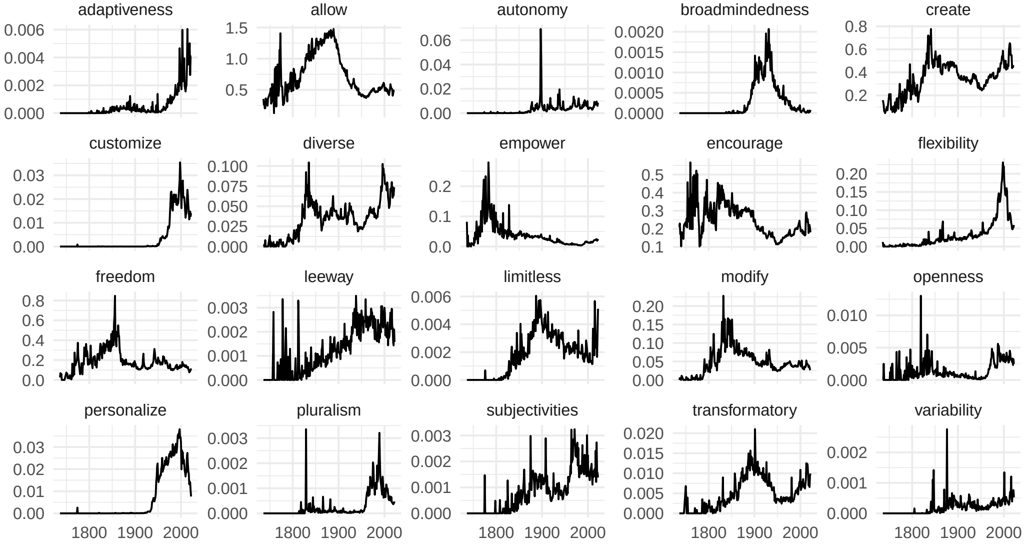
▸ Back



Newspaperarchive sample

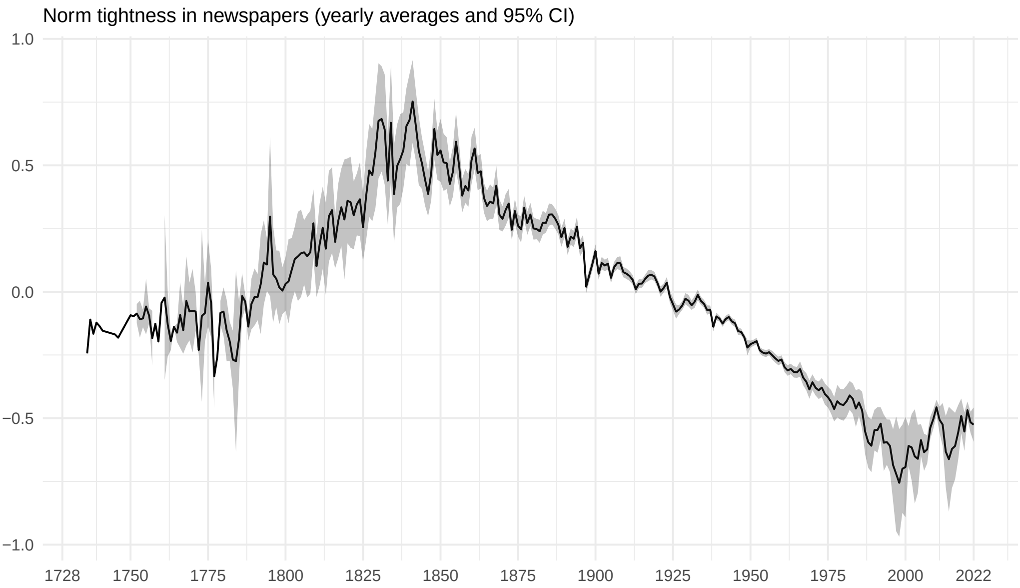
Back

Relative frequencies of loose keywords



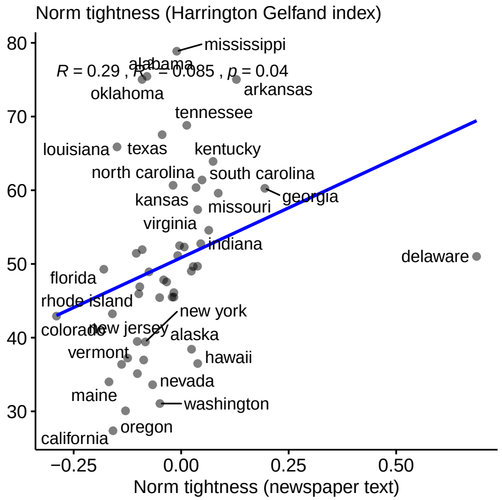
Newspaperarchive sample

▸ Back



Newspaperarchive sample

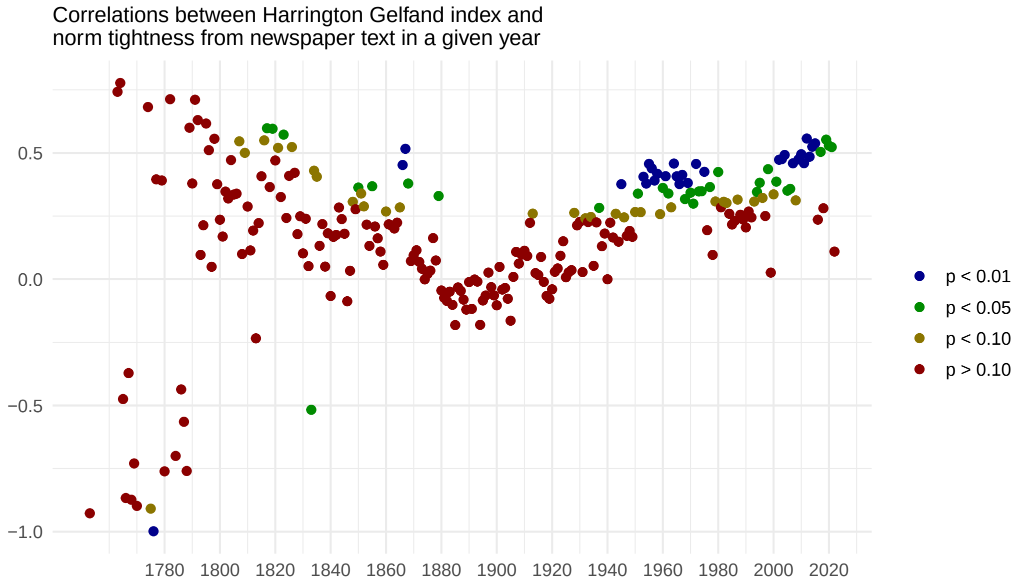
Back



Data shown on y-axis are from Harrington Gelfand (PNAS, 2014).

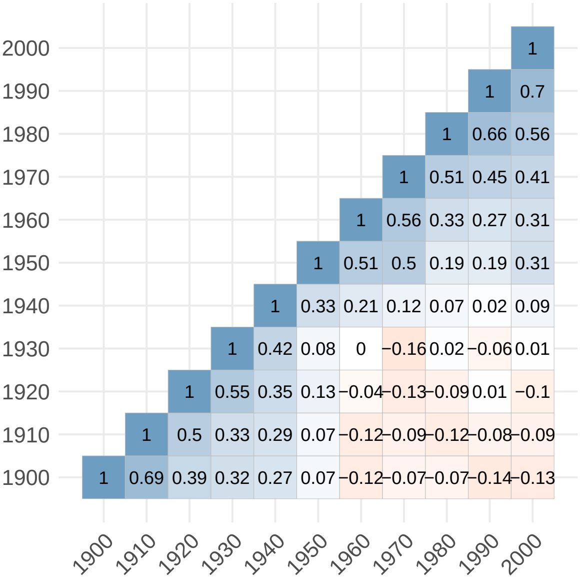
Newspaperarchive sample

▸ Back

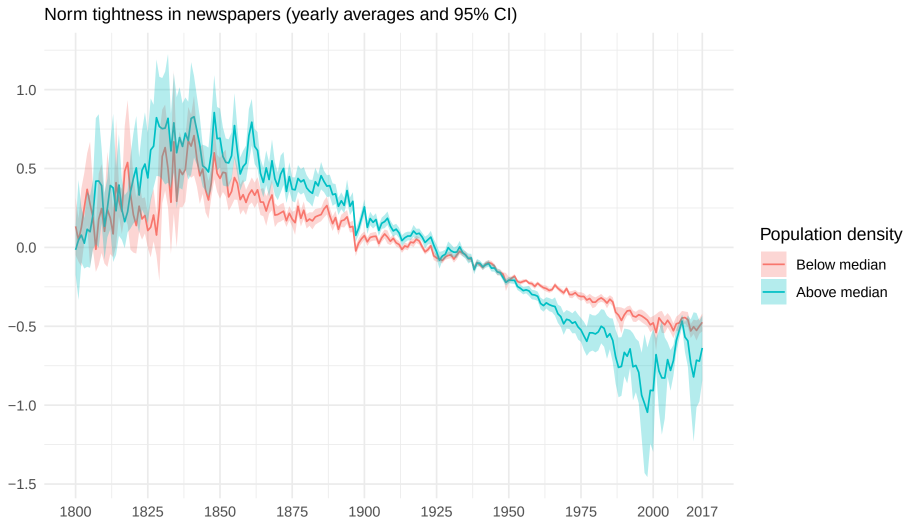


Newspaperarchive sample

Back

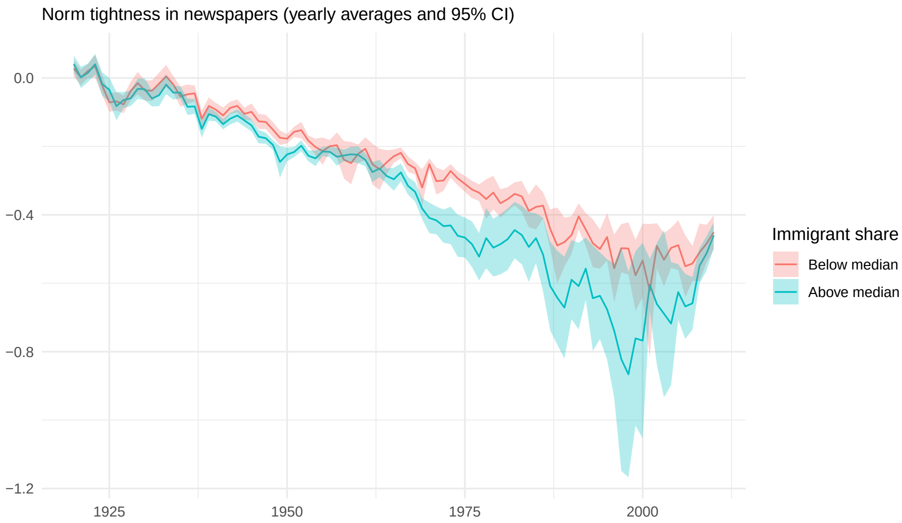


Newspaperarchive sample ‣ Back



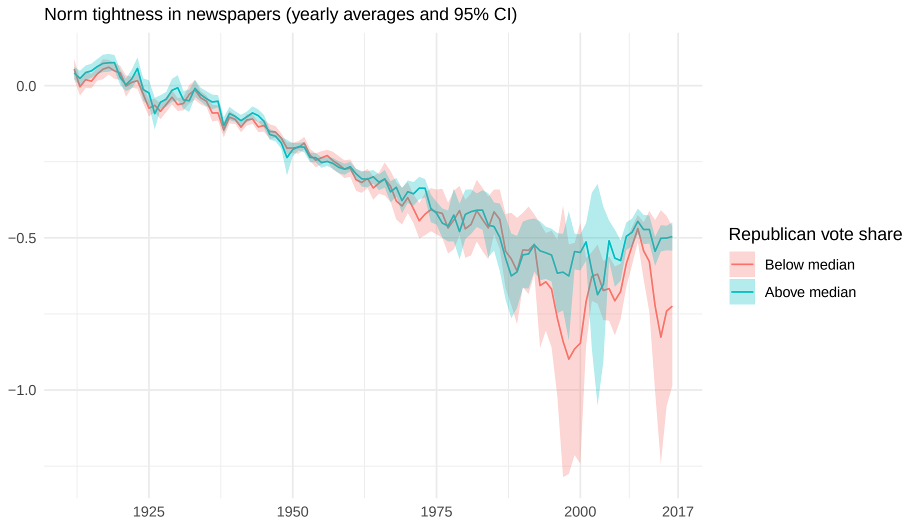
Newspaperarchive sample

Back



Newspaperarchive sample

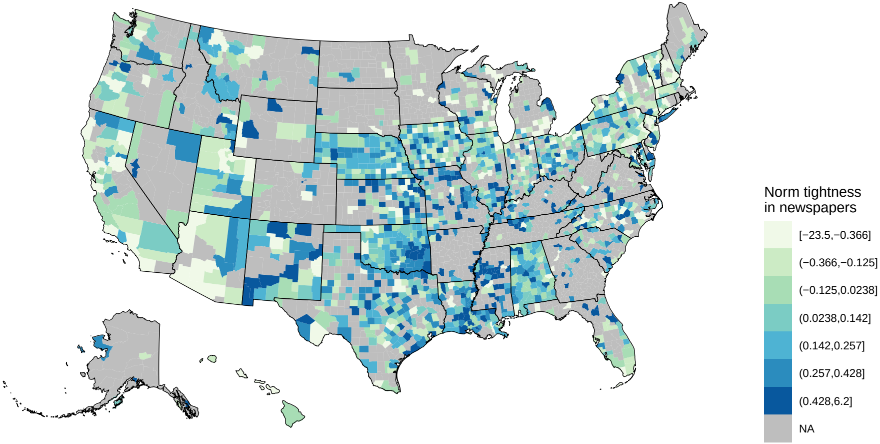
▸ Back

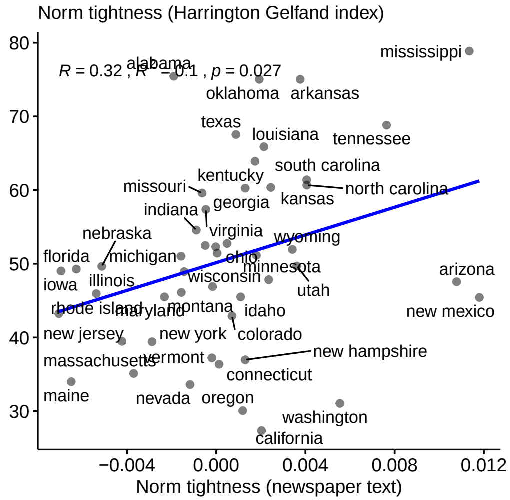


Norm tightness across US counties

averaged over 1950-2022

Back





Data shown on y-axis are from Harrington Gelfand (PNAS, 2014).