Policymaking in the American States, 1787-2020*

Charles Angelucci
MIT Sloan

Elliott Ash
ETH Zurich

Nicolas Longuet-Marx
Stanford University

March 27, 2026

Abstract

The U.S. federal system is often described as a set of "laboratories of democracy," yet the structure of policy diffusion across states and between states and the federal government is difficult to measure. We compile the universe of state statutes from 1787–2020 (2.5 million acts) together with all federal statutes, extract distinct policy adoptions (7.9 million) using large language models with human validation, and group policies into clusters of legally similar policies. We estimate a discrete-choice model of state and federal policy adoption in which each jurisdiction's decision depends on local conditions as well as horizontal (interstate) and vertical (state-federal) diffusion dynamics. Horizontal diffusion is allowed to vary with geographic, economic, and political alignment between states. To address unobserved correlated shocks, we conduct a range of robustness analyses; in particular, our estimates remain stable under progressively stricter fixed effects, including restricting to within state-pair by policy by year variation. We also exploit plausibly exogenous variation in peers' adoption arising from cross-legislature differences in observable characteristics. Geography dominates horizontal diffusion through the mid-twentieth century, economic similarity has a positive and relatively stable effect throughout, and partisan alignment increases sharply after the 1990s but remains comparable to historical levels. Finally, federal legislative activity becomes increasingly predictive of subsequent state legislation after World War II and now exceeds that of any single state, while the predictive relationship from state legislative activity to federal legislation has declined over time.

*We thank Clementine Abed Meraim, Matthew Bonci, Revekka Gershovich, Jairo Gudino-Rosero, Michel Gutmann, Niklas Halfar, and Xianglin Li for excellent research assistance. We are also grateful to Stefano DellaVigna, Mert Demirer, Suresh Naidu, and to participants at seminars at Berkeley, Bocconi, Boston University, Columbia, Harvard, LSE, LUISS, MIT, the Monash-Warwick-Zurich Text-as-Data Workshop, the Paris Empirical Political Economics Seminar, the Third Political Economy Winter Workshop, UChicago, and UCLA Anderson for suggestions that have improved this paper.

1 Introduction

Legislative policymaking in the United States unfolds across a multilayered federal system that divides authority among the federal government, the fifty states, and thousands of local governments. The framers designed this system to balance a limited national center with substantial state autonomy–a principle later codified in the Tenth Amendment and captured in Madison's claim that "the powers delegated by the proposed Constitution to the federal government are few and defined. Those which are to remain in the state governments are numerous and indefinite" (Madison, 1788). This constitutional architecture is understood through two linked dimensions of federalism. *Horizontal federalism* concerns how state legislatures influence one another's policy choices, potentially acting as "laboratories of democracy" in which policy innovations can be developed, tested, and diffused across states (Elazar, 1966, 1987).[1] *Vertical federalism* captures the reciprocal influence between federal and state policies, whereby federal policy choices respond to developments in state policy and, conversely, state policy choices adjust in reaction to changes at the federal level.

Against this backdrop, contemporary scholars and policymakers engage in intense debates over how the United States federal system of legislative policymaking currently operates.[2] One view stresses that the federal government increasingly shapes state policymaking through direct mandates, conditional funding, and regulatory frameworks that blur traditional jurisdictional boundaries (see e.g., Karch, 2006; Karch and Rosenthal, 2016). A second view emphasizes that persistent gridlock in Congress has repositioned state legislatures as central venues for consequential policy change (e.g., Toscano, 2021). Health care expansion under and beyond the Affordable Care Act, the proliferation of state level reproductive rights statutes, and state action on marijuana legalization, gun regulation, immigration enforcement, and data privacy illustrate the renewed centrality of state legislatures in shaping policy. These developments unfold alongside the rising nationalization and polarization of party politics, as voters increasingly anchor their preferences in national rather than local considerations (Hopkins, 2018) and national interest groups intensify their focus on state politics (see, e.g., Hertel-Fernandez, 2019*a*), thereby raising the question of whether states still function as meaningful laboratories of democracy.

These developments motivate three broad questions that guide our analysis. First,

[1] A term famously used by Justice Louis Brandeis in *New State Ice Co. v. Liebmann* (1932).

[2] A large contemporary literature debates the proper role of states and the overall health of the U.S. federal system. Some scholars portray federalism as a "tragic compromise" rooted in managing deep territorial cleavages rather than in any inherent normative virtue (Feeley and Rubin, 2008), while others emphasize the promise of overlapping, "polyphonic" authority structures for protecting fundamental rights (Schapiro, 2009).

how much do state policies diverge across the United States, and how has this divergence evolved over the last two centuries? Second, what forces shape similarities and differences in state policy choices–geographic proximity, economic structure, or partisan alignment? Third, to what extent do state policies drive or respond to federal initiatives? Recasting these questions in terms of vertical and horizontal federalism, we ask: How strongly does federal legislation influence state policymaking, and conversely? Which states exert the greatest diffusion influence on their peers, and along what dimensions–geographic, economic, or partisan–do states most strongly emulate or differentiate themselves from one another?

Addressing these questions empirically requires a way to observe policy choices at a fine level of detail across all states and over a long horizon. We therefore assemble a new corpus comprising the universe of enacted state statutes from the creation of each state legislature through 2020 (about 2.5 million acts), alongside all federal statutes enacted over the same period (roughly 80,000 acts). From this four-million-page corpus, using state-of-the-art AI models and embedding techniques, we extract policy content from these legal texts and create clusters of legally similar policies, which we then validate with legal experts. This approach allows us to measure legal similarity between individual states over time, as well as between states and federal legislation. We obtain roughly 2,500 distinct clusters of legally similar policies. On average, each cluster spans many decades and is populated by statutes from numerous states as well as the federal government, providing a natural unit for tracking policy adoption and diffusion.

We use these clusters to construct a cluster-level panel of legislative activity and to estimate a model of policy diffusion. For each policy cluster c, state s, and year t, we observe whether the state adopts at least one policy in that cluster-year. We model this adoption decision using a discrete-choice framework in which a legislature's net propensity to legislate in (s, c, t) depends on (i) the state's time-varying political and economic conditions, (ii) its own recent history of legislating in that cluster, and (iii) prior adoption activity by other states and the federal government in the same cluster. Interdependence is allowed to vary systematically with pairwise similarity: we estimate separate decade-by-decade amplification effects for geographic proximity, economic similarity, and partisan alignment, so that (for example) adoption by a co-partisan or economically similar state may shift incentives more than adoption by a distant or politically opposed state.

A central challenge is that exposure to other jurisdictions' prior adoptions may be correlated with unobserved shocks or latent similarities. Our baseline identification include high-dimensional state-by-year fixed effects that absorb each state's overall propensity to legislate in year t, and cluster fixed effects that absorb persistent differ-

ences in baseline adoption rates across clusters. We also allow the effects of observed state-year covariates to be cluster-specific, so that the same political or economic shift can reallocate legislative effort across policy areas rather than mechanically raising or lowering adoption everywhere. We then probe the sensitivity of the diffusion estimates using two increasingly demanding strategies. First, we progressively strengthen the fixed-effect structure—adding state-pair × policy × year fixed effects—so that identification comes from within-pair asymmetries in adoption timing (situations in which one state in a pair has adopted recently while the other has not). Second, we instrument other jurisdictions' adoption stocks using plausibly exogenous shifts in their observable characteristics (legislative capacity, political composition, and economic composition), implemented with cluster-specific first stages and a high-dimensional instrument set selected via Lasso. These strategies do not eliminate every threat to a causal interpretation, but they substantially narrow the scope for confounding by common shocks and slow-moving latent similarity. Consistent with this, placebo tests show that future adoptions, instrumented using future observable characteristics, do not predict current adoptions.

Our results point to three broad findings. First, horizontal diffusion is systematically stronger among similar states, but the relevant notion of similarity has shifted over time. In earlier decades, geographic proximity is the strongest predictor of spillovers; over the long run its importance declines, consistent with a weakening role for purely local or regional channels. Economic similarity is positive and comparatively stable throughout the sample. Partisan alignment is modest early on but rises sharply beginning in the late twentieth century, eventually reaching magnitudes comparable to (and, in the most demanding fixed-effect specifications, in some decades exceeding) the geographic and economic channels. Second, diffusion is highly asymmetric across senders: some states consistently act as outsized sources of policy spillovers, and the composition of influential senders changes over time, indicating that horizontal federalism is shaped not only by similarity networks but also by heterogeneous "influence" across legislatures. Third, we document a pronounced rebalancing of vertical federalism. The influence of federal legislation on subsequent state policymaking rises markedly over time: by the 2000s, moving from no recent federal activity in a cluster to sustained federal activity throughout a ten-year window is associated with an increase of roughly 2.5 percentage points in a typical state's probability of adopting in that cluster. In the reverse direction, the responsiveness of federal adoption to aggregate state adoption activity declines over time, so that federal-to-state spillovers dominate state-to-federal spillovers in recent decades. Unlike for horizontal diffusion, we do not have a natural instrumental-variables strategy for these vertical coefficients, so a more cautious inter-

pretation is that federal activity has become increasingly predictive of subsequent state activity within clusters, while the reverse predictive relationship from state activity to federal adoption has weakened over time.

To quantify the role of these interdependencies, we conduct two counterfactual exercises. First, taking adoption histories as of 2000 as given, we shut down similarity-based spillovers (geographic proximity, economic similarity, and partisan alignment) separately and jointly, and trace the implied change in subsequent adoption probabilities. Each channel reduces diffusion when removed, with partisan alignment generating the largest single-channel effect in the post-2000 counterfactual; shutting down all three channels produces a cumulative decline that grows over time as diffusion opportunities compound. Second, we evaluate the influence of key actors by reducing their estimated sender effect to that of an average state legislature starting in 2000. Implementing this for California and for the federal government yields sizeable reductions in subsequent adoption across states, with pronounced heterogeneity across policy domains: federal influence is especially important in nationally oriented regulatory areas (e.g., trade, immigration, international affairs), whereas California's influence is comparatively diffuse and particularly salient in domains such as environment, public lands, health, labor, and education. In several domains, California's estimated influence is comparable to that of the federal government.

These findings come with important caveats. Our outcomes capture statutory policymaking and therefore miss policy change implemented through administrative rulemaking, courts, or direct democracy. We measure adoption rather than effectiveness: policies that diffuse rapidly are not necessarily those that work best, and highly effective policies may diffuse slowly. Our diffusion framework is intentionally myopic—it summarizes how past adoption shifts current adoption incentives, not the equilibrium of a forward-looking dynamic policy game—and the estimated "influence" parameters should be read as descriptive spillover measures rather than strategic returns to innovation. In addition, our instrumental-variables strategy is feasible only from the 1970s onward due to data availability, so causal claims based on IV pertain to the modern period rather than the full historical sample. Finally, despite rich fixed effects, lag structure, and the complementary IV design, we cannot fully rule out remaining correlated shocks or latent similarities that jointly affect adoption propensities and measured exposure, especially along partisan or regional lines. We therefore interpret the estimates as disciplined characterizations of policy interdependence (and how it evolves over time), rather than as uniformly causal effects in every dimension and period.

Related Literature. These findings add to several strands in political economy. First, we contribute to the literature on policy diffusion among U.S. states. Classic work, rooted in Walker (1969), emphasizes learning, competition, and emulation as key drivers of diffusion (see Mooney, 2021, for a recent review).[3] Desmarais, Harden and Boehmke (2015) use timing of adoption data on 187 policies (1960–2009) to infer the directed network that best explains the observed patterns of adoption (see also Boehmke et al., 2020). They find that policy diffusion across U.S. states is shaped more by capacity and similarity than by geography, with partisan alignment representing a modest channel. Recent work by de Paula, Rasul and Souza (2024) shows that tax competition across U.S. states is not confined to immediate neighbors: states often respond to noncontiguous but geographically proximate and demographically similar states, and partisan alignment between governors reduces the likelihood that they compete in taxes. Using a dataset of 602 state-level policies, DellaVigna and Kim (2025) show that geographic proximity best predicts policy adoption before 2000, while partisan alignment dominates thereafter (Mallinson, 2021, see also), consistent with increasing ideological polarization across state legislatures Shor and McCarty (2011); Grumbach (2018).

We contribute to this literature by constructing a dataset covering all policies adopted by U.S. state legislatures and the federal government since their inception, and by explicitly modeling the two-way interaction between state and federal policy-making, alongside horizontal interactions across states. By considering both diffusion across states and diffusion between states and the federal government, we further connect to the recent literature on the increasing nationalization of state politics (Hopkins, 2018; **?**). Complementing our focus on the United States, recent work on China shows how central vs local incentives shape both diffusion and policy fit. Wang and Yang (2025) document that large scale policy pilots are selected into more developed localities and can bias learning and national rollouts. Luo, Wang and Yang (2025) assemble a universe of policies across levels of government and show a sharp move after 2013 toward centralization.

Second, we relate more generally to network-based approaches in political economy. Acemoglu, García-Jimeno and Robinson (2015) model state capacity as a network game between municipalities and the national government. Battaglini and Patacchini (2018) study influence and targeting when legislators are connected by social ties, and Canen, Jackson and Trebbi (2023) develop a model of endogenous social interactions in the

[3]For theoretical analyses of policy experimentation in federal systems, see Volden, Ting and Carpenter (2008), Cai and Treisman (2009), and Callander and Harstad (2015). Bernecker, Boyer and Gathmann (2021) studies the extent to which reelection motives shape governors' incentives to engage in policy experimentation.

legislative process and study implications for legislative activity. Harmon, Fisman and Kamenica (2019) exploit seating rules in the European Parliament to identify peer effects in legislative voting.

Third, we build on prior analyses of state legislation. Previous work has shown that more populous states enact longer bodies of legislation (Mulligan and Shleifer, 2005), and that higher legislative output correlates with stronger economic growth (Ash, Morelli and Vannoni, 2025).

Finally, we add to the literature on text as data in economics (Gentzkow, Kelly and Taddy, 2017; Ash and Hansen, 2023). Prior work documents cross party differences using text scaling and supervised methods (Gentzkow and Shapiro, 2010; Gentzkow, Shapiro and Taddy, 2017; Ash, Morelli and Weelden, 2017; Ash, Morelli and Osnabrügge, 2019; Widmer, Galletta and Ash, 2020); newer work leverages unsupervised embeddings (Gennaro and Ash, 2021). The use of large language models to summarize legal text is recent (Ash et al., 2024), as is the application of transformer based document encoders (Arold et al., 2024). We show how to validate and implement these tools for large historical corpora.

The remainder of this paper is organized as follows. Section 2 provides institutional background on vertical and horizontal federalism. Section 3 describes the legislative corpora, the extraction of policy adoptions, and the construction and validation of policy clusters. Section 5 presents the diffusion model and estimation strategy. Section 6 presents the empirical results on horizontal and vertical spillovers, together with robustness and counterfactual exercises. Section 8 concludes.

2 Institutional Background

A defining feature of the United States is the substantial autonomy of its fifty state governments. Except for Nebraska's unicameral legislature, each state maintains a bicameral body that typically meets less often and with fewer members than Congress. Governors hold veto authority and propose budgets and other major initiatives. Although the scope of their powers varies across states, governors are always key in shaping state-level policymaking.

States exercise primary authority over education, health, taxation, and much of criminal law, while the federal government focuses on defense, interstate commerce, and national taxation. Yet many significant domains, including environmental regulation, labor standards, and civil rights protections, fall under overlapping federal and state authority (Elazar, 1966). States are free to adopt rules that exceed federal minima or

to legislate in areas left unaddressed by Congress.

These institutional arrangements motivate classic public-economics views of federalism, which posit that decentralization exploits local information and preference heterogeneity, with mobility and yardstick competition disciplining governments (Tiebout, 1956; Oates, 1999; Besley and Case, 1995; Besley and Coate, 2003). Political economy accounts emphasize how institutional limits on the center enable subnational experimentation and competition, and how decentralization shapes accountability (Weingast, 1995; Mookherjee, 2015).

Vertical Federalism. Under the Constitution's Supremacy Clause, valid federal statutes preempt conflicting state laws. In practice, however, Washington shapes state policymaking through multiple channels. Congress may displace state statutes when uniform national standards, whether in aviation or banking, are deemed essential. Under its Spending Power (Art. I, §8), federal grants, ranging from tightly conditioned categorical funds such as Medicaid to broader block grants, require states to embed specified program designs and civil rights safeguards (Oates, 1999). Cooperative federalism frameworks, such as Clean Air Act implementation plans and Section 1115 waivers in Medicaid, with analogous waivers under the Every Student Succeeds Act, permit state experimentation within federal guardrails (McCubbins and Schwartz, 1984). Federal courts further delimit state prerogatives by interpreting constitutional provisions, including the Equal Protection Clause and the First Amendment, and by striking down statutes that exceed those bounds. Private suits under Section 1983 and enforcement actions by the Department of Justice impose additional constraints. Agency rulemaking, guidance documents, and model legislation promoted by federal bodies and allied advocacy groups also exert pressure on state policymakers.

State innovations can, in turn, lead to national reform. When a jurisdiction pioneers a novel program, such as Massachusetts's early health insurance exchange model, federal legislators and regulators adapt its features into national policy, as with the Affordable Care Act's federal exchanges. Renewable Portfolio Standards first enacted by Texas and Iowa informed subsequent federal tax credits and grant programs for clean energy. Initial state statutes recognizing same sex marriage generated legislative histories and public testimony that helped shape later congressional debates and, ultimately, Supreme Court consideration (Berry and Berry, 1990).

Horizontal Federalism. Beyond vertical channels, states influence one another directly through policy diffusion. Acting as "laboratories of democracy,"[4] jurisdictions observe the legislation, implementation experiences, and political consequences of policies adopted by their peers and update their own choices accordingly. Diffusion operates through multiple mechanisms: informational learning from early adopters; competition for mobile capital, labor, or federal funds; coordination within regional blocs; and pressure transmitted through interstate interest group networks, professional associations, and party organizations. For example, California's 2004 paid family leave program was followed by similar statutes in New Jersey, Rhode Island, and other states, while the legalization of recreational marijuana in Colorado and Washington in 2012 triggered a cascade of subsequent adoptions. These horizontal processes, shaped by policy learning, strategic interaction, and shared political alignments, generate clusters of policy innovation that extend beyond individual state borders (Rogers, 2003; Berry and Berry, 1990).

Our analysis models how the combined forces of vertical and horizontal federalism interact to produce and propagate clusters of policy innovations across the American federal system.

3 Measurement of Policy Output from Legislative Texts

This section lays out our method for extracting and clustering policies using the raw text of state and federal legislation.

3.1 Data Sources

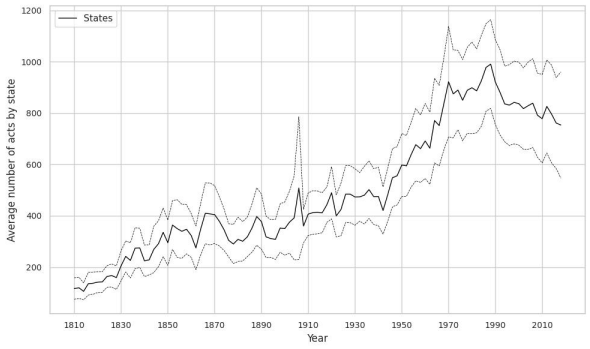
State Session Laws Corpus. Our state legislative records consist of the full text of U.S. State Session Laws, sourced from HeinOnline. The session laws contain legislative acts enacted by each state legislature during each session, among other material. This corpus has been used in several papers in economics and political science (Ash, 2018; Vannoni, Ash and Morelli, 2021; Ash, Morelli and Vannoni, 2025). The corpus comes as plain-text scans of the session laws, which are published as annual or biennial journals (Figure A.1).

The texts are generated by Optical Character Recognition (OCR) applied to printed legislation. When examining samples, OCR is mostly of high quality. Some laws are

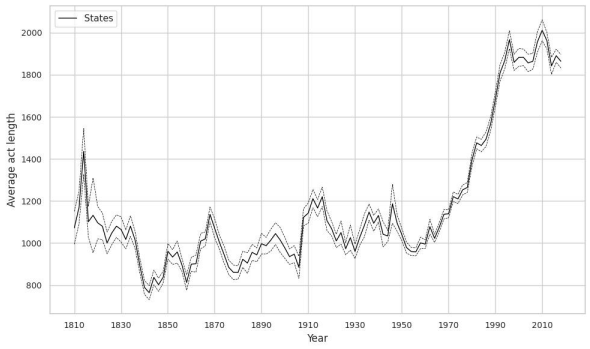
[4] Justice Louis Brandeis in *New State Ice Co. v. Liebmann* (1932).

Figure 1: Evolution of the number and length of acts

(a) Average number of acts per state



(b) Average number of words per act



Notes: The left panel shows the average number of acts per state per biennial legislative period. The right panel shows the average number of words per act per biennial legislative period. We report the 95% confidence interval of the mean.

old, but the quality of digitization is quite good. Nonetheless, as with any historical digitized corpus, there are a non-negligible number of OCR errors. To investigate this, we calculated a proxy OCR error rate as the common (non-proper) noun misspelling rate. The misspelling rate is not that high (about 3% of common nouns), but still significant enough to help justify our use of an LLM for policy extraction, which is quite robust to OCR errors.

The laws are published once a year or every two years. To ensure consistency, the dataset is created every two years.

U.S. Federal Statutes at Large. We use the Federal Statutes at Large, containing equivalent legislative act texts for the U.S. federal government. These are also obtained from HeinOnline, with similar digitization quality.

Text Pre-processing. The first step is to merge and process this raw text. A script serves to append pages, remove headers, footers, tables of contents, indexes, and other non-statute material. Then, it segments the text into individual acts using text markers for the start of new statutes. These include indicators for new Chapters, Articles, or Titles, such as a line with CHAPTER followed by a Roman numeral. Some states have their own standard indicators, such as P.A followed by a number to indicate a new Public Act. The script also uses common text for the beginning of a statute preamble (e.g., An act to...) and for enacting clauses (e.g., Be it enacted that...). Research assistants checked samples of the statute segmenter for each state-biennium

to make sure it worked well. Figure 1 Panel A plots the average number of acts per state over time.

Next, we adjust the text and formatting of each statute to make them more comparable across states and over time. For each statute, we normalize the text to lowercase, remove punctuation, function words, and extra white space. Figure 1 Panel B plots the average number of words per act over time.

Finally, we perform a post-processing step to limit the included texts to *acts*, rather than *resolutions*. Acts are binding legislative measures that establish, amend, or repeal statutory rules, whereas resolutions are nonbinding expressions of intent, commemoration, or internal chamber procedure. By filtering out these symbolic or procedural items, the dataset focuses on substantive legal changes that alter the operative body of law. This classification is done with a combination of dictionary matching and LLM annotation. See Appendix Figure A.9 for the prompt and other details.

Political, economic, and geographic data. We combine several datasets in order to collect information on the demographic, economic and political conditions of each state. Economic data come from the NAICS and SIC sector shares for each year from 1940 onward, from Klarner (2015). We complement this data with NHGIS Employment by industry from decennial census files going back to 1810 (Manson, 2020). The core legislative professionalism components—session length, legislator compensation, legislative expenditures, and multidimensional scaling scores—are drawn from Legislative Professionalism Component Scores (Greene and Bowen, 2020), with additional imputation performed by the authors and indicated by imputation dummy variables. Measures of committee structure, agenda control, leadership power, and seniority are constructed from the replication data of Fouirnaies and Hall (2017), while alternative and extended professionalism measures—including normalized session length, salary, staffing, and composite professionalism scores—are based on the updated Squire Index data (Squire, 2023).

3.2 Policy Extraction

This section describes the AI-based approach for extraction of policies using a large language model and the subsequent encoding using a transformer-based context-sensitive encoder.

LLM extraction of policy content. First, we take the text of acts and extract their policy content. We use a lengthy LLM prompt, with the main instructions listed

in Figure A.2 and supplementary instructions listed in Figure A.10. All LLM prompts are applied to the text with the OpenAI API using GPT-4.1 Mini (gpt-4.1-mini-2025-04-14) model, with temperature zero.

The prompt works as follows. First, it sets out the overall task of transforming a plain text document to a list of policies, defining a "policy" as "a self-contained government action or bundle of closely connected government actions with a socioeconomic (societal and/or economic) goal or impact." This definition is based on standard concepts from the public policy literature (Peters, 2021). Second, for each policy, the model is asked to distinguish introductions of new policies, modifications of existing policies, and repeals of existing policies. We will exclude repeals from our dataset of policies. Third, the extracted policy is summarized in one sentence, which "can be as long as needed, to capture the full detail of the associated policy." Finally, the LLM should supply a short label (up to five words) summarizing the policy.

The LLM summaries produced with this prompt are extremely high-quality. Figure A.3 shows some examples extracted from a 1986 California Statute (top panel) and a 1957 federal statute (bottom panel). The summaries effectively articulate the policy content of a statute, even for historical documents and in the presence of OCR errors. From inspections of samples, we can say that all summaries are reasonable.

We validate the performance of the policy summaries by hiring three legal experts (law school graduates) to evaluate them along four dimensions: missingness (whether the summary omits relevant information), additions (whether it introduces information not present in the text), mistakes (whether it contains inaccuracies), and neutrality (whether it adds meaning or interpretation beyond the original text). Appendix Figure A.14 reports the results of this validation exercise. In 13% of cases, legal experts judged that the summary omitted significant information from the statute. In 4% of cases, they judged that it contained an inaccurate description; in 1% of cases, that it added information; and in 0% of cases, that it added meaning or interpretation.

The prompt can extract multiple policies. On average, each statute contains 3.55 policies (St Dev: 4.53). See Figure A.4 for the distribution of policies per cluster. Overall, we extract roughly 7,900,000 policies (not necessarily distinct). That includes 5,214,671 introductions, 2,920,334 modifications, and 244,334 repeals. Policy repeals are excluded from subsequent analysis.

Anonymization of persons, places, and periods. Next, we want to adjust the content of the summaries to further focus on policy content and remove information about the institutional context, such as information about time and place. We need to do this because we want the policies to be comparable across time and space. We

undertake an additional LLM-based processing step to "anonymize" the content of the policies. We need to do this because we are going to vectorize and cluster the policy texts, and if we did not do the adjustment, the resulting similarity metrics and clusters would load substantially on time and place characteristics. For example, you would get a Texas cluster and an Idaho cluster, rather than a tax cluster and a spend cluster.

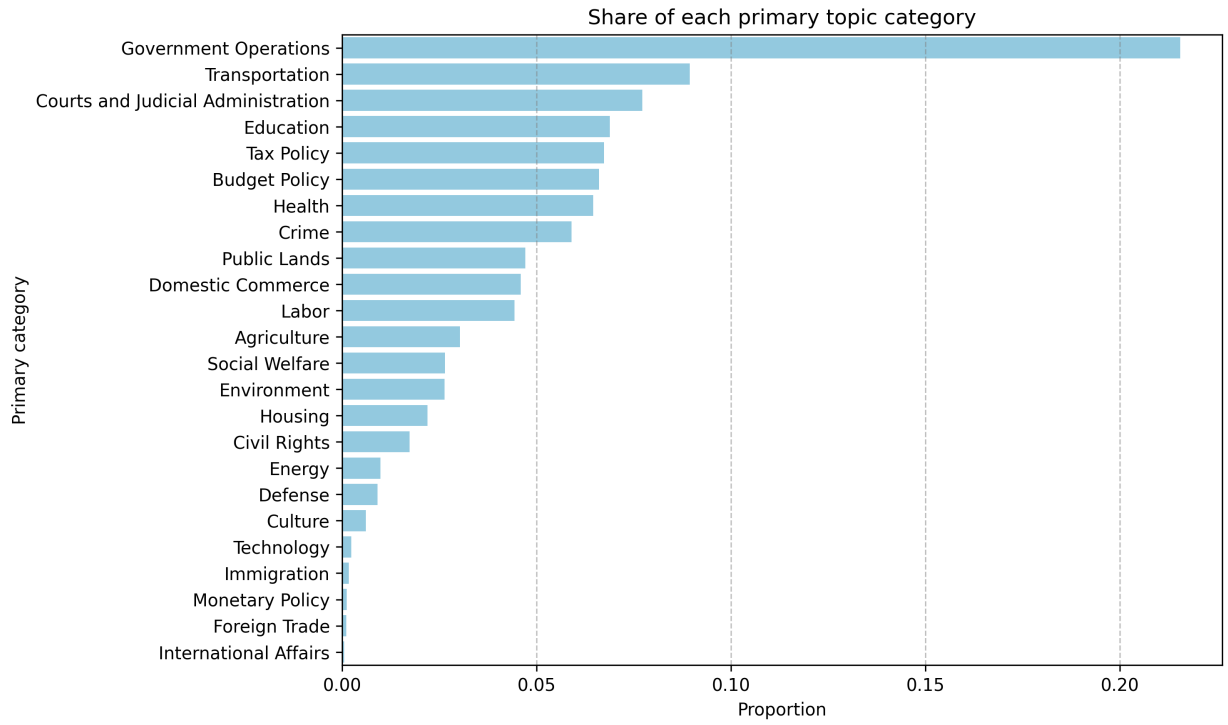
The LLM prompt for anonymizing the policy extracts is shown in Figure A.5. We carefully adapt the prompt through an iterative procedure to remove information about locations, but in a way that maintains the important policy information. In the included example on railroad incorporation, "Four Mile Valley Rail Road Company' becomes "a specific railroad company," "Hamilton" becomes "a specific city", "Ohio" becomes "a specific state" and "Indiana" becomes "another specific state". The policy-relevant information is maintained, but the information about the local context is excised.

Assignment of broad topics to policy extracts. Next, we use LLM prompting to assign a policy topic and subtopic to each statute summary using a set of topics (see Table B.1 for a list of topics and subtopics). Since a policy can address multiple topics, we allow for a primary topic and a potential secondary topic. The topic assignment prompt is listed in Figure A.6 (see Figure A.11 in the appendix for subtopics). Figure 2 shows summary statistics on the assigned topics.

Identifying Similar and Dissimilar Policy Pairs. We want to identify groups of related policies in a data-driven manner. To do so, we first need labels on whether policy texts are related or not. Starting with the extracted and anonymized policies from above, we construct a sample of 250,000 policy pairs using a weighted sampling procedure designed to oversample semantically similar pairs that come from the same legislative topic.[5] We then query an LLM to code the policy pairs as related or not. The prompt (see Figure A.7) defines policies as "similar" if they describe the same policy or very closely related policies. A number of examples are provided to indicate dissimilar policies: regulating different technologies or sectors, taxes on different tax bases, funding on different types of public goods, or policies that go in different directions (e.g. increasing vs. decreasing taxes on the same base). Finally, the prompt

[5] As discussed below, we have vectors for each policy from Qwen3-Embedding-0.6B. Using the pre-trained vectors (before fine-tuning), we compute cosine similarity scores for a large pool of within-topic candidate pairs. We then partition pairs into high-similarity strata defined by empirical percentile thresholds. Within each stratum, we sample across policy topics in proportion to their representation in that similarity band and sample without replacement. This procedure allows us to sample more in the upper tail of the similarity distribution while preserving approximate proportional representation across topics.

Figure 2: Summary Statistics on Legislative Topic Shares



Notes: This figure shows the proportion of policies assigned to each Comparative Agendas Project policy topic, based on labels generated by the LLM.

defines a set of graded similarity categories, ranging from (1) "Definitely yes" to (7) "Definitely not", as the set of valid responses.

Figure A.8 shows the distribution of responses across the policy-similarity categories. Notwithstanding that semantically similar policy pairs were oversampled, we still see that most policy pairs are classified as unrelated, a substantial share are only weakly related, and relatively few are closely related. Cases that are closely related but move in opposing policy directions (Category 6) are rare.

We then validate the similarity scores produced by the LLM using three legal experts (law school graduates). We assign the same task to the legal experts and ask them to provide similarity scores analogous to those generated by the LLM. On average, the legal experts agree more with the LLM than they agree with one another. As shown in Appendix Figure A.15, the average correlation between human and LLM scores is 72.3%, compared with 67.6% between legal experts. The LLM's scores also appear slightly closer to the overall consensus than those assigned by individual human evaluators, as shown in Appendix Figure A.16.

Encoding of policy summaries. Next, we encode the anonymized policy summaries with a pre-trained transformer-based embedding model, Qwen3-Embedding-0.6B. Qwen3 is an open-weights decoder-style transformer architecture, similar to the GPT series. The embedding model is a variant of the decoder model that produces informative 1024-dimensional vector representations of plain-text documents. To fine-tune the document vectors for semantic similarity and retrieval tasks, Qwen3's embedding model is trained with a contrastive learning objective on a large number of text pairs, encouraging semantically related texts to lie close together in embedding space and unrelated texts to be farther apart. An advantage of these models is that they interpret documents using the full context; for example, the word "bank" can have a different meaning for the regulation of financial institutions versus the regulation of river-ways.

Qwen3 vectors are trained on generic language tasks and are not specifically optimized for U.S. state policy similarity. To improve their fit for our purpose, we fine-tune the embeddings using the sample of policy pairs annotated as similar or dissimilar (Dell, 2025). Fine-tuning is performed separately within each broad topic using a weighted contrastive loss with cosine or Euclidean distance. The weights encode information on how similar a pair is according to the LLM's choice of graded similarity. During training, embeddings of policy pairs labeled as related are pulled closer together in vector space, while embeddings of unrelated pairs are pushed apart, up to a specified margin

Table 1: Cluster descriptives for statutes data

	Mean	Std dev.	5th pctile	95th pctile
Number of statutes per cluster	3167.066	5136.805	22.350	10918.350
Time span per cluster	90.583	43.461	26.000	157.000
Number of states per cluster	43.080	14.653	4.000	50.000
HH index	0.431	0.272	0.108	1.000
Share of clusters with more than one state contributing	0.984			
Share of clusters where all states are contributing	0.679			
Share of clusters where Fed Govt contributes	0.978			
Share of clusters where state contributes $\geq 50\%$	0.020			
Share clusters where Fed Govt contributes $\geq 50\%$	0.109			

Note: For statistics related to years, we removed statutes below the 10th percentile and above the 90th percentile per cluster. For statistics on states, we excluded federal government statutes from the clusters.

(for us 0.8). To address class imbalance, in addition to the per-pair loss weights, we also use weighted sampling. For each topic, the fine-tuning processes pushes vectors together for pairs that are similar, and pushes them apart for pairs that are dissimilar. This process reshapes the embedding space, stretching and rotating it via the distance loss, eventually leading to embeddings that are specialized for distinguishing policies.

3.3 Constructing Policy Clusters

The policy embeddings are comparable across years and states, thanks to the anonymization, and they are fine-tuned to distinguish similar policies within the legislative context of our data. Therefore, we can identify discrete policy groups by extracting local clusters of policy summaries in the fine-tuned embedding space.

We apply k-means clustering to the policy embeddings, separately by the broad topic. This algorithm learns centroids in the embedding space and assigns each policy summary to the closest centroid. We evaluate the clustering for different choices of k (number of clusters) based on the LLM-labeled policy pairs, as described in Appendix A.1.1.

In total, we end up with 2,488 distinct clusters across 24 topics. Figure 3 shows examples of pairs of policies that are from the same cluster. For each broad topic, Table A.1 shows the number of clusters chosen through our optimization process. There is quite broad variation in the number of policy clusters by topic, for example with only 8 for monetary policy and 192 for crime. Figure 3 offers a closer look at the composition of the clusters. On average, there are 3,167 policies in a given cluster.

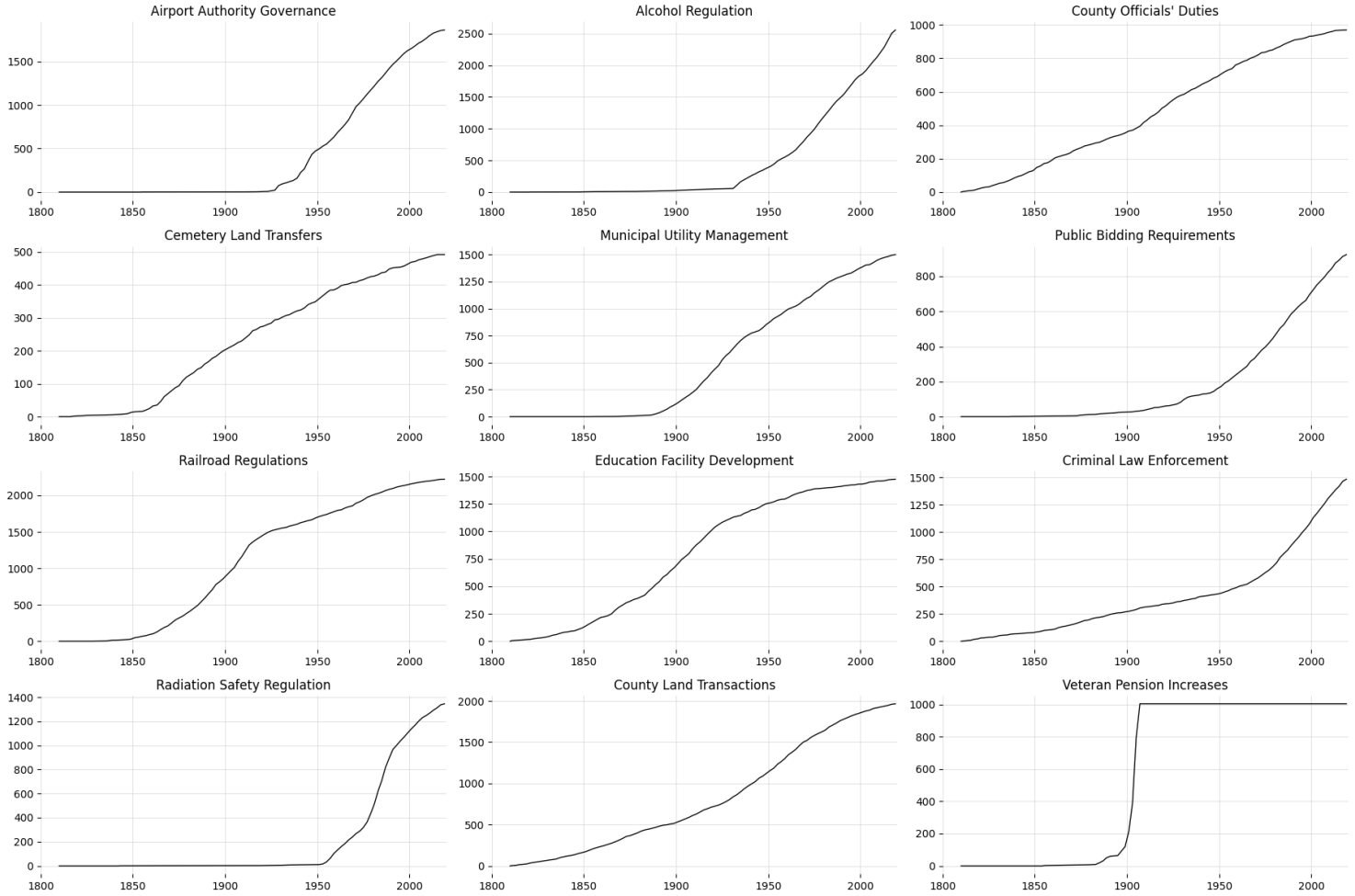
We can then get at policy priorities by state by counting the number of times a policy is enacted for each cluster. Figure 4 shows the evolution of the stock for a sample

Figure 3: Examples of policy pairs that are from the same cluster

Policy	Policy from same cluster
School Pesticide Notification (WA, 2009) A policy that requires schools to notify parents, guardians, and employees about pesticide use, including advance notification, posting of signs with some share of information, record keeping, and exemptions for certain applications, to ensure awareness and safety regarding pesticide applications on school facilities.	**Annual Pesticide Notification (CA, 2006)** A policy that requires school designees to annually notify staff and parents of all pesticide products expected to be applied at school sites, provide opportunities for individual application notifications at least a specific number of hours prior, post warning signs at pesticide application areas, and make efforts to notify during emergency pesticide applications, with certain exemptions and specific provisions for child day care facilities and juvenile justice schools.
City Elections and Voting Rights (TX, 1893) A policy that grants all qualified state electors residing six months in a specific city the right to vote for mayor and city officers, but restricts voting on money expenditure or debt assumption to property tax payers.	**Town Voting Qualifications (IL, 1865)** A policy that qualifies all male white citizens of this jurisdiction over a specific age, not disqualified by law of this jurisdiction, who have resided in a specific town for a specific time period prior to the election, to vote for town officers.
No Points for Seat Belt Violations (OK, 2002) A policy that prohibits the Department of Public Safety from recording or assessing points on a license holder's traffic record for violations of the mandatory seat belt usage section.	**Insurance Coverage Protection (MA, 1993)** A policy that prohibits insurance companies from denying coverage or the right to purchase motor vehicle liability policies to individuals who failed to wear safety belts during accidents resulting in bodily injury or who violated the seat belt use provisions.

of clusters over the time period. The figure shows that different policies were enacted at different times. For example, policies around electoral precincts were introduced already in the early and mid-1800s. Regulations about pesticides were introduced in the late 1900s.

Figure 4: Evolution of the stock of 12 randomly picked clusters



Notes: Each graph shows the cumulative stock of acts adopted in each cluster over the period. Cluster labels are provided by a GPT prompt based on a sample of associated policy summaries.

4 Descriptive Statistics

In this section, we provide descriptive evidence on policy adoption patterns and on the properties of our policy clusters.

We begin with Table 2, which reports moments of the key variables constructed from the cluster data and used in our empirical framework. Some of these variables are mechanically, or nearly mechanically, related to one another, for example when one variable is a scalar multiple of another. We nonetheless report them separately to provide a sense of magnitudes and scaling for the objects that enter the empirical analysis.

The average number of adoptions by a state in a given cluster and biennium, Y_{sct}, is 0.23 (SD = 0.42); the corresponding statistic for the federal government is nearly identical. In a typical biennium, the average state adopts policies in 576 distinct clusters, with substantial dispersion (10th percentile: 243; 90th percentile: 1004).

Turning to cross-state activity within clusters, the average number of states adopting in a given cluster-biennium is approximately 10, while the median is 6, indicating a left-skewed distribution with many cluster-years characterized by limited adoption.

The average state's own stock of adoptions in a given cluster over the previous 10 years, N_{sct}, is 1.11—mechanically equal to five times the mean of Y_{sct}. Over the same horizon, the average stock of other states' adoptions in a cluster is 50.78, and the corresponding federal stock, $N_{FD,ct}$, is 1.16. Finally, in the average year, a state is politically aligned with 9.06 peer states. Summing adoptions by peer states over the previous 10 years (where alignment is constructed at the year level), peer states make an average of 10.97 adoptions.

Table 2: Summary Statistics

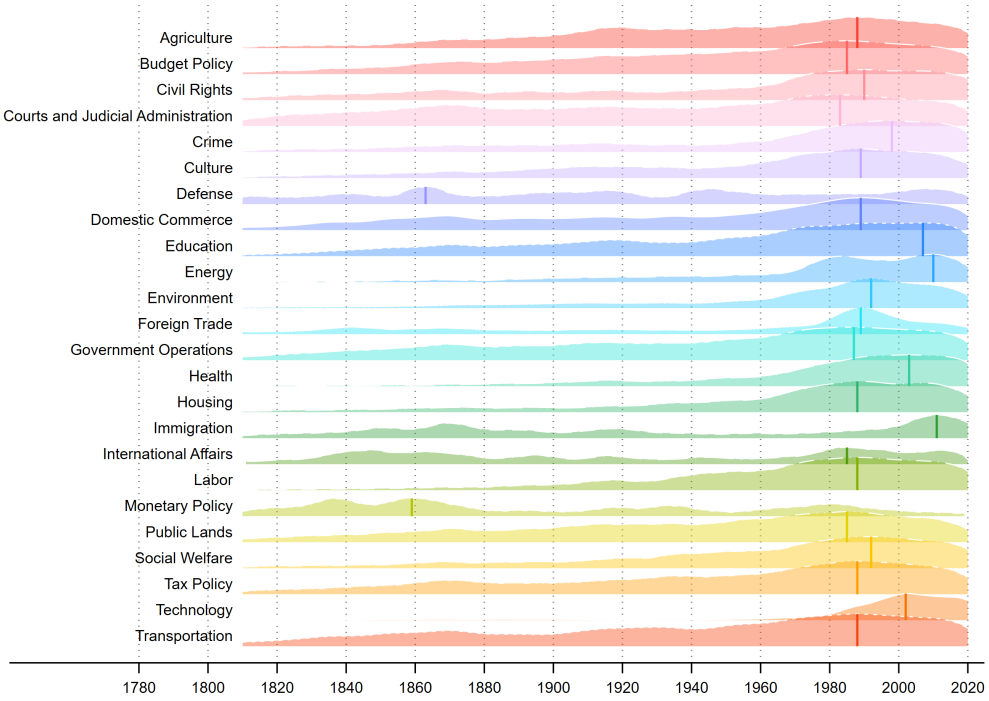
	Mean	SD	P10	P50	P90	N
Y_{sct}	.2316948	.4219151	0	0	1	9561384
Y_{fct}	.23833	.426063	0	0	1	213968
Y_{st}	576.4567	294.0533	243	525	1004	3843
Y_{ct}	10.35352	11.83262	0	6	29	213968
N_{sct}	1.115858	1.568867	0	0	4	9441960
$N_{s'ct}$	50.78271	57.22159	0	29	141	9441960
N_{fct}	1.164938	1.589167	0	0	4	211480
$N^{pol}_{s'ct}$	10.97189	17.99424	0	0	38	8138248
Nr. of political peers by period	9.06	8.28	0	12	20.4	672

The table reports the mean, standard deviation, 10th percentile, median, 90th percentile, and observation count for the analyzed variables.

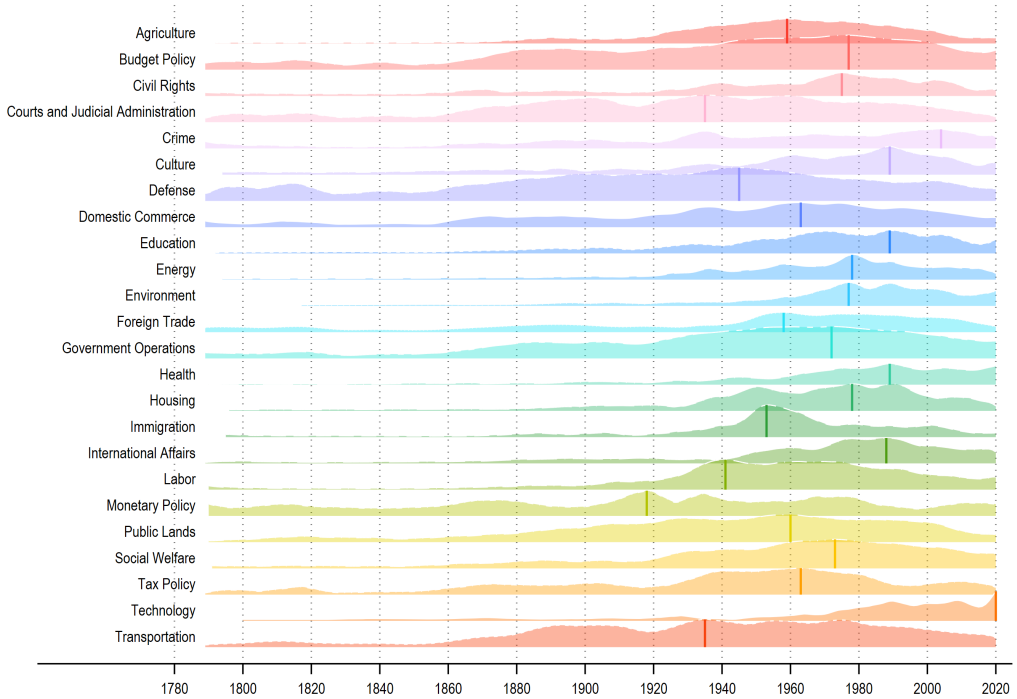
Next, we document the evolution over time of aggregate policy adoptions across primary topics. Figure 5 restricts attention to state legislatures and shows a general increase in adoption activity in recent decades, consistent with the rise in overall legislative volume depicted in Figure 1. At the same time, there is substantial heterogeneity across topics. Some are persistently active throughout the sample, such as Government Operations, Taxation, and Transportation. Others exhibit distinct historical waves: Monetary Policy peaks in the nineteenth century, largely reflecting bank incorporations that required statutory approval; Immigration restrictions surge in the late nineteenth century; Labor legislation expands in the mid-twentieth century; Defense-related activity spikes during the Civil War; Environmental policy rises sharply around the period of the Clean Water Act; and Technology-related policies become prominent only in recent decades.

Figure 5b reports the same exercise restricting attention to the federal government. Several differences relative to Figure 5 are apparent. For example, federal adoption in Education occurs later than in the states.

Figure 5: Distribution of policy adoptions by primary topics over time



(a) All states

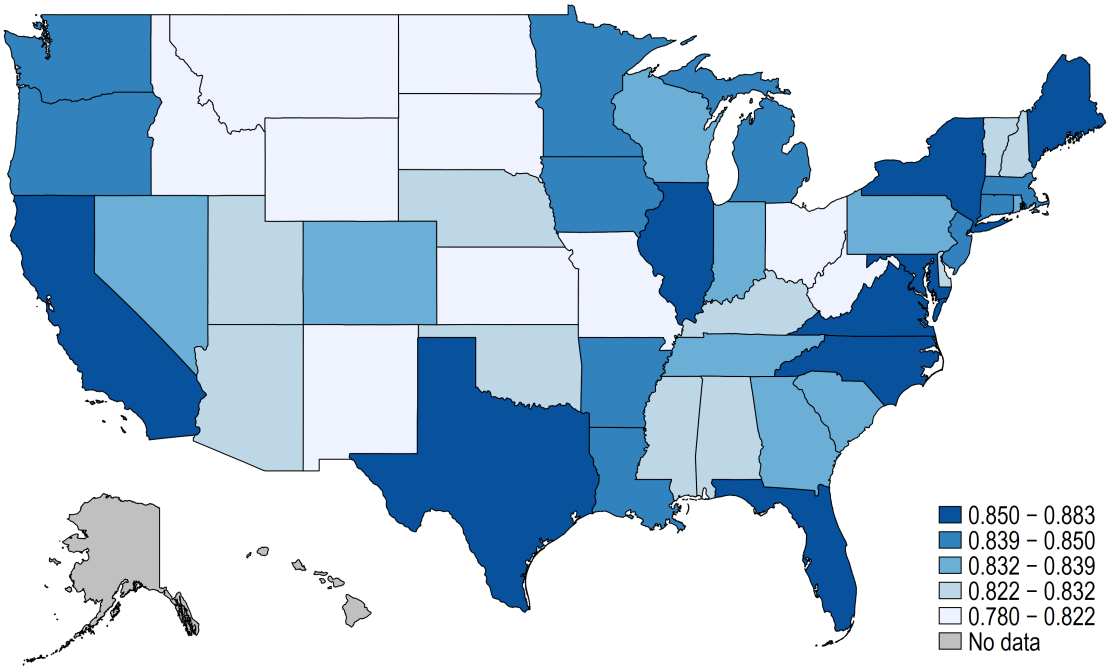


(b) Federal Government

Notes: Each panel aggregates unique policy adoptions—across states excluding the federal government (a) and by the federal government (b)—and clusters by primary topic in each year. For each topic, the time series of adoptions is normalized by that topic's maximum annual value. The vertical line marks the year with the highest total number of adoptions.

Figure 6 reports a state-level measure of legislative diversity by showing, for each state, the share of policy clusters in which it adopts at least one policy over the full sample period. Two features stand out. First, every state has contributed at least once to the vast majority of policy clusters, yet no state has contributed to more than approximately 88.3 percent of clusters. Second, there is meaningful cross state variation. States such as California and Texas have contributed to roughly 85 to 88.3 percent of clusters, whereas states such as Montana and New Mexico have contributed to about 78 to 82.2 percent.

Figure 6: Share of Policy Clusters in which a state has adopted at least once



Notes: The figure shows, for each state, the share of the 2488 policy clusters in which the state has adopted at least one policy over the whole sample period.

Figure A.18 in the Online Appendix shows, for eight topics and for each state separately, the share of total adoptions attributable to each topic over the full period. The resulting spatial patterns are broadly consistent with intuition: agricultural states concentrate relatively more on agriculture, whereas states such as California exhibit greater emphasis on environmental and technology policy.

5 Empirical Analysis

5.1 Framework

We develop a framework for analyzing legislative policymaking in the United States that incorporates both horizontal federalism, referring to interdependencies among state legislatures, and vertical federalism, referring to two way interactions between state legislatures and the federal legislature.

Let $\mathcal{S}$ denote the set of states and let f denote the federal government, and define $\mathcal{A} \equiv \mathcal{S} \cup f$. By a slight abuse of notation, we use $s \in \mathcal{A}$ to index a generic jurisdiction, i.e., either a state or the federal government. Time is indexed by years $t \in \mathcal{T}$. Policies are partitioned into discrete "clusters" $c \in \mathcal{C}$ (reflecting our data construction). For a given (s, c, t), define the binary adoption indicator $Y_{sct} \in \{0, 1\}$, which equals one if state s adopts at least one policy in cluster c in year t. For the federal government, we define Y_{fct} analogously as an indicator equal to one if the federal legislature adopts at least one policy in cluster c in year t.

We interpret Y_{sct} through a random-utility framework. Let $U_{sct}(1)$ denote the utility (payoff) to state s from adopting at least one policy in cluster c in year t, and let $U_{sct}(0)$ denote the utility from not adopting in that cluster-year. The observed adoption indicator satisfies $Y_{sct} = \mathbf{1}\{U_{sct}(1) \geq U_{sct}(0)\}$.

We assume that clusters are technologically independent: conditional on covariates, the costs and benefits of adopting a policy in cluster c at time t do not depend on a state's adoption decisions in any other cluster $c' \neq c$ in the same year. Within a cluster, however, adoption incentives may depend on (i) the state's own past adoption history in that cluster and (ii) the past adoptions of other jurisdictions, including the federal government, within the same cluster.

For a fixed horizon $\tau \geq 1$, let:

$$T_t^-(\tau) \;\equiv\; \{\, t' \in \mathcal{T} \;:\; t - \tau \leq t' < t \,\}$$

denote the set of recent years that potentially shape current adoption incentives.

We parameterize the adoption and non-adoption utilities as

$$U_{sct}^{(1)} = V_{sct} + \varepsilon_{sct}, \qquad U_{sct}^{(0)} = 0, \tag{1}$$

with ε_{sct} i.i.d. logistic with unit scale. Under the normalization $U_{sct}(0) = 0$, $U_{sct}(1)$ can be interpreted as the *net* utility of adoption in cluster c and year t for state s (relative to non-adoption). Further, let

$$V_{sct} \;=\; \alpha_{sct} \;+\; \delta_c \sum_{t' \in T_t^-(\tau)} Y_{sct'} \;+\; \sum_{\substack{s' \in \mathcal{A} \\ s' \neq s}} \sum_{t' \in T_t^-(\tau)} \beta_{ss'ct'}\, Y_{s'ct'}. \tag{2}$$

The index V_{sct} is the systematic (observable) component of state s's net utility from adopting in cluster c at time t, while ε_{sct} captures idiosyncratic state–cluster–year shocks. The baseline component of V_{sct} is

$$\alpha_{sct} \;=\; \alpha_{sc} \;+\; \alpha_t \;+\; \boldsymbol{\omega}_c^{\top} \boldsymbol{X}_{st}, \tag{3}$$

where $\boldsymbol{X}_{st}$ collects state-year covariates (e.g., economic or political conditions) and $\boldsymbol{\omega}_c$ are cluster-specific coefficients. Here, α_c is a cluster-specific intercept common to all jurisdictions and years, while α_{st} captures state-by-year propensity to legislate.

Next, the term $\delta_c \sum_{t' \in T_t^-(\tau)} Y_{sct'}$ captures how a state's own within-cluster adoption history affects its current propensity to adopt in that cluster.[6] Interdependencies across jurisdictions are captured by:

$$\sum_{s' \neq s} \sum_{t' \in T_t^-(\tau)} \beta_{ss't'}\, Y_{s'ct'}, \tag{4}$$

which shifts s's utility in response to past adoptions by other states and by the federal government ($s' = f$). We parameterize these interactions as:

$$\beta_{ss't'} \;=\; \beta_{s'} \;+\; \boldsymbol{\gamma}^{\top} \boldsymbol{SimX}_{ss'}, \tag{5}$$

where β_c captures the baseline spillover strength within cluster c, $\beta_{s'}$ captures the influence of the sender jurisdiction s', and $\boldsymbol{SimX}_{ss'}$ denotes a vector of pairwise characteristics (such as geographic proximity, economic similarity, or partisan alignment) with associated coefficients $\boldsymbol{\gamma}$. These characteristics allow systematic differences across jurisdiction pairs to shape the strength and direction of interdependencies.

We write the probability of adoption by each state in each policy cluster in each year as

$$p_{sct} = \frac{\exp(V_{sct})}{1 + \exp(V_{sct})}. \tag{6}$$

[6] See Boehmke and Witmer (2004) on learning from a state's prior adoptions.

Finally, for notational simplicity, we define the τ-period lagged adoption stock as

$$N_{sct} \equiv \sum_{t' \in T^{-}t(\tau)} Y_{sct'}, \tag{7}$$

that is, the number of recent periods in which jurisdiction s adopted a policy in cluster c.[7] For a given adopter s, we refer to the collection $\{N_{s'ct}\}_{s' \neq s}$ as s's *exposure* to other jurisdictions' prior adoptions in cluster c.

Non-strategic (myopic) interpretation. Equations (1)–(6) admit a utility interpretation under a *non-strategic* behavioral assumption: jurisdictions respond to the history of past adoptions as predetermined state variables, but they do not choose current adoption Y_{sct} in order to influence future adoption paths.[8] In particular, although today's adoption may mechanically enter other jurisdictions' future utilities through the lag structure in V_{sct}, we assume that neither states nor the federal government internalize these dynamic feedback effects when making the current decision. Thus, the model should be read as a myopic discrete-choice framework rather than a forward-looking dynamic game. For instance, when we refer to $\beta_{s'}$ as the "influence" of jurisdiction s', this is a descriptive notion: it measures how past adoption by s' shifts other jurisdictions' *current* adoption utilities, not the strategic return to s' from inducing future adoption elsewhere.

5.2 Estimation and Identification

We estimate the parameters of the adoption model in Section 4.1 by maximum likelihood under the logistic specification for ε_{sct}. The estimating equation is a logit for Y_{sct} with systematic component V_{sct} as defined in equation (2), including cluster fixed effects α_c and state year fixed effects α_{st}, and allowing the effect of observables to vary by cluster through ω_c in equation (3).

The covariate vector X_{st} includes political preferences and institutions (Democratic presidential vote share, party control of the legislature, and governor party), economic structure (employment shares by NAICS2 sector),[9] geography (longitude and latitude of the population weighted state centroid), and legislative capacity (session length, legislator salary, and legislative budget).

[7] In our baseline implementation we set $\tau = 10$.

[8] Allowing for forward-looking strategic behavior would require specifying continuation values and solving for an equilibrium (e.g., Markov perfect equilibrium) in a dynamic adoption game.

[9] NAICS2 sector shares are available from 1950 onward from County Business Patterns (Eckert et al., 2020); earlier years use coarser sector shares from the decennial census.

Horizontal diffusion heterogeneity is captured by $\text{SimX}_{ss'}$ in equation (5). Political similarity is an indicator equal to one when the two states share party control of both the legislature and the governorship (DellaVigna and Kim, 2025). Economic similarity is an indicator equal to one when the cosine similarity of states' sector share vectors is in the top tercile of the pairwise distribution. Geographic proximity is an indicator equal to one when the distance between population weighted centroids is in the bottom tercile. We allow these channels to evolve over time by interacting each element of $\text{SimX}_{ss'}$ with decade indicators, yielding decade specific coefficients γ_d.

Baseline identification includes both state year fixed effects α_{st} that capture shocks that shift a state's overall propensity to legislate in year t and α_c which capture persistent differences in baseline adoption rates across clusters. Cluster specific slopes ω_c allow changes in X_{st} to reallocate legislative effort across clusters rather than mechanically shifting adoption in all clusters.

With horizon $\tau = 10$, we measure own history using N_{sct} from equation (7) and include $\delta_c N_{sct}$ in the adoption index. Peer exposure uses $\{N_{s'ct}\}_{s' \neq s}$, so adoption in year t depends only on adoptions in prior years.

Vertical diffusion is incorporated by including the federal adoption stock N_{fct} in the state equations and the aggregate state adoption stock $\sum_{s \in S} N_{sct}$ in the federal equations. We estimate decade specific coefficients for these vertical terms and do not impose similarity shifters for state and federal pairs.

For inference, we report heteroskedasticity robust standard errors clustered two ways: by adopter state $\times$ year to account for the potentially joint decision of adoption of different clusters and by adopter state $\times$ cluster to account for serial correlation in adoptions.

Under conditional independence of ε_{sct} given $(X_{st}, \alpha_{st}, \alpha_c)$ and the lag structure, coefficients on peer stocks can be interpreted as the effect of prior adoption on current adoption incentives. The main threats are unobserved shocks that jointly move multiple jurisdictions and latent similarities that induce correlated timing even without interdependence. Because correlated shocks and latent similarities may still contaminate exposure even after the baseline fixed effects, we implement two complementary robustness strategies (reported in Section 5.5). First, we introduce progressively more demanding fixed effects (including pair and policy-by-year components) so that identification comes from within-pair asymmetries in adoption timing. Second, we instrument other jurisdictions' adoption stocks using plausibly exogenous shifts in their observable characteristics. A limitation of this instrumental-variables approach is that it can only be implemented starting in the 1970s, due to the availability of the underlying

legislative-capacity covariates used to construct the instruments.

6 Empirical Results

Using the discrete-choice model outlined in Section 5.1, we estimate how local political conditions and the adoption history of other jurisdictions shape and federal decisions to adopt policies in a given cluster.

6.1 Baseline Adoption Propensity and Local Political Conditions

In this subsection, we examine how local covariates X_{st} enter V_{sct} through the cluster-specific coefficients ω_c in the baseline component $\alpha_{sct} = \alpha_{sc} + \alpha_t + \omega_c^\top X_{st}$ (equation (3)), and thereby tilt a state's adoption incentives across policy clusters. Because we include policy by state fixed effects α_{sc}, these estimates should be interpreted as policy-mix effects: holding fixed a state's overall propensity to legislate in a cluster, changes in X_{st} shift the *composition* of legislative activity toward some clusters and away from others.

To illustrate, we focus on voters' preferences, proxied by the state-level Democratic presidential vote share in the most recent election. Figure 11 plots, for some selected clusters c, the estimated coefficient on Democratic vote share in $\omega_c^\top X_{st}$, rescaled so that the horizontal axis corresponds to a 10 percentage point increase in Democratic vote share. Since adoption probabilities satisfy $p_{sct} = \exp(V_{sct})/(1 + \exp(V_{sct}))$ (equation (6)), the coefficients in Figure 11 are in log-odds units. For example, a value of 0.30 implies a rise of the probability to adopt for a state at the average baseline probability ($p_0 = 0.17$) is 0.05 percentage points.

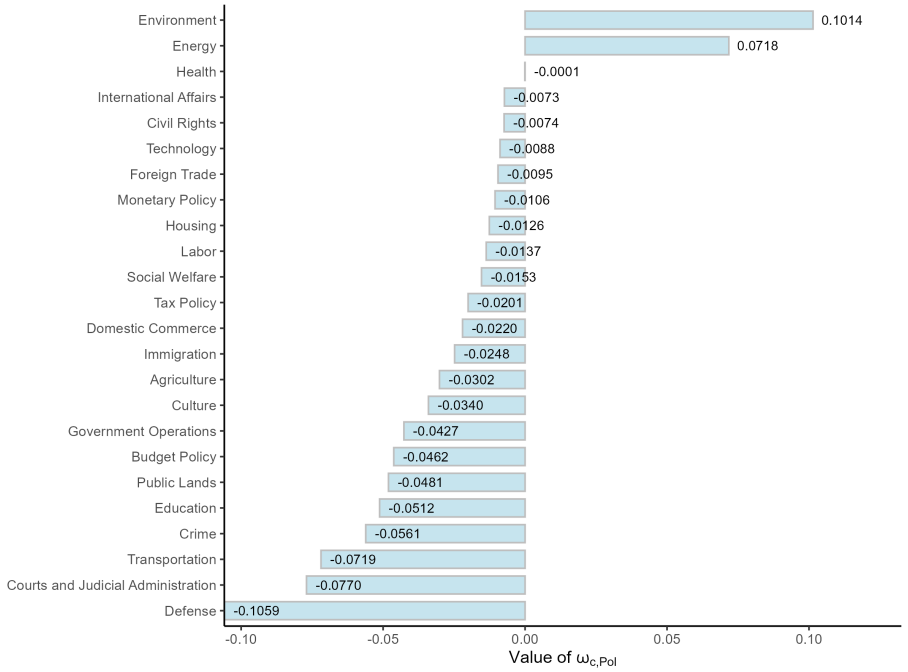
The estimates reveal substantial heterogeneity across policy domains, consistent with intuitive partisan priorities. Positive coefficients correspond to clusters that are disproportionately adopted in more Democratic-leaning states, such as energy policy, education for handicapped children, health care planning, firearms regulation, and affordable housing. Negative coefficients correspond to clusters that are disproportionately adopted in more Republican-leaning states, such as immigration regulation, veterans' pension benefits and bonuses, and restrictions on liquor sales near churches.

To conclude, local political conditions meaningfully shape state-level legislative policymaking, and these effects vary systematically across policy clusters. We next turn to the diffusion components of the model, which capture how adoption histories in other

Figure 7: Average estimates of ($\omega_{c,Pol}$) by policy topic



(a) Estimates of ($\omega_{c,Pol}$) for single policies



(b) Average likelihood to adopt policies from a given topic by Partisanship

Notes: Panel (a) reports the estimated parameters $\omega_{c,Pol}$ for selected clusters for the period 1880–2020. These coefficients capture the association between the linear index V_{sct} and the share of Democratic seats in the state legislature (see equation (2)). Panel (b) reports the averages by topic. Positive values indicate policies that are more likely to be adopted by states with a Democratic majority.

jurisdictions shift current adoption incentives.

6.2 Diffusion Strength of Individual Jurisdictions

In our framework, a state's propensity to adopt in cluster c at time t depends on its exposure to recent adoptions by other jurisdictions (other states and the federal government). Formally, in equations equations (4)–(5), adoption by a sender jurisdiction s' affects an adopter state s through the lagged adoption stock $N_{s'ct} \equiv \sum_{t' \in T_t^-(\tau)} Y_{s'ct'}$.

This subsection focuses on the sender-specific parameter $\beta_{s'}$, which captures how strongly jurisdiction s' "transmits" policies to other states on average, holding fixed (i) the cluster-specific spillover strength and (ii) similarity-weighted pairwise channels (geographic, economic, and political similarity). Intuitively, $\beta_{s'}$ is a sender fixed effect: when s' adopts in a cluster, a larger $\beta_{s'}$ means that other states are more likely to adopt in that same cluster in subsequent years.

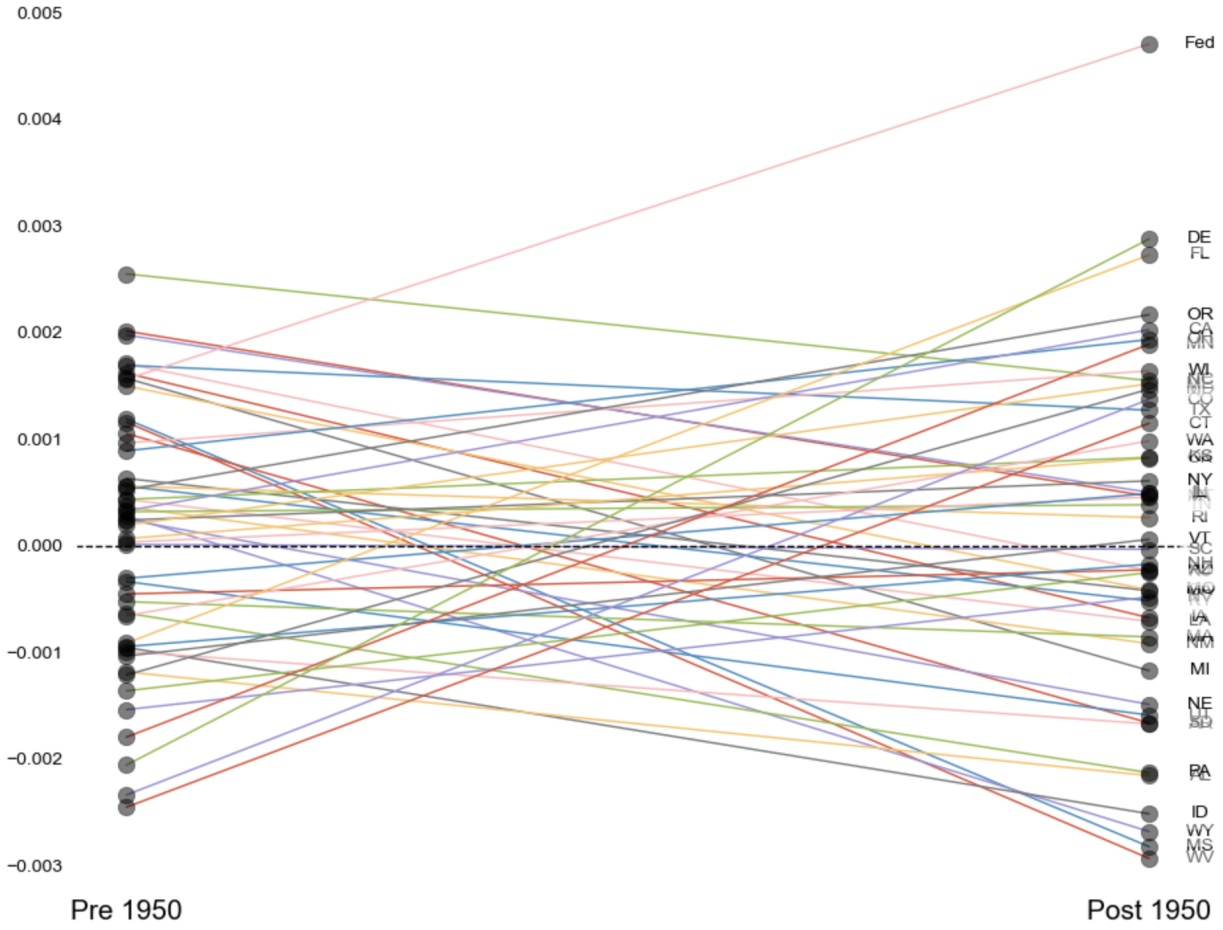
We estimate $\beta_{s'}$ by pooling all clusters while including cluster fixed effects, and we report results separately for the 1880–1950 and 1950–2020 periods. We impose the normalization $\sum_{s' \in A} \beta_{s'} = 0$, so coefficients are interpreted as deviations from the average sender.

Figure 8 ranks jurisdictions by their average diffusion influence.

The coefficient $\beta_{s'}$ is in log-odds units. For example, $\beta_{s'} = 0.004$ translates to a difference of about 0.6 percentage points per additional adoption-year at the baseline adoption probability $p \approx 0.17$.

Three patterns stand out. First, the federal government ("Fed") is among the most influential senders, consistent with an important vertical diffusion channel from federal statutes to state legislation (we return to the evolution of vertical diffusion in Section 6.4). Second, there is substantial heterogeneity across states: some jurisdictions systematically generate stronger follow-on adoption than others, even after conditioning on cluster effects and similarity-weighted spillovers. Third, the identities of influential senders are consistent with plausible institutional mechanisms. For instance, Delaware's influence is consistent with its central role in corporate law and corporate chartering; similarly, large and policy-innovative states (e.g., California or Florida) often rank among the more influential senders.

Figure 8: Jurisdiction-level diffusion influence ($\beta_{s'}$)



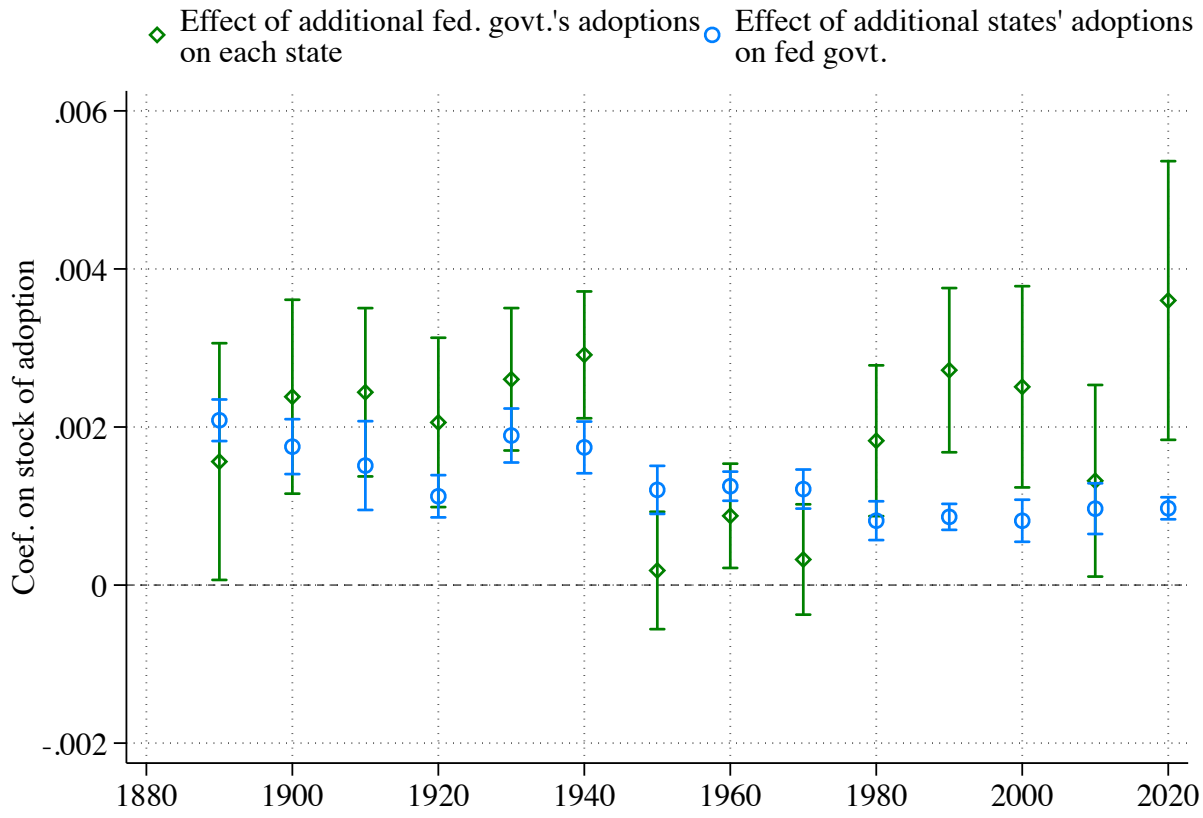
Notes: Each point is a sender jurisdiction s' (states and the federal government). The figure plots the estimated sender effects $\beta_{s'}$ from equation (5), shown separately for 1880–1950 and 1950–2020; lines connect the same sender across periods. A one-unit increase in $N_{s'ct}$ (i.e., one additional year within the last τ years in which s' adopted in cluster c) shifts the adopter's index by $\beta_{s'}$, so positive values indicate more influential-than-average senders and negative values indicate less influential-than-average senders. Confidence intervals are reported in Appendix Table A.4.

6.3 Vertical Federalism: Federal–State Interactions Over Time

A central feature of our framework is vertical diffusion: past federal legislation in a policy cluster can shift states' incentives to legislate in that same cluster, and past state legislation can shift federal incentives. To trace how these two directions evolve over time, we estimate decade-specific vertical diffusion coefficients using the adoption *stocks* $N_{sct} \equiv \sum_{t' \in T_t^-(\tau)} Y_{sct'}$, where $\tau = 10$ in our baseline specification. Thus, increasing N_{sct} by one corresponds to one additional year in the last τ years in which jurisdiction s adopted at least one policy in cluster c.

Figure 9 summarizes the estimated vertical links decade by decade. The green series plots the effect of federal exposure on states: the change in a representative state's adoption index from a one-unit increase in the federal adoption stock N_{fct} (i.e., an additional recent federal adoption-year in the same cluster). The blue series plots the reverse direction, the effect of state exposure on the federal government, rescaled to be comparable: it reports the change in the federal adoption index from an increase of five units in the aggregate state adoption stock $\sum_{s \in S} N_{sct}$, which corresponds to five states each adding one additional adoption-year in that cluster. (This "five-state" scaling is purely for readability; multiplying by any constant would rescale the blue series proportionally without affecting the time pattern.)

Figure 9: Marginal effects of vertical diffusion between the federal government and the states



Notes: The figure reports decade-specific estimates of vertical diffusion using the lagged adoption stocks N_{act} (Section 4.2, $\tau = 10$). Green markers show the estimated marginal effect of federal exposure on state adoption, $\partial p_{sct}/\partial N_{fct}$, averaged across clusters and estimated at the average propensity to adopt $p_0 = 0.17$. Blue markers show the estimated marginal effect of state exposure on federal adoption, $\partial p_{fct}/\partial \left(\sum_{s \in S} N_{sct} \right)$, estimated at the average propensity to adopt $p_0 = 0.17$. Error bars are 95% confidence intervals based on two-way clustered standard errors (adopter × cluster and adopter × year).

Three patterns stand out. First, the influence of federal legislation on states rises sharply over time. In the late nineteenth century the estimated effect is near zero (and occasionally slightly negative), but it becomes distinctly positive by the postwar era and remains large thereafter. By the 2000s, the marginal effects at the baseline propensity to adopt are around 0.0035, meaning that each additional federal government's adoption is associated with 0.35 percentage points greater probability that each state adopts.

Second, the reverse channel weakens over time: the responsiveness of the federal government to state legislative activity (blue series) declines steadily across decades. In the early part of the sample, the effect of each state on the federal government is comparable to the federal-to-state effect but by the 2000s, the federal-to-state effect clearly dominates the state-to-federal effect.

Finally, these estimates reflect averages across all policy clusters. Vertical diffusion may be much stronger in some domains (e.g., civil rights, environmental regulation, health) and weaker in others. In Section 6.4, we return to the mechanisms and implications of this increasing asymmetry in federal–state interactions.

6.4 Horizontal Federalism: Similarity-Based Spillovers

Having documented (i) which jurisdictions are especially influential senders ($\beta_{s'}$) and (ii) how vertical diffusion between the federal government and the states evolves over time, we now turn to how diffusion operates across states. In our model, the spillover from a sender state s' to an adopter state s depends not only on the sender itself, but also on how similar the two states are.

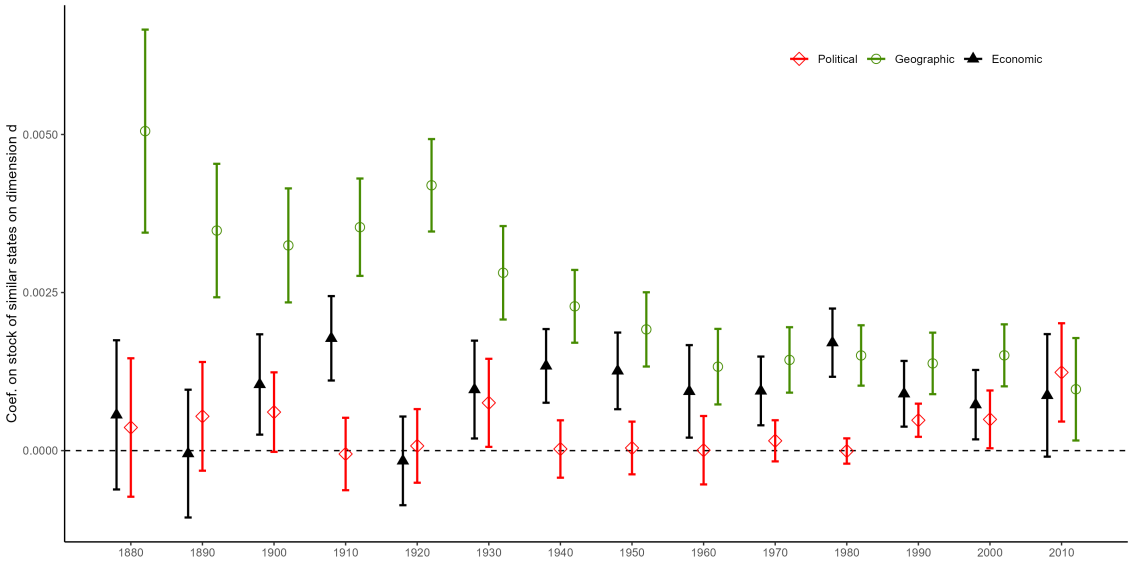
Recall that the contribution of s''s past adoption history in cluster c to s's adoption index can be written as

$$\sum_{t' \in T_t^-(\tau)} \beta_{ss't'} \, Y_{s'ct'} \quad \text{with} \quad \beta_{ss'ct'} = \beta_{s'} + \gamma^\top \text{SimX}_{ss'}, \tag{8}$$

so that γ captures *similarity-based* heterogeneity in horizontal diffusion (equation (5)). Intuitively, γ answers the question: when a state looks to other states as reference points, does it put more weight on geographically proximate states, economically similar states, or politically aligned states? (See Section 4.2 for the construction of the similarity indicators.)

To allow these channels to evolve, we estimate decade-specific coefficients γ_d and report their paths over time. Figure 10 plots the decade-by-decade estimates for the three similarity dimensions. Because the coefficients enter the log-odds index V_{sct}, a positive value for (say) political alignment means that, conditional on the baseline spillover strength and sender effects, adoption by a politically aligned state generates a larger increase in s's adoption propensity than adoption by a non-aligned state.

Figure 10: Marginal effects of similarity-based horizontal diffusion over time



Notes: The figure plots decade-specific marginal effects of the similarity-shifter coefficients γ_d from equation (5). The three series correspond to geographic proximity, economic similarity, and political alignment (as defined in Section 4.2). Coefficients measure how strongly diffusion is amplified among similar state pairs, relative to otherwise comparable dissimilar pairs. Standard errors are two-way clustered by adopter state × year and adopter state × cluster.

Three patterns stand out. First, geographic proximity is a strong predictor of diffusion in the early decades but declines steadily over time, consistent with a weakening role for purely local or regional channels. Second, economic similarity is consistently positive and comparatively stable throughout the sample. Third, political alignment is modest in earlier decades but rises sharply beginning in the 1980s, eventually reaching magnitudes comparable to the geographic and economic channels. Taken together, these estimates suggest a long-run reweighting of horizontal diffusion away from geography and toward partisan alignment, with economic similarity remaining an important and persistent channel.

6.5 Sensitivity to correlated shocks

One might worry that the coefficients estimated above partly reflect common shocks that jointly affect states' policy adoption decisions, rather than a causal influence of one state on another. For example, the Covid-19 pandemic may have led many states to adopt similar healthcare policies over a short period of time, even in the absence of any diffusion mechanism. A related concern is that correlated adoption may arise from latent, shared pairwise characteristics (e.g., culture, educational composition, political

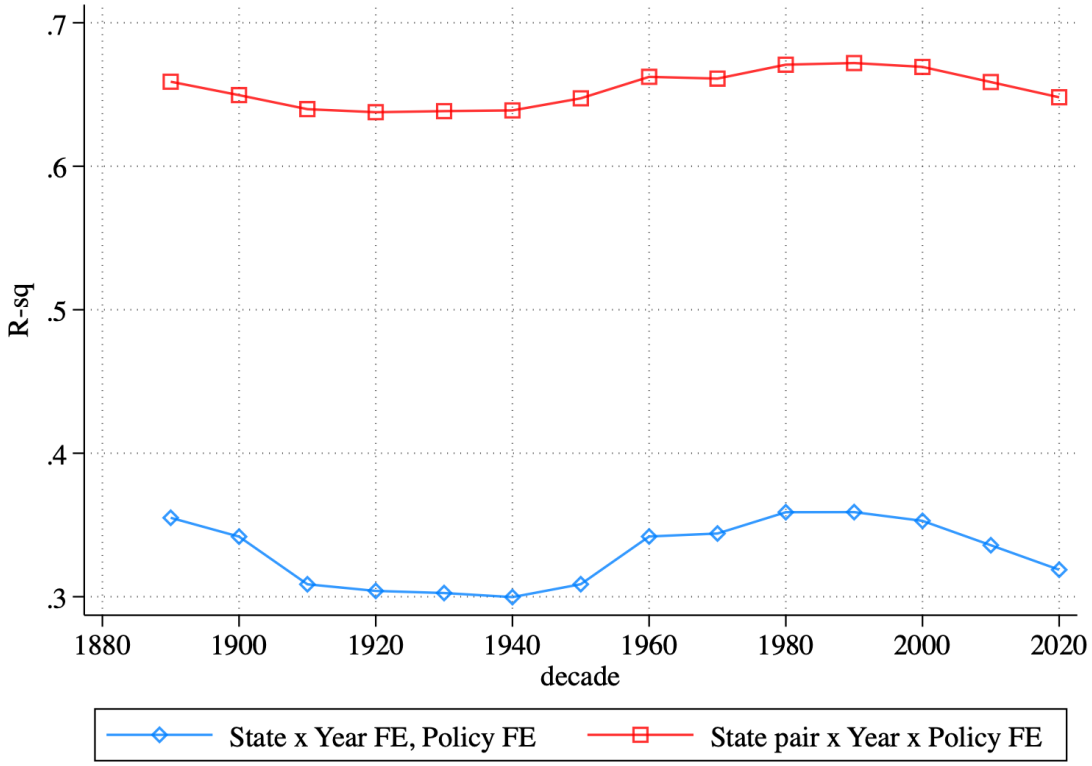
orientation, or legal tradition) that persist over time. In the baseline model, we account for these potential sources of confounding in several ways. First, we do not estimate the contemporaneous effect of state A's adoption in year t on state B's adoption in the same year. Instead, we focus on temporal variation, which helps mitigate concerns related to reflection (Manski, 1993). Second, the main parameters of interest are defined in terms of differences in diffusion intensity across types of state pairs, for example whether diffusion is stronger among politically aligned states than among non-politically states. In this sense, identification comes from comparing the effect of policy adoption by similar states along a given dimension to the effect of adoption by dissimilar states.

Since some common shocks may still affect only subsets of states, and in particular subsets defined along a given similarity dimension (for example, all Democratic states), we adopt two additional strategies described below. The first strategy accounts for any shared characteristics across state pairs by adopting state pair by policy by year fixed effects. The second strategy leverages changes in states' observable characteristics as exogenous shifters of their propensity to adopt policies.

6.5.1 Accounting for shared state-pair characteristics

In our baseline specification, we include state-by-year and policy-cluster fixed effects. In this subsection, we show that our estimates remain stable as we progressively introduce stricter fixed effects, which absorb an increasingly large share of unobservable common shocks. Consistent with this approach, Figure 11 reports the R-squared values for various fixed-effect specifications, estimated separately by decade. The blue line reports the R^2 for the baseline model. The green line reports the R^2 when we augment the baseline with state-pair and policy-by-year fixed effects. The red line reports the R^2 under our most demanding specification, which includes state-pair $\times$ policy $\times$ year fixed effects.

Figure 11: R-squared from decade-specific regressions with varying levels of fixed effects



Notes: The figure shows the R-squared captured by fixed effects from three different specifications, separately for each decade.

In the most stringent specification, described in equation (9), identification of the diffusion parameters relies on within-state-pair asymmetries in adoption timing. State-pair × policy × year fixed effects absorb any shock that is common to a given pair of states for a given policy and year, so the coefficients are identified from situations in which, within a pair, one state has previously adopted a policy while the other has not. The parameter therefore captures whether a state is more likely to adopt following adoption by the other state in the pair, and whether this tendency is stronger for pairs that are more similar along a given dimension (political, economic, or geographic). Because these fixed effects absorb all pair-policy-year common components, we cannot separately identify the effect of a state's own stock of adoptions from the effect of the other state's stock. Since the object of interest is not the level effect of exposure, but rather how diffusion differs across types of state pairs, this lack of separate identification is not a strong limitation. Evidence in favor of diffusion along a given similarity dimension therefore comes from systematic differences in adoption timing across pairs

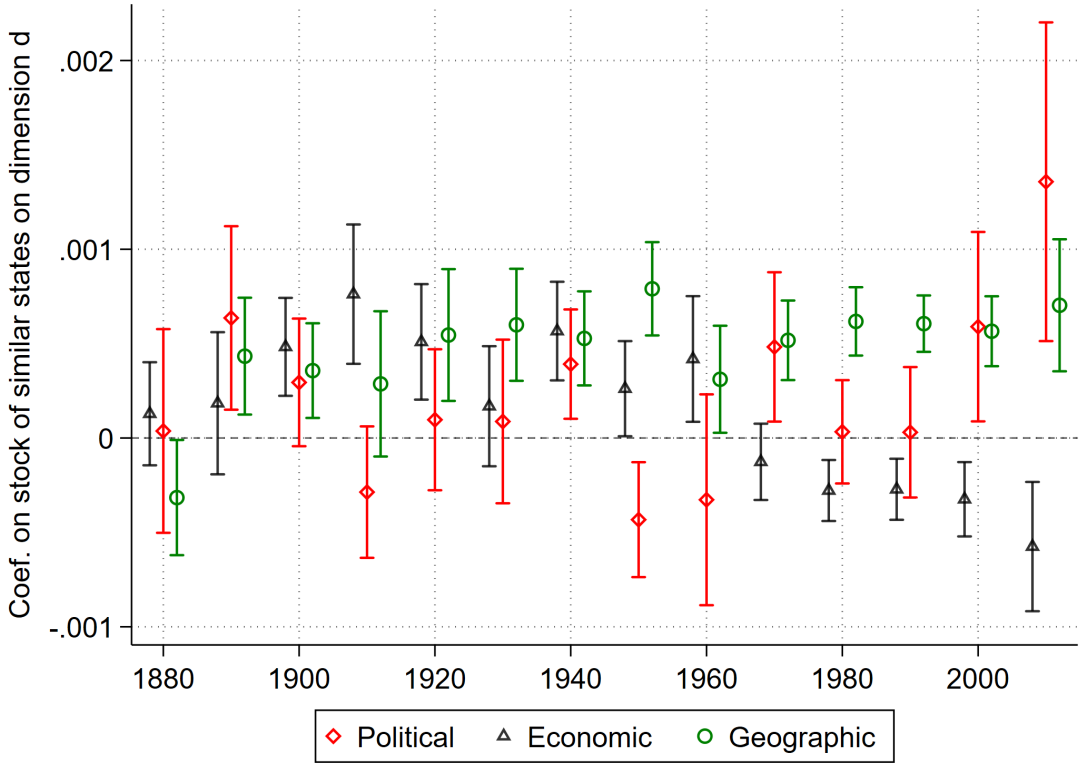
with different degrees of similarity, rather than from coordinated movements across pairs. Specifically:

$$Y_{stc} = \omega^{all} N_{s'tc} + \sum_{sim \in \{\text{pol,econ,geo}\}} \omega^{sim} N^{sim}_{s'tc} + \delta N_{stc} + \alpha_{ss'tc} + \epsilon_{ss'tc}, \tag{9}$$

where each (s, t, c) observation is stacked over potential diffuser states $s' \neq s$. The term $\alpha_{ss'tc}$ denotes undirected state-pair $\times$ policy $\times$ year fixed effects, i.e. $\alpha_{ss'tc} = \alpha_{s'stc}$.[10]

As shown in Figure 12, the three similarity parameters are very close to those obtained in our baseline specification. The main difference is an even sharper decline in the role of geographic proximity, accompanied by substantial increases in the importance of political and economic similarity in recent decades.

Figure 12: Horizontal diffusion parameters with state pair by policy by year fixed effects



Notes: The figure shows the marginal effects from the coefficients ω^{sim} from equation (9), where we estimate separate coefficients by decade. Standard errors are clustered by adopter state by year and by adopter state by policy.

[10]State-pair $\times$ policy $\times$ year fixed effects absorb any unobserved determinants of adoption that are common to the pair in a given cluster-year—for example, a nationwide increase in the salience of a policy

6.5.2 Instrumenting states' adoptions with changes in observable characteristics

The model derived in the previous section naturally suggests using other states' observable characteristics, denoted by X_{st}, as instruments for their policy adoptions. These characteristics include, for example, the length of the legislative session, the salary of state legislators, the composition of the legislature, or NAICS2 employment shares. Such observables shift a state's baseline propensity to adopt policies within a given cluster and can therefore serve as sources of exogenous variation in other states' adoption decisions. Conditional on state A's own observable characteristics, changes in state B's observable characteristics plausibly affect state A's policy adoption only through their impact on state B's policy adoption. As before, our interest is not in the *level* of diffusion, but in heterogeneity across political, economic, and geographic dimensions. The identifying assumption is therefore that *differences* in characteristics of states that are similar and non similar affect adoption decisions only through the resulting differences in those states' policy adoptions. For example, if a nationwide shock would raise State A's propensity to legislate regardless of what other states do, then using the *difference* in (say) legislative-session-length changes between a politically similar state B and a dissimilar state C differences out that common shock, making it more plausible that the remaining variation affects A only because B (relative to C) adopts more policies.[11]

Because the effect of observable characteristics is likely to vary across policy clusters (for example, a Democratic governor may increase the propensity to adopt carbon pricing while reducing the likelihood of adopting a right-to-work policy), we estimate separate first-stage coefficients for each policy cluster. The first stage is therefore

issue (e.g., a Supreme Court ruling, a federal funding program, or a macro shock), a regional event affecting both states in the pair, or a coordinated advocacy/model-legislation campaign that targets both states at the same time. As a result, identification does not come from synchronized co-movements within the pair; it comes from within-pair differences in timing. Even with these fixed effects, our estimates can still be biased if an unobserved driver of policy demand hits the two states at different times and therefore generates within-pair asymmetries in adoption timing that mimic diffusion. This type of contamination is especially problematic if the underlying shock propagates more quickly within politically aligned, economically similar, or geographically proximate pairs, which can spuriously inflate the corresponding similarity coefficient.

[11] The exclusion restriction can still fail if the same underlying force that shifts X_{Bt} relative to X_{Ct} also affects state A's adoption incentives directly in a way that is systematically stronger for the "similar" group. For example, a national party organization or interest group could intensify organizing in co-partisan states, simultaneously expanding B's legislative capacity (raising X_{Bt}) and directly lobbying lawmakers in A on the same policy cluster—so that $X_{Bt} - X_{Ct}$ is correlated with A's adoption shocks even absent any channel through B's adoptions. That said, such a violation requires a fairly specific (and therefore more demanding) pattern: a targeted, similarity-indexed shock that both moves $X_{Bt} - X_{Ct}$ and directly shifts A's policy incentives.

policy-cluster-specific. In all specifications, we control for each state's own characteristics as well as average conditions in other states, both interacted with policy-cluster indicators.

We can write the first-stage equations as follows. As before, N_{-stc} captures the average stock of adoptions in all states other than s in year t for policy cluster c, while N^{sim}_{-stc} captures the stock of adoptions in states other than s that are similar to s along dimension $sim \in \{\text{pol}, \text{econ}, \text{geo}\}$. Let $\mathbf{X}_{stc}$ denote state s's observable characteristics, $\mathbf{X}_{-stc}$ their average across other states, and $\mathbf{X}^{sim}_{-stc}$ the corresponding average over states similar to s along dimension sim.

For each similarity dimension $sim \in \{\text{pol}, \text{econ}, \text{geo}\}$, the first stage is given by

$$N^{sim}_{-stc} = \omega^{all}_c \cdot \mathbf{X}_{-stc} + \omega^{sim}_c \cdot \mathbf{X}^{sim}_{-stc} + \zeta_c \cdot \mathbf{X}_{stc} + \rho_c N_{stc} + \alpha_c + \alpha_s + \alpha_t + \nu_{stc}. \quad (10)$$

The second stage can be written as

$$Y_{stc} = \delta_c N_{stc} + \beta_1 N_{-stc} + \sum_{sim} \beta_{sim} \widehat{N^{sim}_{-stc}} + \omega_c \cdot \mathbf{X}_{stc} + \omega^{all,2s}_c \cdot \mathbf{X}_{-stc} + \alpha_c + \alpha_s + \alpha_t + \epsilon_{stc}. \quad (11)$$

In the second stage, we control for each state's own characteristics $\mathbf{X}_{stc}$, the average conditions in the rest of the country $\mathbf{X}_{-stc}$, both interacted with policy-cluster indicators and the average stock of adoptions in other states N_{-stc}. These controls are also included in the first stage. Given the high dimensionality of the parameter space, we estimate the model in two steps using a control-function approach.

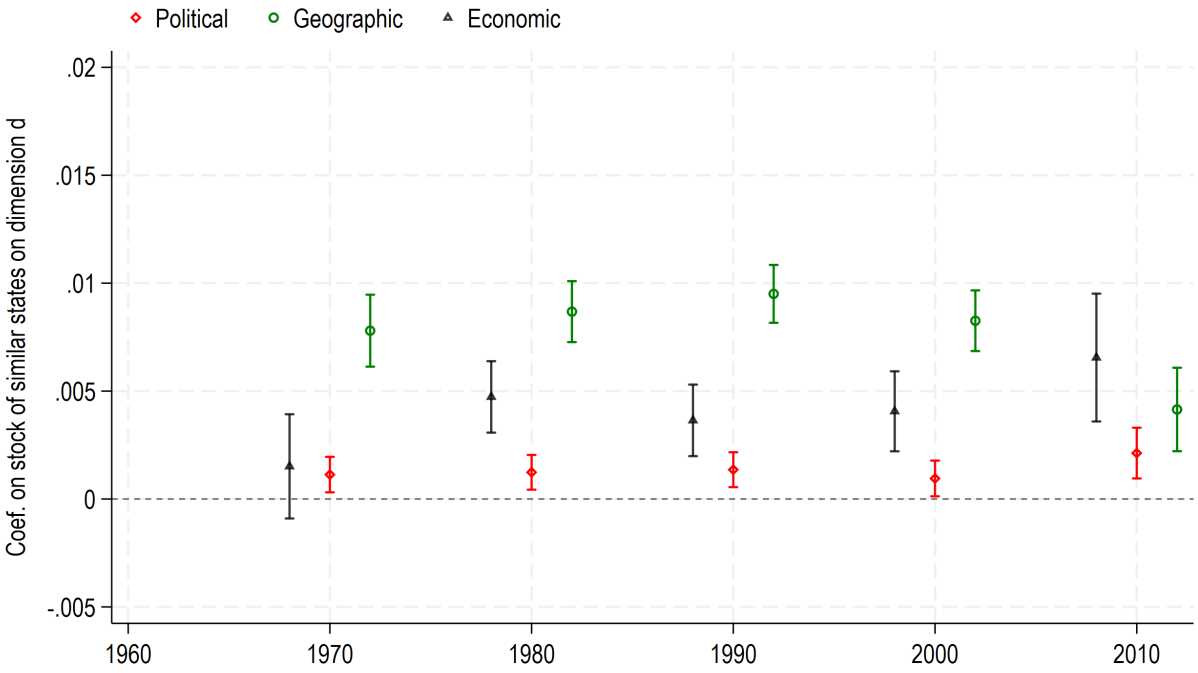
We use two main types of instruments: (i) the legislative capacity of the state, measured by the length of the legislative session as well as the salary and budget of legislators and (ii) the political composition of the state, measured by the share of seats in the legislature held by Democrats and an indicator equal to one if the governor is a Democrat.[12] Because the observable characteristics used as instruments (e.g., legislative capacity, party composition, and sectoral employment shares) are not consistently available before the 1970s, the IV analysis in this subsection is necessarily restricted to the post-1970 period.

We start by reporting the distribution of F-statistics across clusters.[13] 64% of clusters have an F-statistic above 10 (48% when using Lasso-selected instruments). Appendix Figure A.17 shows the distribution of the statistics across specifications.

[12] For now we do not use the economic composition of the state, measured by the first three principal components of the vector of NAICS2 employment sector shares as instruments given that the identifying assumptions for the economic composition instruments may be stronger.

[13] We refer to these as F-statistics for simplicity, but because we have multiple endogenous variables, these correspond to Kleibergen–Paap weak identification statistics.

Figure 13: Horizontal Diffusion with Instrumental Variables



Notes: The figure reports the marginal effects from the coefficients γ, which capture heterogeneity in horizontal diffusion across political, economic, and geographic dimensions. The coefficients are estimated using instrumental variables within a control-function framework. Standard errors are clustered by state by year and by state by policy.

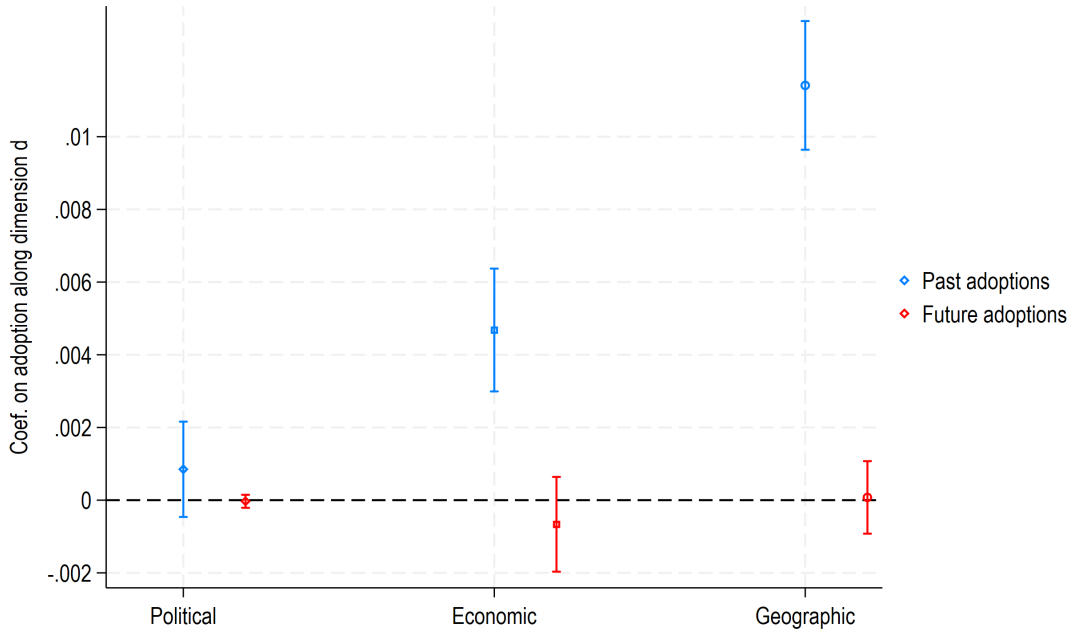
Figure 13 shows the evolution of the IV coefficients over time for the last five decades for the full set of clusters. Taken together, the coefficients are similar in magnitude compared to the earlier analysis without IV and therefore reinforce the baseline conclusion that horizontal diffusion is strongest among similar states, and that the relative importance of different similarity channels evolves over time.

This instrumental-variable strategy also enables us to conduct a placebo test for spurious diffusion. Rather than examining whether the stock of past peer adoptions predicts a given state's current adoption, we test whether the future stock of peer adoptions is correlated with current adoption. In a setting without instrumental variables, this would be uninformative, since a given state's current adoption is might influence peers' future adoptions, and thus be correlated with their future stock. With our instruments, based on changes in observable characteristics, this placebo test end up testing whether future peers' characteristics affect current adoption in a given state, to the same extent than current characteristics do.

Figure 14 reports the results. In this specification, which for simplicity, constrain the coefficients to be time-invariant, all three coefficients on future peer adoption exposure are close to zero and precisely estimated. This finding alleviates concerns that the baseline similarity-based spillovers reflect unobserved similarity-based shocks or slow-moving latent factors that jointly shape current and future adoptions (see Footnote 11 for further discussion).

One should note, however, that this test does not address all concerns regarding the validity of our instruments. In particular, it remains possible that contemporaneous shocks to peer states affect the focal state's adoption decision, while being uncorrelated with peers' future characteristics. For instance, if the focal state responds endogenously to peers' contemporaneous shocks rather than to their policy adoption per se.

Figure 14: Placebo: Future Adoptions by Peer States



Notes: The figure reports estimates from a placebo version of the baseline horizontal diffusion model, in which peer exposure is constructed using future peer adoption stocks over the subsequent $\tau = 10$ years, rather than lagged adoption stocks over the previous ten years, and is instrumented using the future observable characteristics of neighboring states. Under a diffusion interpretation, future peer adoptions should not affect current adoption, implying coefficients of zero.

7 Counterfactual Analysis

7.1 Counterfactual Diffusion Without Similarity Links

In this subsection, we quantify how geographic, economic, and political proximity between states shapes adoption dynamics over both short- and long-run horizons. We take the history of policy adoptions as of 2000 as given and assess the contribution of each proximity channel to subsequent adoption patterns for the average state. Our approach is to shut down each similarity link in turn and compute the resulting decline in the probability of adoption in an average cluster relative to the baseline model. We also report a counterfactual in which all three channels are simultaneously removed, allowing us to measure the aggregate role of proximity-driven diffusion.

Table 3 reports the percentage–point reduction in adoption probabilities at various horizons. Economic similarity has a modest but growing effect over time, geographic proximity matters somewhat less, and political alignment produces the largest single-channel effect. Shutting down all three channels simultaneously produces a substantial cumulative decline, reflecting the compounding nature of diffusion dynamics. Intuitively, these effects grow over time because, as fewer states adopt, the opportunities for cross-state influence shrink, leading to progressively weaker diffusion forces.

Table 3: Percentage–Point Decrease in the Probability of Adoption in an Average Cluster

	t	$t+2$	$t+4$	$t+6$	$t+8$
Economic	0.7	0.8	0.9	1.1	1.4
Geographic	0.2	0.3	0.5	0.7	0.9
Political	0.9	1.0	1.1	1.3	1.7
All	1.8	2.1	2.5	3.0	3.6

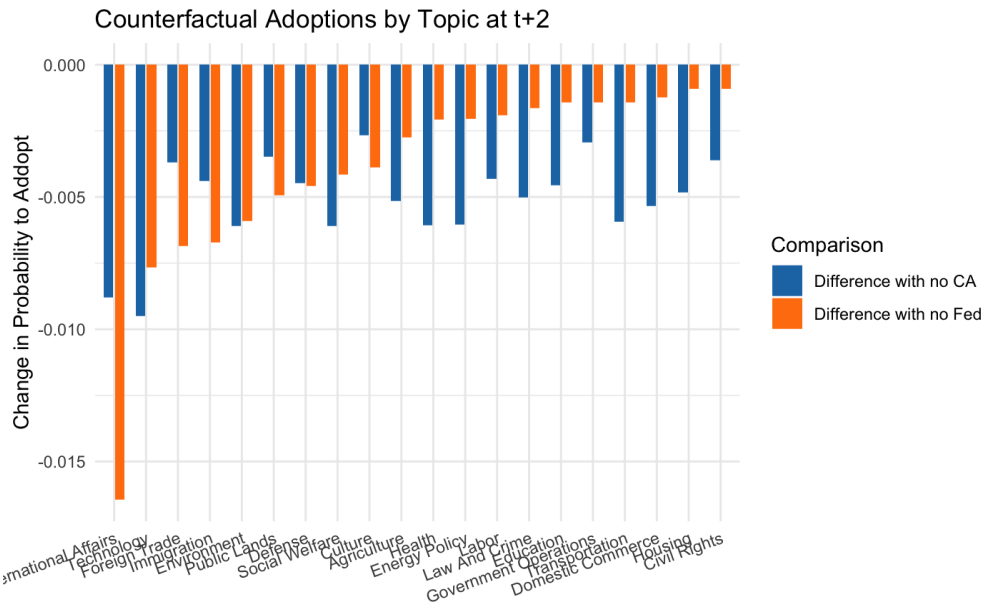
The table reports the percentage-point decline in the probability of adoption in an average cluster for the average state when economic, geographic, and political proximity channels are shut down individually and jointly, with effects shown at multiple horizons. For comparison, note that the baseline probability of adoption is equal to 0.17.

7.2 Counterfactual Diffusion without Major Jurisdictions

In our next counterfactual exercise, we quantify the influence of key legislatures across policy domains. Taking the history of adoptions as of 2000 as given, we compute the percentage-point change in an average state's probability of adopting in each topic's average cluster under two scenarios: (i) California's influence is reduced to that of the average state, and (ii) the Federal Government's influence is reduced to that of the

average state. This allows us to assess how much these pivotal jurisdictions shape diffusion patterns across topics.

Figure 15



The figure reports the percentage-point decline in an average state's probability of adopting in a topic's average cluster when either California's influence or the Federal Government's influence is reduced to that of the average state. For reference, the baseline probability of adoption during this period is 26.33%.

Figure 15 shows substantial heterogeneity across policy domains in how much California and the Federal Government influence subsequent diffusion. For nearly all topics, eliminating either actor's influence reduces the probability that an average state adopts in that domain. The magnitude, however, varies markedly: federal influence is especially pronounced in domains with a strong national or regulatory dimension—such as International Affairs, Technology, Trade, and Immigration–where removing federal spillovers leads to the largest declines. California's influence, by contrast, is more diffuse across a broader range of topics, with particularly noticeable effects in areas such as Environment, Public Lands, Health, Labor, and Education. Across most domains, the reduction in adoption probabilities at $t + 2$ is larger when federal influence is removed, but the pattern also reveals several policy areas in which California plays a comparably central or even dominant role.

8 Conclusion

In this paper, we introduced a new methodology for mapping the universe of all U.S. state and federal laws into clusters of legally related policies. We then develop and estimate an empirical model that allows us to examine how interactions among states, and between states and the federal government, have shaped policy adoption over time.

Partisanship has become increasingly predictive of state-to-state policy similarity in recent decades, complementing earlier results in the literature and extending them further back in time. Contrary to common perceptions, the recent rise in polarization does not surpass historical peaks, indicating that the United States has experienced comparably polarized eras in the past.

We also documented growing federal influence on states. While geography played a substantial role in historical policy diffusion, over time it has shrunk compared to economic and partisan factors, with the federal government now emerging as a major reference point. This pattern is consistent with a nationalizing political environment, though future work is needed to clarify the precise mechanisms of federal–state policy interactions.

Future work could begin to investigate how states enact variations of federal laws, whether they expand or restrict federal provisions. Another interesting angle would be to evaluate the influence of interest groups, for example with model legislation (Burgess et al., 2016; Hertel-Fernandez, 2019*b*). It would also be interesting to explore how innovations in communication technologies – railways, telegraphs, telephones, the internet, and AI – have and will shape the spread of policy ideas.

References

Acemoglu, Daron, Camilo García-Jimeno and James A. Robinson. 2015. "State Capacity and Economic Development: A Network Approach." *American Economic Review* 105(8):2364–2409.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.20140044*

Arold, Benjamin W, Elliott Ash, W Bentley MacLeod and Suresh Naidu. 2024. "Do Words Matter?".

Ash, Elliott. 2018. "What drives partisan tax policy? The effective tax code." *Center for Law & Economics Working Paper Series* 4.

Ash, Elliott, Aniket Kesari, Suresh Naidu, Lena Song and Dominik Stammbach. 2024. "Translating Legalese.".

Ash, Elliott, Massimo Morelli and Matia Vannoni. 2025. "More Laws, More Growth? Evidence from US States." *Journal of Political Economy* 133(7):2139–2179.
URL: *https://doi.org/10.1086/734874*

Ash, Elliott, Massimo Morelli and Moritz Osnabrügge. 2019. "Electoral Systems and Political Attention: Evidence from the Electoral Reform in New Zealand.".

Ash, Elliott, Massimo Morelli and Richard Van Weelden. 2017. "Elections and Divisiveness: Theory and Evidence." *The Journal of Politics* 79(4):1268–1285.

Ash, Elliott and Stephen Hansen. 2023. "Text algorithms in economics." *Annual Review of Economics* .

Battaglini, M. and E. Patacchini. 2018. "Influencing Connected Legislators." *Journal of Political Economy* 126(6):2277–2322.

Bernecker, Andreas, Pierre C. Boyer and Christina Gathmann. 2021. "The Role of Electoral Incentives for Policy Innovation: Evidence from the US Welfare Reform." *American Economic Journal: Economic Policy* 13(2):26–57.
URL: *https://www.aeaweb.org/articles?id=10.1257/pol.20190690*

Berry, Frances Stokes and William D. Berry. 1990. "State Lottery Adoptions as Policy Innovations: An Event History Analysis." *American Political Science Review* 84(2):395–415.

Besley, Timothy and Anne Case. 1995. "Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition." *American Economic Review* 85(1):25–45.

Besley, Timothy and Stephen Coate. 2003. "Centralized versus Decentralized Provision of Local Public Goods: A Political Economy Approach." *Journal of Public Economics* 87(12):2611–2637.

Boehmke, Frederick J., Mark Brockway, Bruce A. Desmarais, Jeffrey J. Harden, Scott LaCombe, Fridolin Linder and Hanna Wallach. 2020. "SPID: A New Database for Inferring Public Policy Innovativeness and Diffusion Networks." *Policy Studies Journal* 48(2):517–545.

Boehmke, Frederick J. and Richard Witmer. 2004. "Disentangling Diffusion: The Effects of Social Learning and Economic Competition on State Policy Innovation and Expansion." *Political Research Quarterly* 57(1):39–51.

Burgess, Matthew, Eugenia Giraudy, Julian Katz-Samuels, Joe Walsh, Derek Willis, Lauren Haynes and Rayid Ghani. 2016. The legislative influence detector: Finding text reuse in state legislation. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 57–66.

Cai, Hongbin and Daniel Treisman. 2009. "Political Decentralization and Policy Experimentation." *Quarterly Journal of Political Science* 4(1):35–58.

Callander, Steven and Bård Harstad. 2015. "Experimentation in Federal Systems." *The Quarterly Journal of Economics* 130(2):951–1002.

Canen, Nathan, Matthew O. Jackson and Francesco Trebbi. 2023. "Social Interactions and Legislative Activity." *Journal of the European Economic Association* 21(3):1072–1118.

de Paula, Áureo, Imran Rasul and Pedro C L Souza. 2024. "Identifying Network Ties from Panel Data: Theory and an Application to Tax Competition." *The Review of Economic Studies* 92(4):2691–2729.
URL: *https://doi.org/10.1093/restud/rdae088*

Dell, Melissa. 2025. "Deep learning for economists." *Journal of Economic Literature* 63(1):5–58.

DellaVigna, Stefano and Woojin Kim. 2025. "Policy Diffusion and Polarization across U.S. States." *The Review of Economic Studies* . Advance article, published online

August 3, 2025.
URL: *https://doi.org/10.1093/restud/rdaf070*

Desmarais, Bruce A., Jeffrey J. Harden and Frederick J. Boehmke. 2015. "Persistent Policy Pathways: Inferring Diffusion Networks in the American States." *American Political Science Review* 109(2):392–406.

Eckert, Fabian, Teresa C Fort, Peter K Schott, Natalie J Yang et al. 2020. "County Business Patterns Database.".

Elazar, Daniel J. 1966. *American Federalism: A View from the States*. New York: Thomas Y. Crowell Company.

Elazar, Daniel J. 1987. *Exploring Federalism*. Tuscaloosa, AL: University of Alabama Press.

Feeley, Malcolm M. and Edward L. Rubin. 2008. *Federalism: Political Identity and Tragic Compromise*. Ann Arbor: University of Michigan Press.

Fouirnaies, Alexander and Andrew B. Hall. 2017. "How Do Interest Groups Seek Access to Committees?" *American Journal of Political Science* 61(1):132–147.

Gennaro, Gloria and Elliott Ash. 2021. "Emotion and Reason in Political Language." *Economic Journal* 2021(02).

Gentzkow, Matthew, Bryan T Kelly and Matt Taddy. 2017. Text as Data. Technical Report 3 National Bureau of Economic Research.

Gentzkow, Matthew and Jesse M Shapiro. 2010. "What drives media slant? Evidence from US daily newspapers." *Econometrica* 78(1):35–71.

Gentzkow, Matthew, Jesse M Shapiro and Matt Taddy. 2017. "Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech.".

Greene, Zachary and Daniel Bowen. 2020. "Legislative Professionalism.". Dataset.
URL: *https://doi.org/10.7910/DVN/27595*

Grumbach, Jacob M. 2018. "From Backwaters to Major Policymakers: Policy Polarization in the States, 1970–2014." *Perspectives on Politics* 16(2):416–435.

Harmon, Nikolaj, Raymond Fisman and Emir Kamenica. 2019. "Peer Effects in Legislative Voting." *American Economic Journal: Applied Economics* 11(4):156–180.

Hertel-Fernandez, Alexander. 2019*a*. *State Capture: How Conservative Activists, Big Businesses, and Wealthy Donors Reshaped the American States—and the Nation*. New York: Oxford University Press.

Hertel-Fernandez, Alexander. 2019*b*. *State capture: how conservative activists, big businesses, and wealthy donors reshaped the American states–and the nation*. Oxford University Press, USA.

Hopkins, Daniel J. 2018. *The Increasingly United States: How and Why American Political Behavior Nationalized*. Chicago, IL: University of Chicago Press.

Karch, Andrew. 2006. "National Intervention and the Diffusion of Policy Innovations." *American Politics Research* 34(4):403–426.

Karch, Andrew and Aaron Rosenthal. 2016. "Vertical Diffusion and the Shifting Politics of Electronic Commerce." *State Politics & Policy Quarterly* 16(1):22–43.

Klarner, Carl. 2015. "State Economic and Government Finance Data." *Harvard Dataverse, Version* 1.

Luo, Kaicheng, Shaoda Wang and David Y. Yang. 2025. Laboratories of Autocracy: Landscape of Central–Local Dynamics in China's Policy Universe. NBER Working Paper 34219 National Bureau of Economic Research.
URL: *https://www.nber.org/papers/w34219*

Madison, James. 1788. *The Federalist Papers*. New York: Modern Library. Federalist No. 45.

Mallinson, Daniel J. 2021. "Who are your neighbors? The role of ideology and decline of geographic proximity in the diffusion of policy innovations." *Policy Studies Journal* 49(1):67–88.

Manski, Charles F. 1993. "Identification of endogenous social effects: The reflection problem." *The review of economic studies* 60(3):531–542.

Manson, Steven M. 2020. "IPUMS national historical geographic information system: version 15.0.".

McCubbins, Mathew D. and Thomas Schwartz. 1984. "Congressional Oversight Overlooked: Police Patrols vs. Fire Alarms." *American Journal of Political Science* 28(1):165–179.

Mookherjee, Dilip. 2015. "Political Decentralization." *Annual Review of Economics* 7:231–249.

Mooney, Christopher Z. 2021. *The Study of US State Policy Diffusion: What Hath Walker Wrought?* Elements in American Politics Cambridge University Press.

Mulligan, Casey B and Andrei Shleifer. 2005. "The Extent of the Market and the Supply of Regulation." *The Quarterly Journal of Economics* 120(4):1445–1473.

Oates, Wallace E. 1999. "An Essay on Fiscal Federalism." *Journal of Economic Literature* 37(3):1120–1149.

Peters, B Guy. 2021. *Advanced introduction to public policy*. Edward Elgar Publishing.

Rogers, Everett M. 2003. *Diffusion of Innovations*. 5th ed. Free Press.

Schapiro, Robert A. 2009. *Polyphonic Federalism: Toward the Protection of Fundamental Rights*. Chicago: University of Chicago Press.

Shor, Boris and Nolan McCarty. 2011. "The Ideological Mapping of American Legislatures." *American Political Science Review* 105(3):530–551.

Squire, Peverill. 2023. "Replication Data for: A Squire Index Update: Stability and Change in Legislative Professionalization, 1979–2021.". Dataset (updated Squire Index components and professionalism scores).
URL: *https://doi.org/10.15139/S3/2VORBN*

Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64(5):416–424.

Toscano, David J. 2021. *Fighting Political Gridlock: How States Shape Our Nation and Our Lives*. Charlottesville: University of Virginia Press.

Vannoni, Matia, Elliott Ash and Massimo Morelli. 2021. "Measuring discretion and delegation in legislative texts: methods and application to US states." *Political Analysis* 29(1):43–57.

Volden, Craig, Michael M. Ting and Daniel P. Carpenter. 2008. "A Formal Model of Learning and Policy Diffusion." *The American Political Science Review* 102(3):319–332.
URL: *http://www.jstor.org/stable/27644523*

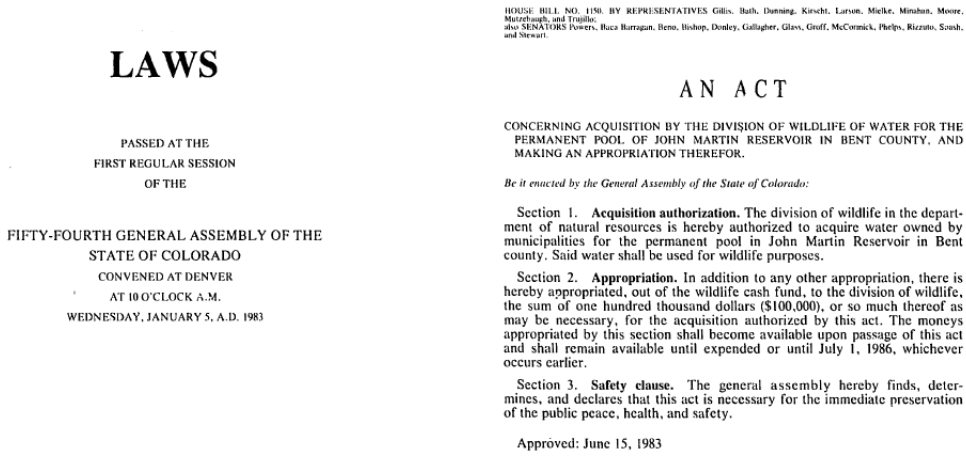
Walker, Jack L. 1969. "The Diffusion of Innovations among the American States." *American Political Science Review* 63(3):880–899.

Wang, Shaoda and David Y. Yang. 2025. "Policy Experimentation in China: The Political Economy of Policy Learning." *Journal of Political Economy* 133(7):2180–2228.

Weingast, Barry R. 1995. "The Economic Role of Political Institutions: Market-Preserving Federalism and Economic Development." *Journal of Law, Economics, & Organization* 11(1):1–31.

Widmer, Philine, Sergio Galletta and Elliott Ash. 2020. "Media Slant is Contagious." *Center for Law & Economics Working Paper Series* 14.

Figure A.1: U.S. State Session Laws Corpus



LAWS

PASSED AT THE
FIRST REGULAR SESSION
OF THE

FIFTY-FOURTH GENERAL ASSEMBLY OF THE
STATE OF COLORADO
CONVENED AT DENVER
AT 10 O'CLOCK A.M.
WEDNESDAY, JANUARY 5, A.D. 1983

HOUSE BILL NO. 1150. BY REPRESENTATIVES Gillis, Bath, Dunning, Kirscht, Larson, Mielke, Minahan, Moore, Mutzebaugh, and Trujillo;
also SENATORS Powers, Ruca Barragan, Beno, Bishop, Donley, Gallagher, Glass, Groff, McCormick, Phelps, Rizzuto, Soash, and Stewart.

AN ACT

CONCERNING ACQUISITION BY THE DIVISION OF WILDLIFE OF WATER FOR THE PERMANENT POOL OF JOHN MARTIN RESERVOIR IN BENT COUNTY, AND MAKING AN APPROPRIATION THEREFOR.

Be it enacted by the General Assembly of the State of Colorado:

Section 1. **Acquisition authorization.** The division of wildlife in the department of natural resources is hereby authorized to acquire water owned by municipalities for the permanent pool in John Martin Reservoir in Bent county. Said water shall be used for wildlife purposes.

Section 2. **Appropriation.** In addition to any other appropriation, there is hereby appropriated, out of the wildlife cash fund, to the division of wildlife, the sum of one hundred thousand dollars ($100,000), or so much thereof as may be necessary, for the acquisition authorized by this act. The moneys appropriated by this section shall become available upon passage of this act and shall remain available until expended or until July 1, 1986, whichever occurs earlier.

Section 3. **Safety clause.** The general assembly hereby finds, determines, and declares that this act is necessary for the immediate preservation of the public peace, health, and safety.

Approved: June 15, 1983

A Text Analysis Appendix

Figure A.9: LLM Prompt: Distinguishing Acts and Resolutions

Prompt

```
I will provide the full text of a statute from a U.S. state legislature or the U.S. Congress.  The
statute text is derived from a digital scan and may contain OCR errors or lack punctuation.

Your task is to distinguish acts from resolutions.  An act is a law formally enacted by a legislature
that has binding legal force.  A resolution is a formal statement of the legislature's opinion,
decision, or intent that lacks the force of law.

Answer only 'Act', 'Resolution', or 'Other'.  Answer 'Other' if the text is not an act or a resolution,
or if there is some other problem with the text that prevents a classification.  If you answer 'Other',
follow up the answer with a semicolon ';' and a short explanation.

Classify the following statute text:  {text}
```

LLM used to annotate legislative documents as acts or resolutions. This is done if the dictionary method does not tag the document. The dictionary assigns "act" if the words 'act' or 'enact' or variants show up in the first 10 words. The dictionary assigns 'resolution' if 'resolution' or variants shows up in the first 10 words. If neither tag hits (about 5% of the time), the LLM approach is used.

Figure A.2: LLM Prompt: Policy Extraction and Summarization

Prompt

I will provide the full text of a statute from a U.S. state legislature or the U.S. Congress. The statute text is derived from a digital scan and may contain OCR errors or lack punctuation.

Your task is to extract the policy or set of policies described in the statute text. A policy is a self-contained government action or bundle of closely connected government actions with a socioeconomic (societal and/or economic) goal or impact. Each policy has an associated measure type: introduction, modification, or repeal.

Carefully analyze the text to identify and extract all policies and associated measure types. For each policy, perform the following three tasks:

(1) Designate the measure type, as follows:

- 'introduction': To be chosen when the statute introduces the policy. This is the default type, as most statutes introduce a policy.
- 'modification': To be chosen if it is clear from the statute text that the policy already exists and is being modified.
- 'repeal': To be chosen if it is clear from the statute text that a policy is being completely removed and/or replaced with a different policy.

(2) Produce a one-sentence summary that describes the associated policy, starting with 'A policy that...'. Depending on the measure type, the summary should:

- for introductions, describe the introduced policy.
- for modifications, describe the resulting policy.
- for repeals, describe the repealed policy -- i.e., what the policy was before being repealed. If the policy being repealed is not described, write '[Undescribed Repealed Policy]' as the summary.

The one-sentence summary should utilize all of the relevant text associated with the policy. The one-sentence summary can be as long as needed, to capture the full detail of the associated policy. The summary should not include procedural or legalistic details such as setting dates, changing names, amending debate rules, re-organizing legal sections, or mentions of specific statute numbers or code sections.

(3) Provide a short label, up to five words, summarizing the measure and policy.

[. . . *additional instructions, see Appendix Figure A.10* . . .]

Here is the statute text: {raw statute}

Figure A.3: Examples of Extracted Policies

Example Extracted Policies: California Statute (1986)

Response

Policy 1: Investigation of Onsite Condition Discrepancies (introduction)

A policy that mandates the Office of the State Architect of the Department of General Services to investigate onsite condition discrepancies and, to the extent that any changes to the contract are considered by the Department of Veterans Affairs to be in the best interests of the State, to make and approve change orders required to make the facility ready for occupancy and to account for the capital outlay as reimbursement for the appropriation.

Policy 2: Federal Grant Application for Construction Costs (introduction)

A policy that requires the Department of Veterans Affairs to apply for a federal grant for the construction cost of the Veterans Home, to ensure that the proceeds from the grant are deposited in the special account for capital outlay in the general fund.

Policy 3: Veterans Home Construction Funding (introduction)

A policy that appropriates $475,000 from a special account to the Department of General Services for expenditure in augmentation of funds to provide for construction cost increases in the project and related master plan projects.

Example Extracted Policies: Federal Statute (1957)

Response

Policy 1: Reporting Requirements for Foreign Currency Agreements (introduction)

A policy that requires a full report to be made to the U.S. Senate and House of Representatives whenever any agreement is entered into for the use of foreign currencies.

Policy 2: Trade Restrictions with Communist Nations (introduction)

A policy that authorizes the President to assist friendly nations by restricting trade with the Soviet Union or areas controlled by the Communist regime in China.

Policy 3: Foreign Currency Loan Provisions (introduction)

A policy that authorizes up to 25% of foreign currencies received by the United States to be used for loans, and allows acceptance of foreign currencies in loan repayment.

Figure A.4: Distribution of the number of policies per statute.

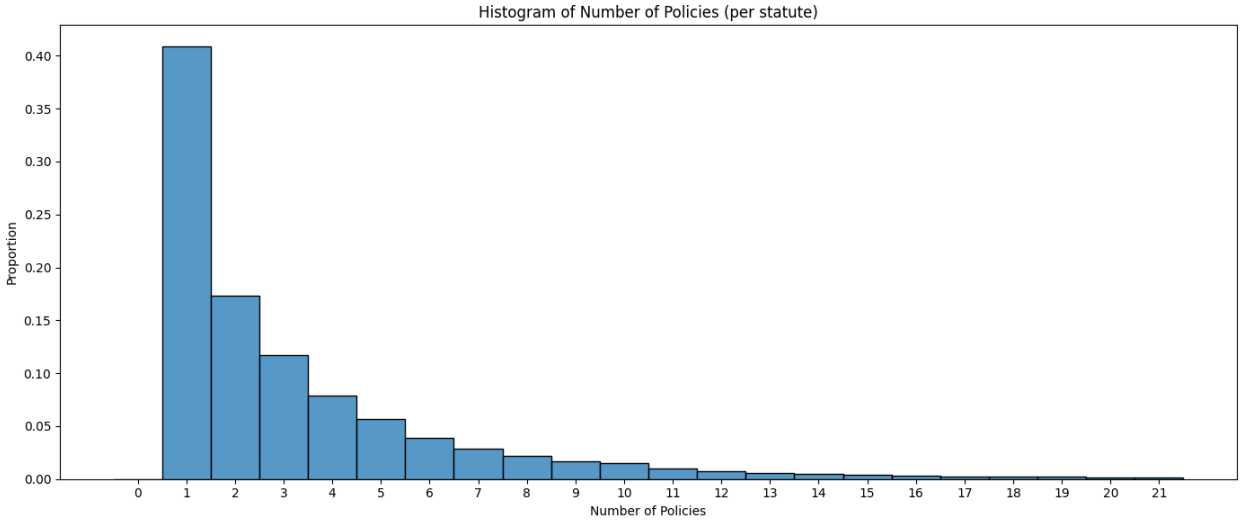


Figure A.10: LLM Prompt Addendum: Policy Extraction and Summarization

Prompt

Notes:

- Make sure that every policy and associated measure is included in the output. Oftentimes, statutes describe a single policy. Avoid treating minor sub-policies as distinct if they logically fall under a broader policy.
- Each label and policy summary in the list should be self-contained and understandable by itself. The summary should not refer to any other policies from the statute text. Each summary should include sufficient detail that it is not necessary to read the summaries of the other policies to understand it.
- If the statute text both introduces some parts of a policy and modifies other parts, indicate ’’modification" as the measure type.
- If the statute text repeals a policy and replaces it with a new one, list the new policy with an ’’introduction" measure type and separately list the old policy with a ’’repeal" measure type.
- The list of policies should not include actions or other legislative material that are not enacted policies. Non-policies include actions that do not have a meaningful socioeconomic impact outside the legislature. Here are some examples of non-policies:
  - Failed or vetoed policies: If it is clear from the context that the listed policy was not legislatively enacted, because it did not pass one the chambers or it was vetoed, that should not be included.
  - Non-legislative material: Table of contents, legislative journal transcripts, etc.
  - Commemorating an historical event or naming a post office;
  - Amending chamber debate rules or creating a legislative committee;
  - Reorganizing code sections, correcting typos, or renaming government agencies.
- If the statute does not describe any meaningful policies, list one item with ’’measure_type" set to ’’not_policy" and use ‘‘label’’ and ’’summary" to briefly explain what the statute is doing.
- If the task cannot be performed (e.g., because the text is empty or does not include legislative material), list one item with ’’measure_type" set to ’’n_a" and use ‘‘label’’ and ’’summary" to briefly explain why the task cannot be performed.

The output format is a list of jsons, where each item corresponds to a policy, with the following items:

’label’: string, // a brief phrase describing the topic of the measure and policy ’summary’: string, // the single-sentence summary of the policy ’measure_type’: string, // the measure type associated with the policy: introduction, modification, repeal, not_policy, or n_a.

For instance: [’label’: ’Drunk Driving Restriction’, ’summary’: ’A policy that bans driving with a blood alcohol concentration above a specified threshold.’, ’measure_type’: ’introduction’, ’label’: ’Backseat Seat Belt Requirement’, ’summary’: ’A policy that requires that backseat occupants wear seat belts.’, ’measure_type’: ’modification’, ’label’: ’Repeal of Speeding Cameras’, ’summary’: ’A policy that requires speeding cameras on all highways.’, ’measure_type’: ’repeal’]

Or if the text does not involve a meaningful policy: [’label’: ’Renaming a Post Office’, ’summary’:

Figure A.5: LLM Prompt: Policy Anonymization

Prompt

I will provide a legislative text coming from a U.S. state legislature or the U.S. Congress. Your task is to fully anonymize the text by removing all specific identifying information.

Instructions:

- Keep the text as close to the original as possible.
- Replace names of specific people, organizations, and places/jurisdictions with generic terms (e.g., ‘‘a specific city,’’ ‘‘a specific individual’’, ‘‘a specific corporation’’).
- Remove mentions of the originating jurisdiction or replace with ‘‘this jurisdiction’’ if needed for clarity.
- Replace mentions of locations outside the originating jurisdiction with ‘‘another specific jurisdiction’’, ‘‘another specific jurisdiction’s city’’, etc.
- Do not anonymize mentions of government agencies or job titles, except to remove the names of jurisdictions. For example: ‘the Texas Department of Transportation’ would become ‘the Department of Transportation’; ‘Texas Secretary of State Smith’ would become ‘the Secretary of State’.
- Anonymize mentions of specific dates or time periods. For example:
  - Replace ’’July 31, 1997" with ’’a specific date"
  - Replace ’’July 1997" with ’’a specific month"
  - Replace ’’1997" with ‘‘a specific year’’
  - Replace ’’July 31 of each year" with ’’a specific date of each year".
- If appropriate and feasible, indicate the time relative to the originating year. For example, if the originating year is 1996, but the text says ’’1997", write ’’next year".
- Anonymize mentions of specific numbers. For example:
  - Replace dollar amounts with "a specific dollar amount".
  - Replace statutory section numbers with "a specific statute section".
  - Replace population thresholds with ’’a specific population threshold".
  - Replace fractions (e.g. ’’one-fourth of") or proportions (’’0.25 of", ’’25% of") with ’’some share of".

Respond only with the fully anonymized text.

The legislative text’s originating jurisdiction: {STATE}. The legislative text’s originating year: {YEAR}.

Here is the legislative text: {text}.

Figure A.6: Topic Assignment of Policy Summaries with GPT

Prompt

Your task is to categorize a legislative text into one of the following legislative topics:

{Tax Policy, Budget Policy, Monetary Policy, Civil Rights, Health, Agriculture, Labor, Education, Environment, Energy, Immigration, Transportation, Crime, Courts and Judicial Administration, Social Welfare, Housing, Domestic Commerce, Defense, Technology, Foreign Trade, International Affairs, Government Operations, Public Lands, Culture}

Fill in the 'primary' field with the primary topic. If the text does not fit into any of these topics, set 'primary' to 'Other', and provide an explanation with an alternative label in the 'explanation' field. If the text cannot be assigned to a topic at all, for example because the text is missing, set 'primary' to 'N/A', and provide an explanation in the 'explanation' field.

If applicable, include secondary topics in a 'secondary' field. List all secondary topics, separated by semicolons, in their order of importance.

Note: Pick topics only from the provided list, or 'Other'. Prioritize selecting only one or the fewest possible topics, focusing on relevance.

[. . .]

Here is the legislative text:{text}.

Figure A.7: LLM Prompt: Identifying Similar and Dissimilar Policy Pairs

Prompt

Your task is to compare the text descriptions of a pair of policies and evaluate whether the policies are related or not, and if they are related, the level of similarity.

In particular:

- A policy is a self-contained government action or bundle of actions pursuing a socioeconomic (societal and/or economic) goal.
- Policies are related if they describe the same policy or very closely related policies, in terms of the described government action(s) { e.g., in terms of which policy instrument is used, how it is used, and the population affected;
- Examples of important differences that imply different policies:
  - If government actions concern or regulate different technologies (e.g., telegraph vs. telephone, or telephone vs. internet), then they are not related.
  - Taxes on different tax bases (e.g. personal income vs. sales) should be considered unrelated.
  - Regulations targeting different sectors (e.g., manufacturing vs. agriculture) should be treated as unrelated.
  - Government expenditures on different categories of public goods (e.g., education vs. sanitation vs. transportation) should be treated as unrelated.
- Policies are not related if the described actions clearly move in different directions relative to the status quo. Examples of opposing actions:
  - increasing sales taxes vs. decreasing sales taxes.
  - increasing funding for a given agency vs. decreasing funding for that same agency.
  - Strengthening regulations on a business activity vs. deregulating that same activity.

Are the provided text descriptions describing related policies? Carefully read them and provide one of the following responses:

1. Definitely yes: exactly or almost identical government actions.
2. Yes: closely related government actions.
3. Probably yes: somewhat related government actions.
4. Probably not: somewhat unrelated government actions.
5. Not: unrelated government actions.
6. Definitely not: related but clearly opposing government actions (e.g. increasing vs. decreasing a tax on the same base; increasing vs. decreasing funding for the same government service).
7. Definitely not: totally unrelated government actions.
8. Not applicable, cannot answer: for example, because one or both of the texts is empty or does not describe a policy.

Answer only with an integer from the range 1 through 8.

Here is the pair of text descriptions, labeled Statute A and Statute B:

Statute A: {statute A}

Statute B: {statute B}

Figure A.8: Distribution of Tagged Similarity Scores for Policy Pairs

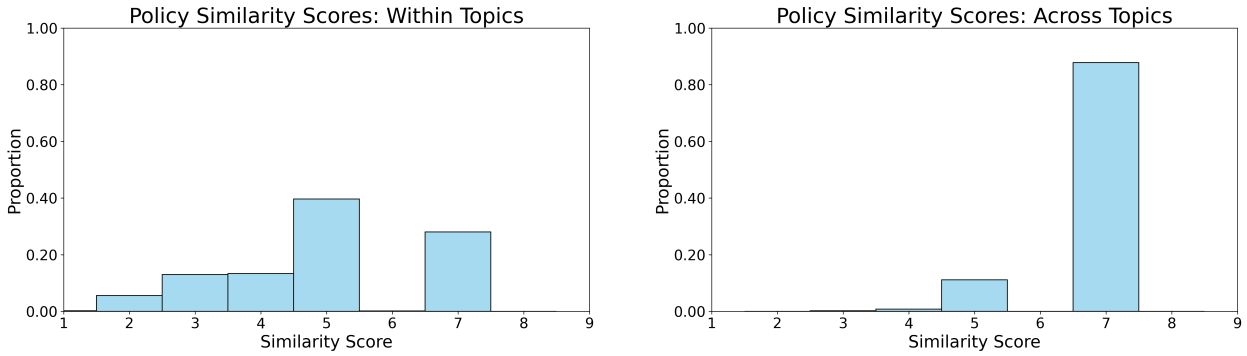


Figure A.11: Subtopic Assignment of Policy Summaries with GPT

User Prompt

Input Example:
Here is a summary of a state statute: "*Turnpike Road Company Organization. This policy authorizes the organization and division of a specific turnpike road company into sections, and the creation of another specific turnpike road company to complete construction.*" Here is the main topic of the policy summary: [primary topic]

Prompt

```
Your task is to categorize this policy summary among a list of possible policy subtopics.
Here is the list of different policy subtopics along with their definition: [subtopics of the primary topic]
Respond with the primary policy subtopic in the format: Primary: [policy subtopic name].
If applicable, you may include secondary subtopics by adding: Secondary: [policy subtopic names].
Prioritize selecting only one or the fewest possible subtopics, focusing on relevance and listing them in order of importance.
```

Response

```
Output Example:
''Highways''
```

Figure A.12: Identification of Distinct Policies in Raw Statutes Using GPT

Prompt

```
Here is a statute from a U.S. state or federal government that enacts one or more policies.  The
statute text is derived from a digital scan and may contain OCR errors or lack punctuation.
Carefully analyze the text to identify and summarize the enacted policy or policies:

   1. If the statute enacts only one policy, summarize it as:

         • Start with:  "This policy..."
         • Combine all sub-policies under one overarching description.

   2. If the statute enacts multiple distinct policies, list them as:

         • "1.  This policy...; 2.  This policy..."
         • Ensure each is distinct, non-overlapping, and less than 20 words.

Avoid treating minor sub-policies as distinct if they logically fall under a broader policy.
```

User Prompt

Input Example:
Here is the statute:

> *"An Act relating to all counties having a population of not less than 56,500 nor more than 59,000 inhabitants, according to the 1970 or any subsequent federal decennial census, to provide an additional expense allowance for members of the county commission.*
>
> Be it enacted by the legislature of Alabama:
>
> **Section 1:** The provisions of this act shall apply to all counties having a population of not less than 56,500 nor more than 59,000 inhabitants, according to the 1970 or any subsequent federal decennial census.
>
> **Section 2:** In addition to any and all other salary, compensation, and expense allowances provided for by law, the members of the county commission in counties to which this act applies shall be paid an additional expense allowance in the amount of $150 per month from the county general fund.
>
> **Section 3:** The provisions of this act shall become effective on the first day of the month next following the date this act becomes law. This act became a law under Section 125 of the Constitution on May 17, 1977, without approval by the governor.
>
> Act No. 622, H. 968, Crawford Sasser."

Response

Output Example:

`` 1. This policy provides an additional monthly expense allowance of $150 for county commission members in specific Alabama counties. ''

Remark

For comparison, here is the statute summary computed with GPT-3.5-turbo:
"*County Commission Expense Allowance. This policy provides an additional expense allowance for members of the county commission in specific population range counties, effective the month following enactment.*"

Notes: The raw statute was corrected from OCR errors using ChatGPT.

Table A.1: Number of clusters by policy topic

Topic	Number of Clusters
Agriculture	128
Budget Policy	32
Civil Rights	96
Courts and Judicial Administration	64
Crime	192
Culture	32
Defense	32
Domestic Commerce	48
Education	96
Energy	48
Environment	128
Foreign Trade	192
Government Operations	224
Health	384
Housing	64
Immigration	96
International Affairs	128
Labor	64
Monetary Policy	8
Public Lands	160
Social Welfare	48
Tax Policy	64
Technology	96
Transportation	64
Total	2488

A.1 Policy Cluster Validation

This section validates the approach outlined in Section 3, focusing on the vectorization and construction of policy clusters. Specifically, we aim to demonstrate that our chosen model is the most robust among existing classical encoders and that the number of clusters effectively groups policies with similar goals.

A.1.1 Cluster Count Validation

Section 3.3 describes our K-means clustering methodology and results, which requires the pre-specification of the number of clusters, denoted as k. Determining the best k is a complex problem, as no universally accepted method exists. Still, there are various metrics that can help in making this choice. It is important to note that there is no "optimal" value of k in clustering, as the best k depends on the specific context: it is contingent upon the data's distribution, inherent structure, and the specific objectives of the analysis. In our case, we are interested in clustering policies within a broad topic. For example, within the context of Agriculture-based legislation, how many meaningful groups can we assign policies to using their semantic content (irrespective of state and time). Is there, for example, a subset of policies which seek to regulate the use of pesticides? Our data-driven approach attempts to find these groupings of related policies at a reasonable granularity.

There are 3 main components to our validation approach: the embedding model, clustering using a set of pre-specified ks, and choosing the correct model and k combination based on an F1 score.

Embedding Fine-tuning. First, for each topic, we fine-tune two separate embedding models using Qwen3 (Section 3.2) which differ in the distance metric chosen for the contrastive loss, i.e., Euclidean versus cosine distance. We do this because we do not know ex-ante which metric will yield a better fine-tuned embedding space.

Clustering using a set of ks. Next, with these fine-tuned models, we encode each extracted policy text and then cluster them using k-means – again within each topic. Our search space of ks is {1, 2, 3, 4, 6, 8, 12, 16, 24, 32, 48, 64, 96, 128, 160, 192, 224, 256, 320, 384, 448, 512, 584, 667, 759}. This set of candidate K values is approximately geometric to reflect the nonlinear way partition structure evolves in high-dimensional embedding spaces. At small K, minor increases can substantially reorganize clusters by splitting large, heterogeneous groups along major semantic dimensions. At larger K, clusters are already smaller and more homogenous, so additional increases produce

only marginal refinements. Geometric spacing therefore concentrates search density where structural changes are most consequential while avoiding redundant evaluation of adjacent high-K values.

Choosing the best F1 Lastly, we validate the model on the set of sampled policy pairs. For each pair, we take the GPT-labeled similarity score as the ground truth and then check to see if the two policies were clustered together. If GPT labeled them as similar and they were clustered together, that is a true positive. If GPT labeled them as similar and they were not clustered together, that is a false negative. If GPT labeled them as dissimilar and they were clustered together, that is a false positive. Using these definitions, we choose an optimal K for each topic based on the F1 score (harmonic mean of precision and recall). That means we calculate 50 F1 scores 2 (number of models) x 25 (number of ks) for each topic. Figure A.13 graphs this process.

A.2 Human Validation

Figure A.13: F1 Score by Number of Clusters, for each Topic

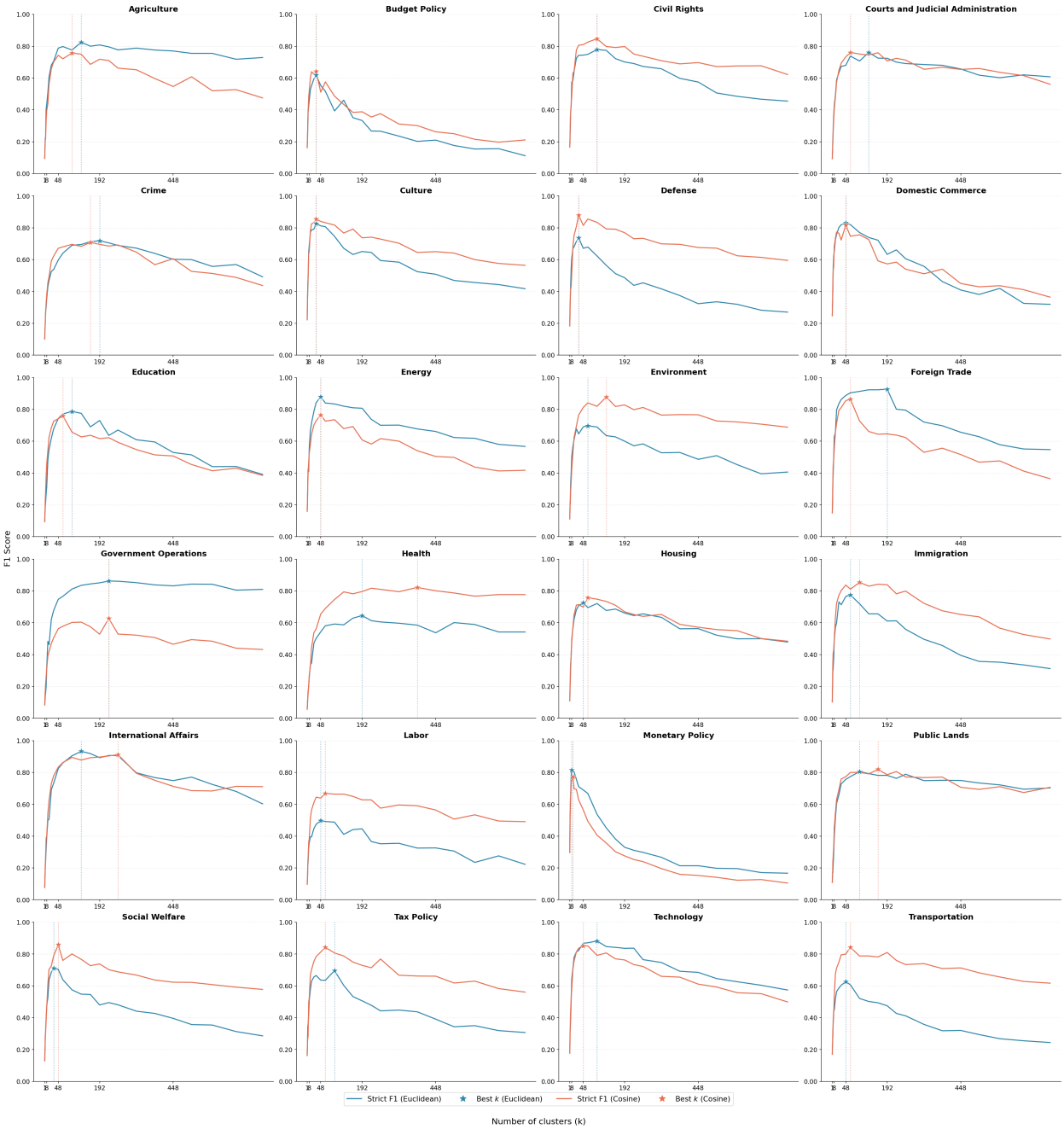
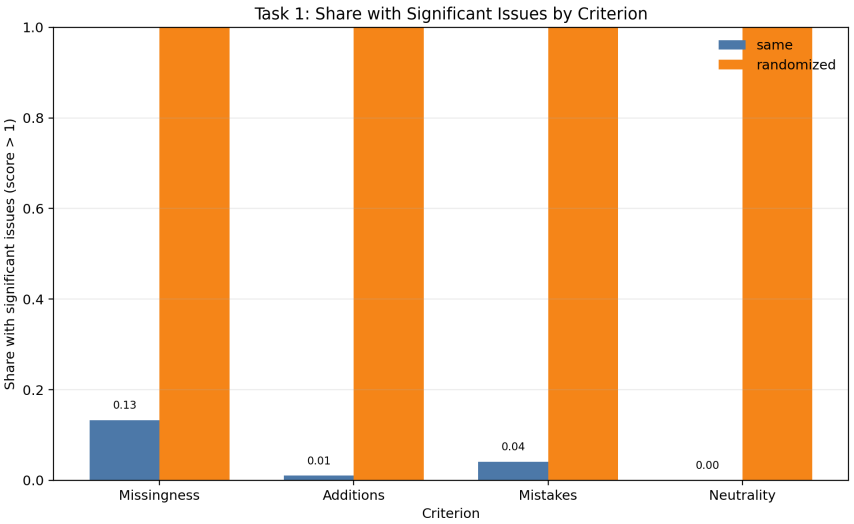


Table A.2: F1 score by policy topic

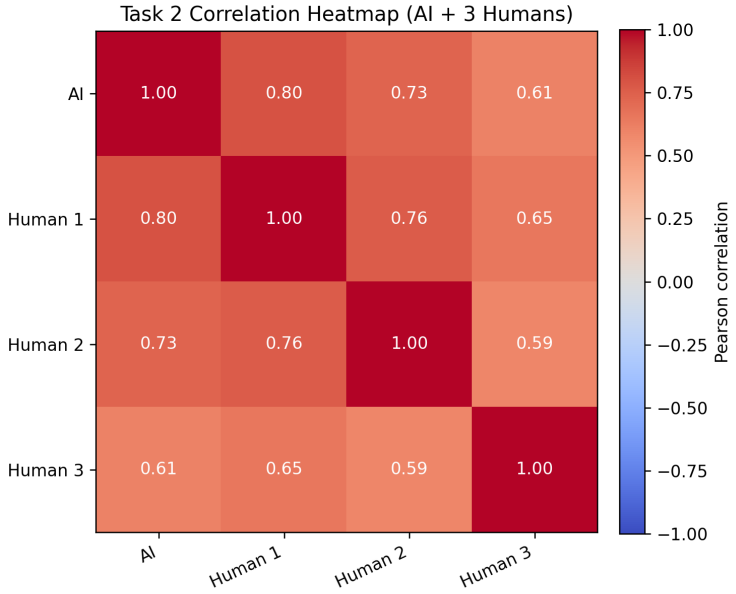
Topic	F1 Score
Agriculture	0.825
Budget Policy	0.642
Civil Rights	0.847
Courts and Judicial Administration	0.761
Crime	0.718
Culture	0.855
Defense	0.879
Domestic Commerce	0.831
Education	0.787
Energy	0.879
Environment	0.878
Foreign Trade	0.927
Government Operations	0.862
Health	0.822
Housing	0.758
Immigration	0.854
International Affairs	0.933
Labor	0.669
Monetary Policy	0.815
Public Lands	0.820
Social Welfare	0.859
Tax Policy	0.842
Technology	0.882
Transportation	0.842
Overall	0.821

Figure A.14: Share of significant issues identified by legal experts



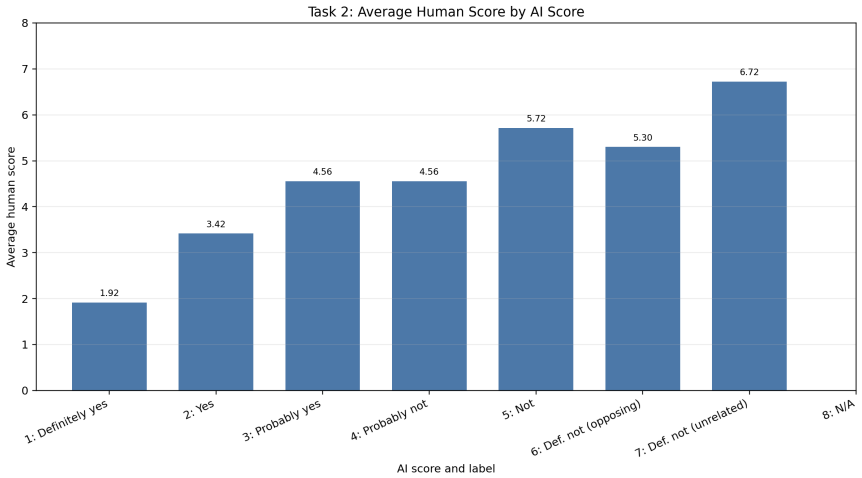
Notes: The figure shows, for each criterion, the share of summaries for which legal experts identified a significant issue. Missingness: the summary omits an important feature of the policy described in the statute. Additions: the summary includes policy features that are not contained in the statute. Mistakes: the summary incorrectly describes features of the policy in the statute. Neutrality: the summary introduces normative framing or motivations not present in the statute.

Figure A.15: Legal Expert-LLM Agreement Matrix



Notes: The figure shows the Pearson correlation in similarity scores assigned by the LLM and the three legal experts across randomly drawn pairs of statutes.

Figure A.16: Average Human Legal Score by LLM Score



Notes: The figure shows the average similarity score assigned by legal experts for randomly drawn pairs of statutes, by the level of similarity assigned by the LLM.

B Supplementary

B.1 Comparative Agenda Project

Main Topic	Subtopics
Macroeconomics	• **Interest Rates:** includes issues related to inflation, cost of living, prices, and interest rates; • **Unemployment Rate:** includes issues related to the unemployment rate and the impact of unemployment; • **Monetary Policy:** includes issues related to the monetary policy, central bank, and the treasury; • **National Budget:** includes issues related to public debt, budgeting, and efforts to reduce deficits; • **Tax Code:** includes issues related to tax policy, the impact of taxes, and tax enforcement; • **Industrial Policy:** includes issues related to manufacturing policy, industrial revitalization and growth; • **Price Control:** includes issues related to wage or price control and emergency price controls; • **Other:** includes issues related to other macroeconomics subtopics.
Civil Rights	• **Minority Discrimination:** includes issues related to minority, ethnic, and racial group discrimination; • **Gender Discrimination:** includes issues related to sex, gender, and sexual orientation discrimination; • **Age Discrimination:** includes issues related to age discrimination, including mandatory retirement age policies; • **Handicap Discrimination:** includes issues related to handicap and disease discrimination; • **Voting Rights:** includes issues related to voting rights, expanding or contracting the franchise, participation, and related issues; • **Freedom of Speech:** includes issues related to freedom of speech, religious freedoms, and other types of freedom of expression; • **Right to Privacy:** includes issues related to privacy rights, including privacy of records, access to government information, and abortion rights; • **Anti-Government:** includes issues related to anti-government activity groups, such as the communist party and local insurgency groups; • **Other:** includes issues related to other civil rights subtopics.

Main Topic	Subtopics
Health	• **Health Care Reform:** includes issues related to broad, comprehensive changes in the health care system; • **Insurance:** includes issues related to health insurance reform, regulation, availability, and cost; • **Drug Industry:** includes issues related to the regulation and promotion of pharmaceuticals, medical devices, and clinical labs; • **Medical Facilities:** includes issues related to facilities construction, regulation, and payments, including waitlists and ambulance services; • **Insurance Providers:** includes issues related to provider and insurer payments and regulation, including other types of benefits or multiple benefits; • **Medical Liability:** includes issues related to medical liability, malpractice issues, medical fraud and abuse, and unfair practices; • **Manpower:** includes issues related to the supply and quantity of labor in the health care industry, training, and licensing; • **Disease Prevention:** includes issues related to disease prevention, treatment, and health promotion, including specific diseases not covered in other subtopics; • **Infants and Children:** includes issues related to infants and children, including coverage and quality of care, health promotion, and school health programs; • **Mental Health:** includes issues related to mental health care and mental health disease; • **Long-term Care:** includes issues related to long-term care, home health care, the terminally ill, and rehabilitation services; • **Drug Coverage and Cost:** includes issues related to prescription drug coverage, programs to pay for prescription drugs, and policy to reduce the cost of prescription drugs; • **Tobacco Abuse:** includes issues related to tobacco abuse, treatment, education, and health effects; • **Drug and Alcohol Abuse:** includes issues related to alcohol and illegal drug abuse, treatment, education, and health effects; • **R&D:** includes issues related to health care research and development; • **Other:** includes issues related to other health care topics.
Agriculture	• **Trade:** includes issues related to the regulation and impact of agricultural foreign trade; • **Subsidies to Farmers:** includes issues related to government subsidies to farmers and ranchers, including agricultural disaster insurance; • **Food Inspection & Safety:** includes issues related to food inspection and safety, including seafood and labeling requirements; • **Food Marketing & Promotion:** includes issues related to the regulation of agricultural marketing and providing information to consumers regarding a healthy diet; • **Animal and Crop Disease:** includes issues related to animal and crop disease, pest control, pesticide regulation, and welfare for domesticated animals; • **Fisheries & Fishing:** includes issues related to fishing, commercial fishery regulation, and conservation; • **R&D:** includes issues related to agricultural research and development; • **Other:** includes issues related to other agricultural subtopics.

Main Topic	Subtopics
Labor	• **Worker Safety:** includes issues related to worker safety and protection and compensation for work-related injury and disease; • **Employment Training:** includes issues related to job training for adult workers, workforce development, and efforts to retrain displaced workers; • **Employee Benefits:** includes issues related to all employee benefits, pensions, and retirement accounts, including government-provided unemployment insurance; • **Labor Unions:** includes issues related to labor unions, collective bargaining, and employer-employee relations; • **Fair Labor Standards:** includes issues related to fair labor standards such as the minimum wage and overtime compensation, and labor law; • **Youth Employment:** includes issues related to youth employment, child labor, and job training for youths; • **Migrant and Seasonal:** includes issues related to migrant, guest, and seasonal workers; • **Other:** includes issues related to other labor policy.
Education	• **Higher:** includes issues related to higher education, student loans and education finance, and the regulation of colleges and universities; • **Elementary & Secondary:** includes issues related to elementary and primary schools, school reform, safety in schools, and efforts to generally improve educational standards and outcomes; • **Underprivileged:** includes issues related to education of underprivileged students, including adult literacy programs, bilingual education needs, and rural education initiatives; • **Vocational:** includes issues related to vocational education for children and adults and their impact; • **Special:** includes issues related to special education and education for the physically or mentally handicapped; • **Excellence:** includes issues related to education excellence, including efforts to increase the quality of specific areas, such as math, science, or foreign language skills; • **R&D:** includes issues related to research and development in education; • **Other:** includes issues related to other subtopics in education policy.

Main Topic	Subtopics
Environment	• **Drinking Water:** includes issues related to domestic drinking water safety, supply, pollution, and additives; • **Waste Disposal:** includes issues related to the disposal and treatment of wastewater, solid waste, and runoff; • **Hazardous Waste:** includes issues related to hazardous waste and toxic chemical regulation, treatment, and disposal; • **Air Pollution:** includes issues related to air pollution, climate change, and noise pollution; • **Recycling:** includes issues related to recycling, reuse, and resource conservation; • **Indoor Hazards:** includes issues related to indoor environmental hazards, indoor air contamination, and indoor hazardous substances; • **Species & Forest:** includes issues related to species and forest protection, endangered species, and regulation of laboratory or performance animals; • **Land and Water Conservation:** includes issues related to land and water resource conservation; • **R&D:** includes issues related to research and development in environmental technology; • **Other:** includes issues related to other environmental subtopics.
Energy Policy	• **Nuclear:** includes issues related to nuclear energy, safety and security, and disposal of nuclear waste; • **Electricity:** includes issues related to general electricity, hydropower, and regulation of electrical utilities; • **Natural Gas & Oil:** includes issues related to natural gas and oil, drilling, oil spills and flaring, oil and gas prices, shortages, and gasoline regulation; • **Coal:** includes issues related to coal production, use, trade, and regulation, including coal gasification and clean coal technologies; • **Alternative & Renewable:** includes issues related to alternative and renewable energy, biofuels, hydrogen, and geothermal power; • **Conservation:** includes issues related to energy conservation and energy efficiency, including vehicles, homes, commercial use, and government; • **R&D:** includes issues related to energy research and development; • **Other:** includes issues related to other energy subtopics.
Immigration	• **Immigration:** includes issues related to immigration, refugees, and citizenship.

Main Topic	Subtopics
Transportation	• **Mass:** includes issues related to mass transportation construction, regulation, safety, and availability; • **Highways:** includes issues related to public highway construction, maintenance, and safety; • **Air Travel:** includes issues related to air travel, regulation, and safety of aviation, airports, air traffic control, pilot training, and aviation technology; • **Railroad Travel:** includes issues related to railroads, rail travel, rail freight, and the development and deployment of new rail technologies; • **Maritime:** includes issues related to maritime transportation, including maritime freight and shipping, safety and security, and inland waterways and channels; • **Infrastructure:** includes issues related to infrastructure and public works, including employment initiatives; • **R&D:** includes issues related to transportation research and development; • **Other:** includes issues related to other transportation subtopics.
Law and Crime	• **Agencies:** includes issues related to all law enforcement agencies, including border customs and other specialized enforcement agencies and their appropriations; • **White Collar Crime:** includes issues related to white-collar crime, organized crime, counterfeiting and fraud, cyber-crime, and money laundering; • **Illegal Drugs:** includes issues related to illegal drug crime and enforcement, criminal penalties for drug crimes, including international efforts to combat drug trafficking; • **Court Administration:** includes issues related to court administration, judiciary appropriations, guidelines for bail, pre-release, fines, and legal representation; • **Prisons:** includes issues related to prisons and jails, parole systems, and appropriations; • **Juvenile Crime:** includes issues related to juvenile crime and justice, juvenile prisons and jails, and efforts to reduce juvenile crime and recidivism; • **Child Abuse:** includes issues related to child abuse, child pornography, sexual exploitation of children, and parental kidnapping; • **Family Issues:** includes issues related to family issues, domestic violence, child welfare, and family law; • **Criminal & Civil Code:** includes issues related to domestic criminal and civil codes, including crimes not mentioned in other subtopics; • **Crime Control:** includes issues related to the control, prevention, and impact of crime; • **Police:** includes issues related to police and other general domestic security responses to terrorism, such as special police; • **Other:** includes issues related to other law, crime, and family subtopics.

Main Topic	Subtopics
Social Welfare	• **Low-Income Assistance:** includes issues related to poverty assistance for low-income families, including food assistance programs, programs to assess or alleviate welfare dependency, and tax credits directed at low-income families; • **Elderly Assistance:** includes issues related to elderly issues and elderly assistance, including government pensions; • **Disabled Assistance:** includes issues related to aid for people with physical or mental disabilities; • **Volunteer Associations:** includes issues related to domestic volunteer associations, charities, and youth organizations; • **Child Care:** includes issues related to parental leave and child care; • **Other:** includes issues related to other social welfare policy subtopics.
Housing	• **Community Development:** includes issues related to housing and community development, neighborhood development, and national housing policy; • **Urban Development:** includes issues related to urban development and general urban issues; • **Rural Housing:** includes issues related to rural housing; • **Rural Development:** includes non-housing issues focused on economic, infrastructure, and other developments in non-urban areas; • **Low-Income Assistance:** includes issues related to housing for low-income individuals and families, including public housing projects and housing affordability programs; • **Veterans:** includes issues related to housing for military veterans and their families, including subsidies for veterans; • **Elderly:** includes issues related to housing for the elderly, including housing facilities for the handicapped elderly; • **Homeless:** includes issues related to housing for the homeless and efforts to reduce homelessness; • **R&D:** includes issues related to housing and community development research and development; • **Other:** includes issues related to other housing and community development subtopics.

Main Topic	Subtopics
Domestic Commerce	• **Banking:** includes issues related to the regulation of national banking systems and other non-bank financial institutions; • **Securities & Commodities:** includes issues related to the regulation and facilitation of securities and commodities trading, regulation of investments and related industries, and exchanges; • **Consumer Finance:** includes issues related to consumer finance, mortgages, credit cards, access to credit records, and consumer credit fraud; • **Insurance Regulation:** includes issues related to insurance regulation, fraud and abuse in the insurance industry, the financial health of the insurance industry, and insurance availability and cost; • **Bankruptcy:** includes issues related to personal, commercial, and municipal bankruptcies; • **Corporate Management:** includes issues related to corporate mergers, antitrust regulation, corporate accounting and governance, and corporate management; • **Small Businesses:** includes issues related to small businesses, including programs to promote and subsidize small businesses; • **Copyrights and Patents:** includes issues related to copyrights and patents, patent reform, and intellectual property; • **Disaster Relief:** includes issues related to domestic natural disaster relief, disaster or flood insurance, and natural disaster preparedness; • **Tourism:** includes issues related to tourism regulation, promotion, and impact; • **Consumer Safety:** includes issues related to consumer fraud and safety in domestic commerce; • **Sports Regulation:** includes issues related to the regulation and promotion of sports, gambling, and personal fitness; • **R&D:** includes issues related to domestic commerce research and development; • **Other:** includes issues related to other domestic commerce policy subtopics.

Main Topic	Subtopics
Defense	• **Alliances:** includes issues related to defense alliances and agreements, security assistance, and UN peacekeeping activities; • **Intelligence:** includes issues related to military intelligence, espionage, and covert operations; • **Readiness:** includes issues related to military readiness, coordination of armed services, air support and sealift capabilities, and national stockpiles of strategic materials; • **Nuclear Arms:** includes issues related to nuclear weapons, nuclear proliferation, and modernization of nuclear equipment; • **Military Aid:** includes issues related to military aid to other countries and the control of arms sales to other countries; • **Personnel Issues:** includes issues related to military manpower, military personnel and their dependents, military courts, and general veterans issues; • **Procurement:** includes issues related to military procurement, conversion of old equipment, and weapons systems evaluation; • **Installations & Land:** includes issues related to military installations, construction, and land transfers; • **Reserve Forces:** includes issues related to military reserves and reserve affairs; • **Hazardous Waste:** includes issues related to military nuclear and hazardous waste disposal and military environmental compliance; • **Civil:** includes issues related to domestic civil defense, national security responses to terrorism, and other issues related to homeland security; • **Civilian Personnel:** includes issues related to non-contractor civilian personnel, civilian employment in the defense industry, and military base closings; • **Contractors:** includes issues related to military contractors and contracting, oversight of military contractors, and fraud by military contractors; • **Foreign Operations:** includes issues related to direct war-related foreign military operations, prisoners of war, and collateral damage to civilian populations; • **Claims against Military:** includes issues related to claims against the military, settlements for military dependents, and compensation for civilians injured in military operations; • **R&D:** includes issues related to defense research and development; • **Other:** includes issues related to other defense policy subtopics.
Technology	• **Space:** includes issues related to government use of space, space resource exploitation agreements, and space exploration; • **Commercial Use of Space:** includes issues related to the regulation and promotion of commercial space use, commercial satellite technology, and related government efforts; • **Science Transfer:** includes issues related to science and technology transfer and international science cooperation; • **Telecommunications:** includes issues related to telecommunications infrastructure, internet access, and regulation; • **Other:** includes additional technology subtopics.

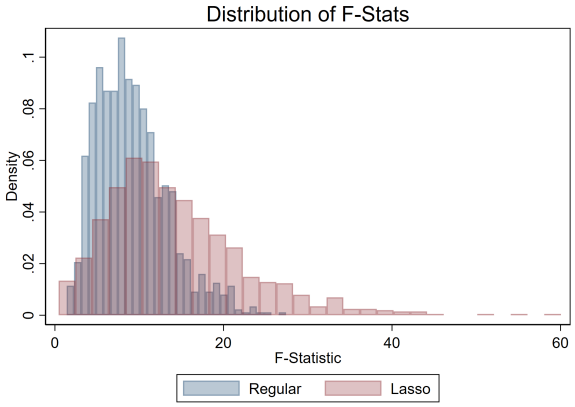
Main Topic	Subtopics
Foreign Trade	• **Trade Agreements:** includes issues related to trade negotiations, disputes, and agreements, including tax treaties; • **Exports:** includes issues related to export regulation, subsidies, promotion, and control; • **Private Investments:** includes issues related to international private business investment and corporate development; • **Competitiveness:** includes issues related to productivity and competitiveness of domestic businesses and balance of payments issues; • **Tariff & Imports:** includes issues related to tariffs and other barriers to imports, import regulation, and impact of imports on domestic industries; • **Exchange Rates:** includes issues related to exchange rates and related issues; • **Other:** includes issues related to other foreign trade policy subtopics.
International Affairs	• **Foreign Aid:** includes issues related to foreign aid not directly targeting international development; • **Resources Exploitation:** includes issues related to international resource exploitation and agreements, law of the sea, and international ocean conservation efforts; • **Developing Countries:** includes issues related specifically to developing countries; • **International Finance:** includes issues related to international finance and economic development, the World Bank, International Monetary Fund, regional development banks, sovereign debt, and implications for international lending institutions; • **Western Europe:** includes issues related to Western Europe and the European Union; • **Specific Country:** includes issues related specifically to a foreign country or region not codable using other codes, including political assessments and relations between countries; • **Human Rights:** includes issues related to human rights, human rights violations, treaties, conventions, UN reports, and crimes such as genocide or crimes against humanity; • **Organizations:** includes issues related to international organizations, NGOs, the United Nations, International Red Cross, UNESCO, International Olympic Committee, and the International Criminal Court; • **Terrorism:** includes issues related to international terrorism, hijacking, piracy, anti-terrorism efforts, and international legal mechanisms against terrorism; • **Diplomats:** includes issues related to diplomats, diplomacy, embassies, citizens abroad, foreign diplomats in the country, visas, and passports; • **Other:** includes issues related to other international affairs policy subtopics.

Main Topic	Subtopics
Government Operations	• **Intergovernmental Relations:** includes issues related to intergovernmental relations and local government issues; • **Bureaucracy:** includes issues related to general government efficiencies and bureaucratic oversight; • **Postal Service:** includes issues related to postal services, regulation of mail, and post-civil service; • **Employees:** includes issues related to civil employees, government pensions, and general civil service issues; • **Appointments:** includes issues related to nominations and appointments; • **Currency:** includes issues related to currency, national mints, medals, and commemorative coins; • **Procurement & Contractors:** includes issues related to government procurement, government contractors, contractor and procurement fraud, and procurement processes; • **Property Management:** includes issues related to government property management, construction, and regulation; • **Tax Administration:** includes issues related to tax administration, enforcement, and auditing for individuals and corporations; • **Scandals:** includes issues related to public scandals and impeachment; • **Branch Relations:** includes issues related to government branch relations, administrative issues, and constitutional reforms; • **Political Campaigns:** includes issues related to regulation of political campaigns, campaign finance, political advertising, and voter registration; • **Census & Statistics:** includes issues related to census and statistics collection by the government; • **Capital City:** includes issues related to the capital city; • **Claims against the Government:** includes issues related to claims against the government and compensation for victims of terrorist attacks; • **National Holidays:** includes issues related to national holidays and their observation; • **Other:** includes issues related to other government operations subtopics.
Public Lands	• **National Parks:** includes issues related to national parks, memorials, historic sites, and recreation, including the management and staffing of cultural sites; • **Indigenous Affairs:** includes issues related to indigenous affairs, indigenous lands, and assistance to indigenous people; • **Public Lands:** includes issues related to natural resources, public lands, and forest management, including forest fires and livestock grazing; • **Water Resources:** includes issues related to water resources, water resource development and civil works, flood control, and research; • **Dependencies & Territories:** includes issues related to territorial and dependency issues and devolution; • **Other:** includes issues related to other public lands policy subtopics.
Culture	• **Culture:** includes issues related to cultural policy and its impact on society.

C Empirical Analysis Appendix

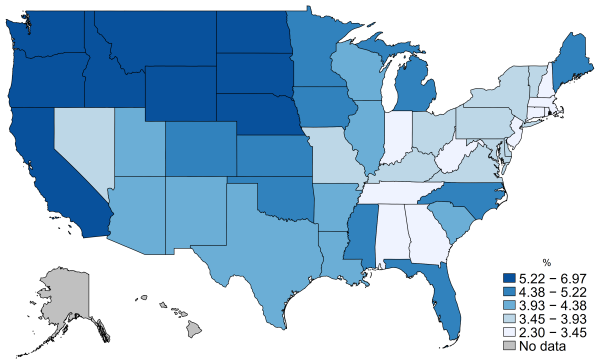
C.1 Additional Tables and Figures not shown in the main paper

Figure A.17: Distribution of first-stage F-statistics across clusters

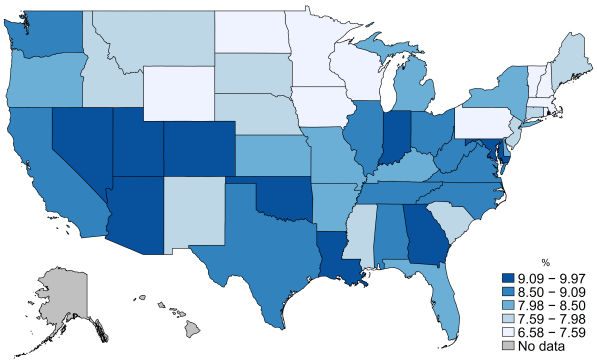


(a) Baseline specification

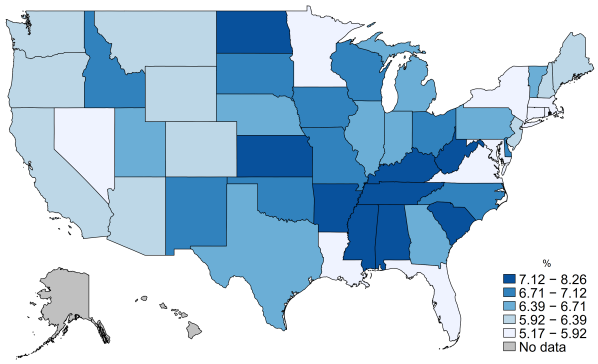
Figure A.18: State's topic-specific adoption share, calculated as the number of adoptions in a given topic divided by the state's total number of adoptions across all topics.



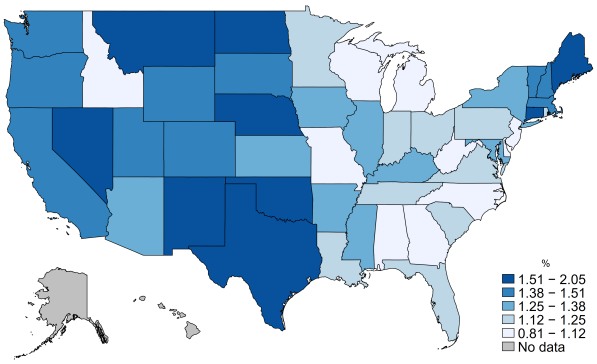
(a) Agriculture



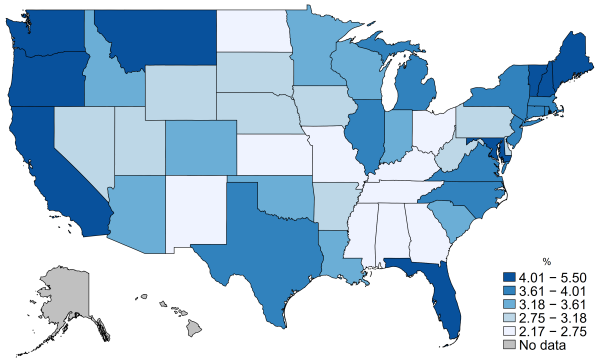
(b) Crime



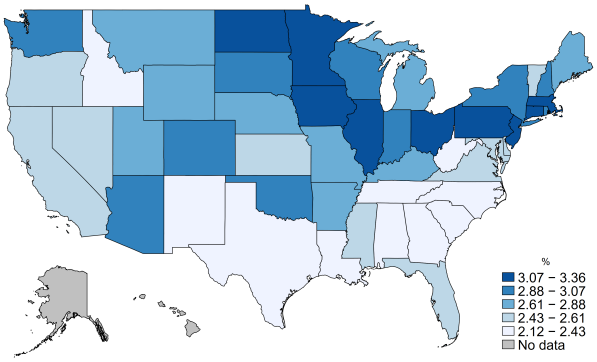
(c) Education



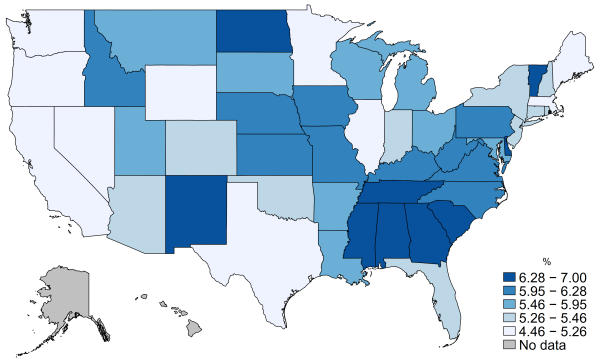
(d) Energy



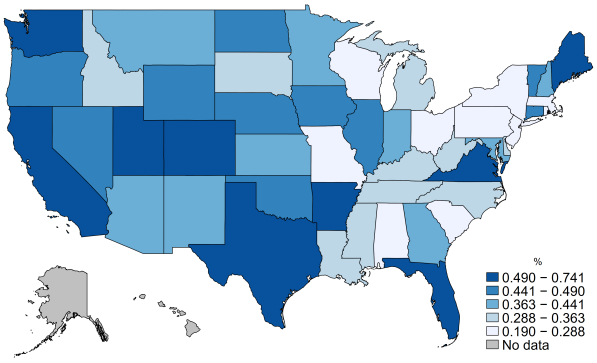
(e) Environment



(f) Social Welfare



(g) Tax Policy



(h) Technology

Table A.4: Propensity to spread laws by state ($\beta_{s'}$) for the pre vs. post 1950 periods

State	Pre 1950	Post 1950
AL	-0.0022	0.0017
	(-0.0029, -0.0014)	(0.0008, 0.0026)
AR	-0.0017	0.0009
	(-0.0025, -0.0009)	(-0.0001, 0.0018)
AZ	-0.0002	-0.0009
	(-0.0011, 0.0006)	(-0.0022, 0.0004)
CA	0.0020	0.0004
	(0.0013, 0.0027)	(-0.0005, 0.0012)
CO	0.0014	0.0016
	(0.0005, 0.0022)	(0.0005, 0.0027)
CT	0.0011	-0.0025
	(0.0004, 0.0019)	(-0.0033, -0.0016)
DE	0.0029	-0.0006
	(0.0020, 0.0038)	(-0.0016, 0.0003)
FL	0.0027	0.0003
	(0.0020, 0.0035)	(-0.0006, 0.0012)
Fed	0.0047	0.0016
	(0.0037, 0.0057)	(0.0010, 0.0028)
GA	-0.0004	0.0006
	(-0.0012, 0.0003)	(-0.0003, 0.0014)
IA	-0.0007	0.0002
	(-0.0015, 0.0002)	(-0.0007, 0.0012)
ID	-0.0025	0.0011
	(-0.0034, -0.0017)	(-0.0001, 0.0022)
IL	0.0005	0.0020
	(-0.0002, 0.0012)	(0.0011, 0.0030)
IN	0.0005	0.0003
	(-0.0004, 0.0013)	(-0.0006, 0.0012)
KS	0.0008	0.0020
	(0.0000, 0.0016)	(0.0011, 0.0029)
KY	-0.0005	-0.0005
	(-0.0015, 0.0004)	(-0.0015, 0.0005)
LA	-0.0007	-0.0023
	(-0.0015, 0.0001)	(-0.0033, -0.0014)
MA	-0.0009	0.0003
	(-0.0017, 0.0000)	(-0.0006, 0.0012)
MD	0.0015	0.0004
	(0.0007, 0.0022)	(-0.0004, 0.0012)
ME	0.0015	-0.0003
	(0.0008, 0.0023)	(-0.0012, 0.0006)
MI	-0.0012	0.0004
	(-0.0020, -0.0003)	(-0.0004, 0.0013)
MN	0.0019	0.0000
	(0.0011, 0.0027)	(-0.0008, 0.0008)
MO	-0.0004	0.0017
	(-0.0013, 0.0005)	(0.0007, 0.0027)
MS	-0.0028	-0.0005
	(-0.0036, -0.0020)	(-0.0014, 0.0005)
MT	0.0005	0.0002
	(-0.0004, 0.0013)	(-0.0009, 0.0013)
NC	0.0015	0.0015
	(0.0008, 0.0023)	(0.0007, 0.0023)
ND	-0.0002	0.0006
	(-0.0010, 0.0006)	(-0.0005, 0.0018)
NE	-0.0015	0.0016
	(-0.0024, -0.0006)	(0.0005, 0.0027)
NH	-0.0002	-0.0018
	(-0.0010, 0.0007)	(-0.0027, -0.0009)
NJ	-0.0003	-0.0009
	(-0.0011, 0.0006)	(-0.0017, -0.0002)
NM	-0.0009	-0.0007
	(-0.0018, -0.0000)	(-0.0018, 0.0005)
NV	-0.0005	-0.0010
	(-0.0012, 0.0003)	(-0.0020, 0.0000)
NY	0.0006	-0.0012
	(-0.0001, 0.0014)	(-0.0019, -0.0005)
OH	0.0019	-0.0014
	(0.0010, 0.0029)	(-0.0022, -0.0005)
OK	0.0008	0.0000
	(-0.0000, 0.0017)	(-0.0011, 0.0012)
OR	0.0022	0.0012
	(0.0014, 0.0029)	(0.0003, 0.0021)
PA	-0.0021	0.0002
	(-0.0030, -0.0013)	(-0.0006, 0.0010)
RI	0.0003	-0.0021
	(-0.0006, 0.0011)	(-0.0030, -0.0011)
SC	-0.0000	0.0001
	(-0.0008, 0.0008)	(-0.0008, 0.0009)
SD	-0.0017	-0.0010
	(-0.0025, -0.0008)	(-0.0021, 0.0002)
TN	0.0004	-0.0003
	(-0.0004, 0.0011)	(-0.0013, 0.0006)
TX	0.0013	0.0006
	(0.0005, 0.0021)	(-0.0004, 0.0015)
UT	-0.0016	0.0010
	(-0.0026, -0.0006)	(-0.0004, 0.0023)
VA	0.0005	0.0012
	(-0.0003, 0.0012)	(0.0004, 0.0020)
VT	0.0001	-0.0012
	(-0.0010, 0.0011)	(-0.0022, -0.0002)
WA	0.0010	0.0025
	(0.0001, 0.0019)	(0.0015, 0.0036)
WI	0.0016	0.0005
	(0.0007, 0.0025)	(-0.0003, 0.0013)
WV	-0.0029	-0.0015
	(-0.0039, -0.0020)	(-0.0026, -0.0005)
WY	-0.0027	-0.0010
	(-0.0037, -0.0017)	(-0.0023, 0.0002)