State Politics & Policy Quarterly (2021), **21**: 4, 430–450
doi:10.1017/spq.2021.17

CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

The Correlates of State Policy and the Structure of State Panel Data

Matt Grossmann[1], Marty P. Jordan[1] and Joshua McCrain[2]

[1]Institute for Public Policy and Social Research, Michigan State University, East Lansing, MI, USA
[2]University of Utah, Department of Political Science, Salt Lake City, UT, USA
Corresponding Author: Matt Grossmann. Email: matt@mattg.org

(Received 24 November 2020; Revised 05 February 2021; Accepted 26 February 2021)

Abstract

The American states offer a wealth of variation across time and space to understand the sources, dynamics, and consequences of public policy. As laboratories of socioeconomic and political differences, they enable both wide-scale assessments of change and studies of specific policy choices. To leverage this potential, we constructed and integrated a database of thousands of state-year variables for designing and executing social research: the Correlates of State Policy Project (CSPP). The database offers one-stop shopping for accurate and reliable data, allows researchers to assess the generalizability of the relationships they uncover, enables assessment of causal inferences, and connects state politics researchers to larger research communities. We demonstrate CSPP's use and breadth, as well as its limitations. Through an applied empirical approach familiar to the state politics literature, we show that researchers should remain attentive to regional variation in key variables and potential lack of within-state variation in independent and dependent variables of interest. By comparing commonly used model specifications, we demonstrate that results are highly sensitive to particular research design choices. Inferences drawn from state politics research largely depend on the nature of over time variation within and across states and the empirical leverage it may or may not provide.

Introduction

Nearly forty years ago, Jewell (1982) lamented that scholars had largely neglected the study of US state governments and politics. Since then, there has been a veritable expansion in the scale, scope, and quality of research on the American states. Scholars made significant theoretical and methodological achievements in part by relying on critical data contributions. Yet, despite these advancements, data are still typically gathered from disparate, individual sources, presenting challenges and increasing transaction costs for researchers. Aptly put by Carsey *et al.* (2008), "[t]he variance that makes analysis of state-level processes so attractive to scholars also makes data collection efforts at the state level difficult" (432). The study of politics and policy making in states has long been overdue for a central repository of key variables.

Here, we formally introduce and detail a free and publicly available database: the Correlates of State Policy Project (CSPP). This dataset includes 2,200 policy, political, and socioeconomic variables at the state-year level spanning the years 1900–2020. Many variables track state policies while others are possible antecedents or consequences of policy adoptions. CSPP provides academics and practitioners a “one-stop-shop” for accurate and reliable US state data, as well as opportunities to better understand the universe of available data, leverage panel data for improved causal inference, and engage the broader scientific community in state research. The breadth and depth of the data enable data exploration and myriad research designs, including single-state studies, cross-sectional designs, within-state panel approaches, or quasi-experimental designs. Moreover, the dataset, statistical package, and associated tools are also useful for practitioner and pedagogical purposes, allowing users to create maps, plot variables of interest, or visualize over-time trends.

We highlight CSPP’s scale, span, accessibility, and sustainability, and we summarize how users are already relying on the data for their research. We also demonstrate the utility of CSPP by portraying the relative stability of familiar state indicators within states, temporal changes for some variables by region, and the correlation of different variables within regions and across time. By showcasing a common application in state politics research—citizen ideology’s influence on policy outputs—we underscore how CSPP facilitates flexibility in research design choices and how those decisions matter for inference. We find that the conclusions drawn from typical state panel analyses largely depend on the nature of temporal and time-series variation available, which only sometimes provides adequate leverage to understand social and political change.

Research opportunities capitalizing on the American states abound. Our hope is that CSPP’s integration of the vast public goods made available in the field will reduce start-up costs, time, and energy for scholars, practitioners, and educators. This resource will allow researchers to improve research design and further our understanding of how institutions and behavior influence policy and political outcomes within and across the US states over time.

Solving Problems in State Politics Research

The CSPP is designed to take advantage of the tremendous opportunity afforded by research on the American states. State politics are important arenas for democratic contestation; state governments are key actors in public policy; and states are principal laboratories for assessing socioeconomic outcomes of political choices. The states offer leverage across both structures and polities, enabling assessments of variation in behavior and institutions in American governance. They are also connected within a federal system, allowing meaningful analyses of interdependence. The long closely-tracked history of roughly similar cases undergoing national, regional, and local changes over time has made states a premier example of how panel data can be used for description and identification in social science.

To help state research expand and improve, CSPP solves four common problems that otherwise limit achievements in the subfield. First, as a repository for myriad data, it enables ease of use and updating. The vastness of variables on states—from numerous policies and outcomes to related social, political, and economic data—can seem overwhelming rather than advantageous to researchers, instructors, and practitioners. This is particularly the case if the data span different time periods and units

https://doi.org/10.1017/spq.2021.17 Published online by Cambridge University Press

of observation, or remain hidden in replication files (or worse, on unconnected personal archives) associated with individual projects. By linking these data, mitigating mistakes in merging, aggregating data by state-year, and updating previously collected data when available, CSPP puts these variables in a unified structure and in commonly used data formats.

A persistent problem with state policy and politics research has been the lack of data availability and access, with researchers frequently reduplicating efforts. The field lags behind some other subfields that have central repositories (e.g., Policy Agendas Project and Correlates of War Project). Fortunately, several scholars have led by example, making their data publicly available (e.g., Walker 1969; Gray and Lowery 1988; Erikson *et al.* 1993; Berry, Fording, and Hanson 1998; Squire 2007, Squire 2008; Brace and Hall 2009). More recent contributions from generous scholars have exponentially added to the scope of available information (e.g., Boehmke and Skinner 2012; Boehmke *et al.* 2020; Caughey and Warshaw 2016, Caughey and Warshaw 2018; Carsey *et al.* 2008; Klarner 2013a; Shor and McCarty 2011; Sorens, Muedini, and Ruger 2008). CSPP would not be conceivable without their important data contributions (which are promoted on the project website). Beyond individual decisions to share data, recent trends such as the open source revolution and expectations of reproducibility have further increased availability. Nongovernmental organizations and universities also have published state-level data (e.g., Open States, Stateminder, National Conference of State Legislators, Ballotpedia, and Council of State Governments). Despite these positive developments, researchers still typically have to cobble together state politics data from scattered, albeit more accessible, sources. That is a central problem we aim to solve.

Social science has also come under fire for cherry picking data to demonstrate relationships. The second large impact of CSPP is to allow researchers not only to select the most relevant policies and outcomes for their analyses, but also to understand the universe of variables from which they are selecting. It is often appropriate to home in on the most closely associated predictors or outcomes of particular policies. CSPP allows the scholarly community to investigate whether those relationships are specific to data choices or part of a broader pattern. For example, a researcher may find that governor partisanship correlates with the relative tax rates paid by Black and White families; CSPP enables a researcher to inspect whether that relationship is part of a pattern across tax policies or racial disparities or simply a singular association, and to assess the durability of that relationship.

Third, the increased emphasis on causal identification in social science has popularized new techniques that require panel data that meets particular assumptions. CSPP allows researchers to assess change over time and variation across states to determine whether they can productively use difference-in-differences, regression discontinuity, fixed effects, quantile regression, or other designs. Researchers may want to study the effects of closely divided legislatures or a particular policy change, for example, but discover that variation is too limited or too closely associated with a broader regional shift. CSPP facilitates searches for instrumental variables, natural experiments, or changes separated from regional or national political dynamics. The comprehensive nature of CSPP's data facilitates researchers' ability to match research design to theory, diagnose problems with a given design, and test the robustness of inferences to specification choices.

Finally, the state politics literature remains segregated from the wider interdisciplinary community interested in variation across the United States. In particular,

https://doi.org/10.1017/spq.2021.17 Published online by Cambridge University Press

scholars interested in outcomes such as income inequality, public health, educational attainment, and environmental degradation may also be concerned with state-level policies and contextual factors, but are sometimes unaware of the standard variables political scientists use to explain variation. In some cases (e.g., education and health), these research communities are larger than the state politics subfield and may connect political scientists with important outcomes of interest in existing well-financed research efforts. CSPP makes it easier for interdisciplinary researchers to find the data subset they need for topic-specific research, increasing awareness of the importance of political variables in socioeconomic trends. Indeed, we later show that dozens of other criminal justice, education, public health, sociology, and public administration scholars have already taken advantage of CSPP.

The CSPP Database

The CSPP database is able to address those common problems for state researchers because it offers four major advantages: its (1) scale, (2) span, (3) accessibility, and (4) sustainability. First, CSPP is remarkable in its scale. It includes 2,200 variables from across the 50 US states and the District of Columbia pooled from multiple reputable academic, governmental, and nongovernmental sources.[1]

Table 1 displays the variable categories, number of variables for each category, and example variables included in the database. There are many variables covering economics, government institutions, and health, but fewer variables covering policy areas like social welfare and labor. Although most variables are directly related to policy, many of the variables in each category could also be broader antecedents or consequences of individual policy choices (e.g., institutional and electoral variables, policy liberalism, interest group density, or demographics). The compendium also allows the state politics research community to see where past researchers have focused (e.g., fiscal and drug policy) and where there may still be holes in our collective understanding (e.g., labor and transportation).

Beyond these policy, political, socioeconomic, and demographic variables, the CSPP also encompasses two additional data resources. Users can access a dataset containing nearly 7,800 ballot measures attempted across US states from 1902 – 2016. Compiled by the National Conference of State Legislatures (NCSL) (2016), the dataset includes the different kinds of measures pursued (i.e., legislative referendum, initiative, popular referendum, and other), election type (i.e., general, primary, and special), passage rates, and topic areas. By linking the datasets, users can also assess how ballot measures and outcomes are related to other state differences. Researchers

[1] Where possible, we rely on the same variable names from primary and secondary sources. We opt for this naming convention strategy for consistency's sake, in case researchers are already familiar with these variable names. If two variables are assigned the same name by different sources, we modify one of the variable names slightly in our database. In addition to copying variable names, we also carry over the variable descriptions and coding notes from the original authorities. We prominently advertise the importance of these notes for users of our data and associated statistical package. Some variables are missing observations for the District of Columbia and occasionally for a few other states (e.g., partisan legislative variables for Nebraska). Federal territories are omitted from the datasets. We do not independently audit data, though we do attend to differences in definitions and missing data, and make note of differences in time period (i.e., fiscal year vs. calendar year used for a variable). We also respond to feedback about errors or oddities, often adding notes for improved understanding.

https://doi.org/10.1017/spq.2021.17 Published online by Cambridge University Press

Table 1. Variables by category

Category	Total variables	Examples
Criminal justice	180	Death Penalty, Crime Rate, Murder Rate
Demographics	41	Population, Refugees, Immigration Laws
Drug-alcohol	122	Medical Marijuana, Smoking Ban, Lottery
Economic-fiscal	410	Gini Coefficient, State Expenditures, State Tax Revenue
Education	77	Charter Schools, Universal Pre-K, Vouchers
Elections	186	Party Control, Expenditure Limits, Campaign Finance Stringency
Environment	54	Solar Tax Credit, Renewable Portfolio Standards, Greenhouse Gas Cap
Government	254	Term Limits, Legislative Professionalism, House Partisanship
Gun control	57	Concealed Carry, Assault Weapons Ban, Permit Costs
Healthcare	219	Health Spending, Medicare Enrollees, Uninsured Population
Labor	46	Minimum Wage, Right to Work Laws, Workers' Compensation
Misc. regulation	190	Occupational Licensing Requirements, Fireworks Legality, Rent Control Prohibition
Policy-ideology	71	Policy Liberalism, Policy Innovativeness, Policy Mood
Rights	105	Abortion Coverage Waivers, Civil Unions and Gay Marriage, Employment Discrimination Laws
Transportation	45	Seat Belt Laws, Auto Insurance Required, Helmet Laws
Welfare	24	CHIP Eligibility, TANF Payments, Medicaid Eligibility

can likewise explore or incorporate relevant state network data in their analyses. Assembled by Olson (2019), the state network data consists of state-to-state relational variables including measures of shared borders, trade and travel between states, and similarities along political, socioeconomic, or demographic dimensions. Researchers can assess the network linkages across states (not only state borders, but also many other social ties) to the policy outcomes they study in the CSPP state attribute data.

Most other large-scale state politics datasets are narrower in focus by design. For example, some datasets (e.g., Boehmke *et al.* 2020) provide binary variables for policy adoptions but lack variables capturing state institutional features. Other datasets (e.g., Carsey *et al.* 2008; Klarner 2013d) cover legislative elections or state partisan control, but no variables measuring elite and mass public attitudes. CSPP builds on these datasets by merging them through common unique identifiers (while providing additional IDs such as Federal Information Processing Standards [FIPS] codes) and including an array of additional covariates and policies. We rely on hundreds of data collectors and aggregators, but the website highlights the five largest contributors who were each responsible for providing more than 50 variables. In constructing the dataset, we made a concerted effort to compile both familiar and obscure state policy and politics variables across diverse categories from multiple reputable sources. The resulting dataset and associated software is a comprehensive if not exhaustive outlet for data originators and researchers seeking information on the US states.

The second advantage of CSPP is its time coverage. The database extends from 1900 to 2020. Although, only a few variables extend as far back as 1900 (e.g., total state population), many of the variables span the last four most recent decades. Figure 1 reports the number of variables available by decade. Coverage is at its height in the 21st century, with the fewest variables available in 1900 and monotonically increasing coverage in each decade of the 20th century and then a spike in the 2000s. Analyses using CSPP can take advantage of tremendous temporal variation, but only if the variables needed are available for the full time series. We have the most evidence

https://doi.org/10.1017/spq.2021.17 Published online by Cambridge University Press

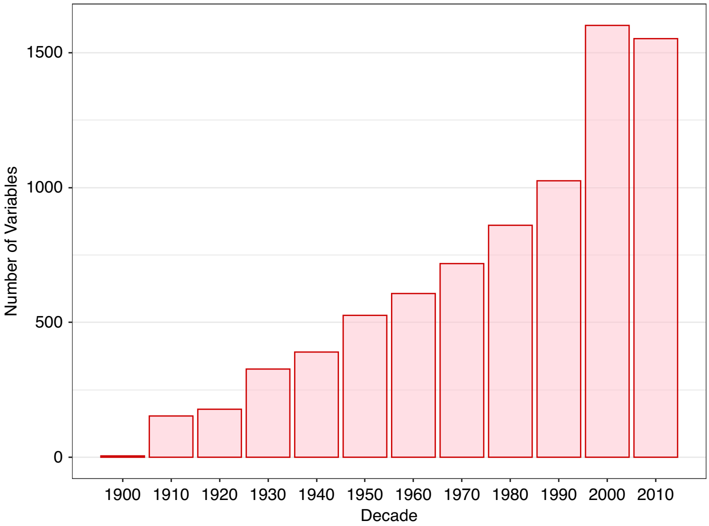


Figure 1. Variable coverage by decade.
Note: The *x*-axis consists of decades covered by CSPP data. The *y*-axis is the raw number of variables with values per decade.

Table 2. Variables by decade

Decade	Average number of states
1900	28
1910	49
1920	48
1930	46
1940	47
1950	47
1960	46
1970	44
1980	42
1990	43
2000	45
2010	44

Note: This table displays the coverage of the CSPP data by decade. The second column is calculated as the average number of states, per decade, that have full coverage in the variables.

about recent changes in the states, which means our analyses might not generalize farther back in time. We view this as partially a product of our data compilation but also indicative of data availability in the field. Many contemporary multivariate panel studies may be applicable only in a narrow time frame.

Table 2 reports the average number of states available for each variable in each decade. Again, we see widespread coverage in the most recent decades, with nearly all states covered for many variables. Still, several of the variables have a few missing subunits, such as Nebraska, Alaska, Hawaii, and the District of Columbia. Fortunately, limited state coverage is less of a problem for the field and the dataset, though our studies' findings may not always generalize to all fifty states.

https://doi.org/10.1017/spq.2021.17 Published online by Cambridge University Press

CSPP's third asset is its accessibility. Variables are coded at the state-year unit of observation (with notes accounting for aggregation and definition differences).[2] Formatting across the variables is standard, making data management and manipulation easier. Moreover, the database is available in multiple formats, including as a Microsoft Excel file with separate content area sheets, a comma-separated values (CSV) file containing all the variables, and an R package (CSPP; Lucas and McCrain 2020). Accompanying the package is a Shiny App web application that allows users to subset, explore, and download CSPP variables and corresponding citations. The R package and associated app also allow users to create visualizations, including maps, time trends, correlation matrices, and network diagrams.[3] CSPP also comes with a searchable, detailed codebook containing complete variable names, time spans, descriptions, original sources, and notes. The codebook and package allow for easy perusal of variables and original sources.

CSPP's fourth benefit is the system in place to ensure the database is maintained, updated, and expanded. Both the Institute for Public Policy and Social Research (IPPSR) and Michigan State University (MSU) have committed resources to compensate undergraduate and graduate students to help ensure appropriate data documentation, update existing variables, add new variables, correct errors, and release new versions of the database in a timely manner. In fact, hundreds of additional variables are planned for future releases of the database. Related, CSPP aims to reward the data producers by encouraging users to cite the original source, as well as tracking and publishing statistics on use so that originators can include these metrics in their citation counts. The R package and web app also allow users to easily compile citations for all of the original variables they utilize.

To be sure, such an enterprise is not without risk. Pooling variables from multiple sources can yield missing data or inaccuracies. These data gaps or inconsistencies may result from missing observations in the original source, coding errors made by the primary source, or data transfer mistakes on our end. For instance, if observations are missing for a particular jurisdiction or year in the original source, these observations are also missing in our database. Furthermore, if mistakes were made in creating or coding data by the primary sources, these errors will also be reproduced in our database. Although we have made considerable effort to ensure data quality and integrity during transfers and aggregation, errors are still possible. We encourage users to check the accuracy of the data, relying on the primary sources. We also appeal to researchers to inform us if they uncover any errors. As we are made aware of data inaccuracies, our team corrects the inconsistencies, promptly updates the database, and informs users of the changes via errata in future release versions. The dataset also allows comparison of multiple measures of the same concept, which we often keep for

[2] Some economic and fiscal variables are averaged across state fiscal quarters (Klarner 2013c). Other variables, such as campaign contributions from industries, are aggregated to the state-year level (as identified in the codebook). The website and statistical package draw attention to the variable notes and differences across year definitions. The first six variables are for identification: a year variable, state abbreviation, state number based on alphabetical assortment, state name, state FIPS code, and the Inter-university Consortium for Political and Social Research (ICPSR) state code.

[3] For users with limited data visualization experience, the host site http://ippsr.msu.edu/public-policy/correlates-state-policy also has embedded Google GeoChart and Line Chart tools to generate graphics or visually explore key variables of interest across states and time.

https://doi.org/10.1017/spq.2021.17 Published online by Cambridge University Press

potential robustness checks as inconsistencies are usually due to subtle differences in coding choices.

Current Use of CSPP

Demonstrating the utility of CSPP and how it is already helping solve problems in US state research, CSPP has been cited by 100 papers, articles, and books (after consolidating near duplicates). Nearly all citations were from the last three years and many were not yet published (21 were from dissertations or theses), suggesting that usage is just beginning to take off. We analyzed these citations to determine how CSPP is currently being used by researchers. Most scholars relied on the CSPP for specific variables (90% of citations), while others made general reference to the database. Most analyses using CSPP made causal inference claims (90% of citations) while others used the variables for descriptive work (6%) or general reviews (3%). Despite attempting to make causal inferences, however, most researchers used a standard panel or cross-sectional research design, with only 47% employing fixed-effects, and only a few others using regression discontinuity or other quasi-experimental approaches.

Although most users have yet to fully leverage CSPP for causal research, many are capitalizing on the breadth and span of the database. Exploring regional variation was a common area of investigation (81% of projects conducted some kind of regional analysis), though it was rarely the main focus of research. Similarly, partisan differences and polarization were common topics (probed in 40% of projects), but they were the focus of only 12 projects. The modal project examined a particular policy area, using panel data to investigate public policy choices or effects. Overall, 30 different policy issue topics were covered by researchers (spanning nearly every major issue area). Slightly more studies use public policy as an independent variable (54%) than a dependent variable (49%). The total exceeds 100% because some studies used policy for both (though others used it for neither). Many of the studies utilize CSPP variables for controls. While some of the researchers' projects extended decades back in time, most investigated contemporary periods, with the average analysis beginning in 1980 and ending in 2013. Underscoring CSPP's inroads in other disciplines, the database has been used by researchers representing diverse subfields. Although most of the authors citing CSPP are political scientists (63%), sociologists (16%), economists (18%), psychologists, criminal justice scholars, public administration researchers, environmental scientists, and many others have also used CSPP.

To see how data compilation enables multiple analyses, we want to highlight two dissimilar uses of CSPP to study the same topic: the effects of state partisanship on socioeconomic outcomes. Grumbach (2018) relies in part on CSPP to examine the polarization of 16 different policy areas across the US states since the 1970s. He finds that the party that controls state government is an increasingly strong predictor of policy outputs in many of the issue areas. Grumbach then shows how that partisan polarization leads to divergent outcomes tied to those policy differences, including incarceration rates and health insurance access. Grumbach's use of CSPP showcases how researchers can pinpoint the effects of consequential policies driven by state politics, revealing significant externalities for states.

Dynes and Holbein (2020) take a broader approach to state outcomes. They use CSPP data to employ both difference-in-differences and regression discontinuity designs to explore how state government party control influences overall socioeconomic outcomes. Rather than select outcomes most likely to be a result of particular party-linked policies, Dynes and Holbein examine dozens of outcomes to assess partisan differences in fulfilling common policy objectives. They find that Democratic- and Republican-controlled governments perform equally well across these objective output measures. Their approach tempers any generalization of party effects on broad state outcomes. Both articles offer important insights: some outcomes are associated with particular policies enacted by partisan governments (as Grumbach finds), but the broader trajectory of states is not dependent on their partisanship (as Dynes and Holbein show).

Regional and Temporal Variation

CSPP's data and tools offer opportunities for researchers to find relevant state variables to answer theory-driven questions employing multiple research designs. But users should also be aware of the dual complexities of using US state panel data. On the one hand, many state attributes remain relatively stable over time. On the other, the nation as a whole has experienced some pronounced regional differences and temporal change. Regional analysis is very common in studies citing CSPP and important in understanding the often-limited sources of variation in state politics. Given that CSPP-enabled studies thus far often employ panel data to assess causal claims (sometimes without quasi-experimental methods), it is critical to understand these patterns and the challenges they produce.

To illustrate this common structure of state panel data, we present the distributional characteristics of four commonly used indices in state politics research: policy mood (where higher scores indicate a more liberal citizenry; Enns and Koch 2013), income inequality (a Gini coefficient, where higher measures indicate greater income concentration; Frank 2014), policy liberalism (where higher scores point to more liberal policy adoptions; Caughey and Warshaw 2016), and union density within a state (measuring the proportion of non-agricultural workforce represented by a union; Kelly and Witko 2014). For these illustrations, we categorize states into the four familiar regions: the Northeast (Maine to Pennsylvania), the Midwest (Ohio to Kansas to the Dakotas), the South (states of the confederacy plus Delaware, the District of Columbia, Kentucky, Maryland, Oklahoma, and West Virginia), and the West (the Mountain and Pacific time zones). Other state divisions show similar effects.

Figure 2 displays the density plots for these four variables standardized across each geographic region from 1990 to 2010. For policy liberalism, the results indicate (unsurprisingly) that the Northeast is the most liberal, while the South is the most conservative. There is considerable variation in policy liberalism in the West and Midwest. Policy mood largely mirrors policy liberalism, with the Northeast displaying more progressive citizens than the other regions, but with little variation across the other regions. Union density fluctuates considerably in all regions besides the South, where it is much lower. Lastly, income inequality shows greater variation at the high end in each region, with limited average regional differences.

Further demonstrating regional variation for key attributes, Figure 3 displays the correlations within regions for several commonly used variables.[4] To the previous

[4] The `corr_plot` function in the R package creates this style of plot.

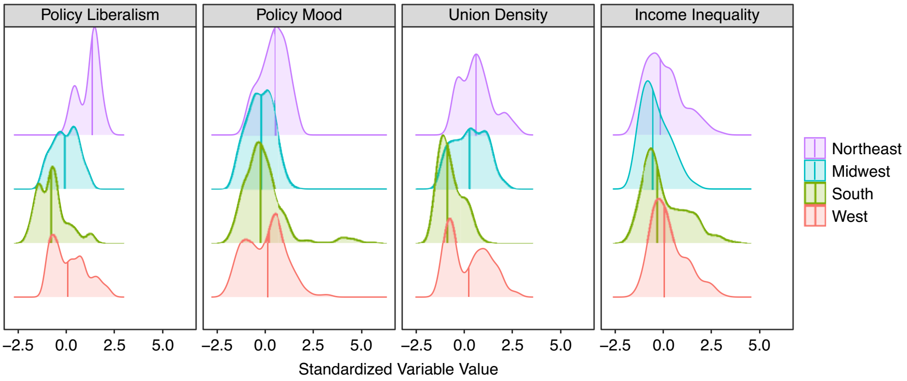


Figure 2. Densities of variables by region.
Note: Each panel is a variable in CSPP data spanning from 1990 to 2010 with sources discussed above. The variables were standardized prior to plotting. The densities of each variable are plotted disaggregated by region, starting from the top: Northeast, Midwest, South, and West.

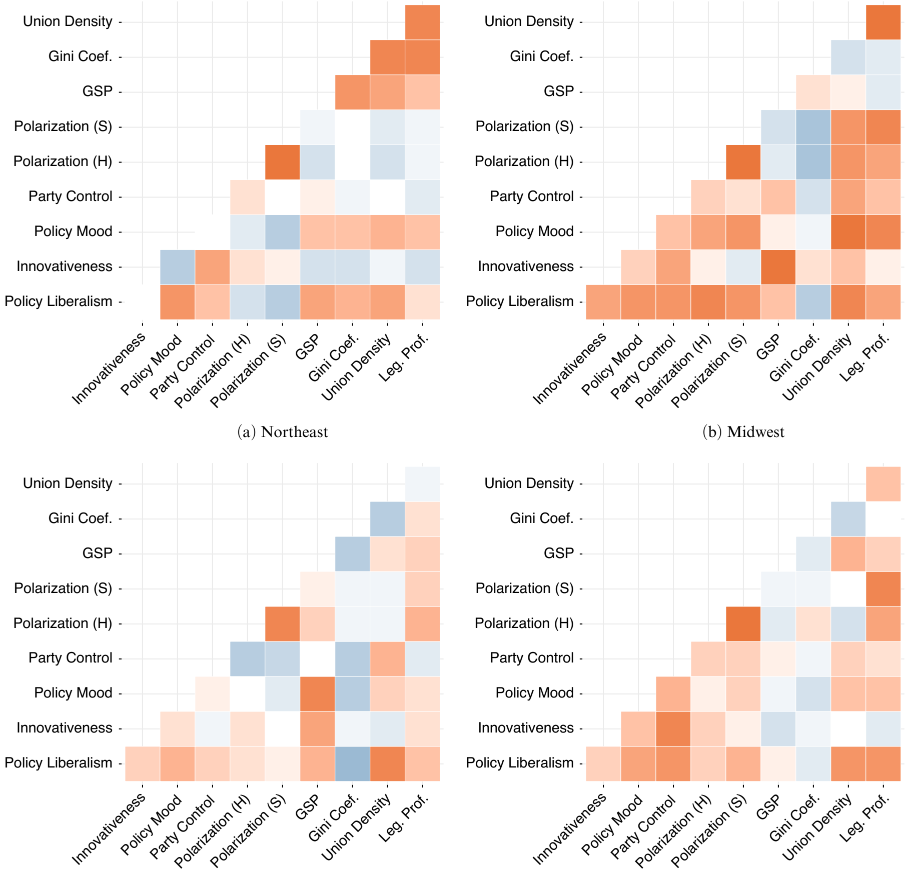
four variables, we add gross state product per capita (US Department of Commerce Bureau of Economic Analysis, 2012), lower (House) and upper (Senate) chamber polarization (Shor and McCarty 2011), party control (where 0 is unified Republican control, 0.5 is neither party full control, and 1 is unified Democratic control; Klarner 2013b), policy innovativeness (indicating a state's willingness to adopt new policies sooner than other states; Boehmke and Skinner 2012), and legislative professionalism (the first dimension scaled from legislator salary and expenditures and session length; Bowen and Greene 2014).

Most analyses look for associations among these variables nationwide, but that could present a misleading picture if there is regional heterogeneity. Policy liberalism, for example, is negatively correlated with income inequality (Gini coefficient) in all regions except the Northeast. Democratic party control is negatively associated with polarization in the South but positively associated with it in other regions. Similarly, union density has a negative correlation with legislative professionalism in the South, but the inverse is true in the other regions. Meanwhile, greater union density in the Midwest is associated with higher polarization, but the opposite holds for other zones. These regional trends reinforce the need to account for potential regional heterogeneity in our investigations.

Most CSPP-enabled panel models also assume that estimated relationships are constant over time. Empirically, however, there is variation in the associations among variables. Figure 4 showcases this using the same variables presented in Figure 3, displaying the changes in these correlations between 1990 and 2010.[5] The results indicate that many of the relationships with inequality are decreasing over time, whereas the associations with party control are increasing over these decades. Likewise, policy innovativeness' relationship to many of these variables has also

[5] The between-variable correlations for 1990–2000 and 2000–2010 are in Supplementary Appendix Figure A1.

https://doi.org/10.1017/spq.2021.17 Published online by Cambridge University Press



(a) Northeast

(b) Midwest

(c) South

(d) West

Figure 3. Variable correlation by region

Note: This figure shows heatmaps of the bivariate correlations of variables of interest, separated by region. Darker colors indicate stronger correlations. Orange shading indicates a positive correlation, while blue shading indicates a negative correlation.

waned during this period. Importantly, what is true in one era may not be true in the next.

Equally important, panel-based statistical analyses require attention to serial correlation and time trends. Figure 5 illustrates temporal differences by region for the original four variables (i.e., policy liberalism, policy mood, union density, and income inequality).[6] Policy liberalism shows steady change by region, with the Northeast trending leftward, the Midwest and South trending a bit rightward, and little change in the West. Policy mood shows notable volatility, with a national decline followed by an increase with regional dispersion. Union density is declining across

[6] In Supplementary Appendix Figures A2 and A3, we show that these regional differences are also evident when plotting state-level coefficients of each variable in equations of the form of Equation (1).

https://doi.org/10.1017/spq.2021.17 Published online by Cambridge University Press

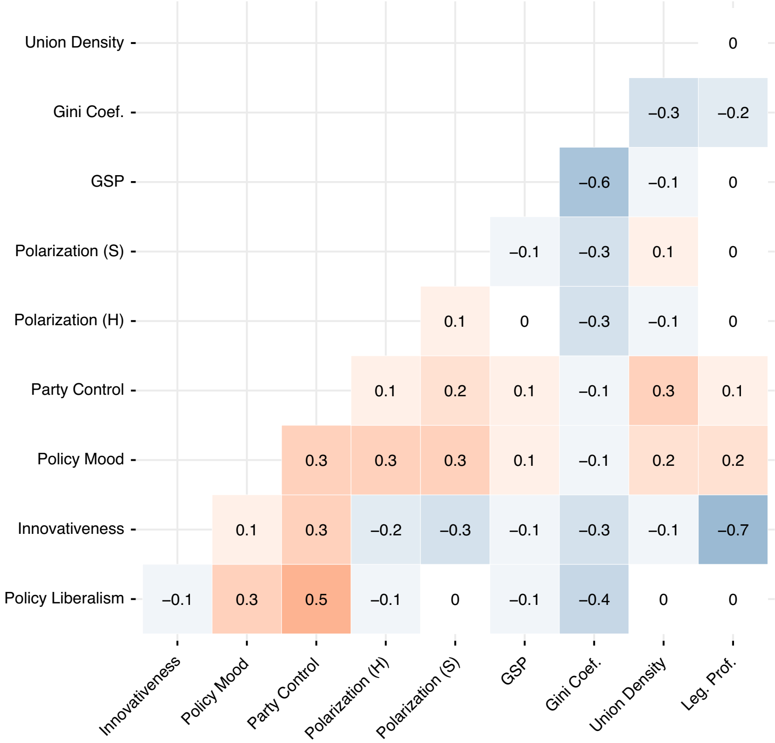
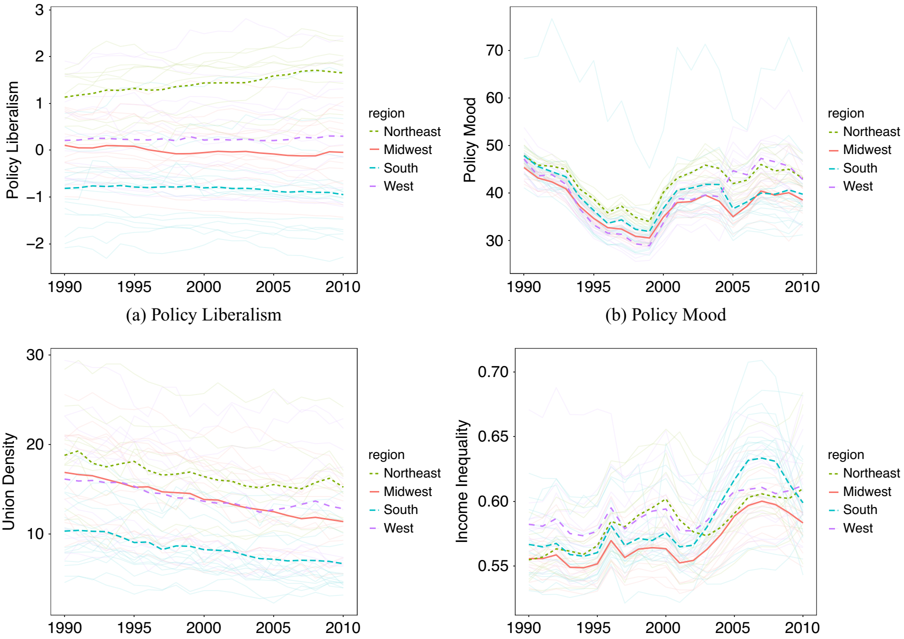


Figure 4. Difference in correlations across decades.
Note: This figure plots the differences in bivariate correlations between 1990–2000 and 2000–2010. The raw change in the correlation is displayed within each cell. For instance, if the correlation between two variables was 0.8 in 1990–2000 and then 0.3 between 2000 and 2010, the value in that cell would take on −0.5. The decade-level correlations are presented in the Supplementary Appendix. Darker colors indicate larger correlations. Orange shading indicates a positive correlation, while blue shading indicates a negative correlation.

regions, with regional differences between the Northeast and South. Income inequality trends are mostly national, with a marked uptick after 2000 and little regional variation. Relating any two of these trends in panel data would need to account for whether the variation was across states or time.

An Application to Public Opinion and Policy Liberalism

To further illustrate the opportunities and challenges of using US state panel data with different regional and temporal patterns, we address a common application: the potential relationship between public opinion and public policy in the American laboratories (see Berry, Fording, and Hanson 1998; Berry *et al.* 2010; Caughey and Warshaw 2018, Caughey and Warshaw 2016; Enns and Koch 2013; Wright Jr., Erikson, and McIver 1987). States with more liberal electorates, such as California,



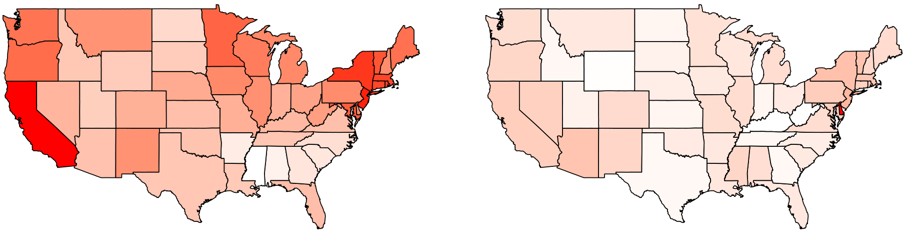
(a) Policy Liberalism

(b) Policy Mood

(c) Union Density

(d) Income Inequality

Figure 5. Temporal variation.
Note: Each panel depicts the over-time values of a variable of interest from 1990 to 2010, both with individual lines per state (transparent) and darker lines reflecting the regional mean.



(a) Policy Liberalism

(b) Policy Mood

Figure 6. Maps of policy liberalism and state policy mood.
Note: Darker shades represent higher values of each variable. Values are averages per state from 1990 to 2010.

also tend to enact more liberal policies. There are, of course, many (political, institutional, socioeconomic, and demographic) factors that may explain this relationship. And since both state public opinion and policies are relatively stable, we cannot discern from the relationship alone that public opinion leads to policy differences.

https://doi.org/10.1017/spq.2021.17 Published online by Cambridge University Press

Figure 6 reveals the overall liberalism of state policy and public opinion, averaging over the 1990s and 2000s, using the same measures we used above (we explore several alternative measures with similar results in the Supplementary Appendix). The results show largely familiar regional and partisan patterns. But to what extent and how do these quantities covary? Although causal identification is clearly complicated, we can use the panel structure of the data to isolate potential confounding and unobserved sources of variation. There are many potential models that might be used to assess the relationship. For illustrations, we include two classes of models that are common in the state politics literature and take the form of:

$$\mathbf{Y}_{st} = \beta_{st} + \mathbf{I}_{st} + \mathbf{S}_{st} + \gamma_t + \epsilon_s \tag{1}$$

and

$$\mathbf{Y}_{st} = \beta_{st} + \mathbf{I}_{st} + \mathbf{S}_{st} + \gamma_t + \theta_s + \epsilon_s, \tag{2}$$

where β_{st} is the independent variable or treatment for a given state s in year t, $\mathbf{Y}_{st}$ is a specific outcome, $\mathbf{I}_{st}$ is a vector of time-varying institutional controls (e.g., legislative professionalism, unified or divided government), $\mathbf{S}_{st}$ is a vector of time-varying socioeconomic and demographic controls (e.g., population, income inequality, and gross state product per capita), γ_t is a year fixed-effect to absorb common year-specific shocks (e.g., recessions and election years), and ϵ_s are state-clustered standard errors. The difference in these two specifications comes through the inclusion of time-invariant state controls or a state fixed-effect θ_s. The former case typically includes a regional dummy, such as a control for the state being in the South. The latter case, with state fixed-effects, absorbs all unobserved time-invariant traits of the state, such as fixed regional or cultural variation, and produces an estimate for β *within* the state.[7]

We run three separate classes of models using variations of these two equations. The first set of models includes only year fixed-effects, the second includes year and region fixed-effects, and the final includes year and state fixed-effects. The institutional set of controls consists of: legislative professionalism (Bowen and Greene 2014), the distance between party medians in the upper and lower chambers (Shor and McCarty 2011), union density (Kelly and Witko 2014), and party control of state government (Klarner 2013b). The demographic and socio-economic controls consist of: logged population total from the Census, the Gini coefficient (Frank 2014), gross state product per capita (Cummins and Weiss 2013), and state consumer price index (Berry, Fording, and Hanson 2000). We lag all controls to account for concerns surrounding post-treatment bias (Montgomery, Nyhan, and Torres 2018), but that does not imply that the version with the most controls is necessarily preferable. Instead, the application is designed to show the complexities and impact of model specification. For all analyses the independent variable of interest is state policy mood (Enns and Koch 2013), and the outcome is policy liberalism (Caughey and Warshaw 2016).

In the pooled models with only year fixed-effects, the comparison produced is across states, assessing, for instance, how California compares to Montana. Introducing regional fixed-effects adjusts the comparison, creating a region-specific

[7] In certain settings, this design approximates a difference-in-differences approach, as we discuss below.

https://doi.org/10.1017/spq.2021.17 Published online by Cambridge University Press

intercept and holding fixed anything unobserved that does not change within region. Finally, in the state fixed-effects specification, the variation is produced *within* states. Substantively what this means is that if there is minimal variation in the outcome within one state overtime, most of the effect of β will be absorbed by the state fixed-effect. For each specification we run the models without controls, with only institutional controls, and with institutional and socioeconomic controls.

Figure 7 plots the coefficient estimate for policy mood (β).[8] This figure directly compares the estimated coefficients from these model specifications, including regional time trends and lagged independent variables. All models are presented in full in the Supplementary Appendix. As is clear in this figure, there is substantial attenuation in the size of the coefficient once accounting for observed and unobserved socioeconomic and institutional characteristics of states.[9] Once all time-invariant state characteristics (observed and unobserved) are isolated out with the inclusion of state fixed-effects, the coefficient estimate is near zero (though some control variables may be partial consequences of citizen ideology). This finding echoes Caughey and Warshaw (2018), who find that increasing specification stringency lowers estimated effect sizes. This is not to suggest that the most constrained model is the correct one, only to acknowledge the common inferential difficulties.

This illustration suggests that modeling choices for this relationship are quite important. Researchers can both justify their choices theoretically and explore specifications to gauge their influence on the estimates of interest. Although model specification is a common concern in social science research, the structure of US state panel data makes it even more important. What looks like a political or policy effect may instead be a consequence of stable state attributes, regional differences, socioeconomic variation, or cross-state comparisons.

The estimated coefficients from specifications using state fixed-effects are worth highlighting. A typical modeling strategy is the use of state (unit) and year fixed-effects to assess the impact of a policy on an outcome of interest. This approach, typically referred to as two-way fixed effects, has an advantage in removing time-invariant unobserved state confounders from the regression mitigating concerns around omitted variable bias. However, as shown in Figure 7, this also removes a substantial amount of variation in key outcomes of interest. If the outcome does not vary within state but only across states, and state fixed-effects are *not* used, then the researcher risks over-estimating the effect of interest by pooling together states. In this setting, much of the variation in policy liberalism is absorbed by state and year fixed-effects. This matters not only for the specific question of public opinion influence on policy, but also how we understand many state policy differences.

[8] In Supplementary Appendix Tables A5 and A6, we include results from the same models using the Caughey and Warshaw (2018) measures of mass citizen ideology (economic and social) as the explanatory variable. The results show an attenuation in the size of the estimated coefficient similar to the main results. However, in these models the relationship remains statistically significant suggesting that policy liberalism does predict variation in citizen ideology, but the magnitude depends on specification choice. In Supplementary Appendix Tables A7 and A8, we present the same models using the Erikson *et al.* (1993) weighted state ideology score and Berry, Fording, and Hanson (1998) citizen ideology score, respectively, as explanatory variables. Again, substantive interpretations remain the same.

[9] Supplementary Appendix Figure A4 plots the predicted values of these regressions, comparing the models with and without controls. It shows the same trend: there is a positive and statistically significant relationship between policy mood and policy liberalism, but the slope is smaller once controls are included.

https://doi.org/10.1017/spq.2021.17 Published online by Cambridge University Press

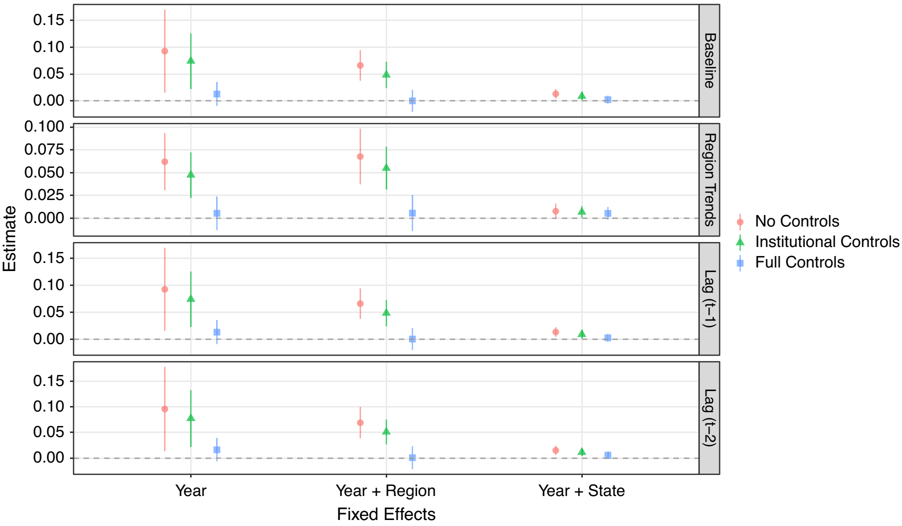


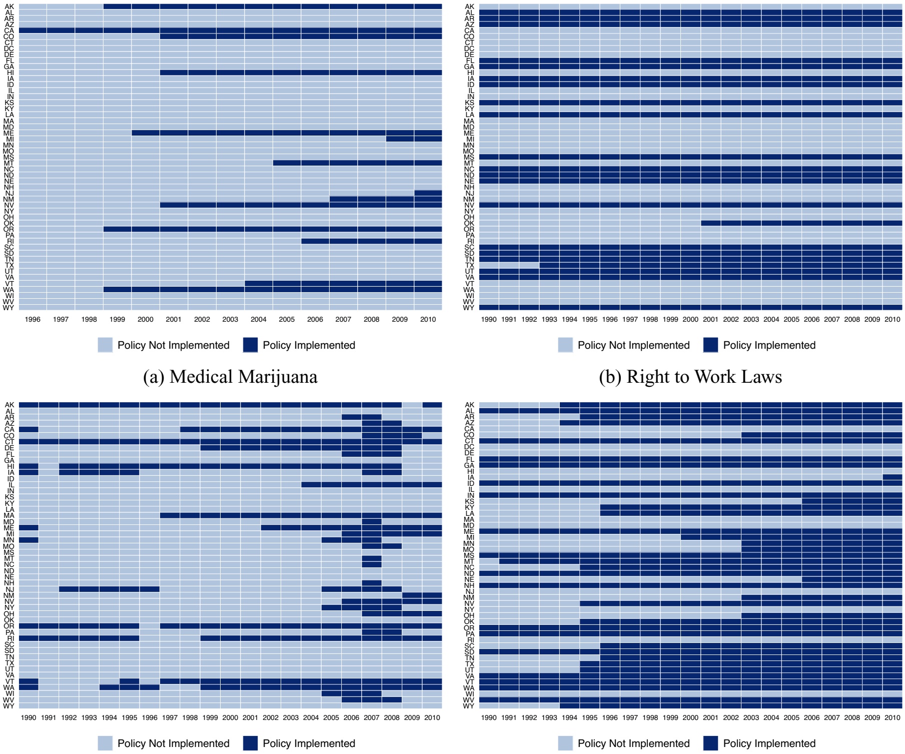
Figure 7. Estimated coefficients by different model specifications.
Note: Models are presented in full in the Supplementary Appendix. The outcome for each model is policy liberalism, and the plotted coefficient is state policy mood. All models include year fixed-effects and state clustered standard errors.

Individual Policies

Many CSPP-enabled studies (as well as those in the broader literature) have been less interested in overall variation in policies along an ideological spectrum than a specific policy output and its potential effect. In fact, the modal study specifies a particular policy of interest (often a dichotomous policy adoption). But the structure of state panel data also matters for these analyses. Further illustrating the caution needed when using state panel data, we turn to examining four individual policies that are of interest to researchers and illustrate the diversity of state policy differences: the legalization of medical marijuana, the presence of "Right-to-Work" laws, whether the state minimum wage is above the federal minimum wage, and whether the state has enacted open carry gun laws. We selected these not only because they are salient policies of interest and span diverse issue areas, but also because they exhibit the differences in temporal patterns of policy change. Each of these policies might be used as an outcome in a study of policy change or diffusion (e.g., Hannah and Mallinson 2018), or as a "treatment" in studying some state-specific outcome produced by the adoption of the policy.

Figure 8 displays variation in these four policies from 1990 to 2010 for all states in our sample. There are three important takeaways from this visualization for applied research design in state politics. First, these examples demonstrate policy-specific differences in the degree to which variation exists both *across* states and *within* states over time. Second, some policies do not change much within states over time, as we see with "Right-to-Work" laws only adopted by two new states during this time period. Third, policies are enacted at different time periods (e.g., open carry gun laws)

https://doi.org/10.1017/spq.2021.17 Published online by Cambridge University Press



(a) Medical Marijuana

(b) Right to Work Laws

(c) Minimum Wage > Federal Minimum Wage

(d) Open Carry

Figure 8. The presence of four policies across states and time.
Note: This figure plots the time-varying status of four separate policies. The *x*-axis is years, and the *y*-axis is all of the states in the data. Light colored cells indicate the policy was not in effect, and dark colored cells indicate years when the policy was in effect.

and some even “turn on” and “off” again (when state minimum-wage is above-federal minimum wage).

This visualization exercise is a useful first step when pursuing a panel research design.[10] The structure of variation in the outcome changes our capacity for inference in different designs. For instance, using a pooled approach with only year fixed-effects to assess the correlation between medical marijuana laws and an outcome such as opioid deaths may bias estimates toward states with substantial numbers of opioid deaths. On the other hand, pursuing a two-way fixed effects strategy (e.g., Harden and Kirkland 2019) using “Right-to-Work” laws as a treatment, the entirety of variation would come from Texas and Oklahoma, the only two states with changes during the time span (treatment in the other states is collinear with state fixed-effects). Of course, researchers often know when they have minimal variation and might extend

[10] The `plot_panel` function in the R package creates this style of visualization.

https://doi.org/10.1017/spq.2021.17 Published online by Cambridge University Press

the time period to address that limitation. These (extreme) examples simply show the value of examining variation.

In the Supplementary Appendix, we demonstrate the sensitivity of the results when using policy adoption as a dependent variable. Table A9 and Figure A6 in the Supplementary Appendix display results from the same specifications as Equations (1) and (2) to assess open carry gun law adoption, thus assessing the relationship between public opinion and this policy enactment. The substantive finding again depends heavily on model specification, ranging from a statistically significant decrease in the probability of adopting open carry gun laws when not using state fixed-effects, to a noisy but positive increase in probability with the inclusion of state fixed effects and time varying controls. Researchers working in this area would consider alternative specifications, possibly including covariates accounting for diffusion or policy-specific measures of public opinion. Here we simply try to connect the results to the prior analyses, showing that similar considerations for state panel data apply when analyzing the adoption of a particular policy.

In policy treatments with high variation, such as the minimum wage or open carry gun laws, a two-way fixed effects approach might be problematic. When treatments turn on and off at different times or some states are always treated or never treated, the two-way fixed effect estimator can produced biased estimates (including estimates in the erroneous direction) of the true treatment effect (Goodman-Bacon 2018). Fortunately, there are approaches for correcting this bias given an adequate number of state-years in which to construct comparison groups (see Imai and Kim 2021).

These research design problems are increasingly well-known and there is software for analyzing the robustness of a chosen model specification (Imai and Kim 2021). Additionally, a longer time series and a variety of covariates can improve estimates and statistical power in many settings (see Harden and Kirkland 2019, for a more thorough discussion of these issues). The CSPP data and accompanying package are valuable tools for applied research by providing access to a variety of outcomes, treatments and covariates that facilitate research design and inference. But a first step is to assess how much variation is available and how much it relies on differences across states or time.

Improving State Policy Research with CSPP Panel Data

The CSPP database and accompanying tools provide researchers, policy practitioners, and instructors a valuable resource in uncovering relationships between socioeconomic, demographic, political, and policy variables and exploring patterns across states and over time. CSPP facilitates different research designs for testing theories and exploring relationships, including cross sectional, state panel, difference-in-differences, regression discontinuity, and other approaches. CSPP even allows for rich single-state studies (Nicholson-Crotty and Meier 2002). Data can be used descriptively across states for productive ends as well. CSPP data are already being employed by researchers, practitioners, instructors, and students across multiple disciplines. We plan to continue supporting CSPP through expanding data infrastructure and creating additional tools that facilitate description and analysis of US states.

CSPP has been and will continue to be useful for analyzing topics of paramount importance to researchers and policy makers in American politics, such as polarization, representation, and the policy-making process. It can also be advantageous for analyzing more niche policy areas, allowing users to explore whether they are

https://doi.org/10.1017/spq.2021.17 Published online by Cambridge University Press

representative of broader trends. An additional value of CSPP is in the construction of new measures and indices that combine multiple variables. Whatever the interest, scholars can both explore determinants of policy adoption and diffusion and exploit policy enactments as independent variables, investigating their effects on state's socio-economic trajectories, institutions, and behavior (e.g., Dynes and Holbein 2020; Grumbach 2018; Harden and Kirkland 2019). Including state policies on the right-hand side of the equation also opens state politics research to other epistemological communities, making it useful to scholars focused on health, education, environmental, economic, or criminal justice outcomes. However, these communities face the same challenges we outline above: understanding the structure of state panel data and taking advantage of available variation to assess relationships and improve inferences. Thus far, much CSPP-enabled research seeks to assess causal claims about policy in particular issue areas. That means it should be attentive to both the structure of variation on particular state policies and the larger regional and ideological trends that could be shaping policies and results.

Regardless of how scholars use CSPP (and state panel data generally), both regionalism and state-level stability over time should be assessed as they present a potential obstacle in the pursuit of credible identification strategies. While it is often advantageous to leverage several covariates over a long period of time (for regression analysis, pretreatment matching, and other applications), there is no substitute for understanding the structure of state panel data and the commonality of state and regional differences. Researchers should be cautious in assuming change is common for outcomes of interest, while capitalizing on such variation when it exists within and across states over time. These difficulties arise in many contexts and are central to the credibility revolution in social science. CSPP highlights the benefits and the difficulties associated with analyzing population, institutional, and policy variation in panel data across any political units. By doing so, it illustrates some fundamental challenges of research and avenues for understanding and overcoming limitations.

CSPP builds on many past and current data contributions. It is a testament to the development of the state politics field. We hope it is a useful resource and public good, greatly expanding our collective ability to better understand how institutions and behavior affect policy, politics, and polities in the American laboratories of democracy.

Supplementary Materials. To view supplementary material for this article, please visit http://dx.doi.org/10.1017/spq.2021.17.

Acknowledgments. We thank Nicolas Bichay, Layla Brooks, Emily Corbeille, Troy Distelrath, Jillian Evans, Nikolaos Frantzeskakis, Babs Hough, Dom Korzecke, Joshua Koss, Caleb Lucas, Zuhaib Mahmood, Shayla Olson, Karl Schneider, Jonathan Spiegler, and Bailey Tjolsen for their outstanding research assistance in helping to maintain, update, and expand the Correlates of State Policy Project database. We thank numerous contributors for sharing data, including Frederick Boehmke. We also appreciate the insightful comments from participants at the 2017 State Politics and Policy Annual Conference in St. Louis, Missouri, especially Dan Butler, Virginia Gray, Michael McDonald, Steve Rogers, and Boris Shor. We are also grateful to Chris Warshaw and to the editors and reviewers for their helpful suggestions.

Funding Statement. The authors thank the Institute for Public Policy and Social Research and the Department of Political Science at Michigan State University and the W. K. Kellogg Foundation for support of their research.

Conflict of Interest. The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement. Replication materials are available on SPPQ Dataverse at https://doi.org/10.15139/S3/DQASUD (McCrain, Grossmann, and Jordan 2021).

References

Berry, William D., Richard C. Fording, and Russell L. Hanson. 1998. “Measuring Citizen and Government Ideology in the American States, 1960-93.” *American Journal of Political Science* 42 (1):327–48.

Berry, William D., Richard C. Fording, and Russell L. Hanson. 2000. “An Annual Cost of Living Index for the American States, 1960-1995.” *Journal of Politics* 62 (2):550–67.

Berry, William D., Richard C. Fording, Evan J. Ringquist, Russell L. Hanson, and Carl E. Klarner. 2010. “Measuring Citizen and Government Ideology in the U.S. States: A Re-appraisal.” *State Politics & Policy Quarterly* 10 (2):117–35.

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):517–45.

Boehmke, Frederick J., and Paul Skinner. 2012. “State Policy Innovativeness Revisited.” *State Politics & Policy Quarterly* 12 (3):303–29.

Bowen, Daniel C., and Zachary Greene. 2014. “Should We Measure Professionalism with an Index? A Note on Theory and Practice in State Legislative Professionalism Research.” *State Politics & Policy Quarterly; Springfield* 14 (3):277–96.

Brace, Paul, and Melinda G. Hall. 2009. “State Supreme Court Data Project.” Harvard Dataverse, V1. Dataset. https://doi.org/10.7910/DVN/Z80F7P.

Carsey, Thomas M., Richard G. Niemi, William D. Berry, Lynda W. Powell, and James M. Snyder Jr. 2008. “State Legislative Elections, 1967 – 2003: Announcing the Completion of a Cleaned and Updated Dataset.” *State Politics & Policy Quarterly* 18 (4):430–43.

Caughey, Devin, and Christopher Warshaw. 2016. “The Dynamics of State Policy Liberalism, 1936–2014.” *American Journal of Political Science* 60 (4):899–913.

Caughey, Devin, and Christopher Warshaw. 2018. “Policy Preferences and Policy Change: Dynamic Responsiveness in the American States, 1936–2014.” *American Political Science Review* 112 (2):249–266.

Cummins, John D., and Mary A. Weiss. 2013. “Analyzing Firm Performance in the Insurance Industry Using Frontier Efficiency and Productivity Methods.” In *Handbook of Insurance*, edited by Georges Dionne, 795–861. New York, NY: Springer.

Dynes, Adam M., and John B. Holbein. 2020. “Noisy Retrospection: The Effect of Party Control on Policy Outcomes.” *American Political Science Review* 114 (1):237–57.

Enns, Peter K., and Julianna Koch. 2013. “Public Opinion in the U.S. States 1956 to 2010.” *State Politics & Policy Quarterly* 13 (3):349–72.

Erikson, Robert S., Gerald C. Wright, Gerald C. Wright, and John P. McIver. 1993. *Statehouse Democracy: Public Opinion and Policy in the American States.* Cambridge: Cambridge University Press.

Frank, Mark W. 2014. “A New State-Level Panel of Annual Inequality Measures over the Period 1916–2005.” *Journal of Business Strategies* 31 (1):241–63.

Goodman-Bacon, Andrew. 2018. “Difference-in-Differences with Variation in Treatment Timing.” *Technical Report*, National Bureau of Economic Research.

Gray, Virginia, and David Lowery. 1988. “Interest Group Politics and Economic Growth in the U.S. States.” *The American Political Science Review* 82 (1):109–31.

Grumbach, Jacob M. 2018. “From Backwaters to Major Policymakers: Policy Polarization in the States, 1970–2014.” *Perspectives on Politics* 16 (2):416–35.

Hannah, A. Lee, and Daniel J. Mallinson. 2018. “Defiant Innovation: The Adoption of Medical Marijuana Laws in the American States.” *Policy Studies Journal* 46 (2):402–23.

Harden, Jeffrey J., and Justin H. Kirkland. 2019. “Does Transparency Inhibit Political Compromise?” *American Journal of Political Science* 65 (2):493–509.

Imai, Kosuke, and In Song Kim. 2021. “On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data.” *Political Analysis* 29 (3):405–15. https://doi.org/10.1017/pan.2020.33.

Jewell, Malcolm E. 1982. “The Neglected World of State Politics.” *The Journal of Politics* 44 (3):638–57.

https://doi.org/10.1017/spq.2021.17 Published online by Cambridge University Press

Kelly, Nathan J., and Christopher Witko. 2014. "Government Ideology and Unemployment in the U.S. States." *State Politics & Policy Quarterly* 14 (4):389–413.

Klarner, Carl. 2013a. "Governors Dataset." Harvard Dataverse. Dataset. http://hdl.handle.net/1902.1/20408.

Klarner, Carl. 2013b. "Other Scholars' Competitiveness Measures." Harvard Dataverse, V1. Dataset. http://hdl.handle.net/1902.1/22519.

Klarner, Carl. 2013c. "State Economic Data." Harvard Dataverse, V1. Dataset. https://doi.org/10.7910/DVN/KMWN7N.

Klarner, Carl. 2013d. "State Partisan Balance Data, 1937–2011." Harvard Dataverse, V1. Dataset. http://hdl.handle.net/1902.1/20403 .

Lucas, Caleb, and Josh McCrain. 2020. "cspp: A Tool for the Correlates of State Policy Project Data." *R package version 0.3.0.* https://cran.r-project.org/web/packages/cspp/index.html.

McCrain, Joshua, Matt Grossmann, and Marty P. Jordan. 2021. "Replication Data for: The Correlates of State Policy and the Structure of State Panel Data." UNC Dataverse. Dataset. https://doi.org/10.15139/S3/DQASUD.

Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres. 2018. "How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It." *American Journal of Political Science* 62 (3):760–75.

National Conference of State Legislatures (NCSL). 2016. "Ballot Measures Database." http://www.ncsl.org/research/elections-and-campaigns/ballot-measures-database.aspx.

Nicholson-Crotty, Sean, and Kenneth J. Meier. 2002. "Size Doesn't Matter: In Defense of Single-State Studies." *State Politics & Policy Quarterly* 2 (4):411–22.

Olson, Shayla F. 2019. *State Networks Database v.1.1*. East Lansing, MI: Michigan State University, Institute for Public Policy and Social Research (IPPSR). http://ippsr.msu.edu/public-policy/state-networks.

Shor, Boris, and Nolan McCarty. 2011. "The Ideological Mapping of American Legislatures." *The American Political Science Review* 105 (3):530–51.

Sorens, Jason, Fait Muedini and William P. Ruger. 2008. "U.S. State and Local Public Policies in 2006: A New Database." *State Politics & Policy Quarterly* 8 (3):309–26.

Squire, Peverill. 2007. "Measuring Legislative Professionalism: The Squire Index Revisited." *State Politics & Policy Quarterly* 7 (2):1–27.

Squire, Peverill. 2008. "Measuring the Professionalization of U.S. State Courts of Last Resort." *State Politics & Policy Quarterly* 8 (3):223–38.

US Department of Commerce Bureau of Economic Analysis. 2012. "NAICS Per Capita GDP by State/SIC Per Capita GDP by State." http://www.bea.gov/regional/downloadzip.cfm.

Walker, Jack L. 1969. "The Diffusion of Innovations among the American States." *The American Political Science Review* 63 (3):880–99.

Wright, Gerald C., Jr., Robert S. Erikson and John P. McIver. 1987. "Public Opinion and Policy Liberalism in the American States." *American Journal of Political Science* 31 (4):980–1001.

Cite this article: Grossmann, Matt, Marty P. Jordan, and Joshua McCrain. 2021. The Correlates of State Policy and the Structure of State Panel Data. *State Politics & Policy Quarterly* **21** (4): 430–450, doi:10.1017/spq.2021.17