Public Opinion in Policy Studies: Lessons from the Logic of Structural Equation Modeling

Nate Breznau, breznau.nate@gmail.com

University of Bremen, Germany
Comparative Research Center, “The Global Dynamics of Social Policy” (SFB 1342)

ABSTRACT

Policy studies scholars regularly investigate linkages between public opinion and policy. The use of public opinion as a variable in empirical research poses special challenges. In this article I suggest that the logic and methods inherent in the art of structural equation modeling provide opportunities to overcome some of these challenges. I describe this type of logic as it pertains to measurement error, context effects and endogeneity. Using General Social Survey data for the United States and taking thermostatic feedback models as an example, I demonstrate why it is important for policy scholars to attend to (a) measurement modeling and cross-level isomorphism, (b) shocks that might bias survey response patterns and (c) endogeneity implied by the theoretically reciprocal nature of opinion and policy feedback. These examples come with discussions of why scholars should pay attention to model specification so that theory and empirics are in unison and how to perform model fitting and testing to better develop theories and models of policy processes.

1 IMPROVING OPINION AND POLICY STUDIES

Public opinion is a quintessential topic in policy studies research (Burstein 1998; Larsen 2018; Mullinix 2011). There is ample theory and evidence suggesting opinion and policy are part of processes leading to changes in each other. For example, policymakers should be interested in public support for themselves and their parties as they seek to gain votes and seats; furthermore, it is literally part of the policymaker's job description to serve the constituents' will (Butler and Nickerson 2011; Soroka and Wlezien 2010). Once in office, opinion can influence policymaker discussions, actions and implementations of policy (Stimson, Mackuen, and Erikson 1995). Public opinion reflects or predicts how the public votes or to which parties and candidates they offer their support (Irwin and Van Holsteyn 2002; Schmitt-Beck 1996) and is a heuristic guiding the actions of those working in bureaucracies and governmental institutions (L. R. Jacobs 1992; Steensland 2006). Therefore, it is no surprise that public opinion is a ubiquitous feature of policy-related research.

The use of public opinion as a variable – a particular attribute of a population whose properties change over time – poses special challenges to policy studies. This is partly an epistemological problem, and philosophical debates throughout history question whether a 'public' is an entity and can have a feature such as 'opinion' (Splichal 1999). It also is myriad methodological and theoretical issues. There are limitations to the instruments available for observing public opinion. A typical approach is to use representative surveys. Acknowledging that media, social movements and networks are also important components of public opinion not easily observed in surveys (Herbst 1992; Huckfeldt 2007), the survey remains the workhorse of public opinion research, for better or worse. Given this reality, the survey is also the focus for measuring public opinion in this article.

Survey data come with inherent errors due to sampling, fielding and coding. The same question in a survey may take on different meanings or cue different attitudes among a population of individuals. The common practice of relying on data obtained from single survey questions as a direct measure of public opinion may therefore be quite unreliable. Unlike policies, opinion change can occur instantaneously. Thus, context effects may suddenly shift opinion, or worse, shift the meaning of the same survey question fielded over time (Nie, Petrocik, and Verba 1999). As opinion represents a property of an entire society, potential unobserved heterogeneity may result from any large-scale disturbances to that society at a given time point. Such shocks could confound scholars' attempts to accurately measure opinion.

Additionally, theories of two-way causality running between opinion and policy suggest the presence of endogeneity that may bias estimates of the impact of one on the other. Scholars tend to aggregate individual survey responses to the mean or median of a given population sample, and then either predict this value as an outcome or treat it as an input to predict a policy or some event along the policymaking process (Burstein 1998; Kang and Powell 2010; Shapiro 2011). Thus, scholars treat opinion and policy as *either* independent *or* dependent variables, and in doing so they implicitly run statistical tests of one-way causal processes. Either X

causes Y or Y causes X . This is a misspecification given theories of the reciprocal relationship between opinion and policy (Brezna 2017; Papadakis 1992; Wlezien 2017).

In this article I offer some solutions to these issues, if not arguments for how to better question public opinion data and research. These derive from the logic of structural equation modeling (SEM). This scientific approach involves thinking about, adjusting for and possibly overcoming biases that result from errors, be they of measurement or confounding. Often thought of as a statistics method, SEM is actually a way of thinking about and modeling unobservable phenomena in the world. It is essentially the art of taking a data generating theory and converting it into statistical models that include known and unknown sources of error as specified by the theory. The SEM practitioner therefore applies any of a variety of methods (Bielby and Hauser 1977; Heckman 2010; Mackenzie 2001; Wolfle 2003). I will not discuss the range of methods in the SEM toolkit¹. I focus here only on demonstrating the way of thinking that characterizes SEM and how this can improve opinion-related policy studies.

I offer three examples where SEM logic questions and improves state of the art policy research. These are: (1) consideration of measurement error in individual survey questions as indicators of public opinion, (2) confounding of opinion due to context and adjusting for this, and (3) potential bias resulting from endogeneity between opinion and policy and how to model and test it. I use three well-known features of policy studies in these examples: the case of the United States, the thermostatic feedback models of Soroka and Wlezien and the data from the *General Social Survey* (GSS)²³.

2 MEASUREMENT ERROR

Errors associated with measurement are a fundamental SEM topic (Bollen 1989; Holbert and Stephenson 2006). One source of error in public opinion research is the survey. As an instrument to measure the complexity of public opinion; the survey has its shortcomings and unreliability because it attempts to fit, "...idiosyncratic attitudes of respondents into cleanly defined and specific questions in a controlled environment" (Perrin and McFarland 2011, 97). Given the complex causal pathways leading to individual survey responses, reliance on a single question to measure an attitude or a single-question mean to measure public opinion for the whole of society should be avoided if possible (Achen 1975; Andrews 1984; Ansolabehere, Rodden, and Snyder Jr. 2008; Diamantopoulos et al. 2012; Groves and Lyberg 2010; Judd and

¹ The reader can find out about these tools in any basic SEM text (Bollen 1989; Bollen and Pearl 2013; Byrne 2001; Kline 2016).

² A note on case selection: The GSS is perhaps the most widely used survey in United States opinion research (Smith et al. 2018). It was central to the development of Soroka and Wlezien's (2010) (hereafter "S&W") thermostatic models based on earlier work by Wlezien (1995, 1996). Selection of the thermostatic model derives partly its prominence with the original 1995 article approaching 1,000 citations and the following articles and book having hundreds each according to Google Scholar 19/06/2019, and from my own in-depth knowledge of the model. Soroka and Wlezien (2005b, 2005a; Wlezien and Soroka 2012) pushed forward the boundaries of policy studies. In fact, their work already contains some SEM approaches including thinking about errors in measurement using factor analysis of GSS questions, treating absolute spending preferences as an unobserved property of the public, and analyzing opinion and policy spending as a feedback system.

³ All replication materials available at the permanent URL <https://osf.io/vu3s5/>.

Krosnick 1989). Even in seemingly innocuous questions, respondents occasionally fail to accurately report things as simple as their age (M. D. R. Evans and Kelley 2008).

Additionally, the questions in a survey may or may not measure similar attitudes across a given population of individuals, groups of individuals or among the same population at different times (Davidov et al. 2014; StegmueLLer 2011). If a survey instrument randomly misestimates an attitude this only impacts reliability, not accuracy of measurement. However, if a given survey instrument consistently over estimates for one group while consistently underestimating for another, the groups could appear to have identical preferences or opinions in the data when in fact they diverge in reality. Policy studies scholars applying an SEM perspective would ask themselves: *Do the individuals in my sample understand the question similarly?*

To demonstrate how to ask such questions, I take S&W's (2010, 67, 77) concept of "net preferences" which is equal to the aggregate percentage of the sample population in the GSS who indicate, 'we are spending too little' on a policy minus the percentage of respondents who indicate, 'we are spending too much' in a given year. This aggregation creates a new variable measuring something at the *public-level*. There is a theoretical assumption by S&W that there are countervailing forces across individual attitudes (pro versus con) that lead to a real-world feature of public opinion that is of interest to elected representatives, candidates for office, scholars, members of the public and the media; and that this feature is part of the causal process leading to, "...public policy, the final and most significant outcome of representative activity" (ibid., 14). Exactly how it gets there is a matter discussed at length in the work of S&W and their literature review (2010).

This means there are *empirical* net preferences in the GSS data and there are *latent* net preferences in the US population at large. The empirical version is nothing more than a statistical product of survey data, but the latent version is a theoretical, but unobserved, quality of the population that is a cause of policymaking. Of course surveys are a signal for policymakers to interpret public opinion, but not the only one as opinion matters in politicians' networks, the media and appeals of constituents (Wlezien 2004). The assumption is that public opinion as a latent variable has a causal impact on all of these signals. Therefore, the better a researcher can measure it in the GSS, the closer that researcher gets to the reality of its effect on policymaking. If there are no systematic errors then the GSS is a perfect measure of latent net preferences, with a degree of reliability proportional to random measurement error. However, survey research rarely has perfectly randomly distributed error. A simple random sample is not possible given that not every member of the population is available to respond to surveys; e.g., the ultra-wealthy and illegal aliens are almost impossible to sample. Moreover, the connotative meanings of questions and their wordings in surveys can produce side effects in response patterns that do not reflect what researchers intended the question to measure (Rasinski 1989; Smith 1987). A researcher cannot change the sample when the data are preexisting, thus questionnaire development is not possible in the current example of the GSS; however, researchers can adjust for measurement error associated with the questions and their wordings. If successful, such adjustments bring the empirical estimates to more accurately reflect the latent estimand.

2.1 *Measurement Modeling in the GSS*

To explore net preferences, S&W engaged in factor analysis – a technique using several questions at once to identify a latent attitude or opinion. I use the word “technique” because factor analysis is not a single method and comes in a variety of formats. Factor analysis represents a fundamental form of SEM logic and practice. The researcher assumes the existence of unobserved variables known as factors – statistical objects meant to capture latent variables. These factors are things shared across a population, like S&W’s net preferences, and theoretically represent some of the data-generation processes leading to the observed data at hand (Jöreskog 2007)⁴. Factors explain variation in individual responses across a survey module. Factor analysis gives researchers the power to investigate how different questions are stronger or weaker indicators of a latent variable, identify the presence of several latent variables and their covariances with one another, and offers the chance to identify and parse out sources of error (Tabachnick and Fidell 2007). If there are different strengths of association between a latent factor capturing public opinion and different survey questions, the common practice of taking the mean of those questions is a misspecification of the latent variable’s properties. Moreover, some questions may not measure the latent variable of interest at all, even if they contain language that theoretically should. In short, each question carries idiosyncratic sources of error and factor analysis is a technique to reduce these. Of course, if not properly applied, factor analysis can also inflate or introduce error.

By analyzing public-level rather than individual-level preferences, S&W define public opinion as something at the level of society rather than individual. This is consistent with S&W’s (2010) review of theory and the standard ontological notion of public opinion (de Boer 2007). Aggregation carries the implicit assumption that the questions have a consistent meaning across individuals over time (Smith 1994), a property known as cross-level isomorphism (Ruelens, Meuleman, and Nicaise 2018). Each survey question might evoke varying heuristics leading to random measurement error, but there should be an underlying common understanding. The logic behind this comes from the SEM perspective, that there is something unobserved but shared across individuals and that the survey question can tap into this. If this logic cannot be empirically substantiated, the use of the survey questions at any level may be biased (Tay, Woo, and Vermunt 2014).

In S&W’s original analysis, they looked only at the public-level thus raising questions of measurement error here. Their method was principal components analysis (PCA) where factors are additive rather than correlated, thus any identified factors are unique, and is a preferable method when data are not multivariate normally distributed, like these GSS questions with only three answer categories (Fabrigar et al. 1999). I follow their lead and conduct PCA analyses on the same nine questions, but I do this at both the individual and public-levels. These questions ask about more, the same or less spending in nine areas that I label with capitalized nouns for spending on: Environment, Health, Education, Welfare, Big Cities, Foreign Aid, Space, Crime and Defense; see Appendix A for wording. At the time S&W (2010) had GSS data pre-2006. Thus, I start by analyzing the same time period; results in Table 1.

⁴ Except for construction of “reflective” indicators, a technique outside of the current theoretical discussion.

Table 1. Principal Components Analyses of Questions about Government Spending, GSS Data 1972-2005.
Standardized, Varimax Rotated Factor Loadings⁵

<i>Individual-Level, N= 28,235</i>				
Variable	Factor1	Factor2	Factor3	Factor4
Environment	0.630	0.016	0.264	-0.281
Health	0.656	0.013	-0.076	-0.052
Education	0.638	0.030	0.147	-0.081
Welfare	0.205	0.688	-0.174	-0.174
Big Cities	0.587	0.241	-0.189	0.051
Foreign Aid	-0.061	0.819	0.151	0.096
Space	0.026	0.034	0.872	0.130
Crime	0.527	-0.070	-0.334	0.492
Defense	-0.112	0.016	0.196	0.831
cumulative explained variance	0.213	0.347	0.469	0.589

NOTE: Only eigenvalues > 1 reported. Shaded cells are at least 0.55 rounded.

<i>Public-Level, N= 24</i>				
Aggregate Variable	FactorA	FactorB	FactorC	FactorD
Environment	0.561	0.291	-0.059	0.674
Health	0.637	0.279	0.438	0.417
Education	0.553	0.738	0.145	0.223
Welfare	0.116	0.220	0.120	0.904
Big Cities	0.920	-0.050	-0.209	0.093
Foreign Aid	0.102	0.539	0.701	0.054
Space	-0.139	0.955	0.021	0.185
Crime	0.221	0.052	-0.921	-0.048
Defense	-0.630	0.172	0.361	-0.451
cumulative explained variance	0.262	0.484	0.678	0.872

NOTE: Four factors retained for comparability with individual-level. Shaded cells are at least 0.55 rounded.

⁵ Loadings diverge slightly from S&W's original study, most likely because they merged two additional data sources into the GSS data when making their aggregate data file. I focus here on the use of a single survey for simplicity.

At first glance, Factor1 at the individual-level looks similar to FactorA at the public-level. Except for a slightly divergent role of Defense, these two factors are determined largely by questions on Environment, Health, Education and Big Cities. Factor2 at the individual-level is driven in particular by the question on Foreign Aid, similar to FactorC at the public-level. Factor3 and FactorB share a strong loading from Space. Factor4 and FactorD are dissimilar, but FactorD has an eigenvalue of only 0.7 meaning it should probably not be included; however, I kept it in for a consistent four-factor structure between levels. The first three factors having a degree of similarity at both levels is quite reassuring, if not surprising, given that there are only 24 cases at the public-level and such small samples make analytical results less reliable. It appears that individuals share similar latent factors and that the meaning of these carries over to the public-level, within a degree of error and varying factor loadings.

These results offer support of S&W's theoretical usage of the GSS at the public-level. Yet a closer look at what they analyzed raises some questions. Rather than run their thermostatic models on three factors – where Factor1 is Environment, Health, Education and Big Cities; Factor2 is Welfare and Foreign Aid; and Factor 3 is Space – they instead analyzed two factors that they derive somewhat in contradiction of their PCA results. First they analyze Defense on its own. The PCA supports this choice as shown in Table 1. Second, they analyze a factor they label “Social” consisting of Education, Welfare and Health (Soroka and Wlezien 2010, 93–96). The combination of Education, Welfare and Health has some theoretical reasoning behind it, related to spending and policymaking domains (*ibid.*); however, the PCA in Table 1 and in their own research suggested that Welfare does not belong with those two. The SEM toolkit offers a way to resolve this contradiction using model fitting.

On the one hand, I have evidence that Environment, Education, Health and Big Cities should be one factor (Factor1 in Table 1). On the other, S&W argue for putting Education, Welfare and Health together. I simplify this test into the question of whether Welfare should be combined with Education and Health or not. This is a legitimate theoretical concern as well given that Welfare is a contentious political concept throughout US history (Béland and Waddan 2017). Table 2 presents an SEM test of models with Welfare in and with Welfare out. It literally takes the logic behind the PCA results and pits it against the S&W model and tests them against each other using the GSS data.

Figure 1. Testing Alternative Models of “Social” Spending Preferences, GSS Data, 1973-2005

In this example, looking only at the public-level it is not possible to tell the factor structures of the two competing models apart when comparing the results of Model 1 and Model 2. First of all, the equal fit tests – a comparison of the observed covariance among the variables versus the model’s implied covariances based on the factor structure with a p-value above 0.05 (Kline 2016) – suggest these models ostensibly fit the data well; however, when comparing the models to each other using an equal fit test, there is no significant difference between them so no way to suggest one is better than the other. However, at the individual-level the S&W measurement model (Model 1) fails an exact fit test ($p < 0.05$) and fits the data significantly worse than Model 2 (difference in chi-square is $(93.48 - 3.13) = 90.35$ with 1 df, $p > 0.05$). This suggests giving preference to Model 2. Ultimately, the Social factor correlates at 0.97 as a latent factor predicted from the two models so there is only small empirical distinction. I return to this ‘small’ difference and how it affects other tests in Table 2. The theoretical implications are nonetheless important: although empirically very similar, Model 2 provides support of Welfare standing alone as a separate factor. Individuals in the United States between 1972 and 2005 did not

have a shared unobserved attitude measurable by combining health, education and welfare spending questions on a survey; welfare is a unique concept in their minds.

2.2 Testing the Implications of Measurement Modeling in the Thermostatic Model

In the thermostatic model, S&W hypothesize that “public responsiveness” takes place where the public’s relative preferences for increasing or decreasing spending in a particular policy domain at the national-level is determined by current policy spending and absolute preferences for policy spending. If the public’s absolute preference is higher than current spending then they prefer an increase, and vice-versa. Thus a change in policy spending produces an inverse change in relative preferences. At the same time, “policy representation” takes place where changes in spending reflect the relative preferences of the public from the previous year, back when the spending budget was made, in addition to being a product of the partisan make-up of the government, with Republicans tending to spend less and Democrats more. All details of their model, including the system of formal equations, statistical methods and replication data⁶ are available elsewhere (Soroka and Wlezien 2010; Wlezien 1995). I do, however, present the formal equations behind the thermostatic model in Section 3 to shed light on endogeneity. For now, this summary suffices to proceed with a test of how results might depend on the measurement of the Social factor.

Using S&W’s original data, I re-run their thermostatic models for “Social”. The left two columns of Table 2 present the original models with the Welfare question in and measured using an additive index (S&W:Table 5.2), and the right two columns the new models suggested by the factor analysis with Welfare out and predicted as a latent variable.

⁶ <http://degreesofdemocracy.net/>

Table 2. Thermostatic Feedback Models Comparing Alternative Measures of “Social” Spending

Variable	Soroka & Wlezien Models ^a		SEM Derived Models ^b	
	Public Responsiveness	Policy Representation	Public Responsiveness	Policy Representation
<i>----- metric results -----</i>				
Social Spending	-0.516***		-0.087	
Net Social Preferences t-1	0.598***	0.314**	0.553***	0.791**
Social Spending t-1		0.050		0.039
Republican President		-1.395		0.199
Republican Seats in Congress		-1.100**		-0.303
Counter^2	0.181***		0.024*	
constant	84.656***	24.463	4.602	15.468
r^2	0.840	0.546	0.666	0.504
<i>---- standardized results ----</i>				
Z - Social Spending	-0.322		ns	
Z - Net Social Preferences t-1	0.593	0.633	0.543	0.569
Z - Social Spending t-1	0.272	ns		ns
Z - Republican President		ns		ns
Z - Republican Seats in Congress		-0.488		ns

^a"Social" = Health, Education & Welfare; "Preferences" calculated as an additive index

^b"Social" = Health, Education & Environment; "Preferences" calculated as predicted values from a factor analysis

ns = not significant

Their original models demonstrate strong negative responsiveness to increases in spending (-0.516***) and strong policy representation as a positive response to public preferences (0.314**). The SEM-derived models show something slightly different. They find evidence of policy representation (0.791**) and it is very strong. This means that, if this model is correctly specified, that policy- and budget-makers are very responsive to ‘public mood’ (Stimson 2018). However, they do not show evidence of public responsiveness to social spending due to an insignificant responsiveness coefficient. This, despite the public expressing a coherent latent attitude (as shown in Table 1), it does not associate in the expected way with spending. This immediately opens up many questions as to why. For example, why does having Welfare in the “Social” measure, and taking the index as a sum score lead to the thermostatic findings. For example, does the Welfare question and portion of the national budget actually drive the model’s findings? Although I cannot fully answer why these findings suggest a lack of thermostatic responsiveness, they point toward the fact that welfare is unique as to such an extent as to potential change the results depending on model specifications.

3 CONTEXT-SPECIFIC DISTURBANCES AND COUNTERFACTUAL LOGIC

Context is a source of potentially unobserved error in public opinion. If unique events in specific times and places have a large impact on opinion, policy or the linkages between them, attempts to isolate a single statistical effect out of context may be biased. This is a phenomenon sometimes known as effect heterogeneity.

Taking a regression to the mean approach assumes that the impact of opinion on policy is constant over time, at least when all other factors in the process are taken into consideration. If an event takes place suddenly shifting survey responses very far in one direction, possibly because the meaning of the survey question temporarily changed in the minds of the public, without a corresponding shift in policy spending for example, this could turn it into an outlier that has a measureable impact on the mean and variance. The result could be a shift in intercept or weakening of the covariance. This potentially reduces or inflates the observed impact of opinion on policy. Thinking about society and political events, scholars can ask counterfactual questions and find solutions to the bias of context-specific disturbances using measurement strategies.

Here are two examples in the US: The Cold War and the events of economic recession and health care reform around 2008 that shocked opinion responses. Figure 2 plots mean-centered net preferences for the GSS questions on Defense and Health.

Figure 2. Mean GSS Responses to Defense and Health Spending, 1973-2016

Wlezien (1995) pointed out a guns-butter tradeoff in public opinion with Health and Defense following inverse trends from 1973 to 1990. These trends end shortly after the collapse of the Soviet Union and then discontinued thereafter as seen in Figure 2. This is a context-effect. The Cold War had an immense impact on rhetoric and policymaking. Often discussed in the public sphere was the trade-off between social and defense-related spending. Thus, it is theoretically likely that public experienced a cognitive trade off, not necessarily that Americans were opposed to spending on Health but saw Defense as a higher priority (Wlezien 2004).

More poignantly, around 2008 the Great Recession took place and discussions about the Affordable Care Act (ACA) populated the public sphere and among candidate rhetoric (Lenz 2009). Suddenly public support for health care spending plummeted. Two new opinion objects that were not previously present shifted individuals' responding to surveys (Crano and Lyrantzis 2015) leading to an apparent retraction of Americans' long-standing majority support of the government providing health care⁷ (Legerski and Berg 2016; Wlezien 2017). Not only are these both context effects, but they likely have unique effects on preferences. The Great Recession effect should be 'global', causing a downward shift in preferences for government spending across most domains. The other effect is 'local' and should only pertain to health spending (Wlezien 2004)⁸. This is similar to what Morgan and Kang (2015) found using a longitudinal component of the GSS. Using similar counterfactual logic and OLS predictions I generate health spending preferences with the ACA-effect removed using the four questions from FactorA in Table 1. Figure 3 plots observed Health preferences (thick dashed line) versus predicted Health preferences (solid line).

Figure 3. Context-Corrected Latent Health Spending Preferences, GSS data 1973-2016

Next I test how well the corrected version of health spending preferences performs in the S&W thermostatic models in Figure 4. Model A is a replication of the original S&W model using data from 1973-

⁷ Gallup, "Healthcare System" In-Depth Topic, <http://news.gallup.com/poll/4708/healthcare-system.aspx>, accessed 1/16/2018

⁸ The Gallup question on the role of the government to ensure that all Americans have health coverage shows a similar dip suggesting that the opinion shift relates to the discussions and then passage of the ACA (see footnote 5).

2005; Model B is an extension of this original model using GSS data through 2016; and Model C is the 1973-2016 time-series using the predicted values for Health (from Figure 3).

Figure 4. Thermostatic Feedback of Public Preferences and Government Spending on Health

NOTE: all coefficients sig. at $p < 0.05$ except the unshaded bar where policy representation is not significantly different from zero ($p > 0.2$) in the 'Extended Model'.

Figure 4 demonstrates that the original policy representation findings of S&W ("Original Model" above) do not hold when extended to the period 2006-2016 ("Extended Model" with an insignificant effect for policy representation). However, counterfactual logic suggests a context-bias in attempts to measure public preferences as demonstrated in Figure 3. Using a corrected measure for this, the extended time series now conforms to the original findings and theoretical expectations of thermostatic feedback ("Context-Corrected Model").

I am not arguing here that this type of solution is the only one suited for shocks or context-bias in time-series, but it provides another example where thinking like a structural equation modeler provides improvements for policy studies scholars interested in public opinion. Moreover, the technique of taking a predicted value of a variable from a simple measurement model is something known to a wide policy research audience, as opposed to highly technical statistical methods which often dominate discussions of context-bias (de Jong and Penzer 1998).

4 ENDOGENEITY OF PUBLIC OPINION AND SOCIAL POLICY

In this section, I start with a discussion of endogeneity and provide theoretical arguments for its presence in opinion and policy research. Next, I focus on S&W's thermostatic models of preferences and spending as an example of opinion-policy research offering a proof of why endogeneity is probably present there as well, and finally I test endogeneity among the S&W models using SEM.

4.1 *The problem of endogeneity*

In formal language, X must be exogenous to Y for accurate estimation of a statistical ‘effect’ of X on Y . Regardless of the wording (‘affects’, ‘effects’, ‘leads to a change in’, ‘predicts’, ‘conditions’, ‘impacts’ and etc.), the logic behind X and Y is that X causes or transmits several causes to Y . To convert a hypothesis that X causes Y into a meaningful statistical model such as a regression, a researcher needs to logically deduce that Y does not also cause X and that nothing else about the observed relationship of X and Y is a product of some third, unobserved variable. If Y causes X or if there is an omitted shared cause, there is a problem of endogeneity.

In statistical modelling, understanding endogeneity necessitates understanding error. The error term in a statistical model is everything about a dependent variable that the model cannot explain. In the familiar case of ordinary least-squares (OLS) regression this is known as the residual. The residual is everything left after taking into account the independent variables’ combined capacity to predict the dependent variable. This means the residual is orthogonal or exogenous to these independent variables; everything left unexplained about the dependent variable. Endogeneity means something is lurking in the real-world that causes one or more of the independent variables to be correlated with the residual, thus violating the exogeneity principle. This correlation could be a product of reverse causality or omitted cause (Pearl 2009), something also known as a directionality problem (Howlett and Cashore 2009).

The only method for identifying whether X does not correlate with the residual is logic. If I run an OLS regression, X will not correlate with the error term by mathematical design, but whether this model specification accurately reflects the real world can only be inferred through logic and perhaps the help of statistical or graphical tools to support this logic (Elwert 2013; Ketokivi and McIntosh 2017). The question is: *are regressions of opinion and policy free of endogeneity?* If each causes the other, then the simple answer is no. Declaring either one as an independent variable when it is in fact caused by the other would violate the exogeneity assumption.

Soroka and Wlezien are only one example among the plethora of scholars arguing for various types of reciprocal causal effects running between opinion and policy (Brezna 2017; Jones and Baumgartner 2004; May and Jochim 2013; Soroka and Wlezien 2010; Wlezien 2004). For example, public opinion shifts in response to changes in policies, or when lack of desired change persists and the public become increasingly frustrated (Campbell 2012; Pierson 1993; Shapiro 2011). Policies impact the distribution of resources (Kenworthy 1999) and the street-level experiences of individuals, both leading to public opinion changes (Lerman and McCabe 2017). Thus, any shift in public opinion toward a given policy does not take place independently of that policy and the policymaking leading up to the implementation of that policy. As much as opinion causes policy (Key 1961), policy also causes opinion. Opinion and policy constitute a feedback loop or cycles of loops and this is

a theoretical reason to suspect models of opinion and policy are at risk of endogeneity bias (see discussion in Breznau 2018, 268)⁹.

4.2 The thermostatic models

The thermostatic model approximates a feedback process based on theories of causality, but S&W are careful not to claim causal inference. They use the term “causal” only once in their book *Degrees of Democracy* (Soroka and Wlezien 2010) on p. 38 to describe the results of the thermostatic feedback model as outcomes of underlying “causal connections”. Again, regardless of language, all coefficients and significance tests in the thermostatic model, and any statistical model for that matter, are only interpretable and accurate with respect to reality *if the model specification is correct* (Greenland et al. 2016). Correct specification is a defining characteristic of a t-test (and subsequent p-value) and a regression coefficient, else these statistics are vacuous numerical symbols. *Correct model specification means that endogeneity is not present*. So whether policy scholars talk about causality or not, their results are only useful if they are drawn from models free from endogeneity (among other problems of confounding, suppression, outlier bias, sampling error, measurement error, collider bias, and more) (Blalock 1964; Elmore 1987; Pearl 2009; Steinberg 2007).

Equations make it much easier to demonstrate how endogeneity might bias the results of thermostatic modeling as described in the abstract in Section 2.2. Written as separate equations the thermostatic models are the following:

$$R_t = a + \beta_1 P_t + \beta_2 W_t + e_t \quad (1)$$

Eq. 1 captures *public responsiveness* where relative preferences R_t for spending in a particular policy area equals an effect of β_1 (i.e., coefficient) for spending P_t plus effect β_2 of the public’s absolute preferences W_t plus an intercept a and error term e_t . The t refers to time, measured in units of one year. Eq. 1 assumes relative preferences are observed in a survey and represent the difference between existing levels of spending P_t and the public’s absolute preferences for spending W_t . However, absolute preferences W_t are unobserved and after years of research and various testing strategies, S&W concluded that a parsimonious instrumentation is to take relative preferences lagged by one year R_{t-1} and include a yearly counter for spending that is squared (C^2), to account for major social spending expansion from the 1970s to the 2000s, in lieu of W_t . After making this substitution (see S&W 2010: p. 92 for Eq. 2 and p. 129 for Eq. 3), Eq. 1 becomes:

$$R_t = a + \beta_1 P_t + \beta_2 R_{t-1} + \beta_3 C^2 + e_t \quad (2)$$

⁹ A relevant point that falls beyond the scope herein is that opinion-policy endogeneity is also a problem created by path dependency (Pierson 2004), something analogous to problems of autocorrelation. Sometimes an opinion or policy influences itself over time, something also known as self-reinforcing or self-undermining feedback (A. M. Jacobs and Weaver 2015).

The other side of the feedback loop, *policy representation*, where a change in P_t equals the coefficient y_1 of relative preferences in the previous year R_{t-1} plus the slope y_2 of partisan control of government in the previous year G_{t-1} and an intercept ρ and error term u_t appears in Eq. 3:

$$\Delta P_t = \rho + y_1 R_{t-1} + y_2 G_{t-1} + u_t \quad (3)$$

For Eq's. 2 and 3 to capture feedback of opinion and spending over time, the error terms e_t and u_t must be exogenous to their respective equations; this is a simple principle of exogeneity. However, as the two equations are not independent and simultaneously describe the same real world thing, here the United States since 1973, the two error terms also must not correlate with one another across equations (Paxton, Hipp, and Marquat-Pyatt 2011). Here is the proof: Consider that e_t in Eq. 2 has a correlation of exactly zero with P_t (and all other independent variables), while u_t must have a correlation with ΔP_t (the residual by definition is what is left of the dependent variable). But if u_t correlates with e_t , then e_t likely correlates with P_t because it is part of ΔP_t (defined as $\Delta P_t = P_t - P_{t-1}$). If e_t correlates with P_t then endogeneity threatens the reliability of any estimates in Eq. 2¹⁰. The question is then: does u_t correlate with e_t ? There is ample theory on reciprocal causality that suggests it does, but now I ask the question empirically.

4.3 Testing endogeneity

I treat S&W's models as correctly specified, i.e., that no other independent variables are necessary to recover thermostatic effects. Based on this assumption, I can test for endogeneity by running their two regression equations simultaneously. To do this I shift from observed error to latent error modeling. For Eq.'s 2 and 3 to produce coefficients that are useful and reliable as representing something in the real world, e_t and u_t must not be correlated. Using SEM practices I can constrain e_t and u_t to correlate at zero – thus I define the structure (the 'S' in SEM) of the error relationship in the model rather than letting the model freely identify it. In simultaneous equation estimation there are two potentially interrelated error terms, thus I cannot estimate OLS

¹⁰ An explanation without math: Current spending and the absolute preferences of the public combine to cause current public preferences for spending as measured in the GSS data (i.e., 'relative preferences' for whether it should be increased or decreased and how much), but these two causes alone cannot explain everything about relative preferences. There are always other factors. Thus, there are parts of relative preferences that are residual because they are not fully explained by the theory. This is fine and normal so long as these residuals are not actually part of the theory in disguise! If these residuals actually contain causes of current spending then whatever these causes are should be explicitly identified in the theoretical description of current spending, not in the 'left overs' which go in the residual. If not, the theory would *de facto* be arguing something along the lines of, 'current spending causes relative preferences *twice*', once as an observed phenomenon and again as an observed phenomenon that the theorist pretends is unobserved and put into the error term. This is not logical, observed spending is either observed and therefore not residual, or it is unobserved and therefore residual. On its own as an observed object in the world it cannot be both.

because this depends on one equation with one residual. Instead this calls for maximum likelihood estimation involving iterations of the software refining results until they best fit both error terms to both equations¹¹.

Focusing only on the models analyzed in this article, Table 3 presents tests of endogeneity for Models A (original S&W Social model, 1973-2005, see Table 2), B (the SEM-inspired Social model, 1973-2005, see Table 2), C (original S&W Health model, 1973-2005, see Figure 4) and E (context-adjusted Health, 1973-2016, see Figure 4). For each model I estimate first a “freely estimated” version where the residuals e_t and u_t freely correlate, and then below that a “constrained” model where e_t and u_t are forced to have zero correlation. Then I provide the size of the correlation when freely estimated, the values of chi-square, the model degrees of freedom (also known as parameters) and a test of equal fit comparing the free and constrained version of each model. The test of equal fit’s p-value indicates whether the constrained model fits the data significantly worse than the freely estimated model.

Table 3. Testing Thermostatic Feedback Models for Endogeneity

Model ^a	Residual Covariance	$\text{corr}(e_t, u_t)$	chi2	df	equal fit test
					p-value ^b
A (Social, S&W original)	freely estimated	0.340	473.0	5	0.102
	constrained	0.000	475.6	6	
B (Social, SEM)	freely estimated	-0.525	98.5	5	0.002
	constrained	0.000	108.3	6	
C (Health, S&W original)	freely estimated	0.338	520.4	8	0.049
	constrained	0.000	524.3	9	
E (Health, context-adjusted)	freely estimated	-0.042	736.9	8	0.800
	constrained	0.000	737.0	9	

^aSee Online Appendix, Table 2 and Figure 4 for full model results.

^bTests if freely estimated model fits the data significantly better than constrained model with $\text{corr}(e_t, u_t) = \text{zero}$.

If the constrained model fits the data worse, this calls into question estimation of public responsiveness and policy representation equations as separate regressions. This is the results in Models B and C ($p < 0.05$). The size and direction of the residual correlations raises additional questions. Although Model A does not fail the equal fit test, a correlation of 0.340 is not-negligible in size. Given that there are only 32 cases in the data, the chance of recovering a true effect of 0.340 at $p < 0.05$ has a power of only 0.617 (see Appendix B). Therefore, any correlation of 0.340 may or may not have a significant biasing impact at p-values up to nearly 0.40. These results, just like the coefficients in the models, need to be interpreted with caution to say the least. Moreover, if the correlation is negative, as with Models B and E, it suggests that when the error in public responsiveness is

¹¹ To help clarify for readers: If I do not constrain the correlation of the error terms nor constrain any other part of the model, maximum likelihood and OLS should be identical except for tiny potential rounding errors. It is only when forcing the correlation to zero that the software must recalculate the regression coefficients so that the other relationships in the model are preserved as closely as possible (i.e., that the zero-constrained implied covariance matrix is as close as possible to the real-world observed covariance matrix).

higher, the error in policy representation is lower, ultimately suggesting that the errors push toward each other and keep the system at equilibrium. However, in Models A and C the errors are positive and this means that when the error is higher in public responsiveness it is also higher in policy representation and this could lead the estimates in the set of equations to systematically expand, or ‘explode’ (Kline 2013); however, as the US context is relatively stable with spending and opinion following reasonable rather than extreme levels, a better explanation is probably that some unobserved variable systematically biases both sides of the equation in the same direction. This suggests that the assumption that models were correctly specified necessary to conduct the endogeneity tests may be incorrect.

5 DISCUSSION

The research presented herein involves three sets of findings representing three ways that structural equation modeling can improve research, or at least improve the questions asked, by policy studies researchers interested in public opinion. Focusing on the state of the art work of Soroka and Wlezien and the well-known *General Social Survey* of the United States, I demonstrated how attention to measurement using factor analysis and theories of public opinion as a latent variable improves the reliability if not results of research. I demonstrated how shocks might bias public opinion surveys and how SEM-logic can overcome some of this bias. Finally, I used theories of reciprocal causality between public opinion and social policy plus formal logic to deduce that any single, one-way analysis of opinion or policy as a dependent variable may suffer from endogeneity bias; and offered one way to test for such a bias in the case of the thermostatic models.

A central component of the SEM approach is consideration of unobserved things like attitudes, public opinion, public mood and other features of individuals and societies. Some policy categories such as social policies include several domains of spending, and therefore combining several questions to measure latent public preferences provides evidence for how to best construct these latent measures. For example, Table 1 demonstrates that not all GSS questions should be given equal weight when constructing a latent scale due to divergent factor loadings. As echoed among SEM practitioners, we should respect the evidence presented in our analyses (Hayduk 2014). Thus, a mean scale in this case is a model misspecification. To be clear, there is a large range of ways to deal with measurement error and the approach taken herein is only one of many. For example, other scholars suggest that GSS survey response categories may not be equidistant suggesting a binomial rather than continuous link function (J. H. Evans, Bryson, and DiMaggio 2001).

The practice of SEM is one concerned with how well a theory explains observed data. As I demonstrated in Table 1 and Figure 1, modeling public opinion toward social spending does not fit the data well when including the GSS question asking about spending more or less on welfare. This is a reminder that welfare Americans loathe “welfare” (Gilens 2009). There are of course unique attitudes associated with each domain (e.g., Space, Defense and Welfare), meaning that each could potentially stand alone, but welfare is particularly unique. Unfortunately, a measurement ideal where a survey fields several questions per-domain does not exist in current nationally-representative time series fielded annually or biennially, neither in the US nor elsewhere.

Despite the idea of unique domains, evidence that social spending is a consistent umbrella concept among the US public (at least from 1973-2005) appeared in Table 1. This is echoed in other research on the government's role in providing socially responsible policies or what some refer to as attitudes toward the "social welfare" state (Hasenfeld and Rafferty 1989; Krosnick 1988; Smith 1987). This points back to arguments made by S&W in their work about global spending preferences and responsiveness across all government domains, and more peculiar or particular local preferences where policymakers may be particularly responsive at certain times and places as with welfare (Soroka and Wlezien 2010, 121–23; Wlezien 2004); thus offering consistent logic aligning with the evidence.

A methodological concern among this particular example is the number of cases. The evidence suggests that the preferred measurement of social spending preferences does not show significant public responsiveness to social spending (Table 2), but there are only 33 observations since 1973. With multiple regressors, small unobserved errors could easily change confidence intervals to cross zero (Brezna 2016). If I relax the significance level in Table 2 to $p < 0.10$, then the responsiveness effect returns that was 'lost' as a result of using the empirically justified measure of social preferences as opposed to S&W's original sum scale. Lessons of the open science movement suggest that scholars should pay less attention to significance levels and focus instead on size of and confidence in effects (Leek et al. 2017). For example, the 'lost' responsiveness coefficient ranges anywhere from zero to -0.18 (the estimated standardized coefficient) and up to -0.36. Thus, most of the possible range is in line with theoretical expectations. Small-N studies carry inherent unreliability that SEM-logic cannot solve, and I would not yet conclude there is no responsiveness in this domain. However, the practice of model fitting demonstrates that the models do not fit the data particularly well, and this may be an indicator of a small-N problem, among others.

A byproduct of the evidence from measurement modeling of social preferences, is the possibility to use questions from other social domains to improve measurements in a single social domain, as I did with health spending in Figure 4. This offers a potential to correct shocks or context-specific effects that may cause sudden shifts in responses to survey questions. In Figures 2 and 3, I demonstrated that American preferences for health spending were suddenly biased because of the discussions of the ACA in the public sphere. The GSS question acted as a heuristic cue to engage attitudes about the ACA, not just health spending. This theory held merit empirically. When extending S&W's original thermostatic model of health spending to include the period 2006-2016, the policy responsiveness coefficient was small and not significantly different from zero (see Figure 4). However, after correcting public opinion for the ACA shock the expected finding replicating S&W's original findings (re)appeared in the extended data. This offers evidence of the size of the impact these events had on public opinion and the great utility of context-corrections. In a related vein, different policy domains, and perhaps different time periods, may lead to heterogeneous relationships of opinion and spending, offering another form of context to consider in these matters (Wlezien 2004). For this SEM logic suggests latent growth models or interaction models, similar to those presented in Wlezien and Soroka (2012) where effects depend on the type of political institutions.

In what is ultimately a theoretical exercise, I used model fit to demonstrate how to test for endogeneity. The logic of SEM comes from theory and attention to correct model specification. In other words, it is an overarching approach to science that aims to best translate theory into models, and then test these models against all possible other models or against specific hypothetical alternatives. This is the essence of model fitting and model testing using chi-square tests among others (Hayduk 2018; Kline 2016) it is also a basic tenet of causal inference and counterfactual logic which has its own universally understandable language of path diagrams or DAGs (directed acyclical graphs) (Dowd 2011; Elwert 2013; Knight and Winship 2013; Pearl 2009). In the thermostatic feedback models there is ample theory to believe that opinion and spending are reciprocally causal. This suggests a risk of endogeneity bias, and like any opinion-policy related model, the thermostatic models are not likely free from this bias. Table 3 reveals that this risk is real as two out of four models show significant ill-fit as a result of endogeneity, and three out of four demonstrate sizeable residual correlations between public responsiveness and policy representation. This is not an indictment of the thermostatic models, but a finding that should raise awareness among all opinion and policy scholars regardless of whether they study spending, roll-call voting, specific policy moments and etc. Opinion and policy are reciprocally causal and any unidirectional attempt to recover an effect between them may be biased by this. These findings bring up questions and ideas for how to take the thermostatic models forward in policy studies research.

In fact, the results of the models suffering from endogeneity bias retain similar results in terms of size and direction as the original S&W models. This suggests not the findings coming into question, but the models. Astute SEM researchers can quickly calculate from the results in Table 3 that none of the models pass an exact fit chi-square test at $p > 0.05$. This means that there is evidence that there are some problems with the models, that they do not fit the observed data particularly well. The models may be misspecified. This may be the source of apparent endogeneity or ill-fit bias. Again, without correct model specification statistical estimates are of no use to theory or practice (Elmore 1987; Steinberg 2007). The chi-square test is a global assessment of the models so it does not reveal exactly what about them seems to cause the troubles. Future research should theoretically explore different model specifications and then again testing if they suffer from endogeneity. For this the SEM toolkit provides many useful tools and it is up to us policy studies scholars to use them such as alternative approaches to estimating systems of reciprocal effect and paying attention to latent variables or latent errors such as autoregression with error correction (Jennings 2013), simultaneous feedback (Brezna 2018) and latent variable cross-lagged models (Allison, Williams, and Moral-Benito 2017).

6 REFERENCES

- Achen, Christopher H. 1975. "Mass Political Attitudes and the Survey Response." *The American Political Science Review* 69(4): 1218–31. <http://www.jstor.org/stable/1955282>.
- Allison, Paul D, Richard Williams, and Enrique Moral-Benito. 2017. "Maximum Likelihood for Cross-Lagged Panel Models with Fixed Effects." *Socius* 3: 1–17. <https://doi.org/10.1177/2378023117710578>.
- Andrews, Frank M. 1984. "Construct Validity and Error Components of Survey Measures: A Structural Modeling Approach." *Public Opinion Quarterly* 48(2): 409–42.
- Ansolabehere, Stephen, Jonathan Rodden, and James M. Snyder Jr. 2008. "The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting." *American Political Science Review* 102(2): 215–32.
- Béland, Daniel, and Alex Waddan. 2017. "Why Are There No Universal Social Programs in The United States: A Historical Institutional Comparison with Canada." *World Affairs* 180(1): 64–92. <https://doi.org/10.1177/0043820017715570>.
- Bielby, William T, and Robert M Hauser. 1977. "Structural Equation Models." *Annual Review of Sociology* 3: 137–61. <http://www.jstor.org/stable/2945933>.
- Blalock, Hubert M. 1964. *Causal Inferences in Nonexperimental Research*. Chapel Hill, NC: The University of North Carolina Press.
- de Boer, Connie. 2007. "Public Opinion." In *The Blackwell Encyclopedia of Sociology*, ed. George Ritzer. Malden, MA: Blackwell Publishers, 3711–15.
- Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York, NY: John Wiley & Sons.
- Bollen, Kenneth A, and Judea Pearl. 2013. "Eight Myths about Causality and Structural Equation Models." In *Handbook of Causal Analysis for Social Research*, ed. Stephen L. Morgan. Dordrecht, Netherlands: Springer Science & Business Media, 301–28.
- Breznau, Nate. 2016. "Secondary Observer Effects: Idiosyncratic Errors in Small-N Secondary Data Analysis." *International Journal of Social Research Methodology* 19(3): 301–18.
- . 2017. "Positive Returns and Equilibrium: Simultaneous Feedback Between Public Opinion and Social Policy." *Policy Studies Journal* 45(4): 583–612. <http://doi.wiley.com/10.1111/psj.12171>.
- . 2018. "Simultaneous Feedback Models with Macro-Comparative Cross-Sectional Data." *methods, data, analyses* 12(2): 265–308.
- Burstein, Paul. 1998. "Bringing the Public Back in: Should Sociologists Consider the Impact of Public Opinion on Public Policy?" *Social Forces* 77(1): 27–62.
- Butler, Daniel M, and David W Nickerson. 2011. "Can Learning Constituency Opinion Affect How Legislators Vote? Results from a Field Experiment." *Quarterly Journal of Political Science* 6(1): 55–83.
- Byrne, Barbara M. 2001. *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. New Jersey: Lawrence Erlbaum.
- Campbell, Andrea Louise. 2012. "Policy Makes Mass Politics." *Annual Review of Political Science* 15(1): 333–51. <http://dx.doi.org/10.1146/annurev-polisci-012610-135202>.
- Crano, William D, and Elena Lyrantzis. 2015. "Structure and Change of Complex Political Attitudes." In *Social Psychology and Politics*, eds. Joseph P Forgas, Klaus Fiedler, and William D Crano. New York and East Sussex: Routledge Psychology Press, 21–40.
- Davidov, Eldad et al. 2014. "Measurement Equivalence in Cross-National Research." *Annual Review of Sociology* 40(1): 55–75.
- Diamantopoulos, Adamantios et al. 2012. "Guidelines for Choosing between Multi-Item and Single-Item Scales for Construct Measurement: A Predictive Validity Perspective." *Journal of the Academy of Marketing Science* 40(3): 434–49. <https://doi.org/10.1007/s11747-011-0300-3>.
- Dowd, Bryan E. 2011. "Separated at Birth: Statisticians, Social Scientists, and Causality in Health Services

- Research.” *Health Services Research* 46(2): 397–420.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3064910/>.
- Elmore, Richard F. 1987. “Instruments and Strategy in Public Policy.” *Review of Policy Research* 7(1): 174–86. <https://doi.org/10.1111/j.1541-1338.1987.tb00036.x>.
- Elwert, Felix. 2013. “Graphical Causal Models.” In *Handbook of Causal Analysis for Social Research*, ed. Stephen L. Morgan. Dordrecht, The Netherlands: Springer Science & Business Media, 245–72.
- Evans, J H, B Bryson, and P DiMaggio. 2001. “Opinion Polarization: Important Contributions, Necessary Limitations.” *American Journal of Sociology* 106(45762): 944–59.
- Evans, M D R, and Jonathan Kelley. 2008. “Trends in Women’s Labor Force Participation in Australia: 1984–2002.” *Social Science Research* 37(1): 287–310.
<http://www.sciencedirect.com/science/article/pii/S0049089X07000051>.
- Fabrigar, Leandre R, Duane T Wegener, Robert C MacCallum, and Erin J Strahan. 1999. “Evaluating the Use of Exploratory Factor Analysis in Psychological Research.” *Psychological methods* 4(3): 272.
- Gilens, Martin. 2009. *Why Americans Hate Welfare: Race, Media, and the Politics of Antipoverty Policy*. Chicago, IL: The University of Chicago Press.
- Greenland, Sander et al. 2016. “Statistical Tests, P Values, Confidence Intervals, and Power: A Guide to Misinterpretations.” *European journal of epidemiology* 31(4): 337–50.
<https://www.ncbi.nlm.nih.gov/pubmed/27209009>.
- Groves, Robert M, and Lars Lyberg. 2010. “Total Survey Error: Past, Present, and Future.” *Public Opinion Quarterly* 74(5): 849–79. <http://dx.doi.org/10.1093/poq/nfq065>.
- Hasenfeld, Yeheskel, and Jane A Rafferty. 1989. “The Determinants of Public Attitudes toward the Welfare State.” *Social Forces* 67(4): 1027–48.
- Hayduk, Leslie A. 2014. “Shame for Disrespecting Evidence: The Personal Consequences of Insufficient Respect for Structural Equation Model Testing.” *BMC Medical Research Methodology* 14(1): 124.
<http://dx.doi.org/10.1186/1471-2288-14-124>.
- . 2018. “Review Essay on Rex B. Kline’s Principles and Practice of Structural Equation Modeling.” *Canadian Studies in Population* 45(3–4).
- Heckman, James J. 2010. “Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy.” *Journal of Economic Literature* 48(2): 356–98.
<http://www.aeaweb.org/articles?id=10.1257/jel.48.2.356>.
- Herbst, Susan. 1992. “Surveys in the Public Sphere: Applying Bourdieu’s Critique of Opinion Polls.” *International Journal of Public Opinion Research* 4(3): 220–29.
- Holbert, R Lance, and Michael T Stephenson. 2006. “Structural Equation Modeling in the Communication Sciences, 1995–2000.” *Human Communication Research* 28(4): 531–51.
<https://dx.doi.org/10.1111/j.1468-2958.2002.tb00822.x>.
- Howlett, Michael, and Benjamin Cashore. 2009. “The Dependent Variable Problem in the Study of Policy Change: Understanding Policy Change as a Methodological Problem.” *Journal of Comparative Policy Analysis: Research and Practice* 11(1): 33–46. <https://doi.org/10.1080/13876980802648144>.
- Huckfeldt, Robert. 2007. “Unanimity, Discord, and the Communication of Public Opinion.” *American Journal of Political Science* 51(4): 978–95. <http://dx.doi.org/10.1111/j.1540-5907.2007.00292.x>.
- Irwin, G A, and J J M Van Holsteyn. 2002. “According to the Polls - The Influence of Opinion Polls on Expectations.” *Public Opinion Quarterly* 66(148): 92–104.
- Jacobs, Alan M, and R Kent Weaver. 2015. “When Policies Undo Themselves: Self-Undermining Feedback as a Source of Policy Change.” *Governance* 28(4): 441–57. <http://dx.doi.org/10.1111/gove.12101>.
- Jacobs, Lawrence R. 1992. “Institutions and Culture: Health Policy and Public Opinion in the U.S. and Britain.” *World Politics* 44(2): 179–209.
- Jennings, Will. 2013. “Error-Correction as a Concept and as a Method: Time Series Analysis of Policy-Opinion Responsiveness.” In *Political Science Research Methods in Action*, eds. Michael Bruter and

- Martin Lodge. London: Palgrave Macmillan UK, 203–28. https://doi.org/10.1057/9781137318268_10.
- Jones, Bryan D, and Frank R Baumgartner. 2004. “Representation and Agenda Setting.” *Policy Studies Journal* 32(1): 1–24. <http://dx.doi.org/10.1111/j.0190-292X.2004.00050.x>.
- de Jong, Piet, and Jeremy Penzer. 1998. “Diagnosing Shocks in Time Series.” *Journal of the American Statistical Association* 93(442): 796–806.
- Jöreskog, Karl G. 2007. “Factor Analysis and Its Extensions.” In *Factor Analysis at 100: Historical Developments and Future Directions*, eds. Robert Cudeck and Robert C MacCallum. Mahwah and London: Lawrence Erlbaum Associates, 47–76.
- Judd, Charles M, and Jon A Krosnick. 1989. “The Structural Bases of Consistency among Political Attitudes: Effects of Political Expertise and Attitude Importance.” In *Attitude Structure and Function*, eds. Anthony R. Pratkanis, Steven J. Breckler, and Anthony G. Greenwald. New York and London: Psychology Press Taylor & Francis Group, 99–128.
- Kang, Shin-Goo, and G Bingham Powell. 2010. “Representation and Policy Responsiveness: The Median Voter, Election Rules, and Redistributive Welfare Spending.” *The Journal of Politics* 72(4): 1014–28. <https://doi.org/10.1017/S0022381610000502>.
- Kenworthy, Lane. 1999. “Do Social-Welfare Policies Reduce Poverty? A Cross-National Assessment.” *Social Forces* 77(3): 1119–39. <http://sf.oxfordjournals.org/content/77/3/1119.abstract>.
- Ketokivi, Mikko, and Cameron N McIntosh. 2017. “Addressing the Endogeneity Dilemma in Operations Management Research: Theoretical, Empirical, and Pragmatic Considerations.” *Journal of Operations Management* 52: 1–14. <http://www.sciencedirect.com/science/article/pii/S0272696317300293>.
- Key, V O. 1961. *Public Opinion and American Democracy*. New York: Knopf.
- Kline, Rex B. 2013. “Reverse Arrow Dynamics: Feedback Loops and Formative Measurement.” In *Structural Equation Modeling: A Second Course*, eds. Gregory R. Hancock and Ralph O. Mueller. Charlotte, NC: Information Age Publishing, 41–79.
- Kline, Rex B. 2016. *Principles and Practice of Structural Equation Modeling*. 4th Editio. New York and London: Guilford Press.
- Knight, Carly R., and Christopher Winship. 2013. “The Causal Implications of Mechanistic Thinking: Identification Using Directed Acyclic Graphs (DAGs).” In *Handbook of Causal Analysis for Social Research*, Handbooks of Sociology and Social Research, ed. Stephen L Morgan. Springer Netherlands, 275–99. http://dx.doi.org/10.1007/978-94-007-6094-3_14.
- Krosnick, Jon A. 1988. “Attitude Importance and Attitude Change.” *Journal of Experimental Social Psychology* 24(3): 240–55. <http://www.sciencedirect.com/science/article/pii/0022103188900388>.
- Larsen, Erik Gahner. 2018. “Policy Feedback Effects on Mass Publics: A Quantitative Review.” *Policy Studies Journal* 0(0). <https://doi.org/10.1111/psj.12280>.
- Leek, Jeff et al. 2017. “Five Ways to Fix Statistics.” *Nature* 551(Nov 28): 557–59.
- Legerski, Elizabeth Miklya, and Justin Allen Berg. 2016. “Americans’ Approval of the 2010 Affordable Care Act: Self-Interest and Symbolic Politics*.” *Sociological Inquiry* 86(3): 285–300. <http://dx.doi.org/10.1111/soin.12121>.
- Lenz, Gabriel S. 2009. “Learning and Opinion Change, Not Priming: Reconsidering the Priming Hypothesis.” *American Journal of Political Science* 53(4): 821–37. <http://dx.doi.org/10.1111/j.1540-5907.2009.00403.x>.
- Lerman, Amy E, and Katherine T McCabe. 2017. “Personal Experience and Public Opinion: A Theory and Test of Conditional Policy Feedback.” *The Journal of Politics* 79(2): 624–41. <https://doi.org/10.1086/689286>.
- Mackenzie, Scott. 2001. “Opportunities for Improving Consumer Research through Latent Variable Structural Equation Modeling.” *Journal of Consumer Research* 28(1): 159–66. <http://www.jstor.org/stable/10.1086/321954>.
- May, Peter J, and Ashley E Jochim. 2013. “Policy Regime Perspectives: Policies, Politics, and Governing.”

- Policy Studies Journal* 41(3): 426–52. <http://dx.doi.org/10.1111/psj.12024>.
- Morgan, Stephen L., and Minhyoung Kang. 2015. “A New Conservative Cold Front? Democrat and Republican Responsiveness to the Passage of the Affordable Care Act.” *Sociological Science* 2(24): 502–26.
- Mullinix, Kevin J. 2011. “Lingering Debates and Innovative Advances: The State of Public Opinion Research.” *Policy Studies Journal* 39(s1): 61–76. http://dx.doi.org/10.1111/j.1541-0072.2010.00389_5.x.
- Nie, Norman H, John R Petrocik, and Sidney Verba. 1999. *The Changing American Voter*. Enlarged E. San Jose, CA: Harvard University Press.
- Papadakis, Elim. 1992. “Public Opinion, Public Policy and the Welfare State.” *Political Studies* 40(1): 21–37.
- Paxton, Pamela, John R. Hipp, and Sandra Marquat-Pyatt. 2011. *Nonrecursive Models: Endogeneity, Reciprocal Relationships and Feedback Loops*. London: Sage Publications, Ltd.
- Pearl, Judea. 2009. *Causality: Models, Reasoning and Inference*. 2nd ed. New York, NY, USA: Cambridge University Press.
- Perrin, Andrew J, and Katherine McFarland. 2011. “Social Theory and Public Opinion.” *Annu. Rev. Sociol.* 37(1): 87–107. <http://dx.doi.org/10.1146/annurev.soc.012809.102659>.
- Pierson, Paul. 1993. “When Effect Becomes Cause: Policy Feedback and Political Change.” *World Politics* 45(4): 595–628.
- . 2004. “Positive Feedback and Path Dependence.” In *Politics in Time: History, Institutions and Social Analysis*, Princeton, NJ: Princeton University Press, 17–53.
- Rasinski, Kenneth A. 1989. “The Effect of Question Wording on Public Support for Government Spending.” *Public Opinion Quarterly* 53(3): 388–94. <https://dx.doi.org/10.1086/269158>.
- Ruelens, Anna, Bart Meuleman, and Ides Nicaise. 2018. “Examining Measurement Isomorphism of Multilevel Constructs: The Case of Political Trust.” *Social Indicators Research* 140(3): 907–27. <https://doi.org/10.1007/s11205-017-1799-6>.
- Schmitt-Beck, R. 1996. “Mass Media, the Electorate, and the Bandwagon. A Study of Communication Effects on Vote Choice in Germany.” *International Journal of Public Opinion Research* 8(3): 266–91.
- Shapiro, Robert Y. 2011. “Public Opinion and American Democracy.” *Public Opinion Quarterly* 75(5): 982–1017.
- Smith, Tom W. 1987. “That Which We Call Welfare by Any Other Name Would Smell Sweeter. An Analysis of the Impact of Question Wording on Response Patterns.” *Public Opinion Quarterly* 51(1): 75–83. <https://dx.doi.org/10.1086/269015>.
- . 1994. “Is There Real Opinion Change?.” *International Journal of Public Opinion Research* 6(2): 187–203. <http://dx.doi.org/10.1093/ijpor/6.2.187>.
- Smith, Tom W, Michael Davern, Jeremy Freese, and Michael Hout. 2018. *General Social Surveys, Cumulative Codebook 1962-2016*. Chicago. http://gss.norc.org/documents/codebook/GSS_Codebook.pdf.
- Soroka, Stuart N, and Christopher Wlezien. 2005a. “Opinion-Policy Dynamics: Public Preferences and Public Expenditure in the United Kingdom.” *British Journal of Political Science* 35(4): 665–89.
- . 2005b. “Opinion Representation and Policy Feedback: Canada in Comparative Perspective.” *Canadian Journal of Political Science* 37(3): 531–59.
- . 2010. *Degrees of Democracy: Politics, Public Opinion and Policy*. Cambridge: Cambridge University Press.
- Splichal, Slavko. 1999. *Public Opinion: Developments and Controversies in the Twentieth Century*. Oxford: Rowman & Littlefield Publishers.
- Steensland, Brian. 2006. “Cultural Categories and the American Welfare State: The Case of Guaranteed Income Policy.” *American Journal of Sociology* 111(5): 1273–1326.
- Stegmueller, Daniel. 2011. “Apples and Oranges? The Problem of Equivalence in Comparative Research.”

- Political Analysis* 19(4): 471–87. <https://www.cambridge.org/core/article/apples-and-oranges-the-problem-of-equivalence-in-comparative-research/8CD1F2F8E41FEABB231F4AE303EF8A51>.
- Steinberg, Paul F. 2007. “Causal Assessment in Small-N Policy Studies.” *Policy Studies Journal* 35(2): 181–204. <https://doi.org/10.1111/j.1541-0072.2007.00215.x>.
- Stimson, James A. 2018. *Public Opinion in America: Moods, Cycles, and Swings*. New York, NY: Routledge.
- Stimson, James A, Michael B Mackuen, and Robert S Erikson. 1995. “Dynamic Representation.” *The American Political Science Review* 89(3): 543–65. <http://www.jstor.org/stable/2082973>.
- Tabachnick, Barbara G, and Linda S Fidell. 2007. *Using Multivariate Statistics*. New York: Allyn & Bacon/Pearson Education.
- Tay, Louis, Sang Eun Woo, and Jeroen K Vermunt. 2014. “A Conceptual and Methodological Framework for Psychometric Isomorphism: Validation of Multilevel Construct Measures.” *Organizational Research Methods* 17(1): 77–106. <https://doi.org/10.1177/1094428113517008>.
- Wlezien, Christopher. 1995. “The Public as Thermostat: Dynamics of Preferences for Spending.” *American Journal of Political Science* 39(4): 981–1000.
- . 1996. “Dynamics of Representation: The Case of US Spending on Defence.” *British Journal of Political Science* 26(1): 81–103. <https://www.cambridge.org/core/article/dynamics-of-representation-the-case-of-us-spending-on-defence/40711196B2FEF74C44B1EA982E1FFA7C>.
- . 2004. “Patterns of Representation: Dynamics of Public Preferences and Policy.” *The Journal of Politics* 66(1): 1–24.
- . 2017. “Public Opinion and Policy Representation: On Conceptualization, Measurement, and Interpretation.” *Policy Studies Journal* 45(4): 561–82. <http://dx.doi.org/10.1111/psj.12190>.
- Wlezien, Christopher, and Stuart N Soroka. 2012. “Political Institutions and the Opinion–Policy Link.” *West European Politics* 35(6): 1407–32. <http://dx.doi.org/10.1080/01402382.2012.713752>.
- Wolfe, Lee M. 2003. “The Introduction of Path Analysis to the Social Sciences, and Some Emergent Themes: An Annotated Bibliography.” *Structural Equation Modeling* 10(1): 1–34.

7 APPENDICES

Appendix A. GSS Source Question

1. First I would like to talk with you about some things people think about today. We are faced with many problems in this country, none of which can be solved easily or inexpensively. I'm going to name some of these problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount. First **(READ ITEM A)** . . . are we spending too much, too little, or about the right amount on **(ITEM)**? **READ EACH ITEM; CODE ONE FOR EACH.**

Too much	Too little	About right	DON'T KNOW
----------	------------	-------------	------------

A. The space exploration program

3 1 2 8

Appendix B. Power Analysis for a Correlation of 0.340 (from Table 3)