Income Inequality and Electoral Theories of Polarization

Dan Alexander[1] and Asya Magazinnik[2*]

[1]*Department of Political Science, University of Rochester, Rochester, NY, USA; dan.alexander@rochester.edu*
[2]*Department of Political Science, MIT, Cambridge, MA, USA; asyam@mit.edu*

ABSTRACT

Both the academic political science literature and the popular discourse are replete with narratives seeking to explain the concurrent rise of income inequality and legislative polarization over the past half century. We focus on a prominent subset of such accounts, which posit the faithful representation of polarizing constituencies as the key causal mechanism linking the two phenomena, and which we therefore refer to as "electoral theories of polarization." We show, however, that constructing a coherent, causal electoral theory of polarization is substantially more complicated than the literature has appreciated. First, we enumerate the necessary ingredients, with special emphasis on the importance of accounting for electoral geography. Second, we develop a causal framework for assessing the effect of income on polarization via a particular electoral channel, and we propose a set of estimation strategies that researchers may tailor to their particular model of how legislative ideology and partisanship are (co)determined. Third, we apply our framework to evaluate how well a model of self-interested "pocketbook voting" can explain patterns of polarization on the economic

dimension observed in the U.S. Senate from 1984 to 2018. We conclude that voters' private benefit from redistribution is unlikely to be a mechanism linking inequality to polarization.

Introduction

The growth of income inequality and the concurrent rise in legislative polarization have been two of the defining forces in the American economy and polity over the last half century. What is more, their coevolution over time has inspired an expansive literature attempting to causally link the two phenomena (Garand, 2010; McCarty *et al.*, 2006; Scheve and Stasavage, 2017). An important subset of such accounts — "bottom-up" explanations for congressional polarization — take polarizing legislators to be motivated by democratic responsiveness to the electorate.[1] Any bottom-up theory linking inequality to polarization has at its core a *theory of elections* that posits some mapping from voters' incomes to their political preferences to their representatives' ideology and partisanship in Congress. Accordingly, we will refer to this prominent class of theories — the focus of our paper — as *electoral theories* of inequality and polarization.

   On the face of it, electoral theories of polarization may be intuitively appealing. Economic voting is perhaps the most direct mechanism that logically connects citizens' incomes to legislative behavior in a democracy. Further, it seems reasonable to believe that a population with an unequal distribution of income will be faithfully represented by a polarized legislature, whereas a more equal society will exhibit less conflict — at least on the economic dimension of preferences.

   In this paper, we explain how this basic intuition may fail. More broadly, we show that the task of positing — much less testing — logically coherent electoral theories of polarization is substantially more complicated than the literature has appreciated. Challenges arise because inequality and polarization are defined over different units of analysis: the former summarizes the distribution of incomes across a population, while the latter summarizes the behavior of legislators representing distinct subpopulations. The analyst must therefore account for both the fixed institutional constraints and the more contested

---

[1]This is in contrast to "top-down" theories (see Layman *et al.*, 2006; Krasa and Polborn, 2014), in which parties and elites drive polarization that may or may not trickle down to the level of the mass public (Fiorina *et al.*, 2010; Iversen and Soskice, 2015; Lee, 2009).

features of democratic politics that structure the mapping from voters' incomes to legislative behavior.

Our work makes several contributions to the study of bottom-up theories of polarization — and, by extension, other phenomena that are connected by a theory of elections. First, we provide the necessary ingredients for a complete and coherent electoral theory linking inequality and polarization, some of which popular folk narratives and academic accounts alike have overlooked. Second, we develop a causal framework for evaluating such theories. Drawing on the Neyman-Rubin Causal Model (Rubin, 2005), we propose an estimand for the effect of changes in income on polarization, as well as a set of estimation strategies that researchers may tailor to their particular model of how legislative ideology and partisanship are (co)determined. We also discuss the conditions under which our approach identifies the causal effect of interest from observational data. Third, we apply our theoretical and statistical frameworks to a foundational, albeit highly stylized, electoral theory of inequality and polarization, one based on self-interested economic voting as formalized by Meltzer and Richard (1981) (henceforth, "Meltzer-Richard"). Our analysis shows that observed patterns of polarization over the past forty years are not likely to be explained by this mechanism.

The rest of the paper proceeds as follows. In Section "On Constructing Coherent Electoral Theories of Inequality and Polarization", we lay out the essential components of an electoral theory of inequality and polarization. First and foremost, we argue, it must incorporate the institutional rules that aggregate citizens into electoral districts, and representatives of electoral districts into a legislature. This inescapable feature of American politics can substantially distort any covariation one might naïvely expect to observe between national income inequality and polarization, even under a radically simple model of economic voting. Consequently, the focus on aggregate-level cotrends that pervades the literature on this topic is something of a red herring for the true causal relationship of interest.

Still, the relevant institutional structure provides some scaffolding that helpfully constrains the universe of plausible theories we may posit. But an electoral theory of polarization requires a host of additional assumptions that analysts must derive not from known institutional constraints, but from theories about representation, many of which are highly contested in the literature. Do candidates converge to the preferences of a decisive voter? Is the median voter decisive? Are politicians' platforms binding? Are political preferences formed on the basis of economic self-interest or are they based on social identities and environments? How does ideology translate into partisanship? Because an electoral theory must map voter income to congressional polarization via a multistage model of democratic politics, it is particularly sensitive to these flexible assumptions, whether or not they are explicitly stated by the analyst. Thus, our schema assists scholars in constructing coherent theories by ensuring

that they do not violate or ignore any fundamental assumptions, by tracking the myriad flexible assumptions they invoke, and by emphasizing the proper units of analysis with which to build a multi-stage model of the democratic process.

In Section "Building an Electoral Theory of Polarization from the Meltzer-Richard Model", we apply this theoretical framework to the canonical Meltzer-Richard model. After briefly explaining the logic of its key prediction, we show how one would properly extend this logic from one polity to a legislature that represents distinct geographic constituencies. We then consider what useful theoretical predictions such an extension might yield. On this front, our findings are not promising: the bridge from a constituency-level comparative static to a prediction about the aggregate behavior of Congress requires a bevy of highly specific assumptions of precisely the variety that elegant modeling seeks to avoid.

Thus, in Section "A Causal Framework for Evaluating Electoral Theories of Inequality and Polarization", we propose a strategy for anchoring the evaluation of electoral theories of polarization not in arbitrary assumptions but in observed empirical patterns. We develop a causal framework and a statistical methodology for assessing whether inequality caused polarization via a particular electoral channel. In Section "Evaluating the Meltzer-Richard Electoral Theory of Polarization", we apply this approach once more to Meltzer-Richard. Our analysis of U.S. Senate data from 1984 to 2018 yields a new, general insight into the plausibility of "pocketbook voting" theories in the American context: while across-state inequality has indeed risen over this period, potentially setting the stage for growing redistributive conflict, state median incomes have almost universally *fallen* relative to the national mean, meaning that all states' median voters now have more to gain from redistribution. These countervailing trends imply that, if voters only considered their private benefit from redistributive policy, a representative Congress should be no more polarized today than it was 40 years ago.

On Constructing Coherent Electoral Theories of Inequality and Polarization

Legislative polarization is perhaps the most straightforward measure we have of the intransigence of political parties, a phenomenon that has vexed observers of U.S. politics from Washington (1796) to Lee (2015). Defined in the American context as the distance between Republicans' and Democrats' average ideologies and usually instantiated as the difference in the parties' mean NOMINATE scores (Poole and Rosenthal, 1985), polarization has risen steadily since World War II in both chambers of the U.S. Congress. The democratic consequences of a polarized legislature — especially absent an equally polarized electorate — are potentially dire: representatives who are unaccountable to their constituencies,

vitriolic and unproductive political discourse, lengthy and wasteful campaigns, and a legislature that cannot govern effectively. It naturally follows that scholars have devoted extensive attention to the rise of polarization as well as its downstream effects.

Perhaps the only recent phenomenon to rival polarization in scholarly concern is income inequality, and it is no coincidence that the two are often part of the same conversation. Both have been seen as symptoms and causes of any number of deeper maladies, and both rebounded after a mid-century slump to higher, historical levels, inspiring the notion that income inequality is to capitalism as polarization is to democracy (Jacobs and Soss, 2010). Observing their concurrent rise over the last half century, McCarty *et al.* (2006) propose a causal "dance" between the two variables. Their Figure 1.1 (p. 6), which overlays the U.S. national Gini coefficient and polarization to emphasize the co-trending behavior, has achieved something of a celebrity, and often evidentiary, status among social scientists of all stripes. McCarty *et al.* pithily summarize the appealing intuition behind a bottom-up theory of polarization based on the growing disparity between the top and bottom of the U.S. income distribution: "People at the top devote time and resources to supporting a political party strongly opposed to redistribution. People at the bottom would have an opposite response."

Intuitive appeal aside, what are the precise causal mechanisms behind this story? The literature abounds with theories of electoral behavior based on voters' own incomes and the distribution of incomes in the electorate at large, many following in the tradition of Meltzer and Richard (1981). Theories concerning politicians' electoral accountability are equally if not more plentiful (see Canes-Wrone (2015) and Ashworth (2012) for excellent reviews). While the supply of pertinent theoretical accounts may not be the limiting factor, we argue that constructing coherent bottom-up electoral theories of polarization is deceptively challenging.

First and foremost among the challenges is the complexity introduced by single-member districts.[2] Polarization is an aggregate phenomenon, summarizing the simultaneous responses of numerous legislators to voters in their distinct constituencies. Thus, in a genuinely bottom-up account of polarization, changes occurring at the constituency level must lead to a set of changes in legislative behavior that, averaged across parties, imply that polarization has increased. National co-trends obfuscate the causal relationship that would be relevant to an electoral theory. A change in national income inequality is not the appropriate input into an electoral theory of polarization; the appropriate

---

[2]More broadly, our discussion applies to geographically distinct constituencies that could be multi-member, as in the U.S. Senate.

input is a set of constituency-level changes in the features of the income distribution that are relevant for voters' political behavior.[3]

We are not alone in appreciating the complications posed by single-member districts. Bolton and Roland (2011) present a model in which a decisive voter in each constituency sets policy for that political unit. Yet, when aggregated into a national structure with each constituency having a single vote, the median of constituency medians becomes the decisive voter. In a similar vein, Borge and Rattsø (2004) argue that a more suitable avenue of research than cross-country regressions relating income inequality to political polarization would be to focus on political outcomes that are determined by a single constituency, and the associated levels of income inequality. This motivates their own study of localities in Norway, which are responsible for a substantial share of the redistributive policy that applies to them.

Most forcefully, Rodden (2010) writes, "In order to sweep geography under the rug, [the] literature often makes one of two simplifying assumptions: Either each district contains an identical distribution of voter ideal points, or, perhaps more realistically, the overall distribution of individuals mirrors the distribution of district medians." He notes, however, that substantial homophily in the actual distribution of voters belies the notion that voters are distributed as-though random. He continues,

> These observations might also have implications for the more abstract literature in the Meltzer-Richard (1981) tradition that assumes political preferences are derived exclusively from one's place on the income spectrum. In many societies the poor live in higher density than the rich, such that the median voter in the median district is wealthier than the median voter in the society as a whole. Moreover, the distribution of income across districts will always be far less right-skewed than the distribution across individuals, and if the districts are large enough, the distribution will not be skewed at all. These simple observations have clear, as-yet-untested implications for redistribution.
>
> Rodden (2010, p. 324)

Accounting properly for electoral geography is not trivial, though it is just the tip of the iceberg when it comes to constructing coherent, causal bottom-up theories of polarization. In the remainder of this section, we formalize what we consider to be the key considerations, which are summarized in Figure 1.

Specifically, we argue, one must contend with two types of assumptions: fundamental and flexible. We think of fundamental assumptions as *given*. The reality of single-member districts, for instance, is a fundamental assumption

---

[3]This observation animated Tullock's (1983) critique of Meltzer and Richard's (1983) "test" of their (1981) theory.
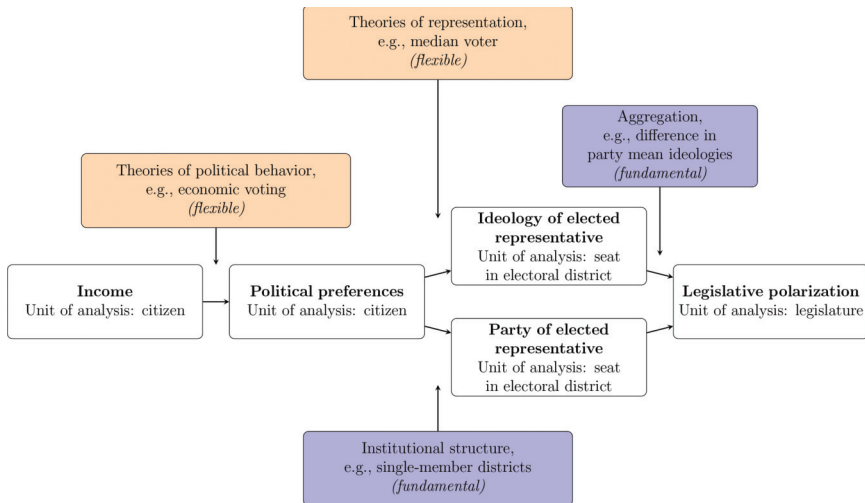
Figure 1: A general framework for constructing electoral theories of inequality and polarization.

that any electoral theory of polarization must account for. The aggregation and summary processes are fundamental assumptions because they are determined by institutional rules (for instance, the number of constituencies sending delegates to the legislature) and by the definition of the outcome of interest (polarization being the difference in parties' mean ideology scores).

Flexible assumptions, in contrast, are *informed* by prior theoretical and empirical work, but they fall short of being universally agreed-upon "facts on the ground." The theory of electoral competition at the core of an electoral theory of polarization is the most obvious flexible assumption to invoke, but it is hardly the only or even the most consequential one. That this theory must take the voting citizen as the unit of analysis is fundamental, but voter preferences, turnout, knowledge, partisan attachment, and even degree of rationality are all flexible assumptions needed to map from the income distribution and a voter's place in it to that voter's political preferences and behavior. Many of these assumptions remain highly contested in the literature. Are voters self-interested or altruistic? Are they motivated by pocketbook considerations, social issues, or partisan attachments and identities? Though our focus is on "pocketbook voting," it would be a grave sin of omission not to note the wealth of research in political behavior suggesting that vote choice in the U.S. context is best explained by social identities and partisan loyalties; just a few prominent recent examples include Achen and Bartels (2017), Mason (2018), and White and Laird (2020).

Still more flexible assumptions are necessary to map from voter preferences to the electoral and ultimately legislative behavior — as well as the

partisanship — of the winning candidate. For example, is the median voter decisive? Do candidates converge to the median's most preferred policy, or do they diverge, perhaps to party medians? When in office, do they faithfully execute on their campaign promises, or do they follow their own preferences? Do politicians trade votes amongst themselves or sell votes for particularistic kickbacks? Finally, the routinely-invoked but rarely-questioned claim that NOMINATE's scoring of equilibrium legislative voting behavior constitutes a representation of legislator ideology, or preferences induced by voters, is certainly an assumption in the "flexible" category (see Clinton (2012) or Fowler and Hall (2012) for alternative approaches to measuring voting behavior).

While flexible assumptions ought to be theoretically motivated and empirically supported, by their nature these assumptions represent findings that scholars are actively debating and refining. Unlike fundamental assumptions, which — once appreciated — are relatively straightforward to accommodate, flexible assumptions continue to present a daunting degree of complexity. In the section that follows, we apply our framework to construct an electoral theory of polarization centered on the Meltzer-Richard model. In so doing, we demonstrate that the flexible assumptions required for even this most basic of electoral theories effectively preclude the possibility of producing robust theoretical predictions about polarization.

Building an Electoral Theory of Polarization from the Meltzer-Richard Model

A Brief Overview of the Meltzer-Richard Model

Although there is a rich and active literature that has produced numerous theoretical accounts of economic voting, we focus on Meltzer-Richard because it formalizes the most basic logic of how rational, self-interested voters form redistributive preferences — and thus captures what is often colloquially meant by "pocketbook voting." In essence, then, we pose the question of to what degree polarization may be linked to inequality via citizens' divergent incomes driving their divergent redistributive preferences.

Meltzer-Richard begins with workers in an economy; income inequality arises out of exogenous differences in their labor productivity. These workers are also citizens of a democratic polity, and they are empowered to select a tax rate by popular vote. With universal and compulsory voting and simple majority rule, the median voter's preference is decisive in setting policy.[4] Thus,

---

[4]One may either imagine voters directly selecting a tax rate by referendum, or an equilibrium in which office-motivated candidates enter and win elections by proposing (with commitment) the median voter's ideal policy, though only the latter scenario will speak to legislative polarization.

the model's key comparative static is defined with respect to the income of this decisive voter — the median in the income distribution — relative to the mean income in the polity, which determines the extent to which that voter is helped or harmed by taxation. The higher the median income relative to the mean, the less this decisive voter gains from taxation, and thus the lower the equilibrium level of redistribution in this society.

Extending the Model to an Electoral Theory of Polarization

Returning to the necessary ingredients for an electoral theory of polarization in Figure 1, Meltzer-Richard provides two essential components: the mapping from citizens' incomes to their political preferences, and an adequate (if unrealistic) theory of representation.[5] Voters choose redistributive policy to maximize their own post-transfer utility, candidates are free to enter with any policy proposal they choose (unconstrained from above by parties), there is a mechanism that commits them to implementing their platforms in office, everyone votes, and in equilibrium the winning candidate proposes and acts on the median voter's preferences.

Extending Meltzer-Richard to a coherent electoral theory of polarization requires yet additional scaffolding. First and foremost, we need to export the model's basic logic to a new institutional setting: a legislature composed of representatives elected from distinct geographic constituencies. We may suppose this legislative body sets a uniform redistributive policy for the nation. A rational voter, then, motivated by the same considerations as in Meltzer-Richard, will determine her optimal tax rate based on her own income relative to the national mean. But the institutional structure will yield as many representatives as there are legislative seats, each determined by the median voter *in that constituency* rather than the nation.[6] Thus, the key comparative static in this setting connects a constituency's median income relative to the national mean with the voting record or latent ideology (along the economic/redistributive dimension) of the representative from that constituency.

To complete the electoral theory of polarization, it remains to specify an aggregation rule that takes a legislature of representatives, each with an induced ideological position, and returns a polarization measure for that group. Our chosen metric, the difference of party means, requires a vector of party

---

[5]Changing some of these flexible assumptions to better align with the American context is a fruitful direction that is fully accommodated by our framework. For instance, one might assume that richer citizens are likelier to vote, implying that the median *voter* is richer than the median *income earner* (see, e.g., Benabou, 2000).

[6]We assume that citizens do not vote strategically with respect to the legislative bargaining progress, for instance by electing an extremist so that they might pull the final outcome closer to their (more moderate) ideal point. See Judd (2019) for a model that incorporates compositional effects within a legislature.

labels to go along with legislators' ideologies. The need to model legislator partisanship in addition to ideology makes predictions about polarization remarkably hard to pin down, as the following theoretical exercise underscores.

Robust Theoretical Predictions from the Extended Model?

Meltzer and Richard (1981) are silent on the partisanship of the winning candidate — an understandable omission given the model's equilibrium convergence to the median voter's ideal policy. Yet partisanship is clearly a central component of polarization. How might we reasonably extend Meltzer-Richard to speak to partisanship? One fairly straightforward approach would be to assume that the greater the income of the median voter in a constituency relative to the national mean,[7] the more likely a Republican candidate is to represent the constituency. More specifically, even, we might suppose that when median income in a district rises relative to the national mean, that district will be weakly more Republican, meaning we allow for a shift from Democratic to Republican representation but rule out a shift in the other direction. The opposite prediction would hold for districts where median income falls: the representative would be weakly more Democratic.

With these two constituency-level comparative statics in hand, describing how changes in a district's median income relative to the national mean relate to the elected representative's partisanship and ideology, what can we say about downstream changes in polarization? In other words, if we can take these comparative statics as given, is it possible to make any general statements relating some *set of changes* in district median incomes to changes in *aggregate polarization*? This turns out to be prohibitively difficult. In Online Appendix A, we run through several simple numerical examples to spell out this point. We show that, holding constant these comparative statics, a given change in constituency incomes may yield more or less polarization, depending on specific assumptions about the mappings implied by the comparative statics. We need to know where ideal points and partisanship fell before any exogenous changes in the income distribution. We also require a host of distributional assumptions about which constituencies experienced which changes in income, again as a function of the initial distribution of ideal points and partisan affiliations. This assumption-laden approach is precisely what modern theoretical work seeks to avoid, so analysts would find themselves trading compelling and elegant modeling for the ability to generate predictions about polarization.

Hence, we conclude that generating theoretical predictions about polarization by augmenting a core model of elections with fundamental assumptions structured by institutions and a bevy of additional flexible assumptions is not just challenging but likely, if we are being honest with ourselves, futile. We

---

[7]See above for an explanation of why the national mean income is the relevant basis for comparisons.

emphasize, however, that our application to Meltzer-Richard still constitutes a coherent electoral theory of polarization. It is unable, though, to produce theoretically the exact output we seek from such a theory: a sufficiently general, testable comparative static relating inequality to polarization.

A Causal Framework for Evaluating Electoral Theories of Inequality and Polarization

How can researchers make progress on what remains an important question in spite of its seemingly intractable complexity? In this section, we pivot to a more promising way forward, approaching the question as an empirical exercise. The theoretical framework depicted in Figure 1 plays a central role in structuring our statistical analysis. We also introduce some additional structure: a causal inference framework to guide us in formulating the estimand of interest and understanding when and how it may be identified from data. Rather than invoking arbitrary distributional assumptions, we show how one would instead anchor the analysis in readily available measures of the key components of the electoral theory: voters' incomes and legislators' partisanship and ideology.

Our aim here is to assess whether changes in the income distribution caused polarization through a particular electoral channel. The rigorous evaluation of such claims is made possible with the help of the potential outcomes framework (Rubin, 2005). To begin, let constituencies, or electoral districts, be indexed by $k = 1, \ldots, K$. To simplify the presentation, let there be one representative elected to the legislature per district. Let the vector $\vec{X} = X_1, \ldots, X_K$ contain the features of the income distribution that are relevant to the pivotal voter in each constituency: for instance, under Meltzer-Richard, this is $\frac{\text{median income}_k}{\text{national mean}}$. While income is a continuous quantity, we simplify the analysis dramatically by considering two particular vectors of incomes: some treatment values, $\vec{X}^T = X_1^T, \ldots, X_K^T$, and a counterfactual, $\vec{X}^C = X_1^C, \ldots, X_K^C$. One is free to specify these vectors according to one's substantive interest. For instance, one could take as $\vec{X}^T$ the distribution of income observed in the U.S. today, and compare it to an $\vec{X}^C$ of the observed income distribution in the U.S. fifty years ago; to a distribution like one currently observed in another nation; to a distribution that would result from a specific tax policy under consideration; or to any purely hypothetical distribution, such as a perfectly flat one.[8]

A theory of elections proposes a causal relationship between X_k and Y_k, the ideology of the legislator elected to Congress from constituency k. The **constituency-level treatment effect of income on ideology** is given by $Y_k(X_k^T) - Y_k(X_k^C)$ for treatment level X_k^T versus the counterfactual level X_k^C.

---

[8]As the subsequent discussion will show, we still make full use of the continuous nature of X; however, restricting our attention to binary treatments greatly simplifies the analysis while still accommodating a wide range of substantively interesting comparisons.

As always, we cannot observe the same constituency under both conditions at the same time. When we are interested in evaluating the theory of elections in its own right, we seek to identify the **average treatment effect of income on ideology**, $\mathbb{E}[Y|X^T] - \mathbb{E}[Y|X^C]$.

Extending such a theory to identify its implied causal effect of income on polarization requires only that we adapt the familiar causal quantity of interest from ideology in a given constituency to a difference of party average ideologies in the overall legislature. The fact that party labels are also likely to move in response to changes in income and ideology, however, complicates the problem significantly. We therefore require a function $R(X, Y)$ that maps legislator ideology, as well as income, to partisanship. As we will discuss in greater detail, R may be completely exogenous; it may be caused by either or both of ideology and income; indeed, it may itself affect ideology. (We will consider each of these causal structures after the initial setup.) We are now ready to define **expected polarization** as the expected difference of average ideologies among Republicans ($R = 1$) and Democrats ($R = 0$) under some income distribution.[9] For income vector $\vec{X}^T$, we would write expected polarization under the treatment distribution as:

$$EP(\vec{X}^T) = \mathbb{E}\left[\overbrace{\frac{1}{\sum_k R_k}\sum_k R_k Y_k}^{\substack{\text{average ideology} \\ \text{for Republicans}}} - \overbrace{\frac{1}{\sum_k (1-R_k)}\sum_k (1-R_k)Y_k}^{\substack{\text{average ideology} \\ \text{for Democrats}}} \middle| \vec{X}^T\right] \quad (1)$$

With this definition in hand, the **treatment effect of income on polarization** may straightforwardly be written as the difference between expected polarization evaluated at the treatment and counterfactual income distributions:

$$\Delta EP(\vec{X}^T, \vec{X}^C) = EP(\vec{X}^T) - EP(\vec{X}^C) \quad (2)$$

The remainder of this discussion focuses on the identification of Equation 2, our causal quantity of interest. In order to make progress on this task, though, we first have to elaborate a theory for how income, ideology, and partisanship are interrelated. Figure 2 presents a number of reasonable scenarios in the form of directed acyclic graphs (DAGs) (Pearl, 2009). Here, arrows represent direct causal effects between two variables; by the same token, an absence of arrows between two variables represents the sharp assumption of no causal effects between them. In every DAG, income affects ideology ($X \rightarrow Y$), encoding

---

[9]The expectation operators are not taking the averages that constitute the components of polarization itself. Rather, we are calculating the expected value of these average party ideologies, given that, often in theory and always in practice, ideology and partisanship are stochastically determined.
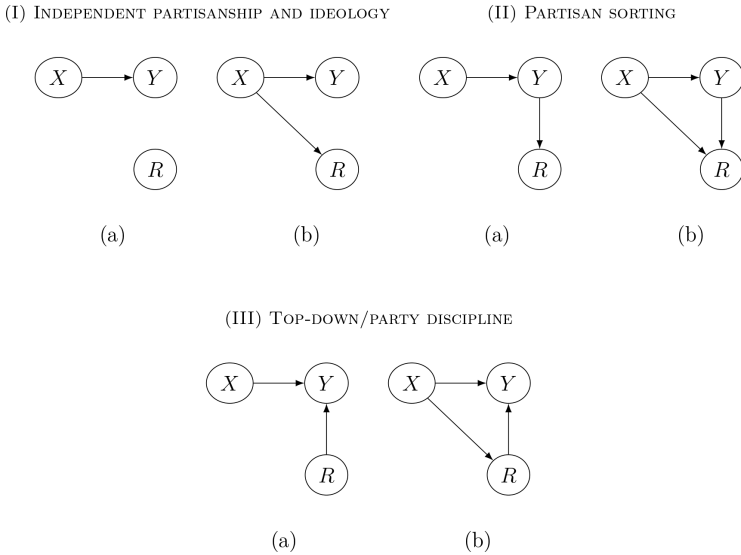
(I) Independent partisanship and ideology

(II) Partisan sorting

$X \rightarrow Y$ $X \rightarrow Y$

R R

(a)   (b)

$X \rightarrow Y$ $X \rightarrow Y$

R R

(a)   (b)

(III) Top-down/party discipline

$X \rightarrow Y$ $X \rightarrow Y$

R R

(a)   (b)

Figure 2: Possible causal DAGs for the relationship between voter income (X), legislator ideology (Y), and legislator partisanship (R).

our core theory of elections. (For concreteness, our discussion will frame this causal relation in terms of the adapted Meltzer-Richard model discussed in the previous section, though of course $X \rightarrow Y$ may represent a wide variety of theories.) However, in each panel we consider a different relationship between income and partisanship, as well as ideology and partisanship.

Panel I represents one simple model: income affects ideology, and it may or may not also affect partisanship (subfigures (b) and (a), respectively). Under this model, the poorer voters get, the more they will support pro-redistribution candidates; in version (b), they also will become likelier to support a Democrat than a Republican with the same ideology. However, there is no direct causal link between Y and R; any association one might observe in DAG I(b) is induced by their common cause, X.

Panel II represents an alternative formulation: income causes ideology, and ideology in turn causes partisanship. Here, voters first form policy preferences — based at least in part on the income distribution and/or their place in it — and vote according to those preferences. Party labels are determined subsequently as a function of ideology. This model is consistent with a *sorting* of pro-redistribution politicians into the Democratic party and anti-redistribution politicians into the Republican party. As before, we may wish to further accommodate a direct causal link from income to partisanship (subfigure (b)).

Panel III offers yet another alternative: rather than ideology determining partisanship, partisanship is a (conditionally) exogenous factor that informs

and constrains the ideological positions legislators can take. This is consistent with a *top-down* understanding of partisan politics where legislators are locally electorally responsive to some extent, but also kept in line by party elites or other national actors. Of course, this model does not require that partisanship be randomly determined; it is simply understood to be exogenous to the redistributive dimension of ideology. This structure can also accommodate an interaction between partisanship and income. For instance, one may respond differently to one's place in the income distribution depending on one's prior partisan commitments: perhaps Republicans demand less redistribution as they get wealthier, while Democrats demand more.[10]

Returning to Equation 2, our causal estimand for the effect of income on polarization, we now rewrite this expression in accordance with the specific structure of each of the six DAGs in Figure 2. In Table 1, we focus on one of its four terms, the expected average ideology among Republicans given income vector $\vec{X}^C$. We take expectations over all the endogenous variables in a given model, and we use the independence relations implied by each DAG — listed in the second column — to reduce the expression.[11] The third column of Table 1 lists the final estimands; all intermediate steps are shown in Online Appendix B. To reconstruct Equation 2, the other three terms may be written similarly, substituting $\vec{X}^T$ instead of $\vec{X}^C$ for treatment income and $1 - R_k$ instead of R_k for Democrats.

Finally, we propose a three-step process for estimating these quantities. First, in the **estimation step**, the analyst uses constituency-level data on income, ideology, and partisanship to obtain estimates of all the component causal effects of the chosen DAG. In DAG I(b), for instance, one would need to estimate the causal effect of X on Y as well as the causal effect of X on R. Next, in the **prediction step**, the analyst uses the estimates obtained in the first step to generate predicted values of ideology and partisanship in every constituency under the treatment and counterfactual income conditions. Lastly, in the **aggregation step**, the predicted ideology and partisanship vectors are used to compute expected polarization under the treatment and counterfactual incomes. The difference between those values is our estimate of the treatment effect of income on polarization.

---

[10]A weakness of the standard DAG framework is that there is no way to distinguish in Panel III whether income and partisanship each exert some separable effect on ideology, or whether partisanship conditions the effect that income exerts on ideology. See Nilsson *et al.* (2021) for an extension of the DAG framework that explicitly accommodates interactions, which in our example would be represented by $R \rightarrow \Delta Y_X$.

[11]Given the simplicity of our models, these are not difficult to see visually, but an algorithm to find all of a DAG's implied conditional independencies is implemented in the R package "daggity" (Textor *et al.*, 2016). This is particularly useful for more complex causal structures.

Table 1: Estimand for expected average ideology among Republicans under the counterfactual income distribution, $\vec{X}^C$.

DAG	Independencies	Estimand: $\mathbb{E}\left[\sum_k \frac{R_k}{\sum_j R_j} Y_k \;\middle|\; \vec{X}^C\right]$
(I) INDEPENDENT PARTISANSHIP AND IDEOLOGY		
(a) $X \to Y$	$R \perp\!\!\!\perp X,\, Y \perp\!\!\!\perp R$	$\frac{1}{N_R}\sum_k R_k \mathbb{E}_Y\left[Y_k | \vec{X}^C\right]$
(b) $X \to Y$	$Y \perp\!\!\!\perp R | X$	$\sum_k \left(\mathbb{E}_R\left[\frac{R_k}{\sum_j R_j} \middle| \vec{X}^C\right] \mathbb{E}_Y[Y_k | \vec{X}^C]\right)$
$\quad X \to R$		
(II) PARTISAN SORTING		
(a) $X \to Y$	$R \perp\!\!\!\perp X | Y$	$\mathbb{E}_Y\left[\mathbb{E}_R\left[\sum_k \frac{R_k}{\sum_j R_j} Y_k \middle| \vec{Y}^C\right] \middle| \vec{X}^C\right]$
$\quad Y \to R$		
(b) $X \to Y$	none	$\mathbb{E}_Y\left[\mathbb{E}_R\left[\sum_k \frac{R_k}{\sum_j R_j} Y_k \middle| \vec{X}^C, \vec{Y}^C\right] \middle| \vec{X}^C\right]$
$\quad Y \to R$		
$\quad X \to R$		
(III) TOP-DOWN/PARTY DISCIPLINE		
(a) $X \to Y$	$R \perp\!\!\!\perp X$	$\sum_k \frac{R_k}{\sum_j R_j} \mathbb{E}_Y[Y_k | \vec{X}^C, \vec{R}^C]$
$\quad R \to Y$		
(b) $X \to Y$	none	$\mathbb{E}_R\left[\mathbb{E}_Y\left[\sum_k \frac{R_k}{\sum_j R_j} Y_k \middle| \vec{X}^C, \vec{R}^C\right] \middle| \vec{X}^C\right]$
$\quad R \to Y$		
$\quad X \to R$		

Though this process is simple, the conditions under which it yields unbiased estimates of the desired causal effect require careful thinking. We conclude this section with a brief discussion of four such conditions.

1. Causal identification of the underlying theory of elections. We note, first and foremost, that the electoral theory of polarization is (at best) as well-identified as its constituent parts. But causal identification of the core theory of elections $(X \to Y)$ is extremely challenging. We assume as a baseline that the analyst brings to the table some identification strategy — perhaps an instrument that perturbs incomes somewhat randomly, or at least a set of fixed effects and/or time-variant controls. Even so, one must grapple with a fundamental trade-off: any empirical strategy that succeeds at causal identification is likely to soak up much of the variation in incomes, leaving little available for explaining polarization.

2. Choosing the right causal structure. A closely related point is the vital importance of choosing the DAG that matches the true data-generating process. We illustrate this with just one familiar example. Suppose the analyst adopts a version of DAG III(a) in which income interacts with partisanship to produce ideology. In that case, it is important to estimate the effect of X on Y *conditional* on party. But if the analyst has gotten the causal model wrong and the data-generating process is actually best described by DAG II(b), then conditioning on party induces collider bias, as R is a common outcome of X and Y (Elwert and Winship, 2014). This leads to a violation of our first requirement: proper causal identification of the core theory of elections.

We further add that getting the model right requires specifying all of the important channels by which income affects polarization. In order to conclude that income has affected polarization via the *particular* bottom-up mechanism the analyst has in mind, one must assume that there are no direct effects of income on polarization. This assumption would be violated if, for example, the fact of rising inequality caused party leaders to adopt more ideologically extreme platforms on both sides. In other words, the DAGs in Figure 2 encode something like an exclusion restriction for the effect of income on polarization: the effect may only travel through the electoral channel of $X \to Y$, and perhaps $X \to R$, but not directly.[12]

3. Non-interference among constituencies. Thus far, we have assumed that the constituency-level causal relationships play out independently in every district, and that one district's income, partisanship, and ideology have no bearing on another district's outcomes. This assumption is easily violated when media markets straddle multiple congressional districts (Snyder and Strömberg, 2010), when successful new campaign strategies inspire imitators, or when social movements sweep the nation. If we allow for these and other spillovers, then it is not enough to just condition on income, ideology, and partisanship in district k, as we have done in deriving the expressions in Table 1; one must condition on the entire income, ideology, or partisanship vectors. Without a way to directly model or simplify such dependencies, this structure quickly becomes unwieldy even with a small number of districts.

4. Good out-of-sample prediction. Another potential pitfall arises when moving from the estimation step to the prediction step. Even if certain causal quantities can be successfully identified from the data, one must carefully consider whether these relationships can be reasonably extrapolated to the proposed treatment or counterfactual income vector. Sufficiently extreme perturbations of the income distribution may bring about tectonic shifts in the economy, partisan realignments, and even revolutions — in other words, completely different equilibria from the ones that produced the data that went into the estimation step. The farther away the hypothetical distribution from

---

[12]We thank Nolan McCarty for bringing this point to light.

observed incomes, the more one should be wary of out-of-sample prediction; that said, given the possibility of feedback loops and tipping points, even small perturbations leave plenty of room for error.

In Online Appendix C, we perform a set of simulations in which we generate treatment and counterfactual potential outcomes according to the DAGs in Figure 2. Then, we apply our proposed three-step method to a subset of the generated data that we assume the analyst can observe. When the assumptions elaborated above hold, we are able to recover the true treatment effects of changes in the distribution of income on polarization.

Evaluating the Meltzer-Richard Electoral Theory of Polarization

Having formulated a complete electoral theory of polarization from the Meltzer-Richard model and elaborated a causal framework for theory testing, we are prepared to bring the electoral theory to data. In this section, we describe how we construct a panel dataset that tracks voter income as well as legislator ideology and partisanship for the U.S. Senate from 1984 through 2018. Then, we proceed to the first stage of empirical testing: evaluating the causal relationships that constitute the electoral theory at the constituency level. We find some evidence in favor of Meltzer-Richard's key comparative static: an increase in a state's median income relative to the national mean is associated with a conservative shift in legislator ideology. Nonetheless, we show that this model does *not* perform well as an electoral theory of polarization. This may be due to the simple fact that constituency median incomes across the board have fallen quite substantially relative to the national mean since 1984, even as across-constituency inequality has risen. Given these countervailing forces, an electoral theory based on Meltzer-Richard would not predict polarization to have risen as clearly as it has over the past half century.

Data

We construct a balanced panel containing two senators per state and two-year congressional session from 1984 through 2018.[13] Following standard practice, we measure legislative ideology using NOMINATE scores downloaded from the Voteview database (Lewis *et al.*, 2021). We focus on estimates of the first dimension, which is understood to capture issues of taxation, spending and redistribution (Boche *et al.*, 2018). We assign relevant features of the state

---

[13]We focus on the Senate because income variables at the House district level are only available going back to 2006, and the 2010 redistricting further cuts down the length of our effective panel. In cases where there were multiple senators per seat in the same legislative session due to deaths or resignations, we keep only the legislator who was elected on the regular election cycle.

and national income distributions (in 2018 current dollars) to each Congress based on the year prior to the start of that session, that is, the election year in which the members of that Congress were conceivably elected.[14] We obtain estimates of state median and national mean income from the Census.

Evaluating the Constituent Causal Relationships

We now proceed to an empirical assessment of the constituency-level relationships between income, ideology, and partisanship. Our purpose here is twofold. First, we need to assess whether Meltzer-Richard holds water as a stand-alone theory of elections, since this is a necessary precondition for being a mechanism linking inequality to polarization. Second, recalling the importance of correctly modeling the relevant constituency-level causal relations, we use the data to inform our choice of DAG from among the options in Figure 2.

A few words on the causal identification of these constituency-level relationships are in order at this point. While our aim in this paper is not airtight identification as much as it is a demonstration our general framework, we emphasize once more that the electoral theory is only as well-identified as its constituent parts, and so we do wish to engage in good faith with the issue of estimation. Throughout, we will use a linear regression with state and Congress fixed effects, which will yield unbiased estimates of the desired effect if there are no excluded state-specific, time-variant confounders. This assumption would be violated for the effect of income on ideology if, for example, states have sorted over time into a knowledge economy or a manufacturing economy, and if this type is associated with both voters' incomes and their support for the Republican party (perhaps for social rather than economic reasons). While we acknowledge the possibility of this sort of confounding, we cautiously proceed with the two-way fixed effects specification as a first pass, with an eye toward strengthening identification in future work.

In Table 2, we report estimates from the regression:

$$\text{NOMINATE}_{skt} = \beta_0 + \beta_1 \frac{\text{median income}_{kt}}{\text{national mean income}_t} + \text{Congress}_t \gamma + \text{State}_k \eta + \varepsilon_{skt} \tag{3}$$

for senator s from state k and Congress t. We gradually build up to this specification (Column 3) from a bivariate regression (Column 1). Although the association between income and ideology is negative under the bivariate and Congress fixed effects models, it becomes positive and statistically significant when we include state fixed effects. Under our most strenuous specification, a one-unit increase in state median income divided by the national mean leads

---

[14]Of course, many members of that Congress would not have been up for election in that year, but it is still reasonable to assume that these are the facts on the ground that inform constituent preferences and thus legislative decisionmaking.

Table 2: First stage test of effect of income on ideology.

	(1)	(2)	(3)
State median/national mean income	-0.930^{***}	-0.933^{***}	0.474^{**}
	(0.079)	(0.083)	(0.173)
Congress FE	No	Yes	Yes
State FE	No	No	Yes
R^2	0.072	0.077	0.524
Adj. R^2	0.072	0.068	0.506
Num. obs.	1800	1800	1800

Notes: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.

Table 3: Tests of implied conditional independencies in the U.S. senate, 1984–2018.

Independency	Regression model	Test	Result
$Y \perp\!\!\!\perp R$	$y_{skt} = \beta_0 + \beta_1 r_{skt} + \varepsilon_{skt}$	$\beta_1 = 0$	$\beta_1 = 0.62\ (p < .01)$
$Y \perp\!\!\!\perp R\|X$	$y_{skt} = \beta_0 + \beta_1 r_{skt} + \beta_2 x_{kt} + \varepsilon_{skt}$	$\beta_1 = 0$	$\beta_1 = 0.62\ (p < .01)$
$R \perp\!\!\!\perp X$	$r_{skt} = \beta_0 + \beta_1 x_{kt} + \varepsilon_{skt}$	$\beta_1 = 0$	$\beta_1 = 0.50\ (p = .06)$
$R \perp\!\!\!\perp X\|Y$	$r_{skt} = \beta_0 + \beta_1 x_{kt} + \beta_2 y_{skt} + \varepsilon_{skt}$	$\beta_1 = 0$	$\beta_1 = -0.16\ (p = 0.10)$

Notes: All models include state and Congress fixed effects. R is a binary indicator for a senator s in state k and Congress t being a Republican; X is the state's median income divided by the national mean; Y is the first-dimension NOMINATE score.

to a 0.47-unit ideological shift in the conservative direction. Substantively, this effect is small but not inconsequential, as the range of the dependent variable is -0.76 to 0.91, and the range of the independent variable is 0.40 to 1.04.

With some preliminary evidence in favor of Meltzer-Richard in hand, we proceed to data-driven DAG selection. We are aided at this stage by the implied conditional independencies shown in the second column of Table 1, which may be tested directly as recommended by Pearl (2010). In Table 3, we report the regression model we use to test each of these independencies, following the same two-way fixed effects approach outlined above, as well as the relevant coefficients from each test.[15] Unsurprisingly, we firmly reject the independence of ideology and partisanship, both unconditionally and conditional on income. The relationship between income and partisanship is somewhat weaker, or at least noisier, but we still interpret p-values under 0.10 as suggestive of some interdependence between them.

Rejecting these independencies leaves only two plausible DAGs for the data-generating process out of the options proposed in Figure 2: II(b) and III(b). This is the extent to which the data can help us, since they can alert

---

[15]The full results are shown in Online Appendix Tables D-4 and D-5.
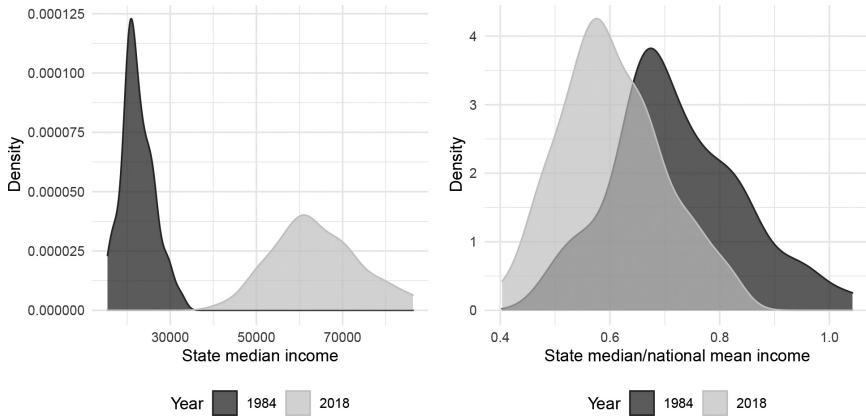
Figure 3: Densities of treatment (2018) and counterfactual (1984) incomes.

us to the existence of an association but not to the direction of the causal link. Thus we are now required to make a substantive judgment: does ideology inform one's choice of party labels, or do preexisting partisan attachments shape one's ideology? Given the undeniable importance of partisan sorting in U.S. politics over the last 50 years (see, e.g., Fiorina, 2017), we proceed with DAG II(b) for our main analysis, leaving other, potentially more complex causal structures as an avenue for future work.[16]

Estimating the Effect of Income on Polarization

To estimate the effect of income on polarization, it remains to specify a vector of treatment and counterfactual incomes. Recalling the importance of choosing income distributions to which one can plausibly extrapolate from the data, we propose a treatment vector of state median income over the national mean as observed in 2018, and a counterfactual vector of the same statistic as observed in 1984.[17] This will allow us to answer the question of what polarization would look like today if relative incomes remained unchanged since 1984. If subsequent trends in income inequality are indeed a major cause of polarization, then expected polarization should be higher under the treatment condition than the counterfactual.

Figure 3 summarizes the treatment and counterfactual distributions. In the left panel, we plot the density of state median incomes (measured in 2018 current dollars) in 1984 (dark gray) and 2018 (light gray). The relative widths

---

[16]If, for instance, we suppose partisanship and ideology simultaneously affect one another, unbiased estimation would require a variable that affects one but not the other, that is, an instrument satisfying an exclusion restriction.

[17]We are indebted to Anthony Fowler for this suggestion.

of these distributions reveal that across-state inequality has increased over this period, potentially setting the stage for growing redistributive conflict in Congress. However, the panel on the right tells a different story. Here, we plot the densities of the key Meltzer-Richard statistic — state median income divided by national mean — in the same two periods. The two distributions have roughly the same shape, but the 2018 distribution is uniformly shifted downward. This is due to an important countervailing trend: at the same time that across-state inequality has grown, nearly *all* state medians have become poorer relative to the national mean.[18] Whereas the former should push Congress toward greater redistributive conflict, the latter should make all legislators more supportive of redistribution under Meltzer-Richard. Which of the two effects dominates is an empirical question that we are now well-positioned to answer.

We estimate the model in a Bayesian framework, which has the advantage of allowing for joint estimation of all the constituent causal relationships in DAG II(b), as well as simultaneous prediction and aggregation. Online Appendix E presents the full hierarchical model specification and further details about the estimation process. After estimating the model on the full dataset, we generate two sets of predictions from the posteriors of the parameter estimates: for each state, a predicted ideology and a probability of electing a Republican under the treatment and counterfactual income conditions.[19] The former is generated using 2018 incomes and a 2018 fixed effect, the latter using 1984 incomes and a 2018 fixed effect — capturing the notion of holding all other conditions constant and only varying incomes.[20]

Figure 4 displays histograms of the predicted ideologies among Democratic states (those with a predicted probability of electing a Republican of less than 0.5) and Republican states (those with a predicted probability greater than 0.5).[21] We also report the posterior means of average Democratic and Republican ideology, as well as the distance between the two — our estimate of expected polarization — under each condition. These estimates are remarkably close to one another: the expected distance between party means is 0.56 under

---

[18]The sole exception is Iowa, where median income divided by national mean income has risen slightly from 0.64 in 1984 to 0.65 in 2018.

[19]Although each state has two senators, we have no senator-level predictors, so it suffices to generate one prediction per state.

[20]If we were to include other time-varying controls, we would similarly hold them at 2018 values; however, the Congress fixed effect is the only time-varying feature of our model other than income.

[21]This may be conceptualized as a modal draw of Congress from the posterior distribution of partisanship but is ultimately purely for visualization purposes. The expected average ideologies reported in Figure 4, as well as our estimates of expected polarization, are the posterior means of those quantities as estimated within our Bayesian framework. This allows us to account properly when calculating expectations for stochastically-realized partisanship given the predicted probability of electing a Republican as well as stochastically-realized ideology.

(a) Treatment: 2018 values
Expected polarization: 0.561
95% CI: [0.431, 0.690]

(b) Counterfactual: 1984 values
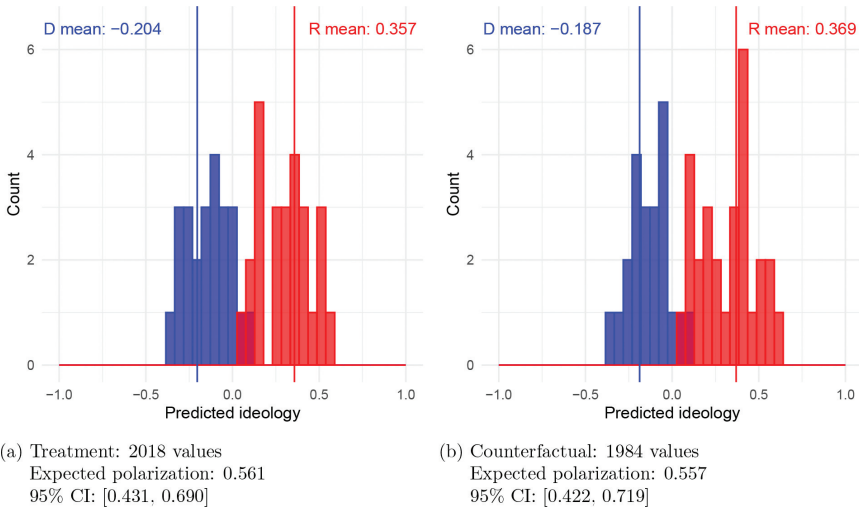Expected polarization: 0.557
95% CI: [0.422, 0.719]

Figure 4: Expected ideologies by expected party under treatment (2018) and counterfactual (1984) incomes and partisanship, U.S. Senate.

both 1984 and 2018 incomes. If anything, both parties, and Republicans in particular, are predicted to be slightly more liberal under 2018 incomes than 1984 incomes, consistent with the locations of the income distributions in Figure 3. We can therefore confidently reject this instantiation of Meltzer-Richard as a causal mechanism linking inequality to polarization.

Discussion

What have we learned about the relationship between income inequality and political polarization from the preceding analysis? In applying and extending the Meltzer-Richard model, we have found that it performs better as a constituency-level theory of political behavior than as an electoral theory of polarization. On average, wealthier constituencies do tend to elect representatives who vote more conservatively and sort into the Republican party. But because median voters across states have grown uniformly poorer relative to the national mean over the past half century, the model does not predict growing redistributive conflict, despite the fact that across-state inequality has also risen.

Enumerating the vulnerabilities of our analysis of a particular electoral theory of polarization leads us to reflect on the endeavor more broadly, so we discuss the limits of our own study in this wider context. The assumptions under which we may claim to have tested an electoral theory of polarization

are, to be blunt, abundant. First, there are the myriad flexible assumptions required to extend an institutionally sparse theory such as Meltzer-Richard to speak to an aggregate phenomenon such as polarization (though, it is worth noting, the frameworks outlined above help greatly in making explicit all of these assumptions for further interrogation). Second, proper identification of the causal quantity of interest — the effect on polarization of today's income distribution compared to the one observed four decades ago — required four rather strenuous assumptions: causal identification of the underlying theory of elections; proper modeling of the causal structure relating income, partisanship and ideology; rejecting the possibility of constituencies affecting one another beyond what is captured by our central summary statistic (state median income over national mean income); and estimates that support out-of-sample prediction. It strains credulity to suppose that these assumptions are fully justified in the present or any other analysis, underscoring the need for transparency around the assumptions. Should future work seek to posit and test electoral theories of polarization, it is our hope that the frameworks above assist in formulating coherent theories and being explicit about the assumptions underlying any claims.

Looking forward, there appear to be several directions for future inquiry. Chief among them would be to move beyond Meltzer-Richard and construct in a similar fashion electoral theories of polarization from other theories of elections. Just a few promising candidates may incorporate the role of altruism (Dimick *et al.*, 2018), differential responses to inequality by the rich and poor (Dimick *et al.*, 2017), the correlation between wealth and political participation (Benabou, 2000), the social distance from the median voter to citizens at the top and bottom of the income distribution (Lupu and Pontusson, 2011), and the use of more accessible proxies for the relevant statistics by voters (Franko, 2017). Perhaps the lowest hanging fruit, though, would be an adaptation of Meltzer-Richard that allows for intra-district divergence in party platforms. Much theoretical work has sought to ground the observation of increasing divergence within individual campaigns (for reviews, see Grofman, 2004; Duggan, 2008), and future work would do well to explore this as a potential source of polarization.

Finally, we note that income need not be the only input into an electoral theory of polarization. Electoral theories of polarization can just as easily be constructed around voter preferences on other dimensions, including social issues. By the same token, our framework accommodates other aggregate outcomes besides polarization, such as intra-party homogeneity (Cox and McCubbins, 2005) or the positioning of key pivots (Krehbiel, 1993), which would only require adjusting the aggregation step of the estimation process to the new aggregate quantity of interest. These adaptations are likely to yield both new substantive insights and refinements of the methodology, an outcome that would be surely welcome.

References

Achen, C. H. and L. M. Bartels. 2017. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton: Princeton University Press.

Ashworth, S. 2012. "Electoral Accountability: Recent Theoretical and Empirical Work". *Annual Review of Political Science*. 15: 183–201.

Benabou, R. 2000. "Unequal Societies: Income Distribution and Social Contract". *The American Economic Review*. 90(1): 96–129.

Boche, A., J. B. Lewis, A. Rudkin, and L. Sonnet. 2018. "The New Voteview.com: Preserving and Continuing Keith Poole's Infrastructure for Scholars, Students and Observers of Congress". *Public Choice*. 176(1): 17–32.

Bolton, P. and G. Roland. 2011. "The Breakup of Nations: A Political Economy Analysis". *The Quarterly Journal of Economics*. 112(4): 1057–1090.

Borge, L.-E. and J. Rattsø. 2004. "Income Distribution and Tax Structure: Empirical Test of the Meltzer-Richard Hypothesis". *European Economic Review*. 48: 805–826.

Canes-Wrone, B. 2015. "From Mass Preferences to Policy". *Annual Review of Political Science*. 18: 147–165.

Clinton, J. D. 2012. "Using Roll Call Estimates to Test Models of Politics". *Annual Review of Political Science*. 15: 79–99.

Cox, G. W. and M. D. McCubbins. 2005. *Setting the Agenda: Responsible Party Government in the U.S. House of Representatives*. Cambridge: Cambridge University Press.

Dimick, M., D. Rueda, and D. Stegmueller. 2017. "The Altruistic Rich? Inequality and Other-Regarding Preferences for Redistribution". *Quarterly Journal of Political Science*. 11(4): 385–439.

Dimick, M., D. Rueda, and D. Stegmueller. 2018. "Models of Other-Regarding Preferences, Inequality, and Redistribution". *Annual Review of Political Science*. 21: 441–460. ISSN: 10942939. DOI: 10.1146/annurev-polisci-091515-030034.

Duggan, J. 2008. "Candidate Objectives and Electoral Equilibrium". In: *The Oxford Handbook of Political Economy*. Oxford: Oxford University Press. 1–21.

Elwert, F. and C. Winship. 2014. "Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable". *Annual Review of Sociology*. 40: 31–53.

Fiorina, M. P. 2017. *Unstable Majorities: Polarization, Party Sorting, and Political Stalemate*. Stanford, CA: Hoover Institution Press.

Fiorina, M. P., S. J. Abrams, and J. C. Pope. 2010. *Culture War? The Myth of a Polarized America*. Boston: Longman.

Fowler, A. and A. Hall. 2012. "Conservative Vote Probabilities: An Easier Method for the Analysis of Roll Call Data". URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2120720.

Franko, W. W. 2017. "Understanding Public Perceptions of Growing Economic Inequality". *State Politics & Policy Quarterly.* 17(3): 319–348.

Garand, J. C. 2010. "Income Inequality, Party Polarization, and Roll-Call Voting in the U.S. Senate". *The Journal of Politics.* 72(4): 1109–1128. ISSN: 0022-3816. DOI: 10.1017/S0022381610000563.

Grofman, B. 2004. "Downs and Two-Party Convergence". *Annual Review of Political Science.* 7: 25–46. ISSN: 10942939. DOI: 10.1146/annurev.polisci.7.012003.104711.

Iversen, T. and D. Soskice. 2015. "Information, Inequality, and Mass Polarization: Ideology in Advanced Democracies". *Comparative Political Studies.* 48(13): 1781–1813.

Jacobs, L. R. and J. Soss. 2010. "The Politics of Inequality in America: A Political Economy Framework". *Annual Review of Political Science.* 13: 341–364.

Judd, G. 2019. "Nominating Legislative Representatives". URL: http://www.gleasonjudd.com/Research/nomlegcand-web.pdf.

Krasa, S. and M. Polborn. 2014. "Social Ideology and Taxes in a Differentiated Candidates Framework". *American Economic Review.* 104(1): 308–322.

Krehbiel, K. 1993. "Where's the Party?" *British Journal of Political Science.* 23(2): 235–266.

Layman, G. C., T. M. Carsey, and J. M. Horowitz. 2006. "Party Polarization in American Politics: Characteristics, Causes, and Consequences". *Annual Review of Political Science.* 9: 83–110.

Lee, F. E. 2009. *Beyond Ideology: Politics, Principles, and Partisanship in the U.S. Senate.* Chicago, IL: The University of Chicago Press.

Lee, F. E. 2015. "How Party Polarization Affects Governance". *Annual Review of Political Science.* 18: 261–282.

Lewis, J. B., K. Poole, H. Rosenthal, A. Boche, A. Rudkin, and L. Sonnet. 2021. "Voteview: Congressional Roll-Call Votes Database". URL: https://voteview.com/.

Lupu, N. and J. Pontusson. 2011. "The Structure of Inequality and the Politics of Redistribution". *American Political Science Review.* 105(2): 316–336.

Mason, L. 2018. *Uncivil Agreement: How Politics Became Our Identity.* Chicago, IL: The University of Chicago Press.

McCarty, N. M., K. T. Poole, and H. Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches.* Cambridge: The MIT Press.

Meltzer, A. H. and S. F. Richard. 1981. "A Rational Theory of the Size of Government". *The Journal of Political Economy.* 89(5): 914–927.

Meltzer, A. H. and S. F. Richard. 1983. "Tests of a Rational Theory of the Size of Government". *Public Choice*. 41: 403–418. ISSN: 00485829. DOI: 10.1007/BF00141072.

Nilsson, A., C. Bonander, U. Strömberg, and J. Björk. 2021. "A Directed Acyclic Graph for Interactions". *International Journal of Epidemiology*. 50(2): 613–619.

Pearl, J. 2009. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.

Pearl, J. 2010. "The Foundations of Causal Inference". *Sociological Methodology*. 40: 75–149.

Poole, K. T. and H. Rosenthal. 1985. "A Spatial Model for Legislative Roll Call Analysis". *American Journal of Political Science*. 29(2): 357–384.

Rodden, J. 2010. "The Geographic Distribution of Political Preferences". *Annual Review of Political Science*. 13: 321–340.

Rubin, D. B. 2005. "Causal Inference Using Potential Outcomes". *Journal of the American Statistical Association*. 100(469): 322–331.

Scheve, K. and D. Stasavage. 2017. "Wealth Inequality and Democracy". *Annual Review of Political Science*. 20: 451–468.

Snyder, J. M. and D. Strömberg. 2010. "Press Coverage and Political Accountability". *Journal of Political Economy*. 118(2): 355–408.

Textor, J., B. van der Zander, M. K. Gilthorpe, M. Liskiewicz, and G. T. Ellison. 2016. "Robust Causal Inference Using Directed Acyclic Graphs: The R Package 'dagitty'". *International Journal of Epidemiology*. 45(6): 1887–1894.

Tullock, G. 1983. "Further Tests of A Rational Theory of the Size of Government". *Public Choice*. 41: 419–421.

Washington, G. 1796. "Farewell Address". *The National Archives*.

White, I. K. and C. N. Laird. 2020. *Steadfast Democrats: How Social Forces Shape Black Political Behavior*. Princeton, NJ: Princeton University Press.