What about the rest of the pie? A dynamic compositional approach to modeling inequality

Ali Kagalwala* Andrew Q. Philips† Guy D. Whitten‡

Forthcoming at Social Science Quarterly

Objective: To demonstrate how a novel method enhances our understanding of determinants of inequality.
Methods: We take advantage of recent advances in dynamic models of compositional dependent variables to simultaneously study trade-offs across multiple slices of the composition of income in the US between 1947 and 2014.
Results: Our analyses demonstrate the utility of dynamic compositional models of income shares. Factors which increase the income share of the top income group often also increase the income share of the second income group and decrease the shares of lower groups at different rates. Polarization and marginal tax rates have large effects on relative income shares, while returns to labor, returns to capital, and partisan control of congress have smaller, but statistically significant effects.
Conclusion: Our suggested approach allows researchers to effectively explore interesting variation across multiple income groups in response to changes in traditional determinants of inequality.

Keywords: Inequality; time series; compositional data

*Ph.D. Student, Department of Political Science, Texas A&M University; 2935 Research Parkway, 4348 TAMU, College Station, TX 77843-4348 (alikagalwala@tamu.edu)

†Assistant Professor, Department of Political Science, University of Colorado Boulder, UCB 333, Boulder, CO 80309-0333 (andrew.philips@colorado.edu).

‡Professor, Department of Political Science, Texas A&M University; 2935 Research Parkway, 4348 TAMU, College Station, TX 77843-4348 (g-whitten@tamu.edu).

There is a rich history of public and scholarly interest in the causes and consequences of economic inequality. This interest has intensified recently with new measures showing the trajectory of the wealthiest citizens relative to the rest of society. Theories about when and why the rich get richer abound. However, this is only part of the picture, and—we argue—only part of what scholars should be investigating. By definition, most measures of inequality are relative measures. Thus, by focusing on the wealthiest segment of society relative to everyone else, we miss a myriad of possibilities for how the same increase or decrease in the relative trajectory of the wealthiest individuals might play out in the lower strata.

Put differently, if we picture national income or wealth as a pie, many recent studies have focused on only one slice of the total pie relative to everything else, by using only the share of the top one percent or the top decile. In this study we take advantage of recent advances in dynamic models of compositional dependent variables to simultaneously study trade-offs across multiple slices of income categories. This approach allows us to ask and answer an important question central to recent debates about rising inequality; what happens to *all* segments of society as the theorized determinants of economic inequality rise and fall? A compositional approach to studying inequality allows us to examine both which segments of society benefit from, and which segments of society lose out to, changes in the Congressional share of Democrats, Congressional polarization, movements in the top marginal tax rate, and changes to returns to capital and returns to labor. That is, a compositional approach provides the full picture of how determinants of inequality increase the income shares of certain segments of society at the expense of others.

In the sections that follow, we begin with a brief overview of extant studies of inequality. We then adapt existing theoretical propositions about what determines the share of the highest income group to a compositional approach. We test these propositions using a dynamic compositional modeling approach with data on income distribution in the United States. We find that an increase in the percent of Democrats in Congress increases the income share of the upper-middle class, and that the rich benefit at the expense of the poor and middle class for increases in polarization, for increases in returns to capital, and for decreases in the top marginal tax rate. For increases in returns to labor, however, the increases in the income shares of the poor and middle class come from decreases in the income share of top earners. We conclude with a summary of how our findings contribute to the literature on inequality and discuss future directions for this research.

Current Approaches to Measuring Inequality

Recent work by scholars—measuring the concentration of income or wealth at the very top of the distribution—has provoked a deluge of both public outrage and scholarly interest (c.f. Piketty 2014; Piketty and Saez 2003, 2006; Atkinson, Piketty and Saez 2011). The main contribution of this work has been the provision of relatively long time series data from developed nations tracking the trajectory of the share of the top of the income and wealth distribution relative to the rest of society. Underlying this measurement strategy is the idea that, as the top group (most often the top one percent) gains a greater share of the income or wealth in a country, the rest lose out, implying a general increase in inequality. Another common approach is the use of a single summary measure of inequality, such as the Gini coefficient. This measure is used in articles ranging from the study of the effect of inequality on political polarization (Garand 2010), to the effect of globalization on income inequality (Ha 2012), or the effect of inequality on nationalist beliefs (Solt 2011).

While this work has made a substantial contribution to our understanding of inequality, it leaves us without the ability to ascertain any information about changes that may be occurring among the lower percentage income ranks of society, as the fortunes of the uppermost group rise or fall. By definition, measures of income share are compositional variables. By considering variation in one category of a compositional variable over time, current studies of the share of income for the top one percent are implicitly lumping the shares of the remaining 99 percent together. While this type of approach is certainly valid, it ignores other forms of potentially interesting variation; trade-offs between the middle class and the working class, for instance. It also assumes, rather than tests, where the breaks are in terms of the groups that are experiencing relative gains and loses. In other words, when the fortunes of the top group rise, how deep is that top group? Do other income categories ever benefit alongside the top 1 or 10 percent?

A rather obvious first answer to these questions is that both the trade-offs across groups and the depth of the top group likely depend on what is driving the type of trajectories observed by these studies. In our initial searches of the literature we were surprised to find relatively few works that empirically tested theoretical claims about the determinants of the top economic groups. Instead, various measures of economic inequality are frequently used as an independent variable to test claims about the consequences of inequality. When economic inequality is analyzed as a dependent variable, Gini indices and other summary measures continue to be the dominant choice of empirical scholars.

To develop a more systematic view of the literature along these lines, we conducted a systematic search of publications in top journals in Economics, Political Science, and Sociology. We initially identified all papers that mention the word "inequality", "income distribution", or "wealth distribution" in the *American Journal of Political Science*, the *American Political Science Review*, the *Quarterly Journal of Economics*, the *American Economic Review*, and the *American Sociological Review*, published between 2005 and 2016. We then winnowed our search down to all articles that tested for an empirical relationship between economic inequality (i.e., income or wealth) and some other variable.

Our findings from this review of the literature are presented in Table 1. Across the sample of 66 papers, we found that just over half of them (36) tested models in which a measure of economic inequality was the dependent variable.

In these papers, the dominant choice of dependent variable was a Gini index (e.g., Lee 2005; Beckfield 2006), or some other type of summary measure of income distribution such as the variance of income or wealth (e.g., Huggett, Ventura and Yaron 2011). Together, summary measures made up over three quarters of the measures of economic inequality employed as a dependent variable across the journals that we examined in our search. The remaining measures analyzed fell into three different categories that can be thought of as more or less representing the logic of Piketty and his colleagues. These included two studies that modeled the share of the top group, three studies that modeled the ratio of income/wealth shares between two groups, and three studies that analyzed the shares of multiple single groups in separate models. Since our study is limited to the United States, we show the number of articles for which exclusively US data are used in parentheses in each cell of Table 1. These articles comprise about 38 percent of the articles we found which test an empirical expectation about inequality.

Table 1: A Quantitative Survey of the Current Literature on Inequality

	AJPS	APSR	AER	QJE	ASR
No. Articles Testing an Empirical Expectation about Inequality	13 (4)	9 (1)	24 (11)	7 (3)	16 (7)
No. Articles Where Inequality is the Dependent Variable	2 (1)	2 (0)	15 (9)	6 (3)	11 (5)
Type of Dependent Variable					
GINI	0 (0)	0 (0)	4 (2)	2 (1)	7 (2)
Other Summary Measure	0 (0)	1 (0)	7 (6)	3 (2)	3 (2)
Top	0 (0)	0 (0)	1 (0)	0 (0)	1 (1)
Single Ratio	2 (1)	1 (0)	0 (0)	1 (0)	0 (0)
Multiple Single	0 (0)	0 (0)	3 (1)	0 (0)	0 (0)

Notes: *American Journal of Political Science*: AJPS, *American Political Science Review*: APSR, *Quarterly Journal of Economics*: QJE, *American Economic Review*: AER, *American Sociological Review*: ASR. Search terms were "inequality", "income distribution", or "wealth distribution", in articles published between 2005 and 2016. The numbers in parenthesis indicate cases in which exclusively US data were used. E.g., only 4 of the 13 AJPS articles tested an empirical expectation about inequality in the US.

The findings from our review of the literature can be summarized in two main points. First, there has been surprisingly little work done to quantitatively model the factors that cause income or wealth disparities in a society to increase or decrease; only about half of all studies estimated an empirical model. Second, of the empirical tests that have been carried out, nearly all of them have operationalized inequality as a single summary measure (e.g., Gini coefficient), thereby overlooking important variation present in the underlying percentiles of the income or wealth distributions.

As shown in Table 1, a few articles have offered a statistical test of factors that affect inequality, in particular political determinants. For instance, Kelly (2005) tests how aggregate policy shifts affect inequality. Bonica et al. (2013) advance a theory about the relationship between political polarization and economic inequality. They argue that during periods of polarization, gridlock ensues, which leads to a lack of innovation in social safety nets and regulatory frameworks (p. 105). In other words, greater polarization slows substantial policy change. This works to the advantage the top one percent at the cost of everyone else, thus contributing to increased inequality.

Like Bonica et al. (2013), Volscho and Kelly (2012) also articulate a theory rooted largely in partisanship. They assume that the working and middle classes possess different, more egalitarian distributional preferences than top earners. Leftist parties and labor unions promote policies aimed towards creating such distributions, rather than the citizens towards the top of the income distribution (Volscho and Kelly 2012, p. 681). The result of this model is that when leftist parties are in power, and when marginal tax rates increase, there is a shift towards greater equality. In contrast, when parties of the right are in government the result is greater inequality.

While these studies have made valuable contributions to our theoretical and empirical understanding of the trajectory of the top economic groups over time, they have provided us with somewhat incomplete answers to the question posed at the beginning of this section; first, what drives trade-offs across groups other than the trade-off between the top and everyone else? By disaggregating the composition of wealth or income into multiple categories—not just the top incomes relative to all other incomes—we may see that not all income categories decrease uniformly in response to an external change. Second, how expansive is the upper income group? Although evidence suggests that income gains have been accruing to the top income category throughout most of the second half of the 20^{th} Century (e.g., the top one, five, or ten percent, Piketty and Saez 2003; Volscho and Kelly 2012), have other top income brackets—the 80 through 90th percentiles, for instance—also experienced a rising share of income? To answer these questions, we propose a series of theoretical claims and test the determinants of trade-offs across all income groups in a society.

Theories About Income Compositions

In this section, we extend theories about income inequality regarding the super-rich relative all other groups in a society to theories about trade-offs across all groups in a society. In other words, we develop theories about income compositions. To do so, we divide society first into income quintiles, and then further divide the top income quintile into the top five percent and those individuals whose income puts them in the sixth to twentieth percentile ranks (i.e., 0-19%, 20-39%, 40-59%, 60-79%, 80-94%, and the top five percent).[1] We then reconsider the collection of independent variables that have typically been theoretically linked to rising or falling income inequality and develop expectations about their impact on the relative shares of income across these six groups.

[1] We divide the income composition into quintiles plus the distinction between the top 5% and the rest of the top quintile since these were the most disaggregated data available over a long enough time series. We follow the lead of Volscho and Kelly (2012) and Philips, Souza and Whitten (2020) and use measures of pre-tax income. As Volscho and Kelly (2012) discuss, we are interesting in examining the impact of factors that shape the abilities and the incentives faced by all income groups. As such, we want our measures to be free of the leveling effects caused by progressive taxation. This allows us to better measure the effects of our independent variables of theoretical interest.

We begin this extension with a reconsideration of the impact of polarization on the composition of income groups. In Table 2 we compare the claims of an approach which studies the trajectory of the top income group relative to all other income groups. This "Top Group Approach" is the extent to which current theoretical claims are able to be empirically tested, as shown by our systematic journal search in Table 1. The right column in Table 2 illustrates our compositional theory. Unlike current approaches that simply hypothesize whether or not inequality will increase or decrease in response to a particular change, we highlight where in the income distribution this change is coming from.[2]

Table 2: Summary of Theoretical Expectations

	Expected Relationships	
	Top Group Approach	**Compositional Approach**
Independent Variable		
Polarization	Increased polarization benefits the top at the expense of all others $A \Uparrow \ (B+C+D+E+F) \downarrow$	Increased polarization benefits the top and bottom at the expense of the middle $A,B \uparrow \ C,D \downarrow \ E,F \uparrow$ or $\rightarrow$
Top Marginal Tax Rate	A cut in the top marginal tax rate leads to increases for the top relative to all others $A \Uparrow \ (B+C+D+E+F) \downarrow$	A cut in the top marginal tax rate decreases for top groups relative to the bottom, with little effect on the middle $A,B \uparrow \ C,D \rightarrow$ or $\downarrow \ E,F \downarrow$
Left Power	Increased left power leads to decreases for the top relative to all others $A \Downarrow \ (B+C+D+E+F) \uparrow$	Increased left power leads to decreases for top groups relative to the bottom $A,B \downarrow \ C,D \downarrow$ or $\rightarrow \ E,F \uparrow$ or $\rightarrow$
Returns to Capital	Increased returns to capital leads to increases for the top relative to all others $A \Uparrow \ (B+C+D+E+F) \downarrow$	Increased returns to capital leads to increases for top groups relative to the bottom $A,B \uparrow \ C,D \downarrow$ or $\rightarrow \ E,F \downarrow$ or $\rightarrow$
Returns to Labor	Increased returns to labor leads decreases for the top relative to all others $A \downarrow \ (B+C+D+E+F) \uparrow$	Increased returns to labor leads to to increases for bottom groups relative to the top $A,B \downarrow \ C,D \downarrow$ or $\rightarrow \ E,F \uparrow$

Note: With our current data, we have 6 income groups: A is the top 5%, B is the 80-94th percentile, C is the 60-79th percentile, D is the 40-59th percentile, E is the 20-39th percentile, and F is the 1-19th percentile.

One factor that may affect inequality is polarization. While the theory given by Bonica et al. (2013) provides a plausible expectation as to how polarization leads to greater concentrations of income among the super-rich, it does not suggest which specific groups experience decreases in relative incomes in such a scenario. If we view polarization from a Downsian perspective, it means a movement from the classic Downsian unimodal scenario depicted on the

[2] As is apparent from Table 1, our theoretical ideas lump the groups into together such that our expectations are about different effects on A+B vs. C+D vs E+F. Given this, it would be reasonable to lump the bottom two quintiles (E+F) and the middle quintiles (C+D) together. We would still want to separate groups A and B from each other to see whether the effects of our variables apply mostly to the very top group or, as we expect whether the effects of variable benefiting or hurting the very top income group extend to the rest of that quintile of the income distribution. In Sections 4 and 5 of our supplemental information document, we present results from models in which we have lumped the income groups into 3 (A+B vs. C+D vs E+F) and 4 (A vs. B vs. C+D vs E+F) categories. As we discuss in the supplemental information document, our main findings are robust to this lumping of income categories together.

left plot in Figure 1 to a bimodal scenario like that depicted on the right. It is reasonable to expect that the income ranks and left-right positions of individuals will generally be highly correlated (c.f. Boix 1998). Under a polarized political scenario, when policies are passed, we expect they will be targeted at either the right or left peak (depending on the party in government) rather than the median voter or middle of the income distribution. Under such conditions, our expectation is that relative increases in income will go to either the top or bottom of the income distribution, and that these gains will come disproportionately from the middle income/ideology groups. Thus, as shown in Table 2, while the top group approach simply predicts an increased income share for the top group, we expect that increased polarization will lead to a "hollowing out" of the relative income shares of the middle incomes, with relative gains instead accruing to the top and bottom of the income distribution.

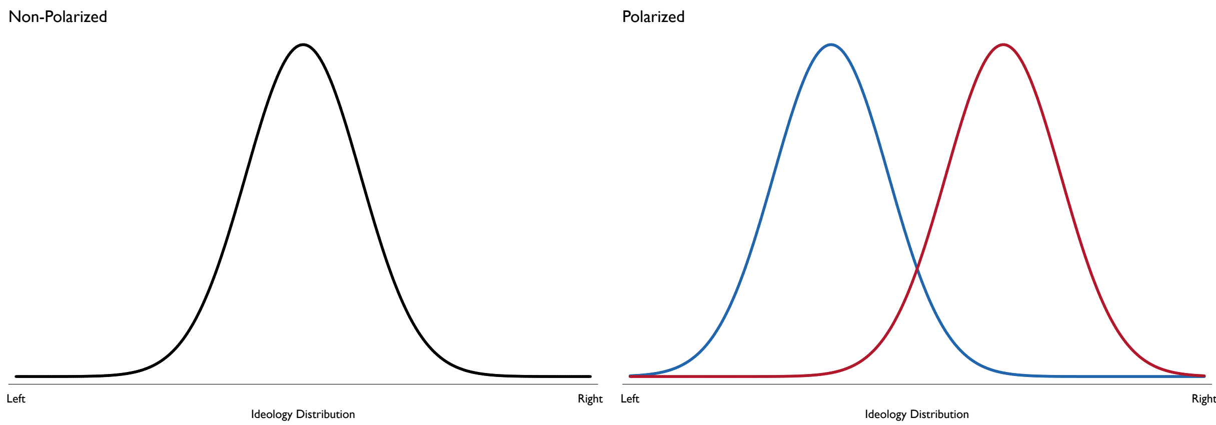


Figure 1: Polarization May Exacerbate Inequality

For obvious reasons, tax rates have been a salient policy topic, since they have the ability, in theory, to reallocate income to lessen the effects of inequality. In the US, income tax rates are progressive, since top earners typically are taxed a larger percentage of their income.[3] Therefore, given a decrease in the top marginal tax rate, a straightforward expectation is that relative income shares of the top incomes will increase. As shown in Table 2, given current approaches, we would expect that in response to a cut in the top marginal tax rate, the top incomes experience a relative increase in income, while the others experience a relative decrease. Yet this is not the entire picture; given our compositional theoretical approach, we expect that not only do top incomes gain in response to a tax cut, the lowest income quantiles—those who would have benefitted from increased redistribution—will experience a relative decrease in their shares of income. We expect the middle quantiles to experience little to no change.

Left-leaning parties are typically associated with a preference for increasing redistributive policies (Kelly 2005; Bartels 2009). The power-resource theory holds that left parties, supported by the middle and lower class, may work

[3]Although there is some evidence that after the 2017 tax cuts, the very highest earners pay a lower rate than other income categories (Saez and Zucman 2019).

to keep the super rich out of important political posts and limit their rent-seeking ability, thus lowering the likelihood that policies get passed that exacerbate inequality (Volscho and Kelly 2012; Kelly and Witko 2012). Although the current top group approach might theorize that top income shares will decrease as a result of increased left power in government, our compositional theory in Table 2 is more nuanced. Not only are top incomes affected, we also hypothesize that the middle income groups might experience a slight decrease in income shares (or no change), and that the bottom two quantiles will experience a slight increase, as redistributive policies should lower inequality.

Last, there has been a renewed focus on how relative differences between the returns to labor and the returns to capital may exacerbate inequality (Piketty 2014). Capital shocks, for instance the loss of businesses during the Great Depression, have been shown to be associated with large shifts in top income shares (Piketty and Saez 2003). We therefore expect that top income categories rise as the returns to capital increase, and that this comes at the expense of relative decreases in incomes for the other categories. Note that in Table 2 our compositional expectation for capital is similar to the top group approach. In contrast, given an increase in returns for labor, we expect increases in the income shares for the bottom two quantiles, no change (or a slight decrease) in the next two quantiles, and a decrease in the top income groups.

A Compositional Approach to Modeling Inequality

As discussed above, we argue that modeling the composition of relative income shares offers a more complete picture of income inequality. In order to implement this strategy, we use a methodological approach first pioneered in geology (Aitchison 1982, 1983), and later applied to political science (Katz and King 1999; Tomz, Tucker and Wittenberg 2002; Basinger, Cann and Ensley 2012; Philips, Rutherford and Whitten 2015, 2016*a*).

Our approach is as follows. Let a society comprised of J income categories (e.g., quantiles) in each year, t, be expressed as a proportion of the total, Y_t such that the income categories sum to unity $\sum_{j=1}^{J} y_{tj} = 1 \; \forall t$. This unity constraint leads to a number of issues as we attempt to model the determinants of each y_{tj}. First, each proportion is bounded between zero and one, making standard unbounded multivariate regression inappropriate (Katz and King 1999).[4] Second, modeling each category separately omits important information given by the compositional nature of the data. For example, given three categories—A, B, and C—an increase in category A must necessarily come at the cost of either category B or C, or both. However, running three separate regressions on categories A, B, and C will omit these correlations. A third problem is that methods designed to account for cross-category correlations—such as a seemingly unrelated regression or vector autoregressive model—face a redundant degree of freedom problem. That

[4]For instance, a regression of y_{tj} might lead to predicted values that are above one or below zero. This is somewhat analogous to how a linear probability model results in non-sensical predictions that are below zero or above one.

is to say, running a seemingly unrelated regression on categories A, B, and C will force one category to drop out of the model, since the other two perfectly predict it (i.e. $1-A-B=C$). Due to this constraint in the composition, only $J-1$ out of the J categories can be modeled, which makes substantive analysis difficult if we are concerned with *all* categories.[5]

In order to "unbound" the composition from the zero-one constraint, and to reduce the number of systems simultaneously modeled from J to $J-1$, we use a log-ratio transformation. First attributed to Aitchison (1982, 1983), this approach involves taking the log of the ratio of each of the j categories relative to a baseline category:

$$s_{tj} = ln\left(\frac{y_{tj}}{y_{t1}}\right) \forall j \neq 1 \tag{1}$$

Note that although the income category y_{t1} is the baseline category, this choice is arbitrary and does not change the results.[6] This transformation has two advantages. First, taking the log-ratio frees the s_{tj} series from the zero-one constraint, allowing standard multivariate approaches to be used (Tomz, Tucker and Wittenberg 2002). Second, since there are now only $J-1$ logged ratio series, s_{tj}, we can include each of them as dependent variables in a system of equations, which gets around the redundant degree of freedom problem. A system of equations approach also provides the ability to take into account the correlated error structure across time, as explicitly modeled using methods such as seemingly unrelated regression (SUR) or vector autoregressive models (VAR) (Brandt, Monroe and Williams 1999; Jackson 2002; Mikhailov, Niemi and Weimer 2002; Basinger, Cann and Ensley 2012; Philips, Rutherford and Whitten 2016*a*).[7]

Data and Empirical Approach

Time series data on income or wealth inequality that disaggregate the distribution into multiple categories are rare. To the best of our knowledge, the most comprehensive source comes from the US Census Bureau's Current Population Survey on household income shares, which are available from 1947 through 2014.[8] The income shares are calculated based on reported taxable income, which includes 42 possible income categories, among them labor income, dividend

[5]For example, in examining party support in the United Kingdom, one could model support for Labour, Conservatives, and the Liberal Democrats simultaneously, which would implicitly lump all other parties into an "other" category.

[6]For a discussion on this, see Philips, Rutherford and Whitten (2016*a*). The log-ratio transformation cannot be used if any one of the categories is equal to zero. There are various approaches to dealing with zero categories (e.g. Martin-Fernandez, Barcelo-Vidal and Pawlowsky-Glahn 2003; Aitchison and Egozcue 2005), but these are outside the scope of this paper.

[7]The log-ratio transformation, sometimes known as the additive log-ratio (alr) transformation, is not the only transformation available for use in modeling of compositional data. For instance, Brandt, Monroe and Williams (1999) use a symmetric (i.e. centered) log transformation to analyze partisanship in the US; the model is given as: $s_{tj} = ln\left(\frac{y_{tj}}{\Pi_{j=1}^{J} y_{tj}^{1/j}}\right) \forall j$. While this approach unbounds the series like the alr transformation, it does not reduce the number of categories that must be estimated, meaning that the degree-of-freedom problem persists.

[8]Although these series are available through 2018, limitations in the data availability of our independent variables restrict us between 1952 and 2014.

income, interest income, trust or estate income, and government assistance income. Since our compositional approach allows us to take advantage of the most disaggregated measures of relative income possible, we divide reported income into the bottom four quintiles of the distribution (0-19%, 20-39%, 40-59%, and 60-79%). We divide the top quintile into the top five percent and the top 80-94 percent. Although further disaggregations would be ideal, these are the most disaggregated income groups available.

It is worth noting a substantial difference between our data and some of the most commonly examined series in the extant literature. While the top income group in our analysis is the income share of the top five percent, other articles have typically examined the income or wealth share of the top one percent (or even a smaller group). We expect this difference to be minimal because our top five percent series is highly correlated ($r = .91$) with the top one percent measure provided by Piketty and Saez (2006) over the period where both measures are available (1947 to 2008).

The over-time trajectories of the six income categories are shown in Figure 2. Each category is shown as a proportion of the total income share. As is clear from the figure, the bottom three quintiles have experienced a relative decrease in their proportion of total income share; this appears to be most pronounced after 1980. In contrast, the second highest quintile (60 to 79 percent) remains at a relatively constant proportion that rises slightly until around 1980 and then trails off slightly thereafter. The largest movement in Figure 2 is in the top quintile, especially for the top five percent. While income shares of the top five percent decreased from around 17.5 percent in 1947 to just under 15 percent in 1980, since then, the relative share of income has steadily increased to over 20 percent by 2014. In contrast, the richest 80th to 94th percent of Americans have slowly increased their share of total income from about 25.5 percent in 1947 to around 28 percent in 2014. These substantial differences in relative income trajectories would have been overlooked had we employed a single measure of inequality.

For an initial test of our hypotheses about the changes in income compositions, we use an error-correction model to estimate determinants of the composition of income. The model is specified as

$$\Delta s_{tj} = \alpha_j + \phi_j s_{t-1,j} + \boldsymbol{\Delta x}_{tj} \boldsymbol{\beta}_{1j} + \boldsymbol{x}_{t-1,j} \boldsymbol{\beta}_{2j} + \varepsilon_{tj} \tag{2}$$

where the change in log-ratio s_{tj} is a function of a constant, its own lag appearing in levels, a vector of the first difference of four weakly exogenous regressors, a vector of the one-period lag of the same four vectors, and an error term that may be contemporaneously correlated across the $J-1$ equations. As discussed in Table 2, we include a number of independent variables to test our theoretical expectations about the determinants of inequality. For our measure of left power, we use the percent of Democrats in Congress, with the expectation that increasing Democratic control will lead to decreases in the top four income categories (the upper and middle class) relative to the lowest two quintiles. To see the effects of redistribution on inequality, we include a measure of the top marginal tax rate.

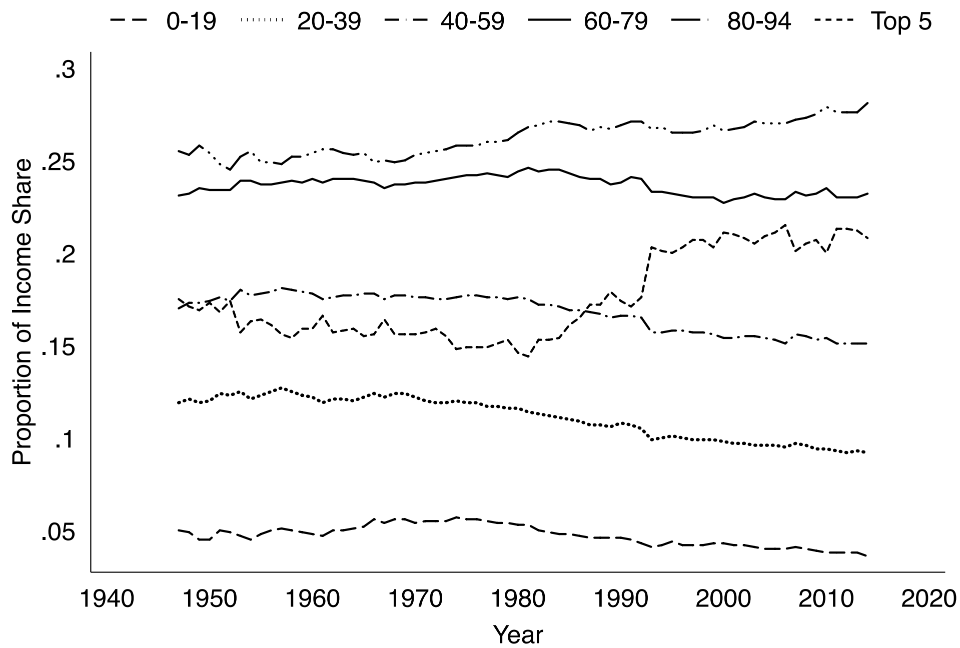


Figure 2: The Proportion of Income in the US, 1947 to 2014

As outlined in Table 2, we expect that increases in the top tax rate will lead to decreases in the share of the income distribution held by the top quintile, no change for the middle quintiles, and a relative increase for the lowest two quintiles. Since we hypothesized that increased polarization has led to a "hollowing out" of the middle class' policy priorities, to the benefit of the poorest and richest, we include a measure of Congressional polarization using D-NOMINATE.[9] To account for returns to capital, we add the annual returns of the S&P 500. We expect that increased returns for capital will lead to increases in the relative incomes of the top income shares, all else equal. Last, to measure returns to labor, we include annual wage growth, with the expectation that increased returns to labor will lead to increases for the lowest income shares at the cost of the highest income shares. Information on the data sources are available in the supplemental information document.

Before estimating Equation 2, we first checked to ensure that all our series contained a unit-root. Dickey-Fuller GLS unit root tests with two lags indicated that all five logged ratios and all independent variables (except for returns to capital) were non-stationary.[10] Next, we performed a number of tests for cointegration. Since there are five equations in our system of equations, we tested each equation separately. The two-step Engle-Granger test finds evidence of cointegration in all models at the five percent level or better. However, Philips (2018) shows that the Engle-Granger

[9] We use the same measure of polarization as Bonica et al. (2013) which is calculated by first calculating the average ideal point for each individual legislator. The average score of Democratic and Republican legislators are then calculated and the difference between them is then used as a measure of polarization. Since these ideal points are constructed so as to be comparable across time, this measure of polarization is well-suited to our purposes. For more details on the use of of D-NOMINATE to measure polarization over time in the U.S., see Hare and Poole (2014).

[10] Returns to labor was a borderline case.

test is susceptible to high Type I error in short series, and instead suggests the use of the ARDL-bounds test for cointegration. Using it we find inconclusive results for cointegration at either the 5% or 10% significance level. Due to the inconclusive results, we check for cointegratoin using the Johansen test, and find that all five equations are cointegrating. Thus, we are relatively confident in our choice of an error correction model.

Although a table of regression results can be found in Table 3 in the supplemental information document, interpreting error-correction models in the context of compositional data (where each dependent variable is the logged ratio of one category divided by some arbitrary baseline category) is neither straightforward nor particularly informative. Instead, we used the Stata program `dynsimpie`, which—similar to an impulse response function in VAR models—graphically depicts how each income category changes over time in response to a counterfactual one-period "shock" to one of the independent variables (Philips, Rutherford and Whitten 2016*b*). The program sets all variables in the system to stable (i.e. sample mean) values, and generates expected values at time $t = 1, 2..., T-1, T$. All first differences are set to zero. At time $t = 4$, we specify a one-period, one standard deviation change in one of the independent variables, holding all others constant. This effect then moves through the system of equations. Last, the program calculates expected values, as well as the upper and lower values of 95 percent confidence intervals.

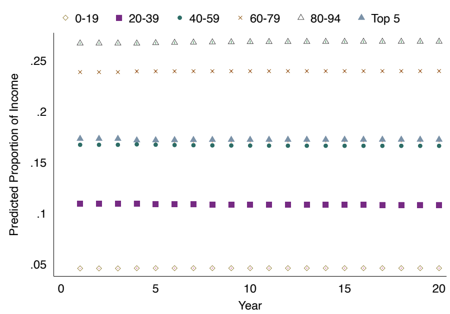
In addition to graphing the expected values using `dynsimpie`, we also interpret our results using effects plots. These show the results from the same simulation, but graph the contemporaneous and long-run changes from an income category's sample mean. This is analogous to a marginal effects plot with compositional data, and is useful for seeing the overall significance of a variable's effect on the compositions, as well as examining which income categories increase or decrease in response to a change.

Results

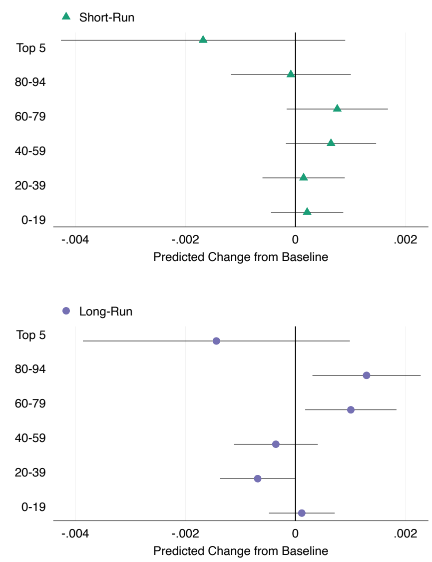
To present the results from our model, we first show the estimated effects of a shift in each independent variable with two different plots in the same figure. In the left panel of each of these figures, we show a dynamic simulation from a one standard deviation shift in the particular independent variable at time $t = 4$. Each of the dynamic simulations shows the expected value (with 95 percent confidence bounds) for the income share of each group. In the right panels of each of these figures, we use effects plots to show the estimated short-run (i.e., one period) effect (top right) and the estimated long-run effect (bottom right) of each shift on the relative share of each income group.

Figure 3 demonstrates the effect of a one standard deviation increase in the percentage of Democrats in Congress. Our theoretical expectation was that increases in Congressional left power would increase the income share of the 0-19 and 20-39 percentiles at the expense of the income share of the middle class and high income groups. From Figure

3a, we observe that there are, at best, slight changes in the income shares of all groups. Figure 3b allows us to more closely scrutinize our results. We find that the income shares of the bottom four quintiles (0-79 percentiles) increase and those of the top two income groups decrease in the short run. In the long run, however, we find that the income shares of the 60-94 percentiles increase and this gain in income comes from losses to the 20-59 percentiles and the top five percent. There are almost no increases in the income share of the poorest income group in both the short- and the long-term. However, our findings for the short-term effects are not statistically significantly different from zero for all income groups. The long-term effects on income shares are statistically different from 0 for only the 60-94 percentiles. Overall, our results do not support our expectations.



(a) Expected Change Over Time



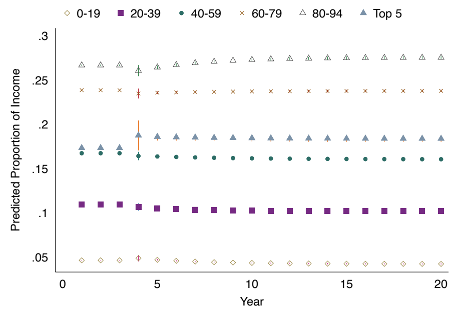
(b) Short and Long Run Effects

Figure 3: 1 Standard Deviation Increase in % Democrats in Congress

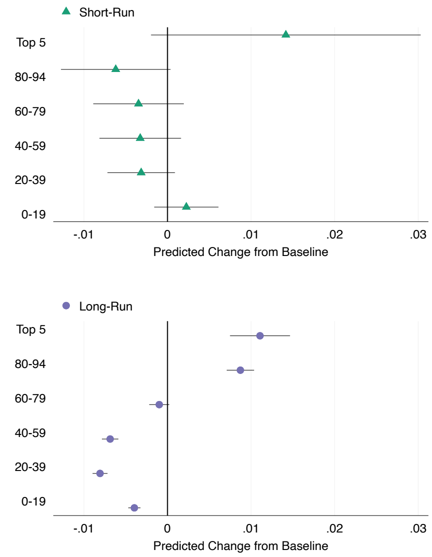
Notes: Left plot shows the expected income shares due to a one-period +1 standard deviation shock to % Democrats in Congress at $t = 4$. Right plot shows the contemporaneous and long-term changes from baseline (sample mean) for each category for the same simulation. All other variables held at sample means. 95% confidence intervals reported.

In Figure 4, we show the estimated effects of a one standard deviation increase in the polarization of Congress. As shown in Figure 4a, when polarization increases at $t = 4$, gains in the expected proportion of income for the top five percent and the 80-94 percentiles appear to come disproportionately from the 20-39th and 40-59th percentiles. In other words, it does not appear that relative gains in income share of the top five percent, in response to increasing polarization, come uniformly from all other income shares; instead, it is the 20-59th percentiles that face the largest relative decreases in income share.

Figure 4b makes the previous observation that gains and losses are non-uniform even more clear. In the short-run, both the top five percent of income earners, as well as the bottom 0-19th percentiles, experience an increase



(a) Expected Change Over Time



(b) Short and Long Run Effects

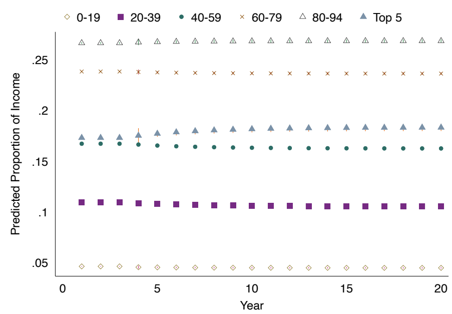
Figure 4: 1 Standard Deviation Increase in Polarization

Notes: Left plot shows the expected income shares due to a one-period +1 standard deviation shock to polarization at $t = 4$. Right plot shows the contemporaneous and long-term changes from baseline (sample mean) for each category for the same simulation. All other variables held at sample means. 95% confidence intervals reported.

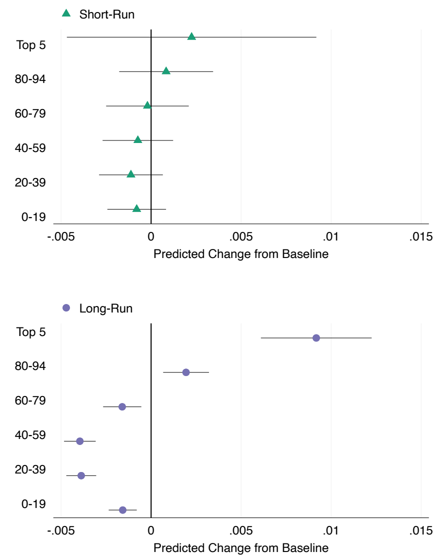
in response to increasing polarization. These short-run changes come approximately evenly from all other income categories. Over the long-run, a one standard deviation increase in polarization results in the top five percent and 80-94 percentiles receiving about two percentage points more of the income share. However, relative losses to the other four categories come mostly from the bottom-middle and bottom income groups. Overall, these results show that the effects of polarization are different across the groups beneath the top income categories, and that the greatest loses accrue to the bottom-middle income groups. This partially supports our theoretical expectations; while the bottom income category experiences a decrease in response to increased polarization, the bottom-middle incomes experience the largest decreases. One reason why the lowest income categories do not experience a relative increase in income shares in response to increased polarization may be that increased gridlock results in less legislative output, including redistributive policies that benefit the lowest incomes (Bonica et al. 2013).

In our theoretical table, we also expected tax rates to affect the composition of incomes. In Figure 5, we show the estimated effects of a one standard deviation decrease in the top marginal tax rate. It is clear that changes in the top marginal tax rate have no statistically significant short-run effect on the composition of income. In contrast, over the long run, the 80th percentile and above enjoy over a one percentage point increase in the proportion of income. Similar to the results for polarization, these relative increases come mostly from the bottom-middle classes (20-59th percentiles), although the substantive magnitude of the changes are around half as large. Those in the 0-19th and

60-79th percentiles also experience decreases in income shares. Thus, as expected, decreases in top tax rates lead to increased income concentration for the top quintile at the expense of the bottom four quintiles.



(a) Expected Change Over Time



(b) Short and Long Run Effects

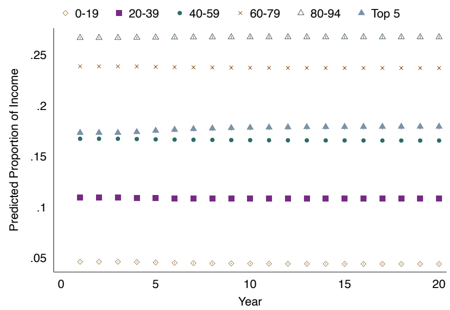
Figure 5: 1 Standard Deviation Decrease in Top Marginal Tax Rate

Notes: Left plot shows the expected income shares due to a one-period -1 standard deviation shock to the top marginal tax rate at $t = 4$. Right plot shows the contemporaneous and long-term changes from baseline (sample mean) for each category for the same simulation. All other variables held at sample means. 95% confidence intervals reported.

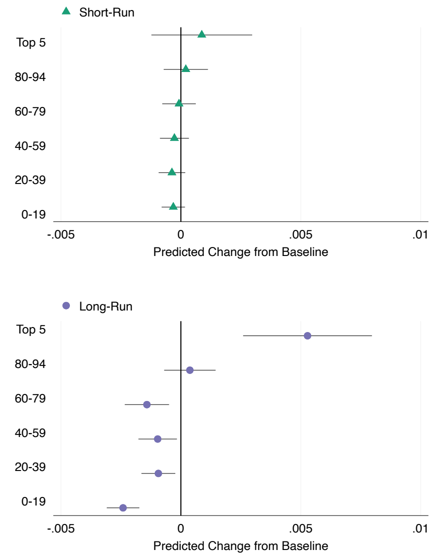
The last two plots focus on the trade-off between labor and capital, and the effect that these factors have on income distributions. In Figure 6, we show the estimated effects of a one standard deviation increase in the returns to capital, holding all else constant. Although there are no statistically significant changes in the short-run, over the long-run, the proportion of income held by the top five percent increases by about 0.5 percentage points.[11] While all other categories—except for the 80-94 percentiles—end up with a slight decrease in their relative income shares, it is the lowest income quintile (0-19), that experiences the largest decrease in income share, albeit less than half a percentage point. This lines up with our theoretical expectations; increases in capital tend to benefit the higher-income segments of society.

We see virtually the opposite effect in Figure 7, which shows the estimated effect of a one standard deviation increase in the returns to labor. In the long-run the proportion of income held by the top five percent decreases by about 0.75 percentage points, while all other income categories experience an increase (although the increase for the 20-39th percentiles is not statistically significantly different from zero). Note too that in the short-run, there is also a

[11] As noted by Piketty and Saez (2003), labor income still accounts for much of the total income of all but the top one percent of incomes. Unfortunately, the top five percent is the most disaggregated income category available in our dataset, so it is difficult to tell if the top one percent is driving these results. Still, we think that capital returns are important since high earners still earn relatively more of their total income from capital than labor than do lower income quantiles.



(a) Expected Change Over Time



(b) Short and Long Run Effects

Figure 6: 1 Standard Deviation Increase in Returns to Capital

Notes: Left plot shows the expected income shares due to a one-period +1 standard deviation shock to returns to capital at $t = 4$. Right plot shows the contemporaneous and long-term changes from baseline (sample mean) for each category for the same simulation. All other variables held at sample means. 95% confidence intervals reported.

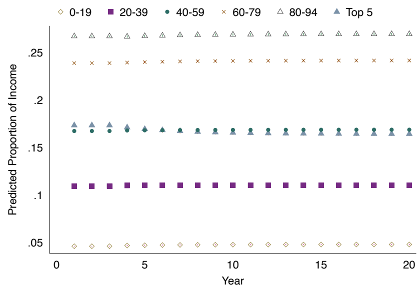
statistically significant, though small, increase in the proportion of income held by the bottom quintile. These findings, coupled with the findings in Figure 6, suggest that there appears to be a trade-off between wage income and capital income, but that these drive the highest and lowest income groups more than the middle.

Discussion and Conclusion

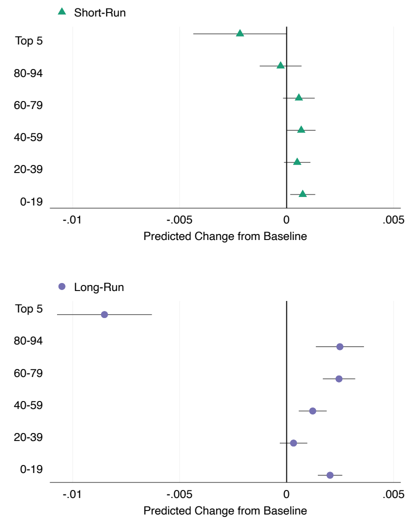
In this paper we have demonstrated the utility of taking a more nuanced compositional approach to the measurement and analysis of income inequality. While previous work has typically either used a single summary measure (such as the Gini coefficient), or examined the over-time trajectory of the top income category, our results show that there is interesting variation worthy of exploration when we think about the more complicated "pie" of the income distribution. We used this compositional approach to develop theoretical expectations that are a better reflection of the shifts that occur across multiple income categories in response to changes in polarization, left control, taxation, and the returns from labor and capital.

In Table 3, we summarize our empirical results together with results from models estimated using approaches common in the extant literature estimated with the same data.[12] From this table, we can see the utility of our suggested

[12] The full results from these additional models are presented in our supplemental information document.



(a) Expected Change Over Time



(b) Short and Long Run Effects

Figure 7: 1 Standard Deviation Increase in Returns to Labor

Notes: Left plot shows the expected income shares due to a one-period +1 standard deviation shock to returns to labor at $t = 4$. Right plot shows the contemporaneous and long-term changes from baseline (sample mean) for each category for the same simulation. All other variables held at sample means. 95% confidence intervals reported.

approach. In the model of a Gini Index constructed using these same data, summarized in the next-to-last column of Table 3, only one variable, polarization, was found to be a statistically significant determinant of income inequality. In the model of the over-time trajectory of the income share for the top five percent, presented in the final column of Table 3, none of our independent variables were found to be significant. In contrast, across the first six of columns of results summaries in Table 3, we can see that our model yields statistically significant estimated effects for each independent variable on at least two income groups. This suggests that analyzing the entire income distribution provides deeper insights into the underlying dynamics of changes in inequality, something we miss using a summary measure such as Gini or by only analyzing the top income group.

As discussed in the previous section, most of our results comport with our expectations as laid out in Table 2. The main exceptions to this are in terms of the estimated long-run consequences of an increase in the percentage of Democrats in Congress and the effects of increased polarization. Our expectation was that an increase the percentage of Democrats in Congress would help the bottom groups at the expense of the top and middle-income groups since the Democrats are traditionally thought of as the party most favoring more generous redistributive policies. Instead, we found that the long-run effect of such an increase is a relative income gain for the top 60-94 percent of the income distribution. Although the estimated long-run effects for the top and bottom are in the expected negative and positive directions, they are not statistically significant. These findings certainly warrant further study. Our long-run estimates of the effects of an increase in polarization did lead to the expected relative income increases in relative income shares

Table 3: Summary of Results

Independent Variable				**Income Group**				**Traditional Measures**	
	Effect	**0-19**	**20-39**	**40-59**	**60-79**	**80-94**	**Top 5%**	**Gini Index**	**Top Group (5%)**
↑ Polarization									
	Short-Run	–	–	–	–	–	–	–	–
	Long-Run	↓	↓	↓	–	↑	↑	↑	–
↓ Top Marginal Tax Rate									
	Short-Run	–	–	–	–	–	–	–	–
	Long-Run	↓	↓	↓	↓	↑	↑	–	–
↑ Left Power									
	Short-Run	–	–	–	–	–	–	–	–
	Long-Run	–	–	–	↑	↑	–	–	–
↑ Returns to Capital									
	Short-Run	–	–	–	–	–	–	–	–
	Long-Run	↓	↓	↓	↓	–	↑	–	–
↑ Returns to Labor									
	Short-Run	↑	–	–	–	–	–	–	–
	Long-Run	↑	–	↑	↑	↑	↓	–	–

Notes: Each cell shows either the short-run or long-run direction of estimated effect from the change to the income variable listed in the first column. A ↑ indicates that we found a statistically-significant increase, a ↓ indicates that we found a statistically-significant decrease, and – indicates that the estimated relationship was not statistically significant. 95% confidence intervals reported. For full details on these models, please see our supplemental information document.

for the top two income groups that we analyzed, but, contrary to our expectations, these gains came at the expense of all other income groups. We had expected that an increase in polarization would either help or have no effect on the bottom two groups. These two sets of findings warrant further investigation.

Despite the oft-heard claim that income inequality has risen in the US, our analysis shows that gains to the top income shares have not come uniformly from all lower-income brackets. In other words, there are theoretically interesting and different consequences to what happens when the rich get richer. We have found that it is often the middle income quantiles that have experienced the largest decreases in their share of income, especially in response to growing polarization and decreases in top marginal tax rates. Our results also demonstrate the utility of thinking more about the scope of gains to the top income groups. While a lot of the scholarly and new-media focus has been on the top 1% or even the top .1%, our analyses show that many of the factors which drive up the relative income share of the top 5% also benefit the rest of the top income quintile (the 80-94% group in our analyses). The combination of this insight with evidence that income is a strong predictor of voter turnout (e.g., Franko, Kelly and Witko (2016)) implies that efforts aimed at changing the income structure might face stronger political resistance than previously imagined.

The analysis in this paper has several extensions worth pursuing. First, future work might compare income distributions across institutional and economic divides by taking advantage of income distribution data in other countries.

Although polarization had a substantial effect on income distribution in the US, this may be very different when examining social democratic countries in Europe, for instance. Second, although current data availability limited us to income distribution, a compositional approach to wealth distributions might lead to very different conclusions. In future work we hope to identify further sources of dynamic compositional data on economic inequality to allow us to further explore these interesting relationships.

References

Aitchison, John. 1982. "The statistical analysis of compositional data." *Journal of the Royal Statistical Society. Series B (Methodological)* 44(2):139–177.

Aitchison, John. 1983. "Principal component analysis of compositional data." *Biometrika* 70(1):57–65.

Aitchison, John and Juan José Egozcue. 2005. "Compositional data analysis: where are we and where should we be heading?" *Mathematical Geology* 37(7):829–850.

Atkinson, Anthony B, Thomas Piketty and Emmanuel Saez. 2011. "Top incomes in the long run of history." *Journal of Economic Literature* 49(1):3–71.

Bartels, Larry M. 2009. *Unequal democracy: The political economy of the new gilded age*. Princeton University Press.

Basinger, Scott J, Damon M Cann and Michael J Ensley. 2012. "Voter response to congressional campaigns: new techniques for analyzing aggregate electoral behavior." *Public Choice* 150(3-4):771–792.

Beckfield, Jason. 2006. "European integration and income inequality." *American Sociological Review* 71(6):964–985.

Boix, Carles. 1998. *Political Parties, Growth and Equality: Conservative and Social Democratic Economic Strategies in the World Economy*. New York, NY: Cambridge University Press.

Bonica, Adam, Nolan McCarty, Keith T Poole and Howard Rosenthal. 2013. "Why hasn't democracy slowed rising inequality?" *The Journal of Economic Perspectives* 27(3):103–123.

Brandt, Patrick T, Burt L Monroe and John T Williams. 1999. Time series models for compositional data. In *Annual Meeting of the American Political Science Association, Atlanta, Georgia, September*.

Franko, William W, Nathan J Kelly and Christopher Witko. 2016. "Class bias in voter turnout, representation, and income inequality." *Perspectives on Politics* 14(2):351.

Garand, James C. 2010. "Income inequality, party polarization, and roll-call voting in the US Senate." *The Journal of Politics* 72(4):1109–1128.

Ha, Eunyoung. 2012. "Globalization, government ideology, and income inequality in developing countries." *The Journal of Politics* 74(2):541–557.

Hare, Christopher and Keith T Poole. 2014. "The polarization of contemporary American politics." *Polity* 46(3):411–429.

Huggett, Mark, Gustavo Ventura and Amir Yaron. 2011. "Sources of lifetime inequality." *The American Economic Review* 101(7):2923–2954.

Jackson, John E. 2002. "A seemingly unrelated regression model for analyzing multiparty elections." *Political Analysis* 10(1):49–65.

Katz, Jonathan N and Gary King. 1999. "A Statistical Model for Multiparty Electoral Data." *American Political Science Review* pp. 15–32.

Kelly, Nathan J. 2005. "Political choice, public policy, and distributional outcomes." *American Journal of Political Science* 49(4):865–880.

Kelly, Nathan J and Christopher Witko. 2012. "Federalism and American inequality." *The Journal of Politics* 74(2):414–426.

Lee, Cheol-Sung. 2005. "Income inequality, democracy, and public sector size." *American Sociological Review* 70(1):158–181.

Martin-Fernandez, Josep A, Carles Barcelo-Vidal and Vera Pawlowsky-Glahn. 2003. "Dealing with zeros and missing values in compositional data sets using nonparametric imputation." *Mathematical Geology* 35(3):253–278.

Mikhailov, Nikolai, Richard G Niemi and David L Weimer. 2002. "Application of Theil group logit methods to district-level vote shares: tests of prospective and retrospective voting in the 1991, 1993, and 1997 Polish elections." *Electoral Studies* 21(4):631–648.

Philips, Andrew Q. 2018. "Have your cake and eat it too? Cointegration and dynamic inference from autoregressive distributed lag models." *American Journal of Political Science* 62(1):230–244.

Philips, Andrew Q, Amanda Rutherford and Guy D Whitten. 2015. "The Dynamic Battle for Pieces of Pie—Modeling Party Support in Multi-Party Nations." *Electoral Studies* 39:264–274.

Philips, Andrew Q, Amanda Rutherford and Guy D Whitten. 2016*a*. "Dynamic Pie: A Strategy for Modeling Trade-Offs in Compositional Variables over Time." *American Journal of Political Science* 60(1):268–283.

Philips, Andrew Q, Amanda Rutherford and Guy D Whitten. 2016*b*. "dynsimpie: A command to examine dynamic compositional dependent variables." *Stata Journal* 16(3):662–677.

Philips, Andrew Q, Flávio DS Souza and Guy D Whitten. 2020. "Globalization and comparative compositional inequality." *Political Science Research and Methods* 8(3):509–525.

Piketty, Thomas. 2014. *Capital in the twenty-first century*. Harvard University Press.

Piketty, Thomas and Emmanuel Saez. 2003. “Income inequality in the United States, 1913–1998.” *The Quarterly journal of economics* 118(1):1–41.

Piketty, Thomas and Emmanuel Saez. 2006. “The evolution of top incomes: a historical and international perspective.” National Bureau of Economic Research. http://unpan1.un.org/intradoc/groups/public/documents/APCITY/UNPAN027111.pdf.

Saez, Emmanuel and Gabriel Zucman. 2019. *The triumph of injustice: How the rich dodge taxes and how to make them pay*. W. W. Norton & Company.

Solt, Frederick. 2011. “Diversionary nationalism: Economic inequality and the formation of national pride.” *The Journal of Politics* 73(3):821–830.

Tomz, Michael, Joshua A Tucker and Jason Wittenberg. 2002. “An Easy and Accurate Regression Model for Multiparty Electoral Data.” *Political Analysis* 10(1):66–83.

Volscho, Thomas W and Nathan J Kelly. 2012. “The rise of the super-rich: Power resources, taxes, financial markets, and the dynamics of the top 1 percent, 1949 to 2008.” *American Sociological Review* 77(5):679–699.