Party Effects on Total and Disaggregated Welfare Spending – A Mixed-Effects Approach

Abstract

Welfare is the largest expenditure category in all advanced democracies. Consequently, much literature has studied partisan effects on total and policy-specific welfare expenditure. Yet, these results cannot be trusted: The methodological standard is to apply time-series cross-section-regressions to annual observation data. But governments hardly change annually. Thus, the number of observations is artificially inflated, leading to incorrect estimates. While this problem has recently been acknowledged, it has not been convincingly resolved. We propose Mixed-Effects Models as a solution, which allow decomposing variance into different levels and permit complex cross-classification data structures. We argue that Mixed-Effects models combine the strengths of existing methodological approaches while alleviating their weaknesses. Empirically, we study partisan effects on total and on disaggregated expenditure in 23 OECD-countries, 1960-2012, using several measures of party preferences.

1. Introduction

Welfare is by far the largest expenditure category in all advanced and many developing democracies. Governments spend between one sixth and one third of GDP on social policies. Consequently, a broad literature has probed determinants of social spending, focusing particularly on the role of governing parties (for an overview see Obinger/Wagschal 2010; full references below). Whereas discussion is still ongoing on whether – and if so, how – parties of different *couleurs* have shaped welfare expansion and retrenchment, the literature almost unanimously agrees on one aspect: Almost all studies use the very same method, namely time-series cross-section (TSCS) regressions applied to annual observation data ('country-years'). While TSCS-regressions have been discussed extensively (e.g., Beck/Katz 2011; Kittel 1999; Kittel/Winner 2005; Plümper et al. 2005), the use of annual observation data still remains largely unquestioned.

We argue that when we are interested in effects of governing parties on expenditure or other policy-outputs, 'country-years' are the incorrect unit of analysis in a TSCS setting because governments usually do not change on a yearly basis. Consequently, existing studies artificially inflate the number of observations, resulting in overconfident estimates. Therefore, we cannot trust the literature's results on partisan effects on public expenditure. A handful of studies has raised similar criticism and proposed using government-terms, i.e. cabinet periods, as the unit of analysis instead (Garritzmann/Seng 2016; Persson et al. 2007; Schmitt 2016; Vis 2011, 2012). While this approach indeed produces more accurate estimates for effects of governments, it empirically introduces new difficulties. First of all, as countries differ with regard to cabinet duration and the time since democratization, some countries *de facto* had twice as many governments as others, which makes it difficult to compare

observations across countries and to set up the data with respect to common trends. Second, aggregating all variables to the periodization of cabinet terms throws away variation in annually changing variables (e.g., unemployment rates, GDP growth). Thus, using government-terms as the unit of analysis to estimate partisan effects is an improvement over simple country-year approaches to estimate partisan effects, but also introduces new problems. Both country-year and cabinet-term approaches are thus imperfect.

Rather than opting for either yearly observations (leading to incorrect estimates of party effects) or cabinet-terms (assuming simultaneity of observations and ignoring variance in other variables) as the unit of analysis, we propose to use Mixed-Effects Models instead, as these models can simultaneously account for the different periodization of observations in one setup (Henderson 1975, 1982; Lindstrom/Bates 1988; McLean et al. 1991). We can therefore use annual observations as the unit of analysis, but develop a model that accounts for the cross-classified, nested structure of annual expenditures within cabinets in countries and within years. To our knowledge no study has used these models in welfare state or public policy research yet.

Empirically, we aim at applying our approach to the widest possible sample to have as much empirical leverage as possible and in order to make a substantive contribution to the literature on welfare state research. Thus, we first investigate the impact of governing parties on total public social spending between 1960 and 2012 in 23 OECD-countries in a Mixed-Effect setup. We compare the findings of our Mixed-Effects Models to results from a 'standard country-year' approach, showing that the standard approach produces incorrect (over-confident) estimates, which suggest incorrect substantive interpretations.

Second, we disaggregate total spending and analyze parties' impact on the five most important welfare areas, namely pensions, health care, unemployment benefits, active labor

market policies, and education, exploiting data from 1980 to 2010 (1970-2012 for the case of education). We can therefore also contribute substantially to the big literatures on specific welfare policies. Moreover, in order to make sure that the results are not driven by operationalization decisions, we use the three most common operationalizations of party preferences: First, we group parties into a leftwing, a center, and a rightwing camp, as is often done in the older literature (see literature review). Second, following more recent studies, we employ a party family approach. Finally, we construct a direct measure of parties' welfare preferences, using a manifesto-based measure.

When using the literature's 'standard country-year approach', we do find highly significant partisan effects. Yet, when moving to our more accurate Mixed Effects Models, the findings do not reveal any systematic differences between leftwing and rightwing parties. We do find effects, however, when using a party family approach: We detect a negative effect of Social Democrats on social security spending, which results from the early period of our analysis and partly stems from country-differences. Moreover, using our manifesto-based operationalization we find that parties that place more emphasis on welfare in their electoral campaigns also deliver on their promises and in fact spend more. This is an encouraging sign for scholars of democratic representation and responsiveness. Yet, this finding only holds for total public spending and we do not find significant partisan effects when analyzing disaggregated welfare spending. These findings imply that public policy scholars and welfare state researchers interested in policy-output should not only pay close attention to the legislative arena, but also to the electoral arena and the respective kind of party competition. Our findings also show that while public expenditure is characterized by strong path dependencies, several short-term socio-economic factors do matter. This finding is robust to a number of alternative specifications.

Our study contributes to welfare state and party politics research, as well as to the public policy and political economy literature more generally. In fact, while we apply our approach to welfare spending, our methodological contribution holds equally for partisan effects on any other policy output, as TSCS-analyses on annual basis have become the 'de facto gold standard' in the quantitative macro-comparative public policy and political economy literature. Future research could thus apply our approach to other policy areas.

We start with a longer state-of-the-art review of existing work because we aim to show that the substantive question ('do parties matter?') goes hand in hand with methodological choices. Put differently, we show that what method you choose affects what results you get. Afterwards, we present Mixed-Effects Models as an alternative approach that allows circumventing the trade-off between country-year approaches and cabinet-term approaches.

2. Literature Review

Parties and total welfare spending

Studies on partisan effects on welfare expenditure abound. While public expenditure is affected by many factors (Obinger/Wagschal [2010] provide a review), political scientists have focused on the question whether – and if so how – governing parties shape spending[1]. Starting from Hibbs' (1977) *partisan hypothesis*, parties should make a difference because they represent different societal groups with different preferences. In the original version, the theory assumed that lower socio-economic strata (SES) voters favor an extensive welfare state and vote for leftwing parties, which respond by increasing welfare expenditure. Higher

[1] Expenditure has been criticized as a measure of welfare generosity (e.g., Allan/Scruggs 2004). While this criticism is valid, we believe – in line with a broad literature – that spending still is an interesting measure of countries' and parties' welfare effort. Future research could apply our framework to other outputs.

SES-voters, vice versa, should favor lower taxes and a lean welfare state; therefore, they would vote for rightwing parties that would retrench expenditure. *Power Resource Theorists* (Huber/Stephens 2001; Korpi 1983; Stephens 1979) developed similar arguments for welfare policies more specifically.

Subsequent studies pointed out that the real world is more complex: It is neither true that all low-SES voters favor welfare (Häusermann et al. 2013), nor that all leftwing parties favor welfare expansion while all rightwing parties favor retrenchment. For example, Christian democrats are strong supporters of some social policies, too (Huber/Stephens 2001; van Kersbergen 1995); and even secular conservative parties support social policies in some circumstances (Jensen 2014). Moreover, partisan influences are likely to vary over time (Pierson 1996; Garritzmann 2016) and across policy-fields because different (re-)distributive mechanisms are at work (Castles 2009). Accordingly, much debate has centered on the question whether parties' impact varies between a "Golden Age" of welfare expansion in the 1950s-70s and a "Silver Age" of retrenchment and/or readjustment since the 1980s (Huber/Stephens 2001; Pierson 1996)[2].

As by now a huge number of studies exist, we cannot discuss these in detail. Instead, we provide a tabular overview of the existing comparative studies in the Online Appendix (Table A) and only highlight some key insights here. The main aim of this detailed review here is to show that the choice of methods is connected to substantive findings. Our review focuses on country-comparative, quantitative studies on OECD-countries that use welfare expenditures as dependent variables and test for partisan effects. Table A summarizes the

[2] Going against common wisdom of disappearing partisan effects in the 'Silver Age', Savage (2018) argued recently that partisan effects have re-appeared after the onset of the Great Recession.

studies' research designs, time period, number of cases, operationalizations of key variables, and findings.

Our main point is that – since the 1990s – (almost) *all* studies have applied time-series cross-section (TSCS) regressions to annual observation data[3]. We will criticize this widely accepted gold standard below. Another important point is that our systematic review reveals that the existing studies use very different model specifications and – arguably as a consequence – draw highly different conclusions. To illustrate, we point at three crucial methodological differences. Our main goal here is to show that these methodological choices are intertwined with the substantive findings, i.e. that the choice of methods shapes the substantive findings.

First, some scholars regress spending *levels* on the partisan composition of government, arguing that only these 'levels-on-levels' analyses yield meaningful results (e.g., Huber et al. 1993; Huber/Stephens 2001). Others analyze how *changes* in government composition affect subsequent *changes* in expenditure ('difference-in-difference' models). Kittel and Obinger even claim "there is little to learn from an analysis of the levels in a panel model" (2003: 29). Interestingly, the methodological choice to study levels or differences has important substantive implications: A count of the analyzed studies shows that 75% of the studies using levels-on-levels report significant partisan effects while only 17% of the studies using DiD-designs do.

Second, the studies' model specifications differ: Many studies rely on Beck and Katz' (1995) dictum to include lagged dependent variables, panel-corrected standard errors, and AR(1)-corrections (e.g., Brady et al. 2005; Swank 2002), but others (especially economists) apply different specifications (e.g., Herwartz/Theilen 2014b; Potrafke 2009). Moreover, some

[3] But see Breunig/Busemeyer (2012) for a composite data analysis, Tepe/Vanhuysse (2010) using event history models, and Loftis/Mortensen (2017) using dynamic linear modeling.

studies include the independent variables both as levels and first-differences ('error-correction models'), whereas others include either levels or differences. Furthermore, there is much discussion on whether country fixed-effects should be included or not. As Table A shows, these methodological decisions again have substantive implications: For example the decision to include country fixed-effects changes not only the interpretation of the model (to within-country variation) as is well known, but also the substantial outcomes in many studies (e.g., Garrett/Mitchell 2001; Swank 2002): Most of the studies (87%) including country fixed-effects do not find evidence for partisan effects, while most studies (85%) excluding country fixed-effects report significant findings. Again, this implies that the choice of methods affects the substantive findings.

Third, the operationalizations of the dependent and independent variables vary considerably: Many analyze social expenditure as a share of GDP, while others focus on specific spending categories, or analyze welfare spending as a share of total spending. There is similar dispute for the operationalization of the independent variable: Some studies rely on simple measures such as left-center-right dummies (e.g., Castles 2009), whereas others use party families (e.g., Busemeyer 2009), or employ more fine-grained manifesto-based measures (e.g., Bräuninger 2005). Moreover, while some studies focus on the current governments' compositions, others cumulatively aggregate parties' historical cabinet seat-shares from democratization to the respective year of analysis. Again, these methodological details are important because different theoretical expectations lie behind these decisions (e.g., 'should we expect short-term or long-term effects?' 'What is the major cleavage over welfare?') and because they have some implications for the findings: Appendix A shows that 66% of the studies using left-right dummies find effects, 58% of those utilizing party families report significant findings, and 66% of the manifesto-based estimation.

Arguably at least partially as a result of these methodological choices, the existing studies differ greatly regarding their substantive findings on partisan effects on welfare expenditure: Half of the reviewed studies report partisan effects whereas the other half reports null findings. Thus, the literature offers contradicting results on the party-expenditure nexus, so we simply do not know whether parties have affected welfare expenditure or not. In this sense, a core question of political science remains unanswered. Our review showed that methodological choices are systematically related to these findings: Studies using difference-in-difference designs or including country fixed-effects are much less likely to find significant partisan effects. That is, methodological and substantive questions go hand in hand, a point we substantiate below.

Parties and policy-specific expenditure

Our state-of-the-art review concluded that the existing literature has produced inconclusive findings regarding the effects of parties on welfare spending. Yet another cause for the inconclusive results is that these studies analyze *total* public social spending, which might conceal divergent effects across different policy areas[4]. Comparing four social spending areas, Castles shows that these are "almost entirely unrelated" (2009: 45). There are two main reasons why partisan effects might vary across policy-areas: First, as the beneficiary groups differ (e.g., pensioners, students, unemployed) and as consequently different (re-)distributive dynamics follow, party effects can be expected to differ across fields. For example, while Social Democrats might be associated with higher unemployment spending, Christian

[4] This is even more problematic because pension and health care consume the largest shares of public expenditure, but especially in these areas the literature expects at best weak partisan effects (Jensen 2014).

Democrats could increase health and old-age spending, as their respective electorates are more likely to benefit. Second, parties' effects might differ across policy-fields because governments' leeway in shaping expenditure varies: Discretionary expenditures (e.g., investments) are easier to change than entitlement spending (e.g., pensions) (Breunig/Busemeyer 2012; Streeck/Mertens 2011).

While hardly any studies compare effects across different social-policy areas (but see Breunig/Busemeyer 2012; Castles 2009; Jensen 2012), broader sub-literatures exist for the respective social policy areas. Hence, before turning towards our argument and findings, we briefly review the literature on party effects on the five most important (i.e. here: largest in terms of budget) welfare policy areas: pensions, health care, unemployment benefits, active labor market policies, and education. Our empirical analysis aims at adding substantive knowledge to these big literatures as well.

Pensions

In all OECD-countries, pensions are the largest welfare expenditure category. Still, spending levels vary considerably across countries. Accordingly, a broad literature analyzes determinants of this variation over time and space (Hinrichs/Lynch 2010 provide an overview). Some studies argue, following Hibbs (1977), that leftwing parties increase pension generosity (Jensen 2012; Myles 1989; Palme 1990), whereas rightwing parties decrease expenditure (ibid.) and generosity (Hicks/Freeman 2009). Newer contributions, however, refine these arguments, claiming that especially Christian democratic parties are proponents of public pensions, because pensions are not (directly) class-based but rather relate to a life-course risk that affects the entire population (Huber/Stephens 1993, 2001; van

Kersbergen 1995). Consequently, Christian democrats are for ideological and electoral reasons inclined to spend more on pensions than other party families (Fernández 2012).

Health-care

Health-care constitutes the second largest category in most governments' welfare expenditures. Like pensions, health care addresses a life-course risk that is less class-related than, for example, unemployment benefits (Jensen 2014). Accordingly, discussions of partisan effects resemble those of pensions: Some studies find that leftwing parties increase public health-care spending (Castles 1998; Fervers et al. 2015; Huber/Stephens 2001; Herwartz/Theilen 2014a). Others claim there is no partisan cleavage over public health-care expenditure as the risk of sickness is universally distributed among the entire population, leading to support across parties (Jensen 2011b, 2014). Accordingly, some studies find no partisan effects (Jensen 2011b, 2014; Potrafke 2010; Reeves et al. 2014); others show that while parties had mattered for an initial expansion period, their impact has declined over time (Jordan 2011); still others even find negative effects of leftwing governments on health-care spending (Jensen 2012).

Unemployment benefits

Allan and Scruggs (2004: 499) argue that unemployment benefits are "the best manifestations of Esping-Andersen's idea of welfare state de-commodification" and that their redistributive patterns are clear: Unemployment benefits benefit the unemployed, who are usually assumed to tilt politically to the left. Consequently, it is commonly assumed that leftwing parties are the major promoter (e.g., Di Tella/MacCulloch 2002). Again, however, reality is more complex. Recent contributions uncover an (increasing) "dualization" between labor

market "insiders" and "outsiders" (Emmenegger et al. 2012) and reason that especially Social Democrats face a dilemma, as they cannot simultaneously satisfy both labor market groups (Rueda 2005). Consequently, Social Democrats might become less disposed to increasing public expenditure on unemployment benefits as they might increasingly focus on insider-protection. This is, moreover, likely as discourse among policy-makers, welfare scholars, and the general public has shifted from compensatory welfare towards "social investment" (Hemerijck 2013; Morel et al. 2012). We therefore next turn towards two ideal-typical social investment policies: ALMPs and education.

ALMPs

Active labor market policies (ALMPs) have recently received much scholarly interest. Thus far, however, inconclusive results have been produced: Boix (1998), Huo et al. (2008), and Iversen and Stephens (2008), among others, find that leftwing parties are strongly associated with ALMP spending, whereas others report no effects (Rueda 2005; Tepe/Vanhuysse 2013; van Vliet/Koster 2011). Newer contributions try to reconcile these contradictory findings by distinguishing several kinds of ALMPs (Bonoli 2010; Vlandas 2013), or by specifying conditions under which partisanship matters for ALMP expenditure (Vis 2011, 2012). Nonetheless, the question whether, and if so how, governing parties shape ALMP spending is not sufficiently answered yet.

Education

Comparing different welfare policies, Wilensky (1975: 3) claims that "education is special". Indeed, education is a complex policy-area, because the redistributive dynamics are more complex than those of other welfare policies. While discussing this in detail lies beyond the

scope of this article, some examples illustrate this point: Busemeyer (2009), Iversen and Stephens (2008), and others find that leftwing parties promote education expansion, arguing that education is redistributive and enhances equality of opportunities. Others object that this is not true for all kinds of education, as different (re-)distributive dynamics apply. *Higher* education, for example, can be regressive, as enrollment rates are not universal and access stratified by parental background (Fernandez/Rogerson 1995). Complicating things further, the redistributive patterns might change over time (Ansell 2010; Garritzmann/Seng 2016) and become even harder to disentangle when we distinguish public and private spending (Garritzmann 2016). Consequently, including public education expenditure in our analysis seems worthwhile, as it is an extremely salient policy field that recently gained an upswing in scientific interest (Busemeyer/Trampusch 2011) and as the literature produces inconclusive results.

In sum, the extensive literature review on parties and welfare expenditure showed that the literature has produced contradictory results. While there is some consensus that generally leftwing and Christian democratic parties might be more supportive of welfare expansion than other parties, the empirical patterns are much less clear. Debate on the strength, size, and even the direction of effects is thus still ongoing. This is true for both total and disaggregated welfare spending. We argue in the next sections that this is at least partly due to methodological choices, which go hand in hand with substantial results.

3. Shortcomings of 'country-year approaches' and initial solutions (using government-terms)

Despite these differences, most existing studies agree in another respect: They apply the same method to investigate the party-expenditure nexus, namely pooled time-series cross-section (TSCS) regressions using annual data ('country-years') as the unit of analysis (see Table A). TSCS regressions have been debated extensively (e.g., Beck/Katz 1995, 2011; Kittel 1999; Kittel/Winner 2005; Plümper et al. 2005). What has been (almost) unequivocally accepted, however, is the use of annual observation data.

We argue that country-years are a problematic unit of analysis in standard TSCS regressions when one is interested in effects of parties because the government composition usually does not change annually. But using country-years as the unit of analysis assumes that we can observe the partisan composition of government every year and that these observations are independent of each other, which is obviously not true. Analyses utilizing country-year data (without controlling for this) therefore artificially inflate the number of observations, resulting in potentially incorrect estimates of the effects and their uncertainty (i.e. standard errors are overconfident).[5] In short, we cannot trust the literature's findings on the impact of governing parties on public welfare expenditure.

A few studies (Garritzmann/Seng 2016; Persson et al. 2007; Schmitt 2016; Vis 2011, 2012) have voiced similar criticism and suggested an alternative procedure: Instead of using country-year data, government-terms (i.e. cabinets) should be used as the unit of analysis. Cabinets are argued to be superior, because we can observe the government's partisan composition only once per government-term. Of course, this reduces the number of observations by country, but the argument is that this is a reasonable reduction, because we can

[5] The standard errors of a regression coefficient are defined as the square root of the diagonal elements of the covariance matrix divided by N. As the average duration of governments is about three years – varying depending on the countries and period – then the reported standard errors will be also three times smaller in a country-year setup than in a cabinet-design.

independently observe each cabinet only once, irrespective of its duration in office[6]. For example, instead of using 70 'observations' (yearly data from 1945 to 2014), one would use 24 cabinets in the UK (from Attlee to Cameron), 22 cabinets in Luxembourg (from Dupong to Bettel), and 53 cabinets in Japan (from Yoshida to Abe). To our knowledge only five studies have investigated spending on cabinet-term basis: Garritzmann and Seng (2016) analyzed education spending, Persson et al. (2007) investigated total government spending, Schmitt (2016) focused on total public welfare expenditure but analyzed a shorter time-period (1980-2009) than we do (1960-2012), and Vis (2011, 2012) concentrated on ALMPs.

In the next section we argue that while these studies have made considerable progress, they in fact introduce new (methodological) problems.

4. Shortcomings of 'cabinet-term approaches'

Using cabinet-terms as the unit of analysis produces more accurate estimates of partisan effects but introduces two new problems. A first problem is that, as the duration of governments differs across countries (and time), comparisons between observations become difficult. Using cabinet terms as the unit of analysis is appropriate when the included countries are similar regarding the duration of democracy and the turnover rate of governments. In the real world, however, both of these assumptions are not met. As exemplified above, the UK and Luxembourg, for example, had less than half the number of governments than Japan since World War II. Thus, if one takes the running number of governments as the time variable in a panel setup (as is usually done), this implies a comparison between a recent government in the UK to one from the 1970s in Japan, which

[6] Consider that someone used monthly data, pretending to 'observe' twelve governments per year. This would inflate the number of observations further and illustrates that we can only observe the government constellation once per term.

does not seem advisable for theoretical reasons. Similarly, as several Southern European countries became stable democracies only in the 1970s, counting the first government as number "1" (as TSCS models usually do) leads to comparing them to observations from the postwar period in other countries. This is problematic given that welfare policies were adopted in certain historical contexts and broader structural changes affect expenditures in manifold ways.

A second problem of using cabinet terms as the unit of analysis is that this approach ignores the time interval of other variables. On the one hand, it ignores variation in annually available variables (e.g., unemployment rates or GDP growth, but also spending), which have to be aggregated to the level of cabinets. On the other hand, some variables change even more seldom than governments (e.g., political institutions). In other words, the problem is that some variables are measured and are changing on a yearly basis, others on a cabinet-term basis, and still others only on a country-basis (time-invariant). Ideally, we thus need a model that can deal with this complex nested structure to avoid the disadvantages of both country-year and cabinet-term analyses.

5. Mixed-Effect Models as a superior alternative

We argue that Mixed-Effects Models are a better tool to deal with this complex, nested, and cross-classified data structure. Mixed-Effects Models can circumvent the trade-off between country-year and cabinet-term approaches. In contrast to standard panel estimators, Mixed-Effect Models allow simultaneous estimation of the effects of variables with different time intervals, i.e. variables that vary annually, only over several years, or not at all within countries. More generally, the main advantage of Mixed-Effects Models is that they allow including fixed effects and random effects while taking the nested and cross-classified data

structure into account (Henderson 1975, 1982; Lindstrom/Bates 1988; McLean et al. 1991). Put differently, we can decompose the variance on several levels (countries, years, cabinets) and study to what extent each of these levels contributes to the overall variance. For these reasons, Mixed-Effects Models solve the arbitrariness of selecting either country-years or cabinet-terms as the unit of analysis, thereby combining the strengths of both approaches.

More specifically, our dependent variable (public spending) as well as several socio-economic control variables vary annually. Our main independent variable, the partisan composition of government, varies on a cabinet-term basis. Finally, some control variables (institutional variables) are largely time-invariant in the period under study. Thus, we design a model with annual observations nested in governments, but also nested in countries as well as in time points (years). The model is thus a cross-classified hierarchical model with random intercepts for governments, countries, and years. We estimate the following equation:

$$y_{igct} = \beta_0 + \beta_1 X_{igct} + \beta_2 W_{gct} + \beta_3 Z_c + \beta_4 S_t + \tau_{00g} + \tau_{00c} + \tau_{00t} + \varepsilon_{igct}$$

$$with\ \tau_{00g} \sim N(0,\sigma_g^2) \quad \tau_{00c} \sim N(0,\sigma_c^2) \quad \tau_{00t} \sim N(0,\sigma_t^2) \quad \varepsilon_{igct} \sim N(0,\sigma_\varepsilon^2)$$

where y is the dependent variable, X is a vector of annually observed (socio-economic) control variables, W is a vector of cabinet-term independent variables, Z is a vector of country-specific independent variables, S is a vector of time-specific independent variables, τ_{00g} is the government-specific variance, τ_{00c} is the country-specific variance, τ_{00t} is the time-specific variance, and ε_{igct} is the idiosyncratic error. The time dimension S is modeled with cubic splines to account for non-linear dynamics (Beck et al. 1998).

6. Data and operationalizations

We utilize the data with the longest and broadest possible coverage for total and disaggregated welfare expenditure for advanced democracies, i.e. covering 23 OECD countries[7]. The longest time-series data (i.e. from 1960 to 2012) is available for SOCIAL SECURITY TRANSFERS (defined by the OECD as "grants and welfare benefits paid by general government [benefits for sickness, old-age, family allowances, etc.]"; cf. Armingeon et al. 2014), which we take as a measure of total public welfare spending. Moreover, we distinguish five specific welfare policies: Public spending on 1. PENSIONS, 2. HEALTH CARE, 3. UNEMPLOYMENT BENEFITS, 4. ALMPs, and 5. EDUCATION (covering all educational sectors from childcare to post-secondary education). This policy-specific expenditure data is available between 1980 and 2010 (1970-2012 for the case of education). As is customary, we analyze spending as a proportion of GDP to allow for temporal and spatial comparisons.

We can thus apply our model for 23 countries over up to five decades in five policy-areas and to total public social spending, which overall provides a strong test of our claims and empirically goes beyond the scope of existing studies, which hardly offer policy-comparative approaches (see Table A). Table B in the Online-Appendix provides a descriptive overview of the variables and their sources.

Our major *independent variable* is the partisan composition of government. As discussed above, there is a broad literature on the operationalization of party preferences. To compare different measures we use the three most common operationalizations of party preferences: First, we use the PROPORTION OF CABINET SEATS HELD BY LEFTWING, CENTER,

[7] Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, United States.

AND RIGHTWING PARTIES, respectively (Armingeon et al. 2014). We estimate models using the leftwing variable versus center or rightwing parties' seat-shares (as the reference category) and check the results using the rightwing variable versus the other categories instead.

Second, we code parties as belonging to one of ten PARTY FAMILIES, relying on the Comparative Manifestos Project's (CMP) "parfam"-variable (Volkens et al. 2011). We use each party family's cabinet seat-share as the second measure, focusing particularly on Social Democrats (following the existing literature's focus).

Third, we use manifesto-based measures of party preferences, based on CMP data (Volkens et al. 2011). The CMP project codes party manifestos in pre-defined issue categories, which can be used as a measure of relative issue emphasis on certain issues. The CMP offers two directly welfare-related items: "Welfare Expansion" (per504) and "Welfare Limitation" (per505). A considerable share of parties' manifestos engages with welfare policies (on average 7.9 percent on welfare expansion and 0.4 on welfare limitation in our sample). We compute government positions by calculating the average issue emphasis on "welfare expansion" (per504) minus the average issue emphasis on "welfare limitation" (per505) of all cabinet parties, weighted by their respective cabinet seat-shares[8]. Although the data is not perfect (e.g. it does not allow differentiating between types of welfare policies), it still offers the best comparative data for a direct measure of party preferences which goes beyond left-right dichotomies or party family approaches. In additional robustness tests (reported below), we also tested the inclusion of many other CMP categories and of manifesto-based left-right scores.

[8] $Government_{ijt}$'s welfare position = ([Government party A's per504 value – Government party A's per505 value] * Government party A's cabinet seat share) + … + ([Government party N's per504 value – Government party N's per505 value] * Government party N's cabinet seat share)

Empirically, the first and second operationalization of party preferences (the left-right categorization and the party-family approach) correlate to a high degree (0.84) as could be expected; yet, the correlation between these operationalizations and the manifesto-based item is much lower (0.3 and 0.28, respectively) indicating that these capture substantially different things. Whether a party holds positive views on welfare is not necessarily determined by its party family or broader left-right camp. Table C in the Appendix descriptively shows the averages and standard deviations of our manifesto-based item for each (governing) party family. A noteworthy finding is that while Social Democrats – as could be expected – do show the highest scores on this measure, Liberal and Agrarian parties also reveal comparatively high values. This underpins that left-right dichotomies and party family approaches are quite rough proxies for party preferences. Accordingly, the manifesto-based item is (although it is far from perfect) our preferred and most direct measure of party preferences.

All partisan composition variables are coded with a one-year lag, i.e. we assume that governments in year t affect expenditure levels in year t+1, which seems reasonable given the lengthy legislative process of passing households[9]. As our model clusters observations of the cabinet level, this implies that we have to worry less about political cycle effects, because in our setup the exact time point of a policy change within a government's term is not relevant – if it happened within a cabinet's duration, our approach will pick it up.

[9] For the same reason, we do not consider governments that have been in office less than one year (yet, we obtain similar results when still including these cases). We control for these cases with an additional dummy variable (SUCCESSION), because due to the exclusions of short governments some governments do not follow each other directly in our sample.

Parties certainly are not the only factor influencing expenditures. Accordingly, we include several controls: To control for economic influences we add GDP PER CAPITA, GDP GROWTH rates, and INFLATION. Moreover, to cover demand-effects, we add UNEMPLOYMENT RATES and the SHARE OF ELDERLY PEOPLE (65+). CAPITAL OPENNESS and DEINDUSTRIALIZATION should cover structural changes. All of these socio-economic variables are included without lags, as we assume that they have immediate effects on budgets as governments use economic forecasts when deciding over their budgets. Finally, following Immergut (1990) we control for INSTITUTIONAL CONSTRAINTS (Schmidt's (1996) 7-point index), because political institutions might affect welfare expansion *and* retrenchment. Additionally, we also include variables for FEDERAL SYSTEMS[10] and SINGLE MEMBER DISTRICTS. Finally, we control for VOTER TURNOUT to pick up the argument that with larger electorates policy-makers might be more responsive to broader welfare demands (Hicks/Swank 1992).

7. Findings

We first present empirical findings for party effects on total social security transfers for the longest possible time period, i.e. 1960-2012. We contrast the results of the literature's standard 'country-year' approach (ignoring cabinet clustering) with our superior Mixed-Effects Models. Subsequently, we investigate effects of the partisan composition of government on specific social policies. Finally, we discuss the robustness.

Social security transfers, 1960-2012

[10] Federalism is one of the seven items in Schmidt's institutional constraint index, so both variables are correlated to a certain degree. Yet, excluding either variable produces similar results.

Before presenting our Mixed Effects Models, we discuss results based on the literature's 'standard country-year' approach. That is, we estimated time-series cross-section regressions using country-years as the unit of analysis. We include the very sample control variables as in our preferred Mixed Effects Models and add country fixed-effects (as indicated also by a Hausman-test). These models (see Table D in the Appendix) show highly significant partisan effects for all three measures of party preferences, i.e. the Left-Center-Right dummies, the party family approach, and the manifesto-based item. All partisan effects are significant on the one-percent level, despite the inclusion of country fixed effects. We do not want to waste a lot of space discussing these results in detail, though, because – as we show next – our Mixed Effects Models reveal that – as argued – these results cannot be trusted, as they produce over-confident estimates.

We start our analysis by decomposing the overall variance of social security transfers between the different levels. The Intraclass Correlation Coefficients (ICCs) of an empty model are 0.207 for governments, 0.183 for years, and 0.587 for countries. The largest part of the variance thus stems from the country-level and is almost three times bigger than the variance from the government level – a component that cancels out in TSCS models with fixed effects or first differences, which is one explanation why these studies are less likely to obtain results for partisan effects. Yet, still about 21 percent of the variance originates from cabinets, indicating potential partisan effects. The smallest part of the variance stems from the time level with about 18 percent. The observed values show much variation, from 3.5 to 24 percent government spending on social transfers, and our predictions of these values are in the same range while the deviation is less than 3 percent over the whole sample.

The results of the Mixed-Effects Models[11] for social security transfers are presented in Table 1[12]. Including the independent variables in the model decreases the government specific variance only slightly to 18 percent. The time specific variance almost cancels out as the cubic splines pick up the dynamics; in other words, there are hardly any year-specific effects in the pooled sample. Consequently, the country specific variance remains the biggest component.

Turning towards determinants, Models 1-3 are similar but include different operationalizations of government composition. Model 1, using the share of leftwing (vs. center and rightwing) parties in government shows no significant effect. That is, in contrast to the common country-year approach, our Mixed-Effects Models reveal that leftwing governments on average do not show a different spending attitude than center or rightwing governments. This might be due to several reasons, one simply being that the measure is very rough and does not pick up more fine-grained differences within the broad ideological camps. Our other two models thus show results for more fine-grained measures.

Model 2 displays findings for the cabinet share of Social Democratic parties. We find a significant within-effect of about half a percent of GDP. The effect is, however, negative, indicating that controlling for variation across countries, Social Democratic government participation is associated with lower social security spending. Investigating this at first glance surprising effect further by looking at country-specific results, we detect that this effect stems from the fact that in some countries (i.e. in the Nordic welfare states, but also in Austria, Belgium, France, the Netherlands, New Zealand, Switzerland, and the UK) social

[11] The estimation was done by restricted maximum likelihood (REML) in order to avoid small sample bias of estimates and confidence intervals, as mentioned by Stegmueller (2013). We additionally corrected the denominator degrees of freedom with a Kenward-Roger approximation for tests in linear mixed models.

[12] Note that we run additional models using a different dependent variable using "public and mandatory private" expenditure as the dependent variable, which produced the same findings (available on request).

democratic governments spent less on social security transfers than governments of other *couleurs*. In these countries, though, the overall expenditures were about 1.5 percent *higher* than in the remaining countries – this positive between-effect is captured by the random intercepts for countries in the model. Put differently, the negative partisan effect in Model 2 stems from the fact that Social Democrats initially focused spending less on social security than other party families, which is in line with Huber and Stephens' (2001) findings that especially Christian Democrats spend more on social security. Probing these findings further, we also split the data in decades in order to test for effect heterogeneity (results available on request). Yet, we do not find any significant party effects, which might also be due to the relatively demanding model specification with fewer observations. Nevertheless, these models show that the negative effect of Social Democrats originates from the early years of our period of investigation and disappears in the 1990s.

Model 3 includes the manifesto-based measure, the – in our view – most direct operationalization of government preferences. Here, we detect a positive significant effect, which indicates that parties that talk more positively about welfare in their manifestos also follow-through with their election promises and increase welfare expenditure. The size of the effect is considerable: The coefficient shows that with each unit-increase in the manifesto-based measure spending increases by 0.04. Substantially, this means that each additional percentage point in positive welfare reference translates into a 0.04 percentage increase in expenditure as a share of GDP. A change from the first to the third quartile is associated with an increase of 0.3 percent. Given that governments' emphasis on welfare (positive emphasis minus negative values) ranges from -9.9 to 43.9 in our sample, the effect can be substantial, especially when adding up over time. These findings are important and encouraging for scholars of democratic representation, as they highlight that party promises are not just

'cheap talk' but in fact do translate into relevant differences in public policy. Substantively, our findings thus show that the electoral and the legislative arena are highly connected. This finding implies that public policy scholars need to pay particular attention not only to the legislative arena, but also to the electoral arena of party competition.

Taking a step back, and comparing the initial findings using the common country-year approach and our more accurate Mixed Effects approach demonstrates that the literature's standard approach produces over-confident estimations and cannot be trusted. As the number of observations is artificially inflated in the common country-year approach, the standard approach suggests highly significant findings where they should not be found. Once we correct for this using Mixed Effects Models, the partisan effects are much less significant. Substantially important, and relieving for proponents of representative democracy, though, we still do find partisan effects, especially when using the manifesto-based item.

Regarding the controls, we find that higher unemployment as well as an older society drive social security transfers irrespective of the government in charge. This clearly supports demand-side arguments and will be further elaborated when looking at specific expenditures below. There are also negative effects of the economic variables indicating that countries with a higher GDP spend significantly less than poorer countries in relation to their GDP. We also find that higher GDP growth leads to lower expenditures, which might have several reasons: First, some areas, like unemployment expenditures, are not likely to rise in times of growth and there might also be less need for additional expenditures. Second, as the dependent variables are measured as a share of GDP, we would find a negative effect even

when holding expenditures constant while GDP is increasing. The effect of the inflation rate is also negative as in times of high inflation additional expenditures would even further increase this rate.

We do not find a significant effect of voter turnout. Moreover, neither the duration in office nor whether one government directly follows the previous one matters (justifying our exclusion of short governments). Moreover, neither institutional constraints in general nor federalism in particular have significant effects, even when dropping one of the two variables in order to account for the fact that they are correlated to some degree. Yet, there is a positive within-effect for single member district voting systems, resulting from the fact that only few countries radically changed their voting systems within the period under study: Italy changed from a proportional to a mixed proportional voting system in 1994 and back to a proportional voting system in 2005, and New Zeeland switched from a single member district voting system to a mixed proportional voting system in 1996.[13] While countries with single member district voting systems spend about 3 percent less than countries with proportional voting systems (the between-effect is captured by the random intercepts) (in line with Iversen/Soskice [2006] and Persson et al. [2007]), the social security transfers as a share of the GDP significantly decreased in New Zeeland after the reform of the voting system, rendering the within-effect positive.

[13] The change in Japan from single non-transferable vote to single member constituencies and regional lists in 1994 did not alter the coding as a "modified proportional representation system".

Table 1: Mixed Effects regression results for social security transfers, 1960-2012

	(Model 1)		(Model 2)		(Model 3)	
	B	*std. Error*	*B*	*std. Error*	*B*	*std. Error*
Fixed Parts						
(Intercept)	-427.42 ***	57.20	-425.78 ***	56.97	-429.83 ***	57.04
Institutional Constraints	-0.03	0.16	-0.04	0.16	-0.03	0.16
Federalism	-0.65	0.42	-0.60	0.42	-0.70	0.42
Single Member District	1.27 ***	0.37	1.25 ***	0.37	1.40 ***	0.37
Voter Turnout	0.00	0.01	0.00	0.01	0.00	0.01
Duration	0.00	0.00	0.00	0.00	0.00	0.00
Succession	0.12	0.20	0.13	0.20	0.09	0.20
Capital Openness	-0.01	0.01	-0.01	0.01	-0.01	0.01
Unemployment Rate	0.23 ***	0.03	0.23 ***	0.03	0.23 ***	0.03
Elder People	0.26 ***	0.07	0.26 ***	0.07	0.28 ***	0.07
Deindustrialization	-0.00	0.03	0.00	0.03	-0.01	0.03
GDP per Head	-0.11 ***	0.02	-0.11 ***	0.02	-0.11 ***	0.02
GDP Growth	-0.07 ***	0.01	-0.07 ***	0.01	-0.07 ***	0.01
Inflation	-0.04 ***	0.01	-0.04 ***	0.01	-0.04 ***	0.01
Spline 1	0.22 ***	0.03	0.22 ***	0.03	0.22 ***	0.03
Spline 2	-0.14 ***	0.02	-0.14 ***	0.02	-0.14 ***	0.02
Left Party	-0.00	0.00				
Social Democrats			-0.52 *	0.24		
Welfare Emphasis					0.04 *	0.02
Random Parts						
σ^2	0.343		0.342		0.342	
$\tau_{00,\ govern}$	2.770		2.762		2.749	
$\tau_{00,\ year}$	0.035		0.034		0.038	
$\tau_{00,\ country}$	11.978		12.072		11.811	

$N_{government}$	379	379	379
N_{year}	51	51	51
$N_{country}$	23	23	23
$ICC_{government}$	0.183	0.182	0.184
ICC_{year}	0.002	0.002	0.003
$ICC_{country}$	0.792	0.794	0.791
Observations	1009	1009	1009
R^2 / Ω_0^2	.988 / .988	.988 / .988	.988 / .988
AIC	3180.130	3166.466	3171.936

Notes: * p<.05 ** p<.01 *** p<.001

Policy-specific social expenditure, 1970-2012

Table 2 presents results for five policy-specific dependent variables using data from 1980-2010 (1970-2012 for education). We only present models for one measure (the cabinet share of Social Democrats), but note that we do not obtain any significant findings when using either the left-center-right categorization or the manifesto-based item. Starting again with the ICCs of the empty models, we see that for all policy-specific spending categories most of the variance lies between countries: for pensions 0.786, for health 0.477, for ALMPs 0.762, for unemployment benefits 0.707, and for education policy 0.628. Also, there is hardly any year-specific variance with less than 0.04 for all categories except for health with 0.255. Most interestingly, the government-specific variance differs across spending categories and is 0.194 for pensions, 0.232 for health, 0.198 for ALMPs, 0.229 for unemployment, and largest for education with 0.315, in line with arguments about discretionary spending being easier to change than entitlements (Breunig/Busemeyer 2012; Streeck/Mertens 2011).

Models 4 through 8 show the results for pensions, health care, ALMPs, unemployment, and education expenditures, respectively. Including the independent variables in the model decreases the government-specific variance for pensions, ALMPs, and unemployment, while

it increases for health and education. For health this is the case because the time-specific variance again almost cancels out with the introduction of cubic splines, while for education the country-specific variance decreases at the same time – which is mainly due to the effect of the single member district variable. Despite these differences, the country specific variance remains the biggest component in all models.

The results do not reveal any significant effects of Social Democrats. We obtain the same findings when using our other operationalizations of partisan governments. That is, we do not detect any partisan effects on spending disaggregated by social policy field in our data, covering the post-1980 (post-1970 for education) period. There might be two main reasons for this. On the one hand, in line with arguments about decreasing political room-for-maneuvre in the 'Silver Age' of the welfare state (Pierson 1996; Garritzmann/Seng 2016) our analysis might simply start 'too late' to detect partisan effects. On the other hand, it might also be a pragmatic problem, as due to the shorter time period there might simply be too few observations to find significant effects. As more data will become available in the future, future research will be able to address this question.

Some of the control variables show significant effects, though: As we would expect, the share of elderly people has the greatest effect on pensions (Model 4) and we also find a significant negative effect for education (Model 8), which seems plausible as there is lower demand with fewer younger people. We also find a positive effect of the unemployment rate on unemployment expenditures (Model 7), again underlining the demand-side argument for public expenditures. Capital openness tends to have a (small) negative effect on some expenditure categories while deindustrialization has a positive effect on all expenditures safe ALMPs. The positive effect of single member districts is also present for social expenditures,

pensions, and ALMP, but not for health-care and education which allows no straightforward conclusion.

Table 2: Mixed Effects regression results for area-specific public social expenditure, 1970/80-2010/12

	(Model 4) Pensions		(Model 5) Health care		(Model 6) ALMPs		(Model 7) Unemployment benefits		(Model 8) Education	
	B	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
Fixed Parts										
(Intercept)	95.43	68.15	195.08 ***	53.60	-185.12***	30.77	-30.99	35.42	122.34 ***	33.49
Institutional Constraints	0.12	0.09	-0.10	0.07	-0.03	0.03	-0.04	0.04	0.06	0.08
Federalism	-0.21	0.25	0.14	0.15	-0.07	0.07	0.08	0.11	-0.23	0.15
Single Member District	1.13 ***	0.18	-0.18	0.12	0.18 **	0.06	-0.11	0.09	-0.42 **	0.13
Voter Turnout	0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	0.00	0.00	0.00
Duration	0.00	0.00	-0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Succession	-0.14	0.11	-0.01	0.08	0.07	0.04	0.03	0.06	0.06	0.09
Capital Openness	-0.01 ***	0.00	-0.01 ***	0.00	0.00	0.00	-0.00	0.00	-0.01 *	0.00
Unemployment	0.07 ***	0.01	-0.01	0.01	0.00	0.00	0.15 ***	0.01	0.01	0.01
Elder People	0.43 ***	0.04	0.07 *	0.03	0.01	0.02	0.03	0.02	-0.12 ***	0.03
Deindustrialization	0.06 ***	0.02	0.05 ***	0.01	-0.00	0.01	0.03 ***	0.01	0.10 ***	0.01
GDP per Head	-0.04 ***	0.01	-0.01	0.01	-0.01 ***	0.00	-0.01	0.01	-0.03 ***	0.01

GDP Growth	-0.02 **	0.01	-0.04 ***	0.01	-0.00	0.00	-0.01 **	0.00	-0.00	0.01
Inflation	-0.01	0.01	-0.01	0.01	-0.00	0.00	0.01	0.01	-0.02 *	0.01
Spline 1	-0.05	0.03	-0.10 ***	0.03	0.09 ***	0.02	0.01	0.02	-0.06 ***	0.02
Spline 2	0.04	0.02	0.11 ***	0.02	-0.06 ***	0.01	-0.03 *	0.01	0.05 ***	0.01
Social Democrats	0.08	0.13	-0.11	0.10	-0.01	0.04	0.03	0.07	-0.05	0.10
Random Parts										
σ^2	0.071		0.046		0.009		0.019		0.063	
$\tau_{00,\ government}$	0.465		0.252		0.042		0.124		0.332	
$\tau_{00,\ year}$	0.000		0.002		0.000		0.000		0.008	
$\tau_{00,\ country}$	4.984		0.530		0.223		0.499		0.479	
$N_{\ government}$	221		221		190		220		252	
$N_{\ year}$	30		30		30		30		41	
$N_{\ country}$	23		23		23		23		23	
$ICC_{\ government}$	0.084		0.303		0.155		0.194		0.376	
$ICC_{\ year}$	0.000		0.002		0.000		0.000		0.009	
$ICC_{\ country}$	0.903		0.639		0.814		0.777		0.543	
Observations	606		613		540		603		623	
R^2 / Ω_0^2	.992 / .992		.977 / .977		.976 / .976		.989 / .989		.975 / .975	
AIC	891.649		571.680		-413.189		67.088		821.096	

Robustness

To further probe the robustness of our findings we run several models with additional control variables and different model specifications. Moreover, we tested different operationalizations using, for example, many other CMP categories (e.g., "free market economy", "market regulation", "Keynesian demand management") or the CMP's ready-made scales ("rile", "welfare", "planeco/markeco"). The only noteworthy finding regards our manifesto-based item: When we only use the "government expansion" item (and disregard the "welfare limitation" issue), we detect a smaller effect (0.03), which is not significant at a five-percent level anymore. Yet, this is plausible given that ignoring the "welfare limitation" item ignores an important part of the variation and is a more imprecise measure of the governments' preferences. In addition to these tests, we checked whether the exclusion of short governments makes a difference for the findings on control variables (it does not). Also, using country fixed effects instead of random intercepts does not affect the results substantially either but is less efficient due to the loss of degrees of freedom. The diagnostics revealed no problems either, as the error terms are homoscedastic and approximately normally distributed (see Appendix Figures A&B). We also tested for time-varying effects by looking at models by decade but again produced similar non-results.

8. Conclusion and discussion

In all advanced economies, public welfare expenditure amounts to between one sixth and one third of GDP. Consequently, scholars have investigated determinants of public social spending, focusing particularly on the impact of governing parties. In this article, we challenged this literature, arguing that their findings might be incorrect due to a crucial methodological misspecification: The common standard is to apply pooled time-series cross-

section (TSCS) regressions to annual observation data ('country-years'). We argued that country-years are misleading when one studies effects of governing parties on policies in a TSCS setting, because governments usually do not change on a yearly basis. The existing studies thus artificially inflate the number of observations, as they do not control for the nested structure of the data, resulting in incorrect estimates. Some recent contributions (Garritzmann/Seng 2016; Persson et al. 2007; Schmitt 2016; Vis 2011, 2012) raised similar criticism and proposed using government-terms as the unit of analysis instead. While this indeed allows more precise estimation of government effects, we argued that it introduces new problems (regarding the comparability of observations and the ignorance towards variables with different time horizons). Scholars thus face a dilemma between two designs that both have strengths and weaknesses but remain imperfect solutions.

Mixed-Effects Models provide a better fit for modeling the nested and complex cross-classified structure of spending data (nested in countries, years, and cabinets) and circumvent the trade-off between the two other approaches. Empirically, we applied these models to total public social spending over more than five decades as well as to expenditure on the five largest welfare policies over three to four decades in 23 advanced democracies. We started by comparing results from a standard country-year setup without taking care of the observations being clustered in cabinets, which produced highly significant findings. We then showed that when moving to a more accurate Mixed Effects setup, significant levels drop a lot. In our setup, we did not find any partisan differences when using broad left-right-center categories. But we still found a significant negative within-effect for social democratic parties on social security transfers that mainly stems from the earlier years of our sample and disappears over time. Most importantly, finally, we found that parties placing more emphasis on welfare in their manifestos (in our view the best available measure of

government preferences) also do increase public welfare expenditure more, supporting assumptions in models of democratic representation.

Yet, when turning to policy-specific spending we did not detect any significant partisan effects. This might result from the fact that – for reasons of data availability – our analysis of policy-specific expenditure focuses on the post-1980 period, which might simply be 'too late' to detect partisan effects: It still might be the case that parties affected spending in the period prior to our investigation, triggering path dependencies in the aftermath (as for example implied by Pierson (1996) or more explicitly argued in Garritzmann's [2016] 'Time-Sensitive Partisan Theory'). Yet, unfortunately comparative data is unavailable to test this claim quantitatively.

We believe these arguments and findings have considerable consequences for the literatures on the welfare state and specific social policies, as well as for the public policy and political economy literature more generally. Substantively, our findings for total public welfare spending imply that scholars interesting in public policy-making should not only focus on the legislative policy-making arena, but also pay close attention to parties' behavior in the electoral arena. Put differently, in order to understand partisan effects on public policy, we also need to consider the type of party competition in the respective field.

Besides the substantial findings, a more general take-away is methodological: We showed how methodological questions are highly intertwined with substantive questions and findings. The methods you chose at least partly affect the results you get. We showed how Mixed-Effects Models can be a very useful alternative to both the country-year and the cabinet-term approach. Using Mixed-Effects Models allows combining the strengths of both approaches while avoiding their respective shortcomings, thus circumventing the trade-off between two other approaches. Needless to say, our approach is also not without flaws. One

difficultly of Mixed-Effects Models is pragmatic, as given a complex data structure, limited data availability, and a quite demanding model specification, these models can quickly run into convergence problems. Another note of caution is that our analysis has focused entirely on the national level (as is customary in this literature), but parties might still affect spending at subnational levels (Kleider et al. 2017).

While we applied the procedure to public expenditure, it equally applies for other welfare measures (e.g., generosity scores) and other policy outputs more generally. Our arguments and approach travel beyond welfare state research and apply to any public policy outcome. Future work could thus explore the use of these models for other outputs. Moreover, future research could add additional layers of theoretical and empirical complexity that we had to disregard in this paper. For example, one could pay additional attention to the topic of time and timing and explore how government durations and sequences play a role; one could add random slopes to test for more complex theoretical expectations; or one could expand the empirical scope of the study by studying other country groups or other outputs.

References

Allan, James P., and Lyle Scruggs. 2004. "Political partisanship and welfare state reform in advanced industrial societies." *American Journal of Political Science* 48(3):496-512.

Ansell, Ben W. 2010. *From the ballot to the blackboard: The redistributive political economy of education*. New York: Cambridge University Press.

Armingeon, Klaus, Laura Knöpfel, David Weisstanner, and Sarah Engler. 2014. Comparative Political Data Set I 1960-2012. Bern: Institute of Political Science, University of Bern.

Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to do (and not to do) with time-series cross-section data." *The American Political Science Review* 89(3):634-647.

Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42(4):1260-1288.

Beck, Nathaniel, and Jonathan N. Katz. 2011. "Modeling Dynamics in Time-Series–Cross-Section Political Economy Data." *Annual Review of Political Science* 14(1):331-352.

Boix, Carles. 1998. *Political parties, growth and equality - Conservative and social democratic economic strategies in the world economy*. Cambridge: Cambridge University Press.

Brady, David, Jason Beckfield, and Martin Seeleib-Kaiser. 2005. "Economic globalization and the welfare state in affluent democracies, 1975-2001." *American Sociological Review* 70(6):921-948.

Bräuninger, Thomas. 2005. "A partisan model of government expenditure." *Public Choice* 125(3-4):409-429.

Breunig, Christian, and Marius R. Busemeyer. 2012. "Fiscal austerity and the trade-off between public investment and social spending." *Journal of European Public Policy* 19(6):921-938.

Busemeyer, Marius R. 2009. "From myth to reality: Globalisation and public spending in OECD countries revisited." *European Journal of Political Research* 48(4):455-482.

Busemeyer, Marius R., and Christine Trampusch. 2011. "Review Article: Comparative Political Science and the Study of Education." *British Journal of Political Science* 41:413-443.

Castles, Francis G. 1998. *Comparative Public Policy. Patterns of Post-war Transformations*. Cheltenham: Edward Elgar.

Castles, Francis G. 2009. "What Welfare States Do: A Disaggregated Expenditure Approach." *Journal of Social Policy* 38(1):45-62.

Di Tella, Rafael, and Robert J. MacCulloch. 2002. "The Determination of Unemployment Benfits." *Journal of Labor Economics* 20(2):404-434.

Emmenegger, Patrick, Silja Häusermann, Bruno Palier, and Martin Seeleib-Kaiser. 2012. *The Age of Dualization. The Changing Face of Inequality in Deindustralized Societies*. Oxford: Oxford University Press.

Fernández, Juan J. 2012. "Explaining the introduction of automatic pension indexation provisions in 17 OECD countries, 1945-2000." *Journal of European Social Policy* 22(3):241-258.

Fernandez, Raquel, and Richard Rogerson. 1995. "On the Political Economy of Education Subsidies." *The Review of Economic Studies* 62(2):249-262.

Fervers, Lukas, Philipp Oser, and Georg Picot. 2015. "Globalization and healthcare policy: A constraint on growing expenditures." *Journal of European Public Policy* 23(2):197-216.

Garrett, Geoffrey, and Deborah Mitchell. 2001. "Globalization, government spending and taxation in the OECD." *European Journal of Political Research* 39(2):145-177.

Garritzmann, Julian L. 2016. *The Political Economy of Higher Education Finance. The Politics of Tuition Fees and Subsidies in OECD countries, 1945-2015*. Basingstoke: Palgrave Macmillan.

Garritzmann, Julian L. and Kilian Seng. 2016. "Party Politics and Education Spending: Challenging Some Common Wisdom." *Journal of European Public Policy* 23(4):510-530.

Häusermann, S., Picot, G., & Geering, D. 2013. "Rethinking party politics and the welfare state–Recent advances in the literature." *British Journal of Political Science* 43(1):221-240.

Henderson, C.R. 1975. "Best Linear Unbiased Estimation and Prediction under a Selection Model." *Biometrics* 31(2):423-447.

Henderson, C.R. 1982. "Analysis of Covariance in the Mixed Model: Higher-Level, Nonhomogeneous, and Random Regressions." *Biometrics* 38(3): 623-640.

Hemerijck, Anton. 2013. *Changing welfare states*. Oxford: Oxford University Press.

Herwartz, Helmut, and Bernd Theilen. 2014a. "Health care and ideology: A reconsideration of political determinants of public healthcare funding in the OECD." *Health Economics* 23:225-240.

Herwartz, Helmut, and Bernd Theilen. 2014b. "Partisan influence on social spending under market integration, fiscal pressure and institutional change." *European Journal of Political Economy* 34:409-424.

Hibbs, Douglas A. 1977. "Political parties and macroeconomic policy." *American Political Science Review* 71:1467-1487.

Hicks, Alexander, and Kendralin Freeman. 2009. "Pension Income Replacement: Permanent and Transitory Determinants." *Journal of European Public Policy* 16(1):127-143.

Hicks, Alexander, and Duane Swank. 1992. "Politics, institutions, and welfare spending in industrialized democracies, 1960-1982." *American Political Science Review* 86:658-674.

Hinrichs, Karl, and Julia F. Lynch. 2010. "Old-age pensions". In *The Oxford Handbook of the Welfare State*, edited by Francis G. Castles, Stephan Leibfried, Jane Lewis, Herbert Obinger, Christopher Pierson. Oxford: Oxford University Press.

Huber, Evelyne, and John D. Stephens. 1993. "Political Parties and Public Pensions. A Quantitative Analysis." *Acta Sociologica* 36:309-325.

Huber, Evelyne, Charles Ragin, and John D. Stephens. 1993. "Social-Democracy, Christian Democracy, Constitutional Structure, and the Welfare-State." *American Journal of Sociology* 99(3):711-749.

Huber, Evelyne, and John D. Stephens. 2001. *Development and Crisis of the Welfare State: Parties and Policies in Global Markets*. Chicago: University of Chicago Press.

Huo, Jingjing J., Moira Nelson, and John D. Stephens. 2008. "Decommodification and activation in social democratic policy: Resolving the paradox." *Journal of European Social Policy* 18(1):5-20.

Immergut, Ellen. 1990. "Institutions, Veto Points, and Policy Results: A Comparative Analysis of Health Care." *Journal of Public Policy* 10(4):391-416.

Iversen, Torben, and David Soskice. 2006. "Electoral Institutions and the Politics of Coalitions. Why Some Democracies Distribute More Than Others." *American Political Science Review* 100(2):165-182.

Iversen, Torben, and John D. Stephens. 2008. "Partisan politics, the welfare state, and three worlds of human capital formation." *Comparative Political Studies* 41(4-5):600-637.

Jensen, Carsten. 2010. "Issue compensation and right-wing government social spending." *European Journal of Political Research* 49:282-299.

Jensen, Carsten. 2011a. "Catching up by transition: Globalization as a generator of convergence in social spending." *Journal of European Public Policy* 18(1):106-121.

Jensen, Carsten. 2011b. "Marketization via compensation: Health care and the politics of the right in advanced industrialized nations." *British Journal of Political Science* 41:907-926.

Jensen, Carsten. 2012. "Two Sides of the Same Coin? Left-Wing Governments and Labour Unions as Determinants of Public Spending." *Socio-Economic Review* 10(2):217-240.

Jensen, Carsten. 2014. *The Right and the Welfare State*. Oxford: Oxford University Press.

Jordan, J. 2011. "Health care politics in the age of retrenchment." *Journal of Social Policy* 40(1):113-134.

Kittel, Bernhard. 1999. "Sense and sensitivity in pooled analysis of political data." *European Journal of Political Research* 35(4):533-533.

Kittel, Bernhard, and Herbert Obinger. 2003. "Political parties, institutions, and the dynamics of social expenditure in times of austerity." *Journal of European Public Policy* 10(1):20-45.

Kittel, Bernhard, and Hannes Winner. 2005. "How reliable is pooled analysis in political economy? The globalization-welfare state nexus revisited." *European Journal of Political Research* 44(2):269-293.

Kleider, Hanna, Leonce Röth, and Julian L. Garritzmann. 2017. "Ideological Alignment and the Distribution of Public Expenditure." *West European Politics* 41(3):779-802.

Korpi, Walter. 1983. *The Democratic Class Struggle*. London: Routledge.

Lindstrom, M.L., and D.M. Bates. 1988. "Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data". *Journal of American Statistical Association* 83(404):1014–1021.

Loftis, Matt W., and Peter B. Mortensen. 2017. "A New Approach to the Study of Partisan Effects of Social Policy." *Journal of European Public Policy* 24(6):890-911.

McLean, Robert A., Sanders, William L., and Walter W. Stroup. 1991. "A Unified Approach to Mixed Linear Models". *The American Statistician* 45(1):54–64.

Morel, Nathalie, Bruno Palier, and Joakim Palme, eds. 2012. *Towards a social investment welfare state: Ideas, policies and challenges*. Bristol: The Policy Press.

Myles, John. 1989. *Old Age in the Welfare State: The Political Economy of Public Pensions*. Kansas: University Press of Kansas.

Obinger, Herbert, and Uwe Wagschal. 2010. "Social Expenditure and Revenues." In *The Oxford Handbook of the Welfare State*, edited by Francis G. Castles, Stephan Leibfried, Jane Lewis, Herbert Obinger, Christopher Pierson. Oxford: Oxford University Press.

Palme, Joakim. 1990. *Pension Rights in Welfare Capitalism: The Development of Old-Age Pensions in 18 OECD Countries 1930 to 1985*. Stockholm: Swedish Institute for Social Research.

Persson, Torsten, Gerard Roland, and Guido Tabellini. 2007. "Electoral rules and government spending in parliamentary democracies." *Quarterly Journal of Political Science* 2(2):155-188.

Pierson, Paul. 1996. "The new politics of the welfare state." *World Politics* 48(2):143-179.

Plümper, T., V.E. Tröger, and P. Manow. 2005. "Panel data analysis in comparative politics: linking method to theory." *European Journal of Political Research* 44(2):327-354.

Potrafke, Niklas. 2009. "Did globalization restrict partisan politics? An empirical evaluation of social expenditures in a panel of OECD countries." *Public Choice* 140(1-2):105-124.

Potrafke, Niklas. 2010. "The growth of public health expenditures in OECD countries: Do government ideology and electoral motives matter?" *Journal of Health Economics* 29:797-810.

Reeves, Aaron, Martin McKee, Sanjay Basu, and David Stuckler. 2014. "The political economy of austerity and healthcare: Cross-national analysis of expenditure changes in 27 European nations 1995-2011." *Health Policy* 115:1-8.

Rueda, David. 2005. "Insider-Outsider Politics in Industrialized Democracies: The Challenge to Social Democratic Parties." *American Political Science Review* 99(1):61-74.

Savage, Lee. 2018. "The Politics of Social Spending after the Great Recession: The Return of Partisan Policy Making." *Governance,* OnlineFirst: https://doi.org/10.1111/gove.12354.

Schmidt, Manfred G. 1996. "When parties matter: A review of the possibilities and limits of partisan influence on public policy." *European Journal of Political Research* 30(2):155-183.

Schmitt, Carina. 2016. "Panel data analysis and partisan variables: how periodization does influence partisan effects." *Journal of European Public Policy* 23(10):1442-1459.

Stegmueller, Daniel. 2013. "How Many Countries for Multilevel Modeling? A Comparison of Frequentist and Bayesian Approaches." *American Journal of Political Science* 57(3): 748-761.

Stephens, John D. 1979. *The transition from capitalism to socialism*. London: Macmillan.

Streeck, Wolfgang, and Daniel Mertens. 2011. "Fiscal austerity and public investment: Is the possible the enemy of the necessary?" *MPIfG discussion paper* 11/12.

Swank, Duane. 2002. *Global capital, political institutions, and policy change in developed welfare states*. Cambridge: Cambridge University Press.

Tepe, Markus, and Pieter Vanhuysse. 2010. "Who Cuts Back and When? The Politics of Delays in Social Expenditure Cutbacks, 1980-2005." *West European Politics* 33(6):1214-1240.

Tepe, Markus, and Pieter Vanhuysse. 2013. "Parties, Unions and Activation Strategies: The Context-Dependent Politics of Active Labour Market Policy Spending." *Political Studies* 61(3):480-504.

van Kersbergen, Kees. 1995. *Social Capitalism. A Study of Christian Democracy and the Welfare State*. London: Routledge.

van Vliet, Olaf, and Ferry Koster. 2011. "Europeanization and the political economy of active labour market policies." *European Union Politics* 12(2):217-239.

Vis, Barbara. 2011. "Under which conditions does spending on active labor market policies increase? An fsQCA analysis of 53 governments between 1985 and 2003." *European Political Science Review* 3(2):229-252.

Vis, Barbara. 2012. "The Comparative Advantages of fsQCA and Regression Analysis for Moderately Large-N Analyses." *Sociological Methods & Research* 41(1):168-198.

Vlandas, Tim. 2013. "Mixing apples with oranges? Partisanship and active labour market policies in Europe." *Journal of European Social Policy* 23(1):3-20.

Volkens, Andrea, O. Lacewell, S. Regel, H. Schultze, and A. Werner. 2011. The Manifesto Data Collection: Manifesto Project. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB).

Wilensky, Harold L. 1975. *The Welfare State and Equality: Structural and Ideological Roots of Public Expenditures*. Berkeley, Los Angeles, and London: University of California Press.