Three Essays on the Measurement of Political Ideology

Phil Swatton

A thesis submitted for the degree of PhD Government

Department of Government
University of Essex

30/09/2022

Contents

List of Figures

List of Tables

Preamble

Acknowledgements

Somewhat unusually, I would like to begin the acknowledgements of this thesis by thanking the British public. While I'm not particularly grateful for the electoral and referendum results of the past few years, they've been crucial in giving me interesting phenomena and events to research. So, if nothing else, I suppose I owe them that much.

More conventionally, I would like to acknowledge the numerous people who have helped me during the process of pursuing this PhD. Like anyone in any stage of academic life, I have benefited enormously from the advice, feedback, and support of others.

I would like to begin by thanking my supervisor, Rob Johns, for all of his help not only during the thesis but at the very beginning, when I first began writing my proposal during my masters'. It is simply impossible to imagine a way in which this document would look the way it does now without Rob's help. From introducing me to factor analysis when I first discussed my concerns about acquiescence bias, through the many conversations and advice sessions over the years to the final help on the framing and tidying up of this thesis, Rob's help has been crucial.

I have been similarly fortunate in the many board members who have also given me advice and support while I have been pursuing my PhD. Dominik Duell, Paul Whiteley, Roi Zur, and Ryan Bakker have all been invaluable sources of feedback and ideas for which I am incredibly greatful. I also wish to additionally thank Roi for his help prior to joining the board, and in particular on introducing me to the simulation modelling approach. It is through academics like Roi who make a particular effort for PGR students that things will get better.

An important set of contributions to this thesis were made during the viva, which led to corrections which have significantly improved this document. I would therefore like to thank my internal and external examiners, Seonghui Lee and Ben Lauderdale, for their time and feedback. I would also like to thank the chair of my viva, Linsey McGoey, for efficiently running the examination.

I also wish to thank Royce Carroll, for whom I have acted as a research assistant and coauthor. While not the only research assistantship or coauthorship I have pursued during the PhD, it is during this one I have been exposed to an incredibly wide range of literature on ideal point estimation, which has dramatically shaped my own view. I also wish to thank Royce for all the opportunities I have had with the Essex Summer School, which has dramatically shaped my post-PhD prospects for the better, and given me many valuable learning opportunities.

Although it has been a long time since I was at Keele University, I wish to thank two figures there who nonetheless had an impact on my journey here at Essex. Jon Herbert, who advised me on my Undergraduate dissertation, was the person who advised me to further pursue my interests here at Essex. It is as a direct consequence of this advice I am here today submitting this thesis. I also wish to thank Matthew Wyman, my personal advisor at Keele, who not only throughout my journey at Keele gave me invaluable support, but continued to do so during my time at Essex - at one stage inviting me to a Keele TA training session in advance of my first teaching job.

As have many others, I have been lucky to be a part of the Essex PhD community. For me, this community has been welcoming, interesting, challenging, engaging, fun, and a valuable source of learning. In particular, there are a few I wish to mention, in reverse alphabetical order[1]. Sarah Wagner has been an invaluable source of both friendship and advice. I continue to miss our coffee sessions, and to benefit from explanations of how to navigate academia. Laura Saveedra-Lux was a brilliant and fun office mate for the first two years, and one that continues to be sorely missed (even if my productivity has enjoyed a boost in your absence). Anam Kuraishi has been a close friend, given indispensable advice on my papers, and co-organised the POPE PhD workshop with me. Lorenzo Crippa has been an excellent flatmate and friend, and I have learned a great deal from our conversations and debates with one another.

If I have forgotten anyone at Essex or neglected to highlight your help and contributions, I can only apologise. In my defence however, I am writing this at a point where my brain stopped working approximately sixty minutes ago.

It would be remiss of me not to mention the love and support of my family

[1]As someone whose surname begins with ‘S’, I am especially sensitive to alphabetical order as a default.

throughout this period. Though doubtless slightly baffled as to my continued studentship, they have been unerringly supportive throughout. Mum, Dad, Emily, Molly, Scott, Sally, and the extended family - thank you.

Finally, I have been fortunate in the later stage of my PhD to have met my partner Samira Diebire. She has been a source of emotional, academic, and career support, and has helped me in numerous ways I frankly could not even begin to recount. I have enjoyed going abroad for the first time in my adult life with her, had her read and comment on every component of my PhD, benefited from her help with the process of applying for jobs for after the PhD, and been taken care of by her. I am not entirely sure how I could have survived the closing stages of my PhD without her, and I am eternally grateful for the privilege of being her partner.

Abstract

In this thesis, I present three essays on the problem of measuring political ideology. Having initially set out to understand how globalisation, demographic changes, and new ideological divides contributed to the ongoing political era we find ourselves in, I quickly encountered measurement problems which could not be avoided. I therefore address various measurement issues, with a particular emphasis on survey measurement, while attempting to address this wider backdrop. I begin first by broadly defining ideology, discussing measurement theory, then discussing how this theory applies to ideology. I outline past approaches, then summarise the three essays that make up this thesis.

In the first essay, I directly address the problem of measuring the ide-

ology of voters through survey research. This task can be difficult, and researchers often rely on ‘off the shelf’ datasets. Many of these datasets contain unbalanced Likert scales, which risk acquiescence bias. This paper proposes a strategy for dealing with this issue. I first demonstrate using two comparable datasets from the UK how unbalanced scales produce distorted distributions and can affect regression results. Then, building on past research that utilises factor analysis to eliminate the influence of acquiescence bias, I demonstrate how researchers can utilise a person intercept confirmatory factor analysis model to obtain factor scores corrected for acquiescence in the case of fully unbalanced scales. I conclude with practical recommendations for researchers and survey designers moving forward.

In the second essay, I address age, period, and cohort effects as possible long-term drivers of both change and stability in political ideology in the electorate. However, the question as to the extent that the explanations these effects offer is consistent across countries has not been addressed. I therefore perform a comparative APC analysis of left-right political ideology. I run two side-by-side APC analyses of left-right positions, constraining effects to be common across countries in one and nesting them within country in the other. I pay special attention to the issue of how measures of ideology can be meaningfully compared, and develop a measure of relative ideology. I find evidence for ageing effects and life-cycle effects. Moreover, I find that while the constraint of common cohort effects is not a strong one, the constrain of common period effects is over-strong. Future research should focus first on better understanding this contrasting result, and second on developing absolute measures to better understand patterns of change and continuity in left-right ideological positions in the public.

In the final essay, I address the problem that social democratic parties have been confronted with vis a vis the rise of second dimension issues. These issues often see social democratic parties facing a choice between competing portions of their own electorate. A particularly prominent second dimension issue is that of the EU: should social democratic parties take pro or anti-EU positions? I look at the case of the UK as an instructive example of this debate. In the build-up and aftermath of the 2019 UK General Election, a debate emerged regarding the optimal Brexit position for the Labour Party. This debate was ultimately without satisfactory conclusion as we do not observe counterfactual versions of reality - we witness only one version of events. I therefore estimate a narrow counterfactual, simulating how the Labour Party's vote share and seat count would have changed as its position on Brexit changes. I call this counterfactual narrow because I only consider the effect of these position changes on vote choice and turnout; and not any broader consequences. I run two simulations to compare the implications of pure proximity and proximity-categorisation models of vote choice. I generate seat predictions from the simulation results by using Uniform National Swing and Uniform Regional Swing. This allows me to assess the specific distributional claims made by those advocating for a more pro-Leave position for the Labour Party.

I conclude this thesis by highlighting the contributions of these three essays not only as standalone papers, but as a cohesive whole. I take a recent example of research in the Financial Times (FT) that attempts the kind of measurement I perform across the three papers, but which falls short in terms of measurement inference. I show how each paper speaks to a different aspect of the FT article, and could have influenced it in a better direction. I conclude by arguing that measurement inference is ultimately a good thing for political science, as it will lead to more secure results and richer substantive interpretation.

Chapter 1

Introduction: What's in a Title?

In 2016, like most people my age, I found myself not for the first time and not for the last on the losing side of a national vote[1]. Unlike the 2015 general election, or the 2014 EU parliament elections however, the consequences of being on the losing side of that vote profoundly shaped my political outlook and interests. In my undergraduate dissertation, I pursued the question of the relationship between age and ideology - a question I return to here. In my masters' dissertation, I pursued the question of the relationship between social class and EU referendum vote. Like many my age, the Brexit vote both captivated and horrified me.

Although a country leaving the European Union is a unique event, Brexit still belongs to a wider set of political trends and changes. Even in the 1960s, Martin voiced the surprise of political socialists at the authoritarianism of the working class (Lipset, 1960). They had expected to be the working classes to be on the side of social liberty, given that it was typically their political par-

[1]To this day, I have not voted for the winning political party in a UK general election.

ties that adopted these positions. Inglehart (1971, 1977, 1990, 1997) long ago posited his theory of cultural evolution, which emphasised a shift from 'materialist' value priorities focussed on the economy and crime towards 'postmaterialist' value priorities focussing on non-material issues.

It was however Kitschelt et al. (1994) who first fully articulated the notion of a non-economic dimension of ideological contestation cross-cutting the older economic one. Kitschelt was interested in this in particular as a problem for social democratic parties, but it has also presented a new set of opportunities and challenges for other political parties. Kriesi et al. (2006, 2008, 2012) defined the new political contestation as being between the 'winners' and 'losers' of globalisation, offering a theory that renders the new divide as much an economic one as not. This perhaps captures a separate but concurrent trend of the decline in the size of the working classes alongside the growth in the number of university graduates as an electoral bloc (Kitschelt and Rehm, 2014; Ford and Jennings, 2020). Importantly - especially for social democratic parties - these two core constituencies of the center-left hold wildly different values on the new dimensions of contestation.

Hooghe, Marks and Wilson (2002) offered a somewhat different account, focusing more explicitly on new ideological dimensions. They discovered that a dimension covering green-alternative-liberal versus traditional-authoritarian-nationalist ideology better corresponded to political party support for the European Union than economic left-right positions (Hooghe, Marks and Wilson, 2002). For Hooghe and Marks (2009, 2018), the rise in the GAL-TAN dimension has been closely associated with the rise in issues of borders, immigration, and national identity that have accompanied the European Union.

I had therefore set out to better understand this set of trends and their competing - if not exclusive - explanations. In particular, I had hoped to begin by examining at the under-explored relationship between age and political ideology. From the moment I begun however, I had concerns regarding the extant measures of political ideology available to me. In particular, I was concerned about the presence of acquiescence bias in the likert scales measuring left-right and libertarian-authoritarian positions in the British Election Study Internet Panel (Fieldhouse et al., 2020). Simply put, acquiescence response style (ASR) is a resonse style wherein some respondents have a tendency towards agreeing with survey items regardless of their content. This results in biased scales. That acquiescence was present in these scales is established: it was noted by the creators of the scales (Evans and Heath, 1995)!. Considerably less established was the question of how to deal with this problem.

This question is not trivial: one well-known correlate of acquiescence response style is education level (Ware Jr, 1978; Winkler, Kanouse and Ware, 1982). Failing to deal with acquiescence bias in the likert scale was a clear potential threat to valid inference. Similarly, on a descriptive level, the scales are biased in a left-wing and authoritarian direction. Yet, it is often reported based on the evidence of these scales (or similar ones) that the UK electorate is left-wing and authoritarian (see e.g. Webb and Bale, 2021, p. 158). This notion of a left-wing and authoritarian[2] electorate is ubiquitous, and has even made its way to mainstream newspapers (see e.g. Surridge, 2019).

The ubiquity of this notion is a direct product of the skew in the scales in the British Election Study Internet Panel and other popularly used datasets.

[2] Sometimes ‘socially conservative’ is used instead of ‘authoritarian’.

This is despite the fact that the bias in these scales has been known since their inception! This is a common theme throughout my work on the thesis: it is often the case that the problems in the measurement of political ideology are well-known. They are simply rarely acted on. It is my hope, therefore, that my work will contribute to a greater willingness on the part of political scientists to engage not only with the contents of this thesis but the broader literature on measurement.

Through my early frustrations with the scales in the British Election Study, this thesis having begun its life as a thesis dealing with a kaleidoscope of themes of ideology, issue salience, realignment, and globalisation morphed into a thesis on the measurement of ideology. The genesis of this thesis remains present throughout; essays 2 and 3 in particular are cases of applied measurement of ideology, with applications in realms directly relevant to the literature I have discussed above. But the problem of measurement has for me become inseparable from any attempt to tackle the broad political trends and changes set out above.

The titular three essays of this thesis are thus not purely papers in the realm of measurement. Indeed, only the first essay is explicitly a methodological papers. But, as I will discuss in more detail below, essays 2 and 3 are also characterised by a particular attention to and concern with the problems of measurement of political ideology. Before I introduce the essays however, I first wish to address two particular problems. These are problems that are continuously present throughout the three essays, an thus also serve to unite them. First, the meaning of ideology. Ideology is a good example of a heavily contested term, and it is important that some common meaning can be established.

Second, measurement models and theory. As I will discuss, measurement is not a straightforward task, and it remains heavily neglected.

1.1 Defining Ideology

Political ideology is a contentious concept. Most political scientists will share some 'common sense' of what ideology is, but few will agree on precise definitions. I therefore aim to begin simply by setting out a reasonably well-known definition that captures many features that we would commonly agree we mean by 'ideology' while being flexible enough to facilitate a wide range of research.

I therefore take as my starting point Converse's definition of political ideology in terms of constraints. Converse (1964) defines ideology as a '*configuration of ideas and attitudes in which the elements are bound together by some form of constraint or functional interdependence*'. In a static sense, 'constraint' implies that knowing one belief enables us to more accurately guess at another belief of that individual; while in a dynamic sense it implies that a drastic change in one belief or attitude will also require changes in other elements of the configuration (Converse, 1964). These constraints can be logical, but also simply psychological - there is no a priori sense in which things go together, only that they do (Converse, 1964). At its simplest level, ideology is then a systematic way in which certain ideas and attitudes 'go together'.

Ideas and attitudes belonging together is however but one aspect of how we intuitively think about ideology. When several ideas and attitudes come together in this way, they form a holistic world-view. This world-view is both all-encompassing and simplifying. Political ideology in one respect acts as a heuristic shortcut for individuals, allowing them to more easily understand the world with less effort. Insofar as the way in which certain ideas and attitudes go together is systematic, we can begin to imagine a higher level of variation driving this lower level of ideas and attitudes. Peffley and Hurwitz (1985) take this notion and develop a hierarchical theory of political ideology, wherein causation flows from a high-level abstract to general domain attitudes, and then from domain attitudes to specific issue attitudes. This model is visualised in figure 1.1:

Figure 1.1: Hierarchical Model of Public Opinion

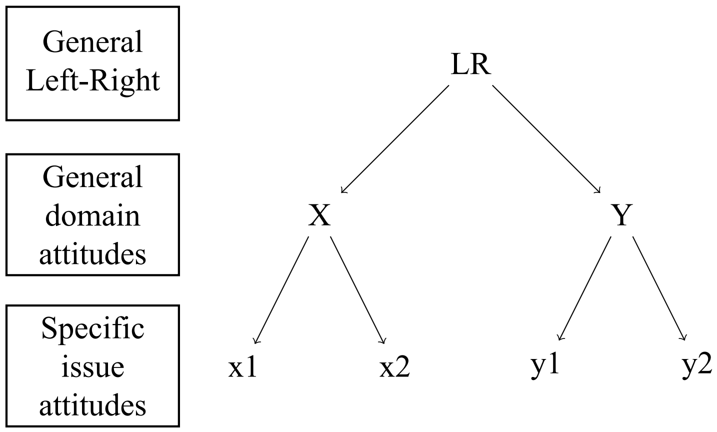


Figure 1.1 is instructive on several fronts. First, the arrows from the higher-level dimensions to the lower-level dimensions represent the constraints in the model. As Peffley and Hurwitz (1985) argue, a failure to properly consider the nature of constraints led to null results such as Converse (1964)'s. Second, the model makes clear that when we speak of ideology as a world-view, there are several levels of variation at play. We might mean the collection of indi-

vidual issue opinions, or we may mean the way the higher levels of variation constrain the issue opinions. Throughout this thesis, by 'ideology' I usually unless otherwise stated am referring to this second notion of ideology.

Holding much in common with Peffley and Hurwitz's hierarchical model is the theory of a basic ideological space. Also built on Converse's theory of constraints, the basic space theory posits that individuals with structured belief systems can have their beliefs represented in a low-dimensional space (Poole, 1998). This low-dimensional space was referred to as the basic space by Ordeshook (1976) and as a predictive dimension by Hinich and colleagues (Hinich and Pollard, 1981). Perhaps the key difference between the hierarchical model and the basic space is around the causal assumptions. Where the hierarchical model describes the relationship between the abstract dimensions and the specific issues in a causal manner, the basic space theory is causally agnostic. It only argues that these beliefs can be represented in a lower-dimensional space.

For a thesis on ideology, it may seem bizzare then to claim that individuals might need not possess this higher level of ideological variation in reality for research focussed on that higher level to be valid. However, it is not a necessary condition for the analysis of voter ideology that voters think in terms of say, left and right, or place themselves on such a dimension prior to being surveyed on that dimension. What matters is that we can accurately represent their views on such a dimension without too much loss of detail.

There are some reasons to imagine that the causal conceptual in the hiarchical model is likely to be correct in the case of voters. Indeed, Zaller and Feldman (1992) go so far to suggest that survey response instability is driven by the fact that few voters have views at the specific level, possess conflicting views at the higher level, and thus draw on these higher views with some degree of randomness. Ideology is not only an attribute of voters however. It also an attribute ascribed to political parties, political elites, and manifestoes. How meaningfully can we claim that a manifesto in reality possesses a higher level position that it draws on to form its specific issue declarations? Even if we suppose that voters necessarily possess a higher level ideology, we cannot assume that inanimate objects also do. For one it is something possessed, for another it is simply a useful summary of positions. As with voters, ideology is arguably a useful description or summary of the positions articulated in a manifesto or by a political party, rather than something that exists in the world prior to our conceptualising it and observing it.

In summary, I take Converse's theory of constraints as the core of my definition. Important here is the relationship between abstract dimensions of ideology and specific issue attitudes. I take the more abstract dimensions to simply be summaries or descriptions of several issue attitudes. This decision relies on the notion that many issue opinions tend to 'go together' in predictable ways. I remain agnostic on the question of whether ideology is an attribute that exists in the world prior to our observing it. Instead, I recognise that the existence of such a causal pathway is not necessary for the analysis of ideology.

1.2 Measurement

How can the fact that ideology may be either a real-world attribute or simply a useful summary be incorporated into a consistent approach to measurement? It is useful to take a step back and consider the task of measurement in the abstract, before then considering specific approaches to the measurement of ideology. Lauderdale (N.d.) defines measurement inference as inference from observed data to quantities describing the observed units. This differs from population inference (inference from observed data to the wider population) and causal inference (inference from observed data to what we would have observed for the same units given counterfactual circumstances).

A particularly useful typology of measurement inference is given by the competing theories of representational versus pragmatic measurement theory Lauderdale (N.d.). As competing approaches, they were first summarised by Hand (1996)[3], although both theories possess an older pedigree. Before beginning, it is useful to quickly define some key terms. In the following section, μ is the theoretical concept of interest, known as the *target concept*. m is the measurement we produce, and I_i is the ith indicator of target concept μ used to construct M. An indicator is defined as a partial or noisy realisation of the target concept. A useful example is the fact that when sports team a beats sports team b, this is an *indicator* for the concept of underlying ability (Lauderdale, N.d., p. 43). Similarly, if manifesto a advocates for redistribution and manifesto b against, this is taken as an indicator of underlying ideological orientation towards redistribution.

[3]Pragmatic measurement at this point in time was known as 'operational' measurement.

1.2.1 Representational versus Pragmatic Measurement

In representational measurement theory, real world objects are related to one another according to an empirical relationship system (ERS). For instance, three rods possess a 'length' attribute. These lengths can possess an order relationship (rod a is longer than rod b) and an additive relationship (the length of rods a and b are the same as the length of rod c). We can thus in turn construct a numerical relation system (NRS) from this ERS, quantitatively representing the empirical relationships between objects (Hand, 1996, 2016). There is therefore a direct causal relationship going from a real-world attribute to the measurement we eventually produce. Figure 1.2 visualises an example of this conceptualisation of measurement (Lauderdale, N.d.):

Figure 1.2: Representational Measurement

$$\begin{tikzcd}
\epsilon \arrow[r] \arrow[dr] \arrow[ddr] & I_1 \arrow[dr] & \\
\mu \arrow[ur] \arrow[r] \arrow[dr] & I_2 \arrow[r] & M \\
& I_3 \arrow[ur] &
\end{tikzcd}$$

Two features of figure 1.2 should be highlighted. First, figure 1.2 displays clear causal effects from μ to the various indicators, and from the indicators to M. If we did not utilise indicators, the causal arrows would flow directly from μ to M. Second, figure 1.2 introduces the concept of ϵ, which represents additional features that cause the indicator variables. Representational theory allows us to recognise that indicators are often imperfect realisations of μ: to use the earlier example, other real-world factors such as whether the sports game is home or away can produce the indicator we observe. It need not be common across indicators: different indicators may enjoy different additional influences.

Pragmatic measurement by contrast is a more epistemologically modest approach to measurement. In the pragmatic theory of measurement, the measure M defines the target concept μ (Hand, 1996, 2016; Lauderdale, N.d.). Instead of real-world objects possessing attributes that we numerically represent, attributes are defined as the outputs of measurement procedures (Hand, 1996). The usefulness of this approach is in the fact it is *discriminative*: real-world attributes may be defined by causal processes as above, but our descriptions of them need not be. One example is in temperature: we could use a representational measure such as kelvins or celsius, or we could simply use more subjective descriptions such as 'hot' and 'cold' that nonetheless better convey a human understanding of the world (Lauderdale, N.d., p. 20). Figure 1.3 visualises the pragmatic conception of measurement (Lauderdale, N.d.):

Figure 1.3: Pragmatic Measurement

$$\begin{tikzcd} & I_1 \arrow[dr] & \\ \mu \arrow[ur, dashed, leftrightarrow] \arrow[r, dashed, leftrightarrow] \arrow[dr, dashed, leftrightarrow] & I_2 \arrow[r] & M \\ & I_3 \arrow[ur] & \end{tikzcd}$$

Two clear differences emerge between figures 1.2 and 1.3. First, in figure 1.3 the causal relationships between μ and the indicators are left ambiguous. Instead, the indicators used to construct M may as much μ as be defined by μ. Second, ϵ has notably disappeared from figure 1.3. Since our measurement procedure defines both measurement and target concept, we can no longer speak of separate influences. Instead, where two measurement procedures produce two different results, they are instead seen to be producing two subtly different theoretical concepts (Hand, 1996).

1.3 Measuring Political Ideology

Many researchers on first encountering the two competing theories of measurement will have a speedy reaction in the direction of representational measurement. Yet, as Lauderdale (N.d.) notes, for social scientists many of our concepts are in many key respects better captured by the notion of pragmatic measurement. Lauderdale highlights the example of democracy: it is difficult to argue that whether countries either posses a binary status of democracy or a continuous amount of democracyness[4]. Instead, democracy is a *summary*, used to describe how the institutions of a nation are organised. The debates between various measures are best seen as debates between rival conceptualisations rather than rival realities of the world.

Ideology is similarly a fit for this pragmatic approach to measurement. The myriad ways of defining and measuring ideology are often more usefully seen

[4] In a module I taught using Lauderdale's work in progress textbook, many of my students opted for the continuous interpretation. Most however began questioning that once it was pointed out they would except that countries such as North Korea were clearly not democracies.

as differing conceptualisations, each offering different ways to discriminate between states in the world. The GAL-TAN measure for instance is a useful way in which to discriminate between the policy programmes and stances of political parties. Another similar definition is the notion of Libertarian-Authoritarian ideology, commonly used in the United Kingdom (Heath, Evans and Martin, 1994; Evans and Heath, 1995; Evans, Heath and Lalljee, 1996). Both are not so much 'correct' as 'useful'. Yet, a purely pragmatic approach is not fully satisfactory either. How can we for instance meaningfully speak about the kind of acquiescence bias I discussed above? In a purely pragmatic concept, measures with and without what I am labelling acquiescence bias are simply subtly different theoretical concepts. Obviously, even with the convincing reasons to favour a pragmatic theory of measurement this is unsatisfactory.

I therefore adopt throughout the essays the 'pragmatic realist' synthesis of both theoretical perspectives advocated by Lauderdale (N.d.). In practice, many measurement procedures sit somewhere between the two poles of representational and pragmatic measurement. It is idea to aim for representational, but we must accept that many measurements are simply summaries of data we have observed in the world. For my purposes, I adopt a particular synthesis: that different measurements of ideology often do simply represent different rather than obviously 'wrong' or 'right' conceptualisations, but that at the same time we can usefully speak about measurement error. There will be many sources of variation in our measurements that regardless of conceptualisation are not welcome.

A pragmatic approach embraces the causal ambiguity around higher levels of variation, but simultaneously need not reject the notion of systematic measurement error. Of the three essays, essay 1 is the most straightforwardly representational. It works with several different realisations of the same conceptual basis, and uses this to focus on measurement biases in likert scales. Essay 2 by contrast takes the most pragmatic approach, remaining entirely agnostic as to whether the left-right ideology it uses exists in reality or is simply a useful summary. Essay 3 finally sits between the two poles, using a priori knowledge to justify the use of a particular set of ideological scales (and thus arguably taking a pragmatic approach in choosing a particular conceptualisation), while on the other hand addressing the problems of differential item functioning and rationalization bias.

1.3.1 Indicators of Ideology

For a researcher interested in ideology in the abstract, the range of potential approaches to its measurement are mind-boggling. The choice will depend in part on which real-world objects we wish to measure ideology for: there is an intimate connection between the research we wish to perform and the data sources we will choose. The most obvious and traditional approach to the measurement of ideology is to simply ask. Survey-based measures of political ideology are as old as the discipline of political science itself. Likewise, survey respondents can be asked to locate the ideological locations of external stimuli such as political parties and elites. These processes however are subject to a myriad of influences external to the problem of ideology, which I discuss in further detail below.

Expert surveys represent an increasingly popular approach for the measurement of political party positions (see Castles and Mair, 1984; Laver and Hunt, 1992; Laver, 1994; Huber and Inglehart, 1995; Laver, 1998*a*,*b*; Laver and Mair, 1999; Ray, 1999; Hooghe et al., 2010; Bakker et al., 2015; Jolly et al., 2022). Here, experts absorb the various real-world signals of party position then are able to place a party on a given scale given a particular conceptualisation with a high degree of knowledge (Ray, 1999). The primary rival approach to the expert survey approach is the manifestos approach, which instead uses the external signals of the manifestos produced by political parties to estimate ideological locations. Approaches to manifestos have included coder-based approaches (Budge et al., 2001; Budge, 2001*a*,*b*, 2002; Elff, 2013) and quantitative text scaling based approaches (Laver, Benoit and Garry, 2003; Slapin and Proksch, 2008).

Coder-based approaches somewhat share the reliance on human intuition and its quantification, though is more based on the interpretation of manifestos alone rather than wider signals. Text scaling approaches can be use either coder based approaches (I include dictionary methods under this heading), latent scaling, or both (Hjorth et al., 2015). Where scaling is utilised, the ideological measure is instead revealed through the data, thus requiring post-estimation interpretation instead of pre-assigned interpretations.

Estimation of the positions of elites has typically focused on three sources. The most direct route is to simply survey the elites, although compared to other approaches this is often expensive and resource intensive. Instead, an alternative approach is the application of scaling procedures to votes in the legislatures (see e.g. Poole and Rosenthal, 1985, 1991, 2000; Poole, 2000; Poole and

Rosenthal, 2001). DW-NOMINATE has been in particular a success in the US, seeing use by organisations such as 538 (Grimmer, 2015). It represents perhaps one of the most successful methodological exports of political science (Grimmer, 2015). The approach of scaling votes has however struggled in high-discipline contexts such as the UK, where instead it seems to extract a pro vs anti-government dimension of variation instead of ideology as we would typically interpret it (Spirling and McLean, 2007). Similarly, quantitative text scaling is a promising approach but so far often suffers from the same output problems as vote scaling (Lauderdale and Herzog, 2016).

More recently, digital trace data have offered new avenues in attempts to measure ideology. There has been particular emphasis on measuring ideology on twitter, in no small part due to the ease of academic access to twitter data. Two approaches have dominated here. The first and most obvious route has been to utilise quantitative text scaling on tweets (Tumasjan et al., 2010; O'Connor et al., 2010). The second has instead focussed on twitter follower networks. Twitter has likewise been used to some degree of success for the measurement of elite ideological positions through text scaling (Barberá, 2015; Barberá et al., 2015). So far, both approaches have produced promising results, but it is not always easy to obtain data for other covariates of interest. Likewise, concerns remain regarding how representative the twitter population is of the wider general population we are often interested in (Mislove et al., 2011).

1.3.2 Ideology in the Electorate

In this thesis, I focus on measurement of political ideology from survey data. This is for the most part not because of any particular preference for survey data on methodological grounds, but instead stems from the research interests established in the introduction of this thesis. If I am to understand the causes and effects of political ideology in the electorate, survey data is the main data source. Twitter and other forms of digital trace data represent a promising avenue, but at present are weak on the availability of other variables of interest and the representativeness from the sample to the general population. Particular problems arise in the use of survey data to measure political ideology for my research goals, which I briefly describe here and address is more detail throughout the essays.

Survey Response Styles

First, and most relevant to essay 1 but a feature through all three essays is the problem of survey response styles. A survey response style, loosely defined, is a tendency to respond to a particular survey item in a particular way regardless of the content of that item. Acquiescence response style is defined as a tendency towards agreement regardless of the content of an item (Ware Jr, 1978; Ray, 1979). Extreme response style is a tendency towards placing oneself on the poles of an ordinal rating scale (Hui and Triandis, 1989; Greenleaf, 1992). Non-extreme response style is instead a tendency towards placing oneself on the center of a scale (or at least systematically avoiding the extremes) (Wetzel, Carstensen and Böhnke, 2013).

In a recent paper, Barnes (2022) used a latent class analysis to analyse zero-sum responses (later used in paper 1 of this thesis). Alongside substantively meaningful classes, her analysis also found several clear reponse styles: acquiescent response style, non-committal response style (always opting for the middle option), and respondents who always opted for 'dont' know'. The fact so many came from a single class analysis of a single survey battery shows how important it is not to neglect this part of the data generating process.

These response styles are most easily seen as problematic from a representational standpoint. However, even accepting that measurement and conceptualisation are one and the same, we are forced to confront the fact it is difficult to include survey response styles in our concepts. As discussed above, a failure to reckon with survey response styles has typically led to incorrect inferences being drawn in the past.

Survey Item Comparability

More relevant to essays 2 and 3 is the problem of survey item comparability. This is a problem that takes several forms. First are survey response differences. The first of these is differential item functioning (DIF), which is defined as a tendency for respondents to give two different responses given the same underlying perception (King et al., 2004). This is a particularly thorny problem for self-placements and placements of external stimuli on external rating scales. Similar is the issue of rationalization bias. Survey respondents tend to place external stimuli they like closer to themselves on a scale, and those they dislike further away (Aldrich and McKelvey, 1977). Raw survey responses cannot thus be taken at face value.

Related is the issue of survey item meaning. In cross-country and cross-temporal surveys, it is reasonable to expect that the substantive interpretation of survey items will change over time and space. One obvious example is left-right placements. A UK respondent in 2019 will likely place much more emphasis on the EU positions of parties in placing them on a left-right scale than will a respondent in 1997. Likewise, a French respondent in placing En Marche to the left of the Rassemblement Nationale today will be placing far less emphasis on redistribution than on cultural issues - a stark difference from say, 20 years ago.

Measuring Voters and Other Objects on the Same Scale

Finally, insofar as we are interested in the electoral implications of the new ideological dimensions, how can voters be related to other objects? How can they be related to political parties or elites? Given the presence of differential item functioning and rationalization bias this is doubly problematic. How can we speak of voter-party spatial distance when there are biases in survey response even given the same underlying perception?

1.4 The Three Essays

These are the kinds of measurement question I address in this thesis. In part, many of these questions have been answered in the past, but these answers have been neglected. This has partly been a general forgetting of these results, but is also arguably driven by the fact that measurement inference is often inconvenient. In essay 1 alone is the development of new methodology the sole focus on the paper. The other two essays are however cases of applied measurement. I am not concerned with establishing the veracity of a new measurement so much as offering a set of results that are consistent with robust approaches to the above set of problems.

A common way of thinking about measurement has been to decompose a given measure M into three constituent variance components. These are variation relating to the concept of interest, variation stemming from measurement method, and random noise (Kenny and Kashy, 1992). The point of the preceding discussion is not that all hitherto research has been pointless, or that the issues of measurement are intractable. Refinement of both concepts and methodological approaches is a normal part of the scientific process. My hope instead is that I hope these essays aid in shifting the respective proportions of these components in the direction of information - that our measures more directly capture our concepts. Without further delay, I therefore proceed to introducing the three essays and their contributions.

1.4.1 Essay One

The first essay, *Agree to Agree: Correcting Acquiescence Bias in the Case of Fully Unbalanced Scales with Application to UK Measurements of Political Beliefs*, develops a solution to a particular problem that can emerge when using Likert scales to measure ideology. Acquiescence bias is a form of survey response bias where irrespective of survey item content, individuals are more likely to agree with that item than they otherwise should be.

When therefore utilising Likert scales - a type of measure built on additively aggregating agree-disagree statements - it is important to ensure that the scale is ‘balanced’ in the sense of having items from both sides of the ideological spectrum. Where this is not done, the scale is unbalanced, and biased in the direction of the imbalance. Where a scale is fully unbalanced, this bias becomes impossible to identify from the indicators alone. We have no way of knowing who is agreeing due to acquiescence, and who is agreeing due to genuine agreement, without the presence of some kind of contradiction in responses.

In this essay, I build on previous work on balanced likert scales to address this problem. I take the work of Evans and Heath as my case study, as their two scales (left-right and libertarian-authoritarian) have been the conceptual underpinning for scales in several UK datasets. I begin by demonstrating the difference between the scales in the British Election Study face-to-face survey and the British Social Attitudes survey. The scales in these surveys have the same theroetical underpinning and several shared indicators.

However, the scales in the British Election Study are (mostly) balanced while the scales in the British Social Attitudes Survey are not. I begin this essay by demonstrating exactly this result. I further show that there is no systematic difference between the common indicators in the datasets - strongly suggesting that the differences must come entirely from scale design. With the difference established, I turn to wave 14 of the British Election Study Internet Panel

I then proceed to develop the person-intercept confirmatory factor analysis model. In this model, a person-specific intercept is added that remains constant across items. I build on past work on this model, removing the unnecessary requirement that the intercept is uncorrelated with other factors and showing two different ways of constructing the model. Once both continuous and ordinal models can be produced, this results in four variations of the person-intercept confirmatory factor analysis model.

To identify the acquiescence bias in the fully unbalanced scales, I introduce the use of a second, unrelated, balanced scale. Without additional information, it is impossible to identify the acquiescence bias within a fully unbalanced scale. I show, by way of comparison to the earlier demonstration and differences between the British Election Study and British Social Attitudes Study, that it is possible to reproduce the results of the British Election Study despite possessing a fully unbalanced scale.

I end with recommendations for researchers and survey designers. Researchers must be alive to the problem of acquiescence bias specifically and survey response bias more broadly. However, acquiescence bias is ultimately best dealt with at the survey design stage. I therefore recommend that going forward, survey designers show greater awareness as to the risks posed by acquiescence bias.

1.4.2 Essay Two

The second essay, *Age Isn't Just a Number: A Comparative Age-Period-Cohort Analysis of Political Beliefs in Europe* enters the realm of applied work in measurement to perform an age-period-cohort (APC) analysis of left-right ideological positions in Western Europe. Where past APC analysis has been performed, it has either been a single country case study or it has not examined

APC trends in left-right ideology. This is an important gap: there are often common political trends in Western European countries. I therefore set out to address this gap.

I begin by delineating the three effect types. Age can be divided between psychological ageing and life-cycle effects. Cohort effects capture long-lasting generation differences developed during the formative years of one's life. Period effects capture the mood of the moment - the transient shifts in ideology at that moment.

A clear problem for comparative research is generating comparable measures of left-right ideology. I therefore establish a theoretical distinction between relative and absolute ideology. Relative ideology embraces the transient nature of left-right ideology, allowing the various issue saliences and position of the overton window to vary. Absolute ideology instead rescales positions such that they become directly comparable to one another, with no variation in meaning.

I develop a measure of relative ideology by using Aldrich-McKelvey scaling on the Comparative Study of Electoral Systems (CSES) dataset. CSES represents an ideal choice primarily because it covers a large span of years and enables exactly this kind of measurement of ideology. The measure has individuals placed 'relative' to the party system in question. As an additional bonus, the methodology also corrects for differential item functioning in the responses. Since I do not have a priori reasons to prefer one model or the other, I run one model with cohort and period effects constrained to be the same for all countries and one model nesting them for all countries.

I find evidence for age, life-cycle, and cohort effects. There is a persistent shift to the relative right both as an individual ages and over the life-cycle. For those belonging to the generations born in the 1940s to the early 1960s, there is a clear left-wing cohort effect. By contrast, those belonging to generations born from the late 1960s to the early 1990s are more supportive of right-wing politics net of other relevant factors.

The fact that these results are reasonably similar between models that constrain cohort effects to be the same across countries is highlighted. By contrast, nesting period effects within country produces a variable set of effects quite different from the constrained model. This implies on the one hand that the socialising influences that shape cohort effects are common from country to country, and on the other hand that period effects typically represent more transient country-specific issues.

The interpretation of these results would be more interesting if there was a satisfactory measure of absolute ideology to compare against. Indeed, at this point we could learn to what extent results change when allowing versus when not allowing for changes in issue salience and in shifts in the center ground. However, no meaningful anchors current exist in the CSES - or any other long-term comparative dataset - that would allow for such an analysis to be performed. I therefore recommend that such a measure is developed via the addition of anchoring vignettes to comparative datasets. It will take decades before they are useful - but I argue that the long-term benefits are clear.

1.4.3 Essay Three

The third and final essay, *Social Democratic Party Positions on the EU: The Case of Brexit*, similarly sits in the realm of applied work in measuring political ideology. Here, my interest is in the UK Labour Party's struggle with its Brexit position as a case study of social democratic parties tackling second dimension issues.

A clear issue in attempting to make a judgement on the strategy of a political party is that we only witness one set of events in reality. We do not get to witness multiple re-runs of events to better understand how different strategies would have played out. Addressing the points made during the Labour Party's internal debate on its positions is then made difficult. I address this difficulty by performing a simulated counterfactual of the 2019 general election based on the Labour Party's EU position.

I begin by setting out the background of the election, highlighting that the primary issue was EU membership, and a secondary issue of redistribution underlined the election. I then survey the literature on spatial vote choice, highlighting in particular the recent addition of categorisation theory and the way in which the UK's electoral system may distort strategically optimal positions.

I then proceed to build a simulated counterfactual in several steps. First, I obtain corrected party positions on the two dimensions alongside voter positions on the same scales by using Bayesian Aldrich McKelvey scaling. From here, I use these positions to estimate a conditional logit model, which estimates vote choice as a function of party-level variables. I then simulate Labour

Party positions across the EU dimension, and predict vote choice based on the estimated parameters of the conditional logit model. Finally, I use Uniform National Swing and Regional National Swing to estimate seat shares for the parties.

I find that broadly, the evidence points towards the Labour Party having likely taken close to the best possible stance during the 2019 general election. This is clear in terms of vote share, but less clear in terms of seat share where the evidence is more mixed. Beyond these results confirming prior results in terms of social democratic party position, this paper also shows how we can measure and use ideology in a meaningful way for the purpose of making recommendations on party strategy. Once again, being alive to the issues inherent in measurement further facilitates gains.

Chapter 2

Agree to Agree: Correcting Acquiescence Bias in the Case of Fully Unbalanced Scales with Application to UK Measurements of Political Beliefs

2.1 Introduction

In political science, a substantive area of interest is the ideology of voters. It follows that a methodologically important area of research is the measurement of voter ideology. A popular approach to measuring voter ideology in surveys is the Likert scale. To construct a Likert scale, survey respondents are given several statements called Likert items[1] to respond to and a range of ordinal responses to choose from; typically ranging from 'strongly disagree' to 'strongly agree'. The responses to the statements are then additively aggregated to produce the Likert scale.

Three types of Likert scale exist. Balanced scales which are built from an equal number of indicators representing both sides of the ideological dimension, partially unbalanced scales which have more indicators from one of the sides of the ideological spectrum than the other, and fully unbalanced scales which have indicators from only one of the sides of the ideological spectrum.

The indicators which are used to build Likert scales are typically subject to acquiescence bias and so where unbalanced scales are used, so too do the final scales. If researchers utilising these scales fail to consider the survey design and response stage of the data generating process (DGP), this can lead to incorrect research conclusions being drawn. In practice, due to the time and expense involved in collecting survey data, most political scientists (and indeed social scientists more broadly) are reliant on 'off the shelf' survey data produced by other researchers. Where unbalanced Likert scales are used in survey design, this typically exacerbates the problem as many researchers are forced knowingly or unknowingly to use the biased scales.

While past research has dealt with modelling acquiescence in balanced scales (see Mirowsky and Ross, 1991; Billiet and McClendon, 2000; Savalei and Falk, 2014; Primi, Santos, De Fruyt and John, 2019; Primi, Hauck-Filho, Valentini, Santos and Falk, 2019), it has not dealt with the more difficult prob-

[1]some scholars may prefer to refer to the individual item as the Likert scale. For the purposes of this paper however, 'Likert scale' refers to the aggregate scale.

lem of fully unbalanced scales. All methods designed to correct acquiescence bias rely on contradiction in survey responses. Without contradiction, it is logically impossible to identify which respondents are expressing agreement due to acquiescence, and which respondents are expressing agreement due to genuine agreement. In fully unbalanced scales, this contradiction does not exist, rendering acquiescence typically impossible to identify. In this paper I therefore discuss the problem of fully unbalanced scales, propose a model-based solution predicated on the introduction of additional indicators, and conclude by providing recommendations to both survey designers and users. The model-based solution I propose is an adaptation of a model previously proposed for balanced and partially unbalanced scales, which I label the person-intercept confirmatory factor analysis (CFA) approach. This approach leverges the common factor model to capture the acquiescence component latent in survey responses.

I begin with a substantive discussion of acquiescence bias in terms of the common factor model. This model divides latent variation between content factors, measurement factors, and unique variation. I use this model to discuss the assumptions underlying Likert scales, and how unbalanced scales introduce acquiescence bias. I proceed with a demonstration of acquiescence bias between two comparable datasets in the form of the British Election Study and the British Social Attitudes survey. These results serve as a baseline for the correction methods I apply. I discuss past work on person intercept CFA, and develop four variations of the model. I discuss the need for empirical identification in the case of fully unbalanced Likert scales. I then apply these to a third dataset in the form of the 14th wave of the British Election Study internet panel. The results broadly show that the correction methods applied succeed in producing results more akin to fully balanced scales. I conclude with recommendations for researchers and survey designers.

2.2 Measuring Voter Ideology

Voter ideology, political beliefs, or political attitudes represent an inherently ambiguous concept. Exactly what it is, what label to give it, how many dimensions it's composed of, and which of those dimensions we should be interested in are all contested. Even once researchers have agreed on a set of answers for the purpose of a given research project, it follows that it is not straightforward how to capture a given definition among survey respondents.

One solution to this issue is the use of Likert scales. For the purposes of this paper, a Likert scale is an aggregate scale constructed from several individual survey items. For each survey item, respondents are shown a set of statements and given a range of responses, often five ranging from 'strongly disagree' to 'strongly agree'. Scores from these responses for each statement are then tallied to produce a final measurement of the concept of interest. The basic idea is to show the respondent many items intuitively related to the concept of interest, then to aggregate over the responses to reveal the respondents' overall preference on the underlying concept of interest.

Likert scales can be balanced, partially unbalanced, or fully unbalanced. Balanced scales are constructed from an equal number of survey items from each end of the dimension of interest (e.g. for a left-right dimension there will be left-wing and right-wing statements). Partially unbalanced items will have more items from one end than the other (but at least one from both), while fully unbalanced items will have items from only one end of the dimension of interest. However, since the individual agree-disagree items will result in acquiescent respondents being more likely to agree than they should be given their position on the underlying dimension, this results in acquiescence bias where scales are unbalanced.

In this section, I express this problem in terms of the common factor model.

2.2.1 The Common Factor Model

For a given set of indicators of the same concept of interest, we can adopt a generative understanding of the measure. This means the indicator is assumed to be 'generated' by the target concept, and variation in the target concept causes variation in the indicator. One method of expressing this is via the common factor model (Brown, 2015)

$$x_{ij} = \lambda_{j1}\eta_{i1} + \ldots + \lambda_{jm}\eta_{im} + \epsilon_{ij} \tag{2.1}$$

where x_{ij} is the jth observed indicator for respondent i, η_{im} is the mth latent factor for respondent i, λ_{jm} is the loading on the mth factor for indicator j, and ϵ_{ij} is the unique factor for the jth observed indicator for respondent i.

The latent factors underlie the observed measurements. They can be further split between content factors capturing substantive variation and measurement factors capturing variation due to the measurement method of choice (Kenny and Kashy, 1992). The unique factor captures variation in that indicator not found in any other indicators, which will be a mix of random noise and unique substantive variation. For my theoretical purposes in this paper, I adopt this common factor model.

2.2.2 Acquiescence Bias

Substantively, acquiescence bias can be described as a tendency to be more likely to 'agree' with survey statements regardless of their content. In terms of the common factor model, we can express acquiescence as a second common factor:

$$x_{ij} = \lambda_{jc}\eta_{ic} + \lambda_{ja}\eta_{ia} + \epsilon_{ij} \tag{2.2}$$

Here the c subscript denotes the concept of interest, while the a subscript denotes acquiescence. When aggregating indicators with this DGP into a Likert scale, we make several implicit assumptions. First, we assume that λ_{jc} is constant across indicators *except* that for some indicators its sign 'flips' depending on the direction of the statement. If for example we have a left-right factor, then we can imagine its sign being negative for left-wing indicators and positive for right-wing indicators. This assumption tends to be incorrect in practice (Billiet and Davidov, 2008, 545), but it need only be a reasonably close approximiation to be successful. Second, in the case of a balanced Likert scale we are assuming that λ_{ja} is also constant across indicators, albeit this time with its sign remaining positive regardless of the direction of the statement. Under these assumptions, acquiescence will 'cancel out' once the indicators are aggregated.

However, when the scale is unbalanced, acquiescence bias will shift the scale in the direction in which the scale is unbalanced. This is because where before the equal number of indicators in opposite directions 'cancelled' out the acquiescence in one another (see Cloud and Vaughan, 1970; Ray, 1979; Evans and Heath, 1995), in the unbalanced case there is leftover acquiescence. The more unbalanced the scale, the more bias leftover. This carries both descriptive and causal implications. Descriptively, this bias will shift the mean of the resultant scale in the direction of the imbalance. If we have a Likert scale comprised of more left-wing indicators than right-wing, then the resultant mean will be further to the left than it would be on a comparable balanced scale. Causally, on the same scale since acquiescence will point in the left-wing direction, variables that causally contribute to a respondent's level of acquiescence will appear to contribute to the scale. This can result in spurious causal associations (if the effect on the concept is 0), inflated causal associations (in the effect on the concept and acquiescence are in the same direction), or hidden causal associations (if the effect on the concept and acquiescence are in opposite directions).

2.2.3 Fully Unbalanced Likert Scales

A case not typically tackled in the literature on capturing acquiescence bias but which often arises in practice (both in political science and the broader social sciences) is that of a fully unbalanced Likert scale. Based on (2.2) and the subsequent discussion, it should become clear that in this case the acquiescence factor is impossible to *empirically* identify[2]. This is because at this point it

[2] I use this term to distinguish from statistical identification of the model. The model may be statistically identified (i.e. a unique solution exists), but that is no reason to believe we

becomes impossible to know which survey respondents are agreeing with the given statements because they sincerely agree with them; and which survey respondents are agreeing with them because they are acquiescent. Empirical identification of a model in this format requires *contradiction* in responses, which does not exist in a fully unbalanced scale.

It is arguably the case that this simple fact has led to some researchers mistakenly arguing that acquiescence bias is not a good explanation for the kind of results described above. For instance, Rodebaugh, Woods and Heimberg (2007) find that removing reverse-scored (i.e. opposite) items improves the psychometric performance of their model; and argue that in fact these items were introducing an additional factor. I do not dispute their second claim, but instead point to the above: that an absence of contradiction in responses means acquiescence bias will become difficult if not impossible to directly identify. That the psychometric performance of the model is better after removing the contradicting items should be unsurprising: acquiescence is well-known among other things to be associated with inflated reliability coefficients and correlations (Winkler, Kanouse and Ware, 1982; Evans and Heath, 1995)[3].

A separate but related line of argument argues that negatively worded items are responsible for the item misresponse, rather than acquiescence bias (Swain, Weathers and Niedrich, 2008). The argument is that negatively worded items introduce additional cognitive complexity (Swain, Weathers and Niedrich,

---

have successfully captured the acquiescence component

[3]Indeed, users of measures of psychometric performance should be careful in how their interpret such results. Cronbach's alpha for instance measures *internal consistency* - butas the common factor model shows consistency can be a function of both measurement variance and content variance!

2008). This point is not straightforwardly wrong: acquiescence bias is not the only potential problem that can occur in the survey design stage. The need for items on both sides of a scale does not mean that a simple negation route should be taken. However, as the demonstration below makes clear - acquiescence bias will still remain if a fully unbalanced scale is utilised. Researchers must therefore be prepared to tackle acquiescence bias within scales they are using.

In this paper, I tackle the specific problem of fully unbalanced Likert scales and offer some solutions for researchers utilising historical data. Given the constraint introduced by the problem of non-contradiction in fully unbalanced scales, these solutions are necessarily data dependent. They do nonetheless offer at minimum a starting point for researchers struggling with the problem of acquiescence in fully unbalanced likert scales.

2.3 Case Selection and Datasets

I use public opinion datasets from Great Britain as my case study. This is because since the 1990s, almost all GB datasets contain 'left-right' and 'libertarian-authoritarian' Likert scales based on the work of Evans and Heath (see Heath, Evans and Martin, 1994; Evans and Heath, 1995; Evans, Heath and Lalljee, 1996). Several operationalisations of the same core concept therefore exist, creating an opportunity to assess how variations in measurement produce variations in research results. I use three datasets - two for demonstration, and one for developing a correction.

For demonstration, I use the British Social Attitudes survey (BSA) (NatCen-

Social-Research, 2017) and the British Election Study (BES) face-to-face survey (Fieldhouse et al., 2017). Both surveys were collected in almost entirely the same time periods. Both surveys were collected in person (although the BSA also used self-completion questionnaires). Given this and the shared conceptual basis of their Likert scales, it is not unreasonable to expect reasonably similar distributions of attitudes in both surveys. Notably however, in the BSA, all items are left-wing or authoritarian. Insofar as acquiescence bias affects these scales, they should have a left-wing and authoritarian bias. The third dataset used is wave 14 of British Election Study internet panel (BESIP) (Fieldhouse et al., 2020), which is used as a cross-sectional dataset. This wave was collected in May 2018. Similar to the BSA, all items are worded in left-wing and authoritarian directions and it should therefore display a similar bias. The item wordings for all three datasets are as follows:

BSA Likert Scales

The statements utilised in the BSA left-right dimension (ranging from Disagree Strongly to Agree Strongly) are as follows:

- Government should redistribute income from the better off to those who are less well off
- Big business benefits owners at the expense of workers
- Ordinary working people do not get their fair share of the nation's wealth
- There is one law for the rich and one for the poor

- Management will always try to get the better of employees if it gets the chance

The statements utilised in the BSA libertarian-authoritarian dimension (ranging from Disagree Strongly to Agree Strongly) are as follows:

- Young people today don't have enough respect for traditional British values
- People who break the law should be given stiffer sentences
- For some crimes, the death penalty is the most appropriate sentence
- Schools should teach children to obey authority
- The law should always be obeyed, even if a particular law is wrong
- Censorship of films and magazines is necessary to uphold moral standards

BES Likert Scales

The statements utilised in the BES left-right dimension (ranging from Strongly Disagree to Strongly Agree) are as follows:

- Ordinary working people get their fair share of the nation's wealth (right)
- There is one law for the rich and one for the poor (left)
- There is no need for strong trade unions to protect employees' working conditions and wages (right)

- Private enterprise is the best way to solve Britain's economic problems (right)
- Major public services and industries ought to be in state ownership (left)
- It is the government's responsibility to provide a job for everyone who wants one (left)

The statements utilised in the BES libertarian-authoritarian dimension (ranging from Strongly Disagree to Strongly Agree) are as follows:

- Young people today don't have enough respect for traditional British values (auth)
- Censorship of films and magazines is necessary to uphold moral standards (auth)
- People should be allowed to organise public meetings to protest against the government (lib)
- People in Britain should be more tolerant of those who lead unconventional lives (lib)
- For some crimes, the death penalty is the most appropriate sentence (auth)
- People who break the law should be given stiffer sentences (auth)

BESIP Likert Scales

The BESIP survey items are shown with their labels from the dataset, which are used in some reporting during the appendix for this paper. The statements utilised in the BESP left-right dimension (ranging from Strongly Disagree to Strongly Agree) are as follows:

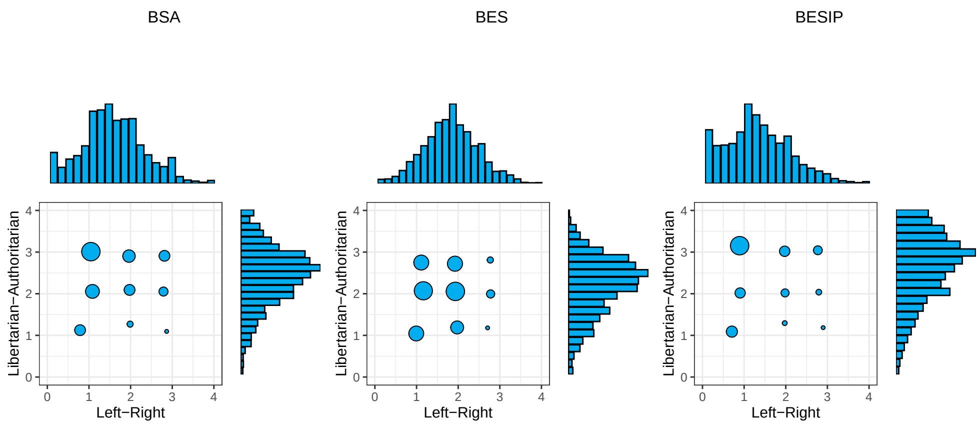
- **lr1:** Government should redistribute income from the better off to those who are less well off
- **lr2:** Big business takes advantage of ordinary people
- **lr3:** Ordinary working people do not get teir fair share of the nation's wealth
- **lr4:** There is one law for the rich and one for the poor
- **lr5:** Management will always try to get the better of employees if it gets the chance

The statements utilised in the BESP libertarian-authoritarian dimension (ranging from Strongly Disagree to Strongly Agree) are as follows:

- **al1:** Young people today don't have enough respect for traditional authority
- **al2:** For some crimes, the death penalty is the most appropriate sentence
- **al3:** Schools should teach children to obey authority
- **al4:** Censorship of films and magazines is necessary to uphold moral standards
- **al5:** People who break the law should be given stiffer sentences

Figure 2.1 presents both the joint and marginal distributions of the Likert scales from all three datasets. The BSA, BES, and BESIP scales are presented from left to right. The scales were constructed to range from 0 to 4. On each x-axis is the left-right scale and on each y-axis is the libertarian-authoritarian scale. The histograms opposite each axis capture the marginal distributions of these scales. To visualise the join distribution of the scales, respondents were divided into 'groups'. Those with scores ranging from 0 to 1.6 were placed in the 'left' and 'libertarian' groups of the respective dimensions. Correspondingly, those with scores ranging from 2.4 to 4 were placed in the 'right' and 'authoritarian' groups of the respective dimensions. Finally, those in-between these values were placed in the 'centre' group for each dimension. These groupings are of course arbitrary, but were chosen in part to resemble similar groupings utilised in research using these scales (see Surridge, 2018). The mean of each group was plotted, while the size of the group's dot corresponds to the number of respondents in that group. Survey weights were used for the graphs.

Figure 2.1: Joint Distribution of BSA and BES Scales



For each plot, the x-axis ranges from left (0) to right (4). The y-axis ranges from libertarian (0) to authoritarian (4). Respondents were grouped along each scale from left/libertarian (0 to 1.6 inclusive), centrist (1.6 to 2.4 exclusive), and right/authoritarian (2.4 to 4 inclusive). The position of the points correspond to the mean positon on both scales of the group, and the size of the point corresponds to the size of the group in the sample. The histograms opposite each axis show the marginal distribution of that particular scale.

Several notable differences emerge between the scales in figure 2.1. In line with the above predictions BSA and BESIP scales show clear left and authoritarian slants as compared to the BES. Indeed, the similarities between the BSA and BESIP plots are striking, both in terms of the marginal and joint distributions of the scales. This is doubly the case given that the BES and BSA surveys share both a time period and method of collection, while the BESIP survey was collected on-line and in a different time period. By contrast, the BES plot shows both less left-wing and less authoritarian respondents. It retains a left-wing slant, but this is driven by the absence of right-wing respondents - it is still more balanced towards the center of its scale relative to the BSA and BESIP plots. While the BES data therefore would offer firmer grounds for believing that the British electorate in 2017 was left-wing, some caution is still required. First, this may plausibly be a quirk of the sample in question. Second, it may be a function of the statements used to construct the Likert scale. Per (2.1), smaller loadings in the right-wing statements could produce a skewed result. Nonetheless, it is better evidence than that available in either of the other scales.

2.4 Demonstration

The impact of acquiescence bias on descriptive inference is straightforwardly demonstrated by the above graphs. However, its impact also extends to explanatory research. It is well-established that acquiescence has a strong negative relationship with education level (Ware Jr, 1978; Winkler, Kanouse and Ware, 1982), and so I take this as my example. Given the similar collection dates and conceptual overlap of the BSA and BES, I regress the scales contained within on a measure of education level. Since education also has some well-established results showing it has a negative relationship with authoritarianism (see Stubager, 2008; Surridge, 2016) but no well-established association with left-right attitudes, some predictions can be made. First, in the BES scales the results will be as described here. Second, in the BSA scale, a spurious positive association[4] between education level and left-right attitudes

[4]Since the scale ranges from left (negative) to right (positive) and acquiescence points in the negative direction

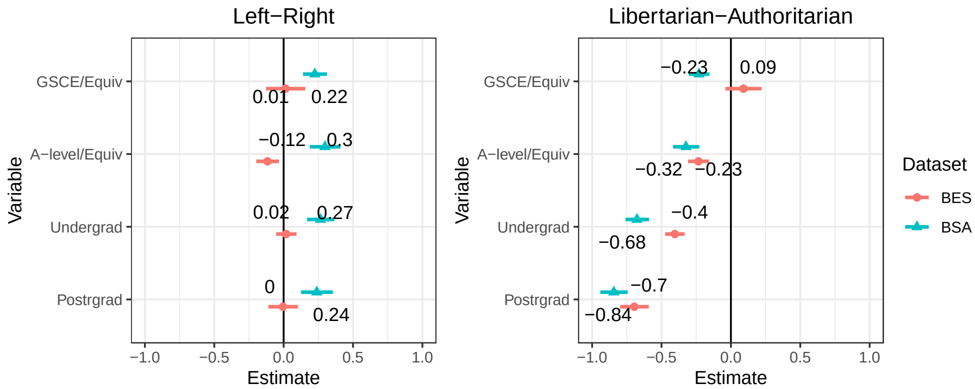
will be observed, while the negative association[5] between education level and libertarian-authoritarian attitudes will be stronger.

My interest here is not in offering some causal explanation of these scales, but rather to offer a clear example of results changing from unbalanced to balanced scales. I therefore do not include control variables as they do not add anything for the purposes of the demonstration. The education variables in both surveys were recoded such that the categories would match (full details of the recodes are in Appendix A). Most of the recodes will be uncontroversial and thus I do not discuss them further here, but the 'foreign' category in the BSA had to be treated as missing as it had no clear placement. Figure 2.2 shows the results of regressing the scales from the BSA and the BES on the recoded education variable. The left coefficient plot shows the results for the left-right scale, while the right coefficient plot shows the results for the libertarian-authoritarian scale. The reference category is possessing no education. 95% confidence intervals are included for each estimate. A full table of regression results is available in appendix B.

Figure 2.2 shows that the pattern of differences between the BSA and BES is as we'd expect given the expectations laid out above. First and least dramatically, the absolute size of the point estimates for the libertarian-authoritarian results are larger in the BSA results. Moreover, two of the confidence intervals for BES coefficients (GSCE/Equiv and Undergrads) have no overlap with those of the BSA, indicating that they are significantly different from one another at the 95% confidence level. By contrast, the decision to use the BSA

[5]Since the scale ranges from libertarian (negative) to authoritarian (positive) and acquiescence points in the positive direction

Figure 2.2: Coefficient Plot of Demonstration Regressions



Plots showing regression of BSA and BES likert scales on education level. No education is the reference category for all coefficients in the plot. Results are demonstrative only, so no further controls were included in the model.

or BES dataset carries profound consequences for the results a researcher will find. The parameters for the BSA are all significant at the 95% confidence level and positive. By contrast, the parameters for the BES are not significant at the 95% confidence level with the exception of the A-levels parameter, which is negative. Only the parameters for GSCEs have overlapping 95% confidence intervals. The point of these results is not to offer some causal interpretation, but rather to highlight how scale construction can be the primary driver of research results.

All of the results demonstrated in this section have rely on an assumption that there should be no predictable differences from the BES to the BSA other than those caused by acquiescence. Given the importance of this assumption to my analysis above, I have performed two robustness checks to verify that these differences are likely driven by acquiescence bias, rather than any particular quirks of the samples.

First, I merged the five indicators common to the BES and BSA into a single dataset and created a binary variable denoting whether a respondent belonged to the BSA. I then regressed this binary variable on the five common indicators, and ran an OLS, Logit, and Probit model to check against model dependency. In all three cases, two indicators were significant[6]. However, their point estimates pointed in opposite directions, strongly suggesting that these were quirks of the samples and not evidence of anything systematic that would explain the observed differences in results.

Next, to verify that there was no temporal instability in results, I regressed the scales from the BES on the survey month of the respondents. The result showed no association between interview month and scale score. I did not do the same for the BSA as the interview date is not included in the publicly available version of the dataset. Taken together, these two checks offer strong evidence that my assumption that systematic differences between these two surveys are primarily driven by acquiescence bias is correct. Regression tables for both of these checks are available in appendix B.

The results of this demonstration - both descriptive plots and differences in regression coefficients - are a striking example of how acquiescence bias can drive research results. This has frequently occurred in the UK context from which these example scales are drawn, with many researchers drawing

[6]For some crimes, the death penalty is the most appropriate sentence; People who break the law should be given stiffer sentences

on the fully imbalanced scales to argue that there is the UK electorate is on average left-wing and authoritarian in the UK (see e.g. Webb and Bale, 2021, p. 158). This argument has made it as far as the public, where researchers have passed on these findings via mainstream newspapers (see e.g. Surridge, 2019). There is therefore a clear need for all survey researchers to be careful with the measurements they use, and where possible to correct them for their biases.

2.5 Methodology

I now turn to the primary task of this paper, which is developing a methodology for the case of fully unbalanced Likert scales. Past research reviewing competing methodologies for modelling acquiescence bias have concluded that one of the most effective is an approach that treats acquiescence as a person-specific intercept across the scale items (Savalei and Falk, 2014; Primi, Santos, De Fruyt and John, 2019; Primi, Hauck-Filho, Valentini, Santos and Falk, 2019). This model was developed in a Confirmatory Factor Analysis (CFA)/Structural Equation Modelling (SEM) context (Mirowsky and Ross, 1991; Billiet and McClendon, 2000) but later extended to a unidimensional Item Response Theory (IRT) context (Primi, Santos, De Fruyt and John, 2019; Primi, Hauck-Filho, Valentini, Santos and Falk, 2019). For the sake of simplicity I focus on the CFA specification in this paper, but the general intuition translates to an IRT context. Here, I only briefly discuss the person intercept CFA model. A more complete background to CFA and its extension to include the person intercept is given in appendix C.

2.5.1 Person Intercept CFA

First used in Mirowsky and Ross' paper *Eliminating Defense and Agreement Bias from Measures of the Sense of Control: A 2 x 2 Index* (1991), the best exposition of the unit-intercept model is in Maydeu-Olivares and Coffman's paper *Random Intercept Item Factor Analysis* (2006). The model is based on (2.2) and traditionally is estimated by setting λ_{ja} to 1. This essentially treats acquiescence bias as a form of *differential person functioning*, where there is a constant difference *between* respondents but not *within* them as to how they respond to the survey items. This assumption is therefore kept from the balanced Likert scales, but the assumption that each item equally captures the concept of interest is relaxed. Although the assumption of equal loadings for the acquiescence component is a strong one, simulations do suggest that the model is robust to violations of this assumption (Savalei and Falk, 2014). Similarly, since the model is being estimated, the unique variation is also stripped from each item - a further relaxation relative to the balanced Likert scale. To identify the scales, the variance of η_{ic} is constrained to 1 while the variance of η_{ia} is freely estimated, producing the following model (Maydeu-Olivares and Coffman, 2006):

$$x_{ij} = \lambda_{jc}\eta_{ic} + 1\eta_{ia} + \epsilon_{ij} \tag{2.3}$$

For the purposes of this paper I label this version of person intercept CFA as CFA1. An alternative specification can be achieved by constraining the variances of both η_{ic} and the η_{ia} to 1, while freely estimating their loadings. However, a constraint is still placed on λ_{ja}, in that it must be equal across indicators. The linear form of this version of the model can thus be given as:

$$x_{ij} = \lambda_{jc}\eta_{ic} + \lambda_a\eta_{ia} + \epsilon_{ij} \tag{2.4}$$

For the purposes of this paper I label this version of person intercept CFA as CFA2. The full set of assumptions for both CFA in general and person intercept CFA are given in appendix C of this paper. However, one crucial difference in my definition of the model to Maydeu-Olivares and Coffman's is that I drop the language of 'random intercepts'. Here, they are drawing a parallel with hierarchical regression modelling in their description of the person intercept. However, the comparison is not necessary and more importantly undermines the utility of the model. In a random-intercepts regression model, the random intercepts are estimated as an error component. The unit-intercept here is not being estimated as an error term - it is being estimated as another common factor. The orthogonality assumption is thus not required for identification purposes (as other assumptions in the model are), but rather is made for the purpose of this comparison. This unnecessarily confuses things and potentially reduces the desirability of the model. In their review, Savalei and Falk (2014) suggest more work is required to explore potential relaxations of the orthogonality assumption. This assumption however is unnecessary to begin with, and I therefore drop it and utilise the terminology person-intercept instead of random-intercept.

In theory, the main difference between the specifications in (2.3) and (2.4) is their interpretability. Since the variances of both factors are the same in (2.4), the main advantage is that the model allows more direct comparison of the respective loadings - it is immediately clear how acquiescence bias compares to the content factors of interest in its effect on the scales. An advantage of the person intercept approach in general is that it does not require a balanced scale to work. Instead, the person intercept merely acts to capture inconsistency in observed responses and thus *in theory* only requires at least one opposite-worded indicator in order to successfully capture acquiescence bias. How many additional indicators are required *in practice* will need to be discovered in future research. I also consider ordinal versions of CFA1 and CFA2 in this paper, and I label them as OCFA1 and OCFA2 respectively. Their specifications are also detailed in appendix C.

2.5.2 Fully Unbalanced Scales

Past simulation studies suggest that unit-intercept models are robust to unbalanced scales where other acquiescence-correction methods require balanced scales (Savalei and Falk, 2014). However, the crucial point made above is that the unit-intercept requires *contradiction* in order to empirically identify the acquiescence component, which is lacking in fully unbalanced scales. If for instance we take the BSA left-right scale, it is impossible to try and tell apart those who are agreeing with left-wing statements because they agree with them and those who are agreeing with the same statements because they are acquiescent. *There is no information available to distinguish the two kinds of agreement.*

To solve this problem and empirically identify the acquiescence component, I use Watson's idea of introducing further information in the model

(1992). Specifically, if a scale which contains statements for which it would be contradictory to agree to all of them, it can be used to identify the acquiescence component in itself and thus also in the fully unbalanced scale. It is unfortunate that a strategy does not exist based on the fully unbalanced scale alone, but it should be clear that it is not possible to identify acquiescence bias in such a scale *without additional information being introduced in some form*. While the same simulations suggest that the person intercept CFA model is robust to differing levels of acquiescence bias in each indicator (Savalei and Falk, 2014), this approach necessarily strengthens the assumption, as it assumes not only the same level of acquiescence for each respondent on one scale, but on all scales in the model.

2.5.3 Identifying Scales in BESIP

The reason I chose the fourteenth wave of BESIP is that it contains two balanced Likert scales which could be used to identify the acquiescence component in the manner described above. This is the May 2018 wave of BESIP and thus some comparability to the other two surveys in this paper is lost. However, as seen in 2.1 there is nonetheless enough similarity in the scales in BESIP and the BSA for the dataset to be suitable for my purposes. The two additional scales cover zero-sum approaches to life and second on personal empathy respectively. The individual item wordings are as follows:

BESIP Extra Likert Scales

The statements on the zero-sum scale are:

- **zero1:** One person's loss is another person's gain (zero-sum)
- **zero4:** There's only so much to go around. Life is about how big a slice of the pie you can get. (zero-sum)
- **zero5:** Life isn't about winners and losers, everyone can do well (everyone can win)
- **zero7:** The only way to make someone better off is to make someone else worse off (zero-sum)
- **zero9:** There are ways to make everyone better off without anyone losing out (everyone can win)
- **zero11:** Everyone can be a winner at the same time (everyone can win)

The statements from the empathy scale are:

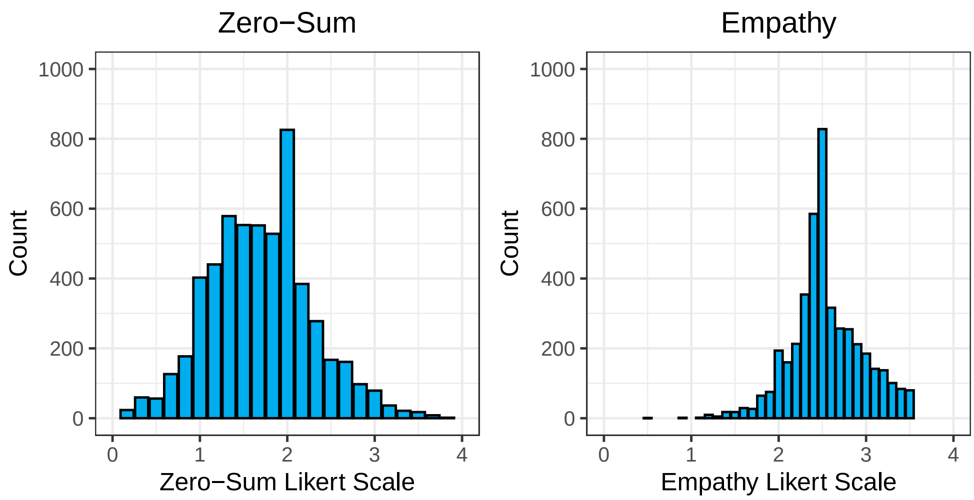
- **empathy1:** I can usually figure out when my friends are scared (empathetic)
- **empathy2:** I can usually realize quickly when a friend is angry (empathetic)
- **empathy3:** I can usually figure out when people are cheerful (empathetic)
- **empathy4:** I am not usually aware of my friends' feelings (unempathetic)
- **empathy5:** When someone is feeling 'down' I can usually understand how they feel (empathetic)

- **empathy6:** After being with a friend who is sad about something, I usually feel sad (empathetic)
- **empathy7:** My friends' unhappiness doesn't make me feel anything (unempathetic)
- **empathy8:** Other people's feelings don't bother me at all (unempathetic)
- **empathy9:** I don't become sad when I see other people crying (unempathetic)
- **empathy10:** My friends' emotions don't affect me much (unempathetic)

The two scales were asked in two separate subsamples of BESIP wave 14. This creates two separate opportunities to test the model, and so I test the four model types across the two BESIP subsamples. Figure 2.3 shows bar plots of the two balanced scales. The zero-sum scale ranges from everyone can win (0) to zero-sum (4), while the empathy scale runs from unempathetic (0) to empathetic (4). After filtering for missing data, the zero-sum subsample has 5836 respondents while the empathy subsample has 4478 respondents.

The empathy scale in figure 2.3 is notably less dispersed than the other scales discussed in this paper. This may be a function of the fact that it is comprised of a higher number of indicators than any of the others (10, as opposed to 5 or 6). It may also be however that given individuals are generally predisposed to view themselves as empathetic that there is less noise - and overall acquiescence - in the empathy scale. To test this second point, I ran

Figure 2.3: Bar Plots of the Balanced Scales



person intercept CFA models on each of these scales alone, the full results for which are available in appendix D. The estimated variance for the acquiescence component in the zero-sum was larger than in the empathy model, suggesting that there is less acquiescence in the empathy scale. The extent to which the corrections are successful likely depends in part on which of these scales is a closer match in terms of acquiescence to the acquiescence in the left-right and libertarian-authoritarian scales. I therefore test all four variations of the person intercept on both subsamples of BESIP wave 14, using the respective additional scales to empirically identify the model.

2.5.4 Estimation

To estimate the CFA1 and CFA2 models, I use robust maximum likelihood (MLR) estimation. MLR returns the same point estimates as ML estimation but adjusts standard errors and test statistics for violations of the normality assumption. A rough rule of thumb suggests it's a reasonable approximation once at least 5 response categories exist. To estimate the OCFA1 and OCFA2 models, I use unweighed least squares estimation (ULS). In a comparison between MLR and diagonally weighted least squares (DWLS) estimation found in favour in DWLS for ordinal data (Li, 2016). However, simulations comparing DWLS to ULS have in turn found in favour of ULS, with the caveat that DWLS may converge in situations where ULS does not (Forero, Maydeu-Olivares and Gallardo-Pujol, 2009). I therefore utilise ULS estimation. All CFA models in this paper were estimated using lavaan version 0.6-9 (Rosseel, 2012) using code adapted from the appendix of Savalei et al (2014). Since lavaan does not currently support survey weights for ordinal CFA models I have not used them in the CFA models themselves, but they were used in producing distributions from the predicted factor scores of the models. Since the associations between variables should be reasonably robust to weighting, this is likely unproblematic.

2.6 Results

In this section I present a series of results demonstrating the comparative performance of the methods. Since my emphasis as throughout the paper is on obtaining corrected measurements, the plots presented here pertain to the predicted factor scores. Tables containing results for the CFA models can be found in appendix D. To verify that the scales were broadly capturing the same content, their correlation matrices were checked. These tables are also available in appendix D. Correlation plots showing the correlations between the recovered measures, the likert scales, and the acquiescence factors are also available in appendix D.

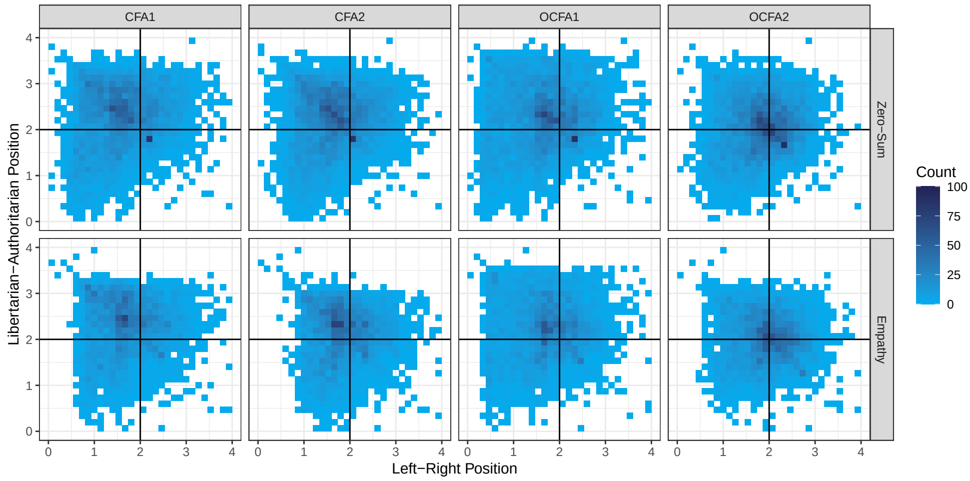
2.6.1 Distributions

Since my demonstration is preceded by distributional differences, I also begin my presentation of correction results with the distributions of the predicted factor scores. Figure 2.4 shows two-dimensional binplots of the resultant measurements from the four variations and the two subsets of the BESIP dataset. The extracted measures were rescaled to range from 0 to 4 to facilitate comparability with one another[7]. The 'left' factor extracted was flipped to range from left to right, rather than right to left. Libertarian-authoritarian factors are on the y-axis, and left-right factors are on the x-axis. The colour of the bins change from light blue to dark blue as the count of respondents in that bin increases. The plots are organised in columns for correction method, and by row for the two BESIP subsets. Plots of the marginal distributions of the predicted factor scores can be viewed in appendix D.

Alongside the marginal distributions available in appendix D, figure 2.4 shows that for all correction methods, relative to figure 2.1 there is a shift towards a more normal distribution. The scale is broadly more evenly distributed (especially in the case of OCFA2). There is a starker effect for the left-right scale, which in some cases appears to remain somewhat left-leaning. These differences are carried into the association between the two scales. The

[7] An identifying constraint on the scales is that they are mean 0. However, this won't necessarily be a meaningful midpoint, especially if the distributions are skewed. Rescaling in this way establishes the midrange point as the central point of each scale, which is no less arbitrary in theory but in practice may be a better approximation to a 'true' midpoint

Figure 2.4: Two-Dimensional Bin Plot of Voter Beliefs



Each plot uses a heatmap to show the joint distributions of the extracted factor scores. The left-right factor scores are on the x-axis, and the libertarian-authoritarian factor scores are on the y axis. The scores were rescaled to range from 0 to 4.

extent to which the joint distribution is even across the four quadrants varies from correction method to correction method, but in all cases similarly appears more evenly distributed than in figure 2.1. These results would therefore indicate that once acquiescence bias is accounted for, the distribution of voter ideology on both scales is closer to a normal distribution. Caution is required in interpreting the midpoint of these scales, but nonetheless the extract factor scores do appear more evenly distributed.

2.6.2 Acquiescence Factor

The question that remains after the initial positive assessment of figure 2.4 is whether the acquiescence factor has been properly captured by the person intercept. It may be the case that other substantive sources of variation are being captured. To verify the match between the acquiescence factor and the observed indicators, I additively aggregated all indicators, assigning a number for the level of agreement with each item (0 for 'strongly disagree' up to 4 for 'strongly agree'). I then assessed the correlations between this additive count and the acquiescence factor extracted from each model.

Table 2.1: Correlations of Acquiescence Factor with Additive Agreement

Subset	Model	Correlation
Zero-Sum	CFA1	0.71
Zero-Sum	CFA2	0.63
Zero-Sum	OCFA1	0.75
Zero-Sum	OCFA2	0.65
Empathy	CFA1	0.50
Empathy	CFA2	0.31
Empathy	OCFA1	0.69
Empathy	OCFA2	0.49

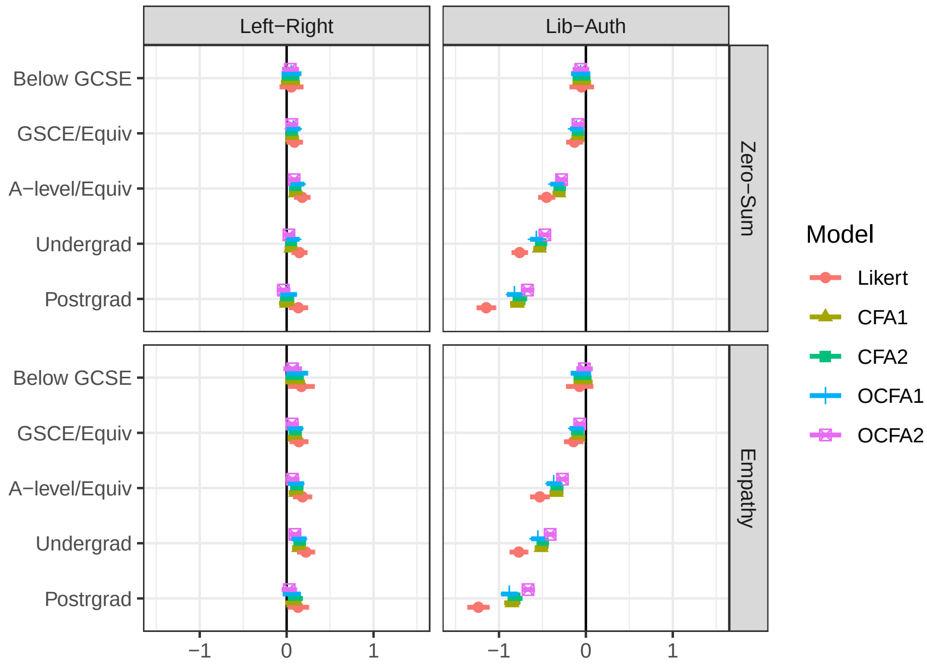
The correlations offer a mixed pattern. Broadly, those for the zero-sum subset of BESIP show reasonably sized correlations ranging from 0.63 to 0.75. By contrast, the correlations for the empathy subset sit in a smaller range of 0.39 to 0.5 except for the OCFA1 model, which achieved a correlation of 0.69. Overall, these medium-sized correlations suggest that the model does indeed capture acquiescence in the model. In line with the difference in acquiescence in the zero-sum and empathy scales, it would therefore appear that the Zero-Sum scale was a more effective identifying scale. It would not be a good thing if these correlations were near 1: there is also left-right and libertarian-authoritarian variation in the additive index, alongside noise generated by the zero-sum and empathy indicators (since aggregation was for agreement, these dimensions of variation will cancel each other in the index instead). We can therefore be increasingly confident regarding the efficacy of the model in properly capturing acquiescence.

2.6.3 Regression Results

With the distributional results of the correction methods established, I now turn to examining how explanatory research results are changed by using the predicted factor scores. I regressed the raw Likert scales and the predicted factor scores in each subsample on the education level variable available in the BESIP dataset. Once again, the reference category for the education level variable is ‘no qualifications’. 95% confidence intervals are included in the plot. Figure 2.5 shows coefficient plots for each of these regression results. Tables for each regression are available in appendix D. For the results displayed in the main body of the paper, I have avoided recoding the education level in the same manner as the demonstration above (i.e. to have the same levels) as this recoding makes some results appear to be somewhat better than they are. I have however included regression results with the recoded education level variable in appendix D.

In the case of the results for the zero-sum left-right scales, the confidence intervals for each model fully overlap in both datasets. However, the point estimates shift towards 0 once correction methods are used; and more importantly inferential differences emerge. If a correction method is used, it becomes the case that a researcher using null hypothesis significance testing will reach the same conclusions using the BESIP data as they would using the

Figure 2.5: Coefficient Plots of Scales Regressed on Education



Plots showing the results of a regression of extracted factor scores on education level in BESIP. For all four plots, the reference category is no education. Plots are organised in columns by scale, and by row for subset of BESIP. Results for likert scale included for comparison. Coefficient values available in appendix.

balanced scales in the BES. The correction is sharper in the zero-sum subsample, as predicted by the differences in acquiescence between the zero-sum and empathy scales. However, the OCFA2 model sufficiently shifts the point estimates in both subsamples that the same inferences will be produced in BESIP regardless of the subsample of choice. In terms of the libertarian-authoritarian scales, the gap in point estimates are considerably larger. In several cases, the 95% confidence intervals do not overlap at all. However, in line with the differences between the BES and BSA, a researcher using null hypothesis significance testing will reach the same inferential results - albeit with smaller point estimates for the corrected scales. The correction methods therefore broadly produce the same inferences as the balanced scales in the BES, while utilising biased data as in the BSA.

2.7 Conclusion

In this paper, I have set out the specific problem of fully unbalanced Likert scales in the context of wider work on acquiescence bias. Fully unbalanced Likert scales carry the particular problem of rendering the acquiescence within impossible to empirically identify without the introduction of additional information. In this paper, I have further clarified the different versions of person intercept CFA relative to Likert scales, relaxed the unnecessary orthogonality assumption, and developed a strategy for identifying a person intercept CFA model in the case of fully unbalanced Likert scales. The OCFA2 approach appears to work best for fully unbalanced scales, but it is not immediately clear why this should be the case. Researchers using these approaches should run all four and compare the results until further research can be conducted on the relative performance of the four methods.

A clear limitation of the correction methods used in this paper is the data requirements they impose on the user in the case of fully unbalanced scales. However, these limtis are necessary: acquiescence bias is not empirically identifiable without some degree of contradiction. Where researchers cannot utilise the corrections, they should at minimum be conscious of the role that acquiescence bias is likely to be playing in their results. Even where Likert scales are fully balanced, their use entails strong assumptions about the data generating process that can be relaxed by the use of person intercept CFA. There is no clear case where the use of these models if possible is not preferable to a raw Likert scale. For survey designers, two main points should be taken from this paper. First, as far as possible they should seek to design Likert scales that are fully balanced. Where this is not entirely possible, whether due to difficulties in designing reverse-keyed items or the need for backwards comparability, they should instead try to include other, substantively unrelated scales for the purposes of identifying person intercept CFA models.

Chapter 3

Age Isn't Just a Number: A Comparative Age-Period-Cohort Analysis of Political Beliefs in Europe

3.1 Introduction

A week is both a long and a short time in politics. It is a long time in the sense of events: unexpected occurrences can dramatically shift electoral realities in a matter of hours. However, politics is also slow: outside of day-to-day drama, patterns are slow to change in politics. Recent examples might include the rise of the radical right or the growing importance of educated voters in the electorate (see Ford and Jennings, 2020). These trends developed in the course of years, even if they are more visible in the drama of particular days.

It is not for nothing therefore that political scientists should be interested in the dynamics of long-term stability and change in political ideology and behaviour. In this paper, I am focussed on the former. One question that has continually perplexed political scientists is whether long-term trends and changes in political ideology are better explained by ageing effects, period effects, or cohort effects. The first describes processes of psychological ageing and predictable changes during the individual life-cycle. The second describes the effect of the ‘mood of the moment’: a transient effect which changes through time. The last describes the long-lasting effect of early influences that remain with a generation.

One reason APC analysis has presented greater difficulty than other statistical analyses is the difficulty involved in disentangling the three effect types. In their linear, continuous forms the three effects are perfectly multicollinear with one another. As a consequence, APC analysis has developed not merely as a form of substantive research but also has an entire sub-methodology associated with it. The primary feature of this sub-methodology has been a strong emphasis on functional form and modelling assumptions has emerged.

Extant APC analyses have thus far broadly found in favour of cohort effects on left-right positions. However, the question remains as to the extent that similar countries enjoy similar cohort and period trends in political ideology. Many of the longer-term events experienced in politics are experienced cross-nationally, from the social democratic moment in the post world war 2 era to today’s present political trends. It is therefore reasonable to wonder as to the extent that these trends are common across nations. Thus far, prior analyses have typically been confined to single country case studies and have made modelling and theoretical assumptions that mean their results do not generalise. In the one case where a comparative analysis has been performed, aggregate cohort effects were not reported, and two cohort effects were included in a single model. I therefore seek to assess in a comparative context the extent to which common cohort and period trends occur, and the extent to which APC analyses help us to understand long-term patterns of change and continuity in left-right ideology.

I begin with a general outline of APC analysis. I discuss the substantive interpretations of the three effect types, the identification problem that arises between them, and the results in past APC analyses of political ideology. I then turn to the issues of defining and measuring ideology in comparative longitudinal research. When making comparisons across contexts, the issue of differences in meaning arises. I discuss two approaches that arise from this: relative and absolute ideology. I further discuss survey measurement issues such as differential item functioning that commonly arise in survey contexts. From here, I proceed to outlining my methodological approach. I use Aldrich-McKelvey scaling to produce a DIF-corrected measure of left-right positions. From here, I use Hierarchical APC models to perform a comparative APC analysis. As part of this, a decision needs to be made regarding the treatment of country contexts. Should cohort and period effects be constrained to be similar across countries, or should they be nested within countries? Lacking a good a priori justification either way, I run both models and compare the results.

I find that there is good evidence for ageing, life-cycle, and cohort effects in terms of relative ideological positions. There are few differences in the inferences made between countries in terms of cohort effects regardless of the model specification. This implies that Western Europe does experience common socialising effects. This goes against the interpretation of past country case studies, which have interpreted cohort effects in terms of country-specific political influences. By contrast, period effects appear to differ much more substantially from country to country once freed to do so, implying that here a country-specific interpretation is more likely correct. I conclude by arguing that future APC and comparative research should focus on the development of absolute measures of ideology, so that the results presented here can be better understood.

3.2 Age-Period-Cohort Analysis

The purpose of age-period-cohort (APC) analysis is to distinguish between age, period, and cohort effects on a dependent variable of interest. In this section I discuss the theoretical distinctions between these effects and give an overview of research results regarding APC effects on ideology.

3.2.1 Theorising APC

'Age effects' broadly captures two theoretical perspectives on the role of ageing. In the first, the physical process of ageing is the causal variable of interest (Glenn, 1974). In political science, this will in typically be understood as psychological ageing. In the second, the individual's progression through stages of the life-cycle represents the causal variable of interest. Here, predictable changes over an individual's life-cycle such as marriage, increases in income, promotions at work, children, and home ownership have an effect on their political orientations (Glenn, 1974; Tilley, 2005; Tilley and Evans, 2014). In both cases, chronological ageing is an imperfect correlate of these processes and never the actual quantity of interest (Glenn, 1974).

'Cohort effects'[1] instead emphasise persistent generational differences. Within the social sciences, these effects are typically considered in terms of 'socialisation', where individuals are socialised into holding certain views as a consequence of influences from their formative years which are retained over time (see Mannheim, 1970; Dawson and Prewitt, 1968). Various potential sources of socialisation have been identified, including historical events (Mannheim, 1970), parental influence (Campbell et al., 1960; Butler and Stoke, 1974), education (Stubager, 2008; Surridge, 2016), and peer groups (Hooghe, 2004). Of primary interest here however is the role of the first: historical events. This is because this type of socialisation represents the closest correspondence to the notion of cohort effects - lasting formative influences unique to a given generation. Past research supports the notion that the long-term influence of historical events is strongest during the formative years between adolescence and young adulthood (Jennings, 2007, p. 35; Rekker, 2016, p. 121).

'Period effects' represent the effect of a given time period on the dependent variable of interest. Similarly to age and cohort effects, interest is not in the chronological time period itself but rather in the predominant features of that time period (Glenn, 2005). Unlike cohort effects however, its influence is taken to be temporary, rather than lasting. While interest is sometimes in age

[1] For clarity, 'cohort' and 'generation' can be considered interchangeable for the purpose of this paper

alone, while in other cases interest is in all three effects as potential explanations for long-term patterns of stability and change, ts inclusion is necessary in all cases. This is because of the fact that the three effects are potential mutual confounders of one another. This mutual confounding gives rise to the identification problem which characterises APC analysis.

In all three cases, the numbers associated with these effects - age, birth year, time period - are simply ways of capturing these effect types. But these numbers are not the source of our interest in themselves, and hence giving these effects careful interpretation during is an important task of APC analysis.

3.2.2 Identification

The APC identification problem emerges from the fact that in their continuous, linear forms

$$C = P - A \tag{3.1}$$

where A is age, P is the time period, and C is cohort membership (i.e. birth year). None of these effects are necessarily mutually exclusive, but in this format they possess perfect multicollinearity with each other. Due to this, a unique solution does not exist and thus the model cannot be estimated. All APC analyses must therefore tackle this identification problem with some set of assumptions regarding some or all of the three effects that allows for statistical identification of the model.

Over time, a variety of 'solutions' have been proposed. Broadly speaking, these typically require that some kind of assumption regarding the APC effects are required (Bell, 2020). The weakest assumptions typically focus on functional form, often assuming that period and/or cohort effects are non-linear in nature. The strongest typically require assuming that one of the three effects is 0 and thus can be ignored (Bell, 2020). A recent controversial methodology is hierarchical APC (HAPC) models, which estimates cohort and period effects as random effects. Although hotly debated (see full discussion below), this requires assuming non-linear effects in cohorts and period.

3.2.3 APC Analysis of Political Ideology

In political science, APC analysis has typically been used to assess long-term patterns of continuity and change in political ideology and behaviour. In this paper, I focus on the former. Broadly speaking, past research has tended to find in favour of the presence of cohort effects. How this should generalise across countries and contexts - if at all - is however not always consistent across studies.

In terms of single-country case studies, a large number of APC analyses on ideology have been performed in the United Kingdom. The earliest of these is an analysis by Tilley (2005) on ‘Libertarian-Authoritarian’ attitudes. Tilley assumes away a psychological process for aging and thus includes only life-cycle indicators rather than age in itself. Tilley finds that cohort effects and not life-cycle effects or cohort composition drive age differences in these attitudes. Newer generations are increasingly libertarian over time - in line with general political changes over time. Here, it is clear that broadly we should expect similar findings in other nations.

More recently, Grasso et al. (2019) similarly find in favour of cohort effects and against ageing effects. Grasso et al. use a generalised additive model (GAM) methodology, first developed by Grasso (2014). Unlike Tilley however, Grasso et al. find increasing right-wing and authoritarian attitudes among those coming of age during the Thatcher and Blair years. Similarly unlike Tilley who uses simple five-year groupings, Grasso et al. group survey respondents' cohort memberships according to political periods distinct to the United Kingdom. In a single-country case study however, it is impossible to distinguish between general versus country-specific cohort trends.

In a comparative context, Down and Wilson (2013, 2017) perform an APC analysis on support for the European Union. Across their two papers, they distinguish between utilitarian and affective attitudes to the European Union and find especially strong cohort effects for the latter, though they are present for both. Generally speaking, more recent generations are ceteris paribus more in favour of European Integration as they have been raised in a context where it was more established.

Shorrocks (2018) examines cohort differences in gender gaps in left-right ideology in Europe and Canada. Shorrocks finds cohort effects in this gap not captured by aggregate-level analysis: older women tend to be more right-wing than older men, while younger women tend to be more left-wing than younger men. However, Shorrocks's model combines two cohort trends: a linear cohort trend interacted with gender, and the cohort random effect more typical of HAPC methodology. The inclusion of a linear cohort trend is justified only on the grounds of creating the interaction, rather than through an assumption of linearity. Nor is it clear how the inclusion of two separate cohort trends should be theoretically understood. Moreover, Shorrocks only reports the changing gap between the genders: not the actual cohort or period effects.

In all of these papers, either comparative analysis is not performed, or where it is performed cohort effects are either constrained to be the same across countries or not but the effect of such a constrain (or its absence is never explored). In many of the single-country case studies, cohort specifications rest on theoretical justifications specific to that country, and so the results cannot be understood to generalise. I therefore aim to fill this gap, by performing an APC analysis of left-right ideology in a comparative setting. I aim not only to learn which the APC effects act as drivers of left-right ideology, but also the extent to which trends are common across countries within Western Europe.

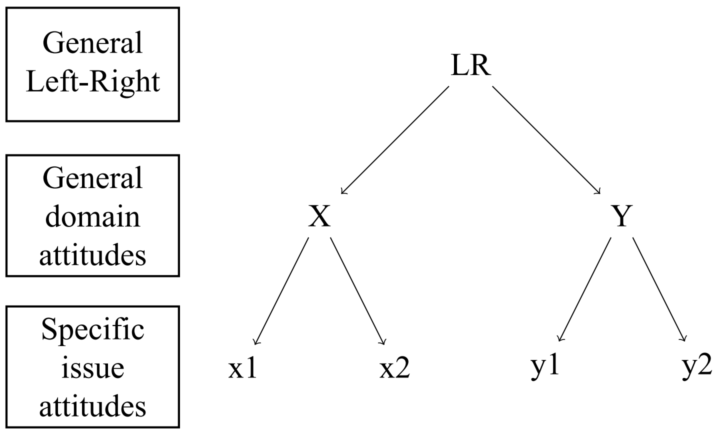
3.3 Political Ideology in a Comparative Setting

Political ideology is a difficult concept to utilise in a comparative setting. Ideology is an abstract concept, and the precise meaning of left and right varies between contexts. Since APC analysis always requires several different time periods (by definition), and the extension to comparative research requires the addition of multiple country contexts, this is a clear problem for APC analysis of political ideology. In this section, I therefore address the multiple issues that arise from attempting to quantitatively measure ideology in a comparative setting.

3.3.1 Defining Ideology

One approach to defining political ideology is to emphasise the role of *constraints*. The seminal work on this is Converse (1964), where ideology is defined as a '*configuration of ideas and attitudes in which the elements are bound together by some form of constraint or functional interdependence*'. If we know one opinion an individual holds, we are better placed to guess another opinion they hold. Peffley and Hurwitz (1985) expands this model of constraints to a hierarchical model of public opinion, wherein an overall left-right dimension drives domain attitudes which in turn drive specific issue attitudes. This is visualised in figure 3.1 below:

Figure 3.1: Hierarchical Model of Public Opinion



Analogous to the hierarchical model of public opinion is the basic space theory of political ideology. Here, also following on from Converse's model, where individuals have structured belief systems they can be represented in a low-dimension space (Poole, 1998). This low-dimensional space was referred to as the basic space by Ordeshook (1976) and as a predictive dimension by Hinich and colleagues (Hinich and Pollard, 1981). The primary difference between the hierarchical model and the basic space theory is that where the hierarchical model assumes causal pathways from more abstract dimensions to more concrete dimensions, the basic space theory is causally agnostic and treats the abstract dimension as a representation of several smaller ones.

There does exist some debate as to the exact extent this holds true. Using an ordinal probit model with random effects, Lauderdale, Hanretty and Vivyan (2018) find that just 1/7th of variation in survey responses corresponds to a summary dimension as described above. Another 3/7ths corresponds to idiosyncratic variation, while the last 3/7ths corresponds to response instability (Lauderdale, Hanretty and Vivyan, 2018). By contrast, using a novel mixture model methodology (Fowler et al., 2022) find that approximately 70% of US citizens are in fact best described by a single ideological dimension, while another 30% are either idiosyncratic[2] or simply random in their responses.

Both studies were conducted in the US, and focus on slightly different quantities - the percentage of variation in responses versus how respondents are best described. Nonetheless, these different methodologies provide vastly different answers as to how well a single dimension does describe variation in ideology. The important point however, is that to some extent or other we can usefully speak of a left-right summary dimension. I do not require that the causal pathways of the hierarchical model hold: only that ideology can be usefully described in in the higher dimensionality of left and right in the manner described above. There are merits in conducting (comparative) APC analyses on other dimensions as have indeed been done - but my focus here is on the left-right dimension.

[2]i.e. consistent in their views, but not in the traditional left-right manner

3.3.2 Relative and Absolute Ideology

In APC analysis and many other comparative analyses, survey respondents from several time periods are pooled together. In a comparative APC analysis, they are pooled from both several time periods and several countries. Given the above conceptualisation of left-right ideology as a summary dimension defined by constraints above, the question emerges how to approach this comparatively. In a given context, the relevant issues, and therefore issue domains will differ. Similarly, the weights on those issues and domains up to the left-right dimension will also differ. The EU is much more important to left-right ideology in the UK in 2019 than in say, 1997. Likewise, French voters who see Marine Le Pen as being to the right of Emmanuel Macron are not doing so on the grounds of their respective economic positions.

I therefore introduce the concepts of relative and absolute ideology. With 'absolute' ideology, the meaning and interpretation of measures of a left-right dimension should be constant over time and space. Following the above discussion, raw survey data does not typically meet this criteria. It will therefore be necessary to rescale the data such that all data points share an interpretation. One method of doing this is via anchoring vignettes (King et al., 2004; King and Wand, 2007; Hopkins and King, 2010). Here, vignettes are provided and scaled by the respondents. Since the information in the vignettes are constant over time, this information becomes the 'anchor' against which all other data can be rescaled. Other options for rescaling the data may be discovered in the future: the important point is that there is some piece of information that is constant over time that raw data can be rescaled against. Likewise, the method of choice for producing this rescaling is immaterial, and many such methods exist (see e.g. Wand, King and Lau, 2011; Bakker, Edwards, Jolly, Polk, Rovny and Steenbergen, 2014; Bakker, Jolly, Polk and Poole, 2014).

By contrast, when measuring ideology in the 'relative' sense, each context is allowed to retain its own meaning and interpretation of the left-right dimension. This should ideally however be relative to a reference point. We may for example seek to use data measured relative to the political center of that context. One of my contentions in this paper is that raw survey data is in fact relative in nature, but implicitly rather than explicitly so. In this paper, I use this type of comparative ideology - largely due to data constraints. I develop my measure of relative ideology by rescaling data with respect to standardised party distributions (not too distant from the idea of scaling against the political center of a given context). I discuss this further in the methodology section.

The conceptual discussion here raises a question: will absolute versus relative ideology produce different results? It seems obvious that it should be so: if the interpretation of a variable changes over time, then surely the nature of the relationship between other variables and itself should also change over time. In APC analysis this seems especially pronounced: are people becoming more right-wing as they age, or is the political system shifting 'left' around them? Without access to measures of both, we cannot answer this question. We can however still gain some interesting results from a single measure, but we must be careful in how we interpret those results. I now turn to a final measurement concern for this paper: differential item functioning.

3.3.3 Differential Item Functioning

Closely related to the problem of item meaning is the issue of scale perception. Differential item functioning (DIF) is a measurement issue that arises when for the same underlying perception, different survey respondents give different answers (King et al., 2004). Even where two respondents come from the same context, and share the same underlying meaning of left and right, and share the same underlying ideological position, DIF means that they will place themselves on the survey scale in different locations to one another. Insofar DIF is purely random, this will result in attenuation bias. Insofar as there are systematic patterns in DIF, this will result in biased results. And indeed, more politically informed respondents may for instance use a survey scale in a different manner to less informed respondents.

Along with survey item meaning, researchers seeking to study ideology must also contend with the problem of DIF. Solutions typically focus on finding some objective external *anchor* on which to rescale responses (King et al., 2004). Several methods for solving DIF exist, but the broad concept is the same. If respondents are asked to locate one or more external stimuli on the same scale, these placements can be used to reveal both the 'true' location of the external stimuli and through this a corrected measure of the respondent's location can be produced. One popular approach to this is the use of anchoring vignettes (King et al., 2004; King and Wand, 2007; Hopkins and King, 2010). Another approach is to use real-world stimuli, such as political parties or elites (Aldrich and McKelvey, 1977; Poole, 1998; Hare et al., 2015). It is this latter approach that I take in this paper.

3.4 Methodology

3.4.1 Case Selection and Data

To perform a comparative APC analysis of left-right ideology, I utilise the Comparative Study of Electoral Systems (CSES) integrated module dataset. This is a dataset collected alongside the election studies of participant countries, meaning that survey responses are always from the context of the heat of an electoral campaign. I chose this dataset for three reasons. First, it spans a good number of years: from 1996 to 2016. This is essential for APC analysis, which requires a large time span in order for analysis to be effective. Secondly, it contains a self-reported measure of left-right positions which can be corrected for differential item functioning. Third and finally, as I discuss below, this correction process can be used to produce a clear measure of relative ideology.

I select Western Europe as a case study because it is a solid testing ground for the notion that cohort and period effects may be similar between countries sharing common political trends. Not only does Western Europe broadly enjoy this reality, it also represents a set of country cases that should be reasonably similar in terms of the relationship between age, generation, and political ideology. Introducing post-soviet countries could potentially introduce very different dynamics in terms of the relationship between age and political ideology. After filtering for required survey questions, the 15 countries included for analysis are Austria, Belgium, Denmark, Finland, France, Germany, Great Britain, Iceland, Ireland, Netherlands, Norway, Portugal, Spain, Sweden, and Switzerland. Any West European countries not included in the analysis are either not included in the CSES, or did not contain the requisite survey questions.

3.4.2 Measuring Left-Right Ideology

As discussed above, when measuring ideology in a comparative context the issue of survey item meaning emerges. I approach both this problem and the problem of DIF simultaneously by utilising Aldrich-McKelvey (AM) scaling (Aldrich and McKelvey, 1977). Aldrich-McKelvey scaling is a methodology developed to correct differential item functioning in respondent placements of political parties. As part of the scaling process, respondent-specific parameters are recovered which can in turn be used to generate DIF-corrected measures for survey respondents on the same scale. In other words, the corrected political party placements are used as external anchors, respondent-specific parameters are generated by regressing these on the respondent's own placements, these parameters are then applied to respondents' self placements.

DIF can thus be corrected by running Aldrich-McKelvey scaling within each country-year subsample. The question then remains how this might become a true measure of relative ideology. When measuring political party positions, it is common practice to standardise them due to the absence of a natural 0 point (see e.g. Hanretty, 2022). Since Aldrich-McKelvey scaling returns political party placements with a mean 0 distribution by construction (Aldrich and McKelvey, 1977), all that remains is to divide party positions by their standard deviation. Since the respondents' recovered placements are on the same scale, respondents' positions can similarly be divided by the party standard deviation. The interpretation of respondents' positions is then their placements relative to their country's standardised party system in that year.

In practice, I do not expect strong differences between models based on the raw survey data and the rescaled relative data. This is in part because DIF is largely treated as a noisy process in Aldrich-McKelvey scaling, and so should not particularly alter the measure other than to remove some noise. It is also because the raw survey data is in itself a form of relative data. Both methods therefore capture a contextual ideology and should not particularly differ in the results they provide. To test the notion that both the raw data and the scaling data provide results relative to that given context, I therefore present models with both the raw data and the rescaled data from the Aldrich-McKelvey procedure outlined above. In the case where the raw data is used, I have standardised it across the entire dataset to be mean 0 and standard deviation 1. This transformation does not affect the relationships between variables, but will make the recovered parameters more similar in size to the rescaled data and thus easier to compare.

For the purposes of Aldrich-McKelvey scaling, the respondents' left-right placements of political parties in CSES were used. Only parties where at least 40% of respondents had placed the party were utilised for this purpose. This was a fairly arbitrary choice. The threshold was chosen to be large enough to remove parties that very few respondents placed, while remaining small enough to avoid removing too many political parties from the scaling procedure. This did however necessitate further filtering of respondents for missing data, as Aldrich-McKelvey scaling requires that respondents place all political parties used.

3.4.3 The Hierarchical Age-Period-Cohort (HAPC) Model

The most recent - and controversial - development in APC research is the hierarchical age-period-cohort (HAPC) model (Yang and Land, 2006, 2008, 2013). In HAPC models, the cohort and period effects are assumed to be non-linear and are modelled as random effects. This is given by (3.2):

$$\begin{aligned} Y_{ijk} = \boldsymbol{X}_{ijk}\boldsymbol{\beta} + cohort_j + period_k + u_{ijk} \\ cohort_j \sim N(0, \sigma_{cohort}) \\ period_k \sim N(0, \sigma_{period}) \\ u_{ijk} \sim N(0, \sigma_u) \end{aligned} \tag{3.2}$$

where Y_{ijk} is the outcome of interest, $\boldsymbol{X}_{ijk}$ is a vector of covariates for individual i, $\boldsymbol{\beta}$ is the vector of fixed effects, $cohort_j$ is the cohort random effects, $period_k$ is the period random effects, and u_{ijk} is the error term in the model. Note that variation occurs across three levels: individuals, cohort, and time period.

An implication of this is that cohort and period membership are treated as contexts within which individual survey respondents are nested (Yang and Land, 2006, p. 85). Strictly speaking, the model is always identified as the three effects are not linear and additive at the same level of analysis (Yang and Land, 2013). However, there has has been substantial controversy as to whether the HAPC model correctly identifies APC effects. In the first critique, HAPC models allow researchers to estimate a model without properly stating their assumptions around the non-linearity of period and cohort effects (Bell, 2020). Where the model's assumption of non-linear effects are correct, it will be correct. However, a model with linear period and cohort effects will still be estimated.

This leads to the second, older, and more serious critique of the model. Here, the concern is that HAPC models misallocate APC effects despite the apparent breaking of linear dependency (Bell and Jones, 2013, 2014*a*,*b*, 2015). The debate that followed generated more heat than light, but broadly some points of consensus do emerge. First, agreement is found on the treatment of cohort and period effects as random effects. Second, in the presence of exact algebraic linear effects the HAPC model will fail (Reither, Masters, Yang, Powers, Zheng and Land, 2015; Bell and Jones, 2015; Reither, Land, Jeon, Powers, Masters, Zheng, Hardy, Keyes, Fu, Hanson et al., 2015). The point on which consensus does not exist is the exact conditions under which the model would enter difficulty beyond exact linear dependency. The critics of HAPC models argue that it is enough that period and cohort effects are monotonic (i.e. always increasing or decreasing, rather than merely linear) for collinearity to occur (Bell and Jones, 2014*a*,*b*, 2015). Moreover, it is argued that HAPC models typically find in favour of period effects because there are typically many more years (and thus groupings) covered by the cohort effects (Bell and Jones, 2018).

I utilise the HAPC model because there are good a priori reasons to expect that cohort and period effects are unlikely to be linear (or even monotonic) in practice. The tide of history does sweep forever in one direction, but ebbs and flows in unpredictable ways. If the post world war 2 era was characterised by greater social democracy, it was also characterised by a level of social authoritarianism that today's center right by contrast would not accept. Likewise, even as during the 1970s the neoliberal turn begun, this was concurrent with a shift in more socially liberal directions. It would be surprising to witness linear cohort and period trends both: it is for this reason I favour the HAPC specification. Some questions however remain. First, how should cohort membership and periods be created? Second, how should the inclusion of multiple countries be incorporated into the HACP model?

There are broadly two competing views on the specification of cohort memberships and time periods. The first suggests we should theoretically specify the cohorts and time periods based on a priori knowledge. The alternative perspective, and the one I endorse, takes the view that however cohorts and periods are defined there will necessarily be arbitrariness at the boundaries at the boundaries between cohorts and periods (Spitzer, 1973, 1355). Hence, many researchers specify cohorts and periods in arbitrary five-year groups. Given the well-acknowledged arbitrariness, I therefore adopt this approach.

3.4.4 Comparative APC

The extension of APC analysis to a comparative setting requires the country contexts survey respondents are nested in to be considered. An advantage of utilising the HAPC methodology is that countries can be treated as another random effect in the model. This is demonstrated in (3.3) below:

$$\begin{aligned} Y_{ijkl} &= \boldsymbol{X}_{ijkl}\boldsymbol{\beta} + cohort_j + period_k + country_l + u_{ijkl} \\ cohort_j &\sim N(0, \sigma_{cohort}) \\ period_k &\sim N(0, \sigma_{period}) \\ country_l &\sim N(0, \sigma_{country}) \\ u_{ijk} &\sim N(0, \sigma_u) \end{aligned} \tag{3.3}$$

where the primary addition is the country random effect $country_l$ and the index denoting that level of variation, l.

However, the question of the relationship between country random effects and the cohort and period random effects arises. In the model above, cohort and period effects do not vary by country and thus are constrained to be the same across countries. Given that cohort membership and time periods are treated as contexts within which individuals are nested, it is not immediately clear whether these should be nested within countries or not. In other words, are cohort and period effects unique to the countries in question? I therefore run both the constrained model in (3.3) and the nested model below:

$$\begin{aligned} Y_{ijkl} &= \boldsymbol{X}_{ijkl}\boldsymbol{\beta} + cohort_{jl} + period_{kl} + country_l + u_{ijkl} \\ cohort_{jl} &\sim N(0, \sigma_{cohort}) \\ period_{kl} &\sim N(0, \sigma_{period}) \\ country_l &\sim N(0, \sigma_{country}) \\ u_{ijk} &\sim N(0, \sigma_u) \end{aligned} \tag{3.4}$$

Note that the only difference in (3.4) relative to (3.3) is that the cohort and period effects are now free to vary by country.

3.4.5 Gender Generation

Although not the focal point of this paper, Shorrocks (2018) establishes the presence of a gender-generation gap in political ideology that varies on cohort lines. However, as discussed above Shorrocks includes an additional linear cohort term in the fixed portion of the HAPC model and interacts it with gender. Instead, I take an approach more in line with the assumptions and theoretical motivations for specifying a HAPC model and utilise cohort random slopes on gender. This allows the effect of gender to vary from cohort to cohort, without requiring that it vary linearly.

3.4.6 Life-Cycles

The final set of modelling decisions regards additional variables to include alongside age. Over an individual's life-cycle, many important changes can occur: university, increases in income, children, marriage, home ownership. I include marital status, income, and university education in the models below as important life-event variables that correlate with age. I do not include children or home ownership as these are not measured in the CSES.

Of these, 'university education' is least straightforwardly interpretable as a life-cycle effect. Higher education is not pursued by everyone, but has been disproportionately pursued by the present younger generation (Ford and Jennings, 2020). Its effect is therefore more a correlate of cohort membership than of age.

3.5 Results

Table 3.1 presents the results of the HAPC regression models. The first two models on the left-hand side are the set of models with shared cohort and period effects. The next two models on the right-hand side are those with nested cohort and period effects. Within these groups, the first left-hand model is that with the raw response data as the dependent variable, while the second on the right-hand side is that with the rescaled response data following the Aldrich-McKelvey procedure outlined above. The reference categories are single for marital status, no education for education level, and the 1st (i.e. lowest) income quintile for income level. Results are reported at 3 decimal points to avoid rounding some effects to 0. 95% confidence intervals are reported alongside the parameter estimates, and a star is used to denote when the null hypothesis of 0 falls outside this interval.

The variances and covariances of the random effects are reported in the tables. Plots of predicted random effects for cohort and period random intercepts are presented throughout the main analysis. Since the gender-generation gap is not a focus of my analysis, I present the plots of the cohort random slopes for gender in the appendix of this paper. Likewise, since I am not substantively focussing on the random intercepts of the various countries in the analysis, these are also presented in the appendix of this paper.

Starting with age effects, in all four models the coefficient for age is positive and significant at the 95% confidence level. Also positive and significant

Table 3.1: HAPC Results

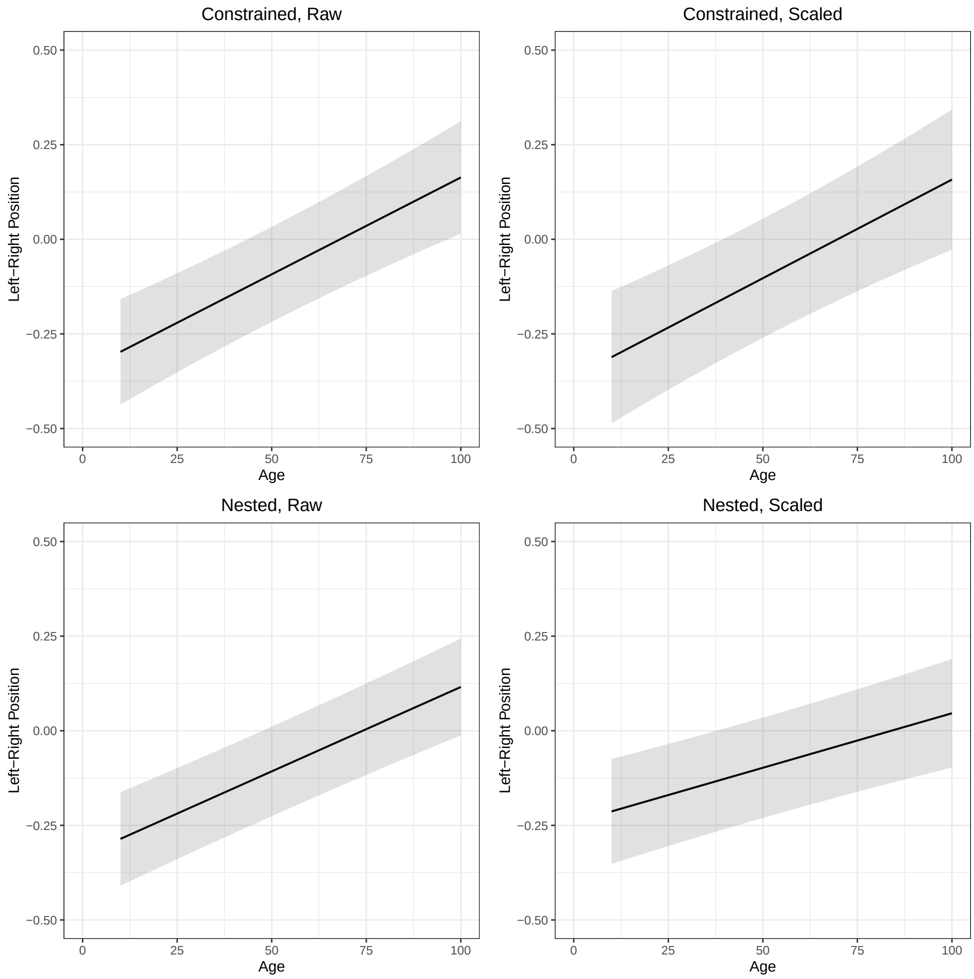
	Constrained		Nested	
	Raw	Scaled	Raw	Scaled
Age	0.005^*	0.005^*	0.004^*	0.003^*
	$[0.004; 0.007]$	$[0.003; 0.007]$	$[0.004; 0.005]$	$[0.002; 0.004]$
Married	0.064^*	0.092^*	0.062^*	0.091^*
	$[0.040; 0.089]$	$[0.066; 0.119]$	$[0.038; 0.087]$	$[0.064; 0.117]$
Divorced/Separated	0.012	0.042^*	0.001	0.029
	$[-0.026; 0.049]$	$[0.002; 0.083]$	$[-0.036; 0.038]$	$[-0.011; 0.069]$
Widowed	0.092^*	0.115^*	0.102^*	0.120^*
	$[0.047; 0.136]$	$[0.067; 0.164]$	$[0.057; 0.147]$	$[0.072; 0.169]$
Primary Education	-0.034	-0.075^*	-0.021	-0.056
	$[-0.099; 0.031]$	$[-0.145; -0.004]$	$[-0.088; 0.045]$	$[-0.127; 0.015]$
Secondary Education	-0.065	-0.164^*	-0.047	-0.124^*
	$[-0.131; 0.000]$	$[-0.235; -0.093]$	$[-0.114; 0.020]$	$[-0.197; -0.051]$
Post-Secondary Education	-0.008	-0.079^*	-0.003	-0.050
	$[-0.075; 0.059]$	$[-0.152; -0.006]$	$[-0.072; 0.065]$	$[-0.124; 0.024]$
University Education	-0.178^*	-0.328^*	-0.172^*	-0.291^*
	$[-0.245; -0.112]$	$[-0.400; -0.255]$	$[-0.240; -0.104]$	$[-0.365; -0.218]$
Other Education	-0.382	-0.247	-0.345	-0.047
	$[-0.916; 0.151]$	$[-0.825; 0.331]$	$[-0.877; 0.188]$	$[-0.623; 0.528]$
2nd Income Quintile	0.022	0.051^*	0.025	0.049^*
	$[-0.006; 0.049]$	$[0.021; 0.080]$	$[-0.003; 0.052]$	$[0.019; 0.078]$
3rd Income Quintile	0.056^*	0.071^*	0.056^*	0.061^*
	$[0.028; 0.084]$	$[0.041; 0.101]$	$[0.028; 0.085]$	$[0.031; 0.092]$
4th Income Quintile	0.096^*	0.107^*	0.097^*	0.095^*
	$[0.066; 0.125]$	$[0.075; 0.139]$	$[0.068; 0.127]$	$[0.064; 0.127]$
5th Income Quintile	0.260^*	0.283^*	0.260^*	0.274^*
	$[0.229; 0.290]$	$[0.250; 0.316]$	$[0.229; 0.290]$	$[0.241; 0.307]$
Var: Cohort	0.007	0.009		
Var: Country:Cohort			0.015	0.016
Var: Cohort (Gender)	0.010	0.008		
Var: Country:Cohort (Gender)			0.023	0.027
Cov: Cohort	-0.003	0.002		
Cov: Country:Cohort			-0.008	-0.006
Var: Period	0.000	0.002		
Var: Country:Period			0.005	0.029
Var: Country	0.036	0.060	0.033	0.035
Var: Residual	0.945	1.112	0.934	1.087
N	55833	55833	55833	55833
AIC	155569.601	164664.656	155335.450	163881.254
BIC	155757.133	164852.189	155522.982	164068.787
Log Likelihood	-77763.800	-82311.328	-77646.725	-81919.627

* Null hypothesis value outside the confidence interval.

at the same level across all models, though relatively small, are the coefficients for those who are married and for those who are widowed relative to those who are single. There is therefore some initial evidence here the notions that there are both ageing and life-cycle effects in terms of relative ideology. In other words - despite potential changes in the political system around them, individuals still move to the (relative) right as they age.

Similarly, across all four models there are positive and significant effects for education and income level. These can less straightforwardly be considered as life-cycle effects, but there are also at present large generational differences - especially in terms of education level. Notably, the effects for university education and the top two quintiles are reasonably large relative to the other effects. Also worth acknowledgement is the fact that unlike other life-cycle effects, education moves individuals to the relative political left rather than right. Overall then, there are strong life-cycle effects in terms of relative ideology. We should not however immediately neglect the coefficient for age on the grounds that it is small: humans enjoy long lifespans, and the shift to the right predicted here will happen over a lifetime. Figure 3.2 plots the predicted values on the relative left-right scale as someone ages, with other variables set to their mean values.

Figure 3.2: Predicted Left-Right Position by Age



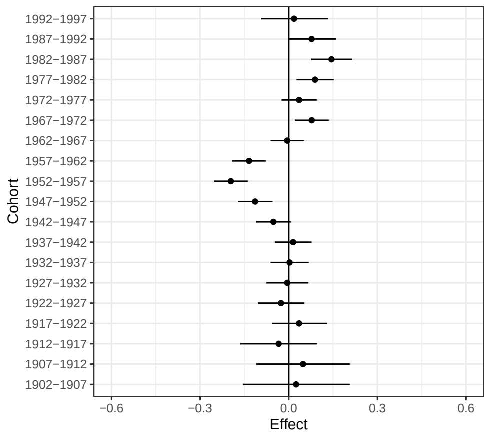
This plot shows the predicted relative left-right position an indivudal will hold as they age. All other variables are set to their means in the sample. The left plot displays this for the constrained model, while the right plot displays this for the nested model.

Once visualised across the span of a human life-time, it becomes clear that there are indeed substantial ageing effects present in terms of relative ideology. Although the year-on-year difference is small, the gradual accumulation over a lifetime results in a shift on average from left to right. Relative to a standardised party distribution, this will be approximate to a notion of moving from center-left to center-right as the average positions here are within the -1 and 1 standard deviations. Of course, it is possible that what is occurring is that the political contexts individuals find themselves in shift left as they age. It will require the creation of an asbolute measure of ideology to better understand the relationship between age an ideology shown here.

Pointed out but not explicitly analysed thus far is the fact that the results from both the raw data and the scaled data from the Aldrich-McKelvey procedure produce the same inferences. This offers evidence that the raw data can indeed be considered as a form of relative data. This does not mean that the exercise in rescaling the data was pointless: instead, it has demonstrated the contention of this paper that we do not know how APC results would look with absolute measures of the data. Moreover, it is *transparently* relative. Instead of providing such results without discussion, the rescaling ensures that it is clear to the readers how they should interpret the results. Since the results continue to be the same between the raw and rescaled data throughout the rest of the analysis, for the goal of concise presentation plots of the random effects from the raw model are presented in the appendix.

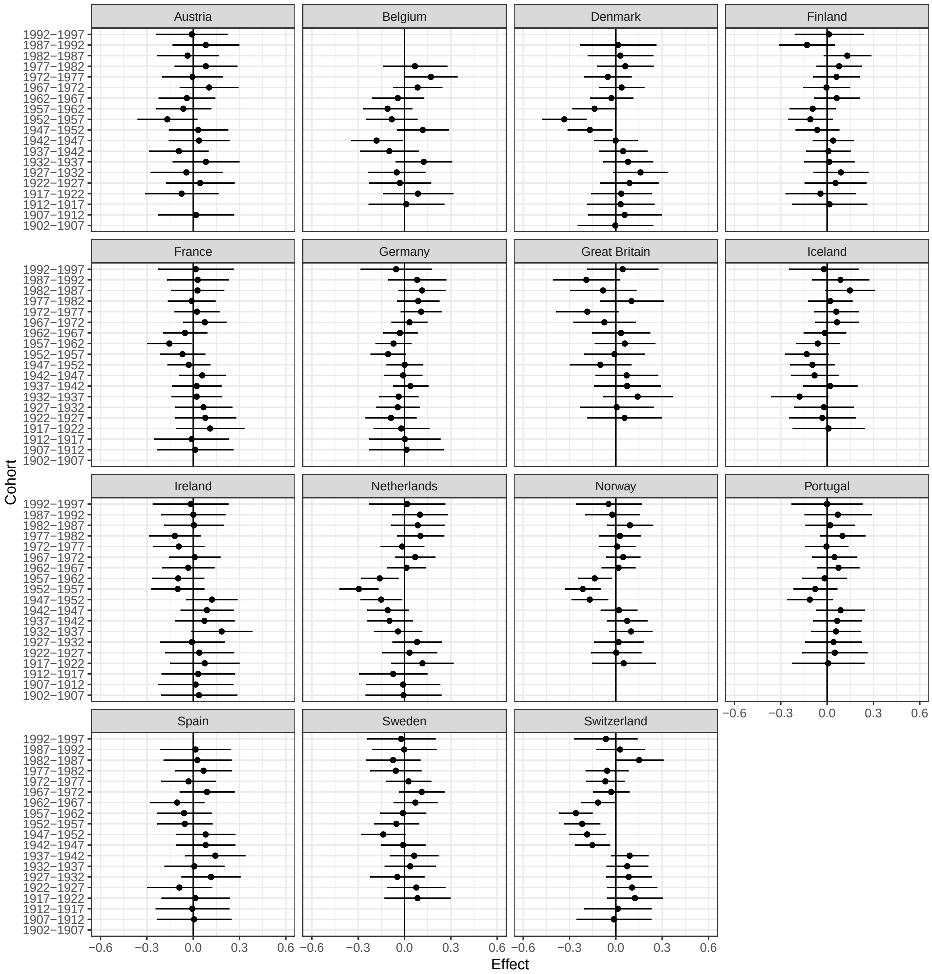
What then about cohort and period effects? I begin by examining cohort effects for the constrained model with the scaled data. The predicted cohort effects are plotted in figure 3.3:

Figure 3.3: Predicted Constrained Scaled Cohort Effects



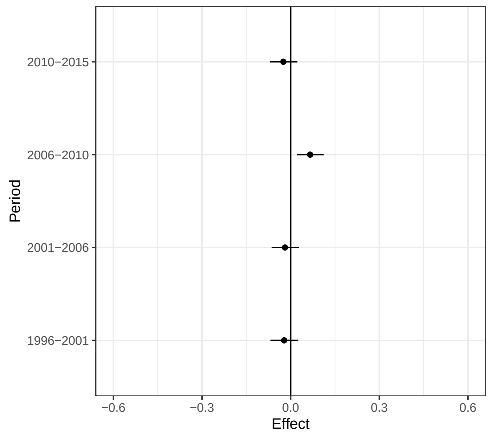
Across the generations, a reasonably clear pattern emerges in figure 3.3. For the earliest generations, there is a clear null effect. However, for the generations born in the 1940s to the early 1960s, there is a clear left-wing cohort effect. Relative to the political system of their day, this model suggests that individuals belonging to these generations are more supportive of left-wing politics net of other relevant factors. By contrast, the generations born in the late 1960s to the early 1990s are more supportive of right-wing politics net of other relevant factors. This could be interpreted as the respective effects of the post-war period of social democracy, followed by the neoliberal turn from the late 1960s onwards. What is fascinating about these effects is that given the relative interpretation of ideology being used, these effects hold even as the nature of left-right ideology changes around them. The question then emerges: how, if it all, does this model change in the nested model where cohort effects are free to vary country by country? Figure 3.4 plots the predicted cohort effects from the nested model with the scaled data:

Figure 3.4: Predicted Nested Scaled Cohort Effects



The trends in figure 3.4 are somewhat mixed. Where individual countries contain statistically significant cohort effects, they follow the same pattern as in figure 3.3. In other countries the observed cohort effects are not statistically distinguishable from 0. The constraint of similar cohorts across countries is therefore not a particularly restrictive one - at least insofar as relative ideology is concerned.. Although the magnitude of cohort effects differ from country to country, the pattern here is clear enough that it brings into question country-specific theories of cohort effects, such as that in Grasso et al. (2019). The final set of effects left to analyse is that of period effects. Figure 3.5 plots the period effects from the constrained model for the scaled data:

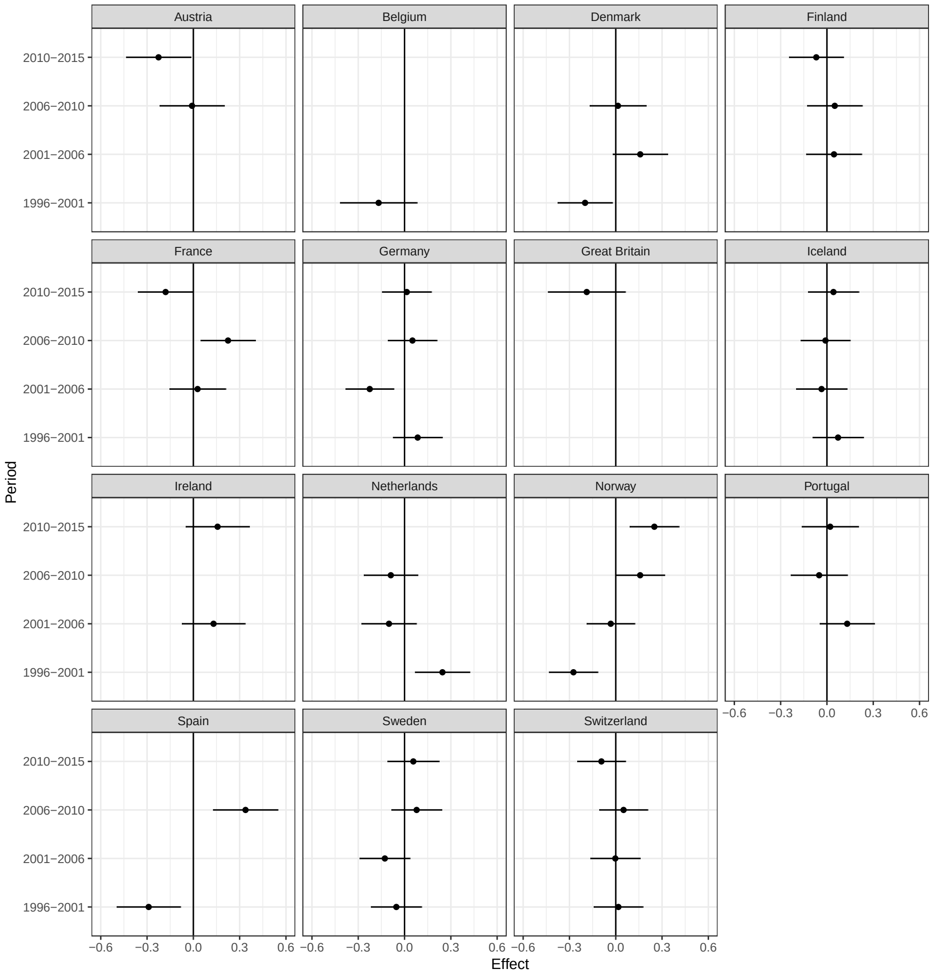
Figure 3.5: Predicted Constrained Scaled Period Effects



With the exception of the 2006-2010 period, the effects here are all null effects. Similarly, the effect for 2006-2010 is small, though perhaps should not be entirely ignored. Overall however, the result here is a fairly simple one: there is little difference from time period to time period in relative left-right ideology - at least for the period in which these surveys were conducted. The evidence here therefore suggests that there is little effect from 'the mood of the moment' on ideological positions. This may of course be driven by the fact that the survey period relative to the wider range of cohorts in the CSES is fairly short. It may be that period effects do exert an effect, but this would only be observable with a longer survey time period. The obvious question that follows is does this result hold if we nest the period effects in countries?

Figure 3.6 therefore plots the predicted period effects from the nested model:

Figure 3.6: Predicted Nested Period Effects



As compared to figure 3.5, the results presented in figure 3.6 offer a more mixed set of results. As in figure 3.5 Belgium, Finland, Britain, Ireland, Iceland, Portugal, Sweden, and Switzerland show null results and thus no difference from time period to time period in the average positioning of the electorate relative to the party system. However, many countries do show interesting trends. Spain and Norway both show a shift from left to right over the years in which they are surveyed. Many countries show their own period trends, such as Germany, France, Austria, and Denmark. These results are therefore in line with a theory of period effects which emphasises the role of political conditions specific to that country. Interpreting an individual set of country results will therefore require country-specific expertise, which is beyond the scope of this paper. Attempting to constrain period effects to remain the same across countries is therefore a much stronger assumption than it is for cohort effects.

3.5.1 Robustness

A key argument in my model specification is that the boundaries between cohorts and periods are arbitrary and thus cohort and periods can also be arbitrarily specified into 5-year periods. There is nothing special per se about 5 year time periods: it is simply a common convention in the APC literature. It is perhaps advantageous because it places enough respondents within a group to be useful, but not so many as to begin losing too much detail. This assumption is easily testable however. I therefore re-ran the analysis with 7-year cohorts and time periods. The plots showing the predicted random intercepts for these are presented in the appendix, and they broadly corroborate the patterns found in the analysis here. The results are therefore robust to the way in which cohorts and periods are specified, as we should theoretically expect.

3.6 Conclusion

In summary, the results presented here show clear evidence in favour of ageing effects, life-cycle effects, and cohort effects in terms of relative ideology. In other words, even as the political system changes around them in terms of the constellation of issue saliences and the 'center' position of the time, people do on average still shift in the direction the political right as they age. They also however retain an initial socialising influence, which is notable given the relative interpretation of the measure here. This would suggest a persistent political influence that remains robust to changes in the precise make-up of political ideology. This in itself is fascinating: it suggests that the socialising effect remains robust to later changes in the nature of political ideology in that country.

There are few differences between cohort effects from the constrained to the nested models: either they are null results or they show the same pattern. This in itself is an important insight: it implies that past theories conceptualising cohort effects in terms of country-specific political socialisation are flawed. Instead, countries in West Europe appear to follow common political trends in terms of socialising influences and cohort effects - at least insofar as relative ideology is concerned. By contrast, the same dynamic is not observed in terms of period effects: where the constrained results are null, in the nested model the countries with significant effects exhibit fairly different patterns to one another. It is probable that were we to compare more different parts of the world: for example Western Europe to Eastern Europe, we would expect the notion of common cohorts to also break down.

This analysis therefore shows that at least within Western Europe, constraining cohorts effects to be the same across countries is not a strong constraint. This implies that, within Western Europe, the early political influences that remain with a generation are not specific from country to country. This may in part explain why we witness common waves of events such as the summer of 1968, or today's rise of the radical right. Identifying exactly what these common influences are should be a field of future investigation.

By contrast, period effects show a more mixed pattern of results once allowed to be free from country to country. This is in line with a transient interpretation of these effects: they are likely driven by local political events, politicians, and elections, insofar as they are present. The fact that there are reasonably common cohort effects but not period effects is in itself interesting, and worthy of further investigation in future research.

Due to the unavailability of such a measure, it is a shame that the results for relative ideology here could not be compared against results using a measure of absolute ideology. Indeed, insofar relative ideology captures how people change or stay in their ideological position as the nature of political ideology changes over time, it would be interesting to compare this against results showing how they change or stay given a single, constant, context-independent measure. Indeed, this would inform us to the extent that the results observed here in terms of ageing effects are driven by changes in the political system rather than actual movement in a right-wing direction per se. Future APC - and indeed comparative research more broadly - should focus on this as a matter of urgency.

Chapter 4

Social Democratic Party Positions on the EU: The Case of Brexit

4.1 Introduction

In recent years, traditional economic issues have reduced in salience as non-economic issues have increasingly emerged as the new issues of political contestation (Ford and Jennings, 2020). The strategic dilemma that emerges for established political parties is how to adapt to the new dimensions. This has been an issue in particular for social democratic parties, which have struggled to understand how best to respond to the emergence of second dimension issues (see Abou-Chadi and Wagner, 2019, 2020). I take as my case the UK, where a second dimension issue in the form of EU membership was the primary issue in the recent 2019 general election. In this election the Labour Party suffered a heavy electoral defeat. The interim between elections, the election itself, and the Labour party's defeat generated a debate on the party's position that remains unresolved.

Within the party, those favouring a more Remain-leaning stance have highlighted the Remain sympathies of the Labour Party's core voters and the existential threat losing these voters would pose. Those favouring a more centrist or Leave-leaning stance have emphasised the fact that the distribution of voters across Labour's safest seats versus more marginal seats meant there was considerably more benefit to winning over pro-Brexit voters. Self-evidentially, the difficulty involved in reaching a conclusion on optimal strategy is a consequence of the fact that we witness only one set of events in reality. In this paper, I therefore set out to simulate a narrow counterfactual of the 2019 general election, estimating Labour Party vote and seat share as its position on Brexit changes. In doing so, I aim to construct the best-available evidence to answer the question as to what the best strategy for the Labour Party with regards to its Brexit position would have been.

Historically, the core voter base for social democratic parties was the working class. However, the emergence of new dimensions of contestation presented an electoral challenge for social democratic parties (Kitschelt et al., 1994). It was initially supposed that the subsequent electoral struggles of social democratic parties was an inability to win over the 'losers' of globalisation (Kriesi et al., 2006, 2008, 2012). With the rise of globalisation has come a rise in the salience of new dimensions. In particular for European politics, the growth and enlargement of the EU has brought issues of nation and immigration to the fore (Hooghe, Marks and Wilson, 2002; Hooghe and Marks, 2009, 2018). However, in recent years the core electorate of social democratic parties has changed. The modern center-left increasingly depends on the votes of the highly educated (Gingrich and Häusermann, 2015), while the raw proportion of votes represented by the blue collar working class has declined (Kitschelt and Rehm, 2014). Moreover, these two segments of the electorate hold very different stances on the new dimensions: the working class favour socially conservative positions on immigration and the EU while the educated typically take a more cosmopolitan stance (Kitschelt and Rehm, 2014; Hakhverdian et al., 2013).

The rise of new dimensions of contestation thus present a clear strategic dilemma for social democratic parties. Initially it was widely supposed that social democratic parties would need to assume authoritarian and anti-EU positions to win over working class voters. However, recent research finds that that broadly social democratic parties do not improve their position by taking anti-EU stances (Abou-Chadi and Wagner, 2020). It has further been found that they perform best by adopting culturally liberal stances alongside investment-oriented economic stances (Abou-Chadi and Wagner, 2019). The debate is however not yet settled. In the case of the radical left, recent research has found that far-left parties benefit from taking anti-EU positions (Wagner, 2021).

Furthermore, the role that electoral systems play needs to be taken into account and doubly so with the context of the United Kingdom. Past research has shown that single member plurality voting systems typically distort results in favour of right-wing parties (Döring and Manow, 2017). Similarly, evidence shows that in the past proportional representation has been associated with moderate parties adopting more pro-redistribution positions (Paulsen, 2022). These strategic aspects of single member plurality systems may therefore alter the strategic incentives for social democratic parties - as Labour's Brexit supporters suggested. All of this gives further impetus to the UK Labour Party as an interesting case study in how social democratic parties handle second dimension issues. On top of this, the specific context of the 2019 general election is valuable as in the UK non-economic issues such as the EU and immigration are very closely bundled together.

To answer this question, I use data from the 17th wave of the British Election Study internet panel (BESIP) and draw on a simulation approach popular in the spatial tradition of political science research. In this approach, a conditional logit model is run regressing vote choice on voter-party distances on the ideological dimensions of interest. Once the model has been estimated, new data are simulated by changing the party's position on a given variable and re-calculating party-voter distances. The simulated data are then turned into predicted vote shares by using the parameters estimated in the earlier stage. This allows for simulation of a narrow counterfactual of party vote shares as the party's position changes (Adams and Merrill III, 1999, 2000; Adams, Merrill III and Grofman, 2005). I call this a narrow counterfactual because I am only considering the effect of changes in the Labour Party's Brexit position on its vote and seat shares and not on anything else. To estimate seat share from the simulated counterfactual, I utilise both uniform national swing and regional national swing to generate predictions. By doing so, I am able to directly assess the particular strategic claims on both sides of the debate.

I proceed with this paper in four steps. First, I briefly outline the context of the 2019 UK general election, the context in which it occurred, and why the debate remains unresolved. Secondly, I discuss the spatial model of vote choice, and introduce the complications of ideological multidimensionality and the UK's party system. Third, I outline my methodological approach. I discuss the use of Aldrich-McKelvey scaling to deal with the twin problems of differential item functioning and placing survey respondents and parties on the same scale. I then proceed to discuss the conditional logit plus simulation approach popular in the spatial tradition of vote choice. I introduce the use of cross-validation methods from the world of machine learning to assess the predictive capabilities of the model. Finally, I discuss generation of seat predictions via uniform national swing and uniform regional swing. Fourth and finally, I present the results of the simulation.

I find that broadly, the evidence of the simulated counterfactual points towards a strategic need for the Labour Party to be a party of Remain. This is in line with past research on how social democratic parties should handle second dimension issues. Points near Labour's 'true' position maximise its vote share and minimise the vote share gap between the Conservatives and the Labour Party. Similarly, a range of points before the mid-point on the scale maximise the Party's seat share. However, some ambiguity does remain in the results in minimising the seat gap between the Labour Party and the Conservatives. This result joins results in past research papers that show once other components of the voting decision have been accounted for, parties can rarely drastically alter their overall results in terms of spatial positioning. The simulation also goes some way to confirming Downs' intuitions regarding the role of voter distributions in shaping optimal positioning, even after a large number of complications over and above Downs' model are introduced.

4.2 Background

With the dramatic pro-Leave result the 2016 EU referendum profoundly reshaped British politics for several years. Although not necessarily a major issue in electoral debates, Brexit nonetheless represented a clear watershed moment in a broader process of electoral realignment in the politics of the UK. In the 2017 general election, the Labour Party won a large percentage of remain voters while the Conservative party won a large percentage of leave voters (Hobolt, 2018). At the time the election was heralded as the return of a two-party dominated politics due to the large combined vote share of both parties after a long period of increasing third party strength in UK politics. The Conservative Party had taken a hardline pro-Brexit stance, while the Labour Party had opted for a more ambiguous and moderate stance rather than appealing directly to its pro-Remain base. Although many argued that the Labour Party's larger gain in vote share and the Conservatives' loss of their majority was a vindication of this strategy, this arguably ignored the fact that the Labour Party gained ground in pro-remain areas and lost ground in pro-leave areas (see discussion in Sobolewska and Ford, 2020, chapter 10).

By the summer of 2019 the picture had drastically changed for both parties. Several rounds of failed attempts at passing May's version of Brexit or to reach any kind of compromise in parliament had deeply damaged the Conservatives' image among Leave voters. The Labour Party meanwhile had lost much of the trust of Remain voters, with Jeremy Corbyn's personal popularity plummeting from its peak in 2017 - plausibly in part due to a lack of a firmly pro-Remain stance on his part. This culminated in the 2019 EU parliament elections which Britain originally should not have participated in. In these elections the Brexit party came first, while the Liberal Democrats came second and the Greens fourth but not too distant from the Labour Party. The election was a disaster for both main parties, with Remain voters abandoning the Labour Party for the Liberal Democrats and the Greens and Leave voters abandoning the Conservative Party for the Brexit Party.

In response to these results, the two parties took different approaches. The Conservatives replaced Theresa May with Boris Johnson, who as the most prominent backer of Vote Leave had clear pro-Brexit credentials. The Labour Party by contrast was less drastic in the changes it pursued, adopting a more clearly pro-second referendum stance - albeit one where Jeremy Corbyn would not take a stance during the second referendum. Like May, Johnson failed to get his Brexit deal past the parliamentary deal past the arithmetic of the hung parliament, so instead called a general election to pass his deal. During the 2019 General Election, the primary issue at stake was Brexit, but economic issues remained important. As before the election, substantial debate during and after the election surrounded the Labour Party's pro second-referendum stance.

Those who favoured a more explicitly pro-Remain stance argued that the Labour Party had faced an existential threat by the summer of 2019 and whatever potential benefits that ambiguity may have had in 2017 no longer existed. For this side of the debate, failure to adopt a clearly pro-Remain stance represented at best a worse strategic position for the party and at worst an existential threat. By contrast, those who favoured a more centrist or pro-Leave position argued that the electoral geography of the UK meant the party needed to win over Leave voters in key constituencies and that it could afford to lose Remain voters in safe seats. Both stances carry a reasonable degree of plausibility to them, although both make many implicit assumptions around the weight of the party's Brexit position and the distribution of voters across the UK's electoral geography. To this day, the debate on the party's stance and the role in its defeat in the 2019 UK general election remains unresolved.

4.3 Theory

In this section I discuss the spatial model of vote choice which I use to build the counterfactual simulation around which this paper is centred. I begin by discussing the spatial model of vote choice in its simplest, two-party and one-dimensional form before introducing additional complications.

4.3.1 Spatial Theory

Spatial models of vote choice are a formalisation of a simple intuition, which is that voters prefer to vote for (and see elected) the political party 'closest' to their own views. In these models, ideological viewpoints are arranged along a numerical dimension (e.g. left-right) and parties and voters are placed along this dimension (see Downs, 1957a,b). To make their voting decision, voters then make a utility calculation, which can be expressed in a general form as

$$U_{ij} \propto \|X_i - P_j\| \tag{4.1}$$

where U_{ij} is the utility vote i receives from party j winning, X_i is voter i's position on the ideological dimension, and P_j is party j's position on the ideological dimension. The function || || is the utility loss function, which shapes the effect of distances between voter i and party j to voter i's utility. A typical choice is the absolute-value norm

$$\| \; \| = | \; | \tag{4.2}$$

although competing choices for || || include the squared distance (or quadratic loss) and a gaussian loss function (Armstrong et al., 2020). For my purposes in this paper, I am using the absolute-value norm because past research has suggested this is the better fit for modelling voter loss functions (see Merrill III, 1995).

An important early contribution to the spatial vote choice literature was a rejection of the median voter theorem by Downs. Early work on spatial vote choice had suggested that where two parties competed on the ideological dimension, as rational actors they would converge to the position of the median voter (Hotelling, 1929; Black, 1948). However, Downs rejected the median voter theorem by introducing the possibility of non-voting. In Downs' model, if both parties are far in distance from a given voter, then they have less reason to vote and thus will abstain (Downs, 1957*b*, 142). It follows from this that the best position for the two political parties is conditional on the distribution of voters along the ideological spectrum. If normally distributed, the parties will converge to the median voter. However, if bi-modally distributed, the parties will move away from each other and towards the two poles (Downs, 1957*a*,*b*).

The result of the median voter theorem therefore does not hold. It follows that our expectations for the optimal Brexit position for the Labour party will be conditional on the distribution of voters along this ideological dimension.

In practice, several additional, overlapping complications for the model exist beyond merely the prospect of non-voting. I have broadly if somewhat arbitrarily divided these between extensions of spatial theory, specific theoretical considerations arising from the UK electoral system, and behavioural theory. The first of the extensions of spatial theory is the existence of multidimensional political ideology. A broad trend in political science over the past decades has dealt with various means of conceptualising new ideological cleavages in the electorate (see Ford and Jennings, 2020) and how challenger political parties have emphasised previously ignored issues (see Hobolt and De Vries, 2015). Including multiple dimensions within vote choice is straightforward, as parameters on the distances can be included representing the salience of a given ideological dimension

$$U_{ij} \propto \sum_{d=1}^{D} \beta_d |X_{id} - P_{jd}| \tag{4.3}$$

where D is the number of dimensions in the voting decision and β_d is the salience parameter for the dth dimension.

More complicated to consider - especially in a simulation context - is the fact that parties may attempt to introduce previously ignored issues to destabilise a previously unfavourable equilibrium. A fundamental result in formal theory is that equilibrium cannot be guaranteed once parties are given this ability (McKelvey, 1976, 1979). However, the fact that it is not guaranteed does not mean that it is not common for equilibrium to exist (Armstrong et al., 2020). For the purposes of the simulation of this paper I therefore make the simplifying assumption that counterfactual changes in party positions in the 2019 UK general election would not have introduced any changes in issue salience. In other words, as discussed above the two ideological dimensions at play in the election were economic issues and Brexit.

The second extension of spatial theory to discuss is that of categorisation theory. While categorisation itself is a theory of mental processing with a long pedigree, a comparatively recent paper introduced it to the realm of voters' understanding of ideological space (Bølstad and Dinas, 2017). In short, voters perceive political ideology - and the relationship of political parties to it - through 'coarse categorisations'. One example would be left-right: most voters will see parties as 'left' or 'right', with finer spatial distinctions mattering more for choosing between multiple parties on the same side as the voter. There is strong evidence in the case of the UK that voters do have strong Brexit identities that likely shape voters' perception of political space (Hobolt, Leeper and Tilley, 2021). I therefore compare simulations of both straightforward proximity models of spatial vote choice and models combining proximity and categorisation.

4.3.2 UK Electoral System

Like any other, the UK electoral system brings its own particular strategic considerations for both voters and political parties. First among these is the single-member plurality (SMP) voting system. In SMP, voters must weight their preferences against the probability of their preferred party actually winning. Often, voters vote strategically for a less-preferred party. It is therefore necessary to consider how likely a party is to win in a given constituency. For political parties however, SMP means that the best ideological position can be distorted by electoral geography. This is because if voters are not distributed randomly, but instead concentrated and dispersed in particular ways, this can separate the tasks of vote maximisation and seat maximisation. In practice, since center-left voters are typically concentrated in cities, electoral geography in SMP tends to skew results to the right (Döring and Manow, 2017). This is the abstract version of the argument made for a shift in a more pro-leave direction for the Labour party described above.

An additional consideration for party strategy in the UK context is the third parties. Until now my theoretical discussion has focussed on the two-party case and I have introduced additional complications to this straightforward competition. However, in the 2019 UK general election several additional parties represented additional key actors: the Liberal Democrats, the Greens, the Brexit Party, Plaid Cymru, and the SNP. In terms of the latter two, in the simulation I consider only English voters. This is because Wales and Scotland introduce an additional issue dimension in the form of nationalism vs unionism. This is especially pronounced in the case of Scotland, where the Labour Party is itself a de-facto third party. England represents the largest constituent nation of the UK by far and elections are broadly decided there, meaning that ultimately the simualted counterfactual should remain reasonably informative. Broadly, these parties should make it harder for both of the main parties to move away from their core voters, as such a move becomes riskier as they no longer need only be concerned with the threat of non-voting. In this particular case, the Liberal Democrats and Greens put a pressure towards remain on the Labour Party, while the Brexit Party puts a leave pressure on the Conservatives.

4.3.3 Behavioural Theory

The spatial and behavioural traditions of vote choice have sometimes been considered as two opposing traditions. However, there are real benefits to integrating both traditions (Adams, Merrill III and Grofman, 2005). The voting decision in practice is deeply habitual. Even in times of increased voter volatility (Fieldhouse et al., 2021, see), most voters will still tend to vote for the same party as before. Similarly, party loyalty matters a great deal; and often identification with a party may overcome the fact that another party may be ideologically closer. Finally, spatial vote choice does not fully encompass all aspects of the voting decision.

The fact that these variables do in fact matter has an impact on party strategy as parties are incentivesed to focus their strategy on voters at least somewhat drawn to them for non-policy reasons instead of all voters equally (Adams, Merrill III and Grofman, 2005). In other words: parties must compete for those voters who might vote for them, not for all voters overall. I therefore include these behavioural elements in the simulation. While it is reasonable enough to acknowledge that voters do prefer to vote when there is a party closer to their own views, and that voters do in fact vote strategically in SMP systems, neither of these phenomena are well-captured by the notions of utility gains. I therefore also dispense with further use of the language of utility, instead simply focussing on the role of spatial proximity and categorisation in the voting decision in its own right.

4.4 Data and Methodology

With the contextual background and theoretical considerations for the simulation established, I now turn to the dataset and methodology for the counterfactual simulation.

4.4.1 Data

In this paper I use the 17th wave of the British Election Study Internet Panel (BESIP) as a cross-sectional dataset (Fieldhouse et al., 2020). The English subsample of this wave contains 22,657 respondents. This wave was the pre-election wave for the 2019 general election collected in November 2019. I use this wave primarily to reduce of threat of reverse causality as for the dependent variable of the analysis I use vote choice in the actual election. To capture the two ideological dimensions of economics and Brexit at play in the election, I use two of the perceptual scales available in most BESIP waves in the form of the redistribution and EU integration scales. These are 0 to 10 self placements and placements of the political parties by respondents with the following item wordings:

- **Redistribution**: *Some people feel that government should make much greater efforts to make people's incomes more equal. Other people feel* *that government should be much less concerned about how equal people's incomes are. Where would you place yourself and the political parties on this scale?*

- **EU Integration**: *Some people feel that Britain should do all it can to unite fully with the European Union. Other people feel that Britain should do all it can to protect its independence from the European Union. Where would you place* ***yourself*** *on this scale?*

I choose these as reasonably close approximations to the ideological dimensions at play in the election. The redistribution variable should reliably capture economic differences in voters and parties, while the EU integration variable should proxy for Brexit positions. I do not use a traditional Left-Right variable because the interpretation of this likely does not so much capture a specifically economic dimension as a mix of the most salient dimensions. For the purposes of this paper, these dimensions are thus more usefully parameterised separately.

4.4.2 Methodology

To construct a simulated counterfactual of how the Labour Party would have performed with different Brexit positions, I proceed in four broad steps:

1. Scale voter and party positions
2. Run model for spatial vote choice
3. Simulate new results based on different Brexit positions for the Labour Party

4. Generalise the results from survey sample to England-wide

Scaling Voter and Party Positions

For the first step, I use Bayesian Aldrich-McKelvey scaling to rescale the voter and party placements (Hare et al., 2015). Aldrich-McKelvey scaling is a method used to correct differential item functioning and rationalisation bias in placements of external stimuli such as political parties along a given ideological dimension and thus to recover a corrected placement for each stimulus (Aldrich and McKelvey, 1977). The parameters recovered can then be applied to respondent self-placements to recover corrected respondent placements on the same ideological dimension as the external stimuli. Bayesian Aldrich-McKelvey scaling improves on the previous iteration of the model by allowing missing data in respondent placements of political parties. This is because the model follows a Bayesian approach to missing data, wherein missing data are treated as parameters to be estimated (Jackman, 2000; Hare et al., 2015). I therefore retain all respondents who reported placements for at least 3 of the 5 parties in BESIP. This has the benefit of reducing data loss in the scaling stages, particularly as data loss here would skew the sample towards more politically informed survey respondents. For each model, ten-thousand burn-in iterations were run across two chains. After that, five thousand draws were taken to construct the posterior distributions.

Spatial Vote Choice

To estimate a model of spatial vote choice, I take use a conditional logit model. The conditional logit model has a close relationship with multinomial models, as both model chooser i's choice Y_{ij} out of J outcomes. Where multinomial logit modes choice outcomes as a function of choice characteristics, conditional logit models choice outcomes as a function of choice characteristics. Substantively, this means I am able to model vote choice (whether individual i voted party j) as a function of party characteristics (such as voter-party distance on a given dimension). The conditional logit model is given by

$$Pr(Y_{ij} = y_j) = \frac{exp(x'_{ij}\beta)}{\sum_{h=1}^{J} exp(x'_{ih}\beta)} \tag{4.4}$$

where x'_{ij} is a vector of choice-specific covariates that may or may not also vary by chooser i, and β is the vector of regression parameters which is constant across choice options. This setup means that in line with the spatial model of vote choice I am able to model vote choice (whether individual i voted party j) as a function of party characteristics (such as voter-party distance on a given dimension). This also constrains each variable to have a single parameter - the effect of distance on a given dimension will be the same across parties.

I estimate two sets of models - one proximity model with distances on the redistribution and Brexit dimensions, and one proximity plus categorisation model. For the categorisation model, I use the 0 point on the rescaled data to determine whether a voter was on the same 'side' on a given distribution as the party or not. There is some arbitrariness in this as the rescaled data are interval and not ratio scale, meaning that the 0 point is not necessarily a meaningful one. However, the stimuli positions are constrained to be mean 0 and so the position should nonetheless be *close enough* to wherever the meaningful center point would be that this is a reasonably good approximation.

In taking this approach, I am simplifying somewhat from the theoretical models of spatial vote choice and categorisation theory. Both traditions - especially the latter - emphasise *perceived* ideological position and categorisation on the part of the voter. I am instead using an 'objective'[1] measure to determine ideological distance and categorisation. This is a simplifying assumption that facilitates the simulation: it would be difficult to design a simulation based on party position without being able to refer to a single, clear position for the Labour Party and to relate it to ideological distance and categorisation. I do believe that in practice this will not be too much of a distortion with respect to actual behaviour.

I further include several controls in both models, including respondent perceived probabilities of the party winning in their constituency, whether the respondent previously voted for that party, whether the respondent identifies with that party, respondent likeability ratings for the party, respondent likeability ratings for the party leader, and party dummies. Most of these control variables follow on from the preceding discussion and as they are straightforward binary variables do not require further explanation. The use of party dummies has a long pedigree in the spatial tradition, fulfilling a definition of valence as 'everything that's not spatial'. I further included the like data and probabilities to better decompose this, with party dummies therefore acting as a baseline for the likelihoods of that party being chosen.

Some further notes should be made explicit for some of the control vari-

[1] I do not mean objective in the sense that this is an unambiguous measurement, but rather in the sense that party positions are held constant across voters and absent their individual perceptions for both distance and categorisation.

ables. For the like data, to prevent missingness a value of '5' was imputed into the 0-10 scales where missing data occurred. This value was chosen as a reasonably good guess as to how someone without a strong opinion on a given party or leader might have responded if forced. The like data were regressed on the recovered ideology values for respondents; and predicted errors from these regressions were used. This is because prior to this the like data will be driven in part by spatial preferences - using the error terms from this regression partials the spatial component of like scores away. The like scores for the Green party leaders were averaged into a single score at the end of this process. For the Brexit Party, I treated past UKIP voters as past Brexit Party voters given the continuity between the two parties. Non-voting was included in the model as an additional choice alongside the five parties. For most variables, the value for this choice was set to 0. The exception was that respondents who previously did not vote were set to this being their previous choice. With this exception, the dummy for non-voting then becomes a sort of threshold which the other choices must overcome if the respondent is to turnout - meaning that turnout patterns will change as the Labour Party's Brexit position changes.

Brexit Position Simulation

Once the models were estimated in the previous step, I turned to the task of predicting choice probabilities across a range of positions for the Labour Party. While this prediction approach to counterfactuals has long existed in the spatial tradition, I adopted a cross-validation approach to modelling the data in the previous stage. I split 60% of the data into a training sample and 40% into a test sample. I ran the conditional logistic models on the training sample, then verified that both models reliably predict the correct vote shares for the test sample.

Once this was verified, I proceeded with the simulation step. Here, I simulated new positions for the Labour Party from -2 to 2 on the Brexit dimension (covering the vast majority of respondent positions) over steps of 0.01, recalculated the relevant party-voter distances and categorisations, then predicted new results from both models. I perform the simulation across such small step sizes to allow for quasi-continuous results. That said, there is no clear interpretation for any particular step-size, other than more or less distance. Indeed, perhaps ironically given the theoretical discussion of this paper it is only be arbitrarily categorising the regions of the plotted political space (e.g. into Remain vs Leave, or relative to the 'true' party positions) that changes in party position can be easily described.

Generalisation

While step 3 is sufficient to learn what position would have been best in terms of maximising vote share it does not answer questions regarding seat share maximisation. Resolving this is crucial to addressing arguments that suggested the Labour Party's best strategically optimal position would have been. I therefore generalise the simulated results from step 3 by utilising Uniform National Swing and Regional National Swing to predict seat shares based on these results.

Uniform National Swing is computed by assuming that each the vote share for each party in each constituency will be exactly the same. While obviously not correct in a strict sense of the term, UNS has a reasonably good predictive record. Uniform Regional Swing by contrast assumes a uniform swing *within* regions, but not nationally. To compute both, the predicted vote shares were aggregated, and the change in vote share was computed for each party nationally and regionally. The changes were then applied to the constituency results from the 2017 general election.

Causal Assumptions of the Simulation

In seeking to construct a counterfactual of the 2019 UK general election, I am by necessity working with observational data from a single run of the election. It is not possible to repeatedly re-run alternative versions of the election in a randomised experiment. The process of estimating a counterfactual thus requires causal assumptions, as does any work that attempts to make causal estimates from observational data.

The counterfactual simulation is built on an individual-level of vote choice. Party positions are then simulated over the estimated model. For the model to produce good counterfactual estimates, the parameters for ideological distance and categorisation need to be causal estimators In this subsection of the paper, I therefore set out explicitly the causal assumptions of the simulation. Many of these happen at the modelling stage, so I begin here. I then discuss some of the other assumptions that may affect the causal estimates of this simulation.

Figure 4.1 presents the relationships this paper assumes exists in the real world. Each node represents a variable, where 'Vote Choice' is the outcome of interest. This figure falls in the tradition of graphical representation of causal assumptions but is not a DAG. While some arrows are directed, others are bi-directional, representing the fact that it is plausible that these variables mutually affect each other.

For simplicity of presentation and discussion, some variables have been combined into a single node. Ideological distance and categorisation along all dimensions are captured in a single node, as are affective ratings towards parties and leaders. Finally, ‘valence’ as described earlier should be understood as capturing ‘everything that’s not spatial’ - in this case including perceived win probabilities (but not other variables explicitly plotted). Where variables share a node, it should be assumed that they mutually confound one another.

Figure 4.1: Data Generating Process

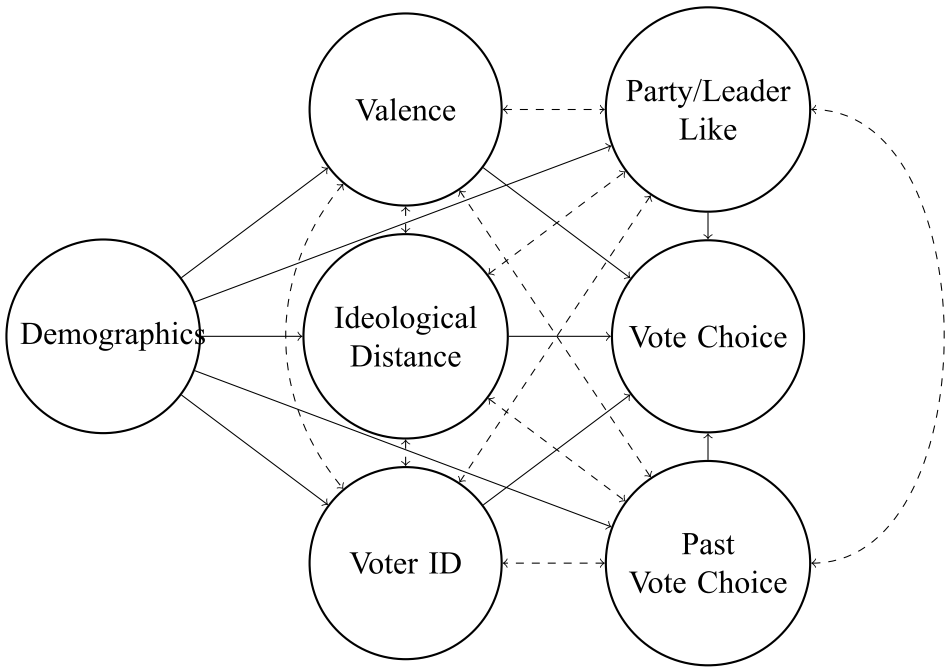


Figure 4.1 presents a complicated set of relationships, graphically representing the theoretical discussion and modelling decisions of this paper.

First and foremost, it assumes that there is no reverse causality or simultaneity bias between vote choice in the 2019 general election *at the time of the survey*, as the survey data was collected prior to the election while vote choice is from the time of the election. Setting up the data in this temporally staggered way should reduce the risk of these forms of causal bias, though this cannot be entirely ruled out (e.g. *anticipation* of one's vote choice may have a causal effect on other variables).

It is more difficult however to decide whether relationships between ideological distance and other variables should be considered confounders or mediators. Ideology may for instance drive voter identification - but voter identification may also drive ideology (even if we require monocausal relations for individuals, there is no reason the causal pathways should be the same across all individuals). Likewise, many of these variables could all plausibly have a causal effect on one another.

This tension is resolved in part by focussing on an explicitly *narrow* counterfactual. If the party changes position, while holding all other relevant factors constant how will individual vote choice change in terms of ideological distance and categorisation? Other variables, whether confounders or mediators, are held constant. Many things such as party identification will affect individual ideology rather at least as much as vice versa. Similarly, a failure to include these variables as controls would be a failure to acknowledge the point made by Adams, Merrill III and Grofman (2005) in highlighting the fact that parties cannot compete for all voters but must compete for the subset of voters who may be willing to support them.

One variable I have assumed cannot be caused by the other variables in this theoretical model is that of voter demographics. I assume that individual's age, education level, gender, etc, have causes independent of political attributes of that individual. I also assume, following the prior discussion, that demographics do not have an effect on vote choice (or, perhaps less restrictive but difficult to graphically represent, that their effect on vote choice is minimal).

Beyond the conditional relationships expressed here, other assumptions also enter the simulation. One assumption is of course that of the functional form of ideological distance, which has been made based on past research. Another assumption is the simplifying assumption moving away from *perceived* distance and categorisation and towards an 'objective' measure of these things. It is my view that this difference will in practice prove a subtle one, and that the model is still reasonably isomorphic to the true data generating process. Nonetheless, it is an assumption that must be explicitly acknowledged.

4.5 Analysis

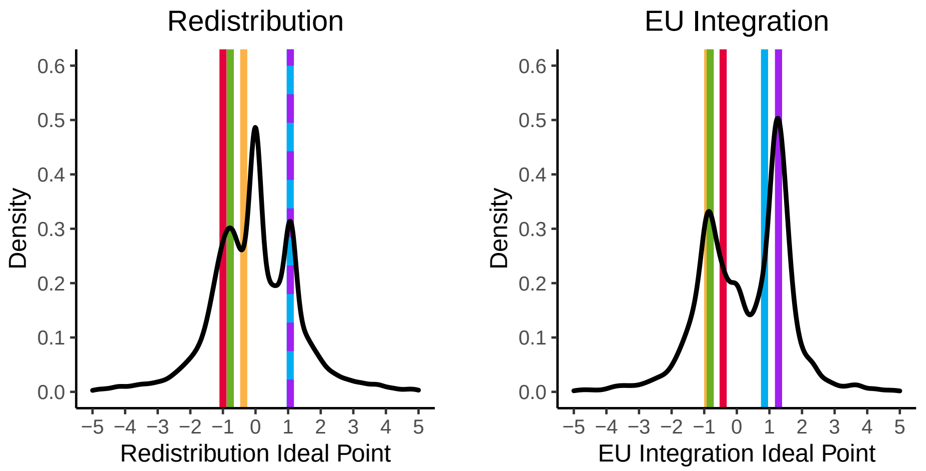
4.5.1 Bayesian Aldrich-McKelvey Scaling

Figure 4.2 shows density plots of the rescaled redistribution and EU integration scales from BESIP wave 17. The vertical lines overlayed on the density plots represent the median positions of the political parties from their posterior distributions. These points were used to calculate the distances in the conditional logit model. For four of the five parties, the lines were coloured using the party's colours[2]. The exception is the Brexit Party, for which purple was

[2] Conservatives as blue, Labour as red, Liberal Democrats as yellow, Greens as green

used to better differentiate it from the Conservative blue. On the redistribution plot, the Brexit Party line was dashed as its position was so close to the Conservatives' as to be overlapping on the plot.

Figure 4.2: Ideology Distributions



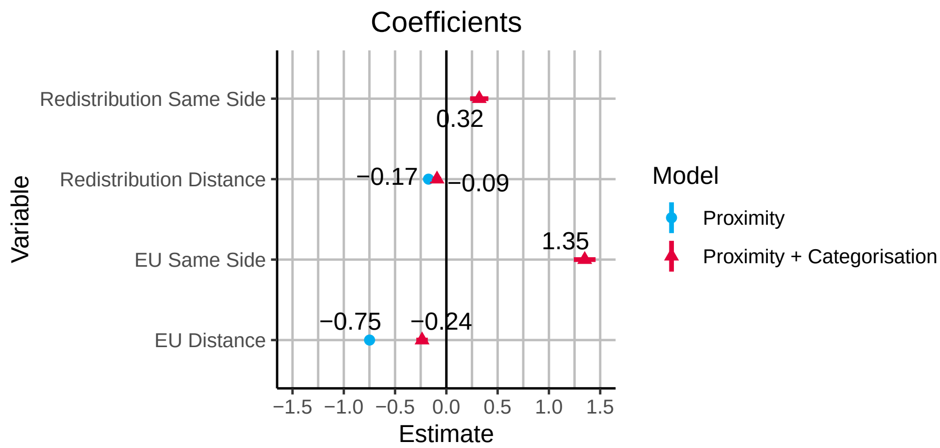
The black line corresponds to the density of the estimated respondent ideal points. The vertical lines represented the estimated locations of the political parties on these scales and are coloured according to political party. The exception is the Brexit Party, which is coloured as purple to better differentiate it from the Conservative Party.

Some clear differences between the distributions of the rescaled redistribution and EU integration variables are made visible by the plots. The redistribution variable has three peaks, but they are close to one another and there is a clear central tendency in-between the parties of left and right. By contrast, there is a clear dip in-between the Remain and Leave parties on the EU integration scale. In the redistribution plot, the Labour Party emerges as the most pro-redistribution, closely followed by the Greens. The Liberal Democrats are reasonably centrist (albeit appearing center-left), while the Conservatives and Brexit Party are equally anti-redistribution. By contrast, the Liberal Democrats and Labour Party essentially switch places in terms of centrism (although still pro-Remain). On the Leave side, the Brexit Party are clearly more pro-Brexit than the Conservatives. Both rank orderings carry a great deal of face validity in terms of the party placements. Similarly, it is unsurprising that the EU integration scale implies a larger divide between voters than the redistribution scale. One implication of this may be that given Downs' theory, there is more benefit in taking a centrist stance on economics and a more pro-Remain stance on Brexit. Of course, such a prediction can only be made prior to fully accounting for the myriad complications discussed above. In the next sections the survivability of Downs' claim through increased levels of complexity is therefore tested.

4.5.2 Conditional Logit

In figure 4.3 I present a coefficient plot from the two conditional logit models estimated. The blue coefficient estimates on the plot relate to the model containing only a proximity component. The red coefficient estimates on the plot relate to the model containing both proximity and categorisation components. 95% confidence intervals are included in the plot. A full table of regression results with all controls is included in the supplementary material. 95% confidence intervals are reported in the table, along with an asterisk where the null hypothesis of 0 is outside the interval.

Figure 4.3: Conditional Logit Coefficients



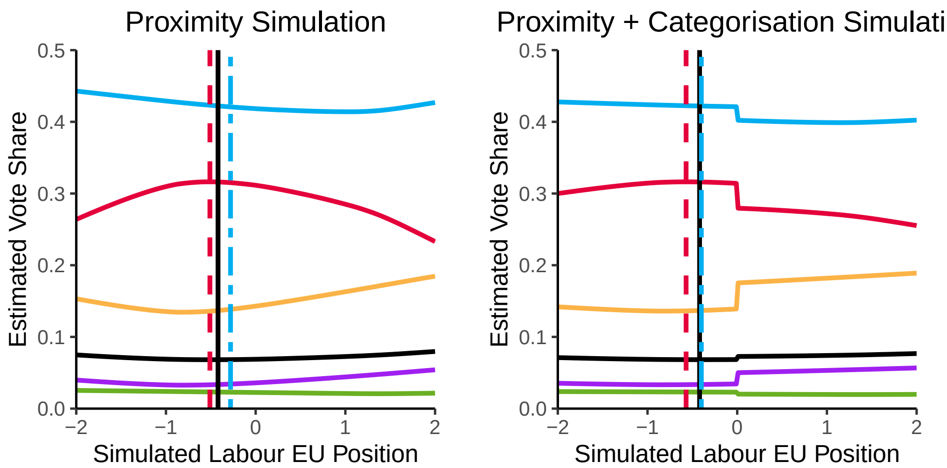
In both models the proximity coefficients for both scales are significant at the 95% confidence level and are negative. The proximity coefficient for the EU integration scale is larger, in line with the fact that the main issue in the election was Brexit. However, the size of the EU integration proximity coefficient varies considerably with the inclusion of categorisation effects. Without categorisation effects, the coefficient is roughly 0.5 larger in the proximity-alone model. The categorisation effects in the second model for both scales are significant at the 95% level and positive, although again in line with the previous model the EU integration categorisation effect is considerably larger than the redistribution categorisation effect. Where these models corroborate one another is in confirming that the primary issue of the election was Brexit. I do not attempt to choose between the models through any formal testing as their comparison is itself useful, but I do highlight that recent results lie in favour of categorisation theory. The categorisation model is thus likely the closer approximation to reality. Cross-validation results for both models are available in the supplementary material.

4.5.3 Simulation

Figure 4.4 contains plots showing how the party and non-voter shares of the sample change as the simulated value for Labour Party vote share changes. The x-axis shows the simulated Labour Party positions along the BESIP EU integration scale, while the y-axis shows the proportion of respondents. The lines along the plot as before are in the party colours and show how the proportion of choices changes with simulated Labour party positions. The black line represents non-voters. The vertical lines represent 3 separate points. The solid black line visualises the 'true' Labour Party position extracted from the Bayesian Aldrich-McKelvey scaling. The red dashed line visualises the simulated position that maximises the Labour Party's vote share. The blue dot-dash line visualises the position that minimises the difference in the Labour-Conservative vote share. The left plot visualises simulation results for the proximity-only model, while the right plot visualises simulation results for the proximity plus categorisation model.

For my primary purpose in this paper, the most salient feature in figure 4.4 is the convergence of evidence showing that the Labour Party optimises its vote share broadly by being a party of Remain. In both plots evidence suggests that Labour Party maximises its vote share by being to a small degree more pro-Remain than it was in the election. There is less convergence in the two models regarding the position that minimises the Labour-Conservative

Figure 4.4: Sample Changes in Vote Share



The horizontal lines represent the estimated vote shares of the political parties in the sample, with the black line being the percentage of non-voters. The vertical lines are various Labour Party positions. The solid black vertical line is the Labour Party's original position, the dashed red vertical line is the Labour Party's vote-maximising position, while the dot-dash blue line is the position minimising the gap between the Labour Party and the Conservative Party.

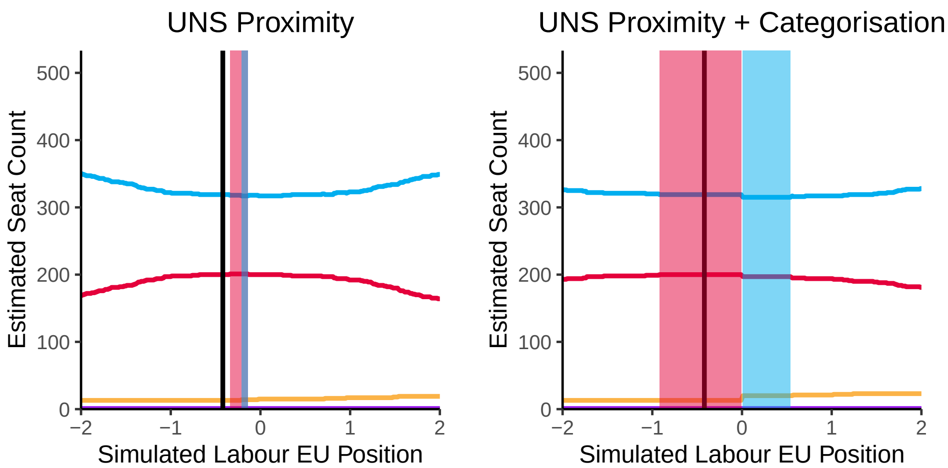
difference in vote share with the proximity-only model favouring a marginally more pro-Leave position (though still overall pro-Remain) and the proximity plus categorisation model suggesting that the 'true' position was in fact approximately best for the party. However, in all cases the 'true' position was not far from the position implied by the simulation to be best for the party's results.

This is a result that cuts in both directions - both simulations also strongly imply that the party would not have benefited from taking a more pro-Remain position than it had already taken. Some degree of moderation was necessary. Matching the Liberal Democrats' position on Brexit would not have been beneficial to the Labour Party. Although this point does vindicate the overall position taken by the party, it also suggests that those looking to Brexit policy to improve the party's vote share were mistaken. In line with recent results regarding centrist parties (see Zur, 2019, 2021), changing the party's spatial position does not drastically improve performance. A wider implication of my results here given others may in fact be that despite common interpretations of politics in terms of ideology, once other components in the voting decision are included changes in party position do not necessarily change electoral results to the degree we might expect. In both cases, the rank order of party vote shares remains the same throughout the simulations.

4.5.4 Generalisation

While the simulation results broadly confirm Downsian intuitions around a polarised electorate requiring the Labour Party to lean Remain (if not on the mode of that side of the distribution), arguments suggesting the party needed to move in a more Leave direction must be addressed. I utilise both uniform national swing (UNS) and uniform regional swing (URS) to this purpose, so as to again check the extent to which results converge. It is however likely that URS will pick up on regional nuances that UNS does not, so insofar as results diverge it may well be the more accurate reference point for discussion. Figure 4.5 contains the UNS results. As before, the x-axis contains the simulated Labour Party positions. The y-axis the seats shares of the parties and the lines show the number of seats that party has won. One again the black solid vertical line shows the 'true' Labour Party position from the Bayesian Aldrich-McKelvey scaling. The red shaded area shows the range of positions where the Labour Party maximises its seat share while the blue shaded area shows the range of values where the Conservative-Labour seat difference is minimised. The left plot shows results for the proximity model while the right plot shows results for the proximity plus categorisation model.

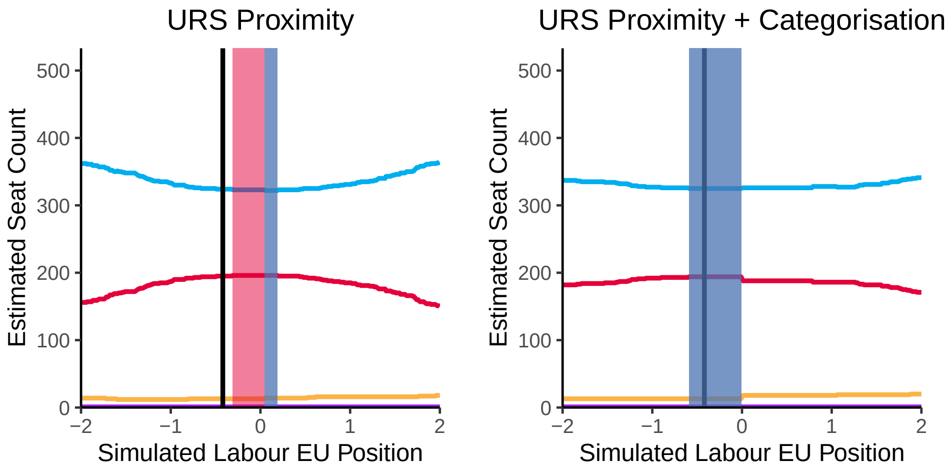
Figure 4.5: UNS Changes in Seat Count



The horizontal lines represent the estimated seat counts of the political parties in the sample, with the black line being the percentage of non-voters. The vertical lines are various Labour Party positions. The solid black vertical line is the Labour Party's original position. The red shaded area is the Labour Party's seat maximising range, while the blue shaded area is the range for minimising the seat gap between the Labour Party and the Conservative Party.

An immediate but clear issue from these results is that the results from both models only half-agree. There is consensus that the Labour Party maximises its seat share on the remain side of the scale, but there is less consensus on minimising the gap between the Conservatives and the Labour Party. On the proximity model, this is very near the 0 point and on the remain side. In the proximity plus categorisation model however, this is after the 0 point. The issue seems in part to be driven by the question of the extent to which the Liberal Democrats benefit from the Labour Party's pro-Leave shift. The move is less drastic in the proximity model, so the difference may be driven by this fact. However, when we turn to the URS results, this becomes less clear. Figure 4.6 visualises these results, following the same structure as figure 4.5.

Figure 4.6: URS Changes in Seat Count



The horizontal lines represent the estimated seat counts of the political parties in the sample, with the black line being the percentage of non-voters. The vertical lines are various Labour Party positions. The solid black vertical line is the Labour Party's original position. The red shaded area is the Labour Party's seat maximising range, while the blue shaded area is the range for minimising the seat gap between the Labour Party and the Conservative Party.

Once again, there is convergent evidence in favour of a pro-Remain stance maximising the Labour Party's seat share. So long as the party stays on the Remain side of the 0 point, it is able to maximise its vote share. However, there is divergent evidence regarding the minimisation of the Conservative-Labour seat gap in the opposite direction. Here, the proximity model favours a Leave stance while the proximity plus categorisation model favours a Remain stance. Following the guidelines set above, one way to approach this issue may be to highlight that the proximity plus categorisation model is a theoretically better approximation to vote choice, while the URS generalisation is a theoretically better approach. Overall though, there is clear convergent evidence of a pro-Remain stance maximising the Labour vote share, minimising the Conservative-Labour vote gap, and maximising Labour's seat share. It is therefore probably the best approach given the available evidence, albeit with some remaining ambiguity around minimising the seat gap between Labour and the Conservatives.

4.6 Robustness of Results

Although each of my modelling decisions in building this simulation have been driven by theory, it is nonetheless important to establish how robust to particular decisions these results are. I verified my results against two important decisions: the functional form of the utility loss function, and the precise placement of the midpoint for the 'same side' approximation. In each analysis, I focus primarily on the robustness of the simulated vote shares result, as it is this result which is most consistent throughout my analysis. All robustness check plots are presented in the supplementary material for this paper.

First, to verify that the predicted election results in the simulation model are reasonably close to reality, I present some tables showing the estimated vote shares from the simulation with the simulated EU position for the Labour party closest to the estimated one from the Bayesian Aldrich-McKelvey scaling. In other words, the one which should be a close match to the real election results. I present two tables: one with non-voters, and one with percentages adjusted to remove non-voters. BESIP like other online surveys is poor at picking up non-voters and thus the first underestimates non-voting. The second however is a close match, with all parties within 3 percentage points of their true position. The divergence for the Greens is possibly the most meaningful, as a drop in two points is half their vote. The main thing to draw from this is that the role of non-voters in optimal party strategy may not be fully captured here.

To examine the robustness of my results to choice of utility loss, I reran the analysis with squared distances between parties and voters. Of the robustness checks, this was most challenging for the results of the paper. The resulting vote maximising and vote-gap minimising positions shift, relative to the main results, in a more pro-leave direction. The difference is most marked for the model without categorisation effects, which shows the vote minimising position as being very near the center. The model with categorisation effects by contrast shows more similar results to the main results of this paper. The theoretical result that absolute distances best capture voter utility loss is thus an important one for the results presented here - especially if we were to ignore categorisation effects.

One important check is on the inclusion of non-spatial variables which nonetheless contribute to the voting decision. Given that I opt to include variables that both confound and mediate spatial vote choice, it is worth examining how the results change with their non-inclusion. I therefore ran models and simulations without these controls (but still with party dummies, to broadly capture valence effects and to act as a baseline for non-voting). The variables removed are the perceived win probability of the party in question, whether the voter voted for the party in the previous election, whether the voter identifies with that party, how much the voter likes the party, and how much the voter likes the party leader.

Somewhat in line with Downs' theory, the Labour Party's vote maximising position is still close to its original position. However, in line with Adams, Merrill III and Grofman (2005) the results change for the *vote-gap minimising* position as it is on the pro-Leave side of the dimension. This is because while the Liberal Democrats' pro-Remain stance results in it taking a large number of voters from the Labour Party, the Labour Party gains many pro-Leave voters for the Conservatives at a fast rate. This robustness check therefore largely shows that in no small part an important part of the reason a pro-Remain strategy was necessary for the Labour party was that non-policy portions of the voting decision such as partisanship meant that Leave voters were unlikely to switch loyalties.

While I have primarily motivated this paper in terms of the spatial model of vote choice, it is worth considering the fact that demographics may contrary to the causal assumptions of the simulation be key drivers of vote choice, or indeed of turnout given the inclusion of non-voting in the model. I therefore ran a conditional-multinomial logit model with individual-level controls with choice-varying parameters (i.e. multinomial model parameters). These variables included age, gender, education, whether someone owns or rents their home, and ethnicity. The results from this simulation broadly corroborate the results of this paper, showing the models are robust to non-inclusion of demographic variables.

Since the selection of a center point is in practice somewhat arbitrary given the fact that the scale extracted from Bayesian Aldrich-McKelvey scaling is interval rather than ratio scale, I ran two further sets of models while varying the choice of center point on the EU dimension. In the first, I set the center point to be -0.1. In the second, I set the center point to be 0.1. Broadly, these results corroborate the main results in the paper - the categorisation model remains robust to selection of this point. Overall, the inferences of this paper are robust, although under a different utility loss function a more moderate though still pro-Remain stance may be recommended for the Labour party.

4.7 Conclusion

The clear conclusion of my counterfactual simulation is that the Labour Party is best off as a party of Remain. This evidence is in line with recent more general research on the positions of social democratic parties with respect to second dimension issues (Abou-Chadi and Wagner, 2019, 2020). On the whole, the party clearly had its stance about right in the election - the vote-maximisation and vote gap-minimisation points from both models were all near the ‘true’ Labour Party position. Becoming as firmly Remain as the Liberal Democrats or Greens would have been a mistake. Similarly, the range of positions in which the party maximises its seats is on the Remain side of the 0 point - even in the proximity-alone models where this point is not explicitly used in the variables of the model. Some ambiguity remains around the range of values where seat gap-minimisation occurs, but broadly it would be inadvisable to build a strategy around this ambiguous result instead of the firmer conclusions set out above.

There are several wider implications for the simulation beyond confirming recent research on social democratic party placement. The first is that it appears once enough aspects of the voting decision are considered that there is little room for political parties to drastically alter electoral results through changes in their spatial position. Although the vote and seat shares could shift in sometimes large amounts, the overall rank order of the parties did not change at any point. There does appear to be some distortion being introduced by the UK's SMP electoral system in that a range of positions produces the same seat count, but the ambiguity in this set of results means that the Labour Party was best off focussing on votes alone. This has some relevance to the ongoing debate on proportional representation in the Labour Party: it may simplify party strategy by more clearly aligning vote and seat maximisation.

However, the counterfactual simulation does have some limitations which require discussion. First and most obvious is that I only consider the effect of changes in Labour Party position on vote choice. In practice, such changes would likely produce new information and arguments in the form of media reactions, and new incentives for political actors. It is not even necessarily clear if Johnson would have been willing to call the election had the Labour Party taken a different stance. However, on this front I argue that the purpose of the simulated counterfactual is not to be a full simulation of reality in all its complexity but rather to be sufficiently informative to a particular political debate. I argue that in terms of this goal it has succeeded.

Some further general limitations should also be acknowledged. First, I do not fully account for all the quirks of the election. I do not account for the fact that the Brexit Party stood down in Conservative incumbent seats. The arbitrariness of the 0 point in Bayesian Aldrich McKelvey scaling is a theoretically important point to acknowledge. It is set by assuming the mean point of the political parties to be approximately 0. In practice, insofar as the scale is a reasonably good approximation to a hypothetical 'true' ratio scale with a meaningful 0 point, the distance (once unit size is accounted for) is probably a reasonably close match by merit of the fact it will be somewhere between the two groups of parties. The fact that the proximity-alone models do seem to capture this in some of the seat share predictions would seem to lend confirmation to this point. For some of the confounding relationships in the model, it is not clear to what extent these variables are confounders versus mediators. Finally, a theoretical point of Downs' that I do not model is the notion of party brands. Downs taxes as axiomatic that where a party moves to the other 'side' of the center point, no one will wish to vote for it because it can no longer be trusted. Insofar as this is true, a move past the center point of the EU integration scale would have resulted in fairly drastic collapse in the Labour Party vote share. If true, this point lends further credence to my conclusion regarding the optimal point for the Labour Party on the scale.

Chapter 5

Conclusion

Shortly before this thesis was due, the UK Conservative party membership elected Liz Truss as their leader and thus also as Prime Minister. In a short two weeks, Truss and her Chancellor of the Exchequer, Kwasi Kwarteng, opted for an economic policy that included major tax cuts for the richest in society. The decision was a catalyst for a crash in the value of the pound and a sudden rise in mortgage interest rates. It also resulted in a piece of research being published in the Financial Times (FT) on the very day this thesis was due to be submitted.

In the FT piece, the FT collects data from the British Election Study Internet Panel (BESIP), the Chapel Hill Expert Survey (CHES), and its own top-up survey of experts. The BESIP left-right and libertarian-authoritarian positions in BESIP are compared and contrasted with CHES economic left-right and GAL-TAN placements[1]. It used a top-up survey of experts to learn about moves in Labour and Conservative party positions since 2019 (Burn-

[1] This is not made explicit, but rather is my own best guess

Murdoch, 2022).

The FT piece makes several arguments. First, it suggests that on the basis of the top-up expert survey, the UK Conservative party has become the most economically right wing in the world (or at least among the parties for which we possess data). Second, it highlights that on the basis of the BESIP scales, the British electorate is primarily left-wing and authoritarian. Third, it therefore concludes that the shift to the right in the Conservative party is primarily a shift away from the core views of the electorate. There are therefore several trends here: the extent to which we can make comparative claims, measurement inference from survey data to an underlying concept, and consideration of parties and voters on the same scale. I treat each of these in the order I raise them in the essays.

5.1 Measurement Inference from Survey Data

Core to the FT's analysis is the notion that the Conservative party has moved dramatically away from the electoral centre of gravity. It's certainly true that the tax cuts represent a shift to the economic right: but is the British public really left-wing? The first essay I present in this paper suggests no - or at least that we cannot readily make this inference from the BESIP Likert scales. The FT is not the only organisation to use the BESIP Likert scales in this way. A very prominent report by UK in a Changing Europe compares the ideological positions voters, party members, and MPs. All three surveys (including BESIP) used within the report use the same Likert scales (Bale et al., 2020).

As essay one makes clear however, we cannot simply assume without investigation that particular patterns in survey responses tell us something about the underlying concept, rather than the way in which respondents use the survey scale. Essay one provides clear evidence not only that the BESIP and BSA scales suffer from acquiescence bias, but that post-correction we are likely to draw very different inferences from the data. Moreover, it seems likely that given MPs are drawn from a disproportionately educated portion of the electorate and their self-evidentially high level of political engagement and information, they are probably less acquiescent than the wider public in their survey responses.

To what extent the gap between voters and MPs shown in Bale et al. (2020) is a question to be pursued in another piece of research. MPs clearly do *respond* to the survey scales in a systematically different way than the public. But do we accept the measurement inference from these scales to the theoretical concept of interest at face value? What this thesis makes clear is that we cannot take the scales used in political science research at face value. Moreover, once we consider the various survey-related biases, we may find that previous results were driven by these biases rather than the underlying concepts we are interested in.

It does not follow that political science is awaiting some large-scale disaster by a failure to fully engage with the measurement literature. In some cases old results will be brought into a new light - in others measurement inference may demonstrate that past results are actually strengthened once measurement bias is removed from the equation. We won't however know until we properly pursue solutions to the measurement problems in our data. As in essay one, many of these solutions are problem-specific. Where the person-intercept

CFA method solves for acquiescence bias, Aldrich-McKelvey scaling used extensively throughout essays two and three solve primarily for differential item functioning. Always, emphasis should be on the inferences we wish to make from our data to specific concepts.

5.2 Comparison of Measures

Prior to demonstrating the distribution of voter positions, the FT piece makes the claim that the UK Conservative party has become the most economically right-wing in the world (Burn-Murdoch, 2022). But has it? The expert survey respondents were all from the UK, all experts on the UK, and broadly they were used to thinking about left-right economics within this specific context. Even assuming longitudinal validity relative to past UK placements, has the Conservative party really become more economically right-wing than say, Bolsonaro's Social Liberal Party in Brazil?

The answer is we don't know. It may be that the FT's survey respondents were over-correcting due to the dramatic events in the news, or it may be that the Conservative party has genuinely shifted dramatically to the furtherest pole of the economic right. This kind of consideration of how observed data from different contexts can be compared forms a core component of essay two.

Substantively, essay two's primary contributions are in terms of the relationship between age-period-cohort (APC) effects and left-right political ideology. It finds in favour of age effects, life-cycle effects, and cohort effects. It finds that cohort effects are reasonably similar from country to country, and that constraining them thus in a model is a relatively weak constraint. Period effects by contrast produce a fairly different set of results once unconstrained. The substantive implication here is that where the socialising influences that last with a generation are reasonably similar from country to country, the transient political moments are not. This deserves further exploration in future research.

The APC results presented here cannot be separated from the choice of measure and measurement method. ‘Relative’ ideology as I have defined it allows the specific constellation of issue salience and political center ground to vary from context to context. Instead, I build on work on party positions and utilise a measure that places respondents relative to the standardised party position in their country-year context. It would however have been more interesting still to compare these results against results based on measure of ‘absolute’ ideology, where all respondent positions had been rescaled to sit on a common scale with a common meaning.

The FT’s analysis runs into a similar set of issues. As raw placements, the expert surveys utilised by the FT are excellent. However, when time comes to compare party system to party system, ambiguities creep into the analysis. In what sense is the UK conservative party more economically right-wing than Bolsonaro’s Social Liberals? Essay two implies that it is in the relative sense. A standardisation of the FT’s findings would likely find the UK conservative party at some distance from the mean position of parties in the UK. The question however is whether an absolute measure would find the same result, or not. It’s a question that for the time being must go unanswered.

As with acquiescence bias, such an approach would be best facilitated at the survey design stage, likely best through the use of anchoring vignettes

(King et al., 2004; Hopkins and King, 2010). This methodology uses a survey vignette, meaning that when respondents locate the vignette stimulus on a scale they do so on the basis of common information. These anchoring vignettes can be mixed with approaches such as Aldrich-McKelvey scaling and its derivatives (Bakker, Edwards, Jolly, Polk, Rovny and Steenbergen, 2014; Bakker, Jolly, Polk and Poole, 2014; Bakker, Jolly and Polk, 2022), or through the use of an anchoring method in its own right (King and Wand, 2007). I therefore strongly recommend the inclusion of such vignettes in comparative datasets, so that in the long term we may begin to easily utilise measures of absolute ideology. By doing so, the issue of survey item comparability is thus addressed. This is not true of political ideology alone: this approach is valid for several other survey measures in which comparability is a concern.

Things are not however hopeless with regards to historic data. Anchoring methods and scaling methods both rely on the notion of rescaling data with respect to an external, 'gold standard' measure of ideological position. One modest proposal I would suggest is to utilise the expert survey methodology, but to have the expert respondents explicitly place historic party platforms against one another on the same scales. At this point, these placements could be used to generate placements against which historical data could be scaled. In the worst-case scenario, this method would confirm the distinction between relative and absolute ideology is not a particularly meaningful one. In the best-case scenario, we will have obtained a valuable route into learning more about the nature of political ideology.

5.3 Voters and Parties on the Same Scale

Finally, the third claim in the FT article is with regards to voter-party distance on ideological scales. The use of BESIP's Likert scales and CHES's ordered rating scales is an interesting one. Implicit in this decision is that these scales capture the same underlying concepts - albeit one at the voter level, one at the party level. It may be that there is a conceptual match here, or maybe not - it is not an easy thing to assess. Essay 3 grapples with the same problem in the process of producing a counterfactual of the 2019 general election in terms of Labour Party strategy on the EU.

In essay three, I use the spatial theory of vote choice to construct a counterfactual of the 2019 UK general election. The election is an interesting case study for three reasons. First, it is a clear example of a social democratic party grappling with the rise of a non-economic issue as the primary issue of political contestation. Second, in this election the issues at stake were clear: the EU as the primary issue, redistribution as a secondary issue. Third, the UK's electoral system may distort the strategic centre of gravity in a more right-wing direction.

By utilising Bayesian Aldrich-McKelvey scaling, I was able to generate DIF-corrected measures of party position and survey respondent position on the same scale. Aldrich-McKelvey scaling and its Bayesian variation both begin by estimating the positions of stimuli and the parameters relating respondent placements of these stimuli to the estimated positions. In Bayesian AM scaling, the parameters are estimated concurrently. In regular AM, they are obtained through a regression of the estimated positions on respondent placements. These parameters are then used to generate a respondent position on the same scale, assuming the respondent used the same DGP to place themselves.

AM scaling is therefore a clear conceptual advance. Not only does it correct for DIF and differences in perceived meaning, it also allows us to be certain of possessing placements on the same scale. In essay three, I leverage this in a conditional logit model, which is invariant to respondent-specific variables (i.e. demographics). I use this model to generate parameters of vote choice given spatial positions, then simulate Labour Party positions and use these parameters to simulate vote choice probabilities given Labour Party EU positioning. Once vote shares were generated from this, I then used Uniform National Swing and Uniform Regional Swing to generate seat shares. On the whole, I found that the Labour Party did best as a pro-EU Remain party, while results for seats were more ambiguous.

Careful attention to how voters and parties can be placed on the same scale thus enabled me to offer a genuine contribution to a debate that thus far had been based solely on speculation. This attention was not limited to merely placements on the same scale, but how the relationship between those voter and party placements could feed into a model of vote choice.

Had the FT given consideration to scaling approaches for its analysis, we could potentially hold more certainty in its arguments regarding the gulf between the electorate and Conservative party. As things stand however, we are forced instead to question whether the same concept is being measured across the scales being utilised. It seems particularly likely that GAL-TAN and Libertarian-Authoritarian are more competing conceptualisations, rather than merely the same scale measured differently.

5.4 The Contribution of this Thesis

As the preceding discussion shows, the contribution of this thesis goes beyond the individual contributions of these essays. As standalone papers, they offer a new methodological approach for acquiescence bias, greater substantive understanding of APC effects on political ideology, a new conceptualisation of relative versus absolute ideology, a resolution to a long-standing debate in British politics, and strategic advice for the Labour Party. Taken as a whole however, they also make a major contribution in highlighting how measurement inference can improve the research we conduct.

Simply by paying attention to the numerous measurement challenges raised and addressed in the course of these three essays, I have been able to adapt this conclusion in the space of less than a day to highlight issues in the FT's analysis. This thesis doesn't just provide a methodological approach to performing measurement inference: it highlights how measurement inference is intimately connected both with the task of conceptualising ideology and substantively interpreting research results.

It will not be the case that all hitherto research that has not fully engaged with measurement inference will be found wanting. In some cases, results will be shown to be driven by methodological artefacts. In others, new results may be discovered by thinking carefully about how our observed data relates to the concepts in which we are interested. In others still we may simply find that measurement error resulted in attenuation bias of past results, or even that it did not really make much of a difference. The point is that on a case by case basis, we simply do not know until we properly apply measurement inference. It is therefore my hope that this thesis makes a significant contribution to the task of refining our approach to the measurement of political ideology - and more- in political science research.

Chapter 6

Bibliography

Abou-Chadi, Tarik and Markus Wagner. 2019. "The electoral appeal of party strategies in postindustrial societies: when can the mainstream left succeed?" *The Journal of Politics* 81(4):1405–1419.

Abou-Chadi, Tarik and Markus Wagner. 2020. "Electoral fortunes of social democratic parties: do second dimension positions matter?" *Journal of European Public Policy* 27(2):246–272.

Adams, James and Samuel Merrill III. 1999. "Modeling Party Strategies and Policy Representation in Multiparty Elections: Why are Strategies so Extreme?" *American Journal of Political Science* pp. 765–791.

Adams, James and Samuel Merrill III. 2000. "Spatial Models of Candidate Competition and the 1988 French Presidential Election: Are Presidential Candidates Vote-Maximizers?" *Journal of Politics* 62(3):729–756.

Adams, James, Samuel Merrill III and Bernard Grofman. 2005. *A unified* *theory of party competition: A cross-national analysis integrating spatial and behavioral factors*. Cambridge University Press.

Aldrich, John H and Richard D McKelvey. 1977. "A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections." *The American Political Science Review* 71(1):111–130.

Armstrong, David A, Ryan Bakker, Royce Carroll, Christopher Hare, Keith T Poole and Howard Rosenthal. 2020. *Analyzing spatial models of choice and judgment*. 2 ed. Chapman and Hall/CRC.

Bakker, Ryan, Catherine De Vries, Erica Edwards, Liesbet Hooghe, Seth Jolly, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen and Milada Anna Vachudova. 2015. "Measuring party positions in Europe: The Chapel Hill expert survey trend file, 1999–2010." *Party Politics* 21(1):143–152.

Bakker, Ryan, Erica Edwards, Seth Jolly, Jonathan Polk, Jan Rovny and Marco Steenbergen. 2014. "Anchoring the experts: Using vignettes to compare party ideology across countries." *Research & Politics* 1(3):2053168014553502.

Bakker, Ryan, Seth Jolly and Jonathan Polk. 2022. "Analyzing the cross-national comparability of party positions on the socio-cultural and EU dimensions in Europe." *Political Science Research and Methods* 10(2):408–418.

Bakker, Ryan, Seth Jolly, Jonathan Polk and Keith Poole. 2014. "The European common space: Extending the use of anchoring vignettes." *The Journal of Politics* 76(4):1089–1101.

Bale, Tim, Aron Cheung, Philip Cowley, Anand Menon and Alan Wager. 2020. "Mind the Values Gap: The Social and Economic Values of MPs, Party Members and Voters." `https://ukandeu.ac.uk/wp-content/uploads/2020/06/Mind-the-values-gap.pdf`. Accessed: 2021-09-14.

Barberá, Pablo. 2015. "Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data." *Political analysis* 23(1):76–91.

Barberá, Pablo, John T Jost, Jonathan Nagler, Joshua A Tucker and Richard Bonneau. 2015. "Tweeting from left to right: Is online political communication more than an echo chamber?" *Psychological science* 26(10):1531–1542.

Barnes, Lucy. 2022. "Taxing the rich: public preferences and public understanding." *Journal of European Public Policy* 29(5):787–804.

Bell, Andrew. 2020. "Age period cohort analysis: a review of what we should and shouldn't do." *Annals of Human Biology* 47(2):208–217.

Bell, Andrew and Kelvyn Jones. 2013. "The impossibility of separating age, period and cohort effects." *Social science & medicine* 93:163–165.

Bell, Andrew and Kelvyn Jones. 2014*a*. "Another 'futile quest'? A simulation study of Yang and Land's Hierarchical Age-Period-Cohort model." *Demographic Research* 30:333–360.

Bell, Andrew and Kelvyn Jones. 2014*b*. "Don't birth cohorts matter? A commentary and simulation exercise on Reither, Hauser, and Yang's (2009) age–period–cohort study of obesity." *Social Science & Medicine* 101:176–180.

Bell, Andrew and Kelvyn Jones. 2015. "Should age-period-cohort analysts accept innovation without scrutiny? A response to Reither, Masters, Yang, Powers, Zheng and Land." *Social Science & Medicine* 128:331–333.

Bell, Andrew and Kelvyn Jones. 2018. "The hierarchical age–period–cohort model: Why does it find the results that it finds?" *Quality & quantity* 52(2):783–799.

Billiet, Jaak B and Eldad Davidov. 2008. "Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design." *Sociological Methods & Research* 36(4):542–562.

Billiet, Jaak B and McKee J McClendon. 2000. "Modeling acquiescence in measurement models for two balanced sets of items." *Structural equation modeling* 7(4):608–628.

Black, Duncan. 1948. "On the Rationale of Group Decision-Making." *Journal of political economy* 56(1):23–34.

Bølstad, Jørgen and Elias Dinas. 2017. "A categorization theory of spatial voting: How the center divides the political space." *British Journal of Political Science* 47(4):829–850.

Brown, Timothy A. 2015. *Confirmatory factor analysis for applied research*. Guilford publications.

Budge, Ian. 2001*a*. "Validating party policy placements." *British Journal of Political Science* 31(1):179–223.

Budge, Ian. 2001*b*. Validating the Manifesto Research Group approach: theoretical assumptions and empirical confirmations. In *Estimating the policy position of political actors*. Routledge pp. 70–85.

Budge, Ian. 2002. "Mapping policy preferences: 21 years of the Comparative Manifestos Project." *European Political Science* 1(3):60–68.

Budge, Ian, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara, Eric Tanenbaum et al. 2001. *Mapping policy preferences: estimates for parties, electors, and governments, 1945-1998*. Vol. 1 Oxford University Press on Demand.

Burn-Murdoch, John. 2022. "The Tories have become UNMOORED from the British people.".
URL: *https://www.ft.com/content/d5f1d564-8c08-4711-b11d-9c6c7759f2b8*

Butler, David and Donald Stoke. 1974. *Political change in Britain: The Evolution of Electoral Change*. St. Martin's Press.

Campbell, Angus, Philip E Converse, Warren E Miller and Donald E Stokes. 1960. *The American Voter*. University of Chicago Press.

Castles, Francis G and Peter Mair. 1984. "Left–right political scales: Some 'expert'judgments." *European Journal of Political Research* 12(1):73–88.

Cloud, Jonathan and Graham M Vaughan. 1970. "Using balanced scales to control acquiescence." *Sociometry* pp. 193–202.

Converse, Philip E. 1964. "The nature of belief systems in mass publics." *Critical review* 18(1-3):1–74.

Dawson, Richard E and Kenneth Prewitt. 1968. *Political socialization: an analytic study*. Little, Brown.

Döring, Holger and Philip Manow. 2017. "Is proportional representation more favourable to the left? Electoral rules and their impact on elections, parliaments and the formation of cabinets." *British Journal of Political Science* 47(1):149–164.

Down, Ian and Carole J Wilson. 2013. "A rising generation of Europeans? Life-cycle and cohort effects on support for 'Europe'." *European Journal of Political Research* 52(4):431–456.

Down, Ian and Carole J Wilson. 2017. "A rising generation of Europeans? Revisited." *European Journal of Political Research* 56(1):199–214.

Downs, Anthony. 1957*a*. "An Economic Theory of Democracy.".

Downs, Anthony. 1957*b*. "An economic Theory of Political Action in a Democracy." *Journal of political economy* 65(2):135–150.

Elff, Martin. 2013. "A dynamic state-space model of coded political texts." *Political analysis* 21(2):217–232.

Evans, Geoffrey A and Anthony F Heath. 1995. “The measurement of left-right and libertarian-authoritarian values: A comparison of balanced and unbalanced scales.” *Quality and Quantity* 29(2):191–206.

Evans, Geoffrey, Anthony Heath and Mansur Lalljee. 1996. “Measuring left-right and libertarian-authoritarian values in the British electorate.” *British Journal of Sociology* pp. 93–112.

Fieldhouse, E., J. Green, G. Evans, H. Schmitt, C. van der Eijk, J. Mellon and C. Prosser. 2017. “British Election Study 2017 Face-to-Face Post-Election Survey.”.

URL: *https://www.britishelectionstudy.com/data-objects/cross-sectional-data/*

Fieldhouse, Ed, Jane Green, Geoffrey Evans, Hermann Schmitt, Cees van der Eijk, Jon Mellon and Chris Prosser. 2020. “British Election Study Internet Panel Waves 1-19.”.

Fieldhouse, Edward, Jane Green, Geoffrey Evans, Jonathan Mellon, Christopher Prosser, Hermann Schmitt and Cees Van der Eijk. 2021. *Electoral shocks: The volatile voter in a turbulent world*. Oxford University Press.

Ford, Robert and Will Jennings. 2020. “The changing cleavage politics of Western Europe.” *Annual Review of Political Science* 23:295–314.

Forero, Carlos G, Alberto Maydeu-Olivares and David Gallardo-Pujol. 2009. “Factor Analysis with Ordinal Indicators: A Monte Carlo Study Comparing DWLS and ULS Estimation.” *Structural Equation Modeling* 16(4):625–641.

Fowler, Anthony, Seth J Hill, Jeffrey B Lewis, Chris Tausanovitch, Lynn Vavreck and Christopher Warshaw. 2022. “Moderates.” *American Political Science Review* pp. 1–18.

Gingrich, Jane and Silja Häusermann. 2015. “The decline of the working-class vote, the reconfiguration of the welfare support coalition and consequences for the welfare state.” *Journal of European Social Policy* 25(1):50–75.

Glenn, Norval D. 1974. “Aging and conservatism.” *The ANNALS of the American Academy of Political and Social Science* 415(1):176–186.

Glenn, Norval D. 2005. *Cohort analysis*. Vol. 5 Sage.

Grasso, Maria T. 2014. “Age, period and cohort analysis in a comparative context: Political generations and political participation repertoires in Western Europe.” *Electoral Studies* 33:63–76.

Grasso, Maria Teresa, Stephen Farrall, Emily Gray, Colin Hay and Will Jennings. 2019. “Thatcher’s children, Blair’s babies, political socialization and trickle-down value change: An age, period and cohort analysis.” *British Journal of Political Science* 49(1):17–36.

Greenleaf, Eric A. 1992. “Measuring extreme response style.” *Public Opinion Quarterly* 56(3):328–351.

Grimmer, Justin. 2015. “We are all social scientists now: How big data, machine learning, and causal inference work together.” *PS: Political Science & Politics* 48(1):80–83.

Hakhverdian, Armen, Erika Van Elsas, Wouter Van der Brug and Theresa Kuhn. 2013. "Euroscepticism and education: A longitudinal study of 12 EU member states, 1973–2010." *European Union Politics* 14(4):522–541.

Hand, David J. 1996. "Statistics and the theory of measurement." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 159(3):445–473.

Hand, David J. 2016. *Measurement: A very short introduction*. Oxford University Press.

Hanretty, Chris. 2022. "Party system polarization and the effective number of parties." *Electoral Studies* 76:102459.

Hare, Christopher, David A Armstrong, Ryan Bakker, Royce Carroll and Keith T Poole. 2015. "Using Bayesian Aldrich-McKelvey Scaling to Study Citizens' Ideological Preferences and Perceptions." *American Journal of Political Science* 59(3):759–774.

Heath, Anthony, Geoffrey Evans and Jean Martin. 1994. "The measurement of core beliefs and values: The development of balanced socialist/laissez faire and libertarian/authoritarian scales." *British Journal of Political Science* 24(1):115–132.

Hinich, Melvin J and Walker Pollard. 1981. "A new approach to the spatial theory of electoral competition." *American Journal of Political Science* pp. 323–341.

Hjorth, Frederik, Robert Klemmensen, Sara Hobolt, Martin Ejnar Hansen and

Peter Kurrild-Klitgaard. 2015. "Computers, coders, and voters: Comparing automated methods for estimating party positions." *Research & Politics* 2(2):2053168015580476.

Hobolt, Sara B. 2018. "Brexit and the 2017 UK general election." *J. Common Mkt. Stud.* 56:39.

Hobolt, Sara B and Catherine E De Vries. 2015. "Issue entrepreneurship and multiparty competition." *Comparative Political Studies* 48(9):1159–1185.

Hobolt, Sara B, Thomas J Leeper and James Tilley. 2021. "Divided by the vote: Affective polarization in the wake of the Brexit referendum." *British Journal of Political Science* 51(4):1476–1493.

Hooghe, Liesbet and Gary Marks. 2009. "A postfunctionalist theory of European integration: From permissive consensus to constraining dissensus." *British journal of political science* 39(1):1–23.

Hooghe, Liesbet and Gary Marks. 2018. "Cleavage theory meets Europe's crises: Lipset, Rokkan, and the transnational cleavage." *Journal of European Public Policy* 25(1):109–135.

Hooghe, Liesbet, Gary Marks and Carole J Wilson. 2002. "Does left/right structure party positions on European integration?" *Comparative political studies* 35(8):965–989.

Hooghe, Liesbet, Ryan Bakker, Anna Brigevich, Catherine De Vries, Erica Edwards, Gary Marks, Jan Rovny, Marco Steenbergen and Milada Vachudova. 2010. "Reliability and validity of the 2002 and 2006 Chapel Hill expert surveys on party positioning." *European Journal of Political Research* 49(5):687–703.

Hooghe, Marc. 2004. "Political socialization and the future of politics." *Acta Politica* 39(4):331–341.

Hopkins, Daniel J and Gary King. 2010. "Improving anchoring vignettes: Designing surveys to correct interpersonal incomparability." *Public opinion quarterly* 74(2):201–222.

Hotelling, Harold. 1929. "Stability in Competition." *The Economic Journal* 39(153):41–57.

Huber, John and Ronald Inglehart. 1995. "Expert interpretations of party space and party locations in 42 societies." *Party politics* 1(1):73–111.

Hui, C Harry and Harry C Triandis. 1989. "Effects of culture and response format on extreme response style." *Journal of cross-cultural psychology* 20(3):296–309.

Inglehart, Ronald. 1971. "The silent revolution in Europe: Intergenerational change in post-industrial societies." *The American political science review* pp. 991–1017.

Inglehart, Ronald. 1977. *The silent revolution: Changing values and political styles among Western publics*. Princeton University Press.

Inglehart, Ronald. 1990. *Culture Shift in Advanced Industrial Society*. Princeton University Press.

Inglehart, Ronald. 1997. *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies*. Princeton University Press.

Jackman, Simon. 2000. "Estimation and inference are missing data problems: Unifying social science statistics via Bayesian simulation." *Political Analysis* 8(4):307–332.

Jennings, M Kent. 2007. Political socialization. In *The Oxford handbook of political behavior*.

Jolly, Seth, Ryan Bakker, Liesbet Hooghe, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen and Milada Anna Vachudova. 2022. "Chapel Hill expert survey trend file, 1999–2019." *Electoral Studies* 75:102420.

Kenny, David A and Deborah A Kashy. 1992. "Analysis of the multitrait-multimethod matrix by confirmatory factor analysis." *Psychological Bulletin* 112(1):165.

King, Gary, Christopher JL Murray, Joshua A Salomon and Ajay Tandon. 2004. "Enhancing the validity and cross-cultural comparability of measurement in survey research." *American political science review* 98(1):191–207.

King, Gary and Jonathan Wand. 2007. "Comparing incomparable survey responses: Evaluating and selecting anchoring vignettes." *Political Analysis* 15(1):46–66.

Kitschelt, Herbert and Philipp Rehm. 2014. "Occupations as a site of political preference formation." *Comparative Political Studies* 47(12):1670–1706.

Kitschelt, Herbert et al. 1994. *The transformation of European social democracy*. cambridge university press.

Kriesi, Hanspeter, Edgar Grande, Martin Dolezal, Marc Helbling, Dominic Höglinger, Swen Hutter and Bruno Wüest. 2012. *Political conflict in western Europe*. Cambridge University Press.

Kriesi, Hanspeter, Edgar Grande, Romain Lachat, Martin Dolezal, Simon Bornschier and Timotheos Frey. 2006. "Globalization and the transformation of the national political space: Six European countries compared." *European Journal of Political Research* 45(6):921–956.

Kriesi, Hanspeter, Edgar Grande, Romain Lachat, Martin Dolezal, Simon Bornschier, Timotheos Frey et al. 2008. *West European politics in the age of globalization*. Cambridge University Press Cambridge.

Lauderdale, Benjamin. N.d. *Pragmatic Social Measurement*. Unpublished Book Manuscript.

Lauderdale, Benjamin E and Alexander Herzog. 2016. "Measuring political positions from legislative speech." *Political Analysis* 24(3):374–394.

Lauderdale, Benjamin E, Chris Hanretty and Nick Vivyan. 2018. "Decomposing public opinion variation into ideology, idiosyncrasy, and instability." *The Journal of Politics* 80(2):707–712.

Laver, Michael. 1994. "Party policy and cabinet portfolios in Ireland 1992: Results from an expert survey." *Irish Political Studies* 9(1):157–164.

Laver, Michael. 1998*a*. “Party policy in Britain 1997: Results from an expert survey.” *Political Studies* 46(2):336–347.

Laver, Michael. 1998*b*. “Party policy in Ireland 1997 results from an expert survey.” *Irish political studies* 13(1):159–171.

Laver, Michael, Kenneth Benoit and John Garry. 2003. “Extracting policy positions from political texts using words as data.” *American political science review* 97(2):311–331.

Laver, Michael and Peter Mair. 1999. “Party policy and cabinet portfolios in the Netherlands, 1998: Results from an expert survey.” *Acta Politica* 34(1):49–66.

Laver, Michael and W Ben Hunt. 1992. “Policy and party competition.”.

Li, Cheng-Hsien. 2016. “Confirmatory Factor Analysis with Ordinal Data: Comparing Robust Maximum Likelihood and Diagonally Weighted Least Squares.” *Behavior research methods* 48(3):936–949.

Lipset, Seymour Martin. 1960. *Political man: The social bases of politics*.

Mannheim, Karl. 1970. “The problem of generations.” *Psychoanalytic review* 57(3):378–404.

Maydeu-Olivares, Albert and Donna L Coffman. 2006. “Random intercept item factor analysis.” *Psychological methods* 11(4):344.

McKelvey, Richard D. 1976. “Intransitivities in multidimensional voting models and some implications for agenda control.” *Journal of Economic theory* 12(3):472–482.

McKelvey, Richard D. 1979. “General conditions for global intransitivities in formal voting models.” *Econometrica: journal of the Econometric Society* pp. 1085–1112.

Merrill III, Samuel. 1995. “Discriminating between the directional and proximity spatial models of electoral competition.” *Electoral Studies* 14(3):273–287.

Mirowsky, John and Catherine E Ross. 1991. “Eliminating defense and agreement bias from measures of the sense of control: A 2 x 2 index.” *Social Psychology Quarterly* pp. 127–145.

Mislove, Alan, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela and James Rosenquist. 2011. Understanding the demographics of Twitter users. In *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 5 pp. 554–557.

NatCen-Social-Research. 2017. “British Social Attitudes Survey, 2017.”.
URL: *https://www.ukdataservice.ac.uk/*

O’Connor, Brendan, Ramnath Balasubramanyan, Bryan R Routledge and Noah A Smith. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *Fourth international AAAI conference on weblogs and social media*.

Ordeshook, Peter C. 1976. “The spatial theory of elections: A review and a critique.” *Party Identification and Beyond,” John Wiley & Sons, New York*
.

Paulsen, Tine. 2022. "Does proportional representation increase redistribution? Evidence from early 20th century Norwegian municipalities." *Electoral Studies* 78:102494.

Peffley, Mark A and Jon Hurwitz. 1985. "A hierarchical model of attitude constraint." *American Journal of Political Science* pp. 871–890.

Poole, Keith T. 1998. "Recovering a basic space from a set of issue scales." *American Journal of Political Science* pp. 954–993.

Poole, Keith T. 2000. "Nonparametric unfolding of binary choice data." *Political Analysis* 8(3):211–237.

Poole, Keith T and Howard Rosenthal. 1985. "A spatial model for legislative roll call analysis." *American journal of political science* pp. 357–384.

Poole, Keith T and Howard Rosenthal. 1991. "Patterns of congressional voting." *American journal of political science* pp. 228–278.

Poole, Keith T and Howard Rosenthal. 2000. *Congress: A political-economic history of roll call voting*. Oxford University Press on Demand.

Poole, Keith T and Howard Rosenthal. 2001. "D-nominate after 10 years: A comparative update to congress: A political-economic history of roll-call voting." *Legislative Studies Quarterly* pp. 5–29.

Primi, Ricardo, Daniel Santos, Filip De Fruyt and Oliver P John. 2019. "Comparison of classical and modern methods for measuring and correcting for acquiescence." *British Journal of Mathematical and Statistical Psychology* 72(3):447–465.

Primi, Ricardo, Nelson Hauck-Filho, Felipe Valentini, Daniel Santos and Carl F Falk. 2019. Controlling acquiescence bias with multidimensional IRT modeling. In *The Annual Meeting of the Psychometric Society*. Springer pp. 39–52.

Ray, John J. 1979. "Is the acquiescent response style problem not so mythical after all? Some results from a successful balanced F scale." *Journal of Personality Assessment* 43(6):638–643.

Ray, Leonard. 1999. "Measuring party orientations towards European integration: Results from an expert survey." *European Journal of Political Research* 36(2):283–306.

Reither, Eric N, Kenneth C Land, Sun Y Jeon, Daniel A Powers, Ryan K Masters, Hui Zheng, Melissa A Hardy, Katherine M Keyes, Qiang Fu, Heidi A Hanson et al. 2015. "Clarifying hierarchical age–period–cohort models: A rejoinder to Bell and Jones." *Social science & medicine* 145:125–128.

Reither, Eric N, Ryan K Masters, Yang Claire Yang, Daniel A Powers, Hui Zheng and Kenneth C Land. 2015. "Should age-period-cohort studies return to the methodologies of the 1970s?" *Social Science & Medicine* 128:356–365.

Rekker, Roderik. 2016. "The lasting impact of adolescence on left-right identification: Cohort replacement and intracohort change in associations with issue attitudes." *Electoral Studies* 44:120–131.

Rodebaugh, Thomas L, Carol M Woods and Richard G Heimberg. 2007. "The Reverse of Social Anxiety is not Always the Opposite: The Reverse-Scored

Items of the Social Interaction Anxiety Scale do not Belong." *Behavior Therapy* 38(2):192–206.

Rosseel, Yves. 2012. "Lavaan: An R package for Structural Equation Modeling and More. Version 0.5–12 (BETA)." *Journal of statistical software* 48(2):1–36.

Savalei, Victoria and Carl F Falk. 2014. "Recovering substantive factor loadings in the presence of acquiescence bias: A comparison of three approaches." *Multivariate behavioral research* 49(5):407–424.

Shorrocks, Rosalind. 2018. "Cohort change in political gender gaps in Europe and Canada: The role of modernization." *Politics & Society* 46(2):135–175.

Slapin, Jonathan B and Sven-Oliver Proksch. 2008. "A scaling model for estimating time-series party positions from texts." *American Journal of Political Science* 52(3):705–722.

Sobolewska, Maria and Robert Ford. 2020. *Brexitland: Identity, diversity and the reshaping of British politics*. Cambridge University Press.

Spirling, Arthur and Iain McLean. 2007. "UK OC OK? Interpreting optimal classification scores for the UK House of Commons." *Political Analysis* 15(1):85–96.

Spitzer, Alan B. 1973. "The historical problem of generations." *The American Historical Review* 78(5):1353–1385.

Stubager, Rune. 2008. "Education effects on authoritarian–libertarian values:

a question of socialization1." *The British journal of sociology* 59(2):327–350.

Surridge, Paula. 2016. "Education and liberalism: pursuing the link." *Oxford Review of Education* 42(2):146–164.

Surridge, Paula. 2018. "The fragmentation of the electoral left since 2010." *Renewal: a Journal of Labour Politics* 26(4):69–78.

Surridge, Paula. 2019. "The difficult truth for liberals: Labour must win back social conservatives.".
URL: *https://www.theguardian.com/commentisfree/2019/dec/06/difficult-truth-labour-social-conservatives*

Swain, Scott D, Danny Weathers and Ronald W Niedrich. 2008. "Assessing Three Sources of Misresponse to Reversed Likert Items." *Journal of Marketing Research* 45(1):116–131.

Tilley, James and Geoffrey Evans. 2014. "Ageing and generational effects on vote choice: Combining cross-sectional and panel data to estimate APC effects." *Electoral studies* 33:19–27.

Tilley, James R. 2005. "Research note: Libertarian-authoritarian value change in Britain, 1974–2001." *Political Studies* 53(2):442–453.

Tumasjan, Andranik, Timm Sprenger, Philipp Sandner and Isabell Welpe. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 4 pp. 178–185.

Wagner, Sarah. 2021. "Euroscepticism as a radical left party strategy for success." *Party Politics* p. 13540688211038917.

Wand, Jonathan, Gary King and Olivia Lau. 2011. "Anchors: Software for anchoring vignette data." *Journal of Statistical Software* 42:1–25.

Ware Jr, John E. 1978. "Effects of acquiescent response set on patient satisfaction ratings." *Medical care* pp. 327–336.

Watson, Dorothy. 1992. "Correcting for acquiescent response bias in the absence of a balanced scale: An application to class consciousness." *Sociological Methods & Research* 21(1):52–88.

Webb, Paul and Tim Bale. 2021. *The modern British party system*. Oxford University Press.

Wetzel, Eunike, Claus H Carstensen and Jan R Böhnke. 2013. "Consistency of extreme response style and non-extreme response style across traits." *Journal of Research in Personality* 47(2):178–189.

Winkler, John D, David E Kanouse and John E Ware. 1982. "Controlling for acquiescence response set in scale development." *Journal of Applied Psychology* 67(5):555–561.

Yang, Yang and Kenneth C Land. 2006. "A mixed models approach to the age-period-cohort analysis of repeated cross-section surveys, with an application to data on trends in verbal test scores." *Sociological methodology* 36(1):75–97.

Yang, Yang and Kenneth C Land. 2008. "Age–period–cohort analysis of repeated cross-section surveys: fixed or random effects?" *Sociological methods & research* 36(3):297–326.

Yang, Yang and Kenneth C Land. 2013. *Age-period-cohort analysis: New models, methods, and empirical applications*. CRC press.

Zaller, John and Stanley Feldman. 1992. "A simple theory of the survey response: Answering questions versus revealing preferences." *American journal of political science* pp. 579–616.

Zur, Roi. 2019. "Stuck in the middle: Ideology, valence and the electoral failures of centrist parties." *British Journal of Political Science* pp. 1–18.

Zur, Roi. 2021. "The Multidimensional Disadvantages of Centrist Parties in Western Europe." *Political Behavior* pp. 1–23.

Chapter 7

Essay 1 Appendix

Since survey weights are used throughout unless otherwise stated, the BES respondents without survey weights were not included all parts of the following analysis.

7.1 Appendix A: Variable Codings

7.1.1 Education Recodes

Table 7.1: BSA Education Recode

Original Coding	New Coding
Postgraduate degree	Postgrad
First degree	Undergrad
Higher educ below degree	A-level/equiv
A level or equiv	A-level/equiv
O level or equiv	GCSE/equiv
CSE or equiv	GCSE/equiv
Foreign or other	Missing
No qualification	No Qualification

Table 7.2: BES Education Recode

Original Coding	New Coding
No qualifications	No qualification
Below GCSE	No qualification
GCSE	GCSE/equiv
A-level	A-level/equiv
Undergraduate	Undergrad
Postgrad	Postgrad

7.2 Appendix B: Demonstration

7.2.1 Regression results

Table 7.3: BSA and BES Scales Regressed on Education

	BSA Left-Right	BES Left-Right	BSA Lib-Auth	BES Lib-Auth
Intercept	1.31^{***}	1.65^{***}	2.82^{***}	2.26^{***}
	(0.04)	(0.02)	(0.03)	(0.02)
GCSE/Equiv	0.22^{***}	0.01	-0.23^{***}	0.09
	(0.04)	(0.07)	(0.04)	(0.07)
A-level/Equiv	0.30^{***}	-0.12^{**}	-0.32^{***}	-0.23^{***}
	(0.06)	(0.04)	(0.05)	(0.04)
Undergrad	0.27^{***}	0.02	-0.68^{***}	-0.40^{***}
	(0.05)	(0.04)	(0.04)	(0.03)
Postgrad	0.24^{***}	-0.00	-0.84^{***}	-0.70^{***}
	(0.06)	(0.05)	(0.05)	(0.05)
R^2	0.01	0.01	0.13	0.12
Adj. R^2	0.01	0.00	0.13	0.12
Num. obs.	3123	1806	3125	1931

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

7.2.2 Demonstration Robustness

Indicators common to both datasets:

- **Ind1:** There is one law for the rich and one for the poor
- **Ind2:** Young people today don't have enough respect for traditional British values
- **Ind3:** Censorship of films and magazines is necessary to uphold moral standards
- **Ind4:** For some crimes, the death penalty is the most appropriate sentence
- **Ind5:** People who break the law should be given stiffer sentences

Table 7.4: Regression of Survey Membership on Common Indicators

	OLS	Logit	Probit
Intercept	0.56	0.25	0.16
	(0.03)	(0.11)	(0.07)
Ind1	0.01	0.05	0.03
	(0.01)	(0.03)	(0.02)
Ind2	−0.00	−0.02	−0.01
	(0.01)	(0.03)	(0.02)
Ind3	−0.01	−0.05	−0.03
	(0.01)	(0.03)	(0.02)
Ind4	−0.02	−0.06	−0.04
	(0.01)	(0.02)	(0.01)
Ind5	0.04	0.16	0.10
	(0.01)	(0.04)	(0.02)
R^2	0.01		
Adj. R^2	0.00		
Num. obs.	5170	5170	5170
AIC		7039.76	7040.06
BIC		7079.07	7079.37
Log Likelihood		−3513.88	−3514.03
Deviance		6841.30	6841.57

Table 7.5: Regression of Scales on Survey Month

	Left-Right	Lib-Auth
Intercept	1.64	2.04
	(0.02)	(0.02)
Aug	−0.03	0.00
	(0.03)	(0.03)
Sep	−0.04	0.04
	(0.04)	(0.05)
R^2	0.00	0.00
Adj. R^2	−0.00	−0.00
Num. obs.	1789	1914

7.3 Appendix C: Unit Intercept Confirmatory Factor Analysis

The standard confirmatory factor analysis model is given in its linear form as:

$$x_{ij} = \lambda_{j1}\eta_{i1} + ... + \lambda_{jm}\eta_{im} + \epsilon_{ij} \tag{7.1}$$

Which is the common factor model discussed in the main body of the paper. The assumptions of this model are:

1. The means of the common factors are 0
2. The common factors are normally distributed
3. The means of the unique components are 0
4. The unique components are normally distributed
5. The unique components are uncorrelated with the common factors
6. The unique components are uncorrelated with each other

The model can be expressed in a more compact matrix form:

$$\mathbf{x} = \boldsymbol{\Lambda\eta} + \boldsymbol{\epsilon} \tag{7.2}$$

Where $\mathbf{x}$ is the $p \times 1$ vector of indicators, $\boldsymbol{\Lambda}$ is the $p \times m$ matrix of factor loadings, $\boldsymbol{\eta}$ is the $m \times 1$ vector of factor scores, and $\boldsymbol{\epsilon}$ is the $p \times 1$ vector of unique components. In turn, we can further express the model in terms of covariance matrices:

$$\mathbf{\Sigma} = \mathbf{\Lambda\Psi\Lambda}' + \mathbf{\Theta_\epsilon} \tag{7.3}$$

Where $\mathbf{\Sigma}$ is the $p \times p$ variance-covariance matrix of the indicators, $\boldsymbol{\psi}$ is the $m \times m$ variance-covariance matrix of the common factors, and $\mathbf{\Theta_\epsilon}$ is the $p \times p$ variance-covariance matrix of unique components which by assumption 6 is a diagonal matrix. When estimated with maximum likelihood (ML), assuming no (further) restrictions are placed on the latent variables means the discrepancy function minimised is:

$$F_{ML} = ln|\mathbf{S}| - ln|\mathbf{\Sigma}| + trace(\boldsymbol{S\Sigma^{-1}}) - p \tag{7.4}$$

Where $\mathbf{S}$ is the model-implied variance-covariance matrix and p is the number of indicators.

7.3.1 Person Intercept CFA

As discussed in the main body of the paper, unit intercept CFA is given by

$$x_{ij} = \lambda_{jc}\eta_{ic} + 1\eta_{ia} + \epsilon_{ij} \tag{7.5}$$

Where factor c would be the common factor and factor a would be the person intercept factor. Maydeu-Olivares and Coffman introduce three further assumptions for this model relative to regular CFA, which deserve discussion. The first two are:

7. The mean of the unit-intercepts is 0

8. The unit intercepts are uncorrelated with the unique components

Thus far, these are simply assumptions 1 and 5 repackaged for treating the unit-intercept factor separately. However, Maydeu-Olivares and Coffman make a further assumption:

9. The unit intercepts are uncorrelated with the common factor(s)

This assumption is explained in part by Maydeu-Olivares and COffman's choice of language for the model. As discussed in the main body of the paper, they specifically refer to the model as a *random-intercept* model and clearly are aiming to draw a parallel with multilevel regression modelling in their description of the unit-intercept confirmatory factor analysis model (indeed, their formulae reflect this too). However, as discussed in the main body of the paper, this comparison is not only unnecessary but arguably limits the utility of the model. I therefore drop this assumption and utilise the terminology person intercept instead.

To identify the scales of the common factors in the person intercept model, the variances of the common factors are constrained to 1 (as opposed to their first loading being constrained to 1). By contrast, the variance of the unit-intercept is freely estimated. The important feature of such a model is that the loading of the unit-intercept factor is constrained across indicators. A method of creating such an intercept while constraining the unit-intercept variance to 1 would simply be to apply equality constraints to the unit-intercept loadings, such that they were equal across all indicators:

$$x_{ij} = \lambda_{jc}\eta_{ic} + \lambda_a\eta_{ia} + \epsilon_{ij} \tag{7.6}$$

As stated in the main body of the paper, the difference between (7.6) and

(7.5) is that instead of a loading of '1' on η_{ia}, there is now a freely estimated loading lacking a 'j' subscript as it is common to all indicators.

7.3.2 Ordinal Confirmatory Factor Analysis

One potential flaw of the person intercept CFA model is that it does not fully take into account the ordinal nature of the indicator variables typical for Likert scales. In ordinal CFA, the relationship between the latent variables and the observed categories are assumed to exist via a threshold relationship:

$$x_{ij}^* = \lambda_{j1}\eta_{i1} + ... + \lambda_{jm}\eta_{im} + \epsilon_{ij} \tag{7.7}$$
$$x_{ij} = K \quad if \quad \tau_{jk} < x_{ij}^* < \tau_{jk+1}$$

Where x_{ij}^* is the latent variable underlying x_{ij}, K is one of the t values x_{ij} can take on, τ_{jk} is the kth threshold for indicator j, $\tau_{j0} = -\infty$ and $\tau_{jt} = \infty$.

Ordinal CFA makes similar assumptions to continuous CFA:

1. The means of the common factors are 0
2. The common factors are normally distributed
3. The means of the unique components are 0
4. The unique components are normally distributed
5. The unique components are uncorrelated with the common factors
6. The unique components are uncorrelated with each other

It follows that x^*_{ij} is normally distributed with mean 0 and the covariance matrix:

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Psi}\mathbf{\Lambda}' + \mathbf{\Theta}_{\boldsymbol{\epsilon}} \tag{7.8}$$

To identify the variances of the unique components, we set

$$\mathbf{\Theta}_{\boldsymbol{\epsilon}} = \mathbf{I} - diag(\mathbf{\Lambda}\mathbf{\Psi}\mathbf{\Lambda}') \tag{7.9}$$

such that the covariance matrix becomes a correlation matrix **P**.

Ordinal CFA is often estimated in a three-step procedure. First, the thresholds are estimated alone using maximum likelihood. The thresholds are often estimated by the corresponding percentage of respondents in each category of the ordinal variable. Second, the polychoric correlation matrix of the observed indicators is estimated via maximum likelihood. Third, assuming no restrictions are placed on the thresholds, a least squares discrepancy function based on the polychoric correlations can be used:

$$F_{LS} = (\hat{\boldsymbol{p}} - \boldsymbol{p}(\boldsymbol{\theta}))'\boldsymbol{V}(\hat{\boldsymbol{p}} - \boldsymbol{p}(\boldsymbol{\theta})) \tag{7.10}$$

Where $\hat{p}$ is the polychoric correlation matrix estimated in the second step, $p(\theta)$ is the model-implied correlation matrix, $\boldsymbol{\theta}$ represents the parameters of the model, and $\boldsymbol{V}$ is a weighting matrix. The choice of weighting matrix determines the exact estimation method being used. If $\hat{\mathbf{\Gamma}}$ is an estimate of the asymptotic covariance matrix of estimated polychoric correlations, then:

- Weighted Least Squares (WLS): $\boldsymbol{V} = \hat{\mathbf{\Gamma}}$

- Diagonally Weighted Least Squares (DWLS): $\mathbf{V} = diag(\hat{\mathbf{\Gamma}})^{-1/2}$
- Unweighted Least Squares (ULS): $\boldsymbol{V} = \boldsymbol{I}$

Similarly to regular CFA, implementing the unit intercept in ordinal CFA is relatively straightforward. We can either set the loadings of the unit-intercept factor to 1 while freeing its variance:

$$x^*_{ij} = \lambda_{jc}\eta_{ic} + 1\eta_{ia} + \epsilon_{ij} \tag{7.11}$$

Or alternatively we can can constrain its variance to 1 while constraining the loadings to be equal but freely estimating their value:

$$x^*_{ij} = \lambda_{jc}\eta_{ic} + \lambda_a\eta_{ia} + \epsilon_{ij} \tag{7.12}$$

Continuing with the convention established above, for the remainder of this paper I refer to these models as (7.11) OCFA1 and (7.12) OCFA2.

7.4 Appendix D: Correction

7.4.1 Identifying Scale CFA

Please note that the empathy scale indicators have been shortened. So ‘empathy1’ is ‘em1’, and so on.

Table 7.6: Zero CFA Check

	Model	
	Estimate	Std. Err.
	Loadings	
Zero		
zero1	0.41	0.01
zero4	0.49	0.02
zero5	-0.53	0.02
zero7	0.61	0.01
zero9	-0.59	0.01
zero11	-0.58	0.02
Acq		
zero1	1.00+	
zero4	1.00+	
zero5	1.00+	
zero7	1.00+	
zero9	1.00+	
zero11	1.00+	
	Latent Variances	
Zero	1.00+	
Acq	0.10	0.00
	Fit Indices	
χ^2(df)	253.63	
CFI	0.96	
TLI	0.93	
RMSEA	0.07	
Scaled χ^2(df)	181.16(8)	

+Fixed parameter

7.4.2 CFA Results

Zero-Sum CFA Results

Table 7.8: Zero-Sum CFA1

	Model	
	Estimate	Std. Err.

	Loadings	
Z		
zero7	0.58	0.01
zero1	0.40	0.01
zero4	0.48	0.02
zero11	-0.59	0.02
zero5	-0.55	0.02
zero9	-0.60	0.01
LeftCorrected		
lr1	0.81	0.02
lr2	0.70	0.01
lr3	0.81	0.01
lr4	0.83	0.01
lr5	0.61	0.01
AuthCorrected		
al1	0.85	0.01
al2	0.99	0.02
al3	0.73	0.01
al4	0.56	0.02
al5	0.72	0.01
Acq		
zero7	1.00^{+}	
zero1	1.00^{+}	
zero4	1.00^{+}	

zero11	1.00^{+}	
zero5	1.00^{+}	
zero9	1.00^{+}	
lr1	1.00^{+}	
lr2	1.00^{+}	
lr3	1.00^{+}	
lr4	1.00^{+}	
lr5	1.00^{+}	
al1	1.00^{+}	
al2	1.00^{+}	
al3	1.00^{+}	
al4	1.00^{+}	
al5	1.00^{+}	
	Latent Variances	
Z	1.00^{+}	
LeftCorrected	1.00^{+}	
AuthCorrected	1.00^{+}	
Acq	0.08	0.00
	Fit Indices	
χ^2(df)	3134.95	
CFI	0.90	
TLI	0.89	
RMSEA	0.07	
Scaled χ^2(df)	2641.83(103)	

+Fixed parameter

Table 7.9: Zero-Sum CFA2

	Model	
	Estimate	Std. Err.
	Loadings	
Z		
zero7	0.67	0.05
zero1	0.50	0.05
zero4	0.60	0.05
zero11	-0.49	0.05
zero5	-0.44	0.05
zero9	-0.50	0.05
LeftCorrected		
lr1	0.83	0.02
lr2	0.74	0.02
lr3	0.85	0.01
lr4	0.87	0.02
lr5	0.65	0.02
AuthCorrected		
al1	0.87	0.02
al2	1.02	0.03
al3	0.75	0.02
al4	0.57	0.02

al5	0.74	0.02
Acq		
zero7	0.32	0.02
zero1	0.32	0.02
zero4	0.32	0.02
zero11	0.32	0.02
zero5	0.32	0.02
zero9	0.32	0.02
lr1	0.32	0.02
lr2	0.32	0.02
lr3	0.32	0.02
lr4	0.32	0.02
lr5	0.32	0.02
al1	0.32	0.02
al2	0.32	0.02
al3	0.32	0.02
al4	0.32	0.02
al5	0.32	0.02
	Latent Variances	
Z	1.00^{+}	
LeftCorrected	1.00^{+}	
AuthCorrected	1.00^{+}	
Acq	1.00^{+}	
	Fit Indices	

χ^2(df)	2705.09
CFI	0.92
TLI	0.90
RMSEA	0.07
Scaled χ^2(df)	2307.25(97)

$^{+}$Fixed parameter

Table 7.10: Zero-Sum OCFA1

	Model	
	Estimate	Std. Err.
	Loadings	
Z		
zero7	0.70	0.01
zero1	0.45	0.01
zero4	0.52	0.01
zero11	-0.57	0.01
zero5	-0.59	0.01
zero9	-0.67	0.01
LeftCorrected		
lr1	0.67	0.01
lr2	0.81	0.01
lr3	0.83	0.01
lr4	0.81	0.01
lr5	0.67	0.01

AuthCorrected		
al1	0.80	0.01
al2	0.70	0.01
al3	0.75	0.01
al4	0.50	0.01
al5	0.79	0.01
Acq		
zero7	1.00^{+}	
zero1	1.00^{+}	
zero4	1.00^{+}	
zero11	1.00^{+}	
zero5	1.00^{+}	
zero9	1.00^{+}	
lr1	1.00^{+}	
lr2	1.00^{+}	
lr3	1.00^{+}	
lr4	1.00^{+}	
lr5	1.00^{+}	
al1	1.00^{+}	
al2	1.00^{+}	
al3	1.00^{+}	
al4	1.00^{+}	
al5	1.00^{+}	
	Latent Variances	

Z	1.00^{+}	
LeftCorrected	1.00^{+}	
AuthCorrected	1.00^{+}	
Acq	0.05	0.00
	Fit Indices	
χ^2(df)	6344.31	
CFI	0.90	
TLI	0.92	
RMSEA	0.08	
Scaled χ^2(df)	1855.01(167)	

$^{+}$Fixed parameter

Table 7.11: Zero-Sum OCFA2

	Model	
	Estimate	Std. Err.
	Loadings	
Z		
zero7	0.69	0.03
zero1	0.49	0.03
zero4	0.61	0.03
zero11	-0.54	0.03
zero5	-0.55	0.03
zero9	-0.70	0.03
LeftCorrected		

lr1	0.78	0.01
lr2	0.86	0.01
lr3	0.90	0.01
lr4	0.85	0.01
lr5	0.68	0.01
AuthCorrected		
al1	0.81	0.01
al2	0.74	0.01
al3	0.76	0.01
al4	0.49	0.01
al5	0.78	0.01
Acq		
zero7	0.35	0.01
zero1	0.35	0.01
zero4	0.35	0.01
zero11	0.35	0.01
zero5	0.35	0.01
zero9	0.35	0.01
lr1	0.35	0.01
lr2	0.35	0.01
lr3	0.35	0.01
lr4	0.35	0.01
lr5	0.35	0.01
al1	0.35	0.01

al2	0.35	0.01
al3	0.35	0.01
al4	0.35	0.01
al5	0.35	0.01
	Latent Variances	
Z	1.00^{+}	
LeftCorrected	1.00^{+}	
AuthCorrected	1.00^{+}	
Acq	1.00^{+}	
	Fit Indices	
χ^2(df)	3190.44	
CFI	0.95	
TLI	0.94	
RMSEA	0.07	
Scaled χ^2(df)	3933.42(97)	

$^{+}$Fixed parameter

Table 7.7: Empathy CFA Check

	Model	
	Estimate	Std. Err.
	Loadings	
Empathy		
em1	0.30	0.01
em2	0.32	0.01
em3	0.30	0.01
em4	-0.34	0.01
em5	0.29	0.01
em6	0.25	0.01
em7	-0.45	0.01
em8	-0.48	0.01
em9	-0.39	0.01
em10	-0.47	0.01
Acq		
em1	1.00^{+}	
em2	1.00^{+}	
em3	1.00^{+}	
em4	1.00^{+}	
em5	1.00^{+}	
em6	1.00^{+}	
em7	1.00^{+}	
em8	1.00^{+}	
em9	1.00^{+}	
em10	1.00^{+}	
	Latent Variances	
Empathy	1.00^{+}	
Acq	0.05	0.00
	Fit Indices	
χ^2(df)	2268.03	
CFI	0.87	
TLI	0.83	
RMSEA	0.12	
Scaled χ^2(df)	1513.06(34)	

$^{+}$Fixed parameter

Empathy CFA Results

Table 7.12: Empathy CFA1

	Model	
	Estimate	Std. Err.
	Loadings	
E		
em1	0.29	0.01
em2	0.31	0.01
em3	0.29	0.01
em4	-0.34	0.01
em5	0.28	0.01
em6	0.25	0.01
em7	-0.46	0.01
em8	-0.49	0.01
em9	-0.40	0.01
em10	-0.48	0.01
LeftCorrected		
lr1	0.83	0.02
lr2	0.70	0.01
lr3	0.84	0.01
lr4	0.82	0.01
lr5	0.65	0.02
AuthCorrected		
al1	0.90	0.02
al2	1.03	0.02

al3	0.77	0.02
al4	0.63	0.02
al5	0.77	0.01
<u>Acq</u>		
em1	1.00^+	
em2	1.00^+	
em3	1.00^+	
em4	1.00^+	
em5	1.00^+	
em6	1.00^+	
em7	1.00^+	
em8	1.00^+	
em9	1.00^+	
em10	1.00^+	
lr1	1.00^+	
lr2	1.00^+	
lr3	1.00^+	
lr4	1.00^+	
lr5	1.00^+	
al1	1.00^+	
al2	1.00^+	
al3	1.00^+	
al4	1.00^+	
al5	1.00^+	

	Latent Variances	
E	1.00^{+}	
LeftCorrected	1.00^{+}	
AuthCorrected	1.00^{+}	
Acq	0.04	0.00
	Fit Indices	
χ^2(df)	4227.66	
CFI	0.89	
TLI	0.87	
RMSEA	0.07	
Scaled χ^2(df)	3470.77(169)	

$^{+}$Fixed parameter

Table 7.13: Empathy CFA2

	Model	
	Estimate	Std. Err.
	Loadings	
E		
em1	0.13	0.05
em2	0.16	0.05
em3	0.14	0.05
em4	-0.49	0.05
em5	0.13	0.05
em6	0.10	0.05

em7	-0.61	0.05
em8	-0.64	0.05
em9	-0.55	0.05
em10	-0.63	0.05
<u>LeftCorrected</u>		
lr1	0.85	0.02
lr2	0.72	0.02
lr3	0.85	0.01
lr4	0.83	0.02
lr5	0.67	0.02
<u>AuthCorrected</u>		
al1	0.92	0.02
al2	1.08	0.03
al3	0.79	0.02
al4	0.65	0.02
al5	0.80	0.02
<u>Acq</u>		
em1	0.27	0.03
em2	0.27	0.03
em3	0.27	0.03
em4	0.27	0.03
em5	0.27	0.03
em6	0.27	0.03
em7	0.27	0.03

em8	0.27	0.03
em9	0.27	0.03
em10	0.27	0.03
lr1	0.27	0.03
lr2	0.27	0.03
lr3	0.27	0.03
lr4	0.27	0.03
lr5	0.27	0.03
al1	0.27	0.03
al2	0.27	0.03
al3	0.27	0.03
al4	0.27	0.03
al5	0.27	0.03
	Latent Variances	
E	1.00^{+}	
LeftCorrected	1.00^{+}	
AuthCorrected	1.00^{+}	
Acq	1.00^{+}	
	Fit Indices	
χ^2(df)	3846.21	
CFI	0.90	
TLI	0.88	
RMSEA	0.07	
Scaled χ^2(df)	3164.99(163)	

+Fixed parameter

Table 7.14: Empathy OCFA1

	Model	
	Estimate	Std. Err.
	Loadings	
E		
em1	0.61	0.01
em2	0.69	0.01
em3	0.66	0.01
em4	-0.53	0.01
em5	0.63	0.01
em6	0.48	0.01
em7	-0.77	0.01
em8	-0.78	0.01
em9	-0.56	0.01
em10	-0.78	0.01
LeftCorrected		
lr1	0.68	0.01
lr2	0.80	0.01
lr3	0.85	0.01
lr4	0.81	0.01
lr5	0.68	0.01
AuthCorrected		

al1	0.81	0.01
al2	0.73	0.01
al3	0.76	0.01
al4	0.50	0.01
al5	0.81	0.01
Acq		
em1	1.00^{+}	
em2	1.00^{+}	
em3	1.00^{+}	
em4	1.00^{+}	
em5	1.00^{+}	
em6	1.00^{+}	
em7	1.00^{+}	
em8	1.00^{+}	
em9	1.00^{+}	
em10	1.00^{+}	
lr1	1.00^{+}	
lr2	1.00^{+}	
lr3	1.00^{+}	
lr4	1.00^{+}	
lr5	1.00^{+}	
al1	1.00^{+}	
al2	1.00^{+}	
al3	1.00^{+}	

al4	1.00^{+}	
al5	1.00^{+}	
	Latent Variances	
E	1.00^{+}	
LeftCorrected	1.00^{+}	
AuthCorrected	1.00^{+}	
Acq	0.04	0.00
	Fit Indices	
χ^2(df)	7500.96	
CFI	0.90	
TLI	0.92	
RMSEA	0.08	
Scaled χ^2(df)	1682.38(239)	

$^{+}$Fixed parameter

Table 7.15: Empathy OCFA2

	Model	
	Estimate	Std. Err.
	Loadings	
E		
em1	0.65	0.04
em2	0.73	0.04
em3	0.71	0.04
em4	-0.49	0.04

em5	0.68	0.04
em6	0.53	0.04
em7	-0.73	0.04
em8	-0.74	0.04
em9	-0.52	0.04
em10	-0.74	0.04
<u>LeftCorrected</u>		
lr1	0.79	0.01
lr2	0.88	0.01
lr3	0.95	0.01
lr4	0.89	0.01
lr5	0.73	0.01
<u>AuthCorrected</u>		
al1	0.87	0.01
al2	0.75	0.01
al3	0.83	0.01
al4	0.59	0.01
al5	0.87	0.01
<u>Acq</u>		
em1	0.33	0.01
em2	0.33	0.01
em3	0.33	0.01
em4	0.33	0.01
em5	0.33	0.01

em6	0.33	0.01
em7	0.33	0.01
em8	0.33	0.01
em9	0.33	0.01
em10	0.33	0.01
lr1	0.33	0.01
lr2	0.33	0.01
lr3	0.33	0.01
lr4	0.33	0.01
lr5	0.33	0.01
al1	0.33	0.01
al2	0.33	0.01
al3	0.33	0.01
al4	0.33	0.01
al5	0.33	0.01
	Latent Variances	
E	1.00^{+}	
LeftCorrected	1.00^{+}	
AuthCorrected	1.00^{+}	
Acq	1.00^{+}	
	Fit Indices	
χ^2(df)	4465.90	
CFI	0.94	
TLI	0.93	

RMSEA	0.08
Scaled χ^2(df)	4111.71(163)

$^+$Fixed parameter

7.4.3 CFA Measure Correlations

Table 7.16: Zero-Sum Left-Right

	CFA1	OCFA1	CFA2	OCFA2
CFA1				
OCFA1	0.991			
CFA2	0.987	0.974		
OCFA2	0.984	0.984	0.984	

Table 7.17: Empathy Left-Right

	CFA1	OCFA1	CFA2	OCFA2
CFA1				
OCFA1	0.982			
CFA2	0.985	0.958		
OCFA2	0.976	0.944	0.981	

Table 7.18: Zero-Sum Lib-Auth

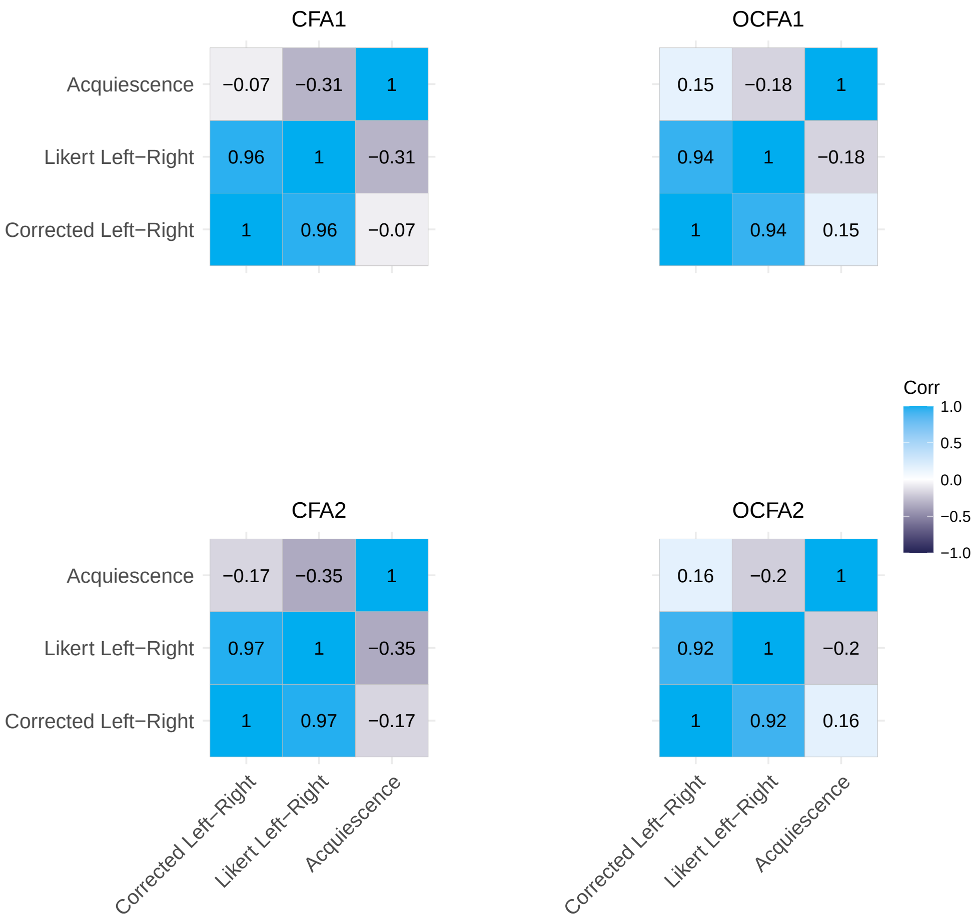
	CFA1	OCFA1	CFA2	OCFA2
CFA1				
OCFA1	0.986			
CFA2	0.989	0.973		
OCFA2	0.979	0.986	0.982	

Table 7.19: Empathy Lib-Auth

	CFA1	OCFA1	CFA2	OCFA2
CFA1				
OCFA1	0.982			
CFA2	0.987	0.963		
OCFA2	0.973	0.941	0.978	

7.4.4 CFA vs Likert

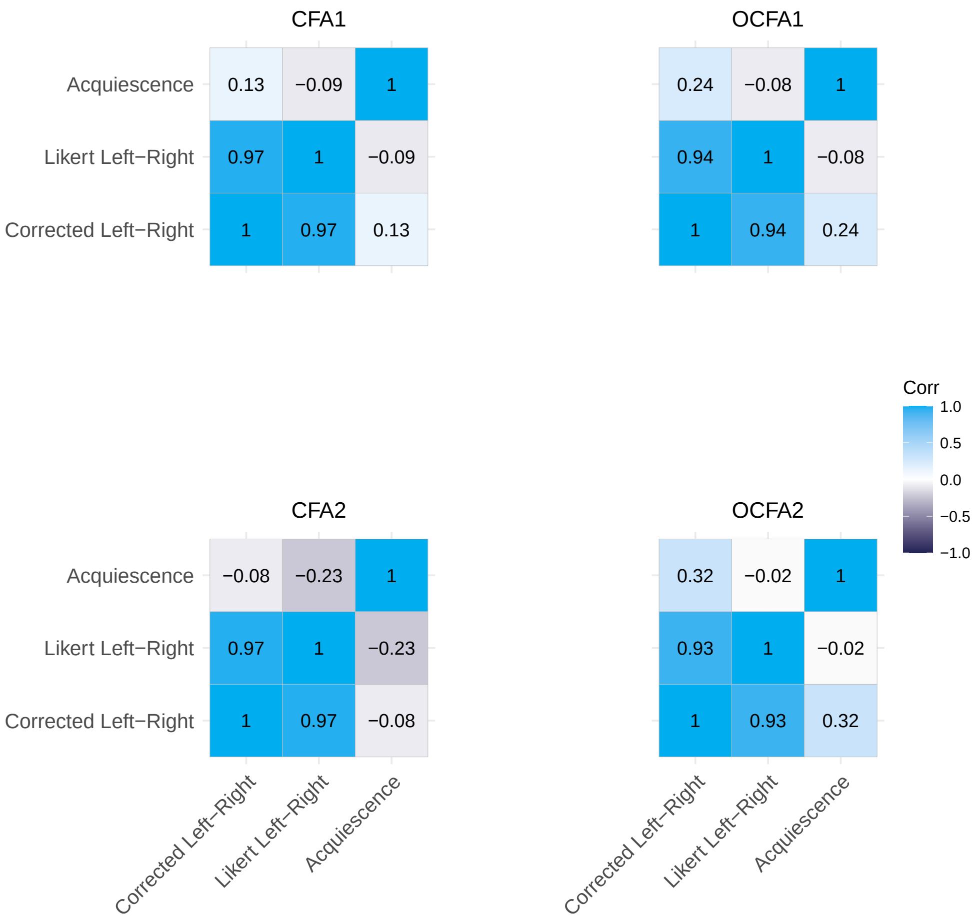
Figure 7.1: Zero-Sum Left-Right Correlations



These correlation matrices compare the corrected left-right scale, left-right likert scale, and recovered acquiescence factor for the zero-sum subset of BESIP.

7.4.5 Marginal Distributions

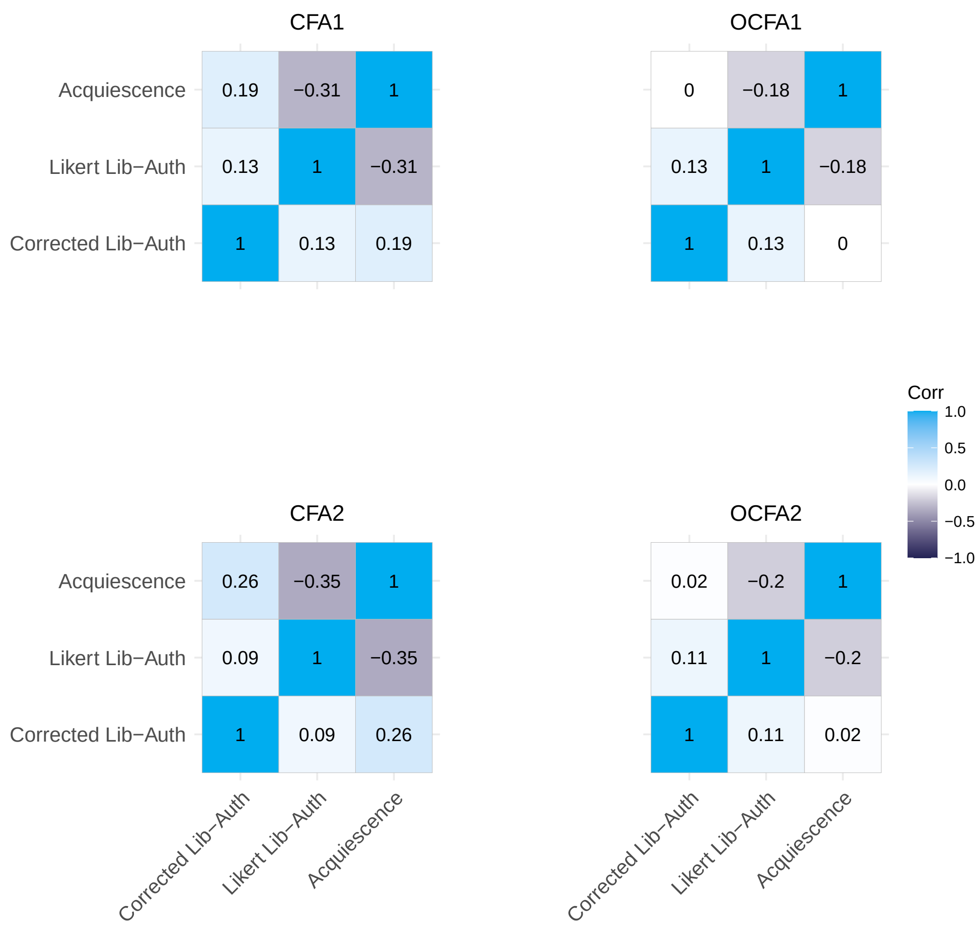
Figure 7.2: Empathy Left-Right Correlations



These correlation matrices compare the corrected left-right scale, left-right likert scale, and recovered acquiescence factor for the empathy subset of BESIP.

7.4.6 Regression Results

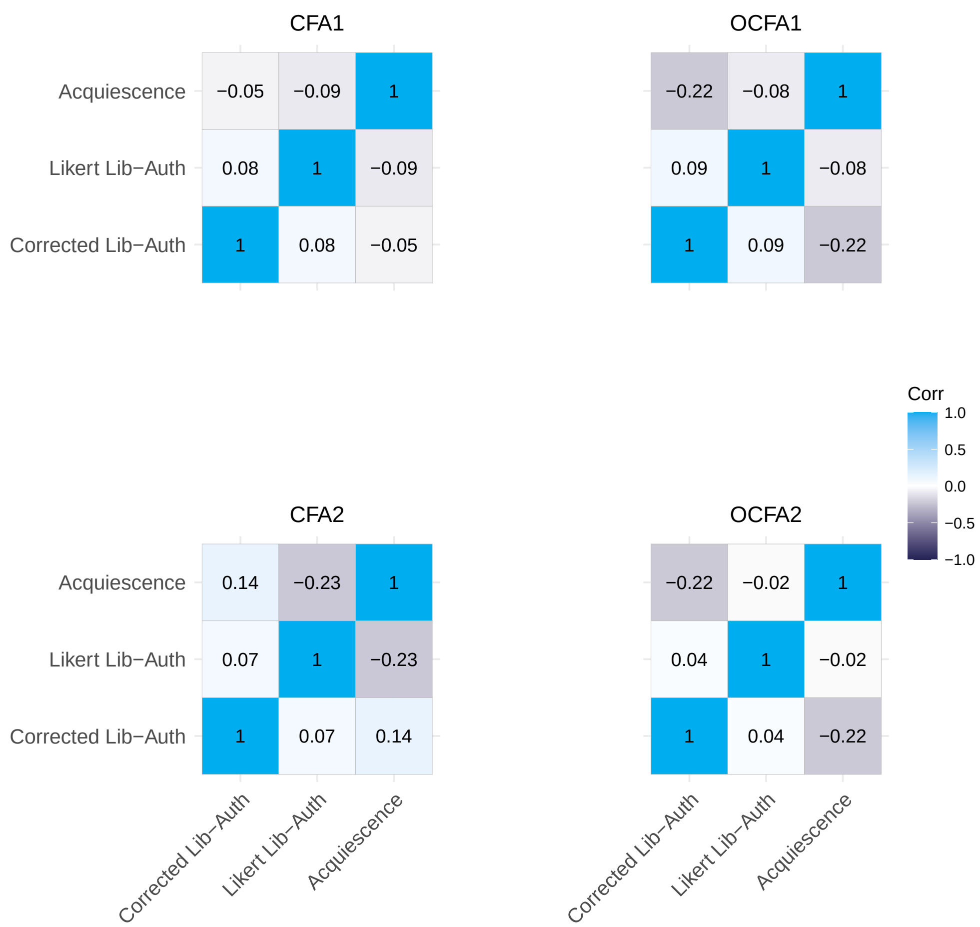
Figure 7.3: Zero-Sum Libertarian-Authoritarian Correlations



These correlation matrices compare the corrected libertarian-authoritarian scale, libertarian-authoritarian likert scale, and recovered acquiescence factor for the zero-sum subset of BESIP.

7.4.7 Education Recode Regression Results

Figure 7.4: Empathy Libertarian-Authoritarian Correlations



These correlation matrices compare the corrected libertarian-authoritarian scale, libertarian-authoritarian likert scale, and recovered acquiescence factor for the empathy subset of BESIP.

Figure 7.5: Density Plots of Left-Right Factors from Correction Models

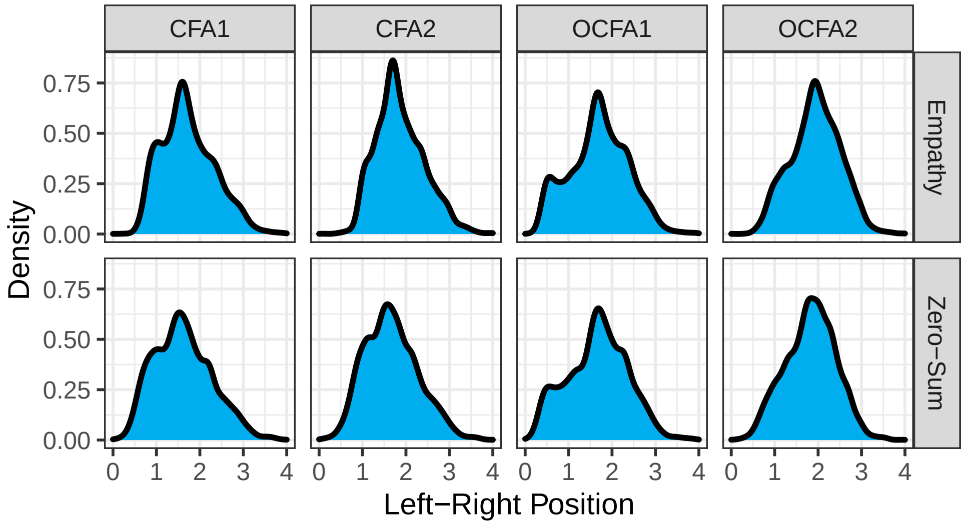


Figure 7.6: Density Plots of Lib-Auth Factors from Correction Models

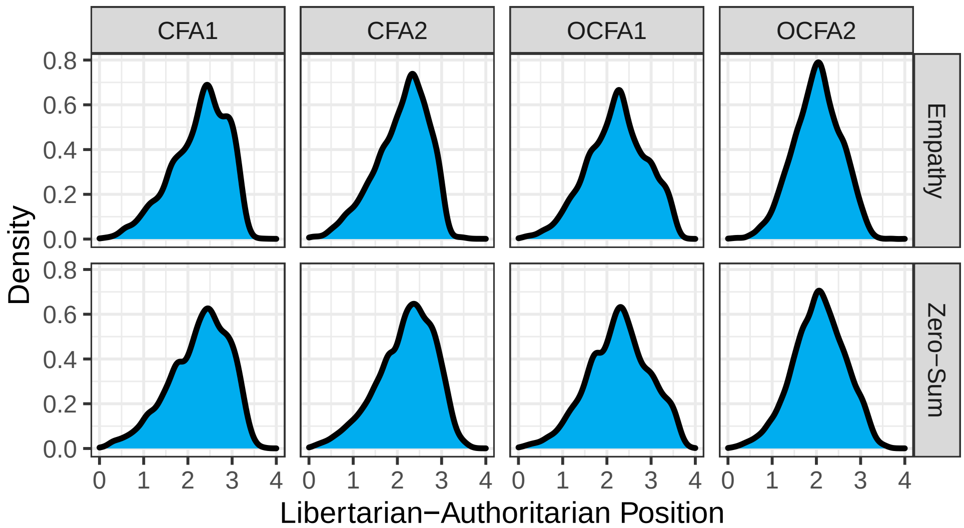


Table 7.20: Zero-Sum Left-Right

	Raw	CFA1	CFA2	OCFA1	OCFA2
Intercept	1.12	1.55	1.61	1.58	1.78
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Below GCSE	0.05	0.04	0.04	0.05	0.04
	(0.07)	(0.05)	(0.05)	(0.06)	(0.05)
GCSE/Equiv	0.09	0.06	0.06	0.08	0.06
	(0.05)	(0.04)	(0.03)	(0.04)	(0.03)
A-level/Equiv	0.18	0.10	0.10	0.12	0.09
	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)
Undergrad	0.15	0.05	0.05	0.07	0.03
	(0.05)	(0.04)	(0.03)	(0.04)	(0.03)
Postgrad	0.13	−0.00	0.00	0.02	−0.04
	(0.06)	(0.04)	(0.04)	(0.05)	(0.04)
R^2	0.00	0.00	0.00	0.00	0.00
Adj. R^2	0.00	0.00	0.00	0.00	0.00
Num. obs.	4965	4965	4965	4965	4965

Table 7.21: Empathy Left-Right

	Raw	CFA1	CFA2	OCFA1	OCFA2
Intercept	1.08	1.62	1.75	1.59	1.87
	(0.05)	(0.04)	(0.03)	(0.04)	(0.03)
Below GCSE	0.17	0.09	0.09	0.12	0.07
	(0.08)	(0.06)	(0.05)	(0.06)	(0.05)
GCSE/Equiv	0.14	0.09	0.10	0.10	0.06
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
A-level/Equiv	0.18	0.11	0.11	0.11	0.07
	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)
Undergrad	0.22	0.14	0.15	0.14	0.09
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
Postgrad	0.13	0.08	0.10	0.06	0.03
	(0.06)	(0.05)	(0.04)	(0.05)	(0.04)
R^2	0.01	0.00	0.00	0.00	0.00
Adj. R^2	0.00	0.00	0.00	0.00	0.00
Num. obs.	3847	3847	3847	3847	3847

Table 7.22: Zero-Sum Libertarian-Authoritarian

	Raw	CFA1	CFA2	OCFA1	OCFA2
Intercept	3.05	2.58	2.51	2.62	2.35
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Below GCSE	−0.05	−0.05	−0.05	−0.06	−0.06
	(0.07)	(0.05)	(0.05)	(0.05)	(0.05)
GCSE/Equiv	−0.13	−0.09	−0.09	−0.11	−0.09
	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)
A-level/Equiv	−0.45	−0.31	−0.30	−0.33	−0.28
	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)
Undergrad	−0.76	−0.53	−0.52	−0.57	−0.47
	(0.05)	(0.04)	(0.03)	(0.04)	(0.03)
Postgrad	−1.15	−0.79	−0.76	−0.82	−0.67
	(0.06)	(0.04)	(0.04)	(0.04)	(0.04)
R^2	0.15	0.13	0.13	0.13	0.12
Adj. R^2	0.15	0.13	0.13	0.13	0.12
Num. obs.	4965	4965	4965	4965	4965

Table 7.23: Empathy Libertarian-Authoritarian

	Raw	CFA1	CFA2	OCFA1	OCFA2
Intercept	3.11	2.60	2.44	2.59	2.29
	(0.05)	(0.03)	(0.03)	(0.04)	(0.03)
Below GCSE	−0.07	−0.03	−0.04	−0.06	−0.02
	(0.08)	(0.05)	(0.05)	(0.06)	(0.05)
GCSE/Equiv	−0.14	−0.09	−0.10	−0.11	−0.07
	(0.06)	(0.04)	(0.04)	(0.04)	(0.03)
A-level/Equiv	−0.53	−0.34	−0.34	−0.37	−0.27
	(0.06)	(0.04)	(0.04)	(0.04)	(0.03)
Undergrad	−0.77	−0.51	−0.50	−0.55	−0.41
	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)
Postgrad	−1.24	−0.85	−0.82	−0.88	−0.67
	(0.06)	(0.05)	(0.04)	(0.05)	(0.04)
R^2	0.16	0.16	0.16	0.15	0.14
Adj. R^2	0.16	0.16	0.16	0.15	0.14
Num. obs.	3847	3847	3847	3847	3847

Table 7.24: Zero-Sum Left-Right Alternative

	Raw	CFA1	CFA2	OCFA1	OCFA2
Intercept	1.14	1.56	1.62	1.60	1.80
	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)
GCSE/Equiv	0.07	0.05	0.04	0.06	0.04
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
A-level/Equiv	0.16	0.09	0.08	0.10	0.07
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Undergrad	0.13	0.03	0.03	0.05	0.01
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Postgrad	0.12	-0.02	-0.01	0.00	-0.05
	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)
R^2	0.00	0.00	0.00	0.00	0.00
Adj. R^2	0.00	0.00	0.00	0.00	0.00
Num. obs.	4965	4965	4965	4965	4965

Table 7.25: Empathy Left-Right Alternative

	Raw	CFA1	CFA2	OCFA1	OCFA2
Intercept	1.14	1.65	1.79	1.63	1.90
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
GCSE/Equiv	0.08	0.06	0.06	0.05	0.04
	(0.05)	(0.03)	(0.03)	(0.04)	(0.03)
A-level/Equiv	0.12	0.07	0.08	0.06	0.04
	(0.05)	(0.03)	(0.03)	(0.04)	(0.03)
Undergrad	0.16	0.10	0.11	0.09	0.07
	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)
Postgrad	0.07	0.04	0.06	0.01	0.00
	(0.06)	(0.04)	(0.04)	(0.05)	(0.04)
R^2	0.00	0.00	0.00	0.00	0.00
Adj. R^2	0.00	0.00	0.00	0.00	0.00
Num. obs.	3847	3847	3847	3847	3847

Table 7.26: Zero-Sum Libertarian-Authoritarian Alternative

	Raw	CFA1	CFA2	OCFA1	OCFA2
Intercept	3.03	2.56	2.50	2.60	2.33
	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)
GCSE/Equiv	−0.11	−0.08	−0.07	−0.09	−0.07
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
A-level/Equiv	−0.44	−0.29	−0.28	−0.31	−0.26
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Undergrad	−0.74	−0.52	−0.50	−0.55	−0.45
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Postgrad	−1.13	−0.77	−0.74	−0.80	−0.65
	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)
R^2	0.15	0.13	0.13	0.13	0.12
Adj. R^2	0.15	0.13	0.13	0.13	0.12
Num. obs.	4965	4965	4965	4965	4965

Table 7.27: Empathy Libertarian-Authoritarian Alternative

	Raw	CFA1	CFA2	OCFA1	OCFA2
Intercept	3.08	2.59	2.42	2.57	2.28
	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)
GCSE/Equiv	−0.12	−0.08	−0.08	−0.08	−0.06
	(0.05)	(0.03)	(0.03)	(0.04)	(0.03)
A-level/Equiv	−0.50	−0.33	−0.32	−0.35	−0.26
	(0.05)	(0.03)	(0.03)	(0.04)	(0.03)
Undergrad	−0.74	−0.50	−0.48	−0.53	−0.40
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Postgrad	−1.21	−0.84	−0.80	−0.86	−0.66
	(0.06)	(0.04)	(0.04)	(0.04)	(0.03)
R^2	0.16	0.16	0.16	0.15	0.14
Adj. R^2	0.16	0.16	0.16	0.15	0.14
Num. obs.	3847	3847	3847	3847	3847

Chapter 8

Essay 2 Appendix

8.1 Plots from the Raw Data Models

The following plots are those from the models using raw data, and present the same random effects as those presented from the scaled models in the main analysis.

Figure 8.1: Predicted Constrained Raw Cohort Effects

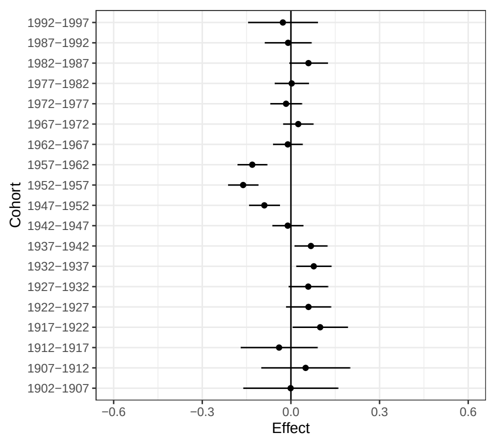


Figure 8.2: Predicted Nested Raw Cohort Effects

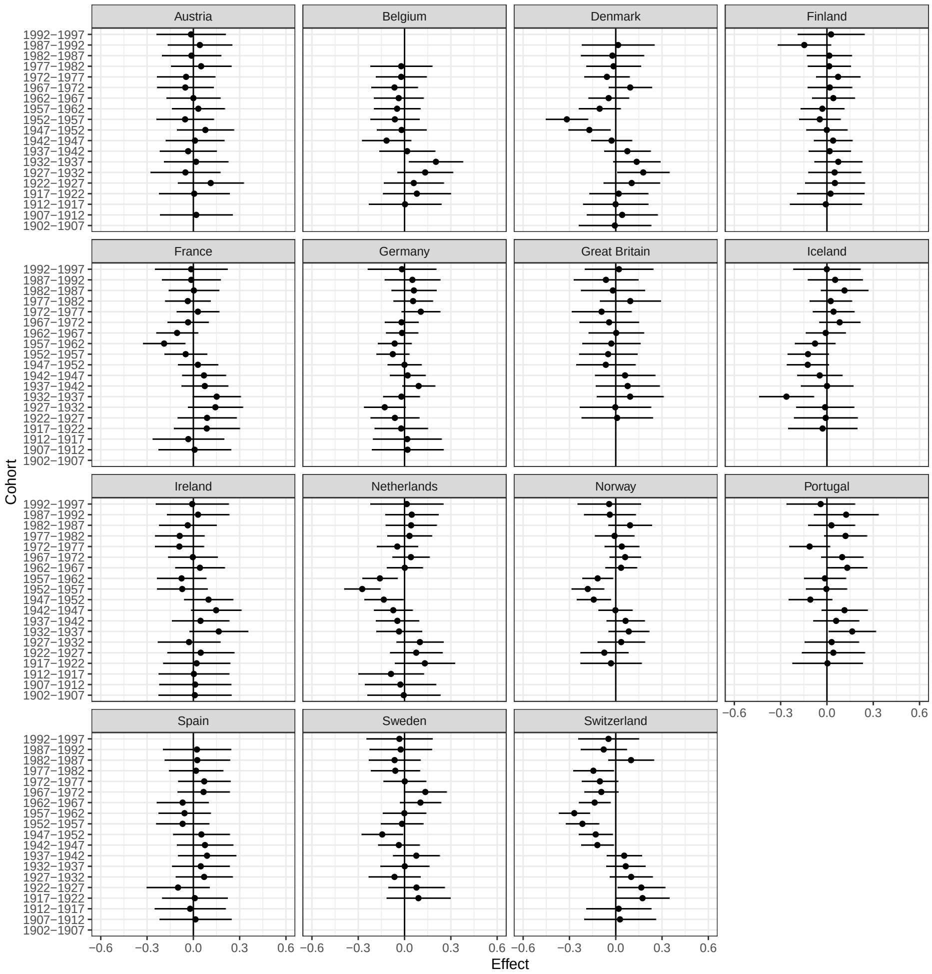


Figure 8.3: Predicted Constrained Raw Period Effects

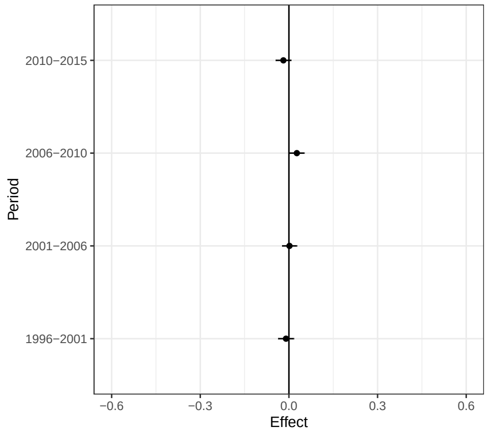
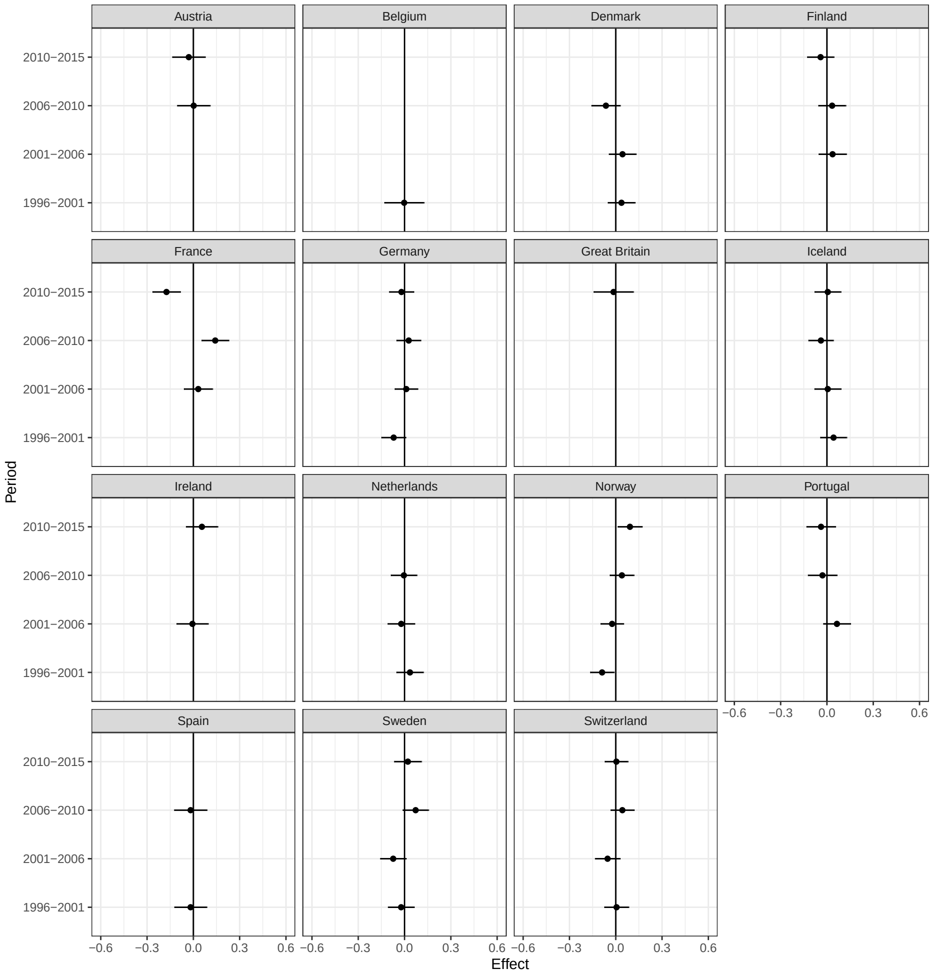


Figure 8.4: Predicted Nested Raw Period Effects



8.2 Additional Results

Figure 8.5: Constrained Cohort Scaled Gender Random Slopes

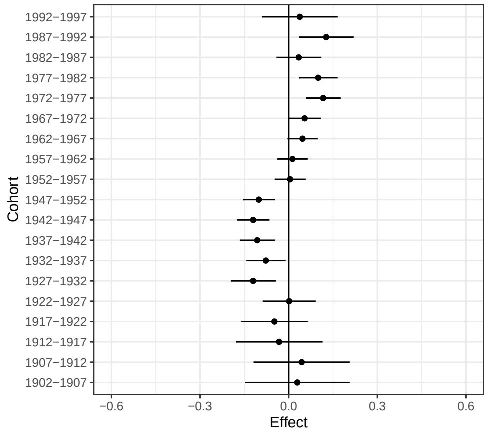


Figure 8.6: Constrained Cohort Raw Gender Random Slopes

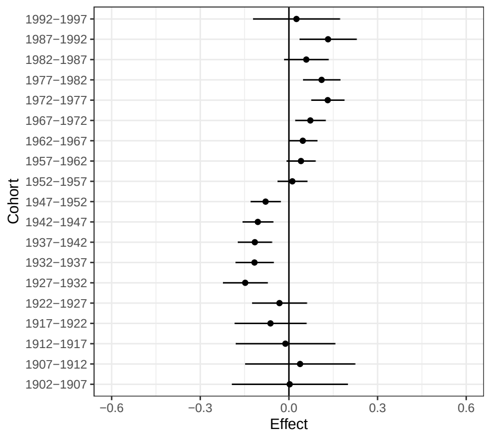


Figure 8.7: Nested Cohort Gender Scaled Random Slopes

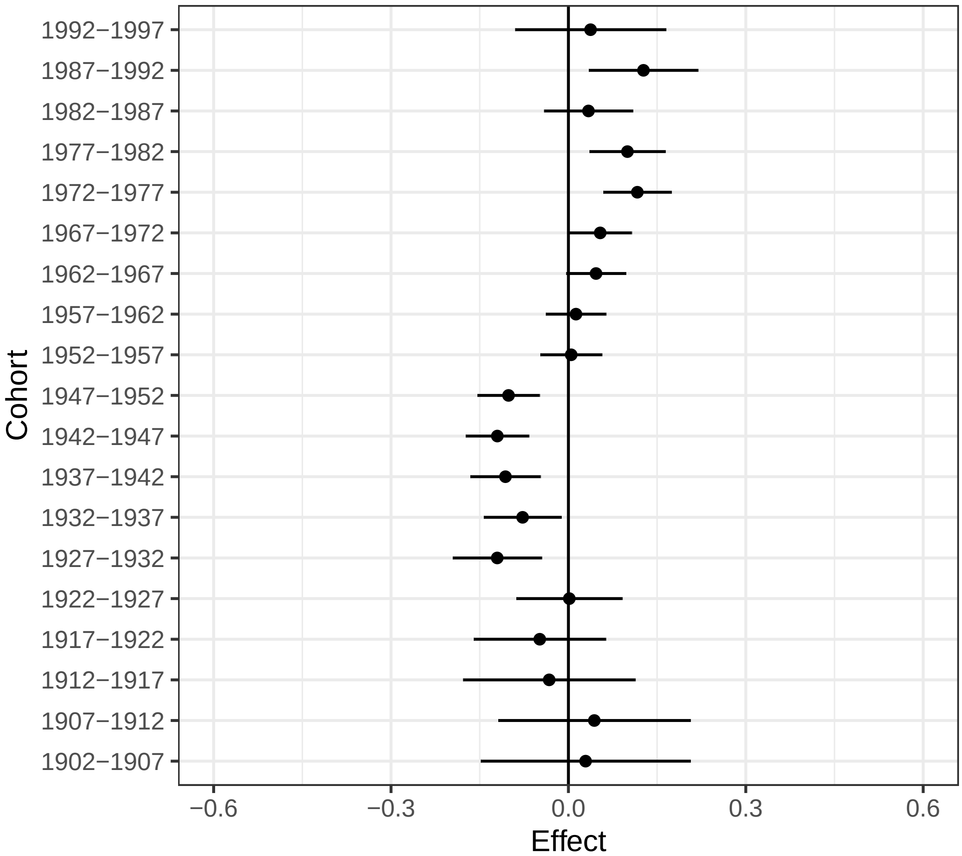


Figure 8.11: Nested Country Scaled Random Effects

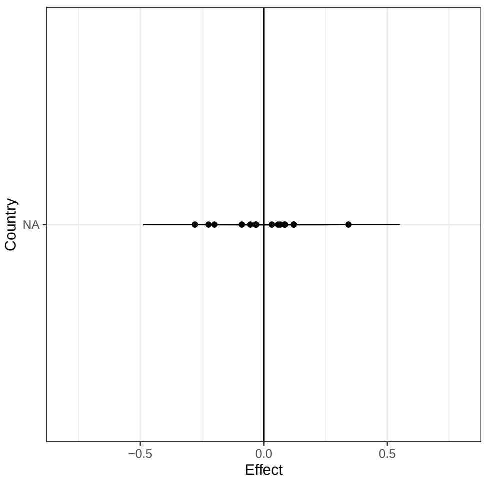


Figure 8.8: Nested Cohort Gender Raw Random Slopes

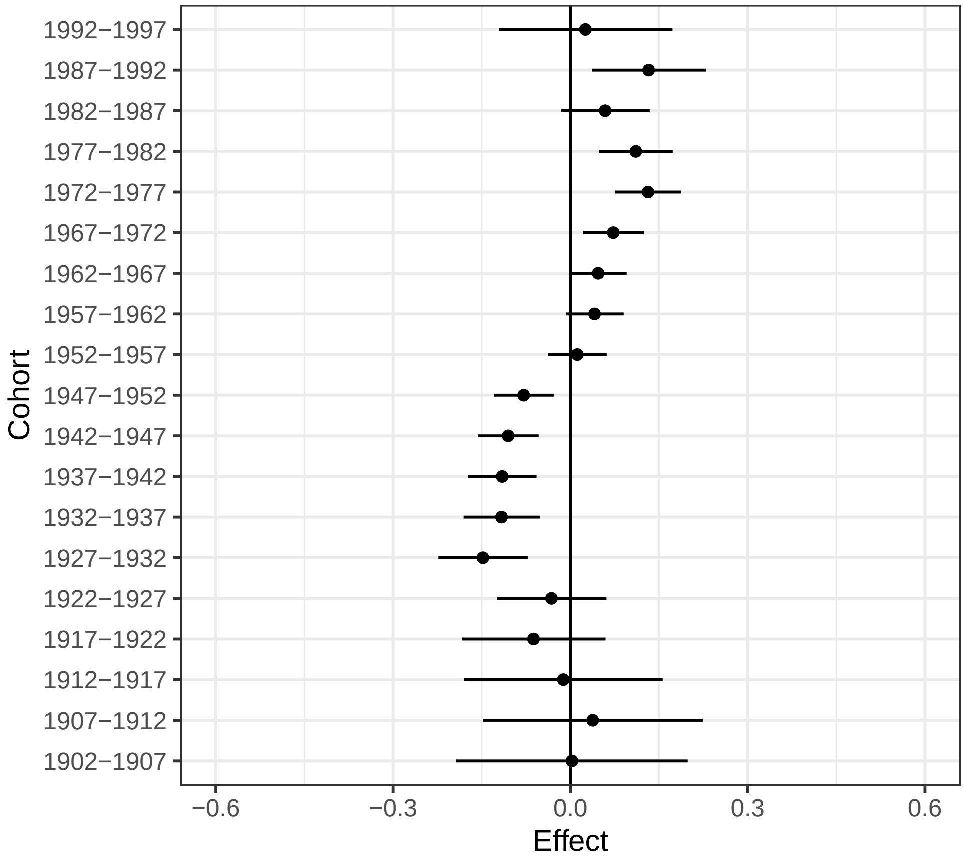


Figure 8.12: Nested Country Raw Random Effects

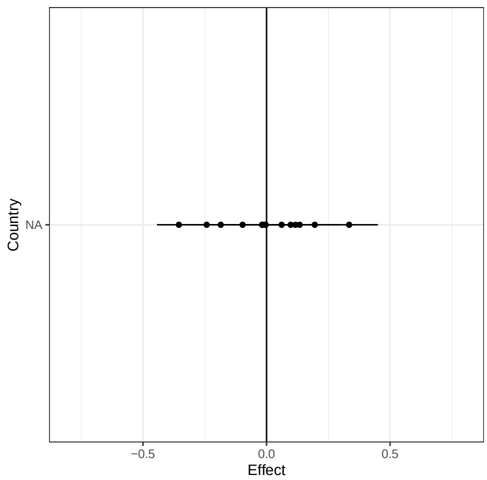


Figure 8.9: Constrained Country Scaled Random Effects

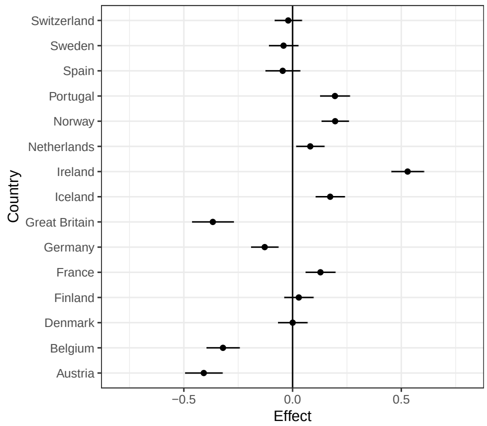
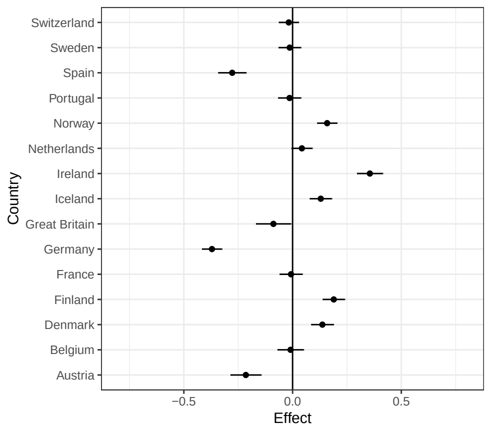


Figure 8.10: Constrained Country Raw Random Effects



8.3 Robustness

Figure 8.13: Predicted Constrained 7 Year Cohort Effects

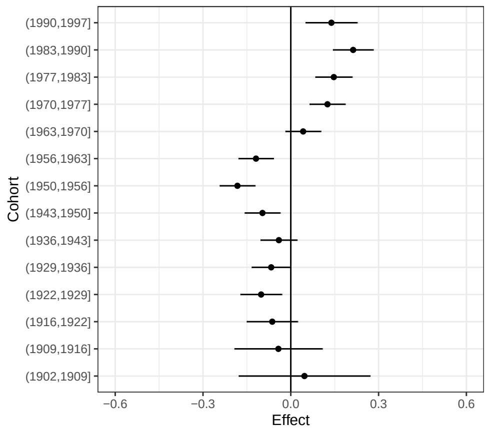


Figure 8.14: Predicted Nested 7 Year Cohort Effects

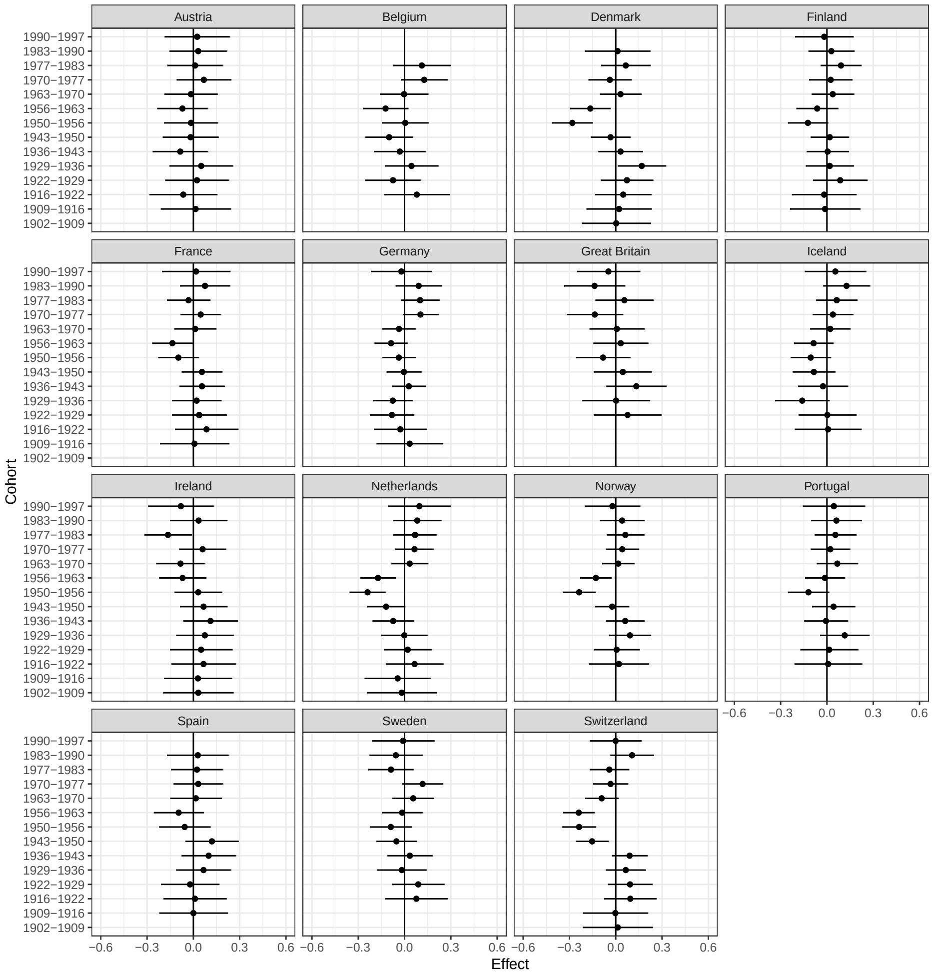


Figure 8.15: Predicted Constrained 7 Year Period Effects

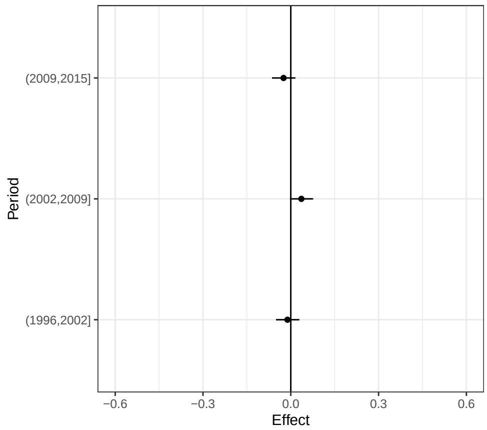
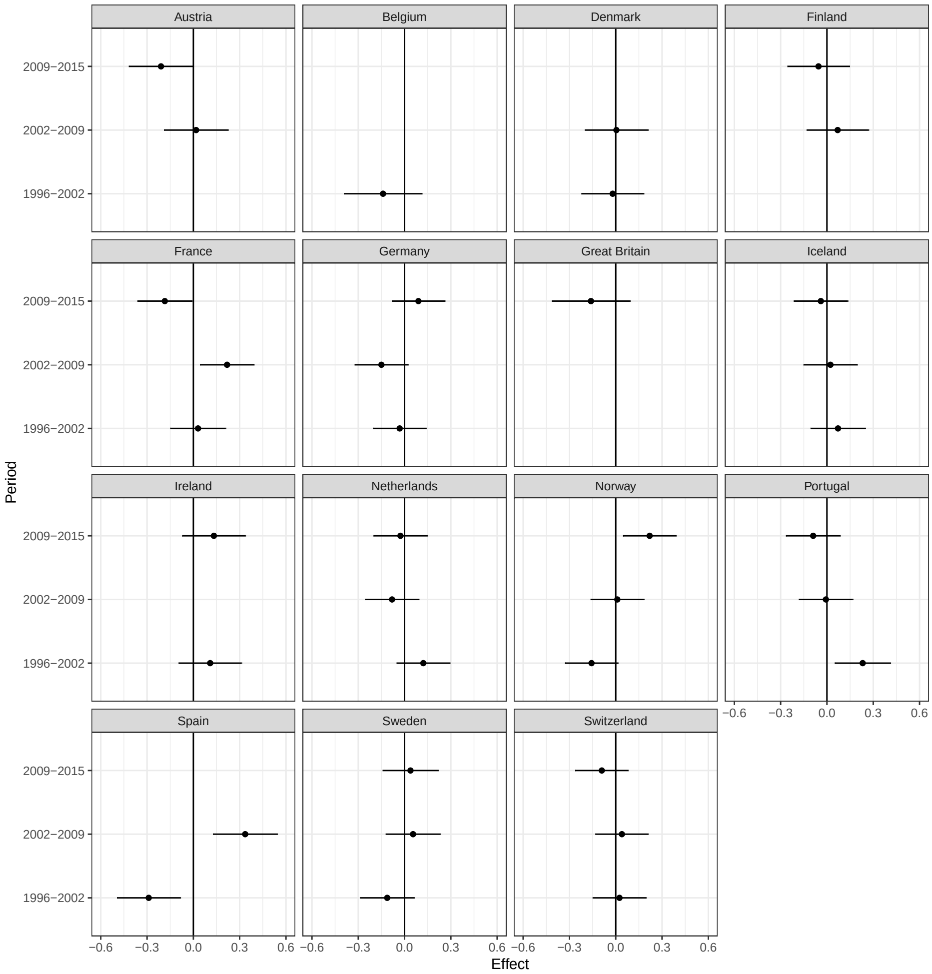


Figure 8.16: Predicted Nested 7 Year Period Effects



Chapter 9

Essay 3 Appendix

9.1 Regression and Cross Validation Tables

Table 9.1: Regression Results

	Proximity	Proximity + Categorisation
EU Distance	-0.75^{*}	-0.24^{*}
	$[-0.79; -0.71]$	$[-0.29; -0.18]$
Redistribution Distance	-0.17^{*}	-0.09^{*}
	$[-0.21; -0.14]$	$[-0.13; -0.05]$
EU Same Side		1.35^{*}
		$[1.25; 1.46]$
Redistribution Same Side		0.32^{*}
		$[0.23; 0.41]$
Win Probability	0.02^{*}	0.02^{*}
	$[0.01; 0.02]$	$[0.02; 0.02]$

	Proximity	Proximity + Categorisation
Like Party	0.13*	0.17*
	[0.11; 0.15]	[0.15; 0.19]
Like Leader	0.18*	0.16*
	[0.16; 0.19]	[0.14; 0.18]
Previously Voted For	1.26*	1.24*
	[1.20; 1.31]	[1.18; 1.30]
Party ID'd With	0.39*	0.31*
	[0.32; 0.46]	[0.24; 0.38]
Conservative Dummy	2.28*	2.31*
	[2.13; 2.42]	[2.16; 2.45]
Labour Dummy	2.15*	2.35*
	[2.00; 2.30]	[2.19; 2.50]
Liberal Democrat Dummy	1.91*	2.06*
	[1.76; 2.06]	[1.91; 2.22]
Non-Voter Dummy	1.13*	3.19*
	[0.95; 1.31]	[2.95; 3.44]
AIC	17608.14	16889.62
Log Likelihood	−8792.07	−8430.81
Num. obs.	12358	12358
K	6	6

* Null hypothesis value outside the confidence interval.

Table 9.2: Test Sample Proportions

Party	True	Proximity	Proximity + Categorisation
Brexit	0.04	0.04	0.04
Con	0.45	0.45	0.45
Green	0.04	0.04	0.04
Lab	0.28	0.28	0.28
LD	0.13	0.10	0.10
Non	0.06	0.08	0.08

9.2 Robustness Checks: Simulated vs Real Vote Shares

The following two tables show the predicted vote shares from the simulation with the position closest to Labour's estimated position from the Bayesian Aldrich McKelvey scaling (i.e. the closest to Labour's 'real' position). It compares these to the real election results in England from the 2019 general election.

Table 9.3: Real vs Simulation Voter Shares, With Non-Voters

Party	Proximity	Proximity + Categorisation	Real
Con	0.42	0.42	0.32
Lab	0.32	0.32	0.23
LD	0.14	0.14	0.08
Brexit	0.02	0.02	0.02
Green	0.03	0.03	0.01
Non	0.07	0.07	0.33

Table 9.4: Real vs Simulation Voter Shares, Without Non-Voters

Party	Proximity	Proximity + Categorisation	Real
Con	0.45	0.45	0.47
Lab	0.34	0.34	0.34
LD	0.15	0.15	0.12
Brexit	0.02	0.02	0.03
Green	0.04	0.04	0.02

9.3 Robustness Checks: Squared Distance

Figure 9.1: Simulated Vote Shares with Squared Distance

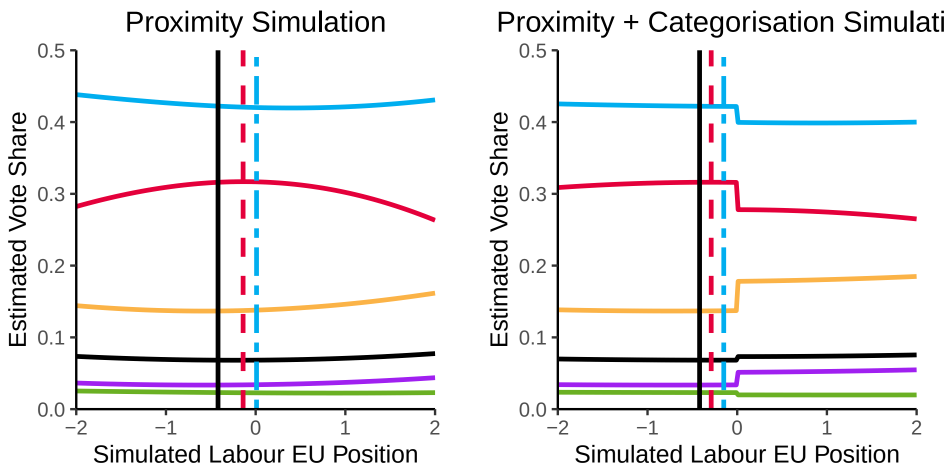


Figure 9.2: UNS Simulated Seat Shares with Squared Distance

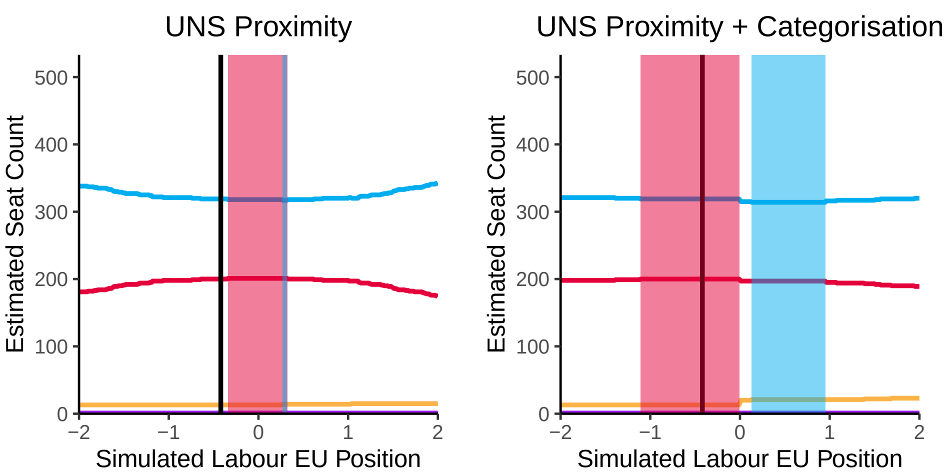
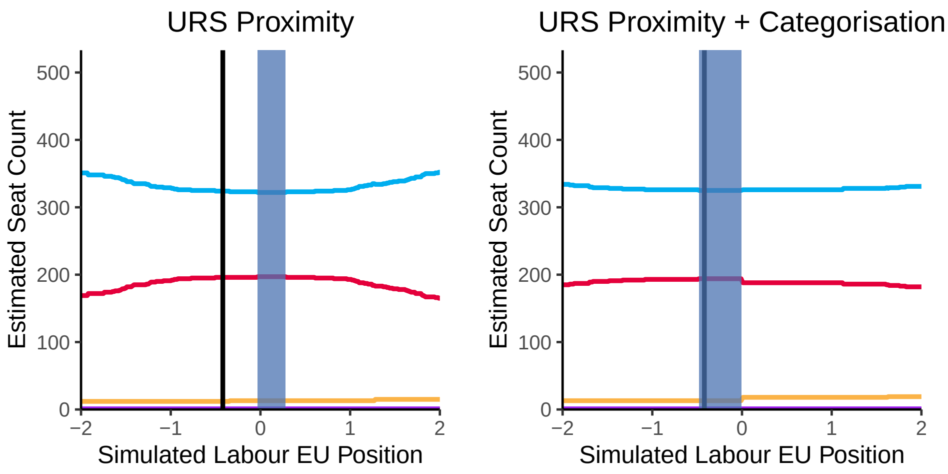


Figure 9.3: URS Simulated Seat Shares with Squared Distance



9.4 Robustness Checks: Demographic Controls

In this model, multinomial variables that vary only by individual with parameters that vary by choice are included in the model. The new multinomial-conditional model is given by

$$Pr(Y_{ij} = y_j) = \frac{exp(x'_{ij}\beta + z'_i\alpha_j)}{\sum_{h=1}^{J} exp(x'_{ih}\beta + z'_i\alpha_j)} \tag{9.1}$$

where Z'_i is the vector of individual-specific and choice-invariant covariates, and α_j is the vector of choice-varying multinomial regression parameters for these variables.

Figure 9.4: Simulated Vote Shares with Demograhpic Controls

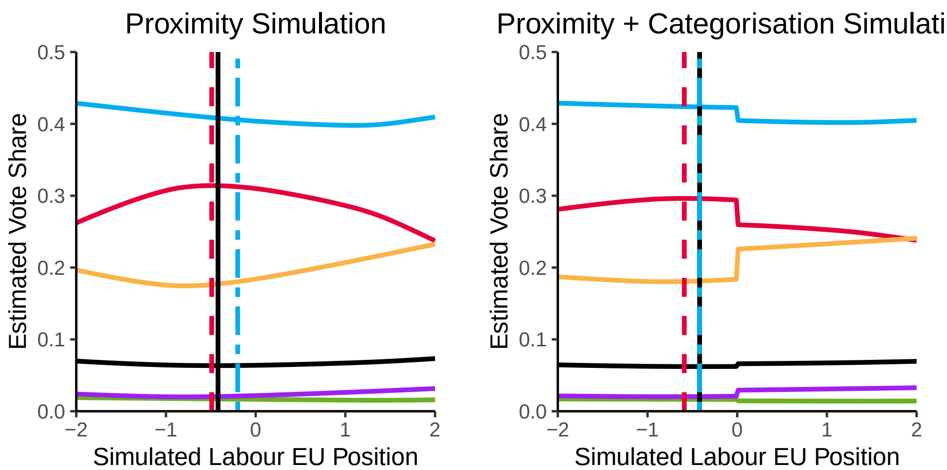


Figure 9.5: UNS Simulated Seat Shares with Demograhpic Controls

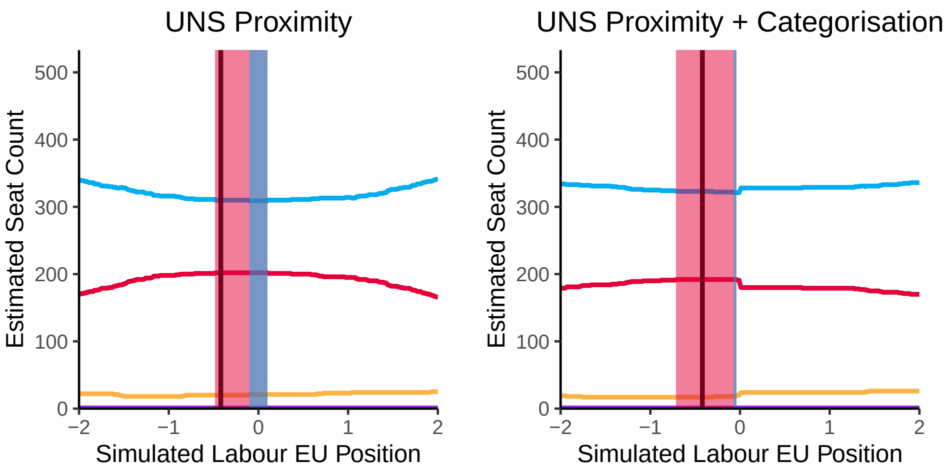
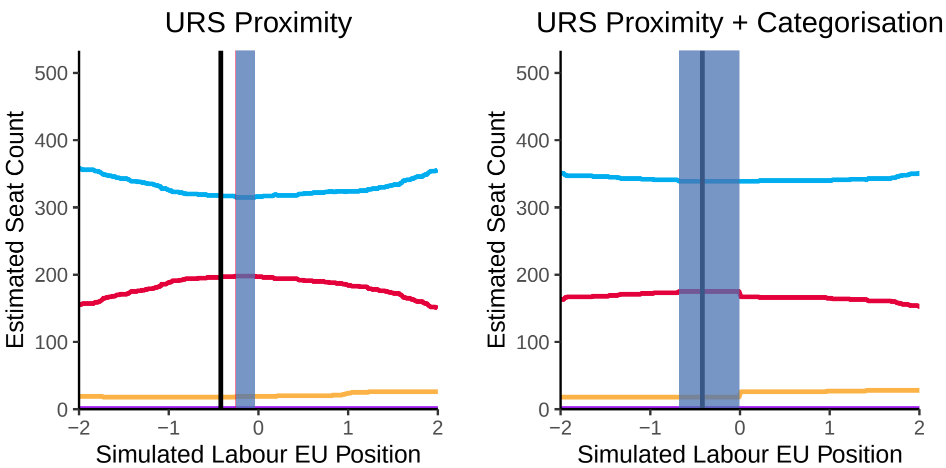


Figure 9.6: URS Simulated Seat Shares with Demograhpic Controls



9.5 Robustness Checks: No Controls

Figure 9.7: Simulated Vote Shares without Controls

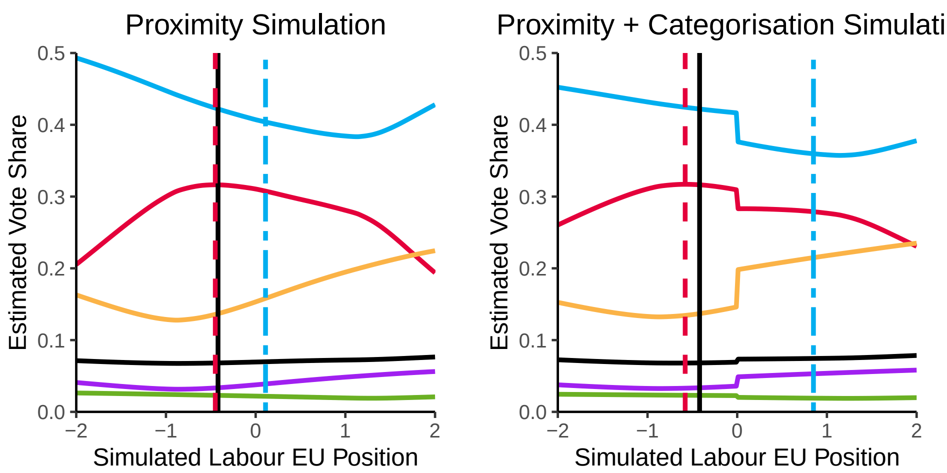


Figure 9.8: UNS Simulated Seat Shares without Controls

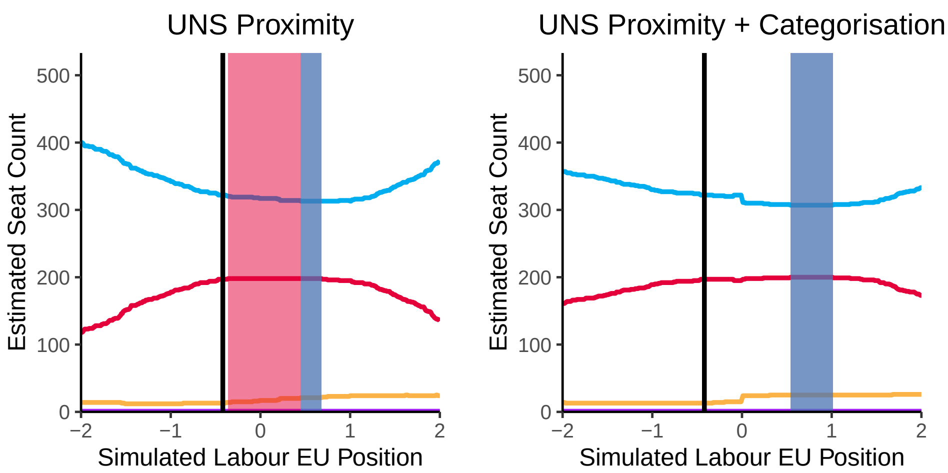
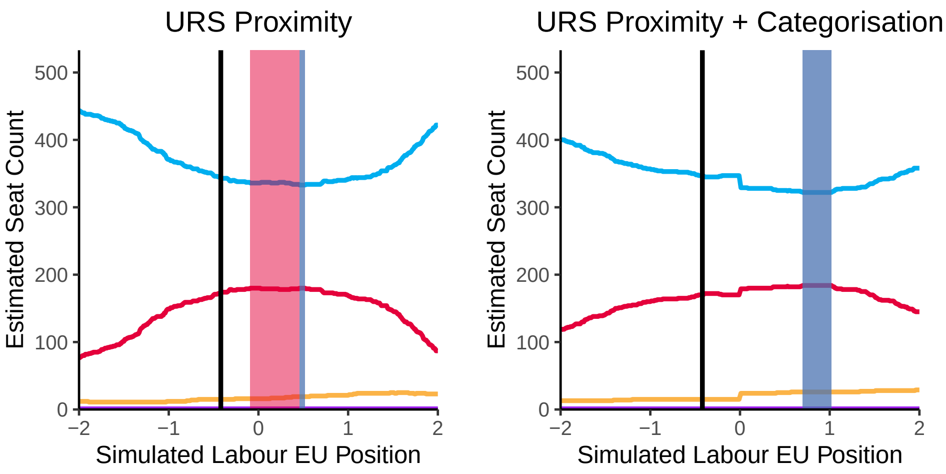


Figure 9.9: URS Simulated Seat Shares without Controls



9.6 Robustness Checks: Center Shifted Left

Figure 9.10: Simulated Vote Shares with Center Shifted Left

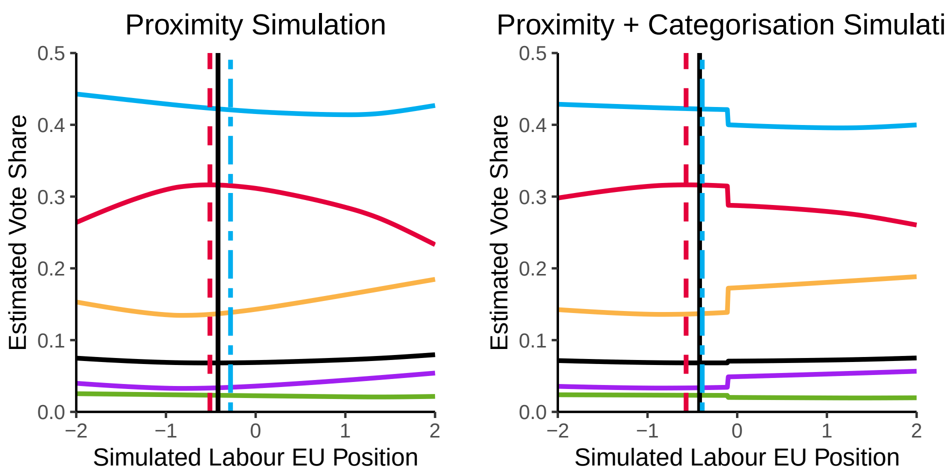


Figure 9.11: UNS Simulated Seat Shares with Center Shifted Left

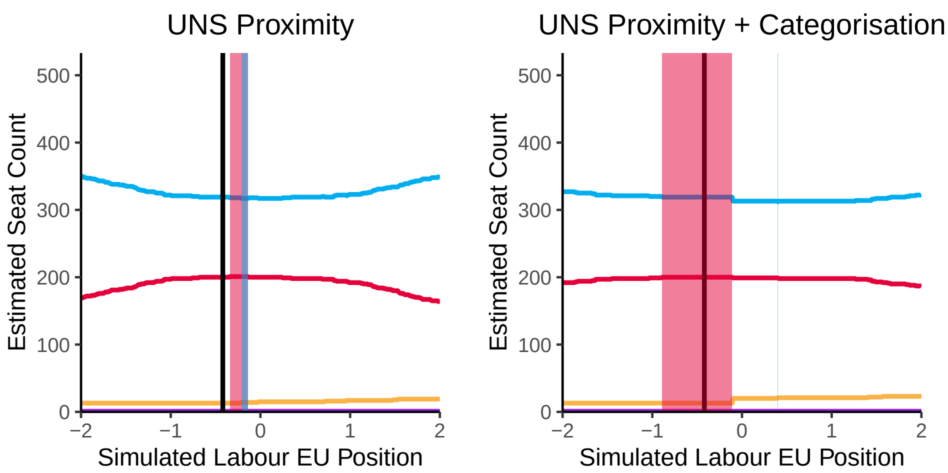
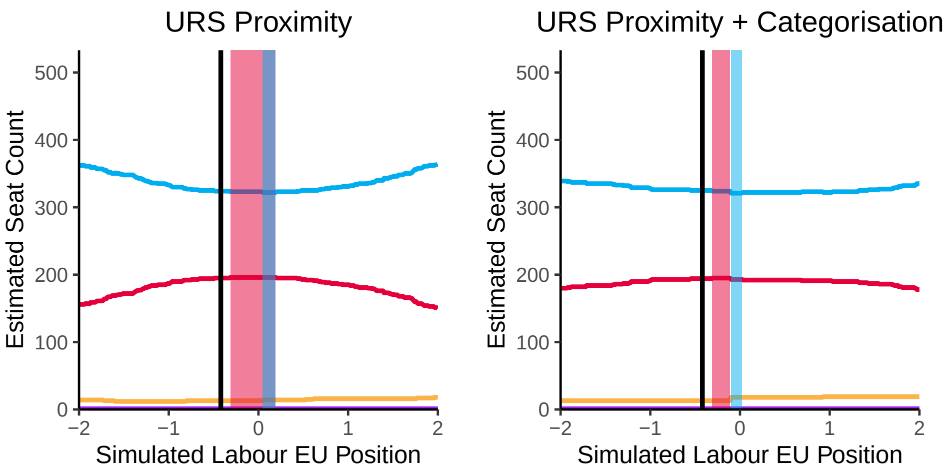


Figure 9.12: URS Simulated Seat Shares with Center Shifted Left



9.7 Robustness Checks: Center Shifted Right

Figure 9.13: Simulated Vote Shares with Center Shifted Right

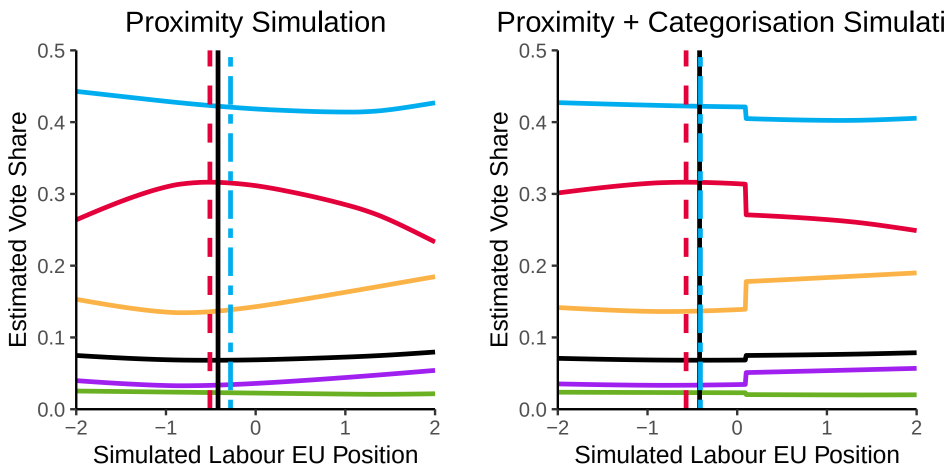


Figure 9.14: UNS Simulated Seat Shares Center Shifted Right

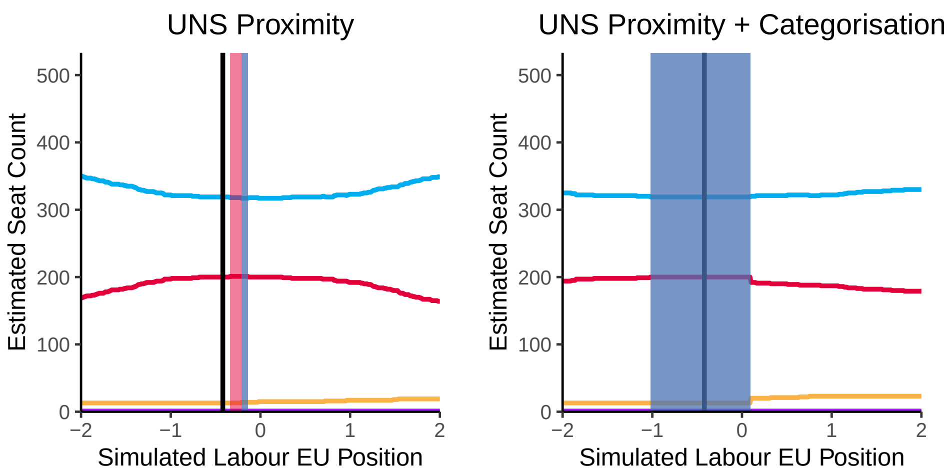


Figure 9.15: URS Simulated Seat Shares with Center Shifted Right