Measuring Attitudes towards Public Spending using a Multivariate Tax Summary Experiment *

Lucy Barnes *University College London*
Jack Blumenau *University College London*
Benjamin E Lauderdale *University College London*

It is difficult to measure public views on tradeoffs between spending priorities because public understanding of existing government spending is limited and the budgetary problem is complicated. We present a new measurement strategy using UK taxpayer summaries as the baseline for a continuous treatment, multivariate choice experiment. The experiment proposes deficit neutral bundles of changes in spending and taxation, allowing us to investigate attitudes towards modifications to the existing budget. We then use a structural choice model to estimate public preferences over 13 spending categories and the taxation level, on average and as a function of respondent attributes. We find that the UK public favours paying more in tax to finance large spending increases across major budget categories; that spending preferences are multidimensional; and that younger people prefer lower levels of taxation and spending than older people. Finally, we report a pre-registered out-of-sample validation of the estimates from the experiment.

Introduction

How much money should the government collect in taxes and how should it spend these resources across different policy areas? These two questions about the public budget are among the most important choices that governments make: they both change the material positions of citizens and enact political and social priorities. With this in mind, pollsters and political scientists have long sought to measure public attitudes regarding how much the government should tax and spend overall, and how much citizens want spent on different things. But despite the importance of taxation and spending itself, and of public opinion about taxation and spending to democratic governance, we lack good strategies for measuring these basic political preferences.

*__This version__: August 21, 2019. We thank colleagues at University College London, EPSA 2019 and Polmeth 2019 for their helpful feedback and comments.

Most studies that engage these questions are based on very general measures of policy mood such as left-right policy liberalism (Kelly and Enns, 2010) or non-budgetary policy preference (Svallfors, 1997), measures that do not engage with the details of budgeting at all. At the other extreme, we could ask repondents to allocate spending, in detail, across the full range of areas in a government budget (Bonica, 2015). But demanding each respondent provide a full government budget requires perhaps unrealistic levels of understanding and engagement. The fundamental tension that scholars face is the collision of the intrinsic complexity of budgeting with the well-known limitations of citizens' understanding and engagement with public policy.[1] How much information can we reasonably expect to extract about the public's views regarding taxation and government spending, when very few members of the public have previously considered what their ideal budget might look like, let alone tried to articulate it? Is it possible to identify a budget that the average citizen might endorse without requiring survey respondents to provide a fully elaborated, internally consistent budget?

This paper proposes a strategy for inferring a distribution of coherent public budget preferences from accessible questions about tradeoffs over subsets of budget categories. We demonstrate this approach using a new survey of UK citizens. Using widely-distributed UK taxpayer summaries as a baseline for status quo levels of taxation and spending, respondents to our experiment face deficit neutral proposed bundles of changes in tax levels and spending categories. These choice tasks allow us to investigate public preferences over the tradeoffs that form the core of public budgeting. Our experimental design ensures that survey respondents have only to consider the relative merits of the status quo versus a proposal that is identical to the status quo on all but three potential spending areas and, possibly, the tax level. This radically simplifies the choice task faced by respondents (versus a full budget allocation), but still maintains the essential quantitative budget constraint that spending changes and tax changes add up.[2]

[1] A separate criticism of using stated preferences to elicit preferences over tax and public spending packages is that surveys ask inherently hypothetical questions of respondents. This may be a particularly important problem in the context of budget preferences given that willingness-to-pay responses exhibit upward biases (Hausman, 2012). Our approach does nothing to improve on this limitation of survey-based measures of preferences. Given the ubiquity of survey-based approaches and the different problems associated with behavioural measurements of budget preferences in public opinion, improvements within the survey method are valuable.

[2] Of course government spending can also be financed through increased borrowing, but in this study we wanted to focus respondents' attention on the key tradeoffs between tax and spending. Our approach easily permits changing deficit levels.

We compensate for the modest demands made on survey respondents with modelling assumptions that enable us to translate responses onto preferences. The limited data we collect from each individual does not contain sufficient information to estimate preferences at the individual level. Instead, we use a structural choice model to translate the choices from our survey into estimates of the distribution of preferences across the UK public, as well as how average preferences vary as a function of demographic and political attributes of citizens. Our general strategy is thus to provide citizens with a set of questions that they could answer if they had well-formed budget preferences, observe the choices they make, and then translate these choices back into the distribution of well-formed preferences across citizens that would have generated those responses. We acknowledge that most respondents may not in fact have well-formed budget preferences. Part of the motivation for our design is that if citizens were asked to articulate such a budget, one strategy they might sensibly use would be to start from the status quo and then contemplate which the tradeoffs involved in changing it.

We identify three major substantive findings. First, on average, UK citizens express a willingness to support higher taxes in return for increased expenditure across a range of spending categories which aggregate to an overall increase of 7% of current levels. Second, preferences over spending areas are multidimensional: it is not simply the case that some people want more spending and others want less. Third, 18-34 year-olds want lower tax and spending levels than those older than 55, both unconditionally and holding constant their voting preferences, degree status and income. This is notable given the very large age gradient in voting in the 2017 election, which saw much higher support for the Labour party among young people than old people. In contrast to the assumptions of much political commentary, young people in the UK are not left-wing in the sense of wanting a substantially larger government.

One reasonable concern about the preference estimates we present is that they fall outside of the support of our experimental design. We therefore conducted a pre-registered validation of the estimates from the main experiment in a second survey that we report below. The validation is a "moderate success" by the standard specified in the pre-registration. Substantially more respondents in the validation study support the average preferred changes we estimate in the experiment than support the status quo, although not by as high a margin as the structural

model analysis of the experiment predicted. The experimental results strongly predict patterns of demographic variation in support for the proposal in the validation study.

Our study contributes to several literatures in political science. First, knowing what policies people support is necessary to answer questions about elite responsiveness to public preferences (Lax and Phillips, 2009*b*). It is striking, however, that this important literature is largely silent on the responsiveness of government spending, focussing instead on issues where spending is a second-order concern: gay rights (Lax and Phillips, 2009*a*), abortion (Arceneaux, 2002) and support for judicial nominees (Kastellec, Lax and Phillips, 2010). It is straightforward to measure levels of government spending across different policy areas, and feasible to measure the budgetary implications of legislative proposals, thus we address a key limiting factor for this literature: the absence of reliable measures of public opinion for levels of spending across policy domains.

Second, public opinion on economic issues in many countries is often treated as though it varies primarily according to a single ideological dimension which captures the classic left-right divide over the size and scope of government (Meltzer and Richard, 1981, Caughey, O'Grady and Warshaw (2017)). We show, by contrast, that that voters do not simply favour more or less government spending (or taxation) overall, but rather favour higher spending in some areas and lower spending in other areas. These preferences have a complex demographic structure, as different groups want more or less spending in different areas of the budget. This claim is distinct from the fact that politics in many countries features a multidimensional mix of economic and social issues: even when focusing solely on fiscal policy, improved measurement reveals that voters' preferences are not closely approximated by a unidimensional structure. These results also speak to recent work on coalitions of support for the reorientation of welfare spending from traditional categories to new ones (Häusermann, N.d.). However, our UK respondents do not align along a budgetary competition between social investment (education spending) and consumption (pensions) (Busemeyer et al., 2018). Instead, our respondents prefer to increase both at the expense of higher taxes and spending reductions elsewhere.

Finally, scholars in comparative political economy are increasingly attentive to budget trade-offs in social and fiscal policy-making at times of austerity. An important body of work asks about

voter preferences on the hard tradeoff between spending cuts and tax increases, if borrowing is off the table, or in need of reduction, so that the government budget constraint is hard (Bremer and Bürgisser, 2019; Hübscher, Sattler and Wagner, 2018). Our approach provides an additional point of evidence – that tax increases may be supported over spending cuts – and one that is anchored in a more precise balancing of the budget than is usual.

What Do We Know About Budget Attitudes and How to Measure Them?

The literature on public attitudes towards the government budget can be usefully separated into two largely distinct traditions, each of which informs key design choices in our study. First, preferences over the government budget are part of the long-running debate about whether—and to what extent—voters have consistent or rational preferences as well as whether they can express them effectively in surveys. This literature motivates our decision to only present budget consistent choices to respondents, but also our decision to only ask about limited numbers of budget categories at once. Second, we build on the recent expansion of forced choice experiments within political science, but highlight that the strict quantitative tradeoffs we are interested in in the context of public finance require us to move beyond the typical conjoint analysis framework.

The Consistency of Public Budget Preferences

The idea that most citizens fail to engage with budget tradeoffs is widely held. Based on the general observation that public spending programmes tend to be popular, while the taxes that pay for them are hated, scholars have focused on whether and why voters want 'something for nothing'. This "irrationality" is sometimes cited as another piece of evidence in the claim that democratic accountability is impossible, given the insatiable demands of the electorate.

This failure of voters to make tradeoffs is more widely lamented than documented. Indeed, the piece coining the idea of 'something for nothing' concludes that the evidence for such preferences 'is mixed' (Citrin, 1979, 128). Where scholars have investigated budget inconsistency, they do not tend to conclude that the public is unable to make budget tradeoffs. Sometimes voters have beliefs that provide good reason for denying a tight tradeoff exists – in particular the perception of waste in the link between taxes paid and services provided (Citrin, 1979; Williamson,

2017). If there is waste to be trimmed, a desire to increase spending without increasing taxes is not inconsistent. In other cases voters may not accurately perceive the costs and benefits on each side of the tradeoff. Scholars have explained a lack of willingness to pay taxes by the inadequate visibility of the spending they support (Downs, 1960; Mettler, 2011), and support for spending programmes by the inadequate visibility of the taxes required to support them (Winter and Mouritzen, 2001). In further cases, patterns of public opinion that appear inconsistent at the aggregate level can arise from preferences that are rational at the individual level where taxpayers prefer higher levels of spending financed by increases in the taxes paid by other people (Edlund and Sevä, 2013).

When survey prompts make budget tradeoffs more explicit, respondents provide more consistent profiles of responses. The most complete existing study of budget preferences under tradeoffs was fielded in the pilot to the 1996 American National Election Study, presenting respondents with choices on exhaustive and mutually exclusive categories (taxation, deficits and domestic and defense spending) that incorporated the relevant tradeoff, asking about the level of support for policy packages which combined increases in one area with offsetting decreases elsewhere. Hansen (1998) finds that these data reveal a high degree of consistency in budget preferences.

Thus, where people are given the chance to express well-structured budget preferences they are able to do so, even if it is easy to demonstrate that people do not give responses that obey budget constraints when given questions that neither enforce nor encourage obeying such constraints. The imposition of coherence is artificial, but so are unconstrained sets of questions about tax and spending preferences. It is not inconsistent or irrational to prefer both high spending and low taxes in a situation where there is no reason you cannot want both. Most people want both, all else equal. The point of the budget constraint is that all else cannot be equal. But if you are answering an survey that only mentions one at a time, or if you are left free to implicitly externalise tax costs to other people or engage in deficit financing, the budget constraints do not obviously bind. This is not a failure of the respondents, this is a failure of the survey instrument. Since our goal is to elicit public preferences over feasible budgets, it makes sense to prompt respondents with feasible tradeoffs.

Exploring how people make these budget-consistent tradeoffs is difficult, both for the respondent and the researcher. There are many spending areas in a government budget (which can be subdivided or clustered in various ways). We are potentially interested in tradeoffs between all of these. This logic leads very naturally to the idea of simply asking each respondent to set spending levels across all budget categories. This approach is taken by Bonica (2015), who uses a computer budgeting task that allows respondents to re-allocate the US federal budget from its existing baseline (holding tax and deficit constant) across 21 spending areas. Unfortunately, this may not be viable as a general solution to the budget preference measurement problem. Bonica's study was fielded to a convenience sample of US undergraduate students, so we have little evidence that the task is accessible to the general population.

Concerns about the difficulty of asking every respondent for a full budget suggests our intermediate solution in which we ask about more limited tradeoffs. This can only work if we still fully explore tradeoffs between all spending areas and incorporate the quantitative information about the magnitude of changes that is contained in a fully specified budget. If (for example) a respondent rejects an increase in education paid for by a decrease in healthcare, we cannot assume that they reject increases in education spending paid for by decreases in pension spending.

Conjoint Analyses and Quantitative Tradeoffs

In recent empirical work, tradeoffs in general have come increasingly under investigation, with the use of conjoint experiments proliferating. In ascertaining preferences over various multidimensional policy areas, conjoint choice analyses have been used to investigate support for different types of policy across a wide range of areas, including: tax progressivity (Ballard-Rosa, Martin and Scheve, 2017), climate change agreements (Bechtel, Genovese and Scheve, 2017) and immigration (Hainmueller and Hopkins, 2015), to name but a few.

Forced choice designs in which respondents are asked to choose a package of fiscal proposals which differ in terms of spending, taxation and government debt (Bremer and Bürgisser, 2019; Hübscher, Sattler and Wagner, 2018) have also been used to elicit preferences over the tradeoffs involved in fiscal consolidation,. In each of these approaches the direction of the changes

are specified – increases or decreases in taxes and spending of different types, which may be 'small' or 'large'. Thus preferences over tax levels and spending within this directionally-balanced budget are elicited. The results of these studies also show little (if any) differentiation across spending types, in terms of the effect of cuts on support for the package (cutting healthcare is equally unpopular as cutting pensions, for example); but this is a hard task for this empirical design as there is no obvious metric by which to convert 'small' cuts in one spending area to 'small' cuts in another. Our approach makes these tradeoffs more concrete by requiring not just directionally but arithmetically balanced budgets.

Our approach draws on the intuition that a multivariate, forced choice experiment like a conjoint experiment is useful for understanding preferences over multidimensional objects like budgets. But the standard conjoint design is not suited to the quantitative tradeoffs involved in budgets. The standard conjoint experiment has fully independent treatments; budget tradeoffs are, by their nature, not independent. Increasing one spending area requires changing other spending areas or tax to maintain balance. These changes are perfectly multicollinear: knowing the changes made in $K-1$ categories of tax and spending determines the remaining category exactly. This makes the resulting data not only incompatible with the standard linear regression analysis of conjoint experiments, but also with the theoretical logic of calculating Average Marginal Component Effects (Hainmueller, Hopkins and Yamamoto, 2014). In order to explore these budget tradeoffs, we need a different sort of multivariate forced choice experiment, and a new strategy to analyse the resulting data.

Experimental Design and Data Collection

We report on two studies in this paper, which we will refer to as the "experiment" and the "validation" study. In the experiment (fielded in October 2018), we presented respondents information mirroring the taxpayer summaries distributed by the UK government. These summaries, which are distributed to every UK taxpayer at the end of the financial year, indicate how individuals' income tax and national insurance contributions are proportionally allocated across a variety of public spending categories. The summaries therefore provide itemized accounts of government spending over the previous year, and are personalised for each recipient by scaling them

to that individual's total tax level. Examples of these summaries and the simplified version that we presented to respondents are provided in the supplemental information.

We used the tax summary spending allocation from fiscal year 2016/17 as a baseline and then – for each respondent – randomly altered the amount of spending in three spending categories and (in some cases) the overall tax level. All proposed changes involved maintaining deficit neutrality: proposed changes in spending and tax always exactly offset. Respondents were then asked to compare the current levels of taxation and spending to the levels we proposed in the experiment, and to indicate which they preferred.

The creation of our new tax and spending proposals followed three basic steps. First, for each proposal we randomly selected whether the overall tax rate for the proposal would change (with probability equal to 2/3), or remain the same. If the proposal was subject to a tax rate change, we multiplied the baseline level of total tax by a factor drawn from a normal distribution with a mean equal to one and standard deviation equal to 0.01. This implies that one third of respondents were presented with a proposed spending profile that was accompanied with an increase in the overall tax rate; one third with a decrease in the overall tax rate; and one third where the tax rate remained unchanged. The resulting treatment distribution included changes ranging from a 4% decrease in total tax to a 4% increase in total tax, with the bulk of the data between -2% and +2%.

Second, we randomly selected three of the spending categories to include in the bundle of expenditure changes. The full tax summaries include information on the amount of government spending devoted to 15 different spending categories. Asking respondents to consider proposed spending levels that included changes to all 15 of spending categories seemed likely to make for a very difficult choice task. Categories included in the new proposals were drawn with equal probability, with the exception of 'National debt interest' and 'Government administration' which we excluded from the experiment.[3]

[3]We excluded the national debt category on the basis that the level of payments on existing national debt is non-discretionary. We similarly excluded government administration as a varying category because the costs of government administration are likely to be mostly non-discretionary given the commitments in other areas. Arguably payments into the EU budget are similarly non-discretionary (conditional on membership in the EU) but we included this category because payments into the EU budget are at the core of a highly salient political cleavage in contemporary British politics. Where to draw this line is debatable: there are reasonable arguments for including government administration as well as for excluding the EU budget contributions.

Third, given the tax rate and the selected categories, we generated new spending levels for the proposal. We drew the proposed spending levels for the three included categories from a dirchlet distribution centred on the status quo spending levels proportionately adjusted to the new tax level drawn previously. This distributional choice meant that we proposed a wider range of spending levels for smaller spending areas. The largest spending category baseline in our experiment was for welfare, and our treatment distribution on this dimension has a central 95% interval from a -13% to a 12% change in spending. By contrast, for proposals that changed spending on the overseas aid budget – one of the smallest categories in the experiment – the same interval ranged from a -53% to 67% changes in spending.

Survey prompt

Respondents were first presented with an introduction screen which outlined the basic idea of the task, and provided information on how government spending is currently allocated across the fifteen spending categories. We expressed all information about tax and spending levels with reference to the income level of an individual who earned £23200 each year, approximately the median income of a UK taxpayer in 2016.[4]

Respondents were then asked to consider a series of comparisons between randomly generated proposed tax and spend levels and the status quo allocation. The prompt used in the experiment is given in figure 1 below. We provided respondents with current and proposed spending levels on each selected category in pounds, and also provided information on the implied change in spending and tax in both pounds and percentage terms. In addition, we provided text to clarify which spending areas would see increased or decreased spending under the proposed plan relative to the status quo, and which areas would see no change in spending. Respondents were then asked to indicate whether they preferred the current levels of spending, the proposed changes, or whether they were not sure. Each respondent faced five different choice tasks, each time comparing a randomly selected spending proposal to the status quo level of taxation and spending.

[4]Our introduction screen, shown in figure 2 in the supplemental information, closely mirrored the existing tax summaries. In particular, respondents are given the titles of the 15 categories and the associated status quo level of spending for each category. Although we did not provide further information to respondents, we provide a description of the decomposition of spending within each category in table 1 of the supplemental information.

YouGov

Consider the current tax and spend levels and the proposed tax and spend levels given below, for an individual with a yearly income of £23200.

	Current level (£)	Proposed level (£)	Change (£)	Change (%)
Welfare	991	981	-£10	-1%
Health	828	864	+£36	+4.3%
Overseas Aid	45	19	-£26	-57%
Total tax	4078	4078	£0	0%

Under the proposed tax plan, compared with current tax and spending levels:

- The total tax paid would **stay the same**.
- Spending on Welfare would **decrease**
- Spending on Health would **increase**
- Spending on Overseas Aid would **decrease**
- Spending on all other categories would **stay the same**.

Which tax and spending levels do you prefer?

- I prefer the current levels
- I prefer the proposed changes
- I am not sure

Figure 1: Experiment prompt

We have suggested above that part of the logic of this approach is that it is more accessible to respondents than a full allocation task. The simplest test of respondent engagement with our prompt is to assess how many people give "I am not sure" responses, and in particular how many give this response to all five proposed sets of changes that they saw. Of 3533 respondents overall, 1846 gave 0 "not sure" responses, 565 gave 1, 305 gave 2, 126 gave 3, 126 gave 4, and 565 gave 5. We would expect that even fully engaged respondents would give this responses to some fraction of the prompts, but the relatively large number of respondents giving this response to 5 items versus 3 or 4 of the items confirms that some fraction of respondents did not engage with the questions.[5] Overall, 16% gave this response to all items. Those with lower levels of education were more likely to not express preferences on any comparison: among those with university degrees, 11% versus 19% among those without. These are moderate levels of differential non-response that are typical of those seen in political survey data (Berinsky, 2008).

Overall, this experiment gives us rich information about how citizens make tradeoffs between different spending areas and tax, but it is not straightforward to analyze. It is not a traditional conjoint experiment, because the experimental treatments are not merely non-independent, but perfectly multicollinear. As a result, we cannot simply compare responses at higher or lower levels of spending on a single budget area, because the spending in other areas will be systematically different. We cannot fit regressions for the responses as a function of the treatments, because of the aforementioned multicollinearity: the "all-else-equal" logic of regression modelling is ill-suited to data where all-else is never equal by design. Instead, we use a structural model that assumes respondents have preferences over spending and tax levels, and make choices accordingly. Given these assumptions and the observed responses, we can infer the underlying distributions of preferences that would be most likely to have generated the data. Branham and Jessee (2017) use a similar approach to US General Social Survey data on "too much", "about right", "too little" assessments across 18 spending areas, mapping these responses back to a unidimensional spatial model, and then estimating the distribution of public preferences in that

[5]The relationship between the number of "not sure" responses and respondents' survey completion time also suggests that these responses are indicative of a lack of engagement, rather than evidence of respondents finding it difficult to adjudicate between similar budget proposals. Each additional "not sure" response is associated with a decrease of 14% in the total time respondents spent on the survey. Respondents who gave 5 "not sure" responses spent, on average, 56% less time on the survey than respondents who provided 0 "not sure" responses.

unidimensional space.

A Model for Respondent Choice

To build a model for respondents' choices, we need to make some assumptions regarding how they will tradeoff changes in spending and taxation. We adopt an additive quadratic loss model for deviations from respondents' preferred spending level in each spending area, which is consistent with typical approaches in the spatial preference modelling literature (Poole, 2001; Clinton, Jackman and Rivers, 2004) and which proves mathematically convenient for defining our estimator below. Each respondent i makes a choice between the proposed alternative (A) and the status quo (S), with an option to say they are not sure (NS) which they prefer. We assume that they do so as a function of the latent utility of the two alternatives, with "not sure" corresponding to cases where their utility difference is small.[6]

- $Y_i =$ proposed alternative if $u_{iA} - u_{iS} > \gamma_A$
- $Y_i =$ not sure if $\gamma_S \leq u_{iA} - u_{iS} < \gamma_A$
- $Y_i =$ status quo if $u_{iA} - u_{iS} \leq \gamma_S$

We assume that respondents i have quadratic loss in deviations of the status quo $S_i = (S_{i1}, \ldots, S_{iM})$ and the proposed alternative $A_i = (A_{i1}, \ldots, A_{iM})$ from their preferred point $\Psi = (\psi_{i1}, \ldots, \psi_{iM})$, and that this is weighted per dimension $j \in 1, \ldots, M$ by a factor λ_j:[7]

$$u_{iA} = -\sum_j \lambda_j \left(A_{ij} - \psi_{ij}\right)^2 \tag{1}$$

[6] In order to identify the latent utility scale relative to the response, we assume that the thresholds that respondents apply for selecting either the proposed alternative or the status quo, as opposed to "not sure", are symmetric around 0 and set to $\gamma_A = +1$ and $\gamma_S = -1$.

[7] This specification means that we assume that spending dimensions are weighted in the same way by all respondents. In order to interpret response variation as reflecting how close the alternatives are to individuals' preferred points, it is necessary to make this homogeneity assumption about how much respondents care about each spending dimension. Work by Abramson, Kocak and Magazinnik (2019) illustrates an analogous point in detail for standard conjoint experiments. We also assume spending preferences are independent in the sense that one's preferred level of health spending does not vary depending on the current level of spending in other areas of the budget, which is necessary in order to identify the multivariate normal covariance in preferences *across* respondents. For these general categories of spending, we think that variation in preferred spending levels is likely to be more consequential than variation in spending dimension weights, and also that covariance in preferred points across persons is more likely to be consequential than across preference dimensions within persons.

$$u_{iS} = -\sum_j \lambda_j \left(S_{ij} - \psi_{ij}\right)^2 \tag{2}$$

$$u_{iA} - u_{iS} = \sum_j \lambda_j (S_{ij}^2 - A_{ij}^2) + 2\lambda_j (A_{ij} - S_{ij})\psi_{ij} \tag{3}$$

The λ_j weights mean that the model allows for respondents putting similar weights on £ deviations in all categories, or similar weights on percentage changes from baseline spending, or some other weighting of categories. We discuss alternative specifications for the quadratic loss, and their consequences for model estimates, below.

The values of S_{ij} and A_{ij} are data and we have plenty of data with which to estimate λ_j for each of the dimensions. The thing we neither know, nor can estimate with useful precision, is a single respondent i's preferred allocation in each spending area j, their ideal point ψ_{ij}.

Estimating the Distribution of Preferences

While we do not have enough data to measure ψ_{ij} at the individual level, we can estimate the distribution of these individual-level ideal points in the population. We assume that ψ_i follows a multivariate normal distribution:

$$\psi_i \sim N\left(\mu, \Sigma\right) \tag{4}$$

where μ is a vector of length M giving the average preferred allocation in each spending area. Σ is the covariance matrix of individual respondents' preferred allocations around that average with elements $\Sigma_{jj'} = \rho_{jj'}\sigma_j\sigma_{j'}$. Equation 3 then defines an affine transformation on this multivariate normal distribution, and thus implies a univariate normal distribution for the utility difference between the two alternatives faced by a given respondent:

$$u_{iA} - u_{iS} \sim N\left(\nu_i, \omega_i^2\right) \tag{5}$$

$$\nu_i = \sum_j \lambda_j (S_{ij}^2 - A_{ij}^2) + 2\lambda_j (A_{ij} - S_{ij})\mu_j \tag{6}$$

$$\omega_i^2 = \sum_j \sum_{j'} \rho_{jj'}\sigma_j\sigma_{j'} \left(2\lambda_j (A_{ij} - S_{ij})\right) \left(2\lambda_{j'} (A_{ij'} - S_{ij'})\right) \tag{7}$$

Given our definition of the response at the outset this yields the following response distri-

bution, conditional on parameters, for the three responses:

$$p(\text{status quo}) = \Phi\left(\frac{\gamma_S - \nu_i}{\omega_i}\right) \tag{8}$$

$$p(\text{not sure}) = \Phi\left(\frac{\gamma_A - \nu_i}{\omega_i}\right) - \Phi\left(\frac{\gamma_S - \nu_i}{\omega_i}\right) \tag{9}$$

$$p(\text{proposed alternative}) = 1 - \Phi\left(\frac{\gamma_A - \nu_i}{\omega_i}\right) \tag{10}$$

This model has a moderate number of free parameters—13 μ_j, 13 λ_j, 13 σ_j, 78 ρ_j = 117 total—but we have 17665 responses with which to estimate those parameters (about 150 per parameter). Because our experiment can only meaningfully test spending levels that are included in the range of provided treatments, we put a prior on our estimated average preferred spending levels that corresponds to that treatment distribution. This prior is $\mu_j \sim N(S_j, sd_j)$, where S_j is the status quo / baseline spending in each area as described above, and sd_j is the standard deviation of the non-zero treatments provided on each spending area.

It is straightforward to extend the model above to enable the description of variation in individual preferences as a function of covariates. Instead of estimating a single vector of average preferred spending levels μ, we instead model individual preferences by replacing Equation 4 with

$$\psi_i \sim N(\alpha + \beta X_i, \Sigma) \tag{11}$$

for some $1 \times K$ vector of covariates X_i describing each respondent i and estimate a $M \times 1$ vector of intercepts α and a $M \times K$ matrix of coefficients β that describe variation in preferences as a function of each of the K covariates on each of the M spending dimensions. For these models, we impose the same prior on α that we put on μ in the model without covariates, and a common shrinkage prior on all β parameters proportional to the category size $\beta \sim N(0, S_j\sigma_\beta)$ to avoid overfitting small subgroups.

We estimate the models with and without covariates using Hamiltonian Monte Carlo as implemented in Stan (Carpenter et al., 2016). Our presented results consist of mean posterior point estimates and central 95% intervals based on five parallel simulation chains of 1000 iterations (after 250 iteration warmup) for each model. In order to maximise the degree of representative-

ness with respect to the UK population, we use demographic survey weights provided by YouGov via a quasi-likelihood approach. The estimates are very similar without using the weights.

Results

Average Preferences

Figure 2 presents the average preferred change in spending for each of the spending categories in our experiment, expressed in percentage terms (where current levels are the baseline). UK respondents, on average, endorse spending increases across a wide variety of policy areas. We see average preferences for spending increases of greater than 10% in education, pensions, health, the environment and housing; we also see smaller increases in transport, welfare and criminal justice. By contrast, UK respondents favour budget cuts in only relatively few spending areas: business and industry, culture, overseas aid, and UK contributions to the EU budget. People want more spending in the four categories with the largest baseline levels: education, pensions, health, and welfare. Likewise, the largest preferred spending cuts – to overseas aid and EU budget contributions – are in the two smallest categories we included in the experiment. One consequence of this is that the average preferred overall spending levels are 1.08 times the current spending levels in the categories where we proposed changes. Once we take into account the two budget categories in which we did not test changes—payments on the national debt and government administration—this implies that respondents would be willing to accept an overall tax increase of 7% to pay for these increases in spending.

Model checks

In the supplemental information, we look in detail at the values of the auxiliary parameters λ, σ and ρ, as well as examining the several alternative specifications of the model. We briefly summarise some key findings here. The dimensional weights λ_j are very close to inversely proportional to the baseline spending levels: the correlation coefficient of the λ_j parameters with $1/S_j$ is $r = 0.985$. This indicates that respondents tend to penalise alternatives in terms of percentages, not pounds (recall that we present both in the experiment). The ρ_j indicate that preferences across spending areas are correlated in expected ways, given how UK politics is or-

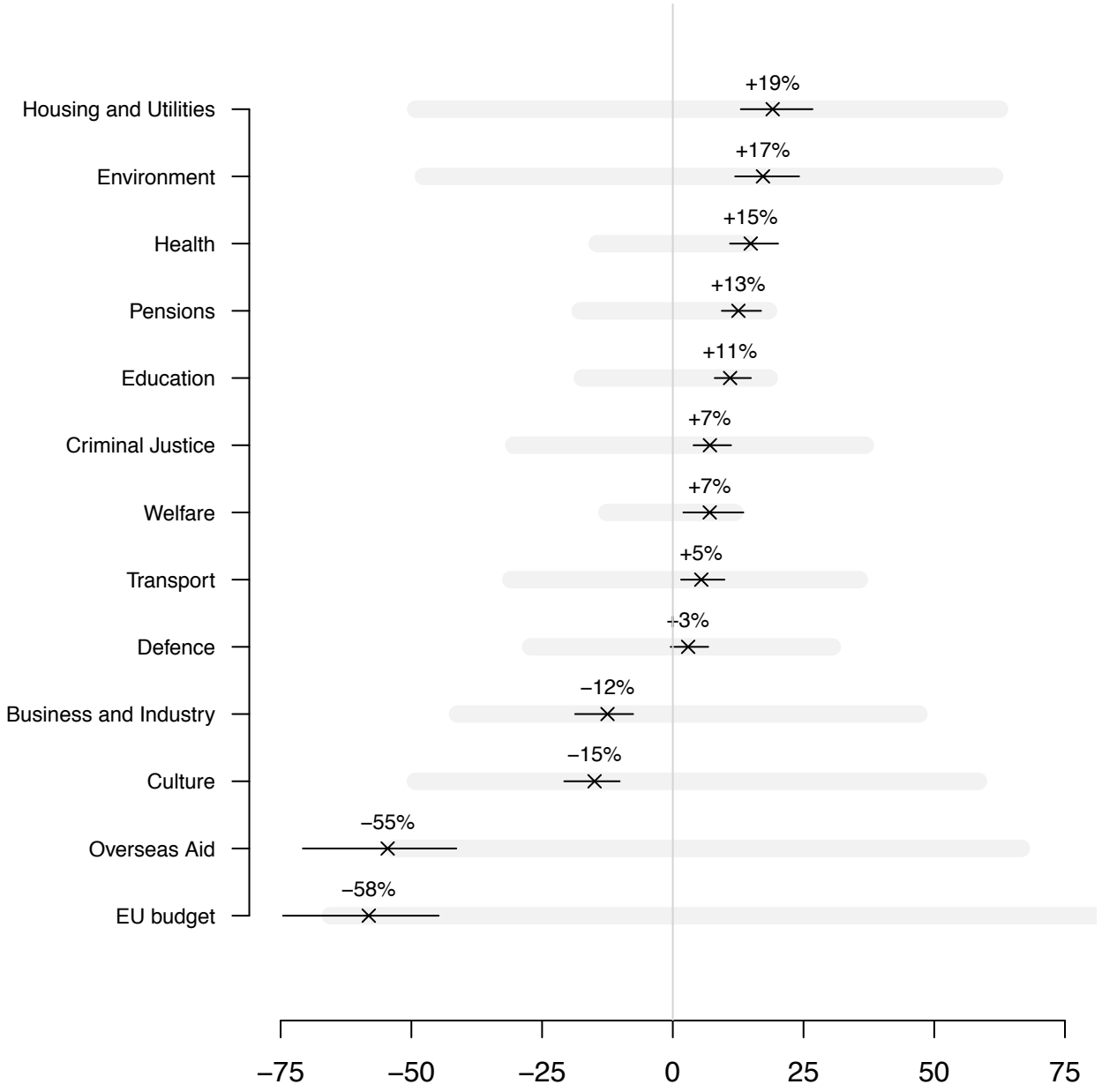

Figure 2: Average Preferred Change in Spending (% of Current Level) with 95% posterior intervals. The light grey bands show the central 95% range of the treatment distribution for that spending area.

ganised, but not very strongly. The most highly correlated preferences are those for the core social welfare categories of welfare, education and health.

In the supplemental information we also report the estimates of μ from our main model compared to (1) a model where we assume quadratic loss in log spending rather than spending, (2) a model where we drop respondents who gave "I am not sure" responses to all five proposed changes, (3) a model where we estimate responses separately for each of those five rounds of responses, and (4) a model where we use only responses to proposals that involve no tax change. Only the last of these model/data variations changes the estimates in a substantively significant way. When we only use the data without proposed tax changes, our estimated preferred spending levels shift down and closely approximate an overall budget reallocation with no tax changes, with similar relative preferences over the 13 budget areas. At the same time, the implied estimate of the preferred level of overall tax becomes more uncertain, because we have excluded all the observations that tell us anything about tax versus spending tradeoffs. Overall, the estimates are robust to these model and data variations, telling a consistent story about respondents' relative preferences for spending in different areas.

Cross-Sectional Preference Variation

As we described earlier, the average preferences of the public can be further disaggregated in terms of covariates. We present a series of simple models (figure 3) with single categorical variables for age (18-34, 35-54, 55+), university degree (no, yes), household income (below £30,000, above £30,000, refused), and combinations of votes in the 2016 EU referendum the 2017 UK general election (Conservative-Leave, Conservative-Remain, Labour-Leave, Labour-Remain, other). In addition, we fit a multivariate model with age, degree, income, 2016 vote and 2017 vote, in order to check whether the univariate patterns that we find for some of these variables might reflect the correlations between the covariates. We present the results from the multivariate model in figure 4.

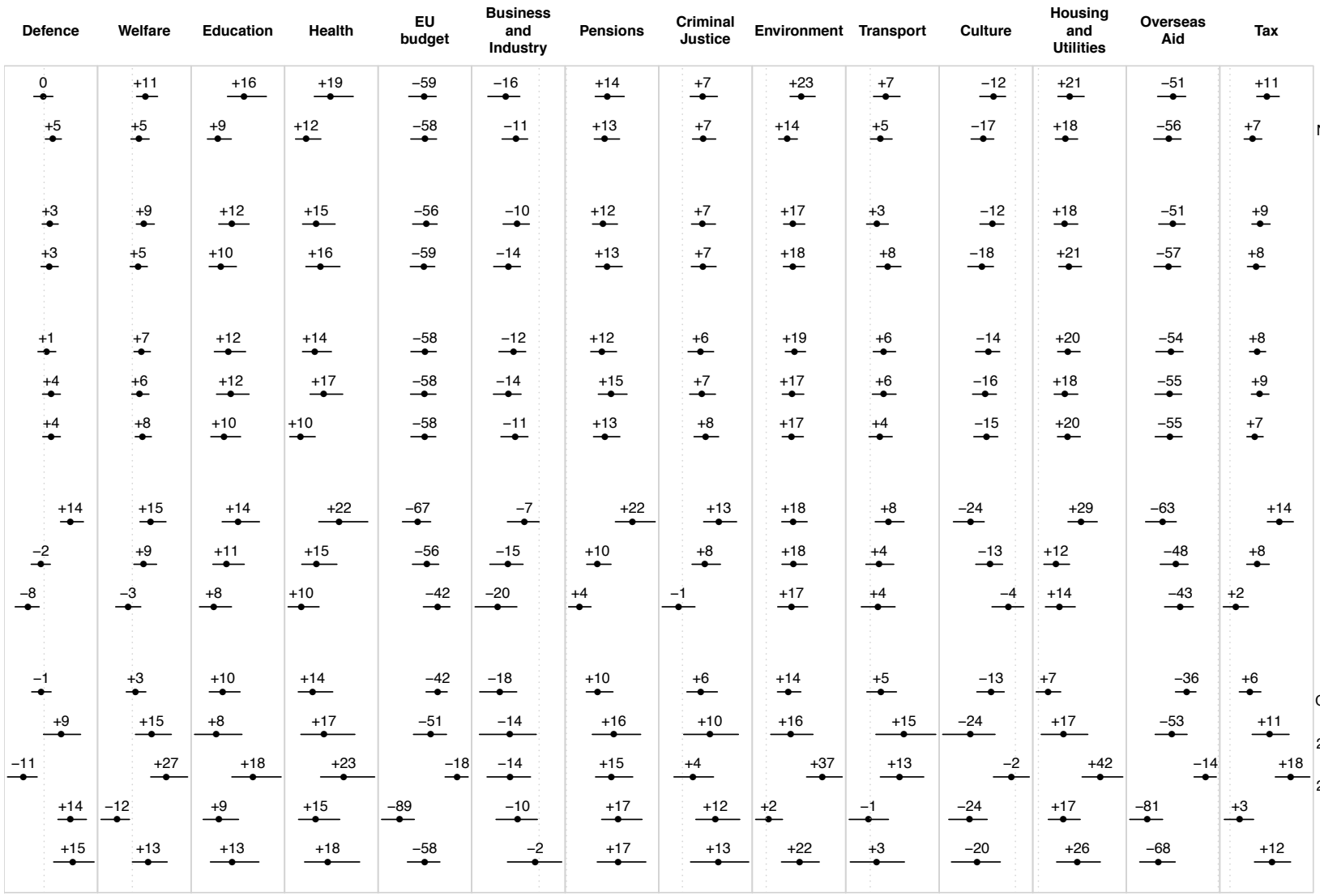

Figure 3: Average preferred changes in each spending area and overall tax (columns), for various subgroups (rows)

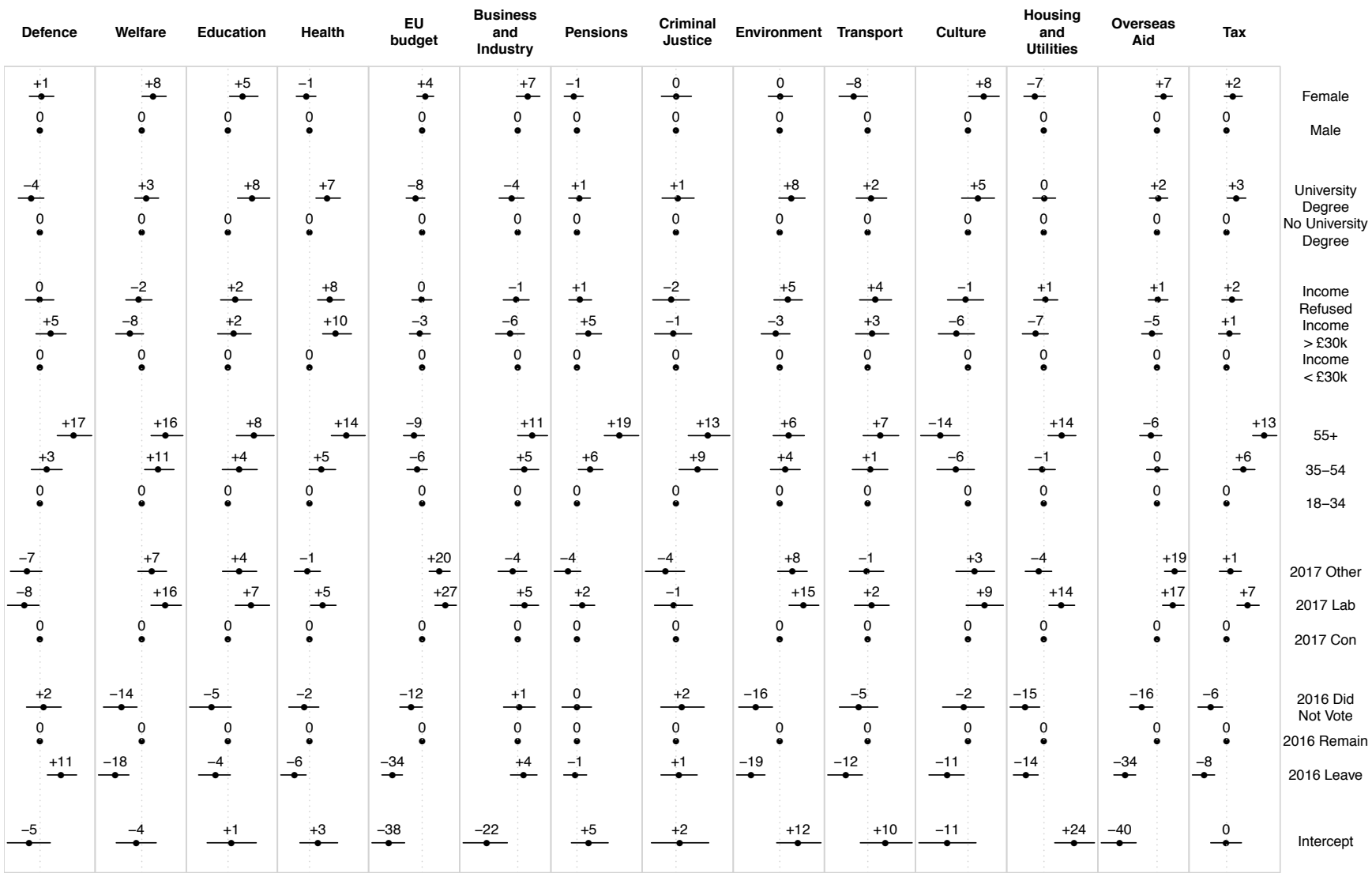

Figure 4: Multivariate model coefficient estimates as a function of covariates (rows) for average preferred spending in % of current baseline for each spending area and for overall tax (columns).

Figure 3 (simple models) and figure 4 (the multivariate model) both provide results that confirm the face validity of the model and also reveal some interesting patterns that would not be obvious given commonly held stereotypes about contemporary British politics. Regarding face validity, we see the basic strong political relationships that we would expect. 2017 Conservative voters want less spending and lower tax overall than 2017 Labour voters. Labour voters want more spending than Conservative voters on most categories of spending, especially welfare, education, health, the environment, and housing. Conservative voters want more defence spending than Labour voters. Leave voters want far lower spending on the EU (and overseas aid) than Remain voters, but also less on the environment, welfare, transport, culture and housing. None of these are surprising, for many of these relationships we would have reason to worry about the validity of the method if it had found otherwise.

What might be more surprising to observers of British politics are the associations of respondents' tax and spend preferences with age. The 2017 UK general election saw a historically remarkable age gradient in voting in the UK, with Labour winning those aged 18-34 by a 60% to 27% margin and the Conservatives winning those aged 55 and older by 54% to 31% (Fieldhouse et al., 2017). Nonetheless, younger respondents to our survey prefer substantially lower levels of taxation and spending overall. This is true both holding constant 2017 vote (figure 4) but also unconditionally (figure 3). In the simple group comparison (figure 3), the average increase in overall spending and tax preferred by 55 year-olds is 14%, versus just 2% for those under 35 years-old. Young people endorse lower spending levels than older people in every spending category except the contribution to the EU budget, spending on culture, and spending on overseas aid, which are all among the smallest budget areas. In some sense this is rational at the individual level: young people have higher tax contributions ahead of them. However the pattern is in stark contrast to the widely held understanding in UK politics that younger cohorts are more amenable to greater public expenditure and taxation than older voters.[8] Unless there is some very large age-associated bias in our measurement strategy, which we find no indications of and think is unlikely, it is the older voters who want to substantially increase government spending

[8] Previous work on generational differences in political attitudes associated with the legacies of Thatcher and Blair has some suggestions of this finding (Grasso et al., 2017), but the measures used in that research are less precisely focused on spending.

not the younger voters.

Overall, a key finding of this analysis is that citizens' spending preferences are multidimensional rather than tightly structured by a single preference dimension. In the supplemental information, we report details of principal components analysis on the estimated distribution of preferences from the model. The variation in preferences that we can explain in terms of covariates is 2 or 3 dimensional, depending on the criterion used for dimensionality. The first dimension is oriented from Conservative-Leave to Labour-Remain, associated with wanting more spending on everything except Defence, Business, Pensions and Criminal Justice. The second dimension is oriented from Conservative-Remain to Labour-Leave, associated with wanting more spending on everything except the EU budget, culture and overseas aid.

However, once we incorporate the residual preference variation, principal components analysis largely fails to identify a sufficient set of dimensions. There is a relatively predictive first dimension (with the same interpretation as above) followed by five additional dimensions with eigenvalues greater than 1. The Conservative-Leave to Labour-Remain dimension is more predictive of spending preferences across areas than any other dimension, but not much more predictive. Preferences over spending in different categories vary in a complex way across individuals and demographic groups that does not reduce to some groups simply being in favour or more or less spending, or more left-wing versus right-wing spending priorities. There is some of this kind of structure, but it does not explain that much of the variation. That preferences are not simply unidimensional should not surprising given what we know about the high level of individual-level idiosyncrasy in citizens' political preferences in other domains (Broockman, 2016; Lauderdale, Hanretty and Vivyan, 2018), even though the idea of a dominant left-right economic dimension is widespread in the political science and political economy literature.

Out-of-sample Validation

The average preferred bundle of spending changes shown in Figure 2 is outside the domain of the experimental design, and thus constitutes a model-based extrapolation, in two senses. First, the experiment never proposed changes in more than three spending categories. Second, while the estimated average preferred changes in every spending category are within the range of

treatments, their cumulative effect on the tax level is not. To test such a large tax hike while only distributing its effects among three spending areas would have required incredibly large changes to those areas. Thus it is reasonable to ask whether we should trust the estimates from our model for such an extrapolation. Might UK citizens' willingness to accept small tax increases in exchange for spending increases in their preferred areas falter when faced with the total tax bill it entails?

To validate our estimates, we conducted a follow-up survey. Because we estimate not only the average preferred preferences of UK citizens (via the μ parameters), but also their distribution and relative weight (via the σ, ρ and λ parameters), we can estimate what proportion of respondents we would expect to support a proposal to change all spending and tax levels to the average preferred level, versus the status quo. According to our structural model, we calculate that 72.7% of respondents should prefer the proposal to the status quo (SE: 0.027; 95% interval: 0.68-0.787), and 27.3% should prefer the status quo, according to whether they have positive or negative latent utility differences $u_{iA} - u_{iS}$. In the original experiment, status quos were preferred nearly twice as often as the proposed changes, so this is a stringent test. Have we found a set of changes that is substantially preferred to the status quo, even when most proposed changes were not preferred in the original experiment?

Assessing changes to all spending categories is a more demanding comparison for respondents than the three change comparisons in the original experiment, and so the not sure response mechanism is unlikely to be exactly the same. To maximise comparability, we define our main quantity of interest for the validation to mirror the original structural model analysis, but allowing for a potentially different rate of "not sure" responses. As before we assume a response model:

- $Y_i =$ proposed alternative if $u_{iA} - u_{iS} > 1$
- $Y_i =$ not sure if $-1 \leq u_{iA} - u_{iS} < 1$
- $Y_i =$ status quo if $u_{iA} - u_{iS} \leq -1$

Since we are only analysing a single, static question, we model $u_{iA} - u_{iS} \sim N(\mu_{validation}, \sigma_{validation})$. Having estimated $\hat{\mu}_{validation}$ and $\hat{\sigma}_{validation}$, we report the implied proportion of respondents with positive utility differences for the alternative over the status quo $\Phi(\hat{\mu}_{validation}/\hat{\sigma}_{validation})$ as our

primary quantity of interest. If "not sure" responses are very rare, this will equal the data proportion supporting the proposed alternative. If not sure responses are moderately frequent, it reallocates "not sure" responses to the status quo and proposed alternative roughly proportionately.

Our validation study was also conducted by YouGov, using the same sampling methods from their UK panel, and we asked a single question of 3,000 respondents. We presented respondents with the status quo figures for each spending category and the overall tax rate from the baseline tax summary, and then asked them to compare that with a proposal that involved changing all spending categories to the average preferences estimated from our model. The entire prompt, as delivered, is provided in supplemental information. We pre-registered this validation study, the statistic described above, and our standards for a successful validation with EGAP before data collection.[9]

Validation results

The results of the validation were as follows. Of the population-weighted respondents, 39% indicated that they supported the proposed changes, versus 25% who indicated that they preferred the current spending levels. 36% indicated that they were not sure. Using the estimator that we pre-registered, we calculate that this response distribution corresponds to 58.3% of respondents preferring the proposed changes to the status quo versus our structural model prediction of 72.7%. According to our pre-registered assessment criteria, this counts as a "moderate success". We were able to use the experiment to identify a profile of changes that was supported by substantially more respondents than opposed it, even though the level of support was not as high as the structural model implied.

In the pre-registration, we identified three reasons to expect that fewer respondents would endorse this proposal than our structural model implied. The first of these was that "the data from our original experiment shows that respondents prefer the status quo twice as often as our generated proposals. To the extent this reflects a status quo bias in responses, as opposed to substantive preferences for elements of the status quo, our proposal may fail to increase sup-

[9] Our pre-registration documents (Barnes, Blumenau and Lauderdale, 2019) can be found at (http://egap.org/registration/5550).

port to the predicted level even though it is in some sense optimised to include the elements that were most supported in the experiment" (Barnes, Blumenau and Lauderdale, 2019, 6). The second of these was that "the tax hike implied by our model is considerably larger than any tax increase we tested in the experiment, and may well have large effects on aggregate approval for the proposal. While the structural model assumes quadratic loss, and therefore builds in some of this aversion to large changes, any extrapolation relies on functional form assumptions that may not hold" (Barnes, Blumenau and Lauderdale, 2019, 6). The third of these was that "this validation study includes changes to thirteen categories rather than the three included in the original experiment. This is a much more difficult task for respondents than our original experiment. More importantly, unlike the original experiment, it is almost guaranteed that every respondent will see at least one change to a spending area that they do not like. If respondents are loss-averse this could significantly depress support for the proposed changes" (Barnes, Blumenau and Lauderdale, 2019, 7).

Moving beyond the basic evaluation set out in the pre-registration, we can examine whether the covariate model estimates from the main experiment that we presented in figure 4 indicate the same patterns of support for the proposal that we observe in the validation study, as a function of the covariates in that model. We make this comparison by post-stratifying the structural model estimates onto the validation survey respondents, and then characterizing the relationship between predicted support levels and actual support for the proposal among those respondents. Figure 5 shows the results of this analysis. While the observed levels of support for the proposal are consistently lower than the predicted levels by the roughly 15 percentage points we observed above, they are strongly associated with those predictions, with a slope close to one and only minor non-linearity in the relationship (left panel). When the individual respondents are aggregated into the same categorical groupings used previously (right panel), there is again a strong association between the predicted and observed levels of support. The demographic groups that the structural model indicated should be more favourable towards the proposal (which was itself the estimated average preference of the whole population) are in fact more favourable towards that proposal.

What do these patterns tell us about the general usefulness of the approach that we have

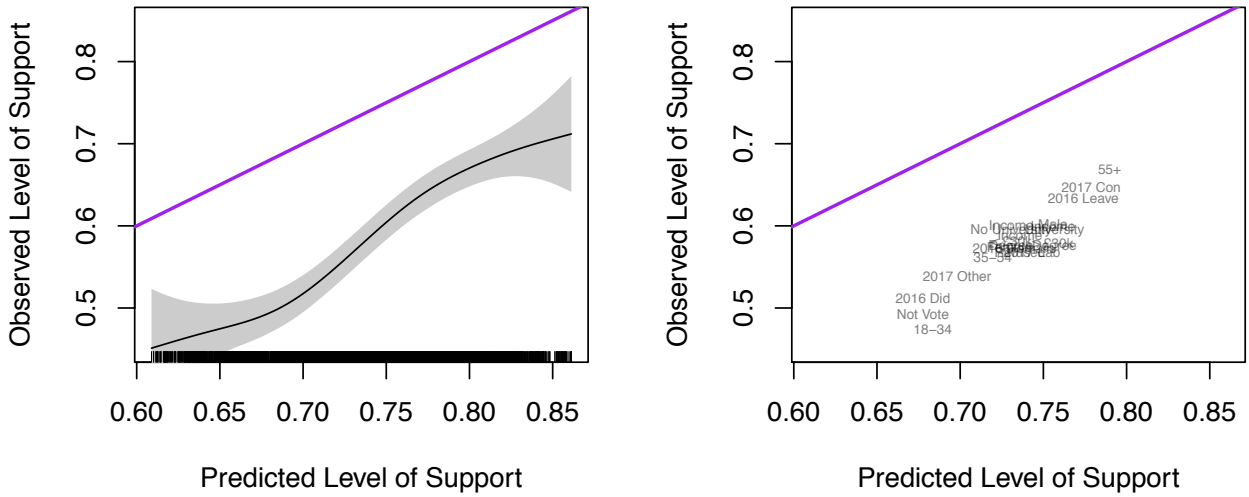

Figure 5: Left: Observed level of support for proposal in validation study as a function of predicted level of support given the set of explanatory variables in the multivariate model presented in Figure 6. Perfect agreement between the predictions and the validation would correspond to the purple line; the black line is a spline fit to the validation data. Right: Observed level of support as a function of predicted level of support for different groups.

developed in this paper? The rate of people stating that they were not sure was higher in the validation (36%) than in the original study (30%). This was part of our motivation for using comparisons involving small subsets of categories in the original experiment; however the increase in the "not sure" rate was modest, and so might indicate that we could have tested proposals involving more (or all) categories in the original experiment. There are still good reasons to think this might not be a good idea, generating a proposal distribution for changes over all categories and getting individual respondents to do more than one of these comparisons would be difficult.

Do these results undermine the estimates we produced of the average preferred spending (and thus tax) levels? To the extent they indicate that respondents do not approximate the choice logic underlying our structural model, that does imply lesser confidence in the overall spending level of the estimates as the "ideal" for the UK population. Nonetheless, our approach did identify a set of spending level changes that received substantially more support than opposition, despite high survey item complexity, potential for loss aversion, status quo bias, and other factors. However, even if we did not in fact identify the ideal spending and tax level overall, the predictive performance with respect to demographic subgroups gives us good reason to be confident in our estimates of the relative preferences for different spending areas.

Design Limitations

In designing our experiment, we aimed to provide the easiest possible comparisons to respondents that would still allow us to estimate a multidimensional distribution of tax and spending preferences. In the theoretical world of the choice model, if we gave respondents more difficult comparison tasks it would be possible to more precisely estimate relevant features of respondents' preferences in several respects. First, our data would be more informative if we did not always use the status quo as an alternative in the choice. However, in practice, even if we did not use the status quo as an alternative, we would probably need to include it as a reference so that people had a reasonable anchor for expressing their views, so this would make the task substantially more difficult. Second, a design that varied the number of spending categories that were changing might enable us to better understand the mechanisms behind "not sure" responses, and perhaps make a more accurate prediction regarding the extent of support for untested profiles of spending changes like the one we considered in the validation. The logical end point of this would be to simply do an entire experiment consisting only of the sort of comparison we did in the validation study, where all categories change. We did not adopt this strategy because we were concerned it would simply prove too difficult for many respondents, but it is something to consider for future research.

We did not adjust the presentation according to the income levels of the respondents. This means that high income people saw smaller (pound) tax increases than they would actually face, and low income people saw larger ones. In principle one could fix this, particularly with an online panel where there are pre-existing measures of income that could be used to condition the presentation of the tax values. However, we noted earlier that the estimated values of λ were inverse proportional to spending levels in each area, which suggests people penalise deviations from their preferred spending levels according to percentage rather than pound differences.

The range of treatments on each dimension is an important design consideration. The reason to include a wider range of values is that this range limits the possible range of average preferred points that can be reliably identified. Figure 2 shows that the estimated average preferred level of health spending is very close to the top of the range of tested treatment levels, and the prior that we put on the average spending level for each category matches that treatment distribution.

Thus it is possible that even higher levels of health spending would have been preferred (despite the necessary tradeoffs) if we had offered them. Several other categories are estimated at levels that approach the edges of the range of spending changes that we proposed. The reason we did not use a more variable treatment distribution is that it would have given us less precision for locating the preferred point because of having a lower density of treatment values in any given range. Extreme change proposals in a given spending area are more likely to be very unpopular than moderate ones, and so we were concerned that widely variable proposals would tell us less about the spending levels that most people want. It was difficult to know ex ante how variable our proposals should be; if we had to do it again we would increase the ranges moderately.

Applications and Extensions

Public spending choices are multidimensional, and subject to real (but endogenous) budget constraints. Having an appropriate strategy to measure preferences in this kind of context, as we have developed in this paper, allows us to improve our understanding of similarly-structured questions across political science.

First, this approach translates readily to contexts other than the British national budget, with appropriate 'domestication' of the categories. In the United States, for example, a parallel implementation might use categories from the President's published budget, as in Bonica (2015). There is no general *right* answer regarding which categories to use, but the choice ought to be as accessible as possible to a typical citizen.

Our approach can also provide a way to gauge preferences over spending tradeoffs *within* policy areas. For example, within the area of education, voters may prefer allocating more resources to early-years intervention, at the cost of either lower K-12 spending or a higher tax bill. Considering preferences over budget allocations within narrower policy areas using a forced choice and structural model strategy can provide richer descriptions of detailed within-policy budget priorities.

More broadly, if we are interested in preferences over realistic government spending choices, then we need to field survey questions which allow for the expression of preferences structured by budget realities. The measurement strategy outlined here could be usefully incorporated

into existing large-scale survey projects in order to measure these preferences in a quantitatively meaningful way over time and across space. Including 'domesticated' versions of the tax and spending allocations in national election studies would provide a consistent way to track relative spending priorities, and the relative priority of tax reductions, within countries over time. Similarly, forced choices over budget allocations in large cross-national surveys like the European Social Survey could be used to measure and study differences in preferences across countries. One natural cross-country version of our experiment would be to tailor the budget categories to reflect spending by the European Union, and to use our design to measure average preferences across these areas in different EU member states.

Second, researchers can use this strategy to measure spending preference outcomes in bespoke experiments, including – and perhaps especially – in the context of studies focused on one area of social policy spending. One problem that studies focused on a given area of policy face when asking about budgetary preferences is that questions about spending in that area typically follow a series of questions about that spending area, creating potential for both framing and demand effects. Our approach has similar advantages to the use of conjoint experiments as a way to address social desirability bias, as it ensures that no one is asked for a preference on a given dimension that could not be alternatively explained in terms of contrasts on other dimensions.

We have not exhausted the kinds of analysis that can be done with the data we have collected already. Because our estimation strategy enables estimation of a linear model for preferences in terms of covariates, the derived preferences can be readily used to generate descriptions of how spending preferences are geographically distributed using multilevel regression and poststratification (MRP) techniques. This would generate geographic estimates of public budget preferences which might be useful for assessing campaigns, intra-party variation in budget preferences of representatives, and other research questions. Finally, there are also variations on the structural model that we have specified which would make somewhat different response assumptions. Exploration of the space of possible modelling strategies is an important avenue for further research to the extent that researchers begin to deploy this kind of experimental design.

Conclusions

We have demonstrated a new strategy to measure spending preferences multidimensionally and quantitatively. We elicit quantitative preferences on narrow spending tradeoffs from a sample of respondents who are representative of the broader public on measurable characteristics, and translate those into estimates of a distribution of public spending preferences. Doing this involves assuming coherent preferences, but we have argued this is a sensible approach because it enables us to find a distribution of coherent budget preferences that is consistent with the choices that citizens make when faced with accessible tradeoffs between spending categories and taxation. Whether or not individual members of the public in fact hold or can articulate consistent budgetary preferences, it is worth knowing which feasible budgets are likely to best match citizens' assessments of the relevant tax and spending tradeoffs.

Our approach to the study of spending preferences looks different from previous strategies in many ways, but also serves as an example of three broader methodological principles which are applicable for the improvement of a much broader range of empirical investigation. First, and most narrowly, our approach highlights the utility of using continuous treatments within multivariate choice experiments, of which conjoint experiments are one prominant example. Since the central tool for analysing the data generated in conjoint experiments is regression analysis, and regression analysis is suited for the analyses of continuous variables, the focus on qualitative attribute levels in conjoint approaches often represents a missed opportunity to generate more precisely quantifiable effects. Second, structural models are a very useful tool for estimating aggregate quantities of interest in contexts where we have insufficient data to generate complete descriptions of the distributions at the individual level and where regression-based approaches are either inapplicable or generate relatively uninteresting causal estimands. Third, this paper emphasises the importance of validation. In any approach based on a parametric model, we should be concerned about the degree to which the assumptions made about functional forms drive the results. The validation exercise not only provides greater confidence in the model and the results obtained, but also highlights specific areas for further development.

Substantively, we find that current levels of government spending in the UK are significantly lower than the average preferences of the UK public, and that this is true across many differ-

ent spending categories. Crucially, as the set-up of our experiment forces respondents to take government budget constraints seriously, the public expresses these preferences for increased spending even when taking into account increases in taxation required to fund them. We showed that a public budget including an implied tax hike of approximately 7% received majority support amongst those expressing an opinion. We are also able to identify demographic variation in preferences for more spending that operates in different directions on different spending areas. Our finding that it is older voters, not younger voters, who want to substantially increase UK government spending is clear in the results of both the original experiment and the validation study and is flatly at odds with the conventional wisdom regarding age and political preferences in the UK. It is not impossible that this is an artifact of the fact that our data comes from an internet panel, but the fact that it persists conditional on gender, education, income, EU referendum vote and general election vote suggests it is not the result of any very simple bias in the sample selection. This too merits further investigation.

References

Abramson, Scott F, Krohan Kocak and Asya Magazinnik. 2019. What Do We Learn About Voter Preferences From Conjoint Experiments? In *2019 Conference of the Society for Political Methodology*.

Arceneaux, Kevin. 2002. "Direct democracy and the link between public opinion and state abortion policy." *State Politics & Policy Quarterly* 2(4):372–387.

Ballard-Rosa, Cameron, Lucy Martin and Kenneth Scheve. 2017. "The Structure of American Income Tax Policy Preferences." *The Journal of Politics* 79(1):1–16.

Barnes, Lucy, Jack Blumenau and Benjamin E Lauderdale. 2019. "Measuring UK Attitudes towards Public Spending: Validation of Predictions from a Multivariate Tax Summary Experiment." *EGAP* (ID: 20190301AA).
URL: *http://egap.org/registration/5550*

Bechtel, Michael M., Federica Genovese and Kenneth F. Scheve. 2017. "Interests, Norms and Support for the Provision of Global Public Goods: The Case of Climate Co-operation." *British Journal of Political Science* pp. 1–23.

Berinsky, Adam J. 2008. "Survey Non-Response." *The SAGE Handbook of Public Opinion Research* pp. 309–321.

Bonica, Adam. 2015. "Measuring public spending preferences using an interactive budgeting questionnaire." *Research & Politics* 2(2):2053168015586471.

Branham, J Alexander and Stephen A Jessee. 2017. "Modeling Spending Preferences & Public Policy." *Electoral Studies* 49:155–172.

Bremer, Björn and Reto Bürgisser. 2019. "Attitudes Towards Fiscal Consolidation: Evidence from Survey Experiments in Four European Countries." Working paper.

Broockman, David E. 2016. "Approaches to Studying Policy Representation." *Legislative Studies Quarterly* 41(1):181–215.

Busemeyer, Marius R., Julian L. Garritzmann, Erik Neimanns and Roula Nezi. 2018. "Investing in education in Europe: Evidence from a new survey of public opinion." *Journal of European Social Policy* 28(1):34–54.

Carpenter, Bob, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A. Brubaker, Jiqiang Guo, Peter Li and Allen Riddell. 2016. "Stan: A probabilistic programming language." *Journal of Statistical Software* .

Caughey, Devin, Tom O'Grady and Christopher Warshaw. 2017. "Policy ideology in European Mass Publics, 1981-2014.".

Citrin, Jack. 1979. "Do People Want Something for Nothing? Public Opinion on Taxes and Government Spending." *National Tax Journal* 32(2):113–129.

Clinton, Joshua, Simon Jackman and Douglas Rivers. 2004. "The statistical analysis of roll call data." *American Political Science Review* 98(2):355–370.

Downs, Anthony. 1960. “Why the Government Budget is Too Small in a Democracy.” *World Politics* 12(4):541–563.

Edlund, Jonas and Ingemar Johansson Sevä. 2013. “Exploring the Ä‘Something for Nothing’Äô Syndrome: Confused Citizens or Free Riders? Evidence from Sweden.” *Scandinavian Political Studies* 36(4):293–319.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9477.2012.00300.x*

Fieldhouse, Edward, Jane Green, Geoffrey Evans, Hermann Schmitt, Cees Van der Eijk, Jonathan Mellon and Christopher Prosser. 2017. “British election study, 2017: Face-to-Face Post-Election Survey.” *UK Data Service* .
URL: *http://doi.org/10.5255/UKDA-SN-8418-1*

Grasso, Maria Teresa, Stephen Farrall, Emily Gray, Colin Hay and Will Jennings. 2017. “Thatcher’s children, Blair’s babies, political socialization and trickle-down value change: An age, period and cohort analysis.” *British Journal of Political Science* pp. 1–20.

Hainmueller, J., D. J. Hopkins and T. Yamamoto. 2014. “Causal inference in conjoint analysis: Understanding multi-dimensional choices via stated preferences experiments.” *Political Analysis* 22(1):1–30.

Hainmueller, Jens and Daniel J. Hopkins. 2015. “The Hidden American Immigration Consensus: A Conjoint Analysis of Attitudes toward Immigrants.” *American Journal of Political Science* 59(3):529–548.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12138*

Hansen, John Mark. 1998. “Individuals, institutions, and public preferences over public finance.” *American Political Science Review* 92(3):513–531.

Häusermann, Silja. N.d. *The Politics of Welfare State Reform in Continental Europe: Modernization in Hard Times*. Cambridge University Press.

Hausman, Jerry. 2012. “Contingent Valuation: From Dubious to Hopeless.” *Journal of Economic Perspectives* 26(4):43–56.

Hübscher, Evelyn, Thomas Sattler and Markus Wagner. 2018. “Voter Responses to Fiscal Austerity.” Paper presented to the EPSA annual conference.

Kastellec, J.P., J.R. Lax and J.H. Phillips. 2010. “Public opinion and Senate confirmation of Supreme Court nominees.” *Journal of Politics* 72:767–784.

Kelly, Nathan J. and Peter K. Enns. 2010. “Inequality and the Dynamics of Public Opinion: The Self-Reinforcing Link Between Economic Inequality and Mass Preferences.” *American Journal of Political Science* 54(4):855–870.

Lauderdale, Benjamin E., Chris Hanretty and Nick Vivyan. 2018. “Decomposing Public Opinion Variation into Ideology, Idiosyncrasy, and Instability.” *Journal of Politics* 80(2):707–712.

Lax, Jeffrey R and Justin H Phillips. 2009*a*. “Gay rights in the states: Public opinion and policy responsiveness.” *American Political Science Review* 103(3):367–386.

Lax, Jeffrey R and Justin H Phillips. 2009*b*. “How should we estimate public opinion in the states?” *American Journal of Political Science* 53(1):107–121.

Meltzer, Allan H and Scott F Richard. 1981. "A rational theory of the size of government." *Journal of political Economy* 89(5):914–927.

Mettler, Suzanne. 2011. *The Submerged State: How Invisible Government Policies Undermine American Democracy*. Chicago: University of Chicago Press.

Poole, Keith T. 2001. "The geometry of multidimensional quadratic utility in models of parliamentary roll call voting." *Political Analysis* 9(3):211–226.

Svallfors, Stefan. 1997. "Worlds of Welfare and Attitudes to Redistribution: A Comparison of Eight Western Nations." *European Sociological Review* 13(3):283–304.

Williamson, Vanessa S. 2017. *Read My Lips: Why Americans are Proud to Pay Taxes*. Princeton and Oxford: Princeton University Press.

Winter, Søren and Poul Erik Mouritzen. 2001. "Why people want something for nothing: The role of asymmetrical illusions." *European Journal of Political Research* 39(1):109–143.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/1475-6765.00572*

Supplemental Information: Measuring Attitudes towards Public Spending using a Multivariate Tax Summary Experiment

Contents

Tax Summary and Experiment Screenshots

How your tax was spent in 2013-14

The information on this page shows you how your Income Tax and National Insurance contributions were spent in 2013-14. This does not include indirect taxes such as VAT and other duties.

For more information about your tax and public spending, go to **www.gov.uk/annual-tax-summary**

How your tax contributed to public spending

	Your contribution
Welfare	£1113
Health	£857
Education	£597
State Pensions	£550
National debt interest	£318
Defence	£241
Criminal justice	£200
Transport	£134
Business and industry	£125
Government administration	£93
Culture eg sports, libraries, museums	£77
Environment	£75
Housing and utilities eg street lights	£75
Overseas aid	£52
UK contribution to the EU budget	£34
Total	**£4542**

All figures are rounded to the nearest pound.

The figures in the table above are intended as a guide to how taxes are spent and not as a direct link between your Income Tax, National Insurance contributions and any specific expenditure.

If you would like to opt out of receiving future tax summaries please register at **www.gov.uk/annual-tax-summary**

Spending information is published by HM Treasury.

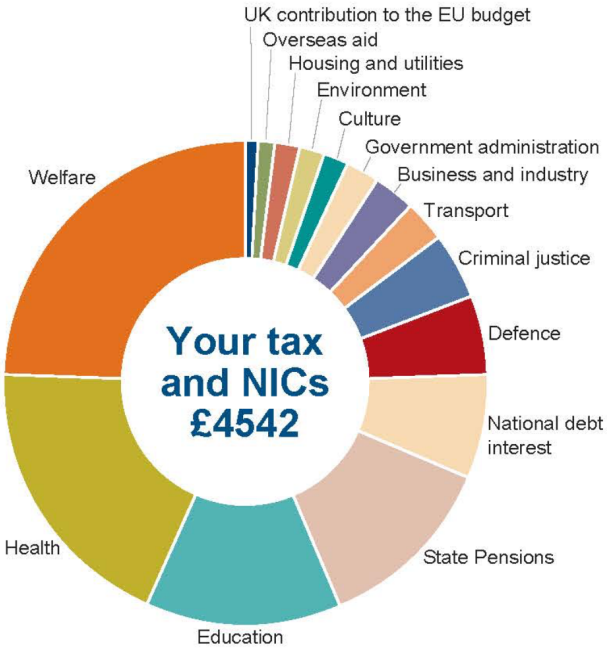

Figure 1: Example of UK Tax Summary

YouGov

We are now going to ask you a few questions about how the UK government spends the money you pay in tax.

Under the current tax system, an individual in the UK earning £23200 each year would typically pay £4078 in income tax and national insurance contributions. The table below shows how these taxes are currently allocated to different spending categories.

	Current level (£)
Welfare	991
Health	828
Education	526
State Pensions	502
National Debt Interest	224
Defence	212
Criminal Justice	171
Transport	171
Business and Industry	102
Government Administration	86
Culture	65
Environment	65
Housing and Utilities	61
Overseas Aid	45
UK contributions to the EU budget	29
Total tax (£)	4078

The purpose of this task is to compare current tax and spending levels to alternative tax and spending proposals. The proposals might change the overall amount of tax paid by individuals, or change the amount spent on different categories, or both. You will be asked to indicate which tax and spending levels you would prefer.

Figure 2: Experiment introduction screen

YouGov

We are now going to ask you about how the UK government spends the money you pay in tax.

Under the current tax system, an individual in the UK earning £23200 each year would typically pay £4078 in income tax and national insurance contributions.

The table below shows how these taxes are currently allocated to different spending categories ('Current') and also an alternative proposal, where the amount of tax and how that tax is spent are different ('Proposed').

	Current (£)	Proposed (£)	Change (£)	Change (%)
Health	828	951	+£123	+15%
Welfare	991	1061	+£70	+7%
State Pensions	502	565	+£63	+13%
Education	526	584	+£58	+11%
Criminal Justice	171	183	+£12	+7%
Housing and Utilities	61	73	+£12	+19%
Environment	65	76	+£11	+17%
Transport	171	180	+£9	+5%
Defence	212	218	+£6	+3%
National Debt Interest	224	224	£0	0%
Government Administration	86	86	£0	0%
Culture	65	55	-£10	-15%
Business and Industry	102	89	-£13	-12%
UK contributions to the EU budget	29	12	-£17	-58%
Overseas Aid	45	20	-£25	-55%
Total tax	4078	4377	+£299	+7%

Under the proposed plan, compared with current levels, both total spending and total taxes would increase by an average of £299 (7%) per person.

You may like some of the proposed changes and dislike others. Overall, which do you prefer?

- I prefer the current levels
- I prefer the proposed changes
- I am not sure

Figure 3: Validation survey prompt

Components of Tax Summary Spending Areas

The tax summaries shown above, and which form the basis of our experiment, are based on information relating to public expenditure on services (both current and capital spend) which is published each year by the Treasury. The categories are based broadly around the UN's Classifications on the Functions of Government (COFOG), but the relevant classifications have been simplified for public consumption.

Table 1 below provides more information on the spending items that constitute the categories that we provide to respondents in the experiment. In particular, it shows how each of the spending categories is decomposed into the different service "sub-functions" as defined by the Public Expenditure Spending Analyses (PESA) produced by the government each year (see table 5.2 in Treasury, 2018). Note that this information is included here for completeness: only the spending category titles were presented to respondents to our survey.

Table 1: Spending categories – PESA definitions

Spending category	Description
Health	Medical services; Medical Research; Central and other health services
Welfare	Personal social services; Incapacity, disability and injury benefits; Family benefits, income support, Universal Credit and tax credits; Other unemployment benefits; R&D social protection
State Pensions	State pension
Education	Primary and pre-primary education; Secondary education; Post-secondary non-tertiary education; Tertiary education; Education not definable by level; Subsidiary services to education; R&D education
Criminal justice	Police services; Fire-protection services; Law courts; Prisons; R&D public order and safety

Spending category	Description
Housing and Utilities	Housing development (local authority and other social housing); Community development; Water supply; Street lighting; R&D housing and community amenities
Environment	Waste management; Waste water management; Pollution abatement; Protection of biodiversity and landscape; R&D environment protection
Transport	National roads; Local roads; Local public transport; Railway; Other transport
Defence	Military defence; Civil defence; Foreign military aid; R&D defence
National Debt Interest	Central government debt interest; Local government debt interest; Public corporation debt interest
Government Administration	Executive and legislative organs, financial and fiscal affairs, external affairs; General services; Basic research; R&D general public services
Culture	Recreational and sporting services; Cultural services; Broadcasting and publishing services; Religious and other community services; R&D recreation, culture and religion
Business and Industry	General economic, commercial and labour affairs; Agriculture, forestry, fishing and hunting; Fuel and energy; Mining, manufacturing and construction; Communication; Other industries; R&D economic affairs
UK contributions to the EU budget	VAT-based and GNI-based contributions; EU receipts; Attributed aid and Common Foreign and Security Policy
Overseas aid	Foreign economic aid

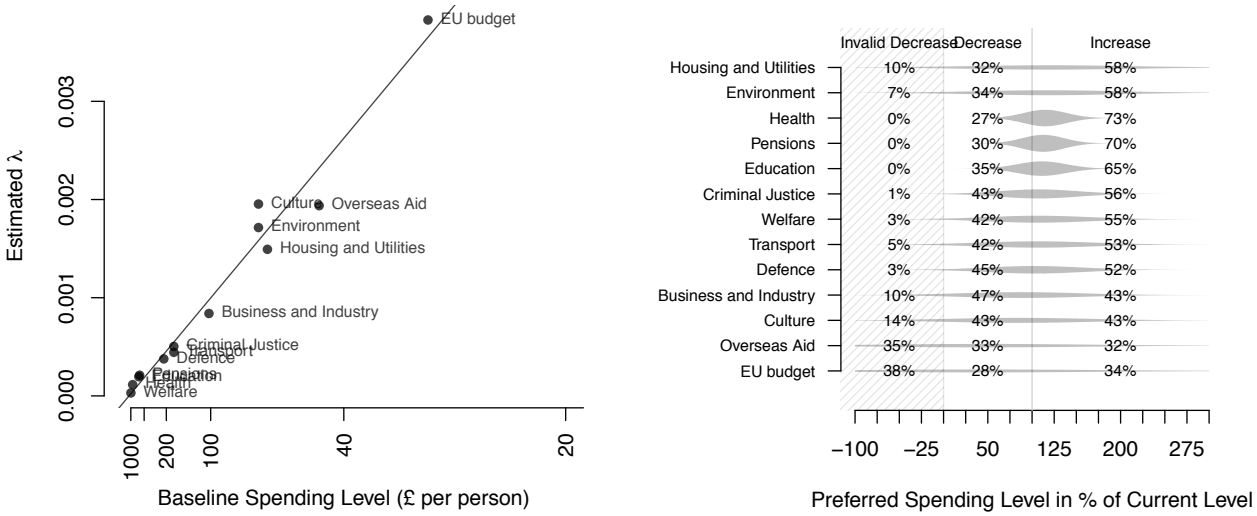

Figure 4: Left: estimated dimension weights as a function of baseline spending level (inverted scale). Right: estimated marginal distribution of preferences on each spending dimension.

Incidental Model Parameters

In the main text of the paper, we focus on the average preference parameters μ_j for each spending dimension j, here we provide additional details about the incidental model parameters λ_j, σ_j and $\rho_{jj'}$. The left panel of Figure 4 shows that the estimated values of λ_j are inversely proportional to the baseline spending levels in each category. This indicates that respondents are thinking in percentage terms, rather than pound terms, in making tradeoffs between deviations from their preferences on different spending areas.

Our model assumes normally distributed preferences regarding spending allocations, but spending cannot be less than zero in reality. It would be problematic if we were estimating σ_j parameters that were large enough relative to μ_j such that a substantial fraction of respondents preferred negative spending in many areas. We do estimate substantial bounds violations in a few of the smallest spending categories, and particularly for the two that are least popular (Figure 4, right panel). Taking the model very literally, we estimate that 38% of respondents want less than zero EU spending, 35% want less than zero aid spending, and 14% want less than zero culture spending. Note, however, that a version of the model in which we assumed quadratic loss in log spending rather than in spending eliminates the possibility of these bounds violations. The results for the main quantities of interest—the average preferred spending levels—are extremely

similar under this variation to the model (Figure 6). These estimates, combined with the results for the λ_j parameters suggest that the model with quadratic loss in log-spending rather than spending may be generally preferable to the main model that we present. We have kept the model for spending as the primary presentation in the paper because the estimates from that model were used as the basis for the validation experiment.

The preference correlation parameters $\rho_{jj'}$ shown in 5 are only weakly identified by the data. It is perhaps more surprising that we are able to estimate these preference correlations from our experiment at all. The "not sure" response option helps make this possible. When we observe an increase in spending on health but a decrease in spending on education, we get more not sure responses because people are conflicted (either because they want more of both or more of neither, but seldom only one).

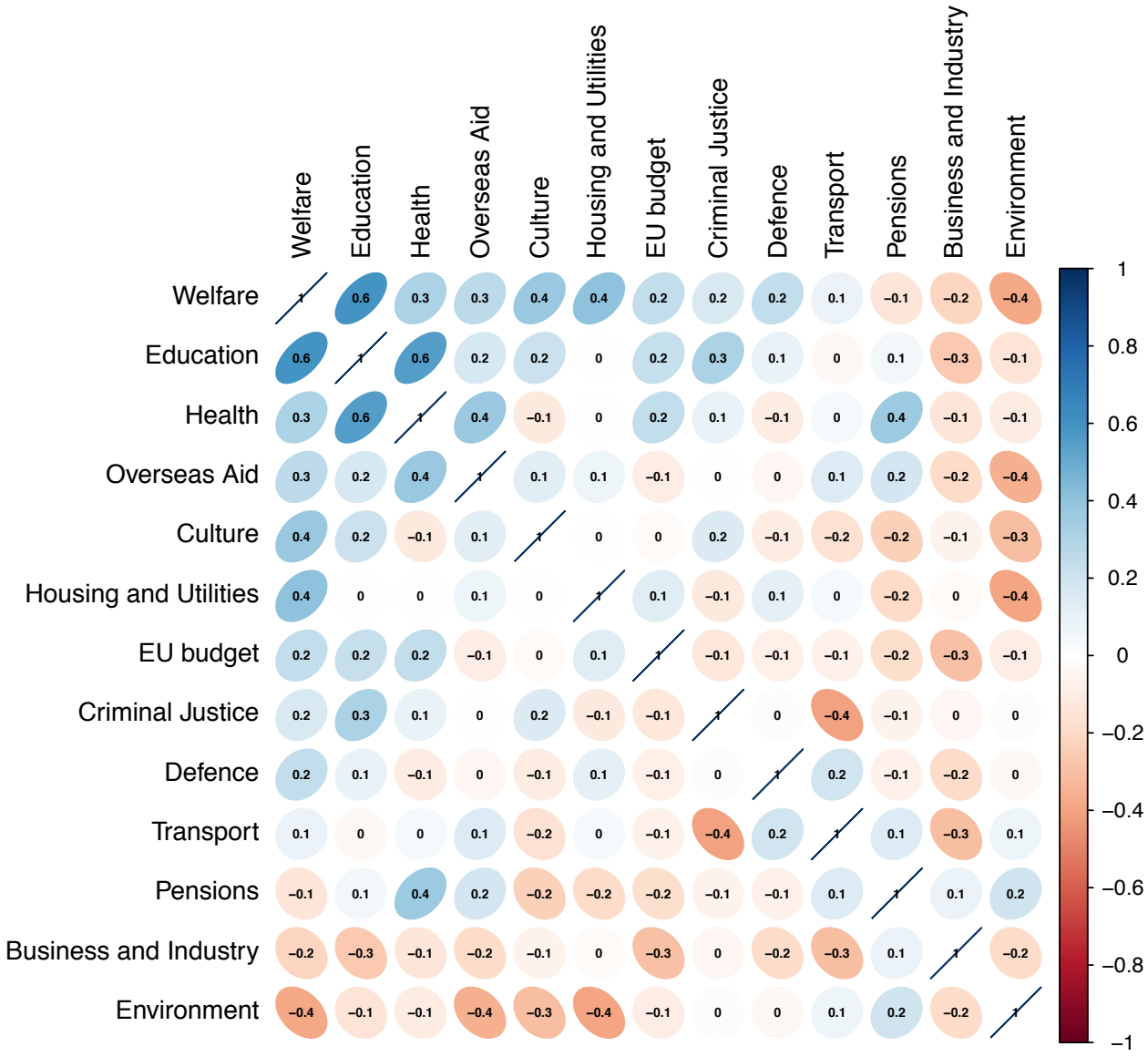

Figure 5: Estimated preference correlations across dimensions. Spending areas ordered by first principal component.

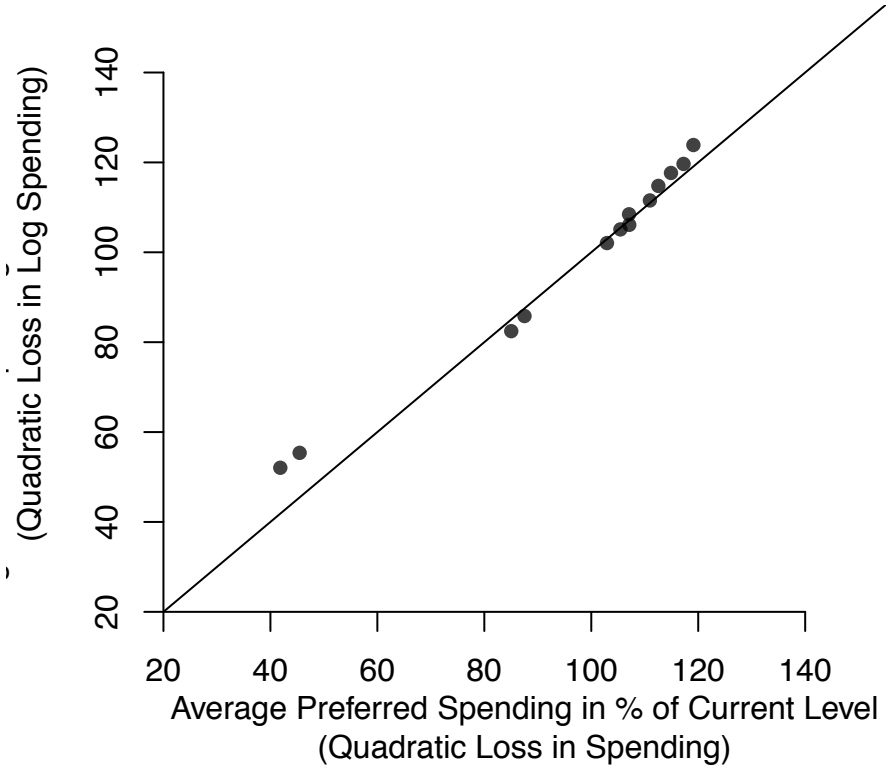

Figure 6: Estimated average spending preferences assuming quadratic loss in log-spending versus spending (the main model estimates).

Estimates Under Alternative Model Specifications and Data Subsets

Table 2: Estimates of average preferred percent changes to spending and taxation levels for the main model, the model with quadratic loss in log spending, the model dropping respondents who gave not sure respondents to all five problems, a covariate model with dummy variables for each response round, and a model using only responses to tax neutral tradeoffs.

	Main Model	Log Model	Drop 5x Not Sure	Response 1	Response 2	Response 3	Response 4	Response 5	Tax Neutral Only
Defence	1.03	1.02	1.03	1.03	1.03	1.03	1.03	1.03	0.93
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)
Welfare	1.07	1.09	1.06	1.06	1.06	1.07	1.08	1.07	0.97
	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.06)
Education	1.11	1.12	1.09	1.11	1.11	1.11	1.12	1.11	1.07
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
Health	1.15	1.18	1.13	1.13	1.15	1.15	1.15	1.16	1.12
	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
EU budget	0.42	0.52	0.43	0.42	0.42	0.42	0.41	0.42	0.45
	(0.08)	(0.04)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)
Business and Industry	0.88	0.86	0.89	0.88	0.88	0.87	0.88	0.87	0.78
	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.06)
Pensions	1.13	1.15	1.12	1.13	1.12	1.12	1.12	1.14	1.10
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
Criminal Justice	1.07	1.06	1.07	1.06	1.08	1.07	1.06	1.08	0.99
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.06)
Environment	1.17	1.20	1.18	1.17	1.17	1.17	1.18	1.17	1.08
	(0.03)	(0.05)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Transport	1.05	1.05	1.05	1.05	1.06	1.06	1.05	1.05	0.95
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Culture	0.85	0.82	0.86	0.85	0.85	0.84	0.85	0.85	0.81
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Housing and Utilities	1.19	1.24	1.18	1.19	1.19	1.19	1.19	1.20	1.09
	(0.04)	(0.07)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Overseas Aid	0.45	0.56	0.53	0.45	0.45	0.46	0.45	0.45	0.52
	(0.08)	(0.04)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)
Tax / Total Spending	1.08	1.09	1.07	1.07	1.08	1.08	1.08	1.08	1.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)

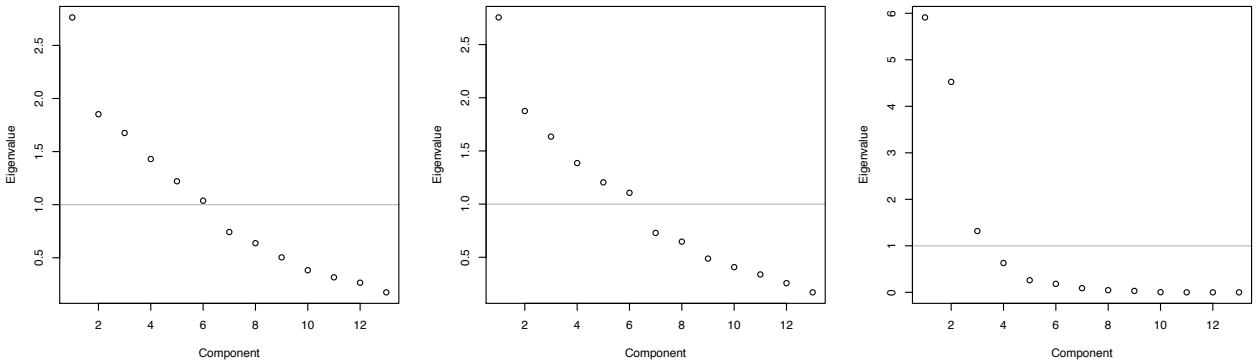

Figure 7: Screeplots for principle components analysis on the estimated distribution of spending preferences from the model without covariates (left), fitted values plus residuals from the model with covariates (center), and just the fitted values from the model with covariates (right).

Principle Components Analysis on Estimated Preference Distributions

In this section, we present results of principle components analysis on the estimated preference distributions from our model without covariates and our model with all covariates. We examine principle components from the model without covariates, the fitted values plus residual distribution from the model with covariates, and the fitted values by themselves from the model with covariates. While the former two analyses reflect the full estimated distribution of preferences, the latter only reflects the components of that distribution that we can predict with the variables in the model.

The screeplots shown in Figure 7 illustrate that while there are two relatively strong dimensions in the analysis of the fitted values, when we include the residual distribution there is a more diffuse structure that includes six dimensions with eigenvalues greater than 1. When we turn to the principle components themselves, we see that the first principle component largely runs from Conservative-Leave to Labour-Remain, associated with wanting more spending on everything except Defence, Business, Pensions and Criminal Justice. Some of the loadings for the first dimension vary in magnitude a bit between the fitted value only model and the full estimated preference distributions, but both reflect a similar dimension of preference dimension. In contrast, and consistent with the screeplots, the second dimension is only clearly interpretable for the fitted value only analysis, where it is oriented from Conservative-Remain to Labour-Leave, associated with wanting more spending on everything except the EU budget, culture and over-

Table 3: First (D1) and second (D2) principle components for the spending areas based on the model without covariates, fitted values plus residuals from the model with covariates, and just the fitted values from the model with covariates.

	D1: No Covariates	D1: Covariate Model	D1: Fitted Values Only	D2: No Covariates	D2: Covariate Model	D2: Fitted Values Only
Defence	0.09	0.05	-0.30	0.02	0.01	0.30
Welfare	0.51	0.52	0.33	-0.10	-0.15	0.20
Education	0.45	0.49	0.27	0.15	0.16	0.26
Health	0.34	0.36	0.18	0.42	0.42	0.35
EU budget	0.22	0.20	0.37	0.01	0.01	-0.14
Business and Industry	-0.22	-0.24	-0.10	-0.23	-0.19	0.26
Pensions	-0.06	-0.03	-0.04	0.52	0.51	0.46
Criminal Justice	0.13	0.19	-0.14	-0.14	-0.10	0.41
Environment	-0.31	-0.24	0.40	0.31	0.34	0.10
Transport	0.02	0.03	0.26	0.36	0.34	0.20
Culture	0.23	0.23	0.32	-0.36	-0.40	-0.23
Housing and Utilities	0.22	0.19	0.22	-0.23	-0.24	0.32
Overseas Aid	0.29	0.27	0.38	0.21	0.14	-0.13

seas aid. The second dimension of the full preference distributions broadly resembles this same dimension, but less clearly so.

References

Treasury, HM. 2018. “Public Expenditure Statistical Analyses.”.