Estimating the Political Center from Aggregate Data: The Stimson Dyad Ratio Algorithm and an Item Response Theory Approach

Anthony J. M[c]Gann

Department of Political Science
University of California, Irvine
Irvine, CA 92697-5100
949-824-6232
email: amcgann@uci.edu

November 9, 2011

Abstract

The Stimson dyad ratio algorithm produces a time series estimate of the political center from aggregate survey data, even when the same questions are not asked in most years. The justification of its claim to be measuring this central tendency, however, is open to question. As an alternative, I lay out an estimation technique that explicitly measures the central tendency of the distribution based on a widely used item response theory model of individual behavior. This has the added advantage of also estimating the dispersion of the distribution of preferences. I compare the results of both techniques using the data on public opinion from the United Kingdom from 1947-2005 from Bartle, Dellepiane-Avellaneda and Stimson (2011). Full code is provided for estimation with free software WinBUGS and JAGS.

1. Introduction

There are many contexts where we wish to estimate the aggregate level of public opinion – the location of the median voter, the position of the political center, the prevailing policy mood, or the overall level of support for a particular policy. Often we do not have individual level panel data for the entire time period we are interested in. This is not an insurmountable problem as we are interested in aggregate, rather than individual, level change. Indeed, in order to increase the coverage of the time period, we may use questions for which we only have aggregate response data. However, we face another problem. Even if we have many question items administered in each year, it is often the case that no item is administered in more than a small proportion of the years. However, if the questions that are asked in different years overlap sufficiently, it is still possible to generate a comparable time series measure for the level of public opinion for all years. The Stimson dyad ratio approach offers one method for doing this. This paper evaluates this method and proposes a different approach based on item response theory.

The Stimson dyad-ratio algorithm has now been used in a considerable number of contexts. It was developed in Stimson's work on American public opinion (Stimson 1991, 1999). It forms an important of the Erikson, McKuen and Stimson's *Macropolity* project (Erikson, MacKuen, and Stimson 2002; Stimson, Mackuen, and Erikson 1995). More recently the approach has been extended to the United Kingdom (Bartle, Dellepiane-Avellaneda, and Stimson 2011; Bartle, Dellepiane-Avellaneda, and Stimson 2011) and France (Stimson, Thiébaut, and Tiberj 2009). There have also been a number of unrelated studies that use the dyad ratio algorithm (Cohen 2000; Chanley, Rudolph, and Rahn 2000; Kellstadt 2003; Voeten and Brewer 2006; Baumgartner, De Boef, and Boydstun 2008).

The ability to measure change in public opinion over long time period, offered by the dyad ratio algorithm and other related approaches, provides the opportunity to address many outstanding problems in comparative politics. This includes not just issues of political behavior, but also political economy and comparative political institutions. For example, to operationalize some rational choice theories, it is necessary to have a measure of the position of the median voter. When studying phenomena such as the growth and retrenchment of the welfare state, it would be hugely helpful to have an independent measure of public demand for such programs. There are various studies that attempt to measure the responsiveness to public opinion of various electoral systems or constitutional arrangements (Powell 2000; McDonald and Budge 2005). The problem they face is that they are forced to use extremely indirect measures of public opinion – left-right self-placement from survey data in Powell; the declared position of the median party in parliament in MacDonald and Budge (see also Kim and Fording 2001). The methods discussed here provide a way to estimate this directly using existing data.

Given the potential of the dyad ratio algorithm to contribute to many different research areas in political science, it is necessary to evaluate it and consider alternatives. The dyad ratio algorithm is an ingenious approach that makes use of the intuition that differences between the responses to a given question item administered in different years provides information about change in public opinion. It had the advantage of being computationally quite tractable – the alternative I propose requires far greater computer resources than were generally available when the dyad ratio algorithm was developed.

The main reason to question the use of the dyad ratio algorithm is that it is not clear exactly what it is measuring. The ratio between the responses between different years surely measures something about public opinion changing; however, it is not obvious that it measures change in the "political center" or in the location of the median voter. There are no individual level micro-foundations for the claim that the position of the median voter is being measured.

As an alternative, I propose an approach that explicitly estimates the position of the median voter based on an individual level model of behavior. This adapts an established item response theory model from psychometrics. This approach has the added advantage of not only estimating the central tendency of the distribution of public opinion, but also the variance of the distribution. The estimate of the variance of the population is interesting in itself (we are often interested in whether the public has become more or less polarized) and also allows for a better estimate of the central tendency (we will see that the effect of increased polarization can disguise changes in the mean). I provide code for implementing the model using freely available software (BUGS and JAGS). I also compare the results of the item response theory approach with those using the dyad ratio algorithm, using the data on British domestic public opinion from Bartle, Dellepiane-Avellaneda and Stimson (2011)

2. The Stimson dyad ratio approach

The Stimson dyad ratio algorithm is an attempt to estimate the central tendency of the policy mood for each year from aggregate data, in spite of the fact that no question is asked in more than a small fraction of the years. In laying out this algorithm, I use the terminology from Bartle, Dellepiane-Avellaneda and Stimson (2011), as this is most recent exposition and the notation is most elegant. However, the fullest explication of the algorithm is in Stimson (1999, 133-7).

The key assumption of the algorithm is that the ratio of left responses between two years in which a given question is asked represents an estimate of the relative policy mood of these two years. It is then possible to piece together the various estimates of the relative magnitudes of policy mood to produce a series of estimates. Thus, each administered survey question is recoded so that answers are classified as either left-wing or not left-wing, and the proportion of left-wing responses is recorded for each question. If a question is asked in years t+i and t+j, then R_{ij} is the ratio of the proportion of left-wing responses to the question administered in year t+i to the proportion of left-wing responses in year t+j. Thus the key assumption of the model (Bartle, Dellepiane-Avellaneda, and Stimson 2011, 268) is:

$$R_{ij} = \frac{x_{t+i}}{x_{t+j}} \qquad (1)$$

where x_{t+i} and x_{t+j} are positions of the political center in years t+i and t+j. As is made clear in the text, what this means is that R_{ij} (for which we have observable data) can be used as an estimate of the relative magnitude of x_{t+i} and x_{t+j}. If we have assumed or imputed a value for x_{t+j}, we can use R_{ij} to estimate x_{t+i}. Of course, as there are usually multiple items that are asked in any pair of years, there are multiple estimates of the policy mood in any year. Furthermore, when estimating policy mood in year t+i we are not limited to using items that are asked in years t+i and t+j, but may use items that were asked in t+i and any other year. The authors use x_{tk} to refer to the k^{th} such estimate of the policy mood at time t.

The algorithm essentially averages together the various estimates of policy mood for each year. It starts by assigning an arbitrary value to the policy mood in the first year, x_1. Then it takes the questions that were asked in years 1 and 2, and uses equation (1) to produce a series of estimates of the policy mood in year 2, $x_{2\ 1...n}$, where n is the number of estimates of x_2. These are then averaged to produce an estimate for x_2, which is denoted P_2. Next we generate a series of estimates for x_3 by applying Equation (1), using the questions that are asked in both year 3 and either of the two previous years, and taking the estimate of x_2 from the previous round (P_2) as the value of x_2. This is repeated iteratively until we have estimates for the policy mood for all years. Thus, if P_t is the aggregate estimate of policy mood at time t, and N is the number of estimates of policy mood using Equation (1), then:

$$P_t = \frac{\sum_{k=1}^{N} x_{t\,k}}{N} \qquad (2)$$

Stimson (1999) notes that smoothing is applied to each estimate of policy mood (P_t) during the algorithm. It is also noted that the different result are obtained depending on whether we start at the first year and work forward, or start at the last year and work backwards. For this reason the results of the forward and backward runs are averaged.

The next stage is to reweight the estimates to take account of the fact that some questions measure policy mood better than others. For each question we calculate a communality estimate, h_i, which is simply the correlation of the percentage giving a left-wing response in each year the question was asked and our estimate of policy mood for that year, P_t. We then repeat the whole estimation procedure all over again using Equation 3 instead of Equation 2:

$$P_t = \frac{\sum_{k=1}^{N} h_i^2 x_{t\,k}}{h^2 N} \qquad (3)$$

where h is the average communality score. This weights each estimate of x_t by the communality of the question that was used to produce it. The new estimates of policy mood are then used to re-estimate the communalities and the process repeated again. This process is iterated until the communalities and policy mood estimates converge.

We can question the basic assumption behind the algorithm. The algorithm rests on the assumption (see Equation 1) that the ratios between the proportions of left-wing answers given in two years provide an estimate of the relative position of the political center in those years. This is taken as an assumption and is never justified. Although the justification of the concept of the policy mood references the idea of the median voter (Bartle, Dellepiane-Avellaneda, and Stimson 2011, 259), there is no individual level theory of survey response. Furthermore, the assumption is not intuitively obvious. It is reasonable to assume that changes in the proportion of people answering in a left-wing way reflect a change in public opinion; however, it is not obvious that this necessarily represents a change in the position of the political center. For example, suppose we have a question where only 10% of the population answers in a left-wing manner. Then suppose that in a subsequent year that only 5% answer the same question in a left-wing way. Do we assume that the median voter has moved sharply to the right (as the .5) ratio would imply? It might simply be the case that a few extremists have changed their views somewhat, and the political center has stayed exactly where it was.

Furthermore, the relationship between the position of the political center and the proportion of left-wing responses need not even be monotonic, unless we make other assumptions. It is possible for the political center to move to the right, but for people to be more likely to answer certain questions in a left-wing way. Again consider a question where only 10% of the population answers in a left-wing manner, that is, a question that is difficult to answer in a left-wing way. Suppose the political center moves to the right, but public opinion becomes more polarized. Because of this increased polarization, there are more left-wing extremists and now 20% of the population gives a left-wing response. We would get a dyad ratio of 2 from this question (implying a strong move to the left), even though the political center has really moved to the right. Of course, if we assume that the shape of the distribution of preferences remains identical (that is every respondent moves leftward or rightward lockstep with the median), then every dyad ratio will indicate a move in the correct direction. However, it seems implausible to assume that some time periods are not more polarized than other.

Even if the final estimates depend mostly on questions of average difficulty (either because there are few extremely easy or hard questions, or because they are assigned low communality scores), we would still expect changes in the standard deviation of the population to change the dyad ratio estimate of the political center. If public opinion is somewhat to the left of center, then an increase in the standard deviation will lead to response rates closer to 50%. This will be interpreted as a move to the right. A decrease in polarization will effectively increase discrimination and produce results further from 50%, which will be interpreted as a move to the left. This, of course, will be reversed if public opinion is right of center. The effect of changes in the standard deviation will take place with a lag, as questions are frequently not asked every year, but only every several years (and the ratio between every administration of a question is taken into account, not just those closest in time).

Equation 1 implies that the dyad ratios provide interval level information about policy mood. For example, Equation 1 implies that a move from 50% give left wing responses to 60% represents the same ratio (1.2) and thus the same change in the political center as a move from 60% to 72%. This implies very precise assumptions about how people with different policy positions respond to the questions, and these assumptions are never laid out explicitly. There is a striking lack of symmetry between left and right. A movement from 20% to 40% gives a ratio of 2, but a movement from 80% to 60% only gives a ratio of 4/3. However, if we counted the probability of right-wing answers instead of left-wing, these ratios would be reversed. This suggests that we may get different results if we calculate from the percentage of right-wing answers. This seems perplexing as the proportion of left-wing answers and the proportion of right-wing answers give exactly the same information.[1] This phenomenon is especially problematic when we consider questions that are extremely hard or extremely easy to answer in a left-wing manner. If 90% of the population answers a question in a left-wing manner in years when the political center is around its long-run median, it is not possible for this question to reflect a sharp move to the left. Even if we have 100% give the left-wing answer, we

[1] This problem could be solved within the framework of the dyad ratio approach by using the ratio of left-wing to right-wing answers = left%/(1-left%), instead of proportion of left responses.

only get a ratio of 10/9. However, if we have a question that 10% answers left-wing in an average year, a move to the right that produced a 1% left-wing response would give a ratio of 1/10.

However, in spite of all these problems it is possible that dyad ratio algorithm may approximate the position of the median voter surprisingly well. We will see in the next section that the estimates of the political center from the item response theory model are actually very similar to those produced by the dyad ratio algorithm. I suspect that reason is that the dyad ratio algorithm does not rely on individual ratios but average over many such ratios. Various biases *may* average out, although this has not been demonstrated.

The weighting of the averages with communalities probably also plays a role. Most of the problems that I have outlined have the strongest effect on questions that are either very easy or very hard to answer in a left-wing manner. Such questions are likely to correlate poorly with the estimates, and thus be assigned low communality scores. The questions that are easy to answer left-wing will correlate poorly because they cannot reflect large movements of policy mood, while question that are very hard to answer left-wing will correlate poorly because they produce ratios that exaggerate small changes. As with factor analysis and other methods based on correlation, items that are extremely easy or difficult correlate poorly with the scale as a whole and are assigned a low weight even when they scale perfectly in the Guttman sense.[2] Thus these questions will be discounted, and the final estimates will be based on questions of moderate difficulty – that is, the

[2] That is to say, if one item is more difficult than the other, the easier item is always answered correctly if the more difficult one is, and the harder one is always answered incorrectly if the easier one is.

questions that are most sensitive to the political center. The algorithm may end up tracking the political center after all.

3. An item response theory model of policy mood

As an alternative I develop a model based on item response theory (IRT). While the dyad ratio approach *may* track the political center, we have no theoretical basis for the claim that we are measuring the position of the median voter. With an IRT approach we can explicitly model the position of the median voter with a widely used model of individual behavior, and draw inferences about it. We can also extract more information from our data. While very hard or very easy questions are problematic for the dyad ratio approach, the IRT approach can use the pattern of responses to these questions to draw inferences about the entire population distribution. Thus we are able to measure the dispersion of the policy positions of the population and not just the central tendency.

Item response theory is a standard approach in psychometrics that allows us to simultaneously estimate the characteristics of test subjects and the difficulty of test questions. Its extension to political behavior is straightforward – indeed Stimson (1999, 50) when describing the dyad-ratio approach at one point falls into the testing metaphor, describing some questions as easy to agree to and some hard. Here I adapt the approach to deal with aggregate data and the estimation of population, as opposed to individual, parameters.

The standard item response theory model uses responses to test questions to simultaneously estimate the ability of the respondents and the characteristics of the questions (Nunnally and Bernstein 1994, 393-409). We assume that each individual has an ability level that is unidimensional, with β_i representing the ability level of individual i. Each question q can have various parameters, but we assume it has two, difficulty and discrimination, represented as λ_q and α_q. A commonly used functional form is the two parameter lognormal:

$$p(\textit{respondent i}\ \text{answer question q correctly}) = \frac{1}{1+e^{-\alpha_q(\beta_i-\lambda_q)}} \qquad (4)$$

This is, of course, closely related to logistic regression, which is commonplace in political science. Figure 1 graphs this response function for three questions with different values of λ and α.

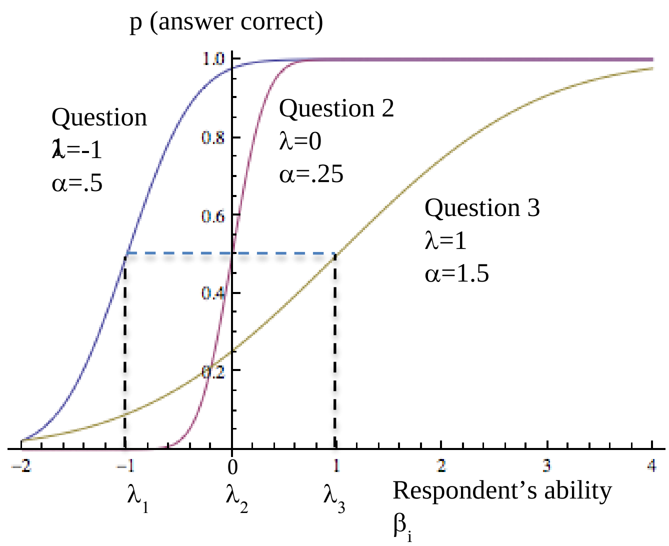

Figure 1 – Probability of correct response to three questions

If the position of the respondent β_i is equal to λ_q, then the probability of giving the correct response is .5. If β_i is greater (less) than λ_q, then the probability of a correct response is greater (less than) 0.5. Thus question 1, with its low value of λ, is easy to answer correctly (a very low level of ability is required to have a high probability answering incorrectly), while question 3 has a high value of λ and is very difficult to answer correctly. How quickly the probability of a correct response increases or decreases as the respondent's ability changes depends on the discrimination parameter α_q. Question 1 has a low value of α, so if the respondent has ability much greater than λ_1, the probability of a correct response rapidly approaches 1. However question 3 has a high α, so as the respondent's ability moves lower than λ_3, the probability of a correct response only falls slowly.

If we had individual level data and the same questions were asked to every individual, this model could be adapted to our problem in a straightforward way.[3] The probability of answering in a left-wing manner would simply replace the probability of answering correctly, and "ability" would be reinterpreted as "policy position", with more left-wing positions being scored higher. The difficulty and discrimination parameters of the questions would work exactly as before. The only change required would be that the same respondent would be allowed to have a different policy position for each year they participated.

We, of course, do not have sufficient individual panel data with which to estimate this model, but instead have to rely on aggregate data. Therefore we cannot estimate the position of each individual, but only the average position of the population in any given year. This, however, is exactly what we are interested in – the position of the median voter. Indeed, if we assume that the population is normally distributed, not only can we

[3] It is not actually necessary for all questions to be asked of all respondents. There simply has to be enough overlap in the questions asked to each respondent.

estimate the population mean (which, of course equals the median given normality) but also the standard deviation, which gives us a measure of how polarized public opinion is.

Let us lay out the model formally. Let us denote the year by y with the range startyear…endyear. Let us denote the number of the question or item being asked as q, which ranges from 1…Q. (The same question will be asked in at least two different years.) We have data on the proportion of respondents who gave a "left-wing" answer to each question that was asked. This we denote as $leftr_{y\ q}$, the proportion that gave a left wing answer to question q in year y. For each year, we estimate the mean policy position of population, which we will call μ_y. This is simply a real number with the convention that low scores stand for more right-wing positions and high scores for more left-wing positions. We will also estimate the standard deviation of the positions of the respondents in each year, σ_y. In addition we need to estimate parameters that characterize each question. As before, the parameter λ_q stands for the "position" of question q, with high scores for question that are hard to answer in a left-wing manner, while parameter α_q represents how effectively the question discriminates between left and right-wing respondents.

Although we only have aggregate data, we need to start with a model of individual choice on responses. Instead of using the logistic specification in Equation (4), I use a cumulative normal distribution function with mean λ_q and standard deviation α_q, implying a random utility model. If $e_{i\ y\ q}$ is the probability of respondent i in year y giving a left wing response to question q, and $\beta_{i\,y}$ is the policy position of respondent i in year y, and Φ is the cumulative normal distribution function, [4]then:

$$e_{i\,y\,q} = \Phi(\beta_{i\,y}, \lambda_q, \alpha_q) \qquad (5)$$

This cumulative normal specifications produce very similar response functions to the logistic ones graphed in Figure 1, and the parameters can be interpreted in the same way. For example, if λ equals β, the probability of a left-wing answer is .5 in both cases, and higher values of α means less discrimination. However, I use the cumulative normal because it is mathematically convenient when we need to integrate over a normally distributed population.

We assume that the respondents in year y are normally distributed with mean μ_y and standard deviation σ_y. Essentially we are assuming that all respondents in a given year are drawn from the same distribution of policy positions, and that each question behaves on average in the same way, whatever year it is asked in. The response function $e_{y\,q}$ (Equation 5) gives us the probability of a respondent with position β giving the left-wing response; we assume that the probability of a respondent having policy position β in year y is given by the normal probability density function with mean μ_y and standard deviation σ_y. Given this, the probability of a randomly selected agent giving the left response to question q in year y (which we will call $m_{y\,q}$) is the product of $e_{y\,q}$ and the normal density function integrated over all possible values of β. Thus if ϕ is the normal probability density function:

$$m_{y\,q} = \int_{-\infty}^{\infty} \Phi(\beta, \lambda_q, \alpha_q)\ \phi(\beta, \mu_y, \sigma_y)\, d\beta \qquad (6)$$

Integrating, we get:

$$m_{y\,q} = \Phi(\beta, \lambda_q, \sqrt{\alpha_q{}^2 + \sigma_y{}^2}) \qquad (7)$$

We may note the symmetry in the effect of the parameters α_q and σ_y. The standard deviation of the normal distribution function that gives the probability of a left response is the geometric mean of the two parameters. Intuitively we would expect the two parameters to have a similar effect. A question that does not discriminate well leads to a significant number of right wing responses, even given a left-wing population. A population with a high variance will also lead to a significant number of right-wing responses, even if the mean of the population is quite left-wing.

Given the function $m_{y\ q}$ for the expected probability of a random respondent giving a left-wing answer, we can model the total number of left-wing responses to a question using the beta-binomial distribution. The reason for using the beta-binomial is to allow some stochastic variation to account for the fact that the same question may be applied or understood in slightly different ways in different years. Thus the final probability of a left-wing response to question q in year y is distributed according to the beta distribution with expectation $m_{y\,q}$. Given that the expectation of the beta distribution with parameters α and β is $\alpha/(\alpha+\beta)$, we can reparameterize the beta distribution in terms of its expectation. Thus the probability of a random respondent giving a left-wing response, which we call $p_{y\,q}$, is distributed thus:

$$p_{y\,q} \sim beta(\frac{m_{y\,q}}{1-m_{y\,q}}\beta, \beta) \qquad (8)$$

The number of left-wing responses expected for question q in year y, where $n_{y\,q}$ is the number of responses to that question, is then distributed binomially:

$$leftr_{y\,q} \sim binomial(p_{y\,q}, n_{y\,q}) \qquad (9)$$

Assigning non-informative uniform priors to the parameter vectors β, σ, λ and α, as well as to the beta parameter β, we can estimate the model using Bayesian inference software such as BUGS or JAGS. Figure 2 gives a graphical summary of the model. Appendix 1 gives complete code for JAGS and BUGS.

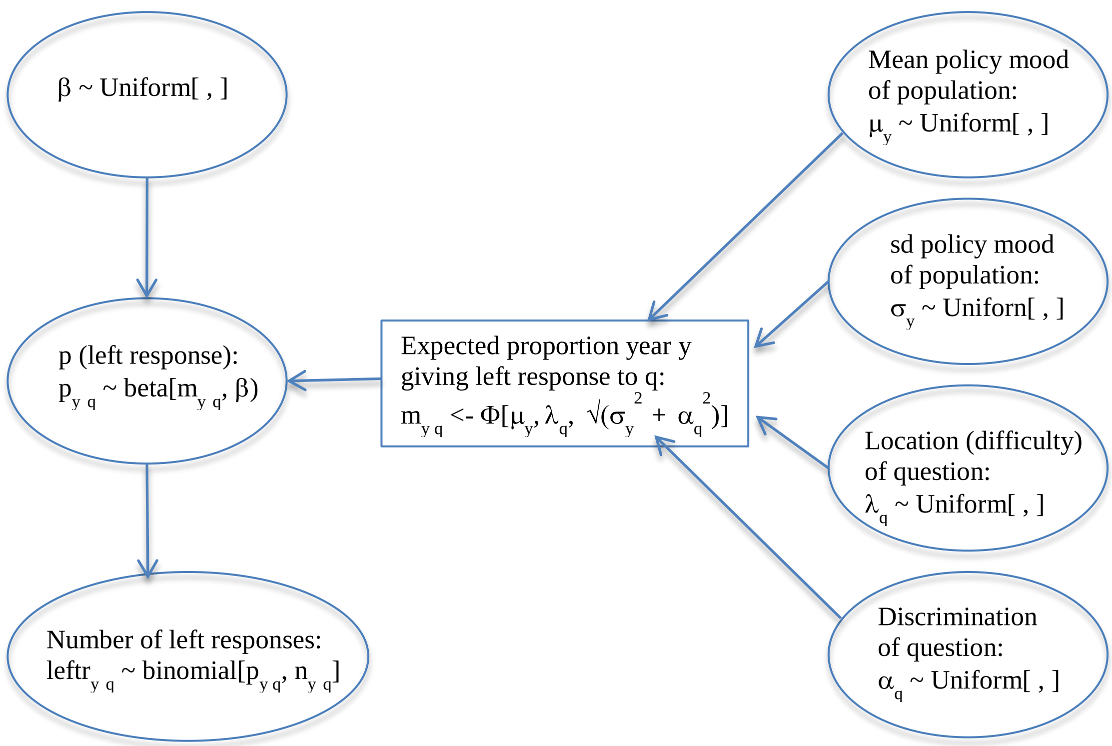

Figure 2 – Graph of the model

We thus have a model to estimate the policy mood parameters of a population based on aggregate data, which nevertheless is derived from a plausible and widely used model of individual choice. We can compare this model with the Stimson dyad-ratio approach. One notable difference is that the IRT model estimates the dispersion of policy mood as well as its central tendency. This is significant for a number of reasons. Firstly, the degree to which public opinion is dispersed or polarized is interesting in itself. Politics is not simply about the location of the median voter. It may matter whether difference in public opinion are relatively minor or whether the public is deeply divided. If there is a move in public opinion (say) to the right, we may be interested in whether just a few people around the median have shifted their position, or whether the entire distribution has shifted right.

The dispersion of policy mood also affects its estimation. In the last section I gave reason to believe that the dyad ratio algorithm will misinterpret changes in dispersion as changes in the political center. Taking dispersion into account also allows use the available information more efficiently. If the standard deviation of the population is small and the mean centrist, then questions that are either very easy or very difficult to answer in a left-wing manner give very little information – virtually everyone will answer left on the easy question and virtually no-one will answer left on the hard one. However, if the population is widely dispersed, these questions will give a great deal of information, while moderately easy questions will be difficult to interpret, as questions over a broad range in the center of the distribution yield left response rates of close to 50%. Thus taking the dispersion of the population into account allows for a better estimate of its central tendency.

Another minor difference is that the IRT approach does not use any form of smoothing. The dyad ratio approach incorporates smoothing of the results iteratively into the estimation process. The results of the IRT model may be smoothed after the fact if this is desired, but there is no smoothing built into the estimation process.

4. Implementation and results

The IRT algorithm was applied to data from Bartle, Dellepiane-Avellaneda and Stimson (2011). The authors collected a database of the aggregate results of survey questions about political matters in the UK between 1947 and 2005. The results were recoded to give the proportion of respondents giving the "left-wing" answer to each question. Following the authors, I have only used the question dealing with domestic policy issues. There were 364 different question items (two of which were dropped for reasons explained below). Each of these was asked in at least two different years, resulting in a total of 2377 question administrations. Each line of the input dataset corresponds to one question administration, with variables for the proportion of left responses, the year the question was asked, the number of the question item, and the number of respondents.

The model was estimated using JAGS. I used 3 chains, each with 30,000 iterations, and a 15,000 iteration burn-in. Convergence was checked for by inspecting the traceplots and by calculated Geweke and Raftery / Lewis diagnostics. These diagnostics are provided in the web/reviewer appendix. The full CODA file is available on request. The model estimates policy position and standard deviation parameters for each of the 59 years from 1947 to 2005, position and dispersion parameters for each of the 362 questions and one coefficient for the beta distribution, which captures the degree of stochastic variation between question administrations. Given we are more interested in the policy positions of the population than the characteristics of the individual questions, the parameters relating to the question items (λ_q and α_q) are given in the web/reviewer appendix, while the parameters for the policy position of the population (β_y and σ_y) are given in the main appendix and graphed in Figures 3 and 4.

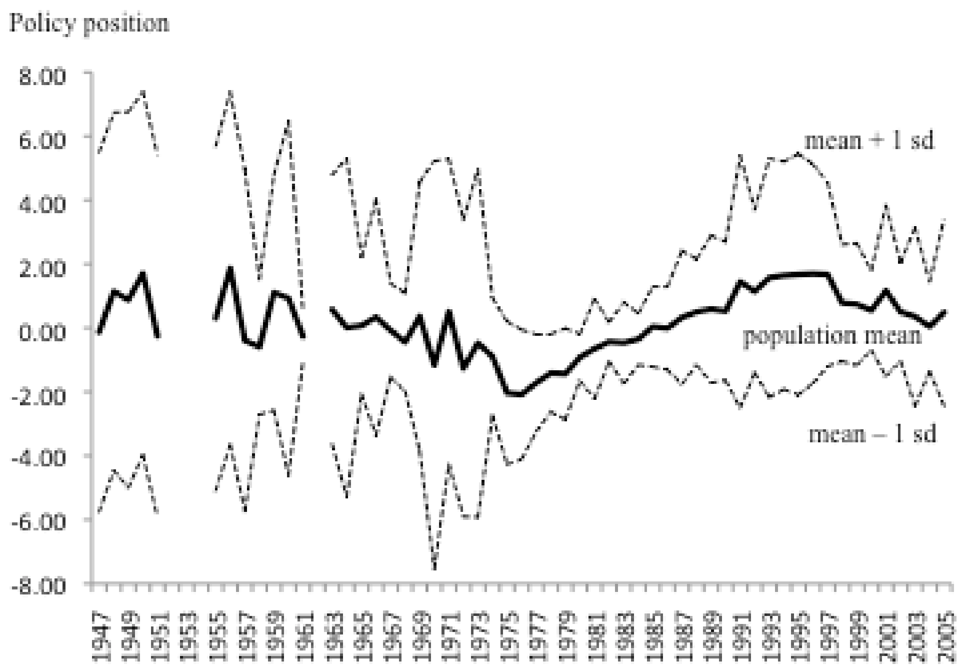

Figure 3 – Estimated policy position of mean respondent, and at mean ± one standard deviation.

Figure 3 graphs the mean policy position of the population for each year (β_y) and the mean policy position plus and minus one standard deviation ($\beta_y \pm \sigma_y$). There was no data for 1954 and only one question for 1952 and 1962, so estimates for these years are omitted. The figure – an in particular the pattern of change in policy mood over time – is quite similar to Figure 1 in Bartle, Dellepiane-Avellaneda and Stimson (2011). However, there are important differences. The grey lines are *not* confidence limits or a measure of the accuracy of our estimate of the population mean (this information is given in the appendix). Rather they are estimates of the position of a respondent one standard deviation from the mean. The IRT model gives an estimate of the dispersion of the population as well as its central tendency. Thus when the grey lines diverge from the black line representing the mean, this means that the population is polarized, with many respondents a considerable distance from the mean. It is notable that the degree of polarization varies considerably over the period studied. An additional difference is that the Stimson algorithm involves smoothing as part of the algorithm, whereas the IRT results are not smoothed.

Both the IRT approach and the Stimson dyad-ratio algorithm assume that questions work the same way when asked in different years. In terms of IRT theory, the response function has to remain constant in terms of the position and discrimination of the question. Violation of this assumption is known as differential item functioning (see Nunnally and Bernstein (1994, 416). One indicator of differential item functioning in our case is that we find very different estimate of policy mood in years in which a particular question is asked compared to years where that question is not asked. For example during the 1960s two question asked by Gallup (question 111, EDUC and question 267, NHS "Do you think the government is spending too much, too little, or about the right amount on: Education / Health?") produced this result. The questions were asked every two years. Years in which the questions were asked produced extremely right-wing estimates of policy mood, which were reversed in the following year when the questions were not asked. The explanation for this is that from the 1970s on, there is near consensus amongst all respondents that not enough is being spent on education and health (and "too little" was classified as the left-wing response). During the 1960s there was actually disagreement on this issue. Thus any year in which the question is asked during the 1960s appears extremely right-wing compared to both the (actually very right-wing) mid-seventies and the previous year when the question were not asked. It is likely that the changes in patterns of response to the two questions between the sixties and the seventies was due to changes in expectation about the performance of the health system rather than a systematic move to the left in policy mood (which would be contrary to all the other evidence). For this reason, I excluded the two questions.

The scale in Figure 3 has no intrinsic meaning. For purposes of interpretation we can calculate the percentage of the population that we would predict to give a left-wing answer to a typical question using equation 4 (the average question had a value of .357 for λ and 6.16 for α). We can also calculate the predicted percentage of left responses for the "left of center" half of the electorate and the "right of center" half of the electorate thus:

$$\text{Left response \% for left half of electorate} = 100 * \int_{\mu_y}^{\infty} \Phi(\beta, \lambda_q, \alpha_q)\, \phi(\beta, \mu_y, \sigma_y)\, d\beta \quad (10)$$

$$\text{Left response \% for right half of electorate} = 100 * \int_{-\infty}^{\mu_y} \Phi(\beta, \lambda_q, \alpha_q)\, \phi(\beta, \mu_y, \sigma_y)\, d\beta \quad (11)$$

These expressions can be evaluated numerically. This gives us an intuitive measure of the degree of polarization of the electorate. These results are graphed in Figure 4.

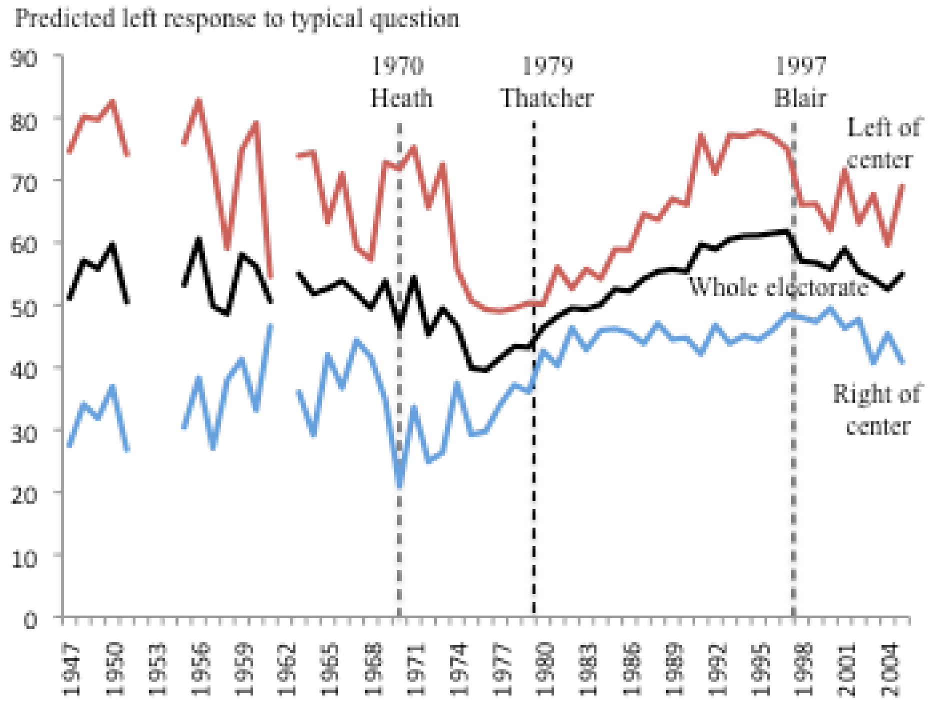

Figure 4 – Predicted percentage giving left response to a typical question for whole population and left-wing and right-wing halves of population

The movement of the "political center" in Figure 4 is very similar to that described by Bartle, Dellepiane-Avellaneda and Stimson (2011, Figure 1). The correlation coefficient between the Bartle, Dellepiane-Avellaneda and Stimson dyad ratio and the IRT estimates from 1950 to 2005 is 0.71. However, this coefficient is depressed by the period 1947-62 for which there is very little data. From 1963 to 2005 the correlation is .91, and for the years 1970-2005 it is .93. We can observe the same patterns in both series of estimates. In general public opinion moves against whoever is in office. The British electorate leans to the left in the early postwar period, but through the fifties and sixties it seems to gradually move towards the center (although data is sparse until the mid-1960s, so we have limited confidence in our measurements). There is instability in the early seventies, followed by a sharp move to the right following the re-election of a Labour government under Harold Wilson in 1974. This movement to the right precedes the Conservative government of Margaret Thatcher. In fact, the electorate moves steadily leftward under the Conservative governments of Thatcher (1979-90) and Major (1990-7). The victory of Tony Blair and the Labour Party in the 1997 general election leads to a gradual movement back towards the right, although this is temporarily halted in the election years of 2001 and 2005.

Where the item response theory estimates do differ from the Bartle, Dellepiane-Avellaneda and Stimson estimates is when there is a sharp change in polarization. We see this during the 1970s. According to both series of estimates there was a move to the right during this period. However, the dyad ratio estimates show a steady move to the right over the whole period. The item response theory estimates show no movement to the right during the Heath government (1970-4), but a sharp increase in polarization during this period. This is followed by a sharp movement to the right combined with a pronounced decrease in polarization during the Wilson administration (1974-6). Thus we find that the low point for left support was 1976, the year of the Sterling crisis when Britain required IMF loans. During the Callaghan government (1976) we actually find that the political center actually move back towards the left a little. As explained earlier, we would expect the dyad ratio approach to interpret changes in polarization as changes in the political center. Similarly, the item response theory estimates show less of a movement to the right after 2000 than the dyad ratio approach, finding instead increased polarization.

The fact that the IRT approach allows us to estimate the polarization of the population not only allows us to better estimate the position of the political center, but also provides additional interesting information. We find periods of high polarization in the early seventies and mid nineties. Measured polarization was high in the fifties and early sixties. While it may be that public opinion really was polarized in this period, it may also be an artifact of the small number of questions asked. Polarization seemed to decrease by the late sixties, but then increased suddenly in the early seventies. Given that the early seventies was a period of extreme industrial unrest, with strikes in many industries, including coal and electricity, and the imposition of a three day working week, it would be a cause for concern if we did not measure high polarization for this period. However, as public opinion moves sharply to the right from 1974, this polarization diminishes rapidly. As we can see from Figure 4, support for left-wing positions amongst the left half of the electorate fell dramatically once Labour was again in government.

It is interesting that polarization by our measure is very low during the 1980s. These are generally assumed to be years in which public opinion was extremely polarized, leading up to the coal miners' strike of 1984-5. The results here suggest that while politicians on both the left and right may have taken more polarized position in this period, they were not responding to contemporary public opinion. We might speculate that they seem to have been reacting to the public opinion of the previous decade. Bartle, Dellepiane-Avellaneda and Stimson (2011) note that governments tend to overshoot public demand for policy change. It appears that this also applies to the polarization of the party system.

The polarization of the 1990s seems to be driven largely by the left of center half of the population. If we look at Figure 4, we can see that by the mid-1980s the right half of the population had moderated considerably from the polarized positions of the 1970s, but did not change its position much after this. The left half of the population, however, became increasing likely to take left-wing positions up until 1997. After 1997, it is the left-wing half of the population that moves towards the center, producing the observed shift of the political center to the right. The observation that the left of center half of the electorate is far less stable than the right of center half is a phenomenon worth further investigation.

5. Conclusion

I outline a model for estimating the central tendency of public opinion using item response theory, and compare this to the dyad-ratio approach used by James Stimson and his collaborators. One advantage of the item response theory approach is that it explicitly models the mean of the population with a widely used psychometric model of individual behavior. The dyad ratio approach, on the other hand, relies on the assumption that the ratio between response rates to the same question asked in different years measures changes in the position of the central tendency. This assumption is only justified if other conditions are met, such as the dispersion of the population not changing. Thus the item response theory approach is theoretically better justified in its claim to be measuring the political center.

The estimates produced by the two methods, however, are extremely similar, at least in the case of the British public opinion data from Bartle, Dellepiane-Avellaneda and Stimson (2011). The correlation coefficient of the two sets of estimates for the years 1970-2005 is .93. In this respect, the item response theory approach validates the results obtained by the dyad ratio algorithm. I have suggested that the reason that the dyad ratio algorithm does track the political center is that it is in effect an average of estimates weighted by iteratively calculated communality estimates of item questions. Questions that are either very easy or difficult to give left-wing answers to will be assigned low communality scores, and thus the final estimates will be dominated by question of medium difficulty. These are precisely the questions that provide the most information about the position of the political center. Where the two sets of estimates do diverge is when there is a substantial change in the dispersion of the distribution of public opinion.

The item response theory approach allows us to estimate the dispersion of public opinion in addition to its central tendency. This is important as it allows us to produce better estimates of the political center. We would expect changes in the standard deviation of the distribution of policy positions to produce changes in the response rate to questions. If we do not take this into account, these changes could be misinterpreted as changes in the position of the political center. The dyad ratio approach appears to react to changes in dispersion in this way. Secondly, the dispersion or polarization of the distribution of preferences is interesting in its own right. We actually find some interesting and surprising results. The 1980s in Britain are commonly seen as an extremely polarized decade. However, the results presented here suggest that public opinion was not very polarized at all – perhaps it was only the political parties that were polarized, reacting with a lag to the very polarized 1970s.

Both the item response theory model outlined here and the dyad ratio algorithm allow us to produce time series estimates of the central tendency of the distribution of preference from aggregate survey data. This is possible even though the same questions are not generally asked every year. As suggested before, this allows us to address numerous questions in political behavior, political economy and comparative political institutions. The code for the item response theory model is provided below, and can be run in the free software WinBUGS and JAGS.

Appendix A: Code

```
##Constants to be set in data or in code: startyear, endyear,
##nquest (number of questions), len (number of administrations of questions)

##variables in data: leftp (proportion answering left), year,
##q (the number of the question asked), n (number of respondents)

##Data statement is for JAGS. Omit next 5 lines if using BUGS
data{
for (i in 1:len){
leftr[i]<-round(leftp[i]*n[i]/100)
}
}

model{

##loop over the data
for (i in 1:len){
## leftr[i]<-round(leftp[i]*n[i]/100)   **Uncomment this line if using BUGS**
p[i]~dbeta(a[i], b)
a[i]<--b*m[i]/(m[i]-1)
m[i]<-phi(x[i])
x[i]<-(mu[year[i]] - lambda[q[i]]) / sqrt((alpha[q[i]])^2+(sigma[year[i]])^2)
}

##priors for this model – make sure uniform priors are wide enough not to constrain
b~dunif(0, 100)
lambda[1]<-0
alpha[1]<-0.25
```

```
for (i in 2:nquest){
lambda[i]~dunif(-10,10)
alpha[i]~dunif(0,10)
}

for (i in startyear:(endyear)){
mu[i]~dunif(-10,10)
sigma[i]~dunif(0,10)
}
}
```

Appendix B: Results

	Mean	SD	Naive SE	Time-series SE
b	2.20E+01	8.39E-01	1.53E-02	2.03E-02
mu[1947]	-1.29E-01	2.77E+00	5.06E-02	5.94E-02
mu[1948]	1.14E+00	3.84E+00	7.01E-02	1.24E-01
mu[1949]	8.79E-01	3.38E+00	6.17E-02	1.19E-01
mu[1950]	1.71E+00	4.01E+00	7.32E-02	1.21E-01
mu[1951]	-2.53E-01	3.83E+00	6.99E-02	1.28E-01
mu[1952]	5.35E+00	2.03E+00	3.70E-02	4.00E-02
mu[1953]	-2.67E+00	2.36E+00	4.30E-02	4.61E-02
mu[1954]	-1.25E-01	5.79E+00	1.06E-01	1.10E-01
mu[1955]	3.05E-01	9.86E-01	1.80E-02	1.92E-02
mu[1956]	1.86E+00	1.98E+00	3.61E-02	3.79E-02
mu[1957]	-3.91E-01	7.18E-01	1.31E-02	1.52E-02
mu[1958]	-5.93E-01	7.98E-01	1.46E-02	1.63E-02
mu[1959]	1.10E+00	8.11E-01	1.48E-02	1.57E-02
mu[1960]	9.32E-01	1.58E+00	2.89E-02	3.11E-02
mu[1961]	-2.63E-01	4.73E-01	8.64E-03	1.61E-02
mu[1962]	-2.40E+00	2.61E+00	4.77E-02	5.30E-02
mu[1963]	5.75E-01	6.04E-01	1.10E-02	1.26E-02
mu[1964]	7.94E-03	4.54E-01	8.30E-03	9.69E-03
mu[1965]	7.16E-02	6.26E-01	1.14E-02	1.28E-02
mu[1966]	3.40E-01	3.56E-01	6.50E-03	7.60E-03
mu[1967]	-7.33E-02	4.69E-01	8.56E-03	1.08E-02
mu[1968]	-4.42E-01	4.41E-01	8.06E-03	1.12E-02
mu[1969]	3.63E-01	7.00E-01	1.28E-02	1.31E-02
mu[1970]	-1.16E+00	7.02E-01	1.28E-02	1.36E-02
mu[1971]	5.14E-01	9.90E-01	1.81E-02	2.12E-02
mu[1972]	-1.27E+00	5.50E-01	1.00E-02	1.12E-02
mu[1973]	-4.70E-01	4.82E-01	8.79E-03	9.78E-03
mu[1974]	-8.98E-01	2.17E-01	3.96E-03	6.88E-03
mu[1975]	-2.03E+00	2.66E-01	4.85E-03	9.48E-03
mu[1976]	-2.08E+00	2.39E-01	4.37E-03	9.65E-03
mu[1977]	-1.72E+00	2.40E-01	4.38E-03	8.54E-03
mu[1978]	-1.41E+00	2.12E-01	3.87E-03	7.83E-03
mu[1979]	-1.44E+00	1.97E-01	3.59E-03	7.17E-03
mu[1980]	-9.19E-01	1.88E-01	3.44E-03	7.87E-03
mu[1981]	-6.51E-01	1.83E-01	3.35E-03	5.96E-03
mu[1982]	-4.34E-01	1.73E-01	3.16E-03	5.48E-03
mu[1983]	-4.68E-01	1.37E-01	2.50E-03	4.65E-03
mu[1984]	-3.50E-01	1.56E-01	2.84E-03	4.85E-03

mu[1985]	4.11E-02	1.34E-01	2.44E-03	3.98E-03
mu[1986]	-1.14E-02	1.35E-01	2.47E-03	3.90E-03
mu[1987]	3.29E-01	1.52E-01	2.78E-03	4.74E-03
mu[1988]	5.02E-01	2.48E-01	4.53E-03	5.35E-03
mu[1989]	5.90E-01	1.69E-01	3.08E-03	4.87E-03
mu[1990]	5.25E-01	1.59E-01	2.90E-03	5.15E-03
mu[1991]	1.44E+00	2.30E-01	4.20E-03	6.39E-03
mu[1992]	1.15E+00	2.06E-01	3.76E-03	5.68E-03
mu[1993]	1.57E+00	2.13E-01	3.88E-03	5.37E-03
mu[1994]	1.64E+00	1.94E-01	3.54E-03	5.42E-03
mu[1995]	1.68E+00	2.15E-01	3.92E-03	5.56E-03
mu[1996]	1.70E+00	2.08E-01	3.80E-03	5.92E-03
mu[1997]	1.68E+00	2.06E-01	3.76E-03	5.77E-03
mu[1998]	7.82E-01	1.79E-01	3.27E-03	4.72E-03
mu[1999]	7.42E-01	2.15E-01	3.92E-03	5.80E-03
mu[2000]	5.51E-01	1.89E-01	3.46E-03	5.47E-03
mu[2001]	1.17E+00	2.29E-01	4.19E-03	6.80E-03
mu[2002]	5.04E-01	2.18E-01	3.97E-03	6.40E-03
mu[2003]	3.56E-01	2.51E-01	4.59E-03	6.16E-03
mu[2004]	4.27E-02	1.62E-01	2.96E-03	3.92E-03
mu[2005]	4.92E-01	2.05E-01	3.75E-03	5.44E-03
deviance	1.92E+04	6.81E+01	1.24E+00	1.28E+00
sigma[1947]	5.63E+00	2.85E+00	5.20E-02	5.45E-02
sigma[1948]	5.60E+00	2.86E+00	5.22E-02	5.14E-02
sigma[1949]	5.88E+00	2.81E+00	5.13E-02	5.25E-02
sigma[1950]	5.68E+00	2.82E+00	5.14E-02	4.36E-02
sigma[1951]	5.66E+00	2.84E+00	5.18E-02	5.26E-02
sigma[1952]	5.40E+00	2.87E+00	5.24E-02	5.07E-02
sigma[1953]	6.49E+00	2.58E+00	4.71E-02	4.65E-02
sigma[1954]	4.94E+00	2.90E+00	5.29E-02	5.80E-02
sigma[1955]	5.39E+00	2.17E+00	3.96E-02	4.46E-02
sigma[1956]	5.52E+00	2.88E+00	5.26E-02	5.05E-02
sigma[1957]	5.34E+00	2.10E+00	3.84E-02	3.93E-02
sigma[1958]	2.12E+00	1.96E+00	3.58E-02	3.71E-02
sigma[1959]	3.68E+00	2.41E+00	4.40E-02	4.51E-02
sigma[1960]	5.54E+00	2.86E+00	5.22E-02	4.84E-02
sigma[1961]	7.76E-01	5.35E-01	9.77E-03	1.36E-02
sigma[1962]	5.47E+00	2.88E+00	5.27E-02	5.29E-02
sigma[1963]	4.22E+00	1.99E+00	3.63E-02	3.63E-02
sigma[1964]	5.31E+00	1.30E+00	2.38E-02	2.59E-02
sigma[1965]	2.14E+00	9.04E-01	1.65E-02	2.16E-02
sigma[1966]	3.71E+00	1.30E+00	2.38E-02	3.14E-02
sigma[1967]	1.47E+00	7.04E-01	1.29E-02	2.07E-02
sigma[1968]	1.54E+00	7.44E-01	1.36E-02	2.25E-02
sigma[1969]	4.22E+00	1.80E+00	3.28E-02	3.35E-02
sigma[1970]	6.38E+00	1.76E+00	3.22E-02	3.41E-02
sigma[1971]	4.77E+00	2.62E+00	4.78E-02	4.94E-02
sigma[1972]	4.64E+00	1.75E+00	3.19E-02	3.30E-02
sigma[1973]	5.45E+00	1.71E+00	3.12E-02	3.29E-02
sigma[1974]	1.83E+00	5.82E-01	1.06E-02	1.57E-02
sigma[1975]	2.24E+00	4.79E-01	8.75E-03	9.89E-03
sigma[1976]	2.03E+00	3.81E-01	6.95E-03	8.22E-03
sigma[1977]	1.52E+00	4.30E-01	7.85E-03	1.06E-02
sigma[1978]	1.22E+00	4.59E-01	8.38E-03	1.27E-02
sigma[1979]	1.42E+00	4.80E-01	8.77E-03	1.28E-02
sigma[1980]	7.33E-01	4.12E-01	7.52E-03	1.42E-02
sigma[1981]	1.56E+00	5.11E-01	9.33E-03	1.56E-02

sigma[1982]	6.13E-01	4.08E-01	7.45E-03	1.96E-02
sigma[1983]	1.27E+00	3.64E-01	6.65E-03	1.68E-02
sigma[1984]	8.08E-01	4.49E-01	8.19E-03	1.73E-02
sigma[1985]	1.25E+00	3.25E-01	5.93E-03	1.64E-02
sigma[1986]	1.30E+00	3.34E-01	6.10E-03	1.66E-02
sigma[1987]	2.09E+00	3.69E-01	6.73E-03	1.70E-02
sigma[1988]	1.66E+00	3.45E-01	6.30E-03	1.28E-02
sigma[1989]	2.30E+00	4.21E-01	7.69E-03	1.71E-02
sigma[1990]	2.17E+00	2.68E-01	4.90E-03	1.27E-02
sigma[1991]	3.95E+00	3.97E-01	7.25E-03	1.37E-02
sigma[1992]	2.55E+00	3.29E-01	6.01E-03	1.26E-02
sigma[1993]	3.74E+00	3.21E-01	5.86E-03	1.19E-02
sigma[1994]	3.56E+00	3.28E-01	5.99E-03	1.07E-02
sigma[1995]	3.77E+00	3.26E-01	5.96E-03	1.09E-02
sigma[1996]	3.43E+00	3.02E-01	5.51E-03	1.19E-02
sigma[1997]	2.88E+00	3.07E-01	5.61E-03	1.14E-02
sigma[1998]	1.83E+00	3.29E-01	6.00E-03	1.30E-02
sigma[1999]	1.91E+00	4.18E-01	7.63E-03	1.39E-02
sigma[2000]	1.24E+00	4.03E-01	7.36E-03	1.80E-02
sigma[2001]	2.68E+00	5.11E-01	9.34E-03	1.80E-02
sigma[2002]	1.54E+00	3.88E-01	7.07E-03	1.90E-02
sigma[2003]	2.80E+00	7.00E-01	1.28E-02	1.82E-02
sigma[2004]	1.40E+00	4.72E-01	8.61E-03	1.99E-02
sigma[2005]	2.94E+00	5.78E-01	1.06E-02	1.89E-02

Table A1 – Parameter Estimates of IRT model (see text)

Bibliography

Bartle, John, Sebastian Dellepiane-Avellaneda, and James Stimson. 2011. The Moving Centre: Preferences for Government Activity in Britain, 1950–2005. *British Journal of Political Science* 41 (2):259-285.

Bartle, John, Sebastian Dellepiane-Avellaneda, and James A. Stimson. 2011. The Policy Mood and the Moving Centre. In *Britain at the polls 2010*, edited by N. Allen and J. Bartle. London: SAGE.

Baumgartner, Frank R., Suzanna L. De Boef, and Amber E. Boydstun. 2008. *The Decline of the Death Penalty and the Discovery of Innocence*. Cambridge: Cambridge University Press.

Chanley, Virginia A., Thomas J. Rudolph, and Wendy M. Rahn. 2000. The Origins and Consequences of Public Trust in Government: A Time Series Analysis. *Public Opinion Quarterly* 64:239-256.

Cohen, Jeffrey E. 2000. The Polls: Public Favourability towards the First Lady. *Presidential Studies Quarterly* 30:575-585.

Erikson, Robert S., Michael MacKuen, and James A. Stimson. 2002. *The macro polity, Cambridge studies in political psychology and public opinion*. New York: Cambridge University Press.

Kellstadt, Paul. 2003. *The Mass Media and the Dynamics of American Racial Attitudes*. Cambridge: Cambridge University Press.

Kim, Hee Min, and Richard Fording. 2001. Extending Party Estimates to Voters and Governments. In *Mapping Policy Preferences: Estiamtes for Parties, Electors, and Governemnts 1945-1998*, edited by I. Budge, H.-D. Klingemann, A. Volkens, J. Bara and E. Tanenbaum. Oxford: Oxford University Press.

McDonald, Michael, and Ian Budge. 2005. *Elections, Parties, Democracy: Conferring the Median Mandate*. Oxford: Oxford University Press.

Nunnally, Jum C., and Ira H. Bernstein. 1994. *Psychometric Theory*. 3 ed. New York.: McGraw-Hill.

Powell, G. Bingham Jr. 2000. *Elections as Instruments of Democracy: Majoritarian and Proportional Visions*. New Haven: Yale University Press.

Stimson, James A. 1991. *Public opinion in America : moods, cycles, and swings*, *Transforming American politics*. Boulder: Westview Press.

Repeated Author. 1999. *Public opinion in America : moods, cycles, and swings*. 2nd ed, *Transforming American politics*. Boulder: Westview Press.

Stimson, James A., Michael B. Mackuen, and Robert S. Erikson. 1995. Dynamic Representation. *American Political Science Review* 89 (3):543-565.

Stimson, James, Cyrille Thiébaut, and Vincent Tiberj. 2009. The Structure of Policy Attitudes in France. In *Annual Meetings of the Midwest Political Science Association*. Chicago.

Voeten, Erik, and Paul R. Brewer. 2006. Public Opinion, the War in Iraq, and Presidential Accountability. *Journal of Conflict Resolution* 50:809-830.