A new approach for estimating parties' positions in voting advice applications*

Kostas Gemenis

Abstract

The primary goal of voting advice applications (VAAs) is to calculate the match between voters' and parties' (or candidates) policy preferences. In order to do so, it is necessary for VAAs to estimate the positions of political parties. In many respects this is a daunting task given that all the commonly used methods for doing so have considerable drawbacks. Most VAAs use a combination of two methods, namely questionnaires sent to parties and expert coding based on party manifestos. The paper will outline the existing approaches in terms of validity and reliability and will propose some improvements in the form of a formal approach to expert coding over multiple rounds with anonymous feedback. The paper will present examples of the advantages of this method using evidence from Choose4Greece, a VAA launched for the May and June 2012 parliamentary elections in Greece.

1 Introduction

The intersection of politics and the internet has prompted the development of many interesting research agendas over the past fifteen or so years. One of the most promising ones is concerned with the advent of voting advice (or voting aid) applications (VAAs). VAAs usually come in the form of internet websites which allow prospective voters to get informed about the policy preferences of candidates and political parties, and provide them with information regarding the congruence between their own preferences and those of candidates/parties. As such, VAAs allow political scientists to engage with the public and influence electoral participation (Fivaz & Nadig 2010, Ladner & Pianzola 2010), to generate data which can be used for testing questions regarding voting behaviour, pledge fulfilment, policy congruence and the dimensionality of political space (Schwarz, Schädel & Ladner 2011, Katsanidou & Lefkofridi 2010, Talonen & Sulkava 2011, Wheatley et al. 2012), but also provide some unique opportunities for methodological explorations.

The core methodological challenge in VAAs is to calculate and present the congruence or agreement between prospective voters and candidates or parties. So far researchers and designers of VAAs have used different visualization techniques such as radar (aka spider) plots and two-dimensional political maps, compared different methods of calculating congruence scores (Louwerse & Rosema 2013, Mendez 2012), and investigated the effect of asking different sets of policy questions (Walgrave, Nuytemans & Pepermans 2009). Although the issues surrounding the positioning political actors have attracted considerable attention in the electoral studies literature[1] and generated several debates and controversies[2] for some time now, it was only very recently that researchers put the methods for generating party positions in VAAs under methodological scrutiny.

Wagner & Ruusuvirta (2012) applied multidimensional scaling on the data of 13 different VAAs and extracted two dimensions. The scores of the first dimension were considered to represent party placements on a general left-right dimension. The scores were also standardized to help with cross-national comparisons and sometimes reversed in order to fit well-established knowledge with regards to parties that are considered to be left-wing. Wagner & Ruusuvirta (2012) correlated these scores with left-right party position estimates obtained from expert surveys and the Comparative Manifestos Project and found that, with a few exceptions, the VAA scores correlated quite well with the established measures. Gemenis (2013a) reached similar conclusions when he compared the EU Profiler left-right and EU integration scale estimates to expert survey data. Nevertheless, the scales examined by Gemenis (2013a) and Wagner & Ruusuvirta (2012) are not the ones which

VAAs routinely use to match voters to parties. If VAAs were to calculate ideological congruence on the basis of distances on such aggregate scales, this would entail a bold theoretical assumption: that voters' (and parties') position on all issues that consist a scale are interchangeable. Since this assumption is problematic in electoral contexts,[3] VAA designers have devised various matching algorithms which are additive functions of the voter/party agreement on each of the issues.[4] *Scaled* VAA party positions may have passed several validity tests (Gemenis 2013a, Wagner & Ruusuvirta 2012), but if voter/party matching is based on agreement on individual issues, to what extent parties' positions on these issues as given by VAAs are valid representations of where political parties actually stand? Unfortunately, it is impossible to give a definite answer to this question in the absence of a valid benchmark to which estimated party positions on individual issues can be compared to. In is not surprising, therefore, that evaluations of party positions on individual issues have focussed on reliability rather than validity (Gemenis 2013a). Nevertheless, reliability has been traditionally viewed as a necessary, but not sufficient, condition for validity in measurement. In this sense, the minimum criterion for a good approach to estimate parties' positions should be the degree to which it provides reliable estimates. In addition, the paper uses a second criterion to evaluate methods for estimation, namely the degree to which they can be employed in the context of a VAA. In other words, any proposed approach to estimation should not be overtly costly, time-consuming or otherwise impractical.

The paper is therefore organized as follows. Section 2 outlines the advantages and disadvantages of using the established methods for estimating parties' positions in the context of VAAs: elite and expert surveys, the content analysis of party manifestos and the analysis of roll-call voting. Recognizing that each of these approaches has several disadvantages which make their application to VAAs problematic, Section 3 outlines an alternative approach which has been devised with VAAs in mind. Although the latter has been used by several VAAs including the EU Profiler, the paper identifies several issues with this approach which make it less than ideal. Section 4 proposes a new approach for estimating parties positions which builds on some of the established methods. Using data from a VAA that has been deployed for the May and June 2012 elections in Greece, the paper illustrates that the proposed method satisfies both criteria set in this paper. While it is practical and not overtly costly or time-consuming, it has the ability to produce reliable and potentially highly valid estimates of parties' positions. The concluding section ends with some recommendations for designers of VAAs as well as with some suggestions with regards to how this approach can be extended to estimating parties' policy positions beyond the context of VAAs.

2 Established methods for estimating parties' positions

As VAAs are designed primarily by political scientists, it makes sense to start the discussion by looking at the methods which have been proposed for estimating parties' positions outside the VAA context. The literature on such methods is extensive and not everything can be convered in detail in this section. For instance, we consider methods based on media sources (Helbling & Tresch 2011) or large scale voter surveys to be prohibitively costly and time-consuming to be applied in the context of VAAs so it becomes meaningless to examine the reliability and validity. The section therefore focuses on three distinct approaches to which various methods can be grouped to: elite surveys, expert surveys, the content analysis of party manifestos and the analysis of roll-call voting. As noted in the introduction, the focus of the discussion is on two issues, the degree to which their application is practical in the context of VAAs and the degree to which they can provide reliable (and potentially valid) estimates of party positions.

2.1 Questionnaires sent to political parties

The starting point is the most obvious approach. Why use complicated or 'fancy' approaches when you can do the obvious, that is, send a questionnaire and ask the parties to position themselves on the statements used by VAAs. *Stemwijzer*, the pioneering VAA in the Netherlands, uses party position estimates exactly from this simple approach: questionnaires are sent to parties and their responses are used to calculate the match between parties and voters. Of course this is not a method pioneered by *Stemwijzer* of VAAs in general, as questionnaires in the form of elite surveys have been targetting party politicians for many decades now. There is nevertheless a practical consideration that makes this approach rather impractical in countries other than the Netherlands. The problem is that political parties are generally sceptical or even hostile to questionnaires which aim to measure political attitudes. Parties are known to prohibit their MPs and activists from answering such questionnaires (Baker et al. 1999, 172) or even threatening with legal action (Trechsel & Mair 2011, 14–15) either because they do not agree with the way the questions are formulated (Nezi, Sotiropoulos & Toka 2010) or because they do not want to openly acknowledge that they hold positions which are generally considered to be unpopular. When the EU Profiler asked parties across the Europe to place themselves on its 30 statements, the response rate across countries was only 37.6%. Trechsel &

Mair (2011, 13) call this rate 'remarkable', and in some sense it is given the previous experience with elite surveys in the context of the European Parliament which have been plagued by non-response (Scully & Farrell 2003), but still is impractical if one considers of using this approach as the sole way to position political parties.

Apart from this practical issue, responses given by parties can be problematic in terms of their reliability and, consequently, validity. Even when parties agree to provide answers to the given questions, the researchers have no guarantees that parties will answer truthfully. In fact, there is considerable evidence that parties sometimes try to manipulate their placement in VAAs in order to get an advantage in terms of voter matches. Wagner & Ruusuvirta (2012, 406) report the case of a VAA in Finland where 'some candidates placed themselves in the middle of the response scale on all statements' which is generally advantageous in terms of the algorithms used for matching. This manipulation became evident when the media criticised them for doing so and the candidates in question were forced to change their responses. Ramonaitė (2010, 133–137) reports a similar case of a Lithuanian VAA where a particular party responded to the questions in a specific way which would give the party a popular with voters but otherwise ideologically inconsistent placement. If researchers have no mechanism to check whether the questionnaire has been answered truthfully and not strategically then we are facing the possibility of unintentionally assisting electoral manipulation. This is particularly true when VAA designers do not ask parties to 'justify' their responses by linking them to published sources or public statements (as it was the case of *StemWijzer* in its first few years). Even when researchers ask for such justifications, there is no guarrantee that the process will work smoothly. citeasnounKrouwel2012 report that in the 2006 *Stemwijzer* contained incorrect information regarding the positions for a particular party as the latter gave the answers that were considered to be the most popular among voters. When the *Stemwijzer* team asked the party to justify the given answers, the party simply sent statements that attacked their main opponent.

In general, there is compelling evidence that much of the data generated from questionnaires sent to parties is unreliable. When the EU Profiler team cross-checked the given responses (Trechsel & Mair 2011, 16–17), they found out that in about 17% of party positions on individual questions there was a discrepancy with what the parties have given and what the researchers had independently found using other sources (including a 5% of substantial discrepancy). For some VAA designers, this cross-checking is a necessary step in party self-positioning. Nevertheless, as soon as researchers begin engaging with the parties and challenging the given positions on the basis of

party manifestos or other statements, then we are already using a different approach in positioning parties which is distinct to simply ask parties for answers. This approach is examined in Section 3 of this paper.

2.2 Content analysis of party manifestos

One of the most popular approaches in estimating parties' policy positions during elections is to content-analyze their election manifestos. The most popular approaches are the hand-coding of units of texts known as 'quasi-sentences' into a pre-defined coding scheme as practiced by the Comparative Manifestos Project (CMP, recently renamed to MARPOR), the Euro-manifestos Project, the Comparative Agendas Project and others, and word counts by computer programmes scaled according to some algorithm or reference text (Laver, Benoit & Garry 2003, Slapin & Proksch 2008). There are at least two problems with this approach which make it prohibitive in the context of VAAs. First, it is well-known that many parties across Europe and elsewhere do not follow the practice of publishing election manifestos, or when they do these might not have the same structure and functions of the election manifestos in countries such as the Netherlands or the UK. Whenever election manifestos cannot be found, researchers often resort to wide variety of proxy documents, such as pamphlets or party leader speeches. Since these proxies often have been produced with different purposes or audiences in mind, the comparisons to party manifestos might produce tenable or even implausible results and give the wrong picture regarding where parties really stand (Gemenis 2012). Secondly, counting words or quasi-sentences may easily reveal the degree to which parties consider the issues associated with these words to be important, but to transform these frequencies into positions additional assumptions are needed (Laver 2001a, Lowe 2008). In particular, researchers need to define which words or coding categories are left or right, progressive or conservative and so on in a scaling model. Some of the most heated discussions have evolved around the adoption of the 'right' scaling model[5] but the debate remains unresolved. Comparisons of different scaling methods, however, reveal these tend to produce radically different results regarding parties positions even when using exactly the same content analysis data (Dinas & Gemenis 2010). If we, as researchers, have not yet agreed with what is appropriate, then we cannot 'sell' the position estimates to the citizens who are uninformed about our methodological assumptions and the surrounding controversies.

2.3 The analysis of roll-call voting

The analysis of roll-call voting as means of estimating the positions of legislators has been very popular in the US context but also in the European Parliament. In the latter case, the scores of the individual legislators have also been aggregated and used as estimates of their respective parties' positions (Hix, Noury & Gérard 2006). Although not widespread, roll-call voting data has also been used for positioning actors in VAAs (Škop 2010). This approach has two drawbacks, however. One practical and one substantive. The practical problem is that one cannot position parties which are contesting the election for the first time or parties which are not represented in the legislature through the analysis of roll-call voting. The substantive problem is that for most of the European legislatures which are characterized by voting patterns where all the opposition parties vote together against the government irrespective of their policy positions. This means that the analysis of roll-call voting in countries such as the UK or Ireland will not reveal parties' policy positions but rather a dimension separating the government from the opposition parties (Hansen 2009, Spirling & McLean 2007).

2.4 Expert surveys

Expert survey estimates have been the main alternative to content analysis with respect to estimating parties' policy positions. One of their main advantages is that they are generally less costly and time-consuming than content analysis, especially of the hand-coding type (Volkens 2007). Moreover, since they consist of mean responses among groups of political science experts, they are less likely to produce results which are implausible, or simply wrong (Marks, Hooghe, Steenbergen & Bakker 2007, Steenbergen & Marks 2007). There was at least one instance which off-the-shelf estimates have been used in a VAA (Wall, Sudulich, Costello & Leon 2009), but there is also the potential of asking experts to estimate parties' positions on the specific statements used in VAAs. This potential use, however, might not work very well in practice. If the typical VAA features 30 issue statements and the typical European party system six important parties, then we will be asking for experts to make 180 different estimates. How valid and reliable would these estimates be? To quote an unnamed researcher who conducted some of the largest expert surveys, let us say that experts have often 'been asked to provide judgements on topics that require information that exists in the ambition of the surveyor, but not in the minds of experts'.

To put this claim to test, I examine the data at the individual expert level of a recent expert survey conducted in Greece (Gemenis & Nezi 2012). Since

it is difficult to examine the validity of estimates in the absence of a commonly accepted valid benchmark, I focus on the reliability of the estimates on the assumption that low reliability increases the likelihood of low validity (Krippendorff 2004, 214). As an indicator of reliability I use the agreement among experts which can be used by van der Eijk's (2001) coefficient of agreement A. Van der Eijk's A has desirable properties for measuring reliability in the content of party position judgements since it includes a measure of unimodality. The coefficient which ranges from -1 to 1, becomes smaller not only when experts disagree with each other, but even less so when this disagreement is clustered around multiple poles of the rating scale.
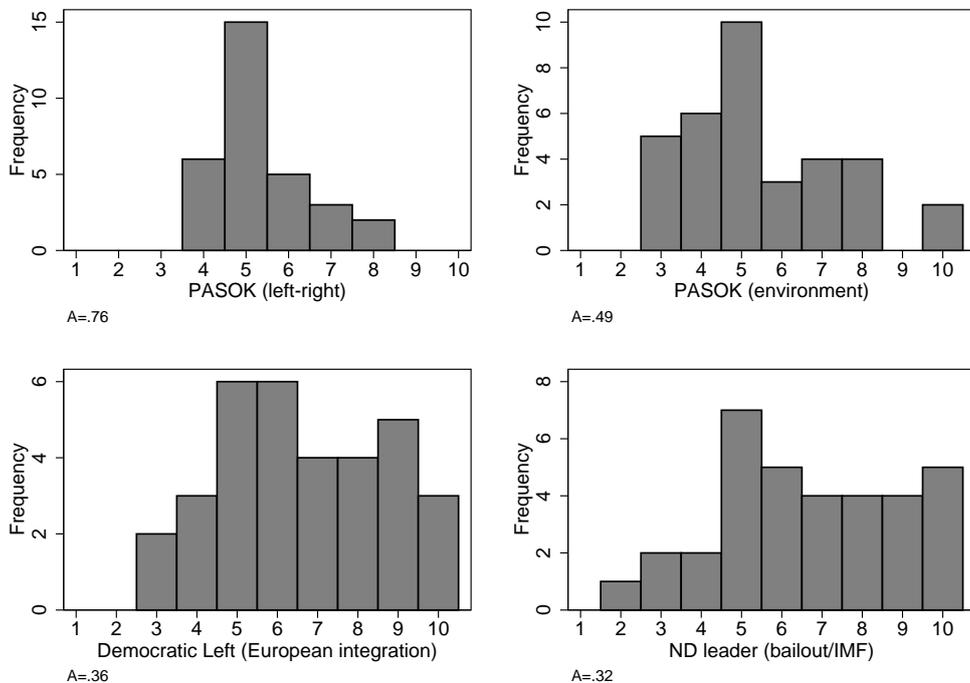

Figure 1: Distribution of judgments in an expert survey.

The following figures show that although the experts generally agree on the left-right placement of parties such as PASOK (A=.76), they have more difficulties placing the same parties on more specific policies (e.g. environmental protection for PASOK A=.49, European integration for Democratic Left, A=.36, position on the bailout for New Democracy, A=.32). This expert disagreement probably appears because different experts evaluate different actors within the party (the leader, the activists, the MPs), because

8

they take different statements of these actors into account, and because they probably also weight them differently in their final estimate (Budge 2000).

To be sure, one could challenge the presented findings on the basis of several arguments. It could be that experts in Greece are not well-informed, or someone could argue that the statements used in VAAs are specific enough for experts to be able to locate parties consistently. The same problem (of unreliability via disagreement) has been observed in the major cross-national expert surveys of party positions. Researchers have generally found that considerable disagreement emerges when experts are asked to estimate the positions of smaller or newer parties and the positions on less-known or more specific issues (Bakker et al. 2012). These results imply that if parties are not to be trusted, experts might not be the most reliable source either.

3   A VAA-specific approach for estimating parties' positions: the *Kieskompas*

Since most of the approached reviewed here are problematic in one way or another, *Kieskompas*, one of the most prominent VAA in the Netherlands, has worked out an approach which combines elements from existing approaches (Krouwel, Vitiello & Wall 2012). One the one hand, parties are asked to place themselves on the VAA statements and justify these placements by providing evidence from their manifestos or other publicly available documents. One the other hand, a group of experts is codes the party placements using the party manifestos. The two estimates are then brought together and compared. In the case of discrepancies, parties are asked to reconsider their placement in light of the expert estimates. In the case when parties can bring additional evidence to justify their position against the experts' estimate, then this information is taken into account. The outcome of this process is a party placement on either direction depending on which side provides the most convincing evidence.

In many respects this approach is an improvement over the situation which parties can self-place without any checks from the VAA designers, or the situation where experts place parties on their basis of their judgements. The combination of the two sources in the *Kieskompas* approach, however, still requires the collaboration of political parties. Although this might be taken as given in the Netherlands where VAAs have been around since the late 1980s, it cannot be assumed in the case of other countries. Moreover, the coding of party positions on the basis of manifestos are subject to similar problems of disagreement. There are at least three sources of disagreement

among expert coders: a) they might use different documents, b) different evidence from the same document, or c) use the same evidence but interpret them differently (e.g. what I consider as 'completely agree' for a colleague might be merely 'agree'). Using a 'hierarchy of documents' like the EU Profiler did which adopted the *Kieskompas* approach does not address point a) adequately as, in many contexts, recent statements might be more valid presentations of where a party stands compared to a party manifesto which was heavily debated in a party congress many months before. Yet even if we accept that different documents should not be a problem, there are still other potential source of unreliability.

To investigate the latter I set up a coding exercise where 80 European Studies undergraduate students attending a course on conceptualization and measurement agreed to participate after being offered partial course credit. The students were assigned to German ($n = 41$) and Dutch ($n = 39$) language groups based on their native or fluency language and where asked to code a selection of Dutch or German parties on one of two sets of eight EU Profiler statements. The parties and sets of statements were randomized among students within each of the language groups. Moreover, the students were given their 2009 EP election manifestos as the only source and were asked to accompany each of their estimates with the exact source they used copy/pasted from the manifesto. As expected, the agreement of their estimates varied widely with van der Eijk's A ranging from -.1 to 1. Averaging A for each statement across parties resulted in figures between .55 and .78 indicating that, even when students use exactly the same documents, there is enough disagreement to render some of their estimates as unreliable and potentially invalid. Of course, some of the disagreement which results in A's should be attributed in mistakes in the interpretation of the scales. Even if such mistakes are taken into account and corrected (by looking at the associated statements that students used for giving each code) we can still find substantial disagreements which are solely due to differences in the interpretation or sources as noted above. Figure 2 presents some such findings in terms of the distribution of the codes. Whereas students generally agree as to where to place Büdnis 90/Die Grünen with respect to EU farmer subsidies, they use different statements from the manifestos or interpret the same statements differently when it comes to estimating their position in EU/Russia relations, or even worse the positions of SPD or FDP in other issues.

How does the *Kieskompas* handle such disagreements? The EU Proler, relies on 'discussions among team members' and consultations with experts and the VAA leadership (Trechsel & Mair 2011). Making decisions by consensus among the VAA team members, however, does not guarantee coding reliability. Armstrong (2006, 5) makes a strong point when he argues against
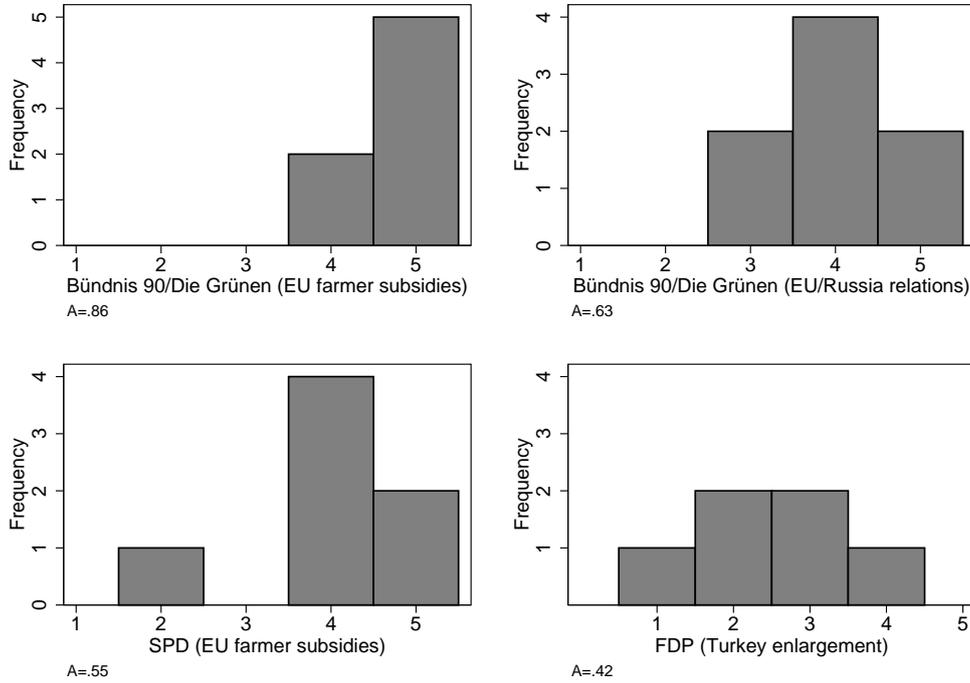
Figure 2: Distribution of estimates in a manifesto coding.

estimating quantities based on face-to-face meetings. Some people are louder than others, some are more powerful or prestigious. Some poeple voice their opinions and do not listen to others, yet others listen to what others say and do not have time to think for themselves. As Krippendorff (2004, 217) warned, 'in groups like these, observers are known to negotiate and to yield to each other in tit-for-tat exchanges, with prestigious group members dominating the outcome [. . .] and coding comes to reect the social structure of the group'. Armstrong (2006, 6) therefore suggests that estimation can be improved when the procedure guarrantees that the opinions can be stated indepedently from one another and when they can be aggregated using a 'pre-determined mechanical scheme'. Moreover, the current practice of group discussions which prevents group members of working indepedently to each other prevents the calculation of formal measures of disagreement which can be used to gauge reliability. A solution to this problem is outlined in the next section where I propose a formalization of the coding process among experts.

4 Learning from other disciplines: the Delphi

The problem of expert disagreement is certainly not unique to political science, let alone VAAs. Psychologists, computer scientists, demographers and professionals in the health sciences all encounter situations where groups of experts are called to given answers to specific questions. Consider the following example. A group of physicians have just examined a patient and need to make an assessment of what the patient suffers from. Each of the physicians has access to a set of medical tests which can be used to make the diagnosis (x-rays, biopsies and so on). These physicians, however, likely to interpret these tests in different ways and also likely to assign different weights to them when making their diagnosis. The physicians could convene in the room and try to agree on a diagnosis by taking to each other but, as Armstrong and Krippendorff warned, the discussion will be most likely dominated by those who have high prestige or a strong personality and not necessarily by those who can make the most persuasive arguments.[6] It has been therefore quite popular for the past 50 or so years to try and solve the problem of expert disagreemet by using a method called 'Delphi'.

The Delphi method was first used in the 1950s to forecast technological changes (Dalkey & Helmer 1963) and has been since applied in many different contexts. It is based on three principles: a) anonymity, b) statistical aggregation, and c) feedback. Regular expert surveys use only the first two principles whereas the coding of party positions on the basis of group discussions (the *Kieskompas* approach) uses only the last two principles. The process of the Delphi method is quite simple. A 'moderator' selects a panel of experts and asks them to give some estimates on the questions of interest and *justify each of their estimates* (see Figure 3). This justification may come at the form of an argument, or could point to the sources that the expert have used for estimation. The experts work independently of each other and without knowing the identities of the other experts involved. The moderator then collects the responses and gives feedback to the experts for a second round of estimation. The nature of the feedback in Delphi can vary. The moderator can give measures of central tendency of the responses in the first round (median, mean), the minimum and maximum score along with the justifications, a combination of both, or can even opt to feed back each and every estimate along with its justification. In the second round, the experts evaluate the other anonymous estimates and the associated justifications and provide a second estimate which might be adjusted based on the feedback information. An important characteristic of this feedback is that it is anonymous. As in the first round, each expert is unaware of the identities of the other experts and cannot tell where each estimate/justification comes from. In addition,

to avoid the moderator being biased in favour or against any of the experts, the process is (quasi-)double blind. The moderator knows the identities of the experts but cannot tell which expert gave what estimate/justification. The same process can be repeated over a number of rounds, which can be predefined or otherwise decided by the moderator.

A considerable body of literature (for a comprehensive review see Rowe & Wright 1999) has shown that the Delphi method gives more accurate estimates compared to mere statistical aggregation across experts (the case of classic expert surveys) or unstructured group discussions (the case of *Kieskompas*) especially when detailed feedback is given from one round to another. The anonymity here plays a crucial role as it guarantees that convergence on the subsequent rounds will be based on the quality of arguments associated with the initial estimates and not of the personalities of the experts involved.
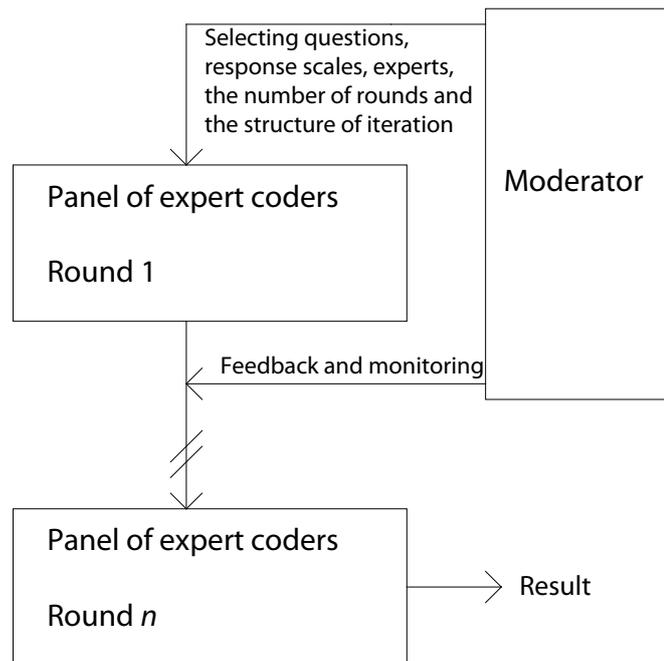
Figure 3: An outline of the Delphi method.

How does the Delphi method work in the VAA context? A first test was conducted for estimating parties' positions for Choose4Greece, a VAA set by the *Preference Matcher* consortium of researchers.[7] The moderator contacted a group of twelve expert coders (PhD students or faculty members) and asked

them to choose the parties which would be most comfortable working with. Five expert coders were assigned to each party and the process took place in an online platform which was designed specifically for the Delphi method.[8] The facilitator supervised the process and participants were remunerated with 70 euros per coded party. Expert coders were given a list links to documents (party manifestos, policy sections in party websites, transcripts of parliamentary debates) that they could use for justifying their estimates. They were also told that they could use alternatives or, in case where they could not find any other source, that they could provide a justification based on their personal knowledge of the party. The coding took place over two rounds and lasted about two working days for each coded party. During the first round they experts had to give an estimated positions for each of the given issues for the party (or parties) they were assigned to. In addition, they had to justify their estimate using the aforementioned options. During the second round the experts were presented with the median response as well as all the individual (anonymous) estimates of the first round and their associated justifications and were asked to estimate parties' positions in light of this feedback. The median response from the second round was taken as the final expert estimate.

How high was the expert agreement on the first round and what was the impact of the second round? Table 1 presents some results from the coding process regarding the solar panels question in Choose4Greece which is indicative of the high difficulty questions expert coders are faced with. The question asked whether parties would permit the installation of solar panels on agricultural land which could be used to produce agricultural products. This was not a big issue in the campaign, but Choose4Greece associated many of its questions with legislation that was proposed over the two years prior to the election, this was one of the issues that came up. As most parties did not include references to solar panels in their manifestos (or did not publish manifestos at all), I consider this as a challenging issue which can be used to evaluate the effectiveness of the Delphi method.

As see from the table, for 10 out of the 14 parties coded, the expert coders could not agree where to place parties as evident from the low ($<$ .7) coefficient of perceptual agreement (van der Eijk 2001). In all of the parties but one the second round improved the degree of agreement. By looking at the justifications to the estimates given by the experts during the first round, it appears that converged during the second round was often achieved by the following mechanism. When experts had little information about the party position (evidenced by the justification they had given during the first round), they changed their estimate in accordance to the estimates that were associated with rich and compelling evidence. In cases of absence

14

Table 1: Expert agreement over two Delphi rounds.

	van der Eijk's A	
	1st round	2nd round
Democratic Alliance	1	1
Recreate Greece!	.9	1
SYRIZA	.88	.7
ANTARSYA	.88	1
LAOS	.63	.8
Golden Dawn	.63	.8
Drassi	.6	.9
Social Agreement	.58	.9
PASOK	.4	.47
ND	.13	.9
KKE	.13	.9
Democratic Left	.06	.75
Independent Greeks	-.14	.43
Ecologist Greens	-.6	.23

of compelling evidence, experts with less information often changed their estimates towards the median response of the fist round. In the only case where agreement was reduced, it did not fall below .7. In the cases of the Independent Greeks and Ecologist Greens, some of the experts confused the direction of the scales and placed parties on the other side of the scale implied by their given justification. The moderator was easily able to detect these instances and alerted the coders collectively (since the moderator could not tell who did the mistake). In the case of the Ecologist Greens one coder seems not to have missed this alert and repeated the mistake in the second round resulting in a low perceptual agreement coefficient (.23). Nevertheless, because the estimate came with a justification which clearly pointed that the expert intended to place the party on the complete opposite side of the 5-point scale, the moderator corrected the mistake at a later stage ($A = .9$).

Did this improvement in perceptual agreement led to any differences in the estimates of party positions between the first and second round? The answer to this question depends highly on the measure of central tendency which is used to aggregate responses. When the mean response is used, there are noticeable differences between the first and second round. Considering that the 5-point response scales are ordinal, however, the median response is the appropriate measure, which was also used by Choose4Greece. Since the

median is robust to outliers, there were very few changes in party positions despite the expert disagreement. In the question about solar panels, for instance, there was a change for KKE (from 'agree' to 'completely agree'), for the Democratic Left (from 'completely agree' to 'agree') and the Independent Greeks (from 'neither agree, nor disagree' to 'completely agree'). Although the differences are small, the noisier estimates of the first round could have implications for the validity of inferences as, *ceteris paribus* there is a smaller chance of estimates being valid when the estimates are unreliable.

5   Conclusions

The paper showed that the established methods for estimating parties' positions (content analysis, expert surveys, the analysis of roll-call voting) have prohibitive disantvantages when applied to the context of VAAs. Moreover, relying on party self-placement is often impractical and/or risky. The approach followed by *Kieskompas*, and other VAAs of the same 'family' such as EU Profiler, is an improvement over the existing approaches but relies too much on the favourable Dutch context (e.g. parties which consistently publish manifestos, parties which always respond to questionnaires) while it does not provide formal measures of reliability.

This paper proposed an alternative approach to estimating parties' positions by using the Delphi method. The results of applying the Delphi to Choose4Greece gave encouraging results. The method was able to get expert coders to reach agreement even in challenging situations where parties do not publish manifestos or when they do not talk about the issues used by VAAs in them. Importantly, the formalization of the Delphi method enabled us to obtain precise measures of agreement which gives an indication of the estimates' reliability. Moreover, the process of two rounds enabled postcoding reconciliation which improved the reliability of data (Krippendorff 2004, 219), while anonymity guaranteed that this reconciliation was based solely on the quality of arguments and sources brought forward and not on the personalities of those involved in the coding process.

There are also added advantages of having formal measures of expert agreement over the two rounds. The Choose4Greece team used the agreement measures of the second round to determine which questions would be used in the public release of the VAA. The experts coded more questions than the VAA intended to use so the team was able to drop those that had a low coefficient agreement ($A < .7$) for more than two parties. Moreover, the agreement measures from round one can be used as proxies for issue importance in a VAA. Most VAAs ask users to indicate how important are

the issues to them and consequently use matching algorithms which weight for the degree of issue importance. None of the available VAAs, however, has tried to introduce issue importance from the parties' side, although this can be easily accommodated in a matching algorithm. The reason is probably due to the fact that no researcher has devised measures of issue importance for parties which can be used in conjunction with specific questions such as those which appear in VAAs. Expert agreement on a party position could be a proxy of issue clarity or issue importance. If parties think that some issues are more important than others, then their positions would be communicated very clearly in the course of an election campaign, something which would drive expert agreement during the coding process. Although some might find such proxying controversial, I suggest that this is a suggestion which would be useful if it is further debated.

Of course one can argue that the method proposed in this paper is more costly and time-consuming than the *Kieskompas* approach. This is certainly true but we need to consider that the Delphi is considerably cheaper and time-efficient compared to other well-established methods such as the hand-coding of party manifestos (Volkens 2007). Moreover, running a Delphi does not require the responses of political parties, although such responses can be solicited and used in addition to the Delphi estimation.[9] These apparent advantages make the estimation method proposed here a candidate for serious consideration among the community of VAA researchers and beyond.

Notes

[1]See Budge (2006), Budge (2001), Budge (2000), Dinas & Gemenis (2010), Benoit & Laver (2007*b*), Benoit, Mikhaylov & Laver (2009), Gemenis (2013*b*), Gemenis (2012), Hooghe et al. (2010), Laver (2001*b*), Laver, Benoit & Garry (2003), Lowe et al. (2011), Marks et al. (2007), Mikhaylov, Laver & Benoit (2012), Pennings (2011), Ray (2007), Slapin & Proksch (2008), Steenbergen & Marks (2007), Volkens (2007).

[2]See Benoit & Laver (2007*a*), Benoit & Laver (2008), Benoit et al. (2012), Budge & Pennings (2007*a*), Budge & Pennings (2007*b*), Budge & McDonald (2012), Martin & Vanberg (2008), Lowe (2008).

[3]Consider the following example where two voters are matched to a party on the basis of binary (left/right) responses to three issues which are combined to form a single 4-point (0–3) scale. Both voters are right-leaning as they score 'right' on two out of three issues and so does the party. By taking their positions on this left-right scale we would assume perfect congruence, since all three actors have exactly the same score. This is not true, however, as it is very clear that only voter A is perfectly congruent with the party. The false impression of congruence stems from the assumption of interchangeable issue positions. We cannot assume that attitudes on, say, taxes, pensions and privatization are interchangeable even when we consider them to be indicators of a single (economic) left-right scale.

	Voter A	Voter B	Party
Issue 1	Left (0)	Right (1)	Left (0)
Issue 2	Right (1)	Left (0)	Right (1)
Issue 3	Right (1)	Right (1)	Right (1)
Score	2	2	2

[4]Some of these are compared by Mendez (2012). Louwerse & Rosema (2013) are right to comment that such algorithms effectively assume a n-dimensional space where n is the number of issues.

[5]Benoit et al. (2012), Benoit & Laver (2008), Budge & McDonald (2012), Franzmann & Kaiser (2006), Gabel & Huber (2000), Grimmer & Stewart (2013), Lowe et al. (2011), citeasnounMartin2008, Slapin & Proksch (2008)

[6]Unless of course, the most prestigious and hard-headed physician also happens to be the most resourcive and qualified in the group. This seem to be happening only in the fictional Dr. House.

[7]See http://www.preferencematcher.org/

[8]See http://armstrong.wharton.upenn.edu/delphi2/

[9]The Choose4Greece solicitied responses from all parties by sending them the VAA questionnaire. Only two parties responded and the team used the responses in fine-tuning these parties' positions on a couple of isues where there was higher than average expert disagreement after the second Delphi round.

References

Armstrong, J. Scott. 2006. "How to make better forecasts and decisions: Avoid face-to-face meetings." *Foresight: The International Journal of Applied Forecasting* 5:3–8.

Baker, David, Andrew Gamble, David Seawright & Katrina Bull. 1999. "MPs and Europe: enthusiasm, circumspection or outright scepticism?" *British Elections & Parties Review* 9:171–185.

Bakker, Ryan, Catherine de Vries, Erica Edwards, Liesbet Hooghe, Seth Jolly, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen & Anna Milada Vachudova. 2012. "Measuring party positions in Europe: The Chapel Hill Expert Survey trend file, 1999–2010." *Party Politics* doi: 10.1177/1354068812462931.

Benoit, Kenneth & Michael Laver. 2007a. "Benchmarks for text analysis: A response to Budge and Pennings." *Electoral Studies* 26:130–135.

Benoit, Kenneth & Michael Laver. 2007*b*. "Estimating party policy positions: Comparing expert surveys and hand-coded content analysis." *Electoral Studies* 26:90–107.

Benoit, Kenneth & Michael Laver. 2008. "Compared to what? A comment on 'A robust transformation procedure for interpreting political text' by Martin and Vanberg." *Political Analysis* 16:101–111.

Benoit, Kenneth, Michael Laver, Will Lowe & Slava Mikhaylov. 2012. "How to scale coded texts without bias: A response to Gemenis." *Electoral Studies* 31:605–608.

Benoit, Kenneth, Slava Mikhaylov & Michael Laver. 2009. "Treating words as data with error: Uncertainty in text statements of policy positions." *American Journal of Political Science* 53:495–513.

Budge, Ian. 2000. "Expert opinions of party policy positions: Uses and limitations in political research." *European Journal of Political Research* 37:103–113.

Budge, Ian. 2001. "Validating party policy placements." *British Journal of Political Science* 31:210–223.

Budge, Ian. 2006. Identifying dimensions and locating parties: methodological and conceptual problems. In *Handbook of party politics*, ed. Richard S. Katz & William Crotty. London: Sage pp. 422–433.

Budge, Ian & Michael D. McDonald. 2012. "Conceptualising and measuring 'centrism' correctly on the left-right scale (RILE) without systematic bias: A general response by MARPOR." *Electoral Studies* 31:609–612.

Budge, Ian & Paul Pennings. 2007*a*. "Do they work? Validating computerised word frequency estimates against policy series." *Electoral Studies* 26:121–129.

Budge, Ian & Paul Pennings. 2007*b*. "Missing the message and shooting the messenger: Benoit and Laver's 'response'." *Electoral Studies* 26:136–141.

Dalkey, Norman C. & Olaf Helmer. 1963. "An experimental application of the Delphi method to the use of experts." *Management Science* 9:458–467.

Dinas, Elias & Kostas Gemenis. 2010. "Measuring parties' ideological positions with manifesto data: A critical evaluation of the competing methods." *Party Politics* 16:427–450.

Fivaz, Jan & Giorgio Nadig. 2010. "Impact of Voting Advice Applications (VAAs) on voter turnout and their potential use for civic education." *Policy & Internet* 2(4):167–200.

Franzmann, Simon & André Kaiser. 2006. "Locating political parties in policy space: A reanalysis of party manifesto data." *Party Politics* 12:163–188.

Gabel, Matthew J. & John D. Huber. 2000. "Putting parties in their place: Inferring party left-right ideological positions from party manifestos data." *American Journal of Political Science* 44:94–103.

Gemenis, Kostas. 2012. "Proxy documents as a source of measurement error in the Comparative Manifestos Project." *Electoral Studies* 31:594–604.

Gemenis, Kostas. 2013a. "Estimating parties' positions through voting advice applications: Some methodological considerations." *Acta Politica* 48:268–295.

Gemenis, Kostas. 2013b. "What to do (and not to do) with the Comparative Manifestos Project data." *Political Studies* 61(S1):3–23.

Gemenis, Kostas & Roula Nezi. 2012. "The 2011 political parties expert survey in Greece [computer file and codebook]." Data Archiving and Networked Services (DANS) [distributor] Persistent identifier: urn:nbn:nl:ui:13-a9zi-p9.

Grimmer, Justin & Brandon M. Stewart. 2013. "Text as data: the promise and pitfalls of automatic content analysis methods for political texts." *Political Analysis* doi: 10.1093/pan/mps028.

Hansen, Martin Ejnar. 2009. "The positions of Irish parliamentary parties 1937–2006." *Irish Political Studies* 24:29–44.

Helbling, Marc & Anke Tresch. 2011. "Measuring party positions and issue salience from media coverage: discussing and cross-validating new indicators." *Electoral Studies* 30:174–183.

Hix, Simon, Abdul Noury & Roland Gérard. 2006. "Dimensions of politics in the European Parliament." *American Journal of Political Science* 50:494–511.

Hooghe, Liesbet, Ryan Bakker, Anna Brigevich, Catherine de Vries, Erica Edwards, Gary Marks, Jan Rovny, Marco Steenbergen & Milada Vachudova. 2010. "Reliability and validity of the 2002 and 2006 Chapel Hill

expert surveys on party positioning." *European Journal of Political Research* 49:687–703.

Katsanidou, Alexia & Zoe Lefkofridi. 2010. "Citizen representation at the EU level: policy congruence in the 2009 EP election." Unpublished paper, GESIS-Leibniz Institute for the Social Sciences.

Krippendorff, Klaus. 2004. *Content Analysis: An Introduction to its Methodology.* Second ed. Thousand Oaks, CA: Sage.

Krouwel, André, Thomas Vitiello & Matthew Wall. 2012. "The practicalities of issuing vote advice: A new methodology for profiling and matching." *International Journal of Electronic Governance* 5:223–243.

Ladner, Andreas & Joëlle Pianzola. 2010. Do voting advice applications have an effect electoral participation and voter turnout? Evidence from the 2007 Swiss federal elections. In *Electronic participation*, ed. E. Tambouris, A. Macintosh & O. Glassey. Berlin: Springer pp. 211–224.

Laver, Michael. 2001*a*. Position and salience in the policies of political actors. In *Estimating the policy position of political actors*, ed. Michael Laver. London: Routledge pp. 66–75.

Laver, Michael. 2001*b*. Why should we estimate the positions of political actors? In *Estimating the policy position of political actors*, ed. Michael Laver. London: Routledge, pp. 3–9.

Laver, Michael, Kenneth Benoit & John Garry. 2003. "Estimating the policy positions of political actors using words as data." *American Political Science Review* 97:311–331.

Louwerse, Tom & Martin Rosema. 2013. "The design effects of voting advice applications: Comparing methods of calculating results." *Acta Politica* doi: 10.1057/ap.2013.30.

Lowe, Will. 2008. "Understanding Wordscores." *Political Analysis* 16:356–371.

Lowe, Will, Kenneth Benoit, Slava Mikhaylov & Michael Laver. 2011. "Scaling policy positions from hand-coded political texts." *Legislative Studies Quarterly* 36:123–155.

Marks, Gary, Liesbet Hooghe, Marco Steenbergen & Ryan Bakker. 2007. "Crossvalidating data on party positioning on European integration." *Electoral Studies* 26:23–38.

Martin, Lanny W. & Georg Vanberg. 2008. "A robust transformation procedure for interpreting political text." *Political Analysis* 16:93–100.

Mendez, Fernando. 2012. "Matching voters with political parties and candidates: An empirical test of four algorithms." *International Journal of Electronic Governance* 5:264–278.

Mikhaylov, Slava, Michael Laver & Kenneth Benoit. 2012. "Coder reliability and misclassification in the human coding of party manifestos." *Political Analysis* 20:78–21.

Nezi, Roula, Dimitri A. Sotiropoulos & Panayiota Toka. 2010. "Attitudes of Greek parliamentarians towards European and national identity, representation and scope of governance." *South European Society & Politics* 15:79–96.

Pennings, Paul. 2011. "Assessing the 'Gold Standard' of party policy placements: Is computerized replication possible?" *Electoral Studies* 30:561–570.

Ramonaitė, Ainė. 2010. "Voting advice applications in Lithuania: Promoting programmatic competition or breeding populism?" *Policy & Internet* 2(1):117–147.

Ray, Leonard. 2007. "Validity of measured party positions on European integration: assumptions, approaches, and a comparison of alternative measures." *Electoral Studies* 26:11–22.

Rowe, Gene & George Wright. 1999. "The Delphi technique as a forecasting tool: Issues and analysis." *International Journal of Forecasting* 15:353–375.

Schwarz, Daniel, Lisa Schädel & Andreas Ladner. 2011. "Pre-election positions and voting behaviour in parliament: consistency among Swiss MPs." *Swiss Political Science Review* 16:533–564.

Scully, Roger & David M. Farrell. 2003. "MEPs as representatives: individual and institutional roles." *Journal of Common Market Studies* 41:269–288.

Škop, Michal. 2010. Are the Voting Advice Applications (VAAs) telling the truth? Measuring VAAs' quality. In *Voting Advice Applications in Europe: the state of the art*, ed. Lorella Cedroni & Diego Garzia. Naples: ScriptaWeb pp. 199–230.

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

Spirling, Athur & Ian McLean. 2007. "UK OC OK? Interpreting optimal classification scores for the U.K. House of Commons." *Political Analysis* 15:85–96.

Steenbergen, Marco & Gary Marks. 2007. "Evaluating expert judgements." *European Journal of Political Research* 46:347–366.

Talonen, Jaakko & Mika Sulkava. 2011. "Analyzing parliamentary elections based on voting advice application data." *Lecture Notes in Computer Science* 7014:340–351.

Trechsel, Alexander H. & Peter Mair. 2011. "When parties (also) position themselves: An introduction to the EU Profiler." *Journal of Information Technology & Politics* 8:1–20.

van der Eijk, Cees. 2001. "Measuring agreement in ordered rating scales." *Quality & Quantity* 35:325–341.

Volkens, Andrea. 2007. "Strengths and weaknesses of approaches to measuring policy positions of parties." *Electoral Studies* 26:108–120.

Wagner, Marcus & Outi Ruusuvirta. 2012. "Matching voters to parties: Voting advice applications and models of party choice." *Acta Politica* 47:400–422.

Walgrave, Stefaan, Michiel Nuytemans & Koen Pepermans. 2009. "Voting advice applications and the effect of statement selection." *West European Politics* 32:1161–1180.

Wall, Matthew, Maria Laura Sudulich, Rory Costello & Enrique Leon. 2009. "Picking your party online: An investigation of Ireland's first online voting advice application." *Information Polity* 14:203–218.

Wheatley, Jonathan, Christopher Carman, Fernando Mendez & James Mitchell. 2012. "The dimensionality of the Scottish political space: Results from an experiment on the 2011 Holyrood elections." *Party Politics* doi: 10.1177/1354068812458614.