PARTY POLITICS VOL 16. No.4 pp. 427–450

MEASURING PARTIES' IDEOLOGICAL POSITIONS WITH MANIFESTO DATA

A Critical Evaluation of the Competing Methods

Elias Dinas and Kostas Gemenis

ABSTRACT

Within the rapidly growing literature on positioning political parties along policy dimensions, the rich data series collected by the Comparative Manifestos Project (CMP) has been widely considered as the most systematic and objective source of information. For estimating parties' positions on the Left–Right dimension alone, several different methods have been proposed which make use of the CMP data. However, unless a new method is proposed, there will seldom be any attempt to check the robustness of the findings across different measurement strategies. In this article, we focus on the parties in Greece, which have been notoriously incorrectly positioned by the 'standard' method proposed by the CMP. We contrast the 'standard' method with various proposed alternatives and show that the latter outperform the former both in terms of face and convergent validity and in terms of reliability. In addition, we show that this cross-checking is essential, since different methods often lead to diametrically opposite results.

KEY WORDS ▪ ideology ▪ left–right ▪ party manifestos ▪ reliability ▪ validity

Introduction

Positioning political actors (and political parties in particular) along the Left–Right (L–R) continuum and other policy dimensions has been an important feature of recent empirical research in comparative politics. Based on a variety of theories and methods, political scientists are now 'able to operationalize a wide range of models within what has become an important sub-discipline of political science' (Laver, 2001a: 6). Three main approaches have been proposed for the study of party location: (a) expert surveys, (b) opinion poll data and (c) content analysis of party manifestos.[1] Although

1354-0688[DOI: 10.1177/1354068809343107]

there is ongoing discussion about the strengths and weaknesses of each approach (see Budge, 2000; Kleinnijnhuis and Pennings, 2001; Mair, 2001; McDonald et al., 2007; Steenbergen and Marks, 2007; Volkens, 2007), the last-mentioned has nevertheless become the most popular for two reasons: first, data from party manifestos attain a greater degree of impartiality. Expert surveys and opinion poll data give us the picture of the party as perceived by political analysts and voters, respectively. Manifestos, on the other hand, provide a more accurate and representative picture of where the parties stand in the policy space, without our requiring further knowledge about their policy record. Second, the Manifesto Research Group (MRG, now renamed Comparative Manifestos Project, CMP) has produced a rich time-series of data (see Budge et al., 2001; Klingemann et al., 2007) that is unrivalled by any other method. Consequently, the MRG/CMP approach has emerged as the prima facie method by which to estimate parties' policy positions against alternative data sources and also among alternative methods of coding party manifestos.

One result of this general consensus has been the proliferation of different analytical methods aimed at measuring parties' ideological positions and use of the CMP data (Budge, 1987; Franzmann and Kaiser, 2006; Gabel and Huber, 2000; Kim and Fording, 1998; Klingemann, 1995; Laver and Budge, 1992; van der Brug, 2001). There are two problems with most of these studies, however. First, they often fail to address the issue of face validity sufficiently, and they only sporadically examine the reliability of the produced estimates. In effect, the only method that has been subjected to any extensive analytical scrutiny is the 'standard' method originally developed by Laver and Budge (1992, for the performance of this method, see Budge and Klingemann, 2001; Klingemann et al., 2007). When the reliability and validity of the estimates are examined, they usually refer to a particular method, i.e. without rigorous examination of how each method stands up against the others. The few attempts that have compared the methods against each other (see Franzmann and Kaiser, 2006; Gabel and Huber, 2000) were motivated by the need to apply empirical support for a newly proposed measurement strategy.

In an attempt to depart from this pattern, the aim of our article is to engage in rigorous examination of the reliability and validity of the various methods of estimating parties' L–R positions through the CMP data. For illustrative purposes, we focus on the case of parties in Greece. First, we show that variability of the estimates calls for sensitivity testing when inferences are based on findings from one of these methods. Second, in the case of Greece at least, we show that empirical evidence does not support the tendency to adopt the 'standard' CMP method for measuring party positions, since it seems that existing alternatives outperform it in terms of both validity and reliability.

Competing Methods for Estimating Parties' Left–Right Positions

In this section, we briefly present the competing methods for estimating parties' positions in the L–R dimension, starting with the 'standard' method developed by the investigators of the CMP (Laver and Budge, 1992). After having reorganized all items within 20 policy dimensions, these authors established two 'marker' Left and Right items, respectively. Further items that constantly loaded highly with either of these groups, and that it made intuitive sense to add to either of the two categories, were also included. This procedure resulted in a fixed number of items forming the Left and Right sides of the dimension. Party scores were taken by the sum of the references of one group minus the other. We refer to this method as the 'standard' method of measuring parties' positions.

For a more case-specific analysis, an important qualification was proposed by Laver and Budge (1992: 26): seven marker variables (consisting of 28 items) were factor-analysed together with all the remaining items. Two (rotated) factors were extracted, corresponding well with the distinction between Left and Right. All items or marker variables loading highly on either the 'Left' or the 'Right' factor were included in the respective scales. Again, the final scale was constructed by subtracting the Left scale from the Right scale, which is known as the 'inductive' method.

Another method that was proposed (Budge, 1987) was based on the following procedure: all items of each domain were factor-analysed and one or two factors were extracted. All factors were then factor-analysed and the first of the new factors was assumed to be the L–R dimension. Here, this is called the 'two-stage factor analysis' (FA) method.

Klingemann (1995) proposed a more theoretically based method. Starting from the assertion that the L–R distinction does not refer to parties' policies about international matters, only items measuring domestic issues were included in the analysis. Two rotated factors were extracted from an FA of all these issues. By observing the pattern matrix, one of the two factors was then deemed to represent parties' positions in the L–R dimension. We refer to this as the 'domestic' method.

Thus far, all proposed models stem from the principal investigators of the CMP. Other scholars, however, have proposed different alternatives. Following Huber and Inglehart (1995), Gabel and Huber (2000) factor-analysed all issues available from the CMP data and then extracted a single common factor, supposedly representing an ideological 'super-issue', i.e. a reliable summary of the positions of parties in the issue space. Following the suggestion of the authors, we refer to this as the 'vanilla' method.

The last method tested here starts with the distinction between valence and position issues (Franzmann and Kaiser, 2006). The relative frequency of references to each item was regressed against party dummies (case-specific approach). Items were divided into position (Left versus Right) and valence issues. Party scores were taken by the ratio of Right minus Left references divided by the total number of references. Frequencies of position issues were adjusted in order to take their valence aspect into account: the minimum party score was subtracted from the original scores (this was implemented only in half the cases, since in all other instances the minimum score was zero). Then, party score for each year was adjusted by taking a weighted mean of the party's score in the previous and the next election (weights are based on the length of the inter-election period). We refer to this as the 'regression' method. Examination of Greek parties' positions involves comparison of the following methods: standard, inductive, two-stage FA, domestic, vanilla and regression.[2]

Before we move forward in the analysis, an important clarifying point needs to be made. As evident, all the attempts to empirically place parties on the L–R dimension with the use of the CMP data follow the same logic: instead of being confined to economic matters, the L–R scale is constructed in ways in which it encompasses both economic and non-economic issues. Therefore, our understanding of L–R throughout this article defines it as an overarching dimension encompassing both economic (regulatory versus neoliberal) and non-economic (authoritarian versus libertarian) issues. Reassuringly, empirical evidence from surveys conducted during the period under study shows that the voters in Greece also perceive the L–R in both economic and non-economic terms.[3]

Reliability of CMP Estimates

Klingemann et al. (2007: 88–97) have recently put the CMP data under rigorous reliability testing, and with favourable results. The Greek parties, however, were missing from this analysis despite the authors' claim that they used 'all established Western democracies in the CMP data-set' (Klingemann et al., 2007: 90). We therefore start our empirical examination by focusing on the reliability of the 'standard' method (Laver and Budge, 1992), which is the one most often used in general and the only one through which the trajectories of Greek parties' stances has been rigorously studied (Konstantinidis, 2004). We do that by following two different paths. First, we subject the underlying assumptions behind the use of the 'standard' L–R scale proposed by Laver and Budge (1992) to rigorous examination. Then we try a different analytical technique by which to compare this method with all other alternatives.

As has been described above, the 'standard' L–R method is constructed by subtracting a set of items denoting policies preferred by the Left indicators from an equal set of Right indicators (for an exact description, see Klingemann et al., 2007). There is a hidden but crucial assumption made here. The addition of a party's scores in all indicators of the two subsets of items implies that the resulting measure reflects the characteristics of a summated rating scale. Each item is given equal weight and all are assumed to belong in the same latent dimension. Thus, all left-wing items are assumed to measure an unobserved dimension referring to 'Left ideology' and all right-wing items are deemed to measure 'Right ideology'. The problem, however, is that a precondition for the incorporation of each item in such a scale is that it is monotonically related to all others (Jacoby, 1991). A rough-and-ready way to examine whether the criterion of monotone homogeneity is satisfied here is to run correlations between each item and a scale constructed by all other items forming the original scale except this. So, to see whether the indicator measuring positive references to military issues (per104) belongs to the right-wing scale, one would need to correlate this item with a scale containing all other right items except per104. Yet, since correlations measure linear functions which constitute a stricter assumption than monotone functions, this is a too demanding and thus problematic diagnostic test for the scalability of each item. The approach adopted here is different and less formal but probably more informative. A locally weighted regression curve (loess) has been fitted into a scatterplot between each item and the scale consisting of all other items. As all non-parametric regression methods, the basic idea behind the loess curve is to trace the salient features of the mean response making only minimal assumptions about its distribution (see Fitzmaurice et al., 2004: 69). Thus, a loess curve showing a monotonic pattern can be considered as a good indication that a given item fits to the scale. The results for each item are presented in Figures 1 and 2 for the Right and Left items, respectively. Although the analysis is confined to

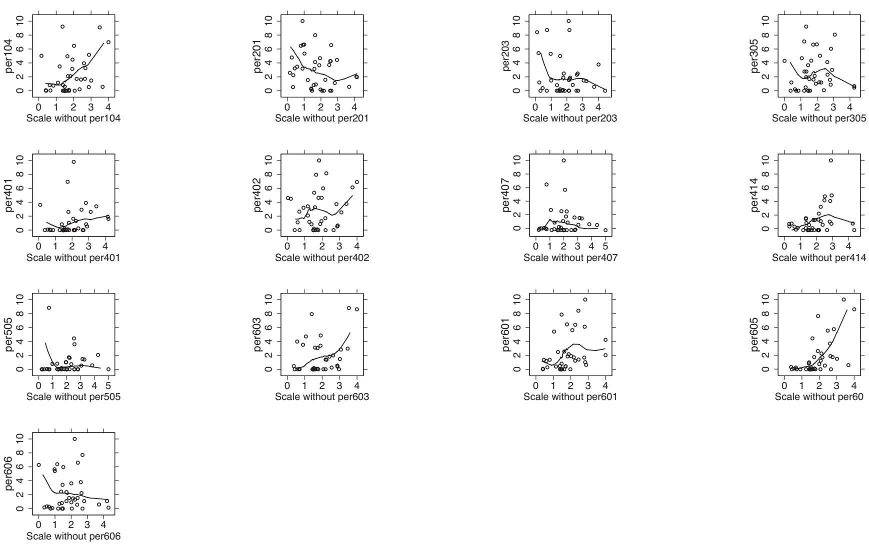

Figure 1. Examining the scalability of the items forming the right-wing indicators according to the 'standard' CMP method. Note: all items have been recoded, ranging from 0 to 10

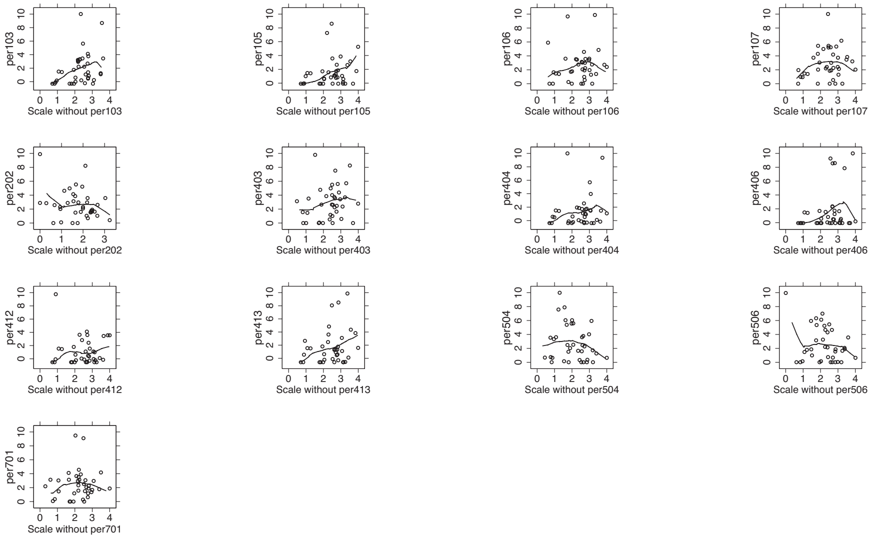

Figure 2. Examining the scalability of the items forming the left-wing indicators according to the 'standard' CMP method. Note: all items have been recoded, ranging from 0 to 10

the cases of Greek parties, the non-parametric nature of the loess curve mediates the problem of distribution assumptions difficult to be met with small *N* (38). Interestingly, in most cases the assumption of a monotonic relationship does not seem to be confirmed, since there is no evidence, whatsoever, that high scores in a given item are associated with high scores in the scale constructed by all other items. This casts doubt on the scalability of the selected items and, consequently, on the reliability of the scale. Importantly, a similar pattern is observed for both sets of indicators.

To be sure, Cronbach's alpha, which is a typical reliability statistic for summated rating scales, is only downward biased if the assumption of monotonicity is not confirmed. Indeed, for the Greek cases the estimate for Left items is 0.56, whereas for Right items this is 0.61. The problem, however, is that Cronbach's alpha is based on the assumption that all items included in a scale belong to the same underlying dimension. This means that construction of the two scales is based on the assumption that fluctuations across the items summed to create each scale are only random. If, however, these observed disturbances reflect systematic deviations stemming from various other underlying dimensions, the scale will appear reliable even when 'true' sources of variation stem from several latent dimensions. A way to explore this possibility is to encompass the items in a single dimension through an FA (factors extracted through the Iterated Principal factor method). Apart from the standard rule of thumb of eigenvalues greater than 1 (implying that we choose factors that account for a greater part of the variance than each single standardized item), a more helpful rule of thumb would be to focus on the angle-point of the screeplot of the eigenvalues associated with each added factor and extract all factors to the left of this point. Both rules of thumb point to a two-factor solution for each group of items.[4]

The analysis thus far has indicated that use of the standard method is not confirmed by the performance of this scale in terms of its reliability in the Greek case. The next step is to see whether other methods perform better regarding this criterion. Certainly, since all other methods rest on a procedure that goes beyond the simple logic of summated rating scales, a different strategy needs to be employed for the examination of their reliability. Given that the CMP data in the Greek case capture a period of 10 elections, we can make use of this time dimension in order to estimate the reliability of the different methods. Previous research on assessing single-item reliability with panel data has been guided by the work of Heise (1969) and Wiley and Wiley (1970). The basic idea of this procedure is that party ideology (as any other measured trait) follows a Markovian [AR(1)] process, such that its current value is a function of its previous value plus some random disturbance. Heise's (1969) model of reliability and stability measurement is based on correlation matrices as opposed to Wiley and Wiley's (1970) covariance matrices. The difference between the two models refers to the identifying restrictions that need to be imposed on the data. Whereas the Heise model sets reliabilities constant across the waves, Wiley and Wiley constrain the measurement error variances. Applying the Heise model without making any of these restricting assumptions requires at least four waves in order to estimate the parameters of interest (Green and Palmquist, 1990). In this case, we can obtain the reliability estimates between the two adjacent non-extreme waves, i.e. wave two and wave three. In order to use this method and at the same time retain most of the cases for the analysis, each L–R variable is correlated with its three lags. This means that parties' positions in the following waves are simultaneously examined:

74–77–81–85
77–81–85–89a
81–85–89a–89b
. . . .
90–93–96–00

Table 1 presents the results for all methods.[5] As can be seen, there is substantial variability in the estimates. Evidently, the original 'standard' method performs less well than its more elaborate counterparts. Regarding the latter, although the 'domestic' method marks out as the most error-laden estimate, there is no procedure that produces estimates as high as those found by Klingemann et al. (2007) in the analysis of established Western democracies. Not surprisingly, the 'regression' method, which already imposes a smoothing restriction on the initial findings, seems to perform better in terms of reliability.

Table 1. Reliability of the estimates according to the Heise test–retest method

	Standard	*Two-stage FA*	*Vanilla*	*Regression*	*Domestic*	*Inductive*
$t_1 \rightarrow t_2$	0.59	0.68	0.67	0.71	0.49	0.71
$t_2 \rightarrow t_3$	0.65	0.67	0.69	0.73	0.55	0.73

N = 24.

An alternative interpretation of the findings could be that, although its key assumptions are met, the Heise model fails to capture changes in parties' positions net from measurement error. To address this possibility, we employ an alternative model which has been deemed to capture all necessary sources of change in parties' ideological stances. The alternation model, as it has been called, assumes that a party's position is a function of its previous position plus a time trend (see Budge, 1994). Employing this specification for each Greek party, the standard deviation of the predicted residuals across all parties has a mean value of 24.4. This value represents the amount of variance in parties' positions not explained by a specification whose theoretical underpinning is general ideological stability (denoted by the lag) and frequent zigzagging (captured by the trend). In the Klingemann et al. (2007) analysis, where Greece was not included, the two outlier countries were Finland and Iceland. Neither of the two reached this value (20.8 and 18.3, respectively).

Validity of CMP Data and Estimates

Construct Validity

According to the Carmines and Woods (2005: 936) definition, a measure is said to be construct valid 'if the empirically observed outcomes are consistent with the theoretical predictions'. In this sense, most attempts to cross-check the validity of CMP data were focused on whether the data fared well against the 'prevalence of valence issues' assumption which the coding of CMP was based on. The first attempt came from Budge and Farlie (1983: 274), who counted each manifesto's references to other parties and policies. The limited number of such references led them to conclude that there were low levels of confrontation and they hence concluded that the CMP's theoretical assumptions were correct. We, nevertheless, do not believe that measuring references to other parties constitutes a valid measure of the degree to which confrontation is marked. Manifestos are by definition texts serving to present a party's positions to the voters rather than texts comparing its capabilities or highlighting their differences with other parties. Moreover, Budge and Farlie's conclusion based merely on data from parties in the United Kingdom and the United States is unwarranted for other countries. Later on, the investigators of the CMP presented a more convincing argument.

By measuring the 'pro versus con' dyads within the CMP data they found 'the overwhelming number of references going to the one of the possible positions' (Robertson, 1987: 50–1), thus confirming that party manifestos are really about 'valence' politics, establishing in this way the construct validity of the CMP data. More recently, the 'salience' hypothesis was also confirmed by computerized word scores showing that different issues are selectively emphasized by different parties (Budge, 2001a: 221). Yet we should point out that these construct validity tests refer to the CMP-produced data (i.e. levels of salience for each coded issue), not to the ability of these data to estimate parties' positions on the L–R dimension. The latter require other types of validity test, such as convergent and face validity.

Convergent Validity

The proliferation of other approaches and methods of positioning political parties (mainly expert surveys and computerized counts) has made it possible to check the convergent validity, i.e. 'the comparison of a measure against one or more measures that are also measures of the same concept' (McDonald, 2005: 944), of CMP policy estimates (but not the data themselves). So far, the results have been rather mixed. Estimates for L–R positions do not seem to correlate with estimates from expert surveys (Benoit and Laver, 2007a; Klingemann et al., 2007: 77–9) or computerized word scores (Budge and Pennings, 2007). The latter, in particular, tend to 'flatten out' party movement across time. Budge and Pennings (2007: 123), however, argue that computerized word scores should be dismissed in favour of CMP data because the latter method has produced unsurpassed 'rich time-series data'. However, we cannot see how a 'richness of data' argument can be brought forward in a discussion about validity. 'Rich data' by no means implies 'valid data', a point taken by Benoit and Laver (2007b), who rightly questioned the notion of CMP being used as a 'benchmark' against which all other approaches should be measured.

Face Validity

Face validity addresses the question of whether a measure appears to be valid. Although it is often dismissed as a measure of validity in expert surveys (e.g. Ray, 1999: 289–90), because we would not expect experts to give us estimates that would not make sense anyway, it is a very good measure of checking the validity of data produced by manifesto content analysis. Laver et al. (2003), for example, discuss face validity extensively in assessing the validity of their computerized word scores estimates. Similarly, Laver and Budge (1992: 22) agree that the 'major check' of the CMP method should be 'the extent to which it generates results that make sense within countries'. Nevertheless, it is readily assumed that the estimates 'pass the test of face validity' (Klingemann et al., 2007: 63) without further inquiry. Whenever there is any discussion about face validity, this concerns only the cases of the United States and the United Kingdom, which seem to fit the 'historical experience' (e.g. Budge, 2001b: 53); the estimates for the remaining cases are simply described as 'quite plausible' (Budge, 2001a: 216). But exactly how plausible is 'quite plausible'?

The recent rigorous examination of individual party systems, such as the Italian, reveals that the CMP data 'do not do a very good job, in terms of *face validity*, of describing parties' locations on the left–right dimension' (Pelizzo, 2003: 67, emphasis added). These problems are evident in other countries 'as diverse as Austria, Belgium, Denmark, France, Germany, Italy and the Netherlands' (Franzmann and Kaiser, 2006: 164) and, most recently, Switzerland (Hug and Schulz, 2007). This criticism can be expanded to include the case of Greece as well, where Budge and Klingemann (2001: 33) awkwardly observed: 'at first sight Greece is not [convincingly represented by our estimates]. Here the Left is occupied by the Socialists (PASOK). The Communists are in the middle, except for 1989. At two points [the Communists] converge with the (conservative) New Democracy Party'.

In explaining this anomaly, Budge and Klingemann (2001: 33–4) maintained that the ideological convergence of the Communist Party of Greece (KKE) with the conservative New Democracy (ND) could be explained by the fact that the two parties formed a coalition government in 1989. This argument, however, could be valid only if the two parties converged in 1989, which was not the case. In the data provided, the two parties converged in 1985 and again in 1996 but not in 1989, where in fact the ideological distance between them is the greatest reached in the entire 1974–96 period. This selective reading of the estimates and the misinterpretation of the 1989 coalition as an explanatory factor of the 'observed' ideological convergence between what has been consistently described as a Stalinist communist party (e.g. Hanley, 2008: 140) and a conservative party, is a prime example of the procrustean use of an argument in order to fit biased data. Similarly, in a more in-depth analysis, Konstantinidis (2004) attempted, but with little success, to reconcile the findings with common knowledge about Greek politics. As is often the case in such instances, the justification for this discrepancy focused on the case rather than on the theory: the failure of the model to depict in a plausible manner the positions and the trajectories of Greek parties during the past three decades was largely attributed to several eccentricities of the Greek party system. To be sure, although such particularities possibly existed until the mid-1980s (see Clogg, 1987), they are unlikely to account for the inconsistency observed until the mid-1990s.

Ever since these problems of face validity were identified, several political scientists have tried to find out 'what went wrong'. Gabel and Huber (2000) believe that the solution lies in using a uniform encompassing L–R dimension for all countries in the dataset. The explanation of Laver (2001b: 73) is different, as he argues that the problem lies in the fact that 'despite being concerned fundamentally with the *salience* of different policy concerns [CMP data] have been used in practice to derive estimates of party *positions*', a point that has been taken by Franzmann and Kaiser, too (2006).

In an attempt to examine these arguments about face validity, we employed all the aforementioned methods in constructing a unifying L–R dimension for Greece, with the CMP data (Budge et al., 2001; Klingemann et al., 2007) for the period between 1974 and 2000.

Before proceeding into the analysis, however, we need to consider the usefulness of employing this intuitive test of face validity. To start with, we have to speculate about whether, and about what extent, a party manifesto should reflect and reproduce prior intuitions. In other words, if parties do not reflect their positions in their manifestos, then any method of measuring party positions with manifesto data will fail to confirm the established patterns, simply because the initial data are not valid.

There are two reasons why a manifesto might not always reflect the party's true positions. The first has to do with distribution of the ideological spectrum within a given political context. As Bartolini and Mair (1990: 199) have argued, in contexts where ideological differences between parties are clear, party manifestos might not prove reliable indicators of their positions because parties that are clearly differentiated on ideological grounds have enough freedom to present a slightly or even substantially different policy image. This is sometimes the case of extreme parties, which, having distinguished themselves from mainstream parties in everyday political discourse, are in a position to present more middle-of-the-road policy stances, avoiding any explicit expression of those policy views that have characterized them as extreme. Inversely, in countries where ideological differences are only small, parties need to differentiate themselves in policy terms, something which is then reflected in their manifestos. For instance, the relatively greater success of the CMP estimates in Great Britain – a country with many valence issues according to Franzmann and Kaiser (2006: 180) – rather than in Italy, can also be attributed to this pattern. We take this argument into account by performing a case study analysis where contextual particularities can be taken into consideration more explicitly.

A second reason might be that parties produce manifestos not in order to present their policies but in order to shape voters' perceptions about their future policy stances (see Pelizzo, 2003). Apart from the fact that this argument is partially based on the contested assumption that voters do indeed read parties' manifestos, at least in the case of Greece it leads to non-distinguishable empirical implications. If parties design their electoral programmes in order to shape voters' attitudes about their future issue stances, we should expect voters' perceptions about parties' ideological positions to correlate more with lagged rather than contemporaneous parties' positions estimated with manifesto data. In testing this possibility, we find that parties' positions correlate equally with voters' both contemporaneous and lagged ideological perceptions about the parties (the results are available in the online Appendix, see note 3). This implies that the counter-argument of shaping voters' attitudes as opposed to reflecting the party's stances cannot explain potential distortions between the findings and what we already know about the parties, since the results under both scenarios would look identical.

For all these reasons, we can expect a broad convergence between parties' manifestos and prior intuitions about their positions. Nevertheless, our prior knowledge should be based on the available secondary literature, which looks at parties' published documents (such as manifestos) and policy record, but also takes into account the general political context in which parties operate. This is exactly how we try to evaluate the findings from the different methods. Certainly, various inconsistencies might be due to the manifestos themselves or simply to random error. This is why we concentrate only on inconsistencies which not only cannot be attributed to the data (since there is inconsistency between the estimates based on the same data), but which in addition contradict previous literature and widespread convictions about the relative locations of parties. Figures 3–8 display the findings according to each of the proposed methods.

Looking at the left of the political spectrum, we would expect to find the Panhellenic Socialist Movement (PASOK) to the right of the communist party (KKE). Although PASOK has been one of the leftist socialist parties in Europe, with evident aspects of populism, anti-Americanism and anti-Western orientations (Moschonas, 2002), KKE managed to earn the title of

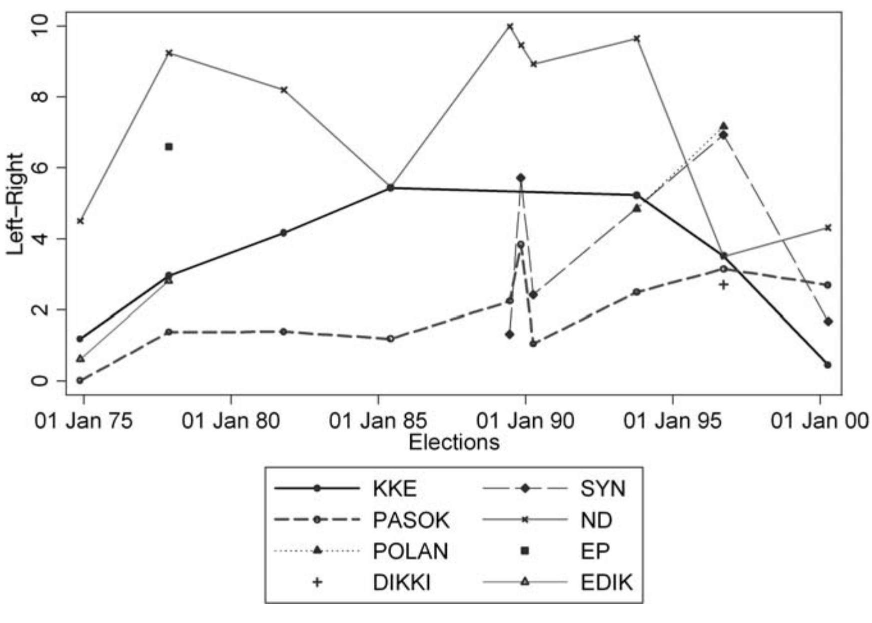

Figure 3. Greek parties' positions according to the 'standard' CMP method (Laver and Budge, 1992). Note: In all methods, the resulting scales have been recoded so as to range between 0 (extreme Left) and 10 (extreme Right)

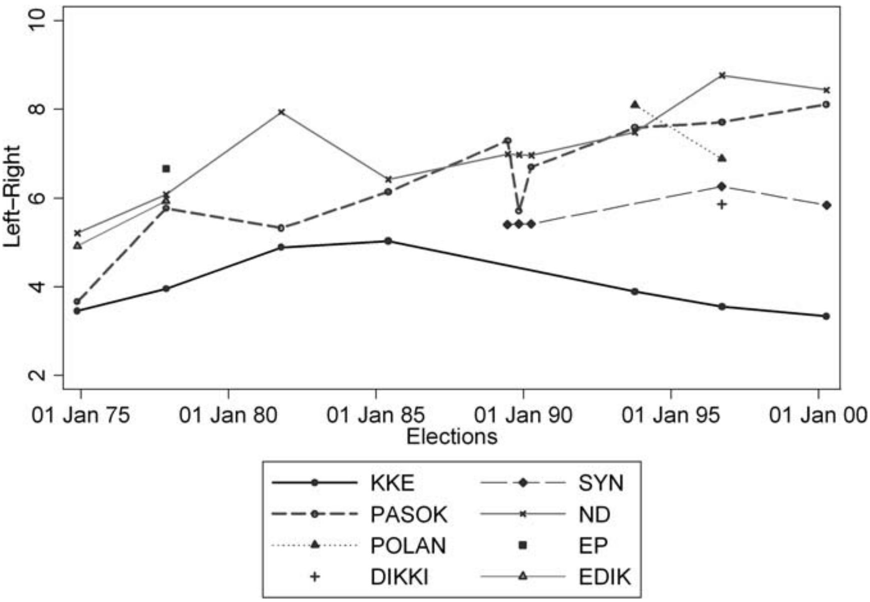

Figure 4. Greek parties' positions according to the 'vanilla' method (Gabel and Huber, 2000)

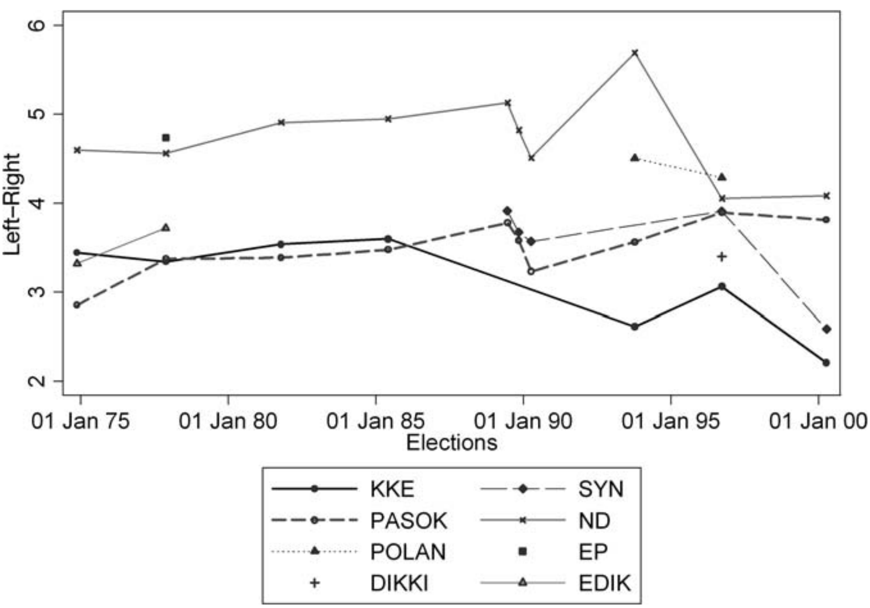

Figure 5. Greek parties' positions according to the 'two-stage factor-analysis' method (Laver and Budge, 1992)

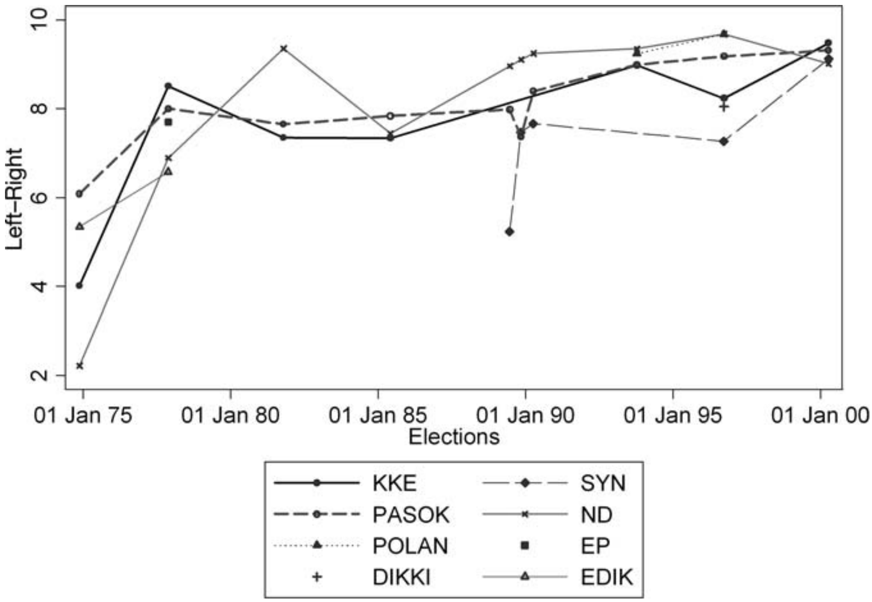

Figure 6. Greek parties' positions according to the 'domestic' method (Klingemann, 1995)

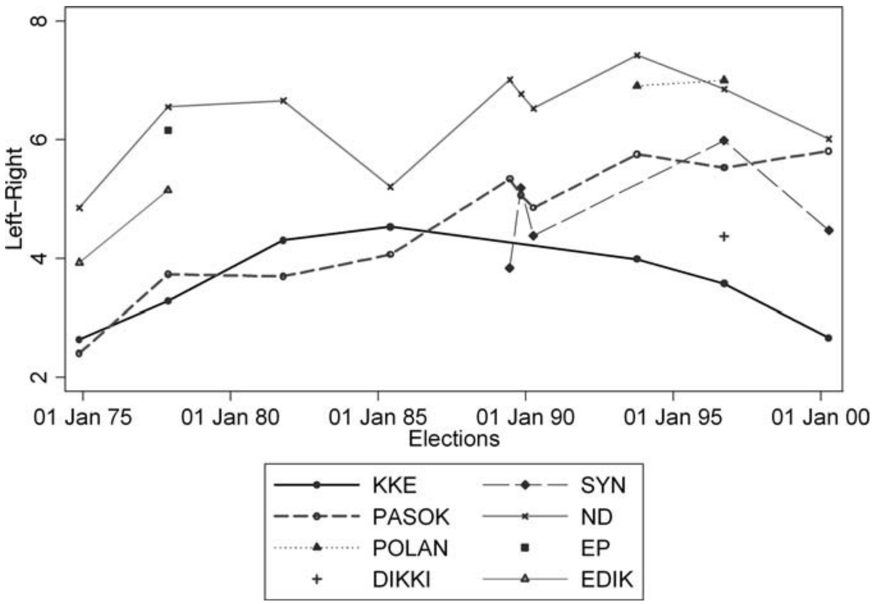

Figure 7. Greek parties' positions according to the 'inductive' method (Laver and Budge, 1992)

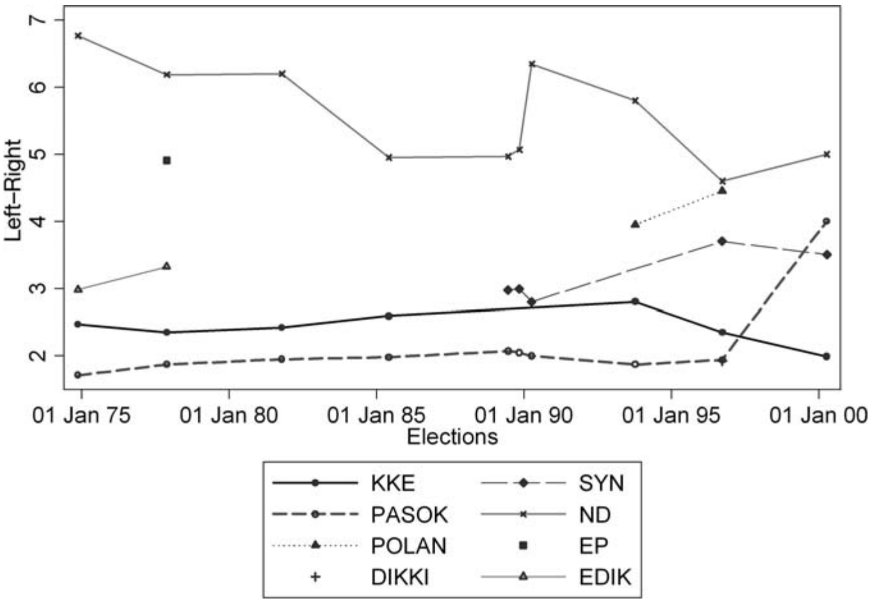

Figure 8. Greek parties' positions according to the 'regression' method (Franzmann and Kaiser, 2006); smoothed solution

one of the most radical communist parties in Western Europe (Bosco, 2001; Hanley, 2008). Moreover, PASOK has lost much of its radicalism since it went into office in 1981 (Spourdalakis, 1988), therefore we can expect to find it placed to the right of KKE at least since the mid-1980s. Surprisingly enough, this finding is consistent only with the 'a-theoretical' 'vanilla method' (Gabel and Huber, 2000), at the cost, however, of equating in various instances the socialists with a typical conservative right-wing party, as ND was at least until 1985. For most of the other methods, there is either constant 'leap frogging' between the two parties ('domestic', 'two-stage FA', 'inductive' methods), or PASOK appears steadily to the left of KKE (as in 'standard' and 'regression' methods). Moving to the period until 1995, according to the 'standard' method, ND, which appears to move its position dramatically from election to election, finds itself in the same ideological position with KKE, both in 1985 and 1996. This is probably because for some reason the communists appear extremely centrist in comparison with what one would anticipate.[6]

As we have already noted, Budge and Klingemann (2001: 33–4) attributed this 'ideological convergence' between KKE and ND to the 1989 coalition government. As Pridham and Verney (1991) amply demonstrated, however, the 1989 coalition among ND, KKE, EAR (a KKE Euro-communist splinter) and DIANA (a short-lived ND splinter party) was necessitated by the surfacing of major financial scandals involving several ministers of the governing PASOK. The short-lived government that was formed (July to October 1989) was essentially a caretaker government with a specific mission to facilitate the criminal proceedings against PASOK ministers. Regardless of the motives that drove this rather unusual coalition, however, there is no indication that it stemmed from a more substantial convergence between the two parties in ideological terms. It is also intriguing how the 'standard' method finds such a zig zag in such a short period (1989–90). It is also confusing that in 1996 it regards Synaspismos as the most right-wing party of the country, together with a nationalist ND splinter, Political Spring (POLAN).

Thus far, the analysis would probably appear biased against the estimates based on the 'standard' method. Besides, it is perhaps too much to expect from a simple coding procedure to capture 'true' parties' positions through their manifestos. In fact, it is worth mentioning some rather unexpectedly positive aspects of some of the estimates presented in the figures. To start with, regardless of the starting point, all methods capture PASOK's right-wing shift especially after 1993. Most of them also indicate ND's shift to neoliberal policies during the 1990–3 term. Furthermore, both Ethniki Parataxi (an extreme right party) in 1977 and DIKKI (a populist PASOK splinter) in the mid-1990s are correctly placed by most methods. POLAN, with some exceptions, is also typically found near to ND, often to its right. Finally, KKE's gradual move towards the extreme left is reflected by all methods.

All that leads to a rather familiar but ambiguous picture. As is usually the case, a mixed pattern is observed for most methods, although for some methods face validity speaks more against than in favour ('domestic', 'standard'). Instead of choosing one of the models regarding our presumptions about which of them generally seems to perform best, we prefer to let the experts and the voters do so. This takes us back to the issue of convergent validity. We try to assess empirically the convergent validity of the methods by evaluating them against measures of party positions taken both by expert surveys and by voters' perceptions. Our goal here is not to evaluate CMP data estimates against other measures. Rather, we are interested in investigating how CMP estimates from different computation methods perform against each other, measured against the same benchmark.

Unfortunately, we could only gather information for 21 cases (parties × elections) of expert survey estimates about Greek parties' positions.[7] Moreover, there is no available source of voters' perceptions about parties' ideological stances during this period. Thus, we resort to a more indirect strategy, a strategy that involves the extrapolation of parties' positions by the position of their voters by using the Euro-Barometer data. The measure we used to infer each party's ideological position is the interpolated median in the L–R item of those respondents who declared their intention to vote for this party.[8]

The first row of Table 2 shows how each of the measures correlates with the findings from the expert surveys. For the moment we focus only on the interval-level correlations shown in the entries of the cells. Importantly, apart from the existing variation, there are some methods that reach rather high correlation with the results from the experts ('inductive regression' methods). The results from the second row are similarly encouraging, since they suggest that with the usual exception of the 'standard' and 'domestic', all other methods attain a relatively high relationship with voters' ideological positions. It is also interesting that, for some methods, CMP estimates correlate better with voter perceptions compared to expert surveys. These differences, however, are relatively small compared to the differences among different methods. The next rows show the correlation between the CMP results across all different methods used in this analysis. With the exception of 'domestic', which seems to correlate poorly with other methods, all other correlations appear significant and, reassuringly, positive. With the exception of a few (6 out of 15) correlations that go beyond 0.75, the level of the association is not as great as might be expected. The mean correlation between all measures is 0.553, and becomes 0.693 when the 'domestic' method is excluded. This implies that although there is enough convergence between the different methods, they do not share, on average, more than half of their variance. Although this threshold should be too high for different variables in a typical regression analysis, it is probably modest when it refers to different measures of the same concept. It is probably problematic to find that by selecting different combinations and implementing different manipulations on the same data one comes up with different outcomes. As

Table 2. Correlations among CMP estimates, expert surveys and voters' perceptions

	Voters	*Standard*	*Two-stage FA*	*Vanilla*	*Regression*	*Domestic*	*Inductive*
Experts	0.918	0.669	0.760	0.768	0.826	0.039*	0.890
		0.673	*0.861*	*0.710*	*0.706*	*0.282**	*0.826*
Voters		0.571	0.796	0.767	0.776	0.197*	0.790
		0.429	*0.731*	*0.654*	*0.674*	*0.428*	*0.668*
Standard			0.778	0.392	0.766	0.579	0.780
			0.803	*0.384*	*0.887*	*0.177**	*0.770*
Two-stage FA				0.623	0.788	0.467	0.802
				0.621	*0.898*	*0.040**	*0.928*
Vanilla					0.458	–0.258*	0.847
					0.469	*0.289**	*0.806*
Regression						0.460	0.705
						*0.093**	*0.918*
Domestic							0.115
							0.197

N = 38 for correlations between CMP estimates; 19 for expert surveys and 29 for voters' perceptions. Entries are Pearson correlations; entries in italics are polychoric (polyserial for the case of experts and voters) correlations. *p > .05.

also made clear by the figures, it matters substantially which measure is to be adopted, since this can lead to considerably different results. This is even the case if we relax the assumption of linear relationships in the data and focus on ordinal-level consistency. Attributing to parties integers from 1 (most Left) to 5 (most Right) according to their observed sequence in the L–R dimension, polychoric correlations (shown in italics) reflect again this partial inconsistency (7 out of 15 correlations greater than 0.75).

Conclusion: Which Method is the Winner?

If there is an answer to this question, then it is probably not the standard CMP method for the measurement of parties' positions in general (Budge and Klingemann, 2001), or, in the case of Greece, more specifically (Konstantinidis, 2004). Beyond that, to argue that either the 'a-theoretical' 'vanilla method' or the 'regression' method (which similarly relaxes the 'valence issue' assumption of CMP coding) performs better, becomes a rather subjective and probably not very useful enterprise. What is most important, however, is that by analysing most of the employed methods in terms of their reliability and validity, we find substantial divergence between the methods. This finding questions the robustness of results based on only one of the measures. Given that there is no particular method that clearly outperforms all others, it seems that it is pivotal for studies which employ the CMP data to subject their analysis to sensitivity testing.

Of course, this is not to say that the CMP data are no good. Quite the opposite is true. They display a remarkable wealth of information which runs through a long time-series and constitutes the most systematic attempt to measure parties' positions during such a long period. In effect, as we already mentioned, this has established 'salience' as the prima facie method of estimating policy positions of political parties using their manifestos. Nevertheless, to paraphrase King (1990: 11), knowing that one is using the most established method in the field to estimate parties' L–R positions 'is comforting but insufficient'. What is needed is to communicate precisely how different methods work and how different methods affect the results we get. As King (1990: 11) suggested, 'we should ask of every new estimator: "what did it do to the data?"'. However, this can only be done on a case-to-case basis.

In applying this 'King criterion' we have studied the reliability and validity of different L–R estimates for political parties in Greece. Our contribution should therefore be viewed as part of the recent critical literature that uses case studies for illustrative purposes in order to highlight the pitfalls and inconsistencies and suggest remedies (e.g. Franzmann and Kaiser, 2006; Pelizzo, 2003). The close inspection of the Greek party system suggested that the implausible CMP estimates might actually stem from the particular method that has been chosen. Our contribution to the general debate is thus our suggestion that researchers aiming to place parties on the L–R scale by using manifesto data should not only inspect the available alternatives, even within the CMP framework, but also supplement and corroborate the estimates with evidence from different sources such as expert and mass surveys (see also Marks, 2007). Furthermore, whether it seems simplistic or not, lacking other means for the evaluation of competing methods for the estimation of parties' positions, an additional criterion for their validation should be the extent to which they reproduce pre-existing patterns or, at least, refrain from contradicting well-established intuitions. The consideration of different methods that make use of CMP data, along with the examination of the estimates' convergent (through triangulation) and face (through case study) validity, is probably the best receipt for finding one's way within the puzzles of party positioning.

Notes

We thank Mark Franklin, Peter Mair, Joost van Spanje and Till Weber, as well as two anonymous reviewers of *Party Politics*, for helpful comments on previous versions of this article. Any remaining errors or omissions remain our responsibility.

1 A fourth approach, the analysis of roll-call data, is becoming increasingly popular. The main problem with it, however, is that in most legislatures outside the United States not all roll-call votes are recorded. Because of the different rules and conditions under which roll-call votes are triggered or requested, cross-country analyses are ridden with problems of selection bias (see Carrubba et al., 2006).

2 To the best of our knowledge, this list covers almost all attempts that have been made to construct an encompassing L–R dimension through the CMP data. Three methods have been excluded from the list; namely, Kim and Fording's (1998) method of ratio differences, van der Brug's (2001) analysis of party dynamics and Warwick's (2005) two-dimensional method. The first was excluded because it correlates very highly with the original 'standard' method (r = 0.98) and yields almost identical results. In regard to the second, although we agree with van der Brug's (2001) logic about the problems stemming from analysing the CMP data with correlation measures, we could not examine his alternative method (which is based on multidimensional scaling) because in this case L–R is not constructed by parties' positions, but rather is extrapolated by voters' perceptions about the parties as they are provided by election studies. The third was excluded because it does not provide an encompassing L–R measure but two different dimensions (L–R economic and postmaterialism/liberal values), and therefore its estimates cannot be compared to the estimates of the other methods.

3 This was tested both with the 1981 Euro-Barometer study and the 2002 European Social Survey. In both instances, both sets of issues appear to contribute significantly to our understanding about how people locate themselves in the L–R continuum (the results can be found in the online Appendix at http://www.keele.ac.uk/kepru). The consistency between voters' perceptions about L–R and the logic guiding the construction of L–R scales with CMP data is important for our empirical assessment of the convergent validity, as will become apparent below.

4 These results can be found in the online Appendix at http://www.keele.ac.uk/kepru.

5 Heise's model is based on the assumption of first-order autocorrelation. Reassuringly, regressions of each L–R variable against its three lags reveal a descending pattern with the first lag typically significant. More importantly, sources of change in parties' ideological positions do not seem to correlate between different time periods, another crucial assumption of the model. For instance, the residuals of the 1985 scores, as predicted from 1974 scores, are not significant predictors of the residuals of the 2000 scores, as predicted by the 1993 scores (the results are available in the online Appendix).

6 A different argument that would also justify the placement of PASOK to the left of KKE has to do with the item composition of the constructed scales. Given that all L–R scales also include non-economic issues, with PASOK being presumably more liberal, one would expect it to emphasize more than KKE leftist non-economic issues. Although this argument contradicts the depiction of PASOK as a populist party, we test it by taking only economic issues into account (in a summated rating scale along the lines of the 'standard' method). Still we find PASOK to the left of KKE. Alternatively, we take the issue of democracy, a typical left non-economic issue. In all this period and with the single exceptions of 1977 and 2000, KKE refers more to this issue than PASOK does (all the results are available in the online Appendix). Thus, at least for the 'standard' method, this argument cannot justify the observed inconsistency in the findings.

7 We used the estimates of Laver and Hunt (1992), Lubbers (2001) and the 1999 Chapel Hill Expert Survey (Steenbergen and Marks, 2007) and recoded them with the method described in Carter (2005: 143). Finally, for the 2000 election we used Lubbers' (2001) survey (in Carter, 2005).

8 In order to validate this method, we used two additional data sets, the 1999 and 2004 European Election Studies (EES). Both surveys ask respondents to locate both themselves and their parties in a 1–10 L–R dimension. In the 2004 survey, the correlation between the median position of voters of a given party and the median position of the positions at which voters attached their party is 0.97. Furthermore, as this could not be identified by a simple correlation but still question the validity of the extrapolation strategy, no significant difference was found in the mean positions between parties and voters as a whole (5.05 for voters, 5.52 for parties, $p > 0.1$). The results are similar for the 1999 EES (r = 0.98, mean values for voters and parties 4.70 and 4.91, respectively).

References

Bartolini, Stefano and Peter Mair (1990) *Identity, Competition, and Electoral Availability: the Stabilisation of European Electorates, 1885–1985*. Cambridge: Cambridge University Press.

Benoit, Kenneth and Michael Laver (2007a) 'Estimating Party Policy Positions: Comparing Expert Surveys and Hand-Coded Content Analysis', *Electoral Studies* 26: 90–107.

Benoit, Kenneth and Michael Laver (2007b) 'Benchmarks for Text Analysis: a Response to Budge and Pennings', *Electoral Studies* 26: 130–5.

Bosco, Anna (2001) 'Four Actors in Search of a Role: the Southern European Communist Parties', in P. Nikiforos Diamandouros and Richard Gunther (eds)

Parties, Politics, and Democracy in the New Southern Europe, pp. 329–87. Baltimore, MD: Johns Hopkins University Press.

Budge, Ian (1987) 'The Internal Analysis of Election Programmes', in Ian Budge, David Robertson and Derek J. Hearl (eds) *Ideology, Strategy and Party Change*, pp. 15–38. Cambridge: Cambridge University Press.

Budge, Ian (1994) 'A New Spatial Theory of Party Competition: Uncertainty, Ideology and Policy Equilibria Viewed Comparatively and Temporally', *British Journal of Political Science* 24: 443–67.

Budge, Ian (2000) 'Expert Opinions of Party Policy Positions: Uses and Limitations in Political Research', *European Journal of Political Research* 37: 103–13.

Budge, Ian (2001a) 'Validating Party Policy Placements', *British Journal of Political Science* 31: 210–23.

Budge, Ian (2001b) 'Validating the Manifesto Research Group Approach: Theoretical Assumptions and Empirical Confirmations', in Michael Laver (ed.) *Estimating the Policy Position of Political Actors*, pp. 50–65. London: Routledge.

Budge, Ian and Dennis Farlie (1983) 'Party Competition – Selective Emphasis or Direct Confrontation? An Alternative View with Data', in Hans Daalder and Peter Mair (eds) *West European Party Systems: Continuity and Change*, pp. 267–305. London: Sage.

Budge, Ian and Hans-Dieter Klingemann (2001) 'Finally! Comparative Over-time Mapping of Policy Movement', in Ian Budge, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara and Eric Tanenbaum (eds) *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments, 1945–1998*, pp. 19–50. Oxford: Oxford University Press.

Budge, Ian, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara and Eric Tanenbaum (eds) (2001) *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments 1945–1998*. Oxford: Oxford University Press.

Budge, Ian and Paul Pennings (2007) 'Do They Work? Validating Computerised Word Frequency Estimates Against Policy Series', *Electoral Studies* 26: 121–9.

Carmines, Edward G. and James A. Woods (2005) 'Validity Assessment', in Kimberly Kempf-Leonard (ed.) *Encyclopedia of Social Measurement*, vol. 3, pp. 933–7. Amsterdam: Elsevier.

Carrubba, Clifford J., Matthew Gabel, Lacey Murrah, Ryan Clough, Elizabeth Montgomery and Rebecca Schambach (2006) 'Off the Record: Unrecorded Legislative Votes, Selection Bias and Roll-Call Vote Analysis', *British Journal of Political Science* 36: 691–704.

Carter, Elisabeth (2005) *The Extreme Right in Western Europe: Success or Failure?* Manchester: Manchester University Press.

Clogg, Richard (1987) *Parties and Elections in Greece: the Search for Legitimacy*. Durham, NC: Duke University Press.

Fitzmaurice, Garrett M., Nan M. Laird and James H. Ware (2004) *Applied Longitudinal Analysis*. Hoboken, NJ: Wiley and Sons.

Franzmann, Simon and André Kaiser (2006) 'Locating Political Parties in Policy Space: a Reanalysis of Party Manifesto Data', *Party Politics* 12: 163–88.

Gabel, Matthew J. and John D. Huber (2000) 'Putting Parties in Their Place: Inferring Party Left–Right Ideological Positions from Party Manifesto Data', *American Journal of Political Science* 44: 94–103.

Green, Donald P. and Bradley Palmquist (1990) 'Of Artifacts and Partisan Instability', *American Journal of Political Science* 34: 872–902.

Hanley, David (2008) *Beyond the Nation State: Parties in the Era of European Integration*. Houndmills: Palgrave Macmillan.

Heise, David R. (1969) 'Separating Reliability and Stability in Test–Retest Correlation', *American Sociological Review* 34: 93–101.

Huber, John D. and Ronald Inglehart (1995) 'Expert Interpretations of Party Space and Party Locations in 42 Societies', *Party Politics* 1: 73–111.

Hug, Simon and Tobias Schulz (2007) 'Left–Right Positions of Political Parties in Switzerland', *Party Politics* 13: 305–30.

Jacoby, William G. (1991) *Data Theory and Dimensional Analysis, Quantitative Applications in the Social Sciences Series*, no. 78. Thousand Oaks, CA: Sage.

Kim, Heemin and Richard C. Fording (1998) 'Voter Ideology in Western Democracies, 1946–1989', *European Journal of Political Research* 33: 73–97.

King, Gary (1990) 'On Political Methodology', *Political Analysis* 2: 1–29.

Kleinnijenhuis, Jan and Paul Pennings (2001) 'Measurement of Party Positions on the Basis of Party Programmes, Media Coverage and Voter Perceptions', in Michael Laver (ed.) *Estimating the Policy Position of Political Actors*, pp. 162–82. London: Routledge.

Klingemann, Hans-Dieter (1995) 'Party Positions and Voter Orientations', in Hans-Dieter Klingemann and Dieter Fuch (eds) *Citizens and the State*, pp. 183–206. Oxford: Oxford University Press.

Klingemann, Hans-Dieter, Andrea Volkens, Judith Bara, Ian Budge and Michael D. McDonald (2007) *Mapping Policy Preferences II: Estimates for Parties, Electors and Governments in Central and Eastern Europe, European Union and OECD, 1990–2003*. New York: Oxford University Press.

Konstantinidis, Ioannis (2004) 'Metavoles sto periechomeno ton programmaton ton Hellinikon kommaton, 1974–2000: He axia tes prosarmostikotetas' [Policy Changes as Pictured in Greek Parties' Electoral Manifestos: the Value of Adaptability], *Greek Political Science Review* 12 (November): 105–39.

Laver, Michael (2001a) 'Why Should We Estimate the Policy Positions of Political Actors?' in Michael Laver (ed.) *Estimating the Policy Position of Political Actors*, pp. 3–9. London: Routledge.

Laver, Michael (2001b) 'Position and Salience in the Politics of Political Parties', in Michael Laver (ed.) *Estimating the Policy Position of Political Actors*, pp. 66–75. London: Routledge.

Laver, Michael, Kenneth Benoit and John Garry (2003) 'Estimating the Policy Positions of Political Actors Using Words as Data', *American Political Science Review* 97: 311–31.

Laver, Michael and Ian Budge (1992) 'Measuring Policy Distances and Modelling Coalition Formation', in Michael Laver and Ian Budge (eds) *Party Policy and Government Coalitions*, pp. 15–40. Basingstoke: Macmillan.

Laver, Michael and Ben W. Hunt (1992) *Policy and Party Competition*. New York. Routledge.

Lubbers, Marcel (2001) *Exclusionistic Electorates: Extreme Right-Wing Voting in Western Europe*. Nijmegen: Center for Social Science Theory and Methodology.

Mair, Peter (2001) 'Searching for the Positions of Political Actors: a Review of Approaches and Critical Evaluation of Expert Surveys', in Michael Laver (ed.) *Estimating the Policy Position of Political Actors*, pp. 10–30. London: Routledge.

Marks, Gary (2007) 'Introduction: Triangulation and the Square-Root Law', *Electoral Studies* 26: 1–10.

McDonald, Michael P. (2005) 'Validity, Data Sources', in Kimberly Kempf-Leonard (ed.) *Encyclopedia of Social Measurement*, vol. 3, pp. 939–48. Amsterdam: Elsevier.

McDonald, Michael D., Silvia M. Mendes and Myunghee Kim (2007) 'Cross-Temporal and Cross-National Comparisons of Party Left–Right Positions', *Electoral Studies* 26: 62–75.

Moschonas, Gerassimos (2002) *In the Name of Social Democracy: the Great Transformation, 1945 to the Present*. London: Verso.

Pelizzo, Ricardo (2003) 'Party Positions or Party Direction? An Analysis of Party Manifesto Data', *West European Politics* 26 (April): 67–89.

Pridham, Geoffrey and Susannah Verney (1991) 'The Coalitions of 1989–90 in Greece: Inter-Party Relations and Democratic Consolidation', *West European Politics* 14 (October): 42–69.

Ray, Leonard (1999) 'Measuring Party Orientations towards European Integration: Results from an Expert Survey', *European Journal of Political Research* 36: 283–306.

Robertson, David (1987) 'Britain, Australia, New Zealand and the United States, 1946–1981: an Initial Comparative Analysis', in Ian Budge, David Robertson and Derek J. Hearl (eds) *Ideology, Strategy and Party Change*, pp. 39–73. Cambridge: Cambridge University Press.

Spourdalakis, Michalis (1988) *The Rise of the Greek Socialist Party*. London: Routledge.

Steenbergen, Marco and Gary Marks (2007) 'Evaluating Expert Surveys', *European Journal of Political Research* 46: 347–66.

van der Brug, Wouter (2001) 'Analysing Party Dynamics by Taking Partially Overlapping Snapshots', in Michael Laver (ed.) *Estimating the Policy Positions of Political Actors*, pp. 115–32. London: Routledge.

Volkens, Andrea (2007) 'Strengths and Weaknesses of Approaches to Measuring Policy Positions of Parties', *Electoral Studies* 26: 108–20.

Warwick, Paul V. (2005) 'When Far Apart Becomes Too Far Apart: Evidence for a Threshold Effect in Coalition Formation', *British Journal of Political Science* 35: 383–401.

Wiley, David E. and James A. Wiley (1970) 'The Estimation of Measurement Error in Panel Data', *American Sociological Review* 35: 112–7.

ELIAS DINAS is a PhD candidate at the European University Institute, Florence. His main field of research is electoral behaviour, and more particularly the dynamic process of partisan formation during early adulthood. He is also interested in party politics, political methodology and Greek politics. His work has been published or accepted for publication in *Electoral Studies*, *West European Politics*, *Comparative European Politics* and *South European Society and Politics*.
ADDRESS: Department of Political and Social Science, European University Institute, Via dei Roccettini 9, I-50014, San Domenico di Fiesole, Florence, Italy. [e-mail: ilias.ntinas@eui.eu]

KOSTAS GEMENIS is a PhD candidate at Keele University. His main field of research is comparative European politics. His dissertation looks at the impact of

European integration on parties' policy positions on environmental issues. He is also interested in conceptualization and measurement in comparative politics and Greek party politics. His work has been published or accepted for publication in *Comparative European Politics*, *Mediterranean Politics* and *Environmental Politics*.
ADDRESS: Research Institute for Law, Politics and Justice, Keele University, Keele, Staffs, ST5 5BG, UK. [e-mail: k.gemenis@ilpj.keele.ac.uk]

Paper submitted 17 January 2008; accepted for publication 4 August 2008.