Proxy documents as a source of measurement error in the Comparative Manifestos Project

Kostas Gemenis

Department of Public Administration, University of Twente, P.O. Box 217, 7500 AE, Enschede, The Netherlands

Abstract

This paper considers the issue of document type diversity in the Comparative Manifestos Project (CMP). For many years the CMP has been collecting and coding a variety of documents, such as speeches, pamphlets, newspaper articles and leaflets, as manifesto proxies. By using previously unexplored archival material to perform controlled comparisons between different types of documents, this paper argues that the coding of such documents introduced considerable measurement error to party position estimates. Statistical analyses indicate that this measurement error is systematic rather than random as it is often manifested as centrist bias in parties’ left-right position estimates. Consequently, the paper argues that random error correction methods cannot always correct for error attributed to the coding of proxy documents. The paper concludes with some recommendations for third-party users of the CMP data and documents and a plea to the CMP research team.

Keywords: Parties, Policy positions, Comparative Manifestos Project, Document selection, Measurement quality

1. Introduction

Political text is an encompassing category which includes the text generated by individual politicians as well as parties. The former includes interviews, parliamentary debates and election campaign speeches whereas the latter consists of documents that represent the party as a whole. These include press releases and pamphlets but most importantly the election campaign documents that are invariably known as programmes, platforms or manifestos. It is widely considered that manifestos constitute the most objective source of information for parties’ policy preferences, proposals, positions and, potentially; parties’ ideology. Manifestos are, after all, documents produced by parties meant to present their positions to the electorate. The fact that they are often adopted or sanctioned at party congresses, allows us to produce estimates of the ideology and policy of the party as a whole and not just of the leadership or a particular faction (Budge, 2001, 211). Therefore, the policy estimates deriving from the analysis of party manifestos can be considered as the most objective source of information (Marks et al., 2007, 27) as manifestos ‘chart the party’s position at each election on the basis of its own authoritative policy pronouncement’ (Budge, 2002).

It was exactly these qualities that led researchers to estimate parties’ policy positions through the content analysis of party manifestos. The early attempt by Borg (1966) who content analyzed the basic programmes of Finnish parties and Robertson (1976) who did likewise for British party manifestos, was later systematized by Ian Budge and his colleagues who established the Manifesto Research Group (MRG) in 1979. The MRG developed into an ambitious project of collecting and hand-coding party manifestos in 19 countries. Ten years later, the MRG was renamed to Comparative Manifestos Project (CMP) under the direction of Hans-Dieter Klingemann and was expanded to include all the OECD countries as well as 24 countries in Central and Eastern Europe. Most recently, the CMP received a major 12-year grant from the German Research Foundation in order to expand its document collection and coding to South American and South Asian countries.1

For the past thirty years, specifically-trained coders have been parsing the text of party manifestos into ‘quasi-sentences’, assigning each of these quasi-sentences into one of the CMP’s 56 issue categories (see Volkens, 2002a). The results of this coding are presented in terms of percentage frequencies, which intend to measure each party’s ‘relative emphasis’ on each of these 56 issues. In addition, the CMP uses a scaling technique to estimate parties’ positions on the left-right (L-R) dimension (see Laver and Budge, 1992) which Budge and Klingemann (2001, 19) called the project’s ‘crowning achievement’. The rich time-series data produced by the CMP have been made freely available to researchers, as part of the Mapping Policy Preferences books (Budge et al., 2001, Klingemann et al., 2006), and as a consequence, the CMP has been established as the prima facie data source for parties’ positions on the L-R dimension as well as on many other policy dimensions.2 Moreover, the CMP data have often been employed as a ‘golden stan-

1 In 2009 the CMP was renamed to Manifesto Research on Political Representation (MARPOR) but since the project is mostly known by its previous title, this paper will use the CMP acronym throughout. For a history of the project see Budge (2002) and Volkens (2002b). The project’s website can be accessed at: http://manifestoproject.wzb.eu/

2 This is evident from the use of data in more than 100 research monographs, peer-reviewed publications and PhD theses (~1,000 citations in Google Scholar) and the 2003 Best Data Set Award by the American Political Science Association.
Proksch and Slapin (2009, 329–330) found that the German party documents collected by the CMP included party congress speeches, many of the so-called Wahlaufruf or short election proclamations and, in one case, an action programme published two years after the election for which it was used for. Again, the authors argued that these documents cannot be considered to be equivalent and comparable to election manifestos since they differ in terms of size, style and authoritativeness.

Denmark and Germany are not the only countries that show considerable diversity in the type of coded documents. For instance, in Belgium we find regional manifestos (e.g. Parti Social Chrétien 1977), in Finland election flyers (e.g. Kansallinen Kokoomus 1979), in France party leader speeches (e.g. Parti Communiste Français 1988), in Ireland journalists’ reports at the national press (e.g. Fianna Fail Nov. 1982), in Italy programme summaries published in the partisan press (e.g. Movimento Sociale Italiano-Destra Nazionale 1992), in Romania hand-written documents (e.g. Convenţia Democrată Română 1992) and in Switzerland summaries of election manifestos (e.g. Sozialdemokratische Partei der Schweiz 1995). Moreover, for Japan the CMP coding was based exclusively on ‘rapid-fire preelection interviews by a national daily newspaper’ which are extremely short and constrained in the range of issues (Proksch et al., 2010, 115), whereas for Israel the CMP collection is almost entirely based on articles, advertisements and interviews in national newspapers.

There is ample evidence that the CMP investigators who collected and coded these documents have expressed their scepticism regarding whether all these documents can be used as proxies to the strictly defined manifestos. This is because the inclusion of material such as reports in newspapers, speeches, pamphlets and leaflets was borne out of necessity, rather than a well-thought strategy based on evidence showing that the different types of documents can be considered to be equivalent to manifestos for the purpose of estimating parties’ policy positions.

Unfortunately, it is difficult to determine the degree of document type diversity by looking at the CMP datasets and appendices with document titles. Both datasets contain a variable (progtype) which indicates the type of coded document by differentiating among ‘manifestos’ (progtype=1), ‘estimates’, that is linear interpolations or copy/pasted data from one year to another (progtype=3); ‘joint programmes’ that is, electoral coalition manifestos which are used for one or more parties after being weighted according to party strength (progtype=2,4,5,7); ‘general programmes’ which are not tied to a particular election (progtype=6) and ‘other’ (progtype=8). The CMP claims that 86.71% of the estimates in the datasets are based on the coding of manifestos but an independent classification of documents in four countries revealed considerable differences.

For illustrative purposes, Figure 1 presents some of the evidence based on an alternative classification of documents based

---

2. The problem of proxy documents in the CMP

As McDonald (2006, 90) put it, ‘few party scholars doubt that noise comes into the CMP scores’ from a variety of sources. Since noise implies random rather than systematic measurement error, most of the critical literature in the context of the CMP has focussed on the former. For instance, researchers have considered the stochastic process of document generation (Benoit et al., 2009b), the parsing of text into quasi-sentence (Däubler et al., forthcoming), the coding of quasi-sentences into issue categories (Mikhaylov et al., 2012) and the scaling of frequency categories into the L-R (Lowe et al., 2011) as sources of random measurement error in the CMP estimates. To the contrary, research so far has not considered the diversity of type in the coded documents as a source of measurement error. In fact, most third-party users of the CMP data are not aware that a considerable portion of the CMP datasets are based on the coding of documents other than manifestos.

In a rather illustrative article about Danish parties, Hansen (2008, 207–210) observed that the CMP has coded, among others, an advertisement/insert from a populist newspaper bundled with a more general manifesto, an issue of a party magazine which included a telegram from Ronald Reagan, bundles of speeches by the party leader, a document composed by a rival party think tank, a manifesto from the previous year’s local election, draft manifestos, advertisements in local newspapers and so on. Hansen (2008) convincingly argued that the Danish documents analyzed by the CMP are very different in terms of breadth and scope and that many of these documents cannot be considered as being equivalent to manifestos. Similarly, Proksch and Slapin (2009, 329–330) found that the German

3Those interested can find such evidence in the contributions in Budge et al. (1987) as well as in the early correspondence between Ian Budge and his collaborators. The correspondence can be found in an extensive collection of materials relating to the development of the MRG/CMP which were deposited by Ian Budge at GESIS-Leibniz Institute for the Social Sciences, Cologne and have since been moved to Social Science Centre, Berlin.
on the inspection of the actual documents and consultations with country experts who have conducted extensive manifesto research in national and party archives. The alternative classification aimed to separate party manifestos, from joint platforms and ‘other’ types of documents (speeches, flyers, etc). Whenever there was uncertainty as to what the document was, the CMP classification was retained. I believe that few researchers would question this alternative classification given that the CMP tendency to label almost every document in the dataset as a manifesto. Most of the documents I classified as ‘other’ are clearly speeches, interviews or flyers and/or I found compelling (sometimes archival or interview) evidence that the party in question did not intend the particular document to serve as their manifesto for the election in question.

4 The findings in Figure 1 confirm the observations made for Denmark and Germany arguing that the CMP is underestimating the degree of diversity in its coded documents. To be fair, the countries mentioned so far (Austria, Belgium, Denmark, Finland, Germany, Greece, Ireland, Israel, Italy, Japan, Romania and Switzerland) show a considerable variation in the coded documents. Conversely, the collection for the British parties consists entirely of manifestos and for a few other countries, such as the Netherlands, the CMP has coded very few documents which are not manifestos. In such countries, however, document quality was ensured because the manifestos came from well-known published collections.

Unlike previous research in the sources of error in the CMP estimates, this paper considers the coding of proxy documents as a source of systematic error. Given that systematic error ‘does not yield easily to analysis and estimation’ (Zeller and Carmines, 1980, 136) because, most often than not, we are agnostic of the true score variance, the following sections adopt three different approaches in identifying document type diversity as a source of measurement error and the extent to which this error can be considered to be systematic rather than random.

3. Controlled comparisons of documents

Estimating measurement error coming solely from variation in the type of coded documents requires keeping error coming from other sources constant. Unfortunately, such controlled comparisons are quite hard to achieve without the use of suitable document material. The inspection of the documents collected by the CMP, however, offers such a unique opportunity. It seems that the Greek CMP coder collected and coded a number of party leader speeches for the social democratic PASOK (Panhellenic Socialist Movement) in addition to its manifestos. Whether this was done purely for the sake of convenience (since speeches are much shorter compared to the notoriously long PASOK manifestos) or because the coder was unaware that the party published manifestos for every election (see Clogg, 1987, Kalogeropoulou, 1989, Spourdalakis, 1988) is difficult to tell. Nevertheless, the copies of the speeches with issue categories assigned to quasi-sentences can be found in the archives of the Social Science Centre, Berlin.

Table 1 presents some descriptive statistics from the comparison of PASOK manifestos to the speeches delivered by the party leader during the election campaigns. It is clear from the
figures, that the two types of documents differ considerably in terms of size or policy coverage. As theorized by Laver and Benoit (2002, 65), manifestos are much larger than party leader speeches (quasi-sentences column) and cover a much broader range of issues (issues column). Moreover, as is evident from the high frequency of uncoded quasi-sentences, speeches feature a different ‘language register’ (Laver and Benoit, 2002, 65) which was not captured effectively by the CMP coding scheme. Speeches are clearly far from being ‘authoritative statements’ that can be used as ‘manifesto surrogates’. The interesting counterfactual question, however, is whether the use of speeches instead of manifestos could have produced a different, and perhaps implausible, view of PASOK’s L-R position.

Table 1: Comparing manifestos to party leader speeches: PASOK 1977–1989.

<table>
<thead>
<tr>
<th>Document</th>
<th>Pages</th>
<th>Quasi-sentences</th>
<th>Percent uncoded</th>
<th>Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manifesto 1977</td>
<td>151</td>
<td>2704</td>
<td>0.3</td>
<td>47</td>
</tr>
<tr>
<td>Speech 1977</td>
<td>20</td>
<td>417</td>
<td>6.95</td>
<td>36</td>
</tr>
<tr>
<td>Manifesto 1981</td>
<td>125</td>
<td>2991</td>
<td>0.67</td>
<td>48</td>
</tr>
<tr>
<td>Speech 1981</td>
<td>10</td>
<td>346</td>
<td>10.98</td>
<td>37</td>
</tr>
<tr>
<td>Manifesto 1985</td>
<td>96</td>
<td>3244</td>
<td>0.46</td>
<td>48</td>
</tr>
<tr>
<td>Speech 1985</td>
<td>13</td>
<td>428</td>
<td>19.16</td>
<td>34</td>
</tr>
<tr>
<td>Manifesto June 1989</td>
<td>80</td>
<td>2215</td>
<td>0.36</td>
<td>47</td>
</tr>
<tr>
<td>Speech June 1989</td>
<td>12</td>
<td>198</td>
<td>22.22</td>
<td>26</td>
</tr>
<tr>
<td>Manifesto November 1989</td>
<td>80</td>
<td>2215</td>
<td>0.36</td>
<td>47</td>
</tr>
<tr>
<td>Speech November 1989</td>
<td>32</td>
<td>417</td>
<td>13.67</td>
<td>36</td>
</tr>
</tbody>
</table>

Source: For manifestos and November 1989 speech, CD-ROM in Budge et al. (2001); for other speeches, own calculations based on the coded documents found in the CMP archive.

The answer to this question can be found in Figure 2 which shows the L-R trajectory of PASOK from November 1974 to the November 1989 election by using the ‘standard’ CMP scaling technique. The dashed line connects the L-R positions as estimated by using party leader speeches and the solid line the L-R positions as estimated by using election manifestos.5 Since the handwriting in the coding of manifestos and speeches is identical, we can be confident that the coding was performed by the same person. In this sense, the comparison controls for measurement error stemming from inter-coder unreliability (see Mikhaylov et al., 2012). Moreover, by using the bootstrapped confidence intervals proposed by Benoit et al. (2009b), the comparison controls for differences reflecting the stochastic process of document generation.6

Figure 2 shows that, by content analyzing different types of documents, we get results which, in most of the cases, differ considerably even when the confidence intervals are taken into account. The PASOK ideological trajectory as given by the coding of manifestos appears to be quite stable, with swings between 1974–1977 and 1985–1989. These swings can be easily explained. Is is well documented in the literature that PASOK almost doubled its share of vote (from 13.58 to 25.34%) between 1974 and 1977 by moderating its extreme leftist discourse (see Spourdalakis, 1988). Moreover, the party became even more moderate after 1985 when it was forced to reconsider its socialist economic policies and introduced a neo-liberal ‘stabilization programme’ (see Tsakalotos, 1998). The party image we get from the coded speeches is quite different. PASOK appears to have moved towards the centre after 1977, whereas for the duration of the 1980s, the speech-based estimates place PASOK consistently, and by almost 20 points on the L-R scale, closer to the centre compared to manifesto-based estimates. To be sure, it can be argued that PASOK was not as radical in the 1980s as its manifestos portray it. However, the speeches place PASOK closer to the centre not because the speeches are centrist in terms of policy emphasis, but they do so because of the high percentage of quasi-sentences that remained uncoded.

To understand why uncoded quasi-sentences move the party to the centre we should consider the way the ‘standard’ CMP L-R scale is constructed. In a review essay, Laver (2001) was one of the first to point out several serious shortcomings of the ‘standard’ CMP L-R scaling technique. According to this technique, the sum of the emphasis on a fixed set of ‘right’ issues is subtracted from another fixed set of ‘left’ issues and divided by the total number of quasi-sentences \( \frac{R-L}{N} \). Laver (2001) cautioned third-party users of the CMP data to be aware of the statistical properties of the scale. According to the ‘standard’ scale, party movement on the L-R scale is not only caused by changes in emphasis on the coding categories included in the Left and Right components but also by changes in the emphasis to all the other categories excluded from the scale (Laver, 2001, 22), including the category of ‘uncoded’ quasi-sentences. In other words, the discourse that cannot be captured by the CMP coding scheme moves parties to the centre. Because between one fifth and one quarter of the PASOK speech content was deemed uncodable, the coding of speeches gives the false

---

5 For 1974 there is a single point because an interview with the party secretary indicated that the particular speech was meant to be used as the party manifesto. In addition, primary sources indicate that PASOK did not issue a manifesto for the November 1989 election but ran the campaign on the June 1989 manifesto hence the identical figures for November 1989 (Gemenis, 2010, 168, 182).

6 I used an R script file kindly provided by Slava Mikhaylov to calculate the confidence intervals for the new documents.
impression of movement towards the centre. The fact that the manifesto and speech estimates move almost parallel to each other provides strong evidence that the measurement error introduced by coding speeches in place of manifestos is systematic rather than random. Researchers have thus far attempted to deal with the problem of uncoded content by using alternative scaling methods (Franzmann and Kaiser, 2006, Kim and Fording, 1998, Lowe et al., 2011). The problem of coding proxy documents, however, is more complex and cannot be effectively solved by scaling alone. The following section considers how the differences in political discourse, as presented in speeches and manifestos, and the specificities of selecting pamphlets as manifesto proxies also introduce systematic error with a centrist direction.

4. Qualitative comparisons of documents

The CMP coding scheme has been devised with British party manifestos (and to a lesser extent American party platforms) in mind. As such, it is very well suited to capture the ‘list-of-pledges’ style of British manifestos, but rather unsuitable for (party leader) speeches which include more complex argumentation, rhetorical devices, emotional appeals (see Beard, 2000, Brader, 2006) or otherwise non-pledge content. The 1996 CMP estimates for the conservative New Democracy (ND), the Communist Party of Greece (KKE) and the Coalition of the Left and Progress (SYN) are illustrative cases in point. For both SYN and KKE the CMP coded party leader speeches, even though both parties published more detailed ‘list-of-pledges’ style electoral manifestos. For ND, the CMP coded a collection of seven policy pamphlets which were published at various dates before the election. The result of the coding was an extreme right position for the ‘greenish’ radical left SYN and a centrist position for the orthodox Stalinist KKE and conservative ND. Both Benoit and Laver (2007, 99) and Dinas and Gemenis (2010) recognized that, on the basis of expert surveys and knowledge of Greek politics, these estimates do not make sense. A closer analysis, however, indicates that these implausible positions reflect systematic error introduced by the choice made by the CMP to code the aforementioned documents in place of the manifestos.

The speech by the SYN leader, Nikos Konstantopoulos used extensively the 1996 election slogan ‘we are not afraid to look into your eyes’ (i.e. we are honest). As SYN failed to pass the newly instituted 3% threshold for parliamentary representation in 1993, the party’s advertisers thought of using this slogan as an emotional appeal. In addition, Konstantopoulos made extensive references to the disproportional of the electoral system. The CMP however, coded most of the numerous emotional appeals as ‘305: Political Authority: Positive’ and the references to the electoral system as ‘606: social harmony’, both of which are right-wing issues in the standard L-R scale. Consequently, according to the CMP, the widely acknowledged as radical left SYN, is placed on the extreme right of the political spectrum because its leader said that his party is honest and that it opposes the majoritarian electoral system. Had the CMP coded the SYN election manifesto instead, we would have been presented with a more valid position of SYN in the policy space. This is because the manifesto featured a different language, less focussed on slogans and emotional appeals, and more focussed on presenting the positions of the party in six policy areas (foreign policy, economic and social policy, development, social justice, environmental protection and political institutions).

The situation with KKE is fairly similar. The speech by the KKE secretary general Aleka Papariga featured many statements attacking the non proportional electoral system and the two-party system, which were were coded as ‘606’ and ‘305’ (right-wing issue categories). Consider the following statement which the CMP considers to consist of eight right-wing quasi-sentences:

‘Shall we accept fatefuly that this country is governed, sometimes by a PASOK majority and sometimes by a ND majority, with the help of an unjust electoral law? [...] It is not appropriate for our ideology and for our politics to sit by and wait for the big changes to come suddenly on a day or overnight.’ [my translation]

It is rather evident that there is nothing right-wing in the statement above so the problem with the 1996 KKE estimate (which appears to be far more right-wing than expected from an orthodox Stalinist party) is due to the coding of a speeches which does not tell us much about KKE’s policy positions. Just like the speech by Konstantopoulos, the speech by Papariga was a mix of emotional appeals, slogans, attacks to the other parties and the electoral system, and only occasionally presented the policies of the party. Conversely, the election manifesto, which was prepared by the party central committee, presented a far more valid (if in a very populist language) picture of KKE’s policy positions.

The situation is different regarding ND since the party does not appear to have issued a manifesto for the 1996 election. This is hardly surprising, however, because ND rarely issued election manifestos before 2000. Instead, the party showed a preference for long-term programmes which were published after party congresses in order to signal major changes in ideology (think of the Bad Godesberg programme) or, since the late 1980s, policy pamphlets on specific issues (Alexakis, 1993). Much like the Scandinavian and German ‘basic programmes’, policy pamphlets are not usually linked to specific elections but feature a detailed treatment of a particular issue. Apart from the apparent differences in their role and the way they are drafted when compared to electoral manifestos, the use of policy pamphlets as manifesto proxies poses a selection problem: which combination of pamphlets can approximate the election manifesto? Consider a party which, for whatever reasons, commissioned the writing of five green/white papers on five different issues. In the lack of an election manifesto, should we assume

---

This is evident by the press conferences held by KKE and SYN to launch the manifestos (eklogikí diákrexis) as reported in the press (see Rizospastis 27 August 1996 and Eleftherotypia 28 August 1996 respectively).
that it would be only these five issues that make up of party’s ideology? Should we assume that these five documents taken together will give a valid placement of the party on the L-R dimension?

To estimate the 1996 ND position, the CMP coded together seven such pamphlets on the subjects of economy, public administration, the environment, education, sports, and culture. I argue that it is a brave assumption to consider these seven pamphlets as a proxy for the 1996 ND election manifesto. First, the pamphlets are simple typescript documents that look nothing like the pamphlets meant for public release. More than anything, they look like internal working papers. Second, there is no evidence that ND wanted to emphasize education, sports and culture as much as the economy for the 1996 election. In fact, apart from the economy, it was foreign policy, defence, law and order and farmers’ issues which featured predominantly in the 1996 ND election campaign (see Featherstone and Kazarinas, 1997, Papathanassopoulos, 2000). As a consequence, the coding of the pamphlets on the environment, education, sports and culture, which consist primarily of issues that do not feature on the coding of the pamphlets on the environment, sports and culture. I argue that it is a brave assumption to consider these seven pamphlets as a proxy for the 1996 ND election manifesto. First, the pamphlets are simple typescript documents that look nothing like the pamphlets meant for public release. More than anything, they look like internal working papers. Second, there is no evidence that ND wanted to emphasize education, sports and culture as much as the economy for the 1996 election. In fact, apart from the economy, it was foreign policy, defence, law and order and farmers’ issues which featured predominantly in the 1996 ND election campaign (see Featherstone and Kazarinas, 1997, Papathanassopoulos, 2000). As a consequence, the coding of the pamphlets on the environment, education, sports and culture, which consist primarily of issues that do not feature on the CMP L-R scale, effectively but spuriously moved ND towards the centre and close to the, equally spurious centrist, position of KKE.

To demonstrate the problems in positioning parties based on proxy documents, Figure 3 compares the standard CMP L-R placement (‘rile’) to the Lowe et al. (2011) logged version of the same scale (‘logrile’) for the 1996 election in Greece. In both scales, the L-R estimates come with bootstrapped 95% confidence intervals which account for the uncertainty associated with the stochastic process of document generation (Benoit et al., 2009b). As most of the standard estimates in Figure 3 have already been dismissed as implausible according to the literature (Dinas and Gemenis, 2010) and the discussion above, we can focus on whether the logged version of the scale and the confidence intervals present any improvements in terms of face validity.8 We see that three out of six placements remain relatively stable between the two scales (ND, PASOK and its leftist splinter party DIKKI). The ‘logrile’ manages to move the radical left SYN from the extreme right to the left but only at the expense of moving the communist KKE which to the extreme right and the conservative splinter POLAN to the left.

Even though the ‘logrile’ scale does not take into account uncoded quasi-sentences which spuriously move parties towards the centre of the scale, the scale does not take into account the problem of misclassification (Mikhaylov et al., 2012). Misclassification does not only result to random measurement error but also leads to the problem of regression towards the mean (Nesselroade et al., 1980, 624–625) where extremist parties are more likely to move away from the extremes of the scale than centrist parties to move away from the centre of the scale. This has been shown empirically by Mikhaylov et al. (2012, 89) who concluded that misclassification is ‘causing a centrist bias in the estimated Rile values even when the misclassification matrix is strictly symmetric’. The source of misclassification-induced bias seems to be the reduction in the number of issue categories found in the coded documents (Mikhaylov et al., 2012, 89). As proxy documents have been shown to contain fewer issue categories than the respective party manifestos (Table 1), we see why proxy documents lead to systematic measurement error which cannot be corrected by scaling procedures such as the one proposed by Lowe et al. (2011). In the next section I offer a more systematic analysis regarding the question of random versus systematic error in the context of document diversity.

5. Statistical comparisons of documents

So far the paper has demonstrated that the error introduced by proxy documents tends to move L-R estimates towards the centre. Since the previous two sections used evidence from a small number of parties in Greece, it would be useful to see whereas we can detect systematic error in the L-R estimates when comparing parties from various countries and whereas this error can be attributed to document diversity. Detecting systematic error from document diversity is akin to detecting error from method artefacts. In this sense, I follow the reasoning of Zeller and

---

8To be sure, one could challenge the view that the CMP estimates have a low face validity on the basis that the late 1980s and early 1990s was a curious period for party placements in Greece because of the 1989 ND and KKE/SYN government coalition. As Pridham and Verney (1991) and Verney (1990) have amply demonstrated, however, the short-lived (June to November) 1989 coalition was essentially a caretaker government which aimed to facilitate the legal proceedings against the former prime minister and several PASOK ministers who were implicated in a major corruption scandal. There is no evidence that the coalition was based on a programmatic convergence between the parties nor there is evidence that it influenced party positions into the 1990s. For instance, in the 1996 election study conducted by the Centre for Social Research in Greece (EKKE), mean party placement (with 95% CIs) on the 1–10 L-R scale was as follows: KKE: 1.88 [1.79, 1.96], SYN: 2.78 [2.69, 2.87], DIKKI: 4.55 [4.46, 4.64], PASOK: 4.87 [4.77, 4.96], POLAN: 7.02 [6.90, 7.14], ND: 8.40 [8.30, 8.51].
Carmines (1980, 107) who advised researchers to compare two sets of observations one which is assumed to be affected by the method artefact and another which is known that is unaffected. I therefore choose to compare the CMP L-R estimates to estimates coming from expert surveys because it is highly unlikely that expert estimates are influenced by the type of document coded by the CMP. The comparison includes parties in France, Greece, Ireland and Switzerland for which I could match L-R scores between the CMP datasets and the available expert surveys (for details on matching, see the Appendix). The choice of countries was dictated by the need to use the data from the independent classification of documents provided in Figure 1. Although the countries used in the statistical analysis represent only a small portion of the CMP datasets, the practice is in line with other approaches which have made similar comparisons using a subset of countries (Klingemann et al., 2006, 77–79) or elections (Benoit and Laver, 2007, 95–100).

Comparing CMP and expert estimates presents us with two problems however. The first is treating expert estimates as the ‘gold standard’. Even though expert estimates are assumed to have high face validity (Benoit and Laver, 2007, Marks et al., 2007, Steenbergen and Marks, 2007) because we should not expect experts to give estimates which do not make sense (Dinas and Gemenis, 2010, 435), they are far from being a ‘gold standard’ (Budge, 2000, Curini, 2010). The second problem is the use of correlation coefficients to assess error when comparing two methods of measurement. As has been pointed out (Krippendorff, 1970, 144), agreement between two methods (Y = X) is not the same as linearity (Y = bX). Agreement implies that points are close to the 45° line which indicates perfect concordance, while correlation coefficients such as Pearson’s r are high when points are close to any straight line (Bland and Altman, 1986, 307). In other words, Pearson’s r is unable to detect the presence of systematic measurement error because it reflects dispersion rather than precision.

To overcome these problems, I transpose the estimates to a common metric and plot the difference between the CMP and expert L-R estimates versus their mean in what is commonly known (at least in medical statistics) as the Bland-Altman plot. The rationale behind the plot is similar to the rationale behind regression residual plots. The difference between the two measures removes much of the substantive variation leaving the measurement error while, in lack of a true measurement or ‘gold standard’, the mean of the two measures serves as the best estimate available (Bland and Altman, 1986, 308). Moreover, the difference is plotted against the mean rather than each of the measures because the latter approach introduces a statistical artefact since the difference is related to each of the measures (Bland and Altman, 1995).

As seen in Figure 4, the dotted line of mean agreement between the CMP and expert estimates (.003) is very close to the line of perfect agreement at 0. This does not imply a near perfect agreement between the two methods however. To the contrary, the dispersion of points from the y = 0 line indicates disagreement. Moreover, the apparent trend in the plot and the .76 correlation between the difference and the mean imply the presence of systematic measurement error. If there was no bias, the points would be scattered randomly between the limits of agreement and no trend would be discernible. This is not the case in this plot, however. The two methods agree most when parties are close to the centre of the L-R scale and disagree most when parties are placed towards the extremes of the scale. This result is consistent with the findings in sections 3 and 4 of the paper which showed that some of the CMP estimates exhibit a centrist bias. In addition, when the ‘rile’ L-R is replaced with the logged version of ‘rile’ proposed by Lowe et al. (2011), one can still observe a trend indicating bias (correlation between the difference and the mean .52). This is a rather telling finding which leads to the following question: Can some of this bias be attributed to the coding of proxy documents? In other words, does the coding of speeches and other types of documents instead of manifestos make parties appear as centrist when experts argue that the same parties should be placed towards the extremes?

An answer to this question can be given by computing the concordance correlation coefficient (Lin, 1989) which has been devised to assess agreement between two methods of measurement. Unlike Pearson’s r, the concordance correlation coefficient \( \hat{p} \) combines measures of precision and accuracy to determine the average point deviation from the 45° line of perfect concordance. Moreover, unlike intraclass correlation coefficients, \( \hat{p} \) can separate random from systematic error (Lin, 1989, 264). This is because, in essence, \( \hat{p} \) is a multiplicative term between Pearson’s r which measures dispersion and the bias correction factor C, which measures the deviation of

\[ \text{Figure 4: Bland-Altman plot of agreement between expert and CMP L-R estimates.} \]
the fit line from the $45^\circ$ line of perfect concordance. In practical terms, a low $r$ indicates the presence of random measurement error and a low $C_b$ indicates the presence of systematic measurement error. The results for the full sample and for the estimates based in the coding of manifestos, joint platforms, speeches and other documents are presented in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>$n$</th>
<th>$\rho_c$</th>
<th>$\text{SE}(\rho_c)$</th>
<th>$r$</th>
<th>$C_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manifestos</td>
<td>80</td>
<td>.666</td>
<td>.039</td>
<td>.824</td>
<td>.808</td>
</tr>
<tr>
<td>Joint platforms</td>
<td>4</td>
<td>.086</td>
<td>.083</td>
<td>.912</td>
<td>.094</td>
</tr>
<tr>
<td>Speeches</td>
<td>13</td>
<td>.561</td>
<td>.119</td>
<td>.849</td>
<td>.660</td>
</tr>
<tr>
<td>Other documents</td>
<td>10</td>
<td>.280</td>
<td>.317</td>
<td>.402</td>
<td>.929</td>
</tr>
<tr>
<td>All documents</td>
<td>107</td>
<td>.628</td>
<td>.038</td>
<td>.779</td>
<td>.807</td>
</tr>
</tbody>
</table>

Note: The figures compare experts’ L-R estimates to CMP’s ‘rile’ scale. The statistics are, concordance correlation coefficient $\rho_c$, product-moment correlation coefficient $r$, and bias correction factor $C_b$.

Since the results for joint platforms and other documents are only informative due to the small number of observations which result to large standard errors, two things are made clear from looking at the table. First, that the overall agreement between the CMP and expert estimates denoted by $\rho_c$ (.628) is poor (according to Lin, 1989, 263), but slightly better in the set which includes only estimates based in the coding of manifestos (.666). Second, coding speeches instead of manifestos contributes to disagreement by means of systematic measurement error (denoted by the decrease in $C_b$) rather by means of random measurement error (denoted by the slight increase in $r$). These findings indicate that at least some of the systematic error evident in the CMP L-R estimates (as shown in the Bland-Altman plot) can be attributed to the coding of proxy documents.

6. Conclusions

This paper argued that the speeches, pamphlets and other documents which were coded by the CMP in place of manifestos cannot be considered as equivalent to manifestos. They are drafted through different processes and thus might not represent the party as a whole; they have different roles in the election campaigns and thus may not cover all policy issues; and they feature different ‘language registers’ which cannot always be captured effectively and reliably by the CMP coding scheme. This paper showed that the selection and coding of such documents can introduce measurement error in addition to the error introduced into the CMP by other means (McDonald, 2006, 90). Most specifically this error seems to be introducing a centrist bias to the L-R estimates and lead to implausible party placements especially for parties that are generally considered to be closer to the extremes of the L-R scale. Moreover, the paper argued that bootstrapped confidence intervals (Benoit et al., 2009b) and alternative scaling methods (Lowe et al., 2011) which have been devised with random measurement error in mind, cannot always account for such bias. These findings have several implications for the third-party users of the CMP data and documents.

Regarding third-party users of the CMP data, the coding of proxy documents implies that, for many parties, the CMP L-R estimates as well as the emphases on individual policy areas might be invalid due to the presence of systematic measurement error. I concur with Benoit et al. (2009b) and Lowe et al. (2011) that researchers should use the proposed confidence intervals and logit scales, but I add that researchers should also acknowledge that these techniques are only partial solutions to a more fundamental problem. Researchers should double check whether their results are driven by influential observations based on the so called ‘estimates’ or other non-manifesto data, and check whether outliers can be explained by problems of data quality.

Regarding third-party users of the CMP documents, the problem identified in this paper implies that a considerable portion of the CMP collection cannot be used for computerized analyses (cf. Pennings, 2006) or alternative coding schemes unless, of course, the collection is used for comparisons across the same type of documents (e.g. Proksch et al., 2010). I therefore concur with Proksch and Slapin (2009, 338) that ‘computer-based position estimation cannot replace the researcher’s judgement [and] that researchers first need to assure the quality of the source documents’. To this end, Benoit et al. (2009a) have set up an on-line database of documents that can be used by those interested in computerized analyses of manifestos. The database builds heavily on the CMP collection but can be easily updated and includes a feedback mechanism with which users can identify problematic documents and offer suggestions for improvements.

Regarding the CMP as a project, I concur with Hansen (2008, 215) that the CMP team could dedicate some of its resources to trace and code the manifestos which are missing from its collection, replaced by proxy documents, or replaced by interpolated/extrapolated data (the so called ‘estimates’). This is probably easier than we might think. Thirty years ago when the bulk of the CMP documents was collected, researchers had problems locating most manifestos and relied instead on proxy documents. Today, with infrastructure such as the internet and social science data archives, the missing manifestos can be found quite easily. For example, the Finnish Social Science Data Archive (FSD) has collected 846 party programmes of various kinds for 58 Finnish political parties. This is an untapped collection which the CMP could use to replace some rather problematic documents in its collection. Moreover, in many countries, parties have well maintained archives which often include full collections of party manifestos, while in many instances researchers have accumulated such collections themselves during the writing of their PhD theses (e.g. Gibbons, 2003). To this end I have communicated to the CMP team a number of ways which document collection can be improved and provided the team with a number of manifestos which were previously missing or replaced by a proxy document (usually a party leader speech).

Despite earlier suggestions to accept the CMP data ‘as is’ (Budge and Pennings, 2007b, Volkens et al., 2009), the current team is committed to addressing the problems of measurement error that have been outlined in the literature. On the one hand,
the CMP is considering changing its policy of leaving document selection at the coders’ discretion and intensifying the various quality checks (Volkens, 2001). On the other hand, however, the CMP remains skeptical as to whether the project should replace any non-manifesto observations in the existing collection by coding alternative documents (such as the ones that have been recently sent to the CMP). The argument against revisiting the original CMP collection and dataset is that the new documents will have to be coded by different coders and this could introduce a different source of measurement error, namely error stemming from inter-coder unreliability (Mikhaylov et al., 2012). I agree that revising the CMP document collection will not solve each and every problem, but by doing so we could at least have a valid basis of documents on which we can implement different coding and scaling methods. This would only enrich the unparalleled collection of the CMP that has contributed to the advancement of our knowledge on democratic representation, coalition formation, programmatic change and party competition.

Acknowledgements

A previous version of this paper was presented at the workshop ‘The Why’s and How’s of Party Manifestos’ at the 2011 ECPR Joint Sessions of Workshops, University of St. Gallen, 12–17 April. I would like to acknowledge a grant from the German Academic Exchange Service (DAAD) which allowed me to access the archives at the GESIS-Leibniz Institute for the Social Sciences, Cologne and to work on the first draft of this paper at the Centre for the Study of Democracy, University of Lüneburg. I wish to thank Elias Dinas, Isabelle Engeli, Isabelle Guinaudeau, Bruno Hopp, Slava Mikhaylov, Roula Nezi, Renée Nieuwenhuis and Sven Regel for their help with documents and data as well as Thomas Däubler, Iannis Konstantinidis, Zoe Lefkofridi, Sofia Vasilopoulou, Anna Katharina Winkler, the editor and the anonymous reviewers for useful comments on earlier drafts of the paper. Any remaining errors or omissions remain my responsibility.

Appendix: Matching expert L-R estimates to CMP’s ‘rile’ scale

To match the CMP L-R estimates to expert estimates I used estimates from the expert surveys presented in Table 3. In particular, I followed the following steps: a) For each of the four countries included in the analysis (France, Greece, Ireland and Switzerland) I used scores for the periods available in the expert surveys indicated in Table 3. This means that I matched the CMP estimates to expert surveys between 1989 and 2003 for Greek and Swiss parties, between 1982 and 2003 for Irish parties and between 1981 and 2002 for French parties. b) Estimates were matched based on the year of the expert survey fieldwork rather than the year of publication. c) As expert surveys were not always conducted during election years, I interpolated between different surveys to acquire estimates for the election years in between. For example, for the 1989 elections in Greece I use the 1989 survey conducted by Laver and Hunt (1992), but because we have no expert survey for 1993, I obtained estimates by interpolating between the 1989 Laver and Hunt (1992) survey and the next available one (1999, Chapel Hill). Unlike the practice followed by the CMP, however, I did not use extrapolations with the exception of the 1981 election in France for which I used the 1982 Castles and Mair (1984) estimates. d) Because the Laver and Hunt (1992), Laver (1994) and Laver (1998) surveys did not score parties on the L-R scale, I used the average placement between the ‘increase services vs cut taxes’ and ‘pro permissive social policy vs anti’ scales instead, as done by Carter (2005, 143) citing communication with Michael Laver. e) Ireland and Greece had two elections in 1982 and 1989 respectively, so I used the same estimates for both elections. f) SYN contested the 1989 and 1990 elections in Greece as an electoral coalition between KKE and the successor party of KKE-interior. For these two elections I used the average placement between the two coalition partners as the SYN estimate. g) Before estimation, all manifesto and expert estimates were transposed to a common scale ranging from 0 to 1 using the well-known Value–Min Value–Max formula.

Table 3: Expert surveys used for the statistical comparison.

<table>
<thead>
<tr>
<th>Survey</th>
<th>Fieldwork</th>
<th>Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laver and Hunt (1992)</td>
<td>1989</td>
<td>FR, GR, IE, CH</td>
</tr>
<tr>
<td>Laver (1994)</td>
<td>1992</td>
<td>IE</td>
</tr>
<tr>
<td>Huber and Inglehart (1995)</td>
<td>1993</td>
<td>FR, IE, CH</td>
</tr>
<tr>
<td>Laver (1998)</td>
<td>1997</td>
<td>IE</td>
</tr>
<tr>
<td>1999 Chapel Hill Expert Survey*</td>
<td>1999</td>
<td>FR, GR, IE</td>
</tr>
<tr>
<td>2002 Chapel Hill Expert Survey*</td>
<td>2002</td>
<td>FR</td>
</tr>
<tr>
<td>Benoit and Laver (2006)</td>
<td>2003</td>
<td>GR, IE, CH</td>
</tr>
</tbody>
</table>

*See Hooghe et al. (2010) and Steenbergen and Marks (2007).

References


