Mapping Policy Preferences with Uncertainty:
Measuring and Correcting Error in Comparative Manifesto Project Estimates

Kenneth Benoit  
Trinity College Dublin  
kbenoit@tcd.ie

Michael Laver  
New York University  
michael.laver@nyu.edu

Slava Mikhailov  
Trinity College Dublin  
mikhailv@tcd.ie

March 21, 2007

Abstract

A well-known source of data on political party positions on policy is the decades-long comparative manifesto project (CMP). Measuring party positions on policy across time and space, this dataset has been the most widely cited source of party positions in comparative political studies of such phenomena as government duration and formation, electoral outcomes, and policy mandate studies. Despite its widespread use, however, the level of error in the estimates of party positions contained in the dataset has never been estimated or even fully characterized. As a remedy, we outline the process of generating CMP codings and positional estimates, identifying measurement error as coming from coder unreliability, as well as fundamental variability in the stochastic process by which latent party positions are translated into manifesto texts when these documents are written. Using actual quasi-sentence codings from the CMP project, we reproduce the error-generating process through simulation of coder unreliability as well as bootstrapping of coded quasi-sentences to reproduce both forms of error. Using the measurements of this error, furthermore, we suggest and demonstrate ways to use these error estimates in subsequent analyses using the CMP data.

*This research was partly supported by the European Commission Fifth Framework (project number SERD-2002-00061), and by the Irish Research Council for Humanities and the Social Sciences. We thank Andrea Volkens for generously sharing her experience and data regarding the CMP, and Thomas Daubler for research assistance. The outline form of this paper was drawn up for the Conference on the Dynamics of Party Position Taking, March 23-24, 2007, University of Binghamton. A more complete version will be prepared for the Midwest Political Science Association Annual National Conference, Palmer House Hotel, Chicago, April 12-15, 2007.
INTRODUCTION
The Comparative Manifestos Project (CMP) has made available a huge and important dataset on party policy in a large number of countries over the entire post-war period, based on the content analysis of party manifestos. reported by MPP to have covered 2347 programmes, issued by 632 parties, in 52 countries (95). Mention how important CMP has been to political science, give indications of its use, etc. Cites will include (Budge et al. 2001), (Klingemann et al. 2006)

Problem however: CMP has no estimates of uncertainty, and such estimates are necessary for effective use of the CMP scores. Can’t distinguish between signal and noise, making it impossible to distinguish between measurement error and “real” movements in party policy positions from one election to another. A rich time series but rich in error as well as information.

In this paper…we pretty much do what it says in the abstract.

THE CMP CODING SCHEME

General description

• Individual categories
• Scales
• Alternative measures

Issues

• Manifesto length and sensitivity of estimates to length
• Unknown stochastic process in generation of manifests by political parties
• Coder error in translating manifesto content into coding categories
• Other issues not dealt with here
  o scheme salience v. position issues
  o scale content
  o correctly identifying what is a “manifesto”

THE STOCHASTIC PROCESS GENERATING CMP DATA

Mathematical treatment introduced here.

Stochastic manifesto generation

• Manifesto generation is a generally mysterious, idiosyncratic process about which our best knowledge is impartial and anecdotal
• Stochastic process whereby frequency counts of categories are generated by party actors with fixed policy positions
• Stochastic total number of quasi-sentences with codeable content
• Stochastic total number of quasi-sentences with uncodeable content.
• We essentially have no firm information, empirically or theoretically, to tell us anything about the nature of the stochastic process that maps actors’ (unobservable) political preferences to observed manifesto content.

Coder error

• Most of the manifestos that form the basis of the CMP dataset are coded once only by a single coder.
• Goal of coding training is to reduce variance in human coders’ application of the scheme to manifesto texts.
• Training consists of coders learning to code a master document, reproduced in MPP2.
• Coder correlations with the “master” text range from .71 to .89 (see Table 3).
• Volkens also reports that the total number of quasi-sentences varies regularly by +/- 10% of the total quasi-sentences in the master document, depending on how coders parsed the manifesto into quasi-sentence units.
• Possibility of mutually cancelling errors also exists, although Volkens (personal communication) believes these are quite rare.
• Possibility of systematic bias as well, that does not get picked up by correlations.¹
• Even with reported correlations, we really don’t know whether .71 or even .90 is really good or really bad in terms of producing noise in the dataset.
• The ideal way to gain more reliable estimates, given the fact that no coder is perfectly reliable, is to have multiple coders code each text and then combine this information to produce agreement. Not only are the resource implications of this suggestion enormous, however, but also there is no practical way to redo the 3000+ manifestos already coded. (Implication for the next section: The only way to discover this is to simulate the effects of this level of coder variation.)

ESTIMATING ERROR THROUGH SIMULATION

Stochastic Text

• Bootstrapping is a method for estimating the sampling distribution of an estimator by resampling with replacement from the original sample. Does not require any assumptions to be met about the distribution of the data being bootstrapped, and can be effectively used with small sample sizes (N<20). (Efron 1979; Efron and Tibshirani 1994)

¹A second source of error not noted by the CMP is that coders could be subject to systematic bias rather than random error. This is important because, even if the CMP had attempted to conceal from its coders the authorship of the texts under analysis, anyone familiar with party manifestos will know that the source of a party manifesto becomes blindingly obvious as soon as the first sentences of it are read. Coder bias would not necessarily be captured in the correlations reported by the CMP. Table A3 gives an extreme example of such potential bias. This uses the actual CMP codings of the 1997 British Conservative manifesto, in the middle column, and applies very strong bias to these.¹ The result is the final column of Table A3 and the biased and unbiased codings look completely different. The category subject to bias, “political authority”, is in fact a right-wing category in the left-right scale, so that the estimated manifesto position would have been shifted massively to the right by coder bias. Yet the correlation between the “correct” set of codings and the systematically biased codings is 0.712, precisely the type of figure that MPP regards as “quite impressive” (100). Thus very strong systematic bias, as opposed to random error, may not be picked up by correlating the variable counts for a coder, by category, with the official solution.
• We bootstrap quasi-sentences from each original manifesto, after converting the PER categories (percentages) into raw frequencies from the consolidated MPP1/MPP2 datasets. Each manifesto is resampled 100 times.
• Possible disadvantage is that zero-frequency categories will always be zero – but we cannot get around that unless we make some assumptions about probability distributions for non-observed categories, and that is what we are trying to avoid.
• Bootstrapping yields standard errors and confidence intervals associated with each category and also with computed scales, such as right-left, welfare, or the planned economy categories.
• Show and discuss results

Coder error

• Tests on coder reliability indicate most likely "seepage" categories
• Volkens dataset indicates seepage through analysis of what mistakes coders make in the training dataset
• Our simulation procedure replicates this process, aiming to produce a "re-coded" simulated dataset where the correlation with the original is .88-.90.
• We simulate the recoding process 100 times for each manifesto to generate confidence intervals based on this process
• The recoding also changes the means, which are shown in relation to the original values in Figure 2.

USING ERROR TO CORRECT APPLICATIONS USING CMP DATA

Over-time mapping of party policy movement with error (Finally!)

• Explain figures 8, 9, and 13
• Compare with expert survey movement in Britain, Ireland, and Japan (not shown in this draft). We have four expert survey time points for each of these countries, timed at election events.

When using CMP as an independent variable in regression models

• Need an application here and a demonstration

When using CMP as a dependent variable in regression models

• Need an application here and a demonstration

CONCLUDING REMARKS AND RECOMMENDATIONS

• Best practice guidelines for using CMP data with error
• Description of availability of correction data from our website
• Suggestions for future tests and investigations of CMP error
• Additional possibilities to consider for this paper:
  o Looking more at non-"rile" measures, or more on environment. This could include the EU for instance, or welfare or planeco.
  o Could look at social and economic left-right as disaggregated in (Benoit and Laver 2007)
  o Could compare CMP rile versus the Kim-Fording left-right
<table>
<thead>
<tr>
<th>Percentage Uncoded</th>
<th>N</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>864</td>
<td>29.0%</td>
</tr>
<tr>
<td>(0,5]</td>
<td>1,121</td>
<td>66.5%</td>
</tr>
<tr>
<td>(5,10]</td>
<td>375</td>
<td>79.1%</td>
</tr>
<tr>
<td>(10,15]</td>
<td>193</td>
<td>85.6%</td>
</tr>
<tr>
<td>(15,20]</td>
<td>119</td>
<td>89.5%</td>
</tr>
<tr>
<td>(20,30]</td>
<td>129</td>
<td>93.9%</td>
</tr>
<tr>
<td>(30,40]</td>
<td>87</td>
<td>96.8%</td>
</tr>
<tr>
<td>(40,50]</td>
<td>50</td>
<td>98.5%</td>
</tr>
<tr>
<td>(50,65]</td>
<td>30</td>
<td>99.5%</td>
</tr>
<tr>
<td>(65,80]</td>
<td>9</td>
<td>99.8%</td>
</tr>
<tr>
<td>(80,95]</td>
<td>7</td>
<td>100.0%</td>
</tr>
<tr>
<td>(95,100]</td>
<td>-</td>
<td>100.0%</td>
</tr>
<tr>
<td>Total</td>
<td>2,984</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Percentage of uncoded quasi-sentences.

<table>
<thead>
<tr>
<th>Total Quasi-sentences</th>
<th>N</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,5]</td>
<td>-</td>
<td>0.0%</td>
</tr>
<tr>
<td>(5,10]</td>
<td>14</td>
<td>0.5%</td>
</tr>
<tr>
<td>(10,15]</td>
<td>14</td>
<td>0.9%</td>
</tr>
<tr>
<td>(15,20]</td>
<td>29</td>
<td>1.9%</td>
</tr>
<tr>
<td>(20,25]</td>
<td>41</td>
<td>3.3%</td>
</tr>
<tr>
<td>(25,50]</td>
<td>277</td>
<td>12.5%</td>
</tr>
<tr>
<td>(50,75]</td>
<td>303</td>
<td>22.5%</td>
</tr>
<tr>
<td>(75,100]</td>
<td>251</td>
<td>30.9%</td>
</tr>
<tr>
<td>(100,200]</td>
<td>711</td>
<td>54.5%</td>
</tr>
<tr>
<td>(200,250]</td>
<td>208</td>
<td>61.4%</td>
</tr>
<tr>
<td>(250,500]</td>
<td>521</td>
<td>78.7%</td>
</tr>
<tr>
<td>(500,1000]</td>
<td>343</td>
<td>90.1%</td>
</tr>
<tr>
<td>(1000,2000]</td>
<td>216</td>
<td>97.2%</td>
</tr>
<tr>
<td>(2000,5000]</td>
<td>79</td>
<td>99.9%</td>
</tr>
<tr>
<td>(5000,Inf]</td>
<td>4</td>
<td>100.0%</td>
</tr>
<tr>
<td>Total</td>
<td>3,011</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Percentage of uncoded quasi-sentences.
<table>
<thead>
<tr>
<th>Test description</th>
<th>Mean Correlation</th>
<th>N</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training coders' solutions with master</td>
<td>0.72</td>
<td></td>
<td>Volkens (2001, 39)</td>
</tr>
<tr>
<td>Training coders' second attempt with master</td>
<td>0.88</td>
<td></td>
<td>MPP2 (2006, 107)</td>
</tr>
<tr>
<td>All pairs of coders</td>
<td>0.71</td>
<td></td>
<td>Volkens (2001, 39)</td>
</tr>
<tr>
<td>Coders trained on 2nd edition of manual</td>
<td>0.83</td>
<td>23</td>
<td>Volkens (2007, 118)</td>
</tr>
<tr>
<td>First time coders</td>
<td>0.82</td>
<td>14</td>
<td>Volkens (2007, 118)</td>
</tr>
<tr>
<td>First test of coders taking second contract</td>
<td>0.70</td>
<td>9</td>
<td>Volkens (2007, 118)</td>
</tr>
<tr>
<td>Second test of coders taking second contract</td>
<td>0.85</td>
<td>9</td>
<td>Volkens (2007, 118)</td>
</tr>
</tbody>
</table>

Table 3. Coder reliability test results reported by CMP.
Sources are (Volkens 2001; Volkens 2007)
Figure 1: Percentage uncoded categories by Date. Line is LOWESS fitted curve.

Figure 2: CMP right-left score versus right-left with simulated coder error. Overall correlation is .88.
Figure 3: Relationship of total quasi-sentences to level of Right-Left error from bootstrapping procedure (log scale).

Figure 4: Relationship of total quasi-sentences to level of Right-Left error from simulated mis-coding procedure (log scale).
Figure 5: Illustrative sample of left-right estimates and 95% confidence interval from bootstrapping, since 1995. A random sub-sample of 100 parties was chosen since all parties would not fit on a single page. The CMP right-left ("rile") scale runs from -100 to 100.
Figure 6: Illustrative sample of left-right estimates and 95% confidence interval from bootstrapping, before 1955. A random sub-sample of 100 parties was chosen since all parties would not fit on a single page.
Figure 7: Illustrative sample of left-right estimates and 95% confidence interval from bootstrapping, since 1990. This random sub-sample of 100 parties excludes parties with zero mention of the environment (PER501).
Figure 8: Movement over time: Two main British parties on the right-left scale.
Bars indicate 95% confidence intervals.
Figure 9. Movement over time: Two main Irish parties on the right-left scale. Bars indicate 95% confidence intervals.
Figure 10. Left-Right placement of Irish parties in 2002.
Bars indicate 95% confidence intervals.

Figure 11. Left-Right placement of German parties in 2002.
Bars indicate 95% confidence intervals.
Figure 11. Left-Right placement of Italian parties in 2001. Bars indicate 95% confidence intervals. The identical mean values for the Marguerita and Polo della Liberta coalitions indicates that a single coalition manifesto was used to estimate the position of all parties in each coalition.

Figure 12. Left-Right placement of French parties in 2001. Bars indicate 95% confidence intervals.
Figure 13. Movement on environmental policy of German CDU-CSU over time. Movement of Dashed line is % environment with 95% CI; dotted line is the number of quasi-sentences per manifesto coded PER501.
BIBLIOGRAPHY


