Codebook for the Manifesto Project data extended to include the logit scales and standard errors

Will Lowe
University of Mannheim
will.lowe@mzes.uni-mannheim.de

Kenneth Benoit
London School of Economics
kbenoit@lse.ac.uk

Slava Mikhaylov
University College London
v.mikhaylov@ucl.ac.uk

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

July 9, 2012

1 Overview

The data described in this paper are available for download from `http://hdl.handle.net/1902.1/17073`. The dataset contains an extension of the Manifesto Project (MP) data Volkens et al. (2011) to include the results from Benoit, Laver and Mikhaylov (2009) and Lowe et al. (2011).
The data presented here includes all of the variables that are available in the Full Manifesto Project Dataset (`https://manifesto-project.wzb.eu/data/download`)[1], standard errors introduced in Benoit, Laver and Mikhaylov (2009), importance and logit policy dimensions with corresponding standard errors introduced in Lowe et al. (2011). Suffixes indicate whether the variable refers to position, the estimate of importance, and the standard errors of position and importance. For the variable `protectionism` (for example), the five associated variables are:

`protectionism`	position estimate
`protectionism_SE`	standard error for position estimate
`protectionism_imp`	importance estimate
`protectionism_impSE`	standard error for importance estimate

2 Computation of the scaling offset

In Lowe et al. (2011), we converted all "per" categories[2] to counts R and L, and then computed

$$\theta^{(L)} = \log \frac{R+o}{L+o} \tag{1}$$
$$= \log \frac{R+.5}{L+.5} \tag{2}$$
$$= \log (R+.5) - \log (L+.5) \tag{3}$$

[1] Downloaded March 21, 2011

[2] The Manifesto Project data variables are prefixed "per" to denote that each variable is a percentage of (quasi-) sentences devoted to a specific category in the total number of (quasi-) sentences in the respective document.

In the MP dataset[3] released after the publication of our article, however, all of the anomalies and inconsistencies related to the "per" categories have been resolved. For the updated dataset, we compute the logit scales directly on the percentages reported in the dataset. This has no effect on the ratio R/L, although it does affect the offset. Using the percentages, we must rescale the offset o to a new offset $o' = \frac{100}{N}o$ according to the number of total quasi-sentences (`total`) so that we can use $L^{PER} = 100\frac{L}{N}$ and $R^{PER} = 100\frac{R}{N}$:

$$o' = 100\frac{.5}{N} = 50/N \tag{4}$$

The equivalency between the form with the count offset o and the percentage offset o' can be shown as follows:

$$\begin{aligned}
\theta^{(L)} &= \log(R+o) - \log(L+o) && (5)\\
&= \log(R+.5) - \log(L+.5) && (6)\\
&= \log\frac{100(R+.5)}{N} - \log\frac{100(L+.5)}{N} && (7)\\
&= \log\frac{100R+50}{N} - \log\frac{100L+50)}{N} && (8)\\
&= \log(R^{PER}+50/N) - \log(L^{PER}+50/N) && (9)\\
&= \log(R^{PER}+o') - \log(L^{PER}+o') && (10)
\end{aligned}$$

To take an example from the dataset, consider the case for `for party==22320 & edate==17may1946`, where we wish to compute `decentralization`, using the MP codes from per302 (R) and per301 (L):

edate	party	per302	per301	decentr~n	total
17-May-1946	22320	0	3.7	-1.079769	81

For this example:

$$\begin{aligned}
N &= 81 && (11)\\
L &= 3.7/100 \cdot 81 \approx 3 && (12)\\
R &= 0/100 \cdot 81 = 0 && (13)\\
L^{PER} &= 3.7 && (14)\\
L^{PER} &= 0 && (15)\\
o &= 0.5 && (16)\\
o' &= 100\frac{0.5}{81} = 0.617284 && (17)
\end{aligned}$$

[3]Full Manifesto Project Dataset Volkens et al. (2011). Downloaded March 21, 2011

So the two forms that yield identical results are:

$$\begin{aligned}
\theta^{(L)} &= \log(R+o) - \log(L+o) &&= \log(R^{PER}+o') - \log(L^{PER}+o') \\
&= \log(0+.5) - \log(3+.5) &&= \log(0+0.617284) - \log(3.7+0.617284) \\
&= \log(.5) - \log(3.5) &&= \log(0.617284) - \log(4.317284) \\
&= -1.94591 &&= -1.945053
\end{aligned}$$

where the only difference of approximately 0.001 is due to rounding error.
For the logit importance scale, the equivalence is directly analogous:

$$\begin{aligned}
\theta^{(L)} &= \log(R+L+2o) - \log N \\
&= \log(R+L+1) - \log N \\
&= \log(\frac{100}{N}R + \frac{100}{N}L + \frac{100}{N}) - \log(\frac{100}{N}N) \\
&= \log(R^{PER} + L^{PER} + 100/N) - \log 100 \\
&= \log(R^{PER} + L^{PER} + 2o') - \log 100
\end{aligned}$$

3 Description of the variables

The dataset contains 11 MP identification variables, 4 MP data quality variables, 5 MP electoral data variables, 114 MP programmatic data variables, 5 MP programmatic dimensions (Volkens et al., 2011). In addition the dataset contains variables introduced in Benoit, Laver and Mikhaylov (2009) and Lowe et al. (2011): 110 standard errors for MP's individual categories, 4 standard errors for MP's programmatic dimensions, 21 position estimates of policy dimensions with corresponding 21 standard errors for position estimates of policy dimensions, 21 importance estimates with corresponding 21 standard errors for importance estimates.
Table 1 presents the list of variables available in this dataset. Detailed description of Manifesto Project variables is available here:
`https://manifesto-project.wzb.eu/data/MPDataset_full_codebook.pdf`. Table 2 and Table 3 describe the variables introduced in Lowe et al. (2011).

Category	Variable name
MP identification variables	country; countryname; oecd; eu; edate; date; party; partyname; parfam
MP data quality variables	coderid; coderyear; manual; testresult
MP electoral data variables	pervote; voteest; presvote; absseat; totseats
MP manifesto programmatic data variables	progtype; datasetorigin; total; peruncod; per101-per706; per1011-per7062 (subcategories used for CEE countries)
MP programmatic dimensions	rile; planeco; markeco; welfare; intpeace
Standard errors for MP individual categories (Benoit, Laver and Mikhaylov, 2009)	per101SE-per7062SE
Standard errors for MP programmatic dimensions (Benoit, Laver and Mikhaylov, 2009)	rileSE; planecoSE; markecoSE; welfareSE
Position estimates of policy dimensions (Lowe et al., 2011)	decentralization; foreignalliances; militarism; internationalism; logeu; constitutionalism; protectionism; keynesian; nationalism; tradmorality; multicult; laborpolicy; freemarket; environment; stateconomy; stateservices; logrile; logplaneco; logwelfare; loglibcons; education
Standard error for position estimate of policy dimensions (Lowe et al., 2011)	decentralization_SE; foreignalliances_SE; militarism_SE; internationalism_SE; logeu_SE; constitutionalism_SE; protectionism_SE; keynesian_SE; nationalism_SE; tradmorality_SE; multicult_SE; laborpolicy_SE; freemarket_SE; environment_SE; stateconomy_SE; stateservices_SE; logrile_SE; logplaneco_SE; logwelfare_SE; loglibcons_SE; education_SE
Estimate of importance of a policy dimension (Lowe et al., 2011)	decentralization_imp; foreignalliances_imp; militarism_imp; internationalism_imp; logeu_imp; constitutionalism_imp; protectionism_imp; keynesian_imp; nationalism_imp; tradmorality_imp; multicult_imp; laborpolicy_imp; freemarket_imp; environment_imp; stateconomy_imp; stateservices_imp; logrile_imp; logplaneco_imp; logwelfare_imp; loglibcons_imp; education_imp
Standard error for importance estimate (Lowe et al., 2011)	decentralization_impSE; foreignalliances_impSE; militarism_impSE; internationalism_impSE; logeu_impSE; constitutionalism_impSE; protectionism_impSE; keynesian_impSE; nationalism_impSE; tradmorality_impSE; multicult_impSE; laborpolicy_impSE; freemarket_impSE; environment_impSE; stateconomy_impSE; stateservices_impSE; logrile_impSE; logplaneco_impSE; logwelfare_impSE; loglibcons_impSE; education_impSE

Table 1: Variables contained in the dataset. Gray colored rows refer to variables from the Full Manifesto Project Dataset (Volkens et al., 2011). Rows colored in pink refer to variables introduced in Benoit, Laver and Mikhaylov (2009). Rows colored in yellow refer to variables introduced in Lowe et al. (2011), with further description of these variables available in Table 2 and Table 3.

Policy dimension	"Left" Position	"Right" Position	Variable name
Foreign alliances	101 Foreign Special Relationships: Positive	102: Foreign Special Relationships: Negative	`foreignalliances`
Militarism	105 Military: Negative	104 Military: Positive	`militarism`
Internationalism	107 Internationalism: Positive	109 Internationalism: Negative	`internationalism`
EU	108 European Integration: Positive	110 European Integration: Negative	`logeu`
Constitutionalism	203 Constitutionalism: Positive	204 Constitutionalism: Negative	`constitutionalism`
Decentralisation	301 Decentralisation: Positive	302 Centralisation: Positive	`decentralization`
Protectionism	406 Protectionism: Positive	407 Protectionism: Negative	`protectionism`
Keynesian policy	409 Keynesian Demand Management: Positive	414 Economic Orthodoxy: Positive	`keynesian`
Nationalism	602 National Way of Life: Negative	601 National Way of Life: Positive	`nationalism`
Traditional morality	604 Traditional Morality: Negative	603 Traditional Morality: Positive	`morality`
Multiculturalism	607 Multiculturalism: Positive	608 Multiculturalism: Negative	`multiculturalism`
Labour policy	701 Labour Groups: Positive	702 Labour Groups: Negative	`laborpolicy`
Welfare state*	504 Welfare State Expansion: Positive	505 Welfare State Limitation: Positive	`logwelfare`
Education spending	506 Educational Provision Expansion: Positive	507 Education Expenditure Limitation: Positive	`education`

Table 2: *Paired policy dimensions and corresponding variable names in the dataset.* *Note: `Welfare state` scale differs from the MP's `welfare` scale in that the MP's version is not confrontational (and does not include `per504`).

Policy dimension	"Left" Position	"Right" Position	Source	Variable name
Free Market Economy	401 Free Enterprise: Positive + 402 Incentives: Positive	403 Market Regulation: Positive + 412 Controlled Economy: Positive + 413 Nationalisation: Positive + 415 Marxist Analysis: Positive	(Proposed)	`freemarket`
Environmental protection	501 Environmental Protection: Positive + 416 Anti-Growth Economy: Positive	410 Productivity: Positive	(Proposed)	`environment`
State involvement in economy	403 Market Regulation: Positive + 404 Economic Planning: Positive + 406 Protectionism: Positive + 412 Controlled Economy: Positive + 413 Nationalisation: Positive + 504 Welfare State Expansion: Positive + 506 Education Expansion: Positive + 701 Labour Groups: Positive	401 Free Enterprise: Positive + 402 Incentives: Positive + 407 Protectionism: Negative + 414 Economic Orthodoxy: Positive + 505 Welfare State Limitation: Positive	Benoit and Laver (2007)	`stateconomy`
State-provided services	504 Welfare State Expansion: Positive + 506 Education Expansion: Positive	505 Welfare State Limitation: Positive + 507 Education Limitation: Positive	(Proposed)	`stateservices`
Left-Right: CMP "Rile"	103 Anti-Imperialism: Anti-Colonialism + 105 Military: Negative + 106 Peace: Positive + 107 Internationalism: Positive + 202 Democracy: Positive + 403 Market Regulation: Positive + 404 Economic Planning: Positive + 406 Protectionism: Positive + 412 Controlled Economy: Positive + 413 Nationalisation: Positive + 504 Welfare State Expansion: Positive + 506 Education Expansion: Positive + 701 Labour Groups: Positive	104 Military: Positive + 201 Freedom and Human Rights: Positive + 203 Constitutionalism: Positive + 305 Political Authority: Positive + 401 Free Enterprise: Positive + 402 Incentives: Positive + 407 Protectionism: Negative + 414 Economic Orthodoxy: Positive + 505 Welfare State Limitation: Positive + 601 National Way of Life: Positive + 603 Traditional Morality: Positive + 605 Law and Order: Positive + 606 Social Harmony: Positive	Laver and Budge (1992)	`logrile`
Planned v. market economy	403 Market Regulation: Positive + 404 Economic Planning: Positive + 412 Controlled Economy: Positive	401 Free Enterprise: Positive + 414 Economic Orthodoxy: Positive	*MPP*	`logplaneco`
Social liberal-conservative	103 Anti-Imperialism: Anti-Colonialism + 105 Military: Negative + 106 Peace: Positive + 107 Internationalism: Positive + 202 Democracy: Positive	104 Military: Positive + 201 Freedom and Human Rights: Positive + 203 Constitutionalism: Positive + 305 Political Authority: Positive + 601 National Way of Life: Positive + 603 Traditional Morality: Positive + 605 Law and Order: Positive + 606 Social Harmony: Positive	Benoit and Laver (2007)	`libcons`

Table 3: *Additive scaled dimensions and corresponding variable names in the dataset*.

References

Benoit, Kenneth and Michael Laver. 2007. "Estimating Party Policy Positions: Comparing Expert Surveys and Hand Coded Content Analysis." *Electoral Studies* 26(1):90–107.

Benoit, Kenneth, Michael Laver and Slava Mikhaylov. 2009. "Treating Words as Data with Error: Uncertainty in Text Statements of Policy Positions." *American Journal of Political Science* 53(2):495–513.

Laver, Michael and Ian Budge. 1992. *Party policy and government coalitions*. New York: St. Martin's Press.

Lowe, William, Kenneth Benoit, Slava Mikhaylov and Michael Laver. 2011. "Scaling Policy Preferences From Coded Political Texts." *Legislative Studies Quarterly* .

Volkens, Andrea, Onawa Lacewell, Pola Lehmann, Sven Regel, Henrike Schultze and Annika Werner. 2011. *The Manifesto Data Collection*. Manifesto Project (MRG/CMP/MARPOR), Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB).