Research Collection School of Social Sciences   School of Social Sciences

9-2012

How to scale coded text units without bias: A response to Gemenis

Kenneth BENOIT
Singapore Management University, kbenoit@smu.edu.sg

Michael LAVER

Will LOWE

Slava MIKHAYLOV

Citation

How to scale coded text units without bias: A response to Gemenis

Kenneth Benoit [a,*], Michael Laver [b], Will Lowe [c], Slava Mikhaylov [d]

[a] Methodology Institute, London School of Economics, Houghton Street, Columbia House, London WC2A 2AE, United Kingdom
[b] Department of Politics, New York University, United States
[c] MZES, University of Mannheim, Germany
[d] Department of Political Science, University College London, United Kingdom

ARTICLE INFO

ABSTRACT

Coding non-manifesto documents as if they were genuine policy platforms produced at election time clearly raises serious issues with error when these codings are used in the standard manner to estimate left-right policy positions. In addition to the long term solution of improving the document base of the Manifesto Project identified by Gemenis (2012), we argue that immediate gains in manifesto-based estimates of policy positions can be realised by using the confrontational logit scales from Lowe et al. (2011), which addresses the problems of scale content and scale construction that are exacerbated by but not unique to the problems found in proxy documents.

© 2012 Elsevier Ltd. All rights reserved.

The core point made by Gemenis (2012) is that a significant number of the documents coded by the Comparative Manifestos Project (CMP) are not party manifestos at all, as widely believed in the profession, but rather "proxy documents" such as party leader speeches and newspaper reports. This raises the level of error in the CMP dataset, Gemenis argues, because the coding scheme is designed for party manifestos and does not capture accurately the policy content of proxy documents. Moreover, this failure cannot be corrected by existing methods. Specifically, Gemenis identifies two problems. First, as demonstrated by comparing Greek substitute documents to genuine manifestos, the proxy documents may contain a disproportionate number of text units (mis)coded in such a way as to bias the resulting policy measures, for instance by making communist parties appear right-wing. Second, because the proxy manifestos tend to contain higher proportions of uncodeable content, standard policy measures such as the CMP's left-right "Rile" measure will suffer attenuation bias, appearing more centrist than their actual policies warrant.

The only solution, Gemenis claims, is to code manifestos not proxy documents, since existing measures such as Benoit et al. (2009) and Lowe et al. (2011) correct neither problem. We agree that the method for computing confidence intervals introduced in Benoit et al. (2009) – designed only to estimate uncertainty while assuming that the correct manifestos have been correctly coded – offers no solution. We argue here, however, that these problems can be significantly mitigated by using the scale correction presented in Lowe et al. (2011). The problems arising from coding non-manifesto documents as if they were genuine manifestos, furthermore, are inherent to the saliency-based scales employed by the CMP and most widely used by third party authors. The confrontational, logit-scaled alternatives offered by Lowe et al. (2011), by contrast, do directly address the problem of bias arising from over-representation of poorly fitting policy codes in the CMP scales – including Rile – and are also unaffected by uncoded content. Our recommendation, bolstered by results presented below, is always to use the logit scales of Lowe et al. (2011), and to substitute more specific confrontational

* Corresponding author. Tel.: +44 20 7955 6812.
E-mail address: kbenoit@lse.ac.uk (K. Benoit).

indexes such as their state economy scale for the more inclusive Rile measure. We agree that the quality of the CMP's manifesto set could be improved by replacing proxy documents with genuine manifestos. But this is a big undertaking, and we show that using better scales on existing CMP data not only avoids some of the problems generated by proxy documents, but also improves policy estimates from hand-coded political text more generally. The problems Gemenis identifies only strengthen this general rationale.

1. Bias and the components of the Rile index

Non-manifesto texts may be coded – inappropriately or not – to categories that are classified as one of the 13 "right" policy categories of the 26 category Rile index. In the case of the far left KKE and the rather leftist SYN in Greece, Gemenis's analysis demonstrates that the right-assigned categories of PER305 ("political authority: positive") and PER606 ("social harmony:" positive) were used heavily in the coding of party leader speeches, treated as proxies for party manifestos, resulting in the (mis)estimation of these parties as having right-wing policy positions. This echoes findings by authors in Laver and Budge (1992) that government policy declarations from cabinets of both right and left tend to contain systematically more references than manifestos to matters such as government authority. The effect of this is that, using Rile, governments tend to be measured as being systematically more to the right than the parties that they comprise. For the same reason, leader speeches may be coded as systematically more right-wing than the manifestos they proxy. This problem is likely aggravated by the fact that the "PER305 Political Authority: Positive" tends to be misused by coders, as demonstrated by coding experiments in Mikhaylov et al. (2012).

Gemenis suggests that the solution is to seek out the genuine manifestos that do not contain so many references to problematic categories, such as "Political Authority: Positive". Even without tracking down real manifestos that were not found by the CMP country specialists, we show here that an alternative solution is to use better scales that do not include these problematic categories. Lowe et al. (2011) propose four economic left-right policy scales, and several others capturing non-economic policy, providing a full dataset of these measures (including bootstrapped confidence intervals). Using one of these measures, for instance the logit scale of statecomy,[2] it can be shown that even with the inappropriately chosen leader speeches from the Greek case many of the parties returned to more

reasonable-looking positions on the dimension of left-right economic policy.

In Fig. 1, we plot the positions over time for both the main Greek parties analyzed in Fig. 2 of Gemenis and otherwise identified as having problematic positions because of the use of proxy documents. Fig. 1 plots three logit scales: state-conomy, Rile, and a combined scale of PER305 + PER606 as the most problematic components of Rile. Fig. 1 clearly shows that for Greek parties, the Rile scale closely tracks its two "right" assigned components, supporting Gemenis's claim that these two categories are primarily responsible for the mis-classification of KKE and SYN as right-wing parties.

In Fig. 2, we plot the positions of the Greek parties from the 1996 election, positions which may be compared directly to the lower half of Gemenis's Figure 3.[3] These results are much more in line with Gemenis's expectations of party placements in Greece in the 1996 election (see fn. 9 in Gemenis). The alignment of main parties KKE, PASOK and ND on the economic dimension corresponds to the expert survey results in Laver and Hunt (1992) and Benoit and Laver (2006). The leftist SYN has a very wide confidence interval around its position estimate which does not allow us to distinguish SYN from either PASOK or ND, but its position is clearly distinct from KKE's. This may in part reflect the fact that out of four main Greek parties in the Benoit and Laver (2006) survey, SYN attached the least importance to the economic dimension, while KKE attached the highest.

2. Centrist bias and the saliency scale

The second problem that Gemenis raises with ersatz manifestos concerns the centrist bias caused by their relatively higher proportions of uncoded text units. This is because the Rile (and indeed all other CMP scales) use what Lowe et al. (2011) call an "absolute proportional difference" or saliency scale that divides the difference of "right"- and "left"-coded content by the total number of text units. The denominator will increase whenever the number of total text units increases, moving the resulting scale position closer to zero, even when the new text units are not in any way related to the scale in question. To take a very simple example, imagine a document from a left-wing party with a total (N) of 100 sentences, in which 50 sentences were coded left (L) and zero coded right (R). The Rile score is $(R - L)/N = -0.5$. Now imagine that 50 sentences are added to the manifesto, consisting of uncodable rhetoric singing the praises of the party leader and trashing the other parties. The Rile score is now -0.33 and the party appears to have moved to the center.

Gemenis's diagnosis of the problem is correct, but simply ensuring that all CMP documents are genuine manifestos is not a complete cure. This is because the root problem of the centrist bias in the CMP's policy scales stems from the CMP's "saliency" approach, which counts all

---

[2] This scale, first introduced in Benoit and Laver (2007), is formed from the manifesto codes identifying the "right" position (PER401 Free Enterprise: Positive + PER402 Incentives: Positive + PER407 Protectionism: Negative + PER414 Economic Orthodoxy: Positive + PER505 Welfare State Limitation: Positive) and the "left" position (PER403 Market Regulation: Positive + PER404 Economic Planning: Positive + PER406 Protectionism: Positive + PER412 Controlled Economy: Positive + PER413 Nationalization: Positive + PER504 Welfare State Expansion: Positive + PER506 Education Expansion: Positive + PER701 Labor Groups: Positive).

[3] We note that using original CMP data and the logit scales published by Lowe et al. (2011), we could not replicate the exact results for the "logit Rile" scale in Gemenis Figure 3.
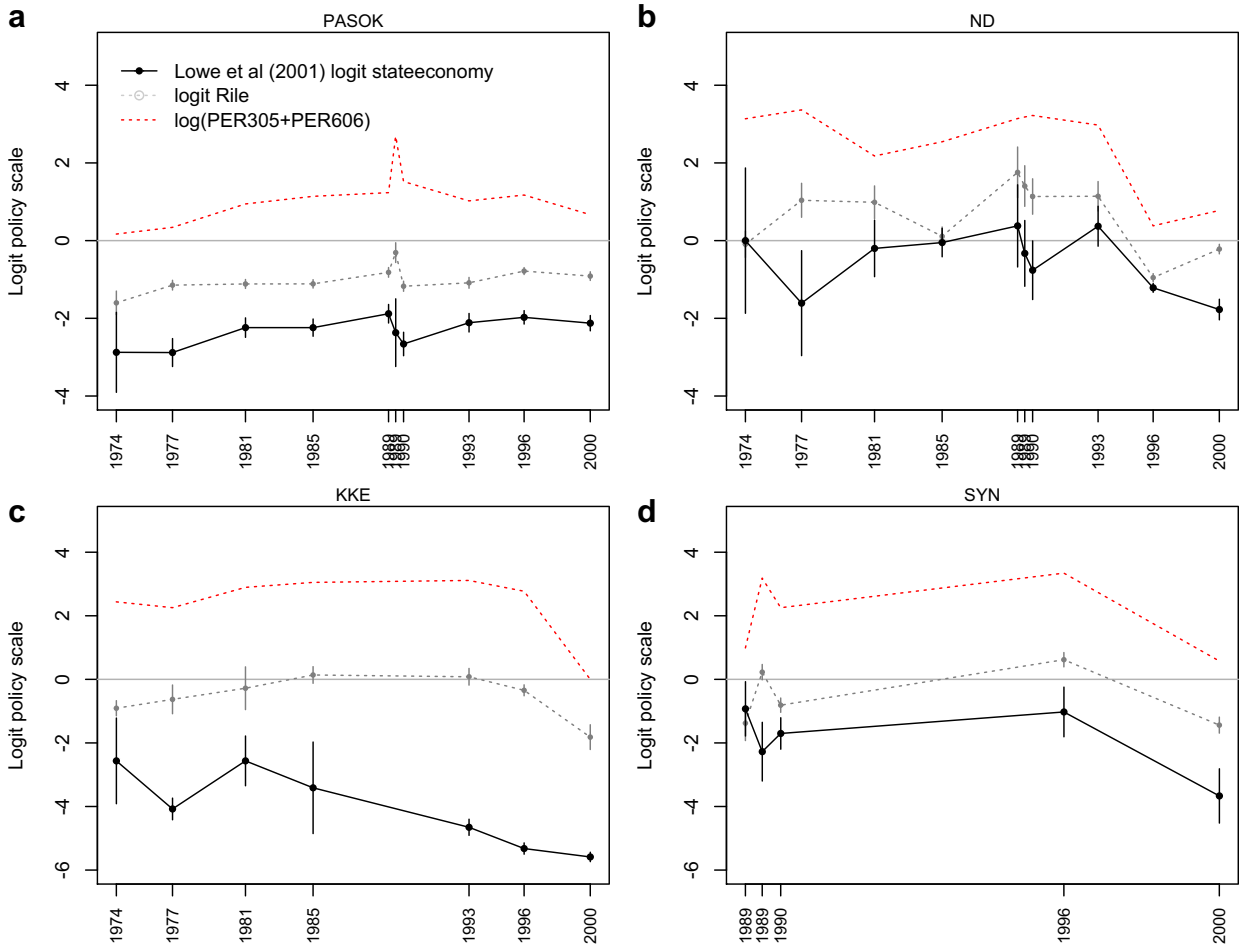
Fig. 1. Trend Differences for Greek Parties in Logit Rile and Logit stateconomy (Lowe et al., 2011). From the Lowe et al. (2011) dataset.

uncoded text units in the denominator. This centrist bias will be greater for documents with more uncoded sentences, but will be present in *any* saliency-based scale, even for documents with *no* uncoded content, because the denominator may still count the text units coded to categories outside the scale. A party will appear more centrist on economic policy, for example, simply because more foreign policy sentences have been added to the manifesto.[4]

Scales that do not count uncoded content are unaffected by this type of centrist bias, but the relative proportional difference scale discussed by Gemenis—the $(R - L)/(R + L)$ scale proposed by Kim and Fording (1998) and Laver and Garry (2000) – is not recommended either. As Lowe et al. (2011) show, the problem is now one of forcing scores toward the extremes. Compare two 100 sentence manifestos: one has a single sentence coded R, with the rest coded as neither R nor L; the other has all 100 sentences

coded R. Both generate the most extreme possible right score of $+1$. Given this, Lowe et al. (2011) recommend scales using only R and L, but based on the log odds ratios of these quantities to reflect the diminishing marginal impact of extra sentence content. Because these logit scales do not
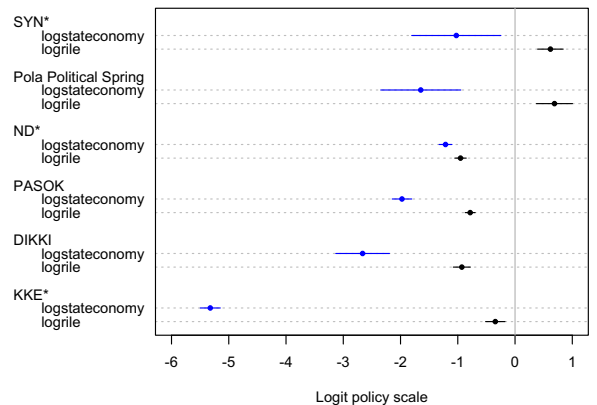

Fig. 2. Party Positioning on Logit Rile and Logit stateconomy, Greece 1996. From the Lowe et al. (2011) dataset. Asterisks denote "proxy" manifestos as identified by Gemenis.

---

[4] This problem is addressed directly and illustrated for non-economic social liberalism versus conservatism in Figure 3 of Lowe et al. (2011, 145). The top panel clearly shows the shrinkage toward zero of the saliency scale approach, while the *y*-axis from the Benoit and Laver (2006) expert surveys clearly spans the full range of its 1–20-point scale.

count uncoded content, if the problem with proxy documents is that they systematically contain more uncoded content and are thus systematically biased toward the center when using the CMP Rile scale, then the Lowe et al. (2011) scale fixes this. Gemenis has in effect demonstrated another reason to use the logit scales rather than either the CMP's saliency scales or the relative proportional difference alternatives proposed by Kim and Fording (1998) and Laver and Garry (2000).

3. Recommendation: use the CMP's codings but not its policy scales

The two problems identified by Gemenis – centrist bias due to systematically higher uncodable content in proxy documents, and right-wing bias due to systematically higher incidence of references to right-wing CMP categories in some proxy documents such as speeches – can each be rectified by using non-rile logit scales from Lowe et al. (2011). That paper addressed both scale construction (replacing the absolute difference scale of the CMP with the logit scale) and scale content (by offering numerous non-Rile alternatives that would be unaffected by the PER305 and PER606 problems). For third party users

of the CMP data, therefore, the best recommendation is still to use Lowe et al. (2011) logit scales, with a specific recommendation to avoid the overly broad Rile index.

References

Benoit, K., Laver, M., 2006. Party Policy in Modern Democracies. Routledge, London.

Benoit, K., Laver, M., 2007. Estimating party policy positions: comparing expert surveys and hand coded content analysis. Electoral Studies 26 (1), 90–107.

Benoit, K., Laver, M., Mikhaylov, S., April, 2009. Treating words as data with error: uncertainty in text statements of policy positions. American Journal of Political Science 53 (2), 495–513.

Gemenis, K., 2012. Proxy documents as a source of measurement error in the Comparative Manifestos Project. Electoral Studies. doi:10.1016/j.electstud.2012.01.002

Kim, H., Fording, R.C., 1998. Voter ideology in western democracies, 1946–1989. European Journal of Political Research 33 (1), 73–97.

Laver, M., Garry, J., 2000. Estimating policy positions from political texts. American Journal of Political Science 44 (3), 619–634.

Laver, M., Hunt, B.W., 1992. Policy and Party Competition. Routledge, New York.

Laver, M., Budge, I., 1992. Party Policy and Government Coalitions. St. Martin's Press, New York.

Lowe, W., Benoit, K., Mikhaylov, S., Laver, M., 2011. Scaling policy preferences from coded political texts. Legislative Studies Quarterly 36 (1), 123–155.

Mikhaylov, S., Laver, M., Benoit, K., 2012. Coder reliability and misclassification in the human coding of party manifestos. Political Analysis 20 (1), 78–91.