Assessing the Validity of the Manifesto Common Space Scores

Jan-Erik Flentje*, Thomas König†, Moritz Marbach‡

Abstract

RILE estimates based on party manifesto data suggest that political parties leapfrog on the left-right scale over time. This implausible finding has raised questions about the efficacy not only of RILE for estimating left-right positions but of coded party manifestos for political science research in general. The recently developed Manifesto Common Space Scores (MCSS), which reduce leapfrogging by accounting for the election-specific character of party manifestos, provide alternative estimates for parties' left/right-positions, but little is known about their validity. This study shows that MCSS estimates exhibit greater convergent validity relative to RILE estimates when compared to other measures of parties' left/right-positions. It also finds that MCSS has greater construct validity relative to RILE estimates in two prominent cases (Greece and Italy). Overall, the findings underscore the election-specific character of party manifestos and demonstrate that MCSS is a useful alternative measure of parties' left-right positions.

*University of Heidelberg, Bergheimer Str. 58, 69115 Heidelberg, Germany; flentje@stud.uni-heidelberg.de.
†University of Mannheim, A5-6, 68159 Mannheim, Germany; koenig@uni-mannheim.de. Corresponding author.
‡University of Mannheim, A5-6, 68159 Mannheim, Germany; mmarbach@mail.uni-mannheim.de.

Contents

1 Introduction

The Comparative Manifesto Project (CMP) provides political scientists with a dataset of more than 3,000 party manifestos, which are coded into a scheme of 56 categories that reflect broad policy issues (Budge et al., 2001; Klingemann et al., 2006). Based on this dataset, the CMP's Right-Left (RILE) score – which persuades by computational simplicity[1] – has become the most frequently used estimator for measuring the ideological left-right positions of political parties in the field of comparative politics (e.g., Martin and Vanberg, 2005; Tavits and Letki, 2009; Adams and Somer-Topcu, 2009).

While frequently used, a closer inspection of the RILE estimates reveals a puzzling ideological volatility (leapfrogging) of political parties over time in many party systems. A typical example is the zigzag movement of parties in Sweden, which is conventionally considered to have a frozen party system due to voters' strong partisan alignment (Lipset and Rokkan, 1967; Foldal, 1989; Vedung, 1988; Sundberg, 1999). Figure 1 (upper panel) illustrates the volatility of the RILE estimates for the most important Swedish parties, which stands in sharp contrast to the system's observed stability and moderate gradual change of parties' positions over time (Hanley, 1999; Mair, 1999; Hug, 2001; Golder, 2003). Even if one expects some change in the Swedish party system with the emergence of the Democrats in 2010 (Pierre, 2015), the sudden collapse of all Swedish parties into a similar ideological left-right position, as indicated by the RILE estimates, is peculiar.

According to Adams (2001) voter bias provides an incentive for political parties to avoid leapfrogging, which has raised questions about the plausibility of the temporal patterns of party positioning and has called left-right estimates based on the CMP into question (Benoit and Laver, 2006, 2007). More specifically, König et al. (2013) (henceforth KMO) question whether counts of left-right categories from the CMP data, and hence the resulting RILE estimates, are comparable over time. According to KMO, the price for the appealing computational simplicity of RILE is to ignore that party manifestos are election-specific statements written for party competition at one particular point in time. Hence, scholars using RILE estimates follow the assumption that parties position themselves, irrespective of their competitors on the left-right scale, when they draft and adopt their party manifestos. Under such a myopic-positioning assumption, leapfrogging (as exemplified in the case of Sweden above) are plausible patterns. However, if one instead believes that parties position themselves relative to their competitors, then neither the left-right category counts nor the resulting RILE estimates are directly comparable over time, and the observed leapfrogging suggests that RILE estimates are contaminated by measurement error. This is worrisome given that measurement error biases coefficient estimates in regression models towards zero.

To construct plausible temporal patterns of party positioning from party manifestos that take into account the relative positioning of parties and consequently are more comparable over time, KMO propose to employ bridge observations in a version of a factor analytical model that integrates each election-specific left-right scale into a common policy space. Using this model and transformed manifesto data, KMO estimate their Manifesto Common Space Scores (MCSS) for parties in 25 European Union (EU) member countries in the period 1945–2010. Their findings show that parties are ideologically much less volatile than suggested by RILE, and that there are significant election-specific changes to the entire left-right scales that distort the comparability of party position estimates if not taken into account.

The lower panel in figure 1 shows their estimated MCSS for Sweden. Consistent with conventional

[1] The RILE score is computed by subtracting the total counts of 13 predefined left categories from the total counts of 13 right categories and dividing by the sum of these categories.

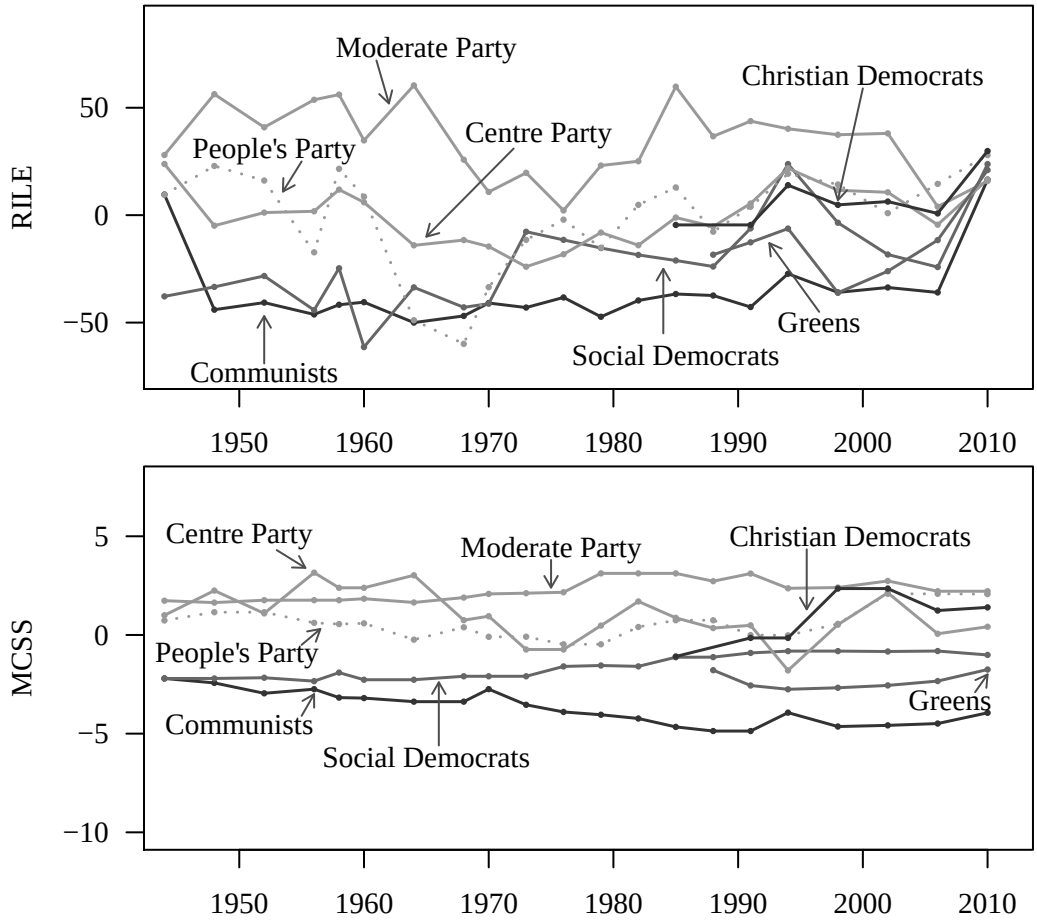

Figure 1: Party positions for Sweden: The upper panel shows the RILE estimates and the lower panel shows the MCSS estimates.

wisdom, the MCSS estimates indicate that the left-right positions of Swedish parties are quite stable over time and do not collapse to a single point in the recent election. Over time, some trends are visible, but there is no leapfrogging from one election to the other, nor does the overall stability of the Swedish party system change (König et al., 2013, 486). These findings, for Sweden, support KMO's approach, but the question is whether MCSS estimates are generally more valid relative to RILE and less prone to measurement error. Compared to the many insights on RILE's convergent validity[2], little is known about MCSS and whether the estimates are more valid relative to the prominent RILE estimates.[3]

In this study, we fill this gap by evaluating the relative performance of RILE and MCSS estimates focusing on the convergent and construct validity. We begin by conducting a country-by-country analysis to examine the convergent validity of MCSS and RILE estimates relative to left-right estimates from other, independent data sources. For this purpose, we compiled a large dataset of validation measures

[2]Recent instances include, for example, Dalton and McAllister (2015) and Bakker et al. (2015) who, inter alia, point out that RILE correlates only weakly with expert- and mass-survey data (relative to the correlation between expert- and mass-survey data).

[3]We focus on RILE because alternative estimators, including the Logit scores (Lowe et al., 2011), the Vanilla scores (Gabel and Huber, 2000), and the Franzmann-Kaiser scores, (Franzmann and Kaiser, 2006) have not been applied as often as RILE.

estimated from text data, as well as mass- and expert-survey data. Since the usage of RILE is typically justified by the provision of a long time-series of party positions, we focus on the country-by-country longitudinal performance of MCSS vis-à-vis RILE and leave the cross-sectional performance to future research. To evaluate the construct validity, we further re-analyze two critical cases: The Greek and Italian party systems. In both cases, scholars have raised serious concerns about the construct validity of RILE (Pelizzo, 2003; Dinas and Gemenis, 2010), so MCSS need to perform better in these two cases if it is to be considered a useful alternative to RILE.

We find that MCSS outperforms RILE in a very large majority of pairwise comparisons with the validation from other data sources. In 10 out of 15 comparisons (67%), MCSS estimates correlate higher than RILE estimates with measures based on automated text analysis. This is also confirmed by our findings on mass surveys, in particular those surveys that capture election-specific conditions. Out of 19 mass surveys, MCSS performs better in 17 cases (89%). This is also corroborated by expert surveys, in which RILE only very rarely outperforms MCSS (13 out of 50 cases). Finally, our analysis of two critical cases indicates that MCSS exhibits a higher construct validity relative to RILE. While the latter results cannot be generalized to other cases, they demonstrate that MCSS ameliorates the problems of RILE in these two prominent, critical instances.

2 CMP, Comparability and MCSS

The Comparative Manifesto Project (CMP) is one of the most important data sources for estimating parties' positions. The CMP classifies each quasi-sentence of a party manifesto into a scheme of 56 categories that capture a predefined set of political issues (Budge et al., 2001; Klingemann and Volkens, 2007). The data, which are available for over 50 countries covering a period of up to 70 years, contain the counts of the categories identified by manual coders for each manifesto. Several methodological debates involve the CMP[4], but the largest is the question about how to estimate left-right positions from these data.[5] In general, there are two approaches: deterministic scaling techniques and latent variable models.

Deterministic scaling techniques are functional transformations of the categories' counts. The right-left (RILE) score, for which the total counts of 13 predefined left categories is subtracted from the counts of 13 right categories and divided by the sum of these categories, is the most prominent estimator for deterministic scaling that represents the initial idea of the CMP's salience-based approach (Budge et al., 2001; Laver and Garry, 2000). Alternative estimators have been developed, most recently the Logit score, which is more suitable for estimating party positions on specific issues before aggregation to broader ideological dimensions (Lowe et al., 2011).[6] Latent variable models come in various forms. Earlier applications of latent-variable models include the "Vanilla" approach by Gabel and Huber (2000). More recent tailored approaches include, for example, the ones by Bakker (2009) and Elff (2013).

[4] Another methodological debate focuses on the inter-coder reliability of the coded manifesto data (e.g., Mikhaylov et al., 2012; Budge, 2013; Lacewell and Werner, 2013; McDonald and Budge, 2014) and the measurement of coding uncertainty (e.g., Benoit et al., 2009; Budge et al., 2013; McDonald, 2013).

[5] For a review of estimating party positions, see Volkens (2007); Laver (2014) and the specific studies by Budge (1999); Laver and Garry (2000); Kim and Fording (2002); McDonald and Mendes (2001); Benoit and Laver (2007); Ray (2007); Lowe et al. (2011).

[6] For Logit and RILE scores we find a Pearson's correlation of about $r = 0.67$ in the complete CMP dataset until 2010, while others report a higher correlation of $r = 0.94$ (Meyer, 2010; Budge and McDonald, 2012; Budge and Meyer, 2013).

2.1 Shifting Spaces

The CMP employs a common coding scheme for all 3,000 party manifestos, regardless of the country and time period. While the purpose for using a common coding scheme is to increase the comparability of parties' positions, the question is whether the meaning and usage of categories changes over time. A typical example is category per415, which codes positive references to Marxist-Leninist ideology and specific uses of Marxist-Leninist terminology in a party manifesto. With the breakdown of the Soviet Union, the meaning and usage of the per415 category has drastically changed.

After all, party manifestos are written by parties for distinct elections, which, following Downs (1957), attempt to position themselves *relative* to their competitors. This implies that common shocks such as the breakdown of the Soviet Union, which may cause common shifts in the parties' positions, are hardly reflected in the parties' manifestos, or the resulting CMP data and RILE estimates. Note that this is not a deficit of using a common coding scheme but rather an inevitable consequence for all estimators that operate on *absolute* frequency counts of words, n-grams, or sentences.

Common shifts in all parties' left-right positions (that is, shifts of the entire party system towards the left or right) across elections can have various sources. Across countries, it might well be that all parties in one country prefer policies that are more to the left compared to all parties in another country. This appears to be a plausible conjecture for parties of a country with a social market economy (e.g., Germany) compared to parties in a country with a liberal market economy (e.g., the United Kingdom). However, even within a country, we may observe shifts from one election to another, i.e., in times of shocks such as an ecological disaster or a terrorist attack, or as gradual shifts occur over long periods of time, e.g., a shift towards liberalism on LGBT issues.

2.2 Bridging Spaces

To remove the distortion that results from unobserved shifts and baselines such as those described above, KMO develop a latent variable model similar to the standard factor analysis model used by Gabel and Huber (2000).[7] In their application, which covers 25 European Union member countries in the period 1945–2010, they use 16 positional issue scales constructed by König and Luig (2012) from the coded manifesto data[8]

The general idea of KMO is to decompose the latent factor into a linear combination of the left-right position (they refer to this score as the Manifesto Common Space Score, or MCSS) and two bias parameters that capture 1) the potential shift of the party system relative to systems in other countries ("country-bias") and, more importantly for this study, 2) the potential shift of a party system relative to preceding elections ("time-bias").

In order to disentangle the time-bias (the unobserved shifts of all parties) from actual changes in parties' left-right positions, the authors identify a series of parties whose relative position across elections is known a priori. These parties serve as bridge observations across elections and are derived from

[7] Similar to Bakker (2009), KMO take a Bayesian perspective for inference and use informed prior densities for the parameters that capture the left-right position of a party. More specifically, they assign each party to a party family and party families to particular subspaces of the left-right scale. The location and variance estimates of party families are estimated from the Chapel-Hill expert surveys (Hooghe et al., 2010; Steenbergen and Marks, 2007).

[8] König and Luig (2012) and assign 36 of the 56 CMP categories to opposing poles of an issue-specific position. Each issue-specific position is constructed by subtracting the logarithms of each pole's quasi-sentence count, as suggested by Lowe et al. (2011).

instrumental assumptions. To estimate the time-bias, they employ an "incentive" hypothesis, according to which parties that gained the largest seat share in comparison to the previous election have little incentive to change their position in the next election and thus compete on the same left-right position in both elections.[9]

In figure 2, we show the estimated time-bias for each country. A positive (negative) estimate implies that, ignoring the time-bias, parties in each respective election would appear to be more right (left) than they are in fact. A closer inspection of the figure shows that for some countries, such as Denmark, the time-bias scatters around zero, whereas it exhibits a persistent trend toward one extreme in others, such as the United Kingdom and Italy. For the latter two countries, the findings indicate that the whole party system has shifted towards the right over time in previously unobservable ways. As a consequence, for example, ignoring the time-bias renders estimates from the British Labour Party of the 1970s incomparable to estimates of the same party in the 1990s, since the whole British party system has moved towards the "third way."

KMO examine the convergent validity with respect to other left-right estimates from the same data sources (CMP and Chapel-Hill expert survey) and conduct two short case studies for Sweden and France. While their approach and findings are plausible, they do not provide further insights as to how well *independent* measures of parties' left-right positions correspond with MCSS.

3 Methods and Data

While the RILE estimates have been subject to several evaluations relative to other expert- and survey data (e.g. recently Bakker et al., 2015), a study on RILE performance relative to MCSS is missing. Since there exists no gold standard for the evaluation of estimated party positions (a set of true party positions), we follow other studies that have examined the validity of party positions using a set of validation measures (a set of imperfect measures of party positions).[10]

We here formalize the logic of this procedure to clarify the assumptions that it involves. We begin with the benchmark case of a gold standard in the form of a vector $\mathbf{Y}$ that collects the (unobserved) true party positions. Our observations are instances of these party positions transformed by some unknown function $f_{\cdot}(\cdot)$ that we call a transformation function. We let $f_M(\cdot)$ and $f_R(\cdot)$ denote the function that maps true party positions to observed MCSS and RILE scores respectively. We then say that MCSS outperforms RILE if $\text{corr}(f_M(\mathbf{Y}), \mathbf{Y}) > \text{corr}(f_R(\mathbf{Y}), \mathbf{Y})$, where $\text{corr}(\mathbf{A}, \mathbf{B})$ is the Pearson correlation coefficient between the bivariate vectors $(\mathbf{A}, \mathbf{B})$.

Unfortunately, $\mathbf{Y}$ cannot be observed directly. However, if we assume that there exists a validation measure that results from applying a linear function, $f_j(\cdot)$, to $\mathbf{Y}$, the two correlations above can be estimated (substituting $f_j(\mathbf{Y})$ for $\mathbf{Y}$) and compared since the Pearson correlation coefficient is invariant to any linear transformation. It is interesting to note that a linear transformation function is entirely consistent with the standard Aldrich-McKelvey scaling (Aldrich and McKelvey, 1977) that estimates party positions from mass-survey data and corrects for differential-item functioning when respondents interpret scales differently.

[9] To estimate the country bias, they invoke the "zero hour" hypothesis, which stipulates that parties pursue the same position in their first European Parliament election in 1979 as in the previous national election.

[10] See, for example, the studies in Electoral Studies' Special Issue "Comparing Measures of Party Positioning: Expert, Manifesto, and Survey Data" (Marks, 2007).

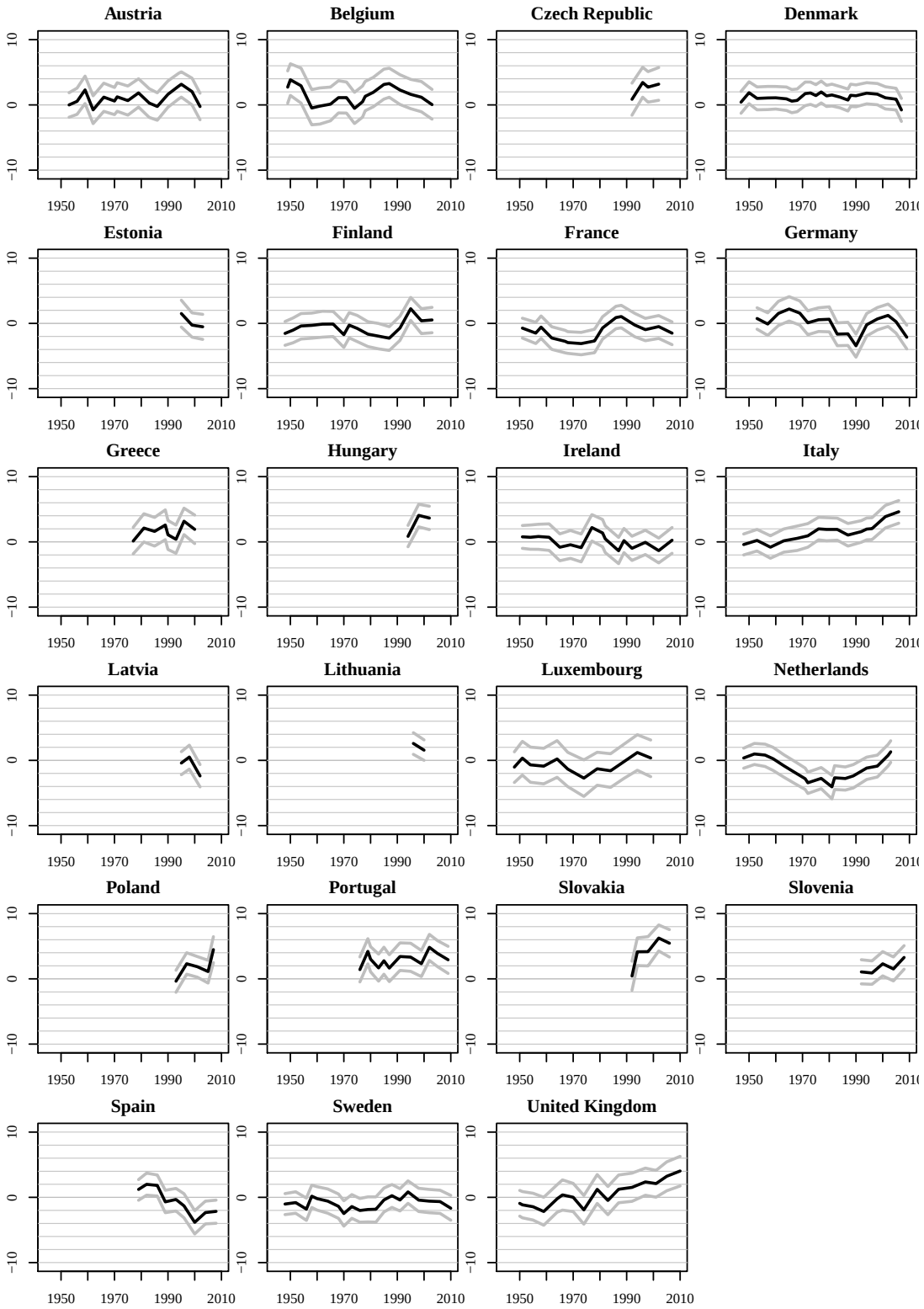

Figure 2: Estimated time-bias parameter from König et al. (2013): Posterior density summary of the time-bias parameter and 95% Bayesian credible interval.

In the case of strongly nonlinear transformations, $\text{corr}(f_M(\mathbf{Y}), f_j(\mathbf{Y})) > \text{corr}(f_R(\mathbf{Y}), f_j(\mathbf{Y}))$ does not necessarily imply $\text{corr}(f_M(\mathbf{Y}), \mathbf{Y}) > \text{corr}(f_R(\mathbf{Y}), \mathbf{Y})$. However, it is typically assumed that if only the minority of a set of validation measures departs from linearity, the majority of comparisons yield an accurate result and consequently allow the analyst to accurately learn which measure performs better. Another approach is to only compare the ranks of the party position estimates (using the Spearman rank correlation) and consequently allow for all nonlinear but monotone transformation functions. In order to take into account that at least some of our validation measures are non-linear transformations of true party positions, we implement both strategies below.

To validate MCSS and RILE, we use validation measures that are estimated from data sources independent of RILE and MCSS (i.e., CMP and the Chapel Hill expert survey). We assemble as many validation measures as possible in a large dataset[11]. This dataset includes Wordfish and Wordscore estimates from raw manifesto texts[12] and estimates from national election studies and Eurobarometer, as well as estimates based on expert survey data. Table 1 provides some descriptive statistics of the raw data for each measure we use in the study, including the MCSS and RILE estimates. Below, we discuss each validation measure briefly and focus on the assumptions required to employ it as a validation measure. Because all measures except for the Eurobarometer estimates have also been used by other authors, we refer the reader to the cited studies for details.

	Mean	S.D.	N	N(Parties)	N(Elections)
MCSS	-0.14	2.32	1832	388	67
RILE	-2.34	21.98	1832	388	67
Wordscore	11.08	4.84	491	87	31
Wordfish	0.02	0.99	104	26	11
NES	5.66	2.05	100	18	18
Eurobarometer	5.25	1.52	476	125	26
Benoit/Laver	10.81	4.81	117	117	5
Castle/Mair	5.14	2.36	75	75	4
Huber/Inglehart	5.40	2.09	93	93	4

Table 1: Descriptive statistics for each left-right measure used in this study. The column "N" refers to the number of observed data points, "N(Parties)" the number of unique parties, and "N(Elections)" the number of unique elections.

Our primary interest concerns the sample differences between alternative validation measures and the two competing measures of parties' left-right positions derived from coded manifesto data (MCSS/RILE). While we focus on sample differences, we also conduct the Meng-Rosenthal-Rubin significance test (Meng and Rubin, 1992) for two overlapping correlations based on dependent groups. However, because this test relies on the assumption of joint normality of the measured variables and large samples, we are cautious in our interpretation of the results and only report them in the appendix (table 3 and 4).[13]

[11] This dataset is available on the authors' websites. We also provide the R-code that assembles the dataset, which might help authors to easily expand the dataset when new measures become available.

[12] We consider the raw-manifesto-text data to be sufficiently independent relative to coded-manifesto data to include them in our analysis and not to rely exclusively on estimates based on surveys.

[13] In the technical appendix, Meng and Rubin (1992) argue that the joint normality assumption "does not appear to be critical in practice." Recent simulation studies along the lines of Hittner et al. (2003) support this argument and also suggest that results based on small samples are not too biased.

3.1 Wordfish and Wordscore

Quantitative text analysis allows us to estimate party positions from the raw text of parties' manifestos. In a nutshell, party positions are generated by first converting the preprocessed party manifesto text into a word-document matrix that captures the word frequency by document. Various models have been suggested to estimate party positions from this matrix, which are essentially variants of the association models developed by Goodman (Goodman, 1979; Lowe, 2016) but differ critically in the amount of information that is included a priori. The Wordfish model (Slapin and Proksch, 2008; Lo et al., 2016) estimates one-dimensional policy positions from a set of texts with minimal prior information. The Wordscore approach instead estimates party positions based on two reference texts that are known to be ideologically different (Laver et al., 2003).

For estimating left-right positions, the primary assumption underlying text analysis is that parties express (signal) their left-right position in their manifesto and that their language usage therein is a function of their left-right ideology. Comparing estimates from these texts over time assumes that there are typical, time-invariant words that are more likely to be used by a left than by a right party and vice versa. Party positions are then a function of the relative usage of these typical words (which are simultaneously identified by the respective model).

We use the Wordfish estimates obtained in the study of Lo et al. (2016) based on the negative-binomial Wordfish model. Their estimates cover four countries (Germany, Ireland, the Netherlands, and Sweden) in the nineties and the beginning of the 2000s. The Wordscore estimates come from the study of Bräuninger et al. (2013), who estimate party positions for most Western European countries, excluding France, Greece, Italy, and Luxembourg, in a period from the early eighties to the mid/late 2000s. They use the manifestos from the most extreme parties in each country as reference texts.

3.2 Eurobarometer and National Election Studies

Mass-survey data can be used to estimate parties' left-right positions in two distinct ways, which we refer to as the indirect and direct approaches. The indirect approach leverages information about a) respondents' ideological self-placement and b) their vote choice in the last election or their party identification. The vote choice information is used to identify respondents as party supporters and then to average their ideological self-placement as an estimate for the left-right position of the supported party. The direct approach, in turn, requires survey items that elicit respondents' perceived party position. An estimate for a party position is obtained by either averaging across all responses or averaging across respondents that report their identification with the respective party (for applications see Adams and Merrill III, 1999, 2000; Adams et al., 2005; Bakker et al., 2015).

Both approaches assume that respondents accurately perceive and place both themselves and the parties. Importantly, it is necessary that all respondents interpret the scale and its endpoints in the same way when answering the survey. If respondents do not interpret the scale in the same way, different self-placements or placements of parties would not necessarily suggest diverging positions but could also be the result of diverging perceptions of the scale.[14] The indirect approach additionally requires an assumption about the interaction of voters and parties. In particular, it is necessary to assume that

[14]This problem is addressed by the above-mentioned Aldrich-McKelvey scaling (Aldrich and McKelvey, 1977); see also Brady (1985); King et al. (2003).

voters choose to identify with and vote for parties in a way that is consistent with the spatial model of voting (Downs, 1957; Hinich and Munger, 1997).

Our estimates constructed with the direct approach are taken from Bawn and Somer-Topcu (2012), who compiled a dataset of national election studies from the United Kingdom, Germany, the Netherlands, Norway, and Sweden for the period from 1971 to 2005. Each of their studies includes a question that asks respondents to locate parties on a 10-point left-right scale. They obtain a measure of party positions by averaging the self-placement of party supporters.

For the indirect approach, we use the Eurobarometer Mannheimer Trendfile (Schmitt et al., 2008) to pool the answers of all respondents in the election year of each country from surveys of the period from 1970 to 2002.[15] This measure is derived from the respondents' self-reported left-right placement on a 10-point scale, averaged across those who intend to vote for a particular party in the next general election.[16]

3.3 Expert Surveys

Expert surveys typically ask insiders, as experts on party politics, to locate the positions of political parties on a left-right scale. The average of the experts' assessments is used as an estimate for the party's position. As in the case of the direct approach with mass-survey data, the primary assumption yet again is that experts locate parties on the same scale. In addition, the set of interviewed experts must be sufficiently knowledgeable to justify the use of the mean as an estimator for the party position. For a broader discussion on expert surveys used to estimate party positions, see, for example, Budge (2000).

We compiled the most prominent expert surveys for our validation study: Benoit and Laver (2006); Huber and Inglehart (1995); Castles and Mair (1984).[17] The Benoit/Laver survey provides us with election-specific party positions across 21 countries from the early 2000s. The other two surveys have no information for the year in which the experts assessed the party positions. We therefore assume that they describe the party position in the year of data collection (1982 and 1993). Accordingly, we match the expert survey estimates with the MCSS/RILE estimates only for the election that is at most three years after the expert survey data collection. The Castles-Mair survey then provides us with estimates from 12 countries, and the Huber-Inglehart estimates from 17 countries.

4 Results

In table 2, we summarize the results from 84 per-country comparisons of MCSS and RILE estimates with our validation measures. Each cell reports the Pearson's correlation coefficient of the estimates. A bold

[15] We converted the "ZEUS" party code from the Eurobarometer file to the CMP party ID using the "Political Parties Link File" by Slava Mikhaylov, available at . We pool the respondents from "Great Britain" and "Northern Ireland," as well as those from "East" and "West" Germany, and also follow the literature in assuming that missing values are completely random (MCAR).

[16] The vote-intention question wording was as follows: "If there were a 'general election' tomorrow ... which party would you support?" We use this question instead of the question about the vote choice in the last election, or the question on party identification, because both of these questions are only included in the waves from 1981/85-1992/94 (party identification) and 1979/82-1995 (last vote choice).

[17] For merging, we again relied on the "Political Parties Link File" by Slava Mikhaylov, as well as the ParlGov Party Table, to link the different datasets (Döring and Manow, 2012). These files were used to convert the different party IDs and match the related party positions for each party.

coefficient indicates a higher correlation for MCSS than RILE estimates. We discuss the results for each validation measure separately.

Across the 11 countries for which we have Wordscore estimates from the study by Bräuninger et al. (2013), MCSS outperforms RILE in eight cases. RILE estimates correlate higher with Wordscore estimates only in the case of Portugal and the United Kingdom, while they perform equally well in Germany. MCSS also correlates highly with Wordfish estimates, and it outperforms RILE strongly in the case of Ireland and the Netherlands. For Germany, RILE performs marginally better, while RILE correlates higher with the Wordfish estimates in the case of Sweden. Overall, in 10 out of 15 comparisons (67%), MCSS estimates correlate higher than RILE estimates with measures based on automated text analysis.

Regarding the similarity to estimates from national election surveys, MCSS outperforms RILE across all four countries. Accordingly, MCSS reflects voters' assessment of the party positions better than RILE does in national elections. A similar picture emerges from the Eurobarometer data, where MCSS outperforms RILE in 13 countries, except for Finland and Sweden. Similar to the national election survey estimates, the differences between the correlations from MCSS and RILE are remarkably high, and on the whole MCSS performs better in 17 out of 19 comparisons with validation measures based on mass-survey data (89%).

Across the three expert surveys, MCSS outperforms RILE in 37 out of 50 comparisons (74%). With respect to the Benoit/Laver estimates, we find that MCSS correlates higher than RILE in 13 out of 20 country comparisons. Interestingly, even in cases in which the correlation between RILE and the Benoit/Laver estimates is extremely low (Estonia, 0.13 and Latvia, 0.11), MCSS correlates much higher (0.7 and 0.9). MCSS also does better in comparison with the Castle/Mair estimates, where MCSS always outperforms RILE in all cases except for Denmark, where the correlations are identical.

The results are less definite for the data from Huber/Inglehart, where MCSS outperforms RILE in 10 out of 17 comparisons. The Hungarian and Slovakian results are particularly peculiar: In Hungary, RILE correlates negatively with the Huber/Ingelhart measure, while MCSS shows a positive correlation; in Slovakia, both measures correlate highly but negatively with the Huber/Inglehart estimates. This suggests that the Huber/Inglehart experts perceive the ordering of Hungarian parties on the left-right dimension very differently than the ordering implied by the estimates from the manifesto data.

Overall, the country-by-country comparisons show that MCSS exhibits a higher convergent validity than RILE. MCSS outperforms RILE in 64 out of 84 cases (76%). Furthermore, in all countries except for Finland, MCSS correlates higher on average than RILE. To the extent that RILE correlates higher with the other validation measures relative to MCSS, the differences between RILE and MCSS are never very large. Since none of our validation measures is a gold standard, these differences are not surprising and are to be expected.

When implementing the analysis using the non-parametric Spearman correlation coefficient, the results are very similar (see appendix table 5). Only when it comes to the data by Huber/Inglehart do we observe that the performance of MCSS and RILE is more similar relative to the results discussed above. However, overall, these results confirm the higher performance of MCSS relative to RILE, even when we only consider the ranking of political parties and ignore their distances on the left-right scale.

	Wordscores		Wordfish		NES		Eurobarometer		Benoit/Laver		Castle/Mair		Huber/Inglehart	
Country	MCSS	RILE	MCSS	RILE	MCSS	RILE	MCSS	RILE	MCSS	RILE	MCSS	RILE	MCSS	RILE
Austria	**0.71**	**0.32**					**0.92**	**0.42**	**0.99**	**0.96**	**0.96**	**0.29**	**0.92**	**0.81**
Belgium	**0.52**	**0.34**					**0.76**	**0.64**	0.76	0.86	**0.86**	**0.64**	**0.81**	**0.62**
Cyprus									0.88	0.97				
Czech Republic									**1.00**	**0.95**			0.44	0.65
Denmark	**0.63**	**0.58**					**0.92**	**0.81**	**0.99**	**0.88**	0.93	0.93	**0.98**	**0.93**
Estonia									**0.70**	**0.13**			**0.99**	**0.89**
Finland	**0.40**	**0.33**					0.71	0.80	0.86	0.91	**0.92**	**0.62**	0.83	0.91
France							**0.91**	**0.88**			**0.94**	**0.92**	**0.99**	**0.99**
Germany	0.68	0.68	0.81	0.82	**0.91**	**0.77**	**0.91**	**0.85**	**0.98**	**0.84**	**0.96**	**0.77**	0.93	0.97
Greece							**0.95**	**0.57**						
Hungary									**0.99**	**0.78**			**0.71**	**-0.23**
Ireland	**0.89**	**0.67**	**0.87**	**0.46**			**0.87**	**0.58**	0.86	0.92	**0.93**	**0.76**		
Italy							**0.86**	**0.45**	0.92	0.92	**0.87**	**0.76**	**0.94**	**0.56**
Latvia									**0.90**	**0.11**				
Lithuania									**0.48**	**0.26**			**0.55**	**0.20**
Luxembourg							**0.74**	**0.62**						
Netherlands	**0.70**	**0.38**	**0.90**	**0.48**	**0.91**	**0.74**	**0.87**	**0.64**	**0.90**	**0.41**	**0.99**	**0.90**	**0.98**	**0.79**
Poland									**0.76**	**0.23**			0.78	0.97
Portugal	-0.04	0.38					**0.84**	**0.41**	**1.00**	**1.00**			0.97	1.00
Slovakia									**0.92**	**0.91**			**-0.86**	**-0.70**
Slovenia									**0.60**	**0.20**				
Spain	**0.67**	**0.57**					**0.84**	**0.57**	**0.80**	**0.80**	**0.83**	**0.75**	0.88	0.90
Sweden	**0.88**	**0.84**	0.74	0.87	**0.92**	**0.87**	0.91	0.95	**0.96**	**0.92**	**0.98**	**0.93**	**0.89**	**0.81**
United Kingdom	0.65	0.69			**0.94**	**0.90**	**0.90**	**0.77**	0.75	0.90	**0.98**	**0.95**		

Table 2: Pearson correlation coefficients for MCSS (RILE) estimates and seven validation measures. Benoit-Laver: Data on France for the left-right position is unavailable. Greece's and Luxembourg's last elections in our data occurred before the Benoit/Laver study and are thus excluded. Huber/Inglehart: Ireland is omitted because the closest election is four years after Huber-Inglehart's data collection.

5 Case Studies: Italy and Greece

Before concluding, we briefly assess the construct validity of MCSS in two critical cases (Greece and Italy) for which other scholars have raised concerns with respect to RILE's construct validity (Pelizzo, 2003; Dinas and Gemenis, 2010). While the results cannot be generalized to other cases, at the very minimum, MCSS should perform much better in these two critical cases to be considered useful to political scientists. We begin by summarizing the criticism from Pelizzo (2003) and Dinas and Gemenis (2010) with respect to RILE and then evaluate whether MCSS provides us with more plausible estimates. Note that in the cases of Italy and Greece, MCSS outperformed RILE across all validation measures in our previous analysis, except for Benoit/Laver's expert survey in the case of Italy, where RILE and MCSS perform equally well.

During the period of study, Greece had essentially a party system with the Panhellenic Socialist Movement (PASOK) and the New Democracy Party (ND) as the dominant parties, capturing almost all votes in the elections. A small percentage share of the votes, around 5-10 percent, went to the two smaller parties to the left: the Communist Party (KKE) and the Progressive Left Coalition (SAP). Figure 3 shows the labeled party positions from all parties included in the CMP for each election.

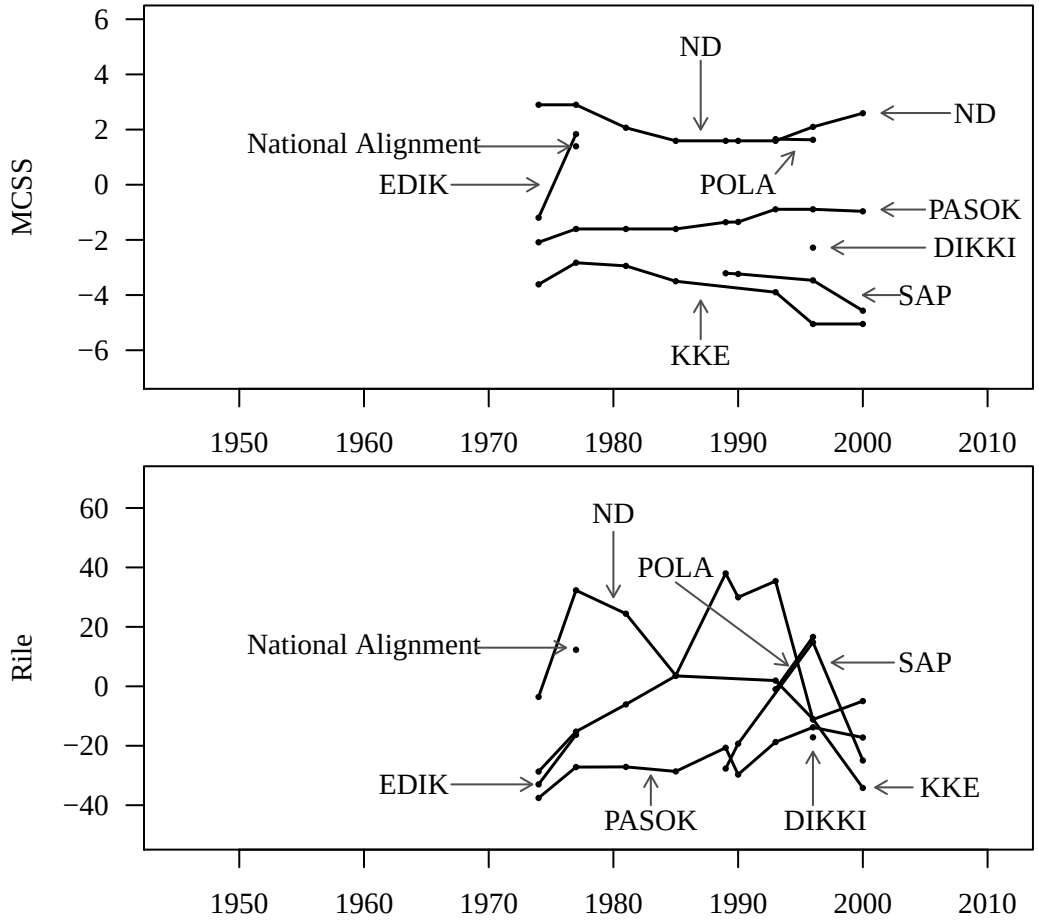

Figure 3: Party Positions for Greece: The upper panel shows the MCSS estimates and the lower panel the RILE estimates.

The study of Dinas and Gemenis (2010) uses the Greek example in their review of various methods to measure party positions from coded manifesto data and criticizes RILE's placement of Greek parties for failing to follow conventional wisdom. In particular, they point out that the relative positions of PASOK, KKE, and SAP do not match the conventional wisdom about these parties' ideological positions. The KKE is typically considered the most extreme party, followed by SP and PASOK. Concerning ND, the authors question both its extreme ideological movement over time and the fact that in 1985 and 1996, RILE places the party at the same position as KKE, essentially implying zero difference between the party farthest to the right and the party farthest to the left in the Greek party system.

Unsurprisingly, MCSS shows a lower ideological volatility in the Greek party system than RILE and can thus overcome the criticism of ND's extreme ideological movement over time indicated by RILE. Furthermore, MCSS locates KKE as the most extreme leftist party in the Greek party system, followed by SAP, while PASOK lies in the middle of the party system with ND occupying the position farthest to the right. This pattern neither changes over time nor do the parties converge into a single ideological position. This suggests that MCSS does indeed exhibit a high construct validity in the case of Greece and addresses the concerns by Dinas and Gemenis (2010) with two estimates from coded manifesto data.

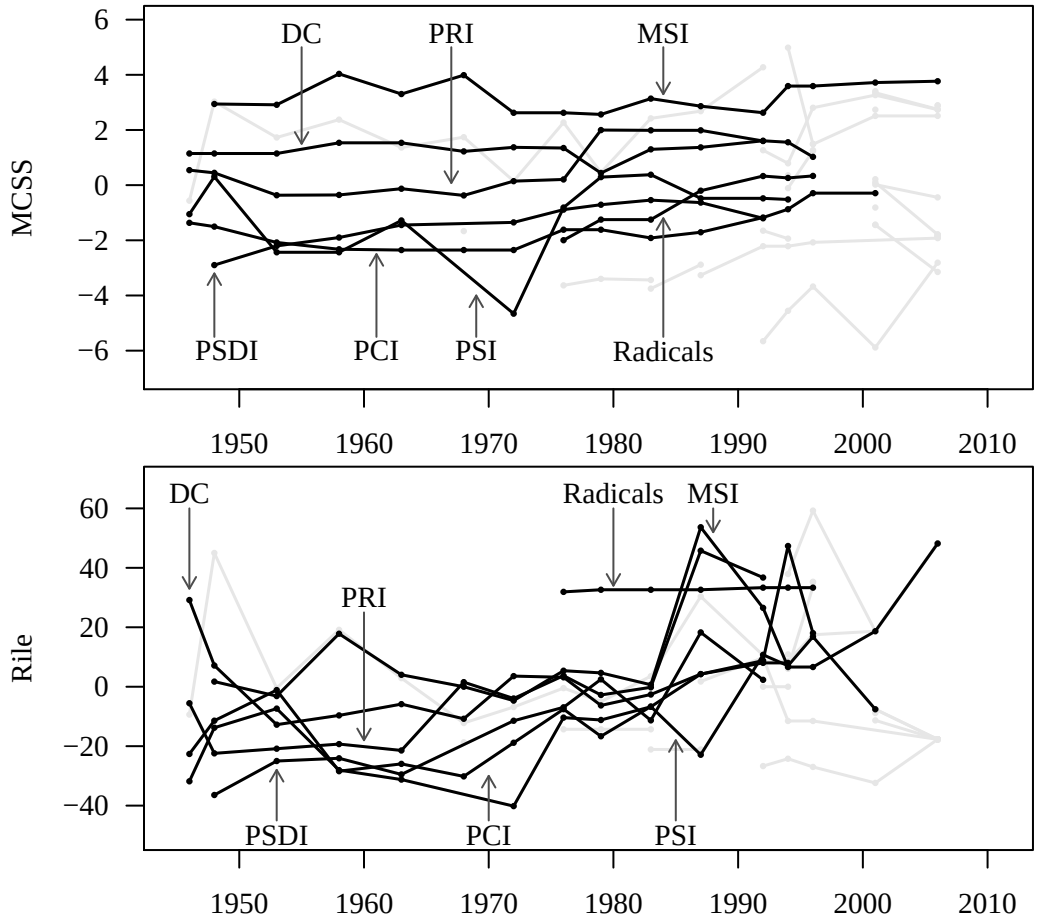

Figure 4: Party Positions for Italy: The upper panel shows the MCSS estimates and the lower panel the RILE estimates. We highlighted the most important parties in black.

Compared to the Greek party system, the Italian is considered a fragile multi-party system that has

undergone significant changes over time, especially during the early 1990s. Pelizzo (2003) discusses the RILE estimates, particularly for the First Republic (1948–87), and also provides some general observations for the Second Republic (1992–96). He notes that the RILE estimates often "do not look very plausible" (Pelizzo, 2003, p.67). He lists a remarkable number of peculiar and leapfrogging patterns that do not square with the conventional wisdom about the Italian party system. In particular, he refers to Sartori (1976), who discusses the Italian party system in more detail.

Following Sartori (1976), the structure of the Italian party system is characterized by main center and extreme parties, with the Christian Democrats (DC) generally located in the center of the party system, the Communist Party (PCI) at the very left, and the neo-Fascist Italian Social Movement (MSI) at the right. However, RILE places DC only at the center in a single election in 1987. PCI is sometimes located as the party farthest to the left but is also placed at the center for a few years, and at the right of the party system for other years. Pelizzo concludes that "It hardly needs noting that this finding is absolutely inconsistent with the results of three decades of studies on the PCI" (Pelizzo, 2003, p.70). Similarly, MSI is not always located as the party farthest to the right, but is sometimes placed side by side with the Communist PCI (in 1948 and 1953). In addition to the leapfrogging of the main Italian parties, Pelizzo also highlights that a number of smaller parties are placed incorrectly. He points out that RILE locates the left Radicals and the left-center PRI always on the right of the spectrum, and that the PSDI, a splinter group of the more moderate PSI party, is placed to the left of the PCI and PSI.

As in the Greek case, MCSS reduces the ideological volatility for most parties. Furthermore, RILE typically places the DC at the center. For Pelizzo's period of study, MCSS always locates the DC between the PCI and MSI. The MSI, in turn, is always placed at the extreme right. Although the PCI is mostly placed at the very left, MCSS shows some overlap with PSI that does not follow Pelizzo's expectation. Only in the late seventies does PSI adjust its positions consistently towards the center. Regarding the smaller parties, we find that MCSS is similar to RILE in placing PSDI inconsistently at the very left, as Pelizzo criticizes. Concerning the PRI and the Radicals, however, MCSS correctly locates both parties in the center of the Italian party system. Overall, MCSS's placement of parties does a much better job of matching the conventional wisdom on the Italian party system. This suggests that, as in the case of the Greek party system, MCSS exhibit a higher construct validity than RILE.

6 Conclusion

RILE estimates are the most frequently used estimates for left-right positions in comparative studies. However, RILE estimates suggest a puzzling ideological volatility (leapfrogging) of political parties over time in several party systems. As a response, some scholars have called into question the usefulness of coded party manifesto data for estimating parties' left-right positions (Benoit and Laver, 2006, 2007). According to KMO, the ideological volatility is the result of the unwarranted assumption that the absolute frequency counts of coded manifesto data and the resulting RILE estimates are comparable over time. While this assumption simplifies the estimation of left-right positions considerably, it ignores that parties write election-specific manifestos to position themselves *relative* to their competitors. KMO introduce a latent variable model with bridge observations to take into account parties' relative positing and to estimate comparable left-right positions, which they refer to as Manifesto Common Space Scores (MCSS).

While MCSS exhibits much less ideological volatility over time and thus seem to overcome the criticism

on RILE, little information existed yet on the relative performance of MCSS as compared to RILE. In this study, we therefore compared the convergent and construct validity of MCSS and RILE. Using a wide range of validation measures from text analysis, expert- and mass-surveys, we find that in 76% of all pairwise comparisons, MCSS estimates correlate higher with the validation measures than RILE estimates. Furthermore, in those instances in which RILE exhibits a higher correlation, the differences between RILE and MCSS are marginal. Finally, by re-analyzing two critical cases (Italy and Greece) we show that MCSS better reflects conventional wisdom about the ideological pattern of these party systems and can be said to exhibit a higher construct validity relative to RILE. Although we do not know whether our construct validity results can be generalized to other countries, they indicate that MCSS estimates can overcome the scholarly criticism on RILE estimates in these two prominent instances.

Overall, our results suggest that MCSS estimates can be recommended when researchers seek to measure left-right positions across time in a comparable fashion. Using bridge observations in a version of a factor analytical model integrates each election-specific left-right scale into a common policy space, which avoids the puzzling leapfrogging of the ideological left/right positions of political parties as observed with RILE estimates and leads to higher convergent validity with measures from other data sources. Using estimates that are less prone to measurement error is important to ensure that coefficient estimates in regression models are not biased towards zero. Although MCSS estimates are certainly not a perfect measure of parties' left-right positions, they are more similar to other validation measures from independent data sources than the RILE estimates. This highlights the importance of taking into account the election-specific character of party manifestos whenever one estimates or uses left-right estimates based on party manifesto data.

References

Adams, J. (2001). A Theory of Spatial Competition with Biased Voters: Party Policies Viewed Temporally and Comparatively. *British Journal of Political Science 31*(1), 121–158.

Adams, J. and S. Merrill III (1999). Party Policy Equilibrium for Alternative Spatial Voting Models: An Application to the Norwegian Storting. *European Journal of Political Research 36*(2), 235–255.

Adams, J. and S. Merrill III (2000). Spatial Models of Candidate Competition and the 1988 French Presidential Election: Are Presidential Candidates Vote-Maximizers? *Journal of Politics 62*(3), 729–756.

Adams, J., S. Merrill III, and B. Grofman (2005). *A Unified Theory of Party Competition: A Cross-National Analysis Integrating Spatial and Behavioral Factors*. New York: Cambridge University Press.

Adams, J. and Z. Somer-Topcu (2009). Policy Adjustment by Parties in Response to Rival Parties' Policy Shifts: Spatial Theory and the Dynamics of Party Competition in Twenty-Five Post-War Democracies. *British Journal of Political Science 39*(4), 825–846.

Aldrich, J. H. and R. McKelvey (1977). A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections. *American Political Science Review 71*(1), 111–130.

Bakker, R. (2009). Re-Measuring Left-Right: A Comparison of Sem and Bayesian Measurement Models for Extracting Left-Right Party Placements. *Electoral Studies 28*(3), 413.

Bakker, R., C. de Vries, E. Edwards, L. Hooghe, S. Jolly, G. Marks, J. Polk, J. Rovny, M. R. Steenbergen, and M. A. Vachudova (2015). Measuring Party Positions in Europe: The Chapel Hill Expert Survey Trend File, 1999-2010. *Party Politics 21*(1), 143–152.

Bawn, K. and Z. Somer-Topcu (2012). Government Versus Opposition at the Polls: How Governing Status Affects the Impact of Policy Positions. *American Journal of Political Science 56*(2), 433–446.

Benoit, K. and M. Laver (2006). *Party Policy in Modern Democracies*. London: Routledge.

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

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

Brady, H. E. (1985). The Perils of Survey Research: Interpersonally Incomparable Responses. *Political Methodology 11*(3–4), 269–90.

Bräuninger, T., M. Debus, and J. Müller (2013). Estimating Policy Positions of Political Actors across Countries and Time. Technical Report 153, Mannheimer Zentrum für Europäische Sozialforschung.

Budge, I. (1999). Estimating Party Policy Preferences: From Ad Hoc Measures to Theoretically Validated Standards. Essex Papers in Politics and Government 139, University of Essex and Department of Government.

Budge, I. (2000). Expert Judgements of Party Policy Positions: Uses and Limitations in Political Research. *European Journal Political Research 37*(1), 103–113.

Budge, I. (2013). Linking Uncertainty Measures to Document Selection and Coding. In A. Volkens, J. Bara, I. Budge, M. D. McDonald, and H.-D. Klingemann (Eds.), *Mapping Policy Preferences from Texts. Statistical Solutions for Manifesto Analysts*, Chapter 7. Oxford University Press.

Budge, I., H.-D. Klingemann, A. Volkens, J. Bara, and E. Tanenbaum (2001). *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments 1945-1998*. Oxford: Oxford University Press.

Budge, I. and M. D. McDonald (2012). Conceptualizing and Measurement and 'Centrism' Correctly on the Left-Right Scale (RILE) - without Systematic Bias. A General Response by MARPOR. *Electoral Studies 31*, 609–612.

Budge, I., M. D. McDonald, and T. Meyer (2013). Validated Estimates Versus Dodgy Adjustments: Focusing Excessively on Error Distorts Results. In A. Volkens, J. Bara, I. Budge, M. D. McDonald, and H.-D. Klingemann (Eds.), *Mapping Policy Preferences from Texts. Statistical Solutions for Manifesto Analysts*, Chapter 4. Oxford University Press.

Budge, I. and T. Meyer (2013). Understanding and Validating the Right-Left Scale (RILE). In A. Volkens, J. Bara, I. Budge, M. D. McDonald, and H.-D. Klingemann (Eds.), *Mapping Policy Preferences from Texts. Statistical Solutions for Manifesto Analysts*, Chapter 5.

Castles, F. G. and P. Mair (1984). Left–Right Political Scales: Some 'Expert' Judgments. *European Journal of Political Research 12*(1), 73–88.

Dalton, R. J. and I. McAllister (2015). Random Walk or Planned Excursion? Continuity and Change in the Left–Right Positions of Political Parties. *Comparative Political Studies 48*(6), 759–787.

Dinas, E. and K. Gemenis (2010). Measuring Parties' Ideological Positions with Manifesto Data: A Critical Evaluation of the Competing Methods. *Party Politics 16*(4), 427–450.

Döring, H. and P. Manow (2012). Parliament and Government Composition Database (ParlGov): An Infrastructure for Empirical Information on Parties, Elections and Governments in Modern Democracies. Version 12/10 – 15 October 2012. http://www.parlgov.org/.

Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper.

Elff, M. (2013). A Dynamic State-Space Model of Coded Political Texts. *Political Analysis 21*(2), 217–232.

Foldal, D. (1989). Sweden. In F. Jacobs (Ed.), *Western European Political Parties: A Comprehensive Guide*, pp. 618–635. Harlow: Longman.

Franzmann, S. and A. Kaiser (2006). Locating Political Parties in Policy Space. *Party Politics 12*(2), 163–188.

Gabel, M. J. and J. D. Huber (2000). Putting Parties in Their Place: Inferring Party Left-Right Ideological Positions from Party Manifestos Data. *American Journal of Political Science 44*(1), 94–103.

Golder, M. (2003). Electoral Institutions, Unemployment and Extreme Right Parties: A Correction. *British Journal of Political Science 33*(3), 525–534.

Goodman, L. (1979). Simple Models for the Analysis of Association in Crossclassifications Having Ordered Categories. *Journal of the American Statistical Association 74*(367), 537–552.

Hanley, D. (1999). France: Living with Instability. In D. Broughton and M. Donovan (Eds.), *Changing Party Systems in Western Europe*, pp. 48–70. London: Pinter.

Hinich, M. and M. Munger (1997). *Analytical Politics*. Cambridge: Cambridge University Press.

Hittner, J. B., K. May, and S. Clayton (2003). A Monte Carlo Evaluation of Tests for Comparing Dependent Correlations. *The Journal of General Psychology 130*(2), 149–168.

Hooghe, L., R. Bakker, A. Brigevich, C. De Vries, E. Edwards, G. Marks, J. Rovny, M. Steenbergen, and M. Vachudova (2010). Reliability and Validity of the 2002 and 2006 Chapel Hill Expert Surveys on Party Positioning. *European Journal of Political Research 49*(5), 687–703.

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

Hug, S. (2001). *Altering Party Systems: Strategic Behavior and the Emergence of new Political Parties in Western Democracies*. Ann Arbor: University of Michigan Press.

Kim, H. and R. C. Fording (2002). Government Partisanship in Western Democracies, 1945-1998. *European Journal of Political Research 41*(2), 187–206.

King, G., C. J. L. Murrary, J. A. Salomon, and A. Tandon (2003). Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research. *American Political Science Review 97*(4), 567–583.

Klingemann, H.-D. and A. Volkens (2007). *Mapping Policy Preferences II: Estimates for Parties, Electors and Governments in Central and Eastern Europe, European Union and Oecd 1990-2003*. Oxford: Oxford University Press.

Klingemann, H.-D., A. Volkens, J. Bara, I. Budge, and M. D. McDonald (2006). *Mapping Policy Preferences II: Estimates for Parties, Electors and Governments in Central and Eastern Europe, European Union and Oecd 1990-2003*. Oxford: Oxford University Press.

König, T. and B. Luig (2012). Party Positions and Legislative Agenda - Estimating Policy Positions for the Study of EU Decision Making. *European Union Politics 13*(4), 604–625.

König, T., M. Marbach, and M. Osnabrügge (2013). Estimating Party Positions across Countries and Time. A Dynamic Latent Variable Model for Manifesto Data. *Political Analysis 21*(4), 468–491.

Lacewell, O. P. and A. Werner (2013). Coder Training: Key to Enhancing Reliability and Validity. In A. Volkens, J. Bara, I. Budge, M. D. McDonald, and H.-D. Klingemann (Eds.), *Mapping Policy Preferences from Texts. Statistical Solutions for Manifesto Analysts*, Chapter 9. Oxford University Press.

Laver, M. (2014). Measuring Policy Positions in Political Space. *Annual Review of Political Science 17*, 207–223.

Laver, M., K. Benoit, and J. Garry (2003). Extracting Policy Positions from Political Texts Using Words as Data. *American Political Science Review 97*(2), 311–331.

Laver, M. and J. Garry (2000). Estimating Policy Positions from Political Texts. *American Journal of Political Science 44*(3), 619–634.

Lipset, S. M. and S. Rokkan (1967). Cleavage Structures, Party Systems and Voter Alignments. In S. M. Lipset and S. Rokkan (Eds.), *Party Systems and Voter Alignments. Cross-National Perspectives*. New York: Free Press.

Lo, J., S.-O. Proksch, and J. B. Slapin (2016). Ideological Clarity in Multiparty Competition: A New Measure and Test Using Election Manifestos. *British Journal of Political Science 46*(3), 591–610.

Lowe, W. (2016). Scaling things we can count. Unpublished.

Lowe, W., K. Benoit, S. Mikhaylov, and M. Laver (2011). Scaling Policy Preferences from Coded Political Texts. *Legislative Studies Quarterly 36*(1), 123–155.

Mair, P. (1999). *Party System Change: Approaches and Interpretations*. Oxford University Press, USA.

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

Martin, L. W. and G. Vanberg (2005). Coalition Policymaking and Legislative Review. *American Political Science Review 99*(1), 93–106.

McDonald, M. D. (2013). Measuring Uncertainty and Error Directly from End Estimates. In A. Volkens, J. Bara, I. Budge, M. D. McDonald, and H.-D. Klingemann (Eds.), *Mapping Policy Preferences from Texts. Statistical Solutions for Manifesto Analysts*, Chapter 6. Oxford University Press.

McDonald, M. D. and I. Budge (2014). Getting it (Approximately) Right (and Center and Left!): Reliability and Uncertainty Estimates for the Comparative Manifesto Data. *Electoral Studies 35*, 67–77.

McDonald, M. D. and S. M. Mendes (2001). The Policy Space of Party Manifestos. In M. Laver (Ed.), *Estimating the Policy Positions of Political Actors*, pp. 90–114. New York: Routlegde.

Meng, Xiao-Li ad Rosenthal, R. and D. B. Rubin (1992). Comparing Correlated Correlation Coefficients. *Psychological Bulletin 111*(1), 172–175.

Meyer, T. (2010). *Party Competition over Time: How Voters and Intra-Party Structure Constrain Party Policy Shifts*. Ph. D. thesis, University of Mannheim.

Mikhaylov, S., M. Laver, and K. Benoit (2012). Coder Reliability and Misclassification in the Human Coding of Party Manifestos. *Political Analysis 20*(1), 78–91.

Pelizzo, R. (2003). Party Positions or Party Direction? An Analysis of Party Manifesto Data. *West European Politics 26*(2), 67–89.

Pierre, J. (2015). Introduction: The decline of swedish exceptionalism? In J. Pierre (Ed.), *The Oxford Handbook of Swedish Politics*, pp. 1–18. Oxford: Oxford University Press.

Ray, L. (2007). Validity of Measured Party Positions on European Integration: Assumptions, Approaches, and a Comparison of Alternative Measures. *Electoral Studies 26*(1), 11–22.

Sartori, G. (1976). *Parties and Party Systems*. Cambridge: Cambridge University Press.

Schmitt, H., E. Scholz, I. Leim, and M. Moschner (2008). The Mannheim Eurobarometer Trend File 1970-2002. GESIS ZA3521 Datenfile Version 2.0.1.

Slapin, J. B. and S.-O. Proksch (2008). A Scaling Model for Estimating Time-Series Party Positions from Texts. *American Journal of Political Science 52*(3), 705–722.

Steenbergen, M. R. and G. Marks (2007). Evaluating Expert Surveys. *European Journal of Political Research 46*(3), 347–366.

Sundberg, J. (1999). The Enduring Scandinavian Party System. *Scandinavian Political Studies 22*(3), 221–241.

Tavits, M. and N. Letki (2009). When Left is Right: Party Ideology and Policy in Post-Communist Europe. *American Political Science Review 103*(4), 555–569.

Vedung, E. (1988). The Swedish Five-Party Syndrome and the Environmentalists. In P. H. Merkl and K. Lawson (Eds.), *When Parties Refuse to Fail: The Case of France*, pp. 76–109. Princeton: Princeton University Press.

Volkens, A. (2007). Strengths and Weaknesses of Approaches to Measuring Policy Positions of Parties. *Electoral Studies 26*(1), 108–120.

	Wordscores			Wordfish			NES			Eurobarometer		
Country	MCSS	RILE	p-value	MCSS	RILE	p-value	MCSS	RILE	p-value	MCSS	RILE	p-value
Austria	**0.71**	**0.32**	**0.008**							**0.92**	**0.42**	**0.003**
Belgium	**0.52**	**0.34**	**0.038**							**0.76**	**0.64**	**0.044**
Cyprus												
Czech Republic												
Denmark	**0.63**	**0.58**	**0.166**							**0.92**	**0.81**	**0.000**
Estonia												
Finland	**0.40**	**0.33**	**0.330**							0.71	0.80	0.794
France										**0.91**	**0.88**	**0.180**
Germany	0.68	0.68	0.507	0.81	0.82	0.613	**0.91**	**0.77**	**0.037**	**0.91**	**0.85**	**0.115**
Greece										**0.95**	**0.57**	**0.000**
Hungary												
Ireland	**0.89**	**0.67**	**0.000**	**0.87**	**0.46**	**0.000**				**0.87**	**0.58**	**0.000**
Italy										**0.86**	**0.45**	**0.000**
Latvia												
Lithuania												
Luxembourg										**0.74**	**0.62**	**0.205**
Netherlands	**0.70**	**0.38**	**0.001**	**0.90**	**0.48**	**0.000**	**0.91**	**0.74**	**0.000**	**0.87**	**0.64**	**0.000**
Poland												
Portugal	-0.04	0.38	0.992							**0.84**	**0.41**	**0.014**
Slovakia												
Slovenia												
Spain	**0.67**	**0.57**	**0.152**							**0.84**	**0.57**	**0.000**
Sweden	**0.88**	**0.84**	**0.126**	0.74	0.87	0.984	**0.92**	**0.87**	**0.059**	0.91	0.95	0.721
United Kingdom	0.65	0.69	0.693				**0.94**	**0.90**	**0.177**	**0.90**	**0.77**	**0.017**

Table 3: Pearson correlation coefficients for MCSS (RILE) estimates and seven validation measures. The p-value is from a Meng-Rosenthal-Rubin test for two overlapping correlations based on dependent groups testing the hypothesis that the MCSS correlation coefficient is larger than RILE.

	Benoit/Laver			Castle/Mair			Huber/Inglehart		
Country	MCSS	RILE	p-value	MCSS	RILE	p-value	MCSS	RILE	p-value
Austria	**0.99**	**0.96**	**0.137**	**0.96**	**0.29**	**0.500**	**0.92**	**0.81**	**0.274**
Belgium	0.76	0.86	0.691	**0.86**	**0.64**	**0.054**	**0.81**	**0.62**	**0.236**
Cyprus	0.88	0.97	0.500						
Czech Republic	**1.00**	**0.95**	**0.500**				0.44	0.65	0.739
Denmark	**0.99**	**0.88**	**0.001**	0.93	0.93	0.532	**0.98**	**0.93**	**0.056**
Estonia	**0.70**	**0.13**	**0.139**				**0.99**	**0.89**	**0.049**
Finland	0.86	0.91	0.672	**0.92**	**0.62**	**0.077**	0.83	0.91	0.774
France				**0.94**	**0.92**	**0.370**	**0.99**	**0.99**	**0.491**
Germany	**0.98**	**0.84**	**0.042**	**0.96**	**0.77**	**0.500**	0.93	0.97	0.705
Greece									
Hungary	**0.99**	**0.78**	**0.500**				**0.71**	**-0.23**	**0.082**
Ireland	0.86	0.92	0.753	**0.93**	**0.76**	**0.500**			
Italy	0.92	0.92	0.504	**0.87**	**0.76**	**0.187**	**0.94**	**0.56**	**0.020**
Latvia	**0.90**	**0.11**	**0.039**						
Lithuania	**0.48**	**0.26**	**0.300**				**0.55**	**0.20**	**0.134**
Luxembourg									
Netherlands	**0.90**	**0.41**	**0.047**	**0.99**	**0.90**	**0.049**	**0.98**	**0.79**	**0.051**
Poland	**0.76**	**0.23**	**0.246**				0.78	0.97	0.500
Portugal	**1.00**	**1.00**	**0.500**				0.97	1.00	0.500
Slovakia	**0.92**	**0.91**	**0.429**				**-0.86**	**-0.70**	**0.500**
Slovenia	**0.60**	**0.20**	**0.196**						
Spain	**0.80**	**0.80**	**0.495**	**0.83**	**0.75**	**0.262**	0.88	0.90	0.541
Sweden	**0.96**	**0.92**	**0.236**	**0.98**	**0.93**	**0.221**	**0.89**	**0.81**	**0.234**
United Kingdom	0.75	0.90	0.667	**0.98**	**0.95**	**0.166**			

Table 4: Pearson correlation coefficients for MCSS (RILE) estimates and seven validation measures. The p-value is from a Meng-Rosenthal-Rubin test for two overlapping correlations based on dependent groups testing the hypothesis that the MCSS correlation coefficient is larger than RILE.

	Wordscores		Wordfish		NES		Eurobarometer		Benoit/Laver		Castle/Mair		Huber/Inglehart	
Country	MCSS	RILE	MCSS	RILE	MCSS	RILE	MCSS	RILE	MCSS	RILE	MCSS	RILE	MCSS	RILE
Austria	**0.74**	**0.32**					**0.81**	**0.51**	**1.00**	**0.80**	**1.00**	**0.50**	0.90	0.90
Belgium	**0.50**	**0.41**					**0.69**	**0.66**	0.71	0.71	**0.87**	**0.73**	**0.79**	**0.43**
Cyprus									0.50	1.00				
Czech Republic									1.00	1.00			0.20	0.54
Denmark	**0.67**	**0.60**					**0.90**	**0.80**	**0.93**	**0.81**	**0.90**	**0.89**	0.93	0.95
Estonia									**0.66**	**0.09**			**1.00**	**0.80**
Finland	**0.37**	**0.37**					0.57	0.79	**0.98**	**0.95**	**0.79**	**0.50**	0.81	0.81
France							**0.88**	**0.77**			**0.94**	**0.83**	0.90	0.90
Germany	**0.71**	**0.67**	0.84	0.86	**0.85**	**0.73**	**0.90**	**0.84**	**1.00**	**0.90**	**1.00**	**0.50**	0.90	1.00
Greece							**0.94**	**0.46**						
Hungary									1.00	1.00			**0.60**	**-0.31**
Ireland	**0.85**	**0.74**	**0.87**	**0.59**			**0.72**	**0.61**	0.77	0.83	0.50	0.50		
Italy							**0.88**	**0.52**	**0.98**	**0.80**	**0.83**	**0.66**	**0.93**	**0.54**
Latvia									**0.71**	**0.26**				
Lithuania									0.11	0.11			**0.54**	**0.09**
Luxembourg							**0.69**	**0.61**						
Netherlands	**0.81**	**0.46**	**0.91**	**0.54**	**0.88**	**0.68**	**0.92**	**0.64**	**0.90**	**0.38**	**1.00**	**0.90**	**1.00**	**0.70**
Poland									**0.70**	**0.40**			0.50	1.00
Portugal	**-0.15**	**0.35**					**0.83**	**0.55**	1.00	1.00			1.00	1.00
Slovakia									**0.89**	**0.79**			**-1.00**	**-0.50**
Slovenia									**0.54**	**0.43**				
Spain	**0.61**	**0.55**					**0.79**	**0.54**	0.70	0.90	**0.85**	**0.65**	0.83	0.83
Sweden	**0.86**	**0.86**	0.68	0.83	**0.92**	**0.81**	**0.93**	**0.86**	**0.96**	**0.89**	**1.00**	**0.90**	**0.81**	**0.52**
United Kingdom	0.66	0.75			**0.88**	**0.85**	**0.85**	**0.76**	0.20	1.00	**1.00**	**0.95**		

Table 5: Spearman's rank correlation coefficients for MCSS (RILE) estimates and seven validation measures.