Measuring cleavage strength

Romain Lachat

Department of political science

University of Montreal

lachat@pwi.unizh.ch

January 2007

Abstract

This paper discusses three distinct problems in the measurement of the strength of social cleavages. First, some properties of the most common measures (Alford, Thomsen, and kappa indices) and some theoretical differences among them are often misunderstood. This paper shows that there are actually only two main strategies for estimating differences in the voting behavior of social groups, relying on either absolute or relative measures, and it discusses their statistical properties. Second, the strength of social cleavages can be affected by two independent processes: changes in the size of social groups (structural dealignment) and changes in their voting behavior (behavioral dealignment). While traditional measures of cleavage strength do not distinguish properly between these two processes, the 'lambda' indices introduced here allow making such a distinction. Finally, the paper suggests improving analyses of cleavage strength by relying on simulation techniques that can be used to estimate the uncertainty surrounding such measures.

Introduction

In recent decades, there has been intense debate in the literature regarding the evolution of social cleavages, focusing especially on the class cleavage. The hypothesis of electoral dealignment argues that social cleavages, and social-structural factors more generally, have lost much of their relevance in structuring voters' choices. Divisions between social classes or religious groups, for example, should weaken in the process of social modernization. As a consequence, voters' choices are less strongly influenced by traditional loyalties. The number of political independents – citizens who do not identify with any party – should increase, and voters should be more strongly influenced by short-term factors.

The dealignment hypothesis, however, has been criticized (e.g., Müller 1997; Evans 1999). Although the changes affecting the social structure, such as the transformation of the employment structure, increases in education level, and secularization, have not really been disputed, their impact on the level of cleavage voting has been reassessed. Dealignment may transform central cleavages, rather than causing them to disappear. In other words, it is expected that an initial phase of dealignment should be followed by a process of realignment, forming new structural divisions and new, stable partisan loyalties. From this point of view, cleavages are still important, but they should also change over time. To show this, however, more detailed classifications of social groups or new measures of cleavage strength must be used. In the case of the class cleavage, many scholars have turned to multi-category class schemas, such as the one developed by Goldthorpe (Erikson and Goldthorpe 1992).

Many analyses have been presented to support either the dealignment or realignment claims. However, comparing the results of different analyses is often difficult, due to the variety of definitions of social groups and measures used to gauge the strength of the corresponding divisions. This study deals with the latter aspect, discussing the merits or drawbacks of various indices of cleavage strength and their ability to capture the main aspects of dealignment and realignment processes. It makes three important arguments.

First of all, this paper shows that the theoretical differences among the most frequently used indices are usually exaggerated. There are actually only two main strategies for estimating differences in the voting behavior of social groups, relying on either absolute or relative differences. The similarities among certain indices have often been overlooked, and the arguments advanced to support or criticize particular indices often concern the underlying theoretical model and the classification of social groups, rather than the statistical properties of these measures.

Second, most analyses do not account for the fact that electoral change may result from two different processes: a change in the voting behavior of citizens or a change in the size of social groups. This distinction between 'behavioral' and 'structural' dealignment is central to explaining and modeling cleavage evolution, but structural changes are not accounted for in many analyses. This is certainly due at least partly to the lack of an adequate measure of cleavage strength, one that is able to distinguish between the respective impacts of the two processes. This paper introduces a new measure of cleavage strength that can make such a distinction.

The third problem is that virtually all analyses ignore the uncertainty surrounding the estimated strength of cleavages. It is suggested here to use simulation techniques that have become more common in political science research over the past few years to compute not only point estimates for various indices of cleavage strength, but also the corresponding confidence intervals.

Structural and behavioral dealignment

The strength or relevance of cleavages can change through two different processes: a change in the voting behavior of social groups that are opposed across a given cleavage, or a change in the size of those groups. Both can contribute to weakening or strengthening the importance of a cleavage, though they are independent processes. The first process simply indicates changes in voting patterns. The traditional class cleavage – the opposition between the working class and the middle class – can weaken if voters in either group become less homogenous in their electoral choices. Part of the working class may turn away from left-wing parties and support a conservative party. Similarly, a religious cleavage can become less salient if differences in the voting behavior of Catholics and Protestants weaken. This form of electoral change is what Flanagan terms ‘sectoral realignment’ (1984: 95) and Zelle ‘dealignment as weakening of intragroup cohesion’ (1998: 55). We designate it as ‘behavioral’ dealignment. The second process of change, termed ‘ecological realignment’ by Flanagan (1984: 96), is labeled here ‘structural’ dealignment.[1] It takes place when the size of the traditional groups of voters changes. In the class cleavage, the opposition between the working class and the middle class will lose importance if the proportion of workers in the electorate diminishes. It is central to emphasize that this aspect of electoral change does not depend on citizens’ voting behavior; even if voting patterns among members of the different social groups do not change, a cleavage can still erode simply because its social-structural basis has changed.

Structural and behavioral dealignments both contribute to changes in the strength of traditional cleavages, and lie at the core of the dealignment hypothesis. A weakening traditional class cleavage can result from both homogenization in the life conditions of social classes and an increase in the size of the new middle class, which is not clearly aligned in the cleavage structure (Dalton 2002). This important distinction is frequently overlooked in

[1] While we refer to both processes of changes as forms of ‘dealignment’, we do not mean that all such changes result in a weakening of cleavages. Transformation of cleavages may also lead to a process of realignment. It would be more precise to speak of behavioural and structural ‘dealignment *or* realignment’ – but we shall use here the shorter expression.

longitudinal analyses of cleavages, which are often limited to behavioral dealignment.[2] This is the case for all measures based on log-odds ratios (discussed and illustrated in the following section). Of course, for some research questions, it may be more appropriate to focus only on behavioral dealignment. If one's goal is to analyze the impact of a given social-structural characteristic, such as race or religious denomination, on voting choices, it makes sense to ignore changes in group size. However, this does not allow a study to answer another equally important question of electoral research, namely, whether the overall relevance of social divisions to explaining voting choices has changed. To this end, both aspects of dealignment must be considered, and as they are independent processes, it must be possible to assess their impact separately. This is all the more true as research on cleavage voting focuses more and more on explaining differences between countries (e.g., Nieuwbeerta and Ultee 1999; Elff 2002b; Nieuwbeerta and Manza 2002; Knutsen 2005). It is important to be able to distinguish between the two forms of dealignment, as they are not likely to be explained by the same factors. In a later section, a new measure of cleavage strength that allows making such a distinction will be introduced. Before this, however, it is necessary to discuss some important characteristics of the measures that are most often used in the literature.

Existing measures of cleavage strength

The most common way to measure the strength of a cleavage is certainly to use the Alford index, computed as the difference in support for left-wing parties between manual and non-manual occupations (Alford 1962). This measure is often referred to in longitudinal analyses of the dealignment hypothesis (e.g., Dalton 2002). While it was designed to assess the strength of the class cleavage, analogous measures can be used for other types of structural oppositions. A simple index of religious voting, for example, can be computed by contrasting

[2] Nieuwbeerta and Manza (2002: 250) recognise the importance of the two dealignment processes, but their analyses do not account for structural changes.

the level of support for Christian parties among frequent churchgoers with support among other voters.

The Alford index is obviously limited, as it focuses only on a dichotomous opposition of social classes and political parties, reflecting the traditional conception of the class cleavage. However, this index has been criticized mainly for another reason: It is affected both by changes in the strength of the association between class and the vote, and by changes in the size of social classes or political parties (e.g., Evans 2000). In other words, it confounds structural and behavioral dealignment. The solution adopted by many scholars is to turn to relative measures of cleavage voting, which assess only the statistical association between class and the vote. In this framework, many scholars have used log-linear or related models (Evans, Heath, and Payne 1991; Hout, Brooks, and Manza 1995; Nieuwbeerta and de Graaf 1999; Nieuwbeerta and Ultee 1999; Weakliem and Heath 1999). Although various measures have been suggested, all share a central characteristic: They are 'margin-free', meaning that they are unaffected by changes in the size of social groups or parties. Thus, they measure only the behavioral aspect of cleavages.

For a simple two-class and two-party model, a relative measure of class voting can be computed by taking the odds ratio, that is, dividing the odds that manual workers support a socialist party by the corresponding odds for non-manual workers. This is the measure used by Evans, Heath, and Payne (1991: 101). However, scholars more usually use a log-odds ratio or 'Thomsen index' instead (Thomsen 1987; Nieuwbeerta and de Graaf 1999; Elff 2002a; Knutsen 2005), which is simply the natural logarithm of the odds ratio.

The distinction between absolute and relative measures is central to the ongoing debate over the evolution of cleavages. Absolute measures, such as the Alford index, rely on differences in the voting probabilities of social groups, while relative measures are based on the odds ratios of voting probabilities. The substantial difference between the two strategies is easiest to grasp when social groups and parties are both dichotomized. In such a case, the Alford index is equivalent to a coefficient from an ordinary least squares (OLS) regression[3] (Korpi 1972), while the Thomsen index is equivalent to a logit coefficient.[4] This correspondence between indices and coefficients makes it quite easy to understand the important differences between the two measures. The value of the Alford index is unaffected by changes in the level of support for left parties, as long as the share of the vote increases or decreases by the same amount in manual and non-manual classes. By contrast, a relative measure of cleavage strength is affected. Similarly, the absolute difference in the share of votes of left parties between two social groups can vary, while the relative difference between their levels of support remains the same.

Although the two types of indices correspond to regression coefficients estimated by different methods, this does not imply that one is always 'better' than the other. It is true that with a dichotomous dependent variable, logistic regression is better than OLS for estimating voting probabilities. How these probabilities are then compared, however, is a different matter. The results obtained for absolute and relative measures are usually very similar to each other. Of course, they cannot be directly compared, as they are expressed in different metrics. But when the strength of a given cleavage is compared between countries or over time, using a relative or an absolute measure usually leads to similar findings, as Nieuwbeerta (1995) shows in his longitudinal analysis of class cleavages in twenty countries. Considering more than 300 elections, he finds that the correlation between the Alford and Thomsen indices is as high as 0.97 (1995: 54). Choosing an absolute or a relative measure leads to substantially different

[3] The Alford index has the same value as an OLS regression coefficient when party choice – coded '1' for socialist and '0' for conservative – is regressed on occupation, coded '0' for non-manual and '1' for manual.

[4] A coefficient from a binary logistic regression is defined as a log-odds ratio (e.g., Long 1997: 51). The logit coefficient and Thomsen index will be equal as long as the difference between the two social groups is coded as one unit.

results only when the distribution of votes across parties is strongly skewed in some social group.

Measures of the strength of cleavages can also be computed with multi-category classifications of social groups, and with more than two parties. One of the most important developments in this respect is the ‘kappa index’, introduced by Brooks, Hout, and Manza (Hout, Brooks, and Manza 1995; Brooks and Manza 1997a, 1997b; Hout, Manza, and Brooks 1999; Manza and Brooks 1999). It is a summary measure of the differences in social groups’ voting behavior in a given election. They first developed it in studying class cleavages, but it has been applied to other cleavages as well, such as those of religion, gender, and race (Manza and Brooks 1999). The index is defined as ‘the standard deviation of class differences in vote choice in a given election’ (Hout, Brooks, and Manza 1995: 813). A central advantage of this measure is that it can be computed with a binomial and a multinomial dependent variable, as well as with different numbers of social groups. It can also be estimated while controlling for other variables. In this way, it is possible to estimate the strength of a cleavage, net of the effects of other cleavages. Furthermore, it is possible to compute either an absolute or a relative kappa index. There are thus two different types of kappa indices. They are maybe best understood as a way to summaries values from several log-odds ratios or differences in probabilities. With two parties but more than two social groups, it is still possible to work with Thomsen or Alford indices, but this involves computing a separate index for each relevant contrast. Brooks, Hout, and Manza suggest summarizing these values by taking their standard deviation. With two parties and four social groups, there are for example three non-redundant log-odds ratios,[5] and the relative kappa index then corresponds to their standard deviation. As coefficients of a binary logistic regression are equal to log-odds ratios, this index can be defined as follows:

[5] More generally, for J parties and S social groups, there are (J – 1)(S – 1) non-redundant log-odds ratios (Powers and Xie 2000: 96).

$$\kappa_{rel} = \sqrt{\frac{\sum_{s=1}^{S}\left(\beta_s^j - \overline{\beta}_s^j\right)^2}{S}}, \qquad [1]$$

where β_s^j is the coefficient from a binary logistic regression for social group *s* and voting outcome *j* (with the β coefficient of the social group chosen as the reference category being equal to 0), and $\overline{\beta}_s^j$ is the average regression coefficient across all S social groups.[6] Alternatively, it is possible to first centre the β_s^j (Hout, Brooks, and Manza 1995: 813; Manza and Brooks 1999: 62). This simplifies the above equation a little bit, as $\overline{\beta}_s^j$ is then equal to zero, but the two procedures lead to identical results. Still, with a dummy dependent variable, the absolute kappa index is defined in a similar way:

$$\kappa_{abs} = \sqrt{\frac{\sum_{s=1}^{S}\left(\pi_s^j - \overline{\pi}_s^j\right)^2}{S}}, \qquad [2]$$

where π_s^j is the probability that a member of social group *s* votes for party *j*, and $\overline{\pi}_s^j = \frac{1}{S}\sum_{s=1}^{S}\pi_s^j$. In other words, it corresponds to the standard deviation of voting probabilities. Such an absolute kappa index has been used by Manza and Brooks (1999: chapter 6) and Nieuwbeerta, de Graaf, and Ultee (2000). It can be extended to a multinomial dependent variable in a fairly straightforward way, by summing across both social groups and parties. This leads to a more general definition of the absolute kappa index, which can be applied to dependent variables with two or more categories:

$$\kappa_{abs} = \sqrt{\frac{\sum_{j=1}^{J}\sum_{s=1}^{S}\left(\pi_s^j - \overline{\pi}_s^j\right)^2}{J \cdot S}}. \qquad [3]$$

[6] The value of this index, of course, does not depend on the choice of the base outcome for the regression. If the dummy dependent variable distinguishes between left-wing and right-wing parties, the relative kappa index has the same value when setting the value 1 for left-wing or for right-wing parties.

Although equations [2] and [3] are very similar, the latter is not a standard deviation, as the term $\overline{\pi}_s^j$ varies from one of the J voting outcomes to the other. Nieuwbeerta and Manza (2002) use this index in their comparative analyses.

Extending the relative kappa index to the multinomial case is less straightforward. It is not possible to extend equation [1] by simply summing across both social groups and parties. This is linked to the difference between odds and relative risks. Odds are defined as the probability of an outcome divided by 1 minus that probability (Long 1997: 51). A relative risk, by contrast, is the probability of an outcome divided by the probability of another outcome. An odds ratio (OR) and a relative risk ratio (RRR) are then simply a ratio of two odds or two relative risks respectively. With only two voting outcomes, these two quantities are the same as π_s^2 (the probability that a member of social group *s* chooses the outcome *j* = 2) is equal to $\left(1-\pi_s^1\right)$. Thus, for any social groups S = 1 and S = 2,

$$OR = \frac{\pi_1^1 \Big/ \left(1-\pi_1^1\right)}{\pi_2^1 \Big/ \left(1-\pi_2^1\right)} = \frac{\pi_1^1 / \pi_1^2}{\pi_2^1 / \pi_2^2} = RRR\,. \quad [4]$$

With a multinomial logistic regression, however, the equality between odds ratios and relative risk ratios does not hold (Borooah 2002: 48f.; Gould 2000: 27). In the latter case, the regression coefficients correspond to the natural logarithm of a relative risk ratio, not to a log-odds ratio. To define the relative kappa index in the multinomial case, it is necessary to start by writing equation [1] again, expressing the regression coefficients in terms of log-odds ratios:

$$\kappa_{rel} = \sqrt{\frac{\sum_{s=1}^{S}\left(LOR_s^j - LOR^j\right)^2}{S}}\,, \qquad [5]$$

where LOR_s^j is the natural logarithm of the ratio of the odds of outcome *j* and of a reference probability,

$$LOR_s^j = \ln\left(\frac{\pi_s^j / \left(1-\pi_s^j\right)}{\pi^{ref} / \left(1-\pi^{ref}\right)}\right), \qquad [6]$$

and where LOR^j is the average value of the s LOR_s^j for outcome j. Equation [5] can be generalized to the multinomial case by summing across both social groups and categories of the dependent variable. The general formula for computing the relative kappa index is then[7]

$$\kappa_{rel} = \sqrt{\frac{\sum_{j=1}^{J}\sum_{s=1}^{S}\left(LOR_s^j - LOR^j\right)^2}{J \cdot S}}. \qquad [7]$$

The choice of the reference probability for computing the log-odds ratios is not important. These reference probabilities do not have to correspond to those of any particular social group included in the model, or to a specific average. As a matter of fact, any probability (different from 0 or 1) can be used.[8] What is important for computing kappa are the standard deviations of the log-odds ratios, not their values.

The values of the absolute and relative kappa indices are expressed in different metrics. The value of the absolute kappa index can range between 0 and 0.5, with a higher value corresponding to larger differences between social groups. However, a value of 0.5 can be reached only when there are two voting outcomes. With a multinomial dependent variable, the highest possible value will be smaller. Figure 1 presents examples of distributions of votes across social groups and parties, as well as the corresponding values of the absolute kappa index.

---

[7] This definition of the relative kappa index differs from the one presented by Gijsberts and Nieuwbeerta (2000: 411), who compute it on the basis of β coefficients. Unless they mean by this log-odds-ratios rather than natural logarithms of relative-risk-ratios, as in a standard multinomial logistic regression, it implies that the value of their index depends on the choice of the reference category for the dependent variable.

[8] The probabilities π^{ref} in both log-odds-ratios cancel each other out. When written in terms of probabilities, the expression $LOR_s^j - LOR^j$ simplifies to $\ln\left(\frac{\pi_s^j}{1-\pi_s^j}\right) - \frac{1}{S}\sum_{s=1}^{S}\ln\left(\frac{\pi_s^j}{1-\pi_s^j}\right)$.

< Figure 1 about here >

These configurations represent the strongest possible polarization, with a number of parties varying between two and six. Though all of these hypothetical cleavages seem to be as strong as possible, the value of the absolute kappa index decreases from 0.50 in the two-by-two case to 0.37 when there are six parties and six social groups. It is especially important to remember this characteristic of the absolute kappa index (or of other absolute indices) when its value is compared among elections with different numbers of parties.

Like the absolute kappa index, the relative kappa index takes a value of 0 when the distribution of votes is the same in all social groups. However, it has no upper bound. While its value is unlikely to be very high in practice, it increases quite sharply as the distribution of votes becomes very unequal in some social groups. This can be seen in figure 2, which presents the value of the relative kappa index in the case of two-by-two configurations, where both parties and social groups are equal-sized. The value of the kappa index is plotted against the distribution of votes in one of the social groups.[9]

< Figure 2 about here >

The value of kappa increases almost linearly, up to a distribution of votes of about 80 to 20 percent. But after this, the increase is exponential. While the values presented here correspond only to a particular configuration of social groups and parties, they show the extreme sensitivity of the relative kappa index (or of other relative indices) to the presence of an extreme distribution of votes in some social groups. The next section discusses this problem in more detail, presenting some important properties of the two categories of indices.

---

[9] In the special case where both parties and social groups have the same size, kappa is a function of a single parameter. If π_1^1 is denoted as *p*, then $p = \pi_1^1 = \pi_2^2 = 1 - \pi_1^2 = 1 - \pi_2^1$, and the relative kappa index can be expressed simply as $\ln\left(\frac{p}{1-p}\right)$.

Properties of absolute and relative measures

One of the main strengths of kappa indices is their flexibility. As they can be computed with different numbers of parties and social groups, it is easy to compare them among elections and party systems. They can also be computed from models that include additional control variables (Hout, Brooks, and Manza 1995; Manza and Brooks 1999). They seem to offer many more possibilities than Alford and Thomsen indices, but the point is not to compare kappa indices with these more traditional measures. As a matter of fact, in a two-party and two-class configuration, the relative kappa index is equivalent to the Thomsen index, and the absolute kappa index is equivalent to the Alford index. More precisely, $\kappa_{rel} = \text{Thomsen index}/2$ and, when the Alford index is expressed as a proportion rather than a percentage, $\kappa_{abs} = |\text{Alford index}|/2$. The Thomsen and Alford indices can thus be seen as special cases of the corresponding kappa indices, where both party choice and social groups are dichotomized. This correspondence has important implications. First, it means that a relative kappa index and an absolute kappa index are very different things. The absolute kappa index does not share the 'margin-free' property of relative indices, contrary to what several authors have argued (Manza and Brooks 1999; Nieuwbeerta, de Graaf, and Ultee 2000: 333f.; Nieuwbeerta and Manza 2002). This also means that, from a statistical point of view, the only important choice is whether to use an absolute or relative measure of cleavage strength. Making this choice requires an awareness of the limitations or potential drawbacks of these two classes of measures.

The first limitation involves the possibility of including additional control variables. With such models, the value of the absolute kappa index depends on the values of the control variables chosen to compute the predicted probabilities, as Manza and Brooks (1999) emphasize. When there are more than two voting outcomes, the relative kappa index is also affected by this choice. This is due, again, to the difference between odds ratios and relative risk ratios. Relative risk ratios do not depend on the values chosen for control variables. This is not the case for odds ratios, unless they are identical, which is only true with a dummy dependent variable. For models including control variables, when using an absolute measure, or a relative one combined with a multinomial dependent variable, one should take care to show how strongly kappa indices vary with different sets of values for the control variables. Relative indices also have an important weakness. As we have seen above, the closer a probability is to 1 (or 0), the sharper the increase (or decrease) in the corresponding log-odds ratios is. When some combination of social category and voting choice has a small number of observations, the value of a relative index will grow very large. This problem has been emphasized by Hout, Brooks, and Manza (1995: 813), who regard this increase in the value of kappa as 'spurious'. They propose to ignore the 'problematic' voting outcome (in their case, third-party candidates in some American presidential elections) and to compute kappa on the basis of the remaining coefficients. Faced with a similar problem, Gijsberts and Nieuwbeerta (2000: 411, 414) choose to exclude social categories in which one of the parties is supported by less than two percent of the observations. Both of these solutions are far from optimal, as the choice of which categories to exclude is rather arbitrary. This is a central problem with relative measures. We have seen above that absolute and relative measures give consistent results as long as the distribution of votes among social groups is not too unequal. The main weakness of absolute measures – their sensitivity to the overall popularity of parties – should thus truly matter only when the strength of parties is very unequal in some social groups. But in such situations, relative measures may also be affected by a weakness of their own. In the most extreme case, it is possible that a party has no supporters at all in a given social group, a problem called quasi-complete separation (Albert and Anderson 1984). With a dichotomous dependent variable and a small number of social groups, quasi-complete separation is only likely to happen with small samples, and should be rather rare when working with standard survey data. However, as the number of voting alternatives or social categories increases, such situations may be encountered more frequently. The consequence is quite clear: The corresponding regression coefficients are infinite in size. While some statistical packages may still provide estimated coefficients, these are ‘almost completely a function of the researcher’s (typically) arbitrary choice of convergence criteria for the estimation routine’ (Zorn 2005: 161). A solution based on penalized maximum likelihood estimation has been proposed in the literature (Firth 1993; Heinze and Schemper 2002; Zorn 2005), which allows one to estimate the underlying parameters in the population. But this does not really solve the problem linked with the value of the relative kappa index, as the corresponding parameters (and log-odds ratios) will still be very large. In such a case, however, absolute measures can still be computed without having to rely on penalized maximum likelihood estimation, as the corresponding probabilities will be equal to 0 or to 1.

Elff (2002a) criticizes another aspect of the kappa indices, noting that because the indices measure all aspects of a given cleavage, their values are affected by changes that may be irrelevant. ‘A decrease of differences between farmers and the service class concerning the choice between green and liberal parties will show up as a decrease in class voting in terms of *Kappa* as well as a decrease of differences between workers and self-employed with respect to the choice between workers’ parties and conservative parties. Thus, changes of *Kappa* measures may also reflect changes in voting behavior that concern anything but class voting in a clear-cut sense’ (Elff 2002a: 10). Elff suggests forming only two groups of parties, and measuring cleavage strength with the help of log-odds ratios. While a kappa index may well be affected by potentially irrelevant differences between social groups, the problem is not linked with the measure itself, but with the definition of the social groups. As Hout, Brooks, and Manza (1995: 814) emphasize, their measure can gauge ‘total class voting’ and can be computed without making any assumptions about the expected relations between parties and social classes. However, their measure can also be applied when focusing on specific associations, such as in the traditional model of class voting.

More important, as Nieuwbeerta (1995: 109; Nieuwbeerta and de Graaf 1999: 29) argues, the estimated log-odds ratio – and thus the value of kappa – may not be robust. The social groups for which they are estimated vary in size, and the uncertainty surrounding the estimated values of the log-odds ratios will be larger for small groups. '[C]alculating log-odds ratios in countries where some classes are small, and where we have a dataset with only a small number of respondents, yields unreliable estimates of the log-odds ratios. Consequentially, this also yields unreliable estimates of the kappa index, and possibly biased descriptions of between-country and over-time differences in levels of […] class voting' (Nieuwbeerta 1995: 109). To avoid this problem, Nieuwbeerta turns to the 'uniform difference model' developed by Erikson and Goldthorpe (1992) and Xie (1992), which is a particular type of log-linear model (see also Evans 2000: 408). This model is much more restrictive, however. It assumes that the relative 'distances' between social groups – that is, the differences in their voting choices – are the same in all years and countries, and that these differences are simply contracted or expanded by a parameter to be estimated.[10] The problem linked with sampling error has also been emphasized by Knutsen (2003), who avoids it by merging very small categories with similar groups.

The best solution to this critical problem is certainly not to turn to more restrictive models, or to change the classification of social groups. As a matter of fact, all measures of cleavage strength represent estimates, characterized by some degree of uncertainty about their exact value. Though the problem may be more acute when some combinations of social group and voting outcome include only a small number of observations, it affects all summary measures of cleavage strength. However, this uncertainty in the parameters of the model can be accounted for when computing a kappa index. Following King, Tomz, and Wittenberg (2000), it has become more common in political science research to present results from statistical models in the form of predicted probabilities and their confidence intervals. The same can be

[10] This parameter is called δ, and the corresponding index is named the 'delta index'.

done when computing measures of cleavage strength. As they are derived from the parameters of a regression model, it is possible to simulate the distribution of these parameters after estimation. For each set of simulated parameters, a predicted value can be computed for each voting outcome, along with the corresponding indices of cleavage strength. This gives information about the distribution of these indices, and allows one to compute their average values and confidence intervals.[11] An example of how to do this for different indices will be presented below. Before this, however, a last problem must be considered.

A new index of cleavage strength

As discussed above, cleavages are influenced by two different processes of dealignment, structural and behavioral. However, none of the indices discussed here can distinguish between them appropriately. Either only behavioral dealignment is accounted for, or the effects of both are confounded.

This paper proposes a new index that makes it possible to examine the evolution of a cleavage's strength due to behavioral and structural changes, or to behavioral changes only. This 'lambda' index is a more general and flexible version of kappa. Like the kappa, it can be computed by focusing either on absolute or relative differences between social groups. The key to doing this is to weight the differences between probabilities or log-odds ratios by the size of the corresponding social groups and parties.[12] In the absolute case, this index is defined as follows:

---

[11] An alternative strategy would be to rely on bootstrapping to estimate the distribution of the regression parameters and the corresponding distribution of the values of the indices (King, Tomz, and Wittenberg 2000: 352).

[12] To avoid any confusion, it must be emphasised that this does not mean including weights in the regression model. Kappa and lambda indices are derived from the same regression model.

$$\lambda_{abs} = \sqrt{\sum_{j=1}^{J}\sum_{s=1}^{S}\omega_j\omega_s\left(\pi_s^j - \overline{\pi}_s^j\right)^2}\ , \qquad [8]$$

where ω_j is the estimated vote share of party *j*, ω_s is the proportion of voters belonging to social group *s*, π_s^j is the probability that a member of social group *s* supports party *j*, and $\overline{\pi}_s^j$ is equal to $\sum_{s=1}^{S}\omega_s\pi_s^j$. Similarly, the relative lambda index is a more general version of the corresponding kappa index and is defined as:

$$\lambda_{rel} = \sqrt{\sum_{j=1}^{J}\sum_{s=1}^{S}\omega_j\omega_s\left(LOR_s^j - LOR^j\right)^2}\ , \qquad [9]$$

where ω_j and ω_j are the same as above, LOR_s^j is the log-odds ratio of outcome *j* versus a reference outcome (as in equations [5] and [6]), and $LOR^j = \sum_{s=1}^{S}\omega_s LOR_s^j$. These two lambda indices share all of the advantages of the corresponding kappa indices. They can be computed with different numbers of categories and parties, allow one to include additional control variables, and can be easily compared between cleavages, over time, and among countries. The kappa and lambda indices differ in how they summaries several log-odds ratios or probability differences. With lambda indices, deviations from the average distribution of votes do not contribute equally to the final index of cleavage strength. They are weighted by the sizes of the corresponding parties and social groups. With kappa indices, the sizes of parties and of social groups are not relevant. In fact, the two kappa indices are only special cases of the more general lambda indices, which can be obtained by setting the following two constraints: $\omega_j = 1/J$ and $\omega_s = 1/S$. In this way, the structural part of the cleavage is ignored, and the value of the indices is only affected by changes in the voting behavior of the social groups.

However, lambda indices allow one to measure the impact of behavioral dealignment, net of structural changes, in a different and potentially more interesting way. If the sizes of social groups are taken into account, as suggested, they can also be held constant over time to isolate the effects of the behavioral component of the dealignment process (e.g., Lachat 2003, 2004). These different results can then be compared easily, as they are derived from the same general index, and expressed in the same metric. Letting the size of groups vary while holding constant the distribution of votes, by contrast, is more problematic. It means constraining the regression coefficients to be equal to some set of values. The standard deviations of the coefficients would not be estimated. While point estimates for the indices of cleavage strength could still be computed, their distribution could not be simulated.

The absolute lambda index can range between 0 and 0.5. However, contrary to the absolute kappa, the highest possible value is not strictly limited by the number of categories for the dependent variable. The reason is that the contributions of the different voting outcomes to the value of the index are weighted by size. Even with four or five parties, a value close to 0.5 could theoretically be reached, if two of them receive each almost half of the votes. The relative lambda index, finally, like the corresponding kappa index, has no upper bound. It is also sensitive to an extreme distribution of votes in some social category. But this problem is less acute, at least if the 'problematic' categories are small, as their impact on the value of lambda is proportional to their size.

An illustration: the transformation of the religious conflict in Germany

The differences between absolute and relative indices, as well as between kappa and lambda indices, will be illustrated with the example of the religious cleavage in Germany. While this cleavage was traditionally based on religious denominations, opposing Catholics to Protestants, several authors have argued that it was increasingly taking the form of a purely religious cleavage, that is, opposing religious and non-religious citizens (Wolf 1996; Elff 2002b: 281; for a similar hypothesis in other contexts, see also Geissbühler 1999; Nieuwbeerta and Manza 2002: 252). This is a typical example of a possible transformation of a traditional cleavage leading to electoral realignment. Although the traditional cleavage has probably become less salient, the evolution of the combined impact of voters' religious denomination and religiosity is rather unclear. Following the hypothesis of a transformation of this cleavage, the degree of polarization may actually increase as the denominational cleavage is replaced by a religious one. However, structural changes are also crucial here. While there may be an increasing gap in the voting choices of religious and non-religious voters, we also know that the process of secularization leads to a strong increase in the proportion of 'non-devout' Christians (e.g., Knutsen 2004: 86ff.).

To see how these contrasting evolutions affect cleavage strength, we consider a multivariate model of voting choice, estimated for each election from 1969 to 2002.[13] To keep the model as simple as possible, this analysis is based only on respondents from (former) West Germany. Voting choice is a dichotomous variable, opposing the Social Democratic Party (SPD) to the Christian Democratic Union (CDU/CSU), and is regressed on a series of dummy variables representing religious groups. The following five categories are distinguished: Catholics with a high level of church attendance, Catholics with a low level of church attendance, two equivalent categories for Protestants, and a fifth group for respondents with no religion or of another religion.[14] The model, estimated with a logistic regression, can be written formally as follows:

$$y_{ij} = \beta_{0j} + \sum_{s=1}^{S-1} \beta_{sj} X_{is} + \varepsilon_{ij}, \qquad [10]$$

where y_i is the logistic transformation of the probability that person *i* supports party *j*, and X_{is} represents dummy variables for four of the five religious groups. The model is analyzed

[13] The list of datasets used for this analysis can be found in the appendix.

[14] Catholics and Protestants with a high level of church attendance are those who go to church 'almost every Sunday' or more frequently (1972 to 1994 election studies), or 'at least once a month' (1969, 1998, and 2002 election studies).

separately for each election to not impose any constraint on the evolution of the differences between religious groups. The estimated parameters were used to compute four indices of cleavage strength, the values of which are presented in figure 3.[15,16]

< Figure 3 about here >

The general pattern of results is the same for all four measures. In a few words, it seems that the cleavage has weakened from the 1970s to the 1990s – but not in a gradual way. Rather, change was more sudden between the 1980 and 1983 elections. The cleavage was already weak in 1969, at about the same level as in the 1990s, but it is difficult to say whether this election is an exception, or whether the period from 1972 to 1980 was characterized by a particularly strong religious cleavage. Still, the comparison reveals interesting differences between the kappa and lambda indices. The lambda indices fluctuate less over time and have smaller standard deviations than the kappa indices do. As the kappa indices give the same importance to all religious groups, their estimated values are more strongly affected by the voting behavior of small groups, and they are characterized by a higher degree of uncertainty. The kappa values do not point unequivocally to a weakening of the cleavage; the point estimate of the 1994 absolute kappa index lies within the confidence intervals of the estimates of the 1972 to 1980 period. Similarly, the values of the relative kappa index in 1987, 1994, and 2002 lie within the confidence intervals of 1976 and 1980.

Differences between absolute and relative measures, on the other hand, are not very pronounced. Their values are expressed in different metrics and cannot be directly compared,

[15] The results from the estimation of these models and of the following ones are not presented here for reasons of space, but are available from the author.

[16] The indices and confidence intervals were computed using a program written for Stata, available from the author.

but both type of indices lead to similar patterns of results.[17] However, as explained above, relative measures may become problematic when social groups or parties are small. To illustrate this with a concrete example, the model of voting choice has been estimated again, distinguishing among four parties instead of two by adding Green and liberal (FDP) voters. This model has the same structure as above, and was estimated with multinomial logistic regressions separately for each election. As far as absolute measures are concerned, the results presented in figure 4 are similar to those of the previous analysis. The values are simply a little bit lower here, and the variations less pronounced, but it must be remembered that the highest possible value of absolute indices becomes smaller when more voting outcomes are distinguished.

< Figure 4 about here >

Relative indices, however, are more problematic: They cannot be computed for four elections, as there are no observations for some combinations of party and social group.[18] Furthermore, the relative kappa index presents quite a different pattern. Its value is highest in 1969 and second-weakest in 1972. The estimated confidence intervals are also much larger than they are for other indices. Clearly, the value of this index is overly sensitive to the presence of small groups. The relative lambda index, by contrast, does not seem to be affected.

Finally, lambda indices can be used to compare the effects of behavioral and structural dealignment. The values presented in figure 3 combined both sources of change. Figure 5 compares them with values of the lambda indices obtained when only behavioral changes are considered. For this, the sizes of the religious groups are kept constant, at the average of the 1969 and 1972 elections.

[17] As relative indices have no upper bound, there is no ‘natural’ range of values to choose for the y-axis. The scale chosen here parallels as closely as possible the evolution of the corresponding absolute indices.

[18] No Protestant with a high level of church attendance voted for the Greens in 1980 or 1990, or for the FDP in 1998. In 1983, the FDP had no supporter in the category ‘other religion/no religion’.

< Figure 5 about here >

The differences between the two series of values are relatively small – and there are no differences between the absolute and relative measures. We can simply observe that cleavage weakening is more pronounced when both sources of changes are accounted for. We thus find no strong evidence for the realignment process postulated by some authors. In recent elections, a divergence between the two series of results appears only in 1994.[19] This may be due to a stronger polarization between religious groups, which is cancelled out by structural changes. However, this does not mark the beginning of a trend, as the divergence is again much smaller in 1998 and 2002.

Conclusion

Four types of indices of cleavage strength have been identified in this paper. Two of them represent the actual standards in the literature: the absolute and relative kappa indices. As we saw, these include not only the measures proposed by Brooks, Hout, and Manza, but also the more restrictive Alford and Thomsen indices. The other two are the absolute and relative lambda indices, which were introduced here. They correspond to more general and flexible formulations of the corresponding kappa indices. However, none of these measures represents a ‘one-size-fits-all’ index of cleavage strength. Each of them has specific characteristics that make it more or less appropriate for different research questions.

More precisely, the four indices differ from each other on two dimensions: which aspects of dealignment or realignment processes they consider (i.e., only behavioral changes for kappa indices; both structural and behavioral for lambda indices), and whether they compare voting choices of social groups in a relative or absolute way. The latter distinction has received much attention in the literature, with most authors advocating using relative measures. However,

[19] It is the only election where the value of the index (both absolute and relative) obtained when controlling for structural changes, does not lie within the 95 per cent confidence interval of the first series of values.

some scholars have already emphasized that, in practice, the two sets of measures lead in most cases to consistent results. They are likely to diverge only when a party receives a very small share of votes in one social group. Yet, this paper has shown that in such cases, relative measures may be problematic as well. Their value can increase sharply for small changes in vote distribution, and they are undefined when a party has no supporters in a given social group. Given these weaknesses, it seems a better idea always to compute both absolute and relative measures – and to compute not only point estimates, but also confidence intervals. The choice between kappa and lambda indices rests on different considerations. The ability to distinguish between both sources of dealignment is an attractive property of lambda indices. We have seen in the case of the religious cleavage in Germany how lambda indices can be used to isolate the impact of behavioral dealignment and compare it with the overall impact of dealignment and realignment processes. Such comparisons are essential to answer some important questions about electoral change. They allow one to make more detailed comparisons of cross-national differences in the strength and evolution of social cleavages. Yet, from perhaps a more sociological perspective, it can be more pertinent to ignore structural changes. In such a case, using kappa indices would be more appropriate.

Appendix: Data sources

The 1969 to 2002 German election studies, analyzed in this paper, are available through the Central Archive for Empirical Social Research at the University of Cologne. The name and reference number at the Central Archive of the datasets used here are listed below:

- ‘Bundestagswahl 1969’ (ref.: 0426)
- ‘Wahlstudie 1972’ (ref.: 0635)
- ‘Wahlstudie 1976’ (ref.: 0823)
- October survey of the ‘Wahlstudie 1980’ (ref.: 1053)
- ‘Wahlstudie 1983’ (ref.: 1276)
- ‘Wahlstudie 1987’ (ref.: 1537)
- ‘Wahlstudie 1990’ (ref.: 1919)
- ‘Nachwahlstudie zur Bundestagswahl 1994’ (ref.: 2601)
- ‘Politische Einstellungen, politische Partizipation und Wählerverhalten im vereinigten Deutschland’ (ref.: 3066)
- ‘Politische Einstellungen, politische Partizipation und Wählerverhalten im vereinigten Deutschland 2002’ (ref.: 3861)

References

Albert, A. and J.A. Anderson (1984). ‘On the Existence of Maximum Likelihood Estimates in Logistic Regression Models’, *Biometrika* 71(1): 1–10.

Alford, Robert R. (1962). ‘A Suggested Index of the Association of Social Class and Voting’, *Public Opinion Quarterly* 26(3): 417–425.

Borooah, Vani K. (2002). *Logit and Probit: Ordered and Multinomial Models*. Sage University Papers series on Quantitative Applications in the Social Sciences, 07–138. Thousand Oaks: Sage.

Brooks, Clem and Jeff Manza (1997a). ‘Class Politics and Political Change in the United States, 1952–1992’, *Social Forces* 76(2): 379–408.

Brooks, Clem and Jeff Manza (1997b). ‘Social Cleavages and Political Alignments: U.S. Presidential Elections, 1960 to 1992’, *American Sociological Review* 62: 937–946.

Dalton, Russell J. (2002). *Citizen Politics: Public Opinion and Political Parties in Advanced Industrial Democracies*. Third Edition. New York: Chatham House.

Elff, Martin (2002a). *An Orthodoxy Disrupted: Cleavage Voting in Western Europe*. Paper presented at the Annual Meeting of the American Political Science Association, Boston, August 29–September 1, 2002.

Elff, Martin (2002b). ‘Parteiensystem, Sozialstruktur und Wahlabsicht. Der Wandel im Zusammenhang zwischen Sozialstruktur und Wahlabsicht in sieben europäischen Nationen und die programmatische Polarisierung zwischen Parteienlagern, 1975–1998’, in Frank Brettschneider, Jan van Deth and Edeltraud Roller (eds). *Das Ende der politisierten Sozialstruktur?* Opladen: Leske + Budrich, pp. 279–313.

Erikson, Robert and John H. Goldthorpe (1992). *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon Press.

Evans, Geoffrey (ed) (1999). *The End of Class Politics? Class Voting in Comparative Context*. Oxford: Oxford University Press.

Evans, Geoffrey (2000). ‘The Continued Significance of Class Voting’, *Annual Review of Political Science* 3: 401–417.

Evans, Geoffrey, Anthony Heath and Clive Payne (1991). ‘Modelling Trends in the Class/Party Relationship 1964–87’, *Electoral Studies* 10(2): 99–117.

Firth, David (1993). ‘Bias Reduction of Maximum Likelihood Estimates’, *Biometrika* 80(1): 27–38.

Flanagan, Scott C. (1984). ‘Patterns of Realignment’, in Russell J. Dalton, Scott C. Flanagan and Paul Allen Beck (eds). *Electoral Change in Advanced Industrial Democracies: Realignment or Dealignment?* Princeton: Princeton University Press, pp. 95–103.

Geissbühler, Simon (1999). ‘Are Catholics Still Different? Catholicism, Political Attitudes and Party Attachments in Switzerland, 1970–95’, *West European Politics* 22(3): 223–240.

Gijsberts, Mérove and Paul Nieuwbeerta (2000). ‘Class Cleavages in Party Preferences in the New Democracies in Eastern Europe: A Comparison with Western Democracies’, *European Societies* 2(4): 397–430.

Gould, William (2000). ‘sg124: Interpreting logistic regression in all its forms’, *Stata Technical Bulletin* 53: 19-29.

Heinze, Georg and Michael Schemper (2002). ‘A solution to the problem of separation in logistic regression’, *Statistics in Medicine* 21: 2409–2419.

Hout, Michael, Clem Brooks and Jeff Manza (1995). ‘The Democratic Class Struggle in the United States, 1948–1992’, *American Sociological Review* 60(6): 805–828.

Hout, Michael, Jeff Manza and Clem Brooks (1999). ‘Classes, Unions, and the Realignment of US Presidential Voting, 1952–1992’, in Geoffrey Evans (ed.). *The End of Class Politics? Class Voting in Comparative Context*. Oxford: Oxford University Press, pp. 83–96.

King, Gary, Michael Tomz and Jason Wittenberg (2000). ‘Making the Most of Statistical Analyses: Improving Interpretation and Presentation’, *American Journal of Political Science* 44(2): 347–361.

Knutsen, Oddbjørn (2003). *Overall and traditional left-right class voting in eight West European countries: A comparative longitudinal study 1975–97*. Paper presented at the Joint Sessions of the European Consortium of Political Research, Edinburgh, March 28–April 2, 2003.

Knutsen, Oddbjørn (2004). *Social Structure and Party Choice in Western Europe: A Comparative Longitudinal Study*. Houndmills: Palgrave Macmillan.

Knutsen, Oddbjørn (2005). *Class voting in eight West European countries: Trends and explanations*. Paper presented at the First Conference of the European Association of Survey Research, Barcelona, July 18–22, 2005.

Korpi, Walter (1972). 'Some Problems in the Measurement of Class Voting', *American Journal of Sociology* 78(3): 627–642.

Lachat, Romain (2003). *Social structure and electoral choice: The impact of the class and religious cleavages in German elections, 1969–1998*. Paper prepared for the Annual Meeting of the Midwest Political Science Association, Chicago, April 3–6, 2003.

Lachat, Romain (2004). *A Heterogeneous Electorate: Individual-Level Differences in the Process of Formation of Voting Choices in Germany and Switzerland*. PhD Thesis. University of Zurich.

Long, J. Scott (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage.

Manza, Jeff and Clem Brooks (1999). *Social Cleavages and Political Change: Voter Alignment and U.S. Party Coalitions*. Oxford: Oxford University Press.

Müller, Walter (1997). 'Sozialstruktur und Wahlverhalten: Eine Widerrede gegen die Individualisierungsthese', *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 49(4): 747-760.

Nieuwbeerta, Paul (1995). *The Democratic Class Struggle in Twenty Countries, 1945–1990*. Amsterdam: Thesis Publishers Amsterdam.

Nieuwbeerta, Paul and Nan Dirk de Graaf (1999). 'Traditional Class Voting in Twenty Postwar Societies', in Geoffrey Evans (ed.). *The End of Class Politics? Class Voting in Comparative Context*. Oxford: Oxford University Press, pp. 23–56.

Nieuwbeerta, Paul, Nan Dirk de Graaf and Wout Ultee (2000). ‘The Effects of Class Mobility on Class Voting in Post–War Western Industrialized Countries’, *European Sociological Review* 16(4): 327–348.

Nieuwbeerta, Paul and Jeff Manza (2002). ‘Klassen-, Religions- und Geschlechterspaltungen: Parteien und Gesellschaften in vergleichender Perspektive’, in Frank Brettschneider, Jan Van Deth and Edeltraud Roller (eds). *Das Ende der politisierten Sozialstruktur?* Opladen: Leske + Budrich, pp. 247–278.

Nieuwbeerta, Paul and Wout Ultee (1999). ‘Class voting in Western industrialized countries, 1945–1990: Systematizing and testing explanations’, *European Journal of Political Research* 35: 123–160.

Powers, Daniel A. and Yu Xie (2000). *Statistical Methods for Categorical Data Analysis*. San Diego: Academic Press.

Thomsen, Søren Risbjerg (1987). *Danish Elections 1920–79: A Logit Approach to Ecological Analysis and Inference*. Aarhus: Politica.

Weakliem, David L. and Anthony F. Heath (1999). ‘The Secret Life of Class Voting: Britain, France, and the United States since the 1930s’, in Geoffrey Evans (ed.). *The End of Class Politics? Class Voting in Comparative Context*. Oxford: Oxford University Press, pp. 97–136.

Wolf, Christof (1996). ‘Konfessionelle versus religiöse Konfliktinie in der deutschen Wählerschaft’, *Politische Vierteljahresschrift* 37(4): 713–734.

Xie, Yu (1992). ‘The Log-Multiplicative Layer Effect Model for Comparing Mobility Tables’, *American Sociological Review* 57(3): 380–395.

Zelle, Carsten (1998). ‘A Third Face of Dealignment? An Update on Party Identification in Germany, 1971–94’, in Christopher Anderson and Carsten Zelle (eds). *Stability and Change in German Elections*. Westport: Praeger, pp. 55–70.

Zorn, Christopher (2005). ‘A Solution to Separation in Binary Response Models’, *Political Analysis* 13(2): 157–170.

parties

groups	a	b
i	1	0
ii	0	1

kappa = 0.50

groups	a	b	c
i	1	0	0
ii	0	1	0
iii	0	0	1

kappa = 0.47

groups	a	b	c	d
i	1	0	0	0
ii	0	1	0	0
iii	0	0	1	0
iv	0	0	0	1

kappa = 0.43

parties

groups	a	b	c	d	e
i	1	0	0	0	0
ii	0	1	0	0	0
iii	0	0	1	0	0
iv	0	0	0	1	0
v	0	0	0	0	1

kappa = 0.40

groups	a	b	c	d	e	f
i	1	0	0	0	0	0
ii	0	1	0	0	0	0
iii	0	0	1	0	0	0
iv	0	0	0	1	0	0
v	0	0	0	0	1	0
vi	0	0	0	0	0	1

kappa = 0.37

Figure 1. Most strongly polarized distribution of votes for different numbers of social groups and of parties, and corresponding values of the absolute kappa index

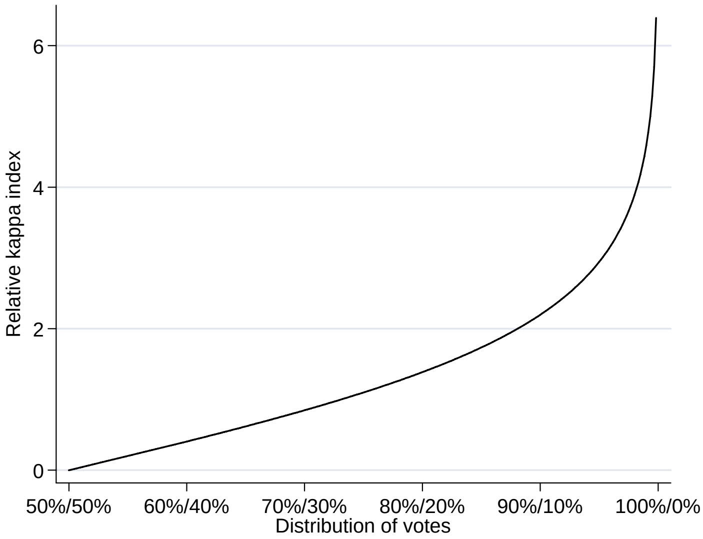

Figure 2. Value of the relative kappa index with two equal-sized parties and two equal-sized social groups, as a function of the distribution of votes

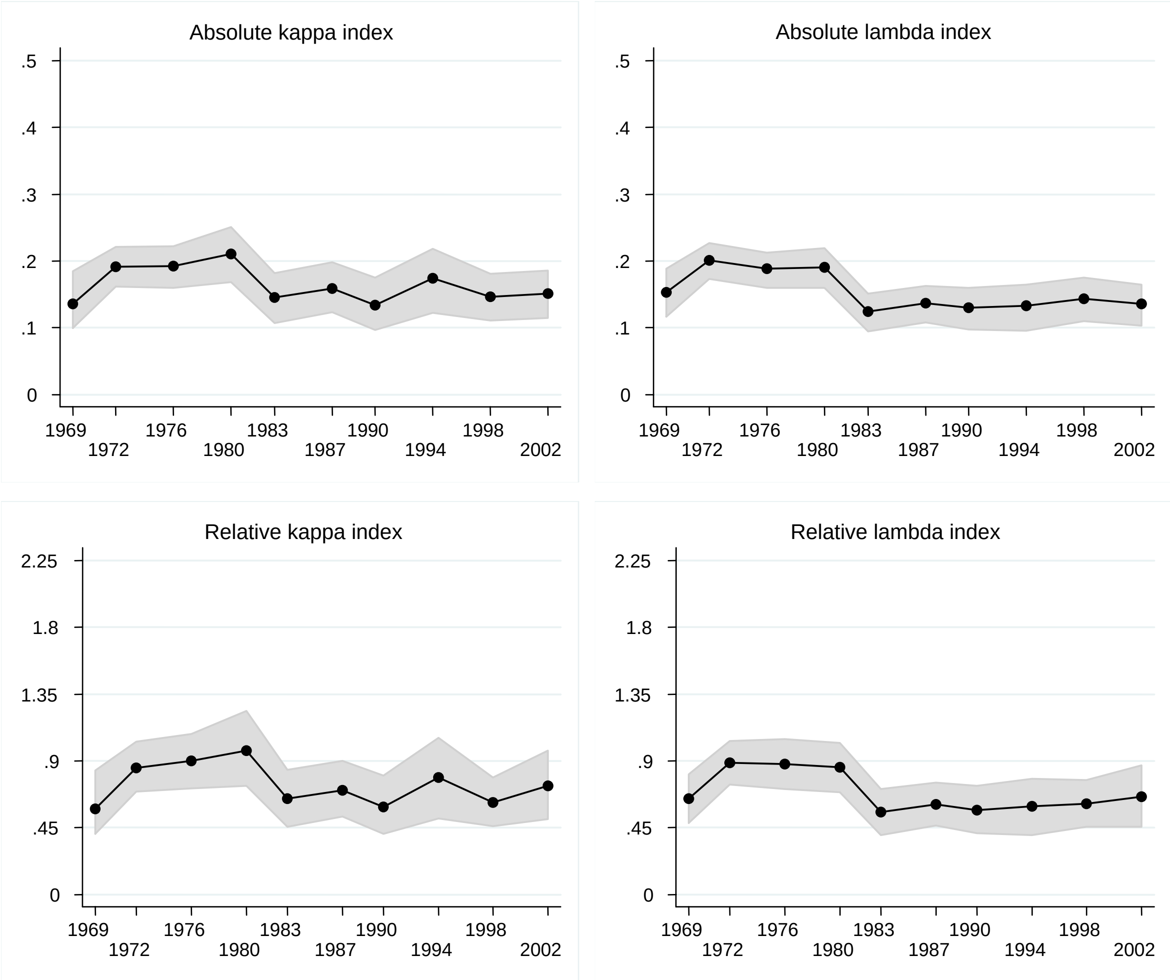

Figure 3. Strength of the religious cleavage with voting choice in two categories (SPD vs. CDU). Four indices of cleavage strength and their 95 per cent confidence intervals

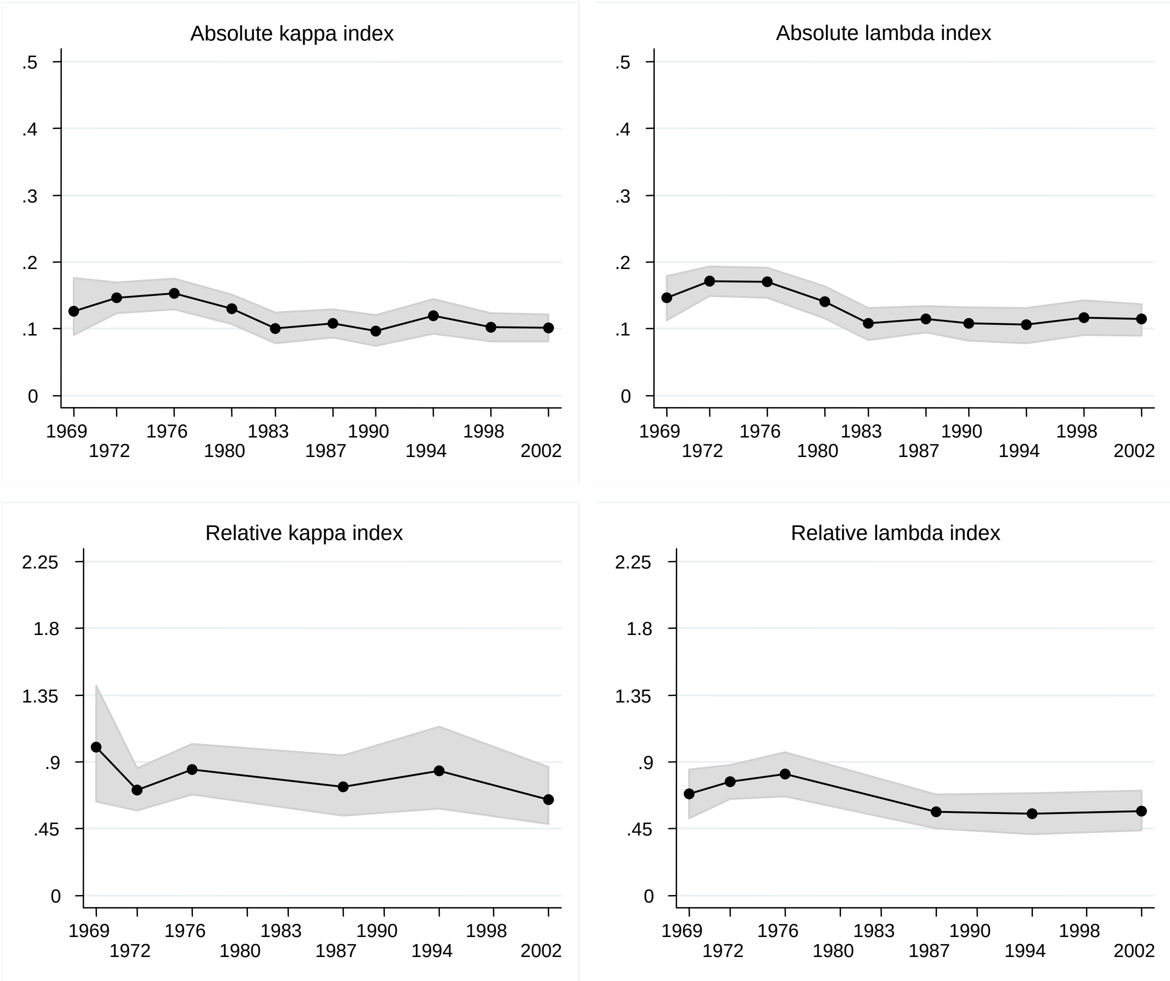

Figure 4. Strength of the religious cleavage with voting choice in four categories (SPD, CDU, FDP, Green). Four indices of cleavage strength and their 95 per cent confidence intervals

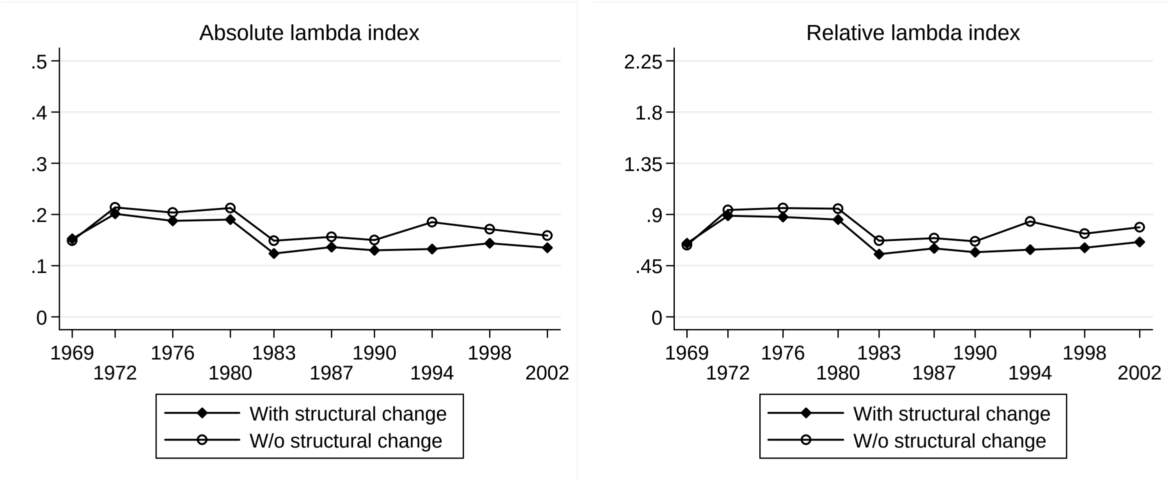

Figure 5. Values of lambda indices with and without structural changes. Voting choice in two categories (SPD, CDU)