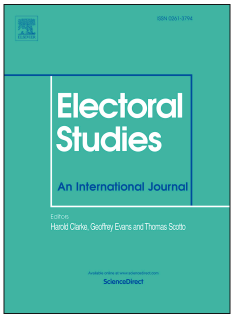
The volatility of volatility:

Casal-Bértoa, Fernando ; Deegan-Krause , Kevin; Haughton, Timothy

DOI:
10.1016/j.electstud.2017.09.007

Accepted Manuscript



The volatility of volatility: Measuring change in party vote shares

Fernando Casal Bértoa, Kevin Deegan-Krause, Tim Haughton

PII: S0261-3794(17)30108-7
DOI: 10.1016/j.electstud.2017.09.007
Reference: JELS 1869

To appear in: *Electoral Studies*

Received Date: 23 February 2017
Revised Date: 9 August 2017
Accepted Date: 20 September 2017

Please cite this article as: Casal Bértoa, F., Deegan-Krause, K., Haughton, T., The volatility of volatility: Measuring change in party vote shares, *Electoral Studies* (2017), doi: 10.1016/j.electstud.2017.09.007.

THE VOLATILITY OF VOLATILITY:

Measuring Change in Party Vote Shares

Fernando Casal Bértoa
School of Politics and International Relations, University of Nottingham
Fernando.Casal.Bertoa@nottingham.ac.uk

Kevin Deegan-Krause
Department of Political Science, Wayne State University
kdk@wayne.edu

Tim Haughton**
corresponding author
Department of Political Science and International Studies
University of Birmingham
B15 2TT, Birmingham,
United Kingdom
Tel: +44 121 414 6360
t.j.haughton@bham.ac.uk

Submitted to the *Electoral Studies*

Revised Version Following Reviewers' Comments

THE VOLATILITY OF VOLATILITY:

Measuring Change in Party Vote Shares

For *Electoral Studies*

Abstract

Volatility is a widely used term in political science, but even the most widely used measure of volatility, Pedersen's index, can mask as much as it reveals. His simple and elegant calculation has become part of the political science toolbox, but scholars employing this tool have tended to produce distinctly different results thanks to a series of decisions about measurement and classification. Using examples from Central Europe the critical role of decisions related to party continuity and threshold of inclusion are identified. The article not only unpacks the underlying questions addressed by different uses of Pedersen's index, but offers standards for choosing particular methods over others and outlines steps that should be followed in creating a more accurate measure of volatility.

1. Introduction

The political world - and particularly the world of political parties - appears to be in the middle of a period of intense, perhaps unprecedented change, but we cannot be sure unless we can have confidence that our measurements capture the type and degree of change. Few indicators in political science are more widespread than Mogens Pedersen's (1979) Index of Electoral Volatility. His straightforward calculation became part of the political science toolbox four decades ago and has remained constantly in use.[1] Nevertheless, scholars employing this tool have tended to produce distinctly different results thanks to a series of decisions about measurement and classification. Cooks following the same recipe should produce similar cakes, but only if that recipe specifies *all* of the important choices. Slightly different ingredients mixed in different ways can yield cakes which look and taste markedly different. Differences in volatility scores, however, are not just a matter of taste. Measurements of volatility matter not just because they form the springboards for theoretical discussions of the causes of continuity and change in party systems and politics more broadly, but because these measurements are used as proxies in research on dozens of other political and economic questions and are used not just as dependent but also independent variables.

Tools for measurement in political science tend to have their limitations and can be accused of bias. In the relatively stable party systems of Western Europe in the twentieth century where there were only occasional new entrants, the limitations of Pedersen's method appeared unimportant, but when applied to the more fluid electoral environments of third wave democracies (or today's fluid Western Europe), those limitations have become clearer. Aware of some of the limitations of Pedersen's calculation, Birch (2003), Mainwaring and colleagues (2016) and Powell and Tucker (2014) produced newer and more complex measures of volatility, which help to specify how much volatility resulted from party entry and exit. These significant advances do not, however, address other factors affecting volatility scores, such as thresholds for including data and different ways of dealing with the changing morphology of party competition such as the decision by Bartolini and Mair (1990) to link together parties with a common origin.

[1] Pedersen's 1979 article had 890 citations in Google scholar as of 1 August 2017.

What these approaches have in common is that they all build on the foundations laid down by Pedersen in the 1970s. Indeed, Pedersen's index remains the starting point for most attempts to capture electoral volatility and party system change. Any attempt to forge a better and more accurate measure, therefore, must begin from analysis of the strengths and weaknesses of Pedersen's formula and its application.[2]

After outlining why measures of volatility matter, drawing on cases from Central Europe - a region notable for both high and varied levels of volatility - this article shows how application of Pedersen's index produces strikingly different scores due to a wide variety of subtle but significant differences in method that authors introduce when applying the index. In particular, we identify the critical role of decisions related to party continuity and the threshold of inclusion. Moreover, we offer standards for choosing particular applications over others and whilst recognizing the merits of Pedersen's tool we conclude by suggesting elements that need to be incorporated into a useable successor to Pedersen that would be able to capture the type and extent of change in party vote shares.

2. Why volatility matters

There is a lot at stake in our measurement of volatility. Scholars agree that volatility in political party systems matters for democracy, and they use calculations of party vote shares as an indicator of a wide variety of phenomena and as an independent variable in studies about the health of democracy.

[2] It is worth noting that Pedersen was not the first to seek to measure volatility in this way. See Delruelle *et al*. (1970), Fraeys (1977) and Przeworski (1975).

Scholars are nearly unanimous in seeing a strong link between changes in the party political landscape and the process of democratization, and they argue that a stable party system is a key element in overall democratic stabilization (e.g. Tavits, 2005) because they 'foster more effective programmatic representation and reduce uncertainty' (Mainwaring and Zoco, 2007, 157). In their respective landmark studies of new democracies in Southern Europe and Latin America both Morlino (1998) and Mainwaring and Scully (1995) maintain that party system stability is a necessary (but not sufficient) condition for the consolidation of democracy, and many subsequent studies link high and persistent levels of volatility with democratic weakness. Volatility measures matter not only for studies of the survival of new democracies but also for research examining the health and development of long-standing democracies which have witnessed the growth of new, anti-establishment parties and experienced "earthquake" elections.

High standards for measuring volatility are particularly important because the Pedersen index is used not only as an indicator of changes in party vote share—the phenomenon that it directly measures—but also as a proxy for many other developments including voter movement (Epperly, 2011; Dassonville and Hooghe, 2015), government alternation (Mair, 2007), party system institutionalization (Weghorst and Bernhard, 2014; Chiaramonte and Emanuele, 2015), elite change (Ishiyama, 2013), declining partisanship (Lupu and Stokes, 2010) and regime stability (Bielasiak, 2001). Volatility may not necessarily suffice as a valid stand-in for all of these phenomena mentioned above, but as long as scholars use it for a wide range of applications across a broad temporal and geographical canvas, it is essential to 'pin down numbers' and make the measure as reliable as possible (Bartolini and Mair, 1990).

3. Why volatility is so volatile

The elegance and simplicity of Pedersen's index of volatility has allowed it to emerge as the dominant measure of change in party systems, but those same qualities have not prevented scholars from disagreeing sharply about volatility scores for specific countries during specific periods. The index goes to the heart of the question of political change by looking at changes in party vote share over time. In its simplest form, it calculates the total amount of change experienced by all individual entities in a closed system. For each entity it calculates the net change of a particular characteristic between two time periods, then takes the absolute value of this change (to prevent positives and negatives from cancelling out), and divides the result by the total amount of the characteristic in the system at the first and second time periods. Phrased in terms of specific variables, it calculates the absolute value of the net change of a particular characteristic (P), for every entity (i) between two time periods (t and t+1) divided by the sum of the same characteristic (P) at both time periods (t and t+1)

$$V = \frac{\sum_{i=1}^{n} \left| \mathrm{P}_{i,t+1} - \mathrm{P}_{i,t} \right|}{\left(\sum_{i=1}^{n} \mathrm{P}_{i,t+1} + \sum_{i=1}^{n} \mathrm{P}_{i,t} \right)}$$

As used by Pedersen, it is the sum of the net change of party vote shares (before *to* after) divided by the sum of all party votes (before *and* after).[3] Since most calculations of this nature employ the vote share of the full party system, the sum of all values of P is equivalent to 1 at both t and t+1. The formula can therefore be simplified to

$$V = \frac{\sum_{i=1}^{n} \left| \mathrm{P}_{i,t+1} - \mathrm{P}_{i,t} \right|}{2}$$

which can be rephrased simply as:

[3] Pedersen 1979, 4.

$$V = \frac{\sum_{i=1}^{n} |\Delta P_i|}{2}$$

The parsimony of the formula does not, however, translate immediately into harmony among those scholars who apply it to complex party systems. Studies of volatility involving Central Europe[4] offer a useful starting point because concern about the region's democratic development has led many scholars to calculate its volatility scores. The rapid changes in the vote shares of parties in Central Europe are unusual by traditional Western Europe standards (Lane and Ersson, 2007; Lewis, 2006; Millard, 2004; Tavits 2005, 2008), but they are not notably different from those of Latin America, Africa and Asia and or indeed from recent electoral periods in Western European countries such as the Netherlands, Belgium, Greece, Italy and Spain.

3.1 Detailing disagreements among volatility calculations

The many calculations of Pedersen volatility index scores in Central Europe often tend to agree in finding high average levels of volatility in the region, but a comprehensive examination of the results shows that they disagree sharply on the scores of particular countries in particular electoral periods. Table 1 displays the starkly different results of thirteen major studies generating Pedersen volatility index scores for Central European countries for electoral periods between 1990 and 2015.[5] In almost every country these scores exhibit wide differences that are not attributable to a small number of outliers. The coefficient of variation among all among authors for particular countries in particular periods averaged 0.31 for all countries and election periods, but for individual countries' election periods it ranged from a low of 0.11 to a high of 0.61. Out of 1,103 possible pairwise comparisons, the authors reached identical scores in fewer than one percent of the cases (9 cases,

[4] Defined for the purpose of this paper as the 10 states from the region which joined the EU in 2004 and 2007: Bulgaria, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia and Slovenia.

[5] Tavits' 2005 seminal study on the sources of party system stability in Eastern Europe draws its figures for volatility directly from Birch (2003).

0.8 per cent) and came within one point in fewer than twelve percent of all cases (128, 11.6 per cent). Pairwise gaps of more than 10 points (452 cases, 41.0 per cent) were actually more numerous than gaps of fewer than 5 points (423 cases, 38.4 per cent). The largest gaps between volatility assessments for any single country were 54.8 points for Estonia's first electoral period and 54.7 for Poland's second electoral period, but all ten countries had at least one gap of more than 15 points and eight of the ten countries had at least one gap of more than 30 points. It is striking that even scholars with strong knowledge of party politics in their native country differ by significant margins. The gap between Gwiazda and Markowski in Poland, for example, averages 4.8 points and is never closer than 3.0 points in any of the five election periods in which they both calculate volatility.

[Table 1 around here]

At first glance the contrasting calculations generated by different scholars focusing on Central Europe might be seen to be of limited interest to the wider field of comparative politics. Nonetheless, not only is the generation of distinctly different scores by different scholars not unique to that region of Europe as calculations of volatility using Pedersen for Africa (Lindberg, 2007; Kuenzi and Lambright, 2001), Latin America (Mainwaring and Scully, 1995; Roberts and Wibbels, 1999) and Southern Europe (Gunther 2007; Gallagher *et al*, 2011) show, but also those calculations for Central Europe have been widely used not just to describe, but also explain both region-wide phenomena and intra-regional differences. Take, for example, O'Dwyer's and Kovalčík (2007) account of why only some countries embarked on second-generation economic reforms. Using volatility statistics as a proxy for institutionalization, they argue that a labile party system may actually insulate state reformers from social and political pressures, thereby allowing politicians to undertake more radical packages of economic reform. For Estonia in the mid-1990s and Slovakia in the early 2000s their volatility scores for those two countries were higher than in Hungary and the Czech Republic where the more radical packages were not introduced. Whilst we are broadly persuaded by their argument if their numbers were replaced by those of other scholars in table 1 such as Bågenholm or Dassonneville and Hooghe, then their argument becomes far less compelling.

Different measures, however, do not just have impact on the explanatory power of volatility, but also on the causes and mechanisms associated with volatility. Bernhard and Karakoc's (2011) assessment of the links between volatility and inequality, Andrews and Bairett's (2014) explanation of the causes of volatility and even one of author of this article's analysis of party system institutionalization (Author) would alter significantly if based on volatility statistics chosen from another author in Table 1. This article does not seek to say which statistics (and therefore which conclusions) are right and which are wrong, but rather to highlight how different calculations affect the consequences, causes and mechanisms of volatility.

Were it simply a question of each author in table 1 having identical patterns but a different baseline or a different scale, it might be easy to harmonize them, but for many countries the relative positions differ sharply in direction and scale and with little consistency over time. Although the coefficient of variation among authors did not increase markedly in any country over time, it also did not show any overall tendency to decline. For many countries there was no consistency at all: Hungary's first election period, for example, appears near the bottom of Mainwaring's and Toka's ranks but near the top of Dassonneville and Hooghe's, whereas Poland's second election period is near the bottom of Bågenholm's but near the top of Mainwaring, España and Gervasoni's and Powell and Tucker's. Other countries produced more consistent evaluations but the consistency lasted only for a few election periods: Bulgaria is near the bottom of every author's scale of volatility for the first election period while the Czech Republic is at the bottom of nearly every scale for the second and third periods and Latvia and Lithuania are almost universally at the top of authors' scales for the first three electoral periods. Among author pairs with more than 20 country-electoral period observations in common, correlations range from a high of r=0.90 (Sikk compared to Lane and Ersson, n=20) to a low of r=0.46 (Sikk compared to Powell and Tucker, n=20). Overall, the average correlation among the pairs with more than 20 observations in common was r=0.73, which is fairly robust for a correlation among different phenomena but shockingly low for results that are intended to measure the *same* phenomenon.

Since all of these authors are working with more or less the same electoral data, the reason for the differences must lie elsewhere. A closer examination shows that there are significant differences in the methods that authors use to incorporate the election data into the Pedersen index calculations. The differences emerge because those who would calculate volatility index scores must wrestle with the interrelated problems of small parties and party continuity. For many authors, especially those engaged in large-scale comparative enterprises, identifying the twists and turns of small parties becomes an overwhelming burden. Identifying continuity is difficult enough with large parties, especially where these face party splits and engage in mergers and name changes. Pedersen's index actually provides several different options for dealing with these problems, but not all authors use the options in the same way, and the various combinations allow Pedersen's seemingly uncomplicated formula to produce a remarkably complex array of results. The three main options for adapting Pedersen's formula to deal with size and continuity are rooted directly in the rules for linking party data across time, for aggregating party data into a single number, and for recalculating individual data points.

3.2 Linkage and aggregation of entities: What counts as a connection?

Perhaps the most complicated aspect of volatility calculations is linking party values over time in the face of party name changes, mergers, splits and various combinations thereof. The category of 'genuinely new' parties (Sikk, 2005) is smaller than it might at first look, and a large number of the entrants in any given electoral period involve partial continuity. Many scholars of volatility respond with case-by-case judgments or with rule-based formulae for deciding which party $P_{i,t}$ to link with which party $P_{i,t+i}$. Methods that use a stricter interpretation of continuity between t and t+1 err on the side of discontinuity and tend to code name changes, splits and mergers and new parties that represent sources of volatility. Methods emphasizing continuity may opt instead for a more relaxed understanding of linkage and pair a party at time t with whatever party it is most closely related—either by name, organizational structure or vote total—at time t+1. Even in these relatively regularized, rule-based systems, however, there may still be considerable human judgment in the assessment of which parties are to be considered within the set of plausible successors. Such decisions become even more sensitive when applying Mainwaring, España and Gervasoni (2016) and Powell and Tucker's (2014) intra- versus extra-system distinction that depends heavily on whether parties are considered as successors or as new.

For those scholars who are reluctant to declare a single predecessor in a merger or a single successor in a split, the Pedersen formula includes a second implicit option: the possibility of aggregating the values of two or more parties at either t or t+1 or both before linking them. For example, if Party A merges with Party B to form Party C, the aggregation method sums Party A and B into a single data point at time t and compares it with party C at time t+1. If, at the same time, party D splits off from party A, the method would compare the sum of Party A and B with the sum of Party C and D. This approach eliminates the need for choice among potential links and emphasizes the continuity that may exist between the original party and multiple successors (or predecessors). Of course in the process it diminishes the number of entities (i) and therefore makes the indicator itself less sensitive. Another common application of this method involves the aggregation of small parties at a particular time period into a single data point commonly referred to as “Other.” Which parties are thus aggregated usually depends on rules for including parties in the main calculation.

3.3 Inclusion and exclusion of information: Which data points count?

Limits of time and information mean that it may not be practical (or even possible) to gather data for small parties. Even though new data storehouses and national websites provide extensive electoral results for most countries, questions of linkage over time are still extremely difficult and time-consuming for very small parties, especially since such parties often face splits by disgruntled members and seek to improve their fortunes through name changes and mergers. Although some authors make ad hoc exclusions based on the availability of data, most seek to avoid selection bias problems by imposing a numerical criterion for inclusion usually based on the magnitude of a party’s election result. Even with a clear threshold, additional questions remain, particularly whether to apply the threshold in a blanket fashion and include values for a particular party if it exceeds the threshold at any point in time, or to include only individual data *points* that exceed the threshold. The former option does little to solve the data collection problem since researchers may still need to find obscure data for parties that only later crossed the threshold; the latter option, by contrast, severs the continuity of party results by relegating some data points—those below the threshold—to the "other" category or dismissing them completely. An intermediate strategy involves the use of transitional pairs of data points to capture each period in which a party's support crosses the threshold, but does not calculate data points that are not adjacent in time to periods above the threshold. Table 2 offers a hypothetical example of how each of these strategies could be applied and how the application would affect volatility scores:

[Table 2 around here]

Using transitional pairs thus involves some recalculation of data points (and the possible transfer of those values to the "Other" category), but not as much as using only individual points. The blanket inclusion method, by contrast, either includes all values or omits a party entirely.

Questions of inclusion are also at the heart of recent modifications to Pedersen's index which seek to distinguish changes in established parties from changes introduced by the entrance and exit of parties from the party system. In an instance of parallel invention, two groups of scholars - Mainwaring, España and Gervasoni (2016) and Powell and Tucker (2014) - simultaneously devised a formula for disaggregating Pedersen that builds on earlier work by Birch (2003).

Within-system/Type B   Extra-system/Type A

$$V = \frac{\sum_{i=1}^{n}\left|\mathrm{P}_{i_c,t+1}-\mathrm{P}_{i_c,t}\right|+\sum_{i=1}^{n}\left(\mathrm{P}_{i_n,t+1}+\mathrm{P}_{i_x,t}\right)}{\left(\sum_{i=1}^{n}\mathrm{P}_{i,t+1}+\sum_{i=1}^{n}\mathrm{P}_{i,t}\right)} = \left(\frac{\sum_{i=1}^{n}\left|\mathrm{P}_{i_c,t+1}-\mathrm{P}_{i_c,t}\right|}{\sum_{i=1}^{n}\left(\mathrm{P}_{i,t+1}+\sum_{i=1}^{n}\mathrm{P}_{i,t}\right)} + \frac{\sum_{i=1}^{n}\left(\mathrm{P}_{i_n,t+1}+\mathrm{P}_{i_x,t}\right)}{\left(\sum_{i=1}^{n}\mathrm{P}_{i,t+1}+\sum_{i=1}^{n}\mathrm{P}_{i,t}\right)}\right)$$

where i_c refers to continuous parties with nonzero values in both t and t+1, i_n refers to new parties with zero values at t but nonzero values at t+1, and i_x refers to parties that exited the political scene yielding non-zero values at t and zero values at t+1. Since i_n by definition has no predecessors and i_x has no successors, these can stand alone in the calculation, implicitly subtracting the previous support or successive support (equal to zero). By then looking only at the left or right hand of the equation, these authors can calculate the total contribution to volatility represented by continuous parties and by those entering or leaving the system and provide distinct figures for within-system and extra-system volatility. The scores generated for within-system volatility can also be added to the extra-system scores to produce an overall volatility score that is equivalent to Pedersen's formula.

Mainwaring et al and Powell and Tucker's disaggregation of 'intra' and 'extra' system volatility is a major advance in our understanding and has generated much fruitful discussion (Crabtree and Golder, 2016a, 2016b; Powell and Tucker, 2016). But it is striking is that the foundation stone on which Mainwaring *et al* and Powell and Tucker's disaggregation is based is the Pedersen index. Because their newer and more sensitive instruments are correspondingly *more* vulnerable to distortion (Crabtree and Golder 2016a, 2016b) caused by small differences in measurement, it is even more important to understand the specific choices required by Pedersen's index and how those choices might best (or at least most transparently) be made.

3.5 Classifying authors' choices of method

Among the fourteen author methods cited here, there are eleven different choices and significant variation in each category: ten without specific thresholds and four with thresholds (ranging from 1 per cent to about 5 per cent) of which one uses recalculation and three use pairwise exclusion; seven of the ten including an "other" categories in their calculations; and seven assessing continuity on the basis of aggregation while four use the linkage method (two using strict linkage and two using relaxed linkage), and three use a mix or other variation.[6]

[following Reviewer 3's suggestion we have moved what was table 3 to an appendix]

4. How method choices affect volatility index scores

How significant a role do these different choices play in producing the different volatility results? Not all authors look at all countries in all electoral periods and so an overall comparison of differences is impossible, but it is possible to calculate pair-wise differences between each author-country pair and then average these to assess the relative magnitude of each author's volatility assessments. Comparing these results to choices about threshold shows no clear pattern. A slightly stronger pattern emerges with regard to the "other" category, with higher volatility scores concentrated among those who do not include "other" in their calculations. Finally, there appears to be a strong pattern related to the choice of aggregation or linkage in establishing continuity. As Figure 1 indicates, with the exception of Birch (who in addition to these choices does not include new parties), the authors choosing the aggregation method have the lowest volatility scores, while those choosing the linkage methods are higher. It is also noteworthy that the authors using the linkage methods also have a far greater range of scores than those using aggregation, which is to be expected since the results of the linkage methods are more affected by differences in binary choices of predecessor and successor that might differ from author to author even in the best of circumstances.

[6] See Table A in the appendix

[Figure 1 around here]

With enough countries and enough combinations of methods, it should theoretically be possible to separate out the effect of particular choices on volatility scores, but at present there are far more variable combinations than there are authors using those combinations. Figuring out the actual impact of each of these variables therefore requires a far more tightly focused method. Central European cases can provide the raw material for a closer analysis that lets us control for specific differences in inclusion, aggregation and linkage by applying every possible combination of those methods to the same sets of full electoral data. The time- and knowledge-intensiveness of this process is quite substantial (and helps to explain the use of selective inclusion methods). For many countries the demands of the process are prohibitive, but for three countries - the Czech Republic, Hungary and Slovakia - we were able to assemble accounts of even the smallest parties (some with fewer than 1,000 votes) and to determine linkage patterns among these parties across each time period. These three cases are also particularly useful because they exhibit a range of volatility characteristics: by nearly all estimates Slovakia experienced relatively high volatility throughout its first 20 years of democracy whereas Hungary's volatility was relatively low in contrast to other Central European cases; the Czech Republic's volatility was also low until the late 2000s but jumped sharply with the 2010 and 2013 elections. These differences both in level and consistency - one low, one high, one low then high - provide a solid initial basis for testing the various permutations of method.

To operationalize every one of the theorized variants of the Pedersen index on these three cases, we performed a full set of calculations for each permutation of methods related to continuity and those related to threshold and inclusion.

Continuity: we calculated volatility in each country according to three methods that affected linkage.

1) Linkage method. There is a wide range of options to deal with the successors and predecessors in the case of splits, mergers and party name changes. We calculated two distinct subsets of linkage according to different guidelines:
    a. Strict linkage: we calculated any split or merger involving a name change or other significant alteration from t to t+1 as a new party with no formal linkage to predecessors or successors.
    b. Relaxed linkage: we linked parties that merely changed names and in the case of split or merger we linked parties to the largest successor or predecessor while treating any other offspring or parents as unlinked across time periods.
2) Inclusive aggregation method. In the case of a party split or merger we aggregated all predecessor and successor parties treating them as a single party in each time period.

Threshold: We calculated Pedersen scores first with no threshold and then with thresholds at ten 0.5 per cent additional increments up to 5.0 per cent (the highest level used by any author in the sample above).

Inclusion: We calculated scores using three threshold-based inclusion methods in three ways, two of which are used by authors in the sample and a third potential option:

1) Individual points: This method includes only data points that stand above the threshold and recalculates others as zero.

2) Transitional pairs: This method includes all at points that stand above the threshold and data points from the periods immediately before and after the point above the threshold.
3) Blanket inclusion: Including all parties whose maximum vote share rose above the threshold at any time period again.

Inclusion of the "Other": For parties that were excluded from the calculations above, we performed volatility calculations with and without an aggregated "other" category.

1) Aggregated and linked other. We produced an aggregated "other" for each election consisting of the sum of data points not included elsewhere, and treated this as a distinct unit, whose changes over time contribute to volatility calculations like those of any other party.
2) Uncalculated. Data points not included elsewhere were not included in volatility calculations.

The combination of three continuity methods, three threshold-inclusion methods, two "other"-aggregation methods and 10 threshold levels produces 180 possible variations per election period in each country. Figures 2a through 2c show the average results of each of these combinations applied to electoral data for the Czech Republic, Hungary and Slovakia.

4.1 Bivariate Results

In analysing the results, we begin in reverse order because the effects of some variations vastly outweigh the effects of others (and the interaction among the variations becomes important).

Continuity: The method of addressing continuity has an extremely strong impact on the measurement of volatility. With the threshold set at zero, over seven electoral periods in Slovakia the strict linkage method produced an average volatility score of 47 whereas the relaxed linkage method averaged just over 31 points and the inclusive aggregation method averaged just under 20 points. The results were similar in the Czech Republic and Hungary, though at lower overall levels and with smaller overall differences. A look at calculations that comprise these results explains this difference: Strict linkage allows for the fewest connections across elections and therefore means that fewer parties have counterparts from t to t+1; relaxed linkage, by contrast, seeks out relationships between t and t+1 but still excludes linkage from all but one designated predecessor or successor. The inclusive aggregation method nearly always produces lower volatility results because it adds together entities that would otherwise be separate, minimizing the differences between units across time periods. Indeed, the inclusion method sometimes goes so far in this direction that it loses the capacity to make meaningful distinctions when parties split from one source and then merge with another. Figure 3 shows an example of split-merger combinations in which the inclusive aggregation cannot detect real change.

[Figure 3 around here]

When this "Z-effect" repeats across the party system, it significantly reduces the basis for calculating volatility. In the example above, if Party D also lost a splinter, Party E, that merged with Party B and this process continued with Party F and beyond, the whole of the party system could be aggregated on each side of the calculation producing zero volatility even in the case of a wildly changing system. This is more than just a theoretical possibility: in the case of Slovakia between 1992 and 1994, mergers and splits produced a single aggregate that stretched across the political spectrum, including nine parties and accounting for over 50 per cent of all votes in both years (Kopeček, 2007; Author).

Threshold Size: A change in the threshold can also produce a significant change in volatility, but its effect is only partially predictable and depends on other factors. As Figures 2a-c show, volatility tends to decrease with an increasing threshold for both the relaxed linkage and strict linkage methods of continuity. In the example of Slovakia, the addition of a 5 per cent threshold to the strict linkage method results in a 10-point drop in measured volatility. This finding fits expectations since in any party system with a significant number of small parties, a higher threshold forces the calculations to ignore changes among smaller parties and thus reveals apparently lower levels of volatility. Thresholds do not produce this effect in the inclusive aggregation method. The underlying party-level calculations show why: where small parties join with larger ones the elimination of small parent or offspring parties from one side of the equation because of thresholds actually decreases the measure's ability to recognize continuity and may produce higher volatility scores rather than lower ones.

Threshold inclusion: The method of applying the threshold matters as well. As Figures 2a-c show, the dotted line representing "blanket inclusion" usually produces a smaller drop in volatility than either the transition pairs or individual points methods since this method expands the number of parties included in the calculations and therefore mitigates the threshold's overall effect.

The individual point method creates a different sort of problem by simply zeroing out any figure below the threshold. The effect has sometimes unpredictable effects, however, because the recalculation produces two countervailing effects: on one hand, it reduces the measured level of volatility by eliminating all pairs of party results that are both below the threshold (a party's rise from 1.0 per cent to 2.0 per cent represents a 1.0 contribution to volatility, but this is not included in volatility calculations when the threshold is set at 2.5 per cent and both points are set to zero); on the other hand it exacerbates volatility in cases where one result in an electoral period is above the threshold and the other is below (a party's fall from 4.0 per cent to 1.0 per cent actually reflects a 3.0 contribution to volatility score but the individual points method recalculates the second of the two figures to 0.0 per cent, producing an apparent contribution of 4.0, a "volatility subsidy" equal to the size of the smaller figure in the electoral period). To understand these effects, Pedersen's original index can be expanded to isolate particular sets of entities, some with values above the threshold in both time periods i_{sa}, some with values below the threshold in both time periods, i_{sb}, and some with values on either side of the threshold, i_{sac} *and* i_{sbc}. This effect can be represented mathematically as

Supra-threshold Sub-Threshold Cross-Threshold(down) Cross-Threshold (up)

$$V = \frac{\left(\sum_{i=1}^{n}\left|\mathrm{P}_{i_{sa},t+1} - \mathrm{P}_{sa,t}\right| + \sum_{i=1}^{n}\left|\mathrm{P}_{i_{sb},t+1} - \mathrm{P}_{i_{sb},t}\right| + \sum_{i=1}^{n}\left|\mathrm{P}_{i_{sac},t+1} - \mathrm{P}_{sbc,t}\right| + \sum_{i=1}^{n}\left|\mathrm{P}_{i_{sbc},t+1} - \mathrm{P}_{sac,t}\right|\right)}{\sum_{i=1}^{n}(\mathrm{P}_{i,t+1} + \sum_{i=1}^{n}\mathrm{P}_{i,t})}$$

The exclusion method sets all i_{sb} to zero, producing

Supra-threshold Sub-Threshold Cross-Threshold(down) Cross-Threshold (up)

$$V = \frac{\left(\sum_{i=1}^{n}\left|\mathrm{P}_{i_{sa},t+1} - \mathrm{P}_{sa,t}\right| + \sum_{i=1}^{n}|0-0| + \sum_{i=1}^{n}\left|\mathrm{P}_{i_{sac},t+1} - 0\right| + \sum_{i=1}^{n}\left|0 - \mathrm{P}_{sac,t}\right|\right)}{\sum_{i=1}^{n}(\mathrm{P}_{i,t+1} + \sum_{i=1}^{n}\mathrm{P}_{i,t})}$$

which can be simplified as

Supra-threshold Cross-Threshold(down) Cross-Threshold (up)

$$V = \frac{\left(\sum_{i=1}^{n}\left|\mathrm{P}_{i_{sa},t+1} - \mathrm{P}_{sa,t}\right| + \sum_{i=1}^{n}\mathrm{P}_{i_{sac},t+1} + \sum_{i=1}^{n}\mathrm{P}_{sac,t}\right)}{\sum_{i=1}^{n}(\mathrm{P}_{i,t+1} + \sum_{i=1}^{n}\mathrm{P}_{i,t})}$$

Raising the threshold has two effects. On one hand, a higher threshold may decrease the number of parties whose volatility counts in the overall volatility calculation (because it is happening in the excluded sub-threshold zone). At the same time, a higher threshold may increase the number of cases in which the result at either t or t+1 is recalculated to zero, and thus may increase the net amount of cross-threshold volatility (up or down). These two countervailing influences are only tangentially related (constrained by the overall party system) and depend heavily on the relative sizes and shifts of party support within a party system during a given time period. Considerable change among sub-threshold parties will generate an artificially low score by increasing the relative importance of i_{sb} while activity that crosses the threshold will have the opposite effect.

How these effects interact will depend on what actually happens in the political sphere. The results of Figure 2a-c suggest that the cross-threshold effect tends to predominate in the countries studied here, causing increases in the threshold to increase the measured volatility. The individual points method may thus inadvertently produce results that more closely resemble a threshold of zero, but this occurs only because of the artificial cross-threshold counter-effect that is largely accidental and cannot be relied upon to consistently push the volatility figure toward the level it would reach without a threshold. It produces something close to the right answer but not necessarily for the right reason and therefore does not represent either a valid or reliable shortcut.

The transitional pair method solves some of the problems above. It excludes data in only those instances where both characteristics in a pair of time periods are below the threshold but permits those where only one is below. The advantage of this is to eliminate the cross-threshold distortions of the individual points method. In the example above, Party A's fall from 4.0 per cent to 1.0 per cent is counted only as a 3.0 contribution to volatility while the subsequent rise from 1.0 per cent to 2.0 per cent does not enter into the calculations at all. In practice this means a lower level of volatility than if the cross-threshold effects were included, but since those effects are artificial and more erratic, it actually produces a smoother line - more directly proportional to the level of the threshold - and a more theoretically justifiable basis for calculation. It is true that it requires slightly more effort in finding data for the time period before and after a party crossed the threshold, but this is more sustainable than gathering linkage information for a small party's entire existence, and it is precisely the kind of information more likely to appear in the popular press ('Party A rose to 4.0 per cent from its previous performance of 1.0 per cent').

<u>Excluded data:</u> Finally, there is the question of how to handle data that is not included by the threshold exclusion methods above (Ocaña, 2007). It is possible simply to omit the data from calculation (in effect recalculating it as zero), but some authors opt for an aggregation method and sum all excluded data into a single "other" category at the bottom of the data table. This aggregation method raises questions about linkage between the aggregated "other" values from one time period to the next. Since the aggregated "other" category may have different composition at each time point, and since, because each item aggregated into the "other" category may be (by definition) below the notice threshold of the scholar gathering the data, circumstances mean that there is no formal way to establish the degree of linkage between two successive "others." As with other questions above, there are several options with relatively clear effects on volatility: to create no linkage between the "other" results and calculate them as new, one-time parties (in effect adding the size of the "other" category to volatility), to refrain from including "other" at all, and to link the "other" results from one time-period to the next. While theoretically possible, the first of these approaches is not a useful option because it assumes that smaller parties have no linkage with any other entity, past or future, thus magnifying volatility by the entire size of the other category. The second of these, by contrast, presumes that there is *no* volatility among the unmeasured entities and

must produce a level that is equal to or lower than the actual volatility level. The final method, which seems to offer a middle way, creates difficulties of another variety. The total size of the "other" from election point to election point is only distantly related to the "other" volatility.

The difference in the size of the "other" category theoretically depends very little on differences related to the overall level of volatility or on the volatility that is lost to calculation through exclusion of data or cases. The grey lines of Figure 1 shows that in practice in exclusion methods, linking data points in the "other" category restores about half the amount of volatility to the calculation that is removed through the imposition of thresholds, but the effect is not strictly linear and appears to increase disproportionately as the threshold approaches 5.0 per cent. The volatility contributed to the overall result by the aggregated other categories, as with the subtractions in the recalculation method of the threshold, is an accidental effect. If all parties under the threshold move in the same direction (they all get smaller or larger) then the "other" category is mathematically identical to the aggregation of the individual elements. If, however, parties move in opposite directions, the volatility is masked and the "other" category obscures the actual pattern. The "other" category may add volatility at the same time that is removed by the threshold but it does not necessarily put back the *same* volatility. In the cases used here it tends to follow the overall curve, but there is no mathematical basis for thinking that it will do so in a reliable fashion and no theoretical justification for aggregating the parties below the threshold into a single, continuous "other" (unlike the more solid basis for aggregating parties involved in mergers and splits in the aggregation alternative to linkage).

4.2 Multivariate results

Looking at threshold, linkage and exclusion separately highlights how these individual ingredients can change the size and shape of the volatility, but what also matters is their relative effects and how they interact. To that end we subjected the data from the three countries to a multivariate regression (table 3). Given the continuous character of our dependent variable we use ordinary least squares with robust standards errors. In order to control for possible country- and election-effects, we use dummy variables for each election per country (not included in the results table for reasons of space).[7]

[Table 3 around here]

In the basic model, the continuity method has large and statistically significant impact (at .001 level). Shifting from the inclusive aggregation method to the relaxed linkage method increases volatility by nearly five points (4.8), and from there to the strict linkage increases volatility by another 10 points for a total of fifteen points above the inclusive aggregation method (15.2), controlling for the other independent variables. Threshold size also has a significant impact: each one percentage point increase in the threshold reduces the amount of volatility by nearly one point (0.81). Including an "other" category is also significant, though the effect of including such a category is relatively small, accounting for approximately a one point increase in volatility (1.3). The threshold inclusion method has no partial effect on the dependent variable, with no significance difference between methods that use individual points, the transitional-pairs, and blanket inclusion.

[7] The reference category for the dummies is the first Hungarian election.

Since Figures 2a-2c suggest that thresholds do not function the same way in all circumstances Model 2 uses interaction terms for threshold size and the other methods. The choice of threshold inclusion method again has no effect, but all of the other methods do interact with the threshold size in statistically significant ways. Model 3 removes the non-significant methods and interaction terms for cleaner results. We will therefore rely on Model 3, rather than Model 2, for the interpretation of the results. The interaction effect between the continuity method and threshold size are substantially and statistically significant. With the inclusive aggregation method, raising the threshold by one percentage point reduces volatility by only 0.5 points, but with the relaxed linkage method each one percentage point increase in the threshold produces a 1.3 unit drop in volatility and with the strict linkage method the result is a 1.4 unit drop, provided that the interaction with "other" is kept at 0. A similar, but less statistically significant (at .01 level) interaction effect occurs between threshold size and the aggregated method of dealing with "other" data. Indeed, provided that the interaction with method is kept at 0, if other is at 0 the model predicts that, holding the other variables constant, each additional threshold point will produce a 0.46 point decrease in volatility. On the other hand, the aggregated other method would produce only a 0.02 decrease.

5. How to evaluate and improve volatility calculation methods

The purpose of the calculations above is not merely to explain the diversity of volatility results found above and the relative impact of specific variations of Pedersen's Index, but also to encourage more accurate and reliable measurement. In reverse order, moving from least to most important we offer three specific recommendations and a point for further consideration.

Firstly, since significantly different variations of the threshold inclusion method have no significant impact on volatility scores, the recommendation must rest on considerations other than results. The blanket inclusion approach gives results closest to the desirable zero-threshold level, but it is both the most demanding in terms of data demands and the most problematic in terms of the constant need to recalculate volatility based on party performance in subsequent years. The individual points method, by contrast, is the easiest to apply, but its underlying mathematics—zeroing out anything below the threshold—may not in every case produce the balance between eliminating one realm of volatility while introducing another that we see in these three cases. The middle ground of the transitional pair approach splits the difference, capturing important cross-threshold changes without the need for massive additional data collection or fundamental recalculation back to the beginning of the dataset if a once-small party makes it big.

Secondly, and with only slightly greater impact is the question of the right method for excluded data; while the decision has statistically significant consequences, it usually does not change volatility scores by more than a single point. Since the use of an aggregated other category introduces as much noise as it removes (like the individual point method of threshold inclusion discussed above) and thereby introduces the possibility (albeit small) of more significant errors, it may be preferable simply to omit the category entirely and acknowledge that doing so will result in a small underestimate of volatility.

Thirdly, and more significant is the question of the threshold itself. The ideal threshold size is zero (especially since this eliminates the above problems connected to threshold inclusion and excluded data), but there is a clear trade-off between accuracy and the costs of gathering information. The experience of data collection for this study suggests that efforts to lower the threshold are accompanied by an almost exponential increase in information costs. For most larger parties the election data and continuity over time was easy to assemble, but as we pushed the threshold below 1.0 per cent and especially below 0.5 per cent, we found parties that were impossible to trace without archival searches and interviews. This difficulty may ease over time with the development of historical databases (especially if scholars gather information about small parties as each election happens), but with the current state of data for most countries, it is probably adequate to use a 1 per cent or 0.5 per cent threshold along with the acknowledgement that this may underestimate volatility by one to two percentage points in the case of the relaxed linkage and strict linkage methods and by a somewhat smaller amount with the inclusive aggregation method.

Finally and most importantly there is the question of continuity method. Unlike the other choices discussed above, this one offers no quick solution. Each method actually represents a distinct and coherent way of thinking about the notion of change in party systems.

The default assumption of the inclusive aggregation method is continuity: splits and mergers do not change voter preferences, and voters can navigate the institutional change to follow leaders or party factions they prefer, even if their favourites move out of an existing party and into a new splinter, or merge into a larger entity. In this light, voters who chose to switch their votes from a small party into its merged successor (or from a big party to one of its splinters) may be engaging in an act of *continuity* rather than change. This method is therefore at its most useful revealing when continuity is *absent*. When scores produced by the inclusive aggregation method are high (by the standards of past elections or neighbouring countries), it signals voter shifts so significant that they fall outside of existing institutional boundaries.

By contrast, the default assumption of the relaxed linkage and strict linkage methods is change: splits and mergers can make it impossible for voters to stick with their past choices, and even choosing to follow a party absorbed by a merger or to follow favoured leaders who split off to form a party of their own represents a rupture with past choices. This method is therefore at its most useful revealing when continuity is *present*. When scores produced by the linkage methods are low, they signal that few voters are shifting despite any pressure imposed by splits and mergers.

Neither method, then, is *intrinsically* better than another. One option is to follow the path taken by Tóka and Henjak in measuring *both* raw (strict linkage) and adjusted (inclusive aggregation) scores to identify the upper and lower bounds of the possible values. Taking this double calculation one step further produces a heretofore unnoticed advantage: a *comparison* of the results produced by both methods offers significant insight into the nature of volatility in particular countries and particular elections.

The comparison in Figure 4 of inclusive aggregation and strict linkage volatility scores in the three countries used in this study accurately reflects the overall history of party change in the three countries. All three alternate between higher and lower amounts of overall volatility among party units and these stand at the upper right and lower left corners of the diagonal line, but certain elections in certain countries also involve an amount of institutional volatility—splits and mergers—over and above the voter shift. Shifts due to a split or merger were relatively rare in

Hungary and the Czech Republic and their data points tend to hug the diagonal (except for a slightly above-average shift related to the splintering and reformation of Hungary's left in 2014 and a major shift following the breakup of the anti-Communist Civic Forum in the Czech Republic in 1992), Slovakia, by contrast, experienced three big merger- and split-related institutional shifts in 1992, 1998 and 2002 which stand out sharply in a comparison of Slovakia's its inclusive aggregation and strict linkage volatility scores.

[Figure 4 around here]

This multiple-method approach is admittedly less useful for those who need a single volatility score as a dependent or an independent variable in a multivariate regression model. In such cases, scholars may need to make an explicit choice of method depending on whether the question at hand refers primarily to voter continuity or institutional change and acknowledge the potential impact of that choice on results. It should also be possible at present to create distinct regression models for volatility based on results from the linkage and aggregation methods and then examine the degree to which the results agree or disagree. In the longer run, however, the better option is to expand and improve volatility measurement so that the partial shifts related to mergers and splits are fully included in the formula and become an integrated part of overall volatility scores.

[Table 4 around here]

6. Conclusion

Pedersen's formula for calculating volatility has been widely used and for good reason. It offers a simple and straightforward way of capturing the extent of electoral change. But it also has its limits and its application can yield a wide variety of results depending on a series of secondary methodological assumptions about inclusion, aggregation and linkage that scholars use to address problems of limited party data and lack of clear party continuity. Even the recent significant advances offered in our understanding of volatility by Mainwaring and colleagues and Powell and Tucker are built on the foundations laid down by Pedersen. As Crabtree and Golder (2016a) assert the study of volatility is an area of research in need of new theoretical development, but satisfactory theorizing needs to be accompanied by the search for a robust measure. This need is all the more pressing given the apparent fluidity of party and electoral politics.

Our analysis does not suggest that Pedersen's tool should necessarily be consigned to a (statistical) retirement home. Rather those employing his formula or using volatility statistics based on an application of his formula should be aware of the limitations and the explicit and implicit assumptions made in the calculations. Nonetheless, we do suggest that our analysis indicates potentially new and fruitful paths towards a more accurate measure of volatility. Space precludes a detailed examination, but three points deserve mention.

Firstly, party systems not just in Central Europe but in many parts of the globe across are increasingly marked not just by the emergence of new parties, but also of splits, splinters and mergers. Even just focusing on the cases in the region in the past few decades highlights the fact that questions of linkage between parties at t and t+1 are not black and white, but shades of grey. Some parties can be categorized as ‘genuinely new’ (Sikk, 2005), but many more fall between the genuinely new and old. The solution here may be to introduce a more callibrated measure recognizing the degree of newness i.e. that party Y is 50% new and 50% of a continuity of party X. Take one example from Slovakia: the creation of the party Most-Hid in 2009. The party’s name, structure and programme were all new, but it was forged by a breakaway group from the Party of the Hungarian Coalition (SMK) including one of that party’s former leaders and most prominent figures, taking with them to the 2010 elections many of the voters who had cast preference votes for those individuals in previous elections. Some statistical recognition of partial continuity and change would intuitively seem more accurate. It is worth stressing that such a callibration could be done as part of a calculation using the Pedersen index, albeit in modified form. Producing a workable measure is unlikely to be easy, but by avoiding a dichotomous choice between successor and non-successor, and moving toward a measurement of partial linkage among multiple entities would help better capture the degree and type of party vote share change.

Secondly, in order to make a more robust calculation of linkage, scholars need to think more carefully about what continuity and change are and how to measure them. It may be helpful to disaggregate parties into constituent components such as Katz and Mair’s (1993) three faces: party on the ground, the party in public office and the party central office when seeking to examine the transfer of “assets” (and liabilities) from party X at t to party Y at t +1. Indeed, the very language of assets and liabilities indicates that with empirical research it would be possible when a party like Most-Hid breaks away to derive a measure of continuity for SMK. Admittedly, it would be impossible to produce a fully satisfactory figure, but when the current binary choices are 0% or

100% continuity even an imperfect calibrated figure between those two extremes is likely to be more accurate.

Thirdly, all this points to the need for more extensive and coordinated data collection and for any new models to be able to identify and accommodate the variations across space and time which close observers of party systems often observe (Mustillo and Jung, 2016). We also recognize that future research should also be directed at seeking to capture voter volatility through survey data. We are aware that such data cannot be retrospectively conjured up and that aggregate-level calculations will continue to be important as well, and we believe that methods of measurement which assess mergers, splits and splinters in party systems can also serve as a new basis for understanding parties' institutional change.

Data collection is time-consuming and expensive and even the most extensive data collection will not yield perfect results. Nonetheless, there are many country experts whose expertise and knowledge of the development of party politics in their cases could be pooled and coordinated to yield rich mines of information. Otherwise continued attempts to generate large n volatility scores on the basis of sketchy and inconsistent data – and analyses exploring other political phenomenon using these volatility calculations - will simply continue to produce edifices built on unstable foundations.

References

Achen, C., 2014. Electoral Volatility: Why the Pedersen Index Succeeds. Annual Meeting of the

Midwest Political Science Association, 3-6 April.

Andrews, J. T., Bairett, R.L., 2014. Institutions and the stabilization of party systems in the new democracies of Central and Eastern Europe. Electoral Studies 33, 307-321.

Bågenholm, A., Johansson H., 2009. Still not stable after all these years: The logic of path dependence in party system stability in Central and Eastern Europe. ISA Annual Convention. New York, 15-18 February.

Bakke, E., Sitter, N., 2005. Patterns of Stability. Party Competition and Strategy in Central Europe since 1989. Party Politics 11 (2), 243-263.

Bartolini, S., Mair, P., 1990. Identity, Competition, and Electoral Availability: the Stabilization of European Electorates 1885–1985. Cambridge University Press, Cambridge.

Bielasiak, J., 2001. On the Institutionalization of Party Regimes in Emerging Democracies, Working Paper 351, Glasgow: University of Strathclyde.

Bielasiak, J., 2002. The Institutionalization of Electoral and Party Systems in Postcommunist States. Comparative Politics 34 (1), 189-210.

Bielasiak, J., 2005. Party Competition in Emerging Democracies: Representation and Effectiveness in Post-communism and Beyond. Democratization 12 (3), 331-56.

Birch, S. 2003. Electoral System and Political Transformation in Post-Communist Europe. Palgrave Macmillan, New York.

Chiaramonte, A., Emanuel, V., 2015. Party System Volatility, Regeneration and De-institutionalization in Western Europe (1945-2015). Party Politics Online First, doi: 10.1177/1354068815601330.

Crabtree, C., Golder, M., 2016a. Party System Volatility in Post-Communist Europe, British Journal of Political Science Online First, doi: 10.1017/S0007123415000344.

Crabtree, C., Golder, M., 2016b. Response to Powell and Tucker's 'Little is Known about Party System Volatility in Post-Communist Europe, but We have Interesting New Methods and Data for Studying It'. Available at: http://mattgolder.com/files/research/bjps_response.pdf (accessed on 10 October 2016).

Dassonneville, R., Hooghe, M., 2011. Mapping electoral volatility in Europe. An analysis of trends in electoral volatility in European democracies since 1945. 1st European Conference on Comparative Electoral Research, 1-3 December.

Dassonneville, R., Hooghe, M., 2015. Economic indicators and electoral volatility: economic effects on electoral volatility in Western Europe, 1950–2013. Comparative European Politics Online First, doi: 10.1057/cep.2015.3

Delruelle, N., Evalenko, R., Fraeys, W., 1970. Le Comportement Politique des Électeurs belges. Enquête de Sociologie Électorale. La Rupture de 1965. Ses Prolongements en 1968. Institut de Sociologie de Bruxelles, Brussels.

Epperly, B., 2011. Institutions and Legacies: Electoral Volatility in the Post-communist World. Comparative Political Studies 44 (7), 829-853.

Fraeys, W., 1977. Les Élections Législatives du 17 Avril 1977: Analyse des Résultats. Res Publica 19, 497-513.

Gallagher, M., Laver, M., Mair, P., 2011. Representative Government in Modern Europe. McGraw-Hill, London.

Gunther, R., 2007. Portuguese Elections in Comparative Perspective: Parties and Electoral Behaviour in Southern Europe. In Freire, A., Costa Lobo, M., Magalhães, P. (Eds.) Portugal at the Polls. Lexington Books Lanham, MD, pp. 11-47.

Gwiazda, A., 2015. Democracy in Poland: Representation, Participation, Competition and Accountability since 1989. Routledge, Abingdon and New York.

Ishiyama, J., 2013. Elite Volatility and Change in North Korean Politics: 1997-2010. Journal of Asian and African Studies 49 (5), 570-583.

Katz, R., and Mair, P. 1993. The Evolution of Party Organizations in Europe: The Three Faces of Party Organization. In Crotty, W. (Ed.) Political Parties in a Changing Age, special issue of American Review of Politics 14, 329-45.

Kopeček, L., (2007). Politické strany na Slovensku: 1989 až 2006. Brno: Centrum pro studium demokracie a kultury.

Kuenzi, M., Lambright, G., 2001. Party System Institutionalization in 30 African Countries. Party Politics 7 (4), 437-468.

Lane, J-E., Ersson, S. 2007. Party System Instability in Europe: Persistent Differences in Volatility between West and East? Democratization 14 (1), 92-110.

Lewis, P., 2006. Party Systems in Post-communist Central Europe: Patterns of Stability and Consolidation. Democratization 13 (4), 562-83.

Lindberg, S., 2007. Institutionalization of Party Systems? Stability and Fluidity among Legislative Parties in Africa's Democracies. Government and Opposition 42 (2), 215-241.

Lupu, N., Stokes, S., 2010. Democracy, Interrupted: Regime Change and Partisanship in Twentieth-century Argentina. Electoral Studies 29 (1), 91-104.

Mainwaring, S., España, A., Gervasoni, C. 2016. Extra- and within-system electoral volatility. Party Politics Online First, doi: 10.1177/1354068815625229.

Mainwaring, S., Scully, T.. 1995. Building Democratic Institutions: Party Systems in Latin America. Stanford University Press, Stanford, CA.

Mainwaring, S., and Zoco, E. 2007. Political Sequences and the Stabilization of Inter-party Competition. Party Politics 13 (2), 155-178.

Mair, P. 2007. Party Systems and Alternation in Government, 1950-2000: Innovation and Institutionalization. In Gloppen, S., Rakner L. (Eds.), Globalisation and Democratisation: Challenges for Political Parties. Fagbokforlaget, Bergen, pp. 135-53.

Markowski, R., 2016. The Polish Parliamentary Election of 2015: A Free and Fair Election that Results in Unfair Political Consequences. West European Politics 39 (6), 1311-1322.

Meleshevich, A., 2007. Party Systems in Post-Soviet Countries: a Comparative Study of Political Institutionalization in the Baltic States, Russia, and Ukraine. New York: Palgrave MacMillan.

Millard, F. 2004. Elections, Parties and Representation in Post-communist Europe. Palgrave Macmillan, New York.

Morlino, L. 1998. Democracy between Consolidation and Crisis: Parties, Groups, and Citizens in Southern Europe. Oxford University Press, Oxford.

Mustillo, T., and Jung, Y. S., 2016. Distinguishing Territorial Structure from Electoral Adventurism: The Distinct Sources of Static and Dynamic Nationalization. Electoral Studies 44, 341-350.

Ocaña, F. 2007. An Approximation Problem in Computing Electoral Volatility. Applied Mathematics and Computation 192 (2), 299–310.

O'Dwyer, C. , Kovalčík, B. 2007. And the Last Shall be First: Party System Institutionalization and Second-Generation Economic Reform in Postcommunist Europe. Studies in Comparative

International Development, 41 (4), 3-26.

Pedersen, M. 1979. The Dynamics of European Party Systems: Changing Patterns of Electoral Volatility. European Journal of Political Research 7 (1), 1-26.

Powell, E. and Tucker, J. 2014. Revisiting Electoral Volatility in Post-communist Countries: New Data, New Results and New Approaches. British Journal of Political Science 44 (1), 123-147.

Powell, E. and Tucker, J. 2016. Little is Known about Party System Volatility in Post-Communist Europe, but We Have Interesting New Methods and Data for Studying It. British Journal of Political Science Online First, doi: 10.1017/S0007123415000769.

Przeworksi, A., 1975. Institutionalization of Voting Patterns, or Is Mobilization the Source of Decay? American Political Science Review 69 (1), 49-67.

Roberts, K., and Wibbels, E. 1999. Party Systems and Electoral Volatility in Latin America: A Test of Economic, Institutional, and Structural Explanations. American Political Science Review 93 (3), 575–590.

Sikk, A., 2005. How Unstable? Volatility and the Genuinely New Parties in Eastern Europe. European Journal of Political Research 44 (3), 391-412.

Tavits, M. 2005. The Development of Stable Party Support: Electoral Dynamics in Post-Communist Europe. American Journal of Political Science 49 (2), 283-298.

Tavits, M. 2008. On the Linkage between Electoral Volatility and Party System Instability in Central and Eastern Europe. European Journal of Political Research 47 (5), 537-55.

Tóka, G., and Henjak, A. 2007. Party Systems and Voting Behaviour in the Visegrad Countries 15 Years after Transition. In Šaradín, P., Bradová, E. (Eds.) Visegrad Votes: Parliamentary Elections 2005-2006. Palacky University Press, Olomouc. pp. 212-244

Weghorst, K., Bernhard, M. 2014. From Formlessnesss to Structure? The Institutionalization of Competitive Party Systems in Africa. Comparative Political Studies 47 (12), 1707-1737

Table 1 Electoral volatility in 10 EU post-communist democracies (1990-2013)

Country	Author(s)[1]	Electoral volatility during election period				
		1st	2nd	3rd	4th	5th
Bulgaria	Bielasiak	17.5	22.8	-	-	-
	Birch	10.0	22.0	-	-	-
	Dassonneville & Hooghe	17.8	25.7	50.5	35.3	49.8
	Lane	21.1	23.1	47.9	-	-
	Powell & Tucker	21.7	39.3	57.2	41.2	54.5
	Sikk	19.1	24.6	-	-	-
Czech Republic	Bågenholm	24.2	14.4	8.6	17.9	33.7
	Bakke & Sitter	24.2	16.3	13	-	-
	Bielasiak	18.5	7.4	-	-	-
	Birch	16.1	7.6	-	-	-
	Dassonneville & Hooghe	24.2	17.2	11.5	17.1	36.6
	Lane	27.0	9.3	16.3	-	-
	Mainwaring et al	31.4	18.5	15.7	-	-
	Powell & Tucker	38.5	16.7	19.4	17.3	-
	Sikk	27.0	15.8	-	-	-
	Tóka & Henjak (Raw)	31.4	18.6	30.9	28.5	-
	Tóka & Henjak (Adjusted)	29.2	16.4	16.6	20.9	-
Estonia	Bågenholm	39.6	22.1	34.1	22.1	-
	Bielasiak	28.4	23.4	42.4	-	-
	Birch	13.2	12.1	-	-	-
	Dassonneville & Hooghe	45.4	27.8	26.6	23.1	13
	Lane	27.9	33.5	35	-	-
	Mainwaring et al	55.6	42	36.4	-	-
	Meleshevich	68.0	49	35.7	-	-
	Powell & Tucker	62.9	44.4	40.8	39.7	-
	Sikk	21.3	24.1	-	-	-
	Bågenholm	22.3	30.4	19.7	4	32.8
	Bakke & Sitter	26.8	31.6	19.1	-	-

[1] Bågenholm 2009; Bakke and Sitter 2005; Bielasiak 2002 and 2005; Birch 2003; Dassonneville and Hooghe 2011; Gwiazda 2015; Lane and Ersson 2007; Mainwaring *et al*. 2016; Markowski 2016; Meleshevich 2007; Powell and Tucker 2014; Sikk 2005; Tóka and Henjak 2007 (original data files and answers to process questions provided in personal follow-up conversation).

Hungary[a]	Bielasiak	26.4	28.5	22.5	-	-
	Birch	23.9	29.4	19.1	-	-
	Dassonneville & Hooghe	44.5	31.6	21.8	8	32
	Lane	26.1	31.4	18.1	-	-
	Mainwaring et al	28.9	33.7	25.1	-	-
	Powell & Tucker	40.9	26.2	28.3	10.4	-
	Sikk	23.7	32.7	-	-	-
	Tóka & Henjak (Raw)	28.3	33.6	54.3	49.5	-
	Tóka & Henjak (Adjusted)	25.8	31.7	18.3	8.4	-
Latvia	Bågenholm	39.8	43.7	44.8	27.3	-
	Bielasiak	33.5	24.5	51.2	-	-
	Birch	25.7	23.0	-	-	-
	Dassonneville & Hooghe	42.1	58.1	49.6	26.3	29.4
	Lane	37.9	45	46.1	-	-
	Mainwaring et al	47.2	56.6	52.1	-	-
	Meleshevich	55.1	59.3	72.6	-	-
	Powell & Tucker	54.8	58.3	62.3	26.6	-
	Sikk	36.6	45.2	-	-	-
Lithuania[a]	Bågenholm	30.4	39.9	50	31.6	-
	Bielasiak	28.9	49.6	-	-	-
	Birch	20.4	36.7	-	-	-
	Dassonneville & Hooghe	35.8	48.8	50.8	33.4	-
	Lane	37.4	48.3	50	-	-
	Meleshevich	64.0	74.1	-	-	-
	Powell & Tucker	73.1	71.8	85.1	47.0	-
	Sikk	35.9	48.5	-	-	-
Poland	Bågenholm	24.4	15.5	36.2	34.1	24.4
	Bakke & Sitter	34.3	21.3	51.1	-	-
	Bielasiak	26.8	22.3	43	-	-
	Birch	17.9	12.3	30.0	-	-
	Dassonneville & Hooghe	34.4	31	51.3	37.9	24.4
	Gwiazda	28.5	16.2	44.8	31.6	21.5
	Lane	19.4	23.6	34.3	-	-

	Mainwaring et al	35.4	51.5	56.5	38.7	-
	Markowski	34.8	19.2	49.3	38.4	25
	Powell & Tucker	41.6	64.8	54.4	35.3	34.1
	Sikk	31.7	21.1	-	-	-
	Tóka & Henjak (Raw)	47.4	67.0	83.5	52.4	-
	Tóka & Henjak (Adjusted)	34.5	19.5	51.6	41.1	-
Romania	Bågenholm	14.4	30.3	19.9	19.7	-
	Bielasiak	12.4	21	-	-	-
	Birch	14.0	16.6	-	-	-
	Dassonneville & Hooghe	15.8	30.5	20.8	21.5	-
	Lane	16.8	34.7	12.1	-	-
	Mainwaring et al	29.1	47.6	36.1	-	-
	Powell & Tucker	49.2	52.6	30.9	49.6	-
	Sikk	14.3	29.1	-	-	-
Slovakia	Bågenholm	20.5	22.1	32.9	20.8	26.1
	Bakke & Sitter	20.6	20.3	30.3	-	-
	Bielasiak	13.3	20.1	-	-	-
	Birch	12.8	9.4	-	-	-
	Dassonneville & Hooghe	31.1	33.7	60.9	24.6	38.5
	Lane	13.7	21.5	28.6	-	-
	Powell & Tucker	45.0	61.0	68.9	39.6	-
	Sikk	13.6	20.2	-	-	-
	Tóka & Henjak (Raw)	43.9	53.9	63.1	33.0	-
	Tóka & Henjak (Adjusted)	23.7	21.4	29.9	32.9	-
Slovenia	Bågenholm	22.0	27.5	23.3	35	-
	Bielasiak	25.4	22.1	-	-	-
	Birch	-	17	-	-	-
	Dassonneville & Hooghe	31.3	27.4	21.4	33.5	-
	Lane	24.4	18.7	22.2	-	-
	Powell & Tucker	39.9	34.4	52.2	41.3	-
	Sikk	23.8	18.8	-	-	-
	Tóka & Henjak (Raw)	39.4	19.7	31.3	-	-

[a]Although it is not always clear from the published sources, our conversations with authors indicate that all of these measures refer to the proportional representation segments within these mixed systems.

Table 2. Volatility calculation Values for hypothetical Party A according to various threshold inclusion methods.

Threshold inclusion method	**Threshold example**	**Data category**	**Election 1**		**Election 2**		**Election 3**		**Election 4**	**Average contribution to volatility**
No threshold	0%	Actual Value	2%		7%		1%		4%	
		Change		*+5*		*-6*		*+3*		*4.7*
Individual points	3%	Adjusted value	0%*		7%		0%*		4%	
		Change		*+7*		*-7*		*+4*		*6.0*
	5%	Adjusted value	0%*		7%		0%*		0%*	
		Change		*+7*		*-7*		*0*		*4.7*
Transition pairs	3%	Adjusted value	2%		7%		1%		4%	
		Change		*+5*		*-6*		*+3*		*4.7*
	5%	Adjusted value	2%		7%		1%		0%*	
		Change		*+5*		*+6*		*-1*		*4.0*
Blanket inclusion	3%	Adjusted value	2%		7%		1%		4%	
		Change		*+5*		*-6*		*+3*		*4.7*
	5%	Adjusted value	2%		7%		1%		4%	
		Change		*+5*		*-6*		*+3*		*4.7*
Maximum difference between methods	3%	Value	2%		0%		1%		1%	
		Change		*2*		*1*		*1*		
	5%	Min. Value	2%		0%		1%		4%	
		Change		*2*		*1*		*4*		

*Value recalculated from actual election value due to threshold inclusion method.

Table 3 (previously table 4. Re-numbered as table 3 moved to the appendix): The relative importance of linkage, threshold and exclusion [column model 3 removed]

Model		**Model 1**	**Model 2**
Description		**Basic Model**	**Basic +Interactions**
Number of observations		3,960	3,960
R-squared		0.673	0.676
Root MSE		0.082	0.082
Constant		0.200***	0.188***
		(0.006)	(0.007)
Continuity method (baseline = Inclusive aggregation)	Relaxed linkage	0.048***	0.070***
		(0.003)	(0.005)
	Strict linkage	0.152***	0.174***
		(0.004)	(0.007)
Threshold size		-0.809***	-0.358
		(0.082)	(0.197)
Threshold inclusion method (baseline = Individual points)	Transitional pairs	-0.003	0.002
		(0.003)	(0.006)
	Blanket inclusion	-0.002	0.001
		(0.003)	(0.006)
Excluded data method	Aggregated linked other	0.013***	0.002
		(0.003)	(0.005)
Interaction: Threshold size with Continuity method (baseline = aggregation)	Relaxed linkage	-	-0.821***
		-	(0.160)
	Strict linkage	-	-0.901***
		-	(0.228)
Interaction: Threshold size with Threshold inclusion method (baseline = Individual points)	Transitional pairs	-	-0.200
		-	(0.201)
	Blanket inclusion	-	-0.098
		-	(0.202)
Interaction: Threshold size with Excluded data method	Aggregated linked other	-	0.435**
		-	(0.164)

Note: OLS with Robust Standard Errors (in brackets). ** $p < .01$; *** $p < .001$

Table 4. (all new, added at the suggestion of reviewer #4) Overall summary of volatility related options and recommendations

Problem	Question	Method	What the Method Does	Effect of Method	Recommendation
Dealing with change	How to treat continuity over time when parties change?	**Inclusive aggregation**	Measures continuity between a party and the sum of all successors or predecessors and ignore name changes	**Produces low volatilty** magnitude because it avoids major institutional shifts in parties.	**Authors should understand and specify the logic and impact of the method they choose.** Each method reflects a distinct understanding of the meaning of change and none is inherently preferable, but the choice of method has a major impact on outcomes and must be specified and justified to prevent confusion. Comparing the results of more than one method may enrich discussion by helping to identify problem areas. Further research is needed on non-binary approaches that reflect partial continuity.
		Relaxed linkage	Accepts continuity between a party and the successor or predecessor with the largest vote share regardless of name change	**Produces higher volatility** magnitude than inclusive aggregation because it is affected to some degree by institutional shifts	
		Strict linkage	Accepts continuity only for parties of the same name and organizational continuity.	**Produces the highest volatility** magnitude because it is extremely sensitive to institutional shifts.	
Dealing with size	Where to set the threshold for excluding very small data points?	**Threshold Level**	Excludes data points below a chosen level which can be set at any point from zero up	**Produces lower volatility magnitude at higher thresholds** because they ignore movement below the threshold.	**Lower thresholds are better than larger ones** because they are less likely to underestimate volatility, but very low thresholds impose very high data-gathering costs for small parties. **A threshold of 0.5% or 1.0%** keeps data-gathering costs reasonable without much loss of accuracy.
	How to deal with data points below the threshold for parties that sometimes rise above the threshold?	**Only Individual points**	Includes only data points above threshold and sets all others to zero	**Minimizes data costs** by looking only above the threshold, and needs no retro-active calculation **but artificially introduces volatility**.	**Transition pair and Blanket inclusion methods are both reasonable choices** because they avoid the tendency of the Individual points method to artificially introduce volatility. Between the two, the Transition Pair method has lower data-gathering costs but a slightly higher chance of introducing small amounts of volatility, while the Blanket inclusion method comes closest to a zero-threshold model with its higher data-gathering costs. Both may require some retro-active recalculation with the addition of new data.
		Transition Pairs	Includes all data points above the threshold and all data points immediately preceding and following those in time.	**Keeps data costs relatively low** by avoiding the most obscure periods of small parties, but may lead to counterintuitive results in rare cases.	
		Blanket Inclusion	Includes all data points for a party if that party exceeds the threshold even once	**Maximizes data cost**, especially if threshold is very low, but produces results closest to zero threshold	
	What to do with data points that do not meet threshold requirements?	**Aggregation**	Aggregates all vote shares excluded by the threshold into a single "other" and treat it as a party in its own right	**Acknowledges all available data, but does not serve as a reliable indicator** unless all changes point in the same direction.	**Exclusion method is more reliable** though it may produce a slight under-estimate of volatility. The Aggregation method is not as reliable because its unified "other" treats multiple independent parties as a single unit and its composition varies in unspecified ways over time.
		Exclusion	Removes from subsequent calculation all vote shares excluded by the threshold	**Avoids unreliable results by ignoring data** below the threshold.	

Table A for Appendix (an amended version of Table 3 in the original submission). Method choices of authors in Pedersen Index calculations: move to appendix

Author	Dealing with size			Dealing with change		
	Inclusion standards and method			Continuity method		
	Threshold for inclusion?	**Threshold inclusion method**	**Excluded data method**	**How to determine successor in case of party split**	**How to determine predecessor in case of party merger**	**Categorization**
Bågenholm	Election threshold	Individual points	Aggregated and linked other	Party of same name; if not, no successor	Sum of predecessors	Mixed method (Inclusive aggregation for predecessor, Relaxed linkage for successor)
Bakke	2%	Transitional pairs	Aggregated and linked other	Sum of successors	Sum of predecessors	Inclusive aggregation
Bielsiak	No	-	Aggregated and linked other	Sum of successors	Sum of predecessors	Inclusive aggregation
Birch	No	-	Aggregated and linked other	Party of same name; if not, no successor	Party of same name; if not, merger is new	Strict linkage
Dassonenville & Hooghe	1%	Individual points	Aggregated and linked other	Sum of successors	Sum of predecessors	Inclusive aggregation
Gwiazdka	No	-	None	Largest offspring party	Sum of predecessors	Mixed method (Inclusive aggregation for predecessor, Relaxed linkage for successor)
Lane & Ersson	No	-	Aggregated and linked other	Sum of successors	Sum of predecessors	Inclusive Aggregation

Mainwaring, España & Gervasoni	No	-	None	Party of same name; if not, largest offspring	Party of same name; if not, largest parent*	Relaxed linkage
Markowski	No	-	None	Sum of successors	Sum of predecessors	Inclusive aggregation
Meleshevich	No	-	None	Author chooses successor	Author chooses predecessor	Mixed linkage (Relaxed linkage and Strict linkage)
Powell & Tucker	2%	Individual points	Uncalculated	Party of same name; if not, no successor	All mergers new unless under 5%	Strict linkage
Sikk	No	-	Aggregated and linked other	Sum of successors	Sum of predecessors	Inclusive aggregation
Tóka (raw)	No	-	None	Party of same name; if not, largest offspring	Party of same name; if not, largest parent	Relaxed linkage
Tóka (adjusted)	No	-	None	Sum of successors	Sum of predecessors	Inclusive aggregation

*All authors use an inclusive aggregation approach for electoral coalitions except for Mainwaring, España and Gervasoni, who in this case adopt a relaxed linkage approach by using the largest predecessor party.

Figure 1 Averages of pair-wise differences in election-period volatility results according to author.

Figure 2a. Volatility in the Czech Republic, 1990-2013

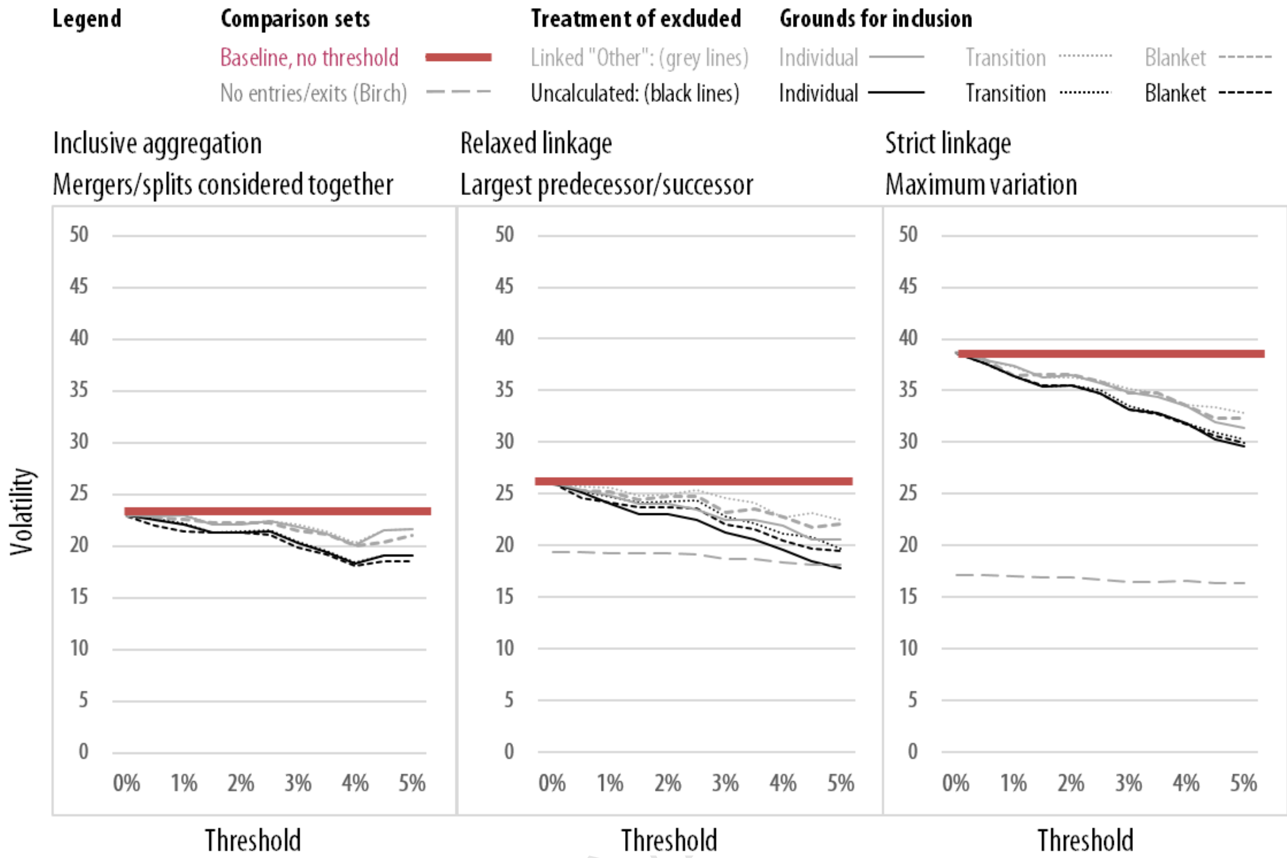


Figure 2b. Volatility in Hungary, 1990-2014

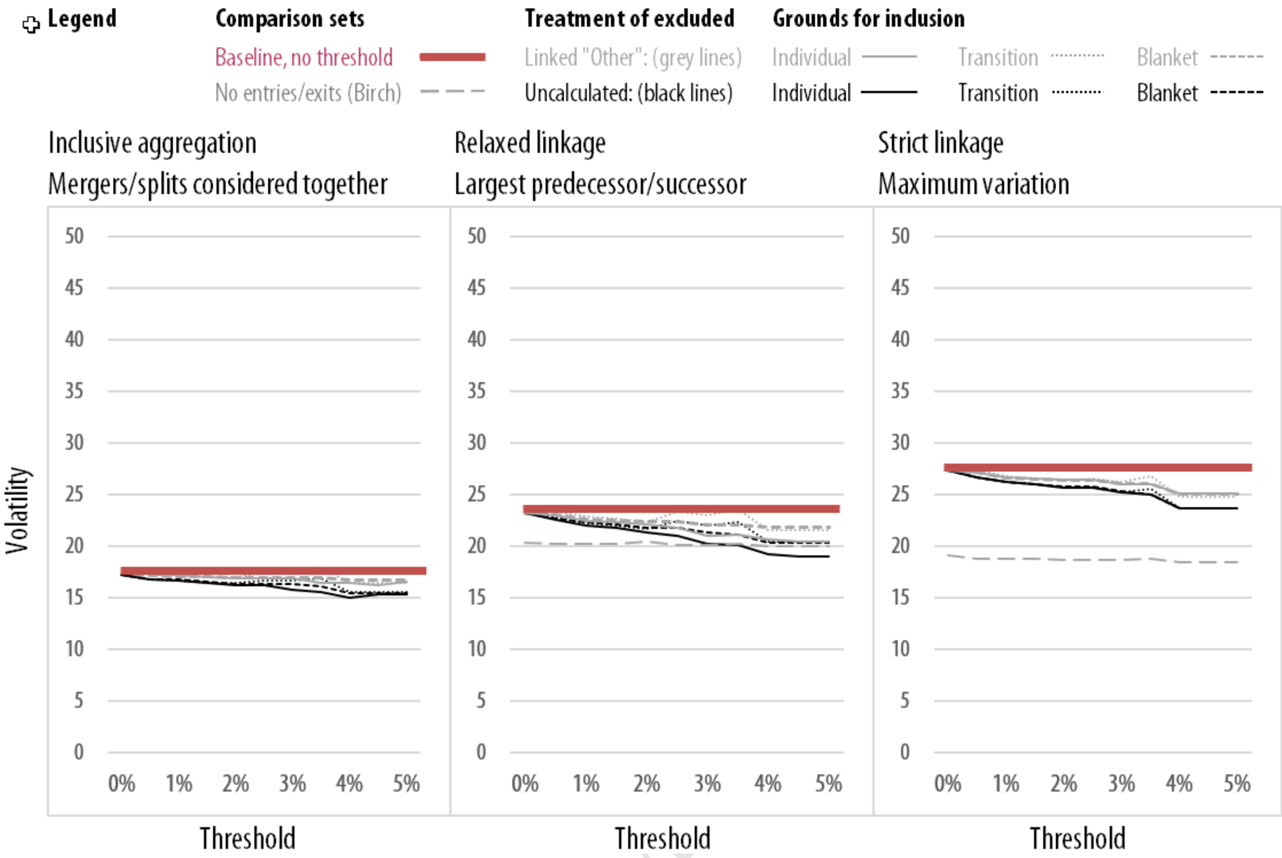


Figure 2c. Volatility in Slovakia, 1990-2012

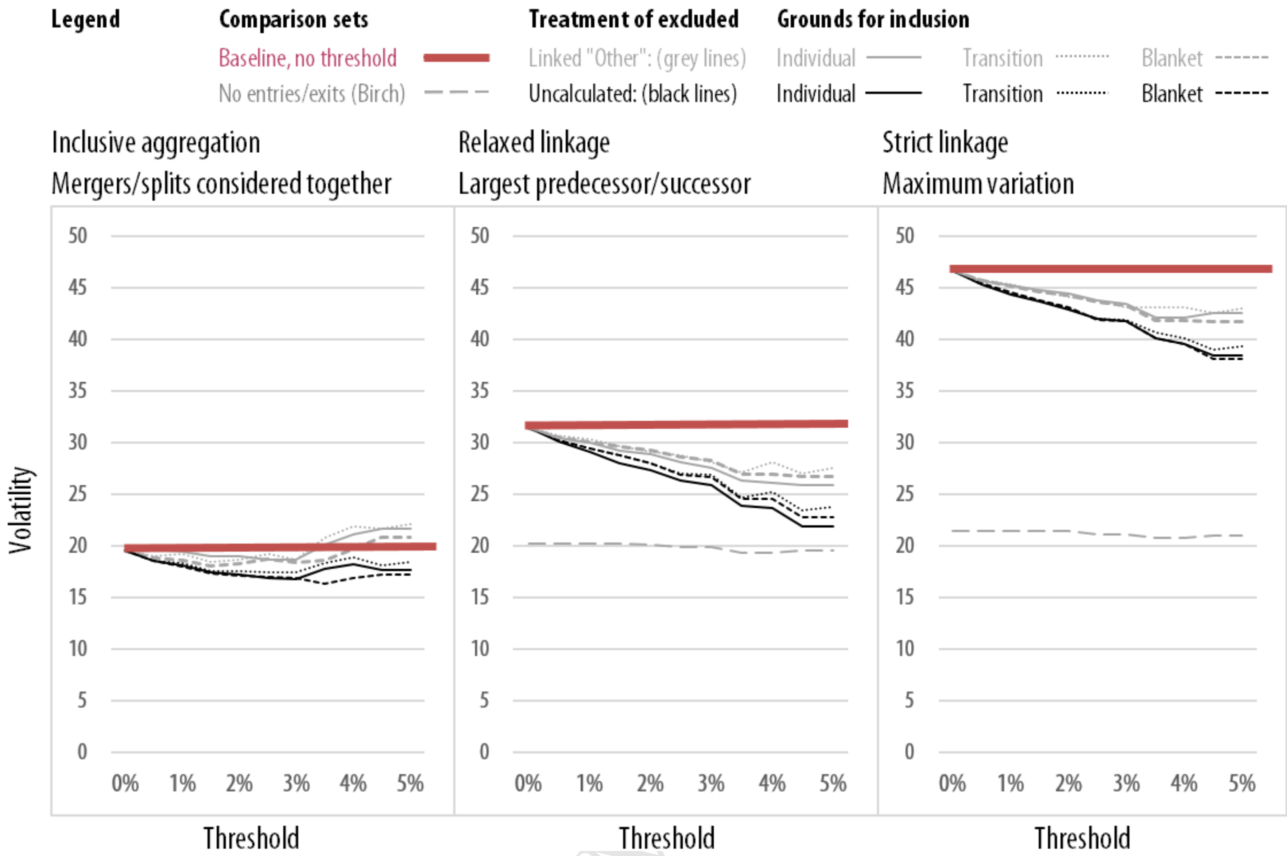


Figure 3 The Inclusive aggregation method's "Z-effect"

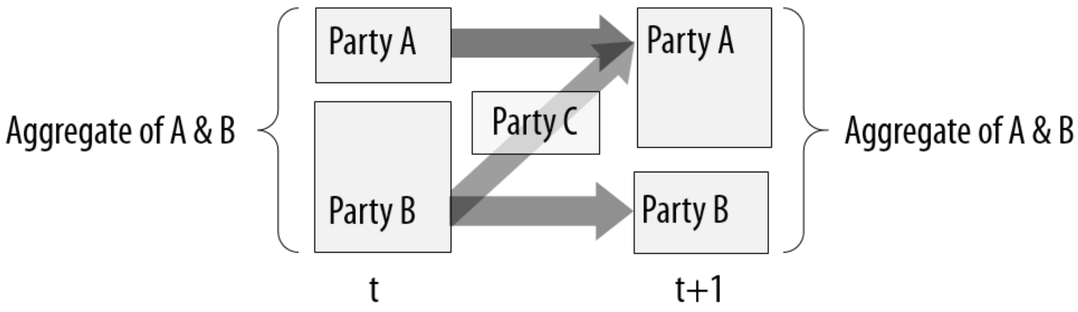


Figure 4 Volatility scores for Inclusive aggregation and Strict linkage methods in the Czech Republic, Hungary and Slovakia, 1990-2014.