Comparing votes and seats with cosine, sine and sign, with attention for the slope and enhanced sensitivity to inequality / disproportionality

Colignatus, Thomas

Thomas Cool Consultancy Econometrics

9 February 2018

Comparing votes and seats with cosine, sine and sign, with attention for the slope and enhanced sensitivity to inequality / disproportionality

Thomas Colignatus
September 16 2017 & February 9 2018

JEL
A100 General Economics: General
D710 Social Choice; Clubs; Committees; Associations,
D720 Political Processes: Rent-seeking, Lobbying, Elections, Legislatures, and Voting Behavior
D630 Equity, Justice, Inequality, and Other Normative Criteria and Measurement

MSC2010
00A69 General applied mathematics
28A75 Measure and integration. Length, area, volume, other geometric measure theory
62J20 Statistics. Diagnotics
97M70 Mathematics education. Behavioral and social sciences

Abstract

Let *v* be a vector of votes for parties and *s* a vector of their seats gained in the House of Commons or the House of Representatives. We use a single zero for the lumped category of "Other", of the wasted vote, for parties that got votes but no seats. Let $V = 1'v$ be total turnout and $S = 1's$ the total number of seats, and $w = v / V$ and $z = s / S$ the perunages (often percentages). There are slopes *b* and *p* from the regressions through the origin (RTO) $z = b\, w + e$ and $w = p\, z + \varepsilon$. Then k = Cos[v, s] = Cos[w, z] = Sqrt[$b\, p$]. The geometric mean slope is a symmetric measure of similarity of the two vectors. θ = ArcCos[k] is the angle between the vectors. Thus Sin[v, s] = Sin[w, z] = Sin[θ] = Sqrt[1 – $b\, p$] is metric and a measure of disproportionality in general. Geometry appears to be less sensitive to disproportionalities than voters, representatives and researchers tend to be. This likely relates to the Weber-Fechner law. Covariance gives a sign for majority switches. A disproportionality measure with enhanced sensitivity for human judgement is the *sine diagonal inequality/disproportionality* SDID = *sign* 10 √Sin[v, s]. This puts an emphasis on the first digits of a scale of 10, which can be seen as an inverse (Bart Simpson) report card. What does disproportionality measure ? The unit of account can be either the party or the individual representative. This distinguishes between the party average and the party marginal candidate. The difference $z - w$ is often treated as a level, and Webster / Sainte-Laguë (WSL) uses the relative expression $z / w - 1$. For the party marginal candidate $z - w$ already is relative, with the unit of account of the individual representative in the denominator. The Hamilton Largest Remainder (HLR) apportionment has the representative as the unit of account. The "Representative Largest Remainder" (RLR) uses a 0.5 natural quota. The paper provides (i) theoretical foundations, (ii) evaluation of the relevant literature in voting theory and statistics, (iii) example outcomes of both theoretical cases and the 2017 elections in Holland, France and the UK, and (iv) comparison to other disproportionality measures and scores on criteria. Using criteria that are accepted in the voting literature, SDID appears to be better than currently available measures.
.

Keywords

General Economics, Social Choice, Social Welfare, Election, Majority Rule, Parliament, Party System, Representation, Sine Diagonal Inequality / Disproportionality, SDID, Proportion, District, Voting, Seat, Euclid, Distance, Cosine, Sine, Gallagher, Loosemore-Hanby, Sainte-Laguë, Webster, Gini coefficient, Lorenz, Jefferson, Hamilton, Largest Remainder, Correlation, Diagonal regression, Regression through the origin, Apportionment, Disproportionality, Equity, Inequality

1. Introduction

1.1. Overlapping chapters in Social Choice and Public Choice

This paper falls in the category of Public Choice, as defined by Gordon Tullock (2008) as *"the use of economic tools to deal with traditional problems of political science"* (JEL D710 and D720). We will also present a new disproportionality index based upon cosine and sine, which falls in a different category (JEL D630). Overall the general content is large (JEL A100).

Economics has a main stream that adopts "methodological individualism". This assumes that explanations are sought in the behaviour of individuals, as opposed to assuming units like *The Family* or *The State*. In Idealism, for example, one might imagine that The State is "an entity" with a life of itself. This would be Idealism and not economics. Similarly for the question whether The Party can be a unit of account in votes and seats. In the USA there is apportionment of seats in the House of Representatives for the *States* of the Union, with the current method of Huntington – Hill. [1] Populations link up with States and it makes sense to use these as units of account. [2] For parties it is different. These are voluntary coalitions of voters and their representatives. Methodology requires that we look at the individual unit of account rather than The Party. In this case, the unit thus is the *party marginal candidate.* For economists, the notion of marginal analysis may have an additional appeal.

Politicians, their bureaucracies and the scientists and mathematicians hired for such purposes, have been adopting particular conceptions or psychological frames, which in science might be called a paradigm. In this case, it is the adoption of The Party as the unit of account. This causes such collusion and confusion, that perhaps even scientists no longer are fully aware of what is happening. Thus we need to construct both (i) what really is at stake and (ii) what that paradigm is, that causes such confusion about what really is at stake.

There is the distinction between ***Equal or Proportional Representation*** (EPR) and ***District Representation*** (DR). Carey & Hix (2009) (2011) discuss the optimal district magnitude, i.e. the number of seats per district. They perceive a trade-off or balance of accountability and representation, and rely on the Euclid / Gallagher Inequality / Disproportionality (EGID) measure for their regressions. There are various issues with their analysis that receive attention in Colignatus (2017d). For this present paper, it suffices to say that C&H overlook that DR appears to be fraught with error and mishap, and that EGID is somewhat dubious. Because of the mishap of DR, Holland in 1917 switched to EPR. Somehow, the UK, USA and much of the *English Speaking World* (like India) missed the boat. This has been causing a lot of trouble to the world. What causes it, that scientists do not clearly explain to politicians and the public that DR is disastrous, and that EPR is the approach that is required by both intellectual integrity and social harmony ? Apparently Carey & Hix (2011) point to accountability, but as said, this is discussed by Colignatus (2017d). The present focus is on inequality / disproportionality measures, that form part of this paradigm of confusion.

Indeed, in the UK there is the Electoral Reform Society (ERS), that advocates a system called "Single Transferable Vote" (STV), that they call EPR but that is not EPR because it still links up to the UK tradition of DR. Such notions at ERS were historically caused by the same concerns that caused Holland in 1917 to switch to EPR. Thus there is scope for progress. It is tragic that this scope gets bogged down in other swamps, in this case at ERS in the confusion about STV. (A confusion is too that voters should use such more complicated methods. Instead, voters best vote for an EPR party list system, while the professionals in parliament might use more complex systems.)

We now look at a part in the literature about measures of inequality / disproportionality. It appears that reseachers have built a cathedral of misconceptions here as well. We need to deconstruct this cathedral and show what really is at stake.

---

[1] Young (2004) in a symposium at the US Census Bureau argues rather convincingly that advanced insight shows that Webster would be better. See also Balinski & Young (1980).
[2] Alexander Bogomolny has online routines: https://www.cut-the-knot.org/Curriculum/SocialScience/SocialChoice.shtml

In the USA, interest for disproportionality only concerned the allocation of seats to *States*, as said, which resulted into the Huntington – Hill method (see footnote 1). The USA has no interest in EPR itself since they have a system with DR. When Arend Lijphart moved from Holland to the USA, he took along a support for the notion of EPR, which he even called *electoral justice*. However, he may also have been sensitive to his new environment of DR, for apparently Lijphart also started to embrace The Party, and thus may have overlooked a key aspect of EPR that we now will highlight, namely *equal representative democracy*.

Obviously, when we would inform the decision makers and public in the USA, UK or France about the advantages of EPR, then the system should be a proper improvement over their current systems. Even Holland is not as EPR as can be.

We consider a vector v of votes, with a total number of votes $V = 1'v$ (turnout). The length of the vector is the number of parties n. The vector s of the seats has a total number of seats $S = 1's$. There are shares or proportions $w = v / V$ and $z = s / S$, unitised to 1 onto the unit simplex so that $1'w = 1'z = 1$. A common use is 100 percent, but this may suggest overly precision, and 10 suffices to get rid of the leading zeros.

Table 1 contains an example. It shows how V = 325 voters elect S = 10 seats. The electoral quota is $Q = V / S$. Dividing the votes by the quota gives a as the average number of seats that a candidate can claim, $a = v / Q = S v / V = S w$. The locus of apportionment of seats s is Floor[a] ≤ s ≤ Ceiling[a], or the averages rounded down or up to the nearest integer. The first batch of seats that can be allocated is Floor[a]. This generates 9 seats that can be assigned directly. There is 1 remainder, and at issue is what principle shall be used to assign this. At issue is not the particular method, but the *principle.* What defines *equal (proportional) representative democracy* ? The question is whether the unit of account is the *party* or the *individual representative,* here the party marginal candidate. The method of *highest averages* would better be called the method of *highest artificial ratios,* since Methodological Individualism requires attention for the individual level and not the assumption of The Party.

Table 1. Unit of account: Voter and (representative versus party)

Methodological Individualism versus *Party as a unit of account*				Largest remainder (marginal candidate)		"Highest average" (party as unit)	
Party	Votes *v*	*v* / *Q*	Floor[*v/Q*]	Remainder	Seats *s*	*v* / (Floor+1)	Seats *s*
John	25	0.77	0	0.77	1	25.00	0
Blue	100	3.08	3	0.08	3	25.00	3
Green	200	6.15	6	0.15	6	28.57	7
Total (*V, S*)	325	10	9		10		10
Q = *V* / *S*	32.5	(Electoral quota)					

Let there be K candidates with v_K votes for the candidates. Elected are s_K. There are n parties. Then G is a {n, K} aggregation matrix with 0 and 1, so that $v = G\, v_K$ and $s = G\, s_K$ are the aggregate outcomes for parties, in which an independent individual can be called a party by himself or herself. This administration should not cause us to use The Party as the unit of account. For we should rather look at the *party marginal candidate*.

Comments w.r.t. **Table 1** are:

(1) In Methodological Individualism, individual candidate John is not opposed to the other parties, but to the candidates in those parties. The remainder seat does not concern the parties but their marginal candidates. The party list determines which particular candidate.

  (a) The relevant principle is "votes per candidate" or the *Remainder Divided By 1 Candidate*. In voting research this is often seen as a level difference, but it is a ratio, with the marginal candidate in the denominator.
  (b) The appropriate principle is that of the Largest Remainder (Hamilton, HLR, but modified in **Appendix L** to *Representative LR* (RLR)). This is often seen and judged as a "method" but then one overlooks that it is a principle.

(c) There are K candidates and V / K votes per candidate on average. The parties manage to collaborate, and collect their remainders, so that more of them can be elected. Their *benefit* should not be an argument to cause *a loss to others*.
(d) It is a difficult issue how to handle inequal votes within a party, but it is accepted practice, and indeed not unreasonable for the formation of coalition governments, that parties can combine their votes into the Floor value. The Floor apportionment fully satisfies the quota and thus can drop from inequality / disproportionality measures, so that these measures can concentrate only on the remainder seats.
(e) It would be another step to accept that parties can also claim more than this. As said, their advantage should not be construed to cause a loss to others.
(f) What about "extreme situations" ? If John got only 2 votes and the other parties had only marginal votes 1 each, then HLR would cause him to be elected (if the electoral quota allowed such small margins, while taking into account that the fractions must add up to the remainder seats). The votes for the larger parties indicate that their remainders are somewhat part of chance processes, while John faces the existential question. To deal with this kind of issue, it is not unreasonable to set a threshold of say 50% of the electoral quota, applicable only to the allocation of the remainder seats. The number is arbitrary, except for the expectation that fractional remainders might tend to average around 50% of a seat. The norm that the fraction ≥ 0.5 fits the convention to round up (compared to the strictness of > 0.5).
(g) This situation also clarifies the rise of parties. Perhaps a short example helps. If such a threshold would be 0.5 Q and there would be 21 candidates who all got a 10 / 21 share of the vote, then none would pass (325 / 21 / Q). There is an obvious advantage that there are parties who collect the remainders. But an advantage for parties should not necessarily work against those not in a party system, who do not need to collaborate and still get at least 0.5 Q. Said in another way: with such 10 / 21 outcome, there would arise bargaining amongst the candidates, who would be elected, and this explains the rise of parties. Yet this should not merit the adoption of The Party as the unit of account.

(2) When parties are taken as the unit of account (with similar methods as taking States as the unit of account), then it makes some sense to look at the *average* votes per seat.

(a) The Floor and the assigned seats already satisfy the electoral quota Q. For each party we can look at what would happen to the ratio of "votes per seat" when that party would get another seat. John has 25 votes and thus a ratio of 25. The Green party has a small remainder of 5 votes, with 5 / 32.5 ≈ 0.15, but the party is so large that the addition of 1 seat still generates a high *average* of "votes per seat" *for that particular party.* In this case Green generates the highest average. This philosophy assigns the remainder to them.
(b) Sometimes the party system is enhanced by requiring a higher threshold. If the number of seats is large then there is much to be said for the natural threshold of the electoral quota. Yet case (1) above shows that candidates below the natural quota can do better than the marginal candidates in parties. One should have a very good reason to exclude the small independent voice.
(c) The argument in favour of parties is that they represent an ideology, and that voting rather concerns party programs than individual representatives. Perhaps. It is not clear though that this is a good reason to exlude the small independent voice that has a higher vote than the party marginal candidate.

For example, Balinski & Young (1976:17-18) present above party-quota rule. What they effectively do is to replace Hamilton's *"votes per seat"* into *"votes per party seats"*. They argue correctly that a choice on electoral systems should be based upon criteria, but they present criteria based upon *parties* and not upon *individual representatives*. They state:

> "In practice one finds that Q [their label for this method] has a tendency to produce solutions that round up the exact quotas of large parties more often than those of small parties. This seems reasonable for the application of Q for proportional

> representation systems in that it inferentially asks for a "large" vote before according any representation at all."

They think in terms of The Party and overlook Methodological Individualism. Their background is in mathematics and not in economic theory.

Appendix L shows that the issue is relevant for Holland 2017. The Dutch system is perhaps the most proportional in the world but can still be improved. **Appendix M** shows the difference between HLR (without threshold) and RLR (with at least 0.5 *Q* for *remainder seats*).

The latter appendix also shows the effect of the number of seats *S.* The literature on inequality / disproportionality measures tends to compare relative allocations of votes and seats, i.e. *w* and *z,* but the number of seats *S* likely affects our interpretation of the behaviour of the measures. Given that HLR / RLR rather *define* equal (proportional) representative democracy, for national parliaments of adequate size, we should not be distracted by supposed paradoxes highlighted by examples with low numbers of seats. For countries with DR, this issue on the number of seats is discussed under the label of "district magnitude", while the relevant discussion is actually about replacing DR with EPR.

The focus on properties of different methods of apportionment is also overdone, while the real discussion concerns the shift from DR to EPR. Beumer (2010:57) looks at 16 Dutch elections in th 50-year period of 1956-2006, and finds for the Dutch House of 150 seats:

> "Coalitions [ftnt] are always formed with a clear margin over the simple majority of 75 seats. Webster apportionments only shift a small number of seats: on average the difference between a Webster apportionment and a Jefferson apportionment is 3.88 seats on a total of 150. Additionally, the difference is virtually always one seat per party; in the period 1956–2006 there is only one occurrence of a two-seat difference. This means that an average coalition of three parties could typically lose three seats. This is not enough to lose the majority. In particular, Webster with minimum requirements would have lead to exactly the same coalitions."

Once you discover a statistical regularity, it is in danger of evaporating. It so happens that the new coalition in Holland 2017 has 4 parties with only 1 seat majority. If apportionment were a real conundrum, then one might also assign representatives the weights of the votes that they actually got, above some threshold. There is no urgent need for this, since HLR / RLR work well, and rather define what *equal (proportional) representative democracy* actually is.

The Political Science literature on apportionment and inequality / disproportionality measures is huge, and they all tend to overlook the above and to adopt the paradigm of The Party. For science it is okay to look at the consequences of a particular hypothesis. However, this literature also comes with the attitude that researchers would be interested in inequality / disproportionality *per se.* Apparently only within their paradigm of The Party. For authors coming from countries with DR, the issue on such measures may only be a theoretical exercise in mathematics, or an application as in the USA on *States*. Subsequently, they present a zoo of possible measures. With such confusion, it is relatively easy for decision makers and the public in DR countries to suffer an information overload on EPR, and then remain with DR. If EPR is so difficult to understand, why change ?

Balinski & Young presented an "impossibility theorem" though assumed The Party. Though this only concerns a mathematical result, the effect of choosing such framing is that it creates the suggestion as if a sound system would be impossible, or that it would be impossible to *define* what *equal (proportional) representative democracy* would be. I am reminded of the story of Ignaz Semmelweis (1818-1865). [3] Current Political Scientists with their lack of criticism of DR are similar to doctors who attend to childbirth without washing their hands. Their defence may be that they are no real doctors treating patients but only academic researchers. They still suffer from the paradigm of The Party and thus create all kinds of confusions that prevent the world to adopt EPR like in Holland in 1917. Obviously one would

[3] https://en.wikipedia.org/wiki/Ignaz_Semmelweis

not indiscriminately copy all that Holland managed to do in 1917, but the present intellectual outrage is the created confusion about the proper formulation of the research issue. It may also be a case of mathematical abstraction running astray, and onlookers not able to perceive what the mathematicians actually are doing but still following their lead.

Our present purpose is to (i) deconstruct the discussion in Political Science about inequality / disproportionality measures, (ii) show what the problem really is (actually already done in the above), (iii) present a clear and simple inequality / disproportionality measure (though with options to make it more complex again). **Appendix L** contains a *representative based* apportionment method, and we continue with the general measurement issue.

1.2. The literature and what we will do

The literature on voting has a rich discussion about potential measures for the inequality / disproportionality of the allocations of s given v. Taagepera & Grofman (2003:673) rightly write about a "zoo of indices proposed and used by various researchers", and they score 19 indices on 12 criteria. [4] Overview discussions are by e.g. Taagepera & Shugart (1989), Gallagher (1991), Taagepera & Grofman (2003), Kestelman (2005), Karpov (2008) and Renwick (2015). Malkevitch (2002) gives a discussion in an AMS feature column. Perhaps for historical reasons this literature has focused on the inequality / disproportionality format, in which 0 indicates *equality* / *proportionality* and in which 1 might be *full disproportionality* - if the measure is normalised on [0, 1]. We also adopt this latter format, though allow for a sign in [-1, 1]. **Section 8.1** gives an nicely designed example overview adopted from Balinski & Young of how inequality / disproportionality measures can score the same vote and different apportionments of seats.

To give a taste of the problem and to test our findings, the next section presents graphs on the outcomes of the (half) elections in 2017 for the House of Commons in the UK, the Legislative in France, and the 2nd Chamber of Parliament in Holland. These bodies compare to the US House of Representatives (rather than the Senate or House of Lords).

The subsequent section presents the newly suggested measure based upon cosine and sine so that it is clear what it amounts to. Subsequently we clarify its motivation and reasoning. Subsequently there are some derivative comments, like on other measures that statistics provides. We close with some theoretical examples.

It obviously matters what we understand by "inequality / disproportionality" and how we interprete the tables and graphs. It is important to highlight some principles here too.

- Partly summarising the above: What do we take as the unit of account: the party or the representative, i.e. the party marginal candidate ? For parties we tend to take the *party average* of the *votes per seat* as a measure. When we take the representative as the unit of account then we would look at the *votes per seat* for the *party marginal representative*. It may be that a small independent candidate below the natural quota threshold still has more *votes per seat* than any party marginal candidate. The voting literature tends to regard the marginal fractions as absolute differences, but in fact these are ratios, with the marginal representative in the denominator. See **Section 6.3**. This confusion has relevance for the discussion about level or relative measures.
- Proportionality is desirable for its *power preserving property*, under majority rule. The base are the possible coalitions amongst the voters. By reflecting their votes into seats, those possible coalitions are also possible in the seats. Thus, even when we take the representative rather than the party as the unit of account, there still is a role for coalitions and thus parties. See **Section** 6.2.

[4] (i) The Taagepera & Grofman (2003) definition of the Gini differs from Colignatus (2017b) which fundamentally affects the major conclusion. (ii) One of their criteria is that there is no explicit need for the number of parties (like in the Chi Square index for the degrees of freedom): this is curious since the measures use s and v that have the length of the number of parties. (iii) Goldenberg & Fisher (2017) criticise the finding on Dalton's principle of transfer, see the discussion below.

- As said, the number of seats is important, see **Appendix M**, though this paper tends to focus on the literature that neglects this.
- A low turnout might generate some proportionality, as voters turn out who have larger chance to win, like may happen more structurally in systems with District Representation (DR). Thus turnout would be a factor in inequality / disproportionality as well. It however is not considered in the comparison of *votes* to *seats.* This is put into **Appendix J**, though it is not off-topic.

1.3. Much Ado About Nothing

There are age-old measures for distances between vectors, namely the *absolute deviation* Sum[Abs[*w* – z]] (a.k.a. the Manhattan distance), and the *Euclidean distance* ||*w* – *z*||, in which we use unitised variables to neutralise the different bases *V* and *S.* For voting and seats, Abs is a relevant measure since seats are integers, so that the difference gives the number of dislocated seats. If $\hat{s}$ = *Ap*[*S, v*] is a method of apportionment to translate votes *v* into optimal seats for a House with *S* seats, then one could use ID[*v, s* ; *S*] = Sum[Abs[*Ap*[*S, v*] – *s*]] = Sum[Abs[$\hat{s}$ – *s*]] with the semi-colon giving the condition that the optimum depends upon *S*. A parliament with more seats might have a better chance at finding a proper fit.

Linear algebra developed, early, an easy expression for the angle θ between two vectors. Thus we also have Cos[θ] = Cos[*v, s*] = Cos[*w, z*]. When θ = 0 then Cos[θ] = 1 and the vectors overlap, and the unitised vectors are exactly the same. Thus Cos is a measure of similarity and both θ and Sin = Sqrt[1 – Cos^2] are a measure of distance. The angle is obviously a metric and Van Dongen & Enright (2012) show that Sin is a metric too. While we can add lengths and angles, we cannot add Sin though, similarly to the logarithm scale of Richter for earthquakes. Here we look into Sin[θ] rather than θ, see **Appendix O**.

For common applications in vector space, Abs and the Euclidean distance suffice, and the literature and textbooks hardly refer to Sin as a measure of distance. Sin is rather superfluous. For votes and seats in the unit simplex this appears to be different.

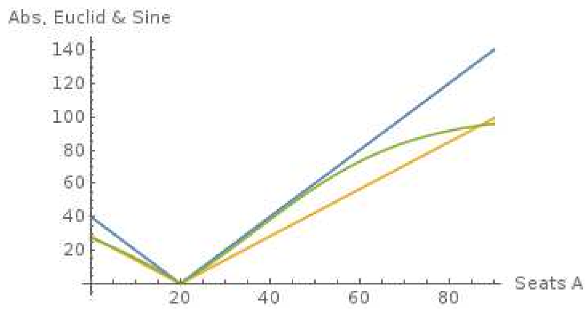
Figure 1 shows an example with votes = {20, 70, 10} for parties *A, B* and *C,* and various seats *sa* for party *A*, assuming that party *C* already got its 10 seats. Thus seats = {*sa,* 90 – *sa*, 10}. We multiply the Sin values by 100 to get up to scale. While we have 3 parties, we fixed 1 party, and for 2 parties the Euclidean measure reduces to a form of Abs. [5] For short distances, Sin follows Abs and for far distances Sin moves to Euclid.

A major conclusion from this figure is that this present paper is *Much Ado About Nothing*. There are already suitable distance measures and the attention for Cos and Sin has limited value. For voting, the Absolute Deviation also has proper interpretation: the relocation of seats from the local optimum. Thus, one wonders why more need be said on this (and even some 100 pages).

Quite aware that this discussion is a bit like discerning the number of angels on top of a pin, the contribution of this paper is:

[5] Below Abs and Euclid will be normalised with factor 2 for double counting relocated seats. Doing so for this particular graph distracts from the intended line of reasoning. For the binary case, the two lines namely overlap when corrected with 2. See the subsequent discussion.

Figure 1. Distance measures Abs (blue), Euclid (yellow) and Sine (green) for votes = {20, 70, 10} and seats = {sa, 90 – sa, 10} [6]



(1) The introduction of Weber – Fechner sensitivity. [7] Geometry is less sensitive to differences in votes and seats than humans are. Thus a transform like the Richter scale can enhance awareness. Rather than the Log scale my suggestion is to use the Square Root, which works out the same, and that is easier to grasp. Perhaps it might be better though to train students to be more aware of small differences.

(2) The paper shows the usefulness of cosine and sine. The Absolute Deviation is insensitive to how seats are relocated. This is very relevant for some purposes, but not for all. If votes are {25, 25, 25, 25} then Abs sees no difference between seats {30, 30, 20, 20} and {35, 25, 20, 20}, while Sin scores them as 19.6 and 23.8 respectively on a scale of 100. Using sensitivity mentioned under (1) has no impact for Abs, since the numbers of relocated seats are the same. Euclid does better here. However, **Figure 1** clarifies that Euclid downplays quite some differences. Euclid is convex to the origin rather than concave. Students (and perhaps also regressions) must be trained to be very sensitive the scores. It appears that (concave) Sin (i) has greater awareness of smaller differences, (ii) can see a difference in this example too. Even at a late stage in writing this paper, I was inclined to advise the use of the absolute deviation, because of its clear interpretation of *relocation of seats*, but given this insensivity it was clear that cosine and sine were an improvement, for this kind of application. In this figure Abs is larger than Sin, but this disappears when we correct for double counting, and for 2D Abs = Euclid. In this figure, there is asymmetry, with Sin following Euclid on the left and Sin following Abs on the right, and this disappears when we correct Abs for double counting.

(3) In standard regression with a constant or data centered around their means, there is the Pearson coefficient of Determination R^2. As well-known, in regression through the origin (RTO) this becomes Cos^2. For comparing votes and seats it is quite natural to have RTO, and thus also Sin.

(4) The paper shows the relation to the slope of $z = b\, w + e.$ Cos is a measure of similarity but can also be interpreted as a geometric average of *slope.* This invites us to consider the deviation in the {*w, z*} scatter plot between the regression line and the diagonal, which represents equality / proportionality $w = z$ (without dispersion). The paper started out with this intuition, but the first three points are more important.

(5) The paper also clarifies how notions of slope are more complicated for the unit simplex.

(6) The paper gives a perspective on the distinction between the mentioned measures that use level data, and the alternative Webster / Sainte-Laguë (WSL) measure that uses relative data. **Appendix O** links up to the Aitchison distance for compositional data.

(7) This paper also clarifies some issues on content that are relevant for measuring inequality / disproportionality of votes and seats. It highlights the question: *what is it, actually, that we try to measure ?* A metric is symmetric by definition, but our topic is fundamentally asymmetric. The votes determine what seats are optimal or not. A focus on {*w, z*} runs the risk that we forget that we actually should have ID[*v, s* ; *S*] = *func*[*Ap*[*S, v*], *s*].

[6] Wolfram Alpha: Plot @@ {v = {20, 70, 10}; s = {sa, 90 - sa, 10}; abs = Plus @@ Abs[(v - s)]; eucl = Sqrt[(v - s). (v - s)]; cos = 1 - CosineDistance[v, s]; sin = 100 Sqrt[1 - cos^2]; {abs, eucl, sin}, {sa, 0, 90}, AxesLabel -> {"Seats A", "Abs, Euclid & Sine"}, PlotRange -> All}

[7] Wikipedia is a portal and no source: https://en.wikipedia.org/wiki/Weber%E2%80%93Fechner_law

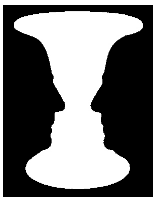
1.4. Statistics and heuristics on the slope

This paper uses descriptive statistics, and we are not in hypothesis testing. The linear algebra in this paper should not be confused with statistical decision methods. The latter methods use the same linear algebra but also involve assumptions on distributions: and we will make no such assumptions. However, when the linear algebra results into a new measure, then this measure can be used for new descriptive statistics again.

A heuristic is this: the voting literature has the Webster / Sainte-Laguë (WSL) measure Sum[w $(z / w - 1)^2$], as the weighed sum of the squared deviations of the ratios from the ideal of unity provided by the diagonal. WSL is not symmetric. If we would try to make WSL symmetric then this causes division by zero, since some parties meet with zero seats (the "wasted vote"). An insight is: the idea to compare each ratio to unity is overdone, because the criterion of equality or proportionality applies to the whole situation, and we should not be distracted by single cases. If b is the slope of the regression of $z = b\, w + e$, then $(1 - b)^2$ is a measure on the regression coefficient. This shifts the focus from individual parties to the *relation* between z and w. Thus we now consider the *slope-diagonal deviation*. This gives scope for symmetry by also looking at the regression $w = p\, z + \varepsilon$, so that we can compare b and $1 / p$.

The latter was the heuristic that started this paper. Taking the geometric average Sqrt[b p] gave the recognition that this gave the same mathematical expression of the cosine as well. In other words Cos[v, s] = Sqrt[b p]. At some point it appeared that the role of the cosine was more important by itself, and thus not regarded as a slope, as it generates the inequality / disproportionality measure Sin, that again was seen as a *slope-diagonal deviation* measure. This double nature of cosine and sine may be illustrated by Rubin's Vase, see **Figure 2**. However, the paper still pays much attention to the role of the slope.

Figure 2. Rubin's Vase [8]



1.5. Electoral justice and inequality

Balinksi & Young (1976:2) quote Daniel Webster:

> *"To apportion is to distribute by right measure, to set off in just parts, to assign in due and proper proportion."*

His emphasis was on "due and proper" and not on "proportion", but the world adopted "proportionality". Arend Lijphart has written about proportionality as "electoral justice". I have considered adopting the term "justice" too but settle for the notion that $z = w$ means equal proportions, or *equality*. A line $y = ½\, x$ in 2D is proportional too. Thus, rather than speaking about "disproportionality" in voting research, it is better to speak about either *unit or diagonal disproportionality* or *electoral inequality.* Similarly, while the standard expression is "Proportional representation" (PR) it would be better to speak about *Equal Representation* (ER). I don't think that there is much chance that the world will rephrase this for the sake of mathematics education, but at least it is important to explain it as a possible source for confusion. [9] Currently I settle for Equal / Proportional Representation (EPR).

[8] https://en.wikipedia.org/wiki/Rubin_vase
https://commons.wikimedia.org/wiki/File:Rubin2.jpg

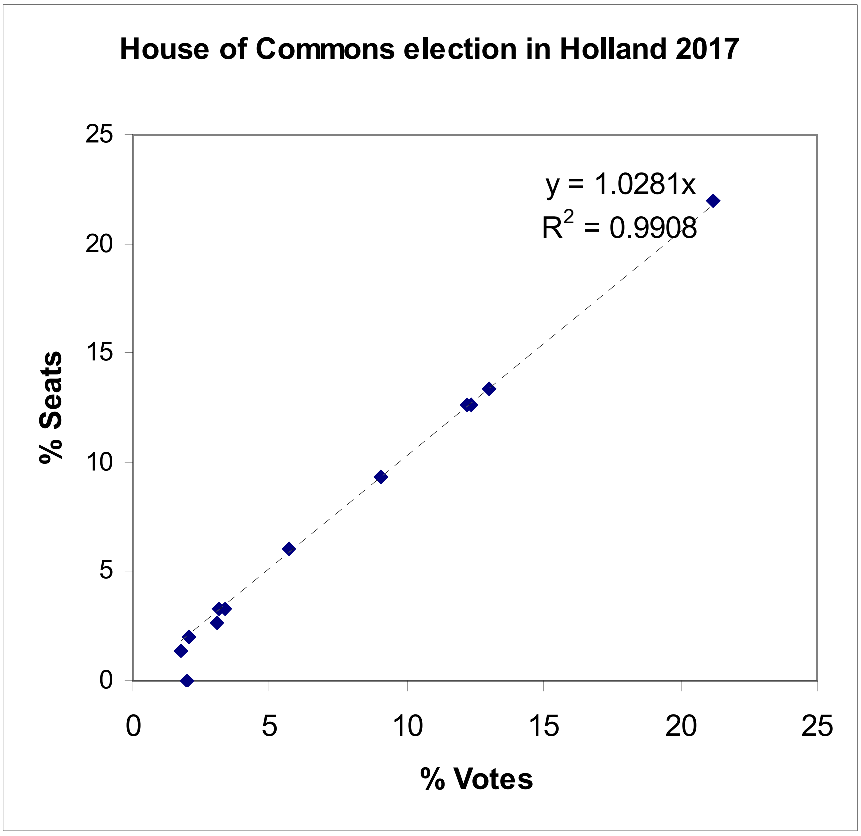
[9] France has the slogan *"Liberté, Egalité, Fraternité"*, and thus it seems important to explain to the French that their system isn't *equal.* Using the word *disproportional* wouldn't ring their bell.

2. *Graphs of (half) election results in 2017 in Holland, France and UK*

For the elections (EPR) in Holland, and half elections (DR) in France and UK in 2017 there is also Colignatus (2017b) for the Lorenz curve and Gini coefficient on electoral inequality. These latter measures are not symmetric, whence this present paper.

On March 15 2017, Dutch voters elected the parties for the 2nd Chamber of Parliament with 150 seats. **Figure 3** gives the regression of $z = s / S$ given $w = v / V$.

Figure 3



Holland has Equal or Proportional Representation (EPR). It uses Jefferson-D'Hondt which favours larger parties. There is the natural threshold that parties must meet the quota of 1/150 = 0.67% to gain a seat. This low threshold invites competition to the establishment. Yet 2% of the vote or 3 seats was "wasted" on small parties that got votes but that did not meet the quota. Holland could achieve better EPR if it left 2% or 3 of the seats empty. An alternative is to adopt a qualified majority rule. Now 76 seats represent a majority of Floor[S / 2] + 1. Turnout V consists of the vote for parties that got elected Ve and the wasted vote W, including invalid ballots. Thus $V = Ve + W$. Majority threshold $V / 2$ should correspond to $S / 2$. A qualified majority qm on Ve gives the equation $qm\ Ve = V / 2$. This gives $qm = ½ / (1 – W / V)$. Thus for a better EPR score in Holland , the 2nd chamber of Dutch parliament should not use a majority of 76 but of 77, at least assuming that the wasted vote might go either way and are not fundamentally opposed to anything that happens when they are not present themselves.

On June 11 and 18, French voters half elected (DR) the Legislative with 577 seats. **Figure 4** gives the regression of z on w. Observe that these regressions do not use a constant, and that the display is around the diagonal (with same scales of horizontal and vertical axes). France has District Representation (DR). The (half) election in districts is in two rounds. We may assume that the first round gives the first preferences of the voters and that the second round involves strategic voting for elimination of the worst option. The horizontal axis uses the votes of the first round and the vertical axis uses seats of the second round.

Figure 4

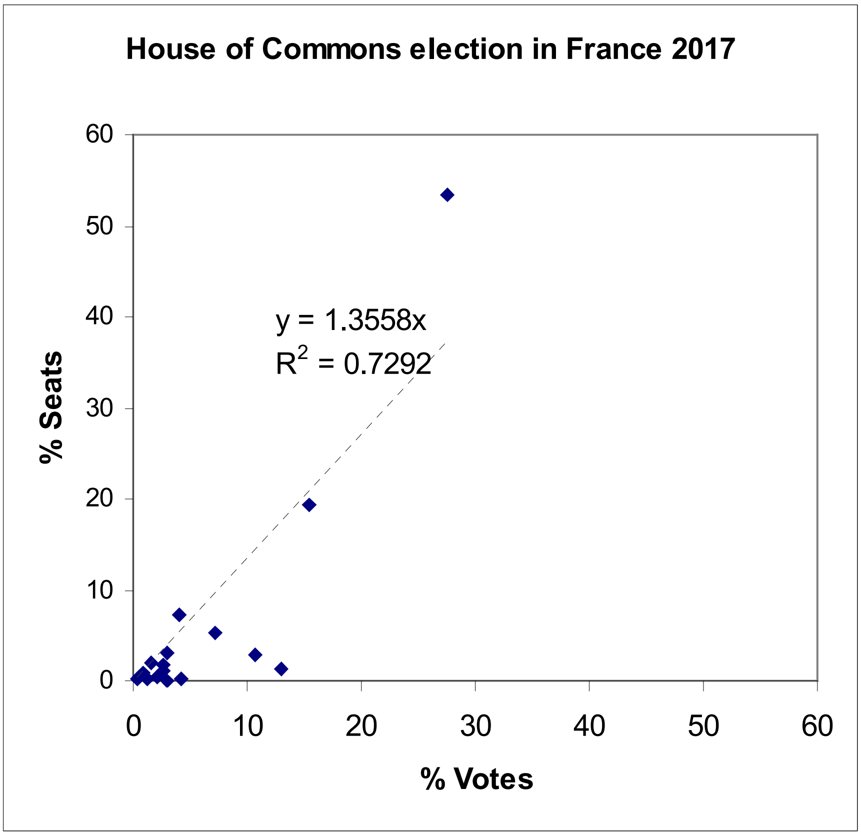
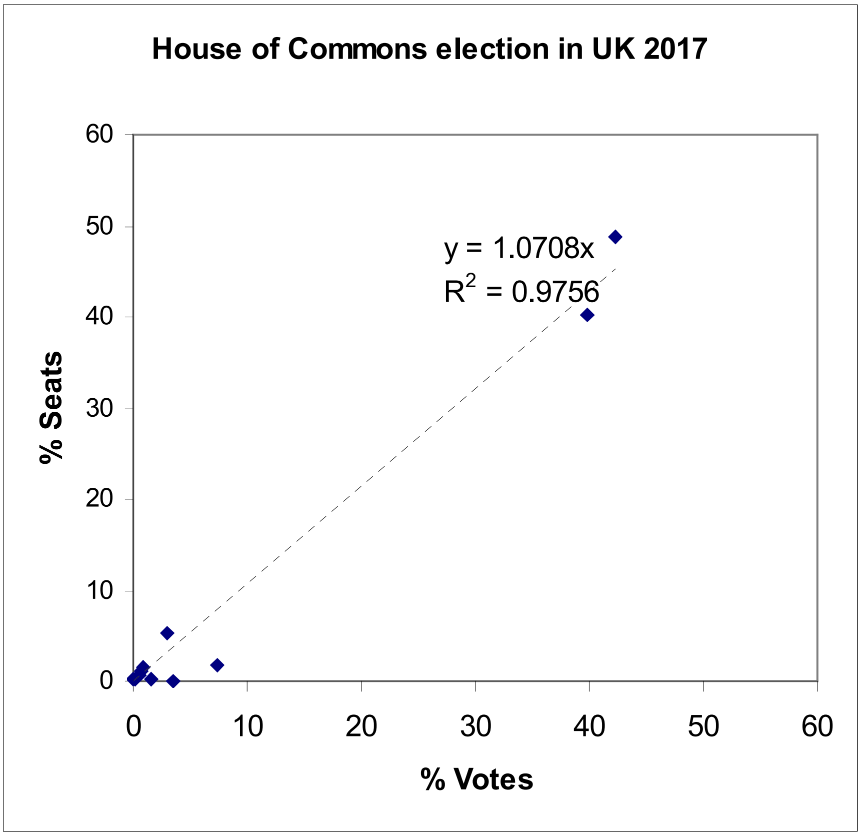


Figure 5



On June 8 2017, voters in the UK (half) elected (DR) the House of Commons with 650 seats. **Figure 5** gives the regression of *z* on *w* The UK has District Representation (DR), in one round, so that voters may adopt strategic voting to keep out the worst alternative that stands a change of winning. The 2017 UK (half) election seems rather proportional since the voters apparently returned to the model of voting either Conservatives or Labour. However, the outcome can be called "masked" because we do not know the first preferences.

3. Different purposes in voting literature and statistics in general

3.1. Descriptive statistics and decisive apportionment

Vectors *s* and $z = s / s'1$ have been created by human design upon *v*, and not by some natural process as in common statistics. A statistical test on *s* / v would require to assume that seats have been allocated with some probability, and this doesn't seem to be so fruitful. The current (inoptimal) inequality / disproportionality measures (WSL, Absolute distance / Loosemore-Hanby (ALHID), Euclid / Gallagher (EGID)) have frequently been used to compare outcomes of electoral systems across countries, but such comparisons have limited value because the designs are different anyway. Taagepera & Grofman (2003) mention also some other reasons for an inequality / disproportionality measure: (i) what they call "vote splitting" (better: votes for different purposes): comparison on President, Senate, House, or regional (half) elections, (ii) what they call "volatility" (better: votes on different occasions): comparison on years in similar settings (both votes and seats).

There appears to be some distance between the voting literature on (unit) inequality / disproportionality and the statistics literature on association, correlation and concordance. A main point is that voting focuses on *unit* equality / proportionality on the unit simplex with $z = b\, w + e$, while statistics has hypotheses tests on general relationships like $s = c + B\, v + u$ and then require stochastics. Some influences are the following:

- The *apportionment of seats* (italics) based upon votes involves some political philosophies that have been adopted by the national parliaments. Researchers on voting may have a tendency to remain with these philosophies when they *measure (unit) inequalities / disproportionalities* (italics) from such apportionments too. Apportionment (deciding) and measuring (describing) have different purposes though.
- There need not be a real distance between the voting literature and statistics, at roots, because (i) the Webster / Sainte-Laguë (WSL) apportionment philosophy compares with the Chi square, and (ii) the EGID index relates to the philosophy of minimising the sum of squared errors (Euclid), which philosophy is applied in the apportionment according to Hamilton / Largest Remainder (HLR). Perhaps this early historical linkage also caused the presumption that voting theory already "had enough" of what was available or relevant in the theory of statistics.
- The wasted vote concern votes for parties who receive no seats. The wasted vote clearly contribute to some (unit) inequality / disproportionality. The Kullback-Leibler (entropy) measure that compares vectors *w* and *z* still works when the *w* are the weights, but when we consider the condition of symmetry then it collapses on the zero in *z.*
- When equal vote-shares {50, 50} translate in equal seat-shares {50, 50} then this would be a unit equal / proportional allocation, yet the conventional Pearson correlation coefficient collapses because of lack of dispersion in the vectors. (Thus **Section** 4 uses the cosine.)
- When vote shares {49, 51} translate into seat shares {51, 49} then this forms a major shift in majority. (Thus **Section** 4 uses the sign of the covariance if negative.)
- There is also a zoo in statistics for measures for all kinds of different purposes on association, correlation and concordance. For voting researchers it might be difficult to find a match on the same kind of purpose.

The voting literature traditionally mentions the Lorenz and Gini too, but apparently in a different format than in Colignatus (2017b) that intends an improvement, see the abstract in **Appendix A** below. It may be observed that Lorenz and Gini are adequate for comparing equal / proportional representation (EPR) and district representation (DR). See **Section** 4.6

on the comparison of Holland, France and the UK. The present discussion on other measures then is not really required for comparing EPR and DR. PM. The Lorenz and Gini in the definition of Colignatus (2017b) require the ordering of the parties on the z / w values. This ordering has drawbacks like: a source of errors, and the need to keep lists with parties in the same order e.g. for comparisons over years. It can be useful to avoid sorting.

When I started looking at this issue, I wondered how the zoo in voting theory linked up to the zoo in statistics. This present discussion indeed improves clarity. It might be seen as a step forward that there is now a better distinction between focusing on the deviation of the slope from the diagonal versus focusing on other (more direct) criteria on w and z themselves. The discussion highlights where voting differs from other purposes in statistics: (i) First the requirement on the diagonal of the scatter plot (with thus a confusing use of the word "proportionality" that means unit vector-proportionality or diagonal vector-proportionality). (ii) Secondly, voting research cannot rely on stochastic assumptions for testing, and thus wants to measure. We use the same linear algebra but for a different purpose. The discussion helps to see that the choice of an inequality / disproportionality measure apparently is not self-evident, at least with the current literature and textbooks so dispersed over various topics. Statistics textbooks advise against RTO, since it is better to test upon modeling errors by means of constant c, yet the situation is quite different for descriptive statistics.

The following repeats explanations that are abundantly available in the literature, yet we should achieve more clarity, while a key point is the emphasis on the wasted vote. I will refer to Wikipedia – a portal and no source – for ease of access to the concepts used.

Before we look at measures of Kullback-Leibler and the Chi-square, it appears to be more fruitful to first discuss the unit simplex before looking at regression and the philosophy underlying the suggested measure.

3.2. Different models and errors

$Q = V / S$ is the natural quota, or number of votes to cover a seat. There may be a threshold to get a seat, or just the natural quota. Voters may vote for parties that do not pass the threshold and that thus get no seats. This sums to the "wasted vote" W. Standardly the wasted vote and zero seats are collected in one category "Other", so that v and s still have the same length. The votes that cause a seat are $Ve = V – W$. For regression it is conventional to write $s = T v$, so that s is explained by v. We call this vector-proportionality because of the lack of a constant. Any such relation also holds for its sum totals, and we can usefully define $T = S / V = 1 / Q$. In reality we have $s = B v + u$ or $z = b w + e$ with proportionality parameter b and scatter e. There is only unit proportionality if $z = w$ or $b = 1$ and $e = 0$. With $a = S v / V = T v = v / Q = S w$ the proportionally accurate average seats that a party might claim, the error will generally be taken as $s – a$. This translates into $S (z – w)$, which shows that the unit simplex is the natural environment to look at this, though we should not forget about the role of S.

The major distinctions are in **Table 2.** Some readers might prefer to look in **Table 9** for a continuation of formulas.

Table 2. Basic models and their errors

$T = S / V = 1 / Q, \quad B = b T$	*Original*	*Unitised*
Without parameter [10])	$s = T v + \tilde{u} = S w + \tilde{u} = a + \tilde{u}$	$z = w + \tilde{e}$ with $\tilde{e} = \tilde{u} / S$
Regression through origin	$s = B v + u$	$z = b w + e$
		$w = p z + \varepsilon$
With constant (centered)	$s = c + B v + \hat{u}$	$z = \gamma + \beta w + \hat{e}$

[10] This is not unreasonable for official data, when $\hat{s} = Ap[S, v]$ is already seen as optimal. Regression then would misstate the error, though b would still be a measure. Also, the cosine measure can also be seen as just a measure of similarity, without such estimation of b involved. This is the double face of the cosine.

We will use regresssion through the origin (RTO) and not regression with a constant (Pearson), but the latter can be mentioned for contrast. The value $a = S\ w$ will be the average, and apportionment of s will tend to be for Floor[a] ≤ s ≤ Ceiling[a]. There are three different error measures:

- $\hat{e}$ from the standard regression with a constant, using centered data (Pearson).
- e from regression through the origin (RTO), with coefficient $b = z'w / w'w$. **SDID uses *e*.**
- $\tilde{e} = z - w$ or "direct error" for the plain difference, using $b = 1$. **ALHID and EGID use $\tilde{e}$.**

Some useful mnemonics directly are:

- $\hat{e}'\hat{e} \leq e'e \leq \tilde{e}'\tilde{e}$ because regression parameters allow the reduction of error.
- SDID picks up a potential source for proportionality that ALHID and EGID do not allow for.
- $1'\ \tilde{e} = 0$ and $b = 1 - 1'e$ because $1'z = 1'w = 1$.
- $b = z'w / w'w$ because we multiply $z = b\ w + e$ with w' while $w'e = 0$. The regression selects the b that minimises $e'e$ and then causes $w'e = 0$.
- Taking the plain differences $\tilde{e} = z - w$ and weighing them by the vote shares and normalising on their squares, gives $w'\ \tilde{e} / w'w = b - 1 = -1'e$. This might perhaps be seen as an "implicit estimate".

3.3. A role for cosine and sine

The suggestion is to use Cos[w, z] = Sqrt[$b\ p$] and Sin = Sqrt[1 – Cos^2] = Sqrt[1 – $b\ p$], where the slopes are found by regression through the origin (RTO). For sensitivity it appears better to take the square root again and multiply by 10, giving the *sine diagonal inequality / disproportionality* (SDID) measure *sign* 10 √ Sin.

- For the unit simplex the proper approach is regression through the origin (RTO). E.g. regression of z = {40, 60} on w = {60, 40} doesn't give a slope of -1 as we would expect from regression with a constant, but a result 12/13.
- Cos^2 is the coefficient of determination R^2 of RTO.
- Koppel & Diskin (2009) have the proposal to use 1 – Cos, a.k.a. the Cosine Distance. This however is not a metric.
- We improve on this by (i) metric Sin, (ii) relation to the slopes, (iii) sensitivity, (iv) sign.

Alongside WSL other common measurea are the Absolute distance / Loosemore-Hanby (ALHID) and Euclid / Gallagher Inequality / Disproportionality (EGID). WSL, ALHID and EGID are parameter-free, and clearly fall under descriptive statistics. My original view was that SDID is parameter-dependent, since it refers to b and p. Paradoxially, however, while we started out looking for a slope measure, Cos is actually parameter-free, since its value as a similarity measure can be found without thinking about slopes at all. It is just a matter of perspective, see **Figure 2**. Thus the bonus of above heuristic is only that it helps to better understand what the measure does. The awareness of this double nature is important for the comparison of ALHID and SDID. For ALHID one might argue that it takes error $z - w = \tilde{e}$ in **Table 2** under the assumption that official regulations have chosen $s = \hat{s} = Ap[S, v]$ with some optimality. Thus, the use of $z = b\ w + e$ can be seen as changing the error. However, when Cos is regarded as a parameter-free measure of similarity, then no errors have been changed. Thus we can compare ALHID and SDID with this argument out of the way. Interpreting Cos as the outcome of RTO is only a bonus, that helps in the development of more perspectives.

Cos and Sin are independent of scale. The use of unitised variables may come with a risk. Dividing s by S and v by V creates the *hidden* base parameters S and V. Since S is constant for a particular parliament, we may also speak about hidden parameters Q or $T = 1 / Q$. I tend to prefer to use T because it reminds of *target* and *tangent*. Though T is a quite small number. For Holland the quota Q is around 70,000 voters.

- One of the objectives is to compare results over various years. A comparison of {$z1$, $w1$} to {$z2$, $w2$} might overlook the effect of turnout, or the change of $Q1$ to $Q2$.

$$(z1 / w1) / (z2 / w2) = (T1 / T2) (s1 / v1) / (s2 / v2)$$

- Above comment on turnout and disproportionality finds a proper place here too.
- Comparison over countries may also overlook the effect that somewhat larger parliaments may have more chance of meeting with equality / proportionality. Thus it might be advisable to have a measure that includes a correction for the base *S*, perhaps w.r.t. some standard. This however is beyond the scope of this discussion.
- Elimination of the base is precisely the purpose of the present exercise. We are not looking at all causes of inequality / disproportionality but only for the relation of *votes* and *seats.* Nevertheless, it remains important to identify the wider scope.

The combination of the *unit simplex* and *regression through the origin* and the use of *nonnegative variables* has particular consequences, that may need some explanation. Comments on this appeared at risk to scatter over the present discussion, and thus I have collected comments on *T* in **Appendix J**. However, the unit simplex still deserves the main **Section** 5. See **Appendix O** for more on compositional data and the Aitchison distance.

3.4. Disproportionality, dispersion and education

An important consideration from education is the following. In my other capacity as teacher of mathematics I want to maintain that we want to explain to students that any line through the origin in 2D represents a proportional relationship. For vectors and their scatter plot we would speak about "vector-proportionality". Thus, also for z and w, a relation $z = a\, w$ is a vector-proportional relationship, and $z = w$ is only *unit* or *diagonal* vector-proportional. It so happens that the unit simplex is defined such that $1'z = a\, 1'w$, thus any pure proportional relation in that space requires $a = 1$ of necessity. I would phrase this as that the space is defined such that those other vector-proportions *do not exist.* I would *not* phrase it as saying that (quote) "$z = a\, w$ for $a \neq 1$ would not be vector-proportional and that only $a = 1$ is vector-proportional" (unquote). It is better to say: we only have $z = b\, w + e$ with vector-proportionality parameter b and scattered e. Thus, overall, it would be didactically preferable to speak about "unit or diagonal proportionality" for voting rather than "proportionality". *Equal Representation* is better anyway. It may be difficult to change a habit, but it would help for mathematics education.

This calls attention to the relation with dispersion. **Table 3** reviews the relations. The key relationship in RTO is that $b = 1 - 1'e$. Thus $b = 1 \leftrightarrow 1'e = 0$. The upper right cell is impossible: Not[$e = 0$ & $1'e \neq 0$]. The lower right case of disproportionality implies dispersion, but dispersion ($e \neq 0$) need not imply disproportionality. The left column with $b = 1$ is in opposition to the right column, but we must distinguish between *unit* or *diagonal* vector-proportionality without dispersion (equality) and *average unit* vector-proportionality with dispersion.

Table 3. Disproportionality and dispersion in Regression Trough the Origin (RTO)

RTO: $b = 1 - 1'e$	$1'e = 0$ & $b = 1$	$1'e \neq 0$ & $b \neq 1$
No dispersion $e = 0$	• *Unit or diagonal* vector proportionality • Equality	• Logically impossible
Dispersion $e \neq 0$	• *Average unit* vector proportionality, can come with dispersion. E.g. $w = \{3, 3, 2\}/8$ and $z = \{2, 1, 1\} / 4$.	• Disproportionality by slope and disperson

With $b = 1$ then $e = z - b\, w = z - w = \tilde{e}$. Taking the plain differences $\tilde{e} = z - w$ and weighing them by the vote shares and normalising on their squares, gives $w' \tilde{e} / w'w = b - 1 = -1'e$. Thus above distinction might also be phrased in terms of an expression on $\tilde{e} = z - w$. Thus $w' (z - w) = 0$ or not, of which $z - w = 0$ is only a special case.

In the lower right hand cell, it may be interesting to decompose disproportionality into its subcomponents of *slope specific disproportionality* 1 – *b* and the *nonspecific disproportionality* given by the dispersion in *e'e.*

3.5. A bit more formal

It is useful to restate the general context so that we are better aware of the (im-) possibilities.

(1) In general, a country has an apportionment method with optimum $\hat{s}$ = *Ap*[*S, v*].
(2) To what extent can *Ap*[*S, v*] rely on some measure *D* as a measure of optimality ?
(3) For measure *D* we require symmetry, because we want a metric. But apportionment may have entirely different considerations. Apportionment is asymmetric by definition.
(4) If the problem is only rounding around *a* = *S v* / *V* = *T v* = *S w* then the issue might be seen as marginal. It is an entirely different issue when seats are allocated via a method of District Representation (DR), for example.
(5) A difference is that *Ap* may have an algorithmic nature in the Floor – Ceiling locus, while *D* may have a simple formula for vectors for general application.
(6) For proportional apportionment, we tend to expect Floor[*T v*] ≤ $\hat{s}$ ≤ Ceiling[*T v*].
(7) When there is rounding, we cannot really eliminate *S* in *Ap*[*S, v*]. The search for a symmetric *D*[*v, s*] = *D*[*s, v*] to replace *Ap* is vain, since in general *Ap*[*S, v*] ≠ *Ap*[*V, s*].
(8) Still, *Ap*[*S, v*] can be seen as relying *in some manner* on some *D*1 for its locus of application. Any other *s* than $\hat{s}$ will be inoptimal according to *Ap*. Let method *Ds*[*s* ; $\hat{s}$] express a degree of inoptimality, in which everything is measured in distance from $\hat{s}$. *Ds* might be a symmetric measure to start with, but this centering around $\hat{s}$ gives *Ds*[*s* ; $\hat{s}$]. Must we take *Ds* = *D*1*,* or might we also take some *D*2, since we generally look outside of the locus of apportionment ?
(9) Given that both *Ap* and *Ds* take the $\hat{s}$ defined by *Ap* as optimal, we might indeed use a wide scope of measures *Ds* rather than the more complicated *Ap* or the *D1* that it might depend upon. Other values are judged as inoptimal, though perhaps not in the same degree. The range of applications could still be limited, but it might suit our purposes.
(10)The error *Ds*[*v* ; $\hat{s}$] = *Ds*[*Ap*[*S, v*], *v*] would be seen as the unavoidable error when *Ds* cannot catch the subtleties of *Ap.* We would expect this to be minimal.
(11)There is the fundamental asymmetry that *v* determines the optimal allocation while the error of any other *s* is taken with repect to this optimum. Thus *Dv*[*v, s*] = *Ds*[*s* ; *Ap*[*S, v*]] is asymmetric. Why look for symmetric measures ? Perhaps only because we only have general data on *v* and *s. Descriptive statistics has other issues than decisive Ap.*
(12)An application as if there were symmetry generates *Dv*[*s, v*] = *Ds*[*v* ; *Ap*[*S, s*]] = *Ds*[*v* ; *s*] since we may assume some consistency that *Ap*[*S, s*] = *s.* However, *Ds*[*v* ; *s*] means that *s* is regarded as optimal and that *v* is judged against this optimum. This is curious, since the intention is to judge *s* against *v,* and not the other way around. Thus *Dv*[*v, s*] as implied by *Ap* and *Ds* is asymmetric in nature. However, if *Ds* originally was symmetric, we might not notice.
(13)A trick is: While we should look at *Ds*[*s* ; $\hat{s}$], it often might be overly accurate to really determine the optimum, and we might approximate it with *Ds*[*s* ; *v*]. If *Ds* originally was a symmetric measure *D* then we are back using *D*[*v, s*] = *D*[*s, v*] again. This requires that we are both aware of the approximation involved, and aware of which of the two is considered optimal.
(14)A major trick is to assume that official {*v, s*} outcomes have been created with some optimal *Ap*[*S, v*], so that researchers may assume this for *s*. In that case *w* – *z* indeed would have an optimal value. The complication arises if *Ds*[*v, s*] allows a lower value for another *s,* so that one might doubt the optimality of *Ap*[*S, v*].
(15)A major method to recover original data is to use *r* = *v* – Floor[*s* Q], which will be negative for parties that received more than their average *a.* To some extent one can thus say something about an official apportionment without knowing its details.
(16)In sum: There is some ground to consider that *Ap* and some *Ds* and the implied *Dv* can satisfy some common notion of vector-proportionality *D*, potentially satisfying symmetry. We then might take a symmetric *D* such that *Ds* is an asymmetric application of *D,* in which the center is taken as $\hat{s}$*.* But we would know that it can't replace *Ap.*

4. Definition of SDID and graphical comparison with ALHID, EGID, WSL

4.1. Definition of the SDID measure

1D variables x and y are considered to be proportional when $y = b\,x$, thus as a ray through the origin without a constant. Seat allocation is a zero-sum game, and thus we are interested in the line $y = x$ and thus with slope $b = 1$. We translate this to vectors. The suggested "sine diagonal inequality / disproportionality" (SDID) measure, see **Table 4**, uses the unit length vectors w and z to define a disproportionality measure around the slope b, in this case of $z = b\,w + e$, also considering the inverted regression $w = p\,z + \varepsilon$. Some readers may directly understand what the measure does. For others the table is an introduction to be alert to the terms that will be explained in the subsequent sections.

Table 4. Definition of the "sine diagonal inequality / disproportionality" measure

$w = v / V$ and $z = s / S$	normalisation to 1, or the unit simplex (frequently 100)
$b = z'\,w / w'\,w$	slope of regression through the origin of z given w
$p = z'\,w / z'\,z$	slope of regression through the origin of w given z
k = Cos[v, s] = Sqrt[$b\,p$]	similarity, also a kind of *geometric average slope* Cos[w, z] independent of scaling, symmetric
θ, φ, ψ	ArcCos[Cos[v, s]], ArcTan[b], ArcTan[1 / p]
d = Sqrt[1 – k^2] = Sin[v, s]	metric, distance of *slope k* from *slope* 1, off-diagonal Sin = Sqrt[1 – $b\,p$]
ID[d] = $(100\,d)^{1/f}$	transform with sensitivity $f \geq 1$, standard $f = 2$
Cov = Covariance[v, s]	covariance of the centered vectors, has a sign. (Vectors v and s are nonnegative, and b and p will be nonnegative since they regress through the origin)
sign = If[sign[Cov] < 0, -1, 1]	zero covariance (cov = 0) should not affect the sign
SDID[v, s] = *sign* ID[d]	include the sign for majority switches (and also not seeing the measure as a slope itself)

In regression analysis we find both slope and dispersion. The present approach has a focus on the slope, but not at full neglect of dispersion. Both Cos and correlation R are measures of association and slope. We get:

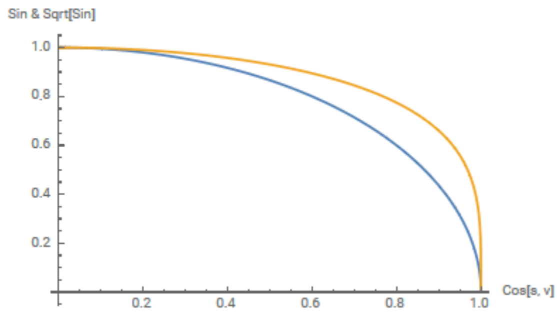
- Vectors in general: Euclidean distance, Cos similarity, Sin distance.
- Centered vectors: The Pearson correlation coefficient R is defined as R = Cos[x – Mean[x], y – Mean[y]].
- When y and x are standardized (by subtraction of mean and division by standard deviation) then $y = R\,x$. Thus $R = 1$ fits the notion of unit or diagonal proportionality (for such standardized variables).
- Unitised variables and RTO: Both Sqrt[1 – R^2] and Sin = Sqrt[1 – Cos^2] are measures of distance. Subsequently, we don't want to use data centered around their means. A vote of {50, 50} and a seat allocation of {50, 50} is perfectly diagonally proportional but causes infinities for R. We still want to keep the regression through the origin (RTO). In that case the cosine is the proper correlation too, see **Section** 10.4. This leads to the definitions in **Table 4**. See Colignatus (2006) on the sample distribution of the adjusted coefficient of determination (for reference only).
- Including Weber-Fechner sensitivity and sign gives: SDID = *sign* 10 √ Sin

4.2. Weber-Fechner consideration

Geometry may be less sensitive to inequality / disproportionality than voters are. A difference of a seat in the Dutch parliament of 150 (0.67%) or the UK parliament of 650 (0.15%) should not be easily overlooked. Such a statement on sensitivity might seem to be a statement from a Social Welfare Function (SWF) and then seem arbitrary. However, it will rather relate to the Weber-Fechner law. [11] Inequality / disproportionality in the higher ranges is not so relevant but small deviations attract our interest.

Figure 6 shows how Sin[*v, s*]^(1 / *f*) follows from estimated Cos[*v, s*]. [12] A sensitivity of *f* = 2 means using √Sin. Looking at some calibrating cases, it appears wise to select the standard value of *f* = 2 indeed. In **Figure 6** the factors 100 and 10 are not included yet.

Figure 6. Sensitivity enhancement with $\text{Sin}[v, s]^{1/f} = \sqrt{[1 - \text{Cos}[v, s]^2]}^{\,1/f}$



Factors 10 or 100 can be introduced to eliminate most leading zeros. **Table 5** reviews two cases. With 100, then *f* = 1 gives the sine values on a scale of 0 to 100. This should not be read as a percentage though. The suggested standard *f* = 2 takes the square root (keeping the sign) and gives values on a scale of 0 to 10. The latter can be read as an inverted (Bart Simpson) report card. Though the index is straightforward math(s), we likely should not attach great value to the second digit. The use of 10 √Sin seems advisable overall.

- Case 1 with large disproportionality is modestly scored as 19.6 by Sin (*f* = 1) on a scale of 100. Yet *f* = 2 better conveys the sizeable disproportionality, with a score of 4.4 on a scale of 10, or 44 or a scale of 100.
- Case 2 with a small disportionality (that flips the majority) gets –4 on a scale of 100 by unadjusted Sin, and gets –2 on a scale of 10. The first is unconvincingly insensitive, and thus we may prefer *f* = 2.

PM 1. The sign indicates the flipping of majority. PM 2. The differences in scales make a design-phase somewhat more complicated. Once the suggested standard *f* = 2 would be adopted in general, then common comparisons can be on the scale of 10 only, without this potential source of confusion on scale 10 or scale 100.

Table 5. Sensitivity and scale 10 or 100

Votes	Seats	*f* = 1 and *sign* 100 Sin	*f* = 2 and *sign* 10 √Sin
{50, 50}	{60, 40}	19.6	4.4
{51, 49}	{49, 51}	-4	-2

[11] https://en.wikipedia.org/wiki/Weber%E2%80%93Fechner_law
[12] Created at Wolfram Alpha as Plot[{Sqrt[1 - x^2], Abs[1 - x^2]^(1/4)}, {x, 0, 1}, AxesLabel -> {"Cos[s, v]", "Sin & Sqrt[Sin]"}]

4.3. The properties for a metric are satisfied

Van Dongen & Enright (2012) discuss metrics [13] that involve cosine and Pearson correlation. Their exposition is like taylored for our purposes. 1 – Cos is a distance too but not a metric. Editing their results to our transform gives **Table 6**. As they did the basic work for us, our own present paper only considers the relevance for the context of electoral studies, and optimises particulars of application. (Though it may perhaps be remarked that I already had the core of this paper before I had the keywords to locate their paper.)

Table 6. Relevant properties of distance and metric

First Quadrant	1 – Cos[*v, s*]	*d* = Sin[*v, s*]	(100 *d*)^(1 / *f*) for *f* ≥ 1
Nonnegative	Yes	Yes	Yes
Zero iff v = s	Yes	Yes	Yes
Symmetric	Yes	Yes	Yes
Triangle inequality	No	Yes	Yes
Thus metric	No	Yes	Yes

Thus |SDID| = Abs[SDID[*v, s*]] is a metric (and deserves the name "measure"). The (negative) sign only is an additional feature to help identify majority switches. If we are not interested in this, then we take the absolute value (and we should avoid squaring).

SDID cannot be turned into a norm, since (i) Cos[*v*, 0] is undefined (though Cos[*v*, 1] might be a standard point of reference), (ii) we don't have SDID[*k v, k s*] = *k* SDID[*v, s*], since obviously we have removed such sensitivity, and Sin has been restricted to [0,1], and (iii) we don't have SDID[*v, s*] = SDID[*v* + *u*, *s* + *u*], e.g. use *u* = -*v*. Especially when vector *u* is in a different plane then the angles have different meanings. Colignatus (2018a) has more on normed spaces.

The property of a metric allows to compare distances (like between Amsterdam, Paris and Rome) but what else ? We can compare earthquakes of 3.2 and 6.1 on the Richter scale, but not really sum them directly. Their sum would be Log[10, 10^3.2 + 10^6.1] for example, if the wavelengths would be additive in such manner. [14] The use of a rule like Sin[A+B] = Sin[A] Cos[B] + Cos[A] Sin[B] seems uninviting. We may consider using ArcSin and add angles, but the vectors are mostly in different planes, and thus the angles have different meanings.

However, Sin and √Sin are on [0, 1] and this would allow using the decomposition of failure rates.

4.4. Example of failure rate decomposition

The following is an example of using the decomposition of failure rates:

- η is a *nonspecific* inequality / disproportionality measure on [0, 1].
- ζ is a *specific* measure, also constructed to fit on [0, 1].
- We allow for overlap.
- Total disproportionality D can be seen as η plus what ζ takes of the remainder $(1 - \eta)$.
- $D = 1 - (1 - \eta)(1 - \zeta)$ or total inequality / disproportionality
- $1 - D = (1 - \eta)(1 - \zeta)$ is a decomposition of equality / proportionality.
- When D is in [0, 1] and we have a specific subcase of ζ also in [0, 1] then we can determine the remainder as $\eta = 1 - (1 - D) / (1 - \zeta) = (D - \zeta) / (1 - \zeta)$, and check on [0, 1].

[13] A portal and no reference: https://en.wikipedia.org/wiki/Metric_(mathematics)
[14] I have considered using Log rather than √ for the transform of Sin, but these functions are quite alike and √ is relatively easier to work with or communicate about.

There are four perspectives that all amount to the same relation. When this decomposition has been defined for Sin then scales using √ Sin must rebase to Sin before using it.

$$\begin{aligned} D &= 1 - (1 - \eta)(1 - \zeta) \\ &= \eta + \zeta - \eta\,\zeta \\ &= \eta + (1 - \eta)\,\zeta \\ &= \zeta + (1 - \zeta)\,\eta \end{aligned}$$

Table 7 gives the decomposition for Holland 2017. Observe that η and ζ define all other results, vertically by addition and subtraction, and horizontally by the transform. The cross term on the right is a remainder from the transformations in the other rows, as $D' - \eta' - \zeta'$.

The total outcome is D = Sin[θ] = Sin[v, s]. The specific factor has been taken as the cosine distance, as an indication of the influence of the slope. We find that the slope in Holland 2017 is fine and contributes little to disproportionality.

Table 7. Specific or Nonspecific, versus Original or Transform, Holland 2017

False data	*Original*		*Transform*	
Measures	Symbol	*sign* 100 Sin	Symbol	*sign* 10 √Sin
Range		1 - 100		1 - 10
Nonspecific	η	6.048	η'	2.459
Specific	ζ	0.194	ζ'	0.441
Cross term	$-\eta\,\zeta$	-0.012	$D' - \eta' - \zeta'$	-0.404
Total	$D = \eta + \zeta - \eta\,\zeta$	6.230	D'	2.496

Comments are:

(1) Sine diagonal inequality / disproportionality in Holland is 6.2% on a scale of 100 using f = 1, and 2.5 on a scale of 10 using f = 2, or 25% by comparison. Thus originally Holland looks fairly proportional, but the Weber-Fechner correction helps us to focus on what needs to be improved in Holland.

(2) For the decomposition in specific or nonspecific causes we must be aware of the cross term so that we do not simply add the scores. This already happens in the original additive scores, and is magnified in the transform. If one mentions only one factor then one might presume that the other part (that is not mentioned) collects all remainder.

(3) The Dutch slope is fairly close to b = 1, and the nonspecific causes dominate. Yet, due to the transform, we see that we should not entirely neglect issues about the slope. The issue rises from 0.2% on a scale of 100 to 0.4 on a scale of 10, or 4% by comparison. There can be a systematic factor, like the use of D'Hondt.

(4) If we start from nonspecific causes with score 2.459 on a scale of 10, and include the specific causes, to arrive at a total of 2.5 on a scale of 10, then it seems as if the slope is fine, and adds little to the disproportionality. If we start from the specific cause with 0.4 on a scale of 10, and then include nonspecific aspects, to arrive at the total of 2.5, then this impression doesn't really change, as the contribution of the slope disappears in the overlap. The transform helps to put matters in perspective, in Weber-Fechner manner.

4.5. Graphical comparison of SDID, ALHID, EGID and WSL

Let us compare the suggested new measure with three main ones in current use. The common measures are not so sensitive, so that a fair comparison with SDID requires a similar Weber-Fechner transform for the common measures.

The Abs / Loosemore-Hanby inequality / disproportionality (ALHID) is Sum[Abs[$w - z$]] / 2 and has a very convincing interpretation. Taking a standard House of 100 seats, it counts the number of seats that are dislocated from the average $w = S\,v\,/\,V$. The factor 2 corrects for double counting. The Euclid / Gallagher (EGID) has the same outcome for binary cases and extreme outcomes as {100, 0, 0} to {0,100, 0}. In the middle ground it reduces differences, as a quadratic function does for values in [0, 1]. The scaling with 100 is for convenience only.

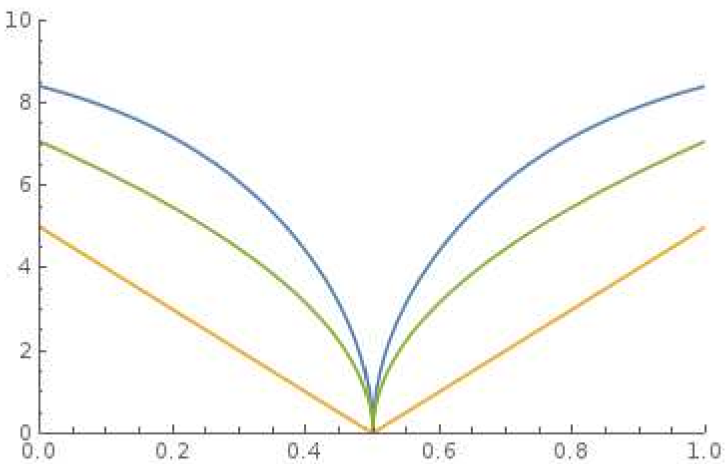
Renwick (2015):

> "The Loosemore-Hanby index is sometimes criticized for, as Lijphart puts it, “exaggerating the disproportionality of systems with many parties” (Lijphart 1994: 60). In fact, however, this index is entirely neutral as to how any given amount of under- or over-representation is divided up among parties. By contrast, the Gallagher index downplays disproportionality the more the votes are fragmented across parties such [that] any given party’s vote–seat deviation is smaller. As the fragmentation of the UK party system has increased over recent years, therefore, the standard measure of disproportionality has, it would appear, increasingly understated the true level of disproportionality."

First consider parties *A* and *B* with votes {*va,* 100 - *va*} and seats {50, 50}. We look at this inversely for a reason explained below.

- SDID: The *sine diagonal inequality / disproportionality* measure, suggested in this paper, is equal to 10 √Sin, where Sin = √(1 – *b p*) where *z* = *b w* + *e* and *w* = *p z* + ε found by regression through the origin (RTO). The factor 10 is for the scale, like a Bart Simpson school report card (the lower the better).
- ALHID: The Abs / Loosemore-Hanby inequality / disproportionality measure, as Sum[Abs[*w* – *z*]] / 2. Conventionally the variables are percentages so that it is in the 0 – 100 range. In this case we divide by 10 to get in the 0-10 range.
- ALHIDT: The transform √ ALHID, so that percentages are in the 0-10 range. The transform focuses attention on the smaller numbers. It still allows the interpretation that squaring the number gives the usual interpretation for dislocated seats. For example a score of 1 on a scale of 10 means also 1 on a scale of 100, and 5 on a scale of 10 means 25% dislocated seats.
- **Figure 7** compares SDID, ALHID / 10 and ALHIDT = √ ALHID.

Figure 7. Votes {*va,* 1 – *va*}, seats {0.5, 05}, with SDID (blue) and ALHID / 10 (yellow) and √ ALHID (Green) [15]



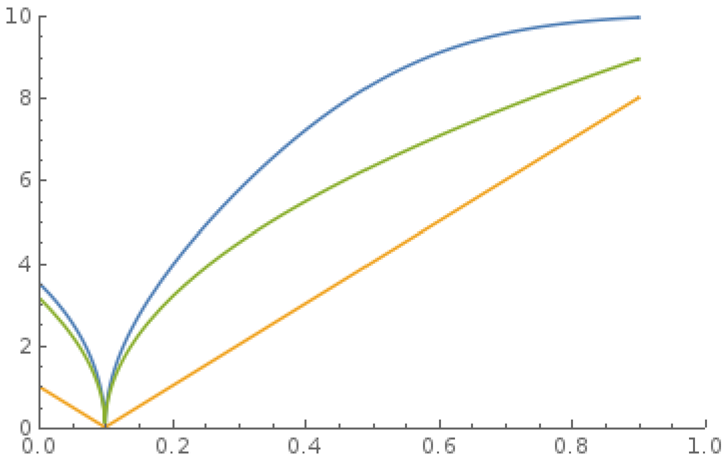
Let us consider parties *A, B* and *C,* with votes {*va,* 10, 90 - *va*} and seats {10, 10, 80}.Party *B* has a proportional allocation. What about the other two parties ?

- EGID: The *Euclidean* / *Gallagher inequality / disproportionality,* √Sum[½ (*z* – *w*)^2], normalised by √2 so that for binary cases it is ALHID (as happens in ternary cases with one fixed). With data in percentages, this ranges from 0 – 100. We scale on 10 (or divide the common percentages by 10).

[15] Wolfram Alpha: Plot @@ {v = {x, 1 - x}; s = {0.5, .5}; 10*{(1 - (1 - CosineDistance[v, s])^2)^(1/4), (ALHID = (Plus @@ Abs[v - s])/2), Sqrt[ALHID]}, {x, 0, 1}, AxesOrigin -> {0, 0}, PlotRange -> {{0, 1}, {0, 10}}}

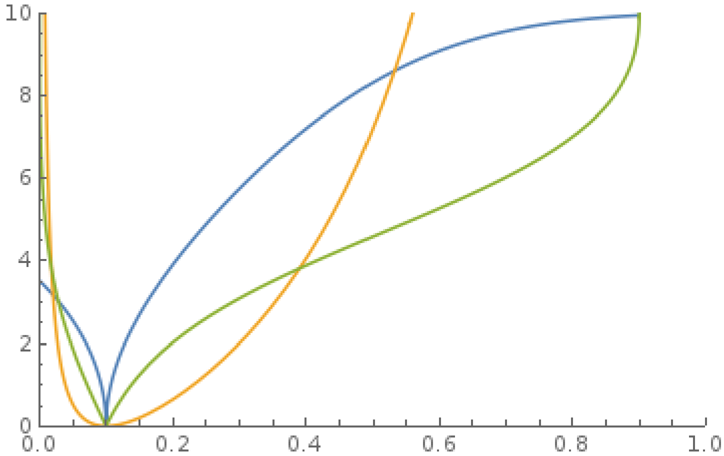
- EGIDT = √ EGID: Since the EGID has a range of 0 – 100, the square root automatically brings it to the 0 – 10 range.
- **Figure 8** thus compares SDID, EGID / 10 and EGIDT = √ EGID. Since this ternary case has one party fixed at some proportionality, EGID reduces to ALHID. The present focus should rather be on the behaviour at 10% rather than the middle 50%.

Figure 8. Votes {*va,* 0.1, 0.9 – *va*}, seats {0.1, 0.1, 0.8}, with SDID (blue) and EGID / 10 (yellow) and EGID Transform (Green) [16]



SDID and EGID are symmetric around *va* = 10%. At 0 they have a simple cut-off. One could argue that if party *A* with 0% of the vote still got 10 seats, that this would be highly disproportional, and that symmetry is no good property. Thus there is also WSL.

Figure 9. Votes {*va,* 0.1, 0.9 – *va*}, seats {0.1, 0.1, 0.8}, with SDID (blue) and WSL / 10 (yellow) and WSL Transform (Green) [17]



- WSL: Webster / Sainte-Laguë as Sum[w (z / w – 1)^2] = Sum[(z – w)^2] / w]. This is asymmetric and has an range from 0 – infinity. With percentages as input it will have even higher values. We will presume percentages, and below graph will divide by 10 again.

[16] Wolfram Alpha; Plot @@ {v = {x, 0.1, 0.9 - x}; s = {0.1, 0.1, 0.8}; 10*{(1 - (1 - CosineDistance[v, s])^2)^(1/4), ((v - s) . (v - s)/ 2)^(1/2), ((v - s) . (v - s)/2)^(1/4)}, {x, 0, 0.9}, AxesOrigin -> {0, 0}, PlotRange -> {{0, 1}, {0, 10}}}

[17] Wolfram Alpha: Plot @@ {v = {x, 0.1, 0.9 - x}; s = {0.1, 0.1, 0.8}; {10*(1 - (1 - CosineDistance[v, s])^2)^(1/4), 10*(wsl = v . (s/v - 1)^2), 10*(1 - 1/(1 + Sqrt[wsl]))}, {x, 0, 0.9}, AxesOrigin -> {0, 0}, PlotRange -> {{0, 1}, {0, 10}}}

- WSLT: The transform 10 (1 – 1 / (1 + √ WSL)) where WSL has been calculated from unit values (percentages divided by 100). (i) The use of the square root makes the measure more sensitive to smaller differences. (ii) The other transform brings it to the 0 – 10 range.
- **Figure 9** compares SDID, WSL / 10 and the transform of WSL. One can now see why we looked at the issue inversely: WSL becomes infinite for *zero votes and still seats*.

We find:

- While we started with ALHID, EGID and WSL and then created SDID as an improvement, we now compare SDID with the transforms ALHIDT, EGIDT and WSLT. Perhaps we might have tried these transforms directly, but it was only because of SDID that this idea surfaced.
- Remarkably, we tend to better appreciate the transforms because they look like SDID.
- The present graphs suggest that voting researchers might simply take the square root of ALHID or EGID that they are used to. [18] We should look at more cases to test this, though.
- I would tend to advise the use of ALHIDT = √ ALHID because of the direct interpretation of dislocated seats, in a standardised House with 100 seats. See below, however.
- WSL remains interesting because (i) the ratio, (ii) the weights, (iii) the sensitivity to when the vote tends to zero.
- The drawbacks of WSL of *insensitivity to small values* and *infinite range* are repaired by WSLT. The only drawback is its asymmetry. This can also be seen as an advantage. If we didn't have an asymmetric measure, then there would be demand for one.

What does this asymmetry mean ? Does the situation of voting and seat allocation really require symmetry ?

(1) Symmetry means: {*a, b*} has the same outcome as {*b, a*}.
(2) An example of asymmetry is a process of growth. Growth of 5% generates a factor 1.05. To get back to 1, we need a factor 1 / 1.05 = 0.9524. A 5% point reduction is only 5% / 1.05 = 4.76% as seen from the new situation. In terms of logarithms, however, we have symmetry again: Log[1.05] = - Log[1 / 1.05].
(3) For votes and seats, the notion of symmetry means the idea that *votes and no seat* is equally bad as *no votes and seats.* This is dubious.
(4) If we had a symmetric measure, then there would arise a demand for an asymmetric measure. Thus it seems optimal that we have both SDID and WSLT. There is no reason to try to create a "symmetric version of WSLT", except perhaps for the weighting scheme.
(5) Balinski & Young (1976:6) mention that Sainte-Laguë formulated his rule but also stated that "one is led to a more complex rule", which would be the method of *Equal Proportions*, or Huntington-Hill (which is somewhat paradoxial w.r.t. footnote 1). I did not look further into this. Though let me remark that the criterion of "relative difference" Abs[*z* – *w*] / Min[*z, w*] is symmetric, and can be made robust on division by zero by looking at the transform Abs[*z* – *w*] / (1 + Min[Abs[*z*]*,* Abs[*w*]]), in which the numerator compares 1 + *w* and 1 + *z,* similar to Exp[*w*] and Exp[*z*].
(6) **Appendix H** shows that using Cos[*z* / *w*, 1] has the undesirable property of multiple local minima. A (dis-) proportionality at one spot would allow a mirroring (dis-) proportionality elsewhere. It seems better to not to allow such compensation.
(7) See Colignatus (2018a) for the relation between WSL and the Aitchison distance.

4.6. Scores for Holland, France and UK in 2017

Table 8 reviews the scores for Holland, France and UK in 2017, for the (half) elections discussed in **Section** 2. It is useful to be aware of *which* is *which*. Presently we are in a design phase. The present discussion is targeted on clarifying that the measure in row 1 is the best measure to adopt in the literature on voting and electoral systems, similar to the Richter scale for earthquakes. The table includes some transforms to highlight some aspects.

[18] The 2015 Canadian election gave a EGID of 12.02 on a scale of 100. The EGIDT gives √12.02 = 3.5 on a scale of 10. This may better convey the inequality / disproportionality. See https://en.wikipedia.org/wiki/Gallagher_index

Table 8. Scores for Holland, France and UK 2017

	2017	Range	UK	France	Holland	Hol-SDID
	EPR or DR		DR	DR 2	EPR	EPR+
	Masked		Yes	No	No	No
1	10 Sqrt Sin	0 - 10	3.7	6.8	2.5	1.5
2	Specific SDID	0 - 10	1.0	3.4	0.4	0.2
3	100 Sqrt Sin	0 - 100	37.1	68.0	25.0	15.5
4	100 Sin	0 - 100	13.8	46.2	6.2	2.4
5	Gini	0 - 100	15.6	41.6	3.6	1.7
6	ALHID	0 - 100	10.5	31.2	3.0	1.3
7	EGID	0 - 100	6.8	21.3	1.7	0.6
8	WSL	0 - inf	12.7	50.1	2.3	0.1
9	ALHIDT transf	0 - 10	3.2	5.6	1.7	1.2
10	EGIDT transf	0 - 10	2.6	4.6	1.3	0.8
11	WSLT transf	0 - 10	2.6	4.1	1.3	0.3
12	#Seats		650	577	150	150
13	#Parties		9	16	13	14
14	CNPs		2.5	3.0	8.1	8.7
15	CNPv		2.9	7.2	8.6	8.6
16	Slope (c = 0)		1.071	1.356	1.028	0.997
17	R^2		0.976	0.729	0.991	0.999

Comments are:

(1) While this paper concerns a design-phase, and while we thus should be alert on the distinction between the scale of 10 or 100, I noticed that I myself was confused on this on some occasions. Thus row 1 uses scale 10 while row 3 multiplies by 10 and thus uses the same measure but on scale 100. Normally we would prefer row 1 over row 3 once confusion is out of the way.

(2) Holland has EPR. The Gini (row 5) has a fine low score of 3.6 (on a scale of 100). The Gini is not symmetric. The Gini includes the wasted vote of 2%. This conforms with the Sin (row 4) of 6.2 (on a scale of 100). However, the plain Gini does not correct for Weber-Fechner.

(3) SDID (row 1) is decomposed in a specific effect on the slope (row 2) and a nonspecific effect of dispersion (not shown). In Holland the slope is fine and there is little contribution to inequality / disproportionality from there.

(4) Originally I was pleased that Holland scored well on equality / proportionality in the Gini, but subsequently this discussion puts more emphasis on the wasted vote. The suggested proper measure (row 1, using $f = 2$) with outcome 2.5 is sensitive to this (25 on row 3). The answer would be that Holland (i) drops using D'Hondt, (ii) allows empty seats or adopts a qualified majority rule that respects this wasted vote, (iii) considers more seats. For readers of Dutch there is Colignatus (2017f).

(5) For comparison, two counterfactual cases for Holland with EPR+ have been created. **Appendix L** contains the *Representative Largest Remainder* (RLR) with a focus on the marginal representative. This still allows the Wasted Vote of blanc votes, invalid votes, and votes for parties that got no seats. One should not neglect this, but consider qualified majority. As an example how this can be dealt with, there is the case in **Appendix F**, of which the results have been included in the last column of **Table 8**. The 2% wasted vote are included as empty seats of a 14^{th} party. The scores improve remarkably. Also remarkable is that the SDID measure still has a relatively high value of 1.5 on a scale of 10. This must be the unavoidable result of squeezing 10 million votes into 150 seats. There is no need to do such calculations for France and UK and produce charts on these, because the outcomes will be similar to the information provided by the scatter plots.

(6) In the decomposition of the Dutch EPR+ score of 1.5 in row 1 and 2, the effect of the slope is almost gone, and the coefficient of determination (row 16) is very close to 1.

(7) France has a staggering electoral inequality of 6.8 (on a scale of 10). The specific score (row 2) helps to diagnose that there is a systematic effect from the slope.

(8) The UK has an electoral inequality of 3.7 (on a scale of 10) This is sizable, see also the Gini of 15.6%. This meaure is quite uncertain however. The UK system of DR masks first preferences. We do not know what the UK voters actually prefer (also because exit polls apparantly do not ask this question on strategic voting).

(9) My own preference originally was for the Gini (row 5). (i) Colignatus (2017b) explains that parties must be sorted on their *s* / *v* scores, to arrive at the proper measurement on the EPR Lorenz and EPR Gini. However, it is not-inconceivable that researchers might make errors on such sorting. Thus a measure is to be preferred that does not depend upon sorting. This leads to SDID (row 1). It is somewhat amazing that considerations on sorting cause such a more fundamental consideration. (ii) However, once SDID had been constructed and then scored on the Taagepera & Grofman (2003) criteria, it was highlighted again that inequality / disproportionality is symmetric and that the Gini inequality measure is not symmetric. See the discussion about **Table 23**. (iii) Colignatus (2017b) doesn't include the Weber-Fechner notion on sensitivity yet. Given this more fundamental suggestion for SDID there is no need to adapt the Gini for sensitivity too.

(10)To rephrase: The Gini tends to correlate with the Sin measure, yet not fully because of asymmetry. Thus we may adopt row 1, that also includes adequate sensitivity.

(11)Webster / Sainte-laguë (WSL) (row 8) tends to correlate with Gini (row 5) or 100 Sin (row 4). WSL is mentioned by Gallagher (1991) as a likely best standard. See the discussion below however.

(12) Paradoxially, the Euclid / Gallagher Inequality / Disproportionality index (EGID) (row 7) has grown to be somewhat of a practical standard in voting literature, contrary to the Gallagher (1991) statement. If we are not uncritical then we may live with this index, that rightly orders Holland < UK < France on a disproportionality scale. However, the proposed index SDID (row 1) better captures distinctions (see below) and has more sensitivity to inequality and nonspecific disproportionality. The WSL has also been transformed (row 11) and remarkably behaves more like EGIDT and ALHIDT.

(13)ALHID is the corner stone for comparisons since it conveys the number of dislocated seats in a House with 100 seats. The ALHIDT is just the square root, and thus is on a scale of 10 again. It is more sensitive, and squaring gives the ALHID, with its interpretation again.

(14)All in all, the ALHIDT, EGIDT, WSLT indeed behave much better – and our yardstick appears to be SDID. They are still not sensitive enough as SDID. For France ALHID finds that 31% of the seats are relocated, and the square root only is 5.6 on a scale of 10, while SDID makes it an earthquake of 6.8 on a scale of 10. Yet one might regard 5.6 as worse-enough, and then prefer ALHIDT for its relation to the number of relocated seats.

(15)The indices on the "concentrated number of parties" are about concentration and not about proportionality. Interestingly, though, they are related to the slopes: $b\ w'w = p\ z'z = w'z$.

5. Spaces: vector and 2D scatter, and unit simplex and its 2D scatter

5.1. Notation

Vectors *v* and *s* contain the data in a sequence. The scatter plots them in parallel.

When we have two vectors *v* and *s* with a common origin, then linear algebra generates the cosine of the angle between them, as k = Cos[θ] = Cos[v, s] = $v's$ / Sqrt[$v'v * s's$]. Thus we also have a slope Tan[θ] = Sin[θ] / Cos[θ]. This angle is between the vectors, in the 2D plane supported by them. In the scatter we might consider Cos[θ] = Sqrt[$b\ p$].

We have been using *v* and *s* both as general vectors in R_+^n and as statistical observations of a point in that space, with the subsequent scatter in {v, s}. This may cause confusion. Properly we choose between either general variables v^* and s^* or observations {v_0, s_0}.

- R_+^n has general vectors *v* and *s*, and allows planes for particular observations. For v_0 and s_0 we get the plane {u ; $u = \lambda\ v_0 + \mu\ s_0$} and the scatter {$v_0, s_0$}.

- R_+^n has general vectors v^* and s^*, and allows planes for particular observations. For *v* and *s* we get the plane $\{u ; u = \lambda v + \mu s\}$ and the scatter $\{v, s\}$.

Using the latter:

- With general variables v^*, s^*, we have diagonal 1 and the observations *v* and *s,* with the three subplanes created by vectors $\{v, s\}$, $\{v, 1\}$ and $\{s, 1\}$.
- The diagonal 1 has only a projection onto a plane $\{u ; u = \lambda v + \mu s\}$, and need not be in that particular plane itself. The angles within the plane with the projection will be different from those with the diagonal in the whole space.
- For more than 2 parties there need not exist weights λ and μ such that $u = 1$ (the unit constant vector). For an arbitrary *u* in this $\{v, s\}$ plane however:

$$\theta v = \text{Cos}^{-1}[\text{Cos}[v, u]] = \theta w = \text{Cos}^{-1}[\text{Cos}[w, u]]$$

$$\theta s = \text{Cos}^{-1}[\text{Cos}[s, u]] = \theta z = \text{Cos}^{-1}[\text{Cos}[z, u]]$$

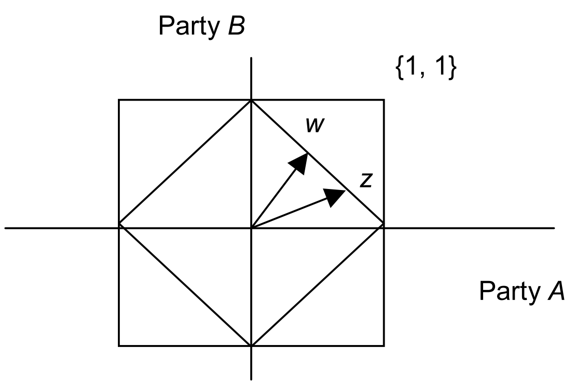
$$\theta = \text{Cos}^{-1}[\text{Cos}[v, s]] = \text{Cos}^{-1}[\text{Cos}[w, z]] = \pm \theta v \pm \theta s \ (\pm \pi) \text{ given } \lambda, \mu$$

5.2. The unit simplex

Division by totals $w = v / V$ and $z = s / S$ transforms to the unit simplex. The unit simplex consists of the vectors themselves, but there is also the *collection of endpoints* of the vectors in a separate plane for themselves.

Figure 10 gives a graphical display for two parties $\{A, B\}$, with approximately $w = \{40, 60\}$ and $z = \{70, 30\}$. While a scatter diagram like **Figure 5** has votes on the horizontal axis and seats on the vertical axis - which suggests looking at the slope *b* or angle φ but also the comparison of individual scores, like in the traditional inequality / disproportionality measures - the graph is now inside-out (and suggests the comparison of the vectors as wholes). This perhaps helps to see that the simplex might be the better way to look at the vectors as wholes. In this case, party *A* gets a higher share of seats than the share of votes. There would be unit vector-proportionality in apportionment if the vectors would overlap (with anchor *w*). Votes and seats are located in the first Quadrant, and we might obliterate the other Quadrants, except that they may help to link up to distance measures that have been developed more in general.

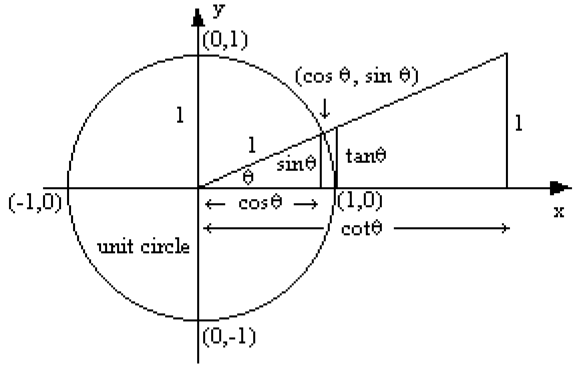
Figure 10. Shares of votes and seats for two parties



5.3. Angles, Cos, Sin and Tan

For the angles and their trigonometric transforms we have **Figure 11** for the 2-dimensional plane with the unit circle and the radius on a proportional line through the origin. If $\tau = \text{Tan}[\theta]$, then $k = \text{Cos}[\theta] = 1 / \text{Sqrt}[1 + \tau^2]$.

Figure 11. Angles φ and ψ in the scatter plot, but we also have θ in the unit simplex [19]



(Colignatus (2009, 2015) proposes to get rid of π for the measurement of angles, and proposes to use the plane itself as a unit of measurement. One may also think of turns, with e.g. ½ turn for 180 degrees and ¼ turn for 90 degrees. Associated with this are the functions xur[α] = cos[α Θ] and yur[α] = sin[α Θ], with Θ written as capital theta but pronounced "archi" (Archimede), with Θ = 2 π. See **Appendix D**.)

For vectors and scatters:

- The regression $z = b\ w + e$ generates b = Tan[φ] or φ = ArcTan[b].
- The regression $w = p\ z + \varepsilon$ generates 1/ p = Tan[ψ] or ψ = ArcTan[1 / p].
- For variables w and z on the unit simplex, $z = b\ w + e$ generates $1'\ z = b\ 1'\ w + 1'\ e$ or $b = 1 - 1'\ e$. For zero error b = 1 or $z = w$.
- It is helpful to think about a target slope τ. For variables z, w in the unit simplex, and plotted in their scatter, τ = 1. If the target proportionality is $z = \tau\ w + e$, for τ = Tan[Φ] for the latter target angle, then a slope specific error is SPE = $(\tau - b)^2 + (\tau - 1 / p)^2$.
- This SPE isn't symmetric. With angles we get symmetric Sin[$\Phi - \varphi$]^2 + Sin[$\Phi - \psi$]^2.
- The angles θ¸ Φ, φ and ψ apply to different planes. The θ is for vectors in the unit simplex (parties in a row) and the Φ, φ and ψ apply to the scatter plot (parties parallel). However k = Sqrt[$b\ p$] can be seen as a kind of geometric average for the scatter plot as well. This would generate the line $y = k\ x$ in the scatter (rather than converting the cos into a tan). We may use k = Cos[θ] = Sqrt[$b\ p$] with error (Tan[θ] – 0)^2 = Sqrt[1 / k^2 - 1]^2 = 1 / k^2 – 1. This is symmetric but not in [0, 1].
- However, if we have Cos as similarity, then we also have Sin as a distance measure.
- When k = 1 or θ = 0, what happens to b ? We shall use regression through the origin (RTO) with nonnegative vectors on the unit simplex, so that $b = w'z / w'w$. Then θ = 0 requires that b = 1 as well. Thus it might suffice to use Cos as similarity and Sin as disproportionality.
- For any vector proportional relation (without dispersion), the use of the unit simplex causes unit or diagonal proportionality there. When $s = \lambda\ v$, then $z = s / s'1 = \lambda\ v / (\lambda\ v\ '1) = w$, so that a target value other than τ = 1 makes little sense. See **Appendix J**.
- However, k = 1 or θ = 0 is a rare occasion, and we want to be able to judge different occasions. Is it the *slope* or the *dispersion* that causes most deviation from θ = 0 ?

5.4. Angular distance measure θ

A proper metric is given by the angular difference between the vectors. For example, the bisector of the angle between w and z generates a vector that can be normalised onto the unit simplex again, that can be said to be "halfway between".

For more dimensions, we get the θ = Cos^{-1}[Cos[w, z]], where the cosine is 1 when θ = 0. We do not have to scale with V, S or n (the number of parties).

[19] http://www.afralisp.net/archive/lisp/bulge.htm

A comparison of $\theta_{Holland}$ and θ_{UK} would be meaningful. Their vectors will be in different planes, but the notion of an angle does not depend upon this. For a distance measure in [0, 1] we can use $d = \theta / (\pi/2)$ with θ in radians or $d = \theta / 90$ in degrees. [20] This is a linear transform that neglects sensitivity. The measure is discussed in Colignatus (2018a), see also **Appendix O** below. Angles and diagonal in the unit simplex, or in the original R_+^n with the nonnegative vectors *v* and *s*, should not be confused with angles and diagonal in the scatter plot (parties parallel).

5.5. Example of the Ternary Plot

It is instructive to look at this in 3D and then project onto 2D again. The case is explained at various locations, and in particular the projection onto the 2D Ternary Plot. [21] The ternary plot differs from the plane of endpoints of the 3D unit simplex. Will Vaughan made an excel sheet for this, [22] and I included some additional triangles, partly in their own spaces.

Let us consider parties *A, B* and *C,* and the votes {50, 49, 1} and seats {60, 40, 0}. One would likely judge this as significantly disproportional, yet not monstrously so (as when *C* would get all seats). **Figure 12** gives the various findings. We take account of:

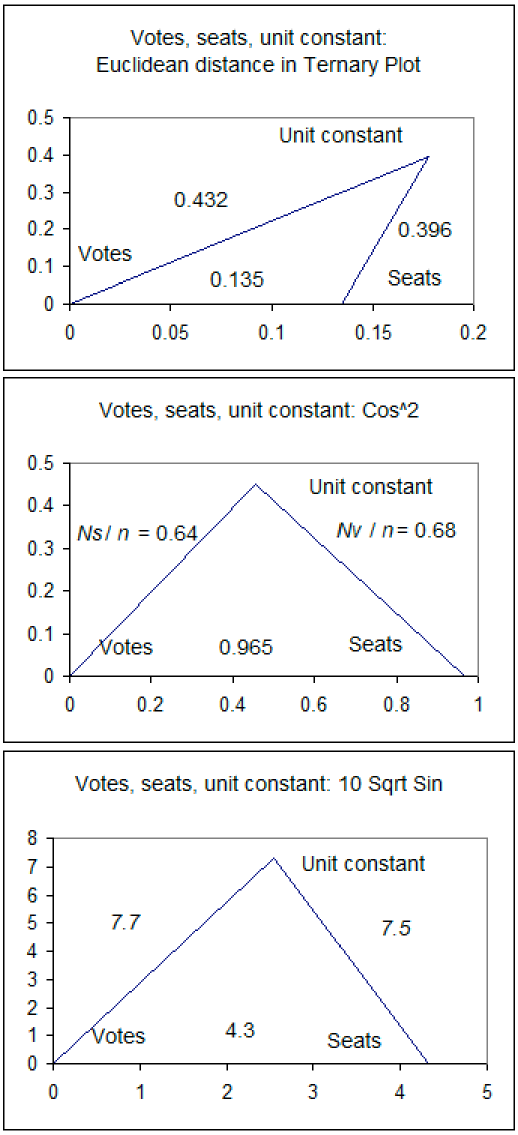
- There are the vectors *v* and *s* in normal space and the vectors *w* and *z* onto the unit simplex (here in percentages). Within these spaces there is also the unit constant vector 1 respectively 1/*n*.
- On the Ternary Plot we can find the end-points of these three vectors. The unit constant vector is at 33.3 points from each side. The 0 seats for *C* are on the AB side. The 60 seats for party *A* can be found by taking distance 60 from the BC side.
- Within the Ternary Plot, we have the Euclidean distance between the end points. The ***top left triangle*** gives these distances (scaled to 1 now). This does not depend upon 3D: for any unit simplex we can find the Euclidean distances between the end points.
- Cos[*v, s*] = Cos[*w,* z] and they give angle θ. With diagonal 1 of R_+^n we have Cos[*v,* 1] and Cos[*s,* 1], and their θv1 and θs1. The unit constant vector will be in the plane of votes and seats only in a special case. Thus only rarely $\theta = \pm$ θv1 ± θs1 (± π), for the relevant selection of signs.
- Nevertheless, with these three separate planes, three cosines and vector lengths, we can use the *law of cosines* to find the mentioned three Euclidean distances for the end points onto the unit simplex: *a*^2 = *b*^2 + *c*^2 - 2 *b c* Cos[*alpha*].
- Having the cosines, it is interesting to observe that the squared cosines Cos[*w,* 1]^2 = 1 / (*n w'w*) = *Nv* / *n*, where *Nv* is known as the "concentrated number of parties, for *v*". In fact *w'w* is the Hirschman-Herfindahl index for concentration. Thus we create the ***middle triangle on the left,*** in its own space of Cos^2, in which two sides of the triangle have this interpretation.
- Subsequently, we transform the Cos^2 into the 10 √ Sin measure. This gives the ***bottom triangle on the left.*** Since this is a metric, there is more meaning than only for the sides of the triangle.
- The three triangles on the left have their own axes (to better insert the text labels), but they help convey the different aspects involved, and how they are related.
- The Euclidean distance on the *endpoints* of the unit simplex already forms a distance measure, and perhaps this might be the one that voting researchers have been looking for. The simplex also allows an interpretation. (i) These distances have been calculated on each pair of vectors with the *law of cosines* applied within their own plane. (ii) If we have the cosine, then it is also possible to directly find the sine. *These measures thus are related, and the sine is simpler.* A major step is rather the Weber-Fechner transform.

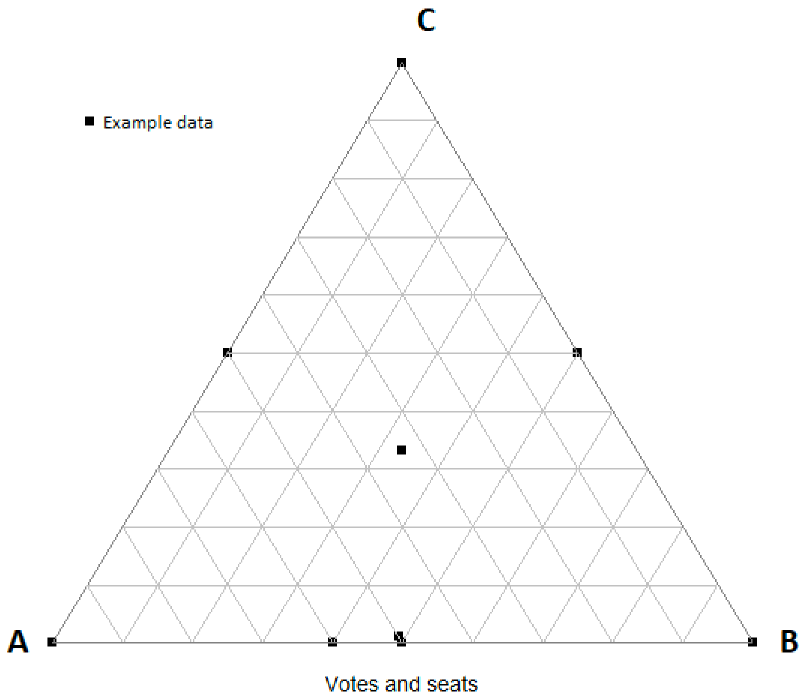
[20] When the angle θ between *s* and *v* is measured using the plane as a unit of account, then $d = \theta / (1/4) = 4\,\theta$ is the transform for [0, 1]. One may appreciate the clarity of this approach.

[21] http://mathworld.wolfram.com/TernaryDiagram.html
https://en.wikipedia.org/wiki/Ternary_plot
http://www.compositionaldata.com/codapack.php

[22] http://wvaughan.org/ternaryplots.html

Figure 12. Ternary plot for votes {50, 49, 1} and seats {60, 40, 0}, with (i) Euclidean Distance on the Ternary Plot, (ii) Cos^2 in its own space, (iii) 10 √ Sin in its own space





Party	*Votes*	*Seats*	*Unit*
A	50	60	1
B	49	40	1
C	1	0	1

Distances	*v to s*	*v to 1*	*s to 1*
Euclid TP	0.135	0.396	0.432
Cos^2	0.965	0.680	0.641
10 Sqrt Sin	4.3	7.5	7.7

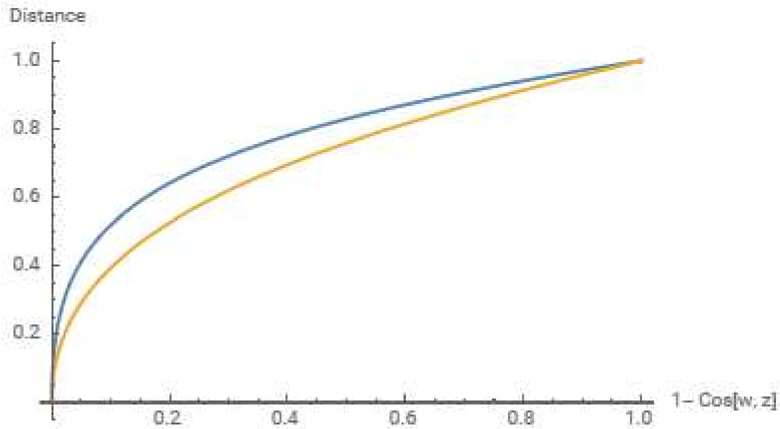
Cos[*w*, 1]^2 =
= 1 / (*n* * *w'w*)
= *Nv* / *n* ((ENPv)

5.6. Weber-Fechner for angles

The Weber-Fechner effect also applies to the angles. Small angles are important for voters but hardly noticed in a linear scale. A transform with logarithmic sensitivity is given in **Figure 13**. [23] One option is Log[1 + θ] and another is Log[1 + 4 θ], both normalised to 1 at θ = π/2 (orthogonality). The latter option scales θ up to the full circle 2π. We already got the same effect with Sin in **Figure 6**. The display uses 1 – Cos, not as a metric but merely for switching the graph. Overall, the square root is easier, because of squaring.

[23] Created in Wolfram Alpha with code Plot[{Log[1 + 4 ArcCos[1 - x]]/Log[1 + 2 Pi], Log[1 + ArcCos[1 - x]]/Log[1 + Pi/2]}, {x, 0, 1}, AxesLabel -> {"1 - Cos[w, z]", "Distance"}]

Figure 13. Distances on [0, 1] with logarithmic sensitivity to the angle between *s* and *v*



5.7. The choice on angle and slope measures

There is no fundamental difference between the angle measure in **Section** 5.4 and the proposed sine diagonal inequality / disproportionality measure that transforms with Sin. Both are based upon k = cosine[v, s] and both are open to a Weber-Fechner transform. The only practical difference is whether we use $\text{Cos}^{-1}[k]$ to have an angle θ as an intermediate variable or whether we use Sin = Sqrt[1 – k^2]. Perhaps I am biased on this, but my impression is:

- It seems to suffice that we have the clarity that θ is best seen as for the whole vectors and that φ and ψ are best seen in the scatter plot.
- For slopes with other values than 1, standards might rather not be formulated in terms of the angle but rather on the slope.
- Likewise for voting we might think that the angle between v and s must be zero, but we rather tend to argue that the scatter should be along the diagonal.
- The discussion about the slope b provides a sound basis in regression (though this would not be needed since we already have the angle).
- The transform for the *slope deviation* 1 - b p to *Sin* seems more tractable than the transform with arccos, log and its normalisation (though we might also use that transform again).
- The perspective on the cosine k as some kind of geometric average slope fits with the discussion on proportionality in terms of slopes, but might also be a somewhat confusing perspective. It works better to emphasize its *double nature.*

Appendix O, **Figure 28**, compares the various measures. Thus one can see that the choice is one of preference only, and likely it is better to remain close to tradition, i.e. not use angles but their transforms (either Sin or slope Tan). **Appendix B** has some residual comments on the unit simplex.

6. Principles: What are we measuring ?

6.1. The Status Quo and the meaning of the majority rule

It appears necessary here to mention a key aspect that is also highlighted in Colignatus 2014), *Voting Theory for Democracy* (VTFD). This is the role of the Status Quo.

The basic situation in voting has a Status Quo. The issue "on the table" is that we consider alternatives to the Status Quo. Only those options are relevant that are Pareto Improving, i.e. that some advance while none lose. The Pareto condition thus gives a minority veto rights against being plundered. Infringements like a rail track in the frontyard will require compensations. Commonly there are more Pareto options, whence there is a deadlock in terms of the Pareto rule, that this rule itself cannot resolve. Then majority voting might be used to break this Pareto-deadlock. Many people tend to forget that majority voting is mainly a deadlock breaking rule. A political body that would use a majority rule without such protection of the Status Quo and minority rights would hardly be called a democracy.

When voting for a new House of Commons then it is generally considered no option to leave the seats empty. In this case there would be no status quo. A situation without a status quo tends to be rather exceptional. National (half) elections however have a high profile, and their lack of a status quo might cause people to forget about the role of the status quo for voting in general.

6.2. Unit proportionality as a power preserving rule

The use of majorities implies the notion of (coalition-) power. Measures are the indices by Shapley-Shubik ("pivotal") or Penrose-Banzhaf ("critical"). The measures have different values but work out the same, though see Laakso (1980). The approach of ordering is also used for the Shapley value, and thus one tends to choose for the SS score, though PB tends to be mentioned more often in the voting literature perhaps because it originated there.

The key point is: to be aware of the mapping from *electorate* to *representatives*:

- The power index would hold for *coalitions amongst the electorate*.
- Electoral equality then is a *power preserving rule,* such that the same coalitions from amongst the electorare are also possible amongst the representative body.

Voters, divided up along their party vote at the time of the (half) election, can form various coalitions, and these possible coalitions are preserved in Equal / Proportional Representation (EPR) by allocating seats to parties in equal proportion to the vote.

The equal representation will not be perfect, also in systems of EPR:

- There are problems of apportionment, e.g. that some parties may get too few votes to meet the natural quota.
- Within EPR there likely will be strategic voting w.r.t. the hoped-for coalition. This is less relevant here, since it would be part of free choice and not something that the system forces. Voters within EPR do not have a *necessity* for strategy, like voters within District Representation (DR) do, giving the risk of wasting their vote.

Though the property of power preservation is imperfect we may still accept that EPR is power preserving in an overall sense. DR has quite different objectives than EPR.

6.3. The unit of account: voter, representative, party, state

The issue has been discussed in the introduction, **Section** 1.1. Are we to judge votes and seats based upon the disaggregate or the party-aggregate level ? The answer is that we should do the first, if only we have the data.

To determine inequality / disproportionality, we should have information about the candidates, and the levels of S and V as well, or Q or T, and this information should not be lost in the presentation of party vectors w and z only. A yardstick might be that only parties and candidates with more than 0.5 Q are recorded in the international databases, see apportionment RLR in **Appendix L**, and that others are collected in a subcategory for the Wasted Vote. If we have this information then we can determine the optimal apportionment $\hat{s}$ = $\hat{G}$ RLR[S, v_K] and $v = \hat{G}\, v_K$ in which the aggregation matrix G is redesigned for the allocation of the wasted vote w.r.t. the optimal $\hat{s}$. The yardstick of measurement is v, and the inequality / disproportionality must be measured on s. We then have the distances:

- SDID[v, $\hat{s}$] the unavoidable error due to approportionment itself (integer representation)
- SDID[s, $\hat{s}$] or *party or other* bias, due to the local convention other than RLR
- SDID[v, s] or *face value* or *conventional* disproportionality (based upon The Party)

- Given the property of a metric: SDID[v, $\hat{s}$] + SDID[s, $\hat{s}$] ≥ SDID[v, s].

The conventional approach thus underestimates true total disproportionality. But conventional scores do not account for the unavoidable error, and then might overstate the problem too.

The discussion about inequality / disproportionality contains some confusion on these issues.

(i) There can be a focus merely on method without looking at principle. Methodological Individualism leads to RLR so that the idea to look at methods that Idealise The Party is dubious. Studies on methods might argue that they apply to v_K rather than v, but there RLR applies, so why consider other methods ? Of course one can always study methods, but then the claim should perhaps not be that one studies voting but some form of Idealism. Relevant would be studies into methods of aggregation, like second rounds and such.

(ii) There tends to be a distinction between authors coming from countries with Proportional Representation (EPR), for which the improvement of proportionality is of marginal value and mostly a political issue, with politicians wondering what marginal change really is worth the effort, and authors coming from countries with District Representation (DR), for which the degree of proportionality is irrelevant (except for USA State seat apportionment like Huntington-Hill (though see footnote 1)). The discussion about the choice of EPR or DR is key for this part of the literature of voting, but the apportionment rule is rather irrelevant in this, because there is already RLR. The Lorenz curve and Gini coefficient show the key difference between EPR and DR, and this is not an apportionment rule but a general measure of inequality, now applied to voting, see Colignatus (2017b).

(iii) There is the distinction between empirical scientists (with also a training in moral philosophy) versus mathematicians (trained on abstraction). Abstract research may draw examples from real life, like the Alabama paradox, but that does not mean that this type of research is empirical, because mathematicians are trained for abstraction and not for empirical research, and they use reality only for examples. If they would study reality, they would not confuse The Party with the party marginal candidate. For example, Balinski [24] and Young [25] have Ph.D.s in mathematics, and one should not be distracted by the fact that they have been working at institutes of economics. We can greatly value their work, and it fits mathematics, but they themselves will agree that it is important to be aware of limitations to application to empirics.

The literature on inequality / disproportionality has a bias in comparing z to w, or comparing z / w to 1, as if such comparison matters on content, while the true issue is that HLR / RLR is the fundamental apportionment method and that countries with DR are better off with EPR. This present paper links up to that very literature and mention of SDID thus suffers from the same bias and distraction. Perhaps it doesn't matter in practice anyway, since we discovered the "implicit regression" $w' \tilde{e} / w'w = b - 1$, so that also WSL must answer to the meaning of this b. Anyway, **Table 1** clarifies that HLR / RLR for the (half) election of representatives is optimal, and the use of other measures would tend to be distractive. However, our findings should bring some clarity. All in all, the reader should not be distracted by our use of the term "party".

6.4. A method of measuring

Consider **Table 1** again. We had votes {25, 100, 200}, found a floor allocation of {0, 3, 6} and allocated a remainder to get at {1, 3, 6} or {0, 3, 7}. Thus we had $s = Ap[S, v] = T\,v + \breve{u}$ or $z = w + \tilde{e}$. Regression through the origin (RTO) gives $z = 0.98\, w + e$ for {1, 3, 6} and $z = 1.09\, w + e$ for {0, 3, 7}. It may be doubted whether the latter comparison of w and z is the relevant and most insightful one. It seems that we should rather focus on $\hat{s}$ = RLR[S, v]. Other seat allocations can be judged against this standard.

- ALHID[{25, 100, 200}, {1, 3, 6}] gives 2.3 and for {0, 3, 7} we get 8.5.

[24] https://en.wikipedia.org/wiki/Michel_Balinski
[25] https://en.wikipedia.org/wiki/Peyton_Young

- ALHID[{1, 3, 6}, {0, 3, 7}] is 10, and symmetry does not allow us to judge which seat allocation is best. We might argue that ALHID should first determine $\hat{s}$, and then measure inequality / disproportionality with that as the base. It is a shortcut to replace $\hat{s}$ by v itself. (We also mentioned this in the above.)
- Comparing {0, 3, 7} to the optimum gives dif = {-1, 0, 1}, which gives 10% of S as error, as Sum[Abs[dif]] / 2 / S, or scaled to 100 by ALHID. Or, with RLR, Green needs only one defector for a stalemate, while in The Party approach it would require 2.
- A criticism of ALHID is (i) that it is not sensitive to small differences, (ii) that it would be neutral for the transfer of a seat from a large party to a small party, or the reverse. The latter comes with insensitivity to the existential question for marginal representatives; but in the above it does a good job.
- Still, the assumption in this paper is that we are presented with v and s, and may not have sufficient data for an apportionment. We try to use these data as best as possible.

6.5. Reasons for measuring

We thus arrive at some observations regarding the reasons why we would want to measure electoral inequality for electoral sytems (and due to the writing sequence with repetition):

- Equal Representation ("proportionality") is a power preserving rule, such that majorities *in the votes v* would be preserved *in the apportioned seats s,* and it suffices that we make sure that the proportions are the same. This still is consistent with requiring that we look at party marginal representatives rather that party averages for *votes per seat.*
- Current methods for apportionment are rather crude and focus on proportions only. It would be a next step in apportionment when parliaments would apportion the seats in such manner that coalition-powers are preserved indeed.
- The discussion should not distract from the importance of HLR or *Representative Largest Remainders* (RLR) (**Appendix L**) compared to other methods of apportionment and suggestions to measure disproportionality.
- Given that EPR and DR have different objectives, the measurement of inequality / disproportionality has limited value in comparing these electoral systems. Yet, when the topic arises, it will still be helpful when this measure is sound.
- A metric can also be used to compare different (half) elections in countries, and across different countries, also over time, with all the restrictions of such comparisons involved.
- On the criterion of *votes per seat,* researchers better highlight the distinction between the *party average* and the *party marginal representative* scores. There appears to exist a general confusion on this, apparently also in the voting research literature. Even if the local standard looks at The Party, researchers would do well in highlighting the alternative, so that readers are alerted to this common confusion.

6.6. Specific and nonspecific factors, inside or on top

In statistics it suffices to test whether b differs from some target T for some level of statistical significance. Presently, we don't have stochastic assumptions. We are looking for a summary measure that includes the error when b differs from T. Phrased a bit differently: Karl Pearson's R combines *association* and *slope,* and we want to disentangle them by including the *target slope T* explicitly. Looking for slope-specific measures, we found the cosine, that has the duplicity of being both a general measure on similarity and a geometric average of slope. This duplicity increases complexity, but also generates scope for flexible application.

(1) Sin can be used stand-alone, and interpreted at will as either general dissimilarity or as an indicator of deviation of slope from diagonal (and then use ArcSin).
(2) If Sin is taken as general dissimilarity, then we can decompose it further into a slope specific component (taking the cosine distance) and remaining nonspecific dispersion. This causes **Section 7**.
(3) If Sin is taken as general dissimilarity, then we can add a slope-specific test on top. This causes **Section 8**. For example, votes and seats might be proportional, but we would want to correct for turnout that indicates a disproportionality.

(4) Considerations about a "target slope" (turnout) differ from the "decomposition w.r.t. the slope" (neglecting turnout). Otherwise there would arise the complexity to add up the two aspects. However, the test example in **Section 8** anyway uses the slope as an example, for this was an easy way to see how it works out. One should not take this test case as a proper example, unless, indeed, one thinks that the sine is insufficiently sensitive to the slope b. With all this flexibility we also have more complexity.

7. Decomposition of "sine diagonal inequality / disproportionality" (SDID)

The crux is in **Section** 7.6 but there are some clarifying steps first.

7.1. Regression through the origin with variables on the unit simplex

The following should assist those readers who are not familiar with regression through the origin or using variables on the unit simplex (as I was on both when starting on this topic). There are some pitfalls to warn about. Eisenhauer (2003) was very helpful for RTO. Kozak & Kozak (1995) discuss the same issue for wood science. We do not want to *explain* s by v, in which case we would be very careful w.r.t. the exclusion of the constant. Instead, we want to design a measure. Yet this uses the same linear algebra. The relevant distinctions are (i) between true values versus estimates, and (ii) between level variables versus unitised variables. The discussion assumes that awareness of the distinction between the scatter plots in **Section** 2 (parties parallel) and the unit simplex in **Section** 5 (parties in a vector).

7.2. True variables v* and s* and particular observations v and s

A proportional relationship for 1D variables is best described by the 2D line $\lambda y + \mu x = 0$, which coefficients may be normalised on the unit circle. For nonzero λ this reduces to $y = T x$ with $T = - \mu / \lambda$, where slope T also is the tangent of the angle of the line with the (horizontal) x-axis. For vectors this generalises into vector-proportionality, with now a plane $u = \lambda y + \mu x$ and then choosing $u = 0$ so that $y = T x$ again. For example, $y = \{1, 2, 3\}$ and $x = \{2, 4, 6\}$, then $T = ½$, and we would see a line without dispersion in the scatter plot.

- The notion of unit proportionality as in the line $y = 1 x + 0$ is a mathematical concept, while in statistics with dimensions we can rebase the variables, so that there need not be a natural base for 1.
- For voting, there are natural bases in the individuals and seats. Larger parliaments may have more scope for a better fit. Still, normalisation onto the unit simplex makes sense.
- At first sight it is not clear where voting differs from other applications, say the ratio of 1 car per 2 persons. Any vector-proportional relation $s = T v$ also holds for the totals. Thus $S = 1's = T 1'v = T V$. In $\{w, z\}$ space it reduces to scatter diagonal with $z = w$ because division gives $z = s / S = T v / S = w$. When any vector-proportional relationship (also non-unity) is transformed onto the unit simplex, then they become unit or diagonal proportional in the scatter, and we lose the original information.
- There also is dispersion. The basic solution is to make the distinction between the true vector-proportionality $s^* = T^* v^*$ that holds for all *elements* on one hand, and on the other hand observations (think of: errors in variables) $s = B v + u$, for which we only have the definition for the *sum totals* as $T = S / V$. This also generates $\tilde{u}$ from $s = T v + \tilde{u}$. (This is a different use of stars as in **Section 5.1** on the notation of vector spaces.)
- For theory we have $z^* = w^*$ but for the data $z = w b + e$. We divided by S and took $b = B / T$ and $e = u / S$. Thus we should not focus only on parameters B and b but also be aware of the hidden $Q = V / S$ or $T = 1 / Q$, and perhaps consider T as an estimate on T^*. In this section we will further neglect T^* and the reader is referred to **Appendix J**.

7.3. Cos, slope and concentrated numbers of parties

Table 9 reviews the relations. For readibility we drop the stars in the theory column, with a risk that we overlook the relations on the norm T^* (**Appendix J**).

Table 9. Norm and estimation, in levels and unitised, regression through origin (RTO)

Norm or true situation [26] [27])	*RTO in levels*	*RTO on unit simplex*
Theoretical s and v	Observed s and v	$w = v / V$ and $z = s / S$
$s = T v$	$s = B v + u$	$z = b w + e, \quad e = u / S$
$1' s = T 1' v$ or $T = S / V$	$1' s = S = B V + 1' u$	$1' z = 1 = b + 1' e$
	$T = S / V$ and $s = T v + \tilde{u}$	$z = w + \tilde{e}$ with $\tilde{e} = \tilde{u} / S$
$s = S / V v$, thus $z = w$	RTO: $1'u$ need not be 0	RTO: $1'e$ need not be 0
	$B = S / V - 1' u / V$	$b = 1 – 1' e = B V / S$
$s's$ = SST = SSX + SSE	Compare B with $T = S / V$	Compare b with 1
Cos = $s'v / \sqrt{(v'v\ s's)} = 1$	Cos = $s'v / \sqrt{(v'v\ s's)}$	Cos = $z'w / \sqrt{(w'w\ z'z)}$
$B = s'v / v'v = T$ Cos	$B = s'v / v'v = \sqrt{(s's / v'v)}$ Cos	$b = z'w / w'w = \sqrt{(z'z / w'w)}$ Cos
$R^2 = T^2\ v'v / s's$ (SSX / SST)	$R^2 = B^2\ v'v / s's$ (SSX / SST)	$R^2 = b^2\ w'w / z'z$ (SSX / SST)
$R = T \sqrt{(v'v / s's)}$ = Cos	$R = B \sqrt{(v'v / s's)}$ = Cos	$R = b \sqrt{(w'w / z'z)}$ = Cos
$T = s'v / v'v$ (alike covar)	1 - R^2 = 1 - Cos^2 = $u'u / s's$	1 - R^2 = 1 - Cos^2 = $e'e / z'z$
$T = s's / s'v$ (alike covar)	Cos = 1 $\leftrightarrow$ $u = 0$	Cos = 1 $\leftrightarrow$ $e = 0$
$T = \sqrt{(s's / v'v)}$ (no covar)	Sin^2 = $u'u / s's$ = SSE / SST	Sin^2 = $e'e / z'z$ = SSE / SST

Key points of RTO on the unit simplex are, using mostly the last column:

- The sum of errors $1'u$ or $1'e$ need not be 0, but $1'\ \tilde{u} = 0$ and $1'\ \tilde{e} = 0$.
- It is a contribution of compositional data to RTO that we now also look at $T = S / V$ and $s = T v + \tilde{u}$.
- Perhaps useful to be aware of: $b\ w'w = p\ z'z = w'z = b / Nv = p / Ns$ while $b / p = Ns / Nv$ is also the square of another geometric average of slope, Sqrt[b / p], see **Section** 10.6.
- The coefficient of determination R^2 applied to non-centered data gives the ratio of SSX / SST, with: SST = sum of squares total, SSX = sum of squares of the explanation = SST - SSE. [28]
- For RTO, Cos takes the role of the coefficient of determination R^2 in OLS with a constant. We might simply use Cos^2. However, it is clearest to write R = Cos.
- $v'\ u = w'\ e = 0$, following the estimates for B and b (or they are chosen such).
- The "analysis of squares" (no deviations and thus no variance) on SST gives: $z'z = (b\ w' + e')\ (b\ w + e) = b^2\ w'w + 2bw'e + e'e = b^2\ w'w + e'e$ with $w'\ e = 0$. Then: $1 = b^2\ w'w / z'z + e'e / z'z = R^2 + e'e / z'z$. We also have R = Cos.
- Sin^2= 1 - Cos^2 = $e'e / z'z = e'e / (b^2\ w'w + e'e)$
- Thus there is a perfect fit $e = 0$ if and only if Cos = 1, when $b = 1$ and $z = w$.
- Observe that $b = 1 – 1'\ e = B\ V / S$ so that B has a relevant dimensional factor.

[26] This column is not to be confused with observations $s = T v + \tilde{u}$, see **Table 2**.
[27] This column drops the stars for readability. For theory we have $s^* = T^*\ v^*$ for the *vectors*, but in estimation we have $s = B v + u$ and thus only $T = S / V$ for the *totals*. Thus there are not only parameters B and b but also an estimate T on T^* (or perhaps institutionally given $T = T^*$). See **Appendix J**.
[28] Often the abbreviation SSR is used, SSR = sum of squares of the regression, but then confusingly with SSR = sum of squares of residuals = SSE = sum of squared errors. The use of SSX has X, of both "explanation" and the variably x in $y = b\ x + e$.

- With $z = s / S$ and $e = u / S$, $\text{Sin}^2 = e'e / z'z = u'u / s's$, and thus it doesn't matter whether we regress on the levels or the unitised variables to find Cos and Sin.
- The Hirschman-Herfindahl concentration indices are $w'w$ and $z'z$. They are known in the voting literature as the inverse "effective number of parties" $Nv = 1 / w'w$ and $Ns = 1 / z'z$. It hasn't been clarified what "effectiveness" would be, and thus we use the better term "concentrated number of parties" (CNP). They now take the place of the covariance in the relation between Cos and b.
- We tend to regard $\text{Cos} = b \sqrt{(w'w / z'z)} = b \sqrt{(Ns / Nv)}$ as the explanation of Cos. We may also see this as an explanation of the slope as $b = \text{Cos} \sqrt{(Nv / Ns)}$.
- The footnote in the table is relevant unless you already saw the distinction between the columns, or the distinction between theoretical s^* and v^* and observed values s and v.

7.4. Analyses of squares for the direct error

Table 9 decomposes SST into SSX and SSE for the model $z = b\, w + e$, but there is also the model $z = w + \tilde{e}$. The square of the Euclidean distance $\tilde{e}'\, \tilde{e} = (z - w)'(z - w) = 2\ \text{EGID}^2$. We would be interested in a relation between EGID and SDID.

(1) A direct result from $z = z$ is that $e = \tilde{e} + (1 - b)\, w$ or $\tilde{e} = e - (1 - b)\, w$
(2) $\tilde{e}'w = e'w - (1 - b)\, w'w$, and with $w'e = 0$, also $\tilde{e}'w / w'w = b - 1$, or $b = 1 + \tilde{e}'w / w'w$. The direct error appears to have an "implicit regression"
(3) Subsequently $\tilde{e}'\, \tilde{e} = e'e + (1 - b)^2\, w'w$ because $w'e = 0$
(4) We already had $\hat{e}'\hat{e} \le e'e \le \tilde{e}'\tilde{e}$ and see this confirmed by the sum of squares
(5) Rewriting $z = w + \tilde{e} = w - (1 - b)\, w + e$ decomposes into: what should be the case (w), subtracts an error on slope ($(1 - b)\, w$), and retains dispersion (e).
(6) Rewriting $z = b\, w + (1 - b)\, w + \tilde{e}$ decomposes into: what has been explained ($b\, w$), the additional error from the ideal 1 ($(1 - b)\, w$), and retains the error for the ideal ($\tilde{e}$)
(7) The present SST is $z'z = w'w + 2w'\, \tilde{e} + \tilde{e}'\, \tilde{e}$. Using (1) gives:
$z'z = w'w + 2\, (w'e - (1 - b)\, w'w) + \tilde{e}'\, \tilde{e}$ or
$z'z = (2\, b - 1)\, w'w + \tilde{e}'\, \tilde{e}$.
When $b < ½$ then that coefficient will be negative. Dividing by SST:
$\text{SSE} / \text{SST} = \tilde{e}'\, \tilde{e} / z'z = 1 - (2\, b - 1)\, w'w / z'z$
where the latter $\text{SSX} / \text{SST} = (2\, b - 1)\, w'w / z'z$ is a form of the coefficient of determination for this "implicit regression", that however might be negative
(8) To find a relation with Sin, we can use $\text{Cos}^2 = b^2\, w'w / z'z$.
Then $\text{SSX} / \text{SST} = (2\, b - 1)\, \text{Cos}^2 / b^2$, or:
$\text{SSE} / \text{SST} = \tilde{e}'\, \tilde{e} / z'z = 1 - (2\, b - 1)\, \text{Cos}^2 / b^2$
(9) We find the same directly with slightly simpler algebra, using (3). The interpretation is that Sin makes the Euclidean distance relative and subtracts a value h (because Cos can see proportionality that EGID doesn't pick up), and then rebases with this h to remain in [0, 1].

$$\tilde{e}'\, \tilde{e} = e'e + (1 - b)^2\, w'w$$

$$\tilde{e}'\, \tilde{e} / z'z = e'e / z'z + (1 - b)^2\, w'w / z'z$$

$$\tilde{e}'\, \tilde{e} / z'z = \text{Sin}^2 + (1 - b)^2\, \text{Cos}^2 / b^2$$

$$\tilde{e}'\, \tilde{e} / z'z = \text{Sin}^2 + h\, (1 - \text{Sin}^2) \text{ using } h = (1 - b)^2 / b^2 = (1 / b - 1)^2$$

$$\tilde{e}'\, \tilde{e} / z'z = h + (1 - h)\, \text{Sin}^2$$

$$\text{Sin}^2 = (\tilde{e}'\, \tilde{e} / z'z - h) / (1 - h) \qquad \text{(* key *)}$$

$$\text{Sin}^2 = (2\ \text{EGID}^2 / z'z - h) / (1 - h)$$

(10) Perhaps this version of the second line is more attractive for some:

$$2\ \text{EGID}^2 / z'z = \text{Sin}^2 + (1 - b)^2\ Ns / Nv$$

$\text{Sin}^2 = 2\ \text{EGID}^2 / z'z - (1 - b)^2\ Ns / Nv$

(11)Recall that we have $\text{Sin}^2 = 1 - \text{Cos}^2 = 1 - b^2\ Ns / Nv$. Reworking the latter gives a relation between EGID, the slope and the CNP terms. Likely **Section** 11.6 on the law of cosines is more relevant though.

7.5. Specific and nonspecific disproportionality

We already discussed the decomposition of Sin and now it is time to look into detail.

The idea is to decompose inequality / disproportionality into a part that depends specifically upon the slope and a part that depends upon nonspecific dispersion. This notion runs counter to the original intuition that $\text{Cos} = \sqrt{(b\ p)} = \gamma$ would be a geometric mean slope itself. Now we look at Cos as a measure of similarity, such that Sin is dissimilarity, and might be decomposed into contributing factors *slope* and *dispersion.* When we regard Sin as a *slope-diagonal deviation* measure itself, then this is a curious "Droste-effect", but when we regard Sin as a plain index then the question on the decomposition is quite valid.

Proportionality requires a proper slope and low dispersion, and their combination generates a multiplication of factors:

$$1 - \text{Sin} = (1 - \eta)\ (1 - \zeta)$$
$$\text{Sin} = \zeta + (1 - \zeta)\ \eta$$

in which:

- $1 - \zeta = \sqrt{(b\ p)} = \text{Cos} = \gamma$ is slope-specific, the positive contribution of a good slope
- $1 - \eta$ is nonspecific, a contribution to proportionality by less dispersion
- ζ are the errors remaining after using a good slope or when not using a good slope
- $\zeta = 1 - \text{Cos}$ is actually the Cosine Distance (not a metric)
- η is the remaining error for dispersion, and from the above $\eta = 1 + \text{Tan}[\theta] - \text{Secant}[\theta]$
- and there is the term for the overlap.

Using $\text{Sin} = \text{Sqrt}[1 - \gamma^2]$:

$$1 - \eta = (1 - \text{Sin}) / (1 - \zeta) = (1 - \text{Sqrt}[1 - \gamma^2]) / \gamma$$

This is nonnegative and $(1 - \text{Sqrt}[1 - \gamma^2]) \leq \gamma$ or $1 - \gamma \leq \text{Sqrt}[1 - \gamma^2]$ for values of γ in [0, 1]. The main question is $\text{Limit}[1 - \eta, \gamma \to 0]$. We multiply numerator and denominator with $1 + \text{Sqrt}[1 - \gamma^2]$ and thus eliminate the Sqrt, and find the limit 0. Thus we have a proper decomposition according to the failure rate model.

Presumably, above decomposition is similar to looking at $e'e$ versus $1'e = 1 - b$. **Appendix K** contains some potential decompositions that failed.

7.6. Decomposing the SDID measure

The former section decomposes the sine into a component for the slope and a remainder for nonspecific dispersion. Clarity is served by putting this explicitly in **Table 10**. The example for Holland 2017 has already been given in **Section** 4.4.

Table 10. Decomposition of SDID measures (standard f = 2)

d = Sin[v, s] = Sin[θ]	Sqrt[1 – Cos[v, s]^2]
d	1 – (1 – η) (1 – ζ) (without *sign* * 10 and f) decomposition of d
ζ	1 – Cos[s, v], cosine distance, Cos = √(b p)
η	1 + Tan – Secant, follows from d and ζ
η' = NonspecificID[v, s, f]	*sign* $(100\ \eta)^{1/f}$ transform with sensitivity $f \geq 1$, standard f = 2, include the sign for majority switches
ζ' = SpecificID[v, s, f]	*sign* $(100\ \zeta)^{1/f}$
SineDiagonalID[v, s, f]	*sign* $(100\ d)^{1/f}$
SDID[v, s] *also denoted as sign 10 √Sin*	SineDiagonalID[v, s] taking f = 2 *sine diagonal inequality / disproportionality*

8. *Investigating a potential bias of SDID for HLR, relation of errors to remainders*

8.1. Comparing on Balinksi & Young 1975 cases

Beumer (2010:31) copies a case by Balinksi & Young 1975 with apportionment of 26 seats. Beware of the effect of the small number of seats. SDID, Gini, ALHID and EGID favour case ZD. A minimax method, also on power, gives case ZE. Apportionment would use constrained minimization anyway. We need to show WSL with more digits to find that it favours case ZB. Thus, we wonder whether SDID has a bias on Hamilton / Largest Remainders (HLR).

Table 11. Comparing JDH, WSL, HH, HLR, Dean, Adams and Minimax

	1	2	3	4	5	6	7	8	9	10	11	12	13
			10√Sin	Specif	100√Sin	100 Sin	Gini	ALHID	EGID	WSL	ALHIDT	EGIDT	WSLT
Ex.	V	S	10	10	100	100	100	100	100	inf	10	10	10
ZA	9061	10	2.7	0.5	27.1	7.4	4.4	3.6	2.9	0.666	1.9	1.7	0.755
JDH	7179	7											
	5259	5											
	3319	3											
	1182	1											
ZB	9061	9	2.6	0.5	26.1	6.8	4.3	3.2	2.6	0.637	1.8	1.6	0.739
WSL	7179	8											
	5259	5											
	3319	3											
	1182	1											

ZC	9061	9	2.5	0.5	25.3	6.4	4.1	2.9	2.3	0.646	1.7	1.5	0.744
HH	7179	7											
	5259	6											
	3319	3											
	1182	1											
ZD	9061	9	2.4	0.4	24.2	5.9	3.7	2.6	2.1	0.713	1.6	1.5	0.779
HLR	7179	7											
Dn	5259	5											
	3319	4											
	1182	1											
ZE	9061	9	2.6	0.5	26.4	6.9	4.5	3.1	2.5	2.363	1.8	1.6	1.332
Ads	7179	7											
MM	5259	5											
	3319	3											
	1182	2											

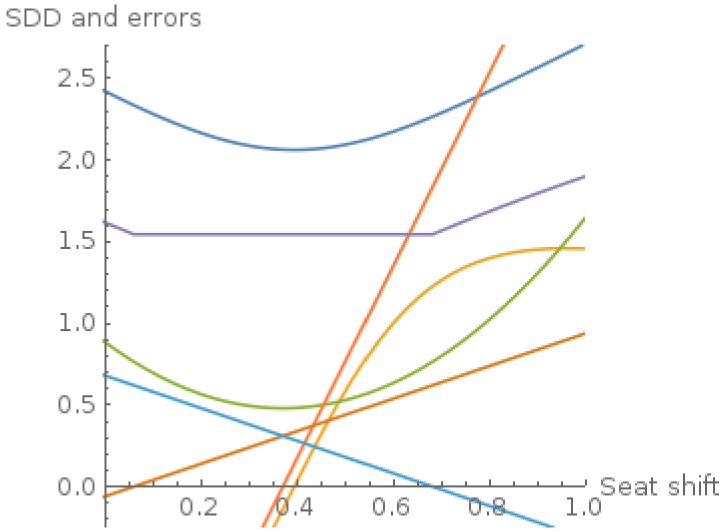
8.2. Fundamental property: remainder → error → Euclidean distance → Sine

This question on the bias has basically been answered by **Section** 7.4 that gave us already:

$$\text{Sin}^2 = (\tilde{e}'\,\tilde{e} \,/\, z'z - h) \,/\, (1 - h) \qquad \text{for } h = (1 - b)^2 \,/\, b^2$$

in which $\tilde{e}'\,\tilde{e}$ = 2 EGID^2, and we also found $b - 1 = \tilde{e}'w \,/\, w'w$. The interpretation is that Sin takes the Euclidean distance relative to SST and subtracts a value h based upon the slope, because Cos may see proportionality that EGID doesn't pick up. Sin rebases with $(1 - h)$ to remain in [0, 1]. The slope divides the area to the diagonal in two parts: b and $1 - b$. This h then is the ratio of their squares. A good fit of EGID close to $b = 1$ requires less correction, a bad fit more. **Figure 14** plots SDID (blue) and $\tilde{e}'\,\tilde{e}$ (green) assuming a gradual rise of A at the cost of D, i.e. $s = \{9 + x, 7, 5, 4 - x, 1\}$. Thus ZD is on the left and ZA on the right.

Figure 14. Transfer of 1 seat from case ZD to ZA, integer values at 0 and 1, errors, SDID (blue), Euclid (green), derivatives w.r.t. *v* (yellow, red), HLIT (purple) [29]



[29] v = {9061, 7179, 5259, 3319, 1182}; s = {9 + x, 7, 5, 4 - x, 1}; w = v/26000; z = s/26; e = z - w; SDID = Simplify[10*(1 - (1 - CosineDistance[w, z])^2)^(1/4), x >= 0]; der = Simplify[D[SDID, x], x >= 0]; euclid = (z - w).(z - w); der2 = Simplify[D[euclid, x], x >= 0]; Plot @@ {{SDID, der, 1000 euclid, 1000 der2, Sqrt[(Plus @@ Abs[e]) 100/2],   26 e [[1]], 26 e[[4]]}, {x, -.2, 1.2}, AspectRatio -> 1,   AxesLabel -> {"Seat shift", "SDID and errors"},   BaseStyle -> {FontSize -> 12}, AxesOrigin -> {0, 0},   PlotRange -> {{0, 1}, {-.25, 2.7}}}

The derivatives of SDID (curved yellow) and Euclid (straight red) show by the intersections with the horizontal axis that the minima of SDID and $\tilde{e}' \tilde{e}$ differ. The errors $\tilde{e}$ of party *A* and *D* at the bottom (brown and light blue) develop in opposite directions, and their intersection is also where their squared sum reaches a minimum.

The eye-catcher is the ALHIDT or Sqrt[100 Sum[Abs[$\tilde{e}$]] / 2] (purple). We see the errors cross the horizontal axis at 0.07 on the left and 0.7 on the right. Between these, ALHIDT is flat, and outside of these it starts to curve up. ALHIDT is flat because the errors compensate each other. Below, we will deduce that if *r* is a remainder for a party, then it gets a remainder seat with error value 1 – *r,* and it does not get a remainder seat with error value –*r.* For the parties the remainder values of *r* do not change here, since in this transfer *w* is constant.

- The brown line gives the error of *A*. It starts with the negative remainder for *A* at *x* = 0, thus below the horizontal axis at –*r,* when it does not get the seat. It enters at 0.07 when ALHIDT goes flat. It moves to 1 – *r* on the right at *x* = 1 when it gets the seat.
- The light blue line gives the error of *D.* It starts on the left at value 1 – *r* when it has the seat and moves to the right, when it hits the horizontal axis at 0.7 where ALHIDT becomes alive again, and ends at the negative value –*r* when it no longer has the seat.
- The remainders determine the errors and their intersection gives the minimum of EGID.
- Given that party *D* apparently has the highest remainder in this story, the allocation of the seat to *D* is appreciated by a lower value of EGID for *x* = 0 than for *x* = 1. Similarly for ALHID and SDID.

It is known that the use of ALHID or EGID in apportionment leads to the adoption of HLR. Given above fundamental relation, there indeed is some bias for Sin as well. Especially when the particular case is already quite proportional (*b* ≈ 1) then the bias will be stronger. That said, Sin also allows for the independent effect of the slope when Euclid is less sensitive to this. Thus its bias is muted.

A muted bias might actually be a problem when we want to enhance the use of HLR, or actually RLR (**Appendix L**). For EGID and HLR the party weight *w* has no role in weighing of the direct errors. This supports the notion of *equal (proportional) representation* of the *party marginal candidate.* If Sin would attach greater value to the weight, like WSL, then it would be less suitable to use in apportionment for HLR, and then perhaps neither in measuring inequality / disproportionality. We would be interested in cases that WSL and HLR have different apportionments, while the SDID of WSL-Ap is lower than the SDID of HLR. I have not been able to find such a case yet. However, for adding a new State, **Appendix N** contains a case in which SDID concurs with WSL and RLR, in opposition to EGID and HLR. The cause is the quota threshold in RLR, and not necessarily the attention for party weights as in WSL.

With bias, we mean that SDID would support HLR, and not quite that it rejects other approaches. But lack of support might be seen as rejection. With the existence of a bias now fairly clear (and desirable), the question causes an interest about "how much bias" there is. Below, we look at some graphs and formulas. It appears that I cannot arrive at a definite statement. The main insight may be the one by Balinski & Young (1980:4): "Third, empirical observation makes clear that the event of a Webster apportionment not satisfying quota is extremely unlikely."

8.3. An aspect of symmetry

If the minimum of EGID is to the left of 0.5 then its value at *x* = 0 will be lower on the left side. A minimum to the right of 0.5 will cause a lower value at *x* = 1. A minimum precisely at *x* = 0.5 will cause a deadlock. Where the minimum of EGID is, depends upon the remainders. If the remainders are equal, then EGID will have a deadlock. If they are unequal then one will be higher and gets EGID's preference. The minimum of SDID is slightly off, and may have a deadlock at another point. It is not clear whether SDID is always on the same side of EGID, or that it is always in the same range as EGID towards the allocation of preference (as in the present figure).

A method like HLR must switch its decision at only one particular point. Parties have experienced the frustration that only a few votes more would have made the difference of the remaining seat. However, we may accept that an integer tilt must be made somewhere. Yet it would be curious that a method of apportionment makes itself dependent upon the values of *x* = 1 or *x* = 0 only. Perhaps a better way to describe the situation is that there is a "range of favour" for *D* and a "range of favour" for *A,* and that the integer solution picks a point in that range. It is doubtful whether this view really helps. The remainders are determined by the floors, and by the *x* = 0 and *x* = 1 values. There is no need yet to rephrase this fundamental given.

8.4. Tie-breaking

There are some stalemates, where a tie-breaking rule must be used when WSL and HLR are neutral, and in which SDID sees a difference. **Table 12** reviews some cases, printing more digits to highlight the tie for some measures. I have not looked at the converse. There might also be cases in which SDID has a tie simply because of the integer seat values, and in which another measure necessarily makes a difference, which supposedly might be "tie-breaking".

Table 12. Stalemate and tie-breaking

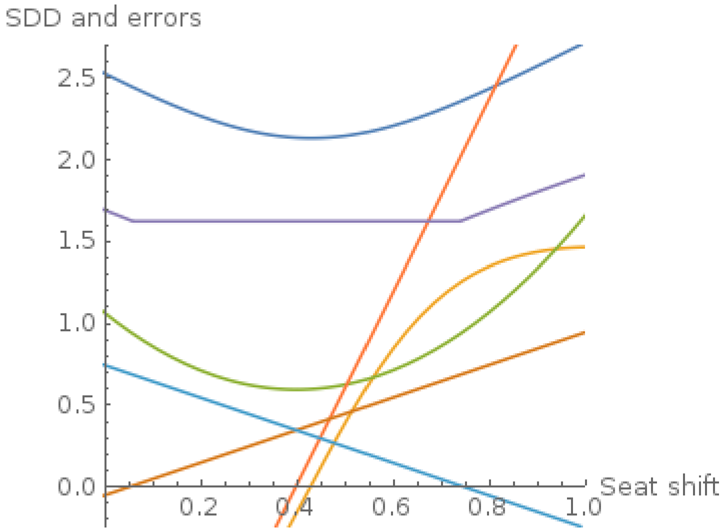
	1	2	3	4	5	6	7	8	9	10	11	12	13
			10vSin	**Specif**	**100vSin**	**100 Sin**	**Gini**	**LHI**	**EGD**	**WSL**	**LHIT**	**EGDT**	**WSLT**
Ex.	V	S	10	10	100	100	100	100	100	inf	10	10	10
TA	150	1	2.529	0.453	25.293	6.397	5.000	5.000	5.000	1.961	2.236	2.236	1.228
	850	9											
TB	150	2	2.650	0.497	26.504	7.025	5.000	5.000	5.000	1.961	2.236	2.236	1.228
	850	8											
TC	250	2	2.769	0.543	27.694	7.670	5.000	5.000	5.000	1.333	2.236	2.236	1.035
	750	8											
TD	250	3	2.882	0.588	28.818	8.305	5.000	5.000	5.000	1.333	2.236	2.236	1.035
	750	7											
TE	350	3	2.982	0.629	29.821	8.893	5.000	5.000	5.000	1.099	2.236	2.236	0.949
	650	7											
TF	350	4	3.065	0.665	30.647	9.392	5.000	5.000	5.000	1.099	2.236	2.236	0.949
	650	6											
TG	45	4	0.738	0.038	7.378	0.544	0.500	0.500	0.500	0.058	0.707	0.707	0.236
	955	96											
TH	45	5	0.741	0.039	7.414	0.550	0.500	0.500	0.500	0.058	0.707	0.707	0.236
	955	95											

8.5. Seat transfer when votes are given

SDID's preference for ZD does not mean that it has special behaviour for it. It treats other cases in the same manner. The only difference is that ZD receives a lower value.

To show this, **Figure 15** reproduces the same plot for the transfer of ZC to ZA. The plots are difficult to distinguish but the derivative of EGID clearly has different intersections at the horizontal axis. For the discussion below I took ZC (HH) merely because of my own temporary interest about HH, as it is not much discussed for parties. ZA has seats {10, 7, 5, 3, 1} and ZC has seats {9, 7, 6, 3, 1}, with the switch of 1 seat between parties *A* and *C,* so that we look at {9 + *x*, 7, 6 – *x*, 3, 1}. The figure shows the same phenomena as already discussed for ZA and ZD. The insensitivity of ALHID has the useful effect that it indicates the range of the errors that are interesting for us now, as remainders.

Figure 15. Transfer of 1 seat from case ZC to ZA, integer values at 0 and 1, errors, SDID (blue), Euclid (green), derivatives w.r.t. *v* (yellow, red), ALHIDT (purple) [30]



8.6. Vote transfer when seats are given: decomposition on w and r

The other way to look at this is inversely, by varying the *votes* around the seats. Thus now we consider *vote transfer* given seats. We assume that the seats have been given by some method of apportionment, and we want to judge how SDID evaluates this. Obviously, the apportionment would change when the votes changes. We try to apply the idea of partial derivatives to a world of integers. Thus, when an apportionment has selected the seats of ZA as optimal, the seats fixed, then we wonder what votes would justify it. And when ZC is optimal, the seats fixed, what votes would justify that. When there is no transfer then this gives the vertical axis.

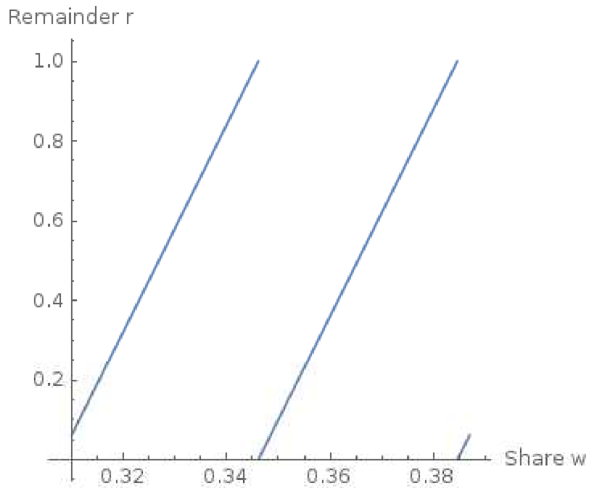
The key point is: when we keep *z* constant and we adjust *w* a little bit, then there will be a large relative effect on remainder *r*. The error is based upon (i) the remainder and (ii) a part outside of the locus of apportionment. Here we focus on *r* because of its relation to optimality, the local minimum of SDID. For the case of *A* considered here, we plot for a larger range, to keep some overview, but the focus of attention is on *r*.

With *Q* = 26000 / 26 = 1000, we can plot over plus or minus one *Q*. Then *w* will change about 1 / 26 or almost 4% points either way. The remainder can vary between 0 and 1. When both *z* and *w* are relatively constant, we are also assuming that *b* is relatively constant. **Figure 16** shows the effect of the share *w* on the remainder *r* of Party *A*.

Figure 17 compares ZA and ZC, but inversely by shifting their votes. A positive shift means that *A* got more votes at the cost of *C*. Thus v = {9061 + *x*, 7179, 5259 – *x*, 3319, 1182}. ZC is on the left with fixed seats {9, 7, 6, 3, 1} and ZA is on the right with fixed seats {10, 7, 5, 3, 1}.

[30] v = {9061, 7179, 5259, 3319, 1182}; s = {9 + x, 7, 6 - x, 3, 1}; w = v/26000; z = s/26; e = z - w; SDID = Simplify[10*(1 - (1 - CosineDistance[w, z])^2)^(1/4), x >= 0]; der = Simplify[D[SDID, x], x >= 0]; euclid = (z - w).(z - w); der2 = Simplify[D[euclid, x], x >= 0]; Plot @@ {{SDID, der, 1000 euclid, 1000 der2, Sqrt[(Plus @@ Abs[e]) 100/2], 26 e [[1]], 26 e[[3]]}, {x, -.2, 1.2}, AspectRatio -> 1, AxesLabel -> {"Seat shift", "SDID and errors"}, BaseStyle -> {FontSize -> 12}, AxesOrigin -> {0, 0}, PlotRange -> {{0, 1}, {-.25, 2.7}}}

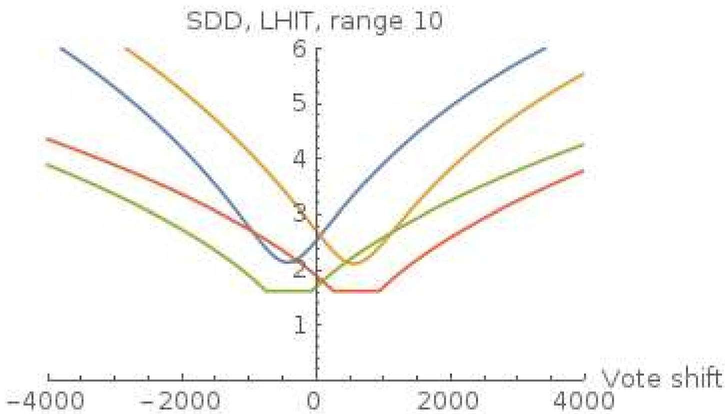
Figure 16. Effect of the vote shift on share w and remainder r of Party *A* [31]



Comments are:

- SDID-blue has a minimum at 9 seats for *A* and SDID-yellow has a minimum at 10 seats for *A,* as SDID takes these seats as optimal. The point of intersection happens to be close to the situation of no transfer (vertical axis). At the left of the intersection, blue is lower, and at the right of the intersection yellow is lower. If SDID would be used for apportionment it would make the switch at the point of intersection. At the vertical axis it prefers blue.
- ALHIDT (green and red) have about the same conclusions as SDID, but with a flat minimum.
- The current minima for blue and yellow are about the same height, but this changes when *A* takes votes from other parties and moves closer to proportionality (not shown).

Figure 17. Party *A* takes votes from Party *C* in cases ZA and ZC, [32]
SDID (blue *C* and yellow *A*), ALHIDT (green and red).
Left fixed seats {9, 7, 6, 3, 1} and right fixed seats {10, 7, 5, 3, 1}.

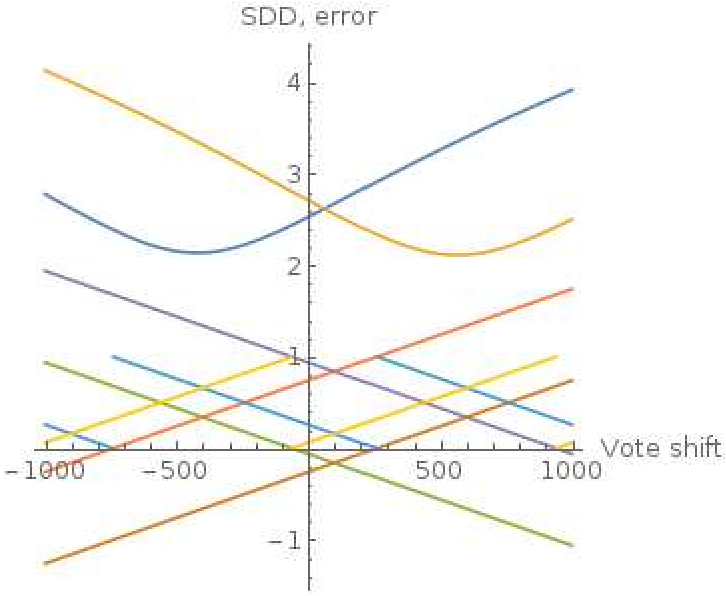


[31] Wolfram Alpha: ParametricPlot @@ {ran = 1000;   v = {9061 + x, 7179 , 5259 - x, 3319, 1182};   s = {9, 7, 6, 3, 1};   w = v / 26000; z = s / 26;   {First[w],   a = 26 w; f = Floor[a]; r = a - f; First[r] }, {x, -ran, ran}, AspectRatio -> 1,   AxesLabel -> {"Share w", "Remainder r"},   BaseStyle -> {FontSize -> 12}}

[32] Wolfram Alpha: Plot @@ {ran = 4000; v = {9061 + x, 7179, 5259-x, 3319, 1182};
s = {10, 7, 5, 3, 1};   s2 = {9, 7, 6, 3, 1};   w = v / 26000; z = s / 26; z2 = s2 / 26;
{10*(1 - (1 - CosineDistance[w, z2])^2)^(1/4),   10*(1 - (1 - CosineDistance[w, z])^2)^(1/4),
Sqrt[(Plus @@ Abs[w - z2]) 100/ 2],   Sqrt[(Plus @@ Abs[w - z]) 100/ 2]
},   {x, -ran, ran}, AxesOrigin -> {0, 0},   PlotRange -> {{-ran, ran}, {0, 6}},   AxesLabel -> {"Vote shift", "SDID, ALHIDT, range 10" },   BaseStyle -> {FontSize -> 12}}

Figure 18 looks at SDID and the errors and remainders. In this case there are two SDIDs, and thus four lines for errors (green and red intersecting on the left, and purple and brown intersecting on the right). We don't plot Euclid's distance, since we know that its minimum is at the intersection of rising and declining errors. Above, we saw the causal chain: *remainder → error → ẽ' ẽ → Sin*. The remainders are in [0, 1], and jumps from 0 to 1, with light blue for party *C* and light yellow for party *A.* The errors in the locus of apportionment are 1 – *r* for gainers and –*r* for losers. Light blue flips to red or brown. Light yellow flips to purple and green. At the vertical axis, no transfer, SDID prefers the blue minimum, and the light blue remainder (of *C*) is higher than the light yellow remainder (of *A*).

Figure 18. Decomposition of Sin, errors and remainders (blue for *C*, yellow for *A*) [33]



SDID-blue has a minimum when *A* loses 422 votes to *C,* and the remainders there are {0.639, 0.179, 0.680, 0.319, 0.182}. [34] The remainders intersect at x = -401 which differs from -422 where SDID-blue reaches its minimum. [35] The difference is explained by regression term *b.* (Also the intersection of ALHIDT (shown earlier) has a slightly different *x,* but it is a transform.)

How do the weight and the remainder (error) relate to the point of intersection and the locations of the minima ? For clarity:

- the remainders and errors have opposite slopes because the error is 1 – *r* (gaining a seat) or –*r* (not gaining a seat) in the locus of apportionment.
- the green line is the error of *A* for ZD (SDID on the left, blue), over the whole range
- the red line is the error of *C* for ZD
- the purple line is the error of *A* for ZA (SDID on the right, yellow), over the whole range
- the brown line is the error of *C* for ZA

---

[33] Plot @@ {ran = 1000; v = {9061 + x, 7179, 5259 - x, 3319, 1182}; s = {10, 7, 5, 3, 1}; s2 = {9, 7, 6, 3, 1}; w = v/26000; z = s/26; z2 = s2/26; e = z - w; e2 = z2 - w; a = 26 w; {10*(1 - (1 - CosineDistance[w, z2])^2)^(1/4), 10*(1 - (1 - CosineDistance[w, z])^2)^(1/4), a = 26 w; f = Floor[a]; r = a - f; 26 e2[[1]], 26 e2[[3]], 26 e[[1]], 26 e[[3]], r[[3]] , r[[1]]}, {x, -ran, ran}, AspectRatio -> 1, AxesLabel -> {"Vote shift", "SDID, error"}, BaseStyle -> {FontSize -> 12}}

[34] Wolfram Alpha: v = {9061 + x, 7179, 5259 - x, 3319, 1182}; s = {10, 7, 5, 3, 1}; s2 = {9, 7, 6, 3, 1}; w = v/26000; z = s/26; z2 = s2/26; a = 26 w; f = Floor[a]; r = a - f; sol = NMinimize @@ {10*(1 - (1 - CosineDistance[w, z2])^2)^(1/4), x}; r /. sol[[2]] // N

[35] v = {9061 + x, 7179, 5259 - x, 3319, 1182}; s = {10, 7, 5, 3, 1}; w = v/26000; a = 26 w; f = Floor[a]; r = a - f; NMinimize @@ {{(First[r] - r[[3]])^2, - 500 <= x <= - 400} , x }

- the light blue line is the remainder of *C* (fitting the best outcome of *C* on the left). It is in the [0, 1] range. The remainder now is not fixed but moves with the votes and errors.
- the light yellow line is the remainder of *A* (fitting the best outcome of *A* on the right). It is in the [0, 1] range. The remainder now is not fixed but moves with the votes and errors.
- When *A* takes votes from *C,* then the remainder of *A* (light yellow) tends to rise gradually, but with a sudden drop to 0 again. The rising remainder is reflected in the falling error (green and purple) (because the error is 1 – *r* or –*r*).
- SDID has no jumps since the remainder of *C* compensates the effect on *A*'s remainder.

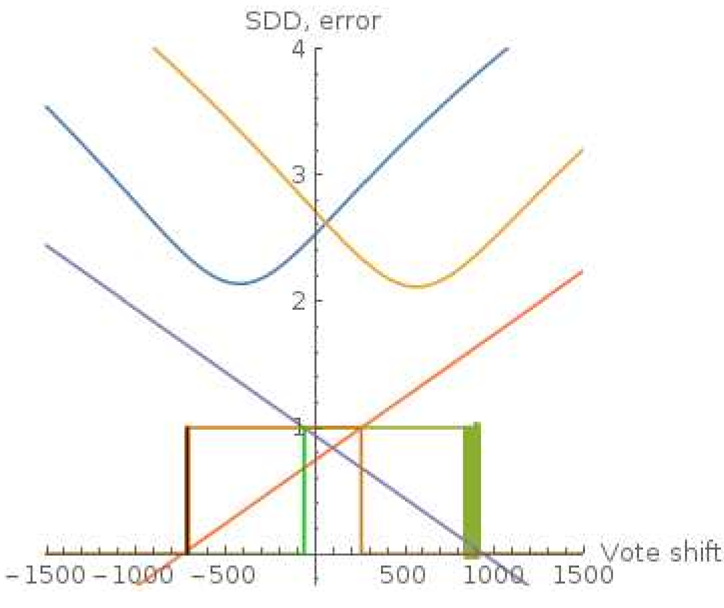
The errors decompose as follows:

(i) the error (say green for *A* and red for *C* on the left) consists of a remainder 1 – *r* or -*r* in the locus of apportionment and a part outside of the locus
(ii) minimum Euclidean distance is at the intersection of the relevant errors
(iii) the relevant intersection concerns the remainders,
(iv) Euclid and slope *b* give SDID.

The graphs are risky to interprete, since we have fixed *z.* Thus there is no generated effect of an apportionment algorithm but merely an assumed outcome. It is tempting to link the remainders to a shift from {9, 7, 6, 3, 1} to {10, 7, 5, 3, 1}, but these are actually fixed. The plot already indicates that the remainders move up and down over the whole horizontal axis, and all without dramatic effect, except for the minima in the locus of apportionment. The answer was given by **Figure 18**, For the overall behaviour the error outside of the locus is relevant, and for the minimum the remainders (errors) within the locus are relevant. The exogenous apportionment method determines whether 1 – *r* or -*r* is an error.

Figure 19 restates the error for *C* on the left (red) and the error for *A* on the right (purple). Thus we compare blue and red, and yellow and purple. Positive errors give *z* > *w,* giving scope for less seats, and negative errors give *z* < *w,* giving scope for more seats.

Figure 19. Locus of apportionment for ZA and ZD [36]



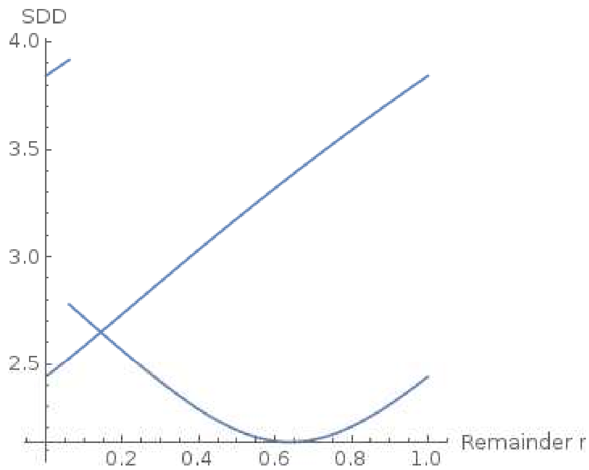
[36] Plot @@ {ran = 1500; v = {9061 + x, 7179, 5259 - x, 3319, 1182}; s = {10, 7, 5, 3, 1}; s2 = {9, 7, 6, 3, 1}; w = v/26000; z = s/26; z2 = s2/26; e = z - w; e2 = z2 - w; a = 26 w; trick = 0; {10*(1 - (1 - CosineDistance[w, z2])^2)^(1/4), 10*(1 - (1 - CosineDistance[w, z])^2)^(1/4), a = 26 w; f = Floor[a]; r = a - f; If[26 e[[1]] == 1 - r[[1]], If[trick == 0, trick = 1; x2 = x; 1, 1], 0], 26 e2[[3]], 26 e[[1]], If[26 e2[[3]] == 1 - r[[3]], x1 = x; 1, 0]}, {x, -ran, ran}, AspectRatio -> 1, AxesLabel -> {"Vote shift", "SDID, error"}, BaseStyle -> {FontSize -> 12}, PlotRange -> {{-ran, ran}, {-.25, 4.}}, Epilog -> Tooltip[{Line[{{x1, 0}, {x1, 1}}], Green, Line[{{x2, 0}, {x2, 1}}]}]}

There are two loci of apportionment, given with a 0 or 1 function, one for ZD (brown) on the left, and one for ZA (green) on the right. Brown must be judged for the seats {9, 7, 6, 3, 1} on the left and green must be judged for the seats {10, 7, 5, 3, 1} on the right. We best see the errors as entering from the negative area (not enough seats), passing through the loci, and then proceeding into the positive range (too many seats). In those loci a seat is allocated when the error is 1 – *r.* (We just restate that the error equals 1 – *r* by defining an If-statement that flips 0 and 1.) Red coming from the left enters the brown locus of apportionment for ZD at about -1000, awards it with a seat, and leaves at about 250. Purple, coming from the right, enters the green locus of apportionment for ZA at about 1000, awards it with a seat, and leaves at -100.This set-up is purely a matter of accounting. The votes will only seldomly hit exactly at integer values, and thus there normally is a remainder. For any seat above *f* it can be said to depend upon the 1 – *r* condition. How do we know that the brown and green loci really concern the one seat that we are transferring from ZD to ZA ? Well, once the errors are out of the [0, 1] range, then they can no longer be equal to some 1 – *r.*

Still, we are actually not relocating the seat, since the seats are fixed. The graph is basically an issue of accounting. The apportionment of seats is exogenous to our discussion. The seats have been given at either of the two locations. We only look at the partial influence of *w* on *r* and *ẽ,* and the definitory relationship between the latter for the locus of apportionment. We may still wonder about the phenomenon that the brown and green loci overlap, and how this relates to the intersection of the two SDID curves. The inference is that this relates to the other parties, whose remainders do not change, but may be higher than the low values of *A* and *C* in this overlap.

Figure 20 gives a parametric plot that cuts up a curve. The parametric plot cannot distinguish the two different loci of the remainders, and treats them as one only. The remainder of *A* covers both the area around the minimum and a leg at a distance. The minimum of SDID depends upon EGID and the slope *b.*

Figure 20. Effect of the vote shift on remainder r and SDID [37]



8.7. Within the locus of apportionment

The formulas below do not resolve our issue on "how much bias" there is, but they might still enlighten it. The following applies to any apportionment within its locus. The errors then depend upon the remainders, whatever the method. Or the remainders can be read as errors.

[37] Wolfram Alpha: ParametricPlot @@ {(ran = 1000; v = {9061 + x, 7179 , 5259- x, 3319, 1182}; s = {9, 7, 6, 3, 1}; w = v / 26000; z = s / 26; {a = 26 w; f = Floor[a]; r = a - f; First[r], 10*(1 - (1 - CosineDistance[w, z])^2)^(1/4) }), {x, -ran, ran}, AspectRatio -> 1, AxesLabel -> {"Remainder r", "SDID"}, BaseStyle -> {FontSize -> 12}}

The apportionment method decides whether the error is important or not. A following subsection looks outside of the locus.
Table 2 gives the notation. We have $s = T v + \tilde{u} = S w + \tilde{u} = a + \tilde{u}$ so that $1' \tilde{u} = 0$. The relation $s = a + \tilde{u}$ in levels gives $z = w + \tilde{e}$, with $\tilde{e} = \tilde{u} / S$ and $a = S w$ or $w = a / S$. Thus the $\tilde{e}$ are small numbers: differences divided by S. For Holland a difference of 1 seat causes the value 0.67% = 0.0067. We now enter into a special application of these errors. We focus on the formula for Cos.

Let Floor[a] ≤ s ≤ Ceiling[a] be the locus. Let f = Floor[a] and $r = a - f$ the remainder. Then $1'a = S = 1'f + 1'r = F + Rm$, with Rm the remainder seats after allocation of F. Perhaps the method of apportionment does not work like this, but we still can do this accounting to specify the error.

Let A be an apportionment matrix, with zeros everywhere except some 1s on the diagonal for parties that get a seat above f. When a method supplies s, then A = diag[$s - f$]. Then $1'A\, 1 = Rm$. Again, a, f and r are given from v and S.

The gains Γ = Max[0, $s - a$] for the electoral lucky can also be written as $\Gamma = A\, (1 - r)$ so that $1'\, \Gamma = Rm - 1'\, A\, r$ where $1'A\, r$ is the sum of the remainders of the gaining parties. With $Rm = 1'r$ we also have $1'\, \Gamma = (1' - 1'A)\, r = 1'\, (\mathrm{I} - A)\, r$.

The losses Λ = Max[0, $a - s$] $= r - A\, r = (\mathrm{I} - A)\, r$ are for those kept to the Floor. We have $1'\, \Lambda = 1'\, (\mathrm{I} - A)\, r = Rm - 1'A\, r = 1'\, \Gamma$. Thus $1'\, \Lambda = 1'\, \Gamma$. Politically, the loss remainders are used to supplement the gain remainders, and losing parties might not enjoy that their votes are used in such manner. The HLR use of maximal r is also the use of minimal $1 - r$, and thus minimal abuse of the losers in this particular sense. For now, we only look at the error.

The key decomposition thus is $\tilde{u} = \Gamma - \Lambda$. The sum of errors $\tilde{u}'\, \tilde{u}$ will be $\Gamma'\, \Gamma + \Lambda'\, \Lambda$. There is also Sum[Abs[$s - a$]] = Sum[$\Gamma + \Lambda$]. For arbitrary method and s, $\tilde{u}'\, \tilde{u}$ need not be minimal. **Table 13** summarizes.

Table 13. Errors for *s* from the optimal *a*, selecting from the remainders

$r = a - f$	*Apportionment*	$\tilde{u} = s - a = s - f - r$	$Rm = 1'r$, $1'\, \Gamma = 1'\, \Lambda$
Given a seat	$s - f = 1$	$\tilde{u} = 1 - r$	Γ = If seat $1 - r$, else 0
Given no seat	$s - f = 0$	$\tilde{u} = - r$	Λ = If 0 in Γ then r, else 0

Thus we have $\tilde{u} = \Gamma - \Lambda = A\, (1 - r) - (\mathrm{I} - A)\, r = A\, 1 - A\, r - r + A\, r = A\, 1 - r$. Also $A\, 1 = s - f$, so that $\tilde{u} = s - a$ again. The relation only shows the switch between error $-r$ and $1 - r$.

For completeness: $\tilde{u}'\, \tilde{u} = \Gamma'\, \Gamma + \Lambda'\, \Lambda = (1 - r)'\, A'\, A\, (1 - r) + r'\, (\mathrm{I} - A)'\, (\mathrm{I} - A)\, r =$
$1'\, A\, 1 - 2\; 1'A\, r + r'\, A\, r + r'\, r - r'\, A\, r = Rm - 2\; 1'\, A\, r + r'\, r$. In itself, $Rm = 1'r$ too, and thus $Rm - 2\; 1'\, A\, r = (1 - 2\; 1'A)\, r$ can be see as a signed sum of r. Thus $\tilde{u}'\, \tilde{u} = r'\, r$ + Sum[r, with negative signs for the gainers and positive signs for the losers].

The relevant decomposition is $s = a + \tilde{u} = S\, w + \tilde{u}$ so that $z = w + \tilde{u} / S = w + (\Gamma - \Lambda) / S$. We had $\tilde{e}'w / w'w = b - 1$, thus also $w'\, (\Gamma - \Lambda) / S / w'w = b - 1$. There are various ways to get at Cos. Assuming some apportionment A or s, and now only decomposing the numerator:

$$\mathrm{Cos} = w'z / \mathrm{Sqrt}[w'w\;\; z'z]$$

$$\mathrm{Cos} = (w'w + w'\, (\Gamma - \Lambda) / S) / \mathrm{Sqrt}[w'w\;\; z'z] \quad (*)$$

Thus, when a method generates an apportionment A within the locus above f = Floor[a], with resulting z, then the cosine obviously directly follows from $w'z$, but this can also be seen as depending upon the error of apportionment $\tilde{u}$, that depends upon the remainders r. (We already had the relation between Sin and $\tilde{e}$, and the present focus is on the remainders.)

PM 1. **Section** 11.2 explains how HLI and EGID support HLR. It is relevant enough to copy the statement. With level error $\tilde{u} = s - a$, minimising $\tilde{u}' \tilde{u}$ effectively means ordering the $r = a - f$ values (on size), and allocate the remainder seats to parties that would otherwise cause most of the error. Let us locally reuse symbols f and c for a single party only. Shifting from Floor f to Ceiling $c = f + 1$ means that the current $\tilde{u}' \tilde{u}$ is reduced by $(a - f)^2$ and rises by $(c - a)^2$. The total change is $(c - a)^2 - (a - f)^2 = 2 (c - a) - 1 = 1 - 2 (a - f)$. Using vectors a and f again: taking the highest values of $a - f$ causes the least increase of the $\tilde{u}' \tilde{u}$.

PM 2. **Section** 7.4 gave us already: $\mathrm{Sin}^2 = (\tilde{e}' \tilde{e} / z'z - h) / (1 - h)$ for $h = (1 - b)^2 / b^2$, which now gives, though with influence from b:

$$\mathrm{Sin}^2 = (\tilde{u}' \tilde{u} / S^2 / z'z - h) / (1 - h) = ((\Gamma ' \Gamma + \Lambda ' \Lambda) / S^2 / z'z - h) / (1 - h) \quad (*2)$$

If we assume that z is given and when we keep b constant, so that all changes come from w (without effect on b), then all change in Sin comes from $\tilde{u}' \tilde{u}$. We can use above decomposition. Let the old value be $r'r + r1 - r2 + ... r\, n$, so that $r2$ is a gainer. Let the new value be $r'r - r1 + r2 + ... + r\, n$. Consider $\mathrm{Sin}^2[\text{new}] - \mathrm{Sin}^2[\text{old}]$. If we maintain the same $r'r$ too, then the difference in the numerator is only $(r1 - r2) - (-r1 + r2) = 2 (r1 - r2)$. Unfortunately, we cannot quite maintain the same $r' r$, since this depends upon the changing w. Above graphs actually have focused on changing r. The reasoning at least shows the paths of changes. (Partial derivatives in an integer setting.) This still assumes that it is a good idea, but it might not be, to fix the value of z and look at the issue in reversed manner, to see the effect of w and the r that it causes.

8.8. General error

We can split the general error $\tilde{u} = s - a$ into the part that pertains to the locus and the part outside of it. Let $s = s1 + g$. Thus $\tilde{u} = s - a = f + A\, 1 + g - a = A\, 1 - r + g$, for $0 \le A \le 1$ and $A = \mathrm{diag}[s1 - f]$, and $g = s - s1$. The latter will redistribute seats, and then have positive and negative integers, while $1'g = 0$ again.

$\Gamma 2 = A\, (1 - r) + \mathrm{Max}[0, g]$,
$\Lambda 2 = (\mathrm{I} - A)\, r + \mathrm{Max}[0, - g]$,
so that $\tilde{u} = \Gamma 2 - \Lambda 2 = \Gamma - \Lambda + g$.

With $z = w + (\Gamma - \Lambda + g) / S$, and again only decomposing the numerator:

$$\mathrm{Cos} = (w'w + w' (\Gamma - \Lambda + g) / S) / \mathrm{Sqrt}[w'w \;\; z'z] \quad (**)$$

The latter can be decomposed with different denominators D:

$$\mathrm{Cos} = (\mathrm{Cos}[*] + w' g / S / D[*]) \;\; D[*] / D[**] \qquad (***)$$

in which $D[*] = \mathrm{Sqrt}[w'w \;\; z'z]$ of (*), using $z = (f + A\, 1) / S = s1 / S$
and $D[**] = \mathrm{Sqrt}[w'w \;\; z'z]$ of (**) using $z = (f + A\, 1 + g) / S = (s1 + g) / S$.

8.9. Behaviour of numerator and denominator of Cos

At issue is whether the equations (*) and (**) or (***) can tell us something about a possible bias of Cos w.r.t. the apportionment method of HLR. [38]

The errors can be looked at from the angle of their relation to the remainders. A can still be any method. The apportionment of s in the locus need not be proportional. It may be that Cos has its own preference. Other apportionment methods may generate other errors, but, when these are in the locus of assignment then Cos might evaluate the remainders on their own merits, in favour of HLR, rather than following the choice of A. Still, it is A that determines the

[38] RLR (**Appendix L)** imposes an explicit threshold of at least 0.5 Q for the allocation of remainders but in practice HLR may also show this property merely because of the random distribution of remainders. Yet we should be cautious with general statistical phenomena.

switches between $-r$ and $1 - r$, and Cos basically evaluates only with w. Yet in this analysis with vote transfer, w can also affect the sizes of r.

As we have deduced that the errors in the locus are $1 - r$ and $-r$, and the the sum of squared errors between two parties reaches its minimum at the intersection, then this also holds for the remainders, as **Figure 18** also shows. The intersection of the remainders that is closer to the vertical axis will determine which EGID will be lower. SDID follows that, except for the slope. (We should have selected an example like France with a slope quite differently from 1.)

(1) A first observation is that (*) is not fully replaced in (**) or (***), but still is present. Thus there is the memory of the apportionment around the optimal a. When g is not large, the effect of (*) would tend to be important. Still, g might relocate to a widely disproportional situation, and this memory might be dim. Thus we would expect a diminishing bias w.r.t. the locus of apportionment around $a = S\, w$.

(2) We have not graphed effects of g (other than extending the error beyond the locus of apportionment). It stands to reason that a z with a large g would generate convex cups in the higher SDID values. If we fix z then also g is fixed, and the only variation in $w'g$ in the numerator of (***) comes from w, so that higher values of w might emphasize positive weights of g with less cost on the negative weights of g, and then the numerator will generate a higher value of Cos. Thus disproportionalities will be less severe on larger parties. But we cannot look only at the numerator here, because w has also an effect in the denominator. The denominator has no combination of g and w though.

(3) For the remainders, it matters whether we assume that w and r are fixed, and there is only variation from z, or whether we assume that z is fixed and the variation comes from w, its r and the implied decision of z on this r.

(4) If we return to flexible z, then we assume for (*) that w and r are fixed and that numerator and denominator depend upon A = diag[$s - f$], with values 1 or 0 that select the relevant $-r$ or $1 - r$. Let us consider the fractional parts. [39]

For the numerator:

- These jumps in the numerator look quite high, namely with a difference of $(1 - r) - (-r) = 1$. For $r = 0.2$, the value of exclusion of -0.2 compares to the gain 0.8 if included. For $r = 0.8$, a value of exclusion of -0.8 compares to 0.2 if included (for, we still allow any apportionment method). Paradoxially, such a switch in the numerator does not matter, for when one party is excluded and causes a change of -1, another is included and causes a change of $+1$.
- The $\tilde{u} = \Gamma - \Lambda$ in the numerator essentially are like switches. If a method of apportionment prefers party 1 over party 2, then this meets with an increase of the numerator of Cos iff $w1\, (1 - r1) + w2\, (-r2) \geq w1\, (-r1) + w2\, (1 - r2)$ or iff $w1 \geq w2$. Thus, for this switching effect, the numerator of Cos favours large parties rather than large remainders. (In (*2) the error in Sin however does not indicate this. The effect may run through b. Derivatives of Cos and Sin are related, and we would explain the one by the other ...)
- Given the above, there is the same property as in (2): When a higher value of $1 - r$ associates with a higher value of w, then the numerator would raise Cos. An apportionment method that assigns such $1 - r$ is favoured only when the weight of that party is higher indeed. That is, for the numerator.
- Still, SDID favours case ZD rather than ZA, while the largest party gets most seats in ZA. Thus this numerator-effect must have a correction in the denominator.
- The number of remainder seats has no role in the numerator either. If it is zero, then the relevant effect is that $r = 0$ and not that the number of remainder seats is zero.

For the denominator:

[39] Squares and Sqrt and division by S would be neutral here

- The complicating factor is that the denominator is also affected by a switch between $-r$ and $1 - r$. This involves the comparison of 1 / Sqrt[$w'w \; z'z$] when a switch between $r1$ and $1 - r1$ finds an opposing switch from another $r2$. In the above, we decomposed the numerator, but this appears to be less relevant, and we should have looked at $z'z$.
- Thus $z'z = (w + (\Gamma - \Lambda) / S)' (w + (\Gamma - \Lambda) / S)$
$= w'w + 2\, w' (\Gamma - \Lambda) / S + (\Gamma ' \Gamma + \Lambda ' \Lambda) / S^2$.
The first and middle relate to the numerator, and for the last part we had $(1 - 2\, 1' A)\, r + r' r = r' r$ + Sum[r, with negative signs for gainers and positive signs for losers].
- It might happen that a high value of $1 - r$ generates a lower 1 / denominator value, or that a high value of r causes a low value of $(1 - r)$^2 but then a high value of 1 / denominator again. And a low value of r causes a low value of $(-r)$^2 and thus a high value of 1 / denominator again. It depends upon the numbers how this balances out.
- We might also neutralise the effect of the denominator by fixing a value of z, and look at the issue in reversed manner. We did so in the graphs, and they were enlightening on w and r. Fixing z does not wholly destroy the relation between r and z, since there still is w.

For both: Heuristically, it may help to have the question in the back of one's mind whether an apportionment method prefers $r \geq 0.5$ above $r < 0.5$. However, the minima of EGID are when the remainders are equal. There are other parties present, so this need not be 0.5.

At this point we have little advancement over **Table 11**. The method of HLR uses the largest remainders and finds a solution, that SDID tends to accept as minimal. There are more examples like that, but no general rule yet except for the relation between Sin and the Euclidean distance, that is already quite revealing. It is not likely that there can be a counterexample that HLR supports an apportionment that SDID rejects, given the core of EGID in Sin. The formulas in the subsections above clarify the relationship of the errors to the remainders. They might be used to develop such a counterexample, or find a class of cases, or prove that it doesn't exist. For the purposes of this paper, it suffices to leave the matter here.

9. Definition of the "sine – target-slope deviation" (STSD) measure (on top of sine)

9.1. When the sine is insufficiently sensitive to a target slope

The use of the sine is not explicitly in the textbooks but follows from them straightforwardly, once they have mentioned the cosine of two vectors.

From the double face of the cosine or sine, we now select the perspective that the sine would only indicate inequality / disproportionality in general, and not the specific value for some target slope. The discussion is fairly long and has been put in **Appendix J**, so that we only look at the main finding now.

For a slope specific error (SPE) we might choose between slopes or angles. Still, heuristically, if we would adopt a slope measure as SPE = $(T - b)^2 + (T - 1 / p)^2$, then a major conceptual question is how to compare this SPE on the [0, infinity) range with the non-specific inequality / disproportionality measure Sin on the [0, 1] range. Obviously we can transform to [0, 1] and then the solution remains the multiplicative failure rate model in **Section** 4.4. It appears that angles are better, since the sine can also handle differences in angles. Thus we can formulate a slope target but a routine will best transform to angles and then to sine again.

Table 14 then defines the "sine – target-slope deviation" (STSD). The transformations used here indicate that STSD would be a metric too. This has not been checked sufficiently thoroughly though. The core measure in this paper remains SDID.

Table 14. Definition of the "sine – target-slope deviation" measure (standard f = 2)

η = Sin[θ] = Sin[v, s]	
η' = NonspecificID[v, s, f]	*sign* (100 η) $^{1/f}$ transform with sensitivity $f \geq 1$, standard f = 2, include the sign for majority switches
T	target slope (for votes and seats: T = 1)
Φ	ArcTan[T], is π / 4 when T = 1
range[Φ]	2 Max[Sin[Φ]^2, Cos[Φ]^2], is 1 when T = 1
ζ = crit[φ, ψ ; Φ] (SPE)	Sqrt[(Sin[$\Phi - \varphi$]^2 + Sin[$\Phi - \psi$]^2) / range[Φ]]
ζ ' = SpecificID[v, s, T, f]	*sign* (100 ζ) $^{1/f}$
d	1 – (1 – η) (1 – ζ) (without *sign* * 10 and f)
TotalID[v, s, T, f]	*sign* (100 d) $^{1/f}$
STSD[v, s]	TotalID[v, s], T = 1, f = 2

9.2. Test application of STSD for Holland 2017

Table 15 gives a test application on Holland 2017. It adds a slope-specific target onto SDID, giving a total value of STSD. The reasoning provides an example how we could handle target aspects of inequality / disproportionality. For a target we rather think about an issue like turnout, but for a test we can use the slope again. If we compare the outcome with **Table 7** then we see that inequality / disproportionality rises by 0.5 points on a scale of 10. Perhaps we have an indication that SDID might perhaps not be as sensitive to slope as we might wish. And perhaps this present test on STSD is an overstatement, since the Dutch slope seems to be fine in itself.

Observe that η and ζ define all other results, vertically by addition and subtraction, and horizontally by the transform. The cross term on the right is a remainder from the transformations in the other rows, as D' - η' - ζ'.

Table 15. Specific or Nonspecific, versus Original or Transform (STSD)

False data	*Original*		*Transform*	
Measures	Symbol	*sign* 100 Sin	Symbol	*sign* 10 √Sin
Range		1 - 100		1 - 10
Nonspecific	η	6.2303	η'	2.496
Specific	ζ	2.0993	ζ'	1.449
Cross term	$-\eta\,\zeta$	-0.1308	$D' - \eta' - \zeta'$	-1.082
Total	$D = \eta + \zeta - \eta\,\zeta$	8.1988	D'	2.863

In the following reasoning, the slope-specific target is added to the SDID value:

(1) Sine-Target-Slope inequality / disproportionality in Holland is 8.2% on a scale of 100 using f = 1, and 2.9 on a scale of 10 using f = 2, or 28.63% by comparison. Thus originally

Holland looks finely proportional, but more focus on the slope and the Weber-Fechner correction help us to focus on what needs to be improved in Holland.
(2) We can look at the situation from the angle of specific or nonspecific causes, but it is useful to be aware of the cross term so that we do not simply add the scores. This already happens in the original additive scores, and is magnified in the transform.
(3) The Dutch slope is quite close to $b = 1$, and the nonspecific causes dominate. Yet, due to the transform, we see that we should not neglect issues about the slope. The issue rises from 2 on a scale of 100 to 1.4 on a scale of 10, or 14.5 by comparison. Yet, 1 point is also overlap.
(4) If we start from nonspecific causes with score 2.5 on a scale of 10, and then include the specific causes, to arrive at a total of 2.9 on a scale of 10, then it seems as if the slope is fine, and adds little to the inequality / disproportionality. But if we start from the specific cause with 1.4 on a scale of 10, and then include nonspecific aspects, to arrive at the total of 2.9, then the two categories seem equally important. The transform helps to put matters in perspective, in Weber-Fechner corrected manner.
(5) Yet, all of this is only a test application, and not the major suggestion on SDID itself.

9.3. Choice in STSD of angles rather than slopes

Formulating a symmetric measure directly on the SPE = $(T - b)^2 + (T - 1 / p)^2$ appears to be feasible but relatively complicated.

- Symmetry is defined *around T.* Symmetry does not mean that we should necessarily also consider $1 / T$ as a target value. We can allow for the value $T = 0$. But symmetry *around T* causes us nevertheless to also flip the axes and consider also $1 / T - 1 / b$. This runs into problems when we allow $T = 0$, $b = 0$ or $p = 0$. This can be resolved by normalising on the unit circle however.
- We still must relate the range [0, infinity) for T and [0, 1] for Sin. For this, we might use weights and calibrate for say Holland.
- The issues are discussed in **Appendix G**. Overall, there seems to be too much arbitrariness in this. The transform √Sin is arbitrary too, but it nicely works out for the scale of 100 versus 10, and there is something to be said for nice things.

Instead, we have $\mathrm{ArcTan}[1 / x] = \pi / 2 - \mathrm{ArcTan}[x]$. Thus, we have immediate symmetry for the angles, for example:

$$(\mathrm{ArcTan}[T] - \mathrm{ArcTan}[b])^2 = (\Phi - \varphi)^2 = (\mathrm{ArcTan}[1 / T] - \mathrm{ArcTan}[1 / b])^2$$

Subsequently, we can shift $(\Phi - \varphi)^2$ directly to the [0, 1] range by using $\mathrm{Sin}[\Phi - \varphi]^2$. This provides a direct link between η and ζ. The choice for angles seems rather logical.

10. Heuristic of the measure: error around the slope of 1

The cosine has the double face of both a general measure of similarity and an interpretation of a slope. Originally this analysis started with the latter interpretation, and used the term "philosophy of the measure". However, with the other interpretation, this appears to be only a heuristic. The same findings remain.

10.1. Symmetry and normalisation of the input: key step in the definition

Taagepera & Grofman (2003: 665-666) correctly require both symmetry and insensitivity to levels of V and S, for the present narrow scope of the intended application. Some first comments are:

- A regression of s given v would generate S / V when the slope $z / w = 1$. Thus a routine that has s and v as inputs should first normalise to z and w.
- Symmetry can be achieved by regressing both z given w (naturally) but also w given z.

- A regression $z = b\, w$ can be regarded as a weighed regression $z / w = b$ 1 with weights w or a weighed regression of $1 = b\, w / z$ with weights z, leaving out the 0 in z. Similarly for a regression $w = p\, z$. Thus there is ample scope for symmetry and there is a wealth of slopes to choose from. A potential development on *averages* that appears to be a digression is discussed in **Appendix E**.
- WSL with Sum[$w\, (z / w - 1)^2$] = Sum[$(z - w)^2 / w$] originated as the first expression, i.e. as a (natural) ratio measure (weighted by the votes). It is said that WSL favours the small parties because of the w in the denominator, but WSL actually attaches more weight to ratios of the larger parties, perhaps at the cost of the smaller parties. Attaching value to the small parties would rather lead to Sum[$(1 / w)\, (z / w - 1)^2$].
- Symmetry would require that WSL with z / w is balanced by w / z, which however conflicts with our objective that the wasted vote are included in the measure. Having a party with 0 seats causes that w / z becomes undefined. It is conceivable to eliminate the 0 though.
- Lovell et at. (2015), Erb & Notredame (2016) and Quinn (2017) propose a statistic based upon the "log-ratio variance" or var[Log[y / x]]. They acknowledge that this is not symmetric and doesn't work for a 0 value. Obviously, Log[y / x] ≈ ($y / x - 1$) and the use of the variance introduces a square, so that the "log-ratio variance" is similar to the WSL. This links up to other approaches for the unit simplex, see **Appendix B**.

10.2. Variance ratio and the F-test

We may regard (unit) (diagonal) proportionality as a condition on the coefficients and then apply the F-test. See **Table 2** for the choice of symbols.

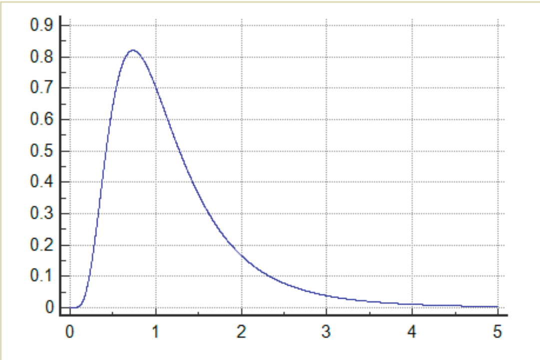
- OLS gives $z = \gamma + \beta\, w + \hat{e}$, with sum of squared errors SSE[γ, β] = $\hat{e}'\, \hat{e} = n\, szz\, (1 - R^2)$ and $n - 2$ degrees of freedom, and szz the variance of z.
- The H0 hypothesis has $z = 0 + 1\, w + \tilde{e}$, with SSE[0, 1] = $\tilde{e}'\, \tilde{e}$ = Sum[$(z - w)^2$] and n degrees of freedom since no estimates are used here. Voting researchers will recognise a transform of Euclidean distance / Gallagher index (EGID).
- The F-ratio test-statistic is $F = F[w, z]$ = (SSE[0, 1] / n) / (SSE[γ, β) / ($n - 2$)), for testing that the true situation is unit proportionality. We reject H0 when the ratio is too large according to some loss function. [40]
- The reasoning is that SSE[γ, β] as OLS estimator already has minimal SSE, so that other values of the coefficients can only raise the ratio. If estimation renders $\gamma = 0$ and $\beta = 1$, then the SSE are equal and $F = (n - 2) / n$.
- In Holland, the elections of 2017 for the House of Commons gave seats to 13 parties, so with the wasted vote lumped into one category, $n = 14$. With unit proportionality $F = 12 / 14 = 0.86$. **Figure 21** [41] gives the density and cumulative F-distribution with degrees of freedom $n = 14$ for the numerator (H0) and $m = 12$ for the denominator (H1).

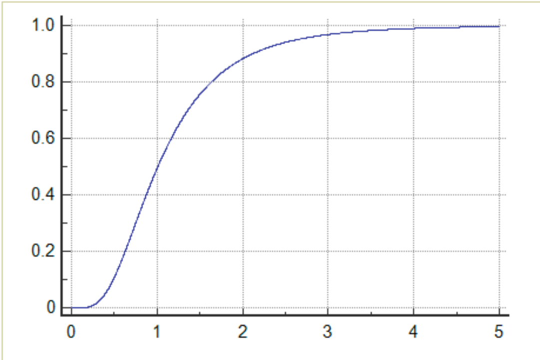
It is not very likely that one can say that the Dutch House of Commons started with the electoral result w and then allocated the seats z under the conditions that satisfy the assumptions of the F-ratio, which involves a degree of normal randomness. Thus the notion of testing is less relevant. But we can consider using F as an inequality / disproportionality index, since the present conceptual setting is sound, F works into the proper direction, and the probability is constrained to [0, 1]. A low value indicates closeness to unit proportionality, a high value indicates disproportionality.

[40] (1) The following exposition is quite nice, though it uses a difference in the numerator: http://www3.wabash.edu/econometrics/EconometricsBook/chap17.htm. (2) Johnston (1972:28) has an explicit expression for entering hypotheses on γ and β, but it is simpler to work directly with the SSE in above manner.

[41] https://www.medcalc.org/manual/f-distribution_functions.php

Figure 21. F-ratio density and distribution for df {14, 12}





Some considerations are:

- Rather than F we would want to use the [0, 1] range, provided by the Fcdf.
- The part of the Fcdf below proportionality with $F = (n – 2) / n$ is not used, and it would be impossible to reach 0 disproportionality. Thus we take a truncated distribution: ID[*x, n*] = Fcdf[*x, n, n* - 2] / (1 – Fcdf[(*n* – 2) / *n*, *n, n* - 2]) for *x* = *F*[*w,* z].
- The Fcdf has the stark slope like the proposed SDID in **Figure 6**. (But when two cars have the same colour, they don't have to be the same brand.)
- The Taagepera & Grofman condition of symmetry however isn't met.
- The Taagepera & Grofman condition that measures of inequality / disproportionality should not be burdened with variables like the number of parties *n*, gains in value. The condition is somewhat superfluous since we use vectors *v* and *s* that clearly have lengths. Yet the dealing with the relevant degrees of freedom per case is something that we rather avoid.

The above gives a clear conceptual setting but also motivates to look further.

10.3. (Dis-) Advantages of Pearson correlation

Inequality / disproportionality might also mean dispersion, as opposed to concentration, but proportionality rather links up with the notion of a *linear* relationship, which is precisely what the Pearson correlation coefficient focuses on.

- Thus $1 – R^2$ (Pearson) would be an interesting point to start with. [42]
- The newer methods of DistanceCorrelation, [43] EnergyDistance [44] and the distance matrix [45] have been designed to discover dependence where Pearson fails, yet Pearson remains interesting precisely because of the notion of proportionality that leads to linearity.

The Pearson correlation however fails when some vector has no dispersion, which would be a common event in voting. A vote of {50000, 50000} and a seat allocation of {50, 50} would be perfectly unit proportional but Pearson correlation generates infinities. The reason for all of this is that Pearson looks at dispersion around the mean, which need not be relevant for (unit) (diagonal) disproportionality.

Concordance correlation (see the NCSS (undated) documentation) [46] has –1 for discordance, 0 for no concordance and 1 for full concordance. It generates 0 for what would be perfect unit proportionality as votes {50000, 50000} and seats {50, 50}, and generates infinities if we enter the data as percentages, for votes {50, 50} and seats {50, 50}.

[42] https://en.wikipedia.org/wiki/Correlation_and_dependence
[43] https://en.wikipedia.org/wiki/Distance_correlation
[44] https://en.wikipedia.org/wiki/Energy_distance
[45] https://en.wikipedia.org/wiki/Distance_matrix
[46] Also https://en.wikipedia.org/wiki/Concordance_correlation_coefficient

Correlation however highlights that votes {49, 51} and seats {51, 49} are opposite. The ALHID / EGID measures give a value of 2, which one tends to regard as low, and which does not convey that there is such a dramatic reversal of majorities. We also get a ALHID / EGID of 2 for votes {35, 34, 31} and seats {37, 34, 29} which looks a fair situation, given that it is a zero sum game. Only outcomes like {40,60} vs {60, 40} give a high ALHID / EGID value of 20.

In other words, we have:

(i) the notion that vector-proportionality focuses on linearity through the origin
(ii) the notion that voting focuses on unit or diagonal proportionality
(iii) the fit of shares, in which {40, 60} and {60, 40} fit fairly well (Kullback-Leibler)
(iv) the implications for majority, in which the latter are quite disproportional.

It is an important insight: the sign of the correlation is a step towards checking whether an allocation reverses a majority. This forms a step towards handling power preservation more accurately. While Correlation collapses for equal outcomes like {50, 50} since correlation requires dispersion in the variables, we can use the covariance or only its sign.

10.4. Regression through the origin: key step in the definition

With these potential sources of confusion out of the way, let us now look closer at what might be done with regression. Just to be sure: we are not *explaining* z by means of w, but we are merely *measuring* a degree of disproportionality, e.g. for cross-country comparisons. Statistically, it may be ill-advised to regress through the origin (RTO), since errors in modeling may show up in the constant which can be tested. But we are not testing here but measuring. For our purposes RTO is mandatory.

If we have variables x and y then we say that these are proportional when $y = b\ x$, thus as a ray through the origin without a constant. Taagepera & Grofman (2003:663) mention that Cox & Shugart in 1991 already proposed to use b though for $y = c + b\ x$.

> "An advantage of [coefficient *b*] over other indices of disproportionality is that it indicates the directionality of the imbalance."

We want to get rid of the constant. [47]

This gives us the OLS estimator of "regression through the origin" (RTO) $b = x'\,y / x'\,x$. See Eisenhauer (2003) for derivation and some comments, notably on calculating the R^2 and dangers of statistical packages. Eisenhauer does not explain that $R = Cos$ though.

A key point to be aware of is that the sum of errors in $y = b\ x + e$ will not necessarily be zero. Thus $ymean = b\ xmean + emean$, so that on average there is an effect that compares to a constant, while there is no explicit constant in the equation.

Eisenhauer shows that a proper R^2 for RTO has $R^2 = b^2\ x'\,x / y'\,y = (y'\,y - \text{SSE}) / y'\,y$, while y are the levels and not the deviations from $ymean$. Substituting b we find: $R^2 = (x'\,y / x'\,x)^2\ x'\,x / y'\,y = (x'\,y)^2 / (x'\,x\ \ y'\,y) = k^2$, with k the cosine. Thus the proper R for RTO is the cosine. Subsequently the sine arises as a measure from $1 - k^2$. This thus differs from the Pearson R^2.

Thus we can take the RTO of x on y and find the other coefficient of determination k^2. Essentially though, the SSE are calculated from the slopes, and basically we are formulating an error measure around these slopes too.

Originally we worried that votes {50, 50} show no dispersion and thus would not allow the statistical estimate for b. However, this particular kind of dispersion is only relevant for a

[47] On the side: Remarkably, Wikipedia (2017-08-07) – a portal and no source - (also) states: "Sometimes it is appropriate to force the regression line to pass through the origin, because x and y are assumed to be proportional." Thus *proportionality indeed*.

regression with a constant, and not for a regression through the origin. (For a regression with a constant - perhaps for a sensitivity analysis – it remains useful to know that for {50, 50} the limit value of *b* is 0.)

10.5. Normalisation of the output: key step in the definition

Yet, *b* found by RTO is rather an equality / proportionality measure and not an inequality / disproportionality measure.

- If *d* = *d*[*b*] is a metric then *f d* and *d* / (1 + *d*) are metrics too, and the latter gives a normalisation to [0, 1].
- We can boost the sensitivity of lower values by a parameter *f,* so that *f d* / (1 + *f d*) = 1 – 1 / (1 + *f d*)
- A problem is that a metric *d* loses the sign of *b,* which we just had discovered as a useful feature. This can be inserted separately.
- The *b* of RTO in the nonnegative Fiirst Quadrant will be nonnegative too. For a sign we must look at the Pearson covariance with centered variables. A zero covariance however might destroy relevant values of *d*. Thus we use the *sign* = If[sign[*cov*] < 0, -1, 1]. (In a way it is remarkable how many small but crucial tricks this present paper requires.)
- When we adopt the cosine then we can find similar adaptations.
- For Sin we can also use √Sin as a way to enhance sensitivity.

10.6. Restatement of diagonal regression, major axis, geometric mean functional relationship, or neutral regression

The following is required for an adequate overview. See Tofallis (2000) for an accessible overview, Lesnik & Tofallis (2005) for more, and Samuelson (1942) for skepticism w.r.t. the underlying model assumptions. [48] Following their references to Draper & Yang (1995)(1997), we there find the following. OLS regression of *y* given *x* with a constant gives coefficient *bols* = *sxy / sxx.* OLS of *x given y* gives *pols* = *sxy / syy.* Flipping the quadrant gives 1 / *pols* as an alternative for *bols.* The diagonal regression (a.k.a. *major axis* or *geometric mean functional relationship* (GMFR) or *neutral regression*) estimate of the slope then is the geometric mean, with *sxy* covariance, *syy* variance, *sy* standard deviation, and *sign* = sign[*sxy*]:

bdiag = *sign* Sqrt[*bols* / *pols*] = *sign* Sqrt[*syy* / *sxx*] = *sign sy* / *sx*

cdiag = *ymean* – *bdiag xmean*

Draper & Yang (1995)(1997) state: "The GMFR estimate is obviously symmetric in that the interchange of *x* and *y* axes leaves the area unchanged." However:

- The numerical value of *bdiag* still depends upon which horizontal axis is taken.
- The intercept will generally not be zero, while we require zero for proportionality.
- Might that area be taken as a measure ? Consider $y = \gamma + \beta\, x + \hat{e}$ so that the absolute vertical distance is Abs[$\hat{e}$]. Then also $x = (y - \gamma - \hat{e}) / \beta$ and the absolute horizontal distance is Abs[$- \hat{e} / \beta$]. The product is Abs[$\hat{e}' \hat{e} / \beta$] = Abs[SSE / β]. The SSE is not the OLS SSE. We minimise over this expression (flipping axes to get rid of the Abs) and can confirm above outcome on *bdiag* and *cdiag.* The method is elegant and it is surprising that it is not used more often (notwithstanding Samuelson's skepticism). Anyhow, Draper & Yang already told us that Abs[$\hat{e}' \hat{e} / \beta$] apparently finds its (symmetric) minimum for *b* = *bdiag.* If I am correct then for nonnegative *bdiag*:

  Abs[$\hat{e}' \hat{e}$ / *bdiag*] = 2 *n* (*sx sy* – *sxy*) = 2 *n sx sy* (1 – *R*)

  The question is whether this would be a disproportionality measure. It would be a transform of correlation. Observe that the LHS doesn't suffer from lack of dispersion in {50, 50} versus {40, 60} while the RHS tranform does (since we divided by *sx sy*). When

[48] https://en.wikipedia.org/wiki/Total_least_squares#Scale_invariant_methods

the outcome would be 0 then we need not find this perfectly unit proportional (slope 1 but also zero constant). As a measure it thus would suffer from the same drawback of the Pearson variation around a mean. This path dies in beauty.

(PM. Both *OLS with a constant* and the *diagonal "major axis" method* differ from OLS through the origin (RTO). We intend to take $b = brto$ and $p = prto$. Obviously we then might take b / p instead of above *bols / pols,* but then we haven't resolved the issue of (numerical) symmetry.)

Tofallis (2000:9) has some useful comments on slope and correlation:

> "(...) reject RMA [diagonal, GMFR] on the grounds that the expression for the slope does not depend on the correlation between x and y. This is rather a strange objection because the correlation is a measure of the strength of the linear relationship and should be independent of the slope. The slope provides an estimate of the rate of change of y with x; why should this value be determined by the correlation (r)? After all, two regression lines can have the same slope but the data sets on which they are based can differ in the correlation; conversely, two sets of data can have the same correlation but have different regression slopes. It is conceivable that their objection may be grounded in the 'OLS conditioning' that most researchers are imbued with (since in OLS the slope is related to the correlation (r), according to: slope = r sy/sx). The eminent statistician John Tukey has indeed described least squares regression as a statistical idol and feels "it is high time that it be used as a springboard to more useful techniques" (Tukey, 1975)."

This should warn us about the interpretation of the OLS formula for the estimates of slope and *R.* The data allow these estimates (or summary statistics), but the data generating processes may have an entirely different structure.

10.7. Relation to the cosine: key step in the definition

Regressing through the origin, there is the slope b of z on w, thus $b = z' w / w' w$, and the slope p of w on z, thus $p = z' w / z' z$. Inverse values arise from flipping the quadrants.

The inequality / disproportionality measure $dmin = 1 – \mathrm{Min}[b, p, 1 / b, 1 / p]$ is symmetric and would be interesting since it uses all possible slopes generated by the data. The values of these slopes are nonnegative since w and z are in the First Quadrant and we force the line through the origin. It turns out that *dmin* is not so sensitive. We would like to see sensitivity. A difference of 1 seat in Holland is 1/150 = 0.67% and in the UK 1/650 = 0.15%, and would not be glossed over. Sensitivity can be enhanced by a transform, though. Somehow the "min" condition doesn't seem so appealing either, but it might perhaps only need getting used to. However, there is no clear linkup to the conditions of a metric, and thus this ends here. (When $b = 0$ then there is no need to consider $1 / b$.)

There are other options. Slope p applies to the Quadrant with the axes flipped, so that we may compare b and $1 / p$ in the original Quadrant. A level distance is Abs[$b - 1 / p$], a relative expression is $b / (1 / p) = b\ p$ and, as said, a geometric average is Sqrt[$b * 1 / p$].

The relative expression is symmetric. It is the same as taking the two slopes by themselves in a more abstract manner. This gives k = Sqrt[$b\ p$] = $z' w /$ Sqrt[$w' w * z' z$] or the Cosine k. [49] We define a deviation measure on it, namely Sin, obviously defined as Sqrt[1 – k^2].

- Thus the (relative, or "quadrant neglecting geometric average") slope estimator is also a measure of correlation. Originally, this paper started with the idea that k = Sqrt[$b\ p$] would be relevant to regard Sin as a sine diagonal inequality / disproportionality measure. However, the arguments eventually caused the distinction between Nonspecific and Specific disproportionality. The main step forward is to accept that the sine has a double face here.
- More conventional is the CosineDistance 1 – k. However, the CosineDistance is not a metric, see Van Dongen & Enright (2012) and a counterexample. [50]

[49] https://en.wikipedia.org/wiki/Cosine_similarity

- Van Dongen & Enright (2012:4-5) find that the sine is actually a metric preserving distance. Also Sin[*x*]^*t* for $0 < t \leq 1$ gives a *distance* transform that we can use. Though they do not explicitly say that this *distance* remains a metric, my inference is that is does.
- The cosine uses an average of the slopes, rather than selecting, like above Min condition, the worst outcome of either regression. Indeed, the CosineDistance generates low values where voters would rather see large values.
- Still, it seems reasonable to use the Cosine anyhow since it has a fair interpretation. Its lack of sensitivity can be adjusted by the transform, since this allows to increase the sensitivity in a more controlled manner. Geometry isn't sensitive to 1% difference but voters are.
- These and other considerations also cause the distinction between Nonspecific and Specific disproportionality, see **Section** 7.5.

Appendix C provides a follow-up on an angle here, using Abs[*k* – 1 / *k*]. PM. This weblog reviews some of the basic relations relevant for the present discussion. [51]

10.8. Relation to Hirschman-Herfindahl or Taagepera-Shugart ENP / CNP

Alternatively, the "quadrant-respecting geometric average" Sqrt[*b* / *p*] may be used. There is no real need for the Sqrt, though. We observe that *b* / *p* = *z'z* / *w'w* = *Nv / Ns*. In the same way as with diagonal regression the crossterm disappears. This gives a remarkable link to the Hirschman – Herfindahl [52] concentration indices *w' w* and z' z, and the Taagepera-Shugart 1989 measure [53] of the "relative reduction in the effective number of parties (ENP)" TS = 1 – *Ns* / *Nv.* Their use of the term "effectiveness" is not really defined, though. We will use the term “concentrated number of parties” (CNP). TS might generate negative values, which is no reduction but an increase. This is reasonable when we see it as a a difference from the slope 1. TS is not symmetric. A potential inequality / disproportionality measure, with symmetric absolute error, thus based upon both *b* / *p* and *p* / *b*, is:

$$d = \text{SymAbsTS} = \text{Abs}[1 - Nv / Ns]\ \text{Abs}[1 - Ns / Nv] = (Ns - Nv)^2 / (Ns\ Nv).$$

It would still have to be normalised to [0, 1]. If all this would suffice then we would save on the calculation of the crossterms. I have not looked further into this. It is likely that there is useful information in the crossterms. SymAbsTS focuses on concentration while there might be (unit) (diagonal) inequality / disproportionality in the original dispersion.

Potentially, the regression through the origin (RTO) is also done in the diagonal regression fashion. We might say this of any regression. There is no reason to continue on this option. since we already found SDID.

10.9. Weighed regression

We encountered weighing in the WSL measure. In principle, we might measure each seat on its share of the vote, cf. single member districts. Summing over districts gives parties, and thus it makes sense to weigh with party shares. Parties also have the opportunity to collect errors to gain seats. The WSL measure isn't symmetric and meets with the problem of division by zero, but we can also use:

$$\text{symWSL} = \text{Sum}[w\,(z / w - 1)^2] + \text{Sum}[z^+\,(w^+ / z^+ - 1)^2] \text{ and } \{z^+, w^+\} \text{ has } z \text{ without } 0$$

[50] Adam Przedniczek gives the points {1, 0}, {1, 1} / Sqrt[2] and {0, 1}, see https://stats.stackexchange.com/questions/198080/proving-that-cosine-distance-function-defined-by-cosine-similarity-between-two-u/198103

[51] https://brenocon.com/blog/2012/03/cosine-similarity-pearson-correlation-and-ols-coefficients/

[52] https://en.wikipedia.org/wiki/Albert_O._Hirschman#Herfindahl-Hirschman_Index

[53] https://en.wikipedia.org/wiki/Effective_number_of_parties

We encountered weighing also in **Section** 10.1. The unweighed RTO $z / w = b$ 1 gives b = Average[z / w], so that this outcome can be seen as $z = b\ w$ with weights 1 / w, that favour the small parties. This appears to be problematic, see **Appendix E**.

It would make some sense to have weighed regressions with weights w for $z = b\ w$ and $w = p\ z$. This would seem to tend to be more neutral to party mergers or splits.

For now, I don't opt for this for the suggestion of SDID. The cosine measure is straight from the first year linear algebra course and from econometrics textbook chapters without weights, and seemed simplest to work with at this stage. If SDID is adopted for inequality / disproportionality measurement, then one might keep in mind that parties (and their marginal candidates) react to rules of apportionment and not to measures of inequality / disproportionality in the research on electoral systems.

10.10. Thus the sine diagonal inequality / disproportionality (SDID) measure

These considerations thus generate the slope deviation measure, defined in **Section 4**.

These considerations arrive somewhat late in this paper, but the comparison of the results for Holland, France and the UK required the early introduction of the proposed measure SDID and the decomposition in *slope-specific* and *dispersion-nonspecific* factors.

The following section first discusses other considerations before we look at numerical examples, since those considerations are relevant for judging the examples.

11. Apportionment, measuring, wasted vote

11.1. Definition of "wasted vote"

This Section considers apportionment that minimises a minimand, then develops potential inequality / disproportionality measures based upon those minimands, and then supplements with additional information on statistics, parts of which we have already seen above.

With Turnout = Wasted Vote + Elected, or $V = W + Ve$, we have the vectors of perfect, though non-integer, allocations $a = S\ v / V = T\ v = v / Q = S\ w$ and $ae = S\ ve / Ve$. The first includes the wasted votes and the latter excludes them. Standardly it is useful to collect all wasted votes into a single variable, not only to avoid a list of unfamiliar names and zero seats (of far away countries). Issues on the number of (contending) parties are rather treated separately.

Formulas will tend to be the same for both apportionment and the definition of an inequality / disproportionality index. It is tempting to choose for ease of exposition, which would mean the use of only *V, v* and *a* rather than *Ve, ve* and *ae.* Common texts on this subject choose for this ease of exposition and then appear to be confusing about the difference between apportionment and measures of disproportionality. Thus, we currently prefer some primness.

However, it would be a major principle for EPR that the wasted vote should be represented by empty seats or be translated into a qualified majority. When we technically fill seats with blanco (dummy) representatives then we can still use the symbols *V, v* and *a* for both apportionment and a measure for disproportionality, see **Table 16**. After finding the relevant formulas, we then can substitute values of *Ve, ve* and *ae,* and then turn those blanco seats into 0, to fit current practice. See **Table 2** for the notation of the variables and errors.

Table 16. Turnout = Wasted Vote + Elected: similarity of formulas

	$w = v / V$, $s[a = S w]$, $z = s / S$	$we = ve / Ve$, se[ae = $S we$], ze
Apportionment: s has no 0. Minimize an error function	(a) Wasted vote get seats with dummy candidates, not counted as 0. The allocation error is either $\tilde{u} = s - a$ or $\tilde{u}r = s / a - 1$.	(b) Same formulas as the left but different base. The allocation error is either $\tilde{u}e = se - ae$ or $\tilde{u}re = se / ae - 1$.
Disproportionality measure = the minimand used in apportionment, or a separate standard	(c) Similar formulas as the above. Potentially one may subtract the value of the minimand found above, since the minimum above would be the (locally) optimal apportionment.	(d) When s is taken from (b), then the measure rises with the effect of excluding the wasted vote (and not using a qualified majority).

11.2. Apportionment (Jefferson, Webster, Hamilton)

Let us consider the situation in (a).

A main consideration is that seats s are taken in the range Floor[a] ≤ s ≤ Ceiling[a]. It is common to minimize a sum of squared errors SSE = Sum[$\tilde{u}^2$] for the given selection range. Yet, shall we take the error and SSE in levels or in ratios ? There is further no objective criterion for what would be unit proportional. Anything in the range Floor[a] ≤ s ≤ Ceiling[a] is a matter of choice, as only a is the proper proportion in a mathematical sense. Any choice might come with a paradox. The gains Γ = Max[0, $s - a$] for the electoral lucky compare to the losses Λ = Max[0, $a - s$] of those sent down to the Floor, with SSE = $\Gamma' \Gamma + \Lambda' \Lambda$. The value of dsa = Sum[Abs[$s - a$]] = Sum[$\Gamma + \Lambda$] gives relative error in allocation dsa / Sum[a] = dsa / S, perhaps a minimal value that is unavoidable in apportionment, but not yet necessarily a measure of disproportionality.

In writing this paper, there developed a view on principle, that has been moved to the introduction, because of its importance. The following is a more traditional discussion of the traditional suggestions on apportionment and measuring disproportionality. The traditional discussion selects The Party rather than the party marginal candidate.

HLR: This method is called the method of "Largest Remainder" or the method of Hamilton (HLR). With level error $\tilde{u} = s - a$, minimising SSE effectively means ordering the a – Floor[a] values (on size), and allocate the remainder seats to parties that would otherwise cause most of the error. Lumping the wasted vote into "Other" may cause that it gets remainder seats.

As **Section 1.1** and **Appendix L** show, HLR is tolerant of parties that fail to meet the natural quota Q. Practical application in Holland suggests the use of a threshold, though Holland has the high threshold of 1 Q rather than 0.5 Q. Thus the better method is the *Representative Largest Remainder* (RLR), defined in **Appendix L**. See **Appendix N** on HLR, RLR and ALHID and the "New States Paradox" (that is no real paradox).

(Let us reuse symbols f and c with a locally different meaning, and neglect a party index. Shifting from Floor f to Ceiling $c = f + 1$ means that the current SSE is reduced by $(a - f)^2$ and rises by $(c - a)^2$. The total change is $(c - a)^2 - (a - f)^2 = 2 (c - a) - 1 = 1 - 2 (a - f)$. Taking the highest values of $a - f$ causes the least increase of the SSE.)

WSL: This is the Webster / Sainte-Laguë (WSL) principle. The latter HLR disregards party sizes. For parties with 2 or 30 seats, an additional seat has different proportional meanings. Small parties are more likely to suffer relative underrepresentation. We could take $s / a - 1 = z / w - 1$ and weigh with the size of the votes w or attach more value to smaller parties with weights $1 - w$ or $1 / w$. These options are:

- minimand Sum[$w (z / w - 1)^2$] = Sum[$(z - w)^2 / w$]
- minimand Sum[$(1 - w) (z / w - 1)^2$] = Sum[$(z - w)^2 (1 - w) / w$]
- minimand Sum[$(z / w - 1)^2 / w$] = Sum[$(z - w)^2 / w^3$] (exponent 3)

WSL uses the first approach. Its solution is to sort the parties [54] on the Floor[a] / v ratio, and assign Ceiling values to the lowest outcomes, so that their position improves to Ceiling[a] / v. Unless the locus Floor[a] ≤ s ≤ Ceiling[a] is strictly imposed, other formulations of WSL might violate it. Lumping the wasted vote into "Other" may not avoid that this is treated as an overall small party, but the effect (supposedly) would be greater if they were not lumped together.

So-called paradoxes on HLR and WSL are: The choice of WSL over HLR is supposed to resolve the paradox that party sizes are overlooked. However, when one adopts WSL then it might be possible that a party with 10 seats might gain votes but lose a seat, while a party of 4 seats loses votes but gains this seat. This is only paradoxical when one would not see the underlying reasoning via the objective function.

With reference to the introduction, one hopes that the reader spotted the fallacy. Since HLR basically *defines* what *equal (proportional) representative democracy* is, it is a fallacy to say, as we did above, following the literature, that " The latter HLR disregards party sizes. For parties with 2 or 30 seats, an additional seat has different proportional meanings. Small parties are more likely to suffer relative underrepresentation." This argument looks at The Party and not at the party marginal candidate. Thus WSL is fundamentally not a relevant criterion for looking at the party marginal candidate.

JDH is Jefferson / D'Hondt (JDH). Gallagher (1991) finds that JDH apparently has the objective function to favour the largest party. Tabular algorithms may look inversely at v / (k Floor[a] + 1) with k = 1 for JDH and k = 2 for WSL.

Given that WSL is a relative measure and we are considering EPR anyway, it would seem that WSL would tend to be the better measure, for The Party. However, for inequality / disproportionality measurement, it is not symmetric. The story that WSL would be better for small parties does not seem to be correct, see **Table 23**.

However, a (political) property of apportionment is that it should not encourage parties to split up merely to gain seats. It is more commonly regarded as acceptable that parties are encouraged to join up and be rewarded by relatively more seats. On the other hand one would not want to overly encourage large parties at the cost of opposition. If the policy makers have these considerations then HLR would beat both WSL and JDH. That is, for apportionment. As said, **Appendix L** suggests RLR, in which parties with at least 0.5 Q get a chance at remainder seats.

HH: The USA uses the Hill-Huntington (**HH**) method to allocate seats to State representatives in proportion to the population in the States, see Malkevitch (2002), [55] (2017), Caulfield (2010), Balinski & Young (1980) and Young (2004). There is little advice [56] to use this system also for the apportionment of seats to parties, and subsequently to turn it into a measure of disproportionality, since the current formulation assigns at least one seat to a state, while parties might belong to the wasted vote. Malkevitch (2017): [57]

> "Much of the theory here was developed by E.V. Huntington a mathematician who taught at Harvard University. Huntington had looked into the 32 ways that the inequality pi/pj > ai/aj (where the population of state i, is pi and the number of seats given state i is ai) could be rewritten by "cross multiplication." He worked out the different measures of "inequity" between pairs of states that could be used in this way. He observed that in a comparison between two states who had average district sizes of [large]100,000 and [small] 50,000 compared with [inbetween large] 75,000 and [very small] 25,000, the absolute difference is the same [namely 50,000]. However, he thought that the inequity was "worse" in the second case because

[54] The various "tabular" and "divisor" methods apparently try to avoid sorting, and seem needlessly complex compared to simple sorting.
[55] http://www.ams.org/samplings/feature-column/fcarc-apportionii1
[56] https://en.wikipedia.org/wiki/Huntington%E2%80%93Hill_method
[57] Pairwise Equity in Apportionment (2017)

> 50000/25000 is 2 while 50000/50000 is 1. The relative difference to Huntington seemed a better measure. (Relative difference between x and y being defined as |x - y|/min(x,y).) Of the absolute differences the two most "natural" are |pi/ai - pj/aj| which is optimal when Dean's method is used and |ai/pi -a/pj| which is optimal when Webster's method is used."

The topic of allocation of seats to states is different from the allocation of seats to parties, though. See also the seats for the Member States of the European Union. [58]

Observing the distinction between *apportionment* and *measuring disproportionality*, we may summarize current findings:

- *Apportionment* takes a decision following the minimisation of an objective function. Apportionment concentrates around the *locus of seat allocation*. This deals with small differences around a roughly proportional allocation, Floor[*a*] ≤ *s* ≤ Ceiling[*a*], in which it may happen that a party with votes may not get a seat due to missing the quota, but in which it should not happen that a party reaching the quota will not get a seat.
- The issue of apportionment is essentially solved by HLR, if one accepts that it defines *equal (proportional) representation,* namely looking at the party marginal candidate.
- *Inequality / disproportionality measures* contain the corresponding minimand, and are intended to judge about *wide differences*. The measures also must deal with cases, as this can happen in District Representation, that a party reaches the quota (by collection over various districts) but still will not get a seat. DR has quite other considerations than EPR.
- Apportionment focuses on parties that get a seat in the House of Commons, so that *s* has no zeros, while measures of inequality / disproportionality would rather include the wasted vote, consisting of invalid ballots and votes for parties that do not meet the quota, so that *s* has a zero (lumped together), since it obviously is disproportional to vote but not to be represented. The shares thus have different denominators, with the rule: Turnout = Wasted + Elected. (If a country would use a qualified majority such that the wasted vote can be replaced by *s* without 0, then this would be measured by rebasing to *s′* such that the wasted votes would have blanco seats, and allowing that this *s′* would not be integer.)
- Thus while apportionment would be perfect under its own rules, it would still show up as somewhat disproportional under its own associated inequality / disproportionality measure.
- For apportionment, a political principe would use HLR / RLR (on *Ve* and *ae*) for neutrality, to not encourage mergers and splitting of parties.
- For measurement one can assume that parties exist as they are. If we would restrict our attention to these traditional methods, then we could tend to use a relative measure as this better fits the relative character of proportionality itself. But a choice for WSL would not be seen as clearly better overall. Given the principle discussed in the Introduction, the remainders above floor assignments are actually also relative values, namely with the party marginal representative in the denominator.
- This difference between apportionment and measurement doesn't need to create a great tension between what Parliaments do and what political scientists measure. (If so, one might subtract a mismatch on a baseline.)
- Measures are sensitive to listing all parties in the wasted vote separately or lumping them into a single "Other". It is advisable that lumping for less than 0.5 *Q* is the standard. For the Gini (in the definition in Colignatus (2017b)) lumping is immaterial but also preferable. The EGID index with its squares on levels would be higher with such lumping.
- Above apportionment neglects issues on majorities. The notion of unit proportionality in apportionment is itself power preserving, but apparently in tradition it plays no role in the allocation of remainder seats (except for JDH apparently). Inequality / disproportionality allows that a party / coalition with a majority in the votes would not get a majority in the seats, or that a party / coalition without such majority in the votes still gets it. Rules of apportionment that impose additional constraints on such preservation might be called "majority (unit) proportional". This would be a more sophisticated manner to improve the power preservation of systems that call themselves EPR.

[58] https://en.wikipedia.org/wiki/European_Parliament#Members

11.3. Inequality / disproportionality measures based upon apportionment

For measuring disproportionality, we would tend to use the *minimand* that is used in the apportionment rule, like Sum[$(s - a)^2$] or Sum[$(s - a)^2 / a$]. There is also the condition *Penalty* * (Max[0, Floor[a] – s]^2 + Max[0, s – Ceiling[a]]^2) for some penalty weight. We repeat some observations from the above but focus a bit more on this condition. Various graphs above suggest smooth trajectories, but we have not imposed strict adherence to the locus of apportionment.

- (a) The Euclid / Gallagher measure (EGID) corrects for the different number of seats in the various countries. With $a / S = v / V$ we get the minimand Sum[$(s / S - v / V)^2$]. The EGID measure then is Sqrt[1/2 Sum[$(z - w)^2$]] for observed values of s or z (and no longer the objective to apportion them). For a binary case with votes {w, 1 – w} and seats {z, 1 – z} with scalars now, the sum of squares is $(z - w)^2 + ((1 - z) - (1 - w))^2 = 2 (z - w)^2$ so EGID reduces to Sum[Abs[$z - w$]] = ALHID, the same value as the Abs / Loosemore-Hanby index.
- PM. The Gallagher measure treats values outside of Floor[a] ≤ s ≤ Ceiling[a] on an equal base. We may take allocation a as a fair estimate of both its floor and ceiling though.
- (b) The Webster / Sainte-Laguë method weighs the proportional errors by the votes, with Sum[$v / V * (s / v - S / V)^2$] = Sum[$(z - w)^2 / w$], so that Gallagher's terms are weighted by the inverses of the vote shares. The method might be more sensitive to values outside of Floor[a] ≤ s ≤ Ceiling[a]. It may be relevant here to stricter impose the penalty. If the penalty uses only squared values and no inverse weights then it would embed the measure within the wider range of the Gallagher measure.
- Gallagher (1991:47): "The Sainte-Laguë index (...) at the theoretical level is probably the soundest of all the measures (...). If it has a drawback, it is that it can be affected by the amount of information available on the fate of small groups." This refers to the lumping of all wasted votes into "Other". The latter can be advised as the standard.
- (c) Gallagher (1991:42) finds that the Jefferson / D'Hondt objective would be to maximise "the seats to votes ratio of the most over-represented party".

PM. Above measures all assume a single vote, with either the first preference or strategic voting. Systems like Single Transferable Vote (STV) allow the transfer of wasted votes to alternative parties. One would still record the first preference as the true vote, and the outcome after transfer might be measured as disproportional, though the idea of STV is that the outcome would be better than wasting the vote.

11.4. Statistical measures: Kullback-Leibler (KL) and Chi square

It remains important to be aware that s has been created by an apportionment on *ve* and that we now measure on s and v. The objective here is not to test apportionment but to see how such apportionment fares w.r.t. a measurement *standard*. In that case we may take greater liberty in choosing an inequality / disproportionality measure. Distance measures may be used in regression analysis or display, see Zand et al. (2015).

Kullback-Leibler (KL) [59] measures how shares or probabilities $z = s / S$ and $w = v / V$ compare. The divergence $D_{KL}[z \,||\, w]$ is also the relative entropy of z with respect to w. Taagepera & Grofman (2003:665) mention the measure simply as "entropy" and refer to an application by Pennisi 1998. It is not a true metric, since the outcome is asymmetric, but this would be okay for votes w and allocation of seats z. Belov & Armstrong (2011) have some results for normally distributed shares. Their lemma 2 has $D_{KL}[z \,||\, w] \approx ½ X^2$ for the *scaled* Chi square. [60]

Theil (1969:522) also showed that the first order approximation of $I[z ; w] \approx ½ X^2$. This is actually also the WSL inequality / disproportionality measure. Theil's suggestion of "weak

59 https://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence
60 https://stats.stackexchange.com/questions/9629/can-chi-square-be-used-to-compare-proportions

proportionality" $z = w^\alpha / (1' w^\alpha)$ is not generally satisfied for outcomes under DR. One may appreciate also Theil's suggestion of strategic voting on entropy.

The Pearson Chi square [61] is often used to test whether observed counts or frequencies are statistically different from the expected values. The allocation of the seats s is *observed* and we *expect* that it would follow the distribution of the votes v with expected value $a = S\ w$. H0 is that the vectors are the same, and we would reject H0 if the p-value is lower than a specified level. A requirement is that all $S\ w$ are at least 5 seats. With n the number of parties with seats, then the degrees of freedom would be $n - 1$, since Sum[$S\ w$] = S. It is debatable whether the assumptions for doing this test really apply though.

$$X^2 = \text{Sum}[\ (s - a)^2 / a\ ] = \text{Sum}[\ (s - S\ w)^2 / (S\ w)\ ] = S\ \text{Sum}[\ (z - w)^2 / w\ ] = S\ \text{WSL}$$

Apparently the Chi square is little used in the voting literature for this purpose, and this is a bit surprising, till we realise that it actually is identical to S WSL. This brings the comparison with the KL measure above to the fore. KL compares to WSL or a *scaled* Chi Square, while the standard Chi square test has S WSL. What to think about this issue on scaling ? Unfortunately this issue is not mentioned by Taagepera & Grofman (2003). However:

- Given that the Chi Square is a standard for comparisons, and given that we want to compare countries with different S, we apparently must also multiply the WSL with the total number of seats S. Gallagher (1991:46) table 4 scores countries on WSL but doesn't mention S. Including S would cause that the correlations with the other indices change. Holland (150 seats) and the UK (650 seats) have scores 2.1 and 23.5. To transform these into comparable standard X^2 we would get 150 * 2.1 versus 650 * 23.5. With such values, we would readily reject H0 that Holland or the UK have a unit proportional distribution.
- Thus the standard Chi Square with S WSL is so sensitive that it will generally cause us to reject H0 even for countries with EPR like Holland. This makes the Chi square less useful.
- Goldenberg & Fisher (2017) argue that WSL was based upon other considerations than the Chi square. This can readily be accepted. Also: WSL compares to KL.
- Thus WSL can still be used to compare countries, based upon WSL itself. WSL links up with the KL measure that also behaves like the scaled Chi square. (Belov & Armstrong (2011) lemma 2 was important for this insight.)

The assumptions of Pearson and WSL do not conflict in all respects. Pearson's assumptions only generate this particular format because the underlying (uniform) normality assumptions allow the derivation of an exact distribution, subsequently baptised with the name of X^2. One can start with the assumptions of WSL and then choose how to deal with the statistical issue. Either one adopts the statistical assumptions for hypothesis testing, or one merely uses the mathematical transform to create the inequality / disproportionality index in probability format. The probability transformation associated with the calculated X^2 still seems to be a tempting measure of (dis-) proportionality. Thus, this would not necessarily be taken as a statistical model but mainly as a useful transformation towards the [0, 1] range. A rising X^2 means a greater disproportionality. Thus $1 - p$[WSL] with p still taken as the $X^2[n - 1]$ would be the measure of disproportionality. Taagepera & Grofman (2003:664) mention that Mudambi 1997 proposed this. It remains to be seen whether it really can be used to compare countries (and their methods) over space and time.

(Remarkably, Pukelsheim (2014:130) puts the seats in the denominator, though correctly calls it a modification. He derives an objective function for a Member of Parliament based upon the number of voters that the MP represents, as Max v / s. Here, we rather look at an overall objective and use the format Min s / v.)

11.5. Correlation and cosine

Perhaps correlation is a more "objective" or "neutral" notion of association. It may be applied to v and s without the level corrections on V and S. The Pearson Correlation has the

[61] https://en.wikipedia.org/wiki/Chi-squared_test

distances from the mean. We may however also consider measurement of the variables just from the origin (as they originally are).

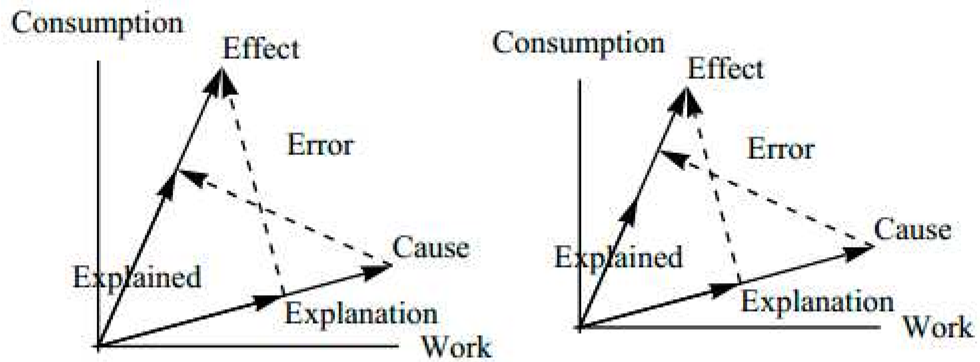
Koppel & Diskin (2009) point to the Cosine as a measure that satisfies key properties. The cosine is also called "similarity". With vote vector *v* and seat vector *s,* Cos[*v* – Mean[*v*]*, s* – Mean[*s*]] is the Pearson coefficient of correlation that reshifts to around the mean of the variables. This connection between cosine and correlation is a well established notion in the statistical literature, but potentially not in the voting literature.

A potential measure is CosineDistance = 1 - cos[*θ*], [62] for *θ* = angle[*s, v*] that does not center the variables around their means. For the link of correlation and cosine to the use of the notion of (unit) "disproportionality" in the voting literature, I benefitted from Koppel & Diskin (2009), who propose to use this CosineDistance. The Wolfram Language adopted the CosineDistance = 1 - Cos in 2007. [63] See also the discussion by Zand et al. (2015). The cosine distance however is not a metric, see Van Dongen & Enright (2012).

Hill (1997) refers to Woodall 1986 in the *Mathematical Intelliger* who mentions that Dr J E G Farina proposed the cosine or correlation to measure the agreement of votes *v* and seats *s*. Hill has the remarkable suggestion to include {-*v, -s*} in the calculation of correlation to test on linearity.

Figure 22 has been taken from Colignatus (2011:143) and shows how correlation and cosine relate.

Figure 22. Correlation is no causation



The model is that the variation in Effect (Consumption) is explained partly by the variation in Cause (Work) and partly by unknown factors (Error). The split is achieved by the projection of effect onto the cause ("Explanation"), while the dashed arrow from there to the Effect is the error. The *Effect* thus comes about by the vector addition of *Explanation* and *Error*. The perpendicularity makes these influences independent. Relative measures compare Explanation to the Effect, or Error to the Effect. A distance measure of 1 means zero Error, and that Cause and Effect fully overlap.

- For correlation the vectors are taken in deviation from their means (centered).
- A step further is to standardise them onto the unit circle, thus divide by the standard deviation, so that the centered vectors have unit Euclidean length.
- For the cosine the vectors are not centered and just taken as they are.
- A step further is to normalise them onto the unit simplex, thus divide by their sum.

With two parties with vote shares {40, 60} and seat shares {60, 40), the correlation is –1 and the Cosine is 0.9231, giving a CosineDistance or 1 - Cos of 7.7%, which is rather insensitive. The ALHID / EGID score is 20%. This actually also shows that the measures might be

[62] https://en.wikipedia.org/wiki/Cosine_similarity
[63] http://reference.wolfram.com/language/ref/CosineDistance.html

insensitive to cases when a false majority is created. Comparing with the cosine for {10, 90} and {90, 10} the correlation remains –1 while the CosineDistance rises to 78%. The change from 7.7% to 78% seems proper, but an original problem was that 7.7% did not capture the false majority anyway. A idea is to use a measure so that the creation of a false majority is better indicated. A heuristic is to try to combine:

Correlation[*v, s*] * (1 – Cos[*v, s*])

A correlation index on votes {50, 50} collapses since correlation cannot deal with variables that have no variation. We can use the covariance then. However, a covariance of zero would be disinformative by itself, and destroys the information coming from the cosine. We are only interested in the switch to –1, and thus adapt the sign accordingly:

If[Covariance[*v, s*] < 0, -1, 1] * (1 – Cos[*v, s*])

Table 17 reproduces the data from Koppel & Diskin (2009), and includes the latter in the last column. The Cos measure is relatively unsensitive and we might apply the sign also to another index, like the Gallagher. PM. It seems curious to include cases with zero votes.

Table 17. Data from Koppel & Diskin (2009) including new last column

	Votes	Seats	D	R	G	G'	1-Cos	Sgn*(1-Cos)
1	{50,50}	{60,40}	.10	.10	.10	.14	.02	.02
2	{50,50,0}	{60,40,0}	.10	.07	.10	.14	.02	.02
3	{50,50,0}	{0,0,100}	1.00	.67	.87	1.00	1.00	-1.00
4	{25,25,25,25}	{30,30,20,20}	.10	.05	.07	.14	.02	.02
5	{25,25,25,25}	{35,25,20,20}	.10	.05	.09	.17	.03	.03
6	{50,50,0}	{50,0,50}	.50	.33	.50	.71	.50	-.50
7	{25,25,25,25,0}	{25,25,25,0,25}	.25	.10	.25	.50	.25	-.25

K&D observe that the Euclidean distance or Gallagher measure would be less sensitive to orthogonal outcomes, whence it would not measure disproportionality. Taagepera & Grofman (2003:666) have an example {25, 25, 25, 25, 0, 0} vs {0, 0, 0, 0, 60, 40}. In their story some Prince appoints people who did not get votes. Or, new (half) elections fully replace parties. They regard this orthogonality as maximally "disproportional" too (or "largest change"). T&P still are inclined to stick with the Gallagher index, which generates 0.62 in their case while they want to see 1. But we are now also more acutely aware that the Prince also switches the majorities. The cosine distance for their Prince becomes 1 but we might wish to see -1.

Looking at the values of the cosine distance for the cases that Koppel & Diskin (2009) consider, we find that the sensitivity is not large. The cosine is not as sensitive to distinctions between *s* and *v* as voters would be. See also **Appendix C**.

11.6. Euclid / Gallagher distance (EGID) and cosine

The Euclidean distance is a metric, and thus Euclid / √2 or the EGID too. The index could be transformed to [0, 10] and made more sensitive, by taking the square root. This transform has been applied for comparison of SDID with EGID in **Section** 12.

The relation between the Euclidean distance and the cosine gives the following (see also the discussion on the unit simplex in **Appendix B**). Let ||*x*|| = Sqrt[*x' x*] or the Euclidean length.

> ||*z* – *w*||^2 = (*z* – *w*)' (*z* – *w*) = ||*z*||^2 + ||*w*||^2 – 2 *z' w* = *z'z* + *w'w* – 2 *z'w*
> ||*z* – *w*||^2 = 2 EGID^2
> *z'z* = 1 / *Ns* or concentration, with *Ns* the concentrated number of parties CNP for *s*
> *w'w* = 1 / *Nv* or concentration, with *Nv* the concentrated number of parties CNP for *v*
> cos[*θ*] = *z' w* / (||*z*|| ||*w*||) = *z' w* Sqrt[*Ns Nv*]
> *z' w* = cos[*θ*] / Sqrt[*Ns Nv*]
> Thus, as the law of cosines:
> 2 EGID^2 = 1 / *Ns* + 1 / *Nv* – 2 cos[*θ*] / Sqrt[*Ns Nv*]

2 EGID^2 = 2 (1 – cos[θ] / Sqrt[*Ns Nv*]) + (1 / *Ns* + 1 / *Nv* – 2)

EGID = Sqrt[(1 – cos[θ] / Sqrt[*Ns Nv*]) + ((1 / *Ns* + 1 / *Nv*) / 2 – 1)]

These corrections cause the EGID to be a metric while 1 – Cos is not. Thus we can see that the EGID measure works in the same direction as 1 - cos[θ], though apparently not to full satisfaction as holds for the sine. As said, for the bipartite case, EGID reduces to ALHID, and then EGID has the insensivity of the Dalton transfer, see **Section** 12.6. PM 1. For the extreme points with *Ns* = *Nv* = 1, we get EGID = Sqrt[(1 – cos[θ]] but with θ either 0 or 90 degrees, so that there is no difference with Cos^2. PM 2. The transformation to the unit circle *v* / ||*v*|| and *s* / ||*s*|| does not change the angle θ, and doesn't change above formula that essentially uses the unit simplex with *z* and *w.* Observe that Sqrt[*z*] and Sqrt[*w*] are on the unit circle too, but with a different angle.

11.7. A measure that uses the determinant

Let me mention the possibility discussed in Colignatus (2007) as well. The measure there is basically an ordinal measure, but the point is made that we tend to be forced to present nominal data in some order too. I have further not looked into a possible application to voting.

11.8. Can a uniform best measure be based upon tradition ?

Applying inequality / disproportionality measures involves counterfactuals. When country *C* uses objective function *O*, then a researcher using method *R* would have to explain why. A major consideration is to compare countries. If countries use the same apportionment then one might use that method. Comparing countries of EPR and DR might involve the idea that there might be a more objective notion of disproportionality, which is somewhat dubious, even though this present paper suggests such a measure.

When we restrict our attention to the traditional methods of JDH, WSL and HLR, then we find:

- WSL, with its relation to the Chi square (divided by *S*), seems to generate the most interesting measure, namely the associated probability transform. This would not be regarded as if the statistical assumptions would apply, but it would merely be a transform that one can attach a [0, 1] meaning to. This requires the mention of the Degrees of Freedom. However, the WSLT presented here seems more useful.
- WSL has a form that suggests higher sensitivity to small parties, yet this form might be deceptive, and the discussion of **Table 23** suggests that this causes the discussion: "when is small also small enough ?"
- If the menu card only contains the traditional measures, then there is no clear advice for a single measure. Seen from the world of the traditional measures, (unit) inequality / disproportionality has Byzantine aspects. Alongside the EGID of current convention and the Lorenz and Gini for display and ease of communication, the discussion ***in this Section*** suggests that if voting research have a preference for WSL then they might consider (with a penalty that dominates outside of the locus) (and then WSL still isn't in the [0, 1] or [-1, 1] range):

  Sign Sum[$(z – w)^2$ / *w* + *Penalty* / *a* * (Max[0, Floor[*a*] – *z S*]^2 + Max[0, *z S* – Ceiling[*a*]]^2)]

  in which Sign = If[Covariance[*s, v*] < 0, -1, 1].

12. Comparison of scores on theoretical cases

The tables have the following layout.

- The first row has the column number.
- The second row states the inequality / disproportionality measure.
- The third row has the range. Negative values that indicate majority switches are neglected for the range.

- Column zero gives labels to the cases.
- Column 1 has votes (V) and column 2 seats (S). There may be a particular numbers but the calculations first turn all vectors into percentages.
- Column 3 gives the proposed measure SDID with *f* = 2, or sign 10 √ Sin. Negative values indicate a switch in majority.
- Column 4 gives the slope-specific component.
- Since there is risk of confusion on the scale, column 5 multiplies column 3 by 10.
- Column 6 uses *f* = 1, or sign 100 Sin.
- The indices ALHID, EGID and WSL and their transforms have been explained above.

12.1. Comparison of bipartite cases around 50%

Table 18 gives scores of the inequality / disproportionality measures on theoretical bipartite cases. For two parties EGID = ALHID = LHI. PM. Some tables are pictures from excel sheets that still use older abbreviations.

Table 18. Scores of measures on theoretical bipartite cases around 50%

	1	2	3	4	5	6	7	8	9	10	11	12	13
			10√Sin	Specif	100√Sin	100 Sin	Gini	LHI	EGD	WSL	LHIT	EGDT	WSLT
Ex.	V	S	10	10	100	100	100	100	100	inf	10	10	10
A	40	50	4.4	1.4	44.3	19.6	10.0	10.0	10.0	4.2	3.2	3.2	1.7
	60	50											
B	50	40	4.4	1.4	44.3	19.6	10.0	10.0	10.0	4.0	3.2	3.2	1.7
	50	60											
C	50	49	1.4	0.1	14.1	2.0	1.0	1.0	1.0	0.0	1.0	1.0	0.2
	50	51											
D	50	50	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	50	50											
E	50	60	4.4	1.4	44.3	19.6	10.0	10.0	10.0	4.0	3.2	3.2	1.7
	50	40											
F	51	49	-2.0	-0.3	-20.0	-4.0	2.0	2.0	2.0	0.2	-1.4	-1.4	-0.4
	49	51											
G	60	40	-6.2	-2.8	-62.0	-38.5	20.0	20.0	20.0	16.7	-4.5	-4.5	-2.9
	40	60											
H	49,000	49	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	51,000	51											
I	49,900	49	1.3	0.1	13.4	1.8	0.9	0.9	0.9	0.0	0.9	0.9	0.2
	50,100	51											

Comments are:

(1) Cases D and H are proportional and are generally recognised as such. Case H is only a test on normalisation to the unit simplex, but helps to compare to **I**.
(2) Cases B and F have already been discussed for **Table 5**, and one now has also the outcomes of the other measures.
(3) ALHID is the corner stone that tells us the number of dislocated seats in a parliament of 100 seats. In case G the ALHID of 20 seats meets with an alarming –6.2 on a scale of 10 of SDID.
(4) WSL is somewhat difficult to judge upon given the infinite range. One rather compares values over the columns than over the rows. Perhaps we might hold that a WSL of 1 already is worrysome and a value of 10 alarming ? Potentially there might be little reason to compare electoral outcomes when we already know ahead that the results will be alarming.
(5) Cases A and B or E are symmetric, and this is generally respected except by WSL.

(6) Case C gives a small difference, and this gets more attention from SDID than the other measures. With a {50, 50} vote and a parliament of 100 seats, scoring an outcome {49, 51} with a low value 1, like the Euclidean distance does, might send the message that there is no reason to worry, except that it could also be much worse with {1, 99} of course. The geometric outcome of 2 by 100 Sin is low too, and one would prefer its square root.
(7) WSL is insensitive to the difference between C and D, and H and I (but we print only 1 decimal). In I, WSL ≈ Euclid^2 / 50 so that the EGID outcome below 1 is squared and divided by a larger number, whence the result disappears in the lower digits, not shown. Multiplying WSL by 100 is not required since we already have done so by using percentages. WSL simply is insensitive to small differences. This is shown again by comparing with WSLT (on a scale of 10). (And see the earlier plot.)

12.2. Comparison of bipartite cases with a small party

Table 19 gives scores of measures on theoretical bipartite cases with a small party.

- The votes for cases J to N are {10, 90} and we look at a range of seats around 10.
- Case M is fully proportional. For reference disproportional case Q has also seats {10, 90},

Table 19. Scores of measures on theoretical bipartite cases with a small party

	1	2	3	4	5	6	7	8	9	10	11	12	13
			10√Sin	Specif	100√Sin	100 Sin	Gini	ALHID	EGID	WSL	ALHIDT	EGIDT	WSLT
Ex.	V	S	10	10	100	100	100	100	100	inf	10	10	10
J	10 90	0 100	3.3	0.8	33.2	11.0	10.0	10.0	10.0	11.1	3.2	3.2	2.5
K	10 90	1 99	3.2	0.7	31.7	10.0	9.0	9.0	9.0	9.0	3.0	3.0	2.3
L	10 90	5 95	2.4	0.4	24.1	5.8	5.0	5.0	5.0	2.8	2.2	2.2	1.4
M	10 90	10 90	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
N	10 90	15 85	2.5	0.5	25.3	6.4	5.0	5.0	5.0	2.8	2.2	2.2	1.4
Q	30 70	10 90	5.4	2.1	53.9	29.0	20.0	20.0	20.0	19.0	4.5	4.5	3.0

Comments are:

(1) All measures pick up the move from inequality / disproportionality with 0 seats to proportionality with 10 seats, and then to disproportionality with 15 seats again.
(2) SDID is indeed more sensitive to inequality / disproportionality than Sin or Gini. When a small party with 10% of the vote doesn't get a seat, it means a score of 3.3 on a scale of 10.
(3) Perhaps 10 seats still is relatively much. There is nothing remarkable about this table, so that one would base decisions upon other results.

12.3. Comparison of bipartite cases with a tiny party

Table 20 looks at the position of a tiny party that might miss the threshold or that might get a bonus from some defectors of a large party (and we don't measure their votes ....).

Table 20. Scores of measures on theoretical bipartite cases with a tiny party

	1	2	3	4	5	6	7	8	9	10	11	12	13
			10√Sin	Specif	100√Sin	100 Sin	Gini	ALHID	EGID	WSL	ALHIDT	EGIDT	WSLT
Ex.	V	S	10	10	100	100	100	100	100	inf	10	10	10
R	1	0	1.0	0.1	10.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9
	99	100											
S	1	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	99	99											
T	1	2	1.0	0.1	10.2	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9
	99	98											
U	5	0	0.7	0.0	7.1	0.5	0.5	0.5	0.5	0.5	0.7	0.7	0.7
	995	100											
V	5	1	0.7	0.0	7.1	0.5	0.5	0.5	0.5	0.5	0.7	0.7	0.7
	995	99											
W	5	2	1.2	0.1	12.4	1.5	1.5	1.5	1.5	4.5	1.2	1.2	1.8
	995	98											

Comments are:

(1) The pattern is remarkably similar to the case for the small party, though with some points.
(2) WSL has a reputation of being more sensitive to small parties, but doesn't really show this, except in case W, where it finds it quite disproportional that 0.5% of the vote generates 2% of the seats. This is also picked up by SDID, with 1.2 on a scale of 10.
(3) Case R: SDID is very sensitive to the case that 1% of the vote still is wasted. This gets 1 point on a scale of 10, or 10.1 on a scale of 100. The other scores also give 1, but WSL also for the infinite range that is hard to judge. Compare with case J in **Table 19** when 10% of the vote is wasted, and finds a SDID score of 3.3 on a scale of 10.
(4) There is the symmetry that wasting the vote in case R is judged as (almost) the same as in case T of having 1% of the vote and getting 2% of the seats. For now we can accept this symmetry. It might be a special subject of how to evaluate vote waste. Penalizing vote waste would impy a negative verdict for most current systems, while we might give that verdict anyhow, and still be interested in how a measure would impose symmetry upon cases that are essentially asymmetric.
(5) Similarly in cases U and V: 0.5% of the vote might either be wasted (0 seats) or be rewarded with 1 seat of 100 seats, and both cases are similarly disproportional.

12.4. Comparison of tripartite cases

Table 21 gives scores for theoretical cases for three parties. If one party has proportionality then again ALHID = EGID.

Table 21. Scores of measures on theoretical cases: 3 parties

	1	2	3	4	5	6	7	8	9	10	11	12	13
			10√Sin	Specif	100√Sin	100 Sin	Gini	ALHID	EGID	WSL	ALHIDT	EGIDT	WSLT
Ex.	V	S	10	10	100	100	100	100	100	inf	10	10	10
AA	48	2	5.3	2.0	53.2	28.3	20.6	17.0	14.7	17.1	4.1	3.8	2.9
	43	3											
	9	0											
AB	48	3	4.4	1.4	44.4	19.7	15.6	12.0	10.8	12.2	3.5	3.3	2.6
	43	2											
	9	0											
AC	50	0	-10.0	-9.9	-100.0	-100.0	99.0	99.0	85.7	9900.	-9.9	-9.3	-9.1
	49	0											
	1	100											
AD	50	50	-9.3	-7.0	-92.6	-85.7	73.5	49.0	49.0	2450.	-7.0	-7.0	-8.3
	49	0											
	1	50											
AE	50	60	4.3	1.3	43.2	18.7	10.4	10.0	9.5	4.7	3.2	3.1	1.8
	49	40											
	1	0											
AF	200	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	400	4											
	600	6											
AG	200	3	4.3	1.3	43.0	18.5	11.1	8.3	8.3	5.6	2.9	2.9	1.9
	400	4											
	600	5											

Comments are:

(1) Case AC shows the usefulness of the non-infinite range. Inequality / disproportionality cannot get very much worse, but the open range of WSL still might suggest this. EGID wastes much attention to even higher degrees of inequality / disproportionality than case AC while one cannot see much value in this.

(2) Case AD was considered by Taagepera & Grofman (2003:672) in search of a notion of "halfway deviation from proportionality". The SDID score indicates strong disproportionality. Their definition of the Gini differs from the one given in Colignatus (2017b). They also allow zero votes, and this is replaced here by 1.
(i) A SDID value of 5.4 on a scale of 10 was found in case Q in **Table 19**, and may be close to the answer to their question what halfway would be.
(ii) T&G write values of EGID and ALHID with a %-sign. ALHID can be interpreted as a % of relocated seats in a House with 100 seats, indeed. EGID however transforms. Though it also is in the 0-100 range the term "percentage" has no clear meaning.
(iii) For comparison, case AE gives a more proportional outcome, and SDID provides more guidance here than the other measures.

(3) Cases AA and AB have been taken from Hill (1997:8), who argues for Single Transferable Vote (STV) - rather than doubling the number of seats, based upon the observed disproportionality. He brings in the notion of affinity of parties for judging disproportionality. This reminds of Zand (2015), though Hill would rather want to see that voters and not researchers determine the affinities of parties.

> Even within strictly party voting, the first-preference measures are unsatisfactory. Consider a 5-seater constituency and several candidates from each of Right, Left and Far-left parties. Suppose that all voters vote first for all the candidates of their favoured parties, but Left and Far-left then put the other of those on the ends of their lists. If the first preferences are 48% Right, 43% Left, 9% Far-left, all the measures will say that 3, 2, 0 is a more proportional result than 2, 3, 0. Yet STV will elect 2, 3, 0 and that is the genuinely best result, because there were more left-wing than right-wing voters. There is no escape by comparing with final preferences, after redistribution, instead of first preferences. That is merely to claim that STV has done well by comparing it with itself. Our opponents may sometimes be dim, but I doubt whether they are dim enough to fall for that one.

(4) Cases AF and AG are also from Hill (1997:7), whose issue with correlation now would be answered with RTO and cosine:

> For example with votes of 200, 400 and 600 and the proportional 2, 4 and 6 seats we get a correlation of 1.0, but the non-proportional 3, 4 and 5 seats equally get 1.0 as those points also fall on a straight line. To get a suitable measure we also need to include the same numbers over again, but negated. Thus 200, 400, 600, −200, −400, −600 with 2, 4, 6, −2, −4, −6 gives a correlation of 1.0 as before, but 200, 400, 600, −200, −400, −600 with 3, 4, 5, −3, −4, −5 gives only 0.983 demonstrating a less good fit.

12.5. Comparison of quadripartite cases

Table 22 gives some cases with four parties. Comments are:

(1) Cases AAA and AAB show a marginal change in inequality / disproportionality given an already high basic level. ALHID is insensitive to the change. Overall it comes as more natural to attach meaning to SDID and Gini as opposed to the other scores.
(2) Case AAC shows the benefit of the sensitivity to majority reversal. Obviously, the sign can also be introduced for the other measures.
(3) ALHIDT, EGIDT and WSLT have been performing fairly well in these comparisons, but their sensitivity leaves to be desired, and SDID does fine. Compare the 10-seat differences in A, B, J, AAA and AAB.

Table 22. Scores of measures on theoretical cases: 4 parties

	1	2	3	4	5	6	7	8	9	10	11	12	13
			10√Sin	Specif	100√Sin	100 Sin	Gini	ALHID	EGID	WSL	ALHIDT	EGIDT	WSLT
Ex.	V	S	10	10	100	100	100	100	100	inf	10	10	10
AAA	25	30	4.4	1.4	44.3	19.6	10.0	10.0	7.1	4.0	3.2	2.7	1.7
	25	30											
	25	20											
	25	20											
AAB	25	35	4.9	1.7	48.8	23.8	12.5	10.0	8.7	6.0	3.2	2.9	2.0
	25	25											
	25	20											
	25	20											
AAC	33	33	-8.6	-5.7	-86.2	-74.2	54.8	33.0	33.0	1122.	-5.7	-5.7	-7.7
	33	33											
	33	0											
	1	34											

12.6. Comparison on the Dalton transfer

We can judge measures on their minimands, and scores on situations, and reversely judge situations on scores, but also look at the marginal change of taking one seat from a party and apportioning it to another. This is callled "Dalton's transfer". Taagepera & Grofman (2003:667) (T&G) give an example. When a seat is transferred from a richer party to a poorer one, the (unit) inequality / disproportionality should decrease. Conversely, an increase.

I am not comfortable with this principle, since apportionment would be restricted to the Floor-Ceiling locus, and it is not clear what we are actually doing when shifting a seat, as this might be *in any direction* from an optimum in the locus, either *to* it or *from* it, or encircling it somewhere at a distance. Presently we do not know about this locus since we use percentages and not a particular number of seats. The cosine is insensitive to scaling but there is still this notion of optimality. I would also prefer mathematics to get at a general statement. Indeed Goldenberg & Fisher (2017) (G&F) provide some relevant formulas on WSL and Dalton's principle.

Table 23 contains the T&G example. Case DA with votes {50, 40, 10} is disproportional with seats {60, 25, 15}. A shift of 1 seat does not necessarily mean an improvement. Case DA, and a transfer of 1 seat from 60 to 15, causes DB. Case DA, and a transfer of 1 seat from 15 to 60, causes DC. Thus we are looking at how disproportionalities compensate each other, rather than in looking at a step towards more proportionality. If we would used SDID and EGID for apportionment then they would follow a different path towards proportionality than if we would use Gini or WSL. Each, taking their steps in their own way, would reach the proportional outcome {50, 40, 10} too. Thus we are rather speaking about *different paths* rather than about comparing what "really" would be disproportionality.

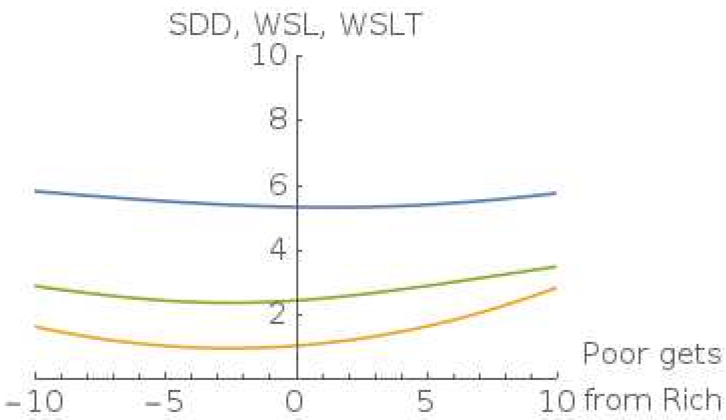
(i) SDID and EGID show differences only in the third digit.
(ii) T&G have a different definition of the Gini but the outcomes here are the same.
(iii) The order of disproportionality for SDID and EGID is **DC > DA > DB**. Gini and WSL have **DB > DA > DC**, i.e. reversed. Is DB or DC the least disproportional ?
(iv) The crucial distinction is that SDID and EGID are symmetric by design and the Gini and WSL are asymmetric by design. They have different objectives. Thus, the wish for a symmetric measure may now conflict with the point that apportionment is inherently asymmetric. But we need not be bothered by the asymmetry, for our present purposes.

Table 23. Taagepera & Grofman (2003:667) on Dalton's principle of transfer

	1	2	3	4	5	6	7	8	9	10	11	12	13
			10√Sin	Specif	100√Sin	100 Sin	Gini	ALHID	EGID	WSL	ALHIDT	EGIDT	WSLT
Ex.	V	S	10	10	100	100	100	100	100	inf	10	10	10
DA	50	60	5.29	2.00	52.9	28.0	16.5	15.0	13.2	10.1	3.9	3.6	2.4
	40	25											
	10	15											
DB	50	59	5.28	1.99	52.8	27.9	17.1	15.0	13.1	10.8	3.9	3.6	2.5
	40	25											
	10	16											
DC	50	61	5.31	2.02	53.1	28.2	15.9	15.0	13.5	9.6	3.9	3.7	2.4
	40	25											
	10	14											

Figure 23 gives the same case. We look at the amount *x* that the poor party receives from the rich party, in the range –10 to 10. For *x* = -10, the rich party receives 10 from the poor party. We find that SDID, WSL and WSLT are fairly flat, meaning that we are circling around proportionality at a distance, and are only looking how disproportionalities compensate each other. SDID has its minimum at around *x* = 1.5 and WSL and WSLT have their minimum at around –2.5. Thus the particular orderings of **DC > DA > DB** for SDID and **DB > DA > DC** for WSL derive from the small steps of 1 seat up or down the slopes of their particular minima. The slopes cannot be interpreted as giving a global gradient. It is remarkable that T&G selected an example that gives the local minimum of SDID at the point DB.

Figure 23. Dalton transfer {60 – x, 25, 15 + x} for {x, -10, 10} [64]
with SDID (blue), WSL / 10 (yellow) and WSLT (green)



Let us continue our comments:

(v) Thus, the Dalton transfer principle is not useful absolutely, but must take account of the gradients and the local minima. The principle is useful if the transfer is into the direction of proportionality indeed, and is less useful for circling around at a distance. This main point has been caught by **Figure 23**.

(vi) T&G explain the Gini outcome by looking at the *z* / *w* "advantage ratio". Their conclusion: "In the numerous cases where the same party (often the largest) has both

[64] Wolfram Alpha: Plot @@ {v = {.50, .40, .10};
s = {60 - x, 25, 15 + x} / 100; {10*(1 - (1 - CosineDistance[v, s])^2)^(1/4),
10*(wsl = v.(s/v - 1)^2), 10*(1 - 1/(1 + Sqrt[wsl]))}, {x, -10, 10}, AxesOrigin -> {0, 0}, PlotRange -> {{-10, 10}, {0, 10}}, AxesLabel -> {" Poor gets\n from Rich", "SDID, WSL, WSLT"}, BaseStyle -> {FontSize -> 14}}

the largest difference and also the largest ratio, of course, [EGID] and Gini pull in the same direction (...)". This is a curious statement since they are reversed.

(vii) T&G: "Yet the mind baulks at adding seats to a party that is already 50 percent overpaid [i.e. *C* with 10% votes and 15% seats] and calling it a reduction in disproportionality. Maybe this is the crucial difference between the notions of 'deviation from proportionality' (stressing the differences) and 'inequality' (stressing the ratio)." I hesitate on this, see the above. Yet it remains true that the present implementation of the Gini sorts on the "advantage ratio", and that the cumulation causes that disadvantaged values at the bottom of the sorted values affect the subsequent calculation with some multiplier. But this is not the only point. The other aspect is that the Gini is asymmetric.

(viii) For SDID, DB is the minimum of these three cases, and apparently SDID attaches value to the larger party and the larger difference between 60 and 50 indeed. One cannot hold that DA should be lower merely because it was the starting point for switching the seat. The relevant information is the opposing reaction of Gini and WSL. Yet the fact that these are asymmetric causes the comparison of apples and pears. Thus the distinction between level and ratio must be balanced by looking at asymmetry. I can concur with this: the Gini is not a measure on (unit) inequality / disproportionality but one of inequality as defined by the Gini. (Gini also looks at the party average and not the party marginal candidate.)

(ix) Overall, this example by T&G is enlightening on this comparison of Gini and SDID.

(x) SDID and Gini are useful for what they do, and the above opposing outcome, under the heading of an overly simplistic "transfer principle", does not disqualify either of them. When SDID defines what rich and poor are, then the ordering is **DC > DA > DB,** and if the Gini defines what rich and poor are, then the ordering is **DB > DA > DC.**

G&F rightly point out that T&G claimed that WSL satisfies Dalton's principle (in levels) while it might sometimes do and sometimes not do.

(xi) G&F point to the UK general (half) election 2010 as a counterexample, while they might also have refered to this very example itself that T&G provided.

(xii) There is also symmetry, and not just the distinction between levels and ratios.

(xiii) In the example of **Table 23**, case DB gives the smallest party relatively more seats. SDID judges this as the least disproportional given the situation on the other parties, while WSL judges it as most disproportional. It is often said that WSL has greater sensitivity for smaller parties, but this thus is not true. Case W in **Table 20** shows also the intolerance of overrepresentation by small parties.

13. Scoring on the Taagepera – Grofman criteria

Taagepera & Grofman (2003) (T&G) give some criteria for (unit) inequality / disproportionality measures, and score the more traditional measures on them. Let us check how SDID does. We already met some criteria like "being a metric" in **Section 4.3**. T&G do not explicitly say so, but they have similar conditions, and if the Dalton transfer can be translated into triangularity, then they require a metric. It is useful to run through the T&G list afresh.

(1) Yes. Informationally complete: makes use of the v and s data for all parties. (This excludes measures that take only the two largest parties.)
(2) Yes. Uniform: uses the data uniformly for all parties.
(3) Yes. Symmetry.
(4) Yes. Range. (Nonnegative.) Varies between 0 and 1 (or 100 percent). (i) The amendment is to prefer [0-10] though. (ii) The amendment is to allow for the negative sign on majority switches.
(5) Yes. Zero base. Has value 0 if $s = v$. (The comment though is that commonly s has a zero because of the wasted vote, and we avoid cases with a 0 in v. We have $s = v$ only in theoretical cases or in comparing s and Ve in **Table 16**.)
(6) Yes. Has value 1 or 100 percent for full disproportionality, when a positive vote matches with 0 (excluding the wasted vote) or conversely. (See their example of the Prince.)

(7) There is a fair effect on Dalton's transfer, see **Section** 12.6. The answer is that symmetry is more important than considerations on ratios, that different measures have different paths towards proportionality, and that the proper comparison should be into the direction of proportionality indeed rather than in comparing how disproportionalities compensate each other.

(8) Yes. Does not include the number of parties. This is a curious criterion since the vectors have a length that gives the number of parties. Yet, we can imagine that there are some practical considerations involved, that cause a desire to not be explicit about this, like the degrees of freedom in the Chi square distribution.

(9) Yes. Insensitive to lumping of "residuals" (individual parties in the wasted vote). This lumping is both a practical consideration and one of fundamental principle, e.g. on the wasted vote that is collected into "Other". (i) T&G claim that the Gini is sensitive to lumping, but they use a different definition, and the Gini in Golignatus (2017b) is insensitive to lumping. (ii) Taagepera & Grofman (2003:668) give the example reproduced in **Table 24**. Cases XA and XB differ because only XB has a wasted vote, so that SDID produces a higher score, and the Gini is quite sensitive to this. Normally, EGID is sensitive to lumping, but in this case the percentage difference in XA for small parties of size 5 is zero, and in XB the difference 1 - 0.1 = 0.9 becomes even smaller when squared. Thus the example is not as strong as can be made. Case XC removes the wasted vote, and SDID and Gini return to the values in XA. (As a standard, it remains advisable to record candidates that have at least 0.5 *Q* of the vote, and to lump the remaining wasted, blanc and invalid votes into one category.)

(10) Yes. Is simple to compute. The notions of cosine and sine are highschool issues, see **Appendix D**. For vectors there is a well established theory of at least some 100 years. The calculation can be done in a simple spreadsheet. Taagepera & Grofman (2003:670) argue that "[EGID] is difficult to calculate", and give it a score of 0.5 on simplicity. It is hard to believe that they are really serious on this. Their yardstick is ALHID of course.

(11) Yes. Is insensitive to a shift from fractional to percent shares. It is somewhat remarkable that this needs to be mentioned. (Some attention to scale adjustment is only fair.)

(12) Yes. Input depends only upon shares $w = v / V$ and $z = s / S$ without use of V and S. (However, researchers would tend to require full data, also for the relation to apportionment. Thus the criterion is practical only in a limited sense. For example, the UK with 650 seats has a greater potentiality to get a proportional fit than Holland with 150 seats. (Check the value for EPR+ in **Table 8**.) Also, it is better to store the data in the form of integers *v* and *s* rather than in fractional form (with at least 6 digits precision too).)

Table 24. Taagepera & Grofman (2003:668) on lumping small parties in "Other"

	1	2	3	4	5	6	7	8	9	10	11	12	13
			10√Sin	Specif	100√Sin	100 Sin	Gini	ALHID	EGID	WSL	ALHIDT	EGIDT	WSLT
Ex.	V	S	10	10	100	100	100	100	100	inf	10	10	10
XA	400	50	4.8	1.6	47.9	22.9	14.0	10.0	10.0	7.5	3.2	3.2	2.1
	300	30											
	200	10											
	50	5											
	50	5											
XB	400	50	5.3	2.0	53.3	28.5	29.1	19.0	12.0	97.5	4.4	3.5	5.0
	300	30											
	200	10											
	90	0											
	1	1											
	1	1											
	1	1											
	1	1											
	1	1											
	1	1											
	1	1											
	1	1											

	1	1											
	1	1											
XC	400	50	4.8	1.6	48.0	23.0	14.0	10.0	10.0	7.5	3.2	3.2	2.1
	300	30											
	200	10											
	10	1											
	10	1											
	10	1											
	10	1											
	10	1											
	10	1											
	10	1											
	10	1											
	10	1											
	10	1											

To summarise:

- Overall, SDID scores perfectly well on these criteria, except that the difference transfer hasn't been worked out mathematically yet, unless it suffices that SDID is a metric.
- Overall, the Taagepera & Grofman (2003) criteria are mostly technical rather than criteria on content. The difference transfer and lumping are the most important on content. The difference transfer was important to highlight the distinction between symmetric SDID and asymmetric WSL and Gini. The example given suffers from the issue that the transfer wasn't into the direction of proportionality but rather encircling it, while a transfer towards proportionality requires a definition, either from EGID and SDID or from WSL and Gini. There would be no discussion about proportionality in the locus of apportionment with remainders zero.
- Remarkably, the regression coefficient, as a criterion at T&G, recieves an overall low score. When one wonders why this is so, then the chain of reasoning and the elimination of the deficiencies can lead to the creation of the SDID.
- We tend to agree with their conclusion p673: "Favouring some other index for any one of these topics would need justification to counterbalance the shortcomings listed here."

14. Conclusion

As stated in the Introduction: Our present purpose is to (i) deconstruct the discussion in Political Science about inequality / disproportionality measures, (ii) show what the problem really is, (iii) present a clear and simple inequality / disproportionality measure (though with options to make it more complex again).

The problem really is: whether we apportion seats for *The Party as unit of account* or for *representatives* or *party marginal candidates as the unit of account*.

Hamilton's Largest Remainder (HLR) method was apparently intended for the application to States, but basically it is a method for apportionment of *representatives*. **Appendix L** contains an adjustment that gives the *Representative Largest Remainder* (RLR) apportionment method. Candidates below 0.5 of the natural quota are fully excluded and those in the 0.5 – 1 range get a chance for partaking in a round of allocation of remainder seats. This actually *defines* what *equal / proportional representation* means. There is variety, of course, say on the threshold, but the principle is clear. The notion of Representative Democracy is also much wider than only apportionment, and each democratic nation has its own definition. Yet on the notion on *equal / proportional representation* this definition of HLR / RLR is possible, without such variety of opinion. If one doesn't want *equal / proportional representation* but something else, like *Equal Party Representation*, then one can adopt another definition, and another apportionment, like WSL that includes party weights. But one should not tell voters that one has *equal / proportional representation* when one actually has something else. One shouldn't sell chalk for cheese.

The *sine diagonal inequality / disproportionality* (SDID) measure defined in **Section 4.1** is rather simple. It is SDID[*v, s*] = *sign* 10 √ Sin[*v, s*]. Perhaps it might look a bit complex but then this is only because of the surrounding zoo of indices both in voting theory and statistics. It is remarkable that SDID hasn't been formulated early at the start when researchers started looking into electoral systems in a quantified approach. (One thinks about Sainte-Lague (1882-1950) and Hill (1860-1938) and Huntington (1874-1952), though the latter authors focused on seats per State rather than seats per party and per party marginal candidate. [65])

In this paper, SDID and RLR were developed independently, on their own considerations. A bit to my surprise, SDID fairly well recognises HLR / RLR outcomes as unit proportional. The difference between HLR and RLR appears in **Appendix N**, on the New State paradox. I arrived at both approaches in parallel and not in deliberate united design. In reflection, the cause of overlap is clear. The fractional remainder is commonly interpreted as an absolute difference. For the party marginal candidate, however, it is a relative measure, with the marginal candidate in the denominator. Nevertheless, this caused **Section 8** on the preference of SDID w.r.t. HLR / RLR. This preference or bias can be observed because of the relation of the sine to the plain Euclidean distance (with a separate influence of the slope).

As an econometrician who has been familiar with the relationship of cosine and correlation since the first year linear algebra course, I found the literature in voting theory about inequality / disproportionality measures bewildering. It comes across as if these researchers try to reinvent statistics without awareness of the wealth of statistical theory. It already starts with the confusing use of the word "proportionality", originally correctly for factor S / V for each seat separately, but now for the unit simplex as only *unit* or *diagonal* proportionality. In that manner, voting researchers also create a barrier for statisticians who want to understand what their problem is. It should help to explain at the very start, as I have done now in this paper:

- The notion of unit proportionality as in the line $y = 1\ x + 0$ is a mathematical concept, but in statistics with dimensions we can always rebase the variables, so that there need be no natural base for 1. For voting and seats there however are natural bases in persons (voters and representatives).
- The use of the unit simplex is a relevant tool but we must also be aware of the creation of hidden parameters V and S, that pop up at other parts of the analysis (e.g. for apportionment that requires these, or for the relation to turnout (over different years)).
- Any vector-proportional relation $s = T\ v$ also holds for the totals, and reduces in the unit simplex to diagonal $z = w$ because $S = 1's = T\ 1'v = T\ V$. Thus $z = s\ /\ S = T\ v\ /\ S = w$.
- For vector-proportionality, deviations from $z = w$ are both $z = w + \tilde{e}$ and $z = b\ w + e$.
- There is a difference between the observations $s = T\ v$ and the true values $s^* = T^*\ v^*$ (potentially with errors in variables) for some external or institutionally given T^*. For votes and seats $s^* = T^*\ v^*$ however might be unattainable because of integer values.
- The proper tool to use here is regression through the origin (RTO), with the coefficient of determination $R^2 = Cos^2$.
- We can distinguish *unit* or *diagonal* equality / proportionality when $b = 1$ without dispersion when Cos = 1, and *average unit* proportionality when $b = 1$ with dispersion when Cos ≠ 1 (and such an average would not count as *equality*).

Voting research has a wealth of insights and its texts have reason and rationality. However, when a text presents a conclusion that such and such measure M behaves such and such, then this may be true, but it becomes irrelevant once it is (already) clear from statistics that this measure M can be discarded. **Section** 12 above has a Byzantine structure of comparing the proposed SDID with inadequate measures, and the Section is rather superfluous once one understands what SDID does. Yet for voting researchers the Section will be highly relevant because it relates to how they understand the world. Let me also mention that Colignatus (2010) fell victim to the same bias for The Party and preference for WSL (though calls attention to qualified majority to keep account of the wasted vote).

[65] http://www-history.mcs.st-andrews.ac.uk/Biographies/Huntington.html

Common measures in voting theory are ALHID (absolute differences), EGID (Euclidean distance) and WSL (ratio deviation from 1, weighted). Originally I hesitated to try to make transforms of these, namely for a "fair" comparison with SDID. With transforms, we don't really compare these measures but only some proposed transform. Eventually, curiosity and the desire for systematic treatment took over, and there are now the suggestions for ALHIDT, EGIDT and WSLT, with T for transform. We find that these transforms are more like SDID and then more satisfactory. The transforms may explain better what is wrong with the original measures.

Colignatus (2018a) extends the analysis with the Aitchison distance for compositional data, in comparison with the angular distance. SDID appears to be better than both too.

Looking back, we can appreciate ALHID, because it measures the dislocated seats in a House with 100 seats, and this is a useful statistic. The drawback is its lack of sensitivity. Transforming what is fundamentally insensitive maintains this insensitivity for much of the intended application. EGID has some relevance, because we have shown the relation of the sine to the Euclidean distance. WSL is relevant, because it highlights the confusion between The Party and the party marginal candidate.

A key point for modesty remains that voting research comes from the tradition exemplified by Jefferson (1743-1826), Hamilton (1755-1804) and Webster (1782-1852), that votes have to be apportioned into seats, and that the choice is not self-evident. It is important that this tradition exists, and we can only respect that it apparently runs its own course with its own traditions, like statistics itself has its traditions. Alexander Hamilton provided the proper method of largest remainders. I had to eliminate some confusions in the Political Science literature before it was clear to me that these were confusions indeed. The key point remains that Hamilton already had a key suggestion.

The advice is to use *Representative Largest Remainder* (RLR) for apportionment in large national Houses of Parliament (see **Appendix L**) and SDID for inequality / disproportionality measurement alongside the quite useful ALHID for the number of dislocated seats. (One should call a measure for what it does and not for who created it (perhaps with another purpose).)

When using mathematics, it is important to be aware that mathematicians have been trained for abstraction, while applications to the real world require a training on empirical science (including adequate mathematics). The only solution is good math and good science. If one of these conditions fails then there will quickly arise confusion. Mathematician Kenneth Arrow's confusion on the interpretation of his Impossibility Theorem – see Colignatus (2014) on the distinction between voting counts and decisions – is another example how research might get lost on fundamental ideas. Or **Appendix D** is sobering on the education of mathematics and its research for the last 5000 years. Fortunately, voting in a free world may help prevent that tradition becomes dogma.

15. Appendix A. Abstract of Colignatus (2017b)

QUOTE

The Lorenz curve and Gini coefficient are applied here to measure and graph inequality / disproportionality in outcomes for multiseat elections held in 2017. The discussion compares Proportional Representation (EPR) in Holland (EPR Gini 3.6%) with District Representation (DR) in France (41.6%), UK (15.6%) and Northern Ireland (NI) (36.7%). In France the first preferences of voters for political parties show from the first round in the two rounds run-off election. In UK and NI the first preferences of voters are masked because of strategic voting in the single round First Past the Post system. Thus the Gini values for UK and NI must be treated with caution. Some statements in the voting literature hold that the Lorenz and Gini statistics are complex to construct and calculate for voting. Instead, it appears that the application is actually straightforward. These statistics appear to enlighten the difference between EPR and DR, and they highlight the inequality / disproportionality in the latter. Two conditions are advised to enhance the usefulness of the statistics and the comparability of results: (1) Order the political parties on the ratio (rather than the difference) of the share of seats to the share of votes, (2) Use turnout as the denominator for the shares, and thus include the invalid and wasted vote (no seats received) as a party of their own. The discussion also touches upon the consequences of inequality / disproportionality by DR. Quite likely Brexit derives from the UK system of DR and the discontent about (mis-) representation. Likely voting theorists from countries with DR have a bias towards DR and they are less familiar with the better democratic qualities of EPR.

UNQUOTE

PM. These references are not in the article: Taagepera & Laakso (2006) and Laakso & Taagepera (2007) plot in the $\{v, s / v\}$ space, calling this a "proportionality profile". This is close to using the Gini.

16. Appendix B. Some comments on the unit simplex

A relevant text is Borg & Groenen (2005) but I did not get to look into this.

16.1. EGID and Origin Concordance (OC)

See also **Section** 11.6 on the relation of Euclid / Gallagher inequality / disproportionality and the cosine. There we use a different notation. See **Figure 10** for the unit simplex.

The Gallagher index has the base in Euclid Sum$[(z - w)^2] = (z - w)' (z - w) = z'z - 2z'w + w'w$. Given that this is nonnegative, we have $2z'w \leq z'z + w'w$.

For voting only Quadrant I with nonnegative vectors is relevant. There we can define "Origin concordance" and its implied distance as:

$$0 \leq \text{OC} = z'w \,/\, (½\,(z'z + w'w)) \leq 1$$

$$0 \leq \text{OCDistance} = 1 - \text{OC} \leq 1$$

- The OC and OCDistance are symmetric, with the proper range, and OCDistance = 0 when $z = w$. I have not checked whether OCDistance is a proper metric though.
- Cosine and OC take variables from the origin while Correlation and Lin's concordance (LC) (**Section** 10.3) [66] take them around their means.
- The Cosine has a geometric average in the denominator and OC an arithmethic.

[66] https://en.wikipedia.org/wiki/Concordance_correlation_coefficient

Looking at the other quadrants, we find that the lower zero bound is also met in Quadrant III, when the negative signs cancel each other. In Quadrants II and IV the minimal value –1 can arise, for example for w = {-1, 0} and z = {1, 0}. (This is not an example for voting.) For general {x, y} on the unit simplex we have:

$$-1 \leq \text{ OC} = x'y \,/\, (\tfrac{1}{2}\,(x'x + y'y)) \;\leq 1$$

It is a bit tedious to figure out how to translate this into a distance. For nonnegative OC the transform 1 – OC already was fine. What to say about a value OC < 0 ? For w = {- ½, ½} in Quadrant II and z = {½, -½} in Quadrant IV, we get (-¼ + -¼) / (½ ((¼ + ¼) + (¼ + ¼))) = -1. Thus we have a directed distance:

$$-1 \leq \text{ OCDistance} = \text{If[OC} < 0\text{, OC, } 1 - \text{OC]} \;\leq 1$$

I have not looked into this further (for the potential use of a (unit) inequality / disproportionality measure), because: (i) the similarity of arithmetic and geometric averages suggests that it would not matter so much, (ii) the cosine is already well established, and I stumbled upon this OC while writing this and have not looked whether there already is some literature on it (perhaps under some other name) (and I might have plainly forgotten about it, simply because it wasn't so relevant until now) (there is already a relation to cosine and the 2D form of the harmonic mean reminds of it).

16.2. Aitchison geometry

Pawlowsky-Glahn et al. (2007) provide an accessible and enlightening overview for the statistical analysis of "compositional data" and the Aitchison geometry. The "compositional data" are the vectors in the unit simplex, a.k.a. relative information, parts per unit, proportions, percentages, and so on. The issue may be more complex than for electoral systems, and for example contains the problem when researcher *A* measures a sample compound on four variables and researcher *B* does this one month later for the same sample on three variables: so that it can appear that a positive correlation in the first study turns into a negative correlation in the second study. (My impression is that this is a question on content and not on geometry.)

The Aitchison geometry subsequently uses positive vectors and logarithmic transforms (with values [0, infinity], also called the log-ratio approach. This is not useful for voting in which the wasted vote has 0 seats. The log-ratio compares to WSL that we found has drawbacks.

This is further discussed in Colignatus (2018a), see also **Appendix O** below. See there too for more on the angular distance.

17. Appendix C. Using Abs[k – 1 / k] and another transform

The following continues the discussion of **Section** 10.7.

First observe that if k is the cosine for the nonnegative quadrant that we are looking at, then Tan = Sin / Cos = Sqrt[1 – k^2] / k = Sqrt[1/k – k] / Sqrt[k]. The denominator might not always be needed. The cosine remains interesting because it already is in the range [0, 1] while other measures are not, and then require a transform.

Also if we use $k = b / p$, then we may consider using d = Abs[k – 1 / k] as a symmetric expression of deviation around slope 1.

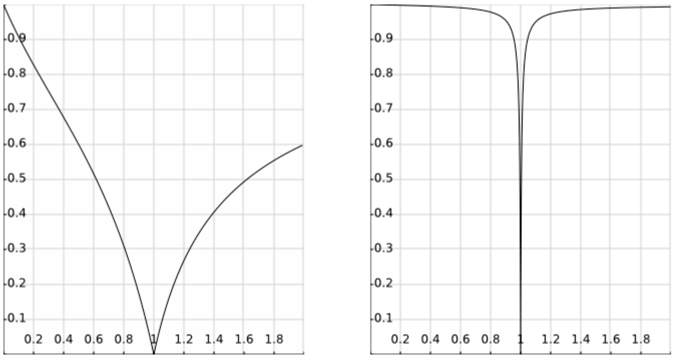
For $k > 0$ we have Abs[1 – k^2] = Abs[k – 1 / k] k and this will work in the same manner as Sqrt[1 – k^2]. Only for Sin it is established that it is a metric.

For a transform of this d we can use a stable function with a fixed parameter f as a norm for sensitivity, $f\,d$ / (1 + $f\,d$), which gives ID[k] = 1 – 1 / (1 + f * Abs[k – 1 / k]). Relevant values for f are 1, 10, 100 and 1000. Figure 24 shows *disp* for d = Abs[k – 1 / k] for values of $0 \leq k \leq 2$

and factors f = 1 and 100. [67] An increasing sensitivity will quicker decide upon higher disproportionality.

NB. This approach assumed $k = b / p$, while the body of the text converged on selecting k = *cosine*[*v, s*], which has only the range [0, 1]. At first it is somewhat remarkable that the graphs for this selection and for the selection of the cosine and sine are quite similar, but the logic behind both approaches is also quite similar.

Figure 24. ID[***k***] **= 1 – 1 / (1 +** ***f*** *** Abs[*k* – 1 / *k*]), 0 ≤ *k* ≤ 2, for *f* = 1 or 100**



18. Appendix D. Xur & Yur

Colignatus (2009, 2015) and (2011) show how sine and cosine can be re-engineered. Figure 25 contains two figures taken from there. The text uses H = -1, so that $x^H = 1 / x$.

- The unit of angular measurement is the whole plane itself.
- Thus angles have a [0, 1] domain on the unit circumference circle. This is also called the "unit measure around" or "unit meter around" (UMA). This compares to 360 degrees (artificial, number of days in a year) or 2π radians (unit radius circle).
- Sine and cosine can also be discussed in terms of xur and yur.
  - They derive from both a ratio and the unit radius circle, and their name can use "ur"
  - The cosine gives the horizontal value or *x,* and thus becomes *xur*
  - The sine gives the vertical value or *y*, and thus becomes *yur*
  - By definition xur^2 + yur^2 = 1
- The constant π derives from the historical interest in the diameter of a circle for practical building with trees, while current theory shows the importance of the radius. Then Θ = 2 π is the more relevant parameter. This is written as capital theta but pronounced as "archi" (from Archimedes).
- Thus xur[α] = cos[α Θ] and yur[α] = sin[α Θ].

For writing this article, I wondered whether I should use xur and yur here too, rather than cosine and sine. However, the re-engineering of mathematics education differs from the re-engineering of voting theory. It remains useful to explain to voting researchers that the use of cosine and sine is not as complicated at they might think.

[67] Created in: http://fooplot.com

Figure 25. Taken from Colignatus (2009, 2015:65), with its labels Figure 4 and 5

Figure 4. Angular circle ($r = \Theta^H$), unit circle (r = 1), x = Xur and y = Yur

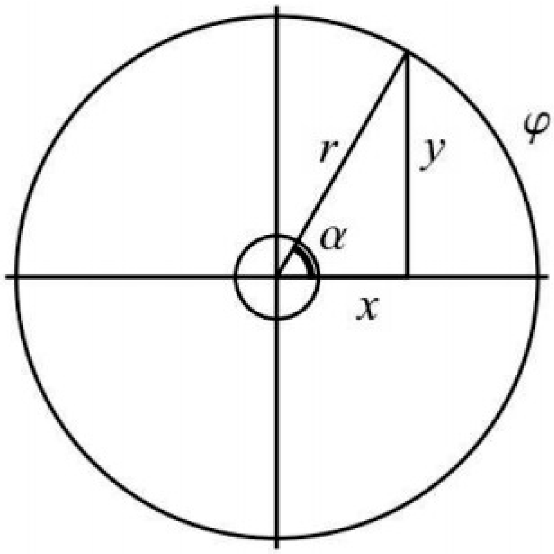
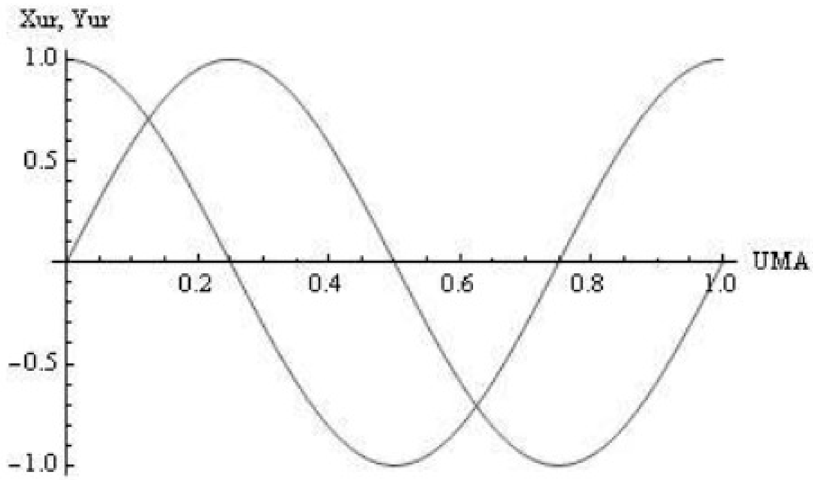


Figure 5. The functional graphs of Xur and Yur



19. Appendix E. The average

The unweighed (RTO) regression z / w = b 1 gives b = Average[z / w]. This can also be seen as the result of regressing z = b w with weights 1 / w, that favour the small parties. [68]

Subsequently, we consider a symmetric expression like A = Abs[Average[z / w] – Average[w^+ / z^+]] in which the division by 0 in the latter denominator is simply removed as a case. The measure looks interesting, as it also has an appeal of junior highschool simplicity.

If w = z (and no wasted vote) then A = 0.

[68] https://stats.stackexchange.com/questions/54794/regression-through-the-origin

However, it is not clear why a inequality / disproportionality above 1 could compensate a inequality / disproportionality below 1, which is what the average does.

The wasted vote causes a major problem too. For votes {99, 1} and seats {100, 0}, the measure becomes Abs[100 / 99 / 2 – 99 / 100 / 1] = 0.485 because the first average divides by length 2 and the second by length 1. We may consider multiplying *A* by 2.

The greatest inequality / disproportionality would arise when one party gets all seats. Consider two parties, a range of votes of {*x*, 100 – *x*} and all seats apportioned as {100, 0}. Then the first average gives 100 / *x* / 2 because there is vector of length 2 with one zero. The second average neglects the 0 and divides by 1. Thus:

A[{*x*, 100 - *x*}, {100, 0}] = Abs[100 / *x* / 2 – *x* / 100 / 1] = Abs[50 / *x* – *x* / 100]

This is large for a small vote (*x* = 1%), becomes zero for *x* = 50 Sqrt[2] and, as said, has a value close to ½ for *x* = 100. It is not clear why the zero for *x* = 50 Sqrt[2] would be the perfect outcome (while it isn't).

It appears to be rather complicated to turn *A* or a variant of it into a disproportionality measure. Since there is already a clear measure based upon the angle between the vectors, there doesn't seem to be a need to pursue this line of reasoning on the average.

20. Appendix F. Potential use for apportionment, Holland 2017

Table 25 gives the potential use of SDID for apportionment, applied to the Dutch election for the 2[nd] Chamber of Parliament in 2017. The scores of the various measures for inequality / disproportionality have already been given in **Table 8**, and see the discussion there. The present table is mainly relevant for the effect on individual parties. The wasted vote will be represented by 3 empty seats or a qualified majority. Large parties lose seats to some smaller ones that were disadvantaged by the use of the method of D'Hondt. This doesn't mean that SDID would be a useful method of apportionment, because the same effect might also be gotten by allowing empty seats and using HLR. SDID only helps to highlight the choice.

Table 25. Potential use of SDID for apportionment, example Holland 2017

	1	2	3	4	5	6	7	8
	Votes	Seats	SDID-Seats	Difference	%Votes	%Seats	%SDID-S	Difference
	10,563,456	150	150	0	100	100	100	0.0
Party								
Other	208,742	0	3	3	2.0	0.0	2.0	2.0
VVD	2,238,351	33	32	-1	21.2	22.0	21.3	-0.7
PVV	1,372,941	20	19	-1	13.0	13.3	12.7	-0.7
CDA	1,301,796	19	18	-1	12.3	12.7	12.0	-0.7
D66	1,285,819	19	18	-1	12.2	12.7	12.0	-0.7
GL	959,600	14	14		9.1	9.3	9.3	
SP	955,633	14	14		9.0	9.3	9.3	
PvdA	599,699	9	8	-1	5.7	6.0	5.3	-0.7
CU	356,271	5	5		3.4	3.3	3.3	
PvdD	335,214	5	5		3.2	3.3	3.3	
50Plus	327,131	4	5	1	3.1	2.7	3.3	0.7
SGP	218,950	3	3		2.1	2.0	2.0	
DENK	216,147	3	3		2.0	2.0	2.0	
FvD	187,162	2	3	1	1.8	1.3	2.0	0.7

21. Appendix G. Normalisation of T, b and p on unit circle

We distinguish Nonspecific disproportionality, for which we have the Sin, and specific inequality / disproportionality around the slope. The body of the text uses angles to define the specific inequality / disproportionality as well. An alternative is to use slopes. Thus, with the target slope T, we might work with T directly rather than first transform into Φ = ArcTan[T].

A problem of working with T directly is that for $T = 0$ and proper $b = 0$, symmetry requires us to consider $1 / b$. Notice that it is okay to have $T = 0$ as the target. We are not speaking about symmetry *of* the target, but of symmetry *around* the target. For a symmetric test, we need not only the slope specific error SPE[T] = $(T - b)^2 + (T - 1 / p)^2$ but also the equivalent value when the axes are flipped; SPE2[$1/T$] = $(1/T - p)^2 + (1/T - 1 / b)^2$. We also want to allow $T = 0$, and then the latter could not be used while it ought to be used. Symmetry around the target also requires us to consider the deviation $1 / T$ to $1 / b$. However we can solve this as follows.

A proportional relationship is best described by the line $\lambda\, y + \mu\, x = 0$, which coefficients may be normalised on the unit circle. For a target $y = T\, x$ and estimates $y = b\, x$ and $x = p\, y$, we better transform to that normalised format.

When we have a general expresion $a\, y + b\, x = 0$ then we normalise on the unit circle by taking $\lambda = a /$ Sqrt[$a^2 + b^2$] and $\mu = b /$ Sqrt[$a^2 + b^2$]. Indirectly we still rely on the angles in the circle. If τ = Tan[θ], then k = Cos[θ] = 1 / Sqrt[$1 + \tau^2$].

The first thing is to show the symmetry. Does $x = p\, y$ give the same result as $y = x / p$? Since we have a regression through the origin of nonnegative vectors, the values b and p are nonnegative too. Then rows 3 and 4 in **Table 26** show that the normalisation generates the same formats for the greeks on the unit circle. The normalisation shows that we need not fear for values like $T = 0$, $b = 0$ or $p = 0$.

Table 26. Formats of the line

Row	*Conventional*	*Parametric*	λ	μ
1	$y = T\, x$	$y - T\, x = 0$	1 / Sqrt[$1 + T^2$]	$-T$ / Sqrt[$1 + T^2$]
2	$y = b\, x$	$y - b\, x = 0$	1 / Sqrt[$1 + b^2$]	$-b$ / Sqrt[$1 + b^2$]
3	$x = p\, y$	$p\, y - x = 0$	p / Sqrt[$1 + p^2$]	-1 / Sqrt[$1 + p^2$]
4	$y = x / p$	$y - x / p = 0$	1 / Sqrt[$1 + 1 / p^2$]	($-1 / p$) / Sqrt[$1 + 1 / p^2$]
			p / Sqrt[$1 + p^2$]	-1 / Sqrt[$1 + p^2$]

Thus we can define the *normalised slope specific error* (NSPE) as in **Table 27**. Now we meet another problem. The specific inequality / disproportionality here uses the sum of squares on the slopes, and we want to combine this with the Nonspecific inequality / disproportionality that uses Sin. These are of different sizes. We can transform again, and calibrate on the Dutch general election in 2017. NSPE normalises and retains the Euclidean metric of a sum of squares, and adds and multiplies with metrix Sin and then transforms properly.Thus Abs[DSD[v, s]] would be a metric. Nevertheless, this approach has too much arbitrariness in the choice of the transforms. The approach with angles in the main body of the text has some arbitrariness too, but arbitrarily less.

This appendix only reports on an approach considered while writing this paper, but the approach has been rejected because of the better findings in the main body of the text. It still gives a perspective on slopes, angles and cos and sine.

Table 27. Working on slopes rather than angles

T	target slope, for votes and seats: $T = 1$
$\text{SPE}[T] = (T - b)^2 + (T - 1 / p)^2$	slope specific error, in the [0, infinity) range Addition for symmetry of nonzero values
$\text{SPE2}[1/T] = (1/T - p)^2 + (1/T - 1 / b)^2$	Flipping the axes for symmetry Doesn't work for $T = 0$ or $b = 0$
NSPE[*i*]	$(\lambda_1 - \lambda_i)^2 + (\mu_1 - \mu_i)^2$ for $i = 2, 3$
sum	NSPE[2] + NSPE[3] for $T = 0$, $b = 0$ or $p = 0$
ζ ζ' = SpecificSlopeID[*v, s, T*]	$1 - 1 / (1 + g * sum)$ default $g = 10$ sign * 10 * ζ
η = NonspecificID[*v, s*]	from main body of the text (without *f*)
d = TotalCriterion[*v, s, T*]	$1 - (1 - \eta)(1 - \zeta)$ Squares of small NSPE are dominated by Sin Calibration of *g* on Holland 2017 (50 - 50) ?
TotalSlopeID[*v, s, T*]	*sign* $(100\, d)^{1/f}$
DSD[*v,* s]	TotalSlopeID[*v, s,* 1] take diagonal $T = 1$

22. *Appendix H.* Cos[z / w, 1] *allows undesirable compensation*

Let us create a test measure as Test[*w, z*] = (1 – Cos[*z* / *w*, 1]^2)^(1/4), similar to SDID but now Cos[*w, z*] replaced by the Webster / Sainte-Laguë idea of comparing *w* / *z* to 1.

Figure 26 takes seats {2, 8} and shows that the two measures work alike. The allocation of 2 seats to the first party is only proportional if it has 20% of the vote.

For three parties, we distinguish between whether the third party is proportional or disproportional. If the third party is proportional, then the allocation of the other two is like before. If the third party is disproportional, then this creates possible disproportionalities in the other two parties. **Figure 27** takes seats {1, 1, 8}, while party *vb* = 1% has a disproportional allocation.

Figure 26. Votes {*va,* 1 – *va*} and seats {2, 8}, with SDID (blue) and Test (yellow) [69]

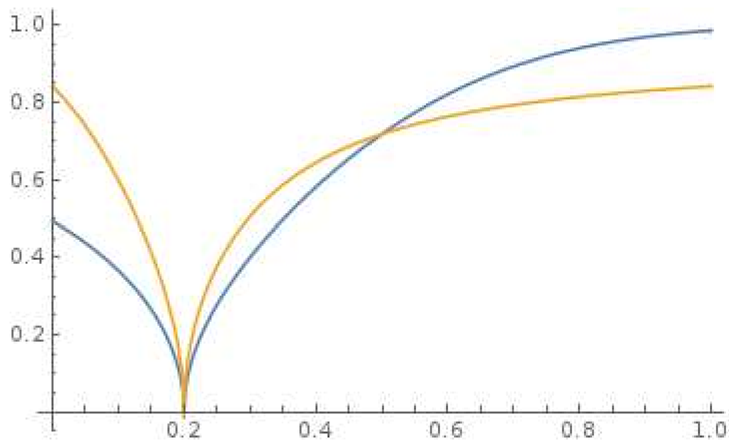
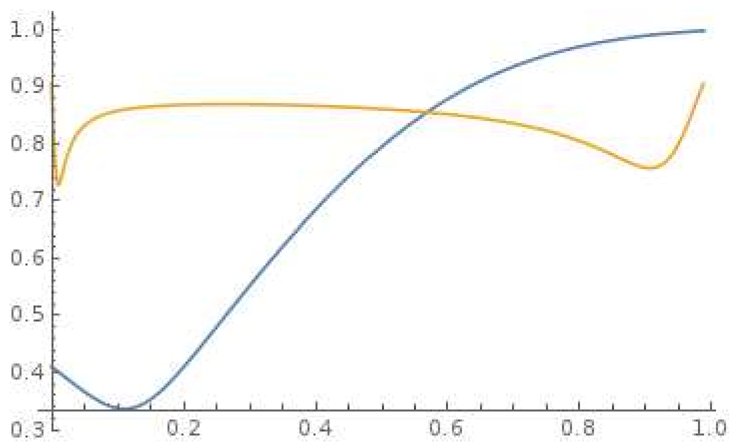


Figure 27. Votes {*va,* 0.01, 0.99 – *va*}, seats {1, 1, 8}, with SDID (blue) and Test (yellow) [70]



- SDID finds that the allocation of 1 seat to *sa* gives minimal inequality / disproportionality if *va* = 0.11. This comes at the cost of a disproportionality of *sc* / *vc* = 8 / (0.99 - 0.11).
- Test has two local minima, one at votes {1, 1, 98) and one at {91, 1, 8}, again given that *vb* = 1. Thus the cosine weights on *z* / *w* allow a balance.
- SDID accepts the inequality / disproportionality for *vb* and *sb,* and concentrates on the proportionality for *va* at 11%.
- WSL itself does not allow this particular balancing of errors. For {1, 1, 98} WSL gives a score around 1.5 and {91, 1, 8} gets a score around 8.

Thus there is an argument to use WSL rather than SDID but there is no argument about using Test (a relative form of SDID) rather than SDID.

[69] Wolfram Alpha: Plot[{ (1 - (1 - CosineDistance[{va, 1 - va}, {2, 8}])^2)^(1 / 4), (1 - (1 - CosineDistance[{2, 8} /{va, 1 - va}, {1, 1}])^2)^(1 / 4)}, {va, 0, 1}]

[70] Wolfram Alpha: Plot[{ (1 - (1 - CosineDistance[{va, .01, .99 - va}, {1, 1, 8}])^2)^(1 / 4), (1 - (1 - CosineDistance[{1, 1, 8} /{va, .01, .99 - va}, {1, 1, 1}])^2)^(1 / 4)}, {va, 0, .99}]

23. Appendix I. Major revision

The third version of the paper with 100 pages is a major revision of the former Colignatus (2017c) (a 2nd version with 40 pages, correcting edit errors). The innovation is:

(1) Name change to "sine diagonal disproportionality" (SDD) from "slope diagonal deviation" (SDD) (same content). (In this 4th version (**Appendix O**) this becomes SDID.)
(2) Focus on the principle of Methodological Individualism, and thus the party marginal candidate instead of the party average candidate.
(3) Formulation of a fitting apportionment method, "Representative Largest Remainder" (RLR), see **Appendix L**.
(4) Relation of Sin to EGID and slope, and thus its similar support for HLR / RLR, **Section** 8.
(5) Highlighting that Cos and Sin have the double face of both general measures and the interpretation for slopes.
(6) Better structure for the distinction between common space and the unit simplex.
(7) Better structure for the different error models and decompositions of the sums of squares.
(8) The failure rate model for decomposition. The approach allows both decomposing Sin (a given case) and including a factor on top (a given example). The former version takes the target for the unit simplex as T = 1, but the present version gives the better explanation that the unit simplex forces this choice anyway. (It reduces the notion of a "target" by applying this term for the unit simplex.) A target value best is seen as applying on top of the Sine measure, for example for turnout. See **Appendix J**.
(9) The transforms for ALHID, EGID, WSL. In the former version I had no proof that ALHID and EGID were restrained to 100, and stated an infinite range, but this is now corrected.
(10)Sharper deconstruction of the Dalton transfer.

It is rather amazing that the Sine measure has not been developed a long time ago and been used for long. My search in the textbooks and literature has its limits however.

The manner how the paper came about leaves traces in repetition. This seems no drawback but rather an advantage. Many readers will not be familiar with both regression through the origin and the unit simplex and the particulars of their combination. Thus it should help to have a repeat warning on the spot where it is relevant. A confusing aspect of variables in the unit simplex is that "proportionality" concerns only the diagonal in the {*w, z*} scatter plot while generally (non-normalised) any line through the origin is proportional. This confusing aspect should now be better clarified in the paper (at repetition).

A comment on Colignatus (2010) is useful too. The latter paper was the first time that I looked at this issue of EPR versus DR. The occasion was the 2011 UK referendum on "Alternative Vote". (This was no referendum on EPR.) The paper explains the major differences between EPR and DR, and focuses on the wasted vote and qualified majority. It takes the traditional "methods" of apportionment and measuring inequality / disproportionality for granted, and tends to prefer WSL. It is only in August – September 2017 that I looked closer at the "philosophies" underlying apportionment and that I have come to realise the confusion in this part of the literature between the party average and marginal candidates. My apologies for not seeing this directly, but my focus was on EPR versus districts.

24. Appendix J. Target base parameters and STSD

The main body of the text establishes that the better standard inequality / disproportionality measure is Sin = $\sqrt{(e'e / z'z)}$. The sum of squared error $e'e$ is lower than for plain error $\tilde{e}'\tilde{e}$ with $\tilde{e} = z - w$ because we allow some slope b to take some of the error. Should be not take $\tilde{e}'\tilde{e}$ as a norm though, like the EGID, or or Sum[Abs[$\tilde{e}$]] like ALHID ? Basically we already answered this question in favour of SDID. Some generality may help both our own perspective and potential application in other fields of research.

24.1. A misleading heuristical idea

The writing of this present paper was hindered for a while by regarding T (in ideal $s^* = T^* v^*$) as a *target slope*, while it mightbe better to see T as a *base.* Though, in that sense, it is a kind of slope again.

- A heuristic idea is: Let us take T as a target slope in $\{w, z\}$. We have $z = b\,w + e$ but want $z = w$. For electoral systems "proportionality" then means $T = 1$, or a slope of 45 degrees or the diagonal in the scatter. The issue may be relevant for more values of T. There, we have a condition of minimal $(b - T)^2 + (1/p - T)^2$, and a nonzero value would indicate inequality / disproportionality with respect to the standard T of the field of study considered.
- The latter idea runs against the properties of the unit simplex. A target value for the unit simplex, as $z = T\,w + \varepsilon$, would actually be a definition of $\varepsilon = z - T\,w$, and it is not clear what theoretical considerations might cause some field of study to make such a definition. If $1'\varepsilon = 0$ then $T = 1$ too. Thus a target T may hold for the levels, but any vector-proportionality in the levels reduces to unit proportionality in the unit simplex. The scatter values around this slope might confuse this view, as they work like "individual constants".
- Still: when we apply RTO to the unit simplex then we also apply RTO for the original level variables, and this reflects on norms on their slope.

24.2. Model with a general target T^*

It helps to formulate the issue more generally. For example y^* are the number of births of baby boys and x^* are the number of all births, scored per day of the week (and doctors causing less births in the weekend). The model is $y^* = T^* x^*$ for the theoretical vectors, with $T^* \approx 0.51$, while we have $y = B\,x + u$ for the estimate on the observed variables for some cities. A slope-specific measure would evaluate $T^* - B$. However, we would also have observational sums $T = S / V$. Thus $B - T$ also requires attention. If we have external information about T^*, then there is a choice to look at $T^* - T$ directly or use the error $T^* - B$. For a voting example, a Federation might wish that States in its Federation House have seats s^* in some *target* proportion T^* to v^* (and now votes rather than population). The target T^* thus may arise from a larger estimate or from institutional norms.

The key relations are summarised in **Table 28**, that adapts **Table 3** for the norm. The norm is formulated for level variables and not unitised variables.

Table 28. Inequality / disproportionality and dispersion in RTO with target T^*

RTO & $(T - T^*)$	T may be T^*	Not so: $T \neq T^*$ or there is no T^*
No dispersion $u = 0$	• Equality / proportionality • $B = T$ • T may be the best estimate of T^* (if the norm is unknown)	• Proportionality in the data but differing fom the target • $B = T$ but error $T - T^*$ • Disproportionality by slope only (deviation from proper proportion)
Dispersion $u \neq 0$	• $1'u = 0$ and thus $B = T$ • *Average* proportionality ($B = T$) can come with dispersion	• $1'u \neq 0$ • $B \neq T$ (and $B \neq T^*$) • Inequality / disproportionality by slope and disperson

The key notions are the errors $T^* - T$, $T^* - B$ and $T - B$ and where they originate.

- Theoretical values: $y^* = T^* x^*$ such that the proportionality holds for all *elements*.
- Observations: $y = B x + u$.
- There is $T = 1'y / 1'x = Y / X$ only for the *sum totals*.
- Thus there is the error $T - T^*$ related to whether $u = 0$ or 1'Abs[u] = 0.
- If $u = 0$ then $B = T$ and this B or T still need not be the best estimate for T^*. The latter might be given on external grounds (norm or larger estimate).
- Summing gives $1'y = Y = B X + 1'u$.
- In regression through the origin $1'u$ need not be 0.
- $(T - B) X = 1'u$, or $B = T \leftrightarrow 1'u = 0$.
- Thus there is the error $T - B$ related to whether $1'u = 0$.
- If $B = T$, then u does not have to be 0. Average proportionality can come with dispersion. However, the lower left cell would still be regarded as inoptimal because of the dispersion.
- If $1'u = 0$, then $B = T$ but we need not have $T = T^*$.

The example for the upper right cell would be that an estimate for some cities would give nice proportionality $B = T$, but the T differs quite a lot from the T^* that one would expect.

The distinctions in **Table 28** are important for who is relatively new to (i) *RTO* and (ii) the *unit simplex in regression* and (iii) the *combination* of these two. As I was relatively new to this, I kept stumbling over aspects without a clear idea of what was happening. In such situations it is always advisable to return to a review of the definitions. Thus the table should help you avoid such stumbling. Part of the answer is: In regression one tends to assume free parameters, but voting might assume some institutional givens.

Table 28 assumes that the top left cell gives unblemished equality / proportionality, because $B = T$ and there is no dispersion. This causes these distinctions:

- Column on the right: Target-specific inequality / disproportionality on T^*
- Bottom row: Target-neglecting inequality / disproportionality on common dispersion
- Lower right cell: Their overlap
- Total inequality / disproportionality as their sum correcting for overlap

The way to handle this decomposition would be as in **Section** 4.4.

24.3. Relevance for votes and seats: the example of turnout

To what extent do these distinctions apply for voting and seats ?

- Lower left cell: Votes and seats are nonnegative variables. The RTO estimate is $B = s'v / v'v$. Namely, u is chosen such that $v'u = 0$. An outcome $B = T = S / V$ then generates the condition that $1 = z'w / w'w$. For nonnegative values on the unit simplex this is only possible when $z = w$.
- Right column on T^*: **Table 16** shows the distinctions between electorate, turnout and elected, with $S = TE\ VE$, $S = T\ V$ and $S = Te\ Ve$. A high level of nonvoting $VE - V$ might be included in overall disproportionality. However, in the present situation, we drop this complexity. For us, it suffices to include the wasted vote $V - Ve$ as the "Other" party, with votes but zero seats. However, one might also opt to have $(T - Te)^2$ as a separate error term, and then add with SDID[ve, se] in the manner of failure rates.

24.4. Is the error on the target additional or already included ?

Our main concern now is whether considerations on the target slope T^* are *additional* to Sin or *already included* as a component. Above **Table 28** sees it as additional, essentially by definition, but it is useful to consider also how it works out in common statistical procedure.

Unitising the variables onto the unit simplex is helpful for the human mind, to compare the ratios. At first sight, it seems that it also helps to force the regression to a target value around

1, with $b = 1 – 1'e$. Researchers on voting might think that this is sufficient for getting information about the *unit* vector-proportionality. This need not be the case though.

Consider a regulation that requires at most 1 car (s) per 2 persons (v). Then $s^* = T^* v^*$ with $T^* = ½$, for theoretical values of the variables. We observe data v and s in various cities. As we know that we do not have to unitise, the regression generates $s = B v + u$. Assume that $u = 0$ and Cos = 1. We can do a statistical test whether B is ½ or not. How are we to evaluate this ? It would depend upon our loss function and then significance level whether we would accept that the regulation works or not. Would we not rather have B close to ½ but Cos different from 1 ? Also, we want to observe developments over the years, and have an index of deviation from the regulation, not just whether it was significant one year and insignificant another year. Essentially we are looking at $(B – T^*)^2$ or $(1 – B / T^*)^2$ (B should rather be smaller) rather than at B or Cos. We want to formulate a more general loss function that does more than assume that all is well when Cos = 1. PM. The issue for $T^* = ½$ is more tractable, since for voting $T^* = S^* / V^*$ tends to generate very low numbers, say 150 seats on 10 million people, which is another reason perhaps why voting researchers might neglect T^*.

It might not matter so much though. If we assume that $S^* / V^* = T^* = \sqrt{(s^{*\prime}s^* / v^{*\prime}v^*)}$ is fairly stable, due to the lack of covariance (its relation to Nv / Ns), then substitution (see **Section 7.3**) gives $R^2 = B^2\, v'v / s's = B^2 / T^{*2}$, and $1 - R^2 = 1 - B^2 / T^{*2} = u'u / s's$. Assuming $T = T^*$ then this reduces to $1 – b^2 = \mathrm{Sin}^2$. Thus, under said assumptions we are already looking at the slope. But those assumptions do not apply in general.

For the issue of cars per person, the values of S and V may have to be estimated as well. For voting they are institutionally given. A question is whether we may substitute the true value T^* for the S / V values in the estimates. If the estimate S / V is unbiased or $T = T^*$ institutionally known, then $b = B / T^*$ indeed. But if we substitute in $B = T^* – 1'u / V ≈ T^*$, then this would only work when $u ≈ 0$. Again, we are not interested in probability assumptions and Expectation, but in designing a (unit) inequality / disproportionality measure.

We calculate U from (externally given) T^* as follows:

$$s = B v + u$$
$$s = T^* v + U$$
$$0 = (B – T^*) v + u – U$$
$$(B – T^*) v = U – u$$

Summing with v and u, with $u'v = 0$ we find: $U'u = u'u$ and:

$$B – T^* = U'v / (v'v)$$

$$b – T^* / T = V / S\ U'v / (v'v) \qquad \text{with } T = S / V$$

$$b – t = E'w / (w'w) \qquad \text{with } E = U / S \text{ and } t = T^* / T$$

$$- 1'e = E'w / (w'w) \qquad \text{if } t = 1\text{, while } 1'e \text{ need not be } 0$$

The value $t = 1$ means $T = T^*$ or that the T in the estimate reflects the same ratio as the required target. In that case E are the plain error $E = \tilde{e} = z – w$. We should not forget about the ALHID criterion Sum[Abs[$\tilde{e}$]] while the present discussion tends to focus on $\hat{e}'\hat{e} ≤ e'e ≤ \tilde{e}'\tilde{e}$. A relation then is also $-1'e = \tilde{e}'w / w'w$. Substitution gives as known $–1'e = b – 1$.

Thus, we have found an expression for the error $B – T^*$, and remarkably $U'u = u'u$ = SSE. Minimising SSE on B might come at the cost of increasing $U'v$ or the difference of B and T^*.

Thus the error on the target is not included in SSE, and must be considered as an additional issue.

All considered, we arrive at this position:

- Assumptions on T^* and thus B and b are external to the RTO estimation of B or b. Statistical procedures assume a "true value" but this doesn't play a part in the estimation (except for Bayesian techniques).
- Statistics has developed an apparatus for hypothesis testing on values of the estimate B with respect to target T^*, which is another way of saying that the target values do not play a role in estimation itself.
- The current least squares estimate uses Minimize $u'u$ over B, but keeping track of the target would give Minimize $u'u + \lambda (B - T^*)$ with a Lagrange condition. For unitised variables this translates into Minimize $e'e + \lambda (b - T^* / T)$ with $T = S / V$ the estimate for T^* in the observations. For example, the imposition of $b = 1$ would give the plain differences (Euclidean distance). Observe though, that when we use b for $z = b\, w$ with $b = 1$, then we have a $b2$ for $z = b2\, w + e$, since that relation does not disappear.
- Part of target T^* seems to be eliminated by using variables on the unit simplex, as we also have $b = B / T$ (for estimated T^*). But this only *seems* to be so. The SSE / SST ratios are the same, and thus involve the same neglect.
- Thus, assumptions on the error between T^* and B come on top of SSE and Sin.
- Thus we have *Target-Neglecting inequality / disproportionality* given by Sin versus *Target-Specific inequality / disproportionality* given by the value of T^* in theoretical $s^* = T^*\, v^*$, and an error function on T^* and B. We can join these into *Total target disproportionality.* For voting, the example is to include a norm on turnout.
- To summarise: For unit vector-proportionality, we cannot assume that the unit simplex is enough. The target T^* holds for B, and thus also for $b = B / T$. The estimation relation $b = 1 - 1'\, e$ cannot be interpreted as an implementation of this target relationship, since it holds for any value of T^*. While the given estimation relation indeed helps to get b close to 1, the considerations on T^* generate a notion of Target Specific inequality / disproportionality that comes on top $b = 1 - 1'\, e$.

25. Appendix K. Failures on failure rate decomposition

Section 7.5 contains a succesful decomposition of Sin into a factor for a slope and a factor for dispersion. This wasn't found so easily, though the cosine distance was obviously available anyway. The following contains some potential ideas that did not work out. The discussion may still be relevant for deeper understanding of some relationships.

25.1. SST = SSX + SSE

The standard regression with a constant generates $1 - R^2 = \hat{e}'\hat{e} / Z'Z$, where $Z = z - zmean$. With RTO we have $\text{Sin}^2 = 1 - \text{Cos}^2 = e'e / z'z$. The regression with a constant will cause a lower error, thus we have $e'e \geq \hat{e}'\hat{e}$.

- Let us write $e'e / \hat{e}'\hat{e} = 1 + \alpha$ for nonnegative α.
- With $z'z \geq Z'Z$ we can write $Z'Z / z'z = 1 - \delta$ for nonnegative δ.
- The total ratio can be higher or lower than 1, depending upon whether the error wins from the effect of centering:

$$\text{Sin}^2 / (1 - R^2) = e'e / z'z / (\hat{e}'\hat{e} / Z'Z) = (e'e / \hat{e}'\hat{e})\, (Z'Z / z'z) = (1 + \alpha)\, (1 - \delta)$$

$$\text{Sin}^2 = (1 + \alpha)\, (1 - \delta)\, (1 - R^2)$$

This decomposition explains inequality / disproportionality in terms of:

- $1 - R^2$ is nonspecific dispersion (Pearson centered dispersion)
- $1 - \delta$ is the impact from rebasing from Z to z
- $1 + \alpha$ is the change by specifically using RTO and forcing the slope

In itself it is possible to choose $\eta = 1 - R^2$ and then determine remainder ζ as in **Section** 4.4. This remains doubtful. The Pearson centered dispersion is different from the dispersion along the RTO regression line. We would be comparing apples and pears. The decomposition has no explicit reference to the slope *b.* This is not the decomposition that we are looking for.

25.2. The Hirschman-Herfindahl concentration index

The Hirschman-Herfindahl concentration index (HHI) for the votes is $w'w$, with a value 1 / n if all parties have an equal size, and a value 1 when one party gets all votes. Nv = 1 / $w'w$ is the "concentrated number of parties", for the votes (not explained what effective would be). Nv has the maximal value n. Thus 1 / (n $w'w$) = Nv / n is in [0, 1]. A value close to 1 means large dispersion, and a value close to 1 / n means concentration.

We have $\text{Sin}^2 = 1 – \text{Cos}^2 = 1 – (1 – \eta) (1 – \zeta)$, in which:

- $\eta = 1 / (n^2\ w'w\ \ z'z)$ is a general measure of dispersion, in the [0, 1] range
- ζ then remains as the contribution of the slope, with $1 – \zeta = \text{Cos}^2 / (1 – \eta)$, and $\text{Cos}^2 = (w'z)^2\ n^2\ \eta$

Unfortunately, we readily find values out of the range. We should actually decompose Sin and not its square, but this causes square roots that are less inviting.

25.3. Cos = √(*b p*) = *b* √(*w'w* / *z'z*)

(a) We might use Min[b, 1 / p, p, 1 / b], exclude division by zero, and continue accordingly.

For a particular application, either b or p will be dominant, so that there will be stability in results. This approach is symmetric. The result need not differ much from the method in the main body of the text. The Min condition is not inviting. Though Max[$z – w$] works too, since a low error on one party causes a high error on another one. We might have overlooked a major approach merely because of convention on smooth formulations.

(b) We also have $\text{Sin}^2 = 1 – \text{Cos}^2 = 1 – b^2\ Ns / Nv = 1 – (1 – \eta) (1 – \zeta)$, in which:

- $\zeta\ = 1 – b^2$ is a slope-specific measure
- $\eta = 1 – w'w / z'z = 1 – Ns / Nv = (Nv – Ns) / Nv$ is a nonspecific dispersion measure

This has the properties that b can be larger than 1 or that Ns can be larger than Nv, so that these are not proper failure rates, though the combination generates the proper Cos and Sin. A high slope finds compensation in a relatively greater concentration of z-values.

However, we also have b w'w = p $z'z$ = $w'z$ and Cos = √(b p) and thus also $\text{Cos}^2 = p^2\ Nv / Ns$. Thus we can use the switch:

- ζ = If $b \leq 1$ then $1 – b^2$ else $1 – p^2$, is a slope-specific measure
- η follows from Sin^2 and ζ, and is a nonspecific dispersion measure

Alternatively, we define a condition on $z'z \leq w'w$ that causes that $b \leq 1$ and otherwise $p < 1$.

There is also Min[b, √($w'w$ / $z'z$)], and then proceed accordingly.

There might be stability in particular applications, but there need not be, and, there will be irrelevant differences over applications. We rather should have a decomposition that is neutral (or symmetric) on b and p. Thus, this does not provide a useful scope for decomposition.

25.4. On overall strategy for looking into this

This Appendix was a useful exercise to see some properties of Cos, but there it stops. The situation is somewhat curious. We originally started out with the traditional inequality / disproportionality measures, and looked for a slope-dependent measure. We recognised that Sin could be used like this, since Cos[v, s] = √(b p). In this Appendix we are repeating the original search for a slope-specific measure, now with the new objective to decompose Cos[*v*, *s*] again. At some point this renewed search stops making sense. It can be agreed that Cos is a general similarity index, so that it is only a *perspective* to regard it as a slope-specific measure. Yet, it satisfies the conditions, like that Sin is a metric, and there is no reason to regard the perspective as invalid. It is more remarkable that **Section** 7.5 found a decomposition anyway.

26. Appendix L. Representative Largest Remainder (RLR), Holland 2017

Table 29 contains the official apportionment of seats in the Dutch 2nd Chamber of Parliament after the elections of March 15 2017, and an alternative apportionment that has no focus on parties but that instead focuses on the *marginal representative*. Holland has 150 seats, uses D'Hondt, and excludes parties that do not attain the natural threshold of 1/150 or 0.67%. At the bottom of the table there are the inequality / disproportionality measures ALHID, EGID, WSL and SDID. Given the Dutch use of D'Hondt that favours large parties, it is not surprising that the alternative apportionment improves on all measures. Only two seats are relocated though.

Since Holland is already quite proportional, this discussion itself is rather marginal, but it helps to highlight that the large literature in Political Science *that does not have this focus on equal representative democracy* rather distracts from this focus. For example, [71] when there is discussion in Canada about the use of the EGID index, then the focus is on EGID, and not on *above clear EPR apportionment*. Political scientists present their measures but we would rather see that they advise to the world that Holland has a sound system (apart from marginal criticism) that is enormously better than systems of DR currently in use in France, UK and USA.

For the alternative apportionment in **Table 29**, the following principles have been used – and this method might be called *Representative Largest Remainder* (RLR) (though perhaps this already was the intention of Hamilton). Independent candidates are called "party" too, but we still look at the party *marginal* candidates. For proper understanding: It is not claimed that SDID will always identify outcomes of RLR as the best. SDID is a measure and RLR is an apportionment method that also considers *quota on marginal parties* in real world applications.

In the literature HLR is tolerant on which candidate has the highest remainder. The Dutch system takes the natural quota Q as a threshold. Thus it excludes a candidate with say 0.75 Q who does not have the backing of a party that has passed the threshold. The following RLR improves on the Dutch model by suggesting a threshold (i) of "at least" or "greater than or equal to" 0.5 Q and (ii) *only for the remainder seats*. Thus RLR is not intended as an improvement of HLR, but only intended as a suggestion for real world application in large National Parliaments. [72] See **Appendix M** for an example that RLR may be too strict for a small number of seats S. The norm "greater than or equal to 0.5" is taken from the general convention of rounding up.

(1) Let the natural quota be $Q = V / S$. Parties have an average score of $a = v / Q = S$ w.
(2) Parties first gain the seats f = Floor[a].
(3) The remainder seats are RemS = S – Sum[f] and the fractional parts are $r = a - f$.
(4) The fractional parts $r \geq 0.5$ are sorted. The remainder seats RemS are allocated from the highest value to the lowest, till there are no seats to allocate or there is no qualified candidate anymore. This gives remainder apportionment *rap*1, and interim total $g = f +$

[71] https://en.wikipedia.org/wiki/Gallagher_index
[72] Beumer (2010:59) on the Dutch fear for small parties in their proportional system: "The political upshot is that a proliferation of small parties is a realistic concern. Jefferson [D'Hondt, JDH] 's bias for large parties largely resolves the issue. On the other hand, imposing minimum requirements achieves the same aim. It is unclear why both measures would have to be taken." And "(...) that Jefferson's side-effects are mainly negative." Indeed. Still, largest remainders follow the philosophy of the marginal candidate, and then a threshold of 0.5 Q that applies only to the remainder seats seems adequate.

*rap*1. For all clarity: in this step all parties partake, also parties that were below the natural quota and did not get a floor assignment. The fractional threshold $r \geq 0.5$ only applies to the fractional remainders after assigning the floor.

(5) If there would not be enough parties above the $r \geq 0.5$ remainder threshold, or if there would not be enough parties at all, so that there are still RemS2 seats to allocate, then the parties h that have seats are selected, their scores on $v - g\ Q$ are sorted (some negative), and the RemS2 seats are allocated from the highest values down, giving *rap*2. The final apportionment is $s = f + rap1 + rap2$.

Table 29. Apportionment with a focus on the marginal representative, Holland 2017

				Official		70107				RepBase	Diff
Dutch Hous	Voters	Per 10^5	% Valid	Seats	Label	Quota	Fr >= 0.5	Floor	Rest	F+R	Off-Rep
VVD	2238351	17360	21.29	33	A	31.93	0.43	31	1	32	1
PVV (Partij	1372941	10648	13.06	20	B	19.58	0.08	19	1	20	
Christen De	1301796	10097	12.38	19	C	18.57	0.07	18	1	19	
Democrate	1285819	9973	12.23	19	D	18.34		18		18	1
Groenlinks	959600	7443	9.13	14	E	13.69	0.19	13	1	14	
SP (Sociali	955633	7412	9.09	14	F	13.63	0.13	13	1	14	
Partij van d	599699	4651	5.70	9	G	8.55	0.05	8	1	9	
ChristenUn	356271	2763	3.39	5	H	5.08		5		5	
Partij voor (	335214	2600	3.19	5	I	4.78	0.28	4	1	5	
50PLUS	327131	2537	3.11	4	J	4.67	0.17	4	1	5	-1
Staatkundig	218950	1698	2.08	3	K	3.12		3		3	
DENK	216147	1676	2.06	3	L	3.08		3		3	
FvD	187162	1452	1.78	2	M	2.67	0.17	2	1	3	-1
VNL (VoorN	38209	296	0.36	0	N	0.55	0.05	0		0	
Piratenpart	35478	275	0.34	0	O	0.51	0.01	0		0	
Artikel 1	28700	223	0.27	0	P	0.41					
Nieuwe We	14362	111	0.14	0	Q	0.20					
Onderneme	12570	97	0.12	0	R	0.18					
Lokaal in d	6858	53	0.07	0	S	0.10					
Niet Stemm	6025	47	0.06	0	T	0.09					
De Burger	5221	40	0.05	0	U	0.07					
Geen Peil	4945	38	0.05	0	V	0.07					
JEZUS LEE	3099	24	0.03	0	W	0.04					
Vrijzinnige	2938	23	0.03	0	X	0.04					
Libertarisch	1492	12	0.01	0	Y	0.02					
MenS en S	726	6	0.01	0	Z	0.01					
StemNL	527	4	0.01	0	AA	0.01					
Vrije Demo	177	1	0.00	0	AB	0.00					
Valid votes	10516041	81561	100	150			1.62	141	9	150	0
PM. Below	161327	1251	1.53	2.30							
PM. Waste	208742	1619	1.98	2.98							
Invalid vote	47415	368	-	0.68							
Blanco	15876	123	-	0.23							
Turnout	10563456	81929	-	-							
Electorate	12893466	100000	-	-							
LHI				3.00						2.31	
EGD				1.09						0.79	
WSL				2.26						2.09	
SDD				1.78						1.70	

For step 5, it might be an alternative to reduce the 0.5 *Q* fraction remainder threshold for parties without seats yet, for participation in the apportionment of the remainder seats. This seems unwise given the general aversion of the fringe. RLR is suggested as an improvement for the Dutch threshold, and not as a suggestion to amend HLR itself for all applications. Another option for step 5 is to use the remainder votes RemV = $v - f\ Q$ to apportion RemS2 in a more traditional manner, within the locus for apportionment, but this causes the question which traditional method to use. The present suggestion is straightforward.

The ALHID, EGID, WSL and SDID scores in the table include the 0 seats for the blanc and invalid votes (0.4% of turnout or 0.68% of a seat). As far as I can see, RLR is recognised by SDID as the lowest score possible, within this framework. However, there is an issue on the wasted vote and qualified majority, see the body of the text. For comparison, **Appendix F**, contains another alternative for Holland, in which the wasted vote are collected in one category, which may be replaced by qualified majority. See the discussion in **Section** 4.6.

Gallagher (1992:491) quotes Balinski & Young:

> "Paradoxes can occur when the Hare quota or the Droop quota is used. [i.e. above natural quota *Q*] Balinski and Young, who outline these paradoxes, argue that they 'occur because at bottom the use of remainders to determine priority is simply not a proportional device. It does not adequately reflect the *relative* size' of the parties. If this implies that LR-Hare is not a genuine EPR method, then it is surely too harsh a judgement."

Gallagher missed the opportunity to observe that the judgement is not only *too harsh* but also *logically absurd*. The very purpose of the apportionment is to arrive at EPR, and mathematicians Balinski and Young only don't see this purpose because they don't see that the *marginal candidate is in the denominator*, because they focus on The Party instead of on the party marginal canditate.

PM. Above RLR scheme improves the position in Holland of small parties with at least 0.5 *Q* that face the existential question by giving them a chance at the allocation of remainder seats. It would seem to be an academic question whether this is called "use of remainders" or "reflect the *relative* size" of parties.

27. Appendix M. The effect of the number of seats on HRL and RLR

The main body of the text mentions that *S* is a key variable but tends to disregard it. **Table 30** indicates the importance. We concentrate on HLR and RLR.

Table 30. The effect of the number of seats

	1	2	3	4	5	6	7	8	9	10	11	12	13
			10√Sin	Specif	100√Sin	100 Sin	Gini	ALHID	EGID	WSL	ALHIDT	EGIDT	WSLT
Ex.	V	S	10	10	100	100	100	100	100	inf	10	10	10
YA	70	4	3.5	0.9	35.1	12.3	11.6	10.0	9.2	9.6	3.2	3.0	2.4
5	22	1											
	8	0											
YB	72	3	4.1	1.2	41.4	17.1	12.0	12.0	10.4	7.1	3.5	3.2	2.1
5	14	1											
	14	1											
YC	1500	7	2.9	0.6	28.6	8.2	5.1	4.5	2.9	1.3	2.1	1.7	1.0
25	1500	7											
	900	4											
	500	3											
	500	3											
	200	1											
YD	1500	8	2.7	0.5	27.2	7.4	4.9	4.3	2.7	1.2	2.1	1.7	1.0
26	1500	8											
	900	5											
	500	2											
	500	2											
	200	1											
YE	60	3	5.3	2.0	53.3	28.4	27.2	17.0	14.7	29.6	4.1	3.8	3.5
5	23	2											
	9	0											
	8	0											
YF	60	3	4.6	1.5	45.8	21.0	19.2	11.0	9.8	21.8	3.3	3.1	3.2
5	23	1											
	9	1											
	8	0											

Cases YA and YB are mentioned by Gallagher (1992:492) as an effect in apportionment, and show that Party *A* can gain votes (from 70 to 72) but still lose a seat. This is a real effect at the level of 5 seats, but for 100 seats the change would be minor. It helps to look at the values of the inequality / disproportionality indices. The indices are relatively high which indicates that the number of seats is small so that it is difficult to get a good fit. The possibilities of allocating the 5 seats over the three parties in the Floor[*a*] ≤ *s* ≤ Ceiling[*a*] locus are limited, and most apportionment methods will regard YB with {3, 1, 1} as optimal.

YC and YD show the "Alabama Paradox". [73] When the number of seats is increased from 25 to 26 while the votes remain the same, then 2 parties lose seats and 3 gain seats (including the extra). Wikipedia doesn't provide the inequality / disproportionality indices, and thus can only state: It is a paradox ! We can see that the addition of a seat and rearrangement helps to reduce the disproportionality. This is about as paradoxical as showing that all even numbers can be divided evenly by 2 and then be surprised that 3 cannot be divided in such manner.

Cases YE and YF are based upon Gallagher (1992:471), and are presented here with the purpose to highlight the difference between HLR (that creates YF) and RLR (that creates YE). The main impact is from the small number of seats, that causes that the quota *Q* = 20. With this high quota, the two small parties have votes below 0.5 *Q*. HLR is tolerant, and RLR is intolerant to this. The reason is that RLR is suggested for National Parliaments with at least 100 seats, and there is no claim that there would be optimality for low values of *S*.

[73] https://en.wikipedia.org/wiki/Largest_remainder_method

28. Appendix N. Relation of HLR, RLR and ALHID, "New State Paradox"

Gallagher (1991:39-40) provides the following statement on HLR as apportionment and ALHID as inequality / disproportionality measure.

What is not always appreciated is that this index is always minimized by the largest remainders method; the two are simply two sides of the same coin (for proof, see Appendix). Loosemore and Hanby's conclusion (1971: 475) that, when measured by their index, disproportionality is lowest in largest remainder systems is thus not very surprising. Consequently, using this index not only pre-judges what is ostensibly to be discovered, but it is also subject to all the anomalies and paradoxes of largest remainders, of which there are several.

The Loosemore-Hanby index's vulnerability to paradoxes can be illustrated by a case involving what is known in the United States as the 'new state' paradox, in which an allocation between two states is disturbed by the arrival of a newcomer. Suppose that there are 90 votes, 2 seats and two parties, *A* with 68 votes and *B* with 22 votes. Both largest remainders and Sainte-Laguë will award both seats to *A*. But if a third party now joins the fray and wins 10 previously uncast votes, so that the distribution is 68-22-10, the Hare quota rises from 45 to 50. Consequently, *A*'s remainder drops from 23 to 18, below *B*'s remainder of 22, so *A* and *B* now receive one seat each, even though the relationship between them has not altered at all. Sainte-Laguë, of course, as a divisor system, still awards both to *A*. The problem is that the Loosemore-Hanby index always by definition slavishly follows the largest remainders method, and so in this case it indicates that 2-0 is the least disproportional allocation when there are 90 votes but 1-1 is least disproportional when there are 100 votes. The index could be used to 'prove' that largest remainders delivers a more proportional result than Sainte-Laguë in the 100-vote case, even though the divergence between them arises only because largest remainders is vulnerable to paradoxes from which Sainte-Laguë is immune.

The Loosemore-Hanby index has the same merits and demerits as the largest remainders formula on which it is based. Its method is straightforward and easy to understand, but it is weakened by its vulnerability to paradoxes. These and other doubts have led to the development of other difference-based indices, to which we now turn.

Table 31 contains the four cases and the scores by the various measures. HLR and ALHID may be related, but also other measures like EGID select the original assignment GA as optimal, and also shift the 1 seat when the New State is added, in GD.

SDID, Sin, Gini, WSL and WSLT keep apportioning the 2 seats to the large party.

While Gallagher rightly points out that HLR will shift 1 seat. RLR (defined in **Appendix L**) will work a bit differently. The parties with seats 22 and 10 have both less than 0.5 *Q,* and will be excluded from both the apportionment of the remainder seats and in the final round. Thus RLR will also give 2 seats only to the largest party.

That said, I find Gallagher's statement difficult to understand. There is nothing paradoxical about these findings. A statement like that WSL is immune to the paradox only means that a paradox was invented in the first place. The measures work like they have been defined. The disqualification of HLR and ALHID based upon so flimsy an misconception as calling this "paradoxical", miscomprehends that HLR basically *defines* what *equal (proportional) representative democracy* entails. See **Appendix L,** also for the difference between HLR and RLR.

Table 31. Case of Gallagher (1991:39-40), the "New State Paradox"

	1	2	3	4	5	6	7	8	9	10	11	12	13
			10√Sin	Specif	100√Sin	100 Sin	Gini	ALHID	EGID	WSL	ALHIDT	EGIDT	WSLT
Ex.	V	S	10	10	100	100	100	100	100	inf	10	10	10
GA	68	2	5.5	2.2	55.5	30.8	24.4	24.4	24.4	32.4	4.9	4.9	3.6
	22	0											
GB	68	1	6.7	3.3	67.5	45.5	25.6	25.6	25.6	35.4	5.1	5.1	3.7
	22	1											
GC	68	2	5.8	2.4	57.9	33.5	32.0	32.0	28.4	47.1	5.7	5.3	4.1
	22	0											
	10	0											
GD	68	1	6.9	3.4	68.7	47.2	33.0	28.0	24.6	50.4	5.3	5.0	4.2
	22	1											
	10	0											

29. Appendix O. Update February 6 2018

The following concerns the update in version 4 in the February 6 2018 update of this paper, compared to version 3 of September 16 2018, Colignatus (2017c).

(1) **Table 3** and **Table 28** have been corrected by a counterexample. A case with b = 1 without dispersion is catalogued as *unit* or *diagonal* equal / vector-proportional. A case with b = 1 with dispersion is catalogued as *average unit* vector-proportional.

(2) **Table 9** now includes a row with $T = S / V$ and $s = T v + \tilde{u}$, and it is emphasized that this is a contribution of compositional data to RTO.

(3) **Section** 11.6 on the cosine rule has been corrected w.r.t. the unit circle.

(4) The abbreviations are now SDID (sine), ALHID (absolute difference) and EGID (Euclid), to express the *inequality / disproportionality*, compared to the earlier SDD, LHI and EGD. The expression *disp*[*v, s*] now is ID[*v, s*]. The abbreviation WSL is not changed into WSLID because of column widths of the tables. Some tables use pictures from excel sheets and still contain the old abbreviations.

(5) The abbreviation EPR reflects *equal or proportional representation*, instead of the earlier PR or ER / PR.

(6) Colignatus (2017e) has been included as a reference, as a short weblog text that summarises the findings here.

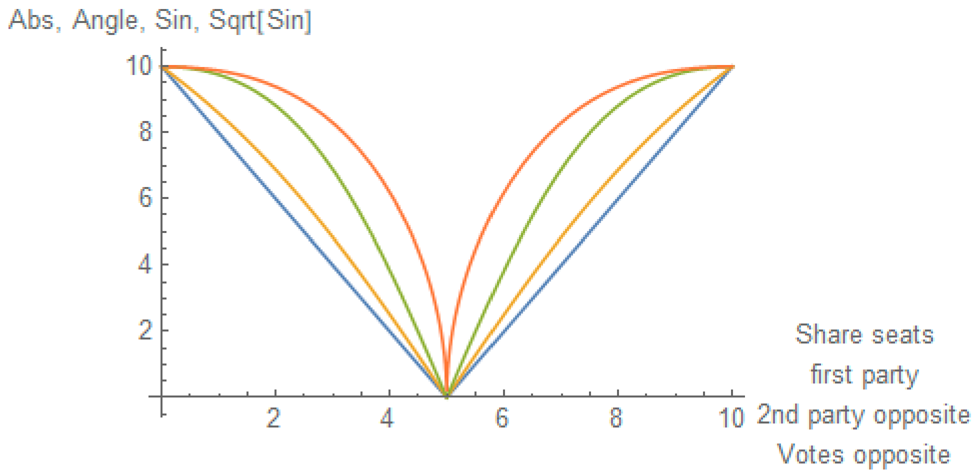
(7) A new reference is Colignatus (2017d), “One woman, one vote. Though not in the USA, UK and France” (1W1V). This discusses equal / proportional representation (EPR) and district representation (DR). These systems have distinct political characteristics, and their properties should not be confused with inequality / disproportionality measures discussed here, even though one might use such measures for diagnostics. 1W1V suggests that the term “election” may be confusing since there are different meanings in EPR and DR, whence there are EPR-elections and DR-elections. An umbrella term is “(half) election”.

(8) Colignatus (2018a) evaluates the Aitchison and the angular distance for use as inequality or disproportionality measures for votes and seats. **Appendix B** above now refers to this.

It appears that the comparison of votes and seats requires conditions that violate the Aitchison distance. For example, with votes {1, 2, 7} and seats {2, 1, 7}, the Aitchison distance essentially compares {1, 2, 7} / {2, 1, 7} = {1/2, 2, 1} with {1, 1, 1}, and neglects outcomes that are 1, for these have Log[1] = 0. In this example, the Aitchison distance essentially compares a two-party case and not a three-party case. Thus, the Aitchison distance is not advisable for the evaluation of votes and seats, though there may be other cases in the voting literature in which the logratio transform remains useful.

Somewhat remarkable, the literature search for this paper found only few references to the angular distance itself. Linear algebra on vectors *x* and *y* generates *k* = Cos[*x, y*] and there are two direct transforms for inequality: (a) θ = ArcCos[*k*] and (b) Sin = Sqrt[1 – *k*^2]. It was a design decision to use the latter as an easier transform, also linking up to the interpretation of R-square, and increasing sensitivity by taking a double Sqrt. Colignatus (2018a) evaluates not only the Aitchison distance but also the AngularID. The conclusion is that the AngularID is the better of the two. However, SDID is better than both, see also **Section** 5.7 above. **Figure 28** contains a plot of the ALHID (blue), AngularID (yellow), the Sine (green) and the |SDID| (red). Since we plot opposite values with majority switches, the SDID becomes negative, and thus we take its absolute value.

Figure 28. Plot of *d*[votes, seats] for votes = 10 – seats and seats = {t, 10 – t}, for *d* = ALHID, AngularID, Sine, and |SDID| (eliminating the latter's negative sign)



(9) While this present paper looks into inequality / disproportionality measures *d*[*v, s*] for votes and seats, political parties can also be scored on policy positions *p* in [0, 10], using the range [0, 10] to avoid leading zeros and avoiding percentages because those might suggest too large a precision. Colignatus (2018b) looks into measures *pd*[*v, s, p*], with a role for *d*[*p v, p s*].

(10)Readers of Dutch might be interested in Colignatus (2017f), a letter to the Dutch State Commission on the Parliamentarian System. This Dutch text collects and summarises various findings, also from this present paper on inequality / disproportionality measures.

30. References

Colignatus is the name in science of Thomas Cool, econometrician and teacher of mathematics, Scheveningen, Holland, http://econpapers.repec.org/RAS/pco170.htm

References in the text to Wikipedia refer to it as a portal and no source.

Balinksi, M. and H.P. Young (1976), "Criteria for proportional representation", IIASA Research Report December, RR-76-020, http://pure.iiasa.ac.at/525/1/RR-76-020.pdf

Balinksi, M. and H.P. Young (1980), "The Webster method of apportionment", Proceedings of the National Academy of Sciences 77(1):1-4, February, https://www.researchgate.net/publication/7186641_The_Webster_method_of_apportionment

Belov, D. and R.D. Armstrong (2011), "Distributions of the Kullback-Leibler divergence with applications", J.Math. Stat. Psychol., May, 64(Pt 2):291-309, https://www.ncbi.nlm.nih.gov/pubmed/21492134, res. report: https://www.lsac.org/docs/default-source/research-(lsac-resources)/rr-09-02.pdf

Beumer, M. (2010), "Apportionment in theory and practice", MSc Thesis in logic, UvA, http://www.illc.uva.nl/Research/Publications/Reports/MoL-2010-07.text.pdf

Borg, I. & P.J.F. Groenen (2005), "Modern Multidimensional Scaling", Springer, http://www.springer.com/gp/book/9780387251509

Carey, J. M., & Hix, S. (2009), "The Electoral Sweet Spot: Low-Magnitude Proportional Electoral Systems", Working Paper (draft, colour graphs), http://www.lse.ac.uk/government/research/resgroups/PSPE/pdf/PSPE_WP1_09.pdf

Carey, J. M., & Hix, S. (2011), "The Electoral Sweet Spot: Low-Magnitude Proportional Electoral Systems", American Journal of Political Science, 55(2), 383-397, Open Source http://onlinelibrary.wiley.com/doi/10.1111/j.1540-5907.2010.00495.x/full and http://personal.lse.ac.uk/hix/Working_Papers/Carey-Hix-AJPS2011.pdf

Caulfield, M.J. (2010), "Apportioning Representatives in the United States Congress", MAA Convergence, https://www.maa.org/press/periodicals/convergence/apportioning-representatives-in-the-united-states-congress-introduction

Colignatus, Th. (2006), “On the sample distribution of the adjusted coefficient of determination (R2Adj) in OLS”, http://library.wolfram.com/infocenter/MathSource/6269/

Colignatus, Th. (2007), "Correlation and regression in contingency tables. A measure of association or correlation in nominal data (contingency tables), using determinants", https://mpra.ub.uni-muenchen.de/3660/

Colignatus, Th. (2009, 2015), "Elegance with Substance", https://zenodo.org/record/291974

Colignatus, Th. (2010), "Single vote multiple seats elections. Didactics of district versus proportional representation, using the examples of the United Kingdom and The Netherlands", https://mpra.ub.uni-muenchen.de/22782/

Colignatus, Th. (2011), "Conquest of the Plane", https://zenodo.org/record/291972

Colignatus, Th. (2014), "Voting Theory for Democracy", Thomas Cool Consultancy & Econometrics, https://zenodo.org/record/291985

Colignatus, Th. (2017a), "The performance of four possible rules for selecting the Prime Minister after the Dutch Parliamentary elections of March 2017", https://mpra.ub.uni-muenchen.de/77616/

Colignatus, Th. (2017b), "Two conditions for the application of Lorenz curve and Gini coefficient to voting and allocated seats", https://mpra.ub.uni-muenchen.de/80297/

Colignatus, Th. (2017c), "Comparing votes and seats with a diagonal (dis-) proportionality measure, using the slope-diagonal deviation (SDD) with cosine, sine and sign", first version of this paper, at MPRA, August 17, no 80833; 1st revision August 24, no 80965; 2nd revision September 16, no 81389. This has become the present paper, as a 4th revision, with the title: “Comparing votes and seats with cosine, sine and sign, with attention for the slope and enhanced sensitivity to inequality / disproportionality”

Colignatus, Th. (2017d), “One woman, one vote. Though not in the USA, UK and France”, https://mpra.ub.uni-muenchen.de/82513/

Colignatus, Th. (2017e), “Statistics, slope, cosine, sine, sign, significance and R-squared”, https://boycottholland.wordpress.com/2017/10/21/statistics-slope-cosine-sine-sign-significance-and-rsquared/

Colignatus, Th. (2017f), “Aan de Staatscommissie parlementair stelsel”, http://thomascool.eu/Thomas/Nederlands/Wetenschap/Advies/2017-12-21-Staatscie-parl-stelsel.pdf

Colignatus, Th. (2018a), “Comparing the Aitchison distance and the angular distance for use as inequality or disproportionality measures for votes and seats”, https://mpra.ub.uni-muenchen.de/84387 also at https://www.wolframcloud.com/objects/thomas-cool/Voting/2018-01-18-Aitchison.nb

Colignatus, Th. (2018b), “Measures of policy distance and inequality / disproportionality of votes and seats”, https://mpra.ub.uni-muenchen.de/84324/ also at https://www.wolframcloud.com/objects/thomas-cool/Voting/2018-02-02-PolicyDistance.nb

Dongen, S. van & A.J. Enright (2012), "Metric distances derived from cosine similarity and Pearson and Spearman correlations", https://arxiv.org/abs/1208.3145

Dumont, P. & J.-F. Caulier (2003), "The "effective number of relevant parties": How voting power improves Laakso-Taagepera's index", http://centres.fusl.ac.be/CEREC/document/2003/cerec2003_7.pdf

Draper, N.R. & Y. Yang (1997), "Generalization of the geometric mean functional relationship", Computational Statistics & Data Analysis, Volume 23, Issue 3, 9 January 1997, Pages 355-372; Preprint 1995 Technical report no 943, http://www.stat.wisc.edu/node/1470

Eisenhauer, J.G. (2003), "Regression through the origin", Teaching Statistics. Volume 25, Number 3, Autumn 2003, p76-80

Erb, I. and C. Notredame (2016), "How should we measure proportionality on relative gene expression data?", Theory Biosci. (2016) 135:21–36, https://link.springer.com/content/pdf/10.1007%2Fs12064-015-0220-8.pdf or https://link.springer.com/article/10.1007%2Fs12064-015-0220-8

Gallagher, M. (1991), "Proportionality, disproportionality and electoral systems", Electoral Studies, 10:1, 33-51, https://www.tcd.ie/Political_Science/draft/staff/michael_gallagher/ElectoralStudies1991.pdf

Gallagher, M. (1992), "Comparing proportional representation electoral systems: Quota's, thresholds, paradoxes and majorities", B.J. Pol.S. 22, 469-496

Gallagher, M. (2007), Electoral Systems Web Site, Dublin, Trinity College, http://www.tcd.ie/Political_Science/Staff/Michael.Gallagher/ElSystems/index.php

Gallagher, M. (2017), "Election indices dataset", accessed at 2017-07-31, http://www.tcd.ie/Political_Science/staff/michael_gallagher/ElSystems/Docts/ElectionIndices.pdf

Goldenberg, J. & S.D. Fisher (2017), "The Sainte-Laguë index of disproportionality and Dalton’s principle of transfers", Accepted for publication in Party, March 2017,

Hill, I.D. (1997), "Measuring proportionality", Voting matters, 8, p7-8, http://www.mcdougall.org.uk/VM/VOL1/ISSUE01-23.pdf

Johnston, J. (1972), "Econometric methods", 2nd edition, McGraw-Hill

Karpov, A. (2008), "Measurement of disproportionality in proportional representation systems", Mathematical and Computer Modelling, Volume 48, Issues 9–10, November 2008, Pages 1421-1438, http://www.sciencedirect.com/science/article/pii/S0895717708001933 Politics

Kestelman, P. (2005), "Apportionment and Proportionality: A Measured View", Voting Matters, 20, p 12-22, http://www.votingmatters.org.uk/ISSUE20/I20P4.PDF

Koppel, M, and A. Diskin (2009), "Measuring disproportionality, volatility and malapportionment: axiomatization and solutions", Social Choice and Welfare, August, 33:281, https://www.researchgate.net/publication/225444815_Measuring_disproportionality_volatility_and_malapportionment_Axiomatization_and_solutions

Kozak, A. and R. A. Kozak (1995), "Notes on regression through the origin", The Forestry Chronicle, May/June, Vol 71 no 3, p326-330

Laakso, M. (1980), "Electoral Justice as a criterion for different systems of proportional representation", Scandinavian Political Studies, Bind 3 (New Series) (1980) 3, https://tidsskrift.dk/scandinavian_political_studies/article/view/32355/30159

Laakso, M., & R. Taagepera (2007), "Proportional representation in Scandinavia: Implications for Finland", Scandinavian Political Studies, https://www.researchgate.net/publication/230000636_Proportional_Representation_in_Scandinavia_Implications_for_Finland

Leznik, M. & C. Tofallis (2005), "Estimating Invariant Principal Components Using Diagonal Regression", http://researchprofiles.herts.ac.uk/portal/en/publications/estimating-invariant-principal-components-using-diagonal-regression(d3379080-bda2-4d76-ab8d-b557c7fabea1).html

Lovell, D. V. Pawlowsky-Glahn, J.J. Egozcue, S. Marguerat, J. Bähler (2015), "Proportionality: A Valid Alternative to Correlation for Relative Data", PLOS Computational Biology, http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004075

Malkevitch, J. (2002), "Apportionment", AMS Feature Column, http://www.ams.org/samplings/feature-column/fcarc-apportion1

Malkevitch, J. (2017), "Pairwise Equity in Apportionment (2017)", https://www.york.cuny.edu/~malk/gametheory/tc-2017-pairwise-apportion.html

NCSS Statistical Software (undated), "Lin's concordance correlation coefficient", Chapter 301 of the documentation, https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Lins_Concordance_Correlation_Coefficient.pdf

Pawlowsky-Glahn, V., J.J. Egozcue, R. Meziat (2007), "The statistical analysis of compositional data: The Aitchison geometry", https://laboratoriomatematicas.uniandes.edu.co/cursocoda/04-Vera-geometry.pdf

Pukelsheim, F. (2014), "Proportional representation. Apportionment methods and their applications", Springer,

Renwick, A. (2015), "Electoral Disproportionality: What Is It and How Should We Measure It?", http://blogs.reading.ac.uk/readingpolitics/2015/06/29/electoral-disproportionality-what-is-it-and-how-should-we-measure-it/

Samuelson, P. (1942), " A Note on Alternative Regressions", Econometrica, Vol. 10, No. 1 (Jan., 1942), pp. 80-83, : http://www.jstor.org/stable/1907024

Taagepera, R. and M.S. Shugart (1989), "Seats and Votes", Yale

Taagepera, R. and M. Laakso (2006), "Proportionality Profiles of West European Electoral Systems", European Journal of Political Research 8(4):423 - 446 · May 2006, https://www.researchgate.net/publication/230041613_Proportionality_Profiles_of_West_European_Electoral_Systems

Taagepera, R. and B. Grofman (2003), "Mapping the indices of seats-votes disproportionality and inter-election volatility", Party Politics, 9(6), p659-677, http://escholarship.org/uc/item/0m9912ff#page-1

Theil, H. (1969), “The desired political entropy”, American Political Science Review, 63: 21-5, https://www.jstor.org/stable/1954705

Tofallis, C. (2000), "Multiple Neutral Regression", Operational Research Paper 14, UHBS 2000:13, http://uhra.herts.ac.uk/bitstream/handle/2299/689/S7.pdf?sequence=1

Tullock, G. (2008), "Public Choice", The New Palgrave Dictionary of Economics, Second Edition, http://www.dictionaryofeconomics.com/article?id=pde2008_P000240&q=rational%20choice&topicid=&result_number=10

Quinn, T. (2017), "An introduction to proportionality", https://cran.r-project.org/web/packages/propr/vignettes/a_introduction.html

Young, H.P. (2004), “Fairness in Apportionment”, Mimeo, Prepared for the U. S. Census Bureau Symposium, http://www.census.gov/history/pdf/Fairness_in_Apportionment_Young.pdf

Zand, M.S., J. Wang, S. Hulchey (2015), "Graphical Representation of Proximity Measures for Multidimensional Data. Classical and Metric Multidimensional Scaling", The *Mathematica* Journal 17, http://www.mathematica-journal.com/2015/09/graphical-representation-of-proximity-measures-for-multidimensional-data/