A Solution to the Seat-Product Problem: The Square Root Rule of Elections*

Jack Bailey†

November 14, 2024

Word count: 3,240

Abstract

Seat-product models of the effective number of parties offer our best explanation for why party-systems fragment. Yet they rely on a simplification. Rather than model the effective number of parties writ large, they models this index using the largest party alone. This induces a positive bias, since small parties have a suppressive effect. I show that we can derive new seat-product models without need for this simplification. The key is to recognise that the effective number of parties is information theoretic in nature. This lets us use well-known information theoretic identities to avoid party shares altogether. Further, this process produces an interesting by-product: a new "root rule" of elections. Just like the cube root of a nation's population approximates the number of seats in its assembly, I show that the square root of the actual number of parties approximates the effective number of parties as well.

*This paper benefitted from useful comments from Chris Hanretty and Stuart Perrett.

†Lecturer in Quantitative Political Science, Department of Politics, The University of Manchester, UK. If you have any comments or questions, please feel free to contact me at jack.bailey@manchester.ac.uk.

Introduction

The "seat-product model" (Shugart and Taagepera 2017; Taagepera and Shugart 1993) offers our best explanation *why* party-systems fragment. They show that the effective number of parties scales with a nation's "seat-product": the number of seats in its assembly times the average magnitude of each of its districts. Yet these models rely on a simplification. Rather than model the effective number of parties *writ large*, they use *the largest party alone*. This is not so unreasonable, as the largest party has the largest effect on the index (Laakso and Taagepera 1979). But smaller parties still matter and have a suppressive effect. Omitting them, thus, induces a positive bias.

In this short paper, I show that we can derive new seat-product models *without recourse to simplification*. The key is to recognise that the effective number of parties is information theoretic in nature. As such, we can use well-known information theoretic identities to break it into sub-quantities that have established bounds. By averaging over these bounds, we can then build new models of the effective number of parties that avoid party shares altogether, solving the "seat-product problem". This also comes with a welcome by-product: a new "root rule" of elections. Just as the cube root of a nation's population approximates the seats in its assembly (Taagepera 1972), I show that the square root of the number of parties approximates the effective number of parties as well.

The "Seat-Product Problem"

Logical models like the seat-product model rely on deduction, not empirical data. To derive them, we first pick some variable to model. Then, we determine its bounds: the largest and smallest values that the variable could possibly take. Next, we take the geometric mean of these two extremes, since this represents our best guess of the variable's true value in the absence of all other data (see Taagepera 2008). Finally, we simplify the resulting expression to arrive at our final model.

When it comes to the *actual* number of parties, this process works rather well. But it soon becomes much more complex, since it is not at all clear how we should convert actual to *effective* parties. In a well-cited, three-decade-old, paper, Taagepera and Shugart (1993) offer a potential solution: to first simplify the problem by modelling the share of the largest party, then use that to derive the effective number instead. As far as assumptions go, this one is not so unreasonable. After

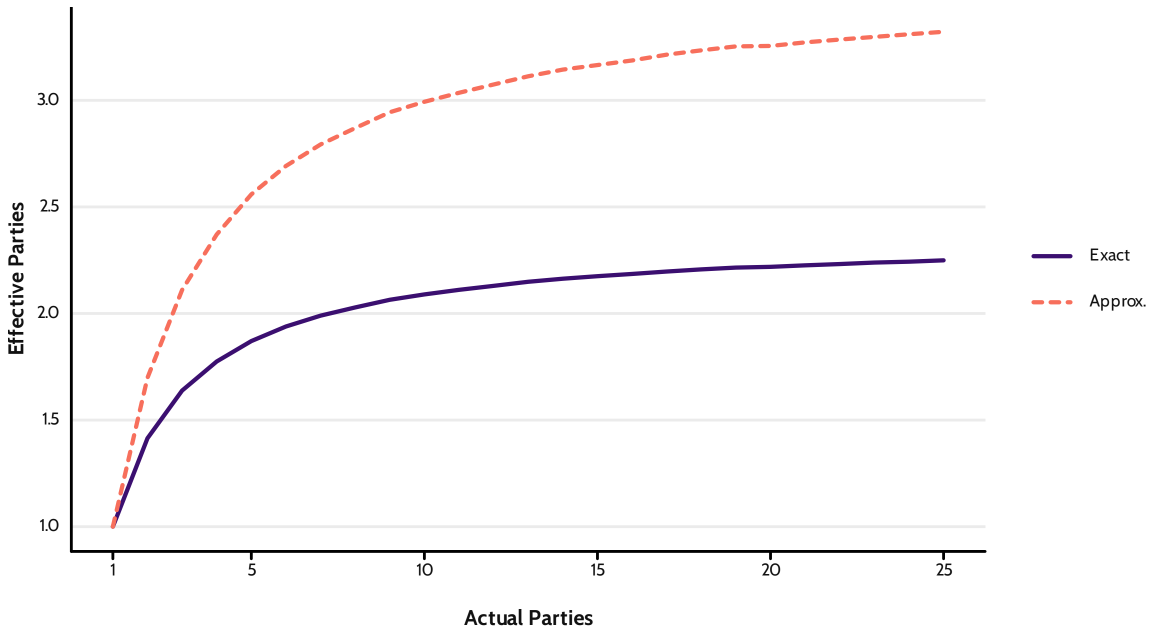

Figure 1: Simulation study showing that it is possible to approximate the effective number of parties using only the largest party share where the actual number of parties is low. However, as this number increases, so too does the the difference between the exact and approximate values.

all, the index acts to favour large parties (Laakso and Taagepera 1979). However, it also introduces a positive bias that affects further models downstream. For instance, Shugart and Taagepera (2017) use this approach to derive seat-product models of the effective number of *seat-winning* parties. These then inform their seat-product model of the effective number of *vote-winning* parties instead.

To understand the positive bias that this simplification introduces, imagine a small party system with only two seat-winning parties. Now assume that the first party wins 75% of the seats and the second whatever is left. If we use these two figures to compute the effective number of parties using the equation that Laakso and Taagepera (1979) give, we get $1/(0.75^2+0.25^2) = 1.6$. Now let us imagine that we have *estimated* only the largest party's seat share, much like Taagepera and Shugart (1993) and Shugart and Taagepera (2017). If we compute the index again, using this single figure alone, we get $1/0.75^2 \approx 1.78$, a difference of 0.18 parties. At first, this may seem only slight. But, as the number of actual parties increases, so too does this positive bias.

Figure 1 shows the results of a simulation study that simulates 20,000 plausible seat shares from a Dirichlet distribution for 1 to 25 actual seat-winning parties. In each case, I also compute the *exact* effective number of parties for each of these 20,000 simulated seat shares and the *approximate* effective number of parties that we would get if we used only the largest party's seat share. As we can see, the difference between the exact and approximate figures at first remains low, implying that the approximation works well where the actual number of seat-winning parties is small. However, as the actual number of seat-winning parties increases, a large difference between the exact and approximate values then soon comes to emerge. Indeed, where there are only ten actual parties, the average difference between the exact and approximate values approaches around one whole party. Clearly, this is less than ideal. Better, therefore, would we to derive a new seat-product model of the effective number of parties that avoids such simplification, thereby avoiding this bias.

Modelling the Effective Number of Seat-Winning Parties, N_{S2}

Laakso and Taagepera (1979) define the effective number of parties, N_2, as:

$$N_2 = \frac{1}{\sum_{i=1}^{N_0} p_i^2} \tag{1}$$

Where p is a set of party shares and N_0 a count of the number of share-winning parties. The index takes the shares, squares them, sums the result, and then takes its reciprocal. It is, therefore, the inverse of a simple weighted average. What makes the index so useful, however, is that it weights each party share *by itself*. As such, it lets us avoid any arbitrary thresholds, boundaries, or rules when measuring how party systems fragment or how many parties should "count".

While introducing the concept, Laakso and Taagepera consider a more general version of their formula that uses arbitrary exponents. The equation that they derive resembles Rényi entropy (1961)—a generalisation of Shannon entropy (1948), also known within political science as the "index of hyperfractionalization" (Wildgen 1971). The measure has a single parameter, α, that allows one to weight probabilities according to their size. Its equation is given as follows:

$$\mathrm{H}_\alpha(X) = \frac{1}{1-\alpha} \ln\left(\sum_{i=1}^{n} p_i^\alpha\right) \tag{2}$$

Further, note that, where $\alpha = 2$, the equation simplifies such that:

$$\mathrm{H}_2(X) = \ln\left(\frac{1}{\sum_{i=1}^{N} p_i^2}\right) \tag{3}$$

https://doi.org/10.33774/apsa-2024-1x8ft-v3 **ORCID:** https://orcid.org/0000-0001-8517-5018 Content not peer-reviewed by APSA.

As we can see, this is just the natural logarithm of the equation for the effective number of parties that Laakso and Taagepera (1979) define in their original paper. As such, the two quantities share an *exact* equivalence: the effective number of parties is just the exponent of the Rényi entropy with an α value of 2. This equivalence is useful, since information theoretic quantities like entropy often have well-known decompositions (Stone 2015). For instance, one relevant identity holds that:

$$\mathrm{H}_2(X) = \ln(N) - \mathrm{D}_2\left(p_X(x) \| P_U(X)\right) \tag{4}$$

Or, in other words, that the Rényi entropy, $\mathrm{H}_2(X)$, of some discrete probability distribution, X, equals the natural logarithm of the total number of elements in the distribution, N, minus the Rényi divergence, $\mathrm{D}_2\left(p_X(x) \| P_U(X)\right)$, which measures how much one probability distribution differs from another. Here, it measures the extent to which the distribution we observe, $p_X(x)$, differs from a uniform one with an equal number of elements, $P_U(X)$.

At this point, we know that, when $\alpha = 2$, the Rényi entropy equals the natural logarithm of the effective number of parties, N_2. Likewise, we also know that the number of elements in our probability distribution, p, is just the actual number of parties, N_0. If we substitute these figures into equation (4), exponentiate both sides, then apply the quotient rule, we can solve for N_2:

$$N_2 = \frac{N_0}{e^{\mathrm{D}_2\left(p_X(x) \| P_U(X)\right)}} \tag{5}$$

Since we know *only* the actual number of parties, N_0, we do not know the divergence term in the denominator. We can, however, derive a logical model of this quantity that tells us what value we should expect it to take in the absence of all other data. If we follow the steps that I set out above, we must first determine its logical bounds. The *smallest* value that this divergence can take is zero, where $p_X(x)$ and $P_U(X)$ are the same (i.e. where they do not diverge). The *largest* value that it can take, instead, occurs where all of the probability mass is concentrated on a single element. Further, note that we can compute the Rényi divergence between any pair of arbitrary distributions, p and q, using the following equation:

$$\mathrm{D}_\alpha(P \| Q) = \frac{1}{\alpha - 1} \ln\left(\sum_{i=1}^{n} \frac{p_i^\alpha}{q_i^{\alpha-1}}\right) \tag{6}$$

Thus, at the upper bound, where $p_1 = 1$ and all $q_i = 1/N$, the equation simplifies such that:

$$\mathrm{D}_2(P\|Q) = \frac{1}{2-1}\ln\left(\sum_{i=1}^{n}\frac{1^2}{\frac{1}{N_0}^1}\right) = \ln(N_0) \tag{7}$$

This tells us, therefore, that the Rényi divergence may vary from 0 to $\ln(N_0)$. Note, however, that the denominator in equation (5) also exponentiates this divergence. That would, then, imply that the denominator can take any value between a lower bound of $e^0 = 1$ and an upper bound of $e^{\ln(N_0)} = N_0$. Having determined these two extremes, we can take their geometric mean. This gives $\sqrt{1 \times N_0} = N_0^{1/2}$. If we then substitute this figure into equation (5), we get:

$$N_2 = \frac{N_0}{N_0^{1/2}} = N_0^{1/2} \tag{8}$$

So, much like the "cube root rule" holds that the cube root of a nation's population approximates the number of seats in its assembly (Taagepera 1972), this equation suggests a kind of "square root rule" that holds that—absent any other data—the square root of the actual number of parties, N_0, represents our best estimate of the effective number of parties, N_2. This is an important finding in its own right, but also establishes a relationship between these two quantities that requires no simplification. The next step, therefore, is to link the actual number of *seat-winning* parties, N_{S0}; the number of assembly seats in the lower house, S; and the average district magnitude, M, to derive a new seat-product model. Thankfully, this relationship is well established due to decades' of tireless work by Taagepera, Shugart, and others (for a thorough overview, see Shugart and Taagepera 2017). They show that the seat-product model that describes N_{S0} in terms of M and S is given as:

$$N_{S0} = (MS)^{1/4} \tag{9}$$

Which, when substituted into equation (8), gives our first new seat-product model:

$$N_{S2} = \left[(MS)^{1/4}\right]^{1/2} = (MS)^{1/8} \tag{10}$$

Modelling the Effective Number of Vote-Winning Parties, N_{V2}

Due to our previous steps, we already know how to convert actual to effective parties. Thus, we can compute the effective number of vote-winning parties with the following equation:[1]

$$N_{V2} = N_{V0}^{1/2} \tag{11}$$

Now we face a more challenging task: to model the number of vote-winning parties, N_{V0}, using the seat-product, MS. The issue is that vote-winning parties face much weaker constraints. At the lower bound, this is not such a problem, as the *smallest* number of vote-winning parties must be the number of *seat-winning* parties, N_{S0}. The upper bound, however, is much less straightforward.

Shugart and Taagepera (2017) argue that "the upper bound on parties that might be viable vote-earners is not knowable" (128). Though I agree that the upper bound on the number of vote-winning parties is *weak*, I am not so sure that it does not exist. The largest value that N_{V0} can take *in principle* is surely just the number of votes, since this would allow only one vote per party. But knowing this does not help much, since it would allow for millions of parties at most large-scale national elections. Instead, a more reasonable upper bound is perhaps the total number of parties *that contest an election*, N_C. As most parties win at least one vote, N_{V0} and N_C will often be equal. But that is not *always* the case, especially at small-scale elections (e.g. those that occur at the district level). Further, the number of contesting parties might also be known in advance, at least when an election is called, much like our other essential predictors: the district magnitude, M, and the number of seats, S.

So, if N_{V0} may vary between N_{S0} and N_C, we can compute the geometric mean of these bounds, then substitute N_{S0} for MS, since $N_{S0} = (MS)^{1/4}$, giving:

$$N_{V0} = (N_{S0} \times N_C)^{1/2} \tag{12}$$

$$= \left[(MS)^{1/4} \times N_C\right]^{1/2} \tag{13}$$

$$= \left(MSN_C^4\right)^{1/8} \tag{14}$$

[1] Shugart and Taagepera (2017) use N_{V0} to denote the "relevant number of vote-winning parties". But this quantity has a similar role in deriving their seat-product model as the actual number of vote-winning parties does here. Further, as it allows for a simple equivalence with the actual number of seat-winning parties, N_{S0}, I maintain this notation throughout.

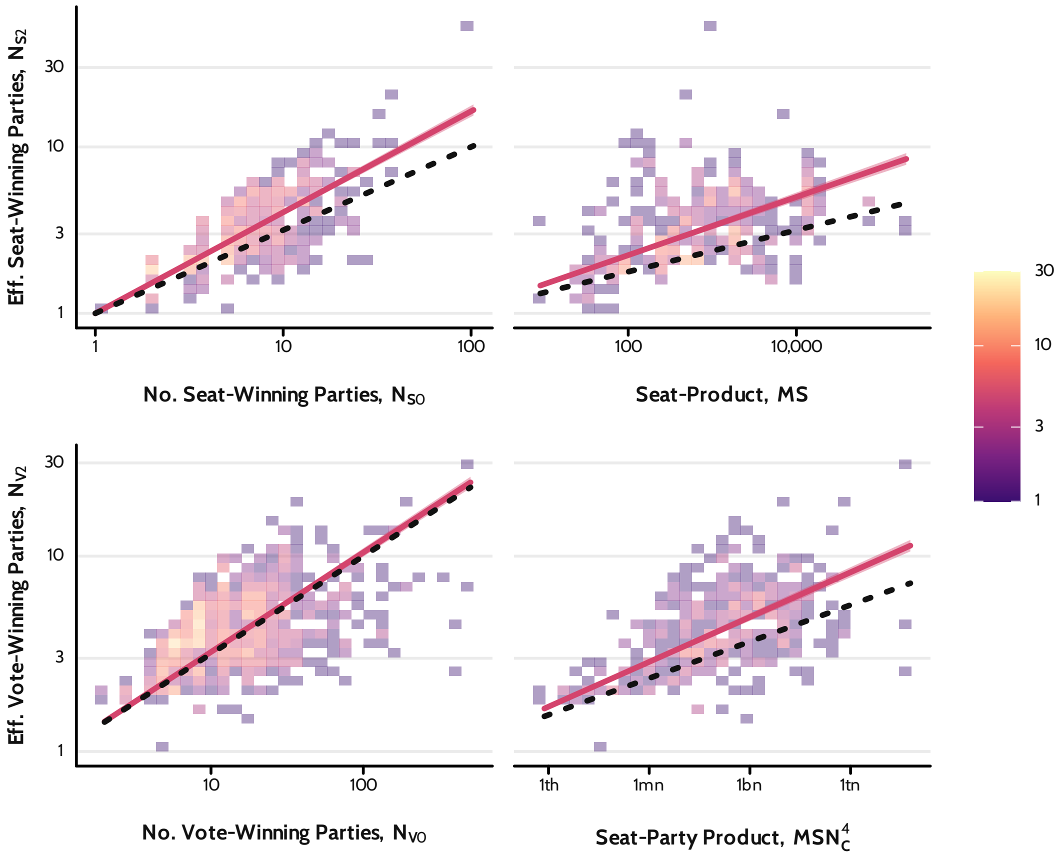

Figure 2: Predictions from four logical models predicting the effective number of parties as a function of the actual number of parties and the seat-product. All show a good fit to real world data from the Constituency-Level Elections Archive, aggregated to the national level. Note that black dashed lines show predictions from logical models and red solid ones from statistical lines of best fit.

Substituting this this model into equation (11) then gives a second new seat-product model of the effective number of vote-winning parties, here in terms of a combined "seat-party product", MSN_C^4:

$$N_{V2} = \left(MSN_C^4\right)^{1/16} \tag{15}$$

Predicting Real World Data

Having derived two new seat-product models, we can now compare them to real world data. For this purpose, I use the Constituency-Level Election Archive (Kollman et al. 2024), subset to include elections that score 0.5 or more on the V-Dem Electoral Democracy scale and that use "simple electoral systems" (Shugart and Taagepera 2017; Taagepera 2007). In total, these data include information on 992 lower house elections, in 93 countries, from 1851 to 2023.

Figure 2 shows how well each model compares to the real world data. The top row shows predictions of the effective number of *seat-winning* parties using the actual number of seat-winning parties (top left) and the seat-product (top right). The bottom row, instead, shows predictions of the effective number of *vote-winning* parties, again made using the actual number of vote-winning parties (bottom left) and the "seat-party product" (bottom right). As a reminder, *none* of these models represent statistical lines of best fit. Rather, *all* rely on only deductive logic and have yet to see any real world data.

Each model shows an impressive fit to the data, which I judge here using their respective root mean squared errors (RMSE), computed using log-scaled differences, then exponentiated to return the figures back to the original scale. If we consider first seat-winning parties, the model shown on the top left has an RMSE of 1.52 and the model shown on the top right of 5.93. This is all well and good, but we need an appropriate baseline against which to compare these figures. As such, I also compute the average "null" RMSE that we would get were we to replace the exponent in each of our equations with some random value between zero and one, again using 20,000 unique simulations. This gives null RMSEs of 2.03 and 42.51, respectively, demonstrating the models' goodness of fit. The same is true for models of vote-winning parties as well. The model shown on the bottom left has an RMSE of 1.57 and the model shown on the bottom right of 1.64. Again, this compares well to the average null of 2.51 and 3,085,307.[2]

Each model also compares well to statistical lines of best fit estimated from the data, despite having seen no data at all. For example, the logical model of N_{S2} as a function of N_{S0} (top left) suggests an exponent of $1/2 = 0.5$ whereas the statistical model estimates an exponent of 0.61 and the logical model of N_{S2} as a function of the seat-product, MS (top right), suggests an exponent of $1/8 = 0.125$ whereas the statistical model estimates an exponent of 0.17. Models of the effective number of vote-winning parties performed better still, with the logical model of N_{V2} as a function of N_{V0} (bottom left) suggesting an exponent of $1/2 = 0.5$ compared to an estimated exponent of 0.51 and the logical model of N_{V2} as a function of the seat-party product, MSN^4 (bottom right), suggesting an exponent of $1/16 = 0.0625$ compared to an estimated coefficient of 0.076.

[2]This extremely large value occurs because the seat-party product, MSN^4, becomes very large very quickly. Taking it to the power of one over sixteen, as equation (15) does, reduces its size by several orders of magnitude. However, for larger random exponents, this is not the case, thereby leading to much larger predictions.

Conclusion

If we hope to understand party-system fragmentation, we need scientific models that explain how this process works. Seat-product models offer our best explanation. But they also require a simplification—that the effective number of parties is a function of the largest party alone—which then introduces a positive bias. I provide a solution to this "seat-product problem", deriving new seat-product models with no need for simplification.

One important by-product is a second "root rule" of elections. The first, Taagepera's (1972) "cube root rule", shows that the cube root of a nation's population approximates the number of seats in its assembly. The second "square root rule" that I identify here, instead, shows that the square root of the actual number of parties approximates the effective number of parties instead. This pleasing symmetry extends further still: any seat-product model of any effective number of parties is simply the seat-product model of the equivalent actual number of parties to the power of one over two.

As well as being our best explanation for party-system fragmentation, seat-product models of the effective number of parties also underpin other logical models (for example, of Gallagher's (1991) deviation from proportionality). As such, the solutions that I identify here have the prospect to feed forward into these models, thereby improving the predictions they make. Likewise, they might also serve as a basis for resurrecting lost data, since they would allow us to impute likely values where incomplete electoral records have been lost to the passage of time.

References

Gallagher, Michael. 1991. "Proportionality, Disproportionality and Electoral Systems." *Electoral Studies* 10 (1): 33–51. https://doi.org/10.1016/0261-3794(91)90004-C.

Kollman, Ken, Allen Hicken, Daniele Caramani, David Backer, and David Lublin. 2024. "Constituency-Level Elections Archive." Ann Arbor, MI.

Laakso, Markku, and Rein Taagepera. 1979. "'Effective' Number of Parties: A Measure with Application to West Europe." *Comparative Political Studies* 12 (1): 3–27. https://doi.org/10.1177/001041407901200101.

Rényi, Alfred. 1961. "On Measures of Entropy and Information." In *Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics and Probability 1960*, 1:547–61. Berkeley, CA.

Shannon, C E. 1948. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27 (3): 379–423.

Shugart, Matthew S., and Rein Taagepera. 2017. *Votes from Seats: Logical Models of Electoral Systems*. Cambridge, UK: Cambridge University Press.

Stone, James V. 2015. *Information Theory: A Tutorial Introduction*. 1st ed. Sheffield, UK: Sebtel Press.

Taagepera, Rein. 1972. "The Size of National Assemblies." *Social Science Research* 1 (4): 385–401. https://doi.org/10.1016/0049-089X(72)90084-1.

———. 2007. *Predicting Party Sizes: The Logic of Simple Electoral Systems*. Oxford: Oxford University Press.

———. 2008. *Making Social Sciences More Scientific: The Need for Predictive Models*. Oxford, UK: Oxford University Press.

Taagepera, Rein, and Matthew S. Shugart. 1993. "Predicting the Number of Parties: A Quantitative Model of Duverger's Mechanical Effect." *American Political Science Review* 87 (2): 455–64. https://doi.org/10.2307/2939053.

Wildgen, John K. 1971. "The Measurement of Hyperfractionalization." *Comparative Political Studies* 4 (2): 233–43. https://doi.org/10.1177/001041407100400205.