Jarosław Flis[1], Wojciech Słomczyński[2], Dariusz Stolicki[3]

# Pot and Ladle: Estimating Seat Shares from Aggregate Vote Numbers under the d'Hondt Electoral System

WORK IN PROGRESS

PLEASE DO NOT CITE WITHOUT THE AUTHORS' PERMISSION

The Jefferson-d'Hondt method – originally devised in 1792 by Thomas Jefferson to apportion seats in the U.S. House of Representatives among the states (Balinski & Young 1978), and later proposed by a Belgian mathematician and lawyer Victor d'Hondt (d'Hondt 1878, 1882) for use in parliamentary elections – is among the most popular formulas for allocating parliamentary seats to party lists in proportional representation electoral systems (see generally Colomer 2004). It is currently used to allocate all parliamentary seats in, inter alia, Argentina, Belgium, Bulgaria, Cape Verde, the Czech Republic, Finland, Israel, Macedonia, Netherlands, Paraguay, Peru, Poland, Portugal, Spain, and Turkey, as well as nearly all seats in Croatia. It is also employed – together with other methods – in Austria, Denmark, Iceland, and Japan, and has been historically used in, among others, Sweden and Germany.

It is well known that the Jefferson-d'Hondt formula is biased in favor of larger parties (see, e.g., Sainte-Laguë 1910; Pólya 1918a, 1918b, 1919; Gallagher 1991; Benoit 2000; Marshall, Olkin & Pukelsheim 2002; van Eck et al. 2005; Pukelsheim 2013). The magnitude of such bias has been estimated for fixed-sized parties by Janson (2014) and for random-sized parties by Schuster et al. (2003), Schwingenschlögl & Drton (2004), and Drton & Schwingenschlögl (2005). Unfortunately, earlier research has focused exclusively on a single-district scenario, while most countries employing the Jefferson-d'Hondt method (with the exception of Netherlands and Israel) conduct allocate seats within each of the multiple electoral districts separately. In those countries, the political effects of the advantage provided by Jefferson-d'Hondt to larger parties can only be assessed on the national scale – when looking at the composition of the legislature as a whole.

Drawing on the article on the benefits of interparty consolidation and on the effects of apparentments by Bochsler (2010), as well as on the earlier works on seat bias, we have developed a robust formula for estimating parliamentary seat allocation solely on the basis of

[1] Jagiellonian Center for Quantitative Research in Political Science / Chair of Media Research, Institute of Journalism, Media and Social Communication, Faculty of Management and Social Communication, Jagiellonian University.
[2] Jagiellonian Center for Quantitative Research in Political Science / Chair of Applied Mathematics, Institute of Mathematics, Faculty of Mathematics and Computer Science, Jagiellonian University.
[3] Jagiellonian Center for Quantitative Research in Political Science / Chair of Constitutionalism and Government Systems, Institute of Political Science and International Relations, Faculty of International and Political Studies, Jagiellonian University.

nationwide electoral results and constant parameters of the electoral system. The expected number of seats for an *i*-th party ($s_i$) is as follows:

$$s_i := p_i \cdot s + p_i \cdot \frac{cn}{2} - \frac{c}{2} \tag{0.1}$$

where $p_i$ is (effective) percentage of votes cast for that party, $s$ is the total number of seats, $c$ is the number of electoral districts, and $n$ is the number of parties participating in the seat allocation process.

Metaphorically, the formula can be thought to consist of three components: a *basic yield* – the number of seats that a party would be obtain under a purely proportional system (without any form of rounding to an integer), a *contribution* to the common *bounty pot*, and a *bounty* drawn from that pot. The key feature of the system lies in the fact that while the contribution to the bounty pot is equal for all parties (i.e., does not depend on the vote share), the size of the bounty depends on the party's *ladle* – which is proportional to the vote share: larger parties get greater servings. Accordingly, small parties are disadvantaged, since they contribute more than they get back from the pot, while large parties receive a bonus. What the formula makes clear, however, is that the size of that bonus depends not only on their vote shares, but also on the size of the bounty pot – which is a function of the number of districts and of the number of parties.

In addition to explaining the magnitude of the nationwide bonus for the winning parties, the formula has a purely practical application. All divisor methods for seat allocation require that district-level results be known, and since the Jefferson-d'Hondt method can be sensitive to small variations in vote shares, they have to be known exactly. In contrast therewith, our formula provides a reasonably good estimate of the nationwide seat allocation results while requiring only aggregate party vote shares to be known. Hence it can be used to accurately model seat allocation on the basis of opinion polls, exit polls, and partial election results, when aggregate vote shares are all that is known.

In part I of this paper we explain how our formula can be derived from the d'Hondt method – and what assumptions must be made for it to work correctly. In part II, we analyze empirical data from six European countries to demonstrate that the formula provides a reasonably accurate estimate of actual seat allocation results and is quite robust against minor violations of its assumptions. Finally in part III, we discuss the formula's systemic consequences for the ordering of national political scenes. All three parts are designed to be sufficiently independent of each other to permit a reader to skip immediately to the one he is interested in.

## I. Mathematical underpinnings

A common formulation of the d'Hondt method of seat allocation is an algorithmic one. Let $s$ be the number of seats to be allocated within a district and $v_i$ be the number of votes cast for the *i*-th party in that district. We define an *m*-th quotient for the *i*-th party as follows:

$$q_{i,m} := \frac{v_i}{m} \tag{1.1}$$

Let $q_s$ be the $s$-th highest quotient overall (i.e., across all parties). The number of seats allocated to the $i$-th party is then defined as[4]:

$$s_i := \max\{m = 1, \dots, s \colon q_{i,m} \geq q_s\}. \tag{1.2}$$

What may be surprising is that this well-known algorithmic version actually differs from the original formula proposed by d'Hondt (1878, 1882), which closely tracked a 1792 proposal by Thomas Jefferson for apportioning House seats among the states in the U.S. Congress (Jefferson 1792) (though it is unclear whether d'Hondt knew of Jefferson's work on the subject). Jefferson's method called for finding such a divisor $D$ that if each party (or state) were to be allocated as many seats as $v_i$ divided by $D$, rounding down to the nearest integer, i.e., $s_i := \left\lfloor \frac{v_i}{D} \right\rfloor$, no seats would remain unallocated (i.e. $\sum_i s_i = s$). D'Hondt even proposed a specific method for arriving at such a divisor (p. 20): let $Q$ be the simple quota, $Q := \frac{v}{s}$, where $v$ is the total number of votes cast for all parties ($v := \sum_i v_i$). In the first allocation step, each party would be allocated the number of seats equal to its number of votes divided by $Q$ and rounded down to the nearest integer, i.e., $s_i := \left\lfloor \frac{v_i}{Q} \right\rfloor$. After this step, the number of allocated seats ($s'$) would be between $s - n$ and $s$, where $n$ is the number of parties taking part in the allocation. If no seats remained unallocated, the algorithm would end with $D = Q$, but otherwise we would compute quotients $q_i := \frac{v_i}{(s_i + 1)}$. The $k$-th highest quotient ($k := s - s'$) would be the highest value of $D$ (and the next highest quotient would be the infimum of the range of possible $D$-s).

The two methods are equivalent, i.e., guaranteed to generate identical allocation of seats: for each $i$ $s_{Alg_i} = s_{Div_i}$, where $s_{Alg_i}$ is the number of seats awarded to the $i$-th party under formula (1.2) and $s_{Div_i} := \left\lfloor \frac{v_i}{D} \right\rfloor$. Let $D \in (q_{s+1}, q_s]$. For each $i$ we know that:

$$\frac{v_i}{s_{Alg_i} + 1} < D \leq \frac{v_i}{s_{Alg_i}}. \tag{1.3}$$

Hence

$$v_i < D s_{Alg_i} + D \tag{1.4}$$

and

$$D s_{Alg_i} \leq v_i; \tag{1.5}$$

therefore

$$D s_{Alg_i} \leq v_i < D s_{Alg_i} + D. \tag{1.6}$$

By dividing both sides by $D$ we obtain

[4] If $\max\{j \in \square \colon q_{s+j} = q_s\} > 0$, the procedure described herein will allocate more than $s$ seats. This condition is equivalent to an election tie and has to be resolved by reference to some rule external to the system being described. From the point of the Jefferson-d'Hondt method, any such resolution is arbitrary, and such cases will therefore be ignored in our analysis. They are in any case extremely rare in real-life elections.

$$s_{Alg_i} \leq \frac{v_i}{D} < s_{Alg_i} + 1 \tag{1.7}$$

which is equivalent to

$$s_{Alg_i} = \left\lfloor \frac{v_i}{D} \right\rfloor \tag{1.8}$$

q.e.d.

Unfortunately, the divisor formula still does not permit a party's seat allocation to be estimated without full knowledge of vote shares of the other parties. A solution to this problem has been proposed by Bochsler (2010), who was drawing on earlier works by Gfeller (1890) and Happacher and Pukelsheim (1996). Let $m := \frac{v}{D}$ (it should be noted that *m* need not be an integer). Then

$$s_i = \left\lfloor \frac{v_i}{\frac{v}{m}} \right\rfloor = \left\lfloor \frac{v_i m}{v} \right\rfloor = \lfloor p_i m \rfloor, \tag{1.9}$$

where $p_i$ is the fraction of total votes cast for the *i*-th party (the "vote share"). Let us denote the remaining fractional part of $p_i m$ as $r_i := p_i m - \lfloor p_i m \rfloor$. We know that allocated seats sum up to *s*, so

$$s = \sum_i s_i = \sum_i (p_i m - x_i) = m \sum_i p_i - \sum_i x_i = m - \sum_i r_i. \tag{1.10}$$

At this point Bochsler assumes that the remaining fractional parts are drawn at random from a continuous uniform distribution $U(0,1)$[5]. In fact, this assumption goes further than necessary, as he only uses one consequence thereof: that

$$E(r) = \frac{1}{2}, \tag{1.11}$$

permitting him to obtain value of *m* from (1.10):

$$m = s + \frac{n}{2}. \tag{1.12}$$

By substituting (1.12) for *m* in (1.9) we obtain

$$s_i = \left\lfloor p_i \left( s + \frac{n}{2} \right) \right\rfloor = p_i \left( s + \frac{n}{2} \right) - r_i\,. \tag{1.13}$$

But since $E(r) = \frac{1}{2}$, the expected number of seats for the *i*-th party can be expressed as:

$$E(s_i) = p_i \left( s + \frac{n}{2} \right) - E(x_i) = p_i \left( s + \frac{n}{2} \right) - \frac{1}{2}. \tag{1.14}$$

It is notable that an identical formula has been obtained by Janson (2014), but in a very different context. While Bochsler treated the number of seats as constant and the distribution of votes across competing parties (with the exception of the *i*-th party) as random, Janson

[5] Such an assumption has been made by many authors treating of apportionment methods, beginning with Jefferson himself: "The probability is that the fractions will generally descend gradually from 29,999 [30,000 being the proposed quota] to 1" (Jefferson 1792).

analyzed a case with a constant distribution of votes, but random number of seats (or, to be exact, an asymptotic case with the number of seats drawn from a uniform distribution $U(1,n)$ for $n \to \infty$).

Unfortunately, the key assumption of the Bochsler formula – uniform distribution of remaining fractional parts – is incorrect. The distribution in question will be uniform if and only if the party vote shares themselves are distributed uniformly. As noted above, however, the uniformity assumption is not necessary for the Bochsler formula to work – it suffices that $E(r_i) = \frac{1}{2}$, and we will proceed to demonstrate that under ordinary circumstances this assumption is satisfied (though in most cases asymptotically rather than exactly).

Let $m := \frac{v}{D}$ . It will be noted that the remaining fractional part of $p_i m$ ($r_i$) will be equal to $x$ if and only if $p_i = \frac{j+x}{m}$ for some $j = 0, \dots, \lfloor m \rfloor$. Therefore, $r_i$ will be distributed according to the following density function:

$$f_X(x) := \frac{1}{s} \sum_{k=0}^{\lfloor m \rfloor} f_i\left(\frac{k+x}{m}\right), \tag{1.15}$$

where $f_i(x)$ is the probability density function of the vote share distribution for the $i$-th party. Since expected value of a continuous probability distribution is equal to $\int_{-\infty}^{\infty} x f(x) dx$, and since each $f_i(x)$ is nonzero only in the $[0,1]$ range, expected value of the remaining fractional part will be:

$$E(r_i) = \int_0^1 \frac{x}{m} \sum_{k=0}^{\lfloor m \rfloor} f_i\left(\frac{k+x}{m}\right) dx. \tag{1.16}$$

By substituting $y = \frac{k+x}{m}$ we obtain:

$$E(r_i) = \sum_{k=0}^{\lfloor m \rfloor} \int_{\frac{k}{m}}^{\frac{k+1}{m}} \frac{my-k}{m} f(y)\, m\, dy \tag{1.17}$$

and

$$E(r_i) = m \int_0^m y f_i(y) dy - \sum_{k=1}^{\lfloor m \rfloor} k \int_{\frac{k}{m}}^{\frac{k+1}{m}} f_i(y) dy. \tag{1.18}$$

But $\int_0^m y f_i(y) dy$ is the expected vote share of the $i$-th party (which we will denote as $\mu_i$). Therefore

$$E(r_i) = m\mu_i - \sum_{k=1}^{\lfloor m \rfloor} k \int_{\frac{k}{m}}^{\frac{k+1}{m}} f_i(y) dy. \tag{1.19}$$

$\int_{\frac{k}{m}}^{\frac{k+1}{m}} f_i(y) dy$ is $F_i\left(\frac{k}{m}\right) - F_i\left(\frac{k+1}{m}\right)$, so we transform the equation:

$$E(r_i) = m\mu_i - \sum\nolimits_{k=1}^{\lfloor m \rfloor} k\left(F_i\left(\frac{k+1}{m}\right) - F_i\left(\frac{k}{m}\right)\right) \tag{1.20}$$

$$E(r_i) = m\mu_i - \sum\nolimits_{k=1}^{\lfloor m \rfloor - 1} k\left(F_i\left(\frac{k+1}{m}\right) - F_i\left(\frac{k}{m}\right)\right) - \lfloor m \rfloor\left(F_i\left(\frac{\lfloor m \rfloor + 1}{m}\right) - F_i\left(\frac{\lfloor m \rfloor}{m}\right)\right) \tag{1.21}$$

$$E(r_i) = m\mu_i - \left(-\sum_{k=1}^{\lfloor m \rfloor} F_i\left(\frac{k}{m}\right) + \lfloor m \rfloor F_i\left(\frac{\lfloor m \rfloor}{m}\right)\right) - \lfloor m \rfloor\left(1 - F_i\left(\frac{\lfloor m \rfloor}{m}\right)\right) \tag{1.22}$$

$$E(r_i) = m\mu_i + \sum_{k=1}^{\lfloor m \rfloor} F_i\left(\frac{k}{m}\right) - \lfloor m \rfloor F_i\left(\frac{\lfloor m \rfloor}{m}\right) - \lfloor m \rfloor + \lfloor m \rfloor F_i\left(\frac{\lfloor m \rfloor}{m}\right). \tag{1.23}$$

And finally we obtain a (relatively) simple formula for expected value of $r_i$:

$$E(r_i) = m\mu_i - \lfloor m \rfloor + \sum_{k=1}^{\lfloor m \rfloor} F_i\left(\frac{k}{m}\right). \tag{1.24}$$

There are several ways to demonstrate that

$$\lim_{m\to\infty} E(r_i) = \frac{1}{2}. \tag{1.25}$$

For instance, it can be noted that

$$\frac{1}{m}\sum_{k=0}^{\lfloor m \rfloor} f_i\left(\frac{k+x}{m}\right) \tag{1.26}$$

is a Riemannian sum and, consequently, that

$$\frac{1}{m}\sum_{k=0}^{\lfloor m \rfloor} f_i\left(\frac{k+x}{m}\right) \sim \int_0^1 f_i(x)\,dx \tag{1.27}$$

as *m* approaches ∞. Since $f_i$ is a density function on [0,1], it integrates to 1. Therefore

$$\lim_{m\to\infty} E(r_i) = \int_0^1 x\,dx = \frac{1}{2}. \tag{1.28}$$

Owens (2014) proves that as long as $f_i$ is differentiable and its first derivative is integrable, for any $x \in [0,1]$ error term $\Delta_n = \int_0^1 f_i(x)dx - \frac{1}{n}\sum_{k=1}^{n} f_i\left(\frac{k-x}{n}\right)$ is asymptotically equivalent to $\left(x - \frac{1}{2}\right)\frac{f_i(1) - f_i(0)}{n}$. Therefore, as long as the distribution of party vote shares is not highly asymmetric at the ends of the [0,1] range[6] or its mean is close to $\frac{1}{2}$ (assumption **A1**), the

[6] Running kernel density estimation on empirical vote share distributions for parties in the European countries using proportional representation systems in lower house elections indicates that party vote share distribution is

expected value of the remaining fractional share will converge to $\frac{1}{2}$ fairly quickly – for instance, more rapidly than $\frac{1}{n}$ converges to 0.

If **A1** is not satisfied and error term is not asymptotically equivalent to 0, we can write

$$\lim_{m\to\infty} E(r_i) = \int_0^1 x\left(1+\left(x-\frac{1}{2}\right)\frac{f_i(1)-f_i(0)}{m}\right)dx = \\ \int_0^1 \left(x^2\frac{f_i(1)-f_i(0)}{m}-\frac{x}{2}\frac{f_i(1)-f_i(0)}{m}+x\right)dx. \tag{1.29}$$

Let

$$A(x) = \int\left(x^2\frac{f_i(1)-f_i(0)}{m}-\frac{x}{2}\frac{f_i(1)-f_i(0)}{m}+x\right)dx = \\ \frac{f_i(1)-f_i(0)}{3m}x^3-\frac{f_i(1)-f_i(0)}{4m}x^2+\frac{1}{2}x^2\,. \tag{1.30}$$

By applying the fundamental theorem of calculus we arrive at

$$\int_0^1 \left(x^2\frac{f_i(1)-f_i(0)}{n}-\frac{x}{2}\frac{f_i(1)-f_i(0)}{n}+x\right)dx = A(1)-A(0) = \\ \frac{f_i(1)-f_i(0)}{3m}-\frac{f_i(1)-f_i(0)}{4m}+\frac{1}{2}-0 = \\ \frac{1}{2}+\frac{f_i(1)-f_i(0)}{12m}, \tag{1.31}$$

leading to a party-specific error estimate of

$$err_i = p_i\left(\frac{n}{2}-\sum_i\left(\frac{1}{2}+\frac{f_i(1)-f_i(0)}{12m}\right)\right)-\frac{1}{2}+\frac{1}{2}+\frac{f_i(1)-f_i(0)}{12m} = \\ p_i\left(-\sum_i\left(\frac{f_i(1)-f_i(0)}{12m}\right)\right)+\frac{f_i(1)-f_i(0)}{12m}. \tag{1.32}$$

While the Bochsler-Janson formula enables us to estimate single-district seat numbers on the basis of an approximate, rather than exact, distribution of votes across parties, it still requires that the distribution of votes across districts be known (excepting, of course, the degenerate case of a single-district electoral system). Yet this difficulty can also be surmounted, as long as certain additional assumptions about such distribution are made.

Let $v_i^k$ and $s_i^k$ be, respectively, the number of votes received and the number of seats awarded for the *i*-th party in the *k*-th electoral district. Let $v_i = \sum_k v_i^k$ and $s_i = \sum_k s_i^k$ be, respectively, the nationwide vote and seat numbers for the *i*-th party, and let $v^k = \sum_i v_i^k$ and $s^k = \sum_i s_i^k$ be,

---

usually close to a truncated Gaussian distribution and in most cases the value of the density function is 0 or very close 0 at both ends of the [0,1] range.

respectively, the district vote and seat totals for the $k$-th district. Finally, let $v = \sum_i v_i = \sum_k v_k$ and $s = \sum_i s_i = \sum_k s_k$ be, respectively, nationwide vote and seat totals. We assume the number of parties ($n$) to be constant across districts (assumption **A2**). In this case:

$$\begin{gathered} s_i = \sum_k \left( \frac{v_i^k}{v^k} \cdot \left( s^k + \frac{n}{2} \right) - \frac{1}{2} \right) = \\ \sum_k \left( \frac{v_i^k}{v^k} \cdot s^k + \frac{v_i^k}{v^k} \cdot \frac{n}{2} \right) - \frac{c}{2} = \\ \sum_k \left( \frac{v_i^k}{v^k} \cdot s^k \right) + \sum_k \left( \frac{v_i^k}{v^k} \cdot \frac{n}{2} \right) - \frac{c}{2}. \end{gathered} \tag{1.33}$$

For the easiest case, where district size (in votes as well as in seats) and distribution of party vote shares are constant across districts (i.e. for each $k$ $v^k = v^1 = \frac{v}{c}$ and $s^k = s^1 = \frac{s}{c}$, and for each $k$, $i$ $v_i^k = v_i^1 = \frac{v_i}{c}$), we can simply write

$$\begin{gathered} s_i = \sum_k \left( \frac{\frac{v_i}{c}}{\frac{v}{c}} \cdot \frac{s}{c} \right) + \sum_k \left( \frac{\frac{v_i}{c}}{\frac{v}{c}} \cdot \frac{n}{2} \right) - \frac{c}{2} = \\ c \left( \frac{v_i}{v} \cdot \frac{s}{c} \right) + c \left( \frac{v_i}{v} \cdot \frac{n}{2} \right) - \frac{c}{2} = \\ p_i s + p_i \frac{cn}{2} - \frac{c}{2}, \end{gathered} \tag{1.34}$$

which is our formula (0.1). Of course, those three assumptions are highly unrealistic and have not been exactly satisfied in any election. However, let us consider whether the formula might still work in other cases. We note that

$$\begin{gathered} p_i s + p_i \frac{cn}{2} - \frac{c}{2} = \\ \frac{\sum_k v_i^k}{\sum_k v^k} \cdot \sum_k s^k + \frac{\sum_k v_i^k}{\sum_k v^k} \cdot \frac{cn}{2} - \frac{c}{2}. \end{gathered} \tag{1.35}$$

The two formulas will be equal if and only if:

$$\sum_k \left( \frac{v_i^k}{v^k} \cdot s^k + \frac{v_i^k}{v^k} \cdot \frac{n}{2} \right) - \frac{c}{2} = \frac{\sum_k v_i^k}{\sum_k v^k} \cdot \sum_k s^k + \frac{\sum_k v_i^k}{\sum_k v^k} \cdot \frac{cn}{2} - \frac{c}{2}. \tag{1.36}$$

This condition will be satisfied if:

$$\sum_k \left( \frac{v_i^k}{v^k} \cdot \frac{n}{2} \right) = \frac{\sum_k v_i^k}{\sum_k v^k} \cdot \frac{cn}{2} \tag{1.37}$$

and

$$\sum_k \left( \frac{v_i^k}{v^k} \cdot s^k \right) = \frac{\sum_k v_i^k}{\sum_k v^k} \cdot \sum_k s^k. \tag{1.38}$$

(Or if neither (1.37) nor (1.38) holds, but the differences cancel each other out.)

Let us first take up condition (1.37):

$$\sum_k \left(\frac{v_i^k}{v^k} \cdot \frac{n}{2}\right) = \frac{\sum_k v_i^k}{\sum_k v^k} \cdot \frac{cn}{2} \tag{1.39}$$

$$\sum_k \frac{v_i^k}{v^k} = \frac{\sum_k v_i^k}{\sum_k v^k} \cdot c \tag{1.40}$$

$$\frac{1}{c} \cdot \sum_k p_i^k = \frac{\sum_k \left(v^k \cdot p_i^k\right)}{\sum_k v^k} \tag{1.41}$$

$$\frac{1}{c} \cdot \sum_k p_i^k \sum_k v^k = \sum_k \left(v^k \cdot p_i^k\right) \tag{1.42}$$

— which is true if party vote shares ($p_i^k$) are not correlated with district magnitude ($v^k$) **in terms of the total number of votes actually cast (not in terms of the number of eligible voters)** **A3**

Condition (1.38) is somewhat more complex:

$$\sum_k \left(\frac{v_i^k}{v^k} \cdot s^k\right) = \frac{\sum_k v_i^k}{\sum_k v^k} \cdot \sum_k s^k \tag{1.43}$$

Relying on (1.42), we know that:

$$\sum_k v_i^k = \sum_k \left(\frac{v_i^k}{v^k} \cdot v^k\right) = \frac{1}{c} \cdot \sum_k \frac{v_i^k}{v^k} \sum_k v^k \,. \tag{1.44}$$

Hence

$$\sum_k \left(\frac{v_i^k}{v^k} \cdot s^k\right) = \frac{\frac{1}{c} \cdot \sum_k \frac{v_i^k}{v^k} \sum_k v^k}{\sum_k v^k} \cdot \sum_k s^k \tag{1.45}$$

and

$$\sum_k \left(p_i^k \cdot s^k\right) = \frac{1}{c} \cdot \sum_k p_i^k \cdot \sum_k s^k \tag{1.46}$$

— which is true if party vote shares ($p_i^k$) are not correlated with district magnitude ($s^k$) **in terms of the total number of seats to be allocated** **A4**

If assumptions assumption **A2**, **A3** and **A4** are satisfied, by aggregating Bochsler-Janson single-district formulas we arrive at formula (0.1), q.e.d.

Before we proceed to the empirical test of the formula (0.1), one additional (albeit somewhat extended) remark is in order. We have been hitherto treating the number of parties *n* as constant. In reality, however, it depends on vote shares due to existence of statutory and natural election thresholds. Statutory thresholds are easy to account for: since parties that fall below them are ignored in the seat allocation process, we simply eliminate them from the electoral results data set before starting any processing. Natural thresholds call for more sophisticated treatment, since they cannot be fixed at arbitrary level without knowing actual election results. Yet we cannot ignore them, since not only applying formula (0.1) for sub-threshold parties can yield negative seat numbers (which are obviously incorrect), but including them in *n* tends to distort results for supra-threshold parties as well.

To avoid this kind of error, we use an iterative algorithm for determining effective *n* (the number of relevant parties) and for identifying supra-threshold parties. The basic strategy is as follows: we sort the parties degressively according to their total number of votes ($v_i$). We then start with only one party in the model (the largest one) and continue to add parties, according to the sort order, until we encounter the first party with negative seat number. At that point, we eliminate such party and end the algorithm.

Let *N* be the whole number of parties (including sub-threshold ones, but excluding those eliminated by the statutory threshold) that participate in seat allocation. In such case, we can define *n* as follows:

$$n = \max\left\{n \in (1, \dots, N): n > \frac{1}{p_n} - 2\bar{s}\right\} \tag{1.47}$$

where $\bar{s} := \frac{s}{c}$ is the mean district size. Of course, since $\sum_{i=1}^{n} p_i$ must be equal 1, we have to renormalize vote totals:

$$v = \sum_{i=1}^{n} v_i \,. \tag{1.48}$$

We then arrive at the full formula:

$$\begin{gathered} n = \max\left\{n \in (1, \dots, N): n > \frac{\sum_{i=1}^{n} v_i}{v_n} - 2\bar{s}\right\} = \\ \max\left\{n \in (1, \dots, N): n > \sum_{i=1}^{n} \frac{v_i}{v_n} - 2\bar{s}\right\}. \end{gathered} \tag{1.49}$$

For our algorithm to be correct, the condition $n > \sum_{i=1}^{n} \frac{v_i}{v_n} - 2\bar{s}$ has to be monotonous. We can transform it into

$$2\bar{s} > \sum_{i=1}^{n} \frac{v_i}{v_n} - n, \tag{1.50}$$

which is equivalent to

$$2\bar{s} > \sum_{i=1}^{n} \frac{v_i - v_n}{v_n}. \tag{1.51}$$

Theorem:

$$2\bar{s} > \sum_{i=1}^{n} \frac{v_i - v_n}{v_n} \quad \text{implies} \quad 2\bar{s} > \sum_{i=1}^{n+1} \frac{v_i - v_{n+1}}{v_{n+1}} \tag{1.52}$$

Proof:

$$\sum_{i=1}^{n} \frac{v_i - v_n}{v_n} < \sum_{i=1}^{n+1} \frac{v_i - v_{n+1}}{v_{n+1}} \tag{1.53}$$

$$\sum_{i=1}^{n} \frac{v_i}{v_n} - n < \sum_{i=1}^{n+1} \frac{v_i}{v_{n+1}} - (n+1) \tag{1.54}$$

$$\left(\sum_{i=1}^{n-1} v_i\right) v_{n+1} - n v_n v_{n+1} < \left(\sum_{i=1}^{n} v_i\right) v_n - (n+1) v_n v_{n+1} \tag{1.55}$$

$$\left(\sum_{i=1}^{n-1} v_i\right) v_{n+1} < \left(\sum_{i=1}^{n} v_i\right) v_n - v_n v_{n+1} \tag{1.56}$$

$$\left(\sum_{i=1}^{n} v_i\right) v_{n+1} < \left(\sum_{i=1}^{n} v_i\right) v_n \tag{1.57}$$

which is true, as parties are sorted degressively by $v_i$ .

We can further note that if

$$\frac{v_n}{\sum_{i=1}^{n} v_i} > \frac{1}{2\bar{s}} \tag{1.58}$$

the condition $n > \sum_{i=1}^{n} \frac{v_i}{v_n} - 2\bar{s}$ will always be satisfied, meaning that we can easily start the algorithm with the first party that does not satisfy equation (1.58). In addition,

$$\delta = \frac{1}{2\bar{s} + n} \tag{1.59}$$

is our estimate of the natural threshold. We can demonstrate that this in accord with earlier works on the subject by Palomares and Ramirez (2003), who estimated the threshold of exclusion to be $\frac{1}{s+1}$ and the threshold of inclusion to be $\frac{1}{s+n-1}$ . We can demonstrate that

$$\frac{1}{2s + n} < \frac{1}{s + 1} \tag{1.60}$$

and that, even though sub-threshold parties are not automatically excluded from our calculations, since

$$\frac{1}{2s + n} < \frac{1}{s + n - 1} \tag{1.61}$$

they never obtain more than ½ seat. Proof: let

$$\frac{1}{s + n - 1} > p > \frac{1}{n + 2s} \tag{1.62}$$

Since $s_i = p\left(s + \frac{n}{2}\right) - \frac{1}{2}$ and $p > \frac{1}{n+2s}$, we know that

$$s_i < \frac{s + \frac{n}{2}}{s + n - 1} - \frac{1}{2}. \tag{1.63}$$

For each $n \geq 2$

$$\frac{s + \frac{n}{2}}{s + n - 1} \leq 1. \tag{1.64}$$

(Proof: $n - \frac{n}{2} \geq 1; \quad n - 1 \geq \frac{n}{2}; \quad s + n - 1 \geq s + \frac{n}{2}$)

This means that

$$s_i < \frac{s + \frac{n}{2}}{s + n - 1} - \frac{1}{2} \leq \frac{1}{2}. \tag{1.65}$$

## II. Empirical test

In the foregoing part, we have made four assumptions which are necessary for formula (0.1) to work correctly:

| | |
|---|---|
| **A1** | for each party, expected value of the remaining fractional shares that are discarded when dividing the number of votes by Jefferson-d'Hondt divisor is ½ |
| **assumption A2** | the number of relevant parties (i.e. parties that reach both the statutory threshold, if any, and the natural threshold) is constant across districts |
| **A3** | party vote shares (party votes / total votes) are not correlated with district magnitude measured by the total number of (effective) votes[7] cast |
| **A4** | party vote shares (party votes / total votes) are not correlated with district magnitude measured by the total number of seats |

While most if not all of these assumptions are intuitive and we would expect them to be approximately satisfied in real-life elections, we still have to test them against empirical data. We would also like to test the formula itself, in order to determine whether it is robust against minor violations of the assumptions (which have to be expected).

As noted above, Jefferson-d'Hondt method is used in numerous electoral systems around the world. Due to data availability limitations, time and length constraints, and preliminary nature of this paper, we restrict ourselves herein to a limited subset of cases that meet the following criteria:

1) National lower house elections (being the most politically salient)…
2) … in EU member states…
3) … and national (rather than fragmented) party systems…
4) … that use Jefferson-d'Hondt method (without any adjustments) to allocate all seats…
5) … and continue to apply it as of June 2016.

There are six countries that satisfy those criteria exactly: Spain, Portugal, Finland, Netherlands, Poland, and the Czech Republic. We also include Croatia, although it does not fully satisfy criterion (5): it uses FPTP method to allocate seats in special districts set aside for ethnic minorities. However, the number of those minority seats is relatively small in comparison to the size of the Croatian parliament (6 out of ca. 150) and elections for those seats are held at different dates, so we just omit them from our calculations. We do not include Belgium, given its use of multitier elections in the Brussels region, but also given its absence of a nationwide party system (our formula would have to be applied separately for Flanders and Wallonia). We also do not include Bulgaria, since its seat allocation outcomes do not fully

[7] By effective votes we mean only those votes that have been cast for a relevant party.

match results obtained by using the Jefferson-d'Hondt algorithm (perhaps due to the fact that available data sets do not include information about the use of apparentments). For Finland, we omit elections prior to 2003 for the same reason – lack of full information about apparentment composition.

For elections before 2007, we have obtained our data from the Global Election Database. For subsequent elections, we have used the Constituency-Level Elections Archive and web sites of the respective national electoral authorities.

Table 1 sets forth the general electoral system parameters of the six countries discussed above:

| Country | Earliest included election | Number of elections | Number of seats | Number of districts | Avg. number of relevant parties |
|---|---|---|---|---|---|
| Croatia | 2000 | 5 | 143-146 | 11 | 6<br>+ 1 regional |
| Czech Republic | 2002 | 4 | 200 | 14 | 5 |
| Finland | 2003 | 4 | 200 | 15 | 7<br>+ 1 regional |
| Netherlands | 1948 | 20 | 100-150 | 1 | 10 |
| Poland | 2005 | 4 | 460 | 41 | 5 |
| Portugal | 1975 | 15 | 226-259 | 20-22 | 4.5 |
| Spain | 1977 | 12 | 350 | 52 | 4<br>+ 5 regional |

**Table 1. General parameters of electoral systems of the test country data set**

What appears troubling is that in three of those seven countries there are regional parties that manage to win parliamentary seats while running only in a limited subset of electoral districts (such as the Convergence and Union (CiU), the Republican Left of Catalonia (ERC), and the Basque Nationalist Party (EAJ / PNV) in Spain, the Swedish People's Party (SFP / RKP) in Finland, and the Croatian Democratic Alliance of Slavonia and Baranja (HDSSB) in Croatia)[8]. Such parties violate our assumption **A2**, since the number of relevant parties is no longer constant. This problem can be solved by introducing **a regional correction**.

We define a region to be a set of electoral district that is defined by a presence of one or more regional parties. To be recognized by us as regional, a party has to register candidate lists in all districts within some region and may not register candidate lists anywhere outside that region. No two regions are permitted to overlap. All districts which have no regional parties constitute the "national region" ($R_0$). In formal terms, a set of all regions $R = \{R_0,\ R_1,\ \dots\}$ is a partition on the set of all electoral districts ($D$).

All those restrictions on the definition on a region ensure:

- that within any region $r$, the number of parties ($n^r$) remains constant, —and
- that no regional party runs in more than one region.

[8] It is interesting to note that the three multidistrict countries that do not have any regional parties – Portugal, Poland, and the Czech Republic – are among the most homogenous countries in Europe in terms of ethnic and cultural composition of the population. Netherlands, of course, has only one electoral district.

Since region boundaries are known before the election, we can describe each region by such parameters as: the number of districts ($c^r$) (which can be one); the total number of seats ($s^r$); and the total number of votes ($v^r$)[9]. Also, since each regional party runs only in one region, its national vote share can be easily translated into regional vote share ($p_i^r = p_i \frac{v^r}{v}$). At this point, we have all information we need to estimate the number of seats allocated for each regional party, using a modified version of formula (0.1):

$$s_i := p_i^r \cdot s^r + p_i \cdot \frac{c^r n^r}{2} - \frac{c^r}{2}. \tag{2.1}$$

After seat allocations have been computed for all regional parties in all regions, we redefine *s* to be the total number of seats **excepting the seats already allocated to regional parties** and allocate them using the standard form of formula (0.1). At this stage, only national parties are taken into account. As we demonstrate below, the regional correction reduces the aggregate error of the formula to levels comparable to those of countries with no regional parties.

Let us now test the remaining three assumptions. For each assumption, we have computed a measure of deviation, defined as:

$$V_{A1,i} = \left|\bar{x}_i - \frac{1}{2}\right| \tag{2.2}$$

(where $\bar{\bar{x}}_i$ is the empirical mean remaining fractional part for the *i*-th party)

for assumption **A1**,

$$V_{A3,i} = \left|\mathrm{Corr}\left(p_i^k, v^k\right)\right| \tag{2.3}$$

(where $\mathrm{Corr}(x, y)$ is Pearson correlation coefficient for *x* and *y*)

for assumption **A3**, and

$$V_{A4,i} = \left|\mathrm{Corr}\left(p_i^k, s^k\right)\right| \tag{2.4}$$

for assumption **A4**.

If all three assumptions are satisfied for the *i*-th political parties, all three measures of deviation shall be equal to 0.

Since the assumptions are party-specific, for each election and assumption we provide two data points per assumption: maximum measure of deviation and median measure of deviation. In all cases, we only consider relevant parties. For regional parties, measures of deviation apply only to those districts in which they have fielded candidates.

| country | year | *n* [10] | deviation from A1 | | deviation from A2 | | deviation from A3 | |
|---|---|---|---|---|---|---|---|---|
| | | | max | median | max | median | max | median |
| Croatia | 2000 | 4 | 0.356 | 0.082 | 0.943 | 0.838 | 0.957 | 0.746 |
| Croatia | 2003 | 7 | 0.467 | 0.110 | 0.776 | 0.337 | 0.846 | 0.409 |

[9] Strictly speaking, $v^r$ is not known before the election, but can be estimated on the basis of the region's number of eligible voters and historical differences in voter turnout.

[10] Without regional parties.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Czech Republic | 2002 | 4 | 0.123 | 0.071 | 0.307 | 0.131 | 0.313 | 0.126 |
| Czech Republic | 2006 | 5 | 0.197 | 0.059 | 0.299 | 0.090 | 0.319 | 0.113 |
| Czech Republic | 2010 | 5 | 0.148 | 0.031 | 0.426 | 0.410 | 0.438 | 0.415 |
| Czech Republic | 2013 | 7 | 0.149 | 0.085 | 0.467 | 0.282 | 0.516 | 0.324 |
| Finland | 2003 | 7 | 0.153 | 0.027 | 0.672 | 0.404 | 0.634 | 0.350 |
| Finland | 2007 | 8 | 0.148 | 0.083 | 0.682 | 0.446 | 0.647 | 0.407 |
| Finland | 2011 | 8 | 0.326 | 0.074 | 0.787 | 0.517 | 0.746 | 0.446 |
| Finland | 2015 | 8 | 0.234 | 0.057 | 0.686 | 0.408 | 0.633 | 0.378 |
| Netherlands | 1948 | 9 | 0.500 | 0.308 | N/A | N/A | N/A | N/A |
| Netherlands | 1952 | 9 | 0.500 | 0.084 | N/A | N/A | N/A | N/A |
| Netherlands | 1956 | 10 | 0.500 | 0.197 | N/A | N/A | N/A | N/A |
| Netherlands | 1959 | 11 | 0.500 | 0.363 | N/A | N/A | N/A | N/A |
| Netherlands | 1963 | 13 | 0.500 | 0.197 | N/A | N/A | N/A | N/A |
| Netherlands | 1967 | 14 | 0.500 | 0.327 | N/A | N/A | N/A | N/A |
| Netherlands | 1971 | 19 | 0.500 | 0.168 | N/A | N/A | N/A | N/A |
| Netherlands | 1972 | 16 | 0.500 | 0.186 | N/A | N/A | N/A | N/A |
| Netherlands | 1977 | 14 | 0.500 | 0.174 | N/A | N/A | N/A | N/A |
| Netherlands | 1981 | 13 | 0.500 | 0.241 | N/A | N/A | N/A | N/A |
| Netherlands | 1982 | 14 | 0.500 | 0.201 | N/A | N/A | N/A | N/A |
| Netherlands | 1986 | 12 | 0.500 | 0.260 | N/A | N/A | N/A | N/A |
| Netherlands | 1989 | 11 | 0.500 | 0.089 | N/A | N/A | N/A | N/A |
| Netherlands | 1994 | 13 | 0.500 | 0.283 | N/A | N/A | N/A | N/A |
| Netherlands | 1998 | 13 | 0.500 | 0.331 | N/A | N/A | N/A | N/A |
| Netherlands | 2002 | 11 | 0.500 | 0.247 | N/A | N/A | N/A | N/A |
| Netherlands | 2003 | 11 | 0.500 | 0.275 | N/A | N/A | N/A | N/A |
| Netherlands | 2006 | 11 | 0.500 | 0.356 | N/A | N/A | N/A | N/A |
| Netherlands | 2010 | 10 | 0.500 | 0.317 | N/A | N/A | N/A | N/A |
| Netherlands | 2012 | 11 | 0.500 | 0.333 | N/A | N/A | N/A | N/A |
| Poland | 2005 | 6 | 0.093 | 0.028 | 0.447 | 0.217 | 0.231 | 0.082 |
| Poland | 2007 | 4 | 0.061 | 0.044 | 0.406 | 0.227 | 0.121 | 0.086 |
| Poland | 2011 | 5 | 0.090 | 0.037 | 0.281 | 0.215 | 0.189 | 0.108 |
| Poland | 2015 | 5 | 0.104 | 0.041 | 0.398 | 0.268 | 0.227 | 0.056 |
| Portugal | 1975 | 5 | 0.105 | 0.057 | 0.390 | 0.156 | 0.402 | 0.152 |
| Portugal | 1976 | 4 | 0.107 | 0.053 | 0.329 | 0.229 | 0.319 | 0.233 |
| Portugal | 1979 | 3 | 0.059 | 0.058 | 0.206 | 0.116 | 0.128 | 0.122 |
| Portugal | 1980 | 3 | 0.046 | 0.027 | 0.262 | 0.250 | 0.268 | 0.155 |
| Portugal | 1983 | 4 | 0.094 | 0.056 | 0.303 | 0.190 | 0.305 | 0.184 |
| Portugal | 1985 | 5 | 0.075 | 0.046 | 0.407 | 0.092 | 0.399 | 0.085 |
| Portugal | 1987 | 5 | 0.179 | 0.047 | 0.239 | 0.152 | 0.226 | 0.157 |
| Portugal | 1991 | 4 | 0.135 | 0.054 | 0.295 | 0.118 | 0.275 | 0.119 |
| Portugal | 1995 | 4 | 0.143 | 0.085 | 0.125 | 0.101 | 0.124 | 0.099 |
| Portugal | 1999 | 4 | 0.107 | 0.076 | 0.124 | 0.072 | 0.119 | 0.076 |
| Portugal | 2002 | 4 | 0.136 | 0.012 | 0.666 | 0.119 | 0.691 | 0.123 |
| Portugal | 2005 | 5 | 0.144 | 0.061 | 0.611 | 0.147 | 0.611 | 0.127 |
| Portugal | 2009 | 5 | 0.146 | 0.055 | 0.275 | 0.173 | 0.248 | 0.168 |
| Portugal | 2011 | 5 | 0.140 | 0.041 | 0.292 | 0.180 | 0.306 | 0.165 |
| Portugal | 2015 | 4 | 0.159 | 0.043 | 0.339 | 0.088 | 0.359 | 0.099 |
| Spain | 1977 | 4 | 0.141 | 0.050 | 0.607 | 0.251 | 0.589 | 0.257 |
| Spain | 1979 | 4 | 0.167 | 0.049 | 0.593 | 0.460 | 0.588 | 0.452 |
| Spain | 1982 | 3 | 0.066 | 0.043 | 0.454 | 0.246 | 0.416 | 0.247 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Spain | 1986 | 3 | 0.045 | 0.019 | 0.354 | 0.222 | 0.381 | 0.222 |
| Spain | 1989 | 4 | 0.078 | 0.020 | 0.531 | 0.179 | 0.582 | 0.192 |
| Spain | 1993 | 2 | 0.099 | 0.065 | 0.523 | 0.229 | 0.574 | 0.199 |
| Spain | 1996 | 2 | 0.088 | 0.063 | 0.932 | 0.371 | 0.943 | 0.383 |
| Spain | 2000 | 2 | 0.146 | 0.080 | 0.633 | 0.204 | 0.671 | 0.209 |
| Spain | 2004 | 2 | 0.005 | 0.005 | 0.247 | 0.127 | 0.249 | 0.152 |
| Spain | 2008 | 2 | 0.044 | 0.032 | 0.137 | 0.103 | 0.175 | 0.103 |
| Spain | 2011 | 3 | 0.127 | 0.052 | 0.640 | 0.243 | 0.626 | 0.233 |
| Spain | 2015 | 4 | 0.056 | 0.044 | 0.385 | 0.241 | 0.403 | 0.256 |

**Table 2. Tests of assumptions A1, A3, and A4**

As Table 2 indicates, assumption **A1** is quite well satisfied for most elections. Assumptions **A3** and **A4** look much worse, though it is important to note that significant violations occur only for individual parties. It is of course possible to devise a correction for the model that would account for correlations between $p_i^k$ and $v^k$ / $s^k$, but they would compromise the formula's predictive function, since application of such corrections would require information that cannot be inferred from typical poll reports or preliminary aggregate results (such as the $V_{A3,i}$ and $V_{A4,i}$ correlation coefficients).

Is a violation of assumptions **A3** and **A4** fatal for the accuracy of the approximation formula itself? This question can be answered by empirically testing the correctness of formula (0.1) itself. For each country, we have compared empirical seat allocation with the results yielded by the approximation formula (with regional correction if necessary). For further comparison, we have also included the results obtained by "naïve proportional" seat allocation, i.e., by allocating to each party exactly $p_i \cdot s$ seats, without accounting for rounding effects, district size variations, and bias in favor of larger parties. Table 3 sets forth the results for the most recent elections in the six test countries.

| election | *s* | *c* | *n*[11] | party[12] | $p_i$ | $s_i$ | formula allocation | | naïve proportionality | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | seats | error | seats | error |
| **Croatia 2015** | 143 | 11 | 5 | DK | **38.7%** | **59** | **58.69** | -0.31 | **53.39** | -5.61 |
| | | | | HR | **37.6%** | **56** | **56.81** | 0.81 | **51.83** | -4.17 |
| | | | | MOST | **15.3%** | **19** | **19.84** | 0.84 | **21.08** | 2.08 |
| | | | | KRS | **3.7%** | **2** | **0.69** | -1.31 | **5.15** | 3.15 |
| | | | | HB | **4.8%** | **1** | **2.38** | 1.38 | **6.56** | 5.56 |
| | | | | IDS | **2.0%** | **3** | **2.87** | -0.13 | **2.81** | -0.19 |
| | | | | HDSSB | **1.6%** | **2** | **1.72** | -0.28 | **2.19** | 0.19 |
| **Czech Republic 2013** | 200 | 14 | 7 | CSSD | **23.4%** | **50** | **51.27** | 1.27 | **46.81** | -3.19 |
| | | | | ANO 2011 | **21.3%** | **47** | **46.14** | -0.86 | **42.68** | -4.32 |
| | | | | KSCM | **17.1%** | **33** | **35.47** | 2.47 | **34.11** | 1.11 |
| | | | | TOP 09 | **13.7%** | **26** | **27.18** | 1.18 | **27.45** | 1.45 |
| | | | | ODS | **8.8%** | **16** | **15.02** | -0.98 | **17.68** | 1.68 |
| | | | | UPD | **7.9%** | **14** | **12.62** | -1.38 | **15.76** | 1.76 |
| | | | | KDU-CSL | **7.8%** | **14** | **12.31** | -1.69 | **15.51** | 1.51 |

[11] Without regional parties.
[12] Regional parties are in shaded rows.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Finland 2015** | 200 | 13 | 7 | Center | **21.6%** | **49** | **47.28** | -1.72 | **43.26** | -5.74 |
| | | | | Nat'l Coal. | **18.7%** | **37** | **39.89** | 2.89 | **37.32** | 0.32 |
| | | | | True Finns | **18.1%** | **38** | **38.48** | 0.48 | **36.18** | -1.82 |
| | | | | SDP | **16.9%** | **34** | **35.59** | 1.59 | **33.86** | -0.14 |
| | | | | Green League | **8.7%** | **15** | **15.23** | 0.23 | **17.48** | 2.48 |
| | | | | Left Alliance | **7.3%** | **12** | **11.70** | -0.30 | **14.64** | 2.64 |
| | | | | SPP | **5.0%** | **9** | **9.31** | 0.31 | **10.00** | 1.00 |
| | | | | Chr. Dems | **3.6%** | **5** | **2.53** | -2.47 | **7.26** | 2.26 |
| **Nether-lands 2012** | 150 | 1 | 11 | VVD | **26.8%** | **41** | **41.22** | 0.22 | **40.25** | -0.75 |
| | | | | PvdA | **25.1%** | **38** | **38.49** | 0.49 | **37.61** | -0.39 |
| | | | | PVV | **10.2%** | **15** | **15.33** | 0.33 | **15.27** | 0.27 |
| | | | | SP | **9.7%** | **15** | **14.66** | -0.34 | **14.62** | -0.38 |
| | | | | CDA | **8.6%** | **13** | **12.85** | -0.15 | **12.88** | -0.12 |
| | | | | D66 | **8.1%** | **12** | **12.11** | 0.11 | **12.16** | 0.16 |
| | | | | CU | **3.2%** | **5** | **4.41** | -0.59 | **4.73** | -0.27 |
| | | | | GL | **2.4%** | **4** | **3.16** | -0.84 | **3.53** | -0.47 |
| | | | | SGP | **2.1%** | **3** | **2.78** | -0.22 | **3.16** | 0.16 |
| | | | | PvdD | **2.0%** | **2** | **2.53** | 0.53 | **2.93** | 0.93 |
| | | | | 50 PLUS | **1.9%** | **2** | **2.46** | 0.46 | **2.85** | 0.85 |
| **Poland 2015** | 460 | 41 | 5 | PiS | **45.2%** | **235** | **233.54** | -1.46 | **207.75** | -27.25 |
| | | | | PO | **29.0%** | **138** | **142.35** | 4.35 | **133.18** | -4.82 |
| | | | | Kukiz 15 | **10.6%** | **42** | **39.04** | -2.96 | **48.69** | 6.69 |
| | | | | Nowoczesna | **9.1%** | **28** | **30.89** | 2.89 | **42.02** | 14.02 |
| | | | | PSL | **6.2%** | **16** | **14.19** | -1.81 | **28.37** | 12.37 |
| **Portugal 2015** | 226 | 20 | 4 | PSD/CDS | **43.0%** | **104** | **104.31** | 0.31 | **97.12** | -6.88 |
| | | | | PSoc | **36.3%** | **85** | **86.56** | 1.56 | **82.04** | -2.96 |
| | | | | BlocEsq | **11.5%** | **19** | **20.47** | 1.47 | **25.89** | 6.89 |
| | | | | PCP-PEV | **9.3%** | **17** | **14.65** | -2.35 | **20.95** | 3.95 |
| **Spain 2016** | 350 | 52 | 4 | PP | **36.7%** | **137** | **132.61** | -4.39 | **119.32** | -17.68 |
| | | | | PSOE | **25.3%** | **85** | **83.18** | -1.82 | **82.14** | -2.86 |
| | | | | Podemos | **23.5%** | **71** | **75.49** | 4.49 | **76.35** | 5.35 |
| | | | | C's | **14.5%** | **32** | **36.55** | 4.55 | **47.05** | 15.05 |
| | | | | ERC | **2.9%** | **9** | **8.42** | -0.58 | **9.51** | 0.51 |
| | | | | CDC (DL) | **2.2%** | **8** | **5.96** | -2.04 | **7.27** | -0.73 |
| | | | | PNV | **1.4%** | **5** | **4.84** | -0.16 | **4.41** | -0.59 |
| | | | | EH Bildu | **0.9%** | **2** | **2.43** | 0.43 | **2.86** | 0.86 |
| | | | | CC | **0.3%** | **1** | **0.52** | -0.48 | **1.09** | 0.09 |

**Table 3. Comparison of predicted and empirical seat allocations in most recent parliamentary elections in six European countries (with regional corrections when necessary)**

It is immediately apparent that in all countries except Netherlands the proposed formula produces a much better approximation of the final result than the "naïve proportionality" approach (in Netherlands both formulas produce very small errors). Indeed, only in 7 cases out of 51 our margin of error exceeds 1% of the national seat total. This is despite quite significant deviations from assumptions **A3** and **A4**.

Six elections and forty parties is still a rather small sample to establish a claim to empirical validity. For this reason, we have repeated the test for all available elections from our six test countries. Due to paper length limitations, we do not present full results for each party. Instead, for each country we have computed a sum of the absolute values of errors:

$$err = \sum_i \left| s_{emp_i} - s_i \right| \tag{2.5}$$

where $s_{emp_i}$ is the number of seats awarded to the $i$-th party under the empirical allocation[13]. Table 4 sets for the values of the error measure for successive elections.

| country | year | $s$ | $c$ | $n$[14] | aggregate error | |
|---|---|---|---|---|---|---|
| | | | | | absolute | % of $s$ |
| Croatia | 2000 | 146 | 11 | 4 | 3.10 | 2.1% |
| Croatia | 2003 | 144 | 11 | 7 | 8.20 | 5.7% |
| Croatia | 2007 | 145 | 11 | 6 | 5.03 | 3.5% |
| Croatia | 2011 | 143 | 11 | 6 | 9.37 | 6.6% |
| Croatia | 2015 | 143 | 11 | 5 | 5.06 | 3.5% |
| Czech Republic | 2002 | 200 | 14 | 4 | 3.98 | 2.0% |
| Czech Republic | 2006 | 200 | 14 | 5 | 6.28 | 3.1% |
| Czech Republic | 2010 | 200 | 14 | 5 | 3.62 | 1.8% |
| Czech Republic | 2013 | 200 | 14 | 7 | 9.83 | 4.9% |
| Finland | 2003 | 200 | 15 | 7 | 10.37 | 5.2% |
| Finland | 2007 | 200 | 15 | 8 | 13.46 | 6.7% |
| Finland | 2011 | 200 | 15 | 8 | 14.28 | 7.1% |
| Finland | 2015 | 200 | 13 | 8 | 9.99 | 5.0% |
| Netherlands | 1948 | 100 | 1 | 9 | 2.01 | 2.0% |
| Netherlands | 1952 | 100 | 1 | 9 | 1.46 | 1.5% |
| Netherlands | 1956 | 150 | 1 | 10 | 2.04 | 1.4% |
| Netherlands | 1959 | 150 | 1 | 11 | 2.22 | 1.5% |
| Netherlands | 1963 | 150 | 1 | 13 | 2.44 | 1.6% |
| Netherlands | 1967 | 150 | 1 | 14 | 3.57 | 2.4% |
| Netherlands | 1971 | 150 | 1 | 19 | 3.50 | 2.3% |
| Netherlands | 1972 | 150 | 1 | 16 | 3.20 | 2.1% |
| Netherlands | 1977 | 150 | 1 | 14 | 2.45 | 1.6% |
| Netherlands | 1981 | 150 | 1 | 13 | 2.67 | 1.8% |
| Netherlands | 1982 | 150 | 1 | 14 | 2.83 | 1.9% |
| Netherlands | 1986 | 150 | 1 | 12 | 2.46 | 1.6% |
| Netherlands | 1989 | 150 | 1 | 11 | 2.31 | 1.5% |
| Netherlands | 1994 | 150 | 1 | 13 | 3.08 | 2.1% |
| Netherlands | 1998 | 150 | 1 | 13 | 2.56 | 1.7% |
| Netherlands | 2002 | 150 | 1 | 11 | 2.46 | 1.6% |
| Netherlands | 2003 | 150 | 1 | 11 | 2.58 | 1.7% |
| Netherlands | 2006 | 150 | 1 | 11 | 3.02 | 2.0% |
| Netherlands | 2010 | 150 | 1 | 10 | 2.99 | 2.0% |
| Netherlands | 2012 | 150 | 1 | 11 | 4.29 | 2.9% |
| Poland | 2005 | 460 | 41 | 6 | 11.87 | 2.6% |
| Poland | 2007 | 460 | 41 | 4 | 12.30 | 2.7% |
| Poland | 2011 | 460 | 41 | 5 | 7.85 | 1.7% |
| Poland | 2015 | 460 | 41 | 5 | 13.47 | 2.9% |

[13] The aggregate error measure $err$ should not be mistaken with **the number of misallocated seats**, which is always equal to $\frac{err}{2}$, since each misallocated seat generates a double error. As seats sum up to $s$, each seat gain to the $i$-th party under the formula (as compared to the empirical allocation) must be paired with a corresponding seat loss to the other parties.
[14] Without regional parties.

| | | | | | | |
|---|---|---|---|---|---|---|
| Portugal | 1975 | 247 | 22 | 5 | 8.95 | 3.6% |
| Portugal | 1976 | 259 | 22 | 4 | 5.76 | 2.2% |
| Portugal | 1979 | 246 | 20 | 3 | 14.99 | 6.1% |
| Portugal | 1980 | 246 | 20 | 3 | 19.00 | 7.7% |
| Portugal | 1983 | 246 | 20 | 4 | 4.17 | 1.7% |
| Portugal | 1985 | 246 | 20 | 5 | 5.54 | 2.3% |
| Portugal | 1987 | 246 | 20 | 5 | 8.79 | 3.6% |
| Portugal | 1991 | 226 | 20 | 4 | 9.16 | 4.1% |
| Portugal | 1995 | 226 | 20 | 4 | 2.90 | 1.3% |
| Portugal | 1999 | 226 | 20 | 4 | 7.15 | 3.2% |
| Portugal | 2002 | 226 | 20 | 4 | 10.43 | 4.6% |
| Portugal | 2005 | 226 | 20 | 5 | 5.76 | 2.5% |
| Portugal | 2009 | 226 | 20 | 5 | 8.96 | 4.0% |
| Portugal | 2011 | 226 | 20 | 5 | 8.86 | 3.9% |
| Portugal | 2015 | 226 | 20 | 4 | 5.69 | 2.5% |
| Spain | 1977 | 350 | 52 | 4 | 37.80 | 10.8% |
| Spain | 1979 | 350 | 52 | 4 | 36.34 | 10.4% |
| Spain | 1982 | 350 | 52 | 3 | 22.11 | 6.3% |
| Spain | 1986 | 350 | 52 | 3 | 30.58 | 8.7% |
| Spain | 1989 | 350 | 52 | 4 | 18.25 | 5.2% |
| Spain | 1993 | 350 | 52 | 2 | 28.11 | 8.0% |
| Spain | 1996 | 350 | 52 | 2 | 30.97 | 8.8% |
| Spain | 2000 | 350 | 52 | 2 | 20.93 | 6.0% |
| Spain | 2004 | 350 | 52 | 2 | 10.92 | 3.1% |
| Spain | 2008 | 350 | 52 | 2 | 10.87 | 3.1% |
| Spain | 2011 | 350 | 52 | 3 | 22.64 | 6.5% |
| Spain | 2015 | 350 | 52 | 4 | 25.87 | 7.4% |
| Spain | 2016 | 350 | 52 | 4 | 18.94 | 5.4% |

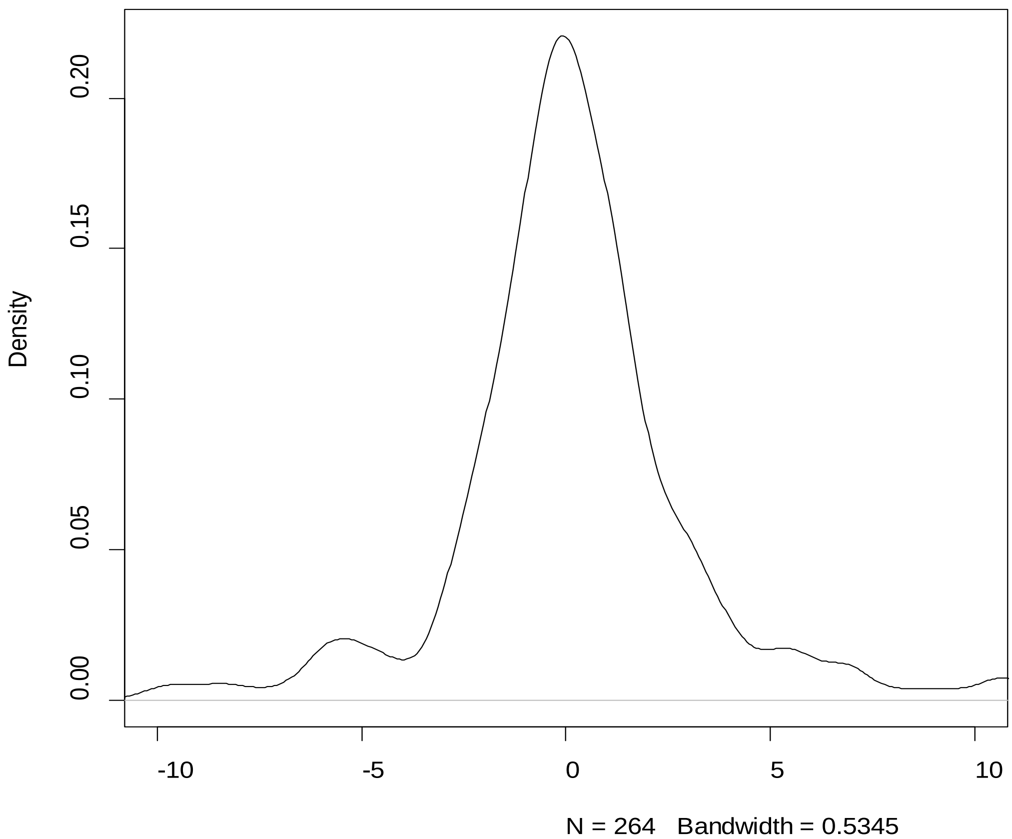
**Table 4. Aggregate errors for available elections**

This much larger sample of elections demonstrates quite well that formula (0.1) indeed works as expected and is robust against violations of its assumptions. Only in 2 out of 65 elections more than 5% seats have been misallocated. It is also interesting to note that errors for Netherlands, where there is only one district, and the errors generated by the transition from the single-district to the multi-district formula are therefore non-existent, are not much smaller than those for other countries. This suggests that the small number of districts, while eliminating some sources of error, introduces other (perhaps related to the fact that average remaining fractional parts of party vote shares are more likely to diverge from the theoretical expected value of ½, something that is highly unlikely in countries like Poland and Spain due to the law of large numbers).

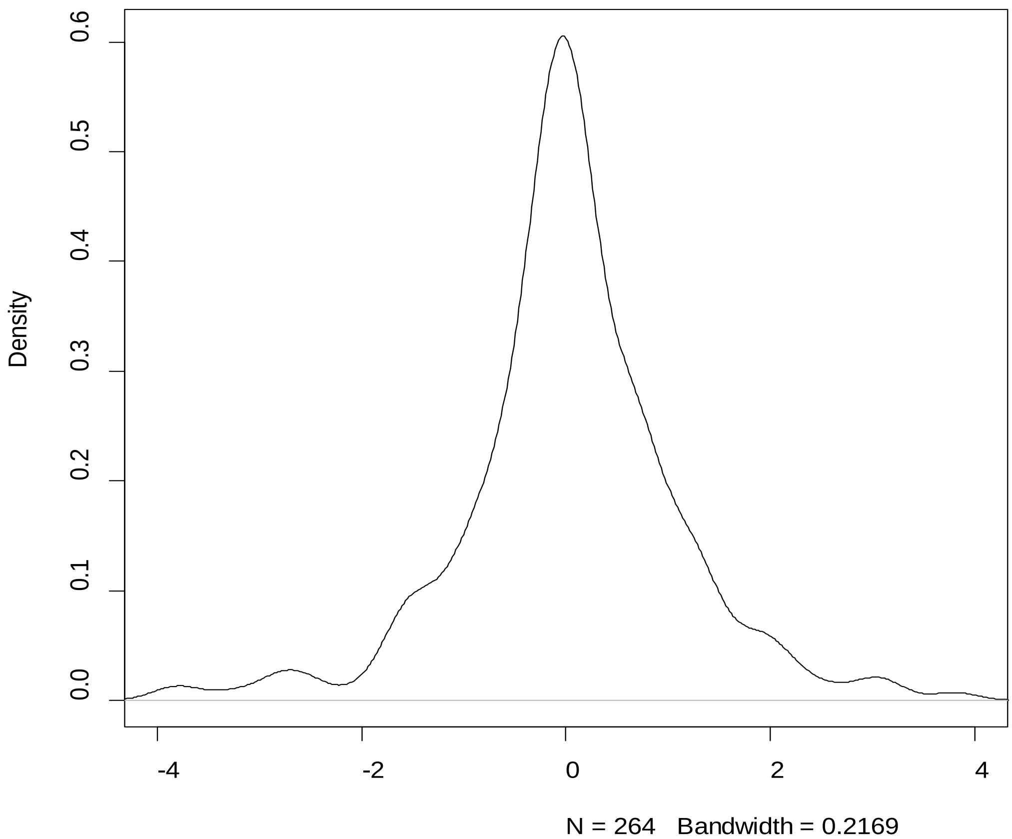
This robustness of the formula undoubtedly requires further investigation, which is beyond the scope of this paper. Nevertheless, our initial study of the subject suggests that under typical conditions encountered in real-life elections (unless the system is gerrymandered or otherwise deliberately skewed in favor of some type of parties), the errors introduced in different stages of approximation tend to largely cancel each other out, thereby making the overall error much smaller than initially expected.

Finally, we have created two density plots, illustrating distribution of absolute party errors (absolute differences between empirical and predicted seat allocations) and relative party errors (absolute party errors divided by national seat totals).

**Distribution of absolute party error**




**Distribution of relative party errors**



## III. Political consequences

Predicting nationwide seat allocation on the basis of party vote shares is practically useful, but from the development of the discipline point of view the utility of our formula lies primarily in its capability to explain the size of the Jefferson-d'Hondt bonus to the larger parties (and the corresponding loss of sets by the smaller parties). We will call it "integration bonus", since it incentivizes integration of the party system by rewarding mergers of smaller parties. The bonus – deviation from proportionality – is nothing else than the number of seats received by the party in excess of pure proportionality, i.e.

$$\Delta_i = p_i \cdot \frac{cn}{2} - \frac{c}{2} \tag{3.1}$$

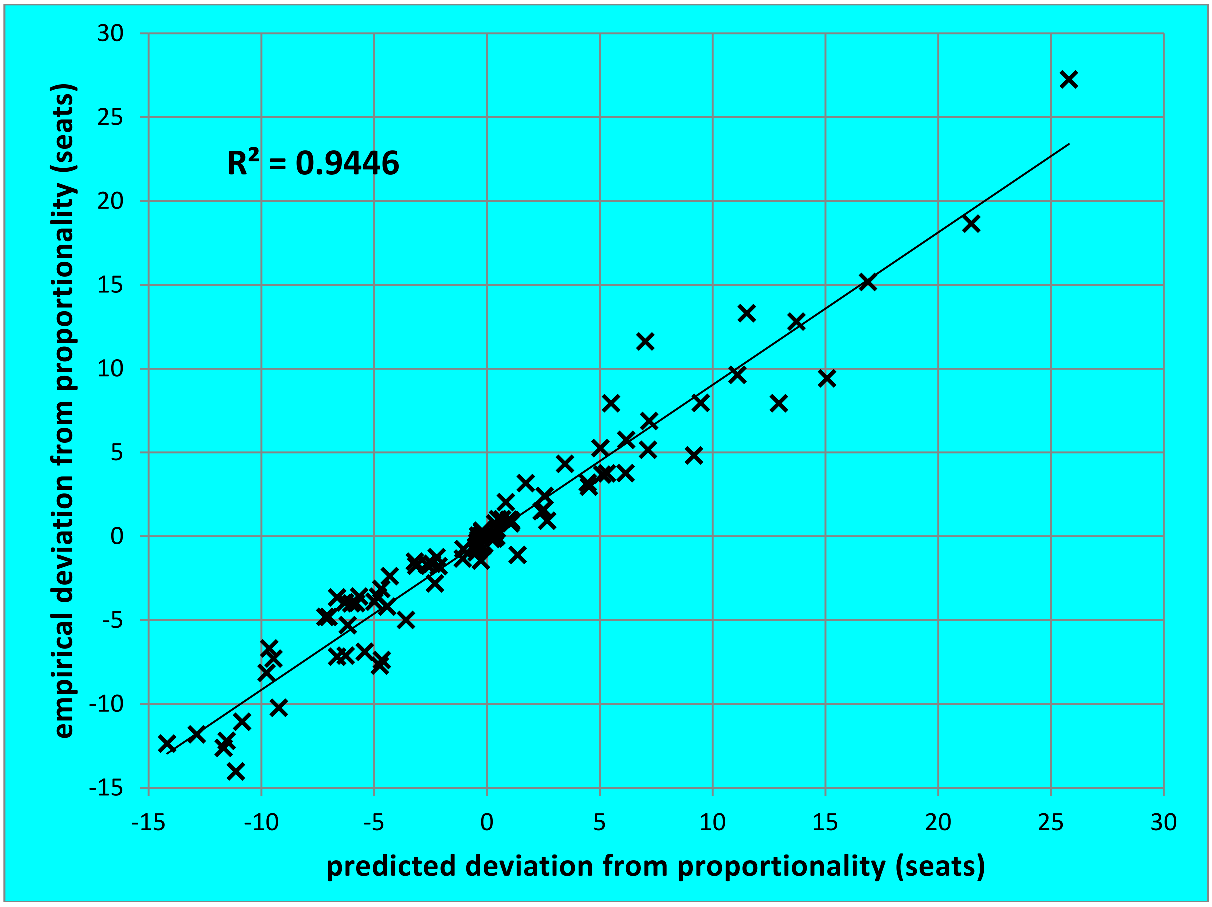
where $p_i$ is the effective vote share of the $i$-th party, $c$ is the number of districts, and $n$ is the number of parties (from formula (0.1)). We will note that the system is neutral (i.e. $\Delta_i = 0$) only towards those parties for which

$$p_i = \frac{1}{n}, \tag{3.2}$$

that is, the mean party vote share. Such parties receive exactly the number of seats corresponding to their vote shares.

To test whether our formula for expected bonus succeeds in predicting the empirical bonus we have tested the two magnitudes for correlation. We have included data from four of our test countries from part II (eliminating those with regional parties, since their existence complicates the bonus mechanism – apart from the national bonus, in regionalized electoral systems there are also bonuses for regional winners). In addition, we have included data from Albanian elections of 2009 and 2013 to illustrate how the bonuses behave when the number of parties in the system is very small (most Albanian parties tend to aggregate themselves into two large party blocks, making the system behave – from the seat allocation point of view – just like a classic two party system, optionally with small third parties).

Figure 1 illustrates how the empirical deviation of proportionality can be explained by the predicted deviation (our $\Delta_i$). We have limited ourselves to the twenty-first century elections in an attempt to equalize the number of data points per country (otherwise, large number of data points for Portugal and Netherlands – on the order of 15-20 – would crowd out 4-5 elections per country from the Central and Eastern European states).



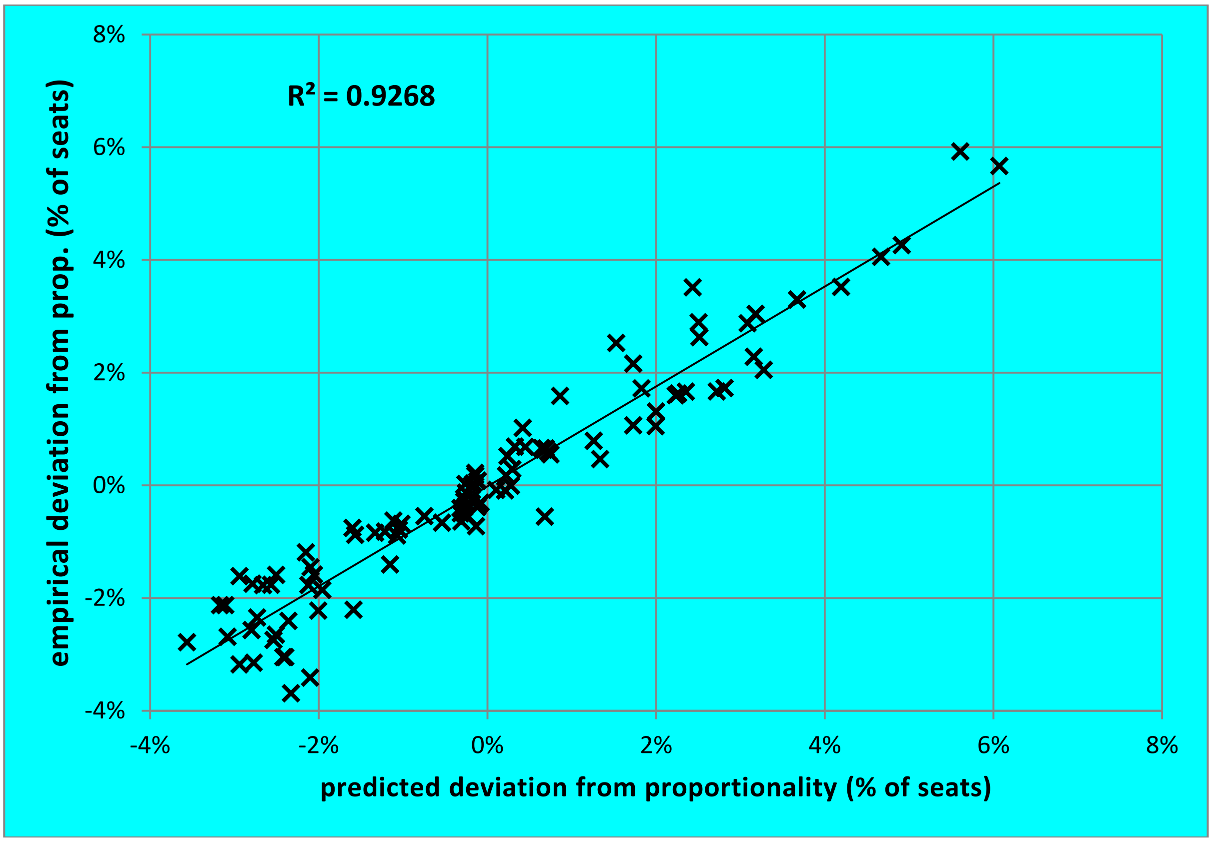




**Figure 1. Relationship between empirical and predicted deviations from proportionality**

The correlation between the two variables is quite high (at least for social sciences), especially when compared with a model that explains the bonus solely in terms of effective vote shares:

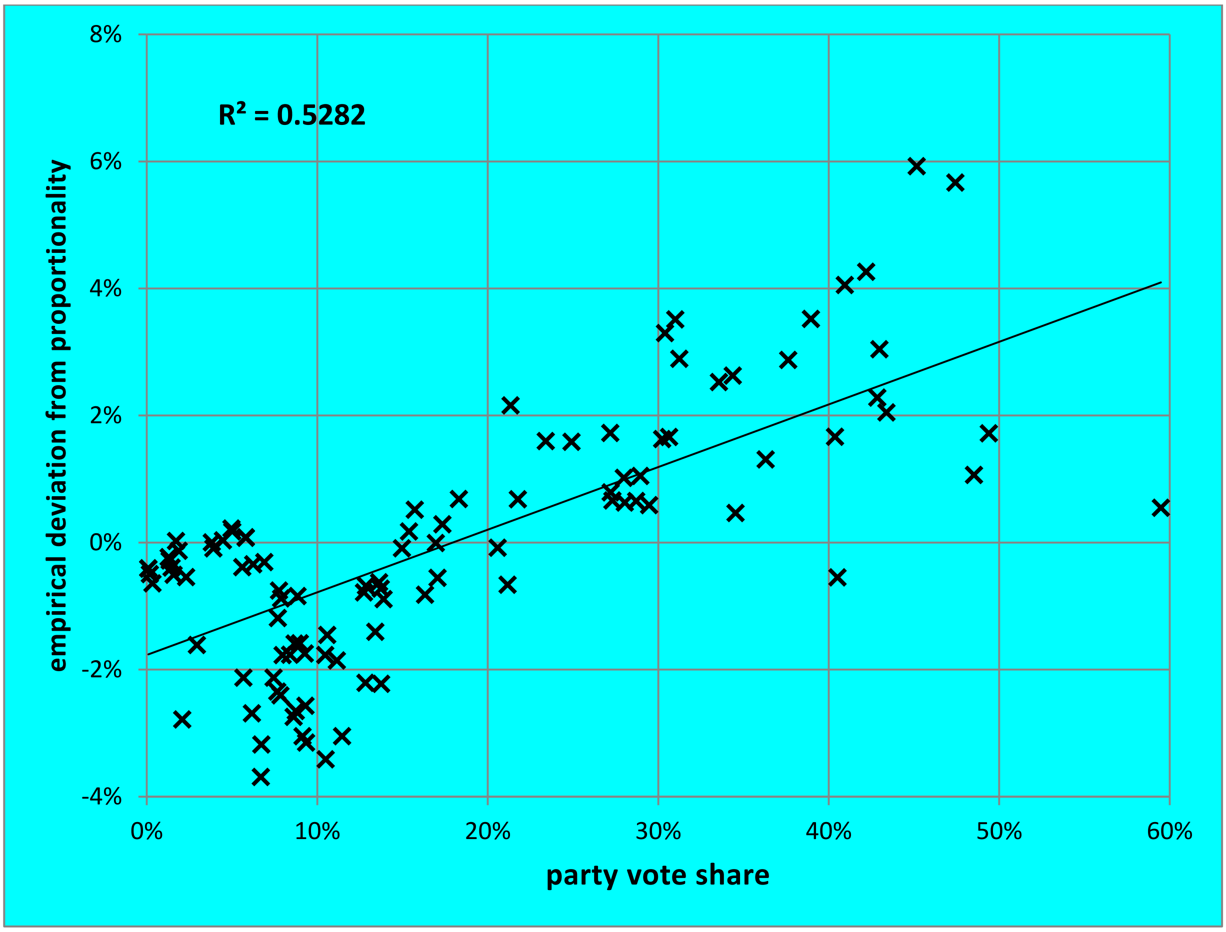



**Figure 2. Relationship between empirical deviations from proportionality and party vote shares**

Although party vote shares are positively correlated with empirical deviations of proportionality and explain some part thereof, only after the remaining factors – district and party counts – are included, the integration bonus generated by the Jefferson-d'Hondt method is explained in full.

Thanks to formula (3.1) providing a good approximation of the integration bonus, we can analyze how does it depend on other parameters of the electoral system. There are four of them: the number of seats, the number of districts, the number of relevant parties and each party's (effective) vote share. The most stable ones are the number of seats (which is often constitutionally fixed) and the number of districts. A change in either is commonly considered to be a major change in the electoral rules (Pilet et al. 2016). For assessing the magnitude of the integration bonus, however, it is only necessary to know the average number of seats per district. For the five countries considered in our model, it varies from 10.5 in Portugal, through 11.2 in Poland, 11.7 in Albania, and 14.3 in the Czech Republic, to 150 in Netherlands (where there is only one district). The relationship thereof with the magnitude of the integration bonus is one of inverse proportionality – the larger the average district, the smaller the potential bonus.

The number of relevant parties is a more variable parameter, though, at least in stable democracies with institutionalized party systems, it rarely varies drastically. Still, not one of our sample of five countries maintained a constant number relevant parties throughout all elections since 2000. For our three model countries – Czech Republic, Poland, and Portugal – it generally varied between 4 and 7. Albania was an extreme case, with only 2 parties in 2009 and 2 in 2013. Netherlands was the other extreme, with 10 to 11 parties running in each post-2000 election.

Finally, the last parameter influencing the magnitude of the integration bonus is party size (its effective share of the total number of votes). While variability of this parameter is related to the number of parties (Taagepera 2007), the relationship is far from trivial (as it is in the case of the expected party size). Similarly sized parties occur in party systems with very different numbers of relevant parties. Yet the party size is ultimately the key determinant of the size of the integration bonus, since it distinguishes the winners and losers of the Jefferson-d'Hondt method, and thereby affects the strength of the incentive to integrate or disintegrate according to the calculation of potential electoral gains.

Importance of the first two parameters – average district size and the number of parties – can be best demonstrated by running a simulation with a fixed-size party. Figures 3 and 4 illustrate the size of the bonus for two parties – one with 40% share in the total number of votes and one with 10% share – as it varies depending on the number of parties (from 3 to 8). Distinct data series are drawn for each country (representing the effects of different average district sizes). Model party sizes (10% and 40%) have not been chosen randomly – for the number of parties in the $[3, 8]$ range a party with 40% of all votes is always guaranteed to have above-average vote share, while a party with 10% of all votes is always guaranteed to have less than average number.

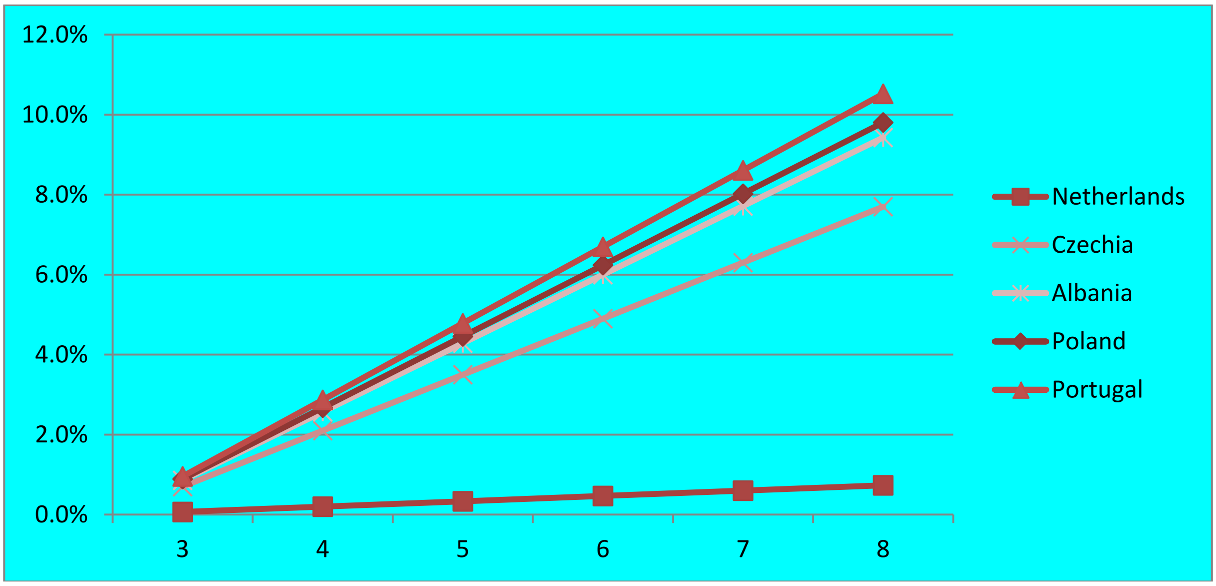


**Figure 3. Integration bonus for a party with effective vote share 40%**

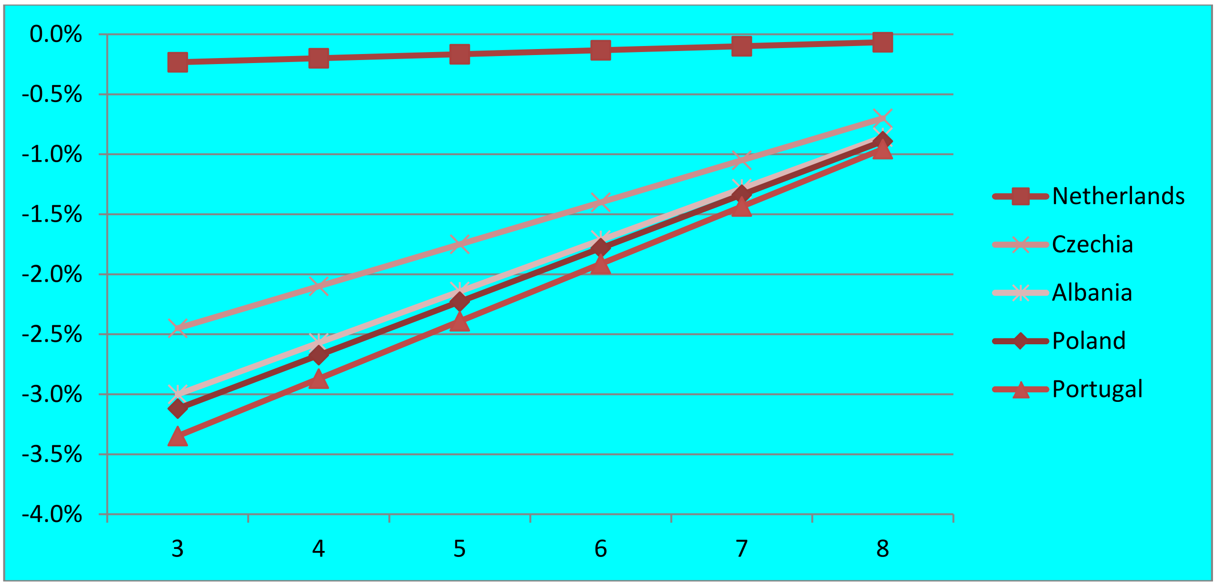



**Figure 4. Integration bonus for a party with effective vote share 10%**

For the party with effective vote share 40% the integration bonus depends primarily on the product of the number of parties and the average district size. If the average district is very large (as in Netherlands) or the number of parties is very small (as in Albania), the integration bonus will be negligible. But in systems with 10- to 15-seat districts and with 4 or more parties, a party with 40% of all votes can expect a gain of at least several percent. It is particularly interesting to note the monotonically increasing relationship between the number of parties and the integration bonus. It creates an interesting feedback mechanism: the incentive for integration is greater where it is most needed.

For small parties (vote shares on the order of 10%) the mechanism is less intuitive. As the integration bonus for large parties increases, the losses are distributed across a larger number of small parties, thereby making each small party lose less seats than in the case of a system with only few parties. From a small party's point of view, a disintegrated party system is therefore better than an integrated one.

## Conclusions

The proposed model has three primary applications. First, as noted above, it provides a reasonably accurate method for estimating nationwide seat allocation on the basis of aggregate data, such as opinion polls or exit polls. Second, it is useful for calculating political strategies (e.g., for estimating consolidation benefits or secession losses). Third, it provides a simple method for estimating effects of changes in electoral system parameters (particularly the number of seats and the number of districts) on particular parties, as well as on general systemic incentives.

## References

Balinski, Michel L. and Young, H. Peyton (1978) 'The Jefferson Method of Apportionment', *SIAM Review*, 20 (3), pp. 278–284.

Balinski, Michel L. and Young, H. Peyton (2001) *Fair Representation: Meeting the Ideal of One Man, One Vote*. Washington, DC: Brookings Institution Press.

Benoit, Kenneth (2000) 'Which Electoral Formula Is the Most Proportional? A New Look with New Evidence', *Political Analysis*, 8, pp. 381–388, doi: 10.1093/oxfordjournals.pan.a029822.

Bochsler, Daniel (2008) 'Who gains from apparentments under D'Hondt?', *Electoral Studies*, 29 (4), pp. 617–627, doi: 10.1016/j.electstud.2010.06.001.

Colomer, Josep M. (2004) *The Handbook of Electoral System Choice*. London: Palgrave Macmillan.

d'Hondt, Victor (1878) *La représentation proportionnelle des parties*. Ghent.

d'Hondt, Victor (1882) *Système pratique et raisonné de représentation proportionnelle*. Bruxelles: Libraire C. Muquardt.

Drton, Mathias and Schwingenschlögl, Udo (2004) 'Asymptotic seat bias formulas', *Metrika*, 62 (1), pp. 23–31, doi: 10.1007/s001840400352.

Gallagher, Michael (1991) 'Proportionality, Disproportionality and Electoral Systems', *Electoral Studies*, 10, pp. 33–51, doi: 10.1016/0261-3794(91)90004-c.

Gfeller, Jules (1890) 'Du transfert des suffrages et de la répartition des sièges complémentaires', *Représentation proportionnelle – Revue mensuelle*, 9, pp. 120–131.

Happacher, Max and Pukelsheim, Friedrich (1996), 'Rounding Probabilities: Unbiased Multipliers', *Statistics & Decisions*, 14, pp. 373–382.

Janson, Svante (2014) 'Asymptotic bias of some election methods', *Annals of Operations Research*, 215 (1), pp. 89–136, arXiv:1110.6369v1.

Jefferson, Thomas (1792) 'Opinion on Apportionment Bill', *The Papers of Thomas Jefferson, Digital Edition*, ed. B. Oberg and J. Looney, Charlottesville: University of Virginia Press, Rotunda, 2008.

Marshall, Albert W., Olkin, Ingram and Pukelsheim, Friedrich (2002) 'A majorization comparison of apportionment methods in proportional representation', *Social Choice and Welfare*, 19 (4), pp. 885–900, doi: 10.1007/s003550200164.

Owens, Lukas (2014) 'Exploring the Rate of Convergence of Approximations to the Riemann Integral', https://www.whitman.edu/Documents/Academics/Mathematics/2014/owensla.pdf.

Palomares, Anotnio and Ramírez, Victoriano (2003), 'Thresholds of the Divisor Methods', *Numerical Algorithms*, 34 (2), pp. 405–415, doi: 10.1023/B:NUMA.0000005353.82970.ce.

Pilet, Jean-Benoit, Renwick, Alan, Núñez, Lidia, Reimink, Elwin and Simón, Pablo (2016) *Electoral System Change in Europe (ESCE): Database of Electoral Systems Codebook*, http://www.electoralsystemchanges.eu/Files/media/MEDIA_767/FILE/Codebook_Database_of_Electoral_Systems_15-02-2016.pdf.

Pólya, György (1918) ‘Über die Verteilungssysteme der Proportionalwahl’, *Zeitschrift für schweizerische Statistik und Volkswirtschaft*, 54, pp. 363–387.

Pólya, György (1918) ‘Sur la représentation proportionnelle en matière électorale’, *L’Enseignement Mathématique*, 20, pp. 355–379.

Pólya, György (1919) ‘Proportionalwahl und Wahrscheinlichkeitsrechnung’, *Zeitschrift für die ges. Staatswissenschaft*, 74, pp. 297–322.

Pukelsheim, Friedrich (2013) *Proportional Representation: Apportionment Methods and Their Applications*. Heidelberg: Springer.

Sainte-Laguë, André (1910) ‘La représentation proportionnelle et la méthode des moindres carrés’, *Annales scientifiques de l'École Normale Supérieure, Sér. 3*, 27, pp. 529–542.

Schwingenschlögl, Udo and Drton, Mathias (2004) ‘Seat allocation distributions and seat biases of stationary apportionment methods for proportional representation’, *Metrika*, 60 (2), pp. 191–202, doi: 10.1007/s001840400347.

Schwingenschlögl, Udo (2008) ‘Seat biases of apportionment methods under general distributional assumptions’, *Applied Mathematics Letters*, 21 (1), pp. 1–3, doi: 10.1016/j.aml.2007.02.005.

Schuster, Karsten, Pukelsheim, Friedrich, Drton, Mathias and Draper, Norman R. (2003) ‘Seat biases of apportionment methods for proportional representation’, *Electoral Studies*, 22, pp. 651–676, doi: 10.1016/S0261-3794(02)00027-6.

Taagepera, Rein (2007) *Predicting party sizes: the logic of simple electoral systems*. Oxford: Oxford University Press.

van Eck, L., Visagie, S.E. and de Kock, H.C. (2005) ‘Fairness of seat allocation methods in proportional representation’, *ORiON*, 21 (2), pp. 93–110, doi: 10.5784/21-2-22.