HAL open science

On party-proportional representation under district distortions

Gabrielle Demange

To cite this version:

Gabrielle Demange. On party-proportional representation under district distortions. 2011. halshs-00623031v2

HAL Id: halshs-00623031
https://shs.hal.science/halshs-00623031v2

Preprint submitted on 28 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

WORKING PAPER N° 2011 – 32

On party-proportional representation under district distortions

Gabrielle Demange

JEL Codes: D70, D71

Keywords: Party-proportional representation, (bi-)apportionment, (bi-)divisor methods, fair shares

PARIS-JOURDAN SCIENCES ECONOMIQUES
48, BD JOURDAN – E.N.S. – 75014 PARIS
TÉL. : 33(0) 1 43 13 63 00 – FAX : 33 (0) 1 43 13 63 10
www.pse.ens.fr

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE – ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES
ÉCOLE DES PONTS PARISTECH – ECOLE NORMALE SUPÉRIEURE – INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE

On party-proportional representation under district distortions

Gabrielle DEMANGE[1]

October 17, 2011

Abstract

The paper presents the problem of choosing the representatives in an assembly when the whole electoral region is subdivided into electoral districts. Because of the two dimensions, geographical (districts) and political (parties), the problem is called bi-apportionment. Often the allocation of seats to districts is pre-determined and furthermore distorted —meaning that the ratios of the number of assigned seats to population size vary significantly across districts. The paper surveys proposed bi-apportionment methods with a focus on the conflict that may arise between party-proportional representation and district distortions.

Keywords Party-proportional representation, (bi-)apportionment, (bi-)divisor methods, fair shares.
JEL D70, D71.

1 Introduction

A variety of rules for choosing the representatives in an assembly aim to achieve proportional representation of the parties while organizing elections at a 'local' level. For example, Article 2 EA for the constitution on the European Parliament (EP) states

> *In accordance with its specific national situation, each member State may establish constituencies for elections to the EP or subdivide its electoral area in a different manner, without generally affecting the proportional nature of the voting system.*

The above statement is rather loose, allowing for diverse interpretations and opening up for a variety of implementations. In particular, what is the proportional *nature* of the voting system? Also, why do we want to achieve proportionality at the *local* level (here local meaning a member State)? Is it to achieve proportional representation of the political forces at the global EU level? (This may explain

[1]PSE-EHESS, address 48 bd Jourdan, 75014 Paris, France e-mail demange@pse.ens.fr. I am very grateful to Margaux Luflade for research assistance. I would like to thank two referees and the editor Jean-Francois Laslier for helpful comments.

the reference to nature as a trait inherited by the global system from its parties.) Though, observing the population-to-seats ratios for the EU state members, Germany currently has 99 deputies (96 in 2014), that is 1 for each 826,285 inhabitants while Malta has 6 representatives, that is 1 for each 68,828 inhabitants. With such distortions in the current distribution of the seats to countries, one may wonder whether party-proportionality at the local level has any chance to achieve some form of party-proportionality at the global level.

The European situation is admittedly a most cumbersome and in progress construction. Electoral rules governing elections at the local level may vary across the member States, in particular each member State is allowed to sub-divide itself by Article 2 EA cited above. Furthermore European parties so far are not well identified organizations, which may explain decentralized electoral rules. For most national assemblies elected on the basis of votes in several separate electoral bodies, rules are identical across these bodies and most parties are active in all local elections. However, the interaction between the local and global levels is still a delicate issue even in well structured settings, as will be illustrated by the complexity of the electoral laws in Italy and Germany. The main focus of the paper is to discuss this interaction, in particular to investigate the proportionality of the parties' representation at the global level when 'regional' considerations matter.

Specifically, the object of study is an *apportionment* that accounts for two dimensions, geographical and political, referred to as *bi-apportionment*. An apportionment is the problem of choosing the representatives of an assembly as a function of some numbers, e.g. votes or populations. The problem is bi-dimensional when the whole electoral region is subdivided into several electoral districts (which can be constituencies, regions, cantons as in Switzerland, countries as in the EU etc.) and the representatives are identified both by their district and their affiliation to a party. (Candidates who are not affiliated to a party are gathered into a 'fictitious' party of 'independents'.) Electoral data is thus collected in a matrix and a bi-apportionment specifies how many seats each party receives in each district as a function of this electoral data.[2] In particular, when some populations are over-represented and others under-represented, how to achieve party-proportionality? If indeed party-proportionality can be achieved, what are the implications on the electoral rules and on the representation at the local level? Of course, there are various arguments against or in favor of party-proportional representation. However our purpose is not to discuss these arguments but rather to analyze how to define and implement proportionality assuming it to be indeed a goal. Similarly, we do not discuss the various arguments against or in favor of a proportional representation of the various populations. The allocation of the seats to the various districts —called hereafter district-apportionment— is taken as given without discussing its rationale. Observe that in most situations in which there is a large variability in the size of the districts, the representation of the populations is far from being proportional. Intuitively district distortions may be a source of difficulty to achieve

[2]This description does not specify which of the party candidates will get the positions. In the case of lists, the number of elected positions is enough to know the identity of the elected candidates.

party-proportional representation. The paper surveys proposed bi-apportionment methods with a focus on this conflict.

So far studies on apportionment methods have mostly focused on uni-dimensional settings, in which data consists either in populations or in votes for parties. In a uni-dimensional setting, putting aside the indivisibility of seats, proportionality is simple to define as it corresponds to a simple ratio, often referred to as '(Hare-)quota'. The difficulty in apportioning the seats thus stems from the fact that seats are indivisible. Because of this indivisibility requirement, a variety of methods can be qualified as 'proportional' as put forward by Huntington (1921). In a bi-dimensional setting, new issues arise that are not due to the indivisibility of seats. This leads us to first address 'bi-allocation' problems, similar to the bi-apportionment ones but without requiring (seat) numbers to assume integer values.

A bi-allocation specifies 'shares' of seats for parties in each district, where shares are positive reals. Thus party-proportionality is defined without ambiguity at the district or aggregate levels, as the common notion that the shares obtained by the parties are proportional to their votes. Assuming away the integral requirement on the outcome makes clear the implications of the distortions on the number of seats per district.

A first implication bears on the implementation of party-proportionality. I show that an electoral rule that always achieves party-proportional representation must be centralized in the sense that the outcome in a district may depend on the votes in other districts. The result, though rather intuitive, explains complicated features of electoral formulae that aim at achieving party-proportionality. Most of them start by allocating seats at the local level and suggest various adjustments which turn out to be insufficient to restore global proportionality and furthermore may induce unexpected behaviors, as illustrated respectively by the Italian and German electoral laws. The Italian 'bug', as called by Pennisi (2006), refers to a serious flaw in the electoral rule of the Chamber of Deputies since the procedure itself is ill-defined and may lead to contradictory results between the computations at the national and the regional constituency levels. In the election of the Bundestag deputies in Germany, increasing the ballots for a party may cause a loss in its seats. The causes of such phenomenon are due to the complexity of the procedure and cannot be explained in a few sentences (see Section 3).

A second implication of the distortions in the district-apportionment is that the notion —or ideal— of proportionality, as reflected by the Hare-quotas in the uni-dimensional setting, is not straightforward. Proportionality is now 'constrained' because the district-apportionment fixes the totals for each district. In addition, in order to achieve party-proportional representation at the level of the whole electorate, priority must be given to aggregate votes for parties, which fixes the allocation to the parties as well. The problem hence becomes one of finding a matrix as much proportional as possible to another one (the votes) under the constraints given by the district and party apportionments.[3] A natural candidate is the *fair share* matrix, which is obtained by scaling the rows and the columns by appropriate values, and enjoys nice properties as characterized by Balinski and Demange (1989-a).[4]

Fair shares provide a natural benchmark or target, which constitutes the basis for bi-apportionment methods. Recall that a bi-apportionment must be integer-valued. Two main approaches have been followed, which parallel those followed in a uni-dimensional setting.[5] The first one is based on the idea of starting with the fair shares and 'rounding' them to an adjacent integer in some way while still satisfying the desired constraints on the row and column sums. Gassner (1991) applies this approach to the Belgian Senate. The second approach performs simultaneously the rounding and the adjustment through scale factors (divisors), which may be perceived as ensuring the rounding in a proportional way. This leads to a variety of *bi-divisor methods*, each one characterized by a distinct rounding method, as introduced and characterized in Balinski and Demange (1989-a, 1989-b). The New Apportionment Procedure developed by Pukelsheim (2006) and adopted in various cantons in Switzerland starting with Zurich is based on a bi-divisor method.

Finally, note that the same type of problems arises in many other contexts than an electoral outcome. Still in the political domain, Hylland (2000) suggests a representation for the Parliament of the Federation of Bosnia and Herzegovina. There are ten Cantons and three 'Constituent people', Bosniacs, Croats and Others. Here the matrix gathers the population of the three groups in the ten cantons according to the census. Thus the three 'Constituent people' play the role of the parties. Another example, not in politics, is based on the formation of the board of a scientific society, as studied by Brams (2008), in which the two dimensions are the regions on the one hand and the scientific specialties on the other. Actually, one can think of many situations in which population is divided into categories characterized by two criteria, say a geographical one and a gender or ethnic one.

The paper is organized as follows. Section 2 describes the standard uni-dimensional apportionment, recalls the main discussion on proportionality due to the indivisibility of seats, and introduces the main methods, those based on quotas (such as the largest remainders) and the divisor methods. Section 3 describes bi-apportionment and bi-allocation problems. The conflict between party-proportionality and decentralization of a bi-allocation is presented and illustrated by the Italian and German examples. Finally, fair shares are introduced. Section 4 presents the two main principles for achieving proportionality in bi-apportionments, the (controlled) rounding of the fair shares on the one hand and the simultaneous adjustment by divisors and rounding (bi-divisor methods) on

[3]Gassner (1991) was the first to make explicit the priority of party-representation by fixing the totals for the parties.

[4]Such an adjusted matrix and the algorithms to compute it have been introduced in various areas under different names, such as bi-proportional matrix or RAS model, see Section 3.

[5]The first approach rounds the quotas and then adjusts them in some way, as in the well known greatest remainders method. The second approach rounds and adjusts simultaneously as in the divisor methods.

the other. The section ends up with some discussion.

2 Uni-Apportionment methods and Proportionality

Apportionment is the problem of assigning the seats of an assembly (chamber, house, parliament). This is a typical problem in the political domain. A uni-dimensional apportionment problem appears in two situations, hereafter referred to as respectively as the party and the district problems. In the party problem, the seats are to be allocated to parties as a function of an election outcome. In the district problem, the seats are to be allocated to districts — which can be constituencies, regions, cantons as in Switzerland, countries as in the EU etc. — as a function of their population sizes, possibly among other factors. I first describe the two situations in a similar way to show the formal similarity (using however different notations useful in the study of the bi-apportionment problem). I discuss then proportionality and recall some well-known methods and their possible drawbacks.

2.1 Party and district apportionment problems

Let H be the total number of seats to be assigned in the house. Let us first describe apportionment problems in terms of parties.

In a party-apportionment problem, the total number of seats has to be assigned as a function of the number of votes received by the parties, as is performed by an electoral rule. Let m be the number of parties, v_j be the total votes for party j, and $\mathbf{v} = (v_1, ..., v_m)$ the vector of all votes. A (party-)apportionment assigns the H seats to the parties. Denoting by s_j the number of seats —integer-valued— received by party j, an apportionment is represented by a vector $\mathbf{s} = (s_1, \ldots, s_m)$ where s_j is an integer (possibly null) and $\sum_j s_j = H$. A *party-apportionment method* assigns to each possible electoral outcome $\mathbf{v}$ an apportionment $\mathbf{s}$.

In the district problem, the total number of seats has to be assigned to districts as a function of the electoral population sizes. Let n be the number of districts, p_i be the electoral population in district i. A (district-)apportionment assigns the H seats to the districts, h_i to district i. The apportionment is thus represented by a vector $\mathbf{h} = (h_1, \ldots, h_n)$ where h_i is an integer (possibly null) and $\sum_i h_i = H$. A *district-apportionment method* assigns to each population $\mathbf{p}$ an apportionment $\mathbf{h}$.

In practice, transparency is often an important difference between the party and district methods. Whereas, in democratic countries, the party-apportionment method is strictly defined through the electoral rule, this is not always the case for the district problems. The trend however is to impose more transparency and to explicit criteria for allocating the seats to districts, leading to explicit methods, as witnessed by the evolution in Europe.

Proportionality, quotas Proportionality can be (one of) the objective(s) to achieve when apportioning seats to parties after an election, or to districts after a census. However, if the idea of proportionality sounds straightforward, proportional representation is not. The difficulty arises because seats are indivisible.

Consider an election in which each 'candidate' represents a single party and assume that the goal of the electoral rule is to achieve a proportional representation (as is the case in Israel for the Knesset[6]). The idea is that the number of seats received by a party-candidate should be proportional to its vote total. To evaluate the distortions due to indivisibilities, it is convenient to introduce the 'ideal' proportional benchmark.

The *quota* for party j is the 'fractional' number of seats that j should receive provided the seats were divisible and strict proportionality was the target of the electoral rule: $q_j = \frac{v_j}{V}H$ where V denotes the total number of votes, $V = \sum_j v_j$.

Similarly, denoting by P the total population, $P = \sum_i p_i$, the (Hare)-*quota* of district i is the theoretical $q_i = \frac{p_i}{P}H$ number of seats that i should receive provided the seats were divisible and strict proportionality of the seats to population sizes was the goal of the apportionment.

2.2 Proportional Apportionment methods

Now the problem is formally stated, let us turn to the review of some methods used to perform uni-dimensional apportionments that aim to achieve proportionality. Quotas constitute benchmark values for proportionality. Recall however that an apportionment must be integer-valued. Apportionment methods that aim to achieve proportionality all differ in the way they approximate this target. Of course they produce the quotas when these are all integer-valued.

Since quotas constitute a kind of target a very natural idea is to 'round' them in such a way that the number of allocated seats is kept equal to H. I describe the most well-known method in this class, the largest remainders. I present then a second class of methods, in which the rounding operates on scaled quotas.

Rounding quotas The largest remainders method has been used in many countries and is known under different names: Hamilton, Hare-Niemeyer or Vinton's. The description is simple: first, give every party its lower quota, and second distribute the remaining seats to the parties that have the largest remainder left.

[6]Israel uses the closed list method of party-list proportional representation, meaning that citizens vote for their preferred party and not for any individual candidates. The election is nationwide. The 120 seats in the Knesset are then assigned (using the D'Hondt method) proportional to each party that received votes, provided that the party gained votes which met or exceeded a 2 % electoral threshold. Parties are permitted to form electoral alliances so as to gain enough collective votes to meet the threshold (the alliance as a whole must meet the threshold, not the individual parties) and thus be allocated a seat. The low threshold makes the Israeli electoral system more favorable to minor parties than systems used in most other countries.

More formally, in terms of parties, the method first consists in computing parties' quotas and allocating to each party as many seats as prescribed by the integral part of its quota: $\lfloor q_j \rfloor$. If all seats have not been allocated during this first step, the unfilled seats are distributed to the parties with the largest remainders, where party j's remainder is defined as the number of its votes not used yet to win a seat: $q_j - \lfloor q_j \rfloor$. Of course, the method can be formulated in the same way in terms of districts.

The largest remainders method however suffers from serious drawbacks, put forward by various 'paradoxes'. These paradoxes are revealed as some parameters vary. One of them is often called the 'Alabama paradox'; as the house size increases, a district (resp. a party) may see its number of seats decrease: the largest remainders method is not monotone with respect to the house size. Another paradoxical phenomenon is known as the 'population paradox'; as its population (resp. number of votes) increases more relative to another state (resp. party), a district (resp. a party) may lose seats to this other state (resp. party). A third possible paradox is known as the 'new state paradox' (or 'Oklahoma paradox'). If, all other things being equal, a new district is added to the union and the house size is increased by the amount of seats corresponding to the new district according to its population, then when re-apportioning the seats in the extended union, former states may see their number of seats vary (while their population has not).

The three paradoxes are due to the fact that remainders vary differently for large and small states. To be more precise, consider the Alabama paradox for example. As the house size varies, although quotas vary in the same *proportion* for all states, the largest remainders method is affected by their variations in *absolute* terms; as a result, since the quotas of large states increase by more than the ones of the small states, the latter may thus lose their priority on extra seats.

The argument extends to any method that rounds the quotas. The population-paradox cannot be avoided if one sticks to the requirement of staying within quotas, that is, of allocating each district/party a number of seats lying between the down-rounding and up-rounding of its quota (Balinski and Young 1982). Specifically there is no method that is monotone with respect to population and always stays within quotas.[7] These drawbacks are avoided by adjusting proportionally the quotas and rounding *simultaneously*, as is performed by the divisor methods.

(Huntington) divisor methods Huntington (1921) elicited a unifying principle behind five common methods and coined the term 'divisor methods'. All are monotone in whatever sense (but none rounds the quotas).

Let us describe one of them, called Adams' method or method of smallest divisors. Each quota is adjusted multiplicatively by the same scale factor so that after rounding down, the sum of the allocated seats is equal to H. Hence the key difference with the greatest remainders method is that instead of rounding down the quota and completing the first apportionment in some way so as to

[7] The result does not hold if one is ready to use random rules, as shown by Grimmett (2004).

allocate all the seats, one adjusts proportionally the quotas and rounds *simultaneously*. It is not hard to see then that the method is monotone.[8]

The same principle applies to each divisor method. They are all obtained by simultaneously scaling and rounding but differ in the rounding method[9]. Formally, a d-rounding is characterized by the thresholds $d(n)$ in $[n, n+1]$ defined for each natural number n. The thresholds serve as a basis for rounding: Each number x in the interval $]d(n), d(n+1)[$ is rounded to n. Threshold $d(n)$ can be rounded to n or $n-1$. Formally the *d-rounding* $[.]_d : \mathbb{R} \to \mathbb{N}$ is defined by:

$$\text{for } x \in [n, n+1],\ [x]_d = \begin{cases} n & \text{if } x < d(n) \\ n+1 & \text{if } x > d(n) \\ n \text{ or } n+1 & \text{if } x = d(n) \end{cases}.$$

The *divisor method (based on d)* associates to any voting results $(\mathbf{v}, H)$ the apportionment(s) $\mathbf{s}$ that satisfy (generically the apportionment is unique):

$$\mathbf{s} = \big(s_j\big) | s_j = [\lambda v_j]_d \text{ for } \lambda > 0 \text{ such that } \sum_j s_j = H.$$

Similarly in a district problem it assigns $\{\mathbf{h} = \big(h_i\big) | h_i = [\lambda p_i]_d$ for a $\lambda > 0$ such that $\sum_i h_i = H\}$. Adams' method is the divisor method based on the divisor criterion $d : n \to n$ (for all $n \in \mathbb{N}$), so that any real number $x \in (n, n+1]$ is rounded up to $n+1$ by the d-rounding. Alternative well-known divisor methods are Jefferson's (or D'Hondt's) based on $d : n \to n+1$ (down-rounding) or Webster's (or Sainte-Laguë's) based on $d : n \to n + \frac{1}{2}$ (standard rounding). Hill's method is based on $d : n \to \sqrt{n(n+1)}$.

Divisor methods avoid monotonicity paradoxes and are the only ones to do so together with some desirable properties as shown by Balinski and Young (1982).

The main lesson to draw is that, assuming that proportionality as embodied in the quotas is a goal, there are different reasonable ways to approximate this goal with integers. Another approach based on optimization rather than properties confirms this. Huntington showed that each of the five main divisor methods and the largest remainders all minimize the distance to the quotas, each of them corresponding to a different measure of inequality.

2.3 Discussion

In practice, proportional representation is often mitigated by additional features even when it is the main goal of the apportionment method.

In the party problem, the stability of the assembly is a concern and the method is likely to be amended and biased in favor of large parties. Such a bias is often implemented through thresholds, according to which parties that do not reach the threshold get no seats but proportionality remains

[8] Furthermore it satisfies consistency requirements, as developed in a bi-dimensional setting in Section 4.2.

[9] I find it more convenient to work with a multiplier rather than with a divisor.

the goal for qualified parties.[10] Also, among the many methods that can be qualified as proportional, some tend to favor more (in the statistical sense) large parties than small ones under rather natural assumptions, as is the case for the Jefferson's rule in the class of divisor methods (see Gallagher 1992 for a comparison of various proportional systems and the impact of thresholds among other features). More drastically, as said in the introduction, proportionality may not be a concern at all but the priority may rather be to provide incentives to form large parties. We will not discuss this situation.

In the district problem, proportionality as required by the United States constitution[11] for the House of Representatives is quite the exception. A most common departure from proportionality is the minimal number of seats received by each district. Even in the US each state must be 'represented', meaning that it is entitled to one representative at least. When districts are fixed for historical reasons, and have very different sizes, the minimal representation introduces a severe distortion with respect to Hare-quotas. There are also explicit claims to favor small populations such as the Treaty of Lisbon for the European parliament which requires the allocation of seats to satisfy a condition of 'degressive proportionality' (see Grimmett (2011) for example).

The rationale behind and the magnitude of the distortions are not necessarily made precise. Distortions may be due to a lengthy unification process that has given power to small entities not in proportion to their sizes. Nevertheless, there is a debate about whether population-proportionality and quotas constitute indeed relevant benchmarks. The argument is that population-proportionality is unfair in terms of the distribution of the 'voting power' of the representatives in the assembly. Such voting power, measured by the ability to change the decisions, would be increasing more than proportionally to the number of representatives. As already noticed by Penrose (1946), this would be due to the 'block' behavior of the representatives who hence do not truly represent their electoral bodies. Proportional representation then induces an inequitable power distribution, distorted in favor of districts with large populations (see e.g. the survey of Widgren (2005)).

3 Bi-apportionment and bi-allocation problems

Bi-apportionment problems combine the district and party dimensions. In many elections for choosing representatives in an assembly, the electoral body is divided into several electoral districts, and the results obtained in each district matter for electing the representatives of that district in the assembly (if the results are aggregated the problem comes down to the party apportionment problem).

[10] Thresholds are present in most systems. Furthermore, a positive value for $d(0)$ in a divisor method plays a similar role as a threshold, though endogenous. Since the d-rounding value $[x]_d$ is null for x smaller than $d(0)$, a party is awarded no seat if λv_j is smaller than $d(0)$ where λ is the (endogeneous) multiplier.

[11] Article 1 of the United States constitution states '*Representatives and direct Taxes shall be apportioned among the several States which may be included within this Union, according to their respective Numbers, which shall be determined by adding to the whole Number of free Persons, including those bound to Service for a Term of Years, and excluding Indians not taxed, three fifths of all other persons.*'

At the same time, the elections in the districts cannot be viewed as separate elections for various reasons. In most circumstances, candidates are affiliated to parties and the outcome determines not only the representatives but also the strength of the different parties and ultimately, in some countries, the prime minister.

In systems with several electoral bodies, the number of seats allocated to a district is fixed, determined prior to the election more or less formally, either by historical factors or negotiations or through a formal constitutional process as in the US. In our terminology, the district-apportionment has been decided and we take it as given. As will be clear, for our purpose, only the distortions with respect to the populations matter but not the process that leads to them. The distortions are summarized by the ratios of the number of seats over the population size in each district. These ratios may not be all equal, even if the apportionment has been conducted through a proportional method, and there can be substantially large distortions due to thresholds or more or less explicit political considerations, as we have seen.

3.1 Bi-dimensional problems

Formally, let H be the total number of seats to be assigned in the house and n be the number of districts. Each district i is represented by a positive number of seats in the house h_i, $i = 1, \ldots, n$. Thus the district apportionment is described by the n-tuple of positive integers $\mathbf{h} = (h_1, h_2, \ldots, h_n)$ which sum to H, $H = \sum_i h_i$. Given p_i the size of the electorate in district i and P the total size, $P = \sum_i p_i$, the distortion with respect to proportionality is described by the ratio of the number of seats received by i to its population $\frac{h_i}{p_i}$. When these ratios are all equal, they are equal to $\frac{H}{P}$. But, most often, the ratios differ.

Definition 1 *Given population sizes* $\mathbf{p}$*, the district-allocation* $\mathbf{h}$ *is said to be distorted if the per district ratios* $\frac{h_i}{p_i}$ *are not all equal.*

In what follows, the population sizes $\mathbf{p}$ and the district-allocation $\mathbf{h}$ are given and this will not be repeated.

Votes are now described by a $n \times m$ matrix $\mathbf{v} = (v_{ij})$ in which row i represents district i, column j party j, and v_{ij} the votes obtained by j in district i. Row i, denoted by $\mathbf{v}_{i.}$, represents the votes in district i for each party and column j, denoted by $\mathbf{v}_{.j}$, represents the votes for party j in each district.

In the following, $\mathcal{V}(\mathbf{p})$ denotes the set of matrices $\mathbf{v}$ with non-negative elements for which the i^{th} row sum is equal to the size of the electoral population in district i, p_i. This implies that turnout is neglected. Accounting for it does not alter the results, as discussed at the end of the proof of Proposition 1.

Given a district-apportionment $\mathbf{h}$, the electoral rule assigns seats to parties in each district so as to satisfy the district constraints described by $\mathbf{h}$. Let b_{ij} represent the number of seats received by j in district i. Formally,

Definition 2 *A* bi-apportionment *is represented by a matrix with non-negative integer elements* $\mathbf{b} = (b_{ij})$ *that satisfies the district constraints* $\mathbf{h}$*:* $\sum_j b_{ij} = h_i$ *for each* i*. A* bi-apportionment method $\mathbf{A}$ *assigns to each voting matrix* $\mathbf{v}$ *in* $\mathcal{V}(\mathbf{p})$ *a bi-apportionment.*

Bi-apportionment methods have so far been relatively little studied. When dealing with proportionality, new issues arise that are not due to the indivisibility of seats. Therefore, I first assume away the integral requirement on the outcome to make clear that the root of the problem comes from the distortions on the number of seats per district relative to their population size. In the following, an *allocation* problem refers to a situation in which the outcome is not required to be integer-valued. (I keep using the term 'seats' to simplify.)

Definition 3 *A* bi-allocation *is represented by a matrix with non-negative elements* $\mathbf{a} = (a_{ij})$ *that satisfies the district constraints* $\mathbf{h}$*:* $\sum_j a_{ij} = h_i$ *for each* i*. A* bi-allocation method $\mathcal{A}$ *assigns to each voting matrix* $\mathbf{v}$ *in* $\mathcal{V}(\mathbf{p})$ *a bi-allocation.*

3.2 Decentralized bi-allocation methods.

For an allocation, which is not subject to integrality requirement, proportionality is defined without ambiguity as the common notion of 'proportional representation' of the votes, and this can be considered either at the local (district) or global level. In what follows, party-proportionality is meant at the global level. This section presents some results on the conflict between party-proportionality and 'decentralization' of the electoral rule at the district level. The conflict arises when distortions are present in the district-allocation. Though simple, these results explain complicated features of electoral formulae that aim at achieving party-proportionality.[12]

A bi-allocation method is said to be *party-proportional* if it assigns a total equal to its quota of the aggregate votes to each party. It is said to be *district-decentralized* if the allocation in district i depends on the votes of citizens in district i only. Formally,

Definition 4 $\mathcal{A}$ *is* party-proportional *if for each* $\mathbf{v}$ *in* $\mathcal{V}(\mathbf{p})$*, each party* j *receives its quota of aggregate votes:* $\sum_i \mathcal{A}_{ij}(\mathbf{v}) = \sum_i v_{ij} \frac{H}{P}$ *for each* j*.*

$\mathcal{A}$ *is* district-decentralized *if for each* $\mathbf{v}$ *and* $\mathbf{v}'$ *in* $\mathcal{V}(\mathbf{p})$ *that have identical* i*'s rows, then* $\mathcal{A}(\mathbf{v})$ *and* $\mathcal{A}(\mathbf{v}')$ *have identical* i*'s rows:* $\mathbf{v}_{i.} = \mathbf{v}'_{i.}$ *implies* $\mathcal{A}_{i.}(\mathbf{v}) = \mathcal{A}_{i.}(\mathbf{v}')$*.*

A simple district-decentralized method computes separately the proportional allocation in each district. This is the subject of the next section.

[12] As far as I can see, they have not appeared in the literature.

3.2.1 The decentralized proportional method

Given the electoral outcome, the decentralized proportional method assigns to parties in each district i its quota in that district given the available number of seats h_i. Specifically, the decentralized method allocates $a_{ij} = v_{ij}\frac{h_i}{p_i}$ to party j in district i.

Proposition 1 *Let the district-allocation* $\mathbf{h}$ *be distorted. The decentralized proportional method is not party-proportional.*

PROOF The argument is very simple. The decentralized method assigns a total to j equal to $\sum_i a_{ij}$ where $a_{ij} = v_{ij}\frac{h_i}{p_i}$, which has to be compared with the quota, $q_j = \sum_i v_{ij}\frac{H}{P}$. Thus the decentralized method assigns a different total to j than its quota if

$$\sum_i v_{ij}\frac{h_i}{p_i} \neq (\sum_i v_{ij})\frac{H}{P}. \tag{1}$$

Both terms are linear functions of the votes v_{ij} and these functions precisely differ when the district-allocation $\mathbf{h}$ is distorted. Hence inequality (1) holds true for almost any $\mathbf{v}$. ∎

The proof uses the fact that the vote outcomes run in $\mathcal{V}(\mathbf{p})$, in particular that their patterns vary across districts. If votes were proportional across districts, inequality (1) would never be satisfied. But proportionality of the votes is unlikely. Even on average it is unlikely to be satisfied. When votes show some regular patterns, not only the decentralized proportional method is likely not to assign their quotas to parties, but also the method may induce systematic bias. The intuition is that a party which tends to receive large shares in districts that are favored by the district-allocation will receive more than its quota. To give a meaningful example, order the districts by decreasing ratios $\frac{h_i}{p_i}$. Let party j's vote share in each district, $\frac{v_{ij}}{p_i}$, be decreasing in i. Thus party j is relatively stronger in districts that are favored, i.e. that have a large ratio seats to population. Since the ratios $\frac{h_i}{p_i}$ are decreasing in i, they first are larger than $\frac{H}{P}$ and then smaller. Hence for some k we have that $(h_i - \frac{H}{P}p_i)$ is non-negative for $i \leq k$ and non-positive for $i \geq k$ with at least one strict inequality. Furthermore, by the assumption that $\frac{v_{ij}}{p_i}$ is decreasing in i, the term $(\frac{v_{ij}}{p_i} - \frac{v_{kj}}{p_k})$ is also non-negative for $i \leq k$ and non-positive for $i \geq k$. Thus

$$\frac{v_{ij}}{p_i}(h_i - \frac{H}{P}p_i) \geq \frac{v_{kj}}{p_k}(h_i - \frac{H}{P}p_i)$$

for each i with a strict inequality for at least one i. Summing over all i's the sum on the right hand side is null and we obtain

$$\sum_i v_{ij}\frac{h_i}{p_i} > (\sum_i v_{ij})\frac{H}{P} \tag{2}$$

hence the decentralized method assigns a higher total to j than its quota. Thus, assuming a distribution on the votes, the method produces a positive bias relative to the quota in favor of a party j that is statistically relatively stronger in favored districts (of course the bias can be the other way).

As said before, the above analysis assumes the number of votes in each district is taken as the electoral population size. This is indeed natural to define distortions with respect to the electoral population size. However, because of turnout, the number of votes in a district i, $v_{i+} = \sum_j v_{ij}$, differs from p_i, possibly by a significant amount. In that case the number of seats obtained by j in district i is given by $\frac{v_{ij}}{v_{i+}} h_i$. When turnout is identical in all districts, nothing is changed in the above proof since distortions are unchanged, namely the ratios seats to population and seats to voters are equal up to a multiplicative factor. Similarly, when turnout differs across districts in a systematic manner, proposition 1 holds true by evaluating the distortions in the district-allocation with respect to these expected electorate size. There is no reason for the distortions with respect to population size to be exactly compensated by turnout (in other words, if the $\frac{h_i}{p_i}$ are unequal, then the $\frac{h_i}{v_{i+}}$ are unlikely to be equal). Actually, it may even be the opposite if the districts with small populations are favored and their electorate tends to have a high turnout. In France for example, the districts with small populations are the rural ones, with relatively old populations who tend to vote more often than younger urban dwellers. When turnout is somewhat unpredictable and varies across districts, i.e. the v_{i+} are random and not proportional to the p_i, the set of electoral outcomes $\mathcal{V}(\mathbf{p})$ is enlarged. Then Proposition 1 a fortiori holds true.

3.2.2 Impossibility result

Proposition 1 shows that proportional representation of the parties at both local and global levels are two conflicting objectives when distortions are present in the district-allocation. The next proposition states the fundamental conflict between district-decentralization and proportional representation at the global level (still under distortions).

Proposition 2 *Let the district-allocation* $\mathbf{h}$ *be distorted. There is no bi-allocation method that is both district-decentralized and party-proportional.*

Proof Let $\mathcal{A}$ be a bi-allocation method that is both district-decentralized and party-proportional. Decentralization implies that the allocation in a district i depends only on i's results $\mathbf{v}_{i.}$. Let us write $a_{ij}(\mathbf{v}_{i.})$ the amount allocated by $\mathcal{A}$ to j. Party-proportionality requires for each $\mathbf{v}$ in $\mathcal{V}(\mathbf{p})$

$$\sum_i a_{ij}(\mathbf{v}_{i.}) = (\sum_i v_{ij})\frac{H}{P} \text{ for each } j. \tag{3}$$

Thus $a_{ij}(\mathbf{v}_{i.}) = v_{ij}\frac{H}{P} + c_j$ where c_j is some 'constant', i.e., is independent of $(\mathbf{v}_{i.})$.

Now c_j must be non negative because otherwise for low enough j's scores v_{ij}, $a_{ij}(\mathbf{v}_{i.})$ would be negative. If the district-allocation is distorted, surely $\frac{H}{P} > \frac{h_i}{p_i}$ for some i. Choose such a district. We have that $a_{ij}(\mathbf{v}_{i.})$ is at least $v_{ij}\frac{H}{P}$ by non-negativity of c_j. Summing over j yields $\sum_j a_{ij}(\mathbf{v}_{i.}) \geq \sum_j v_{ij}\frac{H}{P}$. Since the allocation satisfies the district constraints $\sum_j a_{ij}(\mathbf{v}_{i.})$ is equal to h_i. As $\sum_j v_{ij} = p_i$, we obtain $h_i \geq p_i\frac{H}{P}$, a contradiction. ∎

An implication of these propositions is that party-proportionality requires the number of seats received by the parties in a particular district to depend on the results in other districts. The conflict is magnified by the distortions in the allocation of the seats to districts or the variations in preferences across parties.

Introducing the integer requirement cannot solve this problem. A party-proportional representation must be computed on the basis of the whole matrix of votes. We illustrate this with the Italian and German examples. Although party-proportionality is not the only goal, it is a concern in both cases, but tackled in different ways. Whatever case, the interaction between the local and global computations generates difficulties.

3.3 The Italian and German examples

The Italian 2005 electoral law is a good example of the failure of decentralization (we rely here on Pennisi (2006)). The law aims at allocating the 630 seats of the Chamber of Deputies proportionally to the votes obtained by the parties at the national level and within 27 multi-member regional constituencies with possibly an adjustment so as to to ensure the existence of a majority. The number of seats h_i given to each constituency i is pre-determined. The procedure is as follows. In a first step, the seats are apportioned to parties at the national level on the grounds of their vote totals. This is done by a modified version of the largest remainders method to ensure the existence of a majority in the Chamber, which produces a 'majority' and 'minority' coefficients representing respectively the number of votes per seats for the majority list and for all other parties (aggregated). In a second step, the seats are allocated to the parties at the constituency level. In each district i, this involves the computation of the districts' coefficients again for the majority party and for the others. Multiplying h_i by these coefficients, rounding down and assigning the remaining seats to the largest remainders, one gets the allocation in district i. However, there is no reason why the sum of seats obtained by a party over all constituencies would be equal to the number of seats assigned to the party at the national level! When a party's constituency seats do not sum up to the number prescribed at the national level, a correction mechanism is applied, which modifies the priority order of parties on extra seats. This mechanism does not tackle the true source of seat surplus and deficits —namely decentralization— and thus does not correct anything. In particular, it does not correct what Pennisi (2006) defines as the three paradoxes of such a procedure: the surplus paradox (when a party gets more seats than it is apportioned at the national level although it did not get any extra seat by the largest remainders), the deficit paradox (conversely, when a party gets less seats than it deserves although it got extra seats in every constituency) and the constituency paradox (when, due to the application of the correction mechanism, the number of seats attributed in constituency i differs from h_i).

The German electoral law for Bundestag members is a different example of the possible consequences of decentralization. The procedure is rather complex (see Pukelsheim (2006)): voters cast 2 ballots, one constituency ballot (vote for a person) and one party ballot (vote for a party). It aims at allocating the 598 seats (2 for each of the 299 constituencies) to parties proportionally to the total party ballots they get while enabling voters to choose the candidates who will fill the seats thanks to the constituency ballot. In a first step, the 598 seats are apportioned to the parties proportionally to their total party votes. Then, 299 'direct' seats are attributed to the 299 constituency winners, that is, the candidates getting the most votes in their constituency. Finally, the 299 remaining 'list' seats are distributed among the parties in order to satisfy the national apportionment. Due to partial decentralization and possible misalignment of the constituency and party votes, three main problems arise: 'overhang seats' (when a party wins more direct seats than the number of seats it is apportioned at the national level, additional seats are created on top of the 598 existing ones to satisfy all direct winners, so that the size of the Bundestag varies), 'doubly successful votes' (when the constituency ballot contributes to the victory of a deputy while the party ballot is aggregated in another party list) and 'negative ballot weights' (when an increase in its party ballots causes a party to lose a seat).

3.4 Fair shares

Proposition 2 states the conflict between party-proportionality and a district-per-district procedure. The conflict can be solved once centralization is allowed (it should be clear that centralization here means that the shares obtained by parties in a given district may depend on votes in other districts). Not only are there allocations that achieve party-proportional representation while abiding by the district apportionment, but there are numerous. These allocations are the positive $\mathbf{a}$ that satisfy

$$\sum_j a_{ij} = h_i \text{ for each } i, \sum_i a_{ij} = q_j \text{ for each } j \tag{4}$$

where $\mathbf{h}$ and $\mathbf{q}$ are respectively the district-apportionment and the parties' quotas. Among them, the fair share matrix, which is obtained by a scaling of the rows and the columns, is a good candidate to represent the idea of proportionality while accounting for a priori constraints, as argued by Balinski and Demange (1989-a).

It is useful to formulate the problem in the more general form of adjusting a matrix $\mathbf{v}$ so as to match constraints on row and column sums given by $\mathbf{r}$ and $\mathbf{c}$ respectively (i.e. not necessarily the district-apportionment or the quotas). To simplify the presentation, let us assume all elements in matrix $\mathbf{v}$ to be positive. The possibility of zero elements is discussed in Section 4.3.

Definition 5 *Given a matrix* $\mathbf{v}$ *with positive elements and row and column sum constraints described by* $\mathbf{r} = (r_i)$ *and* $\mathbf{c} = (c_j)$*, there is a unique matrix* $\mathbf{f}$ *of the form* $\left(f_{ij} = \lambda_i v_{ij} \mu_j\right)$ *that matches the row and column sums:*

$$\sum_j f_{ij} = r_i \textit{ for each } i, \sum_i f_{ij} = c_j \textit{ for each } j. \tag{5}$$

Call **f** *the fair share (matrix) to problem* (**v**, **r**, **c**)*. The fair share method assigns to any problem* (**v**, **r**, **c**) *its fair share.*

The fair share corresponds to a common solution for adjusting a matrix used in various areas such as in statistics for adjusting contingencies tables, in economics for balancing international trade accounts (in the RAS model) or for filling missing accounting data as in cross-banking relationships.[13] Observe that the 'multipliers', the λ_i and the μ_j are defined up to a multiplicative constant: given a set of multipliers (λ, μ), multiplying each λ_i by a constant and dividing each μ_j by the same constant gives another set of multipliers. Only their relative values matter.

The fair share method is characterized by three axioms, called exactness, homogeneity, and consistency. Exactness requires that if the exact proportional matrix with global sum equal to H matches the constraints, then it is the solution. Homogeneity requires the simple idea of proportionality between rows or columns: if two rows (or columns) are proportional and assigned the same total, they must be assigned the same allocations. Consistency is a familiar and key property in many fair division problems. It asks that any part of a (fair) allocation is itself a (fair) allocation. The formal definitions are in Balinski and Demange (1989-a) (where the term uniformity is used instead of consistency) who prove the next proposition.[14]

Proposition 3 *The fair share method is the unique allocation method that satisfies exactness, homogeneity, and consistency.*

It is easy to check that the fair share method also satisfies population monotonicity.

The computation of the fair share matrix relies on a well-known procedure of alternately scaling rows and columns, called iterative proportional fitting procedure or RAS algorithm. The procedure starts by allocating shares in a decentralized way district per district. Party j is entitled to $\lambda_i^1 v_{ij}$ shares in district i where λ_i^1 is set equal to district i's ratio. Hence setting the parties' multipliers μ_j^1 to 1, the matrix with general element $(\lambda_i^1 v_{ij} \mu_j^1)$ satisfies the district constraints. If all parties receive their required totals, the matrix is the fair share to problem (**v**, **r**, **c**) and the process stops. Otherwise, some parties get more than their assigned total (for instance more than their quotas if **c** is taken equal to **q**) and some less. Parties' multipliers are adjusted down or up so as to meet the parties' constraints: this defines the new values (μ_j^2) so that $\sum_j \lambda_i^1 v_{ij} \mu_j^2 = c_j$ for each j. If the matrix $(\lambda_i^1 v_{ij} \mu_j^2)$ satisfies the row constraints, it is the fair share. Otherwise the procedure starts over again, alternating row-scaling for odd iterations and column-scaling for even iterations. The

[13] The fair share matrix is sometimes called bi-proportional matrix. Bi-proportionality may introduce some confusion as it may suggest that the row and column sums are proportional apportionments respectively for districts and parties. In addition the fair share can be defined in the more general setting in which row and column sums may be only constrained to belong to some intervals rather than being assigned some values.

[14] More precisely Proposition 3 follows from Theorem 2, the statement of which differs because it considers a more general situation with inequalities bounds (see footnote 13). The proof however involves a step that treats equality constraints on row and column sums as in Proposition 3.

process converges to the fair share matrix (see Bacharach (1965) who also studies carefully the case where the matrix $\mathbf{v}$ has some zeros).

Proposition 3 applies whatever values for the constraints on totals as described by $\mathbf{r}$ and $\mathbf{c}$. Specifying the column sums to be the quotas of the parties at the global level, the fair share method is also party-proportional. Specifically, the method that assigns to results $\mathbf{v}$ and district-apportionment $\mathbf{h}$ the fair share associated to $(\mathbf{v}, \mathbf{h}, \mathbf{q})$ is party-proportional and satisfies the properties stated in Proposition 3. The impact of the distortions in the district-apportionment are channelled through the multipliers. At the fair share, a party that gets relatively large votes in favored districts has a 'small' multiplier, smaller than one that gets relatively few votes in favored districts. This must hold because otherwise the former would receive too many seats relative to its global score or the latter too few seats.[15] Though, by construction, multipliers (or more precisely their relative positions) have no impact on the total number of seats of a party, they affect the distribution of the parties' representatives across districts. An implication of this is that 'reversals' are possible.

A *reversal* arises in an allocation when a category that has less voting support than another gets a larger share: formally if $v_{ij} < v_{k\ell}$ and $a_{ij} > a_{k\ell}$ for some pairs (i, j) (k, ℓ). The absence of reversals within a district is probably the more important issue politically. A reversal within a district means that a party with higher scores than another one gets less shares. Using the expression $a_{ij} = (\lambda_i v_{ij} \mu_j)$, a reversal in district i occurs when $v_{ij} < v_{i\ell}$ and $v_{ij}\mu_j > v_{i\ell}\mu_\ell$. In other words the advantage of ℓ relative to j in district i is not enough to compensate the necessary corrections at the global level between the parties as measured by the ratio μ_j/μ_ℓ, i.e. $1 < v_{i\ell}/v_{ij} < \mu_j/\mu_\ell$.

So, the likelihood of a reversal depends on the relative differences in the parties' multipliers. From the previous analysis, reversals are more likely to arise the more severe the distortions in the district-apportionment and the more diverse the pattern of the parties' votes. Finally observe that the argument carries over, and even has more bite, when the party constraints are not given by the quotas. This is surely the case when considering apportionments and parties' quotas are 'rounded' in some way. Reversals will be illustrated in the next section.

4 Bi-apportionment methods

A fair share matrix represents a kind of ideal proportionality under some constraints. It remains to translate this ideal target into a bi-apportionment with integers. A fair share matrix is defined for general constraints. Without integrality requirements, party-proportionality is easily defined by fixing the column totals to the quotas. With integrality requirements, party-proportionality may be understood in various ways, in particular at the local or global level.

[15] In the general case where vote shares are not necessarily monotone in the distortions in districts, the ordering of the multipliers is not easy to derive. For example, correlations between vote shares and distortions do not give enough information: multipliers are not necessarily ordered (in the reverse sense) as the correlations.

In what follows, whatever method, the priority is given to results at the national level. This is achieved by carrying out the method in two steps. In the first step, the H seats are apportioned to the parties on the basis of their overall vote totals: $\mathbf{s} = (s_j)$ would be the result of a national election if districts did not matter. The second step computes a bi-apportionment assigned to the ($\mathbf{v}$, $\mathbf{h}$, $\mathbf{s}$) where the constraints on districts are given (the h_i's) and those on parties (the s_j's) have been determined in the first step. For this second step, two classes of methods are possible, as in the uni-dimensional setting.

The chosen (uni-)apportionment method in the first step allows to achieve the desired properties on the parties' representation. When party-proportionality is a goal, a 'proportional' method should be chosen among the ones described in Section 2. If party-proportionality is not a goal and another apportionment method is chosen to determine this first party-apportionment, the second step goes through. The only difference is that the bi-apportionment will satisfy some properties that differ from party-proportionality but that will meet other pre-defined requirements.

Before proceeding, it is useful to note that the rounding problem encountered at the second step is not a bi-dimensional extension of the uni-dimensional problem. Once both the party and district apportionments are fixed, there is less flexibility than in the general uni-dimensional problem; for example, in the case of a single district, the second step becomes a vacuous problem.[16]

Running example Gassner (1991), motivated by the severe drawbacks of the Belgian electoral law, expressed the idea of apportioning seats at the global level and then using this apportionment as a constraint when distributing the seats at the local level. She looked at the 1981 Senate election in the former Belgian province of Brabant (see her paper for the law in place at that time and the 'bad' outcome). 24 seats are to be allocated among 3 districts (Brussels, Leuven and Nivelles) and the 10 parties which take part in the election. Prior to the vote, the district-apportionment is fixed to 17 seats for Brussels, 4 for Leuven and 3 for Nivelles. This example will serve as a running example to instance several points in the sequel. [17]Here it is used to illustrate the presence of reversals and that rounding fair shares is not trivial.

The first table corresponds to the matrix of votes $\mathbf{v}$:

	CVP	FDF-RW	PVV	PRL	PS	SP	VU	UDRT	PSC	Ecolo
Br.	109,377	148,928	88,645	106,920	80,644	53,409	63,807	59,730	54,549	34,966
Le.	78,280	2,048	70,273	1,765	0	60,024	32,178	4,203	0	0
Ni.	0	17,879	707	47,579	48,965	0	923	16,984	24,590	13,060

[16]The setting with inequalities constraints as considered by Balinski and Demange (1989-a), (1989-b) provides a uniform framework that encompasses both the uni-dimensional setting and the bi-dimensional setting with equalities constraints.

[17]The matrix has some null elements but the methods are nevertheless well defined for that matrix; see more in Section 4.3.

As a first step, we take the party-apportionment as obtained from the parties' overall votes by Jefferson's method. The following table illustrates this step; the second row indicates the parties' overall vote totals while the third gives the number of seats they are globally apportioned:

CVP	FDF-RW	PVV	PRL	PS	SP	VU	UDRT	PSC	Ecolo
187,657	168,855	159,625	156,264	129,609	113,433	96,908	80,917	79,139	48,026
4	4	3	3	2	2	2	1	1	1

Then, performing the iterative fitting procedure with $\mathbf{v}, \mathbf{h} = (17, 4, 3)$ and $\mathbf{s} = (4, 4, 3, 3, 2, 2, 2, 1, 1, 1)$ gives the fair share matrix:

	CVP	FDF-RW	PVV	PRL	PS	SP	VU	UDRT	PSC	Ecolo
Br.	2.6282	3.5879	1.8926	2.1418	1.9604	1.0991	1.4488	0.7694	0.7175	0.7540
Le.	1.3715	0.0360	1.0939	0.0258	0	0.9006	0.5327	0.0395	0	0
Ni.	0	0.3763	0.0132	0.8326	1.0399	0	0.0183	0.1911	0.2826	0.2460

Observe that a reversal arises in the Nivelles district: FDF-RW, which got $17,879$ votes, is allocated 0.3763 shares while PSC is allocated 0.2826 shares for $24,590$ ballots. The multiplier for FDF-RW relative to PSC is $\mu_{\text{F}}/\mu_{\text{P}} = 1.83$. Since $24,590/17,879$ is clearly smaller than this ratio, the correction performed by the multipliers in favor of FDF-RW relative to PSC at the global level outweighs the advantage PSC had over FDF-RW in terms of votes in the Nivelles district. The explanation for the ratio here is due to the fact that FDF-RW, a large party, is favored by the Jefferson apportionment. With not much more than a double number of seats as PSC, it gets four times as much as seats. Its pattern of votes, with a large proportion of its votes in Bruxelles, which is slightly favored,[18] weakens this effect (arguing as in the previous section) and explains that the ratio is much smaller than 4 !

This example illustrates the importance of the choice in the party-apportionment. In particular here the fair share associated to the parties' quotas, i.e., to $(\mathbf{v}, \mathbf{h}, \mathbf{q})$, has no reversal.

The fair shares constitute a benchmark for the apportionment. Though, rounding the fair share matrix in a standard fashion, that is taking the standard rounding of *each* of its elements, cannot be used as a method to find the apportionment, namely the matrix of rounded elements may very well fail to satisfy the row and column sum constraints:

	CVP	FDF-RW	PVV	PRL	PS	SP	VU	UDRT	PSC	Ecolo	District-Sum
Br.	3	4	2	2	2	1	1	1	1	1	18
Le.	1	0	1	0	0	1	1	0	0	0	4
Ni.	0	0	0	1	1	0	0	0	0	0	2

[18] The ratio of the seats to the number of votes is the largest in Bruxelles, possibly due to low turnout.

We have used here standard rounding. The same problem arises when using another fixed rounding method (not necessarily for this same matrix): d-rounding each element of the fair share of $(\mathbf{v}, \mathbf{h}, \mathbf{s})$ does not always yield a matrix that meets the row and columns constraints. As in the uni-dimensional case (to some extent), two approaches have been followed.

4.1 Rounding the fair shares

Starting from the fair share benchmark, a solution would be to 'round' its elements to an adjacent integer while keeping the same row and column totals, but not necessarily using an identical pre-specified rounding method for each element. The fact that this is indeed possible is not obvious. It is due to the special structure of the linear system described by (5), which makes all extreme points integer-valued when the $\mathbf{r}$ and the $\mathbf{c}$ are integer-valued.[19]

A simple procedure due to Cox (1987) and similar to the idea underlying Birkhoff theorem provides a good proof of why this is true: as long as the matrix $\mathbf{f}$ does not consist only in integers and because of the integer-valued row and column sums, each row or column that contains a non-integer element actually contains at least two of them. In this case, it is therefore possible to find an alternating row-column cycle of fractional elements. All the elements can thus be successively rounded without modifying the totals as follows: pick the fractional element (say e_0) that is the closest to an integer, select an alternating row-column cycle of non-integers starting (and finishing) at e_0, round e_0 to its nearest integer (denoted $R(e_0)$) and distribute the resulting modification $|e_0 - R(e_0)|$ by alternately adding it to or substracting it from the successive elements of the cycle, starting from e_0. However, this procedure turns out not to be an appropriate basis for an electoral rule because of its being indeterminate: the final output depends on the paths that are chosen and the way to select them is not well-defined.

A well-defined procedure, developed by Cox and Ernst (1982) and known as *controlled rounding*, produces a unique result almost surely. The basic idea is to minimize the total error due to rounding each element of the benchmark and each row or column sum. More precisely, starting from a matrix $\mathbf{f}$, the idea of controlled rounding is to produce a matrix $R(\mathbf{f})$ of integers such that (1) each non-integer element of the initial matrix is rounded to an adjacent integer; (2) row and column totals are kept constant (3) any integer element of the initial matrix $\mathbf{f}$ is left unchanged.[20]

Here is the controlled rounding of our fair share matrix:

[19] Such a result, is often referred to as Birkhoff theorem. A well-known example is the matching or assignment game where $n = m$ and all the components of $\mathbf{r}$ and $\mathbf{c}$ are equal to 1.

[20] The procedure applies to more general situations in which the initial row and column sums are not integer-valued, and must as well be rounded to an adjacent integer during the process.

	CVP	FDF-RW	PVV	PRL	PS	SP	VU	UDRT	PSC	Ecolo
Br.	3	3	2	2	2	1	1	1	1	1
Le.	1	0	1	0	0	1	1	0	0	0
Ni.	0	1	0	1	1	0	0	0	0	0

Alternative rounding procedures for producing a bi-apportionment could be contemplated by considering alternative cost of errors.[21]

By definition, a rounding method 'respects the fair shares' in the sense that each element is rounded to an adjacent integer. Finally, it should be noted that the bi-apportionment methods based on rounding the fair shares are not a mere bi-dimensional extension of the uni-dimensional methods rounding the quotas (such as the largest remainders). As already said, since the party-apportionment is fixed, there is less flexibility than in the general uni-dimensional problem. Besides, the drawbacks and (monotonicity) paradoxes suffered by the uni-dimensional quota methods are avoided at the global level by these rounding methods when, in the first step, divisor methods are used to define the (integer) number of seats each party is globally entitled to.

4.2 Simultaneously Adjusting and Rounding: Bi-divisor methods

Alternatively to these controlled rounding procedures, one can use two-dimensional divisor methods to derive a bi-apportionment matrix. Bi-divisor methods were introduced in Balinski and Demange (1989-b and 1989-a) who define and provide an axiomatization for problems in which constraints on row sums and column sums are given and can be inequalities, and propose an algorithm for computing them.

The idea is to use divisors as in the uni-dimensional case, but now there is one for each constraint.[22] The adjustment through the multipliers and the rounding are 'simultaneous' so that the bi-apportionment is not necessarily a rounding of the fair share. Formally, just as in the one-dimensional case, bi-dimensional divisor methods are based on a d-rounding function. I assume that $d(0)$ is positive, so that a party may end up with no seat in a district.[23] One looks for multipliers λ_i for district i, μ_j for party j such that d-rounding each element of the matrix with general element $\lambda_i v_{ij} \mu_j$ is a bi-apportionment, namely the constraints on row- and column-totals are met. Formally

Definition 6 *A* bi-divisor method *based on d assigns to any positive problem* $(\mathbf{v}, \mathbf{h}, \mathbf{s})$ *the bi-*

[21] Also Gassner (1991) describes two procedures and uses them to compute the seat apportionment in the Belgian Senate. However, these procedures are rather complex and it seems difficult to know which properties they enjoy.

[22] The uni-dimensional setting considered in Section 2 has only one overall constraint corresponding to the total number of the seats. When there are additional constraints, say a minimum number of seats per districts, additional divisors are introduced.

[23] When $d(0) = 0$, each party is awarded at least one seat in each district (see footnote 10). This may favor too much small parties, which, in turn, raises existence issues, especially when some districts are allocated a low number of seats.

apportionment(s) $\mathbf{b} = \big(b_{ij}\big)$ *that satisfy*

$$b_{ij} = [\lambda_i v_{ij} \mu_j]_d \textit{ for } \lambda_i > 0,\ \mu_j > 0 \textit{ such that}$$
$$\sum_j b_{ij} = h_i \textit{ for each } i \textit{ and } \sum_i b_{ij} = s_j \textit{ for each } j.$$

For a matrix $\mathbf{v}$ which has all its elements positive, such a bi-apportionment exists. Furthermore it is typically unique.

A bi-divisor method satisfies exactness, monotony, consistency, proportionality. Furthermore if $\mathbf{s}$ is obtained by a proportional apportionment method, the bi-divisor method satisfies party-proportionality.

A method can easily be adapted to the situation where a minimal vote total is required for a party to be represented. It suffices to assign zero to the party votes and to delete the corresponding column in the second step. Also observe that one may very well use the Jefferson method for apportioning the seats to parties, say to favor large parties, and use another rounding to obtain the full bi-apportionment (thanks to the two-step procedure).

As in the uni-dimensional case, several bi-divisor methods are obtained, each one associated to a different rounding. However, the same argument applies as in controlled rounding: since both district and party apportionments are predetermined, it is difficult to predict how each method influences the final outcome and furthermore the differences between them should be quite rare (but this needs some more serious appraisal). In a problem with fixed constraints, the standard rounding sounds like the more appropriate choice.

Algorithms have been designed and implemented to find the bi-apportionment and the multipliers.[24] Balinski and Demange (1989-b) propose a Tie-and-Transfer (TT) algorithm. The basic idea is to translate the matrix $\mathbf{v}$ into a bipartite weighted graph G and to operate seat *transfers* from overrepresented rows (or columns) to underrepresented rows (or columns) and updating the multipliers so as to create *ties* that allow for new transfers because some numbers can be rounded both up or down.

It should be noted that the natural idea of the iterative scaling procedure does not work (but may be useful at the beginning of the procedure before TT). The scaling procedure operates on the rounded values: Start with the matrix $\mathbf{v}$, the first iteration consists in eliciting divisors $\lambda^{\mathbf{1}} = (\lambda_i^1)$ such that. the constraints on row sums are met by the rounded elements: $\sum_j [\lambda_i^1 v_{ij}]_d = h_i$ for all i. Iterating, the algorithm produces matrices of the form $[\lambda_i^t v_{ij} \mu^t]_d$ that satisfy the constraints on row sums and $[\lambda_i^t v_{ij} \mu^{t+1}]_d$ that satisfy the constraints on row sums. However, contrary to the linear case, the procedure may not converge (see Maier, Zachariassen and Zachariassen 2010 for a detailed analysis, and how to combine the two algorithms to fasten the computation).

[24]'BAZI', A Free Computer Program for Proportional Representation provides useful programs at http://www.math.uni-augsburg.de/stochastik/bazi/welcome.html.

As for our example, here is the apportionment obtained by a bi-divisor method with standard rounding (computed with BAZI):

	CVP	FDF-RW	PVV	PRL	PS	SP	VU	UDRT	PSC	Ecolo
Br.	3	3	2	2	2	1	1	1	1	1
Le.	1	0	1	0	0	1	1	0	0	0
Ni.	0	1	0	1	1	0	0	0	0	0

It coincides with the one obtained with controlled rounding.

A direct application of the method is found in the Swiss canton of Zurich: the New Zurich Apportionment Procedure, which has been first used in 2004, has been developed by Pukelsheim (2006) based on a bi-proportional divisor method with standard rounding.[25]

4.3 Discussion and concluding remarks

I discuss first a positive aspect of the methods introduced, and then some difficulties pertaining to the presence of local parties, the complexity of a bi-apportionment method, and political agreement.

A key feature of the two-step methods, which first step consists in an apportionment of the seats to parties on the grounds of their overall vote totals, is to offer a solution to the problem of 'lost ballots' that was highlighted in 2002 by the Swiss Federal Court: in small districts, the number of seats is too small to allow a *proportional* representation of the parties. In fact, if a party does not get any seat in some small district, a voter of this district who supported that party may very well complain that his/her vote was lost and that he/she was not treated on an equal footing with a supporter of the same party but in a large district where the number of seats at stake enables the representation of the party. The first step, which regards the votes at the global (i.e. national or cantonal in the Swiss case) level, responds to this inequality problem since every voter's ballot counts equally in the overall distribution of seats to the parties.

When some parties are local, an electoral matrix has some null elements (as is the case in our running example). The analysis extends provided some conditions are fulfilled that I describe for a bi-divisor method. We want to assign zero seat to a party in a district where the party is not present, i.e. receives no vote. This requires the existence of an apportionment matrix which has a zero element at each entry where $\mathbf{v}$ has. It turns out that the existence of such a matrix is also sufficient for a well-defined apportionment by a divisor method (for a d-rounding function that has $d(0) > 0$ as assumed before, which guarantees zero seat to zero vote). Specifically, the bi-divisor method is defined for matrices $\mathbf{v}$ for which there is a matrix $\mathbf{b}$ integer-valued that satisfies

$$\sum_j b_{ij} = h_i \text{ for each } i, \sum_i b_{ij} = s_j \text{ for each } j, \text{ and } b_{ij} = 0 \text{ if } v_{ij} = 0.$$

[25]The method slightly differs because each voter can cast as many ballots as there are seats at stake in his/her district, so an adjustment of the votes is needed.

From well-known results on transportation problems, the existence of such a matrix is equivalent to the following conditions:

$$\sum_{j \in I} s_j \leq \sum_{i \in I_j} h_i, \text{ where } I_j = \{i | \text{ there exists } j \in J : v_{ij} > 0\}. \quad (6)$$

That is, any subgroup of parties should not be assigned more seats than the total number of seats in the districts in which the parties are qualified.[26] As is intuitive, the conditions are more difficult to be fulfilled, the more zeros there are in the electoral matrix. Applying (6) to a 'local' party, present in a single district, the condition requires that the number of seats in that district is larger than the total number of seats assigned to the party. This is quite a mild condition (if the distortions in the district or party allocations are not too large). However, even if local parties can be handled with, the bi-allocation methods are a priori designed for situations where parties are present in almost all districts. Otherwise, there is not much sense in trying to link the results in the various districts. For this reason, the European Parliament example alluded to in the introduction is currently not appropriate to bi-allocation methods, though the European parties exist and are somewhat encouraged. In a perspective where European parties should play a more prominent role, but with elections still at a member state level, bi-allocation methods could be useful.

One may worry about the complexity of the methods for voters. There are at least two distinct aspects, one linked with the understanding of a method and its logic and the other one with the possibility of checking the outcome.

Understanding the methods requires some attention. All methods rely on fair shares —and the associated multipliers— and on some rounding techniques. (The diversity in rounding techniques is already present in the uni-dimensional problem but most voters seem not to care about it.) There is a chance that a substantial fraction of voters might not want to make the effort of understanding multipliers. On this aspect, the successful implementation of the new Zurich apportionment method is encouraging. As for the rounding techniques, the controlled rounding method seems more intuitive than a bi-divisor method. Given a 'target', just try to round up or down the data so as to satisfy the row and column constraints. A difficulty with a bi-divisor method is that both the multipliers and the rounding must be adjusted simultaneously.

Checking the outcome following an election may be another difficulty. Whatever method that is truly bi-dimensional involves handling matrices and the rounding adds another difficulty in computing the outcomes. As such, both types of methods, although 'transparent' in theory, may appear quite obscure to many voters or politicians, who cannot compute the outcome without an adequate software. There is nevertheless a difference between the controlled rounding and the bi-divisor

[26] For the existence of a fair share, which serves as a basis to a rounding method, the existence of a fair share requires a slight strengthening of the above conditions. When these conditions are fulfilled the rounding method of the fair share works as described. Since the null elements are kept null (as they are integer hence kept unchanged), the method delivers an apportionment that assigns zero seat to zero vote.

methods. For the latter, once the multipliers are made public, the outcome can be checked by hand. For the controlled rounding, this is not the case since the optimal rounding solves an optimization program.

Another difficulty in the adoption of a bi-apportionment method is more political and bears about the conveyed information. The multipliers need to be made public in order to justify the outcome. Even if multipliers are not perfectly understood, they uncover the patent distortions in the district allocations (here distortions have not necessarily a negative connotation). Although such distortions are known, making them repeatedly visible after each election should trigger some reaction and a demand for justification. For the parties, differences in multipliers should not generate the same difficulty since their values are determined so as to respect the party-apportionment, which is changed at each election as a function of the voting results and according to a well-defined rule. The main difficulty for parties is the possibility of reversals. Under a reversal, a candidate can be eliminated by one of an opponent party although he/she received more votes only to respect the party-apportionment. If parties are powerful enough and can impose a discipline on their candidates, this should not be a major issue. After all, a basic premise underlying party's proportionality is that indeed parties are major actors.

Finally, the analysis has taken the district-apportionment as 'given'. The party-apportionment is determined, say to achieve proportional party-representation (modulo the variations in the proportional methods) or another goal determined by the aggregate parties' results. The interaction between the district- and party-apportionments has not been tackled, as far as I know. This issue deserves some discussion. The main theoretical argument against proportionality in the district-apportionment is that the representatives of a district do not fully represent their citizens because representatives vote in 'block', as argued by Penrose (1946). This argument is no longer valid when citizens vote for parties, hence express their preferences related to 'general issues' handled by parties. Citizens are represented not only by their district representatives but also by their parties' representatives. This suggests that the rationale for favoring districts with small population sizes is much weaker. The answer should depend on the type of issues handled by the 'representative' assembly, but in any case is worth studying.

References

Bacharach M. (1965). 'Estimating Non-Negative Matrices for Marginal Data'. *International Economic Review* 6(3): 294–310

Balinski M.L. and G. Demange (1989-a). 'An Axiomatic Approach to Proportionality between Matrices'. *Mathematics of Operations Research* 14: 700–719.

Balinski M.L. and G. Demange (1989-b). 'Algorithm for Proportional Matrices in Reals and Integers'. *Mathematical Programming* 45: 193–210.

Balinski, M.L and H. P. Young (1982). *Fair Representation: Meeting the Ideal of One Man, One Vote* (1st ed.). Yale Univ Press.

Brams S. (2008). 'Mathematics and democracy: Designing better voting and fair-division procedures'. *Mathematical and Computer Modelling* 48(9–10): 1666–1670.

Cox L.H. and L.R. Ernst (1982). 'Controlled Rounding'. *INFOR* 20: 423–432.

Cox L.H. (1987). 'A Constructive Procedure for Unbiased Controlled Rounding'. *Journal of the American Statistic Association* 82(398): 520–524.

Gallagher M. (1992). ' Comparing proportional representation electoral systems: Quotas, thresholds, paradoxes and majorities' *British Journal of Political Science* éé: 469-496.

Gassner M. (1991). 'Biproportional Delegations: A Solution for Two-Dimensional Proportional Representation'. *Journal of Theoretical Politics* 3(3):321–342.

Grimmett G. (2004). 'Stochastic Apportionment'. *The American Mathematical Monthly* 111(4): 299–307.

Grimmett G. (2011). 'European Apportionment via the Cambridge Compromise'. Available at *http://www.statslab.cam.ac.uk/g̃rg/papers/USep3.pdf.*

Huntington E.V. (1921). 'The Mathematical Theory of Apportionment of Representatives'. *Proceedings of the National Academy of Science of the USA* 7(4): 123–127.

Hylland A. (2000). 'Proportionality in two dimensions — Applied to the representation of Cantons and Constituent Peoples in the House of Peoples of the Parliament of the Federation of Bosnia and Herzegovina'. Available at *http://www.sv.uio.no/econ/personer/vit/aanundh/upubliserte-artikler-og-notater/Prop-2d%5B1%5D.pdf.*

Maier S., P. Zachariassen and M. Zachariassen (2010). 'Divisor-Based Biproportional Apportionment in Electoral Systems: A Real-Life Benchmark Study'. *Management Science* 56(2): 373–387.

Pennisi A. (2006). 'The Italian Bug: A Flawed Procedure for Bi-Proportional Seat Allocation'. In B. Simeone and F. Pukelsheim (eds), *Mathematics and democracy: recent advances in voting systems and collective choice* (151–165). Springer.

Penrose L.S. (1946). 'The elementary statistics of majority voting'. *Journal of the Royal Statistical Society* 109: 53–57.

Pukelsheim, F. (2006). 'Current issues of apportionment methods' In B. Simeone and F. Pukelsheim (eds) *op. cit.*. on Aug. 25th, 2011).

Widgren M. (2005). 'Power in the Design of Constitutional Rules'. In G. Demange and M. Wooders (eds) *Group Formation in Economics: Networks, Clubs, and Coalitions* (312–334). Cambridge University Press.