A Note on the Direct Democracy Deficit in Two-tier Voting[†]

Nicola Maaser[*]
and
Stefan Napel[*,‡]

July 6, 2011

Abstract

A large population of voters with single-peaked preferences are partitioned into disjoint constituencies. Collective decisions are taken by their representatives, one from each constituency, according to a weighted voting rule. It is assumed that each representative's ideal point perfectly matches that of the respective *constituency median* and that top-tier decisions are in the voting game's *core*. The resulting representative-democratic voting outcomes generally differ from those of a direct-democratic, single-tier system. The expected discrepancy varies with the voting weight allocation. It is minimized by weights proportional to constituency population sizes only if citizens differ sufficiently more between than within constituencies. Weights proportional to the *square root of population sizes* perform better if all citizens have independent and identically distributed ideal points.

Keywords: weighted voting, two-tier voting systems, square root rules, representative democracy, majoritarianism

1 Introduction

Democratic government of large political units such as modern nation states and supranational entities involves the use of political representatives who make decisions on behalf of the citizens. As democratic principles are being extended from city states to nation states and ever larger units – dubbed the "second democratic transformation" by Dahl (1994) – the question of whether representatives take the right decisions from the point of view of their citizens has been gaining importance.

[†]We thank Helmut Rieder for very helpful conversation and are also grateful for constructive discussion at the Leverhulme Trust's *Voting Power in Practice Symposium* 2011.

[*]Dept. of Economics, University of Bayreuth, Universitätsstr. 30, 95447 Bayreuth, Germany, and Public Choice Research Centre, Turku, Finland; phone: +49-921-552880; e-mail: nicola.maaser@uni-bayreuth.de, stefan.napel@uni-bayreuth.de

[‡]Corresponding author.

Discrepancies between the legislative outcome of a representative system and the policy preferences of citizens can be accounted for from at least two different perspectives: First, various frictions in political markets may leave citizens unable to effectively constrain the behavior of elected politicians. Potentially, these agency problems can be alleviated by more direct participation of citizens in the legislation process. Popular referenda and other direct-democratic institutions are increasingly recommended as a complement or corrective to existing representative systems (see, e.g., Frey and Stutzer 2006 or Kirchgässner and Feld 2004).

Second, representatives are in many democracies elected in disjoint constituencies, and then participate in a governing body, council, or top-tier assembly where each member casts a block vote for his or her entire constituency. We refer to such an arrangement as a *two-tier voting system*. In this case, the agreed policy will often deviate from citizen preferences even if no political market imperfections exist. In a frictionless median voter world, each representative will fully comply with the preferences of his district's citizens in the sense that he will adopt the policy position preferred by the *constituency's median voter* when acting in the top-tier assembly. However, the compromise reached by these ideal representatives need not coincide with the outcome preferred by the overall *population's median voter*. If a weighted voting rule is used in the assembly, it singles out an issue-specific *pivotal representative*, whose most-preferred policy (i.e., ideally that of his district's median voter) is implemented. This is generically distinct from the population median unless ideal points are highly clustered.

In principle, if the goal is to implement the preferences of the population median, then one could simply have an assembly of *all* citizens take the relevant policy decisions.[1] Yet, at least at larger scales, representative democracy offers significant advantages. It relieves citizens from the burden of acquiring information on every issue and avoids potentially high other costs of involving the full population. Another advantage are better negotiation possibilities in small bodies of representatives: political bargaining may bring about Pareto-superior solutions which, due to transaction costs, could most probably not be reached when decisions are taken at the level of a large-scale citizenry (see Baurmann and Kliemt 1993).[2]

The aim of this paper is to study the links between the allocation of block voting rights, i.e., the *voting weights* of constituency representatives, when top-tier decisions are taken according to simple majority rule by one representative each from every constituency, and the congruence of the outcomes produced by this two-tier decision process and by direct democracy. We study the case in which all policy alternatives are elements of a one-dimensional policy space and individual voter preferences are single-peaked. The expected distance between the legislative outcomes of, first, indirect two-tier decision-making and, second, an ideal direct democracy provides a measure of the *direct democracy*

[1] In fact, decision-making by an assembly of all citizens still persists in two Swiss cantons and a number of Swiss and US municipalities.

[2] While trade between representatives reduces the transaction costs of political decisions, it also facilitates the exchange of votes via log-rolling arrangements that result in pork-barrel politics (see Weingast, Shepsle, and Johnsen 1981).

deficit which is implied by a particular allocation of voting weights. We seek to find the deficit-minimizing weight allocation rule when constituencies are – e.g., for geographical, ethnic, or historical reasons – of different sizes.

In this setting, with single-peaked preferences over a one-dimensional policy space, the direct-democratic outcome can easily be identified with the ideal point of the median individual. Finding the outcome of decision-making in the two-tier system is slightly more involved: first, the policy advocated by the representative of any given constituency is supposed to coincide with the ideal point of the respective constituency's median voter (also the unique element of the constituency's *core*). Second, the decision which is taken at the top tier is identified with the position of the *pivotal representative* (corresponding to the assembly's core), where pivotality is determined by the vector of voting weights and a 50% decision quota. Consideration of the respective core is meant to capture the result of possible strategic interaction.[3] As long as this is a reasonable approximation, the actual systems determining collective choices can stay unspecified. They could differ across constituencies.

Because the population size of a constituency typically affects the distribution of its median, the location of the top-tier decision in the policy space becomes a rather involved function of (the order statistics of) differently distributed random variables. A straightforward analytical investigation of the model is therefore possible only for degenerate cases. For this reason, we resort to Monte-Carlo approximations of the expected distance between the outcomes of representative and direct-democratic decision making, considering randomly generated, artificial population configurations as well as recent EU population data.

The main finding of our analysis is that the direct democracy deficit is minimized (in the class of simple power laws) by the use of a *square root rule*, i.e., top-tier weights proportional to the square root of a constituency's population size, if the ideal points of all citizens are *independent and identically distributed* (i.i.d.). If, in contrast, preferences are dependent and have high *positive correlation* within constituencies, then weights should be proportional to population sizes.

The design of weighted voting rules has already received considerable attention in the literature on indirect or representative democracy. Several – potentially conflicting – normative criteria have been applied to the problem of defining the weights of representatives from differently sized constituencies. For instance, the design of voting rules which minimizes the deviation of two-tier decision-making from direct democracy under simple majority rule and votes on *binary alternatives* has been studied by Felsenthal and Machover (1999). Their objective has been minimization of the so-called "mean majority deficit".[4] The latter arises whenever the alternative chosen by the body of representatives is supported only by a minority of all citizens, and can be measured as the difference between the size of the popular majority camp and the number of citizens in favor of the assembly's

[3] See, e.g., Cho and Duggan (2009).

[4] As demonstrated by Felsenthal and Machover (1999), this is equivalent to maximizing the sum of citizens' indirect voting power measured by the Penrose-Banzhaf measure.

decision. Felsenthal and Machover have shown that the mean majority deficit is minimal under a square root allocation of weights.[5] Results on a related notion of majoritarian deficit have been obtained by Kirsch (2007). Feix et al. (2008) also consider majority votes on two alternatives, and seek to minimize the *probability* of situations where the decision taken by the representatives is at odds with the decision that citizens would have adopted in a referendum.[6]

This paper investigates the difference between direct and representative democratic outcomes for *non-binary* choices from a one-dimensional convex policy space. Our model and findings can broadly be viewed as a generalization of the mentioned literature to the context of many finely graded policy alternatives and with strategic interaction captured by the median voter theorem. As in the case of binary alternatives, the degrees of similarity within and across constituencies are critical for whether a linear or a square root weight allocation rule performs better.

While we mainly emphasize the gap between direct vs. representative democracy, our work also makes a contribution to the optimal design of voting rules from a welfare perspective. In the spatial voting model which we consider, minimization of the direct democracy deficit is equivalent to maximization of expected *total utility* of the citizens provided that their preferences over policy outcomes (i) all have the same intensity, and (ii) are representable by a cardinal utility function that decreases linearly in the distance to the respective ideal policy.[7] A utilitarian ideal of maximizing welfare in two-tier voting systems has been studied for binary decision making, e.g., by Barberà and Jackson (2006), Beisbart and Bovens (2007), and Koriyama and Laslier (2011).

Note that an important alternative approach to assessing voting rules from a normative constitutional perspective concerns the indirect *influence* or *power* of citizens. Analytical investigations of the objective to implement the "one person, one vote" principle in two-tier voting systems with a focus on influence date back to the seminal work of Penrose (1946). He identified a square root rule (see fn. 5) as the solution to the problem in binary settings (see Felsenthal and Machover 1998, Sect. 3.4, and also Kaniovski 2008). Maaser and Napel (2007) and Kurz et al. (2011) have extended the analysis to unidimensional spatial voting.

2 Model

Consider the partition $\mathfrak{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_r\}$ of a large voter population into r *constituencies* with $n_j = |\mathcal{C}_j| > 0$ members each. Let $n \equiv \sum_j n_j$ and all n_j be odd numbers for simplicity. The preferences of any voter $i \in \{1, \ldots, n\} = \bigcup_j \mathcal{C}_j$ are assumed to be single-peaked with

[5] They refer to this allocation rule as the *second square root rule* in order to distinguish it from Penrose's (1946) (first) square root rule, which requires representatives' *voting power* – rather than their weight – to be proportional to the square roots of their constituencies' population sizes.

[6] The situation is known in the social choice literature as an instance of the *referendum paradox* (see, e.g., Nurmi 1998 and Lepelley et al. 2011).

[7] The policy corresponding to the median citizen's ideal point maximizes overall welfare under these assumptions (see, e.g., Schwertman et al. 1990). If utility decreases quadratically in distance, the ideal point of the *mean* voter maximizes overall welfare.

ideal point ν^i in a convex one-dimensional *policy space* $X \subset \mathbb{R}$. These ideal points are conceived of as realizations of random variables with a known a priori distribution; any given profile $(\nu^1, \ldots, \nu^n)$ of ideal points is interpreted as reflecting voter preferences on a specific one-dimensional policy issue (a tax level, an emission standard, expenditure on a public good, etc.).

For any realization of voter preferences, let

$$\nu^* = \text{median}\{\nu^1, \ldots, \nu^n\}$$

denote the median ideal point in the population. Under frictionless simple majority voting, the collective decision taken by a full assembly which comprises all voters $1, \ldots, n$ would coincide with the median voter's ideal point. It beats every alternative policy in a pairwise vote and hence is the unique element of the corresponding spatial voting game's *core*. We will refer to ν^* as the *direct democracy benchmark* for the following representative-democratic, two-tier voting process.

Namely, a collective decision $x \in X$ on the issue at hand is actually taken by an assembly or council of representatives $\mathcal{R}$ which consists of one representative from each constituency. Without going into details, we assume that the preferences of $\mathcal{C}_j$'s representative are congruent with its median voter, i.e., representative j has ideal point

$$\lambda_j = \text{median}\{\nu^i \colon i \in \mathcal{C}_j\}.$$

In theory, elected representatives are fully responsive to their constituency's voters, and in particular the median voter. In practice, of course, representatives tend to develop preferences (e.g., concerning their privileges) that differ from those of regular citizens.[8]

In the top-tier assembly $\mathcal{R}$, each constituency $\mathcal{C}_j$ has *voting weight* $w_j \geq 0$. Any subset $S \subseteq \{1, \ldots, r\}$ of representatives which achieves a combined weight $\sum_{j \in S} w_j$ above $q \equiv 0.5 \sum_{j=1}^{r} w_j$, i.e., comprises a *simple majority* of total weight, can implement a policy $x \in X$. So collective decisions are taken according to the weighted voting rule $[q; w_1, \ldots, w_r]$.

Let $\lambda_{k:r}$ denote the k-th leftmost ideal point amongst the representatives (i. e., the k-th order statistic of $\lambda_1, \ldots, \lambda_r$) and consider the random variable P defined by

$$P = \min\Big\{ l \in \{1, \ldots, r\} \colon \sum_{k=1}^{l} w_{k:r} > q \Big\}.$$

Representative $P : r$'s ideal point, $\lambda_{P:r}$, is the unique policy that beats any alternative $x \in X$ in a pairwise majority vote, i. e., it constitutes the core of the voting game in $\mathcal{R}$ with weights $w_1, \ldots, w_r$ and quota q. Without any formal analysis of decision procedures that might be applied in $\mathcal{R}$ (see Banks and Duggan 2000, or Cho and Duggan 2009), we assume that the policy agreed in the council coincides with the ideal point of *pivotal*

[8]See Gerber and Lewis (2004) for empirical evidence. Another problem arises when systematic abstention of certain social groups drives a wedge between the median voter's and the median citizen's preferences.

representative $P : r$. In summary, the policy outcome produced by the two-tiered voting system is

$$x_{\mathcal{R}} = \lambda_{P:r}\,.$$

Usually, some incongruence between the collective decision $x_{\mathcal{R}}$ taken by the representatives and the direct democracy benchmark ν^* is unavoidable. It can be viewed as a price that needs to be paid for the several advantages of representative democracy. Still, large differences

$$\Delta = |\nu^* - x_{\mathcal{R}}|$$

between the two-tier outcome and the direct democracy benchmark may be regarded as undesirable. Democracies create representative systems primarily for efficiency reasons – not in order to purposely implement policies that the popular majority would prefer to be replaced.

A system which produces small average gaps between ν^* and $x_{\mathcal{R}}$ approximates direct democracy better than one which yields big gaps. This raises the question: *Which voting weight allocation rule minimizes the expected difference between the policy outcomes of an indirect, two-tier voting system and a direct-democratic system?* We will call Δ's expected value, $\mathbf{E}[\Delta]$, the *direct democracy deficit* of the weighted voting rule $[q; w_1, \ldots, w_r]$ employed in $\mathcal{R}$ (taking partition $\mathfrak{C}$ as given).[9] By an "allocation rule" we mean a simple mapping W which assigns weights $(w_1, \ldots, w_r) = W(\mathcal{C}_1, \ldots, \mathcal{C}_r)$ to any given partition of a large population. The rule shall single out weights $w_1, \ldots, w_r$ which approximate a solution to the problem

$$\min_{w_1', \ldots, w_r'} \mathbf{E}[\Delta]\,, \tag{1}$$

and our criterion for acceptably "simple" mappings $W : \mathfrak{C} \mapsto (w_1, \ldots, w_r)$ will be that they are *power laws*, i.e., $w_j = n_j^{\alpha}$ for some constant $\alpha \in [0, 1]$. This class of mappings nests the square root and linear rules which have played prominent roles in the previous literature.

3 Analysis

The population size of a constituency generally affects the distribution of its median. Specifically, if the ideal points of voters $i \in \mathcal{C}_j$ are pairwise independent and come from an arbitrary identical distribution F with positive density f on X, then its median position λ_j asymptotically has a normal distribution with mean $\mu = F^{-1}(0.5)$ and standard deviation

$$\sigma_j = \frac{1}{2\, f(\mu) \sqrt{n_j}} \tag{2}$$

(see, e.g., Arnold et al. 1992, p. 223). The variance of the position of $\mathcal{C}_j$'s representative is the smaller, the greater the population size n_j.

[9] Recall that $q = 0.5 \sum_j w_j$. We conjecture that, in general, a greater direct democracy deficit is implied by supermajority requirements than by simple majority, but leave a detailed investigation for future research.

Even in the trivial case of $w_1 = \ldots = w_r$, the top-tier decision $x_{\mathcal{R}} \in X$ has a rather non-trivial distribution when constituency sizes differ because it is an order statistic of *differently* distributed random variables. For non-identical weights $w_1, \ldots, w_r$, it is a combinatorial function of such order statistics. The analytical options to finding or approximating a solution to (1) are, therefore, extremely limited. We will consider only two somewhat degenerate but instructive special cases in order to develop some intuition, comment on the optimal statistical aggregation of $\lambda_1, \ldots, \lambda_r$, and then turn to simulations in Section 4.

The first trivial case is to only have constituencies with equal population sizes. The optimal constant α is then undetermined. Problem (1) is indeed solved by any $w'_1 = \ldots = w'_r > 0$ if constituencies are perfectly symmetric. However, even $n_1 = \ldots = n_r = n/r$ and i.i.d. ideal points in general do not allow for $\mathbf{E}[\Delta] = 0$. Under optimal, i.e., identical weights, the pivotal representative's ideal point is the (unweighted) median of $\lambda_1, \ldots, \lambda_r$. So $(r+1)/2$ constituency median positions are located weakly to the left (right) of $x_{\mathcal{R}}$, which guarantees that at least $(r+1)/2 \cdot (n/r+1)/2$ voter ideal points are located weakly to $x_{\mathcal{R}}$'s left and right, respectively. This allows for realizations of $\nu^1, \ldots, \nu^n$ where the population median ν^* is located up to

$$\left| \frac{(r+1)(n+r)}{4r} - \frac{n+1}{2} \right| = \frac{(n-r)(r-1)}{4r} < \frac{n}{r} \cdot \frac{r-1}{4}$$

ideal points away from $x_{\mathcal{R}}$.

We have so far left unspecified how voter ideal points are jointly distributed. For the kind of constitutional design problem that we are dealing with, it is a common assumption that all citizens should be considered *identical* a priori, i.e., every ideal point ν^i is drawn from the same marginal probability distribution F. This moves the analysis behind a constitutional "veil of ignorance", which ignores knowledge about specific preferences for normative reasons. Such a constitutional a priori perspective does not necessarily entail that preferences of citizens must also be considered as *independent* of each other. An i.i.d. assumption for all ideal points ν^i with $i \in \bigcup_j \mathcal{C}_j$, i.e., the product distribution F^n, is a particularly compelling benchmark. However, the partition $\mathfrak{C}$ may have reasons that need to be acknowledged behind the "veil of ignorance" (e.g., geographic barriers, ethnics, language, religion). These reasons are likely to involve or give rise to closer political connections between voters within constituencies than across them.[10]

An extreme instance of "close connections" within constituencies is the one in which $\nu^i = \nu^h$ whenever $i, h \in \mathcal{C}_j$. This takes the idea that citizens are on average more similar to each other within constituencies than across constituencies to its limit. In this degenerate situation, the direct democracy deficit is minimized by using the linear rule $w_j = n_j$ for $j = 1, \ldots, r$, i.e., by $\alpha^* = 1$. In particular, the directly proportional weight allocation implies $\mathbf{E}[\Delta] = 0$ in this case. This follows from noting that the ordered ideal points of all citizens $i = 1, \ldots, n$,

$$\nu^{1:n} \leq \nu^{2:n} \leq \nu^{3:n} \leq \ldots \leq \nu^{n-1:n} \leq \nu^{n:n},$$

[10]If there were no special connection between the respective citizens within differently sized constituencies $\mathcal{C}_j$ and $\mathcal{C}_k$, then the obvious solution, even behind a veil of ignorance, would be to redistrict partition $\mathfrak{C}$ to a partition $\mathfrak{C}'$ such that $n'_j = n'_k$ for any $j, k \in \{1, \ldots, r\}$.

can be written as

$$\underbrace{\lambda_{1:r} = \ldots = \lambda_{1:r}}_{n_{1:r} \text{ times}} \leq \underbrace{\lambda_{2:r} = \ldots = \lambda_{2:r}}_{n_{2:r} \text{ times}} \leq \ldots \leq \underbrace{\lambda_{r:r} = \ldots = \lambda_{r:r}}_{n_{r:r} \text{ times}}$$

under perfect correlation within constituencies. So representative j is pivotal in $\mathcal{R}$ iff its policy position (and thus that of all $\mathcal{C}_j$-citizens) is also the population median. This optimality of weights proportional to population sizes should be expected to apply *approximately* when voter ideal points are not perfectly, but highly correlated within constituencies (and independent across). The simulations reported in Section 4 indeed confirm this intuition: the direct democracy deficit is minimized by an essentially linearly proportional rule provided that the ideal points of the citizens vary noticeably more across than within constituencies.

An intuition for the case in which all individual ideal points are i.i.d. is much harder to come by. As already mentioned, λ_j has an asymptotically normal distribution with the common mean $\mu = F^{-1}(0.5)$ and a standard deviation σ_j which is inversely proportional to the square root of n_j. The optimal voting weights somehow have to strike a balance between accounting for a large constituency median's on average greater centrality and informational value (suggesting high α is good), and guarding against the possibility that a large constituency's representative can implement its preferred policy even if that constituency happens to be an outlier (warning against high α). Which α strikes the right balance is unclear at the outset.

An admittedly crude intuitive argument in favor of $\alpha = 0.5$ is as follows: if all voter ideal points ν^i are i.i.d. then each individual $i = 1, \ldots, n$ a priori has probability $1/n$ to be the population median; hence the population median is found in constituency $\mathcal{C}_j$ with probability n_j/n. Weights which give all constituencies a top-tier pivot probability proportional to their population size, therefore, are a reasonable starting point. Proportionality between the probability π_j for event $\{j = P : r\}$ and n_j is achieved in the i.i.d. case by selecting weights w_j that are proportional to the *square root* of n_j (under some regularity assumptions – see Kurz et al. 2011).[11] Of course, even if the events $\{j = P : r\}$ and $\{\text{some } i \in \mathcal{C}_j \text{ is the population median}\}$ coincide (they need not), the implemented policy will be λ_j rather than ideal point ν^i of the overall median voter i who happens to belong to $\mathcal{C}_j$. The "internal representation error" $|\nu^i - \lambda_j|$ tends to be greater for small constituencies because λ_j's distribution is less concentrated around the theoretical median. However, if one consequently tried to overweight large constituencies $\mathcal{C}_j$, i.e., to give them a probability $\pi_j > n_j/n$, in order to avoid large internal errors, this risks coming at the price of creating additional "external representation error", i.e., smaller probability for $\{j = P : r\}$ and $\{\text{some } i \in \mathcal{C}_j \text{ is the population median}\}$ to coincide. This suggests that the square root starting point may not be much improved upon.

It is worthwhile to note that if we gave up the assumption that the pivotal representative in $\mathcal{R}$ exploits the ability to make a proposal that beats every other one and if, instead, we

[11] Greater centrality of medians from large constituencies gives them a higher baseline probability for being pivotal under simple majority rule. This advantage is proportional to $n_j^{0.5}$, and it is sufficient to combine it with $w_j = n_j^{0.5}$ in order to obtain the desired proportionality of π_j and n_j.

assumed that the representatives always seek a "weighted compromise" which amounts to their average position $\bar{\lambda} = \sum_j \frac{w_j}{w_\Sigma} \cdot \lambda_j$ with $w_\Sigma = \sum_j w_j$, then $\alpha = 1$ would be optimal. In particular, under the assumption that $\nu^1, \ldots, \nu^n$ are i.i.d. random variables with a distribution F that has median $m = F^{-1}(0.5)$ and a positive density f in a neighborhood of m, the sample median ν^* of a specific realization $\nu^1, \ldots, \nu^n$ satisfies (see Rieder 1994, p. 21, Corollary 1.5.4)

$$\nu^* = m + \frac{1}{n} \sum_{i=1}^{n} \psi(\nu_i) + o_{F^n}\big(\frac{1}{\sqrt{n}}\big) \tag{3}$$

with the so-called *influence curve*

$$\psi(\nu_i) = \frac{\text{sign}(\nu_i - m)}{2f(m)}$$

and where $o_{F^n}(\cdot)$ is the stochastic Landau symbol.[12] This means that one can conceive of sample median ν^* as approximately the result of starting at F's theoretical median m and then, for i ranging from 1 to n, doing equidistant jumps of $\frac{1}{2f(m)}$ to either the right or the left depending on whether $\nu^i > m$ or $\nu^i < m$ (staying put in the zero-probability event $\nu^i = m$). Analogously, we have

$$\lambda_j = m + \frac{1}{n_j} \sum_{i \in \mathcal{C}_j} \psi(\nu_i) + o_{F^n}\big(\frac{1}{\sqrt{n_j}}\big) \tag{4}$$

for every $j = 1, \ldots, r$. Equations (3)–(4) then imply that

$$\bar{\lambda} = \sum_j \frac{n_j}{n} \cdot \lambda_j \approx m + \frac{1}{n} \sum_{j=1}^{r} \sum_{i \in \mathcal{C}_j} \psi(\nu_i) = m + \frac{1}{n} \sum_{i=1}^{n} \psi(\nu_i) \approx \nu^*$$

with a random error term that vanishes as $n \to \infty$. So computing their *n_j-weighted mean* would actually be the correct statistical way of aggregating the representative ideal points (i.e., constituency medians). Unfortunately, this does not answer the question of which value of α yields the smallest democracy deficit when the outcome of strategic interaction and voting amongst the representatives is indeed the position of $\mathcal{R}$'s pivotal member, i.e., their *n_j^α-weighted median.*[13]

4 Simulations

Since we are unable to obtain any useful analytical approximation of $\mathbf{E}[\Delta]$ as a function of α, we turn to the *Monte-Carlo approach.* It exploits that – by the law of large numbers –

[12]Namely, $o_{F^n}\big(\frac{1}{\sqrt{n}}\big)/\frac{1}{\sqrt{n}}$ in F^n-probability converges to 0 as $n \to \infty$.

[13]In our simulations, the direct democracy deficit for the best n_j^α-weighted median of the constituency medians is between one and two orders of magnitude greater than for their n_j-weighted mean. This suggests that letting a bureaucrat average the $(\lambda_1, \ldots, \lambda_r)$-information might be preferable to having political bargaining. But note that a weighted mean rule creates incentives for strategic misrepresentation of λ_j. A weighted median rule avoids this.

the empirical average of s independent realizations of $\Delta = |\nu^* - x_{\mathcal{R}}|$ converges to $\mathbf{E}[\Delta]$ as $s \to \infty$.

In order to obtain a realization of Δ for the case of i.i.d. voter ideal points, we draw n (pseudo-)random numbers from a given distribution F.[14] First, the resulting list $\mathbf{v} = (\nu^1, \ldots, \nu^n)$ is sorted in order to obtain a realization of median ν^*. Second, the original list $\mathbf{v}$ is sorted within consecutive blocks of size $n_1, n_2, \ldots, n_r$ in order to obtain the corresponding realizations of the constituency medians $\lambda_1, \lambda_2, \ldots, \lambda_r$. We then infer the weighted median of these, using weights $w_j = n_j^\alpha$ for values of α which range from 0 to 1 in steps of 0.02, and thus obtain $x_{\mathcal{R}}$ for each value of α. The resulting gaps between population median and pivotal representative are recorded, and the procedure is repeated for one million iterations. Finally, we determine the value of α which produced the smallest average gap.[15]

This level of α is determined for a sizeable number of distinct *population configurations* (different numbers r of constituencies and population sizes $n_1, \ldots, n_r$ of these constituencies) in order to be able to draw sufficiently robust conclusions. Most of the considered population configurations are artificial: they are obtained by first fixing r and then selecting sizes $n_1, \ldots, n_r$ by drawing random numbers from a specified distribution – $\mathbf{U}(10^3, 3 \cdot 10^3)$, for instance, presupposes that each constituency size between 1,000 and 3,000 voters is equally likely.

For each "type" of population configuration (e.g., $r = 15$ and $\mathbf{U}(10^3, 3 \cdot 10^3)$) five independent realizations of $n_1, \ldots, n_r$ have been investigated. So Table 1 reports the respective "optimal" value α^* for altogether 90 different configurations. The 95%-confidence interval around the empirical mean of Δ is typically too wide to rule out that one of the $2k$ neighboring values $\alpha^* \pm 0.02k$ produces a smaller direct democracy deficit. However, this only involves $1 \leq k \leq 3$ and, in particular, differences are significant when sufficiently distinct values like $\alpha = 0.5$ and $\alpha = 1$ are compared. The obtained estimates of $\mathbf{E}[\Delta]$ are in almost all cases strictly decreasing functions of α on $[0, \alpha^*)$ and strictly increasing on $(\alpha, 1]$. This and the overwhelming number of instances where $\alpha^* = 0.5$ – regardless of the number r of constituencies and irrespective of whether population sizes are drawn from *uniform*, truncated *normal*, or (most realistically for population sizes) *Pareto* distributions – provides robust evidence that a square root allocation rule is optimal in case of i.i.d. ideal points (in the class of power laws).[16]

---

[14] Since the considered number of voters in each constituency $\mathcal{C}_j$ is large ($n_j \gg 50$), the respective population and constituency medians will approximately have normal distributions irrespective of the specific F which one considers. For the sake of completeness, let it still be mentioned that individual ideal points were drawn from a standard uniform distribution $\mathbf{U}(0, 1)$ in our simulations. The MATLAB source code is available upon e-mail request.

[15] It is a tempting simulation "short-cut" to directly draw realizations of $\lambda_1, \lambda_2, \ldots, \lambda_r$ from $\mathbf{N}(0, \sigma_j^2)$ with variances $\sigma_j^2 = (4\, f(\mu)^2 n_j)^{-1}$, respectively, and to consider a realization from $\mathbf{N}(0, \sigma^2)$ with $\sigma^2 = (4\, f(\mu)^2 n)^{-1}$ as a proxy for ν^*. This avoids the storage and time-consuming sorting of $(\nu^1, \ldots, \nu^n)$, i.e., it is computationally very efficient. However, it ignores the statistical dependence of ν^* and $\lambda_1, \lambda_2, \ldots, \lambda_r$. Experimenting with this short-cut, we found that the dependence of ν^* and $\lambda_1, \lambda_2, \ldots, \lambda_r$ cannot be ignored (at least for $r \leq 35$).

[16] Moreover, the table indicates that the direct democracy deficit tends to decrease in the number r of constituencies when a fixed a priori distribution for n_j is considered. Comparison between the respective

r	$n_1,\dots,n_r$	**(I)**	**(II)**	**(III)**	**(IV)**	**(V)**
15	$\mathbf{U}(10^3, 3\cdot10^3)$	**0.50** ($1.73\cdot10^{-3}$)	**0.50** ($1.59\cdot10^{-3}$)	**0.48** ($1.61\cdot10^{-3}$)	**0.48** ($1.76\cdot10^{-3}$)	**0.50** ($1.70\cdot10^{-3}$)
	$\mathbf{U}(10^3, 11\cdot10^3)$	**0.50** ($1.03\cdot10^{-3}$)	**0.50** ($1.01\cdot10^{-3}$)	**0.48** ($1.09\cdot10^{-3}$)	**0.50** ($1.02\cdot10^{-3}$)	**0.50** ($1.03\cdot10^{-3}$)
25	$\mathbf{U}(10^3, 3\cdot10^3)$	**0.52** ($1.33\cdot10^{-3}$)	**0.50** ($1.33\cdot10^{-3}$)	**0.50** ($1.34\cdot10^{-3}$)	**0.50** ($1.43\cdot10^{-3}$)	**0.50** ($1.34\cdot10^{-3}$)
	$\mathbf{U}(10^3, 11\cdot10^3)$	**0.52** ($7.64\cdot10^{-4}$)	**0.50** ($8.30\cdot10^{-4}$)	**0.50** ($7.89\cdot10^{-4}$)	**0.50** ($7.46\cdot10^{-4}$)	**0.48** ($7.77\cdot10^{-4}$)
35	$\mathbf{U}(10^3, 3\cdot10^3)$	**0.50** ($1.10\cdot10^{-3}$)	**0.50** ($1.12\cdot10^{-3}$)	**0.52** ($1.14\cdot10^{-3}$)	**0.52** ($1.14\cdot10^{-3}$)	**0.48** ($1.18\cdot10^{-3}$)
	$\mathbf{U}(10^3, 11\cdot10^3)$	**0.50** ($6.69\cdot10^{-4}$)	**0.50** ($6.59\cdot10^{-4}$)	**0.50** ($6.60\cdot10^{-4}$)	**0.50** ($6.66\cdot10^{-4}$)	**0.50** ($6.40\cdot10^{-4}$)
15	$\mathbf{N}(5\cdot10^3, 2\cdot10^3)$	**0.50** ($1.08\cdot10^{-3}$)	**0.52** ($1.12\cdot10^{-3}$)	**0.50** ($1.10\cdot10^{-3}$)	**0.50** ($1.07\cdot10^{-3}$)	**0.50** ($1.20\cdot10^{-3}$)
	$\mathbf{N}(10^4, 4\cdot10^3)$	**0.52** ($7.59\cdot10^{-4}$)	**0.48** ($7.14\cdot10^{-4}$)	**0.50** ($7.84\cdot10^{-4}$)	**0.52** ($8.06\cdot10^{-4}$)	**0.50** ($7.80\cdot10^{-4}$)
25	$\mathbf{N}(5\cdot10^3, 2\cdot10^3)$	**0.50** ($8.86\cdot10^{-4}$)	**0.50** ($8.55\cdot10^{-4}$)	**0.50** ($8.42\cdot10^{-4}$)	**0.50** ($8.15\cdot10^{-4}$)	**0.52** ($8.67\cdot10^{-4}$)
	$\mathbf{N}(10^4, 4\cdot10^3)$	**0.52** ($5.61\cdot10^{-4}$)	**0.50** ($6.15\cdot10^{-4}$)	**0.50** ($5.60\cdot10^{-4}$)	**0.50** ($5.76\cdot10^{-4}$)	**0.50** ($5.89\cdot10^{-4}$)
35	$\mathbf{N}(5\cdot10^3, 2\cdot10^3)$	**0.50** ($6.69\cdot10^{-4}$)	**0.50** ($7.04\cdot10^{-4}$)	**0.50** ($7.25\cdot10^{-4}$)	**0.50** ($7.17\cdot10^{-4}$)	**0.48** ($7.61\cdot10^{-4}$)
	$\mathbf{N}(10^4, 4\cdot10^3)$	**0.50** ($5.11\cdot10^{-4}$)	**0.50** ($4.94\cdot10^{-4}$)	**0.52** ($4.72\cdot10^{-4}$)	**0.50** ($5.44\cdot10^{-4}$)	**0.50** ($5.03\cdot10^{-4}$)
15	$\mathbf{P}(1.0, 10^3)$	**0.50** ($1.26\cdot10^{-3}$)	**0.50** ($3.72\cdot10^{-4}$)	**0.50** ($1.34\cdot10^{-3}$)	**0.50** ($1.03\cdot10^{-3}$)	**0.50** ($8.82\cdot10^{-4}$)
	$\mathbf{P}(1.5, 10^3)$	**0.50** ($8.08\cdot10^{-4}$)	**0.50** ($3.05\cdot10^{-4}$)	**0.52** ($2.95\cdot10^{-4}$)	**0.50** ($6.80\cdot10^{-4}$)	**0.50** ($1.15\cdot10^{-3}$)
25	$\mathbf{P}(1.0, 10^3)$	**0.50** ($1.53\cdot10^{-6}$)	**0.50** ($8.32\cdot10^{-7}$)	**0.50** ($1.73\cdot10^{-6}$)	**0.50** ($1.09\cdot10^{-6}$)	**0.50** ($1.91\cdot10^{-6}$)
	$\mathbf{P}(1.5, 10^3)$	**0.54** ($9.70\cdot10^{-4}$)	**0.54** ($7.18\cdot10^{-4}$)	**0.52** ($1.03\cdot10^{-3}$)	**0.52** ($8.16\cdot10^{-4}$)	**0.52** ($1.09\cdot10^{-3}$)
35	$\mathbf{P}(1.0, 10^3)$	**0.50** ($8.00\cdot10^{-4}$)	**0.50** ($4.86\cdot10^{-4}$)	**0.50** ($7.26\cdot10^{-4}$)	**0.50** ($8.15\cdot10^{-4}$)	**0.50** ($5.32\cdot10^{-4}$)
	$\mathbf{P}(1.5, 10^3)$	**0.50** ($4.93\cdot10^{-4}$)	**0.50** ($3.91\cdot10^{-4}$)	**0.50** ($3.32\cdot10^{-4}$)	**0.50** ($3.87\cdot10^{-4}$)	**0.50** ($3.09\cdot10^{-4}$)

Table 1: Optimal α for i.i.d. voters (with estimated $\mathbf{E}[\Delta]$ in parentheses)

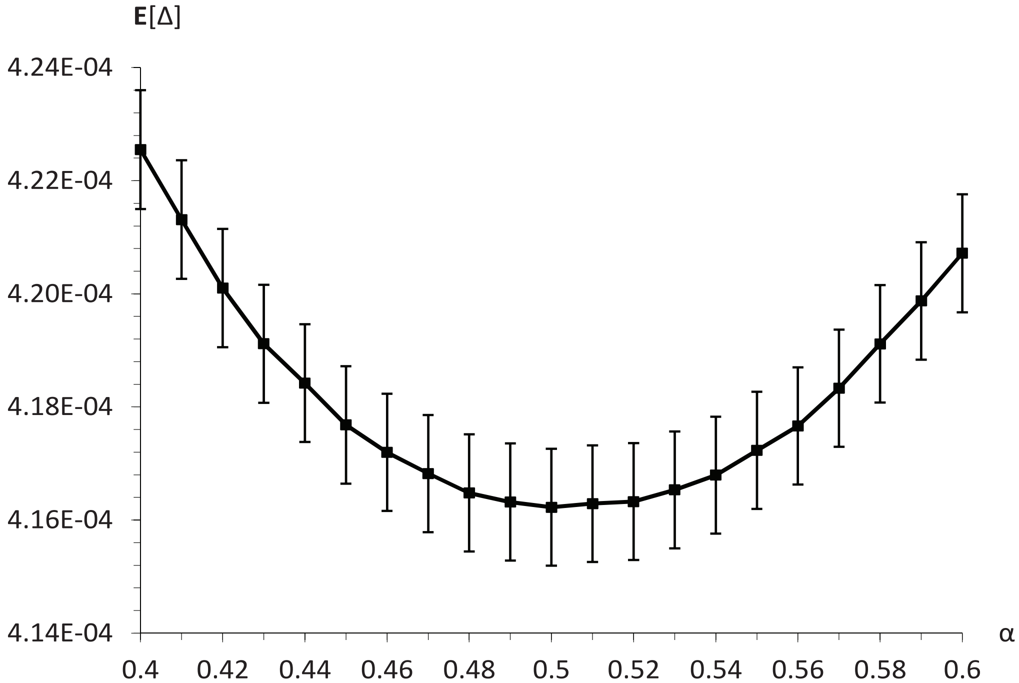

Figure 1: Estimates of $\mathbf{E}[\Delta]$ with 95%-confidence intervals for EU27 population data

Figure 1 illustrates how $\mathbf{E}[\Delta]$ varies as α ranges from 0.4 to 0.6 in steps of 0.01 for recent population data of the European Union (EU27).[17] Its Council, commonly referred to as the Council of Ministers, plays a most important rule in legislation of the EU (see, e.g., Napel and Widgrén 2006). It is a prominent example of a two-tier voting system because its members officially represent national governments and, eventually, the citizenries of the member states.[18] The current weighted voting rules for the Council, and also its future ones as codified in the Treaty of Lisbon, involve supermajority requirements in multiple dimensions. If we, nevertheless, assume a 50% decision quota as in Figure 1, our estimate for $\mathbf{E}[\Delta]$ is minimized by $\alpha = 0.5$ among the considered weight allocation rules. The figure includes the respective 95%-confidence intervals for $\mathbf{E}[\Delta]$. While an even greater number of iterations would tighten these intervals and allow to discriminate between values of α in the immediate proximity of 0.5, the square root rule already does significantly better than $\alpha \leq 0.44$ or $\alpha \geq 0.57$

low-variance and high-variance uniform and normal distribution with fixed r suggests that the deficit and variance of the population sizes are inversely related.

[17]We have used 2010 population data measured in *1,000 individuals* for computational reasons. This corresponds with the "block model" in Barberà and Jackson (2006), which supposes that a constituency can be subdivided into equally sized "blocks" whose members have perfectly correlated preferences within blocks, but are independent across blocks.

[18]Other "canonical examples" include the US Electoral College and the IMF's Board of Governors and Executive Board. To the extent that factions split along national rather than party lines, also the European Parliament might be regarded as the upper tier of a two-tiered system.

The robustness of $\alpha = 0.5$ being optimal for i.i.d. preferences which is indicated by Table 1 and Figure 1 is, in fact, a bit surprising. In particular, the approximate optimality of a square root law under the alternative objective of rendering constituencies' pivot probabilities proportional to their population sizes (in order to equalize each citizen's chances to be doubly pivotal), which we have invoked in order to gain some intuition in Section 3, is much more sensitive to r, the variance σ^2 when $n_j \sim \mathbf{N}(\mu, \sigma^2)$, and the skewness parameter κ when $n_j \sim \mathbf{P}(\kappa, \theta)$ (see Maaser and Napel 2007). That, here, our simulation results are so clear-cut suggests that, despite the problems which we encountered, a straightforward analytical answer to our question in the i.i.d. case may exist after all.

Concerning cases in which the ideal points of citizens are *not* independent and identically distributed, we focus on a special type of positive correlation within constituencies. In particular, we determine individual ideal points ν^i by a two-step random experiment: first, we draw a constituency-specific parameter μ_j independently for each $j = 1, \ldots, r$ from a distribution G with standard deviation σ_{ext}. This parameter captures the degree of *external heterogeneity* between $\mathcal{C}_1, \ldots, \mathcal{C}_r$ for the policy issue at hand. Parameter μ_j is taken to reflect the expected ideal point of citizens from $\mathcal{C}_j$. Each citizen $i \in \mathcal{C}_j$ is then assigned an individual ideal point ν^i from a distribution F_{μ_j} which has mean μ_j and is otherwise just a shifted version of some distribution $F \equiv F_0$ for each constituency $j = 1, \ldots, r$.[19] F's standard deviation σ_{int} is a measure of the *internal heterogeneity* in any constituency. It reflects opinion differences within any given $\mathcal{C}_j$. So, in summary, our second set of simulations has taken the ideal points of all citizens to be identically distributed, with convolved a priori distribution $G * F$, but to exhibit dependencies: citizens in constituency $\mathcal{C}_j$ all experience the same shift μ_j, which is independent of μ_k for any $k \neq j$.

The ratio $\sigma_{ext}/\sigma_{int}$ between external and internal heterogeneity provides a measure of the degree to which citizens are more similar within than between constituencies or, loosely speaking, the *preference dissimilarity* of the constituencies. The ratio $\sigma_{ext}/\sigma_{int} = 0$ corresponds to the i.i.d. case. Figure 2 shows how positive external heterogeneity quickly causes the square root rule ($\alpha^* = 0.5$) to stop being the power law which minimizes the direct democracy deficit. This finding is independent of whether one considers the real-world EU27 population configuration underlying Figure 1 (full line) or artificial configurations from Table 1 (dotted and broken lines).[20] The optimality of $\alpha = 1$ as $\sigma_{ext}/\sigma_{int} \to \infty$, which was derived in Section 3 for fixed $\sigma_{ext} > 0$ and $\sigma_{int} \to 0$, extends to relatively low dissimilarity levels in approximation (say, $\sigma_{ext}/\sigma_{int} \geq 4$). The greater the preference dissimilarity which shapes the policy ideals of representatives in $\mathcal{R}$, the more we approach a situation in which a linear voting weight allocation yields a minimal direct democracy deficit.

[19] Specifically, we draw μ_j from a uniform distribution $\mathbf{U}(-a, a)$ with variance σ^2_{ext}, and then obtain $\nu^i = \mu_j + \varepsilon$ with $\varepsilon \sim \mathbf{U}(0, 1)$.

[20] We display the column (I) configurations $\mathbf{U}(10^3, 3{\cdot}10^3)$, $\mathbf{N}(5{\cdot}10^3, 2{\cdot}10^3)$, and $\mathbf{P}(1.0, 10^3)$ with $r = 15$.

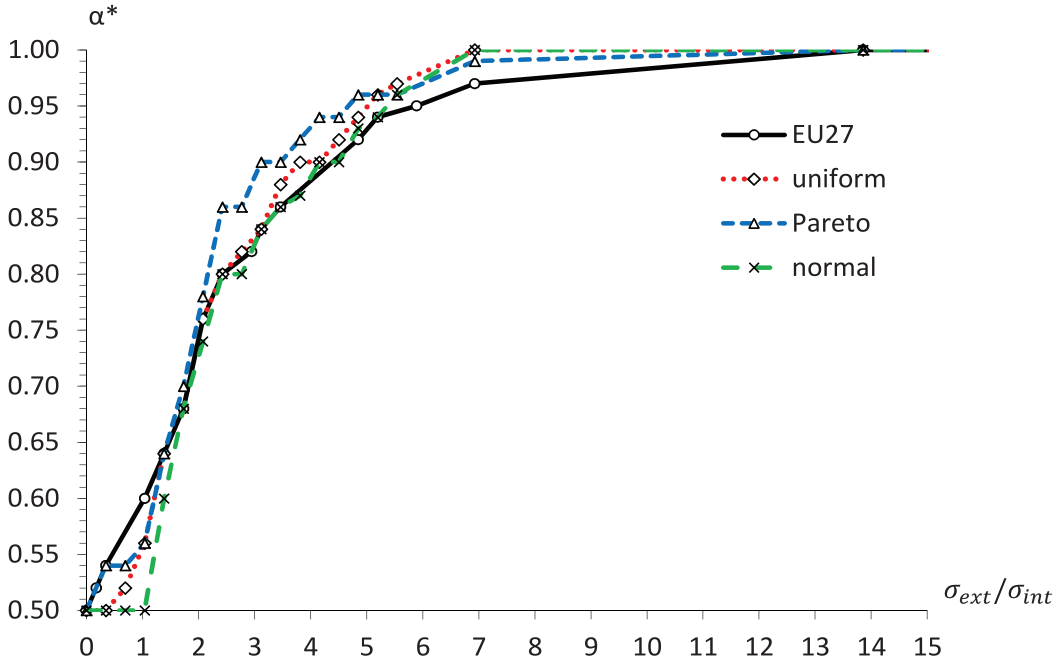

Figure 2: Optimal α for EU27 and artificial configurations with increasing dissimilarity

5 Concluding remarks

Obviously, one needs to be cautious in drawing political conclusions from these findings. Many different criteria play a role when voting weights are assigned to representatives from differently sized constituencies in an assembly such as the Council of the EU. We have here investigated only one – the congruence between the assembly's decision and that which a direct democracy would have arrived at. And this was done under admittedly very stylized assumptions. The presumption of one-dimensional issues with single-peaked preferences and the prominent role given to the (weighted) median in both bottom and top-tier decision making are the most noteworthy.

Both presumptions are direct generalizations from the world of binary decision making, which has so far dominated the analysis of two-tier voting systems: assuming that individual voters are either "for" or "against" a proposal implies single-peakedness on $\{0, 1\}$. The majority preference in a constituency is equivalent to its median position for a binary policy space, and taking a weighted majority decision at the top tier is nothing but adopting its weighted median. Our investigation moves from the dichotomous policy space $\{0, 1\}$ to the interval $[0, 1]$, and other finite or infinite one-dimensional intervals of alternatives. The so-called "second square root rule" of Felsenthal and Machover (1999) is, non-trivially, found to extend to this richer framework if its underlying i.i.d. assumptions are maintained.

However, as a still growing list of investigations of binary collective decisions[21] has high-

[21] See, for instance, Gelman et al. (2002), Barberà and Jackson (2006), Beisbart and Bovens (2007), Kirsch (2007), Feix et al. (2008), and Kaniovski (2008).

lighted, independence is indeed a critical assumption. For binary decisions as well as ones taken in the conventional median voter world[22], and considering a variety of majoritarian, egalitarian or utilitarian objective functions, the "optimality" of *square root rules* tends to break down when voters are acknowledged to be more similar within than between constituencies. *Linear* voting weight allocations then perform better. This pattern has been confirmed in the present paper. Whether a square root or a linear rule should be selected in order to allocate voting weights to the representatives of differently sized constituencies is not so much a question of whether one thinks utilitarian welfare, the principle of "one person, one vote", or majoritarianism in approximation of popular referenda and direct democracy more important. The results so far suggest that it does neither depend on whether political decisions are binary on exogenous proposals or the median voter's position emerges endogenously on the agenda and is then passed. The truly fundamental question for the selection of voting weights is: how great is the heterogeneity of voter interests *across* constituencies, such as member states of the European Union, compared to *within* constituencies?

References

Arnold, B. C., N. Balakrishnan, and H. N. Nagaraja (1992). *A First Course in Order Statistics*. New York: John Wiley & Sons.

Banks, J. S. and J. Duggan (2000). A bargaining model of collective choice. *American Political Science Review 94*(1), 73–88.

Barberà, S. and M. O. Jackson (2006). On the weights of nations: Assigning voting weights in a heterogeneous union. *Journal of Political Economy 114*(2), 317–339.

Baurmann, M. and H. Kliemt (1993). Volksabstimmungen, Verhandlungen und der Schleier der Insignifikanz. *Analyse und Kritik 15*(2), 150–167.

Beisbart, C. and L. Bovens (2007). Welfarist evaluations of decision rules for boards of representatives. *Social Choice and Welfare 29*(4), 581–608.

Cho, S. and J. Duggan (2009). Bargaining foundations of the median voter theorem. *Journal of Economic Theory 144*(2), 851–868.

Dahl, R. A. (1994). A democratic dilemma: System effectiveness versus citizen participation. *Political Science Quarterly 109*(1), 23–34.

Feix, M. R., D. Lepelley, V. Merlin, J.-L. Rouet, and L. Vidu (2008). Majority efficient representation of the citizens in a federal union. mimeo, Université de la Réunion, Université de Caen, and Université d'Orleans.

Felsenthal, D. and M. Machover (1998). *The Measurement of Voting Power – Theory and Practice, Problems and Paradoxes*. Cheltenham: Edward Elgar.

[22]See Maaser and Napel (2007) and Kurz et al. (2011). – A third branch of the literature on two-tier voting has started to look also at distributive collective decisions, i.e., a simplex of monetary or utility payoffs. Examples are Laruelle and Valenciano (2008) and Le Breton et al. (2011).

Felsenthal, D. and M. Machover (1999). Minimizing the mean majority deficit: The second square-root rule. *Mathematical Social Sciences 37*, 25–37.

Frey, B. S. and A. Stutzer (2006). Direct democracy: Designing a living constitution. In R. Congleton and B. Swedenborg (Eds.), *Democratic Constitutional Design and Public Policy. Analysis and Evidence*, pp. 39–80. Cambridge, MA: MIT Press.

Gelman, A., J. N. Katz, and F. Tuerlinckx (2002). The mathematics and statistics of voting power. *Statistical Science 17*, 420–435.

Gerber, E. R. and J. B. Lewis (2004). Beyond the median: Voter preferences, district heterogeneity, and political representation. *Journal of Political Economy 112*(6), 1364–1383.

Kaniovski, S. (2008). The exact bias of the Banzhaf measure of power when votes are neither equiprobable nor independent. *Social Choice and Welfare 31*(2), 281–300.

Kirchgässner, G. and L. P. Feld (2004). The role of direct democracy in the European Union. In C. B. Blankart and D. C. Mueller (Eds.), *A Constitution for the European Union*, pp. 203–235. Cambridge, MA: MIT Press.

Kirsch, W. (2007). On Penrose's square-root law and beyond. *Homo Oeconomicus 24*(3/4), 357–380.

Koriyama, Y. and J.-F. Laslier (2011). Optimal apportionment. mimeo, École Polytechnique.

Kurz, S., N. Maaser, and S. Napel (2011). On the egalitarian weights of nations. mimeo, University of Bayreuth.

Laruelle, A. and F. Valenciano (2008). Bargaining in committees of representatives: The 'neutral' voting rule. *Journal of Theoretical Politics 20*(1), 93–106.

Le Breton, M., M. Montero, and V. Zaporozhets (2011). Voting power in the EU Council of Ministers and fair decision making in distributive politics. CeDEx Discussion Paper 2011-03, University of Nottingham.

Lepelley, D., V. Merlin, and J.-L. Rouet (2011). Three ways to compute accurately the probability of the referendum paradox. *Mathematical Social Sciences* (forthcoming). DOI: 10.1016/j.mathsocsci.2011.04.006.

Maaser, N. and S. Napel (2007). Equal represenation in two-tier voting systems. *Social Choice and Welfare 28*(3), 401–420.

Napel, S. and M. Widgrén (2006). The inter-institutional distribution of power in EU codecision. *Social Choice and Welfare 27*(1), 129–154.

Nurmi, H. (1998). *Rational Behaviour and Design of Institutions*. Cheltenham: Edward Elgar.

Penrose, L. S. (1946). The elementary statistics of majority voting. *Journal of the Royal Statistical Society 109*(1), 53–57.

Rieder, H. (1994). *Robust Asymptotic Statistics*. New York: Springer.

Schwertman, N. C., A. J. Gilks, and J. Cameron (1990). A simple noncalculus proof that the median minimizes the sum of the absolute deviations. *American Statistician 44*(1), 38–39.

Weingast, B. R., K. A. Shepsle, and C. Johnsen (1981). The political economy of benefits and costs: A neoclassical approach to distributive politics. *Journal of Political Economy 89*(4), 642–666.