AUTHOR QUERY FORM

ELSEVIER	**Journal:** JELS **Article Number:** 1040	**Please e-mail or fax your responses and any corrections to:** **E-mail:** **corrections.esch@elsevier.tnq.co.in** **Fax: +31 2048 52789**

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof.

Location in article	**Query / Remark: click on the Q link to go** **Please insert your reply or correction at the corresponding line in the proof**
Q1	As per the journal style Research Highlights is mandatory. Please provide it.

Thank you for your assistance.

Electoral Studies xxx (2010) 1–5

Contents lists available at ScienceDirect

Electoral Studies

journal homepage: www.elsevier.com/locate/electstud

Constraints on the turnout gap between high and low knowledge (or income) voters: Combining the Duncan-Davis method of bounds with the Taagepera method of bounds☆

Bernard Grofman*

Department of Political Science, Center for the Study of Democracy, University of California, 3151 Social Science Plaza, Irvine, CA 92697-5100, United States

A R T I C L E I N F O

Article history:
Received 22 January 2009
Received in revised form 26 April 2010
Accepted 1 July 2010

Keywords:
Turnout
Boundary conditions
Method of bounds

A B S T R A C T

Countries differ quite substantially in mean turnout levels, and it is equally well known that there may be substantial within-country variation as well, for example, between high income and low-income groupings or between high political knowledge and low political knowledge groupings. It has been hypothesized that the size of such between-group gaps will fall as turnout rises, and conversely (Franklin, 2004. Blais, 2000). However, as Franklin (2004) also noted, there are mathematical constraints on the size of the turnout gap that are related to the level of turnout. For example, in the limit, if turnout is 100%, then all groups must have identical turnout. Here we build on this insight by adapting the classic work on boundary conditions done by two sociologists (Duncan and Davis, 1953) show precisely what the boundary constraints look like over the entire range of turnout values. Then we show how these constraints can help make sense of the strong relationship found between overall turnout and the size of the gap between voters above and below the median in political knowledge in the Fisher et al. (2008) cross-national study. To do so we draw on ideas in Rein Taagepera (2007, 2008) about how to use boundary condition information to develop better theoretical models.

© 2010 Published by Elsevier Ltd.

1. Introduction: The Duncan-Davis methods of bounds

There is a huge literature on voter turnout addressing a variety of questions, from the micro-level question of why an individual voter might choose to vote in any particular election, to the macro-level question of how to explain cross-national variation in turnout, to meso-level questions such as how to explain variations across particular types of groups in their mean level of turnout. In the best of all possible worlds, there would be a unified theory of turnout that would allow us to address the full range of such questions. Some recent work has attempted to bridge the gap between levels of analysis by arguing that, for some groups, the size of the turnout differences between them can be linked to aggregate mean turnout levels. In particular, it has been hypothesized that the size of the gap between high income and low-income groupings, or between high political knowledge and low political knowledge groupings, will fall as overall turnout rises, and conversely

☆ We are indebted to Sue Ludeman for bibliographic assistance and to Rein Taagepera for invaluable suggestions. This research was partially supported by the Jack W. Peltason (Bren Foundation) Endowed Chair at the University of California, Irvine and under Social Sciences and Humanities Research Council of Canada (SSHRCC) Grant #410-2007-2153 to study "Political Competition" (co-PIs: Stanley Winer, and J. Stephen Ferris).

* Tel.: +1 949 824 6394; fax: +1 949 824 8762.
E-mail address: bgrofman@uci.edu.

0261-3794/$ – see front matter © 2010 Published by Elsevier Ltd.
doi:10.1016/j.electstud.2010.07.002

Please cite this article in press as: Grofman, B., Constraints on the turnout gap between high and low knowledge (or income) voters: Combining the Duncan-Davis method of bounds with the Taagepera method of bounds, Electoral Studies (2010), doi:10.1016/j.electstud.2010.07.002

(Franklin, 2004; Blais, 2000). Most recently, Fisher et al. (2008) test this latter hypothesis using Wave 1 and Wave 2 CSES data on countries with a 1 or 2 score on Freedom House democracy measure in 2001, but excluding two compulsory voting countries, Australia and Belgium.[1]

Franklin (2004: 206–7) sounded a note of caution in studying the link between turnout differences between groups and overall turnout by pointing out that there were mathematical constraints on the magnitude of the possible differences in turnout across groups linked to the overall turnout level. In particular, in the limit, if turnout is 100%, then all groups must have identical turnout. If we divide a sample into two non-overlapping groups on any threshold variable, and then compare the turnout differences between the two groups, we will find that there are upper bounds on the size of the turnout gap (or the log-odds turnout ratio) between the two groups based on the overall mean turnout level. The key to understanding what is going on is the notion of a weighted average, combined with the observation that a variable like turnout is a bounded variable such that $0 \leq t \leq 1$. We will focus on the maximum and minimum gaps.

If there are two groups of size p and $1 - p$, respectively and each has turnout t_1 and t_2, respectively, then, by definition[2]

$$\bar{t} = pt_1 + (1 - p)t_2 \qquad (1)$$

But, in particular, if we construct the two groups as those above and those below a median on some single threshold variable, it must be the case that

$$\bar{t} = (t_1 + t_2)/2 \qquad (1)'$$

The sociologists, Otis Dudley Duncan and James Davis (Duncan and Davis, 1953) noticed the fact that Equation (1) and some simple algebra involving the bounded- ness of t could be used to set bounds on t_1 and t_2 as p was varied. Without loss of generality we may take $t_2 \leq t_1$. First consider the case where $\bar{t} > p$. Clearly the *maximum* value of t_1 in such a case is 1; but if $t_1 = 1$, then that leaves $(\bar{t} - p)$ units of turnout to be allocated to the $(1 - p)$ members of group 2, which mean that the *minimum* value of $t_2 = (\bar{t} - p)/(1 - p)$. Similarly, in the case where $\bar{t} \leq p$, the *minimum* value of t_2 in such a case is 0; but if $t_2 = 0$, then that leaves $\bar{t}$ units of turnout to be allocated to the p members of group 1, which means that the *maximum* value of $t_1 = \bar{t}/p$.[3]

When we deal with the special case where $1 - p = p = .5$, e.g., where we have divided into two groups of equal size above and below the median on some variable, the above formulas simplify further to get the results shown in Table 1 below. Once we know the minimum and maximum turnout values of the two groups, the maximum gap between them is straightforward to calculate, as shown in Table 1. But the minimum gap between them is even easier to calculate; it is 0, since we may have identical turnout levels in the two groups. Thus Table 1 allows us to set both minimum and maximum bounds on the size of the turnout gap between the two groups.

The reader may verify that, in case (a), if $\bar{t} = 1$, we get the limiting case where the both groups have identical turnout. Similarly, in case (b), if $\bar{t} = 0$, we hit the other limiting case. The reason these maximum and minimum values are important to know is that they set constraints on what is possible, and thus can be used to give rise to expectations of what is likely.

2. Theoretically derived bounds on turnout differences between groups

We show in Fig. 1, the maximum values and minimum values of the gap between the turnout levels of two groups defined as above or below the median level of some variable (e.g., income, knowledge, education, etc.). In this figure, following Taagepera (2008) we also take a first cut at predicting the likely values of the turnout gap between the two groups by appealing to the principle of ignorance, i.e., if we do not know what to expect, we predict something in between the best case and the worst case scenario. While Taagepera normally uses a *geometric* mean for this purpose (see e.g., Taagepera and Shugart, 1989; Taagepera, 2007, 2008; see also Grofman, 2004), because our lower bound is zero we first use a simple *arithmetic* average. The lower triangle shown in the graph indicates these "averaged" bounds.

[1] Fisher et al. (2008) are particularly concerned to show that, even controlling for turnout, the gap in turnout between high and low knowledge groupings will be higher in countries using plurality than in countries using proportional representation. This is not a question we will look at here. We would simply note that these authors present strong evidence that, on average, plurality systems do exacerbate the turnout gap between high knowledge and low knowledge groupings (see e.g., Fig. 1, p. 96).

[2] The knowledgeable reader will recognize this equation as the basis for Goodman (1953, 1959) ecological regression.

[3] The Duncan-Davis method was long used by expert witnesses testifying about vote dilution and racially polarized patterns of voting in elections where inferences had to be made about the voting behavior of minority and non-minority voters from aggregate election data at the voting tabulation unit (precinct) level combined with census or registration data on the racial composition of the electorate in the very common situation where no survey or exit poll data was available (see e.g., Loewen and Grofman, 1989). King's method of ecological inference (King, 1997) builds on the Duncan-Davis method of bounds by treating its maximum and minimum values as determining the slope and intercept of a line. Then observations from each ecological unit determines a line that represents the range of values that is consistent with the data in the unit, and the tomographic plot of all such lines can be used (using MLE methods) to estimate the most likely values for unknown parameters – such as t_1 and t_2, Where t is taken to be support for a black candidate in a black versus white contest and p is the proportion of the electoral constituency that is African-American, this approach has been used to directly assess evidence of racial bloc voting using aggregate data on voting tabulation units (precincts). It offers a substantial improvement on both the theoretical underpinnings and (often) the descriptive accuracy of the Goodman ecological regression approach and the simple use of the Duncan-Davis method of bounds (Grofman, 2000).

UNCORRECTED PROOF

Please cite this article in press as: Grofman, B., Constraints on the turnout gap between high and low knowledge (or income) voters: Combining the Duncan-Davis method of bounds with the Taagepera method of bounds, Electoral Studies (2010), doi:10.1016/j.electstud.2010.07.002

Table 1
Duncan-Davis Bounds on turnout for two groups respectively above and below the median level of some variable.

	t_1	t_2	gap
(a) $\bar{t} > .5$			
max gap case	1	$2\bar{t}-1$	$2-2\bar{t}$
min gap case	$\bar{t}$	$\bar{t}$	0
(b) $\bar{t} \leq .5$			
max gap case	$2\bar{t}$	0	$2\bar{t}$
min gap case	$\bar{t}$	$\bar{t}$	0

Perhaps the most obvious (and perhaps not especially intuitive) feature of Fig. 1 is the remarkable non-monotonicity it displays. For turnout values above .5 we get the pattern that we might expect, with the turnout gap necessarily declining as we move toward 100% turnout. But for turnout values below .5 we get the opposite pattern of the turnout gap necessarily declining as we move toward 0% turnout. While this seemingly peculiar feature of the graph is, no doubt, obvious upon reflection, it is useful to call attention to it, since if we see this kind of curvilinearity in empirical data on gap magnitudes as a function of total turnout this suggests that the boundary constraint does matter. And if the boundary constraint matters, then at least some of what we might be attributing to empirically-driven regularities is actually a kind of purely statistical artifact. A second feature of the mathematical results represented in Fig. 1 is that it shows clearly that simply eliminating the tails of the turnout distribution, e.g., cases with very very high or very very low turnout, does not deal with the problem of the boundaries shaping what is possible. Everywhere in the graph, except just around 50% turnout, there are non-trivial constraints on what is possible. Moreover, if we were to *randomly* assign cases onto the feasible space (the triangle), using a uniform distribution over that bounded space, and then plot the resultant values using LOESS, [4] what we would get would necessarily mimic the triangular distribution pattern. And, of course, if we confine ourselves to turnout values above .5, such a *random* distribution over the *feasible* space would *necessarily* show a pattern of gap magnitude decreasing with turnout.

There is, however, an alternative way to think about an “ignorance-based” model of what the distribution of gap sizes might look like as we vary overall turnout. Instead of using an arithmetic average, what we do is assume the expected maximum gap will occur at the center of gravity of the space of feasible alternatives.[5] Because our bounds are in the form of a triangle, the location of the center of gravity is related to the means on the bounding edges of that triangle. In particular, the center of gravity of the bounding triangle may be determined by finding the intersection of the line between the vertex of the triangle at (0, 0) and the midpoint of the right edge of the triangle (.75, .50) and the line between the vertex at (1, 0) and the midpoint of the left edge of the triangle (.25, .50). These lines are $y = .667$ x and $y = -.667x + .667$, respectively. They intersect at (.5, .333). Thus, absent any other information we would expect that the maximum turnout gap would occur when turnout=.50 and would have a value of .33.

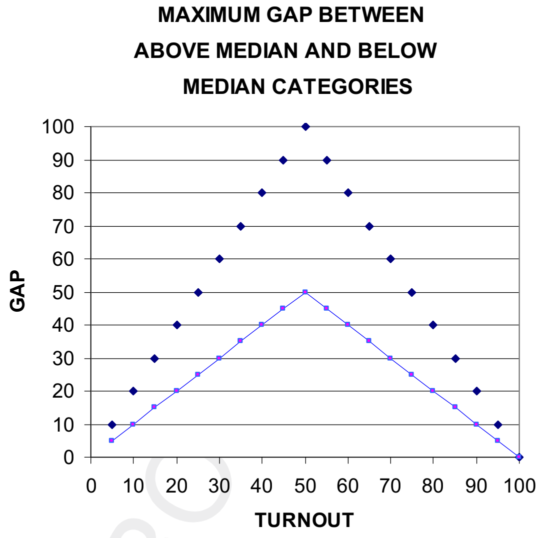

Fig. 1. Theoretically derived expectations about turnout differences.

3. Empirical application of methods of bounds

Now we will use the data from Fisher et al. (2008) to test the predictions that the maximum turnout gap would occur when turnout=.50 and would have a value of .33 and that the general pattern would be roughly triangular. Looking at Fisher et al. (2008: Fig. 1, p.96) we see that the mean turnout gap between high knowledge and low knowledge groups at a 50% level of turnout is roughly 28%.[6] According to our statistically derived expectations, *ceteris paribus*, the gap at 50% turnout should be the *largest* gap. While the actual (mean) gap at the 50% level of turnout is nowhere near the 100% that is mathematically possible it is remarkably close to the 33.3% predicted by our purely theoretical “ignorance-based” model.

We may then take that single parameter, 1/3, as a correction factor on the maximum values shown in Fig. 1, i.e., we specify the *expected* gap as running from 0 to $.1/3 \times (2\bar{t})$ for $\bar{t} < .5$ and from 1/3 to $1/3 \times (2 - 2\bar{t})$ for $\bar{t} > .5$. Fig. 2 shows what this ignorance-based graph looks like, while Fig. 3 superimposes the values of the graph shown in Fig. 2 (in percentages rather than fractions, so as to be consistent with the style of presentation in Fisher et al., 2008) on the empirical cross-national data in Fisher et al. (2008: Fig. 1, p. 96) and the LOESS best fit curve calculated by them.

Considering that we are estimating the model we report in Fig. 3 based on purely mathematical considerations of bounds *with no empirical input whatsoever*, a visual

[4] LOESS (sometime more mnemonically spelled as LOWESS), is “locally weighted scatter plot smoothing” It is based on calculating regression values over small areas (the size of that local area is called “bandwidth”) and combining the results into a single graph.

[5] We are indebted to Rein Taagepera (personal communication, January 15, 2009) for suggesting this approach to determining bounds.

[6] We derive the figure of 28% from visual inspection of the graph.

Please cite this article in press as: Grofman, B., Constraints on the turnout gap between high and low knowledge (or income) voters: Combining the Duncan-Davis method of bounds with the Taagepera method of bounds, Electoral Studies (2010), doi:10.1016/j.electstud.2010.07.002

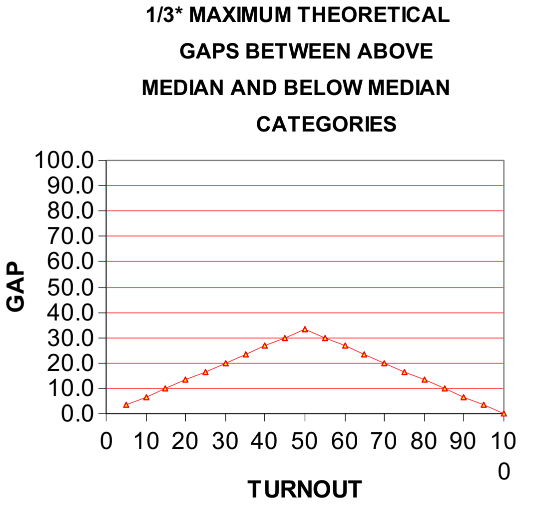

Fig. 2. Revised theoretically derived expectations about turnout differences.

comparison of our theoretical estimates and that of the LOESS fit to the empirical data in Fisher et al. (2008) shows concordance to a remarkable degree! Such a fit of a purely statistical "ignorance-based" model suggests that the fundamental mathematical relationships identified earlier in the paper are driving the differences in *mean* gap sizes as a function of overall turnout found in the Fisher et al. (2008) data. In excluding the compulsory voting countries Fisher et al. (2008: 94) show their awareness of this potential statistical pitfall. However, as noted earlier, excluding extreme cases does not fully address the problem. Also, contrary to what Fisher et al claim (2008: 94), shifting to a log-odds relationship rather than looking at the raw turnout gap does not solve the problem either.

There is also a third approach that we might use to estimate relationships based on the bounds, what we might call a "partial ignorance" approach. In modifying the parameters of a theoretical model to better fit empirical data, Taagepera (2007) uses one single value that is empirically determined to estimate a larger pattern, e.g., using the size of the largest party as a parameter in estimating the distribution of party sizes. In our context, it seems natural to take the value of the gap at the theoretical maximum gap point, a turnout level of 50%, to adjust the maximum gap parameters. Taking this value, 28%, as what we might call our "shrinking parameter" also gives us a very good fit to the data. However, the empirical fit to the observed LOEES curve of the purely theoretical model using the 1/3 center of gravity correction is actually marginally better than if we fit the triangular set of bounds using as our shrinking parameter the roughly 28% gap value at a turnout level of 50% that is empirically observed by Fisher et al. [7]

[7] For space reasons we have omitted the figure showing this .28 shrinking factor superimposition. It is available from the authors upon request. That fit is also incredibly good, very close to the fit we get from using 1/3 as our shrinking factor based on the theoretical bounds.

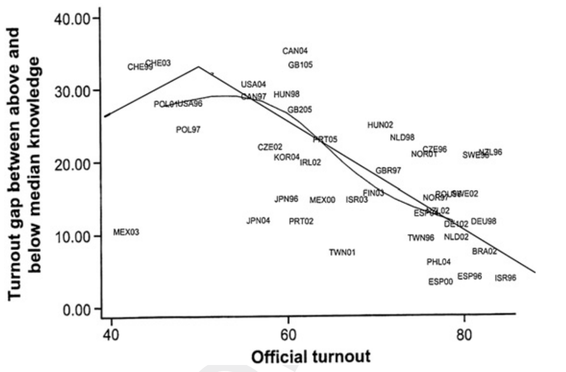

Fig. 3. Superimposition of triangularly shaped theoretically expected gap values on fisher et al. (2008: Fig. 1, p. 96) cross-national data and LOESS best fit curve.

4. Discussion

We are pleased to have been able to combine a way of specifying boundary constraints from classic sociological work written more than half a century ago with ideas about how to use boundary information to develop theory taken from a leading contemporary electoral systems theorist. In sum we believe this work to be promising in indicating areas where ideas in Taagepera (2008) and in his earlier work can be applied to move us away from pure curve fitting toward models that are informed by (algebraic) considerations of what is mathematically possible.

Of course, more sophisticated models could no doubt allow us to improve the fit between our theoretical expectations and the actual data, e.g., by using some non-linear model rather than the simple linear model we fit – such as with a curve that fell off more slowly from its peak value;[8] and/or by incorporating additional variables to account for the substantial variations in the turnout gap among countries with more or less identical levels of mean turnout that we observe in the data in Fig. 3 (taken from Fig. 1 in Fisher et al., 2008). In this context, in addition to the role of electoral system differences to which Fisher et al. (2008) call attention, we do have one particular suggestion, namely looking at "difference of mean" effects. For example, it seems obvious that if the knowledge levels of high knowledge and

[8] In this context Rein Taagepera (personal communication, January 18, 2009) has suggested that the simplest smooth function, G, that satisfies the boundary conditions that the gap must be zero at $t = 0$ and must be zero at $t = 1$ is $G = T(1 - T)$. That function has a smooth peak at (.5, .25), a bit below either the .33 value derived from the center of gravity approach or the .28 figure that is empirically observed. The fit of the G curve to the Fisher et al. (2008) data is also quite good, even if not quite as good as that obtained from the other two approaches (figure omitted for space reasons and available from the author upon request). Taagepera has also pointed out an important (and non-obvious) link between the $G = T(1 - T)$ function and the center of gravity approach, namely that the area under the triangular figure with peak at .33 and the smoothed figure with peak at .25 is actually the same, since integrating $G = T - T^2$ leads to $T^2/2 - T^3/3$. For T ranging between 0 and 1 this gives us $(1/2 - 0) - (1/3 - 0) = 1/6$. For the triangle the area is of course, simply, $½ \times 1/3 = 1/6$. We have deliberately not sought to explore the issues of best fit in this short essay since we see its main point as a very simple one: without taking boundary constraints into account, one will not understand what the data are actually showing us about causal relationships.

UNCORRECTED PROOF

Please cite this article in press as: Grofman, B., Constraints on the turnout gap between high and low knowledge (or income) voters: Combining the Duncan-Davis method of bounds with the Taagepera method of bounds, Electoral Studies (2010), doi:10.1016/j.electstud.2010.07.002

low knowledge groups are not that far apart then, *ceteris paribus*, we ought not to expect that the turnout differences between the two groups would be that far apart. Similarly, *ceteris paribus*, if a society is characterized by a high degree of socioeconomic equality, then we might expect the turnout gap between low SES and high SES groups would not be that large. Combining such empirically derived expectations with the theoretical bounds calculations should allow us considerable leverage in making sense of cross-national variation in the size of between group differences in turnout – or many other variables, for that matter.

References

Blais, André, 2000. To Vote or Not to Vote? The Merits and Limits of Rational Choice. University of Pittsburgh Press, Pittsburgh.

Duncan, Dudley, Davis, Beverley, 1953. An alternative to ecological correlation. American Sociological Review 18, 665–666.

Fisher, Stephen D., Lessard-Phillips, Laurence, Hobolt, Sara B., Curtice, John, 2008. Disengaging voters: do plurality systems discourage the less knowledgeable from voting. Electoral Studies 27, 89–104.

Franklin, Mark N, 2004. Voter Turnout and the Dynamics of Electoral Competition in Established Democracies since 1945. Cambridge University Press, Cambridge, MA.

Goodman, Leo, 1953. Ecological regression and the behavior of individuals. American Sociological Review 18 (6), 663–664.

Goodman, Leo, 1959. Some alternatives to ecological correlation. American Journal of Sociology 64, 610–625.

Grofman, Bernard, 2000. A primer on racial bloc voting analysis. In: Persily, Nathaniel (Ed.), The Real Y2K Problem: Census 2000 Data and Redistricting Technology. The Brennan Center for Justice, New York University School of Law, New York.

Grofman, Bernard, 2004. Rein Taagepera's approach to the study of electoral systems. Journal of Baltic Studies 35 (2), 167–185.

King, Gary, 1997. A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data. Princeton University Press, Princeton.

Loewen, James W., Grofman, Bernard, 1989. Comment: recent developments in methods used in voting rights litigation. Urban Lawyer 21 (3), 589–604.

Taagepera, Rein, 2007. Predicting Party Sizes: The Logic of Simple Electoral Systems. Oxford University Press, Oxford.

Taagepera, Rein, 2008. Making Social Sciences More Scientific: The Need for Predictive Models. Oxford University Press, Oxford.

Taagepera, Rein, Shugart, Matthew, 1989. Seats and Votes: The Effects and Determinants of Electoral Systems. Yale University Press, New Haven, CT.

UNCORRECTED PROOF

Please cite this article in press as: Grofman, B., Constraints on the turnout gap between high and low knowledge (or income) voters: Combining the Duncan-Davis method of bounds with the Taagepera method of bounds, Electoral Studies (2010), doi:10.1016/j.electstud.2010.07.002