AMERICAN ASSOCIATION OF WINE ECONOMISTS

AAWE WORKING PAPER
No. 97
Economics

A MATHEMATICIAN MEDDLES WITH MEDALS

Neal D. Hulkower

February 2012

www.wine-economics.org

A Mathematician Meddles with Medals

Neal D. Hulkower[1]

Abstract

A method to determine medal winners in wine competitions by aggregating rankings of judges using the Borda Count is presented. Each judge inserts three markers within his or her ranking of the wines that separate them into four clusters, those deemed worthy of a gold, silver, or bronze medal or unworthy of any medal.

JEL Codes: D70, L66
Keywords: Borda Count, wine judging, medals

Introduction

Scores of wine competitions are held annually to bestow gold, silver or bronze medals upon a select number of bottles. Only within the past few years has the reliability of the awards come under rigorous scrutiny and been found wanting. The focus of these studies has been on the decision makers. For major competitions, these are typically expert judges who may or may not have been trained and who may or may not have been screened for consistency. Since ultimately the credibility of the decisions stems from the credibility of the judges, how they are selected is critical. But once reliable judges have been empanelled, how their opinions are recorded and aggregated becomes of paramount importance. It is well-known that an outcome of any vote depends as much on the choice of method used to combine the votes as on any other factor (Saari 2001b). After examining a number of procedures for comparing wines, Amerine and Roessler (1983) concluded that "[r]anking procedures are then usually preferred" (p. 168). Ashenfelter and Quandt (1999) used rank values, which they called "Points Against," introduced in Amerine and Roessler (1983) to reassess the famous Judgment of Paris red wine competition. This method is equivalent to the Borda Count which Hulkower (2009, 2011) emphasized is the most mathematically defensible for combining individual rankings of wines to arrive at an aggregate ranking. In addition to the unique properties summarized in the third section of this paper, the Borda Count avoids distortions introduced by summing or averaging points assigned by individual judges which can diminish the influence of tougher graders thereby violating "one judge, one vote." The purpose of this paper is to offer a ranking procedure based on the Borda Count that can be used to award medals in a manner that most reliably and completely reflects each judge's opinion while preserving "one judge, one vote."

A summary of recent studies exposing problems with wine competitions is presented in the next section. The case for the Borda Count is made in the third section and a method for awarding medals based on it comprises the fourth section. The fifth section is a discussion that compares

[1] McMinnville, OR 97128. nhulkower@yahoo.com

the method proposed in this paper to an alternative in the literature. Conclusions are contained in the sixth section.

Judging the Judges

In any competition, the awarding of medals to wines reflects the opinion of each judge and the method used to combine these opinions. So the credibility of the medals derives from the credibility of the judges and the validity of the voting scheme. Hodgson (2008, 2009a, 2009b) has called into doubt the reliability of judges and the validity of medals awarded to wines in major competitions. His analysis of wine judge performance over four years concluded that "[i]n 30 cases [out of 65], about 50 percent, the wine and only the wine was the significant factor in determining the judges' score" (Hodgson, 2008, p. 111). In the remaining cases, judge bias and judge inconsistency dominate. In a follow up study (Hodgson, 2009b), Hodgson offers a metric that quantifies the consistency of a judge and suggests a threshold value for classifying a judge as "expert." Based on values of this metric computed from data in two studies, he concludes "that less than 30% of expert wine judges studied are, in fact, 'expert'" (Hodgson 2009b, p. 241). But his most damning finding based on a study of over 4000 entrants into 13 wine competitions is "that (1) there is almost no consensus among the 13 wine competitions regarding wine quality, (2) for wines receiving a Gold medal in one or more competitions, it is very likely that the same wine received no award at another, (3) the likelihood of receiving a Gold medal can be statistically explained by chance alone" (Hodgson, 2009a, p. 8).

Cao and Stokes (2010) considered three quantifiable characteristics of judges: bias, discrimination ability, and random variation. Bias is the difference between a judge's score and the average of all the judges' scores. A judge's ability to differentiate among wines based on some assessment of quality is called discrimination. Random variation is the magnitude of the random component of variability in an assessment of wine quality. This is determined from the variance of the scores given to three samples each of four wines. The authors conclude that "…bias and a low level of discrimination may be easier to overcome than a high variance. Judges with these two issues still rank the wines in a similar order compared to the other judges, and they only need to adjust the center and scale of their scores to be more consistent with the rest of the panel" (Cao and Stokes, 2010, p. 140). This finding suggests that a method of aggregating judges' assessments expressed as individual rankings will not be tainted by two negative characteristics exhibited by some of the judges.

A recent peripherally related study found "that individuals who are unaware of the price do not, on average, derive more enjoyment from more expensive wine. In fact, unless they are experts, they enjoy more expensive wines slightly less" (Goldstein, et al., 2008, p. 3). This begs two questions: For whose benefit are medals awarded? Who should decide which wines get them? If the intended audience is the average consumer, there is now evidence that expert judges may not share the same preferences and perhaps panels including experts and nonexperts should be tested. "Our results suggest that non-expert wine consumers should not anticipate greater enjoyment of the intrinsic qualities of a wine simply because it is expensive or is appreciated by experts" (Goldstein, et al., 2008, p. 8-9). This finding is ironic given the increased attention to selecting more credible and consistent judges. Clearly much more research is needed. In any

case, regardless of how the issue of how to select judges for a wine competition is resolved, the method of aggregating rankings detailed in the fourth section should be used.

Why Borda is Better

The Borda Count is named for eighteenth century French mathematician Jean-Charles de Borda who proposed this method to select new members of the Paris Académie des Sciences. He documented this positional voting scheme in (Borda, 1781). A positional voting scheme requires each voter to rank order all candidates, with ties permitted, and then assigns points to each candidate based on its position in the ranking. The points are summed and the candidate with the largest total is declared the winner. The simplest and most common positional voting scheme is the plurality vote in which each candidate ranked in first place is assigned 1 point while all others are assigned 0. Note that only the candidate in first place is considered by this method while the ranking of the remaining candidates is ignored. In contrast, for *n* candidates, the Borda Count assigns a Borda score of *n*-1 points to the top ranked choice, *n*-2 for the second and so on down to 0 for the last choice. In other words, the Borda score = *n* – the rank. Each of the candidates given the same rank is assigned the average of the points for the positions they occupy. If two candidates out of 6 are tied and occupy the third and fourth ranking, each is assigned (3 + 2)/2 = 2.5 points. Unlike plurality voting, the Borda Count utilizes the complete ranking provided by each voter to determine the societal outcome.

Borda's method has other highly desirable characteristics that make it unique among all positional voting schemes. It is the only one that satisfies all four of the following properties that are generally regarded as rational for a voting method:

- Any strictly transitive ranking can be chosen by a voter (Unrestricted Domain)
- The transitivity of the individual rankings is preserved in the aggregate ranking. In other words, if $A \succ B$ (i.e., if A is preferred to B) and $B \succ C$ then $A \succ C$.
- If each voter picks $A \succ B$, then the societal outcome has $A \succ B$ (Pareto Condition)
- The ranking of two candidates in the societal outcome is determined by the relative ranking by each voter and the number of candidates between the alternatives. This is known as Intensity Form of Independence of Irrelevant Alternatives (IIIA) or Intensity Form of Binary Independence (IBI) and was introduced by Donald Saari to "depose" Kenneth Arrow's dictator (Saari, 2000a, 2000b, 2001a, 2001b, 2008).

While unexpected outcomes or paradoxes can result from any voting method, Saari proved that the Borda Count is the only positional voting scheme whose inconsistencies over subsets of alternatives are due solely to cycles of rankings that result in ties known as Condorcet n-tuples (Saari 2000a, 2000b, 2008). What this means is if an alternative is either removed or added, the tie is broken and rank reversal may occur in the aggregate outcome. While there are other potential sources of paradoxes only the Borda Count is immune to them.

Hulkower (2007) introduced the idea of Borda with None as a binding option to give individual voters the ability to reject all or part of a slate. In other words, each voter ranks None along with the other candidates thereby partitioning them into two categories, acceptable and unacceptable. If None is in first place in the aggregated ranking, none of the candidates is elected.

A Borda-based Method for Awarding Medals

Adapting the idea in Hulkower (2007), the Borda Count can be used to aggregate the rankings of judges in a wine competition at the same time delineating those bottles worthy of each medal in a manner that best reflects the opinions of the judges. It works as follows. Each judge rank-orders each of the wines and inserts three markers, N_G, N_S, and N_B. Wines ranked above N_G are the judge's selections for Gold medals. Similarly, those ranked between N_G and N_S are the judge's choice for a Silver medal and those between N_S and N_B, for a Bronze medal. Those ranked below N_B would go without any medal. Each of these markers is treated in the same manner as a candidate wine and is assigned a score based on its position in the ranking. The Borda scores of the wines and the three markers resulting from each judge's ranking are summed yielding the societal outcome as an aggregate ranking. Based on this aggregate ranking, gold medals are awarded to those wines that rank above with N_G, silver is given to those ranked between N_G and N_S, and bronze is presented to those between N_S and N_B. Any wine ranked below N_B is not medaled. If a wine is tied with a marker, it can be awarded the higher medal. Any wine ranked above N_G by each judge can be awarded Double Gold. The highest ranked wine can be named Best in Show in its category.

This method is applied to a hypothetical case of 6 judges evaluating 9 wines. Table 1 contains each judge's ranking of the 9 wines and the three markers with 1 denoting first place down to 12 for last place. Note that Judge 3 ranked wines E and H in a tie in third place.

	Options											
Judge	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**N_G**	**N_S**	**N_B**
1	**12**	**3**	**9**	**8**	**2**	**11**	**7**	**4**	**6**	**1**	**5**	**10**
2	**11**	**3**	**10**	**7**	**1**	**12**	**8**	**2**	**5**	**4**	**6**	**9**
3	**10**	**1**	**9**	**6**	**3.5**	**11**	**12**	**3.5**	**7**	**2**	**5**	**8**
4	**10**	**2**	**7**	**6**	**1**	**12**	**9**	**3**	**4**	**5**	**8**	**11**
5	**11**	**4**	**10**	**8**	**2**	**12**	**7**	**1**	**5**	**3**	**6**	**9**
6	**5**	**3**	**6**	**10**	**2**	**8**	**12**	**9**	**11**	**1**	**4**	**7**

Table 1. Rankings of the 9 Wines and 3 Markers of Each Judge

These rankings readily translate into the Borda scores shown in Table 2.

	Options											
Judge	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**N_G**	**N_S**	**N_B**
1	**0**	**9**	**3**	**4**	**10**	**1**	**5**	**8**	**6**	**11**	**7**	**2**
2	**1**	**9**	**2**	**5**	**11**	**0**	**4**	**10**	**7**	**8**	**6**	**3**
3	**2**	**11**	**3**	**6**	**8.5**	**1**	**0**	**8.5**	**5**	**10**	**7**	**4**
4	**2**	**10**	**5**	**6**	**11**	**0**	**3**	**9**	**8**	**7**	**4**	**1**
5	**1**	**8**	**2**	**4**	**10**	**0**	**5**	**11**	**7**	**9**	**6**	**3**
6	**7**	**9**	**6**	**2**	**10**	**4**	**0**	**3**	**1**	**11**	**8**	**5**
Sum	**13**	**56**	**21**	**27**	**60.5**	**6**	**17**	**49.5**	**34**	**56**	**38**	**18**

Table 2. Borda Scores (= 12 – Ranking in Table 1) of the 9 Wines and 3 Markers of Each Judge

The aggregation of the 6 judges' rankings using the Borda Count yields the following:

$$E \succ B \approx N_G \succ H \succ \mathbf{N_S} \succ I \succ D \succ C \succ N_B \succ G \succ A \succ F$$

Wines E and B (which is tied with the gold marker, denoted by ≈) are awarded gold medals. Only wine H gets a silver medal and wines I, D, and C receive bronze medals. Wines G, A and F earn no medals. Since no wine was ranked above N_G by all six judges, none receives Double Gold. Depending on the rules of the competition, wine E may be declared Best of Show.

Some judges may find it difficult to rank a large number of wines. If so, pairwise comparisons can be made and scored in a manner that is equivalent to the Borda Count (Dym, Wood, Scott, 2002; Saari and Sieberg, 2004). To illustrate using the case above, we construct a pairwise comparison chart (Table 3) that records for each pair of options the number of wins in rows and the number of losses in columns. Options that tie each receive a 0.5. For example, four of the judges preferred option G to option F and two preferred option F to option G so 4 is entered in row G, column F and 2 appears in row F, column G. For each pair, the total number of wins is equal to the number of judges, in this case, 6. For each option, total wins + total losses = 66 which is the number of judges times the number of pairwise comparisons each must make, in this case, 6 x 11. Finally, note that the Total Wins for each option equal the sum of its Borda scores in Table 2 thus resulting in the same aggregate ranking of the 12 options.

Options	A	B	C	D	E	F	G	H	I	N_G	N_S	N_B	Total Wins
A		0	1	1	0	5	2	1	1	0	0	2	13
B	6		6	6	1	6	6	4	6	3	6	6	56
C	5	0		1	0	6	3	1	1	0	1	3	21
D	5	0	5		0	5	4	0	2	0	1	5	27
E	6	5	6	6		6	6	4.5	6	3	6	6	60.5
F	1	0	0	1	0		2	1	1	0	0	0	6
G	4	0	3	2	0	4		0	0	0	0	4	17
H	5	2	5	6	1.5	5	6		6	3	5	5	49.5
I	5	0	5	4	0	5	6	0		1	3	5	34
N_G	6	3	6	6	3	6	6	3	5		6	6	56
N_S	6	0	5	5	0	6	6	1	3	0		6	38
N_B	4	0	3	1	0	6	2	1	1	0	0		18
Total Losses	53	10	45	39	5.5	60	49	16.5	32	10	28	48	

Table 3. The Pairwise Comparison Chart for the Case of 6 Judges and 9 Wines

Discussion

Hodgson's findings have naturally spawned recommendations for alternative means of deciding medal winners. Cicchetti (2009) offers "a two-tiered method that considers an award on the basis of a total medal score averaged across four judges per each evaluated wine; and a reliability algorithm that grades the degree of inter-judge reliability…" (p. 242). The reliability algorithm mirrors the 100 point scale favored by several prominent critics. Each judge assigns a score of 0 for no medal, 1 for bronze, 2 for silver and 3 for gold. A table was constructed with the 35 possible outcomes, the associated scores, and the medal to be awarded. The reliability score expressed as a percent is assigned, apparently not calculated, based on the scatter of the judges' scores with lower values being given to those assessments with wide ranges of scores. This method has a number of shortcomings including the arbitrariness of the reliability scores, the lack of a clear description as to how they are used, and the need for the use of Preponderance of Opinion (POO) in three instances when three of judges assigned the same score and the fourth score differed by at least 2 points. In contrast to the method proposed in this paper, the two-tiered method does not consider the relative rankings of wines within any particular medal category. So, for example, all wines that a judge assigns a 3 are treated as equally ranked. As such, important information that can and usually does influence the societal outcome and final selection of medal winners is ignored.

Cicchetti attempts to use a measure of reliability of the judges along with the medal scores to determine winners, though precisely how is unclear. Robin Goldstein also weighs higher the scores of judges who perform better on identifying two samples of the same wine out of six (Taber, 2011). While the selection of reliable judges is critical to the credibility of awards made by them, there are problems associated with weighing the scores of judges or panels based on any measure. Since the number of judges tends to be small, any weighing of scores can make a

single judge effectively a dictator. It is this author's opinion that vetting judges should be separated from the process of awarding medals. Once a judge is deemed reliable, his or her ranking should be weighed equally with all others. While the method described in this paper can be adjusted to weigh rankings of one or more judges by simply multiplying the Borda scores by the desired weight, this is not recommended if "one judge, one vote" is to be preserved.

Conclusions

While the use of the Borda Count or its equivalent to rank wines is not novel, having individual judges insert markers into their rankings to separate those deserving gold, silver or bronze, or no medal at all, is. This method uses all information about each judge's preferences to arrive at a societal outcome that best reflects the collective opinions of the judges. It is transparent, easy to use and to score, and is mathematically sound. What remains is to apply this method in a real competition.

References

Amerine, Maynard and Roessler, Edward (1983) *Wines: Their Sensory Evaluation*, New York: W.H. Freeman.

Ashenfelter, Orley and Quandt, Richard (1999) Analyzing a Wine Tasting Statistically, *Chance* 12, 16-20.

Borda, J. C. (1781) Mémoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences, Paris*, as translated in: McLean, Iain and Urken, Arnold B. (eds) (1995) *Classics of Social Choice*, Ann Arbor: The University of Michigan Press, 83-89.

Cao, Jing and Stokes, Lynne (2010) Evaluation of Wine Judge Performance through Three Characteristics: Bias, Discrimination, and Variation, *Journal of Wine Economics*, 5, 132-142.

Cicchetti, Domenic V. (2009) A Proposed System for Awarding Medals at a Major U.S. Wine Competition, *Journal of Wine Economics*, 4, 242-247.

Dym, Clive L., Wood, William H., Scott, Michael J. (2002) Rank ordering engineering designs: pairwise comparison charts and Borda counts, *Research in Engineering Design*, 13, 236-242.

Goldstein, Robin et al. (2008) Do More Expensive Wines Taste Better? Evidence from a Large Sample of Blind Tastings, *Journal of Wine Economics*, 3, 1-9.

Hodgson, Robert T. (2008) An Examination of Judge Reliability at a major U.S. Wine Competition, *Journal of Wine Economics*, 3, 105-113.

Hodgson, Robert T. (2009a) An Analysis of the Concordance Among 13 U.S. Wine Competition, *Journal of Wine Economics*, 4, 1-9.

Hodgson, Robert T. (2009b) How Expert are "Expert" Wine Judges? *Journal of Wine Economics*, 4, 233-241.

Hulkower, Neal D. (2007) The Power of None, unpublished PowerPoint charts.

Hulkower, Neal D. (2009) The Judgment of Paris According to Borda, *Journal of Wine Research*, 20, 171-182.

Hulkower, Neal D. (2011) Borda is Better, *Oregon Wine Press*, October 2011, Issue no. 308, p. 36. Available at http://oregonwinepress.com/article?articleTitle=borda+is+better--1317231621--959--guest_columns

Saari, D. G. (2000a) Mathematical structures of voting paradoxes I: pairwise vote, *Economic Theory*, 15, 1-53.

Saari, D. G. (2000b) Mathematical structures of voting paradoxes II: positional voting, *Economic Theory*, 15, 55-101.

Saari, D. G. (2001a) *Chaotic Elections! A Mathematician Looks at Voting*, Providence, RI: American Mathematical Society.

Saari, D. G. (2001b) *Decisions and Elections: Explaining the unexplained*, Cambridge, UK: Cambridge University Press.

Saari, Donald G. (2008) *Disposing Dictators, Demystifying Voting Paradoxes, Social Choice Analysis*, New York: Cambridge University Press.

Saari, D. G. and Sieberg, K. K. (2004) Are partwise comparisons reliable?, *Research in Engineering Design* 15, 62–71.

Taber, George M. (2011) *A Toast to Bargain Wines: How innovators, iconoclasts, and winemaking revolutionaries are changing the way the world drinks*, New York: Scribner.