A New Method to Obtain a Consensus Ranking of a Region's Vintages' Quality

José Borges, António C. Real, J. Sarsfield Cabral and Gregory V. Jones

Journal of Wine Economics / Volume 7 / Issue 01 / May 2012, pp 88 - 107
DOI: 10.1017/jwe.2012.7, Published online: 31 July 2012

Journal of Wine Economics, Volume 7, Number 1, 2012, Pages 88–107
doi:10.1017/jwe.2012.7

A New Method to Obtain a Consensus Ranking of a Region's Vintages' Quality*

José Borges[a], António C. Real[b], J. Sarsfield Cabral[c] and Gregory V. Jones[d]

Abstract

An impartial assessment of the quality of the wine produced over the years in a region (vintage quality) is an essential tool for producers, consumers, investors, and wine researchers to understand factors influencing quality and make purchasing or investing decisions. However, scoring the overall wine quality over the years does not necessarily produce a consensus of which year or years are best. Several critics, magazines, and organizations publish vintage charts that assign a score to each vintage, representing the corresponding perception of the wine quality. Often, the scores given by different institutions reveal little consensus with respect to the relative quality of the vintages.

In this work, we propose the utilization of a rank aggregation method to combine a collection of vintage charts for a region into a ranking of the vintages that represents the consensus of the input vintage charts. As a result, we obtain an impartial ranking of the vintages that represents the consensus of an arbitrary number of independent vintage charts. We illustrate the method with the scores from three wine regions.

The proposed method produces a ranking of vintage-to-vintage quality that represents an impartial consensus of a collection of independent sources, each using a different rating format, scale, or classification. Such a ranking has the potential to be useful for the research community, which needs a relative measure of wine production quality over the years. Therefore, we make publicly available a software tool that implements the method (Borges, 2011). (JEL Classification: C38, C61, C88)

*We thank Associação para o Desenvolvimento da Viticultura Duriense (ADVID), which partially supported the work. The authors also thank the editor, Karl Storchmann, and the anonymous reviewer for the constructive comments that helped improve the manuscript.
[a] INESC TEC, Faculty of Engineering, University of Porto, Portugal, Rua Dr. Roberto Frias, 4200–465 Porto, Portugal, email: jlborges@fe.up.pt
[b] INESC TEC, Faculty of Engineering, University of Porto, Portugal, Rua Dr. Roberto Frias, 4200–465 Porto, Portugal, email: acreals@gmail.com
[c] INESC TEC, Faculty of Engineering, University of Porto, Portugal, Rua Dr. Roberto Frias, 4200–465 Porto, Portugal, email: jacabral@fe.up.pt
[d] Southern Oregon University, Department of Environmental Studies, 1250 Siskiyou Boulevard, Ashland, OR 97520, email: gjones@sou.edu

1. Introduction

Wine production is a highly variable agricultural endeavor with yield and quality variations driven largely by climate (Jones and Davis, 2000). Understanding vintage quality variability and its influences are important in the economic sustainability of producers, consumer purchasing decisions, investor portfolio holdings, and researchers examining the myriad drivers of quality. However, the process of finding an adequate measure of vintage quality is a challenging task due to the paucity of information and the inherent subjectivity in assessing quality. One option is to use the yearly vintage charts published by internationally recognized critics, magazines, or organizations, which compare and contrast wines from different properties, different regions, or different vintages. Examples of influential vintage charts include the *Wine Spectator* Vintage Chart (Spectator, 2011) or the *Michael Broadbent*'s Vintage Wine Companion (Broadbent, 2007). A vintage chart assigns a score to each year representing an overall rating for the quality of the vintage typically for an entire region or a category of wine (i.e., red or white), and, in general, no information is provided on the scores given by the tasting panel to the individual wines tasted.

Vintage ratings have been used in numerous studies examining a wide range of economic, consumer, and scientific topics. For example, Wanhill (1995) found that vintage ratings were significant predictors of the **hammer prices** for a collection of vintages from a single chateau in Bordeaux. Other research by Landon and Smith (1998) examining Bordeaux wines found that reputation far exceeds current quality (ratings) in terms of the price premium achieved. Also in Bordeaux, Jones and Storchmann (2001) found differences in the sensitivity of ratings between cultivars with Cabernet Sauvignon fruit quality being more influential than Merlot on ratings. Schamel and Anderson (2003) found that regional reputations in Australia and New Zealand have become increasingly differentiated over time and that vintage ratings by James Halliday and *Winestate* magazine have a highly significant effect on the price premium paid by consumers. Exploiting a delay in the publishing of the ratings by Robert Parker in 2003, Ali *et al*. (2008) estimated the "Parker effect" to be 2.80 euros per bottle for Bordeaux wines. Gergaud and Ginsburgh (2008) studied the role of technology and terroir in wine quality, finding that technological choices in wine production affect quality much more than natural endowments (e.g., aspects of terroir). In addition, Gokcekus and Nottebaum (2011) found that consumer scores on wine quality tend to correlate more highly with certain experts, but that the correlation between consumer scores and expert ratings are less than those observed between the expert ratings alone.

A large area of study in vintage ratings includes the relationship between weather and climate. Research using vintage ratings has found that they accurately reflect the weather factors long known to determine wine quality and ultimately influences market or futures prices (Ashenfelter *et al*., 1995; Ashenfelter and Jones, 2012; Corsi and Ashenfelter, 2001). Examining numerous regions and chateaux in

Bordeaux, Jones and Davis (2000) found strong connections between climate, grapevine phenology, fruit composition, and vintage ratings during 1952–1997. Examining vintage ratings for Napa Valley, Nemani *et al*. (2001) also showed how *Wine Spectator* ratings affect price, with an average rating increase of 10 points (on a 0 to 100 scale), resulting in a 220% increase in price per bottle for the 1995 vintage. Furthermore, Jones *et al*. (2005) found that vintage ratings for 27 wine regions worldwide have shown trends of increasing overall quality with less vintage-to-vintage variation and that vintage ratings are strongly correlated with growing season temperatures. In 24 wine regions in Australia, Sadras *et al*. (2007) also found that higher vintage ratings and a reduction in vintage-to-vintage quality variability were related to temperatures during the growing season. While varying some by the region and wine type, the average marginal effect of growing season temperatures shows that a 1.0 °C warmer **vintage** results in an average 13-point increase (Jones *et al*. 2005). Examining climate variability mechanisms, Jones and Goodrich (2008) found significant variability in *Wine Spectator* vintage-to-vintage ratings and that much of it could be explained by conditions in the El Niño–Southern Oscillation and the Pacific Decadal Oscillation.

The analysis of such vintage charts reveals, however, that there is no widespread consensus on the vintage quality of a given region over the years. Each publisher has its own tasting panel, with its own criteria and perception of quality, which tastes a different set of wines, at different times and under different conditions. In addition, a variety of rating scales are used. Although some publishers use a 5-star rating scale, others use a 10-point or 20-point scale, and still others a 100-point rating scale (Cicchetti and Cicchetti, 2009). The difficulty of combining the judgment of several vintage charts is even greater in view of the fact that some publishers use the same rating scale but with different criteria. For example, both the *Wine Spectator* Vintage Chart (Spectator, 2011) and the *Robert Parker Vintage Guide* (Parker, 2011) use a 100-point scale in which ratings below 50 are not considered. However, while the former splits the top half of the scale into 7 intervals, the latter splits the same top half of the scale into 6 intervals (see Cicchetti and Cicchetti, 2009). As a result, for the *Wine Spectator*, 95 points correspond to a rating in the top tier while for the *Wine Advocate* the same rating is in the second tier. Therefore, combining the ratings provided by a set of vintage charts into a single absolute score that represents the production quality of a vintage is a process that has to be based on a set of questionable and arbitrary assumptions. Such assumptions are necessary to define the process of converting every rating scale into a common range of values. Also, it would be difficult to generalize such a process to an arbitrary collection of vintage charts.

In order to assess the degree of consensus among the ratings provided by a set of vintage charts, Table 1 gives the correlation coefficient for the scores given by several vintage charts for the three wine regions that we use to illustrate our method (DC: Decanter; (Decanter, 2011); WS: *Wine Spectator* (Spectator, 2011); WA: *Wine Advocate* (Parker, 2011); VC: Vintages (Spirits, 2011); AB: Addy Bassin's (AB)

Table 1
Correlation Coefficients for the Scores Given by Several Publishers to Three Wine Regions

Piedmont 1985–2006					
	DC	**WS**	**WA**	**VC**	**AB**
DC	1.00	0.77	0.76	0.77	0.84
WS		1.00	0.95	0.84	0.93
WA			1.00	0.90	0.89
VC				1.00	0.88
AB					1.00

White Burgundy 1982–2005						
	DC	**WS**	**WA**	**VC**	**AB**	**MB**
DC	1.00	0.80	0.61	0.78	0.73	0.53
WS		1.00	0.68	0.75	0.77	0.59
WA			1.00	0.47	0.50	0.52
VC				1.00	0.80	0.53
AB					1.00	0.62
MB						1.00

Champagne 1982–2003				
	DC	**WS**	**VC**	**MB**
DC	1.00	0.17	0.59	0.54
WS		1.00	0.48	0.52
VC			1.00	0.79
MB				1.00

(DC: Decanter; WS: *Wine Spectator*; WA: *Wine Advocate*; VC: Vintages; AB: Addy Bassin; MB: Michael Broadbent).

(MacArthur, 2011); MB: Michael Broadbent's (Broadbent, 2007)). For the sake of this example, the correlations were calculated with the original scores, that is, without normalizing the scores. The results show that for the Piedmont region, the correlations vary between 0.76 and 0.95, for white Burgundy between 0.47 and 0.80, and for Champagne between 0.17 and 0.79. The higher the correlation coefficient, the higher is the consensus among publishers; the results show that in some cases, the consensus is low.

Therefore, we propose the use of a rank aggregation method to combine a collection of vintage chart ratings into a ranking of the vintages that represents the consensus of the input vintage charts. The method takes advantage of the information available from a set of independent sources and combines it into an impartial ranking of a region's vintages over the years. The resulting ranking provides a relative measure of a given region's vintage quality. The method is general in the sense that it can be used with an arbitrary set of distinct input vintage charts, each having its own ordinal rating scale. We illustrate the method with the scores given by up to six different vintage charts to three different wine regions.

Several papers have been published using rank aggregation methods to study wine classifications. For example, Balinski and Laraki (2011) proposed the Majority Grade method, which can be used to induce a ranking of the wines tasted by a given panel of judges using the same classification language. In our context, this requirement is not always met, because several vintage chart publishers use different rating scales. Another example is the work by Quandt (2006) that proposes a method based on statistical tests to analyze the significance of group tasting preferences, expressed as a consensus ranking, obtained by ranking individual rank sums, which were obtained by transformation of a judge's grades into ranks.

However, the resulting ranking induced by the sum of the ranks does not respect the Condorcet property.

The method we present in this work has the advantage of making use of the information available in the form of vintage charts for a given wine region, each potentially using a different rating scale. Thus, we believe that the proposed method has the potential to be a useful tool for researchers who need an impartial measure of the wine production quality for a given region over the years, and, therefore, we make publicly available a software tool that implements the method (Borges, 2011).

The rest of the paper is organized as follows. In the next section, we describe the proposed method; in the subsequent section, we illustrate its utilization with data from three wine regions; and in the final section, we discuss the results and present our concluding remarks.

2. Materials and Methods

We propose a method that combines a set of input vintage chart ratings for a wine region into a ranking of the vintages over the years. The combined ranking gives an ordering of the vintages' quality that represents the consensus of the ratings given by the set of publishers' vintage chart.

We describe the method by means of an example. In Table 2a, we give the scores for white wines from the Burgundy region between 1983 and 1988 according to three publishers, Decanter (DC) (Decanter, 2011), the *Wine Spectator* (WS) (Spectator, 2011), and the *Wine Advocate* (WA) (Parker, 2011). The analysis of the scores reveals that the three publishers give the top score among the six years to the 1985 vintage. Also, the DC gives an identical score to the 1986 vintage, the WS gives the second-best score to 1986, and WA gives that vintage the third-best score . Thus, we can say that there is a consensus regarding the best year but not regarding the second-best year.

2.1 *Converting the Vintage Charts' Scores into Rankings*

The goal of the method is to induce a relative measure of the vintages' quality that takes into account the information given by the publishers in an impartial way. One possibility could be to convert the scores given by the publishers into a common scale and to compute the average score. Such a process would require some undesirable and arbitrary assumptions on which scores in one scale correspond to which scores in another scale. We propose to convert the scores given by each publisher into a ranking of the vintages. Therefore, for each publisher we construct a ranking of the years in which the ranks represent their preferences with respect to the vintages' quality, originally expressed as a score. The year with the highest score

Table 2
An Illustration of the Conversion of the Vintage Chart Scores (a) into Ranks with the Scores for White Wines from the Burgundy Wine Region (b)

(a) The Vintage Chart Scores

	DC	WS	WA
1983	3	85	85
1984	2	78	
1985	4	94	89
1986	4	92	82
1987	2	84	79
1988	3	86	82

(b) The Rankings Corresponding to the Scores

	DC	WS	WA
1983	3	4	2
1984	5	6	6
1985	1	1	1
1986	1	2	3
1987	5	5	5
1988	3	3	3

Note: Acronyms are as given in the note to Table 1.

is assigned the top rank and the year with the lowest score the bottom rank. If the same score is assigned to two different years, such years are assigned the same rank (e.g., 1985 and 1986 from DC in Table 2). We note that the rank of a given year gives the number of years that are better than it plus one. Thus, because in the ranking for DC two years are tied for first place, the number two is not assigned to any year, while years 1983 and 1988 are tied for third place. The option to give years the same ranking instead of the common practice of adopting the average rank is due to the more natural interpretation in our particular context of having two years tied for first place than of having two years tied for 1.5th place. However, the proposed method can be modified to adopt the average rank without any loss of generality.

As a result of this first step, we obtain a set of input rankings such that each ranking represents an ordering of the vintages' quality over the years as perceived by the corresponding publisher's tasting panel. Table 2b gives the rankings of the years corresponding to the scores given by each of the publishers.

The only assumption of this first step is that each vintage chart's publisher uses, at least, an ordinal scale and has evaluating criteria that remain stable over the years, in such a way that it is possible to compare the perceived quality of the vintages by comparing their scores. We believe that this is a reasonable assumption.

We also note that vintage charts occasionally have missing values. Missing values occur if there are years for which a score is not provided. Several methods are available to deal with missing values. One option is to assume that if a provider

did not rate a given year it was because it was decided that the perceived quality of the corresponding harvest was not sufficiently good to justify the tasting. In this case, such years are assigned the bottom rank. In the context of vintage charts, such an assumption is often reasonable because when some vintages are perceived as uninteresting at early stages, the publishers decide that its quality does not justify the effort associated with its evaluation. For example, the description of the *Wine Advocate* vintage chart, (Parker, 2011) refers to the fact that wines with a score below 50 are not reviewed, and *Vintages* (Spirits, 2011) states that wines with a score below 4 are not rated. In the example of Table 2, WA did not provide a score for the year 1984. A closer look shows that 1984 gets low scores from the other two publishers, and therefore it is reasonable to assign the bottom rank to 1984.

In cases in which the assumption of a missing year being the lowest quality is not defensible, traditional methods can be used to deal with missing values. More precisely, if most of the publishers do not provide a score for a particular year that year can be removed from the analysis, because there is insufficient information to rank that year. In the case in which a publisher does not provide a score for most of the years, this publisher could be removed because it does not provide sufficient information for the method. Finally, there is the possibility of filling one particular missing value with the average ranking of the other available publishers for that particular year. In this case, it may be necessary to re-rank the years below the year that was filled with the average rank. We note, however, that the method we propose is able to handle any of the above options without any loss of generality.

2.2 *Aggregating the Input Rankings into a Consensus Ranking*

The rank aggregation problem is defined as the task of combining many different rank orderings into the ranking that is closest to the set of input rankings (Lin, 2010). This is a classic problem from voting theory that has gained interest recently because of its application to the problem of combining the search results of a collection of web search engines.

In his work, Arrow (1950) has shown that when voters have more than two alternatives, it is not possible to convert the individual rank preferences into a ranking that meets a set of reasonable axioms. Some methods are available for combining ranks that meet some of the axioms.

A simple way to approach the rank aggregation problem is to assign each vintage year a number of points equal to its rank in the input ranking of each publisher. The resulting total score of a vintage year is the sum of the individual ranks, thus, the vintage years can then be ranked by their total scores. This method is equivalent to computing the average ranking, however, the average of the ranks is known for not respecting the Condorcet property (Condorcet, 1785), as described in the following example.

Table 3
An Example to Illustrate the Limitations of Using the Average Rank to Induce a Ranking of the Years

(a) The Judges Rank Preferences and the Average Rank							(b) Average Ranking
	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	Avg. rank	Resulting ranking
Year 1	1	3	1	5	2	2.4	Year 1
Year 2	2	1	6	1	5	3.0	Year 3
Year 3	5	2	2	4	1	2.8	Year 2
Year 4	6	6	4	3	6	5.0	Year 6
Year 5	4	4	5	6	3	4.4	Year 5
Year 6	3	5	3	2	4	3.4	Year 4

The Condorcet property states that if an alternative X is preferred by more judges than the alternative Y, in the optimal ranking the alternative X should be ranked higher than alternative Y. In Table 3, we give an example to illustrate that the average ranking does not respect this property. In Table 3a, six years are ranked by five distinct judges, and the average rank for each of the years is given. In Table 3b, the ranking of the years when ordered by the average ranks is given. In the resulting ranking, year 3 is preferred to year 2 because it has a lower average for the ranks. However, the inspection of the ranks provided by the judges reveals that three out of the five judges preferred year 2 instead of year 3, thus, according to the Condorcet property year 2 should be preferred to year 3.

One alternative to the average ranking could be the majority grade method proposed by Balinski and Laraki (2011), in which the goal is to find a final grade for each element, given the individual messages of all judges, that corresponds to the highest grade approved by an absolute majority of the judges. This method,

however, has the limitation in our context of relying on a language of evaluation that is common to all the judges of a jury conducting the evaluation. That is not the case of vintage chart ratings provided by a set of different publishers.

Here, we propose the use of a method that meets the Condorcet criterion (Condorcet, 1785). Methods based on the Condorcet criterion rank each candidate by measuring the number of competitors that would be beaten by it in a two-candidate election.

In order to respect the Condorcet property, the rank aggregation problem has been defined as the task of minimizing the number of pairwise disagreements between the input rankings and the resulting ranking (Kemeny, 1959). A disagreement between an input ranking and the final ranking occurs if two years are ranked in a different order in the two rankings. For example, if one of the rankings prefers 1986 to 1983 and the other ranking prefers 1983 to 1986, we say the two rankings disagree with respect to the pair 1983 and 1986. This formulation is known as the Kemeny rank aggregation (Young, 1988), and it has been shown to verify the Condorcet property.

We now give a definition of the rank aggregation problem in our context. Consider a set of input rankings from K publishers, $R_1, \ldots, R_K$, each with an ordering of the same n elements, and the resulting combined ranking R_c. We let $R_k(i)$ for $i{:}1, \ldots, n$ be the rank of according to the ranking provided by publisher k and $R_c(i)$ be the ranking of according to the combined ranking. The distance between an input ranking and the combined resulting ranking is given by the Kendall-tau metric, which is defined as:

$$K(R_k,\, R_c) = \sum_{i=1}^{n} \sum_{j=1}^{n} \bar{K}_{i,j}(R_k,\, R_c)$$

where $\bar{K}_{i,j}(R_k,\, R_c) = 0$ if i and j are in the same order in R_k and R_c and $\bar{K}_{i,j}(R_k,\, R_c) = 1$ if i and j are in the opposite order in R_k and R_c.

In the case in which there are no ties in the rankings, the Kendall-tau metric can be normalized to the [0,1] interval by dividing it by $n(n-1)/2$, in which 1 implies maximum disagreement. The rank aggregation problem seeks to minimize the normalized measure (Adler, 1957).

Because the problem is NP-hard, there is no algorithm for finding the optimal solution in polynomial time (Schalekamp and Zuylen, 2009). However, several algorithms are available for finding close to optimum solutions. To illustrate the process, we adopt the Quicksort with local search optimization approach described in Schalekamp and Zuylen (2009).

In the first step of the algorithm, a matrix of weights w, is defined such that $w_{i,j}$, with $i, j{:}1, \ldots, n$, gives the number of input rankings that prefer i to j. In Table 4 we

Table 4
The *w* Matrix of Weights Corresponding to the Example in Table 2

	1983	*1984*	*1985*	*1986*	*1987*	*1988*
1983	0	3	0	1	3	1
1984	0	0	0	0	0	0
1985	3	3	0	2	3	3
1986	2	3	0	0	3	2
1987	0	2	0	0	0	0
1988	1	3	0	0	3	0

Table 5
The Resulting Combined Ranking Obtained by Application of the Proposed Method

Year	R_c *(combined rank)*
1985	1
1986	2
1988	3
1983	3
1987	5
1984	6

give the *w* matrix corresponding to the example in Table 2. Each cell of the matrix gives the number of publishers that prefer the row-year to the column-year. For example, row 1985 and column 1983 have the number 3, which means that three publishers prefer 1985 to 1983.

In the second step of the process, a classic Quicksort algorithm (see Hoare, 1962) is applied that sorts elements in the ranking in such a way that i is preferred to j when $w_{ij} \geq w_{ji}$ and j is preferred to i when $w_{ij} \geq w_{ji}$.

The final step of the algorithm is a local search that consists of swapping pairs of elements in the ranking given by the Quicksort step and verifies whether it improves the Kendall-tau metric. Because the Quicksort algorithm has a random nature, the algorithm should be run several times and the best solution is chosen according to the Kendall-tau metric.

The final combined ranking (R_c) of the years obtained by the proposed method for the running example is given in Table 5.

The method can be adapted to incorporate weighting schemes if there is a need to differentiate the relative importance attributed by the publishers. That can be done by modifying the way in which the average Kendall-tau metric is computed. For example, if, for each of the K publishers, we define the weights $p_1, \ldots p_K$

such that $\sum_{l=1}^{K} p_l = 1$, the average pairwise disagreement can be computed as follows:

$$\sum_{l=1}^{K} (p_1 \cdot K(R_l, R_c)/(n(n-1)/2)).$$

The method is applied in a predefined time window. If the analyst needs to extend the time window, for example, by including more recent years, the overall ranking has to be recomputed. In such a case, the ranks of the previously existing years might be altered. However, the Condorcet property ensures that the relative positioning of such years is maintained. Finally, we note that the algorithm described above was chosen for its simplicity, but any of the rank aggregation algorithms described in Schalekamp and Zuylen (2009) can be used for the second step of the proposed method.

3. Results

In this section, we illustrate the application of the method with data for three wine-producing regions: Piedmont, white Burgundy, and Champagne. The regions were chosen in order to illustrate the method when using data sets with different proportions of missing values. For each wine region, we use up to six publicly available vintage charts:

- *Decanter Vintage Guide (DC)* (Decanter, 2011),
- *Wine Spectator vintage chart (WS)* (Spectator, 2011),
- *Wine Advocate 1970–2008 vintage guide (WA)* (Parker, 2011),
- *Vintages vintage chart (VC)* (Spirits, 2011),
- *Addy Bassin's vintage chart (AB)* (MacArthur, 2011),
- *Michael Broadbent's pocket vintage wine companion (MB)* (Broadbent, 2007).

The method is generic in the sense that an arbitrary number of publishers using any ordinal rating scale could be included. For this experiment, we considered that all vintage charts are equally important and, therefore, are assigned the same weight (as referred to in Section 2, the input scores could be weighted according to the publishers' importance, as perceived by the analyst).

Table 6 gives the scores and the corresponding ranks according to five publishers for the Piedmont Italian wine region between 1985 and 2006. The first publisher (DC) uses a 5-point scale for the scores, the second publisher (VC) uses a 10-point scale, and the last three publishers (WS, WA, and AB) use a 100-point scale. It is interesting to note that for this particular region all publishers provide a score for every vintage year in the selected time window, that is, there are no missing values.

In order to apply our method, the scores were converted into rankings, which are given in the bottom half of Table 6. As described in Section 2, for a given source, the rank of a year gives the number of years that have a higher score than itself plus one. Therefore, years with rank 1 are the years that were given the highest score by the corresponding vintage chart.

The analysis of the rankings for the Piedmont region reveals a relatively high consensus for the worst vintage years (1991 and 1992). From the ranking in Table 6, 1992 is clearly the worst year, as three of the publishers assign it the bottom rank and the other two publishers assign the penultimate rank. With respect to the best years, there is some consensus. On the one hand, WS gives 100 points to the 2000 vintage, while the other publishers do not rank that same year in the top four. On the other hand, both 1989 and 1990 are given the top rank by three publishers. This is a good illustration of the difficulties in combining the ratings from different publishers.

We have applied the method to the scores of the Piedmont region in order to induce the combined consensus ranking. The Quicksort with a local search algorithm was run 500 times, and the best solution was chosen. Table 7 gives the resulting combined ranking (Rc) for the Piedmont region. The Kendall-tau value for the solution is 0.160. The Kendall-tau metric is a value between 0 and 1, which measures the average distance between the input rankings and the combined ranking. When the value is 0, there is a total agreement.

As a result of the application of the method, we conclude that, according to the combined scores of five independent vintage charts, the years 1989, 1990, and 1996 are the three indistinguishable best vintages for the Piedmont region between 1985 and 2006. Similarly, the quality of the 1997 vintage is equivalent to the quality of the 2000 vintage, and they are both better than the 2001 vintage.

In order to illustrate another property of the method, Table 8 gives the combined ranking for the Piedmont region when only the vintages from the decade of 1990 are considered. It is important to note that the relative ordering of the vintages is maintained when compared to the results given in Table 7. For example, according to the ranking given in Table 7, 1996 is ranked above 2000, and this is verified in the ranking given in Table 7. Also, 1997 and 2000 are rank-tied in both rankings.

Table 9 gives the scores and corresponding ranks for white wines from the Burgundy wine region according to six different vintage charts. For this particular region, there are four missing values.

Table 10 gives the resulting combined consensus ranking that represents the ordering of the vintages by their quality according to the six vintage chart scores when the missing values are assigned the bottom rank. In this case, the assumption seems reasonable for the missing values in 1984 and in 1987, because the scores given by the other publishers are very low, but not reasonable for the missing value in 1989, because the other publishers give good scores to this year. The Kendall-tau

Table 6

The Scores for the Piedmont Wine Region and the Corresponding Ranks

	Year	*85*	*86*	*87*	*88*	*89*	*90*	*91*	*92*	*93*	*94*	*95*	*96*	*97*	*98*	*99*	*00*	*01*	*02*	*03*	*04*	*05*	*06*
Scores	DC	5	3	2	4	5	5	2	1	2	2	2	5	5	5	3	4	4	2	4	4	2	5
	VC	9	8	7	8	10	10	5	5	8	7	9	10	9	9	9	9	9	7	8	10	8	9
	WS	94	86	83	92	97	97	77	76	87	77	88	98	99	93	90	100	95	72	88	93	94	89
	WA	95	89.5	86	90	97	96	76	74	90	77	87	97	93	92	93.5	95	96	75	89.5	94.5	93	91.5
	AB	92	80	86	92	97	98	78	76	87	80	88	96	97	93	90	94	92	80	88	91	88	90
Ranks	DC	1	13	15	8	1	1	15	22	15	15	15	1	1	1	13	8	8	15	8	8	15	1
	VC	5	13	18	13	1	1	21	21	13	18	5	1	5	5	5	5	5	18	13	1	13	5
	WS	7	17	18	11	4	4	19	21	16	19	14	3	2	9	12	1	6	22	14	9	7	13
	WA	5	15	18	13	1	3	20	22	13	19	17	1	9	11	8	5	3	21	15	7	9	12
	AB	7	18	17	7	2	1	21	22	16	18	13	4	2	6	11	5	7	18	13	10	13	11

Note: Acronyms are as given in the note to Table 1.

Table 7

The Combined Consensus Ranking for the Piedmont Wine Region

Year	*90*	*89*	*96*	*97*	*00*	*01*	*85*	*04*	*98*	*88*	*99*	*06*	*05*	*03*	*95*	*93*	*86*	*87*	*94*	*02*	*91*	*92*
R_c	1	1	1	4	4	6	7	8	8	10	11	12	13	14	15	16	17	18	19	20	20	22

Table 8
The Combined Consensus Ranking for the Piedmont Wine Region for the Years Between 1991 and 2000

Year	*96*	*00*	*97*	*98*	*99*	*95*	*93*	*94*	*91*	*92*
R_c	1	2	2	4	5	6	7	8	9	10

value is 0.229, meaning that there is a higher degree of disagreement relative to the consensus ranking for white Burgundy than for the consensus ranking for the Piedmont region (which has a value of 0.160).

As discussed in Section 2, other alternative methods could be used for dealing with missing values. One method consists of assigning to a missing value the average rank of the scores in the rankings induced by the other publishers. If this method is used, the 1989 vintage's new rank according to MB is $(1+6+5+2+1)/5=3$. In order to maintain the consistency of the MB ranking, the ranks that are below the third position and not below twenty-second position (the 1989 previous rank) are decreased by one position. For example, the rank of the 1982 vintage is changed from fourth place to fifth place. The complete updated ranking for the MB source is given in Table 11 and Table 12 gives the resulting combined ranking when this method is applied to all missing values. The Kendall-tau value is now 0.222, which is almost identical to the value obtained for the solution given in Table 10. In this case, the results are not significantly affected by the method used to deal with the missing values. This is justified by the relatively low percentage of missing values, which have a low impact due to the number of input vintage charts used.

Table 13 gives the scores according to four different publishers for the Champagne wine region. This example was chosen because of its larger percentage of missing values. Two reasons to justify the occurrence of a missing value are (1) the publisher decided not to taste the wines that year for operational reasons or (2) the crop quality was uneven, and the number of "declared" vintages was not sufficient to justify the tasting. Declaring vintages is most commonly done with Champagne and Port, for which producers declare a vintage only when the quality is extraordinary.

The method was run on this data set with both methods for dealing with the missing values. Table 14 gives the results when missing values are given the bottom rank of the corresponding source, and Table 15 the results when missing values are given the average rank of the available publishers. While the first solution has a Kendall-tau of 0.189, the second solution has a Kendall-tau of 0.150, meaning that the solution obtained with the average ranking is improved slightly. It is interesting to note that the two solutions are very close, being identical in the top 13 positions. When the average rank is used to fill the missing values, the vintage of 1992 is ranked three places higher, the vintage of 1994 is ranked two places higher and the vintage of 2001 two places higher. The other 19 vintages maintain their relative position. Finally, we note that of the 10 missing values observed in Table 13, only

Table 9

The Scores for White Wines from the Burgundy Wine Region and the Corresponding Ranks

	Year	*82*	*83*	*84*	*85*	*86*	*87*	*88*	*89*	*90*	*91*	*92*	*93*	*94*	*95*	*96*	*97*	*98*	*99*	*0*	*01*	*02*	*03*	*04*	*05*
Scores	DC	4	3	2	4	4	2	3	5	4	3	4	3	3	5	5	4	4	4	4	4	5	3	3	4
	WS	83	85	78	94	92	84	86	92	92	85	89	82	87	93	95	88	88	88	90	89	95	87	90	93
	WA	88	85		89	82	79	82	90	87	70	90	72	77	93	92	89	84	89	88	86	92	84	91	90
	VC	7	7	6	9	8	6	7	9	9	5	8	6	8	10	9	9	7	9	7	7	9	7	6	9
	AB	89	86	82	89	92	81	85	94	92	74	92	76	88	92	94	89	86	88	88	87	93	87	90	93
	MB	4	4		4	5		3		4	3	3	3	3	4	5	4	4	4	3	3.5	4	3.5	4	5
Ranks	DC	5	16	23	5	5	23	16	1	5	16	5	16	16	1	1	5	5	5	5	5	1	16	16	5
	WS	22	19	24	3	6	21	18	6	6	19	11	23	16	4	1	13	13	13	9	11	1	16	9	4
	WA	11	15	24	8	18	20	18	5	13	23	5	22	21	1	2	8	16	8	11	14	2	16	4	5
	VC	13	13	20	2	10	20	13	2	2	24	10	20	10	1	2	2	13	2	13	13	2	13	20	2
	AB	10	18	21	10	5	22	20	1	5	24	5	23	13	5	1	10	18	13	13	16	3	16	9	3
	MB	4	4	22	4	1	22	16	22	4	16	16	16	16	4	1	4	4	4	16	14	4	14	4	1

Note: Acronyms are as given in the note to Table 1.

Table 10

The Combined Consensus Ranking White Wines from the Burgundy Wine Region When the Missing Values Are Given the Bottom Rank of the Source

Year	*95*	*96*	*02*	*05*	*89*	*85*	*90*	*86*	*04*	*97*	*99*	*92*	*00*	*82*	*01*	*98*	*94*	*03*	*83*	*88*	*87*	*93*	*91*	*84*
R_c	1	2	2	4	4	6	7	8	9	10	10	12	13	13	15	16	17	17	19	20	21	22	23	24

Table 11

The Ranks Given by MB to White Wines from the Burgundy Wine Region If Missing Values Are Given the Average Rank of the Available Publishers

Year	*82*	*83*	*84*	*85*	*86*	*87*	*88*	*89*	*90*	*91*	*92*	*93*	*94*	*95*	*96*	*97*	*98*	*99*	*0*	*01*	*02*	*03*	*04*	*05*
Ranks	5	5	24	5	1	22	17	3	5	17	17	17	17	5	1	5	5	5	17	15	5	15	5	1

Table 12

The Combined Consensus Ranking for White Wines from the Burgundy Wine Region When the Missing Values Are Given the Average Rank of the Other Available Publishers

Year	*95*	*96*	*02*	*05*	*89*	*85*	*90*	*86*	*04*	*97*	*99*	*92*	*82*	*00*	*01*	*98*	*94*	*03*	*83*	*88*	*87*	*93*	*91*	*84*
R_c	1	2	2	4	4	6	7	8	9	10	10	12	13	14	15	16	17	17	19	20	21	22	23	23

Table 13

The Scores for the Champagne Wine Region

Year	*82*	*83*	*84*	*85*	*86*	*87*	*88*	*89*	*90*	*91*	*92*	*93*	*94*	*95*	*96*	*97*	*98*	*99*	*00*	*01*	*02*	*03*
DC	5	4			3		5	5	5					4	5	4	4	4	4	3	4	3
WS	94	83	79	96	86	81	95	90	97	79	84	87	82	94	96	87	91	89	89		94	88
VC	9	7	4	9	7	5	9	8	10	6	7	8	7	9	10	7	7	8	8	6	9	7
MB	5	3		5	2.5	2	5	4	5	2	4	3		4	5	4	4	3	2	1	4	2

Note: Acronyms are as given in the note to Table 1.

Table 14

The Combined Consensus Ranking for the Champagne Wine Region When the Missing Values Are Given the Bottom Rank of the Source

Year	*90*	*96*	*82*	*85*	*88*	*02*	*95*	*89*	*98*	*00*	*99*	*97*	*93*	*83*	*86*	*03*	*92*	*94*	*91*	*87*	*01*	*84*
R_c	1	2	3	3	3	6	6	8	9	10	10	12	13	14	15	15	17	18	19	19	21	22

Table 15

The Combined Consensus Ranking for the Champagne Wine Region When the Missing Values Are Given the Average Rank of the Other Available Publishers

Year	*90*	*96*	*82*	*85*	*88*	*02*	*95*	*89*	*98*	*00*	*99*	*97*	*93*	*92*	*83*	*94*	*86*	*03*	*01*	*91*	*87*	*84*
R_c	1	2	3	3	3	6	6	8	9	10	10	12	13	14	15	16	16	18	19	20	20	22

the 1985 score from DC corresponds to a vintage in the top half of the resulting combined ranking.

4. Discussion and Conclusions

Although vintage ratings like those used in this research may not be perfect, efficient measures of a year's worth of weather, farming, and winemaking, consumers have come to use these numbers as a general rule of thumb for purchasing wine. Producers have also incorporated vintage ratings into numerous aspects of the economics of their businesses (e.g., winemakers are given bonuses based upon achieving higher scores) and the marketing of their wines or regions. In addition, much important economic and scientific research is also based upon quality metrics such as ratings. Therefore, the problem of assessing the vintage quality over the years for a given wine region is an important research topic.

In this work, we propose a method that takes advantage of the numerous vintage charts published yearly by renowned wine-rating critics, magazines, and organizations. The method converts the ratings of each individual source into rankings and uses a rank aggregation algorithm to combine the input ranking into a consensus ranking. As a result, we are able to calculate a ranking of the vintages' quality that can be seen as a measure of their relative quality. The ranking represents an impartial consensus of the collection of input vintage charts, in the sense that no assumption is made on how each vintage chart was formulated. The method effectively incorporates input from numerous publishers, using different types of scoring formats (ordinal, interval, or ratio), with different scales (e.g., 0–5, 0–10, 0–100), and with different assumptions with respect to the bottom of the scale and the interval that constitutes an extraordinary wine (see Cicchetti and Cicchetti, 2009). Furthermore, the applicability of the method is illustrated with the analysis of the scores for three wine regions with different levels of missing values. However, it should be noted that a limitation of the method is that it provides only a relative measure instead of an absolute measure of the vintage quality.

In summary, we believe that the proposed method has the potential to be a useful tool for wine research that requires an impartial assessment of the vintage quality for a given wine region. Thus, we make available a software tool with the full implementation of the method and all the data files used in the tests (Borges, 2011).

References

Adler, L.M. (1957). A modification of Kendall's Tau for the case of arbitrary ties in both rankings. *Journal of the American Statistical Association*, 52, 33–35.

Ali, H.H., Lecocq, S. and Visser, M. (2008). The impact of gurus: Parker grades and en primeur wine prices. *Journal of Wine Economics*, 5, 22–39.

Arrow, K.J. (1950). A difficulty in the concept of social welfare. *Journal of Political Economy*, 58, 328–346.

Ashenfelter, O. and Jones, G.V. (2012). The demand for expert opinions: Bordeaux wine. *Cahier Scientifique de l'Observatoire des Conjonctures Vinicoles Européennes, Journal of Wine Economics*, 7(2), forthcoming.

Ashenfelter, O., Ashmore, D. and LaLonde, R. (1995). Bordeaux wine vintage quality and the weather. *Chance*, 8, 7–14.

Balinski, M. and Laraki, R. (2011). *Majority judgment: Measuring, ranking, and electing*. Cambridge: MIT Press.

Borges, J. (2011). Wine studies links. Retrieved from http://paginas.fe.up.pt/~jlborges/wine_studies.html.

Broadbent, M. (2007). *Pocket vintage wine companion*. London: Pavilion Books (in association with Christie's).

Cicchetti, D.V. and Cicchetti, A.F. (2009). Wine rating scales: Assessing their utility for producers, consumers, and oenologic researchers. *International Journal of Wine Research*, 1, 73–83.

Condorcet, M.D. (1785). *Des décisions rendues à la pluralité des voix. Essai sur l'application de l'analyse à la probabilité*. Paris: Imprimerie royale.

Corsi, A. and Ashenfelter, O. (2001). Predicting Italian wine quality from weather data and expert "ratings." Presented at the *Cahier Scientifique de l'Observatoire des Conjonctures Vinicoles Européennes*, (4).

Decanter. (2011). Vintage guide. Retrieved from www.decanter.com/wine-learning/vintage-guides/, April 2011.

Gergaud, O. and Ginsburgh, V. (2008). Natural endowments, production technologies, and the quality of wines in Bordeaux. *Does terroir matter? Economic Journal*, 118, F142–F157.

Gokcekus, O. and Nottebaum, D. (2011, September). The buyer's dilemma—Whose rating should a wine drinker pay attention to? American Association of Wine Economists, Working Paper no. 91.

Hoare, C.A. (1962). Quicksort. *Computer Journal*, 5, 10–15.

Jones, G.V. and Davis, R.E. (2000). Climate influences on grapevine phenology, grape composition, and wine production and quality for Bordeaux, France. *American Journal of Viticulture and Enology*, 51, 249–261.

Jones, G.V. and Storchmann, K.-H. (2001). Wine market prices and investment under uncertainty: An econometric model for Bordeaux crus classes. *Agricultural Economics*, 26, 115–133.

Jones, G.V. and Goodrich, G.B. (2008). Influence of climate variability on wine region in the western USA and on wine quality in the Napa Valley. *Climate Research*, 35, 241–254.

Jones, G.V., White, M.A., Cooper, O.R. and Storchmann, K. (2005). Climate change and global wine quality. *Climatic Change*, 73, 319–343.

Kemeny, J.G. (1959). Mathematics without numbers. *Daedalus*, 88, 571–591.

Landon, S. and Smith, C.E. (1998). Quality expectations, reputation, and price. *Southern Economic Journal*, 64, 628–647.

Lin, S. (2010). Rank aggregation methods. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2, 555–570.

MacArthur, A.B. (2011). Vintage guide. Retrieved April 2011, from www.bassins.com/resources/vin_chart.html.

Nemani, R.R., White, M.A., Cayan, D.R., Jones, G.V., Running, S.W. and Coughlan, J.C. (2001). Asymmetric climatic warming improves California vintages. *Climate Research*, 19, 25–34.

Parker, R. (2011, April). Wine advocate. Retrieved April 2011, from www.erobertparker.com/newsearch/vintagechart1.aspx.

Quandt, R.E. (2006). Measurement and inference in wine tasting. *Journal of Wine Economics*, 1, 7–31.

Sadras, V., Soar, C. and Petrie, P. (2007). Quantification of time trends in vintage scores and their variability for major wine regions of Australia." *Australian Journal of Grape and Wine Research*, 13, 117–123.

Schalekamp, F. and Zuylen, A.V. (2009). Rank aggregation: Together we're strong. In *Proceedings of the Eleventh Workshop on Algorithm Engineering and Experiments (ALENEX)* (pp. 38–51). New York: SIAM.

Schamel, G. and Anderson, K. (2003). Wine quality and varietal, regional and winery reputations: Hedonic prices for Australia and New Zealand. *Economic Record*, 79, 357–369.

Spectator, W. (2011). Vintage chart. Retrieved April 2011, from www.winespectator.com/servefile/serve/id/41618

Spirits, V.F. (2011). Vintage guide. Retrieved April 2011, from www.vintages.com/circular/vintage-chart.pdf

Wanhill, S.R.C. (1995). A test of claret price formulation in auction markets. *International Journal of Wine Marketing*, 7, 17–22.

Young, H. (1988). Condorcet's theory of voting. *American Political Science Review*, 82, 1231–1244.