AMERICAN ASSOCIATION OF WINE ECONOMISTS

AAWE WORKING PAPER
No. 180
Economics

REDUCING QUALITY UNCERTAINTY FOR BORDEAUX *EN PRIMEUR* WINES: A UNIFORM WINE SCORE

Jean-Marie Cardebat and
Emmanuel Paroissien

June 2015

www.wine-economics.org

Reducing Quality Uncertainty for Bordeaux *en primeur* Wines: A Uniform Wine Score

Jean-Marie Cardebat
Larefi, University of Bordeaux – INSEEC
Bordeaux Wine Economics
jean-marie.cardebat@u-bordeaux.fr

Emmanuel Paroissien
Larefi, University of Bordeaux – CREST
Bordeaux Wine Economics
emmanuel.paroissien@u-bordeaux.fr

Abstract

In this paper we provide a simple and transparent non parametric methodology to express the scores of each wine expert (15) on the same rating scale. We discuss the advantage of this methodology over a linear transformation. The non paramatric method ensures the comparability of scores among experts and allows for a relevant average calculation of available wine scores. This approach may be usefuel to wine professionals who seek to reduce uncertainties leading to improved market efficiency. Uniform scores for many Bordeaux *en primeur* wines can be freely accessed at **globalwinescore.com**.

Key words: Wine, quality, experts.

1. Introduction

As an experience good, the quality of a wine is only known after its consumption. This information asymmetry justifies the emergence of wine experts providing information on wine quality. The contingent information market is particularly well-developed in the wine sectors where numerous experts coexist. The subjectivity of the wine quality assessment, the regional segmentations[1] or their (supposed) preferences (Storchmann, 2012) partly justify this high number of experts. Moreover, the grading systems and habits could differ from one expert to another. In particular, the European experts are used to rating wine on a 20-point scale whilst US experts use 100 points (Masset and Weisskopf, 2015). The heterogeneity of the rating systems can increase the consumer's perceived uncertainty. The question of the rating homogenization on the same scale of preferences is therefore at the heart of the uncertainty debate about wine quality.

This uncertainty on wine quality is particularly high during the *en primeur* campaign in the Bordeaux Region. The *primeur* market can be seen as a forward market dedicated to fine Bordeaux wines. The *en primeur* campaign takes place during the spring, starting with a huge multi-day tasting organized by the chateaux in the first week of April. Wine merchants, wine enthusiasts and of course wine experts are involved in this event. They all taste the wine from the latest harvest. Therefore, the wine is not yet vinified and the quality assessment is particularly difficult and uncertain. The aim of this campaign is to sell (chateaux) and buy (wine merchants)[2] before the wine is effectively released in bottles (about 18 months later). The prices and quantities exchanged are determined during the *en primeur* campaign and the wine will be delivered once it is available (in a bottle).

The economic stakes of the tasting are therefore extremely high because the prices and quantities exchanged are influenced by the experts' scores. The wine economics literature has provided ample evidence of the link between *en primeur* wine prices and the experts' scores (see notably Hadj Ali and Nauges, 2007; Hadj Ali et al., 2008; Masset et al., 2015). Another strand of the literature deals with the information contained in the experts' grades (see for example Ashenfelter et al., 1995, Ashenfelter, 2008, or more recently Cardebat et al., 2014), the divergence between experts (notably Ashton, 2012, 2013; Hodgson, 2008; Masset et al., 2015; Olkin et al., 2015) or the randomness of the tastings (Quandt, 2007; Bodington, 2015).

However, no paper has tried to express the experts' scores on the same scale of preference or in the same rating system before analyzing the grades divergence or bias or impact on prices. As noted by Masset et al. (2015, p.80) "Comparisons are difficult to make, as not all experts use the same scale to establish their scores". Furthermore, as far as we know, there is no paper trying to provide a uniform score aggregating all the marks released by experts during the *en primeur*

[1] We mean by regional segmentation the fact that not only are certain experts more or less specialized in wines coming from specific regions, but also that some experts target specific consumers (at least as regards the choice of the language in which they edit their comments).

[2] The wine merchants (called *negociants* in Bordeaux) are free to buy or not, but they receive allocations (the right to buy in a certain amount) from the chateaux and if they not buy a specific year, the chateaux can take back their allocations for the following year.

campaign, although a demand exists for such a uniform score from the professionals. However, if no academic papers exist, in the wine industry, most of the web merchants provide such aggregated scores (see, for example, wine decider or wine searcher). The website of Bertrand Leguern is also dedicated to the calculation of an aggregated score which is used by wine professionals. Nevertheless, we cannot find any information on the way these scores have been calculated. There is no transparency in their calculation, thereby reinforcing the information asymmetry instead of reducing it.

Wine professionals, mainly the *negociants* who buy the *en primeur* wines, request aggregated and transparent information on wine quality rather than numerous grades emanating from a variety of experts. What is at stake is to reduce the information asymmetry and therefore increase the *en primeur* market efficiency (Mahenc and Meunier, 2006). This request is particularly important in a context of the retirement of the main expert, Robert Parker. The disappearance of this expert will reinforce the uncertainty and the need for a reference score.

The aim of this paper is therefore to suggest a relevant methodology for calculating a single uniform score aggregating the grades released by 15 experts who have traditionally intervened on this market since the beginning of the last decade. Based on an exhaustive database of *en primeur* Bordeaux wines experts' scores, we propose a methodology to translate the rating scale of one expert into the rating scale of another, thereby ensuring the comparability of all the experts' scores. The uniform score is then basically calculated as a simple arithmetic average of these transformed scores. By its construction, this uniform score has the potential to be considered as a new reference score on the fine wine market.

The interest of this study is twofold. First, academics will benefit from a methodology ensuring proper experts' scores comparison by taking into account the different rating systems existing among experts. Secondly, based on this methodology, we provide wine professionals with a unique global wine score aggregating the information coming from all the experts operating on the *en primeur* market. **Based on the methodology presented in this paper, the most recent uniform wine scores can be freely accessed at globalwinescore.com.**

The remainder of this paper is structured as follows: the next sections present our dataset, while section 3 displays the methodology of the global wine score; section 4 displays the uniform scores and discusses the results following different robustness checks; the last section concludes.

2. Data

Our exhaustive dataset contains the scores given by 15 well-known wine experts[3] during the *en primeur* campaign over the period 2000-2014. All the wines rated by these experts are present in

[3] The term "expert" is used here indifferently to designate a person (James Suckling, Jancis Robinson, etc.) or an organization (i.e. a journal like Wine Spectator or La Revue des Vins de France – RVF, etc.). Decanter has a special status in the sense that we split its scores into two categories: Decanter 20 and Decanter 100. This case is very interesting because Decanter chose to change its traditional 20-point scale for a 100-point scale during the period studied. We have therefore decided to consider its scores on 20 points and 100 points as two different experts. The following statistics will be made on the basis of 13 experts.

the dataset. It represents 447 chateaux and 4333 chateau-vintage pairs, meaning than on average, each chateau is rated 9.7 times over the period.

The first column in Table 1 shows the number of wines effectively rated by each expert. Rene Gabriel appears as the most productive expert with 3639 scores over the period. Six experts display similar profiles with a high activity on the wine opinion market. They all have rated more than 2000 *en primeur* wines between 2000 and 2014. On the contrary, the last four experts of this list exhibit a significantly weaker activity with less than 500 scores. The following columns display the traditional descriptive statistics on the experts' scores. Among the 16 (15+1, see note 1) experts, seven use a 20-point grading scale, they are all European, and nine use a 100-point scale, they are overwhelmingly American, the exceptions coming from the Chinese J. Cho Lee and the British Tim Atkin.

Table 1: Descriptive statistics on experts' scores

	Freq.	**Min**	**Max**	**Mean**	**Med.**	**Std dev.**
Rene Gabriel	3639	12	20	17.12	17	1.14
Wine Spectator	2886	77	98.5	90.2	90	3.5
Robert Parker	2609	71.5	99.5	90.4	90.5	3.5
Jancis Robinson	2538	12	20	16.4	16.5	1.0
Jacques Dupont	2156	13	20	15.8	16	1.3
Bettane&Desseauve	2113	10	20	16.6	16.5	1.3
Neal Martin	1711	70	99	90.0	90	3.5
Decanter20	1615	14.5	20	16.9	17	1.0
Jean-Marc Quarin	1497	10	20	15.74	15.75	1.07
James Suckling	1059	84.5	100	91.3	91.5	2.7
Decanter100	1026	81	95	88.2	88	2.9
Tim Atkin	1011	82	100	91.37	92	3.35
La RVF	484	11.5	20	16.3	16.25	1.4
Jeannie Cho Lee	219	80	99	91.9	92	2.8
Antonio Galloni	210	79	95.5	89.2	89.5	2.8
Jeff Leve	158	83	99	90.3	90	3.0

Source: authors' calculation based on Wine Services data.

The scores given by the experts seem relatively homogenous with average scores comprised between 15.74 and 17.12 for the European ones and between 89.24 and 91.87 for the US experts. Interestingly, we can see that the Europeans have all given the 20 maximum grade while only J. Suckling and Tim Atkin have given the maximum 100 grade. The score range defined as the difference between the maximum and the minimum score for each expert is comprised between 14 and 29 for the US experts and 5.5 to 10 for the European experts. Two remarks are to be made at this stage. Firstly, all the experts only use a fraction of their scale. In proportion, the fraction used by US experts seems particularly small (20 points on average). But this proportion is larger in absolute value than the spectrum used by European experts (7.8 points on average), giving the former a potentially higher accuracy in their rating. Secondly, in both cases (US and European),

these experts exhibit significant differences in the way they rate the wines: there is no homogeneity among them concerning the size of the score range they use; it varies from single to double. Therefore, the direct comparison among experts' scores is fallacious, even if they use the same rating scale. Each expert has his/her own preferences' space and our aim is to express all the scores in the same space of preferences.

The medians also offer interesting information as it can be interpreted as a threshold between good wines and less good/bad wines. 90 points (16.5) for the US (European) experts appears as the borderline between these two catagories.

Table 2 presents the number of wines that have been tasted by each expert pair. With 2698 wines rated both by Ren Gabriel and Wine Spectator, these two experts exhibit the highest number of same wines tasted. On average, Robert Parker, Neal Martin, Jancis Robinson, Wine Spectator, Bettane&Desseauve, , Jacques Dupont, la Revue des Vins de France and Rene Gabriel have rated more than 1000 same wines over the period with other experts.

Table 2: Wine pairing (number of same wines tasted by both experts)

	RP	NM	JR	WS	AG	BD	JD	JS	JC	JL	RVF	JMQ	RG	TA	D20	D100
RP		1361	1833	2168	168	1637	1667	706	231	568	1317	1041	2443	750	811	198
NM	1361		1549	1419	160	1422	1353	714	243	556	1294	996	1578	787	842	170
JR	1833	1549		2049	171	1929	1946	730	247	561	1440	1268	2361	838	898	194
WS	2168	1419	2049		168	1753	1803	663	221	529	1330	1151	2698	743	832	178
AG	168	160	171	168		161	164	177	1	173	157	167	184	158	13	183
BD	1637	1422	1929	1753	161		1756	687	232	570	1427	1185	2007	773	856	200
JD	1667	1353	1946	1803	164	1756		650	230	515	1316	1139	2039	738	796	186
JS	706	714	730	663	177	687	650		219	535	618	652	867	731	565	211
JC	231	243	247	221	1	232	230	219		158	211	186	250	227	246	1
JL	568	556	561	529	173	570	515	535	158		484	427	599	524	403	203
RVF	1317	1294	1440	1330	157	1427	1316	618	211	484		959	1546	690	756	191
JMQ	1041	996	1268	1151	167	1185	1139	652	186	427	959		1366	657	538	206
RG	2443	1578	2361	2698	184	2007	2039	867	250	599	1546	1366		918	929	214
TA	750	787	838	743	158	773	738	731	227	524	690	657	918		670	181
D20	811	842	898	832	13	856	796	565	246	403	756	538	929	670		0
D100	198	170	194	178	183	200	186	211	1	203	191	206	214	181	0	
Average	**1262**	**1054**	**1340**	**1329**	**154**	**1231**	**1215**	**602**	**202**	**473**	**1008**	**878**	**1495**	**656**	**653**	**180**

Source: authors' calculation based on Wine Services data.

WS :Wine Spectator ; RP : Robert Parker; JR: Jancis Robinson; JD: Jacques Dupont; BD:Bettane&Desseauve; NM: Neal Martin; D20: Decanter20; JS: James Suckling; D100: Decanter100; RVF: La RVF; JCL: Jeannie Cho Lee; AG: Antonio Galloni; JL: Jeff Leve; JMQ: Jean-Marc Quarin; TA: Tim Atkin; RG: Rene Gabriel.

Table 3: Experts' scores correlation matrix

	RP	NM	JR	WS	AG	BD	JD	JS	JC	JL	RVF	JMQ	RG	TA	D20	D100
RP		0,57	0,43	0,61	0,41	0,59	0,50	0,69	0,68	0,77	0,60	0,68	0,58	0,57	0,71	0,58
NM	0,57		0,49	0,62	0,56	0,58	0,50	0,69	0,59	0,74	0,59	0,65	0,58	0,57	0,67	0,61
JR	0,43	0,49		0,51	0,17	0,50	0,39	0,48	0,55	0,42	0,46	0,54	0,45	0,57	0,63	0,36
WS	0,61	0,62	0,51		0,59	0,62	0,47	0,74	0,70	0,75	0,60	0,64	0,61	0,64	0,69	0,62
AG	0,41	0,56	0,17	0,59		0,45	0,35	0,47		0,60	0,44	0,52	0,54	0,35	0,32	0,56
BD	0,59	0,58	0,50	0,62	0,45		0,50	0,65	0,70	0,67	0,65	0,69	0,55	0,63	0,75	0,74
JD	0,50	0,50	0,39	0,47	0,35	0,50		0,59	0,60	0,62	0,56	0,62	0,47	0,52	0,63	0,67
JS	0,69	0,69	0,48	0,74	0,47	0,65	0,59		0,66	0,75	0,70	0,70	0,66	0,54	0,71	0,60
JC	0,68	0,59	0,55	0,70		0,70	0,60	0,66		0,71	0,58	0,65	0,64	0,62	0,72	
JL	0,77	0,74	0,42	0,75	0,60	0,67	0,62	0,75	0,71		0,73	0,79	0,69	0,56	0,68	0,68
RVF	0,60	0,59	0,46	0,60	0,44	0,65	0,56	0,70	0,58	0,73		0,72	0,57	0,56	0,71	0,78
JMQ	0,68	0,65	0,54	0,64	0,52	0,69	0,62	0,70	0,65	0,79	0,72		0,65	0,67	0,75	0,75
RG	0,58	0,58	0,45	0,61	0,54	0,55	0,47	0,66	0,64	0,69	0,57	0,65		0,57	0,65	0,62
TA	0,57	0,57	0,57	0,64	0,35	0,63	0,52	0,54	0,62	0,56	0,56	0,67	0,57		0,66	0,61
D20	0,71	0,67	0,63	0,69	0,32	0,75	0,63	0,71	0,72	0,68	0,71	0,75	0,65	0,66		
D100	0,58	0,61	0,36	0,62	0,56	0,74	0,67	0,60		0,68	0,78	0,75	0,62	0,61		
Average	**0,59**	**0,60**	**0,45**	**0,62**	**0,46**	**0,59**	**0,51**	**0,65**	**0,64**	**0,69**	**0,60**	**0,65**	**0,58**	**0,57**	**0,66**	**0,63**

Source: authors' calculation based on Wine Services data.

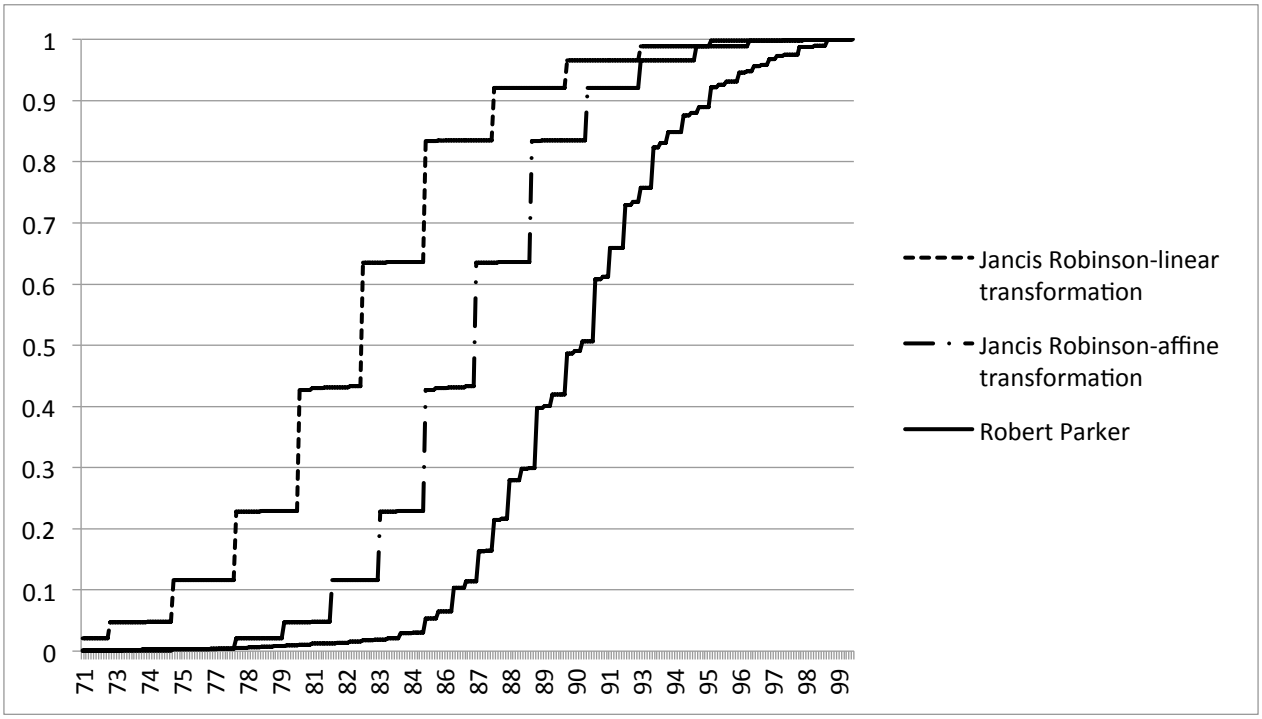
WS :Wine Spectator ; RP : Robert Parker; JR: Jancis Robinson; JD: Jacques Dupont; BD:Bettane&Desseauve; NM: Neal Martin; D20: Decanter20; JS: James Suckling; D100: Decanter100; RVF: La RVF; JCL: Jeannie Cho Lee; AG: Antonio Galloni; JL: Jeff Leve; JMQ: Jean-Marc Quarin; TA: Tim Atkin; RG: Rene Gabriel.

Table 3 shows a systematic positive correlation between each expert pair but the average correlation among experts does not exceed 0.59. The highest correlation can be found between Jean-Marc Quarin and Jeffe Leve. Jancis Robinson and Antonio Galloni exhibit the lowest correlation and therefore the lowest consensus with the other experts. On the contrary, Jeff Leve and Decanter 20 display the highest correlation and therefore the best level of consensus with the other experts. In particular, these two experts have a strong correlation with Robert Parker. The US experts seem to have higher correlation among themselves compared to the European ones. All these results are perfectly in line with the work of Masset and Weisskopf (2015), even if they conclude on a strong consensus while an average correlation of 0.59 suggests a mixed conclusion for us, particularly according to the high volatility of the correlation pairs.

3. Methodology

Two experts best embody the issue of transforming the grading scales. Robert Parker and Jancis Robinson are influential experts, respectively in the U.S. and in England. Robert Parker scores out of 100 whereas Jancis Robinson scores out of 20. Our method addresses a very common problem of quality assessing. We can imagine a comparison between two wines. The first is graded by the two experts, but the second only by Robert Parker. The key issue is to use the information given by Jancis Robinson on the first wine to qualify its global evaluation.

Figure 1: Distribution functions for each transformation and Robert Parker's scores distribution



Source: authors' calculation based on Wine Services data.

The naïve solution is the linear function, which consists simply in multiplying Jancis Robinson's scores by a factor of five. It is not satisfactory, as it does not use the minimum grades. In order to use the minima of the intervals used by each expert, one can use an affine function of the scores of Jancis Robinson from the interval [12,20] into the interval [70,100]. The best way to judge the relevance of the transformation is to compare the distribution functions. Figure 1 displays the distribution functions of the Jancis Robinson scores after each transformation, compared to Robert Parker's scores distribution function.

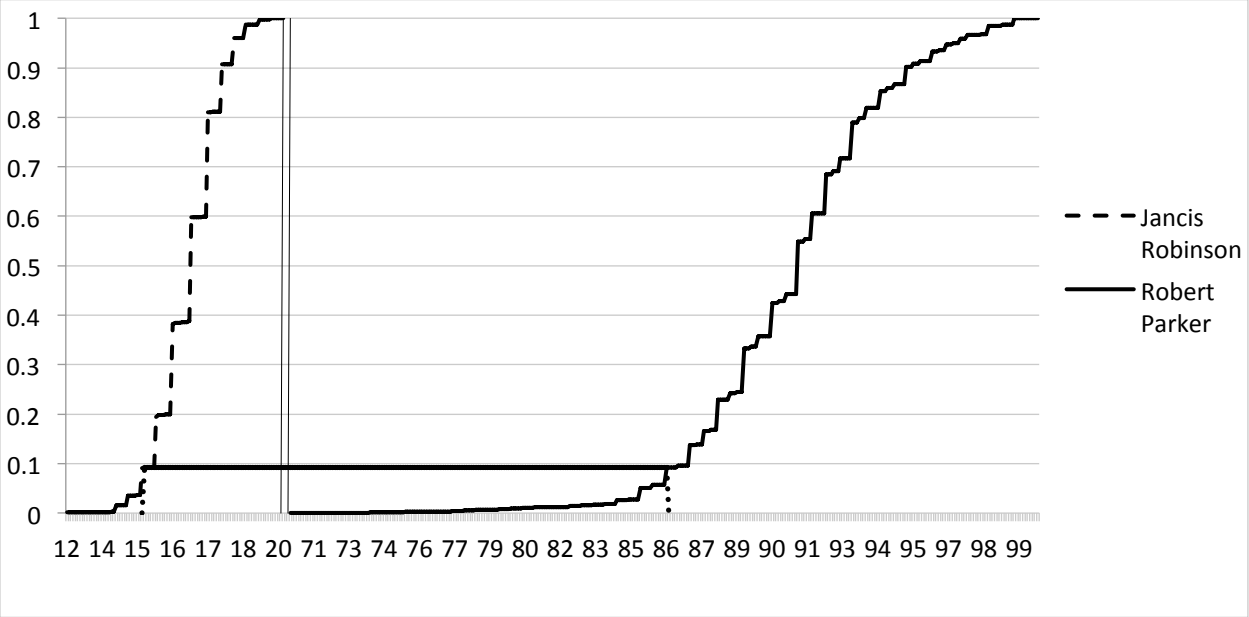
The distribution of Jancis Robinson's transformed scores is closer to Robert Parker's distribution with the affine function. Still, one might argue that the Jancis Robinson transformed scores are still underrated compared to the grading system of Robert Parker. More than half of the Robert Parker scores are above 90/100, against only 8% for the scores computed with the affine function. As a result, a 90/100 for Robert Parker is a much lower evaluation of quality than a 90/100 for Jancis Robinson with the affine function. A satisfactory transformation of the scores should both put the scores on the same scale and convey the same value to each score. Transformed scores should then have the same distribution function as the Robert Parker scores. Such a function exists and is non-parametrically tractable.

The theoretical framework is the following. Posit that quality of Bordeaux wines is a random variable. The experts evaluate this quality along a scale of their choice, according to their preferences and to their use of their scales. Let F be the distribution function of the Jancis Robinson scores, and G be the distribution function of the Robert Parker scores. These functions express both their respective ways of using their grading scales, and their respective appreciation of Bordeaux wines. Both causes tackle the comparison between two Bordeaux wines. Recall that our main goal is to use the information given by Jancis Robinson to qualify a score given by Robert Parker, and to take into account that Jancis Robinson usually gives lower scores.

Our idea is to apply the function $G^{-1} \circ F$ in order to obtain the same distribution function for the Jancis Robinson transformed scores and Robert Parker raw scores. This uses the following classical property of probability distribution. Let F_X and F_Y be the distribution of the continuous random variables X and Y, then the random variable ${F_Y}^{-1} \circ F_X(X)$ has the same probability distribution as Y, ${F_Y}^{-1}$ being the generalized inverse of F_Y. To avoid any selection bias, the two empirical distributions are computed on a common sample, which contains all wines with a score from each expert. For the chosen couple of experts, the sample includes 1833 observations.

Figure 2 provides a graphical illustration of our method. As an example, we evaluate the image of a 15/20 from Jancis Robinson on the Robert Parker scale. 15/20 is the quantile of order 0.092 for the Jancis Robinson distribution function, which means that 9.2% of the Jancis Robinson scores are less than or equal to 15/20. On the Robert Parker distribution function, we read that this quantile is 86/100. We obtain that a 15/20 given by Jancis Robinson is worth a 86/100 given by Robert Parker for Bordeaux wines. In the situation previously stated, this method allows the Jancis Robinson score to be turned into the Robert Parker scale. The average of the two scores is a synthetic indicator of all available information, and it can be directly confronted to single Parker scores when Jancis Robinson scores are missing.

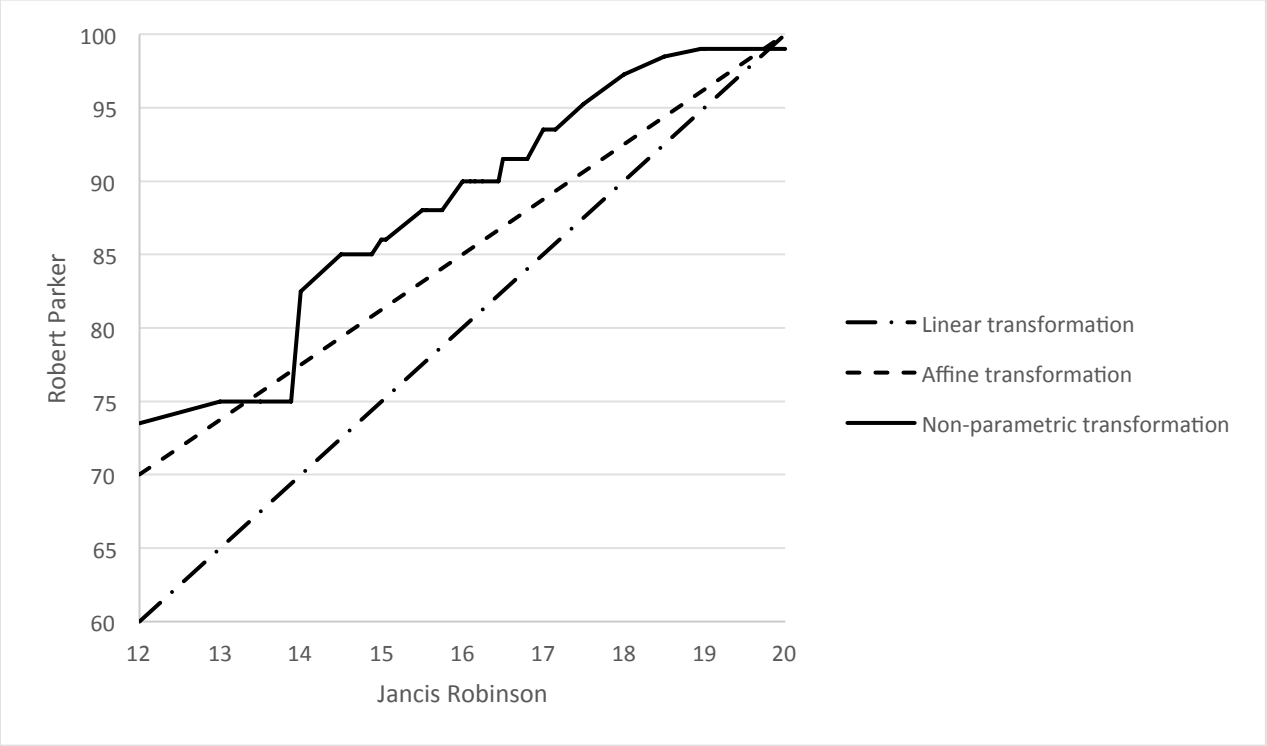
Figure 2: Original method using the empirical distribution functions



Source: authors' calculation based on Wine Services data.
Note:The double vertical lines stands for the gap on the x-axis between 20 and 70.

Doing so for all existing scores from Jancis Robinson, we obtain a non-parametric function which ensures that the image scores have the same distribution as the Robert Parker scores. Figure 3 compares the plots of the three functions.

Figure 3: Plot of the three functions



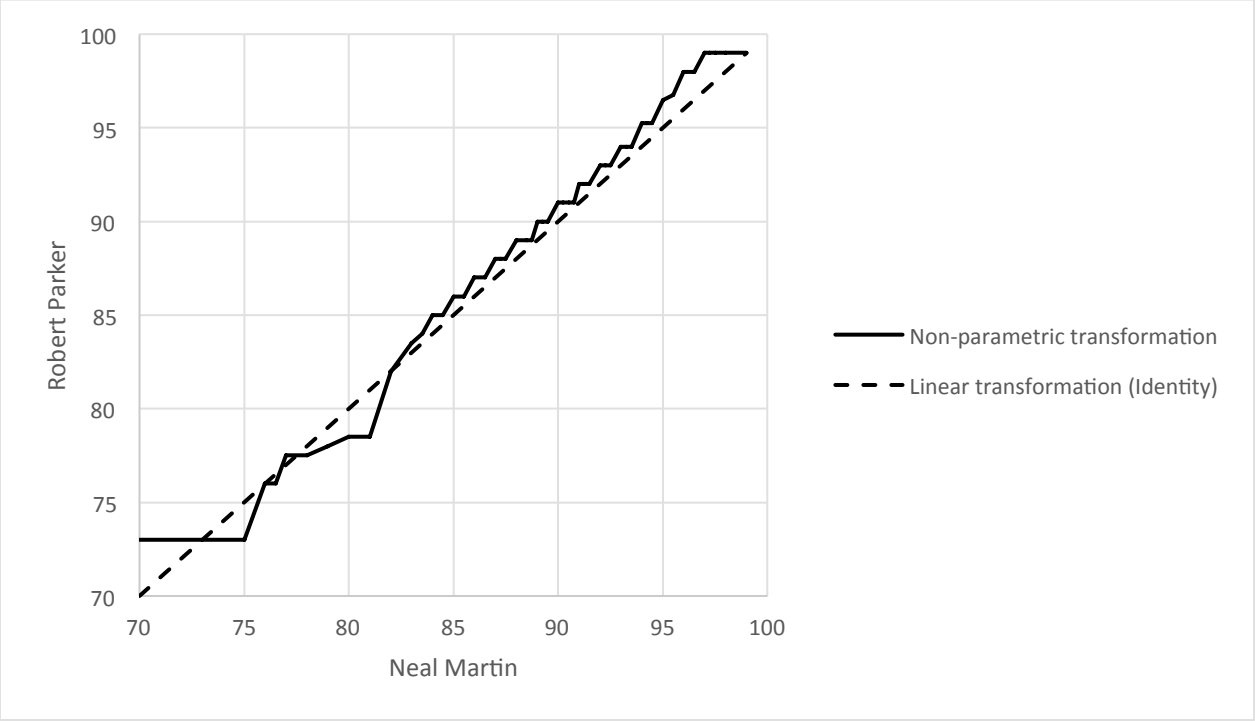
Source: authors' calculation based on Wine Services data.

The non-parametric function is irregular on the half-open interval [12,14[. Indeed, this interval only concerns 5 observations and 0.4% of the distribution of the Jancis Robinson scores. It corresponds to the half-open interval [70,86[for Robert Parker.

This method can also be applied for two experts who both score out of 100. Figure 4 plots the non-parametric function which turns Neal Martin scores into the Robert Parker scale[4]. We find the same regularity issue within the interval [70,82[, but the function reveals that Robert Parker has been less reluctant than Neal Martin to grant scores above 95/100. A 97/100 by Neal Martin is as scarce as a 99/100 by Robert Parker.

[4] **All the experts' transformations curves are presented in Annex 1.**

Figure 4: Conversion of Neal Martin scores into Robert Parker scale



Source: authors' calculation based on Wine Services data.

4. Example of outcomes

Our conversion method facilitates all kinds of comparison between scores, whether among winemakers, appellations or vintages. We hereafter provide an insight into the possible outcomes. While the general method allows the scores of any expert to be converted into any other expert's scale, we have chosen to convert all scores into the Robert Parker scale. Since he is commonly referred to as the most influential expert for Bordeaux wines (see notably Hadj Ali et al. 2008; Masset et al., 2015), we assume that his scale is the most familiar for the reader.

Table 4 displays all available 2013 *primeurs* scores for a subsample of twenty famous Bordeaux properties. Columns 2 to 4 display the average of the available scores respectively transformed by the linear, the affine and the non-parametric function. Our non-parametric method gives the highest scores, as it transposes the scores on the scale of Robert Parker, used to giving high scores compared to his peers. On the whole, the other experts mitigate the negative opinion of Robert Parker on vintage 2013, as the mean score is often above Robert Parker's grade.

The last column of Table 4 provides the standard deviation of the scores for each wine. As our method puts all scores on the same scale, it is now relevant to compute a standard deviation for each wine across experts. It gives an indication of the level of consensus on each wine: the lower the deviation among the scores, the more reliable the mean score is. Château Clinet shows the highest level of consensus with a standard deviation of only 1.2 while Château Le Gay shows the largest dispersion with a standard deviation of 2.64.

Another possible outcome is to facilitate the comparison between vintages for two experts. Table 4 shows the mean scores of vintages 2003 to 2013 for Robert Parker and Jancis Robinson with and without applying our function to the Jancis Robinson scores. The comparison is much easier as the scores are given on the very same scale. Our transformation highlights that Jancis Robinson was much more lenient with vintages 2007 and 2013 than Robert Parker, and that she really enjoyed vintage 2012.

Table 4: Raw *primeur* scores for a subsample of vintage 2013 and mean scores computed for the three methods

Wine	Score - Linear	Score - Affine	Score - Non parametric	sd	RP	NM	JR	WS	AG	BD	JD	JS	JL	RVF	D	JMQ	RG	TA
Angelus	**89.4**	**91.7**	**92.7**	**1.87**	91.5	91	17.5		91	18.75		92.5	92	16.25	90.25	16.5	17	95
Ausone	**90.3**	**92.6**	**93.7**	**2.27**	94	92	17.5		91	19	17	91.5	94	16	92	16.75	19	94
Cheval Blanc	**89.1**	**91.7**	**92.6**	**1.25**	90	92	17	92.5	91	18	16.5	93.5	93.5	16.25	92	16.5	18	93
Clinet	**86.7**	**89.7**	**90.4**	**1.2**	92	91	16	88.5	91.5	16		90.5	92	16	89	15.5	17	91
Eglise Clinet	**91.2**	**93,0**	**94,0**	**1.93**	93	95	17.5	90.5	91.75	17.5		93.5	95	17.25	91	16.75	19	96
Evangile	**88.1**	**90.8**	**91.9**	**1.98**	88.5	92	18	91.5	93	17.75	16	90.5	91	16.25	90.25	16	17	92
Gazin	**85.6**	**89.3**	**90.2**	**1.5**	91	90	16	87.5		15.5	16.75		90.5	16.75	89	15.75	17	91
Grand Vin de Latour	**89.1**	**91.8**	**92.8**	**1.55**	89		17	91.5	92	17.25	16.75	92.5	92	17.5	94	16	18	95
Haut Brion	**90.1**	**92.2**	**93.1**	**1.85**	91	90	16.5	92.5	92.5	18	16.75	92.5	93	16.5	94		19	93
La Conseillante	**86.7**	**89.9**	**90.8**	**1.7**	90	91	15	89.5	91.5	17	15.5	90.5	92	17	90	16	17	92
La Violette	**88.6**	**91.1**	**91.7**	**2.49**	87	93			88	17.75			93.5	16.75		16.25	18	93
Lafite Rothschild	**88.6**	**91.3**	**92.4**	**1.8**	88	92	17	90.5	91.5	17.75	16.25	92.5	91	16.25	94	16	18	95
Lafleur	**90.6**	**92.8**	**93.8**	**1.84**	90	94	18		93	18		93.5	93.5	17.5	93	16.75	17	95
Le Gay	**86.5**	**89.6**	**90.3**	**2.64**	86	91	15		90.5	17		91.5	93	15	88	16	18	93
Margaux	**89.9**	**92.1**	**92.9**	**1.53**	89	92	16.5	91.5	92.5	17.25	16.5	94.5	93.5	17.5	94		18	94
Mouton Rothschild	**89.6**	**92,0**	**92.8**	**1.79**	92	93	17	92.5	89	18.25	17	92.5	93	17.5	94	16.75	17	92
Pavie	**88.6**	**91.3**	**92.1**	**1.9**	93	92	16		92.25	18		91.5	93.5	15.5	93	16.25	18	90
Petrus	**90.2**	**92.5**	**93.5**	**1.82**	91.5	91	18.5		91	18	17.5	92.5	94	17	92	16.5	18	94
Trotanoy	**87.5**	**90.7**	**91.7**	**1.6**	92.5	91	16	91.5	91	17	15	90.5	93.5	16	91	16	18	95
Vieux Chateau Certan	**88.0**	**90.6**	**91.5**	**1.93**	87.5	93	17.5	91.5	91.5	17.5	15	90.5	91.5		91	16.25	17	92

Source: authors' calculation based on Wine Services data.

Legend: sd : standard deviation of the scores obtained with the non-parametric method. RP:Robert Partker ; NM:Neal Martin; JR:Jancis Robinson; WS: Wine Spectator; AG:Antonio Galloni; BD:Bettane et Desseauve; D:Decanter; JD: Jacques Dupont; JS: James Suckling; JL: Jeffe Leve; RVF: Revue des Vins de France ; JMQ : Jean-Marc Quarin ; RG : Rene Gabriel; TA: Tim Atkin.

Table 4: Mean vintage score for Robert Parker and Jancis Robinson with and without transformation

Vintage	Number of observation	Robert Parker	Jancis Robinson - Non-parametric	Jancis Robinson - Raw scores
2003	126	**90.5**	**89.9**	16.1
2004	69	**91.3**	**92.0**	16.6
2005	174	**91.8**	**91.7**	16.6
2006	116	**91.6**	**92.2**	16.7
2007	196	**88.7**	**90.5**	16.2
2008	198	**91.0**	**91.9**	16.6
2009	195	**92.6**	**92.4**	16.7
2010	201	**92.4**	**92.4**	16.7
2011	186	**90.0**	**91.2**	16.4
2012	194	**90.5**	**92.3**	16.7
2013	168	**88.9**	**91.1**	16.3

Source: authors' calculation based on Wine Services data.
Note: We lack Jancis Robinson *primeurs* scores for vintages 2000, 2001, 2002 and 2014.

5. Conclusion

This paper offers a simple and transparent methodology to express the scores of each wine expert on the same rating scale. It ensures perfect comparability of the scores among experts and then allows for a relevant average calculation of the available wine scores. In that sense, this research work meets the requirement of wine professionals by reducing their uncertainty and could therefore contribute towards optimal market efficiency**. In this line this score is available for free for each Bordeaux *en primeur* wine on the website globalwinescore.com. The website presents at this point the scores of the 2014 vintage. It will be automatically actualized when new experts' scores will be released thanks to the involvement of the company Wine Services (a French startup located in Bordeaux and specialized in wine marketing data).**

Nevertheless, several issues still have to be addressed. Who has to be the expert of reference? Robert Parker seems to be the natural candidate but he has now retired and stopped tasting the Bordeaux *en primeur* in 2015. How to interpret the standard deviation in the cases where wines are not tasted by the same number of experts? Does a standard devation calculated on the basis of 2 scores provide the same information as a standard devation calculated on the basis of 12 scores in terms of consensus? Other questions will certainly have to be addressed and we hope that this paper will induce some reactions to improve our methodology.

References:

Ali, H. H., & Nauges, C. (2007). The pricing of experience goods: the example of *en primeur* wine. american Journal of Agricultural Economics, 89(1), 91-103.

Ali, H. H., Lecocq, S., & Visser, M. (2008). The Impact of Gurus: Parker Grades and *En Primeur* Wine Prices. The Economic Journal, 118(529), F158-F173.

Ashenfelter, O. (2008). Predicting the quality and prices of Bordeaux wine. The Economic Journal, 118(529), F174-F184.

Ashenfelter, O., Ashmore, D., & Lalonde, R. (1995). Bordeaux wine vintage quality and the weather. Chance, 8(4), 7-14.

Bodington, J. C. (2015). Evaluating wine-tasting results and randomness with a mixture of rank preference models. Journal of Wine Economics, 10(1), 31-46.

Cardebat, J. M., Figuet, J. M., & Paroissien, E. (2014). Expert Opinion and Bordeaux Wine Prices: An Attempt to Correct Biases in Subjective Judgments. Journal of Wine Economics, 9(3), 282-303.

Mahenc, P., & Meunier, V. (2006). Early sales of Bordeaux grands crus. Journal of Wine Economics, 1(1), 57-74.

Masset, P., Weisskopf, J. P., & Cossutta, M. (2015). Wine tasters, ratings, and en primeur prices. Journal of Wine Economics, 10(1), 75-107.

Olkin I., Lou Y., Stoks L., & Cao J. (2015). Analyses of wine-tasting data: A tutorial. Journal of Wine Economics, 10(1), 4-30.

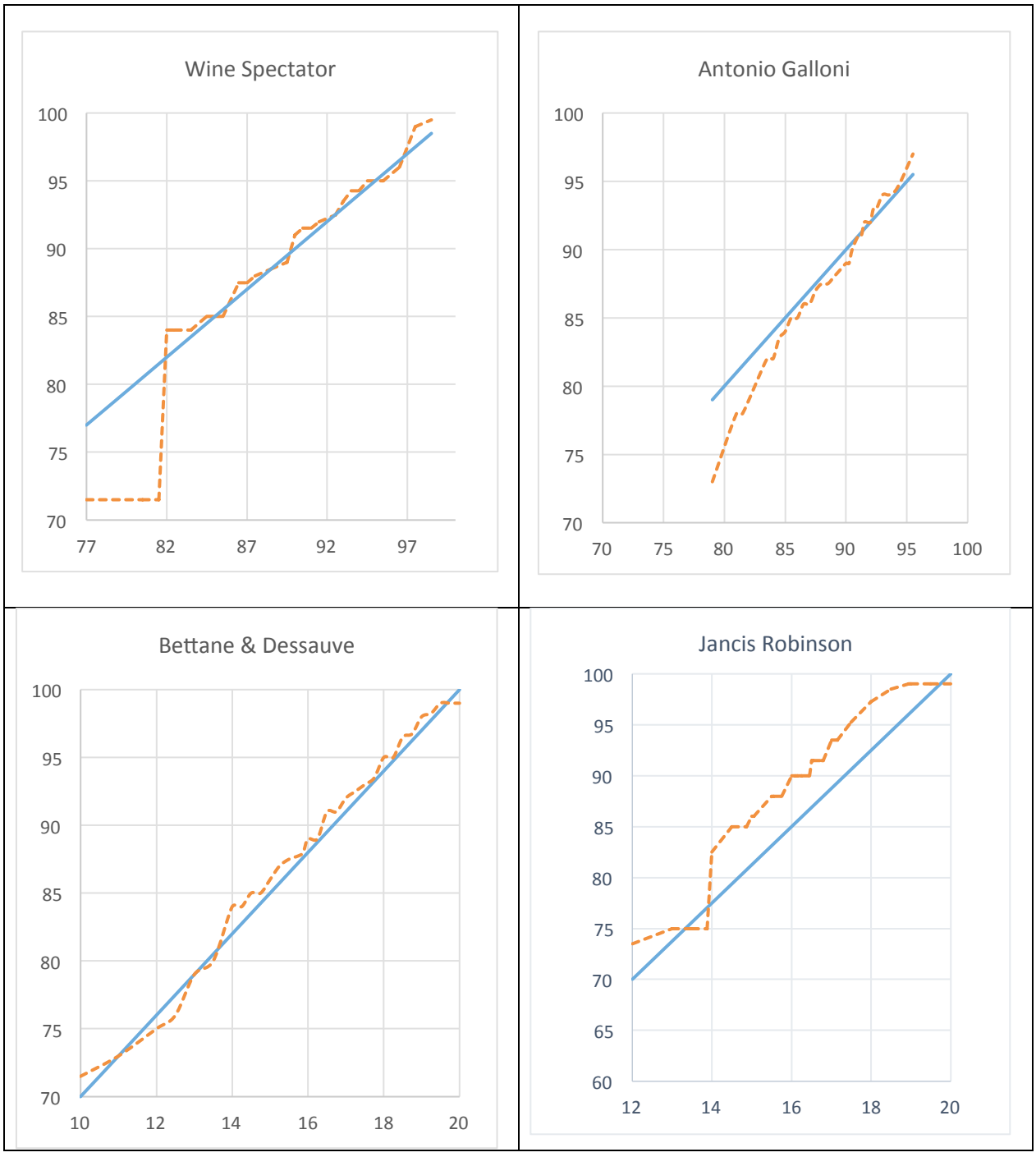
Quandt, R. E. (2007). On wine bullshit: some new software? Journal of Wine Economics, 2(02), 129-135.

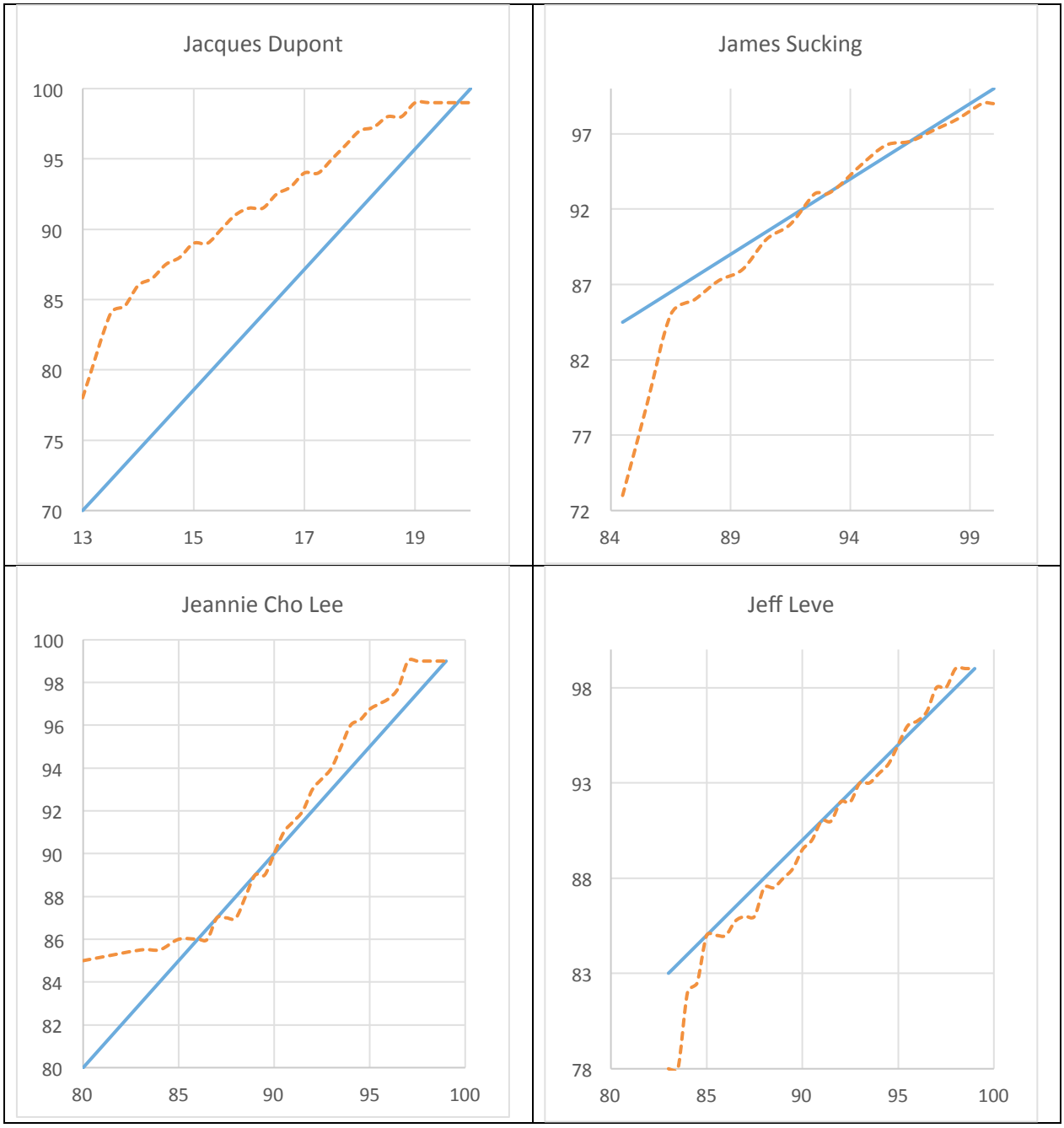
Storchmann, K. (2012). Wine economics. Journal of Wine Economics, 7(1), 1-33.

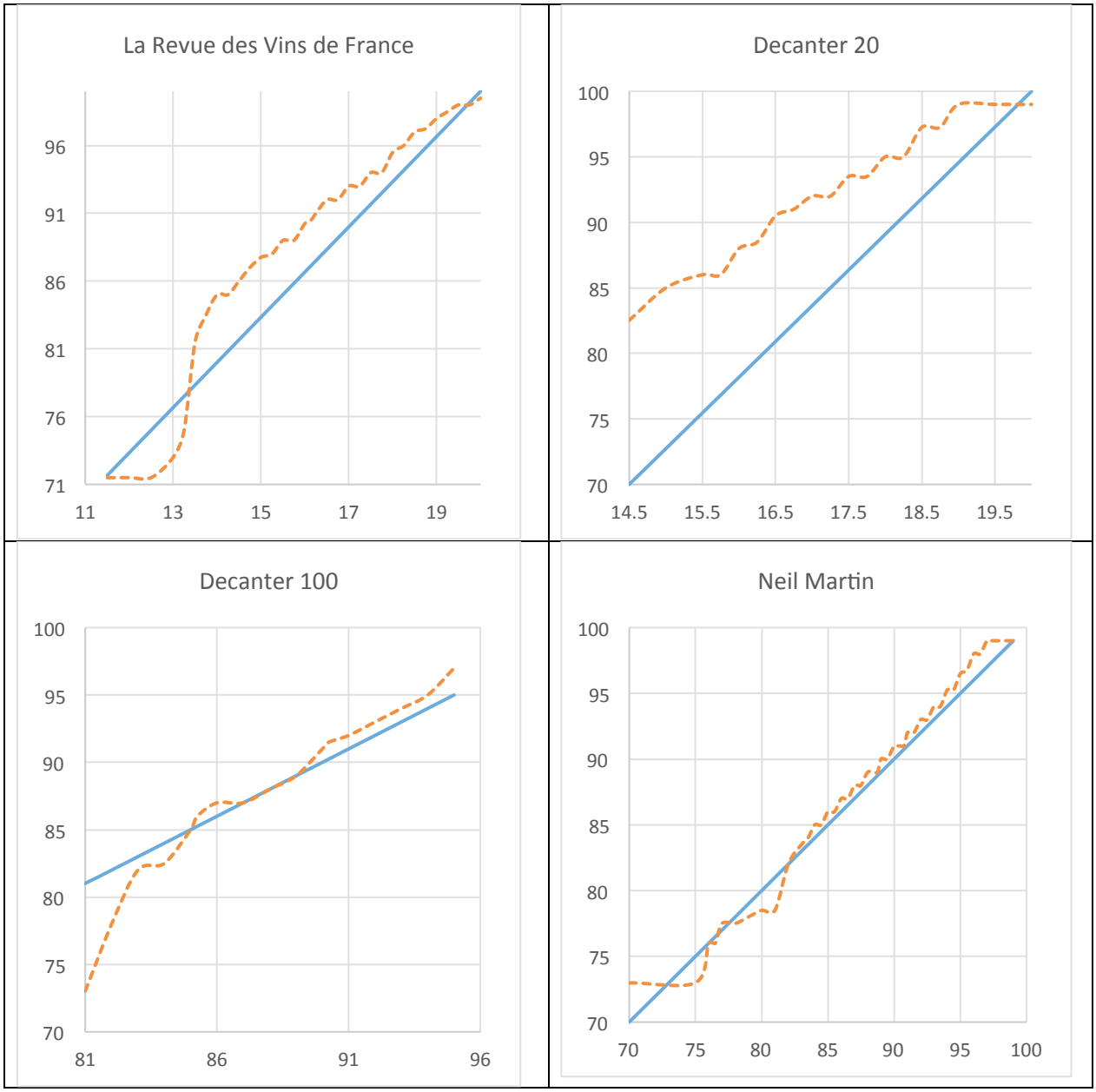
Annex 1: Conversion of expert *i*'s scores (x axis) into Robert Parker scale (y axis)

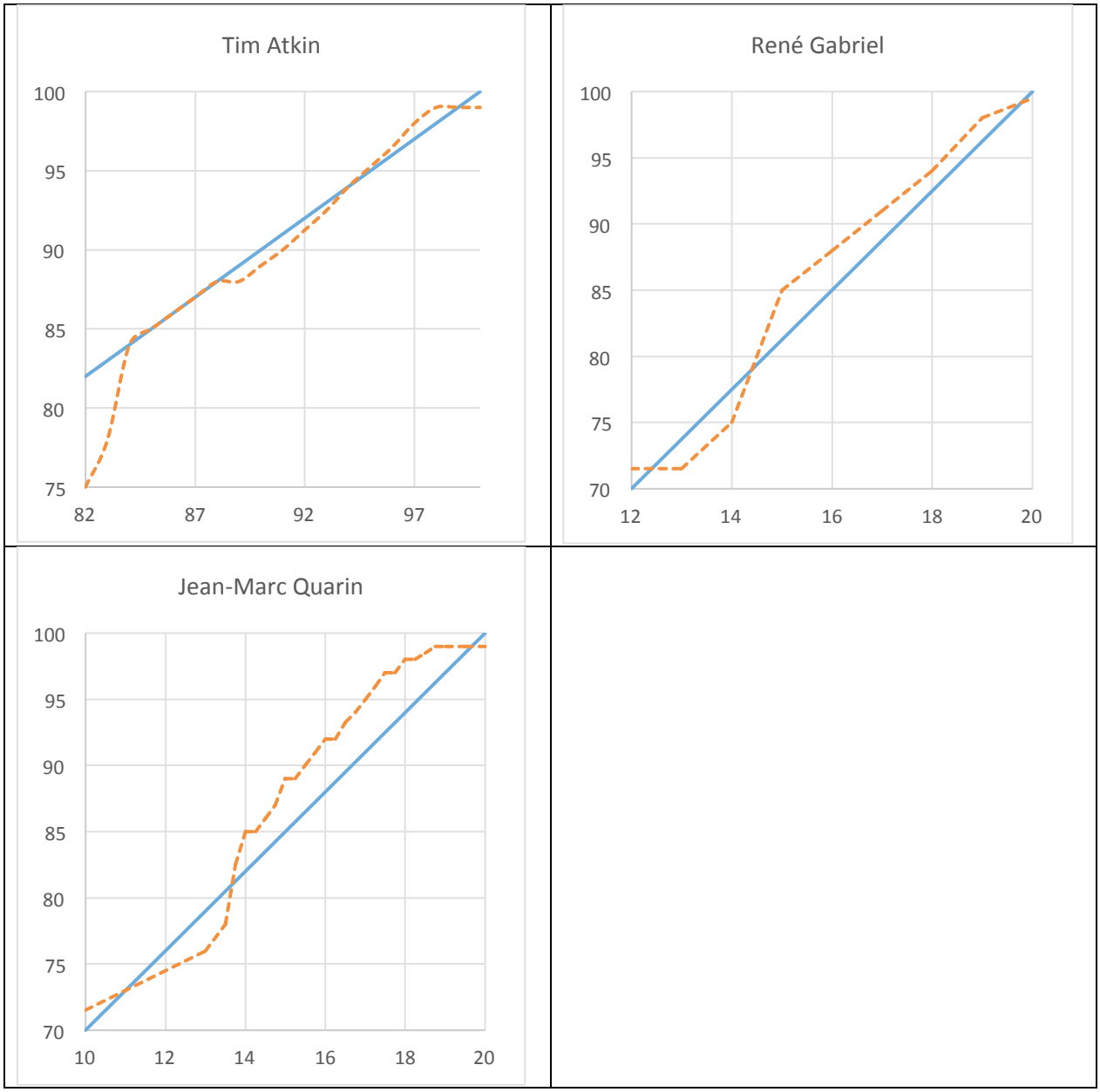
Affine transformation

Non parametric transformation




Jacques Dupont
100
95
90
85
80
75
70
13
15
17
19
James Sucking
97
92
87
82
77
72
84
89
94
99
Jeannie Cho Lee
100
98
96
94
92
90
88
86
84
82
80
80
85
90
95
100
Jeff Leve
98
93
88
83
78
80
85
90
95
100


La Revue des Vins de France
96
91
86
81
76
71
11
13
15
17
19
Decanter 20
100
95
90
85
80
75
70
14.5
15.5
16.5
17.5
18.5
19.5
Decanter 100
100
95
90
85
80
75
70
81
86
91
96
Neil Martin
100
95
90
85
80
75
70
70
75
80
85
90
95
100



Source: authors' calculation based on Wine Services data.