Economics Bulletin

Volume 32, Issue 2

Do consumers learn from tasting scores set by experts?

Nicolas Gérard Vaillant
LEM, U Catholique de Lille (FLSEG-CRCH and ISTC)

François-charles Wolff
LEMNA, Université de Nantes and INED, Paris, France.

Abstract

Using data from whisky tastings and fixed effect regressions, we find that net of any composition effect, experts report different tasting scores on average. This indicates that consumers have little to learn from absolute scores.

We thank one anonymous reviewer for helpful suggestions on a previous draft and Philippe Frocrain for research assistance with the data.
Citation: Nicolas Gérard Vaillant and François-charles Wolff, (2012) "Do consumers learn from tasting scores set by experts?", *Economics Bulletin*, Vol. 32 No. 2 pp. 1378-1384.
Contact: Nicolas Gérard Vaillant - nicolas.vaillant@icl-lille.fr, François-charles Wolff - francois.wolff@univ-nantes.fr.
Submitted: November 20, 2011. **Published:** May 09, 2012.

1. Introduction

The judgment and ratings of experts matter in a wide range of markets including arts (Ginsburgh, 2003), wine (Ashenfelter and Jones, 2000) or restaurants (Chossat and Gergaud, 2003). Experts are not only expected to reduce consumer search costs, they also have a large influence on market outcomes like sales or prices (Hadj Ali et al., 2008). There is also a large number of guidebooks and related consumer reports for experience goods like wine or cigars, since their intrinsic quality cannot really be appreciated before consumption (Nelson, 1970). For such goods, consumers truly benefit from additional qualitative information when reading experts' ratings (Vaillant and Wolff, 2012).

In this paper, we assess the relevance of experts' opinions by focusing on the case where two experts give a mark after tasting the same product. While two ratings should provide more information to consumers, we argue that multiple ratings from different experts may be counter-productive if the experts have different reference scales in mind. For a given product quality, if some experts set high scores on average while other experts set much lower scores, then consumers may be mislead into thinking that tasting scores correctly reflect quality. Some normalization accounting for the average score set by experts would be required for guidebooks to be more informative.

Specifically, we draw on tasting notes reported in the issues of Whisky Magazine to study whether experts report different tasting scores on average. A difficulty is that differences in the mean marks may simply be the result of a non-random allocation of whiskies among experts. But as the selected whiskies are tasted by two experts, we account for the unobserved and observed whisky features using fixed effect regressions. Our results show that net of any composition effect, on average experts give different tasting scores.

2. Data and descriptive statistics

Our empirical analysis is based on data collected from back issues of Whisky Magazine, which defines itself as "the perfect complement to the dram in your glass". Every issue of Whisky Magazine, published eight times a year, brings "articles on the art, science and romance of the 'water of life', plus page after page of tasting notes". Given data availability, we rely on the previous issues published from January 1999 (n° 1) till July 2008 (n° 73)[1].

In each magazine, a set of whiskies undergoes tastings by experts. There are 27 tastings per magazine on average. For each tasting, we have the following information. First, the review includes an objective description of the product: brand, type (blended, single malt, etc), alcohol by volume, age, vintage (if any), place of production (distillery), region of production and availability. Secondly,

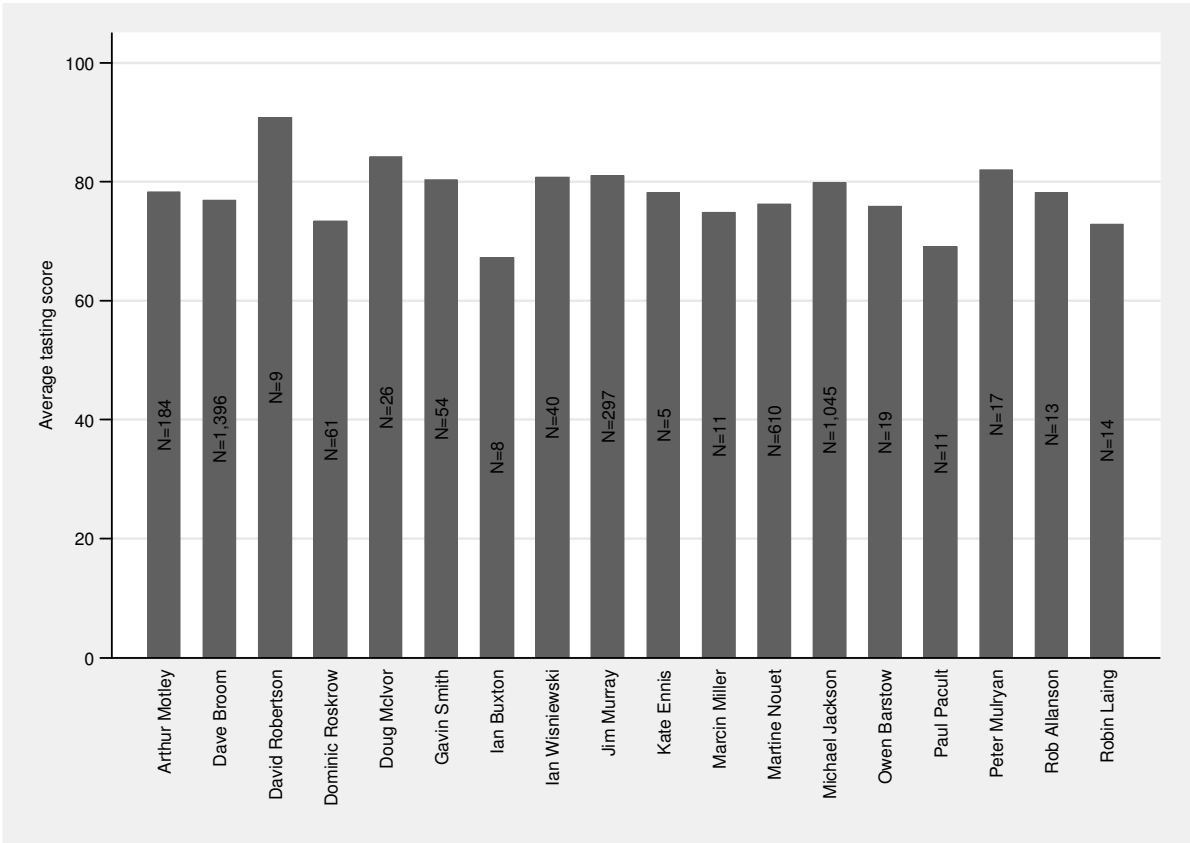
[1] Data were collected in May 2011. All the information is available online at http://www.whiskymag.com/. Tastings of the more recent issues are not available online.

there are tasting notes given by experts. In almost all cases (99.6%), each whisky was tasted by two different experts. Each taster provides a description of nose, palate, finish and indicates a general comment on the tasting along with a score over 100.

Our sample comprises 3820 tastings corresponding to 1914 whiskies. Statistics on the mean tasting scores are presented in Figure 1. On average, the mean score delivered by an expert is equal to 78, with a standard deviation of 8.4. Figure 1 shows large differences in the mean tasting notes. David Robertson is the most generous with a mean score of 90.8 (9 tastings), followed by Doug McIvor with a mean score of 84.2 (26 tastings). Conversely, Paul Pacult and Ian Buxton set lower scores on average, respectively 69.1 (11 tastings) and 67.3 (8 tastings).

A simple explanation of these low and high average scores could be that these experts taste a small number of whiskies with very specific characteristics, either bad or good. However, a look at the most frequent contributors to Whisky Magazine suggests that on average experts tend to report different tasting notes. The mean scores provided by Jim Murray (297 tastings) and Michael Jackson (1,045 tastings) are much higher than those reported by Dave Broom (1,396 tastings) and Martine Nouet (610 tastings). In what follows, we turn to an econometric analysis to account for the whisky features when explaining the tasting scores.

Figure 1. Mean tasting scores from experts



Source: authors' calculations, data from Whisky Magazine.

3. Econometric analysis

Let $eval_{iw}$ be the note given by the expert i $(i = 1, \dots, I)$ when tasting the whisky w $(w = 1, \dots, W)$. The score is expected to depend on objective characteristics of the whisky such as age, vintage or production region, on unobservable traits (to the econometrician) like aroma or palate, and on the personal appreciation of both sets of characteristics by the taster. Denoting by $expert_i$ a set of dummy variables such that $expert_i = 1$ for taster i and $expert_i = 0$ otherwise, then estimating the following linear regression will shed light on differences in evaluation:

$$eval_{iw} = \beta X_w + \sum_i \theta_i * expert_i + \varepsilon_{iw} \quad (1)$$

with X_w a set of whisky features, β and θ_i are parameters to estimate, and ε_{iw} an error term. Considering one expert j as reference, the various coefficients θ_i indicate whether the average score given by the various experts i is significantly different or not from that of expert j (with $i \neq j$).

A difficulty with the OLS estimates of (1) is that many of the whisky features are likely to remain unobserved. Hence, the coefficients θ_i are expected to be biased because of the problem of omitted variables. Since we have evaluations from two different experts for almost all whiskies, we are able to account for both observed and unobserved heterogeneity at the whisky level by estimating the following fixed effect model:

$$eval_{iw} = \sum_i \theta_i * expert_i + \delta_w + \varepsilon_{iw} \quad (2)$$

where δ_w is a whisky specific heterogeneity term. The fixed effect estimates indicate whether experts set different marks on average net of any composition effect (either on the basis of observable or unobservable characteristics) in the whisky they taste.

Estimates from OLS and fixed effect regressions are reported in Table 1. Without control of whisky features (column 1), we find significant differences in experts' average scores. Among the most frequent tasters, marks given by Jim Murray are on average 4.2 points higher compared to those of Dave Broom[2]. Ian Wisniewski (+3.9), Gavin Smith (+3.5) and Michael Jackson (+3.0) also provide, on average, better scores than Dave Broom when tasting a whisky. Conversely, tasting notes given by Dominic Roskrow are significantly lower (-3.5). Nevertheless, the role of experts remains limited as it explains about 6.4% in the variation in scores.

Obviously, these differences may be due to a non-random allocation of the whiskies tasted. Experts may give lower scores because they taste whiskies of lower quality. We thus include the following whisky features in the linear regression explaining the score set by experts: six age groups, a vintage dummy, alcohol content (with a quadratic profile) and a set of brand dummies. These covariates significantly influence the experts' ratings and the R^2 is now about 30%. The coefficients

[2] We find larger differences in the mean score set by experts with very few tastings : +13.9 for David Robertson (9 tastings), +7.3 for Doug McIvor (26 tastings), -9.6 for Ian Buxton (8 tastings), -7.8 for Paul Pacult (11 tastings).

associated to the experts' dummies are reduced when whisky features are controlled for (column 2). The difference in the average mark between Dave Broom and Ian Buxton is for instance no longer significant. Nevertheless, a Wald test shows that the assumption of null expert coefficients remains rejected.

Table 1. OLS and fixed effect estimates of experts' marks

Variables	(1)		(2)		(3)	
	coef	s.e.	coef	s.e	coef	s.e
Constant	76.858***	0.227	63.356***	12.264	77.258***	0.172
Experts (ref: Dave Broom)						
Arthur Motley	1.424***	0.541	1.060*	0.572	1.036*	0.616
David Robertson	13.920***	1.259	6.208***	1.771	2.503	2.655
Dominic Roskrow	-3.498***	0.906	-3.500***	1.032	-3.639***	1.167
Doug McIvor	7.334***	1.841	6.262***	1.715	1.341	1.581
Gavin Smith	3.475***	0.777	3.303***	0.930	1.722	1.077
Ian Buxton	-9.608***	1.667	-1.202	2.788	4.000	2.798
Ian Wisniewski	3.867***	0.712	3.087***	1.113	2.385*	1.278
Jim Murray	4.186***	0.625	4.119***	0.619	2.948***	0.552
Kate Ennis	1.342	2.967	-0.950	1.843	0.600	3.539
Marcin Miller	-2.040	1.827	-1.433	2.241	-3.093	2.405
Martine Nouet	-0.678**	0.335	-0.649*	0.346	0.291	0.331
Michael Jackson	2.980***	0.278	2.768***	0.295	1.725***	0.304
Owen Barstow	-1.069	1.867	-2.635	1.950	-0.595	1.917
Paul Pacult	-7.767*	4.320	-11.444***	4.049	-10.547***	2.405
Peter Mulryan	5.142***	1.745	4.275**	1.669	4.588**	1.919
Rob Allanson	1.296	1.322	1.209	1.158	0.769	2.195
Robin Laing	-4.001*	2.174	-3.166*	1.898	-4.282*	2.415
Whisky features	NO		YES		NO	
Whisky fixed effects	NO		NO		YES	
F-test of null expert effects						
Value ; prob	19.93***	0.000	8.61***	0.000	5.56***	0.000
Number of observations	3820		3820		3820	
Number of whiskies	1914		1914		1914	
R^2	0.064		0.292		0.771	

Source: authors' calculations, data from Whisky Magazine

Note: standard errors are adjusted for clustering at the whisky level. Whisky features include six dummies for age, vintage, alcohol content and brand dummies.

We finally estimate a fixed effect regression that takes into account all the unobserved and observed whisky features. Focusing on experts with at least 40 tastings (8 experts), we still observe significant differences in the average mark given by experts (column 3). Jim Murray (+2.9), Ian Wisniewsky (+2.4, significant at the 10% level), Michael Jackson (+1.7) and Arthur Motley (+1.0, significant at the 10% level) give more generous scores than Dave Broom on average, while the reverse pattern is found for Dominic Roskrow (-3.6). Conversely, there is no significant difference in the average score reported by Dave Broom, Martine Nouet and Gavin Smith.

So, our results emphasize the limitations of absolute rankings that arise because of the absence of a universally agreed on boundary condition. It is hence interesting to investigate how a

normalized ranking alters the evaluation from experts. To account for differences in the average score set by experts, we propose from (2) the following normalized score:

$$eval_{iw}^{norm} = eval_{iw} - \sum_i \theta_i * expert_i \quad (3)$$

Estimation of (2) with $eval_{iw}^{norm}$ as dependent variable would give null values for the experts' coefficients θ_i, but note that for two expert $i \neq j$ we have $E(eval_{iw}^{norm}) \neq E(eval_{jw}^{norm})$ given the non random allocation of whiskys.

Now, consider the list of the top whiskys defined as those with an original (non-normalized) score above 90. Among the 101 products, 36 (35.6%) were tasted by Michael Jackson and 31 (30.7%) by Jim Murray, Dave Broom coming third with 12 tastings (11.9%). We notice that both Michael Jackson and Jim Murray are over-represented in this top list since over the 3,820 tastings, their respective contributions were equal to 27.4% and 7.8%. As the same time, the fixed effect estimates reported in column (3) of Table 1 show that these two experts set higher marks on average net of any composition effect.

As shown in Table 2, results from normalized scores substantially differ for the best whiskys. When considering the 91 bottles with a normalized score above 90, we find a reduced contribution from Jim Murray, with 17 tastings (18.9%) instead of 31. At the same time, the top list now includes tastings from both Paul Pacult (5 tastings) and Dominique Roskrow (2 tastings). In fact, these two experts are the most severe ones in terms of marks according to our fixed effect estimates. Once their tasting scores are corrected for this severity bias, we conclude that they contribute to the list of the best whisky tastings.

Table 2. Number of top tastings (with a score above 90), by experts

(1) Non-normalized score (N=101)		(2) Normalized score (N=91)	
Expert	N	Expert	N
Michael Jackson	36	Michael Jackson	36
Jim Murray	31	Jim Murray	17
Dave Broom	12	Dave Broom	12
Martine Nouet	7	Martine Nouet	7
David Robertson	5	Paul Pacult	5
Doug McIvor	4	Doug McIvor	4
Arthur Motley	4	Arthur Motley	4
Peter Mulryan	2	David Robertson	2
		Dominic Roskrow	2
		Owen Barstow	1
		Robin Laing	1

Source: authors' calculations, data from Whisky Magazine

4. Conclusion

The aim of this paper was to estimate the influence of the preferences of experts on their evaluation. Our main result is that net of any composition effect in the whisky they taste, on average experts give different marks. This finding, which is robust to a fixed effect estimation, suggests that

the informative content of marks delivered by experts is poor. Reading that an expert A gives a higher mark than B does not mean that A appreciates the whisky more than B: A may simply give higher marks than B for all tastings. Similarly, a whisky with high scores given by both experts A and B is not necessarily of higher quality than another whisky with lower scores given by both experts C and D if the latter are not generous in their scores. Since readers learn little from absolute notes, a recommendation of our empirical study is that experts should provide a relative assessment in the form of demeaned marks.

References

Ashenfelter O., Jones G., (2000), 'The demand for expert opinions: Bordeaux Wine', *Les Cahiers Scientifiques de l'Observations des Conjonctures Vinicoles Européennes*, 3, 1-17.

Chossat V., Gergaud O., (2003), 'Expert opinion and gastronomy: The recipe for success', *Journal of Cultural Economics*, 27, 127-141.

Ginsburgh V., (2003), 'Awards, success and aesthetic quality in the arts', *Journal of Economic Perspectives*, 17, 99-111.

Hadj Ali H., Lecocq S., Visser M., (2008), 'The impact of gurus: Parker grades and en primeur wine prices', *Economic Journal*, 118, pp. F158-F173.

Nelson P., (1970), 'Information and consumer behavior', *Journal of Political Economy*, 78, 311-329.

Vaillant N., Wolff F.C., (2012), 'Understanding how experts rate cigars: An "havanometric" analysis', *Applied Economics*, forthcoming.