AMERICAN ASSOCIATION OF WINE ECONOMISTS

AAWE WORKING PAPER
No. 98
Economics

TASTERS' BIAS IN WINE GUIDES' QUALITY EVALUATIONS

Stefano Castriota, Daniele Curzi and Marco Delmastro

February 2012

www.wine-economics.org

Tasters' Bias in Wine Guides' Quality Evaluations

Stefano Castriota[1]
Università di Perugia

Daniele Curzi[2]
Università degli Studi di Milano

Marco Delmastro[3]
Autorità per le Garanzie nelle Comunicazioni,

Abstract

Using a new database on Italian wines we show that a guide's quality evaluations are affected by two sources of personal bias, namely generosity and personal preferences towards certain types of wine characteristics.

[1] Via A. Pascoli 20, 06123 Perugia (Italy).
[2] Via Celoria 2, 20133 Milano (Italy).
[3] Via Isonzo 21b, 00198 Roma (Italy).
JEL codes: L15.
Keywords: wine, guide, quality, rating, bias.

1. Introduction

Over the last years the field of wine economics has seen a surge in the amount of research carried on. Apart from the charm exerted by the item itself, the wine market is ideal to conduct economic research since it is characterized by the presence of thousands of small and medium enterprises, enormous variety of products, abundance of information and consequent huge information asymmetries among producers and buyers.

Wine is an experience good because consumers learn only after purchase about the actual quality of the product. As pointed out by Akerlof (1970), this feature can lead to market failures due to the information asymmetry between the producer, who followed each step of the production process (Dubois and Nauges, 2010), and the consumer. In extreme cases this can prevent agents from buying the good.

In order to reduce information asymmetries and avoid market failures wine guides have assumed the function of rating agencies (Hay, 2010)[4]. An extensive literature relying on hedonic price models has shown that the judgment of experts strongly affects the final price of wines (see, among others, Arias-Bolzmann *et al.*, 2003; Lecocq and Visser, 2006; Dubois and Nauges, 2010) since market prices are most often determined before consumers obtain any direct and personal information about the quality of the wine of the current year (Castriota and Delmastro 2011, Landon and Smith, 1997; 1998)[5]. As shown by Ali et al. (2008), famous gurus like Robert Parker can affect even *en-primeur* wine prices. Therefore, wine guides represent a key factor that affects the market price mechanism (Oczkowski, 2001).

In spite of its prominence in shaping market equilibria, a number of studies have started pointing out that wine experts may differ in their opinion (Cliff and King, 1997; Ashenfelter, 2006; Hodgson, 2008). This paper aims at providing new evidence on wine experts' behaviour when forming quality ratings. Using a unique database on

[4] "*Wine critics are the bond-rating agencies of the wine market. Their scores and written reviews give many wine buyers the confidence they need to make what really is a risky purchase*" (M. Veseth, 2011).

[5] A number of other variables, ranging from market to socio-economic context, affect the level and the dispersion of wine prices. For a recent and exhaustive study on wines sold in the United States see Jaeger and Storchmann (2011).

Italian wines we show that tasters' quality evaluations are affected by personal bias, which leads judges to be systematically more or less generous than their colleagues, and by personal preferences towards certain wine characteristics. Our results are even more interesting if we think that the tasters under scrutiny here work for the same wine guide and share a set of common and agreed tasting and rating rules.

2. Data and econometric methodology

We collected data over all the wines present in the Italian guide *I Vini di Veronelli* from 2004 to 2009 and tasted by the two wine experts in charge of carrying on the evaluations. The final database is made of nearly 50,000 bottles produced by more than 4,000 firms and provides detailed information on:

- taster's ratings (from one to four stars);
- wine characteristics like being rated for the first time (*new entry*), colour (*white/rosé/red*), type of wine (*sweet* or *dry*), quality classification established by national decree (*VDT/IGT/DOC/DOCG*)[6], age of the bottle (*age*), ageing procedure (use of wooden *barrel* and/or *barrique*), and the natural logarithm of the total *number of bottles* of the wine (label) under scrutiny produced by the firm;
- winery size (measured as the natural logarithm of the *number of hectares* of vineyard cultivated) and the ownership structure (*private* or *cooperative* firm).

The aim of this paper is to study the determinants of wine quality and verify whether there is a taster's bias in terms of higher/lower generosity and/or personal preferences towards certain type of products. In order to do this we rely on ordered logit regressions with robust standard errors where the dependent variable is the number of stars awarded by the guide: this variable ranges from one (minimum grade)

[6] Wine classifications established by national or international decree represent an institutional response aimed at guaranteeing consumers a minimum expected level of quality. In the EU there exists a classification of wines based on two broad categories, quality wines (i.e. VQPRD, *Vins de Qualité Produits dans les Regions Determinées*) and table wines, where quality wines are mainly identified with the origin of grapes. Italian wines are classified into four categories (from the lowest to the highest level of quality): *Vini Da Tavola* (VDT – table wines), *Indicazione Geografica Tipica* (IGT – typical geographic indication), *Denominazione di Origine Controllata* (DOC – controlled denomination of origin) and *Denominazione di Origine Controllata e Garantita* (DOCG – controlled and guaranteed denomination of origin).

to four (maximum). Regressors include all the variables mentioned above, in addition to region and year dummies to control for unobserved elements (see the first two columns of Table 1 for a detailed description).

Wine ratings are provided by two tasters, renamed A and B to protect their privacy. As usual, tastings are blind and wines are randomly assigned to the two judges, which avoids self-selection and reverse causality problems.

3. Results

Given the discrete nature of the dependent variable in Table 1 we start our empirical investigation with a standard ordered logit model, that is a model which imposes the parallel lines restriction (see Williams, 2006). All the specifications include region, year and denomination (IGT, DOC and DOCG) dummy variables. Results of control variables are not shown for reasons of space but are available upon request. *White*, *Dry*, *Steel*, *Private* firm, *Taster B*, *North-East*, year *2004* and region *Calabria* are the omitted variables to avoid perfect multicollinearity. The number of available observations differs among variables and is smallest for the use of wooden *barrel* and *barrique*. Therefore, the use of the full set of regressors reduces considerably the sample size. Nevertheless, the results obtained in the four specifications appear to be very stable. The first regression contains all the control variables, except *barrel* and *barrique*. Then, in the second specification we move to the complete set of regressors which includes the last two variables. Finally, in the last two columns we include slope dummy variables to control for tasters' bias due to generosity and personal preferences.

Looking at the complete set of regressions we can see that wines rated for the first time (*new entry*) receive a lower score. With respect to wine characteristics, *red* and *sweet* wines receive higher grades, contrary to what happens to *rosé* ones. *Age*, *barrel* and especially the use of *barrique* have a positive impact on quality rating, meaning that costly investments to increase quality are rewarded. Turning to firm characteristics, size, measured as *number of hectares* of vineyard, is positively correlated to quality rating. On the opposite, the *number of bottles* of each label

produced by firms is inversely related to quality[7]. Finally, there is no significant qualitative difference between wines produced by *cooperatives* and other firms.

In the last two columns we move to the core of our research. In the third regression we add the dummy variable *taster A* which turns out to be positive and strongly significant: taster A is systematically more generous than taster B[8]. Next, in the last column we add interaction variables to verify whether the two tasters have different preferences towards most common wine characteristics which can contribute to produce systematically different judgments. It turns out that, interacting the dummy variable *taster A* with variables for the wine colour and sweetness and the macro-region of production (*North West*, *North East* and *Centre*), taster A shows a clear preference for red and, especially, sweet wines. Furthermore, with respect to taster B, taster A seems to prefer wines from North-East. No difference emerges with respect to wines from the South, while those from North-West and Centre of Italy display a relative preference by taster B.

4. Robustness

To check the robustness of our results we first run separate regressions by type of wine: white vs. red and sweet vs. dry. Results are similar, even for sweet vs. dry where the number of observations in the two categories is very different (the production of sweet wines is a limited niche).

Next, in order to control whether the parallel line assumption imposed by the ordered logit model is violated, we run a Brant test. As it often happens, the Brant test rejects the parallel lines assumption. Even if this result is largely accepted in the literature, to go ahead with wrong parameter restrictions can lead to incorrect, incomplete or misleading results (Williams, 2006). Therefore, we repeat the exercise

[7] This variable should not be confounded with the total number of bottles of all labels produced by the company as whole which is strongly correlated with the number of hectares of land.
[8] Cao and Stokes (2010) analyze data from a major U.S. Wine Competition and show that (i) there are a number of judges with positive or negative bias with respect to the other evaluators and (ii) that a number of judges have a limited ability to discriminate between wines of different quality. In their case, however, jury members come from different institutions and backgrounds while in our the evaluation criteria should be identical.

by estimating the same specification with the partial proportional odds model, where we relax the assumption of parallel lines and allow those coefficients which significantly differ among the three contrasts to vary. We obtain three binary logistic regressions, where we compare 1 vs. 2-3-4 stars, 1-2 vs. 3-4, 1-2-3 vs. 4. Table 3 provides the results of these estimates: as assumed by the partial proportional odds model, some coefficients are identical while others are not. Main results still hold true, Taster A being systematically more generous than Taster B and preferring red, sweet and North-Eastern wines and disliking those from North-West and Centre.

5. Conclusions

We build a new database that provides detailed information over the characteristics and quality ratings of nearly 50,000 bottles of wine produced by more than 4,000 wineries and tasted by two experts of the eldest Italian guide (*I Vini di Veronelli*). We provide clear evidence that quality ratings are affected by two sources of tasters' bias, the first related to the systematic difference in generosity between the two judges and the second to the personal preferences of each taster towards certain types of wine. These results are even more interesting if we think that the two tasters work for the same wine guide and share a set of agreed tasting and grading rules. Results can be easily generalized to other contexts, such as the financial market, where information asymmetries and rating agencies play a fundamental role.

Acknowledgements

We thank Alberto Nastasi for precious suggestions and gratefully acknowledge financial support of MIUR – FISR project INTERAGRO (metodi e strumenti per la supply chain INTEgRata nell'AGROalimentare). The views expressed herein by Marco Delmastro are the sole responsibility of the author and cannot be interpreted as reflecting those of the Autorità per le Garanzie nelle Comunicazioni.

References

[1] Akerlof, G., 1970. The Market for 'Lemons': Quality Uncertainty and the Market Mechanism. The Quarterly Journal of Economics. 84, 488-500.

[2] Ali, H.H., Lecocq, S., Visser, M., 2008. The Impact of Gurus: Parker Grades and En Primeur Wine Prices. The Economic Journal, 118, 158-173.

[3] Arias-Bolzmann, L., Sak, O., Musalem, A., Lodish, L., Báez, R. K., De Sousa, L.J., 2003. Wine Pricing: The Influence of Country of Origin, Variety, and Wine Magazine Ratings. International Journal of Wine Marketing. 15 (2), 47 – 57.

[4] Ashenfelter, O., 2006. Tales from the Crypt: Bruce Kaiser Tells us about the Trials and Tribulations of a Wine Judge. Journal of Wine Economics. 1, 173–175.

[5] Cao, J., Stokes, S.L., 2010. Evaluation of Wine Judge Performance through Three Characteristics: Bias, Discrimination, and Variation. Journal of Wine Economics, 5 (2), 1-11.

[6] Castriota S., Delmastro M., 2011. Seller Reputation: Individual, Collective and Institutional Factors. Journal of Wine Economics, forthcoming.

[7] Cliff, M.A., King, M.C., 1997. Application of Eggshell Plots for Evaluation of Judges at Wine Competition. Journal of Wine Research. 8, 75–80.

[8] Dubois, P., Nauges, C., 2010. Identifying the Effect of Unobserved Quality and Expert Reviews in the Pricing of Experience Goods: Empirical Application on Bordeaux Wine. International Journal of Industrial Organization. 28 (3), 205-212.

[9] Hay, C., 2010. The Political Economy of Price and Status Formation in the Bordeaux en Primeur Market: the Role of Wine Critics as Rating Agencies. Socio-Economic Review. 8 (4), 1-23.

[10] Hodgson, R.T., 2008. An Examination of Judge Reliability at a Major U.S. Wine Competition. Journal of Wine Economics. 3, 105–113.

[11] Jaeger, D.A. and Storchmann, K. (2011). Wine Retail Price Dispersion in the United States: Searching for Expensive Wines? American Economic Review: Papers and Proceedings, 101(3), 136-141.

[12] Landon, S., Smith, C., 1997. The Use of Quality and Reputation Indicators by Consumers: The Case of Bordeaux Wine. Journal of Consumer Policy. 20, 289-323.

[13] Landon, S., Smith, C., 1998. Quality Expectations, Reputation, and Price. Southern Economic Journal. 64, 628-647.

[14] Lecocq, S., Visser, M., 2006. Spatial Variations in Weather Conditions and Wine Prices in Bordeaux. Journal of Wine Economics. 1, 114–124.

[15] Oczkowski, E., 2001. Hedonic Wine Price Functions and Measurement Error. Economic Record. 77, 374-382.

[16] Veseth, M., 2011. *Wine Wars. The Curse of the Blue Nun, the Miracle of Two Buck Chuck, and the Revenge of the Terroirists,* Rowman & Littlefiled Publishers.

[17] Williams, R., 2006. Genealized Ordered Logit/Partial Proportional Odds Model for Ordinal Dependent Variables. The Stata Journal. 6 (1), 55-82.

Table 1: The determinants of wine quality grades, full sample

Variable name	Description of the variables used	(1)	(2)	(3)	(4)
New entry	DV=1 if the wine is rated for the first time	-0.889***	-0.788***	-0.753***	-0.755***
		(0.0247)	(0.0402)	(0.0404)	(0.0404)
Red	DV=1 if the wine is red	0.837***	0.396***	0.444***	0.378***
		(0.0223)	(0.0370)	(0.0376)	(0.0416)
Rosè	DV=1 if the wine is rose	-0.523***	-0.337***	-0.305***	-0.322***
		(0.0717)	(0.0988)	(0.101)	(0.113)
Sweet	DV=1 if the wine is sweet	0.406***	0.370***	0.462***	0.217***
		(0.0491)	(0.0753)	(0.0749)	(0.0809)
Age	Age of the wine in years	3.228***	3.667***	3.545***	3.545***
		(0.0940)	(0.146)	(0.147)	(0.147)
(ln) bottles	Natural log of the bottles of the label produced	-0.264***	-0.125***	-0.139***	-0.144***
		(0.00952)	(0.0152)	(0.0154)	(0.0155)
(ln) hectares	Natural log of the hectares owned by the firm	0.169***	0.124***	0.131***	0.134***
		(0.00968)	(0.0153)	(0.0154)	(0.0155)
Cooperative	DV=1 if the wine is produced by a cooperative	0.291	0.0228	-0.0222	-0.0561
		(0.198)	(0.270)	(0.269)	(0.269)
(ln) hectares* coop	Interacted term	-0.137***	-0.101**	-0.0915**	-0.0863*
		(0.0341)	(0.0464)	(0.0465)	(0.0463)
Barrel	DV=1 if the wine is aged in wooden barrels		0.529***	0.553***	0.556***
			(0.0385)	(0.0391)	(0.0392)
Barrique	DV=1 if the wine is aged in wooden barriques		1.388***	1.369***	1.373***
			(0.0361)	(0.0363)	(0.0364)
Taster A	DV=1 if the wine has been evaluated by taster A			1.344***	1.372***
				(0.0373)	(0.0810)
Taster A * red	Interacted term				0.414***
					(0.0757)
Taster A * rosè	Interacted term				0.124
					(0.235)
Taster A * sweet	Interacted term				1.403***
					(0.175)
Taster A * North-West	Interacted term				-0.398***
					(0.108)
Taster A * Centre	Interacted term				-0.690***
					(0.101)
Taster A * South	Interacted term				-0.104
					(0.108)
Cut 1		-2.632***	0.046	0.070	0.039
		(0.129)	(0.325)	(0.333)	(0.334)
Cut 2		0.638***	3.679***	3.895***	3.884***
		(0.128)	(0.326)	(0.334)	(0.335)
Cut 3		2.126***	5.276***	5.565***	5.564***
		(0.130)	(0.327)	(0.335)	(0.336)
N		47,227	21,103	21,103	21,103
Pseudo R-sq		0.119	0.174	0.203	0.206

Notes: results come from ordered logit regressions with robust standard errors (in brackets). Regressions are run with the *ologit* STATA command which estimates the proportional odds/parallel lines model. Regressors with *** are significant at 1% level, with ** at 5% and with * at 10%. All the specifications include region, year and denomination (IGT, DOC and DOCG) dummy variables. Results are not shown for reasons of space but are available upon request.

Table 2: The determinants of quality grades, by type of wine

Regresors	(1) Red	(2) White	(3) Sweet	(4) Dry
New entry	-0.746***	-0.690***	-0.741***	-0.747***
	(0.0528)	(0.0717)	(0.157)	(0.0418)
Red				0.350***
				(0.0422)
Rosè				-0.323***
				(0.113)
Age	3.422***	3.641***	4.241***	3.507***
	(0.139)	(0.801)	(0.651)	(0.152)
(ln) bottles	-0.195***	-0.122***	-0.0322	-0.149***
	(0.0204)	(0.0286)	(0.0706)	(0.0162)
(ln) hectares	0.145***	0.162***	0.104*	0.136***
	(0.0203)	(0.0280)	(0.0622)	(0.0161)
Cooperative	-0.298	0.618	-2.584*	0.0614
	(0.348)	(0.496)	(1.545)	(0.275)
(ln) hectares* coop	-0.0562	-0.200**	0.349	-0.107**
	(0.0608)	(0.0846)	(0.255)	(0.0476)
Barrel	0.566***	0.957***	0.295	0.584***
	(0.0440)	(0.113)	(0.252)	(0.0401)
Barrique	1.479***	1.444***	0.391***	1.460***
	(0.0446)	(0.0810)	(0.146)	(0.0381)
Taster A	1.694***	1.598***	2.882***	1.358***
	(0.125)	(0.116)	(0.379)	(0.0825)
Taster A * red				0.427***
				(0.0766)
Taster A * rosè				0.108
				(0.237)
Taster A * North-West	-0.375**	-0.378	-0.737	-0.375***
	(0.146)	(0.233)	(0.452)	(0.112)
Taster A * Centre	-0.757***	-0.552***	-0.0772	-0.705***
	(0.142)	(0.179)	(0.480)	(0.104)
Taster A * South	-0.111	0.207	-1.376***	-0.0397
	(0.163)	(0.174)	(0.518)	(0.111)
Cut 1	-1.672***	2.244***	-1.150	0.054
	(0.405)	(0.907)	(1.475)	(0.349)
Cut 2	2.101***	6.697***	2.627*	3.941***
	(0.406)	(0.903)	(1.470)	(0.350)
Cut 3	3.772***	8.750***	4.263***	5.631***
	(0.407)	(0.896)	(1.467)	(0.351)
N	12,881	6,495	1,235	19,868
Pseudo R-sq	0.206	0.168	0.200	0.211

Notes: results come from ordered logit regressions with robust standard errors (in brackets). Regressions are run with the *ologit* STATA command which estimates the proportional odds/parallel lines model. Regressors with *** are significant at 1% level, with ** at 5% and with * at 10%. All the specifications include region, year and denomination (IGT, DOC and DOCG) dummy variables. Results are not shown for reasons of space but are available upon request.

Table 3: The determinants of wine quality grades, without parallelism restriction

Regressors	(1) 1 vs. 2-3-4 stars	(2) 1-2 vs. 3-4 stars	(3) 1-2-3 vs. 4 stars
New entry	-0.810***	-0.530***	-0.158
	(0.0426)	(0.0688)	(0.115)
Red	0.209***	0.791***	1.035***
	(0.0457)	(0.0807)	(0.158)
Rosè	-0.320***	-0.320***	-0.320***
	(0.111)	(0.111)	(0.111)
Sweet	0.138	0.502***	0.933***
	(0.0850)	(0.128)	(0.206)
Age	3.143***	3.739***	3.614***
	(0.132)	(0.108)	(0.121)
(ln) bottles	-0.101***	-0.210***	-0.210***
	(0.0181)	(0.0232)	(0.0367)
(ln) hectares	0.0873***	0.196***	0.241***
	(0.0174)	(0.0205)	(0.0311)
Cooperative	-0.115	-0.115	-0.115
	(0.277)	(0.277)	(0.277)
(ln) hectares* coop	-0.0729	-0.0729	-0.0729
	(0.0476)	(0.0476)	(0.0476)
Barrel	0.793***	0.386***	0.152*
	(0.0515)	(0.0553)	(0.0850)
Barrique	1.378***	1.378***	1.378***
	(0.0368)	(0.0368)	(0.0368)
Taster A	1.911***	1.025***	0.730***
	(0.122)	(0.140)	(0.218)
Taster A * red	0.167	0.518***	0.813***
	(0.126)	(0.135)	(0.221)
Taster A * rosè	-0.129	-0.129	-0.129
	(0.286)	(0.286)	(0.286)
Taster A * sweet	1.512***	1.512***	1.512***
	(0.212)	(0.212)	(0.212)
Taster A * North-West	-0.239*	-0.239*	-0.239*
	(0.125)	(0.125)	(0.125)
Taster A * Centre	-1.138***	-0.511***	-0.379**
	(0.150)	(0.124)	(0.152)
Taster A * South	-0.188	-0.188	-0.188
	(0.123)	(0.123)	(0.123)
Constant	-0.236	-3.770***	-5.798***
	(0.353)	(0.377)	(0.465)
N	21,103	21,103	21,103
Pseudo R-sq	0.214	0.214	0.214

Notes: results come from ordered logit regressions with robust standard errors (in brackets). Regressions are run with the STATA *gologit2* command which relaxes the parallel lines assumption and estimates the partial proportional odds model. Regressors with *** are significant at 1% level, with ** at 5% and with * at 10%. All the specifications include region, year and denomination (IGT, DOC and DOCG) dummy variables. Results are not shown for reasons of space but are available upon request.