

Crowdsourcing the assessment of wine quality

Authors: Orestis Kopsacheilis, Pantelis P. Analytis, Karthikeya Kaushik, Stefan M. Herzog, Bahador Bahrami and Ophelia Deroy

Introduction

The task of inferring a wine's quality, prior to actually tasting it, can appear daunting. Consider a prospective consumer, staring down the wine-section aisle of a supermarket, confronted with a seemingly endless variety of options. What information does she have in her disposal to navigate through this choice overload? The casual wine consumer has traditionally relied heavily on either the price of the bottle at hand and/or, when available, on the score of an expert wine critic that the label features.

Despite their seeming allure, both of the aforementioned mechanisms come with considerable limitations. Using the market mechanism of price to infer the quality of a wine has been shown to be a useful but imperfect heuristic. In a relevant meta-analysis, Oczkowski and Doucouliagos (2015) report a 0.30 price-quality correlation, a positive but moderate effect. Moreover, price is by definition an expensive rule-of-thumb that can lead to disappointments (relatively low quality at high prices) and restrict the possibilities of finding good bargains (relatively high quality at low prices).

Likewise, experts' opinions can be problematic for a variety of reasons. For example, there is no abundance of reliable experts and most wine labels do not feature such information. Moreover, experts' scores can be mis-aligned with consumers' expectations. This can be because experts' and non-experts' tastes often differ (Goldstein, 2008). Or, it might be because experts' scores capture the wine's potential rather than immediate value. Many wines have aging potential, reflecting ongoing chemical processes that persist well after fermentation ends (Goode, 2005), and would have further changed if the consumer had not abruptly stopped its maturation process. Those wines would feel astringent and potentially unpleasant if judged for their hedonic value prematurely.

Recently, thanks to the proliferation of the internet, the perplexed wine buyer has yet another source of information on her disposal. Online wine apps such as Vivino, Cellar Tracker and Wine Searcher crowdsource every-day consumers' reviews providing instant information for a wine's quality, as this has been assessed by large numbers of non-expert consumers. Instead of relying on a few expert critics, such platforms outsource the task of wine evaluation to a large and heterogeneous network of people, leveraging the principle of crowd wisdom (Surowiecki, 2005). Crowdsourcing ratings has shown promising results in a variety of settings, ranging from biomedical applications (Ranard et al., 2014) to launching new entrepreneurial endeavors (Mollick, 2014).

In this paper we explore the benefits of harnessing the wisdom of the oenophilic crowds by exploring the determinants of such online scores. Specifically, we focus on the following two dimensions of Vivino average ratings: how they compare with experts' scores and how they respond to the 'weather conditions' of the year the grapes that made the wine were harvested and 'ageing' of the wine in the bottle prior to it being consumed. Although in matters of taste the role of subjectivity cannot be overstated (Müller-Trede et al., 2018), those two parameters enjoy a broad consensus regarding their importance and their relation to the quality of a wine (Jones and Storchmann, 2001; Ashenfelter, 2008). Therefore, examining how online ratings respond to such established quality markers can be interpreted as an indirect test for their reliability.

Analysis

To address these questions, we collect more than 200,000 reviews from Vivino - a wine-app that was founded in 2010, and as of today approximately 46 million users have provided more

than 62 million evaluations of 12 million wine labels. Our analysis focuses on red wines from Bordeaux with vintages ranging from 2001 to 2017. For every wine name on this list, we obtained Vivino links for every available vintage in the targeted period. We scraped these vivino links for reviews given by amateur raters. Reviews range on a 1-5 scale, with 1 being the lowest and 5 the highest possible score. We compliment this data set with information regarding weather conditions for the year that each wine was produced (vintage) and age.

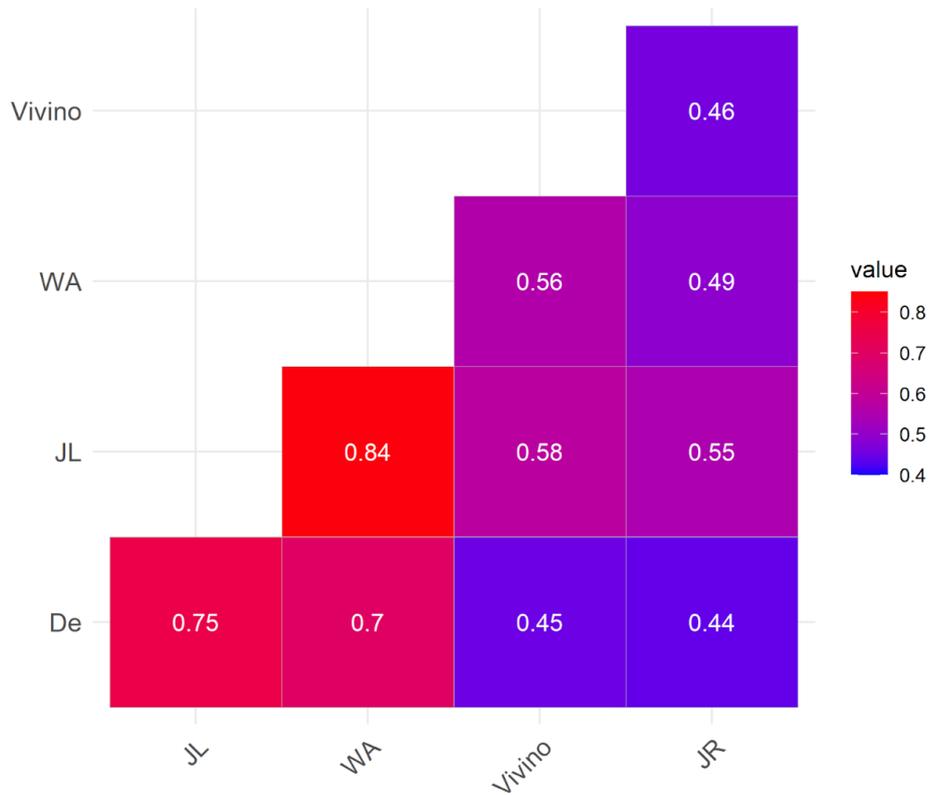


Figure 1. Spearman's rho correlation coefficients between Vivino users' scores and experts' opinions from professional wine critics such as: Jancis Robinson (JR), the Wine Advocate (WA), Decanter (De) and Jeff Leve (JL). Scores have been normalized.

First, we turn our attention to the relation between Vivino's ratings and experts' scores. Figure 1 reports partial correlations between Vivino users' scores and a representative list of experts'. Although the partial correlations reveal a significant overlap in tastes between expert and non-expert opinions, the relation is far from perfect.

Overall, Vivino's average ratings correlate the most with Jeff Leve's scores (Spearman's rho=.58). To shed more light on the source of variation between experts' and non-experts tastes, we compare yearly averages of Vivino scores with Jeff Leve's overall assessment of a vintage quality. The results of this analysis can be seen in Figure 2 where bar-plots capture Vivino ratings (scores are arranged in ascending order) while colour shading captures Jeff Leve's assessment for each vintage's overall quality (lighter colours correspond to higher scores).

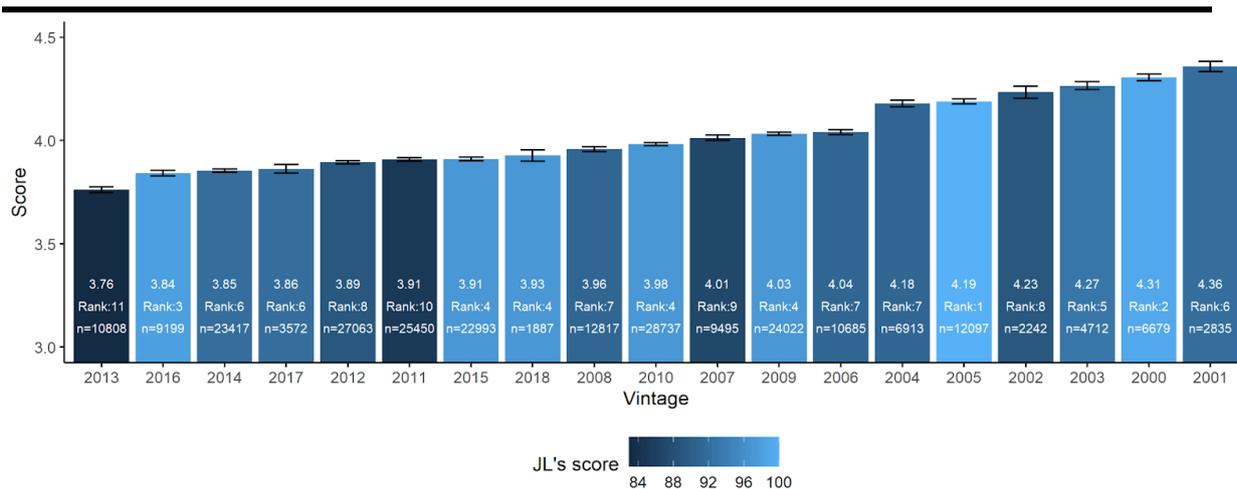


Figure 2. Average scores across vintages. Bar-plots capture Vivino ratings (scores are arranged in ascending order) while colour shading captures Jeff Leve's assessment for each vintage's overall quality (lighter colours correspond to higher scores).

If tastes were in perfect overlap we would expect to see lighter colours towards the right-end of the plot. Instead, we find that Vivino users -unlike Jeff Leve- exhibit a systematic preference for older vintages whose peak performance is closer to the consumption date.

A potentially confounding factor to this analysis is price. Older vintages are likely to be more expensive and therefore reflect aspects of quality beyond the maturation cycle of tannins. To avoid such confounds we use a multiple linear regression for our main analysis that allows us to control for such confounding variables. This is a similar approach as that followed by Ashenfelter (2008). Ashenfelter's econometric model, often referred to as the Bordeaux equation, explains auction prices of wines based on the wine's age as well as key weather conditions during the year in which the grapes were grown. This model has proven surprisingly effective at assessing the quality of Bordeaux vintages and predicting prices of mature wines. A key difference with our estimation method is the choice of the dependent variable where instead of price, we use average online ratings. A second important difference with Ashenfelter's analysis relates to the observation period which is in our data set is shifted by approximately 60 years.

Dependent Variable:	log(Price)	Rating
<i>Variables</i>		
(Intercept)		2.010*** (0.1027)
age	0.023*** (0.007)	0.0194*** (0.0008)
AvgTempGrow	0.608*** (0.116)	0.0704*** (0.0063)
RainAug	-0.003*** 0.0009	-0.0008*** (8.42×10^{-5})
RainPrecedGrow	0.0015*** 0.0005	0.0001*** (3.61×10^{-5})
TempSept	0.0076 (0.0565)	-0.0175*** (0.0027)
log(price)		0.2100*** (0.0029)
Observations		8,244
R ²	0.828	0.48766

Table 1. Left column: output from the regression model in Ashenfelter (2008). The (natural logarithm of) price is regressed on the wine's age and a set of climate variables. Right column: output from the regression model in current study. The annual average rating of a wine is regressed on climate variables, the (natural logarithm of) price and age of the wine. 'AvgTempGrow': average temperature during the growing season (April - September); 'RainAug': precipitation during August; 'RainPrecedGrow': precipitation during the months preceding growth (October - March). 'TempSept': temperature during September.

The results of our regression analysis are displayed in Table 1. Similar to Ashenfelter (2008), we find aging to have a beneficial impact on a wine's appreciation as older vintages collect on average higher scores - even after controlling for price. We also include variables capturing important weather conditions. Again, similar to Ashenfelter (2008) we find that subjective perceptions of wine quality (as proxied by review scores rather than prices in this analysis) react positively to rain preceding growth but negatively when rain stretches out during veraison. Moreover, we find a significantly positive impact of higher temperatures during the growing season. These results are in line with past literature (Jones and Davis, 2000; Jones and Storchmann, 2001; Ashenfelter, 2008; van Leeuwen et al., 2009). However, unlike previous findings, we find a significant negative impact on the wine's rating when high temperatures persist until September.

Discussion and future work

Is there value in delegating the assessment of wine-quality to the wisdom of the oenophilic crowds? Exploring a novel, rich data set from online Vivino reviews we find important insights that can begin to address this important question. In the first part of our analysis we examine how such aggregated amateur ratings compare to experts' opinions. Our analysis revealed substantial but imperfect correlation between experts and non-experts, which in turn suggests that there is novel information to be harnessed by crowdsourcing amateur reviews.

To fully understand the value of crowdsourced reviews we need to better understand the mechanism leading to differences in the ratings of critics and amateur raters in crowdsourcing platforms. The most straight-forward explanation is that the observed differences between crowdsourced reviews and experts reflect differences in tastes between experts and non-

experts. After all, even experts' opinions often diverge with Jancis Robinson's scores being the more divergent ones in our data set.

We propose that there might be more to this story. Specifically, our more detailed analysis suggests that yet another source for this apparent dissonance between experts' and non-experts' opinions is one of scope. Experts' assessments are more forward-looking while non-experts' ones tend to focus mostly on the immediate hedonic aspects of consumption rather than projections of quality development. Younger wines are scored systematically lower by Vivino users compared to experts, presumably because tannins in young wines render their flavor astringent. Conversely, experts' scores are likely to be incorporating forecasts of future value potential at the time of tasting.

However, there are two alternative explanations that are also consistent with our findings. First, if the wines that are sold in Vivino are only the best wines that survive from a vintage, our results could reflect a selection bias. Second, the results could be the product of psychological biases in consuming a wine that is rare and expensive. For example, the rating may not only reflect the quality of the wine, but also the joy or surprise related to opening a bottle much later than when it was produced. Our goal in this project is to extend our analysis in order to disentangle between those competing mechanisms.

In the second part of our analysis, we explore how Vivino ratings respond to "weather conditions" and the "age" of the wine. Using the econometric model introduced by Ashenfelter (2008), but replacing the dependent variable with subjective ratings rather than prices, we find that amateur ratings respond to these variables in qualitatively similar ways as prices do in Ashenfelter's analysis. Specifically, we find aging to have a beneficial impact on a wine's appreciation as older vintages collect on average higher scores - even after controlling for price. Moreover, non-experts' reviews seem to be able to discern meteorological conditions that are widely accepted to influence the quality of the grapes that make the wine. Specifically, review scores react positively to rain preceding growth but negatively when rain stretches out during veraison. Moreover, we find a significantly positive impact of higher temperatures during the growing season. These results are in line with past literature (Jones and Davis, 2000; Jones and Storchmann, 2001; Ashenfelter, 2008; van Leeuwen et al., 2009) and so, to the extent that weather conditions and wine-ageing reflect near-objective quality markers, our findings suggest that the wisdom of the oenophilic crowd is surprisingly insightful.

A noticeable exception to previous literature's findings is with respect to the effect of high temperatures when these persist past the veraison period. Unlike the analysis of Ashenfelter (2008), we find that higher temperatures in September have a detrimental impact on the assessment of wines whose grapes were harvested that year. We suspect that the cause for this apparent disparity is the fact that our observation period is set 60 years past that of Ashenfelter. Over the past decades, climate change has raised average temperatures and extended the periods in which they persist. Previously, global warming has been associated with positive effects on central European wines (Jones and Storchman, 2001). From this perspective, our result can be interpreted as early evidence that this relation may start reversing.

References

- Ashenfelter, O. (2008). Predicting the quality and prices of bordeaux wine. *The Economic Journal*, 118(529):F174–F184.
- Goldstein, R., Almenberg, J., Dreber, A., Emerson, J. W., Herschkowitsch, A., Katz, J., et al. (2008). Do more expensive wines taste better? evidence from a large sample of blind tastings. *Journal of Wine Economics*, 3(1):1–9.
- Goode, J. (2005). *The science of wine: From vine to glass*. Univ of California Press.
- Jones, G. V., & Davis, R. E. (2000). Climate influences on grapevine phenology, grape composition, and wine production and quality for Bordeaux, France. *American journal of enology and viticulture*, 51(3), 249-261.
- Jones, G. V. and Storchmann, K.-H. (2001). Wine market prices and investment under uncertainty: an econometric model for bordeaux crus classés. *Agricultural Economics*, 26(2):115–133.
- Jones, G. V., White, M. A., Cooper, O. R., and Storchmann, K. (2005). Climate change and global wine quality. *Climatic change*, 73(3):319–343.
- Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of business venturing*, 29(1), 1-16.
- Müller-Trede, J., Choshen-Hillel, S., Barneron, M., & Yaniv, I. (2018). The wisdom of crowds in matters of taste. *Management Science*, 64(4), 1779-1803.
- Oczkowski, E. and Doucouliagos, H. (2015). Wine prices and quality ratings: A meta-regression analysis. *American Journal of Agricultural Economics*, 97(1):103–121.10
- Ranard, B. L., Ha, Y. P., Meisel, Z. F., Asch, D. A., Hill, S. S., Becker, L. B., Seymour, A. K., and Merchant, R. M. (2014). Crowdsourcing—harnessing the masses to advance health and medicine, a systematic review. *Journal of general internal medicine*, 29(1):187–203.
- Storchmann, K. (2012). Wine economics. *Journal of Wine Economics*, 7(1):1–33.
- Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.
- van Leeuwen, C., Trégoat, O., Choné, X., Bois, B., Pernet, D., & Gaudillère, J. P. (2009). Vine water status is a key factor in grape ripening and vintage quality for red Bordeaux wine. How can it be assessed for vineyard management purposes?. *Oeno One*, 43(3), 121-134.