Crowdsourcing the assessment of wine quality - evidence from Vivino ratings*

Orestis Kopsacheilis[†] Pantelis P. Analytis[‡] Karthikeya Kaushik[§]
Stefan M. Herzog [¶] Bahador Bahrami[‖] Ophelia Deroy [**]

November 28, 2022

Abstract

Prospective wine buyers have mainly relied upon professional critics' reviews and prices to infer wine quality. Crowdsourcing platforms, such as Vivino, have brought about a third source of information, leveraging the opinions from large numbers of amateur wine reviewers and averaging their ratings. We assess how well the 'wisdom of the wine-loving crowds' matches the conventional wisdom of the wine-world by analyzing approximately 85,000 Vivino ratings for a portfolio of Bordeaux red wines. We find that ratings from Vivino amateurs correlate substantially with the ratings of professional critics and are sensitive to weather conditions in much the same way that auction prices of wines are. However, amateurs' correlations with different critics are on average lower than the correlations observed among critics. We show that this is partly due to differences in scope: amateurs focus on the immediate pleasure whereas critics focus on the potential of a wine once it has matured. In addition, our analysis suggests that global warming may already have a discernible negative effect on wine quality as ratings are negatively affected by increased temperatures late in the growing season. In sum, our results demonstrate that crowdsourced ratings are a reliable source of information that can complement existing ones and help generate valuable insights for both consumers and producers.

Keywords: crowd-sourcing; wine quality; wine-ageing; global warming; Vivino

*Orestis Kopsacheilis acknowledges the American Association of Wine Economists for their scholarship. Ophelia Deroy was supported by the NOMIS foundation (acronym—DISE) and the Volkswagen Stiftung grant Co-Sense. Bahador Bahrami and Karthykeya Kaushik were supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (819040—acronym: rid-O).

†Corresponding author. Technical University of Munich. email:orestis.kopsacheilis@tum.de
‡University of Southern Denmark
§University of California, Berkeley
¶Max Planck Institute for Human Development
‖Ludwig Maximilian University
**University of London

1 Introduction

Inferring a wine's quality, prior to actually tasting it, can be a daunting task. Consider a prospective wine consumer staring down the wine-section aisle of a supermarket, confronted with a seemingly endless variety of options. What information can they use to navigate this informationally challenging task and pick a wine that they will enjoy? In the absence of prior experience with a particular wine label, the casual wine consumer has mainly relied on either the price of the bottle at hand and, when available, on the ratings of professional wine critics.[1]

Using prices as a signal for quality is a well-established market mechanism in the literature of behavioral marketing (Rao and Monroe, 1989; Bagwell and Riordan, 1991). However, for the case of wine in particular, the idea that higher prices correspond to better (self-assessed) quality is of limited use. According to a meta-analysis by Oczkowski and Doucouliagos (2015), there is only a moderate, positive price-quality correlation. Moreover, price is by definition an expensive rule-of-thumb that can lead to disappointments (i.e., experiencing relatively low quality when buying more expensive wines) and restrict the opportunities to find good bargains (i.e., relatively high quality when buying less expensive wines).

Likewise, relying on professional critics' opinions comes with its own set of caveats. First, it is not always easy as there are only a few well-recognized ones, and they tend to be selective about the wines they taste. Further, many renowned critics have monetized their reviews by running subscription websites where consumers have to pay to access their reviews. Crucially, critics' scores can be mis-aligned with consumers' expectations. For example, in a sample of blind tastings, Goldstein et al. 2008 show how people's subjective judgments of wine quality varies substantially depending on the tasters' level of training. People with no prior wine-training rate more expensive wines lower but this correlation reverses among people who have had such training (e.g. having

[1]Arguably, sensory information pertaining to the shape of the bottle or the design of the label can also influence consumers' perceived quality (see for e.g. Drichoutis et al. 2017) - but such factors are outside the scope of this analysis.

attended a sommelier course). The authors argue that '*non-expert wine consumers should not anticipate greater enjoyment of (...) a wine simply because it is expensive or is appreciated by experts*'. In fact, differences between what experts and amateurs value in matters of taste have been also observed in other domains of taste, such as in films and Broadway plays (Hirschman and Pieros, 1985; Holbrook, 1999).

In recent years, the proliferation of the Internet has equiped the prospective buyer with yet another tool to inform their choice: crowdsourced online ratings (e.g. Chevalier and Mayzlin 2006). In the world of wine, freely-available, crowdsourcing apps such as Vivino, CellarTracker and Wine-Searcher, have given a voice to amateur wine consumers and have extended the task of wine evaluation to a large and heterogeneous network of people, leveraging the principle of crowd wisdom (see Surowiecki 2005, for a popular book summarizing the benefits of this principle; Analytis et al. 2018; Müller-Trede et al. 2018, for applications in matters of taste). These apps have attenuated the scarcity of information about wines, bringing information about a wine's quality at people's fingertips.

Relying on the opinion of relatively inexperienced crowds instead of a few specialized experts, has shown promising results in domains ranging from political and economic forecasting (Mellers et al., 2014; Budescu and Chen, 2015) to entrepreneurial endeavors (Mollick, 2014) and medical diagnostics (Ranard et al., 2014), to name just a few. Nonetheless, crowdsourced ratings are prone to social influence effects (Le Mens et al., 2018; Muchnik et al., 2013) or even fraud (Luca and Zervas, 2016), that can dilute their value. Thus, assessing the quality and properties of crowdsourced online ratings remains an open scientific challenge that needs to be addressed in different consumer domains.

In this paper we assess how well the 'wisdom of the wine-loving crowds' matches the 'conventional wisdom' of the wine-world. We do this by collecting and analyzing approximately 85,000 individual reviews from Vivino. Vivino is a wine-app that was founded in 2010 and –according to their own webpage– it is the world's most downloaded wine app, featuring millions of reviews of wines from around the world. We focus our analysis on a portfolio of red wine labels from Bordeaux and track their ratings on Vivino from the 2004 to the 2016 vintage (i.e. the year when a wine's grapes were harvested). Bordeaux is arguably the most famous wine region and home to many of the world's best known wine estates. Its world-renowned status renders wines from that region especially popular and frequently reviewed, equipping our analysis with a rich dataset and allows us to connect our results with an extensive body of literature that has also focused on that region (e.g. Jones and Davis 2000; Ashenfelter 2008; Ashton 2013).

In the first part of our analysis we examine whether crowdsourced ratings reflect the established wisdom of wine critics. We find that despite there existing a substantial correlation between Vivino and critics' ratings, this correlation is smaller than that between professional critics' scores. Although this can be partly attributed to differences in taste, we provide evidence for yet another source of this dissociation. Namely, we conjecture that critics judge the future potential of a wine once it has matured while amateurs judge the immediate pleasure of drinking a wine.

In the second part of our analysis we focus on how amateur ratings respond to the wine's age and the weather conditions in the year the grapes grew. Being able to assess a wine's quality based only on such objectively observable parameters has been a key question in the relevant literature for the past 40 years. Ashenfelter's seminal work in the 80's and 90's provided a highly successful econometric model for assessing wine quality of Bordeaux vintages and predicting their prices in auctions based on weather conditions and the wine's age (e.g. Ashenfelter 1987, 1988, 1990; but see Ashenfelter 2008, for an updated version). This research programme, questioning the role of expert-opinion in favor of a statistical model, instigated a heated debate in the circles of the oenophilic world that has been characterized as 'somewhere between violent and hysterical' (Passell 1990, but see also Swets et al. 2000 for a relevant discussion).

In the spirit of Ashenfelter (2008) we estimate a multiple linear regression model, but instead of auctioned prices, we use averaged Vivino online ratings as our dependent variable. Although our observation period succeeds that of

Ashenfelter by more than half a century and although we use individual wines rather than vintages as the unit of analysis, our results are qualitatively very similar.

We find that averaged Vivino ratings react positively to the wine's age. At face value, this finding seems to be in line with the conventional wisdom that (red) wines improve over time as their initially astringent tannins evolve and become smoother. However, it also suggests that Vivino users do not fully incorporate the wine's potential when reviewing the product and focus instead on the basis of its current hedonic value. This corroborates our conjecture that experts' and amateurs' reviews differ in scope.

Furthermore, we show that averaged Vivino ratings mirror the weather conditions that determined the grapes' quality. They respond positively to rain preceding growth and negatively when rain stretches out during the month of August - which is the typical period of veraison, when grapes change color and start the accumulation of sugars. Moreover, we find a positive impact of higher temperatures during the growing season. These results are in line with past literature (see Storchmann 2012, for a relevant review). However, unlike previous findings reporting an unequivocally positive relation between higher temperatures in Bordeaux and wine quality (Jones and Davis, 2000; Jones et al., 2005; Ashenfelter, 2008; Van Leeuwen et al., 2009) we find a potential 'back-fire' effect of increasingly higher temperatures. Specifically, we find that averaged ratings are negatively affected when higher temperatures persist during the month of September. We discuss this novel result in the context of global warming.

2 Analysis

We collect our data on amateur online review ratings from Vivino's public data. We restrict our dataset to a list of red Bordeaux wines that are featured in En Primeur events. These are wine tastings in which critics and merchants are invited to taste wines from the barrel when the wines are only 6–8 months old. The wine can be purchased during this event already, but it will become available to buyers at a later date, once the wine is bottled. This is why En Primeur events are often referred to as 'wine futures'.[2] We further require that these wines are reviewed by Vivino users for every year between 2004 and 2016.[3] Our final dataset consists of 84,751 ratings for 832 wines (64 chateaux that are all observed over 13 years). Our dataset includes all reviews that were provided for each wine at the time of data collection (2020-10-20).

Figure 1 depicts some important, vintage-level summary statistics of our dataset such as vintage-level averaged ratings (left panel) and median prices of the wines in our portfolio (right panel). When comparing Vivino ratings averaged[4] at the vintage level with Jeff Leve's vintage assessment[5], we find a striking resemblance(Figure 1). For example, both Vivino-amateurs and Jeff Leve agree that the 2013 vintage was the worst in recent years. Similarly, both agree that the 2005 vintage was overall exceptional. Both of these claims are widely shared beliefs among the wine community.[6]

Panel B in Figure 1 summarizes information about retail prices across different vintages and their relation to averaged Vivino ratings. The median retail price of our portfolio of Bordeaux reds is €41.95 and does not exhibit extreme fluctuations over the years (min=€40.80; max=€44.92). There is a (slightly) positive relation between prices and ratings, suggesting that produc-

[2]Wines featured in En Primeur events tend to be more popular and therefore more often reviewed, making our dataset richer. Moreover, since critics are frequently attending such events, this selection increases our chances of matching a wine that is reviewed in Vivino with critics' reviews.

[3]Even though Vivino has been in operation only since 2013, the site includes reviews of vintages going further back in time. We selected the period between 2004-2016 for our data collection in anticipation of providing denser data. Given the funding date of Vivino, vintages before 2004 are less likely to be widely reviewed in the app. Similarly, since vintages are typically released in the market 2-3 years after being bottled, we anticipated a relative scarcity of reviews for vintages younger than 2016.

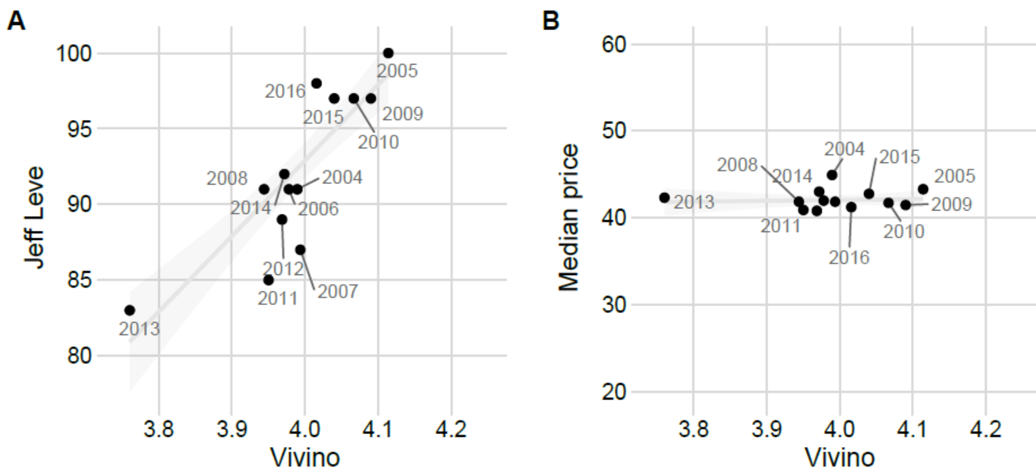
[4]Unless specified otherwise, 'average' always refers to the arithmetic mean.

[5]Jeff Leve is an established professional wine-critic specializing in the Bordeaux region. We collected Jeff Leve's vintage-level ratings from his personal website: http://www.thewinecellarinsider.com

[6]For example, Stéphane Derenoncourt, a French vigneron working as a consultant for numerous estates in Bordeaux, described producing the 2013 vintage as a 'war against nature' (Willsher, 2013). In contrast, the 2005 vintage is often characterised as 'majestic' (Asimov, 2021).

ers anticipated that certain vintages would be in higher demand. However, as we show in subsequent analysis, retail price alone does not capture all the variation in Vivino's averaged ratings.

Figure 1: Vintage-level summary statistics.



Notes: A: Vivino averaged ratings against Jeff Leve's assesment for each vintage's overall quality. Review ratings for Vivino range on a 1-5 scale, with 1 being the lowest and 5 the highest possible score. Jeff Leve's ratings range between 0 (lowest) and 100 (highest). B: Information about the median price of our wine-portfolio across different vintages. Vivino's information about price is based on the average retailer's price and is, therefore, only approximate.

Next, we take a closer look at the relation between ratings from professional critics and amateurs by merging the Vivino-reviews dataset with a dataset containing ratings from several professional critics. We collect information about critics' scores from Bordoverview - a website aggregating critics' ratings provided in Bordeaux - En Primeur events.[7] Table 1 reports on the exact number of wines, from each critic, that we were able to match to a Vivino average. The maximum overlap is between Vivino and Decanter –for which we were able to match 360 wine ratings out of the total 832 wines– while the median overlap between Vivino and a critic is 200. Nonetheless, this sub-portfolio of wines has a very similar median price to that of our full dataset

[7]Link to Bordoverview: https://www.bordoverview.com/

(€41.20 vs. euro41.95) and is therefore representative of the initial portfolio.

Table 1: Number of wines reviewed by Vivino and critics

	Vivino	TA	JR	DE	WA	JA	RG	JS	NM	JL
Vivino	832	168	351	360	223	146	200	208	189	153
TA	168	168	159	164	48	74	52	62	150	141
JR	351	159	351	340	208	139	188	198	179	142
DE	360	164	340	360	214	141	192	202	184	149
WA	223	48	208	214	223	101	147	132	68	44
JA	146	74	139	141	101	146	90	52	82	51
RG	200	52	188	192	147	90	200	126	75	48
JS	208	62	198	202	132	52	126	208	91	67
NM	189	150	179	184	68	82	75	91	189	142
JL	153	141	142	149	44	51	48	67	142	153

Notes: Each cell reports the number of wines from our initial portfolio of Bordeaux, reds that were reviewed in Vivino, for which we were able to find a rating from a professional critic. Vivino: averaged ratings from Vivino users. JR: Jancis Robinson; WA: the Wine Advocate; Jancis Robinson (JR), the Wine Advocate (WA); TA: Tim Atkin; B.T: Bettanne and Desseauve; JS: James Suckling; JL: Jeff Leve; De: Decanter; JA: Jane Anson, RG: Rene Gabriel.

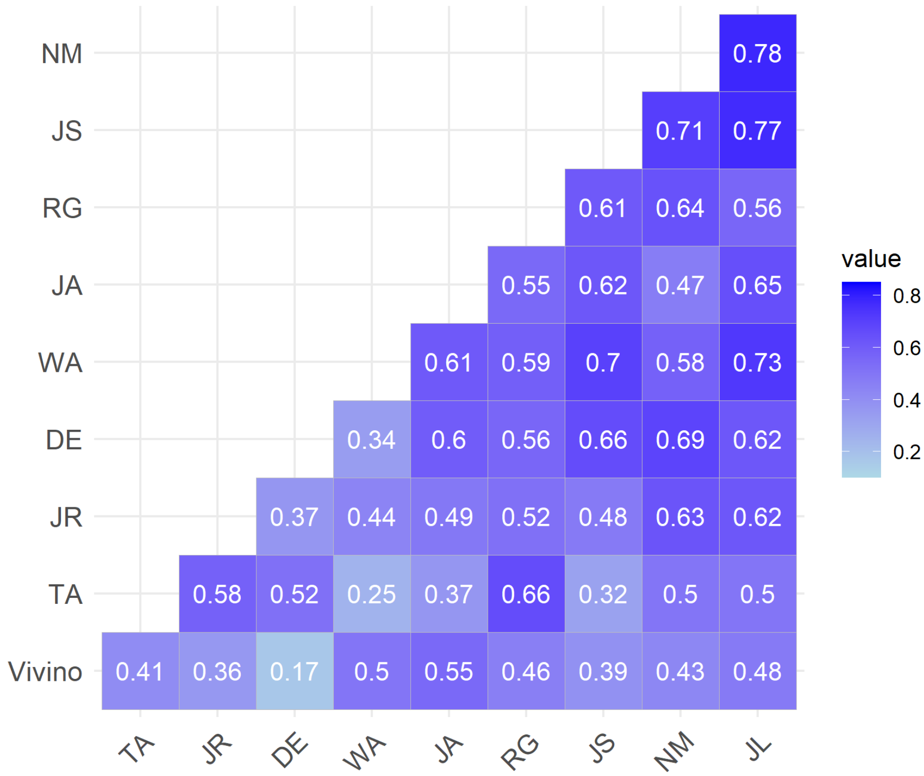
Figure 2 reports on the two-way correlations across pairs of critics.[8] Here, we treat Vivino averaged ratings as an independent critic. To account for differences in scoring systems, we transform all ratings into z-scores. Vivino averages display sizable correlations with professional critics' ratings, ranging from 0.17 with Decanteur to 0.55 with Jane Anson. However, on average, Vivino's average tastes are the furthest from the tastes of professional critics as they exhibit the lowest average correlation (0.42).[9] Among professional critics, Jeff Leve is the one with the highest average correlation (0.63), while Tim Atkin and Jancis Robinson's are the ones with the lowest (0.45 and 0.50

[8]Correlation analysis is commonly employed to study the consensus in the judgments of panels of experts (e.g. Lawless et al. 1997; Ashton 2012).

[9]In matters of fact, average correlations with other experts have been used to identify highly accurate individuals ex-ante (Kurvers et al., 2019), and in matters of taste to identify individuals with a large potential to influence others (Analytis et al., 2020).

respectively).[10]

Figure 2: Correlations between different wine-raters



Notes: Reported values correspond to Pearson's correlation coefficients. All ratings have been transformed to z-scores. Raters have been ordered in ascending order according to their average intercorrelation. Vivino: averaged ratings from Vivino users. JR: Jancis Robinson; WA: the Wine Advocate; Jancis Robinson (JR), the Wine Advocate (WA); TA: Tim Atkin; B.T: Bettanne and Desseauve; JS: James Suckling; JL: Jeff Leve; De: Decanter; JA: Jane Anson, RG: Rene Gabriel

We now turn to the question of how, if at all, averaged Vivino ratings reflect important, observable determinants of wine quality such as the age of the wine and the weather conditions that were present the year that the wine was made. For this part of the analysis we include information about

[10] Also see Masset et al. 2015, for a similar analysis of consensus among critics from a different range of years).

the weather in Bordeaux from Merignac's weather station.[11] Since we do not use critics' reviews in this analysis, we use the full dataset of 832 wines. We also construct a variable for age for each review. 'Age' counts the number of years between the wine's vintage and the year in which a particular review was posted on Vivino.

Our methodology in this section is influenced by Ashefelter's econometric model (2008), often referred to as 'the Bordeaux equation'. Ashenfelter regresses a vintage-level price index (obtained from auctions of a specific wine-portfolio) on a set of weather variables and the wine's age. This model has proven surprisingly effective at assessing the quality of Bordeaux vintages and predicting prices of mature wines (Storchmann, 2012).

Similarly to Ashenfelter (2008) we implement a multiple-regression approach to examine how (perceived) quality is affected by observable criteria such as weather conditions and the age of the wine. However, it is worth pointing out 3 important differences. First, we use the averaged Vivino ratings as our dependent variable, instead of auctioned prices.[12] Second, our observation period succeeds that in Ashenfelter (2008) by approximately 50 years. The observation period in Ashenfelter's analysis covers the years 1952 to 1980 while that in the current paper covers the more recent period between 2004 and 2016. Third, Ashenfelter uses the vintage-level as the unit of analysis (and aggregation), while we aggregate at the level of the age in which the wine was reviewed, using random effects at the level of each wine.

The results of our regression analysis are displayed in Table 2. The difference between models (1) and (2) is that the latter includes (the natural logarithm of) price as an additional control while the former does not. As we see, the inclusion of this additional control does not affect our main results, suggesting that each of our independent variables accounts for significant variation of averaged ratings scores that the retail-price, alone, does not capture.

[11] Source: `http://Infoclimat.fr`

[12] We do include, however, information about the price of a wine as an additional control in our model. Unlike Ashenfelter's analysis, our price-variable is not based on secondary markets (auctions) but rather on the average retailers' price at the time that the wine became first available in the market.

Table 2: Determinants of averaged Vivino ratings

	Dependent variable:	
	Averaged Rating	
	(1)	(2)
log(price)		0.216^{***}
		(0.010)
Age	0.035^{***}	0.032^{***}
	(0.002)	(0.002)
AvgTempGrow	0.188^{***}	0.184^{***}
	(0.032)	(0.025)
RainAug	-0.001^{*}	-0.001^{**}
	(0.0003)	(0.0003)
RainPrecedGrow	0.001^{***}	0.001^{***}
	(0.0002)	(0.0001)
TempSept	-0.049^{***}	-0.049^{***}
	(0.012)	(0.009)
Constant	0.849	0.138
	(0.498)	(0.397)
Observations	5438	5424
Log Likelihood	-2450.181	-2260.451
Akaike Inf. Crit.	4916.363	4538.902
Bayesian Inf. Crit.	4969.172	4598.289
Note:	*p<0.05; **p<0.01; ***p<0.001	

Notes: The (arithmetically) averaged Vivino-rating of a wine is regressed on climate variables, (the natural logarithm of) price and the age of each wine at the time when it was reviewed. The level of aggregation is at the age (measured in years) in which the wine was reviewed. The regression includes random effects at the level of the wine. We used R's lmer package from the lme4 library (Bates et al., 2014) to estimate this regression. The dataset we use for this regression comprises of a total of 5,438 observations(14 of which do not have information about price and, therefore, drop out in the second model). This includes a total of 832 wines (64 Chateaux observed in 13 vintages) that we observe for a period of 6.5 years (on average). Each Vivino rating ranges between 1 and 5 stars. Age is measured in years of distance between the vintage of a wine and the date that it was reviewed. 'AvgTempGrow': average temperature during the growing season (April - September) - measured in °C; 'RainAug': precipitation during August- measured in mm of water accumulated over the entire period; 'RainPrecedGrow': precipitation during the months preceding growth (October - March) -measured in mm of water accumulated over the entire period;'TempSept': average temperature during September - measured in °C. *p<0.05; **p<0.01; ***p<0.001

We find that the older the wine at the time of review, the better the rating, suggesting that aging has a beneficial impact to amateurs' subjective appreciation. This is similar to Ashenfelter (2008) who also finds a positive effect of aging. As we discuss in the next section, the fact that averaged ratings for a wine improve over time, suggests that at the time of evaluation, Vivino users do not fully account for the wine's potential, once the tannins have reached their optimal phase of maturation.

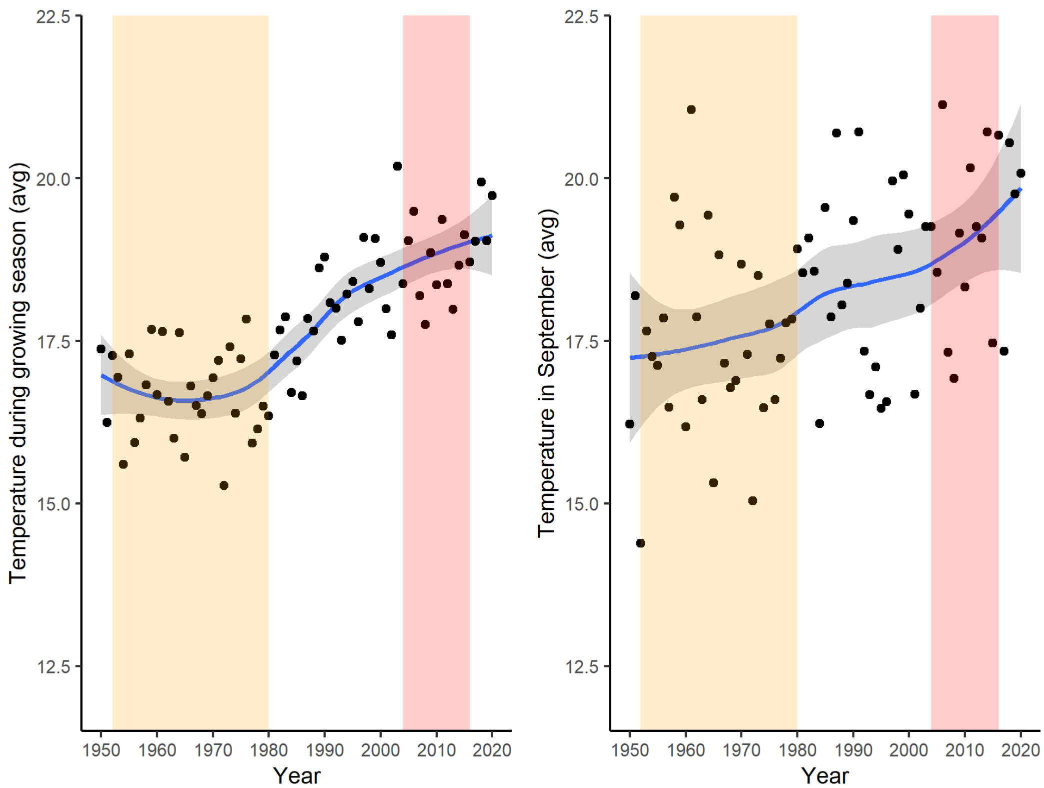
The coefficients of the weather variables also tell a very similar story as those in Ashenfelter (2008). Averaged ratings react positively to rain preceding growth but negatively when rain stretches out during August. Moreover, we find a significant positive impact of higher temperatures during the growing season. However, unlike Ashenfelter's analysis that finds a positive but statistically insignificant coefficient for higher temperatures in September, we find that coefficient to be negative and statistically significant.

Figure 3 shows the evolution of average temperatures in Bordeaux over the past 60 years. The average temperature between the period of Ashenfelter's analysis has increased by 1.6C when compared to our observation period: from 12.5 (average temperature in Bordeaux/Merignac between 1950 and 1980) to 14.1 (average temperature in Bordeaux/ Merignac between 2004 and 2016). As we discuss in the next section, the significant negative coefficient of our variable capturing average temperature during September, may be foreshadowing a 'back-fire' effect of persistently increasing temperatures.

3 Discussion and conclusion

Freely available wine-apps such as Vivino have democratized the production and consumption of information in the wine world, by bringing wine ratings a few clicks away. For many wine consumers, such apps have turned into the main source of information when considering which wines to buy. But how reliable is this crowdsourced assessment of wine quality and how does this 'wisdom of the wine-loving crowds' compare to the 'conventional wisdom' of the wine world? In this paper, we addressed these questions by studying a

Figure 3: Evolution of average temperatures in Bordeaux



Notes: Left panel: average temperatures during the growing season (March - October). Right panel: average temperatures during September. Source: `http://Infoclimat.fr`, weather station: Merignac. Highlighted in orange is the observation period for Ashenfelter (2008), ranging from 1952 to 1980. Highlighted in red is the observation period of the current paper, ranging from 2004 to 2016. Curves and confidence bands show robust LOESS curves (locally estimated scatterplot smoothing using re-descending M-estimator with Tukey's biweight function) and their 95% confidence band.

new and rich dataset of online Vivino reviews, and uncovered a number of new insights about the information captured in these crowdsourced ratings.

First, we showed that crowdsourced ratings of red wines from Bordeaux correlate substantially with ratings from renowned professional critics. Among those, some critics such as Jeff Leve, Jane Anson and the Wine Advocate, seem to be better at getting the pulse of the broader wine-loving audience, as compared to Decanteur or Jancis Robinson. Overall, the observed correlation range is close to those reported in studies investigating experts' consensus in blind-tastings of wine (Cicchetti, 2006; Brien et al., 1987) but slightly lower than those reported in correlational studies using critics' published ratings (Stuen et al., 2015; Ashton, 2013) as well as the average correlation that critics exhibit with each other in our dataset.

There are several mechanisms that can account for this mis-alignment between the ratings of Vivino users and professional critics. Perhaps the most straight-forward one relates to systematic differences in taste. Professional critics tend to share similar background and training which are often very different from those of the general population. As Goldstein et al. (2008) suggest, differences in wine-training can lead to systematic differences in taste. Yet another explanation may relate to asymmetric peer-pressure effects between professional critics and amateur reviewers. According to Stuen et al. (2015), critics who deviate too much from the opinions of others run the risk of being perceived as inept, off-beat, or both and so have an incentive to comply with their peers' ratings. Conversely, amateurs are unlikely to have such image concerns.

We contribute to this literature, partly by identifying another source of systematic divergence. Namely, we suggest that this apparent discrepancy between amateurs' and professionals' tastes is partly due to differences in scope, in that amateurs focus on the immediate pleasure of drinking a wine whereas critics focus on the potential of a wine once it has matured.

Many wines have aging potential, reflecting ongoing chemical processes that persist well after fermentation ends (Goode, 2005), and would thus have further improved had the consumer not abruptly stopped the maturation process by opening the bottle. This aging potential is particularly pronounced for red wines from Bordeaux, which are typically rich in tannins and would feel astringent and potentially unpleasant if judged for their drinking pleasure at a young age. On the other hand, critics' evaluations are prospective, judging how wines will develop over time, rather than how much they can be immediately enjoyed. This is especially true for the wines in our dataset that featured in En Primeur events, whose character emphasizes the wine's potential.

Indeed, our ensuing analysis provides evidence corroborating this hypothesis. Specifically, in the second part of our analysis we explore how averaged Vivino ratings respond to the weather conditions that were present the year the wine was produced as well as to the age of the wine. Mirroring the econometric approach introduced by Ashenfelter (2008), but replacing the dependent variable with subjective ratings rather than auctioned prices, we find that amateurs' ratings respond to these variables in qualitatively similar ways as the auction prices do in Ashenfelter's analyses. 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.

We interpret this finding as supporting evidence to our conjecture that professionals' and amateurs' ratings differ in scope, with amateurs focusing on the immediate hedonic experience and professionals on the wine's potential. If Vivino users were to fully incorporate their forecast about the wine's potential, then the aging coefficient should on average be null rather than significantly positive. Although we do not observe critics' ratings across time within our dataset, evidence supporting the claim that critics' ratings are forward-looking can be found in an analysis of En Primeur re-notations that was published online by the Global Wine Score team (GlobalWineScore, 2017). These re-notations capture critics' adjustments to their En Primeur assessment in subsequent years. If critics' did not account for the wine's improvement with age, we would expect to see –on average– systematic positive adjustments. Instead, the report finds that the net adjustments, averaged across all wines and across all critics for each subsequent year, are either zero or slightly negative.

Our finding has interesting implications for prospective wine buyers. If their intention is to find a good bargain that they can invest for the future, then the critics' rating might be a better guide than crowdsourced ratings. However, if they are invited to a dinner party by peers that same night, then consulting the average online rating of an app like Vivino is more likely to help them make an impression.

With respect to our weather-conditions analysis, we find that amateurs' ratings mirror the meteorological conditions that were present the year in which the wine was produced. Specifically, ratings react positively to rain preceding growth but negatively when rain stretches out during August, when grapes change color and start the accumulation of sugars. Moreover, we find a significant positive impact of higher temperatures during the growing season.

Although acknowledging subjectivity is key in matters of taste (Analytis et al. 2018; Müller-Trede et al. 2018) there is a broad consensus that the weather conditions during growing season are pivotal for the ensuing quality of the wine (Jones and Davis, 2000; Jones and Storchmann, 2001; Ashenfelter, 2008; Van Leeuwen et al., 2009). From this perspective, the sensitivity of amateurs' ratings towards such meteorological phenomena can be interpreted as evidence for their validity.

A noticeable exception to the agreement between our findings and those of previous literature is the effect of high temperatures when these persist later in the growing season. Unlike Ashenfelter (2008), who reports a positive (but statistically not significant) effect, 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 part of the reason for this apparent disparity might be due to the fact that our observation period is set more than half a century past that in Ashenfelter's original analysis.

Over these past decades average temperatures in Bordeaux (and also, globally) have been constantly rising. Literature focusing on viticulture has already pointed out that sufficiently high temperatures can be detrimental for the quality of the grapes. For example, during flowering and throughout the growth of the berries (a period that often includes the month of September), high temperatures can inhibit certain biochemical pathways or physiological processes essential for the production of quality grapes (Deloire et al., 2004). In extreme cases, such high temperatures can cause premature veraison, high grape mortality through abscission, enzyme inactivation and partial or total failure of flavor ripening (Mullins et al., 1992). From this perspective, our analysis might be foreshadowing a 'turn of the tide' to the hitherto unequivocally positive relation between higher temperatures in Bordeaux and wine quality

(Jones and Davis, 2000; Jones et al., 2005; Ashenfelter, 2008; Van Leeuwen et al., 2009).

The trend of increasing temperatures and its potential detrimental effects to wine quality has not gone unnoticed by local producers. In 2019, the Bordeaux Wine Council issued a press release delineating strategies to adapt to the further increase in temperatures in the coming years with practices that include delaying pruning dates and choosing later-ripening varieties (BordeauxWineCouncil, 2019). For the moment, Bordeaux takes pride for being 'the first French vineyard to introduce new grape varieties of interest for adapting to climate change into its AOC specifications'. But given how influential the Bordeaux region is in terms of vinification methods, we expect other regions will soon follow suit. Overall, these adaptations to the wine-making bioeconomy not only promise to make the wine-maker's long-term production goals easier to achieve, but based on our analyses, they are also likely to increase consumers' enjoyment of the wines.

References

Analytis, P. P., Barkoczi, D., and Herzog, S. M. (2018). Social learning strategies for matters of taste. *Nature human behaviour*, 2(6):415–424.

Analytis, P. P., Barkoczi, D., Lorenz-Spreen, P., and Herzog, S. (2020). The structure of social influence in recommender networks. In *Proceedings of The Web Conference 2020*, pages 2655–2661.

Ashenfelter, O. (1987). Vintage advice: Is 1986 another outstanding bordeaux vintage?(no! and you heard it here first). *Liquid Assets*, 2:1–7.

Ashenfelter, O. (1988). A new objective ranking of the chateaux of bordeaux (based on auction prices, of course!). *Liquid Assets*, 5:1–6.

Ashenfelter, O. (1990). Just how good are wine writers' predictions?(surprise! the recent vintages are rated highest!). *Liquid Assets*, 7:1–9.

Ashenfelter, O. (2008). Predicting the quality and prices of bordeaux wine. *The Economic Journal*, 118(529):F174–F184.

Ashton, R. H. (2012). Reliability and consensus of experienced wine judges: Expertise within and between? *Journal of Wine Economics*, 7(1):70–87.

Ashton, R. H. (2013). Is there consensus among wine quality ratings of prominent critics? an empirical analysis of red bordeaux, 2004–2010. *Journal of Wine Economics*, 8(2):225–234.

Asimov, E. (2021). A wine worth waiting for. *The New York Times*.

Bagwell, K. and Riordan, M. H. (1991). High and declining prices signal product quality. *The American Economic Review*, pages 224–239.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

BordeauxWineCouncil (2019). Bordeaux in the face of climate change. *Press Release*.

Brien, C., May, P., and Mayo, O. (1987). Analysis of judge performance in wine-quality evaluations. *Journal of Food Science*, 52(5):1273–1279.

Budescu, D. V. and Chen, E. (2015). Identifying expertise to extract the wisdom of crowds. *Management Science*, 61(2):267–280.

Chevalier, J. A. and Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, 43(3):345–354.

Cicchetti, D. V. (2006). The paris 1976 wine tastings revisited once more: Comparing ratings of consistent and inconsistent tasters. *Journal of Wine Economics*, 1(2):125–140.

Deloire, A., Carbonneau, A., Wang, Z., and Ojeda, H. (2004). Vine and water: a short review. *OENO One*, 38(1):1–13.

Drichoutis, A. C., Klonaris, S., and Papoutsi, G. S. (2017). Do good things come in small packages? bottle size effects on willingness to pay for pomegranate wine and grape wine. *Journal of Wine Economics*, 12(1):84–104.

GlobalWineScore, T. (2017). Do the en primeur scores predict well the future quality of wines? *Medium*.

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.

Hirschman, E. C. and Pieros, A. (1985). Relationships among indicators of success in broadway plays and motion pictures. *Journal of Cultural Economics*, pages 35–63.

Holbrook, M. B. (1999). Popular appeal versus expert judgments of motion pictures. *Journal of consumer research*, 26(2):144–155.

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 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.

Kurvers, R. H., Herzog, S. M., Hertwig, R., Krause, J., Moussaid, M., Argenziano, G., Zalaudek, I., Carney, P. A., and Wolf, M. (2019). How to detect high-performing individuals and groups: Decision similarity predicts accuracy. *Science advances*, 5(11):eaaw9011.

Lawless, H., Lui, Y.-F., and Goldwyn, C. (1997). Evaluation of wine quality using a small-panel hedonic scaling method. *Journal of Sensory Studies*, 12(4):317–332.

Le Mens, G., Kovács, B., Avrahami, J., and Kareev, Y. (2018). How endogenous crowd formation undermines the wisdom of the crowd in online ratings. *Psychological science*, 29(9):1475–1490.

Luca, M. and Zervas, G. (2016). Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science*, 62(12):3412–3427.

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

Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., Scott, S. E., Moore, D., Atanasov, P., Swift, S. A., et al. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological science*, 25(5):1106–1115.

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of business venturing*, 29(1):1–16.

Muchnik, L., Aral, S., and Taylor, S. J. (2013). Social influence bias: A randomized experiment. *Science*, 341(6146):647–651.

Müller-Trede, J., Choshen-Hillel, S., Barneron, M., and Yaniv, I. (2018). The wisdom of crowds in matters of taste. *Management Science*, 64(4):1779–1803.

Mullins, M. G., Bouquet, A., and Williams, L. E. (1992). *Biology of the grapevine*. Cambridge University Press.

Oczkowski, E. and Doucouliagos, H. (2015). Wine prices and quality ratings: A meta-regression analysis. *American Journal of Agricultural Economics*, 97(1):103–121.

Passell, P. (1990). Wine equation puts some noses out of joint. *The New York Times*, 4:1.

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.

Rao, A. R. and Monroe, K. B. (1989). The effect of price, brand name, and store name on buyers' perceptions of product quality: An integrative review. *Journal of marketing Research*, 26(3):351–357.

Storchmann, K. (2012). Wine economics. *Journal of Wine Economics*, 7(1):1–33.

Stuen, E. T., Miller, J. R., and Stone, R. W. (2015). An analysis of wine critic consensus: A study of washington and california wines. *Journal of Wine Economics*, 10(1):47–61.

Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.

Swets, J. A., Dawes, R. M., and Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological science in the public interest*, 1(1):1–26.

Van Leeuwen, C., Trégoat, O., Choné, X., Bois, B., Pernet, D., and 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.

Willsher, K. (2013). French winemakers count cost of devastating hailstorms. *The Guardian*.