Explaining Wine Scores Through Stochastic Frontier Analysis

Gustavo Ferro[1] and Nicolás Gatti[2]

Abstract

Experts give scores to wines, which are quality proxies for marketers and buyers. The production of wine quality is explained by a set of observable objective attributes, plus another set of unobservable and subjective (sensory) features, and randomness. We use a Stochastic Frontier Approach (SFA) to understand whether objective and subjective (sensory) characteristics of wines explain the differences in wine scores. We estimate a wine quality stochastic frontier production function, using a database of 1800 top-scored wines, in an 18 years-window encompassing objective determinants (price, production, year, grape, country, etcetera), being sensory aspects related to wine grading unobservable. We find that the variables included explain half of the "efficiency" in attaining scores and our results suggest that sensory variables may have a role in explaining inefficiency.

Keywords

wine, quality, stochastic frontier analysis, SFA

JEL Codes: D61, L66

1. Introduction

Quality is a concept difficult to quantify due to its very nature. The economic literature on product differentiation deals with the task of decomposing qualitative features into quantified attributes (Ashenfelter and Quandt, 1998; Cicchetti, 2004; Lindley, 2006). Wine is not an exception, and some differences in quality can be treated as quantifiable attributes (production, price, vintage, organoleptic features, years to start to drink, years to store, etcetera) as well as other identifiable variables which characterize a wine (such as region or country origin, grape, varietal wine, or a blend of grapes, etcetera). However, other differences are related to sensory characteristics, which quantification is elusive. As Ashton (2017) points out, wine evaluation rests on both sensory and cognitive mechanisms. The former elements are physiologically based, pertaining not only to taste and smell but also to the visual and tactile senses, while the latter group is experience based.

Specialized publications give quantitative scores to wines based on expert tastings, which function as proxies for quality both to marketers ("this is a 100-point wine, drink it!") and to buyers ("this is a 100-point wine, I will buy it because it seems to be a high-quality one").

Our objectives are twofold. First, we apply a well-known technique to a somewhat non-conventional goal. We use a stochastic frontier analysis to determine the importance of objective and subjective wine characteristics in a "quality" production function, which is rare in the literature. Second, we build and utilize a comprehensive database of 1800 top wines in an 18 years-window from one of the most famous specialized publications' rankings, Wine Spectator. Third, our results contribute to determining efficiency in quality production -understood as scores- and point to sensory unobservable characteristics as potential drivers of inefficiencies.

We apply frontier econometric techniques to the "production" of scores, being inputs of all the quantitative (or quantifiable) information we have of the top 100 best wines in 18 years of Wine Spectator publications. Some wines do better in the rankings than others given the

[1] Universidad del CEMA (UCEMA) and CONICET. Corresponding author. Email: gaf97@ucema.edu.ar, gferro05@yahoo.com.ar

[2] Universidad del CEMA (UCEMA) and Centro de Investigación en Economía y Prospectiva CIEP-INTA. Email: nicolasgatti1987@gmail.com

inputs they use to produce scores. The residuals of the frontier models are normally interpreted as "inefficiency" (since randomness is conveniently isolated), we interpret them as having poor non-quantifiable attributes (which are in the nose and in the mouth of the expert that tastes the wine) that impede wines to reach higher scores, given their inputs. This way, we have a form to quantify the elusive sensory attributes of wines. Our contribution is one novel way to proxy otherwise unquantifiable quality attributes, by employing a well-known technique.

A stochastic frontier approach has advantages over traditional OLS estimations. As opposed to Ordinary Least Squares (OLS), frontiers are a representation of the best performing units, in our case, the quality of wines. By estimating a production function frontier, we can obtain wine quality efficiency scores to evaluate the performance of the different wines. Suppose you have one unit of output produced with 10 units of input, and ceteris paribus, another unit of the same output is produced by another firm using 8 units of inputs. The second one is in principle deemed as more efficient than the first one, i.e., produces the same with less input usage. However, not all the residuals of the units not in the frontier in relation to the best practices are deemed as inefficiency. The reason for lower than optimal outcomes could be explained by shocks out of the control of wine producers such as bad or good luck, randomness, and changes in consumer preferences. We use the SFA to produce indicators of which one of the two sources of variations -observable or unobservable- is more important in the residuals of the estimates.

The research article is organized as follows. Section 2 discusses the literature on wine scores. Section 3 presents the method, the database, and the models. Section 4 displays and discuss the estimates, and section 5 concludes.

2. Literature review

The use of scores as a measure of wine quality has been widely studied in the wine economics literature: a substantial oenological literature exists on the opinions of experts and neophytes translated to interpretations about wine quality (Ashenfelter and Quandt, 1998; Cicchetti, 2004; Lindley, 2006). These opinions can be contrasted with factual binary questions about wine, as Cicchetti (2017) addresses, to apply a biostatistical methodology to compare wine judgments with wine facts. Using hypothetical examples, wine judges' classifications of wines as oaked or unoaked were analyzed for their degree of accuracy. The results show that "Overall Accuracy" is a poor measure of the accuracy of binary judgments relative to Sensitivity, Specificity, Predicted Positive Accuracy, and Predicted Negative Accuracy.

Differences in scores could be due to intrinsic and extrinsic wine characteristics. Veale and Quester (2008) investigate the respective influences of price and country of origin as extrinsic cues on consumer evaluations of wine quality when all intrinsic cues are experienced through sensory perception. Price and country of origin were both found to be more important contributors to quality perception than taste. Reliance on extrinsic cues was found to remain extremely robust. Consumer belief in the price/value schema dominates quality assessment. Consequently, marketers cannot assume that intrinsic attributes will be weighted and interpreted accurately by consumers. On the other hand, Ashton (2014 b) provides evidence that an important segment of wine consumers does not consider price a useful cue to quality. Specifically, in blind tastings, average wine drinkers consider less expensive wines to taste better than more expensive wines. In three out of four tastings of high-quality French wines in a club, there was no relationship between price and enjoyment, while in the other the relationship was negative. In the same vein, Lelocq and Visser (2006) are concerned with the lack of correlation between price and pleasure. They apply the hedonic technique to wines, encompassing objective, as well as sensory characteristics and a grade assigned by expert tasters. The results are used to make comparisons between two of the most important wine regions in France, and comparisons over time. Fried and Tauer (2019) specify a two-tier stochastic frontier model to estimate a price equation for U.S. Rieslings and determine the overpricing or underpricing of specific wines for 2000–2016. Using tasting score and tasting score squared as sole independent variables, they find evidence of overpricing and underpricing of wines. When additional location variables were entered, both underpricing and overpricing do not occur simultaneously in this market.

Wine score reliability is an issue to judge wine quality. Nevertheless, reliability, defined as the correspondence of repeated ratings by one critic, and consensus, defined as the correspondence of ratings between critics, have been moderate or low in some empirical studies. Luxen (2018) observes that wine consumers and producers make decisions partly based on the ratings of wine critics. Ashton (2013) found a correlation of around 0.60 for important critics of red Bordeaux analyzing vintages 2004–2010. Luxen (2018) extends the analyses to the years 2011–2016 for the same wines, yielding a correlation close to Ashton's (2013) values. In addition, he observes that critics agree more about what they do not like. In his sample, wines score below-average ratings when they cost less than the threshold of 35 euros, and higher ratings when they cost between 35 and 100 euros, while in more expensive wines there is no correlation between ratings and price. Stuen et al. (2015) evaluate whether the 100-point scale score is a reliable quality measure. Using data on 853 wines, they find a moderately high level of consensus, measured by the correlation coefficient, between most pairs of publications, as in Ashton (2013). Rank and intraclass correlations are similar thus opinions seem consistent. Cao and Stokes (2017) find that the ranking based on the *scoring* average is generally more accurate than that based on the *rank* average. The ranking puts less burden on judges in wine tasting and may outperform the other methods in certain conditions. Marks (2020) suggests that consumers use expert ratings to help choose wine, that is, the former proxy quality from scores and economists find correlations between ratings and transaction prices. People taste differently, thus ratings are often unreliable hedonic markers of enjoyment. Thus, hedonic measurements attempt to adjust for differences in perceived sensory sensitivity and offer clues.

Some studies have provided direct quantification of how much consensus, as opposed to randomness, exists in wine ratings. Cao (2014) suggests that a significant lack of consensus exists in wine quality ratings. A permutation-based mixed model tries to quantify randomness versus consensus in wine ratings. The study shows that the model can provide excellent model fit, which indicates that wine ratings, indeed, consist of a mixture of randomness and consensus. A direct measure is easily computed to quantify randomness versus consensus in wine ratings. The method is demonstrated with data analysis from a major wine competition and a simulation study. Bodington (2020) argues that much literature shows that the ratings assigned by wine judges are uncertain, some authors have proposed that judges be tested, and a few wine competitions do test judges. Thus, he uses multiple correlation coefficients to rate 54 judges who assigned ratings to 2,811 wines in a commercial competition. Results show that there is a strong and positive correlation between the ratings assigned by most judges to most wines. However, the ratings assigned by approximately 10% of judges are indistinguishable from random assignments.

Another branch of the literature looks at wine tasting experience as a driver for differences in scores. Ashton (2017) explores the distinction between novices and experts in wine evaluation. The research examines expert/novice differences at both the chemical component level (detecting, discriminating among, and describing wine-relevant chemical components) and the holistic level (hedonic evaluation of wine as an integrated manifestation of its components). In conclusion, experts tend to outperform non-experts in wine evaluation. Bodington (2017) points to the stochastic nature of ratings and proposes and tests a conditional-probability model that yields maximum-likelihood estimates of judges' latent consensus, idiosyncratic, and random assignments of scores to wines. Using data for a tasting of eight Sauvignon Blanc wines that contain a blind triplicate, the conditional-probability model detects the similarity between the triplicates and the model results also show that the scores that a judge assigns to replicates may not be a robust guide to the accuracy of the scores that the judge assigns to other wines. Ashton (2013) examines the level of agreement among the wine quality ratings of six prominent wine critics for seven consecutive vintages of red Bordeaux. The level of consensus among these critics is contrasted with that among non-prominent wine professionals and several other field professionals. Agreement (or disagreement) seems close in different professional disciplines besides wine tasters. Ashton (2014) investigates the influence of expectations on the sensory perception of wines. Prior to tasting, only the color of the wine was known, and neither wine club members nor experienced wine professionals can distinguish between New Jersey and California wines in terms of personal enjoyment. Another study in which tasters were informed that some of the wines were from New Jersey, received lower enjoyment ratings than when the identical wine is believed to be from California. This connects with the next issue discussed in the literature. Scores could be related to judges' biases in their tasting experience. Ashton (2012) considers the levels of reliability (or consistency) and consensus of wine quality judgments found in studies of experienced wine judges. Reliability and consensus levels found in wine judging are compared to those in six other fields. In all of them, reliability is greater than consensus. Both reliability and consensus are, on average, substantially lower in wine judging than in other fields. Cardebat et al. (2014) assess the role of expert opinion on the pricing of Bordeaux wines from France, Spain, and the United States by nine experts from 2000 to 2010. Expert wine scores are explained by objective fundamentals of wine production (the quality of soil, producers' skills, and climate conditions) and subjective individual opinions of the experts. Comparing the impacts of both factors, the study finds that prices are influenced more deeply by the fundamental quality of the wine than by the judge's subjectivity. Oczkowski (2016) examines a framework developed by Cardebat et al. (2014) for identifying the impacts of both objective and subjective quality on wine prices, which is applied to Australian premium wines. Results indicate that the price impact of expert personal opinions is comparable to the impact of objective quality as estimated via weather, vintage, and producer fixed effects. In addition, the selection of judges may have a role in determining the scores. Ashton (2011) demonstrates that the wine quality judgments of the experts who participated in the 1976 Judge of Paris tasting would have been improved simply by averaging the quality ratings of two or more of the judges. The results have implications for choosing judges to include in tasting panels that award prizes or provide expert advice to consumers, as well as for a better understanding of the variability in the price-quality association.

Stochastic frontier analysis and impact evaluation techniques have been widely used to solve selection bias. Bravo-Ureta's (2014) work links stochastic frontiers with impact evaluation methods. A major hurdle is resolving biases that might arise from the selection of observables and unobservable characteristics. Bravo-Ureta (2014), provides an overview of how impact evaluation and stochastic frontiers, two well-established areas in applied econometrics, are being brought together to shed light on the productivity effects of agricultural development interventions.

3. Method, database, and models

In this section, we briefly explain the method we follow for the empirical assessment thus we present and describe the database, and later we show the model variants we estimate.

3.1 Method

There are two general methodologies to construct efficiency frontiers: the econometric approach and the mathematical programming method. The econometric approach is stochastic (distinguishing randomness from inefficiency or management decisions within the residuals of the estimates), and parametric (assumes a specific functional form for the relations it studies). On the other hand, the mathematical programming method is generally deterministic (not distinguishing between pure randomness and efficiency, that is, assuming that all the residuals of the estimates can be deemed as inefficiency), and non-parametric (not assuming a functional form between the variables).
Data Envelopment Analysis (DEA) is the most popular non-parametric method. It is a very flexible method; it builds the frontier with only a subset of the sample (the observations lying on the frontier), it is very sensitive to outliers, allowing it detection in large datasets, and it does not allow statistical tests of the results. Instead, the SFA method, which can be statistically tested is sensitive to the selection of the adequate functional form, and to the criteria for separating stochastic noise from deterministic components, which requires the selection of a statistical distribution for the error term (Coelli et al., 1998).
We use the standard model for estimating efficiency using a Stochastic Frontier Analysis (SFA), in cross-sectional databases, as described in Kumbhakar et al. (2015). Its general formula is:

$$Y_i = Y(x_i; \beta) + v_i - u_i \qquad (1)$$

Where Y_i is the observed output for each Decision-Making Unit (DMU)i; x_i is the input vector; selective "environmental" variables can be included in the preceding basic models (also known as hedonic or control variables), z_i is the environmental variable vector; β is the unknown parameter vector to estimate; v_i is a random error (independently and identically distributed), u_i is an inefficiency parameter (which distribution is assumed exponential). Besides, ui and v_i are independently distributed from each other and from the model's covariates. In logarithms,

$$lny_i = lny_{i*} - u_i \qquad (2)$$

And the efficient unit i output is:

$$lny_{i*} = f(x_i; \beta) + v_i$$

The term ui is the log difference between the maximum lny_{i*} and the actual output lny_i or the percentage by which actual output can be increased using the same inputs if production is fully efficient. It gives the percentage of output that is lost due to technical inefficiency. The estimated value of ui is referred to as the output-oriented (technical) inefficiency, with a value close to 0 implying closeness to fully efficiency.

$$\exp(-u_i) = \frac{lny_i}{lny_{i*}} \quad (3)$$

Because $ui \geq 0$ the ratio (3) is bounded between 0 and 1 (being 1 fully efficient), with a value equal to 1 implying that the firm is fully efficient technically.
The real functional form is unknown, being the more common choices the Cobb-Douglas, which is simple and easy to interpret, and the Trans logarithmic (or “Trans log”), which addresses squared and interaction terms of the variables. The Trans logarithmic function has the advantage of being more flexible than Cobb-Douglas. It does not impose a priori constraints on input substitution feasibility and allows scale economies to varying together with the output level.
Time can affect productivity due to technical change. A temporal variable can also be included to capture the technological progress or frontier shift occurring in time. We assume that technological progress directly affects the production function; that is, all DMU are subject to the same technological shocks over time. These shocks include a linear indication of time in the Cobb-Douglas (T) and a quadratic polynomial of time in the Trans logarithmic equation because this functional form is a second order approximation (including the T term as well as the T^2 term). The rate of technological change is given by: $T^* = \partial y/\partial t$. If $T^* > 0$, technical change is positive, indicating productivity growth, and conversely, if $T^* < 0$, signaling a decrease in productivity.

3.2 Data

The database comprehends information from top-100 wines from Wine Spectator’s yearly rankings. Our dataset goes from 2003 to 2020 and contains wine scores and their objective characteristics. Wines are rated by specialists, and they provide scores that represent a proxy for quality. Because there are always new wines and changes in wine quality, wines drop and join the ranking every year. Therefore, we structure the dataset as a pooled cross-section of wines. Each wine score is determined by name, country of origin, position in the ranking, vintage year, production volume, price, recommended years to storage, and release year. Moreover, we distinguish varietal wines from blend ones, as well as wines by type of grape. All the definitions of the variables are shown in Table 1. To relate the categories of wines to specific grapes and origins, see the examples in Table 2.

Table 1. Database Variables Description

Name	Meaning	Unit
YEAR RANKING	Year in which the wine was ranked	Year (2003-2020)
POSITION	Position in the ranking	Unit (1 to 100)
VINTAGE	Vintage year. It can be the same as YEAR RANKING and normally it is some previous year.	Year
NAME	Wine denomination	Qualitative

SCORE	Points awarded by Wine Spectator's Top 100 Annual Rankings	Points (90 to 100)
PRICE	Price in constant US dollars of 2010	Current prices in US$ in the American Market, deflated with US CPI
PRODUCTION	The quantity of wine produced in, or imported to, the USA	Cases
RELEASE	Suggested years from vintage until starting drinking (Minimum years of storage)	Years
STORAGE	Suggested years of storage since vintage (Optimum years of aging)	Years
EURO	If European = 1, else = 0.	Dummy
VARIETAL	If Varietal = 1, else (blends) = 0	Dummy
LIGHT WHITE*	If Light White = 1, else = 0.	Dummy
RICH WHITE*	If Rich White = 1, else = 0.	Dummy
ELEGANT RED*	If Elegant Red = 1, else = 0.	Dummy
BIG RED*	If Big Red = 1, else = 0.	Dummy
ROSÉ*	If Rosé = 1, else = 0.	Dummy
SPARKLING*	If Sparkling = 1, else = 0.	Dummy
COUNTRY	Country of origin	18 Country dummies

*See Wine Spectator's Classification for details and paradigmatic grapes, synthesized in Table 2. We also generate a Generic Classification and a Varietal Classification (see Table 3) to group the former categories.
Source: Authors' Elaboration on Wine Spectator's Top 100 Annual Rankings.

Table 2. Wine Spectator's Classification of Wines

Type	Features	Paradigmatic grapes	Typical origins
Light Whites	Fresh, light	Sauvignon Blanc, Pinot Grigio, non-oaked Chardonnay	Almost all producer countries; fresh or coastal areas
Rich Whites	Fresh, strong body	Burgundy, oaked Chardonnay, Chenin Blanc, Godello, Friulano	France, South Africa; USA, Spain, Italy
Elegant Reds	Fresh, light, fruited, low tannins	Pinot Noir, Barbera, Dolcetto, Sangiovese	USA, New Zealand, Italy
Big Reds	Powerful, with a strong body, alcoholic, tannic, complex, and with enough acidity.	Malbec, Shiraz, Cabernet Sauvignon, Tempranillo, Douro	Warm and sunny places. Argentina, Australia, USA, France, Portugal
Rosé	Light, fruity	Typically blends	France, Spain, Italy, USA
Sparkling	Sparkling, with or without added sugar	Typically blends	France, Italy, USA, and New Zealand

Source: Wine Spectator.

Our dataset, whose descriptive statistics are presented in Table 3, contains 1,800 observations from 18 years of wine rankings. The average score (points) of each wine is 92.9 and they range from 90 to 100 every year. Wine prices in the US markets vary from 8 to 535 constant 2010 dollars per bottle; wine production is 11,078 cases on average and ranges from 0 to 589,000 cases. The average time since the wine was first released is 3.6 years, and the mean time in storage is approximately 10 years. We also generate several dummy variables to characterize grapes, country of origin, if the wine is European or not, if it is a blend or a varietal, and combinations of some grouping of the precedent variables (see below, in the model subsection).

Table 3. Wine Main descriptive statistics

Objective characteristics	Mean	Sd	Min	Max
Score	92.92	2.23	90	100
Wine price in US$ (adjusted by the US CPI)	45.01	37.53	8	535
Production	11,079	27,513	0	589,000
Years since first released	3.59	2.45	0	29
Storage (years)	9.86	7.18	0	49
Grapes	Mean	Sd	Min	Max
Wine Spectator's Classification				
Big Red	0.52	0.50	0	1
Elegant Red	0.21	0.41	0	1
Light Whites	0.19	0.39	0	1
Rich Whites	0.05	0.22	0	1
Sparkling	0.03	0.16	0	1
Rose	0.01	0.08	0	1
Generic Classification				
Red (Big + Elegant)	0.73	0.44	0	1
White (Light + Rich)	0.24	0.43	0	1
Different from Red or White	0.03	0.17	0	1
Varietal Classification				
Varietals	0.67	0.47	0	1
Varietal & Red	0.46	0.50	0	1
Varietal & White	0.21	0.41	0	1
Varietal & Different from Red or White	0.001	0.03	0	1
Country of Origin	Mean	Sd	Min	Max
USA	0.300	0.46	0	1
France	0.206	0.40	0	1
Italy	0.166	0.37	0	1
Spain	0.078	0.27	0	1
Australia	0.063	0.24	0	1
Portugal	0.033	0.18	0	1
Argentina	0.030	0.17	0	1
Chile	0.027	0.16	0	1
New Zealand	0.027	0.16	0	1
Germany	0.021	0.14	0	1
South Africa	0.020	0.14	0	1
Austria	0.014	0.12	0	1
Greece	0.006	0.07	0	1
Hungary	0.004	0.06	0	1
Israel	0.004	0.06	0	1
Canada	0.001	0.03	0	1
North Macedonia	0.001	0.02	0	1
Uruguay	0.001	0.03	0	1
Observations	1800			

Source: Authors' elaboration from Wine Spectator.

Most wines in the rankings are reds (73 percent), followed by whites (24 percent), and lastly, a small portion of wines different from white or red is 3 percent of the dataset. Varietal wines are 67 percent of the sample, while 46 percent are varietal reds and 21 percent are varietal whites. Only 0.1 percent are other varietal wines, different from white or red. Regarding grapes, and following the Wine Spectator's classification, big reds are more than 50 percent of wines in the sample while 21 percent are elegant reds; light whites represent 19 percent and rich whites are 5 percent.

The main countries represented in the wine rankings are USA and France, the most important in the American market, where the rankings are assessed. Together French and American wines account for 50 percent of the dataset. From the remaining 50 percent of the sample, 32 percentage points are wines from European countries other than France, that have been increasing their presence in the rankings. The rest of the wines are coming from countries in the southern hemisphere where growing conditions allow for good quality wines; 9 percent of them are from Oceania, approximately 6 percent from South America, and the rest are from South Africa.

3.3 The models

We estimate five log-log SFA variants, presented in Table 4. A frontier estimate normally has the "core" or textbook determinants (for example, in a production function of a good, capital and labor explain quantity) plus "environmental" features (to address specifics, such as, in our case country of origin, grapes, etcetera).
SFA model 1 contains only the "core" variables in the production of scores, while SFA models 2 to 5 vary in the inclusion of different classifications of grapes, and in the last one, the country of origin is also included.

Table 4: Variants of SFA estimated models

Dependent variable: ln(Score)	SFA Model 1	SFA Model 2	SFA Model 3	SFA Model 4	SFA Model 5
Ln(Ranking position)	X	X	X	X	X
ln(Wine Price in U$S)	X	X	X	X	X
Ln(Production)	X	X	X	X	X
Ln(Years since first released)	X	X	X	X	X
Ln(Storage years)	X	X	X	X	X
Varietal (=0 Blend)	X	X	X	X	X
Time Trend	X	X	X	X	X
Constant	X	X	X	X	X
White		X			
Red		X			
Rosé		X	X	X	X
Sparkling		(Omitted)	(Omitted)	(Omitted)	(Omitted)
Varietal and Red			X		
Varietal and White			X		
Light White				X	X
Rich White				X	X
Big Red				X	X
Elegant Red				X	X
17 Country dummies (France is Omitted)					X

Source: Authors' elaboration.

4. Stochastic Frontier Estimations

Table 5 presents several quality production function estimations, being Scores as the dependent variable. In equation 1, we present the pooled OLS regression of the Scores production function as a benchmark for the Stochastic frontiers, SFA Models 1 to 5. From left to right, we present several specifications that add dummy variables for wine characteristics showing that coefficients from the main inputs explaining scores are stable. In the last column, we add country-fixed effects to control for economic, and contextual differences between countries.
All the stochastic frontier estimations show that the log of the variance of the inefficiency component is statistically significant. To properly address how much of the error term is related to inefficiency we present the inefficiency to statistical noise ratio called lambda, which gives us the proportion of the standard deviation of inefficiency to the randomness standard deviation. Across specifications, we find that there is around 0.50, meaning that one-

third of the standard deviation of the scores is explained by inefficiency, which is our proxy for sensory attributes.

To analyze the coefficients, the production function expanded by wine characteristics and country fixed effects (equation 6) is our preferred specification since it has the lowest Akaike Information Criteria (AIC) value. We find that there is a negative association between the position in the ranking and wine scores. The reason for this is that both scales are not monotonic: hypothetically, one wine can be the first on the list and its score is 95 or less. The relationship between scores and price and quantities (positive and negative, respectively) is reasonable. The higher the wine price, the more points a wine would receive. Likewise, an increase in production is associated with lower scores. Years in storage and years since first released are positively associated with wine scores.

Wine characteristics' coefficients show that scores are heterogeneous. Red, Rosé, and Sparkling wines have statistically significant negative coefficients. These kinds of wines show lower scores than wines that do not have these characteristics. Varietal wines seem to have a discount in points compared to blends, this discount offsets when the varietal is red, and it is exacerbated for white varietals. However, these coefficients are not ever statistically significant.

The trend coefficient is positive and statistically significant showing that scores have been increasing over time. The base comparison for scores and country fixed effects is France, which is a historical reference in terms of high wine quality globally. In that respect, we see that wines from the USA, Portugal, New Zealand, and Germany have statistically significant positive coefficients showing that they are receiving differentially higher scores than French wines. In contrast, we find that wines from Israel, North Macedonia, and Uruguay have differentially lower scores than France. However, it is worth noticing that wines from these countries have a small presence in the rankings. We use the distribution of TE to rank wines and learn about the characteristics of the most efficient wines. In Table 6, we present a t-test for mean differences across quantiles of the technical efficiency distribution. The first 4 columns have the mean value for each quantile ordered from smallest to largest TE scores. Secondly, we provide the t-tests by quantiles. We are especially interested in learning what makes an efficient wine based on our definition of a score production function. If we compare statistical differences between the selected top TE wines compared to the 3^{rd} quantile, we find that there are only a few features that differentiate them. Top wines in terms of efficiency are on average 6.73 dollars more expensive, have been released 0.4 years before, are being stored 0.6 years more, and have 728 cases less of production than the following quantile in the TE distribution. In terms of wine types, the salient characteristics are that there are 5.8 percent more white varietal wines and 5.1 percent less rich and light wines than in the following TE quantile. The remaining characteristics are not statistically significant.

Table 6 shows the Test for differences in characteristics by TE quantiles. Columns 1 to 4 show the average and standard deviation for each quartile of the TE distribution, going from the least to the most efficient wines. The following columns have the t-tests for the differences by quantile and, lastly, we present a joint orthogonality test for all the distribution. Wines have significant differences in terms of main inputs and objective and subjective wine characteristics when looking at the joint orthogonality test for all the variables by quartiles. When looking at individual differences, the test over the 3rd and 4th quantile is showing the differences between the two most efficient groups of wines. We have a negative difference in prices and vintage years which are both productive characteristics. Likewise, we observe that varietal whites are in the most efficient group. In contrast, light white wines are more present in the third quartile than in the most efficient group of wines. In terms of wine origins, the presence of New Zealand in the third quartile is statistically significant.

5. Conclusions

Experts have a direct role in proposing measures intended to proxy wine quality. Marketers and buyers use the scores that experts determine to feed changes in the perception of wine quality. The quality of wines or their proxy, the “scores”, are related to objective inputs used in wine production, plus a set of unobservable and subjective (sensory) features, plus randomness. Our contribution is one novel way to quantify otherwise unquantifiable quality attributes, by employing a well-known technique. Stochastic Frontier Models are useful to explain the distance between scores of different wines given the observable and unobservable inputs, and at the same time leaving aside randomness.

We estimate a wine quality production function using the best wines in the world based on the Wine Spectator ranking. We apply SFA techniques to a database of 1800 top-quality wines, in an 18 years-window encompassing objective determinants (price, production, year, grape, country, etcetera) and attributes not explained differences in sensory aspects to wine grading.

Our results show that the variables included in the OLS analysis explain 83% of wine scores. Second, we obtain that 50% of the residuals are explained by inefficiencies and not pure exogenous shocks. However, because objective measures included in the analysis already explain a good portion of the variability of the scores, the efficiency improvement may come from sensory attributes.

Table 5. Production function and production frontiers estimations

	Ln(Scores)					
	OLS regression	SFA Model 1	SFA Model 2	SFA Model 3	SFA Model 4	SFA Model 5
Ln(Ranking position)	-0.00916***	-0.00975***	-0.00973***	-0.00974***	-0.00974***	-0.00950***
	(0.000376)	(0.000464)	(0.000459)	(0.000462)	(0.000466)	(0.000512)
Ln(Wine price in U$S)	0.0227***	0.0231***	0.0232***	0.0231***	0.0232***	0.0231***
	(0.000472)	(0.000481)	(0.000488)	(0.000482)	(0.000490)	(0.000514)
Ln(Production)	-0.00207***	-0.00218***	-0.00213***	-0.00217***	-0.00210***	-0.00209***
	(0.000234)	(0.000214)	(0.000215)	(0.000215)	(0.000217)	(0.000223)
Ln(Years since first released)	0.00215***	0.00211***	0.00242***	0.00224***	0.00242***	0.00278***
	(0.000495)	(0.000495)	(0.000520)	(0.000513)	(0.000518)	(0.000551)
Ln(Storage years)	0.000529*	0.000486	0.000553*	0.000491	0.000606*	0.000553
	(0.000319)	(0.000317)	(0.000326)	(0.000323)	(0.000326)	(0.000339)
Varietal (=0 Blend)	-0.000412	-0.000462	-0.000913*	-0.00173	-0.00105**	-0.00106*
	(0.000510)	(0.000497)	(0.000526)	(0.00308)	(0.000535)	(0.000613)
White			0.00272**			
			(0.00127)			
Red			0.000849			
			(0.00118)			
Rosé			-0.00378**	-0.00495***	-0.00393**	-0.00367**
			(0.00167)	(0.00142)	(0.00176)	(0.00172)
Varietal & red				0.000905		
				(0.00307)		
Varietal & white				0.00192		
				(0.00310)		
Light White					0.00263*	0.00230
					(0.00136)	(0.00143)
Rich White					0.00229	0.00229
					(0.00154)	(0.00156)
Elegant Red					0.00115	0.000716
					(0.00135)	(0.00138)
Big Red					0.000372	0.000130
					(0.00125)	(0.00128)
USA						0.00126*
						(0.000737)
Italy						0.000621
						(0.000859)
Spain						0.00007
						(0.000960)
Australia						0.00171
						(0.00106)
Portugal						0.00424***
						(0.00144)
Argentina						0.000794
						(0.00111)

Chile						-0.000659
						(0.00128)
New Zealand						0.00455***
						(0.00146)
Germany						0.00299
						(0.00192)
South Africa						-0.00115
						(0.00131)
Austria						-0.000566
						(0.00182)
Greece						-0.00226
						(0.00149)
Hungary						0.00319
						(0.00410)
Israel						-0.00648*
						(0.00384)
Canada						-0.00749
						(0.0261)
North Macedonia						-0.00490***
						(0.000858)
Uruguay						-0.00719***
						(0.00122)
Trend	0.000282***	0.000304***	0.000311***	0.000312***	0.000306***	0.000316***
	(0.00005)	(0.00005)	(0.00005)	(0.00005)	(0.00005)	(0.00005)
Constant	4.495***	4.501***	4.499***	4.501***	4.499***	4.497***
	(0.00361)	(0.00361)	(0.00391)	(0.00369)	(0.00398)	(0.00425)
lnsig2v		-9.537***	-9.553***	-9.545***	-9.554***	-9.547***
		(0.0841)	(0.0842)	(0.0837)	(0.0858)	(0.105)
lnsig2u		-10.87***	-10.84***	-10.85***	-10.85***	-10.97***
		(0.264)	(0.257)	(0.260)	(0.262)	(0.384)
sigma_v		0.0084912	0.0084253	0.0084586	0.0084218	0.0084522
		(0.0003571)	(0.0003546)	(0.000354)	(0.0003611)	(0.0004448)
sigma_u		0.0043695	0.0044188	0.0043993	0.0044157	0.0041446
		(0.0005775)	(0.0005683)	(0.0005708)	(0.0005788)	(0.0007953)
sigma2		0.0000912	0.0000905	0.0000909	0.0000904	0.0000886
		(0.000004)	(0.000004)	(0.000004)	(0.000004)	(0.000004)
lambda		0.515	0.524	0.520	0.524	0.490
		(0.0009)	(0.000862)	(0.0008632)	(0.0008805)	(0.0011961)
R^2	0.838					
AIC	-11609.48	-11618.47	-11627.08	-11618.76	-11624.94	-11627.12
Observations	1800	1,800	1,800	1,800	1,800	1,800

Source: Authors' elaboration.

Table 6. T-test for differences in wines characteristics for quantiles of the TE distribution

	Mean / SE Quantile (1)	Mean / SE Quantile (2)	Mean / SE Quantile (3)	Mean / SE Quantile (4)	T-Test by Quantile (1)-(2)	T-Test by Quantile (1)-(3)	T-Test by Quantile (1)-(4)	T-Test by Quantile (2)-(3)	T-Test by Quantile (2)-(4)	T-Test by Quantile (3)-(4)	F Test
Wine price (adjusted by the US CPI)	49.132	40.198	41.983	48.709	8.935***	7.149***	0.424	-1.785	-8.511***	-6.73***	0.000***
	[1.873]	[1.635]	[1.731]	[1.795]							
Production	6928.95	10004.00	14055.09	13327.11	-3075.05***	-7126.14***	-6398.156***	-4051.09**	-3323.10*	727.98	0.000***
	[511.50]	[800.93]	[1765.51]	[1627.20]							
Years since first released	3.673	3.449	3.431	3.811	0.224	0.242	-0.138	0.018	-0.362**	-0.380**	0.056*
	[0.102]	[0.106]	[0.123]	[0.129]							
Storage (years)	10.029	8.638	10.060	10.718	1.391***	-0.031	-0.689	-1.422***	-2.080***	-0.658	0.000***
	[0.315]	[0.281]	[0.367]	[0.374]							
Varietal	0.673	0.656	0.667	0.678	0.018	0.007	-0.004	-0.011	-0.022	-0.011	0.904
	[0.022]	[0.022]	[0.022]	[0.022]							
Red	0.256	0.222	0.260	0.218	0.033	-0.004	0.038	-0.038	0.004	0.042	0.311
	[0.021]	[0.020]	[0.021]	[0.019]							
Rose	0.729	0.729	0.707	0.753	0.000	0.022	-0.024	0.022	-0.024	-0.047	0.478
	[0.021]	[0.021]	[0.021]	[0.020]							
Varietal & red	0.002	0.009	0.011	0.004	-0.007	-0.009	-0.002	-0.002	0.004	0.007	0.340
	[0.002]	[0.004]	[0.005]	[0.003]							
Varietal & white	0.431	0.451	0.442	0.500	-0.020	-0.011	-0.069**	0.009	-0.049	-0.058*	0.170
	[0.023]	[0.023]	[0.023]	[0.024]							
White & Light	0.240	0.204	0.227	0.176	0.036	0.013	0.064**	-0.022	0.029	0.051*	0.093*
	[0.020]	[0.019]	[0.020]	[0.018]							
White & body	0.198	0.184	0.200	0.176	0.013	-0.002	0.022	-0.016	0.009	0.024	0.761
	[0.019]	[0.018]	[0.019]	[0.018]							
Red & light	0.056	0.040	0.060	0.042	0.016	-0.004	0.013	-0.020	-0.002	0.018	0.428
	[0.011]	[0.009]	[0.011]	[0.009]							
Red & body	0.242	0.173	0.196	0.224	0.069**	0.047*	0.018	-0.022	-0.051*	-0.029	0.055*
	[0.020]	[0.018]	[0.019]	[0.020]							
France	0.487	0.551	0.509	0.527	-0.064*	-0.022	-0.040	0.042	0.024	-0.018	0.259
	[0.024]	[0.023]	[0.024]	[0.024]							
USA	0.236	0.178	0.209	0.200	0.058**	0.027	0.036	-0.031	-0.022	0.009	0.194
	[0.020]	[0.018]	[0.019]	[0.019]							
Italy	0.262	0.344	0.278	0.316	-0.082***	-0.016	-0.053*	0.067**	0.029	-0.038	0.031**
	[0.021]	[0.022]	[0.021]	[0.022]							
Spain	0.198	0.136	0.156	0.173	0.062**	0.042*	0.024	-0.020	-0.038	-0.018	0.077*
	[0.019]	[0.016]	[0.017]	[0.018]							
Australia	0.076	0.078	0.071	0.089	-0.002	0.004	-0.013	0.007	-0.011	-0.018	0.785
	[0.012]	[0.013]	[0.012]	[0.013]							
Portugal	0.064	0.056	0.071	0.060	0.009	-0.007	0.004	-0.016	-0.004	0.011	0.799
	[0.012]	[0.011]	[0.012]	[0.011]							
Argentina	0.031	0.042	0.031	0.027	-0.011	0.000	0.004	0.011	0.016	0.004	0.599

	[0.008]	[0.009]	[0.008]	[0.008]							
Chile	0.018	0.042	0.036	0.024	-0.024**	-0.018*	-0.007	0.007	0.018	0.011	0.134
	[0.006]	[0.009]	[0.009]	[0.007]							
New Zealand	0.018	0.024	0.044	0.022	-0.007	-0.027**	-0.004	-0.020	0.002	0.022*	0.068*
	[0.006]	[0.007]	[0.010]	[0.007]							
Germany	0.033	0.016	0.031	0.027	0.018*	0.002	0.007	-0.016	-0.011	0.004	0.355
	[0.008]	[0.006]	[0.008]	[0.008]							
South Africa	0.022	0.027	0.011	0.024	-0.004	0.011	-0.002	0.016*	0.002	-0.013	0.374
	[0.007]	[0.008]	[0.005]	[0.007]							
Austria	0.018	0.024	0.022	0.016	-0.007	-0.004	0.002	0.002	0.009	0.007	0.769
	[0.006]	[0.007]	[0.007]	[0.006]							
Greece	0.013	0.018	0.018	0.007	-0.004	-0.004	0.007	0.000	0.011	0.011	0.438
	[0.005]	[0.006]	[0.006]	[0.004]							
Hungary	0.000	0.011	0.009	0.002	-0.011**	-0.009**	-0.002	0.002	0.009	0.007	0.077*
	[0.000]	[0.005]	[0.004]	[0.002]							
Israel	0.004	0.000	0.007	0.004	0.004	-0.002	0.000	-0.007*	-0.004	0.002	0.437
	[0.003]	[0.000]	[0.004]	[0.003]							
Canada	0.004	0.002	0.002	0.007	0.002	0.002	-0.002	0.000	-0.004	-0.004	0.665
	[0.003]	[0.002]	[0.002]	[0.004]							
North Macedonia	0.002	0.000	0.000	0.002	0.002	0.002	0.000	N/A	-0.002	-0.002	0.573
	[0.002]	[0.000]	[0.000]	[0.002]							
Uruguay	0.000	0.000	0.002	0.000	N/A		N/A	-0.002	N/A	0.002	0.392
	[0.000]	[0.000]	[0.002]	[0.000]							
TE by quartile (upper bound)	99.5%	99.6%	99.7%	99.9%							
Obs	450	450	450	450							

Notes: The value displayed for t-tests are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels.

6. References

Ashenfelter, O., and Quandt, R. (1998). Analyzing a wine tasting statistically. Chance, 12(3), 16–20.

Ashton. R. H. (2012). Reliability and Consensus of Experienced Wine Judges: Expertise Within and Between? Journal of Wine Economics 7(1), 70–87. doi:10.1017/jwe.2012.6

Ashton, R. H. (2014 a). "Nothing Good Ever Came from New Jersey": Expectations and the Sensory Perception of Wines. Journal of Wine Economics 9(3), 304–319. doi:10.1017/jwe.2014.28

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. doi:10.1017/jwe.2013.18

Ashton, R. H. (2011). Improving Experts' Wine Quality Judgments: Two Heads Are Better than One. Journal of Wine Economics 6(2), 160–178.

Ashton, R. H. (2017). Dimensions of Expertise in Wine Evaluation. Journal of Wine Economics 12(1), 59–83. doi:10.1017/jwe.2016.27

Ashton, R. H. (2014 b). Wine as an Experience Good: Price Versus Enjoyment in Blind Tastings of Expensive and Inexpensive Wines. Journal of Wine Economics 9(2), 171–182. doi:10.1017/jwe.2014.7

Battese, G. and T. Coelli (1988). Prediction of firm levels of technical efficiencies with a generalized frontier production function and panel data. Journal of Econometrics, 38(3), 387-399.

Battese, G. and T. Coelli (1992). Frontier production functions, technical efficiency, and panel data: with application to paddy farmers in India. Journal of Productivity Analysis, 3(1-2), 153-169.

Bodington, J. C. (2017). Disentangling Wine Judges' Consensus, Idiosyncratic, and Random Expressions of Quality or Preference. Journal of Wine Economics 12(3), 267–281. doi:10.1017/jwe.2017.21

Bodington, J. C (2020). Rate the Raters: A Note on Wine Judge Consistency. Journal of Wine Economics 15(4), 363–369. doi:10.1017/jwe.2020.30

Bravo-Ureta, B. E. (2014). Stochastic Frontiers, Productivity Effects, and Development Projects. Economics and Business Letters 3(1), 51-58.

Cao, J. (2014). Quantifying Randomness Versus Consensus in Wine Quality Ratings. Journal of Wine Economics 9(2), 202–213. doi:10.1017/jwe.2014.8

Cao J. and Stokes, L. (2017). Comparison of Different Ranking Methods in Wine Tasting. Journal of Wine Economics 12(2), 203–210. doi:10.1017/jwe.2017.22

Cardebat, J.-M., Figuet, J.-M. and Paroissien, E. (2014). Expert Opinion and Bordeaux Wine Prices: An Attempt to Correct Biases in Subjective Judgments. Journal of Wine Economics 9(3), 282–303. doi:10.1017/jwe.2014.23

Cicchetti, D. (2017). Opinions versus Facts: A Bio-statistical Paradigm Shift in Oenological Research. Journal of Wine Economics 12(4), 354–362. doi:10.1017/jwe.2017.14

Fried H. O. and Tauer, L. W. (2019). Efficient Wine Pricing Using Stochastic Frontier Models. Journal of Wine Economics 14(2), 164–181. doi:10.1017/jwe.2019.16

Kumbhakar, S. C., and C. A. K. Lovell. 2000. Stochastic Frontier Analysis. Cambridge: Cambridge University Press.

Lecocq, S. and Visser, M. (2006). What Determines Wine Prices: Objective vs. Sensory Characteristics. Journal of Wine Economics 1(1), 42-56.

Lindley, D. V. (2006). Analysis of a wine tasting. Journal of Wine Economics 1(1), 33–41.

Luxen, M. F. (2018). Consensus between Ratings of Red Bordeaux Wines by Prominent Critics and Correlations with Prices 2004–2010 and 2011–2016: Ashton Revisited and Expanded. Journal of Wine Economics 13(1), 83–91. doi:10.1017/jwe.2018.1

Marks, D. (2020). Erring Experts? A Critique of Wine Ratings as Hedonic Scaling. Journal of Wine Economics 15(4), 386–393. doi:10.1017/jwe.2020.42

Oczkowski, E. (2016). Identifying the Effects of Objective and Subjective Quality on Wine Prices. Journal of Wine Economics 11(2), 249–260. doi:10.1017/jwe.2016.1

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 doi:10.1017/jwe.2015.3

Veale, R., and Quester, P. (2008). Consumer Sensory Evaluations of Wine Quality: The Respective Influence of Price and Country of Origin. Journal of Wine Economics 3(1), 10-29.