Journal of Wine Economics (2024), 1–21
doi:10.1017/jwe.2024.22

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Willingness to pay of Portuguese sparkling wine consumers: Econometric and machine learning approaches

Lina Lourenço-Gomes[1], Mário Gonzalez Pereira[2], Norberto Jorge Gonçalves[3], Tânia Gonçalves[1] and João Rebelo[1]

[1]University of Trás-os-Montes and Alto Douro (UTAD),Department of Economics, Sociology, and Management (DESG), Centre for Transdisciplinary Development Studies (CETRAD), Vila Real, Portugal; [2]Centre for Research and Technology of Agro-Environmental and Biological Sciences (CITAB), Inov4Agro, University of Trás-os-Montes and Alto Douro (UTAD), Vila Real, Portugal and [3]Department of Physics, School of Science and Technology, University of Trás-os-Montes and Alto Douro (UTAD), Vila Real, Portugal

Corresponding author: Lina Lourenço-Gomes; Email: lsofia@utad.pt

Abstract

Understanding consumer choices and their drivers of willingness to pay (WTP) for a bottle of wine has been a research challenge in wine economics, particularly in niche markets such as sparkling wine. This study investigates the determinants of WTP for sparkling wine based on data from Portuguese consumers. The results provided by two alternative methodologies are compared: a traditional econometric model, based on the estimation of an ordered probit model; and a modelling approach based on data-driven and using machine learning algorithms. Both approaches present similar results, highlighting the relevance of some determinants including income, Champagne brand, not being a protected designation of origin and being a red wine consumer as main predictors of WTP for sparkling wine in Portugal.

Keywords: consumer behavior; wine; willingness to pay; ordered probit model; machine learning

JEL classifications: C25; C45; D12

I. Introduction

Wine is widely perceived as an experience good whose market structure is typically a monopolistic competition, where information plays a key role in consumers' purchasing decisions and their willingness to pay (WTP). Research into wine prices or WTP allows for a better understanding of consumer behavior, including their reaction and sensitivity to price changes, to understand market trends, to anticipate changes in supply and demand, and to provide support for decision-making, in order to increase the economic performance of wineries (Le Fur et al., 2023). Therefore, knowledge about

the predictors, or determinants, that influence consumers' WTP is strategically important for wineries to increase their profitability and to develop more effective pricing and marketing strategies aligned with consumer preferences.

The WTP has been generally estimated through the hedonic price function that models the WTP as a function of predictive objective and subjective product attributes, as well as control variables as a set of consumers' socioeconomic characteristics. Thus, the estimation of WTP has been dominated by a theory-driven paradigm, in which the researcher imposes a structure on data, and models consumer decisions based on utility theory. The best model is selected by comparing the econometric results based on alternative functions differing both in terms of functional form and in the selected explanatory variables.

Nevertheless, alongside the restrictive assumptions about consumer behavior, it remains uncertain whether the chosen econometric model adequately represents the data generation process and if it can be used for predictive and assertive inferences (Rodrigues et al., 2022). To overcome these potential drawbacks, recent alternative modelling and estimation techniques have emerged based on data-driven analysis supported by machine learning (ML) algorithms, complementing and enlarging the traditional choice modelling approach (e.g., van Cranenburgh et al., 2022). This approach has been recently applied to the field of wine (e.g., Niklas and Rinke, 2020; Rinke and Ho, 2023). In this line of research, van Cranenburgh et al. (2022) point out the need for additional work integrating and comparing the results of the two modelling paradigms, the econometric approach and ML one, to which this paper aims to contribute. Associated with this objective, the main research questions are: (i) what are the main predictors of WTP for Portuguese sparkling wine consumers? and (ii) how similar are the results obtained with the econometric and ML approaches?

Thus, the contribution of this article is threefold: (i) identifying the main drivers of WTP of Portuguese sparkling wine consumers; (ii) exploring the determinants of WTP for wine, employing a dual lens of traditional econometric methods and contemporary ML techniques; and (iii) providing useful information for wineries to outline better marketing strategies for sparkling wine.

This article is structured as follows: section II presents an overview of the main potential predictors of WTP for wine, including the sparkling wine category; section III describes the data and methods, i.e., the ordered probit model and the ML approach; section IV describes and discusses the results; and, finally, section V presents the conclusions of the study.

II. Theoretical background

The WTP for wine has been the subject of extensive research in several countries over the last 30 years. The literature has revealed a series of determinants that influence WTP, using econometric models based on Lancaster's (1966) theory of consumer behavior, which emphasizes the importance of product characteristics. Generally, WTP is assumed to be influenced by intrinsic and extrinsic wine cues, which consumers use according to experiential and psychological factors, such as wine knowledge (Dodd et al., 2005) and involvement (Cox, 2009), as well as socioeconomic and demographic variables, including income, age, education, and socioeconomic status

(Elliot and Barth, 2012; Lange et al., 2002; Lerro et al., 2020; Skuras and Vakrou, 2002). Outreville and Le Fur (2020) present a descriptive review of empirical studies on the determinants of wine price during the 1993–2018 period. In the same line, Le Fur et al. (2023) include a recent and detailed literature review and a bibliometric analysis of academic research on wine prices in economics, covering 180 articles published in journals between 1992 and 2022. Both literature reviews provide useful knowledge on price predictors and the methodologies that have been used, highlighting the role of the theory-driven paradigm and the associated econometric estimation of hedonic price functions.

Early studies usually employed hedonic pricing models to analyze the formation of wine prices (Combris et al., 1997; e.g., Oczkowski, 1994), evaluating the price premium associated with specific characteristics. More recent econometric efforts have expanded their framework to include experimental auctions (Vecchio, 2013) and discrete choice experiments (D'Alessandro and Pecotich, 2013; Gonçalves et al., 2020; Palma et al., 2016), offering a deeper understanding of consumer preferences and their impact on WTP.

The results of previous studies obtained for sparkling wines suggest distinct preferences and WTP for appellations, Prosecco in this case (Onofri et al., 2015; Rossetto and Gastaldello, 2018; e.g., Thiene et al., 2013), brands (Vecchio et al., 2019), and loyalty (Bassi et al., 2021). Culbert et al. (2017) reveal that the production method influences the sensory profile of Australian sparkling white wine styles, and the Charmat is the preferred method. In turn, Lange et al. (2002) compare two mechanisms, the hedonic test and the Vickrey auction, to reveal consumers' WTP for Champagne, concluding that, in general, participants are willing to pay more for wines from big brands, and older consumers are willing to pay higher prices than younger people for reserve wines. Pickering et al. (2022) focused on the effect of label information, compare the evaluated WTP and quality perception in different information scenarios for a set of two simulated sparkling wine labels (Champagne and Prosecco). They conclude that those who consider themselves more knowledgeable about sparkling wines are willing to pay more for the Prosecco wine style and that WTP increases with the amount typically paid for both wine styles. Researchers have also found that when sparkling wine is purchased for a special occasion (e.g., a celebration), consumers are often willing to spend more (Morton et al., 2004; Velikova et al., 2016). Verdonk et al. (2017) point out that when consumers buy wine as a gift, they are more willing to purchase expensive, prestigious brands of sparkling wine, including Champagne.

In parallel with advances in econometric models, the development of data science and ML, well described in the literature, has opened new avenues for the analysis of wine WTP. ML techniques, known for their ability to deal with extensive data sets and discover nonlinear relationships, have demonstrated high potential in the wine industry. For example, some studies have applied various ML algorithms to predict wine quality (Jain et al., 2023), taste preferences (Cortez et al., 2009), wine price (Niklas and Rinke, 2020), and indirectly shedding light on WTP.

In summary, the literature review allowed us to conclude that the main predictors of WTP can be organized into four groups of variables: (i) socioeconomic characteristics (gender, age, education, marital status, place of residence, income); (ii) personal and behavioral aspects (consumer knowledge, who buys, motives of purchase, knowledge

of production method, consumption frequency, consumption of other beverages); (iii) collective (region of origin, collective brand) and individual (brands and awards) reputations; and (iv) attributes of the sparkling wine (e.g., category, sweetness, production method, and organic production).

III. Data and methods

a. Data

Based on the findings provided by the literature review and advice from experts in the sparkling wine market, the authors developed a survey that includes three groups of questions: (i) purchasing and consumption patterns; (ii) discrete choice experiment; and (iii) sociodemographic information. Data collection was carried out online and managed by a company specialized in market research in Portugal, in September 2022. This company guaranteed a representative sample in terms of gender, age, income, and professional status. The sample has information from 800 people over 18 years old who consumed sparkling wine at least once a year. Table A1 (Appendix A) describes the variables used in this study and briefly characterizes the data collected. Next, we will present the main characteristics of the database.

Regarding the price of a bottle of sparkling wine (PriceRan), only 8% of the respondents revealed willingness to spend more than €15, while the majority (>70%) said they would spend less than €10 per bottle.

Concerning the socioeconomic variables, 51% of respondents are male; 28.9% of the sample is aged between 55 and 64, followed by 18.4% in the 35 to 44 age group; 83.6% of the sample lives in urban areas (residence); 42% earn between €1,000 and €1,999.99 net income per month. The analysis of the professional status of the interviewees reveals that 58% are employees, 16% are self-employed, 15% are retired, 6% are unemployed, and 5% are students. More than 90% of respondents purchase sparkling wine for celebrations, which is in line with evidence reported in other studies.

Concerning the personal and behavioral variables, 61% of respondents are responsible for purchasing sparkling wine (BuySparkling variable). When asked about their knowledge about sparkling wines (SparklingKnow), 51% of those interviewed assume they know a little about sparkling wines, almost 30% consider themselves to have moderate knowledge, 16.6% are new to this market, and only less than 3% state to know a lot or consider themselves experts in sparkling wine. More than 70% claim to know the traditional/classic production method (ClassTrad), 16.4% know the Charmat method, 14.9% know both methods, and 26.9% do not know either of them. More than 90% buy sparkling wine for celebrations (BuyCeleb), 77.1% for other events or parties (BuyParty), 33.6% to consume during meals (BuyFood), and 28.5% buy sparkling wine as a gift (BuyGift). Around 37% consume sparkling wine between 5 and 10 times/year and 2.4% consume it at least once a week (SparklingCons). Sparkling wine is consumed as an aperitif (32%), in cocktails (45%), during (54%) and after (59%) meals. The majority of sparkling wine consumers in the sample report consuming other alcoholic beverages, such as red (ConsRed, 82%) and white wine (ConsWhite, 85%), beer (ConsBeer, 86%), spirits (ConsSpirits, 60%), sangria (a wine-based drink flavored with fruit and spices originating in Portugal and Spain, ConsSangria 64%), and soft drinks

(ConsSoft, 71%). Concerning the wines' collective reputation, i.e., the region of origin, the choices were Bairrada (35%), Champagne (15.4%), and Távora (14.5%), but 20.6% have no preferred region (NotPDO). Concerning the attributes of the sparkling wine and personal preferences, the individual brand, the importance of awards and being organic are highlighted. In summary, the data collected a heterogeneous e sociodemographic profile of Portuguese sparkling wine consumers, considering their preferences, consumption and purchasing habits, as well as the evaluation of the product's attributes for decision-making.

IV Methods

a Ordered probit model

Taking into account the nature of the dependent variable the price range or the WTP for sparkling wine ordered by classes, an econometric ordered probit model is estimated. Under this model, there is a latent continuous metric underlying the ordinal responses, being the latent continuous dependent variable y_i^* determined by a vector of explanatory variables (x_i) and a disturbance term (ε_i). Therefore, the ordered probit regression method allows modelling of the ordered outcome based on J categories (in our case four categories) of WTP, as a linear function of the observed vector of x_i. The latent regression is specified as follows:

$$y_i^* = \beta' x_i + \varepsilon_i, \varepsilon_i \sim N(0,1), \forall i = 1, \ldots, N. \quad (1)$$

where i is the Nth observation, y^* is the unobserved $N \times 1$ dependent variable, β' is the vector of $K \times 1$ estimated parameters, x ($N \times K$) are the covariates (predictors) assumed to be independent of ε, and ε ($N \times 1$) is the error term including unobservable factors. The probabilities underlying this model are given by

$$Prob[y = 0] = \Phi(-\beta' x)$$

$$Prob[y = 1] = \Phi(\mu_1 - \beta' x) - \Phi(-\beta' x)$$

$$Prob[y = 2] = \Phi(\mu_2 - \beta' x) - \Phi(\mu_1 - \beta' x)$$

$$\ldots$$

$$Prob[y = J] = 1 - \Phi(\mu_{J-1} - \beta' x)$$

where $\Phi(\bullet)$ stands for the cumulative distribution function, and $\mu_j, j = 1, \ldots, K$, are the unknown threshold parameters, between which the categorical responses are estimated. The model estimation through the likelihood function is based on the implied probabilities.

b Machine learning

In the data-driven approach, the identification and ordering of statistically significant predictors of WTP, among a vast set (pool) of potential explanatory variables, named as features in this context, is a category of supervised ML named classification, in which an algorithm "learns" to classify new observations from examples of input and output labelled/categorical data (Kotsiantis et al., 2007; Mohamed, 2017; Nasteski, 2017; Osisanwo et al., 2017; Singh et al., 2016). This approach includes the use of feature selection or feature ranking algorithms (FRA) followed by the development (training/calibration) of a WTP classification model.

In this study, features were ordered and classification models were trained and tested with (i) Orange, which is an open-source ML and data visualization software, that creates data analysis workflows visually, with a large and diverse toolbox and, (ii) MATLAB Classification Learner (CL), which allows the user to explore the data, select features, train, validate, and tune classification models for binary or multiclass problems using supervised ML statistics and ML toolbox 12.4. The choice to use both software packages depends on the user's specific needs and technical proficiency. Both Orange and MATLAB CL are widely used because they are particularly appealing to users seeking an accessible and intuitive tool for data mining and ML (Ciaburro, 2017; Demšar et al., 2013).

Different FRAs were tested and used to rank the features, namely the minimum redundancy maximum relevance (mRMR), univariate FRA for classification using chi-square tests (χ^2), ReliefF algorithm with k nearest neighbors, one-way ANOVA for each predictor variable grouped by class, Kruskal–Wallis test (KW), information gain (InforGain), gain ratio, Gini index, fast correlation based filter. These FRAs support categorical and continuous features and are well described in the literature (e.g., Guyon and Elisseeff, 2003; Radovic et al., 2017). Nevertheless, we present a brief description of the FRA in Table B1 (Appendix B). The selected features will be presented and sorted in descending order of scores. For χ^2, ANOVA and KW, the features are ranked using the $p - values$ since scores correspond to $-log\,(p)$.

The 42 different classification models used are of nine types: decision trees, discriminant analysis, support vector machines, logistic regression, nearest neighbors, naive bayes, kernel approximation, neural network, and ensemble classifiers. A detailed description, comparison and review of these classifiers are easily found in the MATLAB Help Center and in the literature (e.g., Guyon and Elisseeff, 2003; Kotsiantis et al., 2007; Mohamed, 2017; Nasteski, 2017; Osisanwo et al., 2017; Radovic et al., 2017; Singh et al., 2016). However, we present a brief description of these algorithms in Table C1 (Appendix C). The accuracy ratio (AR), defined as the ratio of cases correctly predicted during calibration, is used as a measure of the performance of the classification models calibrated with the features/variables selected by each FRA. The AR of each FRA is the maximum performance achieved by the classification models.

V. Results and discussion

The results of an ordered probit model with ordered price intervals (in euro) as the dependent variable, which is a proxy for WTP for a bottle of a sparkling wine, include the ordered list of the 34 covariates or predictors, their respective regression coefficients

and standard errors (Table 1 and **Table A2**, Appendix A). It is important to highlight that the null hypothesis of errors normally distributed is observed, the value of the likelihood ratio test also confirms that the parameters are globally significant at a 1% level and the ratio of correctly predicted cases (AR) is 45% (Table 1). Based on the decreasing level of significance (1%, 5%, and 10%), the 15 predictors statistically significant of the ordered probit WTP (Table 1) are the Champagne brand, income, importance of awards (ImpDCE2), being the one who buys or purchases the product, the importance given by the respondent to a set of variables related to organic production (ImpDCE6), production method (ImpDCE5), and sweetness (ImpDCE4), to be a consumer of red wine, to buy sparkling wine as a gift, the importance of the brand, gender, local of residence, region of production and not being protected designation of origin (PDO).

For the sake of simplicity, only the results of five FRAs that present the highest AR among the ML algorithms available in CL and Orange (Table 1) will be presented, although it is important to note that the results of different FRAs are very similar. For example, in CL, the results obtained with χ^2, ANOVA and KW are very alike. The features ranked by ANOVA and KW in the first eleventh positions are the same and sorted in the same order. Three of the last four features selected by these two algorithms are also the same, although they are ordered differently. Features selected with χ^2 are also similar and only differ in the order between some pairs of successive features. The other algorithms (ReliefF, mRMR, InforGain) selected only some of the same features, not always in the same order. However, approximately the same set of features tend to be selected for the top seven positions, namely Income, Champagne, NotPDO, ClassTrad, ImpBrand, BuySparkling, and Tavora. The features selected with the other FRAs of Orange are relatively similar. For example, the set of features in the top seven positions is very similar, except for BuySprakling and Tavora, which are only selected once. On the other hand, BuyGift, ImpDCE1, and ImpDCE3 are selected twice by the Orange FRA. The IngorGain and Gini select the same variables but in slightly different order. The other three used FRAs present some similarities in the common features selected.

Regarding the performance of classification models, the AR is slightly higher for the ordered probit model (45%) than for the ML methods, which ranges between 41% and 43%. AR is only slightly higher for models available in Orange than in CL. In the case of CL FRA, the classifiers with the highest performance are the linear discriminant analysis model for χ^2 and KW, weighted KNN (nearest neighbor classifier) for ReliefF, Ensemble Subspace Discriminant (ensemble classifier) for mRMR, Kernel Naïve Bayes for ANOVA. For Orange, the SVM provides the highest AR for χ^2, Neural Networks for InofrGain e Gini, and Naïve Bayes for gain ratio and ReliefF.

Since, for χ^2, ANOVA and KW, the scores correspond to—log(p), it is important to mention that for these FRA the number of features with p-values below 5% is 11 for ANOVA, 12 for χ^2 and 14 for KW. However, classification models were calibrated with several predictors ranging between 12 and 15 with very similar AR. These results suggest that the inclusion of additional, but less important predictors does not lead to better-performing models. Results obtained with ML methods suggest a tendency toward similarity in the main predictors/features selected by the different methods. This finding is likely a consequence of the fact that the initial features (predictors) in

Table 1. Top 15 features selected using the algorithm mRMR, χ^2, ReliefF, ANOVA, Kruskal–Wallis (KW) Test and InforGain

Theory-driven model	Data-driven approach(Machine learning ML)									
	MATLAB Classification Learner					Orange				
Ordered probit	mRMR	χ^2	ReliefF	ANOVA	KW	χ^2	Gain ratio	InforGain	Gini	ReliefF
Champagne	Income	Income	Champagne	Champagne	Champagne	Champagne	Champagne	Income	Income	ConsSpirits
Income	Residence	Champagne	Income	Income	Income	NotPDO	NotPDO	Champagne	NotPDO	NotPDO
ImpDCE2	BuySparkling	NotPDO	Charmat	NotPDO	NotPDO	Income	Income	ImpBrand	ImpDCE3	ImpDCE2
BuySparkling	Champagne	ClasTrad	Marital	ImpBrand	ImpBrand	Tavora	ClasTrad	NotPDO	ImpBrand	ConsSoft
ImpDCE6	BuyGift	ImpBrand	Gender	ClasTrad	ClasTrad	ImpBrand	Tavora	ImpDCE3	Champagne	ImpDCE3
ImpDCE5	ConsSoft	BuySparkling	ImpDCE1	BuySparkling	BuySparkling	Charmat	ImpBrand	ClasTrad	ClasTrad	ImpDCE4
ImpDCE3	ClasTrad	Tavora	ConsBeer	Tavora	Tavora	BuyGift	BuySparkling	ImpDCE1	ImpDCE1	BuyGift
ImpDCE4	Prosecco	BuyGift	ConsSpirits	BuyGift	BuyGift	ClasTrad	Charmat	ImpDCE5	Tavora	SparklingKnown
ConsRed	ConsRed	Charmat	Sparklingcons	Charmat	Charmat	BuySparkling	ConsRed	Sparklingcons	ImpDCE5	ImpBrand
BuyGift	Tavora	ConsSoft	ConsSoft	ConsSoft	ConsSoft	ImpDCE3	BuyGift	ImpDCE2	ImpDCE6	ImpDCE1
ImpBrand	ImpBrand	ConsRed	Age	ConsRed	ConsRed	ImpDCE2	ConsSoft	ImpDCE6	ImpDCE2	BuyParty
Gender	Charmat	ImpDCE3	BuyCeleb	ConsSpirits	ImpDCE3	Residence	ImpDCE3	BuySparkling	Sparklingcons	Income
Residence	NotPDO	ConsSpirits	SparklingKnown	ImpDCE3	ImpDCE2	Cava	Residence	Tavora	BuySparkling	ClasTrad
ImpDCE1	BuyCeleb	Residence	ImpDCE4	ImpDCE2	ImpDCE6	AgreFam	SparklingKnown	SparklingKnown	AgreFam	Prosecco
NotPDO	ImpDCE4	SparklingKnown	ImpDCE2	Residence	ConsSpirits	Prosecco	BuyCeleb	AgreFam	SparklingKnown	Sparklingcons
AR = 45%	AR = 41%	AR = 42%	AR = 41%	AR = 42%	AR = 42%	AR = 43%	AR = 42%	AR = 42%	AR = 42%	AR = 42%

AR = accuracy ratio defined as the ratio of cases correctly predicted

https://doi.org/10.1017/jwe.2024.22 Published online by Cambridge University Press

the ML approach are the same as the covariates in the probit model, the inclusion of which is supported by previous studies.

In general, the findings highlight the relevance of the same variables, namely income, the Champagne brand, not being PDO, the traditional/classic production method and the importance of the brand. Additionally, the ordered probit stresses the significance of the six variables that express the personal importance of the respondents given to a set of variables related to production region, awards, categories, sweetness, production method, and organic. Although all ML models confirm the importance of income, Champagne and ClasTrad, the evidence reported for the other variables is different, which draws attention to the need to choose the best method based on a given performance measure.

The results are in line with the findings of previous studies. The brand, Champagne appellation and income are strong determinants of Portuguese WTP for sparkling wine (e.g., Pickering et al., 2022). The greatest importance of these three characteristics for modeling WTP was evidenced with the ordered probit model, four of the five MATLAB classifiers and one Orange model, although two other Orange classifiers select champagne and income in the first three positions. Individuals are willing to pay more for big brands, confirming that Champagne has a strong collective reputation as an indicator of status (Combris et al., 2006; Dal Bianco et al., 2018; e.g., Lange et al., 2002; Pickering et al., 2022; Verdonk et al., 2017). Pickering et al. (2022) compared Champagne and Prosecco wine style labels and found that respondents with higher incomes are willing to pay more for both sparkling wine styles than their counterparts. In the same line, as expected, the WTP for sparkling wine is influenced by the absence of a PDO, suggesting that the *terroir* collective reputation that comes from the designation of origin, affects the WTP for sparkling wine.

Additionally, there seems to exist some differentiation between male and female consumers, in line with findings from previous studies. Female consumers are found to consume significantly more sparkling wine than men. Women are slightly more willing to pay for sparkling wine than men, which could reflect the perception that sparkling wine is "feminine" or a "women's drink," possibly due to its connotations of glamor and romanticism (Bruwer and McCutcheon, 2017; Stephen Charters, 2005). Furthermore, women tend to be the main shoppers in their households, which may also explain differences in the WTP (e.g., Marshall and Anderson, 2000). Our results show that being responsible for purchasing wine increases the WTP of sparkling wine because it enhances the involvement with wine, which is related to the amount typically spent on a bottle of wine, as spending consumers are the most involved with wine in general (Thach and Olsen, 2015).

A relationship is also observed between the consumption of sparkling wine and the consumption habits of other drinks, in the sense that red wine consumers seem to be positively correlated to the WTP of sparkling wine. This relationship may be associated with the strong presence of a traditional consumption model, in which families buy more wine, especially still red wine, to consume with meals (Dal Bianco et al., 2018). In this sense, there is a complementary effect between the consumption of red wine with meals and the WTP for sparkling wine for consumption outside meals.

In line with other studies (e.g., Stephen Charters, 2005; Steve Charters et al., 2011, for still wine), our results suggest that sparkling wine is perceived as a separate product

type from other beverages (e.g. still white wine, beer, soft drinks, spirits, and sangria). One reason for this result may be that households that consume significant quantities of wine tend to purchase cheaper products, and sensitivity to branded prices is not a significant determinant of their WTP for sparkling wines (Dal Bianco et al., 2018). As demonstrated previously, the present study also confirms that consumers are usually willing to spend more on sparkling wines purchased for special occasions, such as festive events/seasons and offers/gifts, which attest to the importance of the purchasing context (e.g., Morton et al., 2004; Velikova et al., 2016; Verdonk et al., 2017).

Finally, broad consistency was observed for personal preferences underlying the choice in the decision-making process and WTP highlighting the importance of wine cues (e.g., ImpDCE1 to ImpDCE6) (e.g., Ferreira et al., 2021); production region (e.g., Verdonk et al., 2017); awards or reputation; categories; sweetness (e.g., Combris et al., 2006; Verdonk et al., 2017); production method (e.g., Culbert et al., 2017); and being organic (e.g., Schäufele and Hamm, 2017). This outcome reinforces the importance of the characteristics revealed to the consumer (on the bottle and label), to explain differences in the consumers' price (e.g., Combris et al., 2006; Lecocq and Visser, 2006).

VI. Conclusion

This study analyzes consumers' WTP for sparkling wine in Portugal. Research on this topic has typically been backed by a theoretical model grounded in utility theory, employing econometric estimation methods tailored to the data structure and study objectives. This study deepens the analysis by comparing the results of a traditional model used to understand the determinants of sparkling wine WTP (ordered probit model) of Portuguese consumers, with the evidence produced by recently emerging alternative methods rooted in ML algorithms.

The results suggest that there is no absolute supremacy of any of the approaches in terms of global performance, although the ordered probit model presents a slightly better performance in the accuracy rate of correctly predicted cases. The two approaches tend to select the same main predictors, highlighting the relevance of the income variable, the Champagne brand, not being a PDO and being a red wine consumer as the main predictors of WTP for sparkling wine in Portugal. However, the advances suggested by ML are quite variable depending on the algorithm and platforms used, which draws attention to the need to choose the best method based on another specific performance measure. In this sense, it should be pointed out that ML classification models are characterized by being parameterizable algorithms. Thus, although a large number of methods were used in this study, this number could have been much higher if other options/parameterizations had been chosen. On the other hand, although the use of ML tools does not require prior knowledge of the relationship to be modelled, it benefits from knowledge of the characteristics of each algorithm and the relationships to be modelled.

This paper contributes to consolidating knowledge on the modelling of consumer behavior and provides useful information for wineries' marketing strategies. Specifically, the results indicate that to increase sparkling wine sales at a higher price, wineries should segment the market according to income, focusing on higher-income niches that are also red wine consumers. At the same time, they should follow the

Champagne strategy as a benchmark, create dynamics of collective and individual reputation, reinforce and benefit from the PDO, boost wine routes, carry out visits and tastings in the vineyard and in the cellar, participating in competitions and tastings, and developing cooperative actions with hotels and restaurants that value the sparkling wine. Moreover, since the ordering of WTP predictors changes with the used method, a detailed analysis and weighting of the different ordering, i.e., a quantified sensibility analysis, is recommended for the robustness of the winery's marketing plan outlined to a target market.

The authors are aware that the used methods and, consequently, the obtained results of this study can be extended, for instance, by applying and integrating the two analytical paradigms in the choice modelling perspective, as highlighted by van Cranenburgh et al. (2022). Moreover, in this study, the departing features of the ML methods are the same as the covariates (predictors) of the ordered probit model, which is supported by previous studies, remaining the research question as to whether ML methods are not especially suited for unstructured big data with limited knowledge about the influence on WTP. In this way, it is suggested as a research trend to apply ML techniques to data from digital platforms such as Google Trends and Vivino.

Acknowledgments. The authors thank an anonymous reviewer and the editor for insightful and constructive comments.

This work was supported by national funds, by the FCT—Portuguese Foundation for Science and Technology under the project UIDB/04011/2020 (https://doi.org/10.54499/UIDB/04011/2020) and the project UIDB/04033/2020 (https://doi.org/10.54499/UIDB/04033/2020).

Availability of data and materials. The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Competing interests. The authors declare they have no competing interests.

References

Bassi, F., Pennoni, F., and Rossetto, L. (2021). Market segmentation and dynamic analysis of sparkling wine purchases in Italy. *Journal of Wine Economics*, 16(3), 283–304. doi:10.1017/JWE.2021.20.

Bruwer, J., and McCutcheon, E. (2017). Marketing implications from a behaviourism perspective of consumption dynamics and socio-demographics of wine consumers. *Asia Pacific Journal of Marketing and Logistics*, 29(3), 519–537. doi:10.1108/APJML-06-2016-0095.

Charters, S. (2005). Drinking sparkling wine: An exploratory investigation. *International Journal of Wine Marketing*, 17(1), 54–68. doi:10.1108/eb008783.

Charters, S., Velikova, N., Ritchie, C., Fountain, J., Thach, L., Dodd, T. H., Fish, N., Herbst, F., and Terblanche, N. (2011). Generation Y and sparkling wines: A cross-cultural perspective. *International Journal of Wine Business Research*, 23(2), 161–175. doi:10.1108/17511061111143016.

Ciaburro, G. (2017). *MATLAB for Machine Learning*. Packt Publishing Ltd.

Combris, P., Lange, C., and Issanchou, S. (2006). Assessing the effect of information on the reservation price for Champagne: What are consumers actually paying for? *Journal of Wine Economics*, 1(1), 75–88. doi:10.1017/S1931436100000109.

Combris, P., Lecocq, S., and Visser, M. (1997). Estimation of a hedonic price equation for Bordeaux wine: Does quality matter? *The Economic Journal*, 107(441), 390–402. doi:10.1111/j.0013-0133.1997.165.x.

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., and Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47(4), 547–553. doi:10.1016/j.dss.2009.05.016.

Cox, D. (2009). Predicting consumption, wine involvement and perceived quality of Australian red wine. *Journal of Wine Research*, 20(3), 209–229. doi:10.1080/09571260903450963.

Culbert, J. A., Ristic, R., Ovington, L. A., Saliba, A. J., and Wilkinson, K. L. (2017). Influence of production method on the sensory profile and consumer acceptance of Australian sparkling white wine styles. *Australian Journal of Grape and Wine Research*, 23(2), 170–178. doi:10.1111/ajgw.12277.

Dal Bianco, A., Boatto, V., Trestini, S., and Caracciolo, F. (2018). Understanding consumption choice of prosecco wine: An empirical analysis using Italian and German Homescan data. *Journal of Wine Research*, 29(3), 190–203. doi:10.1080/09571264.2018.1506322.

D'Alessandro, S., and Pecotich, A. (2013). Evaluation of wine by expert and novice consumers in the presence of variations in quality, brand and country of origin cues. *Food Quality and Preference*, 28(1), 287–303. doi:10.1016/j.foodqual.2012.10.002.

Demšar, J., Erjavec, A., Hočevar, T., Milutinovič, M., Možina, M., Toplak, M., Umek, L., Zbontar, J., and Zupan, B. (2013). Orange: Data mining toolbox in Python. *Journal of Machine Learning Research*, 14(1), 2349–2353. https://jmlr.csail.mit.edu/papers/volume14/demsar13a/demsar13a.pdf.

Ding, C., and Peng, H. (2011). Minimum redundancy feature selection from microarray gene expression data. *Journal of Bioinformatics and Computational Biology*, 3(2), 185–205. doi:10.1142/s0219720005001004.

Dodd, T. H., Laverie, D. A., Wilcox, J. F., and Duhan, D. F. (2005). Differential effects of experience, subjective knowledge, and objective knowledge on sources of information used in consumer wine purchasing. 29(1), 3–19. 10.1177/1096348004267518.

Elliot, S., and Barth, J. E. (2012). Wine label design and personality preferences of millennials. *Journal of Product and Brand Management*, 21(3), 183–191. doi:10.1108/10610421211228801/FULL/XML.

Ferreira, C., Costa Pinto, L. M., and Lourenço-Gomes, L. (2021). Effect of region of origin on willingness to pay for wine: An experimental auction. *Applied Economics*, 53(32), 3715–3729. doi:10.1080/00036846.2021.1885611.

Gonçalves, T., Lourenço-Gomes, L., and Pinto, L. (2020). Modelling consumer preferences heterogeneity in emerging wine markets: A latent class analysis. *Applied Economics*, 52(56), 6136–6144. doi:10.1080/00036846.2020.1784389.

Guyon, I., and Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3(Mar), 1157–1182. https://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf.

Han, J., Kamber, M., and Pei, J. (2012). *Data Mining: Concepts and Techniques*. 3rd Edition, Waltham: Morgan Kaufmann Publishers.

Herbrich, R. (2001). *Learning Kernel Classifiers: Theory and Algorithms*. The MIT Press. https://books.google.com/books/about/Learning_Kernel_Classifiers.html?hl=pt-PT&id=el_cDjHdP0cC

Hogg, R. V., and Ledolter, J. (1987). Engineering statistics. In: *Engineering Statistics*. Michigan: Macmillan. https://cir.nii.ac.jp/crid/1130000798294433024.

Jain, K., Kaushik, K., Gupta, S. K., Mahajan, S., and Kadry, S. (2023). Machine learning-based predictive modelling for the enhancement of wine quality. *Scientific Reports*, 13(1), 1–18. doi:10.1038/s41598-023-44111-9.

Kotsiantis, S. B., Zaharakis, I., and Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, 160(1), 3–24. https://datajobs.com/data-science-repo/Supervised-Learning-[SB-Kotsiantis].pdf.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74(2), 132–157. doi:10.1086/259131.

Lange, C., Martin, C., Chabanet, C., Combris, P., and Issanchou, S. (2002). Impact of the information provided to consumers on their willingness to pay for Champagne: Comparison with hedonic scores. *Food Quality and Preference*, 13(7–8), 597–608. doi:10.1016/S0950-3293(02)00059-9.

Lecocq, S., and Visser, M. (2006). What determines wine prices: Objective vs. sensory characteristics*. *Journal of Wine Economics*, 1(1), 42–56. doi:10.1017/S1931436100000080.

Le Fur, E., Thelisson, A. S., and Guyottot, O. (2023). Wine prices in economics: A bibliometric analysis. *Strategic Change*. doi:10.1002/JSC.2561.

Lerro, M., Vecchio, R., Nazzaro, C., and Pomarici, E. (2020). The growing (good) bubbles: Insights into US consumers of sparkling wine. *British Food Journal*, 122(8), 2371–2384. doi:10.1108/BFJ-02-2019-0139.

Marshall, D. W., and Anderson, A. S. (2000). Who's responsible for the food shopping? A study of young Scottish couples in their "honeymoon" period. *The International Review of Retail, Distribution and Consumer Research*, 10(1), 59–72. doi:10.1080/095939600342406.

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia Medica*, 23(2), 143–149. doi:10.11613/BM.2013.018.

McKight, P. E., and Najab, J. (2010). Kruskal-Wallis Test. *The Corsini Encyclopedia of Psychology*, 1–1. doi:10.1002/9780470479216.CORPSY0491.

Mohamed, A. E. (2017). Comparative study of four supervised machine learning techniques for classification. *International Journal of Applied Science and Technology*, 7(2). www.ijastnet.com

Morton, A.-L., Rivers, C., and Healy, M. (2004). Beyond the bubbles: Identifying other purchase decision variables beyond country of origin effect that make Australians buy champagne. *ANZIBA Conference Proceedings 2004: Dynamism and Challenges in Internationalisation*.

Nasteski, V. (2017). *An overview of the supervised machine learning methods*. 10.20544/HORIZONS.B.04.1.17.P05.

Niklas, B., and Rinke, W. (2020). Pricing models for German wine: Hedonic regression vs. machine learning. *Journal of Wine Economics*, 15(3), 284–311. doi:10.1017/jwe.2020.16.

Oczkowski, E. (1994). A HEDONIC PRICE FUNCTION FOR AUSTRALIAN PREMIUM TABLE WINE. *Australian Journal of Agricultural Economics*, 38(1), 93–110. doi:10.1111/j.1467-8489.1994.tb00721.x.

Onofri, L., Boatto, V., and Bianco, A. D. (2015). Who likes it "sparkling"? An empirical analysis of Prosecco consumers' profile. *Agricultural and Food Economics*, 3(1), 1–15. doi:10.1186/s40100-014-0026-x.

Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., and Akinjobi, J. (2017). Supervised machine learning algorithms: Classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3), 128–138. doi:10.14445/22312803/IJCTT-V48P126.

Outreville, J. F., and Le Fur, E. (2020). Hedonic price functions and wine price determinants: A review of empirical research. *Journal of Agricultural and Food Industrial Organization*, 18(2). doi:10.1515/jafio-2019-0028.

Palma, D., Ortúzar, J. D. D., Rizzi, L. I., Guevara, C. A., Casaubon, G., and Ma, H. (2016). Modelling choice when price is a cue for quality: A case study with Chinese consumers. *Journal of Choice Modelling*, 19, 24–39. doi:10.1016/j.jocm.2016.06.002.

Pickering, G. J., Duben, M., and Kemp, B. (2022). The importance of informational components of sparkling wine labels varies with key consumer characteristics. *Beverages*, 8(2), 27. doi:10.3390/beverages8020027.

Radovic, M., Ghalwash, M., Filipovic, N., and Obradovic, Z. (2017). Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinformatics*, 18(1), 1–14. doi:10.1186/s12859-016-1423-9.

Rinke, W., and Ho, S.-T. (2023). Have consumers escaped from COVID-19 restrictions by seeking variety? A machine learning approach analyzing wine purchase behavior in the United States. *Journal of Wine Economics*, 1–10. doi:10.1017/JWE.2023.25.

Robnik-Šikonja, M., and Kononenko, I. (2003). Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning*, 53(1–2), 23–69. doi:10.1023/A:1025667309714.

Rodrigues, F., Ortelli, N., Bierlaire, M., and Pereira, F. C. (2022). Bayesian automatic relevance determination for utility function specification in discrete choice models. *IEEE Transactions on Intelligent Transportation Systems*, 23(4), 3126–3136. doi:10.1109/TITS.2020.3031965.

Rossetto, L., and Gastaldello, G. (2018). The loyalty structure of sparkling wine brands in Italy. *Journal of Wine Economics*, 13(4), 409–418. doi:10.1017/jwe.2018.43.

Schäufele, I., and Hamm, U. (2017). Consumers' perceptions, preferences and willingness-to-pay for wine with sustainability characteristics: A review. *Journal of Cleaner Production*, 147, 379–394. doi:10.1016/j.jclepro.2017.01.118.

Singh, A., Thakur, N., and Sharma, A. (2016). A review of supervised machine learning algorithms. *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 1310–1315.

Skuras, D., and Vakrou, A. (2002). Consumers' willingness to pay for origin labelled wine: A Greek case study. *British Food Journal*, 104(11), 898–912. doi:10.1108/00070700210454622.

Thach, L., and Olsen, J. (2015). Profiling the high frequency wine consumer by price segmentation in the US market. *Wine Economics and Policy*, 4(1), 53–59. doi:10.1016/j.wep.2015.04.001.

Thiene, M., Scarpa, R., Galletto, L., and Boatto, V. (2013). Sparkling wine choice from supermarket shelves: The impact of certification of origin and production practices. *Agricultural Economics*, 44(4–5), 523–536. doi:10.1111/agec.12036.

van Cranenburgh, S., Wang, S., Vij, A., Pereira, F., and Walker, J. (2022). Choice modelling in the age of machine learning - Discussion paper. *Journal of Choice Modelling*, 42, 100340. doi:10.1016/j.jocm.2021.100340.

Vecchio, R. (2013). Determinants of willingness-to-pay for sustainable wine: Evidence from experimental auctions. *Wine Economics and Policy*, 2(2), 85–92. doi:10.1016/j.wep.2013.11.002.

Vecchio, R., Lisanti, M. T., Caracciolo, F., Cembalo, L., Gambuti, A., Moio, L., Siani, T., Marotta, G., Nazzaro, C., and Piombino, P. (2019). The role of production process and information on quality expectations and perceptions of sparkling wines. *Journal of the Science of Food & Agriculture*, 99(1), 124–135. doi:10.1002/jsfa.9153.

Velikova, N., Charters, S., Fountain, J., Ritchie, C., Fish, N., and Dodd, T. (2016). Status or fun? A cross-cultural examination of young consumers' responses to images of champagne and sparkling wine. *British Food Journal*, 118(8), 1960–1975. doi:10.1108/BFJ-12-2015-0497.

Verdonk, N., Wilkinson, J., Culbert, J., Ristic, R., Pearce, K., and Wilkinson, K. (2017). Toward a model of sparkling wine purchasing preferences. *International Journal of Wine Business Research*, 29(1), 58–73. doi:10.1108/IJWBR-10-2015-0048.

Williams, C. K. I. (2003). Learning kernel classifiers. *Journal of the American Statistical Association*, 98, 489–490. doi:10.1198/jasa.2003.s270.

Appendix A

Table A1. Summary of the data collected including the names of the dependent and independent variables, the value or range of values for each class, and the proportion of the data set in each class

Variables	Scale	#respondents	%
Dependent (WTP)			
Price range (€/bottle) PriceRan			
Class 1	0–6.99€	296	37.0
Class 2	7€–9.99€	272	34.0
Class 3	10€–14.99€	168	21.0
Class 4	More than 15€	64	8.0
Explanatory/Predictors			
Socioeconomic characteristics			
Gender	1 if male	408	51.0
Age group (%)	<25	78	9.7
	25 – 34	84	10.5
	35 – 44	147	18.4
	45 – 54	145	18.1
	55 – 64	232	29.0
	>64	114	14.3
Education	High school	376	47.0
	Bachelor's degree	309	38.6
	Master or PhD	115	14.4

(*Continued*)

Table A1. (*Continued.*)

Variables	Scale	#respondents	%
Dependent (WTP)			
Price range (€/bottle) PriceRan			
Marital status	1 if married or equivalent	520	65.0
Residence	1 if rural	144	18.0
Income (household monthly net income €)	Less than 650€	20	2.5
	650–999.99	97	12.2
	1000–1999.99	336	42.0
	2000–2999.99	225	28.1
	3000– 3999.99	87	10.8
	Over 3999.99€	35	4.4
Personal and behavioral aspects			
SparklingKnow (subjective sparkling wine knowledge)	A new entrant in the sparkling world	133	16.6
	Know a little	408	51.0
	Moderate	236	29.5
	Know a lot	21	2.6
	Expert	2	0.3
Buy sparkling	1 if is responsible for purchasing wine	488	61.0
Motives of purchase			
BuyGift	1 if to offer	228	28.5
BuyFood	1 if for meals	269	33.6
BuyCeleb	1 if for celebrations	727	90.9
BuyParty	1 if for other events	617	77.1
Known production method			
ClasTrad	1 if knows Traditional/Classic	573	71.6
Charmat	1 if knows Charmat	131	16.4
SparklingCons (frequency of consumption)	Once a year	25	3.1
	2–4 times/year	245	30.6
	5–10 times/year	295	36.9
	Once a month	135	16.9
	2–3 times/per month	81	10.1
	One or more times a week	19	2.4

(*Continued*)

https://doi.org/10.1017/jwe.2024.22 Published online by Cambridge University Press

Table A1. (*Continued.*)

Variables	Scale	#respondents	%
Dependent (WTP)			
Price range (€/bottle) PriceRan			
Consumption of other beverages			
ConsRed (red wine)	1 if yes	656	82.0
ConsWhite (white wine)	1 if yes	680	85.0
ConsBeer (beer)	1 if yes	688	86.0
ConsSoft (soft drinks)	1 if yes	568	71.0
ConsSpirits (spirits)	1 if yes	480	60.0
ConsSangria ("sangria")	1 if yes	512	64.0
Willingness to pay (bottle for celebration)	Mean (€)		18.5
Collective reputation			
Region of origin of purchased sparkling			
Bairrada	1 if Bairrada (Portugal)	283	35.4
Cava	1 if Cava (Spain)	104	13.0
Champagne	1 if Champagne (France)	123	15.4
Prosecco	1 if Prosecco (Italy)	760	95.0
Tavora	1 if Távora-Varosa (Portugal)	123	15.4
NotPDO	1 if does not consider this cue	165	20.6
Attributes of the sparkling wine			
	Scale	Median	
ImpBrand (importance of the brand)	1 (not) to 5 (highly important)	3	
Importance of personal preferences			
ImpDCE1 (importance of production region)	1 (more) to 7 (less important)	3	
ImpDCE2 (importance of awards)	1 (more) to 7 (less important)	5	
ImpDCE3 (importance of categories)	1 (more) to 7 (less important)	4	
ImpDCE4 (importance of sweetness)	1 (more) to 7 (less important)	1	
ImpDCE5 (importance of production method)	1 (more) to 7 (less important)	5	
ImpDCE6 (importance of being organic)	1 (more) to 7 (less important)	6	

Table A2. Results obtained with the ordered probit model for price range as the dependent variable, including the explanatory variables and the values of the coefficient and standard error

Explanatory variables	*Coefficient*	*Std. Error*
Gender	−0.1967**	0.0897
Age	0.0011	0.0038
AgreFam	−0.0103	0.0370
Marital	−0.0078	0.001
Income	0.2110***	0.0426
Residence	−0.2395**	0.1126
SparklingKnown	−0.0999	0.0608
Sparklingcons	−0.0599	0.0423
BuySparkling	0.3350***	0.0883
ConsRed	0.2941**	0.1169
ConsWhite	0.0365	0.1268
ConsBeer	−0.1794	0.1380
ConsSoft	0.0530	0.1009
ConsSpirits	0.0738	0.0873
ConsSangria	−0.0214	0.0963
BuyCeleb	0.1726	0.1586
BuyParty	−0.0638	0.1096
BuyFood	−0.0689	0.0950
BuyGift	0.2186**	0.0936
Tavora	0.2263	0.1436
Bairrada	0.1187	0.1287
Champagne	0.6324***	0.1560
Prosecco	−0.1799	0.1208
Cava	−0.0350	0.1051
NotPDO	−0.2552*	0.1447
ImpBrand	0.1040**	0.0470
ClasTrad	0.1596	0.1028
Charmat	0.1088	0.1107
ImpDCE1	−0.0642**	0.0308
ImpDCE2	−0.1403***	0.0300
ImpDCE3	−0.1148***	0.0312
ImpDCE4	−0.1212***	0.0386
ImpDCE5	−0.1004***	0.0310
ImpDCE6	−0.1323***	0.0345
cut1	−1.9116***	0.6982

(*Continued*)

Table A2. (*Continued.*)

Explanatory variables	*Coefficient*	*Std. Error*
cut2	−0.8984	0.6934
cut3	0.0992	0.6953
Mean dependent var	1.0000	
Log-likelihood	−917.9055	
Schwarz criterion	2083.1420	
S.D. dependent var	0.9493	
Akaike criterion	1909.8110	
Hannan-Quinn	1976.3970	
Number of cases "correctly predicted"	360 (45.0%)	
Likelihood ratio test	Chi-square (34) = 187.33 [0.0000]	
Test for normality of residual	Null hypothesis: error is normally distributed Test statistic: Chi-square(2) = 3.4834, with asymptotic *p*-value = 0.175	

***significant at the 1%
**significant at the 5%
*significant at the 10%

Appendix B

Table B1. Brief description of the machine learning (ML) feature ranking algorithms (FRAs). Based on the information provided by MATLAB help center (https://www.mathworks.com/help/stats/classificationlearner-app.html)

Feature ranking algorithm	Description
Minimum redundancy maximum relevance (mRMR)	The mRMR (minimum redundancy maximum relevance) algorithm selects an optimal feature set that maximally represents the response variable by maximizing relevance to the response while minimizing inter-feature redundancy. It employs mutual information to measure feature importance and prioritizes features based on a balance of relevance and redundancy. A feature's significance is determined through a scoring system, with higher scores indicating greater importance (Ding and Peng, 2011).
Chi2 (χ^2)	Test whether each predictor variable is independent of the response variable using individual χ^2 tests, and then rank the features using the p-values from the χ^2 test statistics (McHugh, 2013). Scores correspond to $-log\,(p)$. A small p-value (large score) indicates that the feature depends on the response variable and therefore, is an important feature. If $p - value < eps\,(0)$ ($eps \equiv floating - point\ relative\ accuracy$), then the output is *Inf*.

(*Continued*)

Table B1. (*Continued.*)

Feature ranking algorithm	Description
ReliefF	The ReliefF and RReliefF algorithms rank features by estimating predictor weights and distinguishing between categorical and continuous variables, respectively. They assess the importance of predictors based on their ability to differentiate between nearest neighbors of the same or different classes, penalizing or rewarding accordingly. Weights are iteratively updated, reflecting a predictor's relevance; higher weights signify greater importance, while negative weights suggest a predictor's inefficacy. RReliefF adapts this process for continuous outcomes (Robnik-Šikonja and Kononenko, 2003).
Analysis of variance (ANOVA)	Classify the features using the *p*-values from a one-way ANOVA performed for each predictor variable, grouped by class (Hogg and Ledolter, 1987). For each predictor variable, the hypothesis that the predictor values grouped by response classes are drawn from populations with the same mean is tested against the alternative hypothesis that population means are not all equal. Scores correspond to $-log\,(p)$ and the predictor variables are ordered according to scores.
Kruskal–Wallis	The algorithm utilizes Kruskal–Wallis Test *p*-values to rank features, offering a nonparametric alternative to ANOVA for comparing medians across multiple groups. This method evaluates whether sample groups derive from populations with identical distributions, employing ranks for analysis rather than direct numeric values. It distinguishes between populations based on median differences, substituting the F-statistic with an χ^2 statistic for significance testing. Predictor variables are assessed to determine if grouped values indicate differing population medians, with scores indicated by $-log\,(p)$ (McKight and Najab, 2010).

Table B2. Brief description of the machine learning (ML) feature ranking algorithms (FRAs) implemented by Orange software package, Version 3.35.0

Feature ranking algorithm	Description
InforGain (IG)	After observing the feature, IG measures the reduction in entropy or uncertainty about the target variable. Entropy is a metric that quantifies the amount of uncertainty or randomness in a dataset's class distribution. A feature with high IG helps distinguish between the classes, thus reducing entropy or unpredictability. The formula for IG, given a feature F and target variable T, is: $IG\,(T, F) = H\,(T) - H\,(T \mid F)$ where: $H\,(T)$ is the entropy of the target variable before the feature is observed, $H\,(T \mid F)$ is the conditional entropy of the target given the feature, which represents the uncertainty in the target after observing the feature. In feature selection or while building decision tree nodes, the goal is to maximize IG. We aim to choose features that provide the most information about the classification or outcome (Han et al., 2012).

(*Continued*)

Table B2. (*Continued.*)

Feature ranking algorithm	Description
Gain ratio (GR)	It was introduced as an enhancement over Information Gain to address its bias toward features with many values. IG tends to prefer attributes that split the dataset into many small partitions, which might not always be helpful. The GR compensates for this by normalizing the IG by the intrinsic information of a split, which measures the potential information generated by splitting the training set into multiple partitions, regardless of the actual outcomes. The GR for a feature F concerning the target variable T is calculated as: Gain Ratio $(T, F) = \frac{IG(T,F)}{IV(F)}$ where: IG(T, F) is the information gain of T given F, which measures the reduction in entropy or uncertainty about T after observing and IV(F) is the intrinsic value of F, which measures the information generated by splitting the dataset based on F, defined as $IV\,(F) = -\sum_{i=1}^{n} \frac{|S|}{|S_i|} \log_2 \frac{|S|}{|S_i|}$, where S_i are the subsets created by splitting S based on F (Han et al., 2012).
Gini	It measures the degree of impurity or purity of a set of elements, indicating how often a randomly chosen element from the set would be incorrectly labelled if it was randomly labelled according to the distribution of labels in the subset. A Gini impurity score of 0 signifies perfect purity, meaning all elements in the subset belong to a single class. In contrast, a score closer to 1 indicates high impurity, with elements distributed across multiple classes. The formula for calculating the Gini impurity of a set S containing instances of multiple classes is: Gini $(S) = 1 - \sum_{i=1}^{n} p_i^2$ where p_i is the proportion of instances of class i in the subset S and n is the number of classes (Han et al., 2012).

Appendix C

Table C1. Brief description of the machine learning (ML) classification models implemented by MATLAB[1] and Orange 3[2], Version 3.35.0 (mainly based on Bishop 2006)

Classification models	Description
Decision trees (DT)[1]	DT is a versatile and useful tool for classification. It features a hierarchical structure with root, internal, and leaf nodes, where each leaf represents a response. Require minimal data preparation, and are easy to interpret, but vulnerable to high variances and can be expensive to fit.
Discriminant analysis (DA)[1]	DA classifies observations by leveraging Gaussian distributions to model different classes with linear or nonlinear thresholds to separate them. DA are simple, efficient, and able to handle high-dimensional data and multicollinearity, but assume shared mean distributions across classes.

(*Continued*)

Table C1. (*Continued.*)

Classification models	Description
Logistic regression classifiers (LRC)[1]	Logistic regression shares similarities with linear regression but focuses on the probability of categorical outcomes. Its interpretability is a key advantage, but it demands larger, representative samples and can overfit with too many predictors.
Naive bayes classifiers (NBC)[1,2]	NBC models input distributions within classes without prioritizing feature importance. Assumes predictor independence, but performs well even when this condition isn't met. Its advantages include simplicity, scalability, and effectiveness with high-dimensional data, but it faces challenges with zero frequency and its core independence assumption.
Support vector machines (SVMs)[1,2]	SVMs linearly and nonlinearly classify data with varying flexibility by maximizing the margin between classes in N-dimensional space. SVMs are preferred for high-dimensional data, outperforming NBC, LRC, and DT in certain scenarios, but require hyperparameter tuning and can be computationally intensive compared to NN.
Nearest neighbor classifiers (K-nearest neighbors, KNN)[1,2]	KNN is a nonparametric method that is based on the proximity of a data point to its 'K' nearest neighbors. Flexibility, bias and variance depend on K. KNN has high predictive accuracy in lower dimensional spaces, but requires more memory and may be less interpretable in higher dimensions.
Kernel approximation classifiers (KAC)	KACs are used for nonlinear classification of data with many observations (Herbrich, 2001; Williams, 2003) when they tend to train and predict faster than SVM classifiers with Gaussian kernels. KACs' flexibility is medium, but increases as the Kernel scale decreases, both for the SVM and LR types.
Ensemble classifiers (EC)	EC enhance accuracy by combining multiple models: bagging (averages decisions from different data samples), stacking (combines different models' predictions), and boosting (improves predictions of each member). Bagging reduces variance but can be computationally demanding and less interpretable; boosting minimizes bias but risks overfitting.
Neural network classifiers (NN)	NN are inspired by the human brain's functioning and consists of interconnected (input, hidden, and output) layers of nodes. NN are easily adaptable, vary in complexity, offer medium to high flexibility, and present high predictive accuracy but are challenging to interpret.

Notes: The results of DT, DA, LRC, and NBC are easy to interpret while KNN, KAC, EC, and NN classifiers are difficult to interpret. The results of SVMs are easy to interpret if linear but hard for all other kernel types; All classification models accept exclusively numerical or categorical predictors and partly numerical and partly categorical predictors, except DA and EC. However, Classification Learner only offers users the models available according to input data type.

Cite this article: Lourenço-Gomes, L., Pereira, M. G., Gonçalves, N. J., Gonçalves, T., and Rebelo, J. (2024). Willingness to pay of Portuguese sparkling wine consumers: Econometric and machine learning approaches. *Journal of Wine Economics*, 1–21. https://doi.org/10.1017/jwe.2024.22