AMERICAN ASSOCIATION OF WINE ECONOMISTS

AAWE WORKING PAPER
No. 28
Economics

UNOBSERVED HETEROGENEITY IN THE WINE MARKET: AN ANALYSIS OF SARDINIAN WINE USING MIXED LOGIT

Maria Bonaria Lai, Teresa Del Giudice, Eugenio Pomarici

November 2008

www.wine-economics.org

Unobserved heterogeneity in the wine market: an analysis on Sardinian wine via Mixed Logit

Maria Bonaria Lai , Teresa Del Giudice, Eugenio Pomarici[1]

1. Introduction

The evolution experienced by the food sector in the last ten years has greatly changed the models which analysts employ to explore the purchase and consumption patterns of post-industrial society. New consumption patterns are based partly on undifferentiated food products and partly on high quality ones, like high quality wine, which are strongly characterised by a close link to specific geographical areas, and their local traditions in food production.
The present study attempts to offer more insights into the Italian wine, in particular the Sardinia wine market with emphasis on wines produced by local vineyards by relating wine choice to consumers' preferences.

We assess the effects of such characteristics of the wine produced in Sardinia on consumer preferences, using a discrete choice model. Perceived quality in wine is complex and often operationalized by multi-dimensional constructs, whose measurement requires sophisticated approaches. Wine is a food category where consumers' quality perception is particularly difficult, because, among other things, wine is a highly differentiated product.

The paper presents some results from a stated-preference study on data collected by a web-based survey of 138 Italian consumers. A series of multinomial logit models are estimated from choice experiment responses and tested for unobserved heterogeneity for some wine attributes. The consequences of such form of heterogeneity are flashed out with respect to issues of market segmentation on the basis of the pattern of correlation across preferences as estimated from mixed logit models.

[1] Lai Maria Bonaria is contract professor at the Economics Faculty, University of Cagliari, V.le S. Ignazio, 17, 09123 Cagliari, Italy (mblai @unica.it). Del Giudice Teresa is researcher at Agricultural Economics and Policy Department, University of Naples Federico II, Via Università, 96, 80055 Portici (NA), Italy (agriqual@unina.it). Pomarici Eugenio is professor at Agricultural Economics and Policy Department, University of Naples Federico II, Via Università, 96, 80055 Portici (NA), Italy (pomarici@unina.it). Paragraph 2 and 5 have been written by Lai; paragraph 3 and 4 by Del Giudice. Introduction and conclusion are common to the three authors.

The paper is organized as follows: the following section illustrates the survey and data; section 3 shows the research methods and theory behind our approach. Section 4 discusses the characteristic of the econometric model, while section 5 presents the results of the choice model. At the end, section 6 discusses some conclusions.

2. The survey and the data

The data used to analyze the consumer preferences for wine in Italy come from an electronic questionnaire. It has been sent to a sample of Italian wine consumers that buy wine through e-commerce and consequently through the electronic wine shop. The sample is composed by 138 consumers.

The questionnaire used is divided into three sections.

The first one is dedicated to the preference analysis. Customers have been asked to choose between different wines (9 possible choices), characterized by some different attributes. In detail, here customers have been asked to order the product from the most preferred to the less favourite item.

Color, vineyard, denomination and price are the attributes used to specify the different product profiles (table 1) [2].

Table n. 1- Attributes and levels used in the ranking survey

Color	Denomination	Vine	Price €/0.75 l
Red	DOCG o DOC	Typical	5
White	Igt	International	7.50
Light red	without		9

The second part aims at determining purchase behavior and the customer preferences about geographic origin, color, taste and age of wines. The third part aims at determining socio-economics characteristic of the sample.

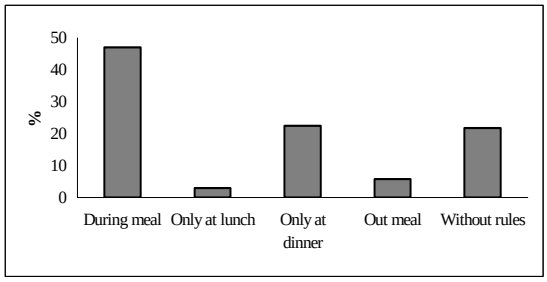
The sample

Purchase and consumer behaviour

87% of the sample judged the wine as a constant element of one's food shopping, although the percentage of subjects that daily consume it results a bit lower (64%). 67% prefer to consume Italian wines, in particular red, old and dry taste wines. They drink it during their main meals (lunch and/or dinner).

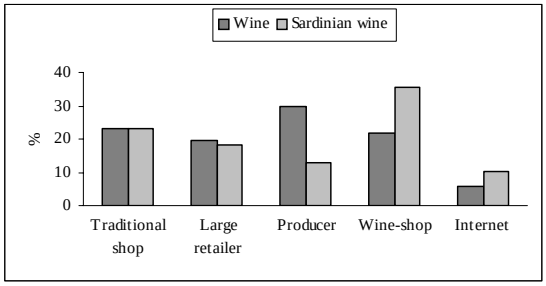
[2] The choice of the attributes was made during five interviews and one focus group.

Figure n. 1 Consumption of wine



The place to buy seems to change on the basis of the typology of wine purchased. For example, the producer is the most favorite place to buy "common"[3] wine for 30% of the interviewees. In general, the wine-shop is the most preferred place to buy Sardinian wine (36% of consumers in the sample), followed by traditional food shops (23%) and large retailers (18%).

Figure n. 2 Place to buy wine



Two other aspects seem to be quite interesting. The first one is related to the purchases in the traditional food shops. For this category, the percentage of customers is the same (23%) for the different kind of wines purchased (Sardinian or not). While the latter aspect concerns virtual shops and their increasing market shares. In this research 10% of respondents have declared to buy Sardinian wine through virtual shops.

Socio-economic characteristics

In the sample, there is a strong male presence (79%). 81% was between 18-50 years old and 68% belong to a family composed of three or more persons.

Education level is middle high, 47% has got a high school diploma, 38% has a degree, while 12% has achieved the middle licence, the others did not answer to this question.

67% of the interviewees are an independent professional, manager or employee. As for their income, the group with incomes between 30.000 and 60.000 euro is the most important one (59%).

[3] The word *common* has been used here to indicate the wine without denomination.

With regards to the origin, almost all Italian regions are presented in the sample.

3 Theory and methods

Much empirical work from qualitative choice theory has made use of random utility-based discrete choice models (Ben-Akiva and Lerman, 1985; Train, 2003). This approach is vastly adopted when modelling choices come from sets of multi-attribute alternatives. The consumer's selection of wine from the large and varied range of available alternatives can be cast in this framework and analysed with multinomial logit models, or some of its more flexible extensions (e.g. G.E.V., nested logit etc.). However, one important limit of these models is that they do not allow for taste heterogeneity unconditionally on socio-economic covariates (i.e. unobserved heterogeneity). In other words, changes in taste seem to be poorly explained by the socio-economic covariates for which one typically has some form of (often self-reported) measure. And even when socio-economic covariates are adequately measured, it may well be that a large part of taste variation is independent of these. This may well be one of the reasons why "unobserved" taste heterogeneity models often seem to explain much of residual taste variation, even when large numbers of socio-economic covariates are included in the random utility model of choice (Scarpa *et al.*, 2003). More importantly, consumers may display regularity in the form of interdependence existing between tastes for certain attributes. In other words, the distribution of tastes may be correlated, and we show how the existence of this correlation may be exploited to investigate the magnitude of preference-based market segments. This may, for example, be a fruitful avenue to characterize niche markets (Loureiro *et al.,* 2002, Scarpa and Del Giudice, 2004).

Of great relevance in this context is a new category of models with unprecedented flexibility in representing patterns of taste variation: the mixed logit models (MXL). In fact, it is shown (McFadden and Train, 2000) that MXL may be used to approximate *any* form of preference. The amount of literature that has rapidly developed around this category of models is a measure of both the relative simplicity with which they can be estimated and the appeal they exercise over applied researchers (Train, 2003) with applications ranging from choice of angling destinations (Train, 1998), to the estimation of public benefits from traffic calming (Garrod *et al.,* 2002), to trading decisions in the pastoralists cattle markets of Kenya (Scarpa *et al.*, 2003), to choice of road transport modes in New Zealand (Hensher and Greene, 2003). MXL are basically multinomial logit models (MNL) with one or more random parameters in the indirect utility function (Revelt and Train, 1998) or some additional error components (Brownstone and Train, 1999). We focus here on the random parameter specification, where randomness affects taste parameters and model results

can be interpreted in the conventional way. When randomness is in the error component, the error means are constrained to be zero, so only the spread parameters need estimation and flexible substitution patterns are introduced across alternatives. Hybrid forms including both error components and random taste parameters may be adequate in some choice contexts (Termansen *et al*., 2004). In estimation of MXL the emphasis is moved from the estimation of the values of taste parameters to the estimation of the parameters of the distributions that regulate the stochastic behaviour (randomness) of these values. For example, in our study the appeal for "Sardinia quality wine" is thought to vary in the population according to a normal distribution with mean μ and variance σ, then the task of MXL estimation is to find the values for these two parameters, which in turn are assumed to describe the random behaviour of taste for "Sardinia wine" in the population. The implication is that one can attach probability statements to the values of taste parameters, and as a consequence, to functions of these, such as part-worths and consumer surplus estimates.

4. The econometric model

In this analysis, we use ranking data. These represent consumer preferences on the basis of ordered sets of alternatives. In our study each alternative is represented by a type of wine and characterised by the presence or absence of certain product attributes and price levels. The sets of alternatives were created by means of a partial factorial orthogonal design, which guaranteed the identification of all taste parameters during estimation (Kuhfeld, 2003). Each respondent (i) ordered the 9 alternatives from the most favourite to the least, thereby giving rise to 8 choices. Under independence the joint probability of the ranked choices in the sample of 138 respondents is therefore:

$$L = \prod_{n=1}^{N=138} \Pr_{ji} \left(U_{j,i=1} \succ U_{j,i=2} \succ ..... \succ U_{j,i=9}\right) \qquad (1)$$

Under linear random utility with Gumbel distributed errors the probability inside the productory operator is a product of logit probabilities. As such the MNL model can be used to obtain, under the correct specification, asymptotically consistent estimates of the taste parameters by maximizing the following log-likelihood based on the so called "exploded logit" model for ranked choices (Hausman and Ruud, 1987):

$$\ln L = \sum_{n=1}^{N=138} \ln \left[\prod_{t=1}^{t=8} \frac{e^{\lambda \beta_{Xt}}}{\sum_{j=1}^{J} e^{\lambda \beta X_{jt}}} \right] \text{where } i, j \in J \qquad (2)$$

In the above $\mathbf{x}_{it}$ represents a vector of price and dummy variables.

MXL models are just an extension of MNL models whereby some taste parameters $\tilde{\beta}$ are random and distributed in the population according to a predetermined law. In our case the probability of each respondent having a particular value of the taste parameter $\tilde{\beta}_n$ when making a choice is assumed to be normal. Conditional on this event the selection probability of an alternative is logit; so computing the marginal probability requires integration over all the possible values of $\tilde{\beta}_n \in$ of its distribution:

$$\Pr(U_j \succ U_{\neq j}) = \int_{\beta_n \in} \prod_{t=1}^{t=8} \frac{e^{\lambda \beta_n x_{it}}}{\sum_{j=1}^{J} e^{\lambda \beta_n x_{jt}}} f(\beta_n | \mu, \Omega) d\beta_n \qquad (3)$$

Such probability does not have a closed form, hence simulation methods need to be employed for the estimation of the set of parameters μ and Ω regulating the value of the random parameters $\tilde{\beta}$. Such methods are well illustrated by Train (2003) and need not be repeated here.

5. Preference analysis

The data regarding the ranking of 9 alternatives have been analysed in two steps. The first one is the application of a Multinomial logit with fixed parameters, while the second one concerns applications of mixed logit[4].

The Mixed logit or random parameters logit, based on the equation (3), has been estimated assuming the dependence among individual choices, the correlation between the different attributes and, assuming for these a normal distribution.

In detail, the first assumption has allowed to use the panel data approach considering 8 periods equal to customer choice moments. Meanwhile, the correlation between the different characteristics allows a more real preference analysis.

In table 2, it is possible to observe the estimates obtained by the fixed parameters model (MLFP) and by the random parameters model (RLP) in table 3.

[4] LIMPDEP (nlogit2) is the software used to estimate the models.

The significant of the estimated attributes variance, the value of the likelihood function and the adjusted R-square represent the indexes used to verify and to compare the different models.

Table n. 2 Multinomial logit

Variables	Value	μ[5]	b/St.Er
Log-L	-1650.44		
Adjusted R^2	30.15		
red		0.8772	9.899
doc		0.8402	9.796
typical		0.4576	6.203
white		0.7183	8.336
igt		0.4332	5.344
price		-0.4762	2.477

Table n. 3 Mixed logit

Variables	Value	μ	b/St.Er	σ	b/St.Er
Log-L	-2425.74				
Adjusted R^2	0.33871				
Red		1.2762	7.214	1.6584	8.705
Doc		1.2833	10.053	0.7783	6.061
Typical		0.6708	5.25	0.9049	4.439
White		1.0228	6.251	1.3307	7.966
Igt		0.614	6.089		
Price		-0.605	6.251		

In order to analyze how each attribute changes inside the population, the normal distributions of the parameters “red” and “doc” are drawn in figure 3.

Computing at zero the density function of these random variables *F(0)*, it is possible to obtain a percentage value of the customers for those the considered attributes is non-preferred, in other words their influence on the consumers utility function is negative.

In detail, 19% of consumers don’t prefer the red wine, while only 13% do not prefer the doc denomination.

[5] It is possible to consider the estimate coefficient for each attribute as the average and, the standard deviation as scrap from the average. So for this reason they have been indicated inside the tables with the letters μ and σ.

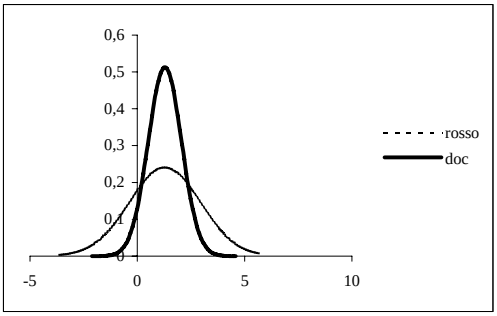


Figure n. 3 Distribution of parameters red and doc

We have also analyzed consumers separately on the basis of their wine purchase point. The results related to consumers that buy wine in specialized shop and the other consumers have shown in table 4.

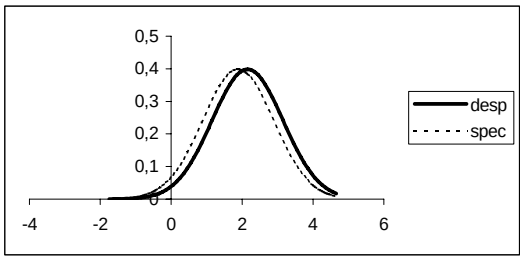
Table n. 4 Mixed Logit

Variables	Specialized shops					Non specialized shops				
	Value	μ	b/St.Er	σ	b/St.Er	Value	μ	b/St.Er	σ	b/St.Er
Log-L	-1212.9					-1212.9				
Adjusted R^2	0.3219					0.3584				
red		0.8695	3.237	1.4989	6.208		1.8664	7.3930	1.4227	5.291
doc		1.2237	6.467	0.5633	2.868		1.3541	6.3870	1.0010	4.6666
typical		0.5135	2.575	0.9022	2.747		0.7414	3.6060	0.9360	3.281
white		0.7168	2.859	1.2881	1.838		1.3292	5.8150	1.2395	4.549
igt		0.4331	3.122				0.8846	5.3500		
price		-0.8196	2.09				-0.5336	0.1680		

Consumers of these two different segments choose Sardinian wine on the basis of the colour attribute. Red is a favourite wine and its implicit price is 3.50 €. Consumers that buy in specialized shops mainly prefer the attribute denomination (doc); again also in this channel the red wine results to be favourite over the white one.

Considering red and doc as random parameters, the normal distributions show that customers seem to prefer more the doc attribute. In this trade channels the percentage of consumers that don't prefer this attribute is equal to 16% for the not specialized shops, and 7% for the specialized shops.

Figure n. 4 Distribution of doc parameter



If we consider the red attribute, the percentage is 12% for the customers of the despecialized channel and 23% for customers of the specialized channel.

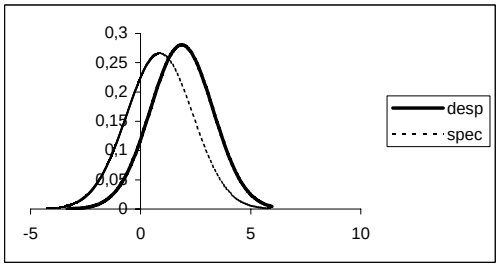


Figure n. 5 Distribution of red parameter

We have also divided the sample on the basis of the geographical origin (North and Centre, South and Islands). The results show that the Northern consumers mainly prefer the denomination attribute, while the Southern customers consider the colour attribute, preferring the purchase and consumption of red wine.

Table n. 5 Mixed Logit

Variables	North and Centre					South and Island				
	Value	μ	b/St.Er	σ	b/St.Er	Value	μ	b/St.Er	σ	b/St.Er
Log-L	-1546.8					-878.89				
Adjusted R^2	0.34157					0.33767				
red		1.1900	5.501	1.3672	5.424		1.4402	3.605	2.0996	5.837
doc		1.3349	8.513	0.6991	4.784		1.0852	4.793	0.7732	3.298
typical		0.6383	4.077	0.7694	2.909		0.6692	2.992	0.9345	3.533
white		1.1096	5.446	1.1586	4.981		0.6645	1.617	1.6443	3.65
igt		0.7013	5.888				0.5248	2.312		
price		-0.2673	0.973				-0.7632	2.139		

Due to a relevant heterogeneity in the taste, the distribution for consumers of the South of Italy is more plate than the same distribution for consumers of North of Italy.

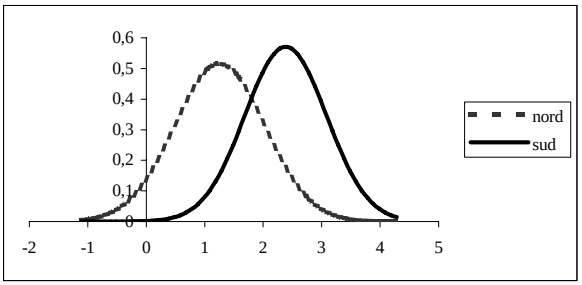


Figure n. 6 Distribution of doc parameter

All customers interviewed and resident in the South of Italy like the attribute doc, while a small percentage (9) of North customers don't like the same attribute.

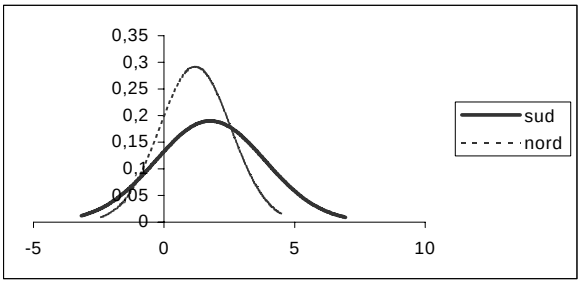


Figure n. 7 Distribution of red parameter

6 Conclusions

The present study carried out on the preferences of Italian consumers gives some interesting results from the methodological point of view and concerning the results obtained. The applied methodology confirms his attitude to reveal the heterogeneity in preferences, and this indicates that the Mixed Logit allows the understanding of how in large samples preferences for different attributes are distributed; this gives a useful indication in the discovery of large or small segments to reach with a specific supply.

The results show a strong interest in the sample for the denomination of origin as a first quality signal. The simple geographical indication is also associated with some utility but at a lower degree. The interest for a quality signal associated with origin is consistent with the preference shown for typical wines. The red and the white wines are actually preferred rather than the rosè, and the red appears more appreciated with respect to the white one. Preferences appear rather independent from the type of outlet where the wine is bought; the consumer's place of residence seems to determine some changes in the degree of utility of the preferred attributes.

7. References

Baker, G. A. and Burnham T. A. "Consumer Response to Genetically Modified Foods: Market Segment Analysis and Implications for Producers and Policy Makers," *Journal of Agricultural and Resource Economics*, 26(2) (2001) 387-403.

Ben-Akiva, M. E., and S. R. Lerman *Discrete choice analysis: theory and application to predict travel demand*. Cambridge: MIT press, 1985.

Bonnet, C., and M. Simioni, "Assessing Consumer Response to Protected Designation of Origin Labelling: A Mixed Multinomial Logit Approach", *European Review of Agricultural Economics*, 28(4) (2001): 433-449.

Brownstone, D., and K. Train, "Forecasting New Product Penetration with Flexible Substitution Patterns", *Journal of Econometrics*, 89(1) (1999): 109-129

Burton, M., Rigby, D., Young, T., and S. James, "Consumer attitudes to genetically modified organisms in food in the UK", *European Review of Agriculture Economics*, 28(4) (2001): 479-498.

Cicia, G., Del Giudice, T., and R. Scarpa, "Consumers' Perception of Quality in Organic Food: a random utility model under preference heterogeneity and choice correlation from rank-orderings", *British Food Journal:* 104 (3/4/5) (2002): 200-213.

Enoteca Italiana, *The list of Doc and DOCG wine*, XV Edition, Siena, Ente vini Enoteca italiana, 2002.

Garrod, G. D., Scarpa, R., and K. G. Willis, "Estimating The Benefits of Traffic Calming on Through Routes: A Choice Experiment Approach", *Journal of Transport Economics and Policy*, 36(2) (2002): 211-231.

Hanley, N., G. Koop, B. Álvarez-Fariz, R. E. Wright and C. Nevin, "Go climb a mountain: an application of recreation demand models to rock climbing in Scotland", *Journal of Agricultural Economics*, 52(1) (2001): 36-52.

Hausman, J. and Ruud P. "Specifying and Testing Econometric Models for rank-ordered data." *Journal of Econometrics*, 34 (1987): 83-104.

Hensher, David A. and William H. Greene, "The mixed logit model: the state of practice", Transportation 30 (2003): 133-176.

Kuhfeld, W. F. *Marketing Research Methods in SAS: Experimental Design, Choice, Conjoint, and Graphical Techniques*. Cary, NC, USA: SAS Institute Inc., 2003

Loureiro, M. L. and Hine S. "Discovering Niche Markets: A Comparison of Consumer Willingness to Pay for Local (Colorado Grown), Organic and GMO-Free Products," *Journal of Agricultural and Applied Economics*, 34(3)(2002): 477-487.

McFadden, D., and K. E. Train, "Mixed MNL models of discrete response", *Journal of Applied Econometrics*, 15 (2000): 447-70.

Rabobank, *Wine is business, shifting demand and distribution: major drivers reshaping the wine industry*, Food & Agribusiness Research, Utrecht, Rabobank International, 2003.

Revelt, D., and K. E Train, "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level", *The Review of Economics and Statistics*, LXXX (4) (1998): 647-657.

Scarpa, R., E.S.K. Ruto, P. Kristjanson, M. Radeny, A. Drucker, and J.E.O. Rege, "Valuing indigenous cattle breeds in Kenya: An empirical comparison of stated and revealed preference value estimates", *Ecological Economics*: 45(3) (2003): 409-426.

Scarpa R., Del Giudice T., Market Segmentation via Mixed Logit: Extra-Virgin Olive Oil in Urban Italy, Journal of Agricultural & Food Industrial Organization, vol. 2, No. 1, Article 7, http://www.bepress.com/jafio/vol2/iss1/art7, 2004.

Termansen, M., McLean C. and R. Scarpa. "Economic Valuation of Danish Forest Recreation combining mixed logit models and GIS" Paper presented at the EAERE conference Budapest, June 25th-28th, 2004.

Train, K. E, "Recreation Demand Models with taste differences over people", *Land Economics*, 74(2) (1998): 230-39.

Train, K. E. *Discrete choice methods with simulation*. Cambridge: University Press, 2003.