Modelling choice when price is a cue for quality

A case study with Chinese wine consumers

Abstract

Experience products are those the quality of which cannot be ascertained until after consumption, forcing consumers to base their purchase decision on an expectation of the product's quality. This expected quality is based on cues available before purchase, among which price is noteworthy, as consumers tend to believe that higher prices imply higher quality. But price also stresses the consumers' budget restriction, inducing a double -and conflicting- global effect on purchase probability. Using the traditional formulation of Random Utility Models for experience goods (i.e. introducing all attributes directly in the utility function) can lead to an endogeneity problem due to the omission of expected quality, introducing bias on the results.

Using a stated wine choice experiment conducted in China as a case study, we correct for endogeneity by modelling each alternative's expected quality as a latent variable, explained by all available quality cues, including price. Then we explain choice as a trade-off between price and expected quality. This allows us to separate both effects of price and correct for at least one source of endogeneity while being consistent with behavioural theory; this has either been ignored or not treated correctly in previous literature. Moreover, as the model requires only a quality indicator for each alternative to achieve identification, the respondents' burden increases marginally.

Our results show that the use of latent variables reduces endogeneity and effectively allows to measure both effects of price separately, obtaining higher significance and correct signs for its parameters.

Keywords: endogeneity, hybrid choice models, latent variables, experience good, wine.

1 Introduction

Price is a key attribute in choice experiments. It is not only relevant for consumers and producers, but from a modelling perspective it is also used to calculate willingness to pay (WTP) estimates and the price elasticity of demand. However, the effect of price on consumers can be twofold. The quality of certain products, such as new foods and beverages, is uncertain before purchase because it cannot be fully evaluated until after consumption (Nelson, 1970; Grisolía et al., 2012). Other products, such as jewellery and some medicines have uncertain qualities even after purchase, as consumer do not have the means or knowledge to determine them. In these cases, consumers resort to extrinsic cues to determine product quality (i.e. they construct an expected quality). Any attribute that can be perceived before purchase, such as packaging, publicity, health claims, store advertising, etc., can constitute an extrinsic cue for quality. Among these, price may become a highly relevant cue for quality (Leavitt, 1954), as consumers tend to assume that higher prices are associated with higher quality in the case of many products. In these cases, price has a double effect: a positive one due to its role as a cue for quality, and a negative effect due to the consumers' budget constraints.

In discrete choice models, the double effect of price can generate an endogeneity problem, causing coefficient estimates to be biased. This happens because as modellers, we do not observe consumers' expected quality for the product, and as this variable correlates with price, omitting it from the utility function makes price endogenous.

Even though the literature offers several approaches to deal with endogeneity in discrete choice models (Guevara 2015), the latent variable approach is particularly suitable for cases where quality is uncertain to the consumer at the time of purchase. It also provides a reliable framework both from a methodological and a behavioural perspective, as it employs a tested econometric approach (hybrid choice models, Walker & Ben-Akiva 2002) and a well-developed behavioural theory (signalling mechanisms, Milgrom & Roberts 1986).

The approach consists in modelling expected quality as a latent variable, explained by the product's observable attributes (including price), while the actual purchase choice is explained by the trade-off between price and expected quality. This easily fits the frame of a stated preference (SP) experiment, where besides recording participants' choices, only an additional indicator of quality is required. Under the appropriate structure, the modeller can correct for endogeneity and measure both the positive and negative effects of price, while keeping the analysis in line with behavioural theory and not overwhelming respondents with excessive additional tasks.

In this paper, we use wine as a case study to test the latent variable approach to correct for endogeneity, in accordance to behavioural theory. To this end, we use a computer-based stated choice experiment which was responded by a particular sample of Chinese wine consumers, experts and students. We find that the method provides promising results, and propose further topics for future research.

The rest of the paper is structured as follows. Section 2 presents a brief literature review about the double effect of price and endogeneity in discrete choice models and their treatment in the foods and beverages literature. Section 3 provides details of the survey, the sample of participants and the models used. Results are presented on section 4 and discussed in section 5.

2 Literature review

2.1 Double effect of price

In traditional economic theory, price is expected to have a negative effect on the purchase probability due to consumers' budget constraints; however, under some circumstances a positive effect may also exist. Scitovsky (1945) proposes that higher prices can be attractive if consumers assume price to be a cue for quality (i.e. they assume that price and quality are positively correlated), a rational assumption in perfect markets. Leavitt (1954) did one of the first experimental measurements of this phenomenon, discovering a tendency to choose the most expensive product when there were no other cues for quality, especially on product categories with heterogeneous levels of quality (i.e. vertical differentiation).

Later studies confirmed the association between price and quality, and therefore the positive effect of higher prices on choice probability. Rao & Monroe (1989) showed that the price-quality association grew stronger as the price difference between alternatives increased, through a meta-analysis. Caves & Greene (1996) found a positive correlation between price and expert's quality ratings in 200 products, while controlling for other variables. They also found that the magnitude of the price-quality correlation depended on the product category and its vertical differentiation. Dodds et al. (1991) proposed and estimated a model where price positively influences perceived quality, and negatively influences willingness to buy, while controlling for brand and store information in the case of calculators and stereo headset players.

Another possible explanation for the positive effect of price on purchase probability is what Lichtenstein et al. (1993) call prestige sensitivity, i.e., a "favourable perception of the price cue based on feelings of prominence and status that higher prices signal to other people about the purchaser”. This concept has been employed mainly in the area of fashion, and found to be strongly related with brand perception (Deeter-Schmelz et al. 2000), as other people see brands, not prices. This phenomenon is also known as Veblen Effect (Veblen 1899/1994), and is directly related with the status provided by the consumption, and only indirectly related with price. Bagwell & Bernheim (1996) claim that “… in a theory of conspicuous consumption that is faithful to Veblen’s analysis, utility should be defined over consumption and status, rather than over consumption and prices”. Therefore, this effect could be controlled for, to a reasonable degree, by including brand in the analysis.

The positive effect of price on perceived quality has also been studied in the case of wines. Plassman et al. (2007) showed that higher prices can positively influence markers of pleasure in the brain activity, even though the wine itself remains unchanged. Aqueveque (2006; 2008) found a negative effect of price on perceived risk and a positive effect on perceived quality, though this effect tended to disappear when experts’ ratings were present and the consumption occasion did not involve other people. Lewis & Zalan (2014) showed that higher prices increased both reported enjoyment and willingness to pay among wine consumers.

2.2 Endogeneity in discrete choice models and some ways to deal with it

From an econometric perspective, the double effect of price generates an endogeneity problem. In this sub-section we present a simple framework to understand how endogeneity is caused by the price – quality association, and review some alternatives to deal with. The different approaches to deal with endogeneity are discussed and evaluated based on their applicability to the problem at hand, that is, a stated choice experiment where the main source of endogeneity is the price – quality association.

Endogeneity occurs when an explanatory variable is correlated with the error term of the model. This can be due to many reasons: omission of an explanatory variable correlated with an included variable, measurement errors in explanatory variables, simultaneous determination of both the dependent and one or more of the explanatory variables, self-selection bias, among others (Guevara 2015). Endogeneity is a serious problem as it renders the estimated parameters inconsistent (see Wooldridge 2002, section 15.7.2 for a proof on binary choice models).

One important source of endogeneity in the case of price's double effect is the omission of perceived quality as an explanatory variable. The omission of other unobservable attributes correlated with price can also play a role in the endogeneity problem (Guevara & Ben-Akiva 2012); however, if these attributes are relevant, they should also be correlated with perceived quality. Simultaneous determination is likely not a severe problem at the microscopic scale, because price is exogenous for each individual, as s/he does not influence price.

More formally, consider the following true model for the utility U of individual n, for alternative j on choice scenario t.

$$U_{njt} = X_{njt}\beta_X + Y_{njt}\beta_Y + \varepsilon_{njt}$$

where X_{njt} and Y_{njt} are attributes of alternative j, ε_{njt} is an independent identically distributed error among alternatives, scenarios and individuals, and β_X and β_Y are parameters to be estimated. Now suppose the modeller does not observe Y_{njt}, therefore she estimates the following model.

$$U_{njt} = X_{njt}\beta_X + \eta_{njt}$$

where $\eta_{njt} = Y_{njt}\beta_Y + \varepsilon_{njt}$. If X and Y are correlated, then so are η_{njt} and X, introducing endogeneity in the model and therefore rendering the estimated $\hat{\beta}_X$ inconsistent. In our particular case, if we consider X to be a vector of attributes including price, and Y to be perceived quality, then the price-quality association would induce correlation between X and Y, generating an endogeneity problem. Then we would say that the explanatory variable price is endogenous.

For discrete choice models, the most popular five ways to correct for endogeneity are the BLP method proposed by Berry et al. (1995), the use of proxies, the control function approach (CFA), the multiple indicators solution (MIS) and the use of latent variables (Guevara 2015).

The BLP method requires market level data in the form of market shares for several different markets. This data is used to capture the endogeneity in constants for each market. This data requirement makes the method unsuitable for models estimated only with consumer-level information, such as our case study.

The Proxy approach consists in including proxies of the unobserved variable in the utility function. A proxy must satisfy two requirements: (i) it must be independent of the choice model's error term and (ii) the difference between the proxy and the unobserved variable should be independent of all other explanatory variables. Both requirements can be fulfilled if the proxy is exogenous to the choice, it is measured with no error, and it is the cause of the unobserved variable (i.e. it is both exogenous to the unobserved variable and it correlates with it). Therefore, the main difficulty of this method is to find an appropriate proxy. For example, a proper proxy for the comfort experienced by a new passenger on a train is the density of passengers in the train before s/he boards.

A proxy for perceived quality should be able to explain it while not being correlated with price. An objective measurement of quality should be a good proxy for perceived quality only if the objective quality does not correlate with price; however, it is not clear that such a measurement exists. In the case of wine, expert ratings may not be appropriate either as their ability to measure objective quality has been seriously questioned (Lawless 1984, Hodgson 2009), as well as their relationship with consumer's quality perception (Lattey et al. 2009, Gokcekus & Nottebaum 2011, D'Alessandro & Pecotich 2013, Hopfer & Heymann 2014). And even though consumers do use experts' ratings as a proxy for quality when available in hypothetical situations (Aqueveque 2006, Mastrobuoni et al. 2014), several studies have indicated that consumers are not really aware of them in real conditions (Channey 2000, Johnson & Bruwer 2004, Atkin & Thach 2012). Furthermore, it is likely that if an objective measurement of quality exists, it would correlate with price due to production costs.

As our experiment used fictional wines, no real experts' quality ratings were available, neither did we include fictional ratings as an extra attribute because Chinese consumers do not seem to consider experts' ratings (at least in the form of prizes or written recommendation) among the most relevant cues for quality (Goodman 2009).

In the particular case of wine, the weather during growth and harvest could be used as a proxy for quality, as wine quality is expected to depend largely on them. But the weather only influences the sensory (or intrinsic) quality of wine, and therefore it would not reflect the expected quality before purchase, when the consumer has not tasted the wine yet. Also, the weather is not available for fictional wines in a SP context.

Another method to correct for endogeneity is the Control Function (CF) approach (Villas-Boas & Winer 1999, Petrin & Train 2010), which is analogous to the Instrumental Variables approach on linear models (Wooldridge 2002, chapter 5). The CF approach requires the modeller to identify instrumental variables for the endogenous explanatory variable (in our case: price). The instrumental variables must fulfil two requirements: (i) correlate with the endogenous explanatory variable and (ii) be independent of the error terms. The estimation procedure has two stages: first, the endogenous variables are regressed on the instrumental and other exogenous explanatory variables, and then the residuals of this regression are included in the choice utility along with the endogenous and exogenous explanatory variables. This way the new extended model is consistently estimated. Estimation can also be performed in a single step using Full Information Maximum Likelihood (Villas-Boas & Winer 1999, Train 2009 section 13.5, Guevara 2015). The main difficulty with this procedure is finding adequate instrumental variables.

Production costs are useful instruments (Villas-Boas & Winer 1999), but they are hardly available for real products, and do not exist in the case of fictional ones. The price of similar alternatives can also be used (Guevara & Ben-Akiva 2006), but once again, they do not exist in the context of hypothetical choices. And even though it is possible to design a stated choice experiment where the weather, the price of similar alternatives or other instruments are fictionally developed, its implementation would be convoluted and probably unrealistic. In summary, CF is hardly applicable on stated choice datasets, such as ours.

A Multiple Indicator Solution (MIS) is yet another way to correct for endogeneity in discrete choice models (Guevara & Polanco 2016). This approach is a mixture between the use of a proxy and a control function. The method requires two indicators of the omitted variable. Indicators are only required to correlate with the unobserved variable, and not to be exogenous to the choice. The idea is to include the first indicator in the utility function, using it as a proxy for the omitted variable and therefore transferring the endogeneity from the original endogenous variable to the indicator. Then, the second indicator serves as an instrument to correct the endogeneity of the first indicator, using the CF approach. The second indicator is a valid instrument for the first indicator, as both are correlated because both are explained by the omitted variable; it is also uncorrelated with both the original error tem of the utility function and the first indicator's error term, under the assumption that both indicators are redundant in the structural equation of utility if the omitted variable is included (Guevara 2015).

In the case of price-quality associations, one would only require two indicators of quality to apply the MIS approach. Unlike proxies, indicators can be noisy and they do not need to have a causal relation with the omitted variable, but quite the contrary, it is the omitted variable that causes and explains both indicators. Therefore, simple quality ratings from the consumers or experts could be used. The former would be preferable though, as they measure expected quality directly. When applied to solve the endogeneity problem due to the price-quality association, the MIS approach could effectively provide consistent estimates for both the positive and negative effects of price, through the first indicator and price coefficients, respectively. However, two reliable and independent (given the omitted variable) indicators must be available. As we only had a single quality indicator in our dataset we could not apply the MIS approach.

Finally, the Latent Variable approach to endogeneity correction consists in explicitly modelling the omitted variable as a latent variable. To do this, two pieces of information are required: (i) at least one indicator of the omitted (latent) variable, and (ii) one or more exogenous explanatory variables for the omitted variable. This method requires strong distributional assumptions, as the structural relation between the omitted variable, its explanatory variables, and the choice is explicitly (and parametrically) formulated. However, its data requirements (at least in the context of this study) are easier to fulfil, as it does not require hard-to-find proxies or instrumental variables, and it only requires one quality indicator.

The Latent Variable approach is the only method that provides a consistent behavioural model in the context of price-quality associations. Therefore, it allows to clearly separate the positive and negative effects of price, and to separately model the perception of quality, and the willingness to buy. This is particularly useful when consumers cannot perceive the quality of a product and therefore must infer it from observable attributes.

2.3 Endogeneity in the foods and beverages literature

In the foods and beverage choice literature, endogeneity has been considered mainly in the context of price's simultaneous determination due to supply and demand equilibration. Using a panel of scanner data at the household level and discrete choice models, Villas-Boas & Winer (1999) applied the CF approach to test and control for endogeneity in the yoghurt and ketchup market. They found evidence of endogeneity, which they explained on the simultaneous determination of price. Also using household data, but analysing it through a discrete-continuous model, Richard & Padilla (2009) analysed the impact of promotions in fast food consumption. They also found evidence of price endogeneity using a CF approach, which they again explained on the simultaneous determination of price. O'Neill et al. (2014) recognized that their analysis of food choices could be affected by endogeneity, but did not explicitly control for it.

In the wine choice literature, endogeneity has been explicitly controlled for mostly in the context of aggregate demand models. Cuellar & Huffman (2008) used aggregate data to estimate the price elasticity using linear models with grape prices as instrumental variables to correct for endogeneity. Stasi et al. (2011) used Italian market aggregate data and simultaneous equation modelling to measure the impact of geographical indicators, while correcting for endogeneity using several instrumental variables, such as lagged prices and seasonal dummies. Michis & Markidou (2013) used aggregate data from Cyprus and a system of simultaneous equations to identify the determinants of wine price, and took market concentration and competitors' prices as instrumental variables to correct for price endogeneity.

To the best of our knowledge, only two papers deal with the endogeneity problem when modelling wine demand at the individual level using stated choice experiments. In particular, although Appleby et al. (2012) do not mention endogeneity explicitly, their approach can be seen as using Wine Spectator's ratings as a proxy for quality, yielding reasonable results. However, as discussed in the previous sub-section, the use of experts' ratings as proxies for quality is highly questionable.

Mastrobuoni et al. (2014) used a two-stage process (somewhat similar to our approach) to separate the positive and negative effects of price in a SP experiment. However, they mixed the Proxy and Latent Variable approaches to correct for endogeneity. Their experiment appears to yield reasonable results, but the method is not applicable to situations without tasting, it resorts to experts' ratings as a proxy for quality and uses a sequential estimation process, which could lead to new endogeneity problems as the deterministic part of the first stage logit's utility is a noisy (and therefore endogenous) proxy for quality.

In this paper, we use the latent variable approach to correct for endogeneity. Our particular application is a stated wine choice experiment where consumers provided a single quality indicator per alternative, additionally to their choices. Due to the way our data was collected, we are not able to offer any comparison of the Latent variable approach with other methods. BLP requires market-level data, which does not exist in a SP experiment. The proxy method in a SP setting implies providing an expert ranking for consumers to use as a proxy for quality, but as consumers do not seek this information in real settings we did not include it the experiment. The CFA is not applicable as there are no available instruments in a SP setting, and we only have one quality indicator in our dataset, therefore the MIS approach cannot be applied either (as it requires at least two indicators).

3 Materials and methods

3.1 Survey design

In association with a private Chilean Vineyard, we designed a computer-based Stated Choice (SC) experiment (Rose & Bliemer 2009; Rose et al. 2008; Ortúzar & Willumsen 2011, section 3.4) that was applied to a sample of Chinese wine consumers, including experts, students and regular consumers. Respondents were presented with six choice scenarios (also called choice exercises) with three alternatives each (Caussade et al. 2005), plus a non-purchase alternative if they rather wished to opt out.

We considered four attributes in the SC experiment (Table 1): label design (6 levels), grape variety (3 levels), name and “story” of the brand (3 levels) and price (3 pivoted levels). In addition, in every choice scenario we also stated one out of two consuming occasions (formal and informal). Attributes were selected after a literature review (see, for example Lockshin & Corsi 2012), focus groups, previous experience with Chilean consumers (Palma et al. 2013), and advice from experts on the Chinese wine market. The “story” attribute, in particular, was proposed by these experts, and included both the name of the wine and a short statement describing its origin (the name and the statement were not shuffled, instead they were always paired in the same way). The objective was to provide a narrative for the product, for example, one story presented the wine as an old family tradition, while another presented it as the last innovation of a young entrepreneur.

Table 1 - Attributes and their levels (levels’ order have been altered)

	Label	Grape variety	Story	Price	Consuming occasion
0	Label 0	Red Blend	Hacienda	Informal low	Informal:
1	Label 1	Shiraz	Don Juan	Informal mean	“an informal dinner
2	Label 2	Cabernet Sauvignon	Union	Informal high	with friends"
3	Label 3			Formal low	
4	Label 4			Formal mean	Formal:
5	Label 5			Formal high	"a formal dinner”

Before facing the SC scenarios, participants provided the minimum and maximum amounts of money they would be willing to pay for a bottle of wine on a formal and on an informal occasion. The phrasing of the question was: “Imagine that you need to buy a wine for the following occasions. How much would you be willing to spend? Please indicate a minimum and a maximum amount of money you would be prepared to pay for each occasion”. Price levels of the SC experiment were pivoted based on these values at the individual level, i.e. each participant saw prices based on his/her own reported buying range for each occasion. This allowed us to make sure that participants did not see alternatives with prices outside their regular buying range, therefore avoiding them ruling out alternatives considered either too cheap or too expensive.

As participants provided different buying ranges for formal and informal occasions, six different price levels were calculated for each participant: informal low (the minimum price the participant would pay for a wine to drink at an informal occasion), informal high (the maximum price in the same case as above), informal mean (the midpoint between the previous two) and three more levels analogous to the previous ones, but for formal occasions. The occasion associated with each scenario determined which set of prices (formal or informal prices) were used.

Consuming occasion only varied between scenarios. Introducing more than one consuming occasion per scenario would have made the experiment unrealistic, as individuals seem to choose differently based on the consuming occasion (Dubow 1992; Quester & Smart 1998; Martínez-Carrasco et al. 2006, Jaeger & Rose 2008).

We generated a D-efficient balanced design assuming a simple MNL model using N-gene (http://choice-metrics.com/). We used null priors for the experts’ design, who answered the experiment first, and then used the experts’ results as priors for the design for the rest of participants. The experimental design had twelve choice scenarios divided into two blocks of six choice scenarios each, to which respondents were assigned randomly. The presentation orders of both scenarios and alternatives were randomized.

Before choosing the wine they would buy in each scenario, respondents had to provide their level of agreement with the phrase "I believe this wine is excellent" for each alternative presented, using a 5-point Likert scale. This information was used as an indicator of quality for each alternative. Then, respondents were told about the consuming occasion, and asked to make their choices (including the opt-out option). Figure 1 shows an example of a choice scenario.

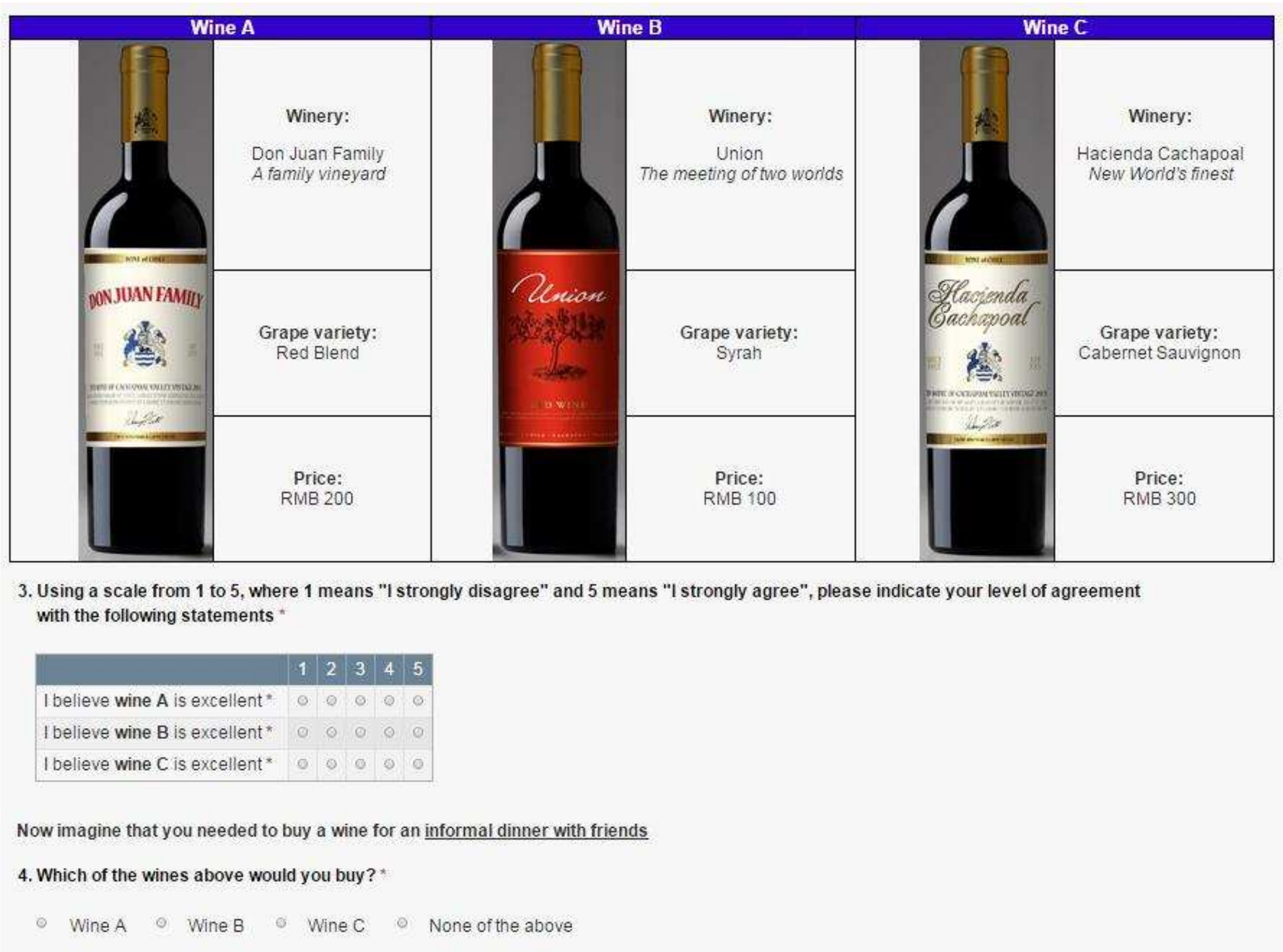

Figure 1 – Example of choice scenario with quality indicator for each alternative (labels have been altered)

Before facing the choice scenarios, participants also had to rate each of the considered grape varieties using a 5-points Likert scale. Based on these ratings, we built a grape variety ranking for each participant excluding ties; that is, when a participant gave the same rating for two or three grape varieties, we excluded them from the ranking.

These rankings were exploded (Chapman & Staelin 1982, Ortúzar & Willumsen, 2011, section 8.7.2.3), creating "grape variety choices" in our dataset generating up to two new observations per respondent. As an example, let us consider a participant whose ranking was: (1st) Cabernet Sauvignon, (2nd) Shiraz and (3rd) Red Blend. In this case, the first "grape variety choice" would be between three wines with the same attributes, except grape variety: wine A would be a Cabernet Sauvignon, wine B would be a Shiraz and wine C would be a Red Blend, and the participant would choose wine A. The second "grape variety choice" would be between wines B and C only (wine A would not be available), and the participant would choose wine B. For participants whose rankings where shorter (due to ties), only one or none "grape variety choice" were generated. When modelling, we multiplied the utility of the "grape variety choices" by a scale factor, so we could control for differences in variance among the traditional choices and the "grape variety choices". However, this scale parameter turned out to be not significant in the ML model, so we removed it from its reported version.

3.2 Sample

A total of 180 participants answered the survey; however, after data cleaning only 168 responses were considered valid. The main reason to eliminate respondents was unreasonable price ranges, that were either too low (maximum was less than 1.6 USD) or too high (minimum was more than 10% of their monthly income).

The sample was divided into three groups: experts (21), regular consumers (81) and students (66). We introduced this classification, as it would help the private vineyard developing a more detailed strategy aimed at connoisseurs (experts), regular consumers and millennials (students). Most experts worked in the wine industry, mainly in marketing or trade departments, while others were wine critics. Regular consumers were mostly professionals and office clerks from different industries, including some scholars. All students were enrolled in some of the wine-related courses taught at the College of Horticulture at CAU.

We used a convenience sample; therefore, there is no guarantee that it represents the average Chinese wine consumer, nor any particular segment of the Chinese wine market. Experts and consumers received a small monetary incentive for their participation and performed the experiment in a laboratory, in a controlled environment. Students, on the other hand, were invited to participate in the experiment during classes, and answered the survey later using their own computers in an uncontrolled environment. Most students (97%) were under 30 years old; more details about the sample are shown in Table 2.

Given the age and profile of the students, their answers for the formal occasion were removed from the analysis, as their self-reported price ranges tended to be unreasonable. On average, students set a minimum price of 17% of their income and a maximum of 129% for formal occasions; instead, experts and consumers set an average price range for the same occasions between 4% and 11% of their income. Therefore, at the end 810 wine choices were collected (126 from experts, 486 from consumers and 198 from students).

Table 2 - Sample description

	Experts	Consumers	Students	Total
Respondents	21	81	66	168
Gender				
Female	11	43	50	104
Male	10	38	16	64
Age				
18 - 24	0	0	1	1
25 - 30	2	20	64	86
31 - 35	10	35	1	46
36 - 40	4	10	0	14
41 - 50	2	8	0	10
51 - 60	2	7	0	9
>60	1	1	0	2
Maximum level of education attained				
12th grade or less	1	2	0	3
Graduated high school	1	2	2	5
Some college, no degree	0	7	62	69
Associate degree	3	2	0	5
Bachelor's degree	3	61	2	66
Post-graduate degree	13	7	0	20
People in household				
Unknown	1	0	0	1
1	1	3	0	4
2	4	12	1	17
3	9	37	48	94
>3	6	29	17	52
Household monthly income (USD)				
<800	0	5	12	17
<1600	4	23	30	57
<2400	3	19	15	37
<3200	6	6	4	16
<4000	3	13	4	20
<4800	2	1	0	3
<5600	1	6	0	7
>5600	2	8	1	11
Average buying price range (USD)				
Min Informal	24	18	17	19
Max Informal	83	69	64	69
Min Formal	51	79	57	66
Max Formal	159	221	345	262

All participants rated the three grape varieties included in the experiment, giving rise to a personal ranking, which was exploded providing up to two additional choices per participant (as mentioned above, ties were excluded). Experts provided 27, consumers 111 and students 59 of these choices. Considering all choices (both wine and grape variety choices), 1007 observations were used for estimation.

3.3 Modelling

Two models were estimated with the available data: a traditional Mixed Logit (ML) model with random coefficients without considering an endogeneity correction (McFadden & Train 2000, Train 2009, chapter 6) and a Hybrid Choice (HC) model using random coefficients and the latent variable approach to correct for endogeneity (Ortúzar and Willumsen, 2011, section 8.4.3; Bolduc & Alvarez-Daziano 2010, Guevara 2015). Comparing both models allows determining how effective the latter is in dealing with endogeneity.

In the ML model, all attributes explain choice by entering the utility function directly (Figure 2).

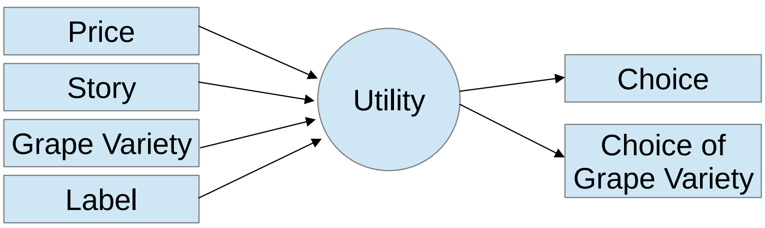

Figure 2 - ML model structure (for each alternative)

The deterministic utilities of the alternatives, their full utilities and the model's likelihood for one individual are shown in equations (1), (2) and (3), respectively.

$$V_{jtn} = X'_{jtn}\beta_{Xn} + \left(\beta_{price} + \beta_{price}^{expert} expert_n + \beta_{price}^{student} student_n\right) price_{jtn} \quad (1)$$

$$U_{jtn} = V_{jtn} + \epsilon_{jtn} \quad (2)$$

$$L(\vec{\iota}_n) = \int \left(\prod_t \frac{e^{V_{itn}}}{\sum_j e^{V_{jtn}}} \right) \prod_g \frac{e^{V_{ign}}}{\sum_j e^{V_{jgn}}} \varphi\left(\beta_{Xn} \middle| \mu_{\beta_X}, \Sigma_{\beta_X}\right) d\beta_{Xn} \quad (3)$$

where V_{jtn} is the deterministic part of the choice utility for alternative j in scenario t for respondent n. X'_{jtn} is a row vector of alternative j's attributes (excluding price); β_{Xn} is a vector of random parameters representing respondent n's preferences for attributes other than price; $expert_n$ and $student_n$ are dummies which take the value 1 if respondent n is an expert or a student respectively, and 0 otherwise; $price_{jtn}$ is the alternative's price. U_{jtn} is the alternative's full random utility, and ϵ_{jtn} is an iid Extreme Value type 1 random error that gives the choice probability its logit form. $\vec{\iota}_n$ is the vector of choices made by respondent n; V_{itn} is the deterministic part of the utility of the chosen alternative i in choice scenario t by respondent n; V_{ign} is the deterministic part of the utility of the chosen alternative i, by respondent n, on the "grape variety choice" g; $\varphi\left(\beta_{Xn} \middle| \mu_{\beta_X}, \Sigma_{\beta_X}\right)$ is the multivariate normal density function of all random coefficients included in the β_{Xn} parameter vector, with vector μ_{β_X} as mean and the diagonal matrix Σ_{β_X} as variance. Finally, $\mu_{\beta_X}, \Sigma_{\beta_X}, \beta_{price}, \beta_{price}^{expert}$ and $\beta_{price}^{student}$ are parameters to be estimated.

No additional error components were included to model the pseudo-panel effect. We did test a specification with error components, as proposed by Daly & Hess (2010), but the error components' standard deviations were non-significant, so we removed them from the final specification. However, as the randomness in β_{Xn} is between and not within participants (Revelt & Train 1998), correlation between the observations of each respondent is present, even though some confounding effects could occur (Daly & Hess 2010). Consuming occasion was not considered in the final specifications either, as we tested several ways to interact it with the different attributes, and none was significant.

Unlike the ML model, the HC model explains the choices made by consumers as a trade-off between an alternative's expected quality and its price. Each alternative's expected quality is modelled as a latent variable, which is explained by its attributes including price (Figure 3).

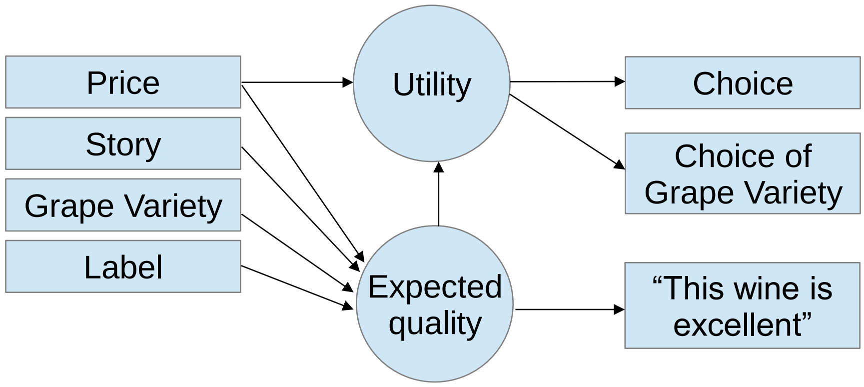

Figure 3 - HC model structure (for each alternative)

Price was included as an explanatory variable both in the structural equation of the expected quality and in the choice utility in the HC model. Its first coefficient was expected to capture the positive effect of price as a cue for quality, while the second intended to measure the negative effect of price owing to the consumers' budget restrictions. Therefore, the price coefficient was expected to be positive in the expected quality's structural equation and negative in the choice utility function.

In the HC specification, a Multinomial Logit (MNL) model was used to link the utility with choices, and an ordered logit model (Greene & Hensher 2010) to link expected quality and level of agreement with the phrase "This wine is excellent". The expected quality's structural equation (4), its measurement equation (5), the ordered logit probability function (6) and the deterministic part of the choice utility (7) are as follows:

$$EQ_{jtn} = X'_{jtn}\alpha_{Xn} + \left(\alpha_{price} + \alpha_{price}^{expert}expert_n + \alpha_{price}^{student}student_n\right)price_{jtn} + \eta_{jn} + \omega_t \quad (4)$$

$$measurement_{jtn} = \lambda EQ_{jtn} + \varepsilon_{jtn} \quad (5)$$

$$P\left(IQ_{jtn} = l\right) = \frac{1}{1 + e^{\lambda EQ_{jtn} - \delta_l}} - \frac{1}{1 + e^{\lambda EQ_{jtn} - \delta_{l-1}}} \quad (6)$$

$$V_{jtn} = \beta_{EQ} EQ_{jtn} + \left(\beta_{price} + \beta_{price}^{expert} expert_n + \beta_{price}^{student} student_n\right) price_{jtn}$$
$$W_{jtn} = (1 + \mu_{student} student_n)\left(1 + \mu_{gVarRnk} gVarRnk_{jtn}\right) V_{jtn} \tag{7}$$

where EQ_{jtn} is participant n's expected quality of alternative j in scenario t; X'_{jtn} is a row vector of alternative's attributes (except for price); α_{Xn} is a vector of normally distributed random parameters representing participant n's preferences, with vectors μ_{α_X} as mean and the diagonal matrix Σ_{α_X} as variance. η_{jn} is a normally distributed error component with mean 0 and standard deviation fixed to 1 (this is a requirement for identification in the structural equation model). These error components capture the expected quality's determinants that are not observed by the modeller (Bahamonde-Birke et al., 2015) and correlate observations of the same respondent by being invariant across choice scenarios (Daly & Hess 2010). ω_t is an iid normal error component with mean zero and variance σ_ω^2 to be estimated, correlating the expected quality of all wines observed on the same choice situation. $measurement_{jtn}$ is the ordered logit's latent variable depending on expected quality and its iid Extreme Value type 1 error component ε_{jtn}, that gives the measurement its ordered logit form. $P\left(IQ_{jtn} = l\right)$, is the ordered logit probability of quality indicator IQ_{jtn} (level of agreement with the phrase "This wine is excellent") being equal to l and V_{jtn} is the deterministic part of the choice utility. The dummies $expert_n$ and $student_n$ take the value 1 if respondent n is an expert or student, respectively, and 0 otherwise. W_{jtn} is the deterministic part of the utility scaled by factors $\mu_{student}$ and $\mu_{gVarRnk}$ when the observation belongs to a student or is a "grape variety choice". The dummy variable gVarRnk takes the value 1 if the observation is a "grape variety observation" and 0 otherwise. Scale factors analogous to these ones were tested in the ML model, but were not significant, therefore they were removed from the final model. Finally, μ_{α_X}, Σ_{α_X}, α_{price}, α_{price}^{expert}, $\alpha_{price}^{consumer}$,λ, δ_l, β_{EQ}, β_{price}, β_{price}^{expert}, $\beta_{price}^{student}$,

$\mu_{student}$ and $\mu_{gVarRnk}$ are parameters to be estimated. Note that $\delta_0$and δ_5 were set to $-\infty$ and $+\infty$, respectively, for identification purposes. Finally, just as in the ML model, no error components or interactions with consuming occasion were included in the utility, as they both were not significant.

The likelihood function of the HC model is presented in equation (8).

$$L(\vec{\imath},\vec{l}) = \int_{\overrightarrow{\eta_n},\omega_t,\alpha_{Xn}} \left[\prod_t \left(\prod_j P(IQ_{jtn} = l_{jtn})\right)\frac{e^{W_{itn}}}{\sum_j e^{W_{jtn}}}\right]\prod_g \frac{e^{W_{ign}}}{\sum_j e^{W_{jgn}}}\varphi(\overrightarrow{\eta_n}|0,1)\varphi(\omega_t|0,\sigma_\omega^2)\varphi\left(\alpha_{Xn}|\mu_{\alpha_X},\Sigma_{\alpha_X}\right)d\overrightarrow{\eta_n}d\omega_t d\alpha_{Xn} \qquad (8)$$

where $\vec{\imath}$ represent the vector of choices and $\vec{l}$ the vector of quality indicators; W_{itn} is the deterministic part of the utility of the chosen alternative i in choice scenario t by respondent n; W_{ign} is the deterministic part of the utility of the chosen alternative i by respondent n on the “grape variety choice” g; $\overrightarrow{\eta_n}$ is the vector containing all three η_{jn} associated with the expected quality of each of the three alternatives; $\varphi(\overrightarrow{\eta_n}|0,I)$ is the multivariate normal density function for the vector of error components associated with expected quality, with mean a vector of zeros and a 3x3 identity matrix for variance. $\varphi(\omega_t|0,\sigma_\omega^2)$ is the normal density function with mean 0 and variance σ_ω^2. Finally, $\varphi\left(\alpha_{Xn}|\mu_{\alpha_X},\Sigma_{\alpha_X}\right)$ is the multivariate normal density function with the vector μ_{α_X} as mean, and the diagonal matrix Σ_{α_X} as variance.

Both models were estimated using the Python version of Biogeme (Bierlaire 2003). Monte Carlo techniques were used to estimate the integrals on the likelihood functions, as these do not have a closed analytical form. This consists in randomly drawing a large number of points from $\varphi\left(\beta_{Xn}|\mu_{\beta_X},\Sigma_{\beta_X}\right)$ (in the case of the ML model) or $\varphi(\overrightarrow{\eta_n})$, $\varphi(\omega_t|0,\sigma_\omega^2)$ and $\varphi\left(\alpha_{Xn}|\mu_{\alpha_X},\Sigma_{\alpha_X}\right)$ (in the case of the HC model) and then evaluating $\left(\prod_t \frac{e^{V_{itn}}}{\sum_j e^{V_{jtn}}}\right)\prod_g \frac{e^{V_{ign}}}{\sum_j e^{V_{jgn}}}$ (in the case of the ML) or $\left[\prod_t\left(\prod_j P\left(IQ_{jtn} = l_{jtn}\right)\right)\frac{e^{W_{itn}}}{\sum_j e^{W_{jtn}}}\right]\prod_g \frac{e^{W_{ign}}}{\sum_j e^{W_{jgn}}}$ (in the case of the HC), for each of these points. The average of all these evaluations is a consistent estimator of the integral's value (Train 2009, chapters 9 and 10).

4 Results

Table 3 and 4 show the ML and HC models' estimated coefficients as well as their goodness of fit measures. All reported t-test are robust (i.e. they were calculated using the "sandwich estimator" clustering by respondent). Both models were estimated using 1000 Modified Latin Hypercube Sampling draws (Hess et al. 2006). Even though we tested interactions with consuming occasion in both models, none turned out to be significant, so we removed these from the final specifications reported in this document. The same holds true for the scale factors for "grape variety choices" and participant classes (experts and students) in the ML model. In the HC model, instead, the "grape variety choices" and the students' scale factors were significant and, therefore, kept in the model. We also kept the non-significant main effects of attributes to facilitate comparisons between models, and to avoid endogeneity problems due to the omission of relevant attributes.

Results indicate that the HC model works as expected. The Price coefficients have the expected signs (i.e. positive in the expected quality's structural equation and negative in the choice utility). The quality indicator (level of agreement with the phrase "this wine is excellent") strongly correlates with expected quality, as reflected by a positive and significant parameter λ. Finally, expected quality has a positive and significant effect on choice utility.

In the ML model none of the Price parameters are significant. Instead, in the HC model all Price parameters have the expected sign: positive in the expected quality structural equation, and negative in the choice utility; and most of them are significant using a one-tail t-test (t-test critical value of 1.645 at 95% significance). In particular, only students exhibit a significant use of price as a cue for quality, while all classes exhibit a significantly negative effect of price, though with different intensities: students are the most sensitive, followed by experts and regular consumers.

Table 3 - Coefficients and goodness of fit measures for the ML model (robust t-test are reported)

		Main effect		Standard deviation	
		Value	t-ratio	Value	t-ratio
Choice utility	Grape variety 1	0.000	0.00	0.646	4.71
	Grape variety 2	-0.036	-0.32	0.445	2.52
	Label 1	-0.338	-2.20	0.586	2.40
	Label 2	-0.008	-0.06	0.069	0.84
	Label 3	0.045	0.23	0.990	3.09
	Label 4	-0.375	-2.05	0.892	2.38
	Label 5	0.171	1.06	0.317	0.73
	Story 1	-0.101	-0.85	0.338	1.40
	Story 2	0.145	1.32	0.267	0.90
	Price	-0.001	-0.97		
	Price x experts	-0.005	-1.57		
	Price x students	-0.006	-1.36		
	Center position	0.212	2.41		
	No purchase	-2.010	-7.30		
Goodness of fit indicators	Number of parameters				23
	Number of observations (respondents)				1007 (168)
	Loglikelihood				-1170.974
	ρ^2				0.114
	Adjusted ρ^2				0.096
	Corrected ρ^2				0.022
	First Preference Recovery (FPR)				0.331
	FPR Expected value				0.387
	Chance recovery (CR)				0.274

Concerning attributes other than price, even though there are similarities between both models, results are not always consistent between them. The main effects of Grape variety are zero in both models, meaning that –on average- there is no particular grape variety preferred over others. However, as all standard deviations of Grape variety are statistically significant, preferences for grape varieties are highly heterogeneous among participants. Both models agree on labels 1 and 4 being –on average- less preferred than the base label, though with significant variability in the population. Both models also agree on labels 2, 3 and 5 to be –on average- equivalent to the base label. But both models disagree on how preferences for labels 2, 3 and 5 distribute among the population, with the ML model implying that only preferences for label 3 have significant variability, while the HC model suggests that the preferences for all three labels do. Finally, the effect of Story is also different in both models: while the ML results imply that all stories are equivalent, the HC model recognizes story 1 to be the least preferred on average, and preferences for story 1 and 2 have significant variability among the population.

Table 4 - Coefficients and goodness of fit measures of the HC model (robust t-tests are reported)

		Main effect		Standard deviation	
		Value	t-ratio	Value	t-ratio
Expected quality	Grape variety 1	0.086	0.420	1.250	2.520
	Grape variety 2	-0.247	-1.140	1.190	3.620
	Label 1	-0.617	-2.600	0.946	3.560
	Label 2	-0.070	-0.370	0.771	2.710
	Label 3	-0.411	-1.400	1.450	2.700
	Label 4	-0.863	-2.480	1.410	4.280
	Label 5	0.069	0.350	0.914	2.880
	Story 1	-0.548	-2.710	0.794	2.030
	Story 2	-0.188	-1.290	1.050	3.780
	Price	0.001	1.540		
	Price x experts	0.003	1.360		
	Price x students	0.009	2.130		
	σ_{ω}	0.621	2.020		
	λ	0.854	4.240		
	Threshold 1	-5.100	-15.100		
	Threshold 2	-3.000	-11.770		
	Threshold 3	-0.263	-1.130		
	Threshold 4	2.320	8.800		
Choice utility	Expected quality	0.640	4.490		
	Price	-0.002	-2.460		
	Price x experts	-0.008	-2.650		
	Price x students	-0.028	-1.880		
	Center position	0.308	2.880		
	No purchase	-2.400	-9.410		
	$\mu_{gVarRnk}$	-0.684	-3.670		
	$\mu_{student}$	-0.608	-4.680		
Goodness of fit indicators	Number of parameters				35
	Number of observations (respondents)				1007 (168)
			with indicators		without indicators
	Loglikelihood		-4184.61		-1181.42
	ρ^2		0.714		0.106
	Adjusted ρ^2		0.712		0.079
	Corrected ρ^2		0.072		0.013
	AIC		4254.6		1251.4
	BIC		8611.2		2604.9
	First Preference Recovery (FPR)		0.331		
	FPR Expected value		0.385		
	Chance recovery (CR)		0.274		

Both models indicate that most of the main effects are statistically equivalent to zero. This is probably due to preferences being highly heterogeneous among consumers, cancelling out on average. As there is no single grape variety, label or story clearly superior to the others, preferences are only a matter of taste. This reflects on the relatively high values of the standard deviations estimated for most parameters, a phenomenon better captured by the HC model than by the ML model. This variability seems to be inherent to all consumers, and not an artefact arising from mixing different classes of them (i.e. experts, regular consumers and students). We tested removing students -probably the most eccentric class- and found no evidence of a decrease in preference variability, nor an increase of t-tests on their average effects.

We tested the effect of alternatives' position on choice by including constants for the left and central alternative (see Figure 1) in the utility function. Results were consistent in both models, with only the central position achieving significance. We therefore kept a constant for the central alternative in the final model, effectively controlling for presentation order bias.

Both scale parameters are negative, meaning that the grape variety choices, as well as all choices by students, have more variability than those by experts and regular consumers (see equation 7). This is to be expected, as preferences for grape variety are highly heterogeneous and students are the less knowledgeable class of respondents. Scale factors for regular choices and consumers are normalized to one, i.e. $(1+\mu_{choice}) = (1+\mu_{consumer}) = 1$ for identification purposes. We tested a scale factor μ_{expert} for experts, but $(1+\mu_{expert})$ it was not significantly different from one.

The goodness of fit indices of both models must be compared with care, as their structures are different: while the ML model takes into consideration only the consumers' choices, the HC model also includes the expected quality indicators. Therefore, only the choice part of the HC model must be taken into account when comparing fit indices (Table 4 presents goodness of fit indices differentiated for the whole HC model and its choice component). As expected, the ML model fits choices better, as all its parameters are exclusively dedicated to fit them, unlike the HC model, where the grape variety, label and story parameters must reproduce the respondents' answers for both the expected quality indicators and the choices. This extra restriction implies a difference of 10 points between their log-likelihoods, and a global loss of fit as the ρ^2, adjusted ρ^2, corrected ρ^2, and Akaike and Bayesian information criterions (AIC and BIC) point out. This loss of fit is significant ($p<0.01$) according to Horowitz (1983)'s test for non-nested models.

In principle the prediction capacity of both models could be tested in-sample and out-of-sample. The First Preference Recovery (FPR) or "percent correctly predicted" is an index of prediction accuracy, which assumes that the alternative with the highest predicted probability is chosen, and then it compares this prediction with the actual choices to determine how many times the prediction was "accurate". However, this is a poor index, as Train (2009, page 69) explains: "The researcher has only enough information to state the probability that the decision maker will choose each alternative. (…). This is quite different from saying that the alternative with the highest probability will be chosen each time." Following Gunn & Bates (1982) we present the actual FPR, its expected value and the value of Chance Recovery (CR), i.e. the prediction by chance. Results are as expected, because market shares in unlabelled experiments tend to be similar to chance recovery, as otherwise the experiment would be unbalanced towards a particular alternative. Furthermore, the FPR is expected to improve significantly if we used individual level parameters for prediction (Train 2009, chapter 11).

5 Discussion

In this study, the HC model using expected quality as a latent variable allowed us to successfully reduce price endogeneity. This reflects on the increased t-test of the price coefficients in the choice utility for all groups of participants. This improvement is caused by the separation of the positive effect of price due to its role as a cue for quality, and its negative effect due to the participants' budget restrictions. While the positive effect is captured in the structural equation of perceived quality, the negative effect is captured in the choice utility.

Comparing our results with other wine studies can only be done in general terms. Most comparable studies were not performed on the same market as ours, so price sensitivities are expected to change. However, it is possible to analyse the general behaviour of price coefficients estimated in studies both with and without endogeneity correction.

Among the studies that do not correct for endogeneity, most tend to find non-linear effects of price, such that mid-range prices provide higher utilities than lower and higher prices. This is likely due to the double effect of price: people may think that wines below some price are of low quality, therefore utility increases with price for a certain interval, but after overcoming a given price threshold, the budget restriction outweighs the price-quality association and the choice utility decreases again. Lockshin et al. (2006) do not report the coefficients of their estimated model, but plot simulations showing how market shares first increase with price, reach a peak at about US$ 11 and then decrease again after that point. Similarly, Mtimet & Albisú (2006) used a quadratic form for price finding a similar concave shape, with the peak utility at about US$ 7. Using dummies for price levels and latent classes, Remaud et al. (2008) and Mueller et al. (2010) found that some classes had this same concave behaviour.

Unlike other studies, Barreiro-Hurlé et al. (2008) and Stasi et al. (2014) obtained monotonic decreasing effects for price in the choice utility without correcting for endogeneity. Stasi et al. (2014) used customised (pivoted) prices for alternatives, varying among 90% and 140% of the average wine price in the area and obtained a negative and significant price coefficient; but it is possible that their strategy for determining alternatives' prices only allowed them to capture the decreasing part of the price-utility curve (i.e. where the budget effect overweighs the price-quality association). Something similar might had happened in the work of Barreiro-Hurlé et al. (2008), who found a negative and highly significant price coefficient using four price levels: 3, 7, 10 and 14 Euros (about 4, 9.5, 14 and 19 US$). According to Mtimet & Albisú (2006), who also studied the Spanish market, three of these levels would fall into the part of the price-utility curve where the budget effect overweighs the price-quality association. Palma et al. (2013) also found a negative coefficient for price; however, they explicitly pivoted the alternatives' prices above the participants self-reported willingness to pay for the considered occasion (from 100% to 160%).

Papers that do correct for endogeneity at an individual level yield results as expected. Appleby et al. (2012) used experts' ratings as a proxy for quality when modelling a stated purchase decision. They found a negative and significant effect of price, as well as a positive effect for the experts' ratings. However, as very few attributes were included in their study, it is possible that participants relied on the experts' ratings more than they would under more realistic conditions (Channey 2000, Johnson & Bruwer 2004, Goodman 2009, Atkin & Thach 2012).

Mastrobuoni et al. (2014) estimated both the positive and negative effects of price separately, finding a negative and significant coefficient for the budget effect of price, and a positive and significant effect for the price-quality association, though only up to 5

Euro (about US$ 7). Their approach to endogeneity correction could be considered as mixed, as they explicitly separated the modelling of both effects (i.e. used a latent variable approach), but also included experts' ratings as a proxy for quality. In their experiment, consumers tasted a set of wines, then chose their preferred alternative, and finally chose the one they would buy. With the first answer the authors modelled the perceived quality using experts' ratings (which consumers do not see) as a proxy for sensory quality. Then, they explained the (hypothetical) purchase decision as a trade-off between price and perceived quality. This approach has three main limitations. First, as it includes tasting, the method is not suitable for situations where the consumer has not tasted the wine (e.g. a first buy). Secondly, and as mentioned before, the use of experts' ratings as a proxy for sensory quality has been questioned (Hodgson 2009). Finally, the proposed estimation method neglects the inherent noise of perceived quality, therefore introducing endogeneity (the measurement of perceived quality becomes a noisy proxy). We tested an analogous procedure to Mastrobuoni et al. (2014) with our dataset, yielding only positive coefficients for price.

In our application, significant coefficients were obtained for the positive effect of price on students, and on all classes of participants for the negative effects of price. Consumers and experts' positive effect of price had the expected sign, with (one-sided) t-tests' p values of 0.06 and 0.09. Results seem to be robust to the particular model structure, as we estimated models without random parameters and with random price parameters, and results remained analogous (i.e. the sign of the price coefficients remained the same, and their t-tests did not decrease significantly).

Students were found to be the most price-sensitive group, but also those who more strongly associated price with quality, as it would be expected for lower-income and less knowledgeable consumers. Experts appeared to be more price sensitive than regular consumers, probably because they purchased wine more frequently than regular consumers and therefore looked for cheaper alternatives. This, however, is only a possible explanation, as we did not record consuming nor purchase frequency in our questionnaire. We tested for other possible explanations, such as income effect and non-linearity in the effect of price, but none of them turned out to be significant.

The positive effect of price could be overstated because of our experimental design. By asking participants at the beginning of the experiment what their minimum and maximum willingness to pay for wine were, and then using these values throughout the SC scenarios, we might have reinforced the use of price as cue for quality[1]. Let us consider the following situation. Participants, when asked for their minimum WTP, think of the lowest quality wine they would be willing to buy and state their WTP for it. Then, when asked for their maximum WTP, respondents think of the best quality wine they have tried, and state their WTP for it. This would lead them to associate immediately the low price with low quality and the high price with high quality during the SC experiment. This, however, could not happen if participants did not use price as a cue for quality. If that was the case, then their willingness to pay range would be completely determined by their budget constraint and the lowest wine price in the market. Therefore, even though our experimental design might have artificially increased the positive effect of price to some degree, it could not have artificially induced it. To avoid this potential problem, we recommend using predetermined price ranges when applying the endogeneity correction method.

Concerning attributes other than price, results of the model with (HC) and without (ML) endogeneity correction are generally aligned, but the HC model seems to provide more information. Preferences for grape variety are highly heterogeneous in the sample,

[1] We are grateful to an unknown referee for having made us note this.

making it impossible to declare a single variety as preferred –on average- over the others. Labels 1 and 4 are less preferred than the base label, while preferences for labels 2, 3 and 5 seem to be equivalent to the base label on average, though both models disagree on the (significant) level variability of these preferences. Finally, while the ML model makes no difference among preferences for stories, the HC model suggests that story 1 is –on average- significantly less preferred than the base story. These results indicate that price endogeneity might not only affect the price parameters, but also other attributes' parameters, though to a lesser degree.

We used random coefficients to capture preference heterogeneity, but latent classes models are also an interesting approach to capture it. Latent classes are easier to interpret than random coefficients, but when the number of different classes is big, they require a higher number of parameters to be estimated. In the case of wine, the heterogeneity of preferences is such that many different classes would be required (as confirmed by some preliminary estimations). Given that our sample had a limited size, we decided to use random coefficients instead.

Contrary to some published literature (Quester & Smart 1998, Hall 2003, Martinez-Carrasco et al. 2006), we found that the effect of consuming occasion was non-significant for Chinese respondents. Several factors may have influenced this result. First, we may have described the consuming occasion without enough detail, making it difficult for participants to picture themselves in it. Secondly, it may be that simply stating the consuming occasion is not enough to evoke such a context in the mind of Chinese consumers; therefore, more compelling methods should be tried in the future. Finally, in formal occasions the Veblen (or snob) effect is more likely to play an important role, but this effect usually manifests itself through brand value. As brands were fictional in our experiment, this effect was probably absent, therefore diminishing the effect of consuming occasion.

Several simple improvements could be applied to the method employed in this paper in order to correct for endogeneity. First, more than one expected quality indicator could be used, though it remains to be determined what indicators would be best. Secondly, it is not necessary to collect an expected quality indicator for each alternative, as it would be possible to separate the survey into two parts: one where only quality indicators are collected (i.e. a series of wines the expected quality of which had to be assessed), and another one where only choices are required (i.e. as in a traditional SC experiment). This could allow optimizing the data collection method by using different efficient designs for each stage, but might decrease the correlation between expected quality and choice.

Despite the limitations of this particular application, modelling quality as a latent variable seems to be a promising approach to deal with endogeneity while being consistent with commonly accepted behavioural frameworks and not demanding excessive extra effort from respondents. Additionally, this method does not require difficult–to-find proxies or instruments. Finally, the method seems to be fairly robust, as it worked on a relatively small and very heterogeneous sample, using a single quality indicator, and on a choice experiment that was not incentive compatible, where choosing an expensive wine had no actual consequence on participants.

Even though this particular case study was concerned with wine choice, the modelling structure can be applied to any product the quality of which is uncertain to the consumer, even after considering observable attributes. Most food and beverage products fit this description, but also many leisure activities do too (e.g. selecting a travel company, choosing a show or a play, etc.), as well as some sparsely bought products or services the quality of which is hard to determine by the consumer even after purchase (e.g. jewellery, some medicine, broadband providers, etc.).

Our approach should not be of much use in cases were the main source of endogeneity is the Veblen effect. In such cases, endogeneity is caused by the unobserved social benefits of conspicuous consumption, which are correlated with price, but are not related with perceived or expected quality. Therefore, modelling quality as a latent variable in such cases would not provide any new information; even more, it might lead to the wrong conclusion that price itself has a positive effect on consumers, when in reality it is conspicuous consumption that provides utility to the consumer. In these situations, including the brand of the product or a measure of its social appreciation might be more useful.

It is very likely that consumers present both the Veblen effect and the use of price as a cue for quality at the same time. This is probably more problematic in a revealed preference context, were brand and quality uncertainty go hand in hand, but less so in a SP experiment with fictional brands, such as the one analysed in this paper. As no real brands were presented, there is no benefit to be obtained from conspicuous consumption, beside that provided by the observable attributes (e.g. a particular label looking more luxurious than another).

Modelling quality as a latent variable assumes that prices are exogenous. Therefore, this method corrects endogeneity only due to the use of price as cue for quality, but does not correct endogeneity due to price’s simultaneous determination (i.e. supply and demand equilibrium). If this later effect is to be considered, then an additional endogeneity correction method especially suited for it should be used.

The latent variable approach for endogeneity correction shows highly promising results, but its real performance should be measured against a revealed preference study, an area we are currently working on. The method should also be compared with other available approaches to correct for endogeneity, notably the Control Function Approach (CFA) and Multiple Indicator Solution (MIS).

6 Acknowledgements

We are grateful to the Chilean Fund for the Development of Scientific and Technological Research (FONDECYT) through Projects 1121058 and 1150590, and to Anakena winery. Thanks are also due to the Millennium Institute in Complex Engineering Systems (ICM: P05-004F; FONDECYT: FB016) for having partially financed this work. We also wish to thank the useful comments of Stephane Hess and four anonymous referees who helped us producing a much better paper. Of course any errors are our responsibility alone.

7 References

Appleby, C., Costanigro, M., Thilmany, D. and Menke, S. (2012). Measuring consumer willingness to pay for low-sulfite wine: a conjoint analysis. Working Paper No. 117, American Association of Wine Economists.

Aqueveque, C. (2006) Extrinsic cues and perceived risk: the influence of consumption situation. Journal of Consumer Marketing **23**, 237 – 247.

Aqueveque, C. (2008) The effect of extrinsic cues on imported wine evaluations: an experimental approach. Journal of Food Products Marketing **14**, 33-47.

Atkin, T. and Thach, L. (2012) Millennial wine consumers: risk perception and information search. Wine Economics and Policy **1**, 54 – 62.

Bagwell, L.S. and Bernheim, D. (1996) Veblen effects in a theory of conspicuous consumption. The American Economic Review **86**, 349 – 373.

Bahamonde-Birke, F.J., Kunert, U., Link, H. and Ortúzar, J. de D. (2015) About attitudes and perceptions – finding the proper way to consider latent variables in discrete choice models. Transportation, 1-19. doi: 10.1007/s11116-015-9663-5.

Barreiro-Hurlé, J., Colombo, S. and Cantos-Villar, E. (2008). Is there a market for functional wines? Consumer preferences and willingness to pay for resveratrol-enriched red wine. Food Quality and Preference **19**, 360 – 371.

Berry, S., Levinsohn, J. and Pakes, A. (1995) Automobile prices in market equilibrium. Econometrica **63**, 841 – 890.

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models, Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Bolduc, D. and Alvarez-Daziano, R. (2010). On estimation of hybrid choice models. In S. Hess and A. Daly (eds,), Choice Modelling: The State-of-the-Art and the State-of-Practice. Emerald Group Publishing, Bingley.

Caussade, S., Ortúzar, J. de D., Rizzi, L.I. and Hensher, D.A. (2005) Assessing the influence of design dimensions on stated choice experiment estimates. Transportation Research **39B**, 621-640.

Caves, R.E. and Greene, D.P. (1996) Brand's quality levels, prices, and advertising outlays: empirical evidence on signals and information costs. International Journal of Industrial Organization **14**, 29 – 52.

Chaney, I. (2000) External search effort for wine. International Journal of Wine Marketing **12**, 5 – 21.

Chapman, R.G. and Staelin, R. (1982) Exploiting rank ordered choice set data within the stochastic utility model. Journal of Marketing Research **19**, 288 – 301.

Cuellar, S.S. and Huffman, R. (2008) Estimating the demand for wine using instrumental variable techniques. Journal of Wine Economics **3**, 172 – 184.

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**, 287 – 303.

Daly, A. and Hess, S. (2010) Simple approaches for random utility modelling with panel data. In European Transport Conference, October 2010.

Deeter-Schmelz, D.R., Moore, J.N. and Goebel, D.J. (2000) Prestige clothing shopping by consumers: a confirmatory assessment and refinement of the Precon scale with managerial implications. Journal of Marketing Theory and Practice **8**, 43 – 58.

Dodds, W.B., Monroe, K.B. and Dhruv, G. (1991) Effects of price, brand, and store information on buyers' product evaluation. Journal of Marketing Research **28**, 307 – 319.

Dubow, J.S. (1992). Occasion-based vs. user-based benefit segmentation: a case study. Journal of Advertising Research **25**, 72 – 82.

Gokcekus, O. and Nottebaum, D. (2011) The buyer's dilemma – whose rating should a wine drinker pay attention to? Working Paper N° 91, American Association of Wine Economists.

Goodman, S. (2009). An international comparison of retail consumer wine choice. International Journal of Wine Business Research **21**, 41 – 49.

Greene, W.H. and Hensher, D.A. (2010). Modelling Ordered Choices: A Primer. Cambridge University Press, Cambridge.

Grisolía, J.M., López, F. and Ortúzar, J. de D. (2012) Sea Urchin: from plague to market opportunity. Food Quality and Preference **25**, 46-56.

Guevara (2015) Critical assessment of five methods to correct for endogeneity in discrete-choice models. Transportation Research **82A**, 240 – 254.

Guevara, C.A. and Ben-Akiva, M. (2006) Endogeneity in residential location choice models. Transportation Research Record **1977**, 60 – 66.

Guevara, C.A. and Ben-Akiva, M. (2012) Change of scale and forecasting with the control-function method in logit models. Transportation Science **46**, 425 – 437.

Guevara, C. A. and Polanco, D. (2016). Correcting for endogeneity due to omitted attributes in discrete-choice models: the multiple indicator solution (MIS). Transportmetrica A (in press).

Hall, E.J. (2003). The Influence of Occasion on Consumer Choice: An Occasion Based, Value Oriented Investigation of Wine Purchase, Using Means-End Analysis. Ph.D. Thesis, School of Agriculture and Wine, University of Adelaide.

Hess, S., Train, K. and Polak, J. (2006) On the use of a modified Latin hypercube sampling (MLHS) method in the estimation of a mixed logit model for vehicle choice. Transportation Research **40B**, 147 – 163.

Hodgson, R.T. (2009) How expert are “expert” wine judges. Journal of Wine Economics **4**, 233 – 241.

Hopfer, H. and Heymann, H. (2014) Judging wine quality: do we need experts, consumers or trained panelists? Food Quality and Preference **32**, 221 – 233.

Horowitz, J. (1983) Statistical comparison of non-nested probabilistic choice models. Transportation Science **17**, 319 – 350.

Jaeger, S. and Rose, J.M. (2008). Stated choice experimentation, contextual influences and food choice: a case study. Food Quality and Preference **19**, 539–564.

Johnson, T. and Bruwer, J. (2004) Generic consumer risk-reduction strategies (RRS) in wine-related lifestyle segments of the Australian wine market. International Journal of Wine Marketing **16**, 5 – 35.

Lattey, K.A., Bramley, B.R. and Francis, I.L. (2009) Consumer acceptability, sensory properties and expert quality judgments of Australian Cabernet Sauvignon and Shiraz wines. Australian Journal of Grape and Wine Research **16**, 189 – 202.

Lawless, H.T. (1984) Flavor description of white wine by "expert" and non-expert wine consumers. Journal of Food Science **49**, 120 – 123.

Leavitt, H.J. (1954) A note on some experimental findings about the meaning of price. The Journal of Business **27**, 205 – 210.

Lewis, G. and Zalan, T. (2014). Strategic implications of the relationship between price and willingness to pay: evidence from wine-tasting experiment. Journal of Wine Economics **9**, 115-134.

Lichtenstein, D.R., Ridgway, N.M. and Netemeyer, R.G. (1993) Price perceptions and consumer shopping behavior: a field study. Journal of Marketing Research **30**, 234 – 245.

Lockshin, L. and Corsi, A.M. (2012). Consumer behaviour for wine 2.0: a review since 2003 and future directions. Wine Economics and Policy **1**, 2 – 23.

Lockshin, L., Jarvis, W., d'Hauteville, F. and Perrouty, J.P. (2006). Using simulations from discrete choice experiments to measure consumer sensitivity to brand, region, price, and awards in wine choice. Food Quality and Preference **17**, 166 – 178.

Martínez-Carrasco, L., Brugarolas, M., Del Campo, F.J. and Martínez, A. (2006). Influence of purchase place and consumption frequency over quality wine preferences. Food Quality and Preference **17**, 315–327.

Mastrobuoni, G., Peracchi, F. and Tetenov A. (2014) Price as a signal of product quality: some experimental evidence. Journal of Wine Economics **9**, 135 – 152.

McFadden, D. and Train, K.E. (2000). Mixed MNL models for discrete response. Journal of Applied Econometrics **15**, 447-440.

Michis, A.A. and Markidou, A.G. (2013) Determinants of retail wine prices: evidence from Cyprus. Empirical Economics **45**, 267 – 280.

Milgrom, P. and Roberts, J. (1986) Price and advertising signals of product quality. Journal of Political Economy **94**, 796 – 821.

Mtimet, N. and Albisu, L.M. (2006). Spanish wine consumer behavior: a choice experiment approach. Agribusiness **22**, 343 – 362.

Mueller, S., Lockshin, L., Saltman, Y. and Blanford, J. (2010). Message on a bottle: the relative influence of wine back label information on wine choice. Food Quality and Preference **21**, 22-32.

Nelson, P. (1970) Information and consumer behavior. Journal of Political Economy **78**, 311 – 329.

O'Neill, V., Hess, S. and Campbell, D. (2013). A question of taste: recognising the role of latent preferences and attitudes in analysing food choices. Food Quality and Preference **32**, 299 – 310.

Ortúzar, J. de D. and Willumsen, L.G. (2011). Modelling Transport. John Wiley and Sons, Chichester.

Palma, D., Ortúzar, J. de D., Casaubon, G., Rizzi, L.I. and Agosin, E. (2013) Measuring consumer preferences using hybrid discrete choice models. Working paper N° 137, American Association of Wine Economists.

Petrin, A. and Train, K. (2010) A control function approach to endogeneity in consumer choice models. Journal of Marketing Research **47**, 3 – 13.

Plassmann, H., O'Doherty, J., Shiv, B. and Rangel, A. (2008) Marketing actions can modulate neural representations of experienced pleasantness. Proceedings of the National Academy of Sciences of the United States of America **105**, 1050 – 1054.

Quester, P. and Smart, J. (1998) The influence of consumption situation and product involvement over consumers' use of product attribute. Journal of Consumer Marketing **15**, 220 – 238.

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

Remaud, H., Mueller, S., Chvyl, P. and Lockshin, L. (2008, July) Do Australian wine consumers value organic wine? 4th International Conference of the Academy of Wine Business Research, Siena.

Revelt, D. and Train, K. (1998) Mixed logit with repeated choices: household's choices of appliance efficiency level. The Review of Economics and Statistics **80**, 647 – 657.

Richard, T.J. and Padilla, L. (2009) Promotion and fast food demand. American Journal of Agricultural Economics **91**, 168 – 183.

Rose, J.M. and Bliemer, M.C.J. (2009). Constructing efficient stated choice experimental designs. Transport Reviews **29**, 1-31.

Rose, J.M., Bliemer, M.C.J., Hensher, D.A. and Collins, A. (2008). Designing efficient stated choice experiments in the presence of reference alternatives. Transportation Research **42B**, 395-406.

Scitovsky, T. (1945) Some consequences of the habit of judging quality by price. The Review of Economic Studies **12**, 100 – 105.

Stasi, A., Nardone, G. and Viscecchia, R. (2011) Italian wine demand and differentiation effect of geographical indications. International Journal of Wine Business Research **23**, 49 – 61.

Stasi, A., Bimbo, F., Viscecchia, R. and Seccia, A. (2014) Italian consumers' preferences regarding dealcoholized wine, information and price. Wine Economics and Policy **3**, 54 – 61.

Train, K. (2009). Discrete Choice Models with Simulation. Cambridge University Press, Cambridge.

Veblen, T. (1899/1994) The Theory of the Leisure Class: An Economic Study of Institutions. Unwin Books, London (1899); reprinted by Dover Publications, New York (1994).

Villas-Boas, J.M. and Winer, R.S. (1999) Endogeneity in brand choice models. Management Science **45**, 1324 – 1338

Walker, J. and Ben-Akiva, M. (2002). Generalized random utility model. Mathematical Social Sciences **43**, 303–343.

Wooldridge, J. (2002) Econometric Analysis of Cross Section and Panel Data. The MIT Press, Cambridge, Mass.