Modelling Wine Choice: Investigating the determinants of wine choice among of the "Black Diamonds"

By

Leah Z.B. Ndanga[1], André Louw[2], Johan van Rooyen[3] & Davison Chikazunga[4]

1. *M.Sc. Student:* Dept. of Agricultural Economics, Extension and Rural Development, University of Pretoria, Pretoria 0002 South Africa. E-Mail: lzbndanga@tuks.co.za Tel: (+27 12) 420 4584 Fax: (+27 12) 420 4958 **(Corresponding Author)**

2. *Professor & ABSA Chair in Agribusiness*: Dept. of Agricultural Economics, Extension and Rural Development, University of Pretoria, Pretoria 0002 South Africa. E-Mail: andre.louw@up.ac.za Tel: (+27 12) 420 5772 Fax: (+27-12) 420 3247

3. *Professor & ABSA Chair in Agribusiness:* Dept. of Agricultural Economics, Extension and Rural Development, University of Pretoria, Pretoria 0002 South Africa. E-Mail: cjvr@sun.ac.za Tel: (+27 21) 808 2374 Fax: (+27 21) 882 9510

4. *PhD. Student:* Dept. of Agricultural Economics, Extension and Rural Development, University of Pretoria, Pretoria 0002 South Africa. E-Mail: dchikaz110@yahoo.com Tel: (+27 12) 420 5767 Fax: (+27 12) 420 4958

For

The International Food & Agribusiness Management Association

19TH Annual World Symposium

Budapest, Hungary

June 20-21, 2009

Modelling Wine Choice: Investigating the determinants of wine choice among of the "Black Diamonds"

EXECUTIVE SUMMARY

This paper uses a choice based conjoint analysis in an attempt to develop a consumer profile for the new market for black consumers. In this study all the respondents are combined, as in the alternate hypothesis which asserts that there are no differences and therefore no segments; and by studying subsets defined by specific market segments, such as gender and other differences in the null hypothesis.

Although the different statistical packages used variants of the MNL model, the results showed no significant contradictions in their results. Despite the models' imminent statistical insignificance, they suggested valuable notions about black consumers' wine choice determinants. The main effects model suggests that women prefer red wine; white and sparkling wine drinkers are willing to spend less for a bottle of wine; Baronne wine drinkers prefer white and sparkling wines and educated wine drinkers prefer red wine.

In terms of the marginal effects models, with respect to red wines over the other wines, the study asserts that consumers' choice of their favourite red wine, age, income and frequency of consumption are significant determinants of their choice. In terms of white wine over the other wines, age and favourite red wine are statistically significant determinants of the choice of white wines. Age, income and frequency of consumption are statistically significant determinants of consumers' choice of sparkling wines over other wines.

Age, gender and the choice of favourite red wine may be used to segment the market as they are often significant determinants of wine choice. The other significant coefficients affect the marketing and distribution choices to be followed by wine companies. The study illustrates the need for further research in the areas of wine choice modelling and market segmentation.

Modelling Wine Choice: Investigating the determinants of wine choice among of the "Black Diamonds"

ABSTRACT

This paper uses a choice based conjoint analysis in an attempt to develop a consumer profile for the new market for black consumers. Although the different statistical packages used variants of the MNL model, the results were significantly similar with no contradictions in their results. Despite the models' imminent statistical insignificance, they suggested valuable notions about black consumers' wine choice determinants. Age, gender and the choice of favourite red wine may be used to segment the market and the other significant coefficients will affect the marketing and distribution choices to be followed by wine companies.

Key words: random utility models, wine choice

Modelling Wine Choice: Investigating the determinants of wine choice among of the "Black Diamonds"

INTRODUCTION

This paper forms part of a Master's study by Ndanga (2009) which sought to develop a framework of reference to assist with the formulation of marketing strategy recommendations for South Africa in terms of the generally untapped emerging black middle class market by identifying and characterizing existing and potential wine consumers and their preferences in order to shift more consumers from beer, and other beverages, to wine consumption. In this paper, as it was in the study, the choice based conjoint (CBC) analysis was undertaken in an attempt to develop a consumer profile for the new market for black consumers, as well as changing consumer attitudes toward wines. CBC was used because it can reveal the interactions of the attributes as well as the consumer's characteristics and the purchase situation through discrete choice experiments (Louviere & Woodworth, 1983 in Gil & Sanchez, 1997).

Johnson, *et al.* (1991) employed conjoint techniques to benefit segmentation in the Australian wine market (Engels, *et al.*, 2004; Gil & Sanchez, 1997), as did Mtimet and Albisu (2006) in their segmentation of the Spanish consumer market. In the last years, the use of choice experiments to analyze wine consumption and wine consumer behaviour has been growing as can be seen from the studies of Berti, 2003; Lockshin, Jarvis, Perrouty, & d'Hauteville, 2006; Perrouty, d'Hauteville, & Lockshin, 2006; Rasmussen, 2001 (Mtimet & Albisu, 2006:3). The discrete choice analysis was also used to gain insight into consumer preferences for New Mexico wine in the study by Allimova, *et al.*, (2006) and by the US firm Tragon, (Penn, 2007). Applications of conjoint analysis to food products can be found, among others, in Johnson *et al*. (1991) for Australian wine, Loader (1990) for fruit and vegetables in the UK, and Ness and Gerhardy (1994) for British eggs (Gil & Sanchez, 1997).

In choice-based conjoint (CBC) analysis the respondent expresses preferences by choosing concepts from sets of concepts, rather than by rating or ranking them. In this study all the respondents are combined, as in the alternate hypothesis which asserts that there are no differences and therefore no segments, and by studying subsets defined by specific market segments, such as gender and other differences in the null hypothesis. "Utility values" are

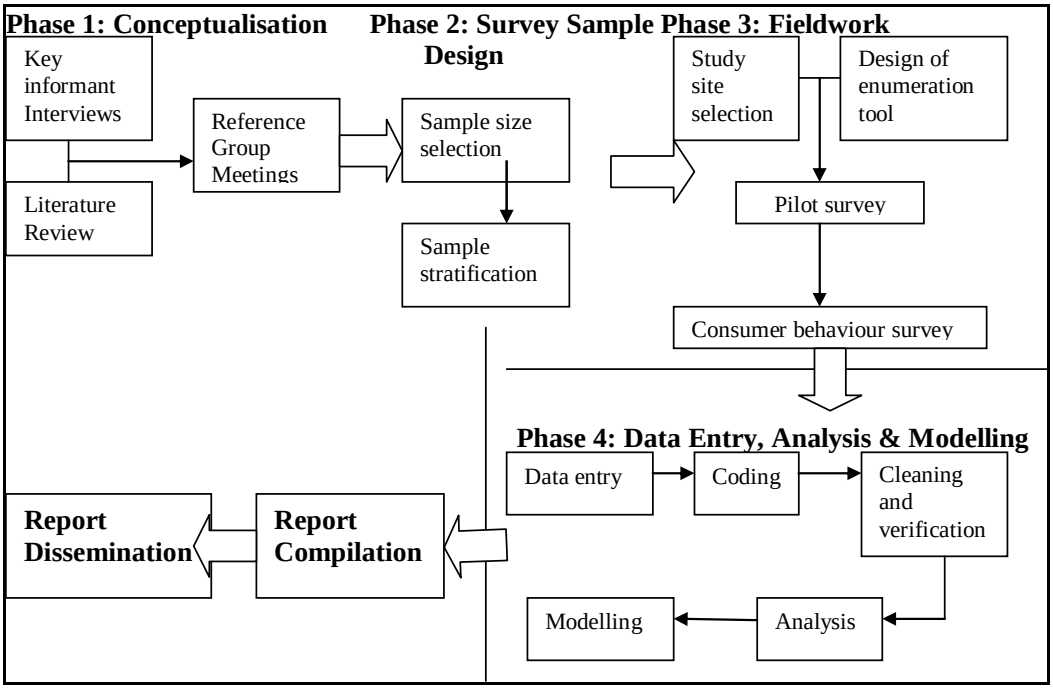
produced for each group of respondents that summarize the choices made by those individuals. And, as in other conjoint methods, the utility values can be used to simulate and predict respondent reactions to product concepts that may not have actually appeared in the choice tasks (questions). The calculation of utilities is completed across the respondent base, typically using aggregate multinomial logit. This operational version of our proposed random utility model (PRU) generalizes the widely MNL model of wine choice (Sawtooth, 1999:2; Pazgal, *et al.*, 2005: 12; Poynter, 2005:7).

The paper seeks to assess the different methods by which a Random Utility Model (RUM) can be constructed and interpreted in order to determine the determinants of wine choice among South Africa black middle class consumers. The next section describes the data used for the different models discussed in this paper. This paper discusses the conjoint analysis and random utility modelling undertaken on the results obtained from the analysis of the data. It discusses the assumptions made in the modelling process, the methodology and interpretations of the random utility model, the findings of the different statistical packages, as well as the limitations of conjoint analysis, random utility modelling and the different statistical packages, and conclusions made from the scholarly trial and error process discussed here. The paper will show that regardless of what statistical package used, it is still very difficult to clearly objectively ascertain the determinants of wine choice, or any other qualitative variable.

DESCRIPTION OF THE DATA

The data and information used in this paper was collected from an integration of a consumer behaviour survey as it was in the study by Engels, *et al* (2004); as well as personal interviews with industry stakeholders and focus group discussions, as in the annual US Wine Market Council consumer surveys and the study by Schmidt (2001). Consumer behaviour questions and subsequent analysis provided answers related to peoples' behaviour and attitudes towards wine; the interviews determined industry stakeholders' perceptions on the current state and future outlook of the South African wine industry; and the focus group discussions provided a basis for the analysis for qualitative data. A summary of the conceptual framework and implementation plan is illustrated in Figure 1.

Figure 1: Conceptual Framework



The data and information used in this study was collected from a consumer behaviour survey using a mall-intercept survey at the 2007 Soweto Wine Festival. The target population was selected on the basis of age, gender, income, race and wine drinking history. The study asserts that the black middle class are different from the white middle class and within the Black Diamonds different segments exist on the basis of factors selected. The sample represents a cross section of the black emerging middle class in South Africa (Tzimitra-Kalogianni, *et al*, 1999:886; Engels *et al.*, 2004). Gauteng was the chosen province for the consumer behaviour survey as various studies have shown it to be the province with the highest concentration of "Black Diamonds".

Table 1: Sampling unit requirements

Criteria	**Specific Requirements**
Race	Black
Age	Must fall into any one of the 4 distinct super-segments for Black Diamonds
Gender	An equal number of Females and Males
Income	Must be either be a student (receiving an allowance) or have some form of income – formal or otherwise
Wine Drinking History	Must have tasted white, red and "pink" wines at least once

The consumer behaviour survey followed a non–probabilistic, quota sample selection process based on the available marketing data and findings from Phase one. The sampling procedure was a non random multi-level stratification of the black middle class wine consumers in the Guateng province of South Africa. The target population of the study is South Africa's black middle class, increasingly referred to as Black Diamonds. All black South Africans present at the festival meeting the afore-mentioned criteria will form the target population. This forms the first level of the sampling frame.

Given that the Black Diamonds' population is approximately 2.6 million and that Research Surveys identified four distinct super-segments for Black Diamonds, four age based segments were also used in this study. This study sought to test the aptness of these segments. Quota sampling in which a stratified sample based on non random selection of sampling units was used for the study.

Given a confidence level of 95 percent and the confidence interval of five percent, a sampling frame of a total of 384 respondents and 91 respondents for each segment should be interviewed in accordance with the formula given in Equation 1. However, to allow for non random sampling errors, a total of 400 respondents and 100 respondents for each super-segment should be interviewed.

Equation 1

$$n = \frac{Z^2*(p)*(1-p)}{c^2}$$

Source: Bartlett, *et al*., 2001:47

where: n is the sample size

Z = Z value (e.g. 1.96 for 95% confidence level)

p = percentage picking a choice, expressed as decimal (.5 used for sample size needed)

c = confidence interval, expressed as decimal (e.g., .05 = ±5)

In actuality, four hundred and two respondents were interviewed and only three hundred and eighty seven were acceptable. These 387 questionnaires were analysed, the results of which are presented in this paper.

DATA MODELLING

Random utility (RU) models are well-established methods for describing discrete choice behaviour. Utility maximization is the objective of the decision process and leads to observed choice in the sense that the consumer chooses the alternative for which utility is maximal. Individual preferences depend on characteristics of the alternatives and the tastes of the consumer. A RU model defines a mapping from observed characteristics into preferences. All the factors affecting preferences are treated as random variables (Baltas & Doyle, 2001:116). The Multinomial Logit (MNL) model is the appropriate treatment of unobserved product attributes. Although in theory, other models (e.g. a restricted probit) can be cast as members of the same class, but in practice, only the MNL has been used. MNL regression is used when the dependent variable in question is nominal (a set of categories which cannot be ordered in any meaningful way) and consists of more than two categories. For example, in this study MNL regression is deemed appropriate for trying to determine what factors affect black consumers' choice of wines, in terms of whether they prefer red, white or sparkling wines.

ASSUMPTIONS OF THE RU MODEL

In accordance with the hypotheses of the study, the MNL model assumes that:

i. The emerging black middle class as a consumer segment are heterogeneous
ii. Various independent factors affect black consumers' wine choice, each of which has a single value for each case, is not linearly correlated to another and of which the odds of wine choice do not depend on other alternatives that are available (i.e., that including additional alternatives or deleting alternatives will not affect the odds on the dependent variable among the alternatives that were included originally)
iii. There are significant differences in terms of wine choice according to gender
iv. Women prefer sparkling and white wines
v. The new emerging "black diamond" consumer market are willing to pay for their wine
vi. Black consumers are willing to become wine drinkers and engage in the ensuing lifestyle
vii. Wine choice variable cannot be perfectly predicted from the independent variables for any case.

THE RU MODEL

In CBC, the utility that the i^{th} person (i = 1,….,I) derives from the j^{th} alternative may be represented as U_{ij}. This utility is considered a linear function of the alternative product attributes, represented by

$$U_{ij} = \beta\, x_{ij} + \varepsilon_{ij}$$

Where β is a vector of coefficients, x is a vector of attributes represented by choice *j* and respondent *i*, and ε is a stochastic error term. The probability P_{ij} the i^{th} respondent chooses the j^{th} alternative from choice set C is the probability that the utility for the j^{th} choice is greater than the utility for all other *k* choices in the choice set. This can be represented mathematically as follows:

$$\Pr(y_i = j) = \frac{\exp(X_i\beta_j)}{1 + \sum_j^J \exp(X_i\beta_j)}$$

and assuming that the error terms (ε_{ij}) are independent and identically distributed with an extreme value distribution (also referred to as Weibull, Gumbel and double exponential distributions) and scale parameter equal to 1, the probability that respondent *i* chooses alternative *j* is:

$$\Pr(y_i = 0) = \frac{1}{1 + \sum_j^J \exp(X_i\beta_j)},$$

Where for the i^{th} individual, y_i is the observed outcome and X_i is a vector of explanatory variables. The unknown parameters β_j are typically estimated by maximum likelihood. It is noteworthy that different distributional assumptions yield different operational versions of the traditional random utility model. For example, in this study, the errors are assumed to be distributed IID Gumbel with an unknown scale parameter μ (and location parameter equal to zero), this renders the traditional random utility model to be the MNL (Pazgal, *et al.*, 2005:20; Mtimet & Albisu, 2006:346).

INTERPRETING THE RU MODEL

When using MNL regression, one category of the dependent variable is chosen as the comparison category. In this study, the choice of red wines as the favourite wine choice was chosen as the comparison category. Separate relative risk ratios are determined for all independent variables for each category of the independent variable with the exception of the comparison category of the dependent variable, which is omitted from the analysis. Relative risk ratios, the exponential beta coefficient, represent the change in the odds of being in the

dependent variable category versus the comparison category associated with a one unit change on the independent variable. This results in a set of numbers comparable to conjoint "utilities," except that they describe preferences for a group rather than for an individual.

CBC's MNL regression reports logit coefficients as well as t and chi square statistics. The regression estimates all main effects (default) and two-way interactions optionally. CBC analysis allows for the selection of main effects and interactions to be included in each logit analysis. When only main effects are estimated, a value is produced for each attribute level that can be interpreted as an "average utility" value for the respondents analyzed. When interactions are included, effects are also estimated for combinations of levels obtained by cross-classifying pairs of attributes (Bierlaire, 1997; Sawtooth, 1999:19).

The main effects model consists of different estimated coefficients. Identification of the wine choice model parameters requires one of the discrete choice indicators in the MNL model to be normalized to zero. Therefore, the structural parameters consist of marginal utilities of attributes of the selected coverage levels relative to the excluded alternative. Initial parameter values for this model were obtained by specifying a “null” model where all wine consumers prefer red wine except for the choice-specific intercept value. The coefficients pertain to alternative specific constants and these constants are estimated relative to the red wine choice alternative which has an implicit value of 0. The rest of the attribute coefficients were estimated relative to one of the attribute levels. That attribute level is omitted from the model since its effect can be defined from the estimated effects of the other three attribute levels.

For example, for the gender attribute, females are omitted. The estimated effects of gender are relative to the wine choice. Any statistical differences that occur are estimated relative to the attribute level that is omitted. The other omitted attribute levels in this model are very low expenditure on wine for personal consumption, favourite red wine and participation in a wine course (Lockshin & Haelstaed, 2005; Mayen & Marshall, 2005:11; Mtimet & Albisu, 2006:350).

The discrete choice data was analysed using three different statistical packages; the SPSS 15.0 MNL program, STATA 8.0 and SAS. The various programs ran different models using various attributes to ascertain the essential attributes to the model. Using SPSS, of the attributes selected, two separate models (with the intercept only and with all the coefficients)

were run using the same MNL analysis. The STATA program ran marginal effects regressions. The SAS model used the main effects model. The programs ran different models using various attributes to ascertain the essential attributes to the model. Of the attributes selected, two separate models (with the intercept only and with all the coefficients) were run using the same MNL analysis. However, it should be noted that there are other variables that were not captured in this model.

This model assumes that:

Wine choice (in terms of red, white or sparkling) = f (gender, expenditure on wine for personal consumption, engagement in any form of wine education)

The pertinent null and alternate hypotheses are given as:

H_0 = consumers prefer red wine, there are significant differences according to gender; the type of red wine preferred as well as the attendance to a wine course affects wine choice.

H_A = consumers are homogenous and prefer white and sparkling wines.

The variables used within the model, as well as their definitions, expected signs and interpretations for these signs are given in Table 2. It should be noted that the first three variables are the dependant variables and the rest are the independent variables. The independent variables included in this model have been found through a process of trial and error and other results can be obtained if other explanatory variables different from those included in this model are used.

Table 2: Variables used within the MNL model

Variable	Definition	Expected Sign	Interpretation
fav_wine = 0	red wines		The more positive the sign on the variable coefficient means that consumers prefer red wines
fav_wine = 1	white wines		As the variable coefficient moves towards zero it means the consumers prefer white wines
fav_wine = 2	sparkling wines		The more negative the sign on the variable coefficient means that consumers prefer sparkling wines
gender=0	females	negative	More likely to favour white and sparkling wines
gender=1	males	positive	More likely to favour red wines
own_spen=0	R50 - R100	positive	More likely to favour red wines
own_spen=1	< R20	negative	More likely to favour white and sparkling wines

own_spen=2	R21 - R35	negative	More likely to favour white and sparkling wines
own_spen=3	R36 - R49	positive	More likely to favour red wines
own_spen=4	> R100	positive	More likely to favour red wines
own_spen=5	Do not purchase	negative	More likely to favour white and sparkling wines
own_spen=6	Free	negative	More likely to favour white and sparkling wines
fav_rw=0	Baronne	positive	More likely to favour red wines
fav_rw=1	Do not drink red wine	negative	More likely to favour white and sparkling wines
fav_rw=2	Pinotage	positive	More likely to favour red wines
fav_rw=3	Shiraz	positive	More likely to favour red wines
fav_rw=4	Rose	positive	More likely to favour red wines
fav_rw=5	Cabernet	positive	More likely to favour red wines
fav_rw=6	Red blends	positive	More likely to favour red wines
fav_rw=7	Merlot	positive	More likely to favour red wines
fav_rw=8	Cabernet Sauvignon	positive	More likely to favour red wines
fav_rw=9	Pinot Noir	positive	More likely to favour red wines
wine_cou=1	Attended wine course	positive	More likely to favour red wines
wine_cou=2	Have not attended wine course	negative	More likely to favour white and sparkling wines

The results of the three various models are presented below, in order of their acceptability with respect to the statistical significance.

THE SAS MODEL

The discrete choice data was analysed using the SAS program. The program ran different models using various attributes to ascertain the essential attributes to the model. The results are given in the table below. It should be noted that there are other variables not captured in this model.

Table 3: Results of model log likelihood tests

	Model fitting criteria	**Likelihood ratio test**		
Model	**-2 Log Likelihood**	**Chi-Square**	**Degrees of Freedom**	**Significance**
Intercept Only	542.737			
Final	499.961	42.776	45	0.567

The data indicates that the said attributes are not viable as they do not provide the best fit to the data. The coverage model is not a good fit to the data as the p value is far greater than 0.05 at 48 degrees of freedom. The model has debatably acceptable Pseudo R squared values.

This means that the model has a relatively low explanatory power as it explains only about 10% of the wine choice preferences.

Table 4: Pseudo R-Square

Cox and Snell	0.105
Nagelkerke	0.122
McFadden	0.056

Table 5 provides the parameter estimates from this stage.

Table 5: SAS output for MNL model

	Model fitting criteria	Likelihood ratio test		
Effect	**-2 Log Likelihood of reduced model**	**Chi-Square**	**Degrees of Freedom**	**Significance**
Intercept	499.961	.000	0	
Age	510.293	10.332	12	.587
Gender	504.562	4.601	3	.203
Wine drinking years (proxy for experience)	508.390	8.429	15	.905
Frequency of consumption	509.985	10.024	12	.614
Wine course (proxy for wine education)	502.365	2.404	3	.493

Link function: Logit.

The model could not be interpreted as it was insignificant and all the independent variables were also insignificant. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom. The unexpected singularities in the Hessian matrix experienced indicate that either some independent/predictor variables should be excluded or some categories should be merged. Further work was deemed necessary.

THE STATA MODEL

Due to the inadequacies of the prior model, the discrete choice data was analysed using the STATA program. Table 6 provides the results of the model log likelihood tests.

Table 6: Results of model log likelihood tests

Final Model	**-2 Log Likelihood**	**Chi-Square**	**Degrees of Freedom**	**Significance**
	-354.16388	43.13	24	0.0096

The results of the multinomial logistic regression indicate that the said attributes are tentatively viable as the coverage model provides a good fit to the data. This is because the p value is less than 0.05 at 48 degrees of freedom. However, the model has an undeniably low Pseudo R squared value of 0.0574. This means that the model has a relatively low explanatory power as it explains only about 5% of the wine choice preferences.

Table 7: STATA output for MNL model

	fav_wine	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
Red wine	age	41.72312	7452098	0	1	-1.46E+07	1.46E+07
	gender	-104.193	3.86E+07	0	1	-7.57E+07	7.57E+07
	income	-9.94782	7714057	0	1	-1.51E+07	1.51E+07
	wine_yrs	-165.374	.	.	.	.	.
	fav_rw	139.0835	.	.	.	.	.
	freq	88.29327	9457440	0	1	-1.85E+07	1.85E+07
	own_spen	41.09792	.	.	.	.	.
	wine_cou	256.1151	.	.	.	.	.
	_cons	-811.13	.	.	.	.	.
White wine	age	-0.17524	0.135578	-1.29	0.196	-0.4409632	0.090493
	gender	0.033518	0.243474	0.14	0.891	-0.4436819	0.510719
	income	-0.07748	0.136477	-0.57	0.57	-0.3449681	0.190014
	wine_yrs	0.058409	0.131028	0.45	0.656	-0.1984011	0.315219
	fav_rw	**0.789277**	**0.333054**	**2.37**	**0.018**	**0.1365029**	**1.442051**
	freq	-0.01657	0.11717	-0.14	0.888	-0.2462168	0.213081
	own_spen	-0.12014	0.119992	-1	0.317	-0.3553158	0.115043
	wine_cou	-0.22863	0.267067	-0.86	0.392	-0.752068	0.294816
	_cons	0.413545	0.602775	0.69	0.493	-0.7678722	1.594962
Sparkling wine	**age**	**0.303204**	**0.157**	**1.93**	**0.053**	**-0.0045095**	**0.610917**
	gender	0.281695	0.308995	0.91	0.362	-0.3239228	0.887313
	income	**-0.33594**	**0.171249**	**-1.96**	**0.05**	**-0.671578**	**-0.00029**
	wine_yrs	-0.14934	0.156021	-0.96	0.338	-0.4551326	0.15646
	fav_rw	0.291575	0.460952	0.63	0.527	-0.6118753	1.195024
	freq	**-0.38231**	**0.171049**	**-2.24**	**0.025**	**-0.7175616**	**-0.04706**
	own_spen	0.048039	0.143343	0.34	0.738	-0.2329072	0.328985
	wine_cou	-0.48608	0.33982	-1.43	0.153	-1.152118	0.179953
	_cons	0.128966	0.762468	0.17	0.866	-1.365444	1.623376

In this model fav_wine =1 which is the base outcome. The bold variables are significant at a less than 10% level of significance.

Where

Coefficients	**Interpretation**
[fav_wine = 1]	red wines
[fav_wine = 2]	white wines
[fav_wine = 3]	sparkling wines
Dummy variables	
[gender=0]	females
[gender=1]	males
[fav_rw=0]	Baronne

[fav_rw=≥1]	The other categories
Other variables	
Age	Age of respondents
Income	Average monthly income
Wine_yrs	Average number of years consuming wines
Freq	Frequency of wine consumption, irregardless of volume
Own_spen	Average expenditure on a standard (750ml) bottle of wine for personal consumption
[wine_cou=1]	Attended wine course
[wine_cou=2]	Have not attended wine course

Explanation

Favourite red wine is a significant determinant of whether or not respondents choose white wines as their favourite wines. The respondents' decision to drink white wines is affected by whether or not they choose Baronne as their favourite red wine. The positive coefficient suggests that respondents that choose Baronne as their favourite red wine are more likely to choose white wines over red wines as their favourite wines.

Age, income and frequency of consumption are statistically significant determinants of consumers' choice of sparkling wines over red and white wines. The positive coefficient on the age variable suggests that the older consumers get the more likely they are to choose to sparkling wines. The negative coefficients on the income and frequency variable suggest that consumers with lower incomes and those who consume wine less often are more likely to choose sparkling wines over red and white wines.

The following three outputs provide the marginal effects of red, white and sparkling wines, respectively.

MARGINAL EFFECTS OF RED WINE

. mfx, predict(p outcome(1))

Marginal effects after mlogit

y = Pr (fav_wine==1) (predict, p outcome (1))

=.58343415

Table 8: STATA output for the marginal effects of red wine

variable	dy/dx	Std. Err.	z	P>z	[95%	C.I.]	X	Std. Err.	z
age	.0041337	0.0279	0.15	0.882	-0.05054	0.058811	2.72656	0.0279	0.15
gender*	-.0281764	0.05196	-0.54	0.588	-0.13001	0.073661	0.481771	0.05196	-0.54
income	.039604	0.02886	1.37	0.17	-0.01697	0.096173	2.5625	0.02886	1.37
wine_yrs	.0025022	0.02738	0.09	0.927	-	0.056168	2.96615	0.02738	0.09

					0.05116				
fav_rw*	**-.159003**	**0.0751**	**-2.12**	**0.034**	**-0.3062**	**-0.01181**	**0.143229**	**0.0751**	**-2.12**
freq	.0334238	0.02549	1.31	0.19	-0.01653	0.08338	2.10417	0.02549	1.31
own_spen	.0156805	0.02528	0.62	0.535	-0.03387	0.065229	3.66406	0.02528	0.62
wine_cou	.0762584	0.05671	1.34	0.179	-0.0349	0.187414	1.32552	0.05671	1.34

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Explanation

By choosing Baronne as the favourite red wine the probability of choosing red wine as your favourite wine increases by 0.16.

MARGINAL EFFECTS OF WHITE WINE

. mfx, predict(p outcome(2))
Marginal effects after mlogit
y = Pr (fav_wine==2) (predict, p outcome (2))
= .27880149

Table 9: STATA output for the marginal effects of white wine

variable		dy/dx	Std. Err.	z	P>z	[95%C.I.]	X
Age	**-.0468805**	**0.02611**	**-1.8**	**0.073**	**-.098062 .004301**	**2.72656**	
gender*	-.0041394	0.04718	-0.09	0.93	-.096613 .088334	0.481771	
Income	-.0026755	0.02651	-0.1	0.92	-.054642 .049291	2.5625	
wine_yrs	.0174803	0.02539	0.69	0.491	-.032292 .067252	2.96615	
fav_rw*	**.160829**	**0.07434**	**2.16**	**0.031**	**.015129 .306529**	**0.143229**	
Freq	.0113528	0.02298	0.49	0.621	-.03369 .056396	2.10417	
own_spen	-.0260011	0.02314	-1.12	0.261	-.071348 .019346	3.66406	
wine_cou	-.0273002	0.05196	-0.53	0.599	-.129133 .074533	1.32552	

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Explanation

Age and favourite red wine are statistically significant determinants of the choice of white wines. The negative age coefficient suggests that younger consumers are more likely to choose white wines over red wines. If age increases, the probability of choosing white wine as the favourite wine reduces by 0.05. The positive coefficient on the favourite red wine variable suggests that consumers that choose Baronne wine as their favourite red wine are more likely to choose white wines over red wines. This implies older consumers will more likely choose red wines over white wines and consumers that choose any of the other red wines, besides Baronne, as their favourite red wine, will choose red wines over white wines.

MARGINAL EFFECTS OF SPARKLING WINE

. mfx, predict(p outcome(3))

Marginal effects after mlogit

y = Pr (fav_wine==3) (predict, p outcome (3))

= .13776436

Table 10: STATA output for the marginal effects of sparkling wine

variable	dy/dx	Std. Err.	z	P>z	[95%	C.I.]	X
age	**.0427468**	**0.01769**	**2.42**	**0.016**	**0.008079**	**0.077414**	**2.72656**
gender*	.0323158	0.03565	0.91	0.365	-0.03755	0.102179	0.481771
income	**-.0369285**	**0.01943**	**-1.9**	**0.057**	**-0.07501**	**0.00115**	**2.5625**
wine_yrs	-.0199824	0.01782	-1.12	0.262	-0.05492	0.014951	2.96615
fav_rw*	-.0018259	0.0512	-0.04	0.972	-0.10217	0.098517	0.143229
freq	**-.0447766**	**0.01907**	**-2.35**	**0.019**	**-0.08215**	**-0.0074**	**2.10417**
own_spen	.0103206	0.0163	0.63	0.527	-0.02162	0.042265	3.66406
wine_cou	-.0489582	0.03879	-1.26	0.207	-0.12499	0.027069	1.32552

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Explanation

Age, income and frequency of consumption are statistically significant determinants of consumers' choice of sparkling wines over red and white wines. The positive coefficient on the age variable suggests that the older consumers get the more likely they are to choose to sparkling wines. The negative coefficients on the income and frequency variable suggest that consumers with lower incomes and those who consume wine less often are more likely to choose sparkling wines over red and white wines. This implies that younger consumers are more likely to choose red and white wines over sparkling wines and consumers with higher incomes and those that consume wine more frequently will more likely choose red and white wines over sparkling wines.

Major findings from the STATA model

Age and favourite red wine are ineffably determinants of wine choice, income and frequency of consumption may also be determinants of the choice of white and sparkling wines over red wines. Although this model is acceptable, the low R squared brings its statistical significance into question and necessitates the use of yet another statistical package, the SPSS program.

THE SPSS MODEL

The discrete choice data was analysed using the SPSS 15.0 MNL program. The program ran different models using various attributes to ascertain the essential attributes to the model. Of

the attributes selected, two separate models (with the intercept only and with all the coefficients) were run using the same MNL analysis. The results are given in Table 11.

Table 11: Results of model log likelihood tests

Model	-2 Log Likelihood	Chi-Square	Degrees of Freedom	Significance
Intercept Only	553.3845484			
Final	469.5927298	83.79181853	48	0.001060119

The data clearly indicated that the said attributes were indeed viable and provide the best fit to the data. The null model serves as a benchmark against which we compare the fit of the final choice model and because the null model is nested in the more complete model with other wine choices, a likelihood ratio test statistic is valid. By this statistic, the coverage model provides a good fit to the data as the chi-square value of 83.79 (given in Table 12) is far greater than the critical value of -30.015 at 48 degrees of freedom.

Table 12: Model goodness-of-fit

	Chi-Square	Degrees of Freedom	Significance
Pearson	922.3862492	495	3.6212E-28
Deviance	350.4496196	495	0.999999844

The model also has acceptable Pseudo R squared values as illustrated in Table 13. This means that although the model has a relatively low explanatory power, it explains at least 20% of the wine choice preferences.

Table 13: Pseudo R-Square

Cox and Snell	0.195132209
Nagelkerke	0.227144096
McFadden	0.110783242

This model was accepted as the valid model. Table 14 provides all parameter estimates from this stage. In this study, the structural parameters are interpreted as marginal utilities with respect to each explanatory variable (Richards, 1998:19; Minbo K, 2001:5).

Table 14: SPSS output for MNL model

Coefficients	Interpretation	Estimate	Standard Error	Significance
[fav_wine = 0]	red wines	-30.015	1.812	0.00
[fav_wine = 1]	white wines	-22.574	1.318	0.00
[fav_wine = 2]	sparkling wines	-20.903	1.306	0.00

[gender=0]	females	0.439	0.222	0.05
[gender=1]	males	0.000	.	.
[own_spen=0]	R50 - R100	-0.866	2.155	0.69
[own_spen=1]	< R20	-2.319	1.216	0.06
[own_spen=2]	R21 - R35	-1.556	1.116	0.16
[own_spen=3]	R36 - R49	-2.211	1.089	0.04
[own_spen=4]	> R100	-2.507	1.085	0.02
[own_spen=5]	Do not purchase	-2.645	1.090	0.02
[own_spen=6]	Free	0.000	.	.
[fav_rw=0]	Baronne	-20.467	0.922	0.00
[fav_rw=1]	Do not drink red wine	-21.599	0.931	0.00
[fav_rw=2]	Pinotage	-21.475	0.940	0.00
[fav_rw=3]	Shiraz	-20.873	0.907	0.00
[fav_rw=4]	Rose	-21.120	0.905	0.00
[fav_rw=5]	Cabernet	-21.533	0.997	0.00
[fav_rw=6]	Red blends	-18.927	0.000	0.00
[fav_rw=7]	Merlot	-20.291	0.917	0.00
[fav_rw=8]	Cabernet Sauvignon	-20.500	0.942	0.00
[fav_rw=9]	Pinot Noir	-20.568	1.295	0.00
[wine_cou=1]	Attended wine course	0.403	0.242	0.10
[wine_cou=2]	Have not attended wine course	0.000	.	.

Link function: Logit.

Major findings from the SPSS model

The model has five main findings, on the basis of wine choice, gender, expenditure on wine for personal consumption, choice of favourite red wine and engagement in wine education.

i. Wine Choice:

The model findings assert that wine choice (in terms of red, white or sparkling) is influenced by gender, expenditure on wine for personal consumption and engagement in any form of wine education.

The null hypothesis tests that consumers prefer red wine, there are significant differences according to gender; the type of red wine preferred as well as the attendance to a wine course affects wine choice. Few other authors have empirically studied possible market segments in the wine industry. Some authors segment the market by consumption (eg. Judica & Perkins, 1992; Gluckman, 1990), by geographical region (eg. Sánchez & Gil, 1997), or consumers' behaviour (Johnson, Ringham & Jurd, 1991; Dodd, Pinkleton & Gustafson, 1996). There

have even been cases of segmentation according to commercial restraints by Johnson, Ringham and Jurd (1991) but the aforementioned authors offered little empirical background and assumed that red and white wine drinkers were mutually exclusive groups. This study asserts the same premise and the model confirms this.

ii. Gender:

The model finds that there is a positive relationship between red wine as a favourite wine and females. The significance of this attribute means that gender could be a significant segmentation attribute. It also means that there is a significant difference in wine choices according to gender and women prefer red wine more than men. This could be due to the fact that females drink wine less often and this consumption is frequently on special occasions where a glass of red wine is more preferred.

iii. Expenditure on wine for personal consumption:

The null attribute for personal expenditure is statistically insignificant. However, the negative relationship between red wine choice and expenditure for own consumption means that red wine drinkers tend to spend more on wine for personal consumption than white wine and sparkling wine drinkers. This is highly plausible given that white wines are significantly cheaper than red wines and white consumers spend less on a 750ml bottle of wine for their own consumption than red wine drinkers.

iv. Favourite red wine:

The negative relationship between red wine as a favourite wine and the choice of red wine means that Baronne wine drinkers are more likely to favour white and sparkling wines. This can be explained by the dominance and Mzansi Youth and Start-Me-Ups in this group who prefer sweeter wines.

v. Engagement in wine education:

There is a positive relationship between the choice of red wine and attendance at a wine course. This means that educated wine drinkers prefer red wine significantly more. This could be explained by the perception that with more experience one develops a taste for the drier red wine types such as the Shiraz, Merlot and Pinotage.

LIMITATIONS OF THE MODEL

The model was run in three different statistical programmes (STATA, SAS and SPSS) all of which were either statistically insignificant or had very low R squared statistics. The model described here as the accepted model had the highest of these low statistics. Possible reasons for these results could be the dominance of ordinal and discrete data which makes statistical modelling difficult. Statistical inferences were also particularly difficult due to the categorical and multi-nomial nature of the dependant variable. Another possible reason for the low statistical significance could be the inconsistencies in the respondents' responses due to their need to avoid exposing their inexperience or limited knowledge regarding wines and their reluctance to divulge personal information.

There is room for further studies which could possibly reduce the statistical insignificance of the results. In future studies, possible upgrades may include more nominal and continuous responses to the questions, as well as a wider, more diverse sample taken from various different sites, instead of focusing on a single study site. The latter will increase the possibilities of more varied and less biased responses and the former will ensure easier statistical modelling.

CONCLUSIONS

This paper has clearly shown that accurately putting a statistical and/or numerical value to qualitative variables is nearly impossible. Although the different statistical models have been made available for the determination of qualitative modelling, the different statistical packages still need more work to statistically validate these qualitative variables, as has proven to be nearly impossible in this case, The different statistical packages discussed in this paper used variants of the MNL model, but the results were significantly similar with no contradictions in their results. Despite the models' imminent statistical insignificance due to other data inconsistencies, they suggested valuable notions about black consumers' wine choice determinants. The main effects model suggests that women prefer red wine; white and sparkling wine drinkers are willing to spend less for a bottle of wine; Baronne wine drinkers prefer white and sparkling wines and educated wine drinkers prefer red wine.

In terms of the marginal effects models, with respect to red wines over the other wines, the study asserts that consumers that choose Baronne as their favourite red wine are more likely to choose white wines over red wines as their favourite wines; the older consumers get the more likely they are to choose to sparkling wines and consumers with lower incomes and those who consume wine less often are more likely to choose sparkling wines over red and white wines. In terms of white wine over the other wines, age and favourite red wine are statistically significant determinants of the choice of white wines; older consumers will more likely choose red wines over white wines and consumers that choose any of the other red wines, besides Baronne, as their favourite red wine, will choose red wines over white wines. Age, income and frequency of consumption are statistically significant determinants of consumers' choice of sparkling wines over red and white wines; the older consumers get the more likely they are to choose to sparkling wines and younger consumers are more likely to choose red and white wines over sparkling wines and consumers with higher incomes and those that consume wine more frequently will more likely choose red and white wines over sparkling wines.

In conclusion; it is interesting to note that age, gender and the choice of favourite red wine may be used to segment the market as they are often significant determinants of wine choice. The other significant coefficients affect the marketing and distribution choices to be followed by wine companies. However, although the study asserts notions about black consumers with respect to wine choice, more research needs to be undertaken and the data collection tool upgraded to ensure more reliable results. This study signals the beginning of a new era in the marketing of wine in South Africa and the world; it illustrates the need for further research in the areas of wine choice modelling and market segmentation, and the necessary statistical tools and packages, as these are indeed integral tools in identifying target markets. By understanding the local markets and providing solutions for their problems the industry is one step further towards solving global challenges through modelling and replication.

REFERENCES

BALTAS, G. & DOYLE, P. (2001). Random utility models in marketing research: a survey. *Journal of Business Research.* Vol. 51(2): 115-125. February 2001. England, UK

BARTLETT, J.E., KOTRLIK, J.W. & HIGGINS, C.C., (2001). Organizational Research: Determining Appropriate Sample Size in Survey Research. *Information Technology, Learning & Performance Journal.* Vol. 19(1):43-50. Spring 2001.

BIERLAIRE, M., (1997). *Discrete choice models.* URL: http://roso.epfl.ch/mbi/papers/discretechoice/paper.html Published by the Intelligent Transportation Systems Program, Massachusetts Institute of Technology, May 1997.

ENGELS, J.E., SARDARYAN, G. & HEBOYAN, V. (2004). *Consumer's Attitudes and Preferences for Armenian Wines.* Paper presented at the International Food and Agribusiness Management Association 14th Annual World Food and Agribusiness Forum & Symposium Montreux, Switzerland.

GIL, J.M. & SANCHEZ, M., (1997). Consumer preferences for wine attributes: a conjoint approach. *British Food Journal*. Vol. 99(1): 3-11.

LOCKSHIN, L. & HALSTEAD, L. (2005). *A Comparison of Australian and Canadian Wine Buyers Using Discrete Choice Analysis*. Paper presented at Second Annual International Wine Marketing Symposium, Sonoma State University, California, July 2005. Available on CD-ROM.

MAYEN, C. & MARSHALL, M.L. (2005). *Consumer Preferences for a Fresh-Cut Melon Product – A Potential Value Added Product for Melon Growers.* Paper presented at the International Food and Agribusiness Management Association 15th Annual World Food and Agribusiness Forum & Symposium, Chicago, USA.

MINBO K, (2001). *Qualitative and Limited Dependent Variable Models Using the New QLIM Procedure*. Paper 279-25. SAS Institute Inc., Cary, NC

MTIMET, N. & ALBISU, L.M. (2006). Spanish Wine Consumer Behavior: A Choice Experiment Approach. *Agribusiness.* Vol. 22 (3): 343–362. Special Issue on Wine Marketing. Issue Edited by Lockshin, L. & Albisu, L.M. Wiley Periodicals, Inc. Published online in Wiley InterScience DOI: 10.1002/agr.20088

NDANGA, L. (2009). *Increasing domestic consumption of South African wines: Exploring the market potential of the emerging Black middle class.* Unpublished MSc thesis.

PAZGAL, A., SEETHARAMAN, P.B. & BATSELL, R.R., (2005). *Incorporating Probabilistic Choice Rules within Random Utility Models of Brand Choice: Theory and Empirical Illustration.*

POYNTER, R. (2005). *The power of conjoint analysis and choice modelling in online surveys.* Published by Virtual Surveys

RICHARDS, T.J. (1998). *A Two Stage Model of the Demand for Specialty Crop Insurance.* URL: http://agb.east.asu.edu/workingpapers/9804.pdf Morrison School of Agribusiness and Resource Management. Faculty Working Paper Series. Arizona State University

SAWTOOTH TECHNOLOGIES, (1999). *The CBC System for Choice-Based Conjoint Analysis.* Choice-based Conjoint (CBC) Technical Paper. Technical Paper Series

TZIMITRA-KALOGIANNI, I., PAPADAKI-KLAVDIANOU, A., ALEXAKI, A. & TSAKIRIDOU, E., (1999). Wine routes in Northern Greece: consumer perceptions. *British Food Journal.* Vol. 101(11): 884-892