Quaderni di Statistica
Vol. 11, 2009

Consumer preferences and territorial certification: the case of four Val d'Orcia food products

Rossella Berni
Dipartimento di Statistica "G.Parenti", Università di Firenze
E-mail: berni@ds.unifi.it

Annalisa Romani
Dipartimento di Scienze Farmaceutiche- Laboratorio di Merceologia, Università di Firenze
E-mail: annalisa.romani@unifi.it

Riccardo Rivello
Dipartimento di Statistica "G.Parenti", Università di Firenze
E-mail: rivello@ds.unifi.it

Summary: This paper deals with the application of binary response choice models to study four food products of the Val d'Orcia area. It aims at addressing two main issues: i) the analysis of food products of high quality by planning a combined study of choice and conjoint designs in the presence of certification criteria and quality characteristics; (ii) the application and the comparison of the conditional logit, the mixed logit and the heteroschedastic extreme value models, in order to study the consumer preferences taking into consideration both the heterogeneity of respondents and the heteroschedasticity of the error term.

Keywords: Choice Modeling, Conjoint Analysis, Mixed Logit, Heteroschedastic Extreme Value Model

1. Introduction

Preference measurements are considered one of the main methods to study and improve consumer behavior, defined as the consumer's decision

to improve his/her utility by choosing a service or a product. In particular, the preference theory must be evaluated according to the nature and definition of preference, namely revealed or stated preferences; in the latter case we may distinguish between contingent valuation, conjoint analysis and choice modeling (for an articulated review see Netzer et al., 2008). Since the fundamental elements of distinction are positively overlapped or interchangeable, attention must be exerted to the fact that considering conjoint analysis and choice modeling the classification is not so clear–cut; therefore, these methods are generally defined as multi-attribute valuation methods.
However, the preference measurements are usually related to a new product and the main distinction between conjoint analysis and choice modeling is the monetary evaluation. To this end, we refer to the willingness to pay, i.e. the quantitative expression of the respondents about their willingness to accept a change in the product concerned or in a single attribute.
Although the corresponding statistical models have long been developed, the last developments in this field were mainly directed at improving features common to all models, such as the heterogeneity of respondents or the heteroschedasticity and the complexity of the alternatives (profiles); for a recent review and an updated bibliography see Berni and Rivello (2009).
More precisely, in recent years, the definition of the Mixed Multinomial Logit has involved the evaluation of the respondent's heterogeneity in the choice modeling context (Train, 1998; McFadden and Train, 2000; Hensher and Greene, 2003); on the other hand, the Heteroschedastic Extreme Value model (Bhat, 1995) was introduced to study the heteroschedasticity of the error term. When considering conjoint analysis, specific statistical models, such as the hierarchical Bayes models of Lenk et al. (1996), were defined in order to take into account the respondents' heterogeneity.

By considering the marketing and economic point of view, the importance of origin trademark has been analyzed in several ways. The main issues may be outlined by discriminating between: i) the use of Protected Geographical Indication (henceforth P.G.I.) and/or Protected Designation of Origin (P.D.O.) certification as a brand; Fotopoulos and Krystallys (2001) deal with the real worth of certification and its importance through customer's consideration by analyzing the willingness to pay for certified food product; ii) the consideration of possible ethnocentric influences in preferring some products (Orth et al., 2003; Stefani et al., 2006; van der Lans et al., 2001). Orth et al. (2003) consider the influence of the ethnocentrism as an auxiliary information for predicting utilities of product evaluation, while van der Lans et al. (2001) explore the influence of ethnocentrism in preferring products originating from the same area of living.

Besides, products with an origin trademark have a higher possibility to be purchased when they meet a customer's idea of the place of origin, identified through history, tradition and territory (van Ittersum et al., 2003). In addition, the influence of the origin trademark is significant when the purchaser lives close to the area of production (van der Lans et al., 2001) and when this is a small and well-defined area (Stefani et al., 2006).

Within the framework of conjoint analysis and choice modeling, we synthesize our theoretical and empirical contributions as follows:

1. we carry out a combined study of conjoint and choice designs through several survey-steps in order to evaluate a set of quality Val d'Orcia (V.O.) local food products;

2. the evaluation of these four V.O. food products takes into account both the main territorial peculiarities, considering P.D.O. and P.G.I. quality protocols, the certification criteria and the consumer preferences; in addition, an ad-hoc questionnaire for each product is supplied to respondents together with a panel-test;

3. in order to deal with the two points mentioned above, discrete choice models with a binary response variable are applied and compared; in particular, the simple Conditional Logit model is compared with the Mixed Multinomial Logit model (Mixed MNL) and the Heteroschedastic Extreme Value model (HEV), by considering the relaxation of the Independence of the Irrelevant Alternatives (IIA) assumption, the respondents heterogeneity and the heteroschedasticity of the error term.

The organization of the paper is the following: in Section 2 the fundamental elements of the utility theory are briefly outlined; in Section 3 the theory behind the three discrete choice models considered is explained; Section 4 includes the case study, the material and methods, the data and variables description; the model results and the discussion are reported in Section 5; final remarks follow.

2. A general framework for utility modeling

In order to define the discrete choice models applied in this paper, we briefly introduce the fundamental elements of the utility theory (Haab and Hicks, 1997).

As first step, the class of Random Utility Models (RUM) is defined. In general, every alternative is indicated by j ($j = 1, ..., J$), while i denotes the consumer/user $(i = 1, ..., I)$. Each alternative will be characterized by a vector of characteristics; in what follows, P.G.I., P.D.O., certification criteria, organic and typical features. Thus, the following expression is characterized by a stochastic utility index U_{ij}, which may be expressed, for each unit i, as:

$$U_{ij} = V_{ij} + \epsilon_{ij} \tag{1}$$

where V_{ij} is the deterministic part of utility, while ϵ_{ij} is the random component. The random component is in general supposed independent and Gumbel or type I extreme value distributed. In the following formulas, (2) and (3), the probability density function and the cumulative distribution function (CDF) of the Gumbel distribution are defined:

$$\lambda(\frac{\epsilon_{ij}}{\theta_j}) = exp^{-\frac{\epsilon_{ij}}{\theta_j}} exp^{-exp^{-\frac{\epsilon_{ij}}{\theta_j}}} \tag{2}$$

$$\Lambda_{ij}(\frac{\epsilon_{ij}}{\theta_j}) = exp(-exp(-\epsilon_{ij}/\theta_j) \tag{3}$$

where θ_j is the scale parameter related to the j-th alternative.

In the RUM, the individual is assumed to choose the alternative j that gives the highest level of utility, where the alternative j belongs to the choice-set C. Let the individual's indirect utility function for the alternative j be represented by:

$$U_j(q_j, y - p_j, \epsilon_j) = V_j(q_j, y - p_j) + \epsilon_j \tag{4}$$

From the researcher's perspective, the indirect utility function has two components. The first, $V_j(q_j, y - p_j)$ represents the observable portion of the individual's indirect utility function, with vector of quality characteristics q_j, income y, and price of the single product p_j. The second component of indirect utility is ϵ_j, the unobservable part of the individual's indirect utility function.
For a given choice occasion, the individual will choose the alternative j if:

$$V_j(q_j, y - p_j) + \epsilon_j \geq V_k(q_k, y - p_k) + \epsilon_k; j \in C, \forall k \in C. \tag{5}$$

Note that, just because a part of the indirect utility function is not observable, indirect utility must be expressed by:

$$v(q, y - p) = E[\max\{V_k(q_k, y - p_k) + \epsilon_k; \forall k \in C\}] \tag{6}$$

where the expectation of the right-hand side of (6) is the researcher's expectation across the random unobservable portion of the individual's utility function. Therefore, the probability of an individual i choosing the product according to the j alternative is modelled as:

$$P_i(j) = P(j|k \in C, w_i) \tag{7}$$

where w_i represents a vector of individual specific characteristic.

For the purposes of the subsequent analysis we can consider the Multinomial Logit model, which can be seen as the basic model for the conditional logit described in the next section; this probability can be written as:

$$P(y_i = j) = P_i(j) = P(j|k \in C, w_i) = \frac{exp^{v_j}}{\sum_{k \in C} exp^{v_k}} \tag{8}$$

3. The discrete choice models

In order to define the discrete choice models applied and discussed in this paper, we briefly introduce the fundamental elements of the related theory; for further details see the previous cited references (Sect.1).

The class of RUM, which aims to achieve the utility maximization for the consumer, enlarges the characteristics of the Logit and Multinomial models where the IIA property is hypothesized. The relaxation of this assumption (Train, 1998) is a relevant improvement because the IIA means that the choosing probability in one choice-set is independent of the presence of other attribute values or any other alternative; on the other hand, we may say that IIA derives from the hypothesis of independence and homoschedasticity of the error terms. In addition, this can also be interpreted by considering the cross-elasticity term. In fact, IIA implies an equal proportional substitution between alternatives, (Scarpa et al., 2007). Furthermore, the Logit and Multinomial models do not allow to evaluate a different behavior of the consumer; i.e. each respondent, with different baseline characteristics, is treated in a similar way (the same estimate values of attributes) according only to his/her judgement.
In order to deal with the above issues, the statistical analysis is carried out through three discrete choice models belonging to this class, and, in particular, through the conditional logit model, the mixed logit and the HEV model (Bhat, 1995); note that, in this case, the discrete choice models are not multinomial because the response variable is the binary choice.

The conditional logit model is the simple model which can be defined as the Multinomial Logit Model (MNL) when choice specific data are available and the choice-set is formed by two alternatives.

The term ”conditional” highlights that the unit i chooses the alternative j, which belongs to a set of alternatives called choice-set C_i and then the model applied is called Conditional Logit (CL). Thus, the probability of the unit i to choose the alternative j is defined as:

$$P(y_i = j) = P_{ij} = \frac{exp(x^{'}_{ij}\beta)}{\sum_{k \in C_i} exp(x^{'}_{ik}\beta)} \qquad (9)$$

where x_{ij} denotes the value of the attribute for the alternative j and the unit i. Note that, the difference is expressed through the J values of the random variable y, which indicates the choice made from the unit i. The CL model is the basic discrete choice model applied in this paper and we remark that this model assumes the IIA property; in addition, in this case, the error term is distributed according to formula (3) without the evaluation of the scale parameter θ_j, i.e. the error terms are supposed identically distributed.

When a Mixed MNL model is considered, the general expression for a RUM model becomes:

$$U_{ij} = V_{ij} + \psi_{ij} + \epsilon_{ij} \tag{10}$$

The main feature of the Mixed MNL model, or of the Mixed logit model when the choice is binary, is the possibility to assume a general continuous distribution for the ψ_{ij} called also mixing term. In fact, a density for ψ_{ij} is defined as in the following:

$$g(\psi \mid \Phi) \tag{11}$$

where the space parameter Φ contains the fixed parameters of the distribution, such as Normal, Uniform, Log-Normal. If ψ is not evaluated, then the mixed logit reduces to the simple conditional logit; in general, the unconditional probability is equal to:

$$P(y_i = j) = P_i(j) = \int_{\psi} L_i(j \mid \psi_{ij}) g(\psi_{ij} \mid \phi) d\psi_{ij} \tag{12}$$

$$L_i(j \mid \psi_{ij}) = \frac{exp(x'_{ij}\beta + \psi_{ij})}{\sum_{k \in C_i} exp(x'_{ik}\beta + \psi_{ik})}$$

Note that the unconditional choice probability $P_i(j)$ is the integral of the conditional probability of the logit model integrated over the distribution of $\psi_{ij}, \forall i, j$ and weighted according to the fixed parameters of the mixing term. Therefore, the mixed logit model allows to treat the heterogeneity of respondents through the random parameters associated to a specific attribute of an alternative. Nevertheless, the error term across alternative in not weighted, as in the following model.

The Heteroschedastic Extreme Value (HEV) model (Bhat, 1995) is the third discrete choice model applied in this paper and belongs to the RUM class as defined in formula (1). The main feature of this model, which differentiates it by the CL model and the Mixed Logit, concerns the modified assumptions on the random component. In this model, the random component, supposed distributed as a type I extreme value distribution, formula (3), is assumed independently but not identically distributed.

This different hypothesis on the random component allows us to treat differently the relaxation on the IIA property with respect to the Mixed Logit model, because, in the HEV model, the homoschedasticity hypothesis of the error terms is not assumed and, therefore, different scale parameters across alternatives are estimated. This last consideration implies that cross-elasticities are not supposed to be all equals, as in the MNL and the logit models.

The main evident advantage is that the scale parameters may be defined as the weights in order to measure the uncertainty related to the alternatives and to the attributed there involved. Furthermore, the presence of large variances for the error terms influences the effects of changing of systematic utility for the generical alternative j.

Therefore, the probability for a respondent i to choose the alternative j from a choice-set C_i is:

$$P(y_i = j) = P_i(j) = \int_{\epsilon} \prod_{k \in C_i; k \neq j} \Lambda\{\frac{x'_{ij}\beta - x'_{ik}\beta + \epsilon_{ij}}{\theta_k}\} \frac{1}{\theta_j} \lambda(\frac{\epsilon_{ij}}{\theta_j}) d\epsilon_{ij} \quad (13)$$

where θ_j is the scale parameter for the j alternative and $\lambda(\cdot)$ is the probability density function of the Gumbel distribution, as in formula (2); the term $x'_{(\cdot)}\beta$ denotes the deterministic part of utility of formula (1). Note that the integral function is defined on the domain $[-\infty, +\infty]$ of the random component ϵ related to the unit i and the alternative j.

The theoretical framework of these three discrete choice models allows us to outline useful comparisons in the following case study (Sect.4 and Sect. 5), when consumer preferences are evaluated. Furthermore, the CL model is seen as the basic and simple model which does not take care of respondent's heterogeneity due to baseline variables (such as gender and age); thus, heterogeneity is modelled in the Mixed Logit through the mixing term, $g(\psi \mid \Phi)$, where the expressed preference of respondent i, $(L_i(j))$, is measured conditioning to the personnel characteristics (ψ_{ij}).

The HEV model is considered as a further and different improvement to the CL model with respect to the Mixed Logit model. In this case the consumer preferences of respondent i are evaluated by considering a scaling term θ_j for the alternative j in the choice-set C_i, i.e. the heteroschedasticity of the error term.

It's not straightforward matter to say that the HEV and the Mixed Logit models could be considered as competitive models in order to identify and to measure the presence of an over-dispersion when modeling the consumer preferences, with respect to the CL model.

In what follows, Sect. 5, the consumer preferences are evaluated by considering respondent's heterogeneity or the heteroschedasticity of the alternatives.

The discrete choice models are evaluated through the following goodness-of-fit criteria: the maximum gradient element, the number of iterations to reach convergence, the Likelihood Ratio (LR), the Akaike's index (AIC) and the McFadden's LR index (McFadden LRI), bounded in $[0, 1]$, which is defined as the complementary to one of the LR.

4. The case study

The aim of this study is the evaluation of four food local products: the pasta V.O., in particular the production protocol related to the "pici", the "cinta senese" ham, the "cinta senese" capocollo and the Tuscan extra virgin olive oil.

More precisely, the Research Foundation of Monte dei Paschi di Siena financed a project named "Increase in value of typical products of Siena: a study of characteristics and flavors in consumer preferences". Into this framework, we evaluated four different kinds of typical food products from Siena. These products could be inserted in an ideal consumer set in order to characterize local or typical products as: "pici" from organic production, P.G.I. "cinta senese" products and Tuscan P.D.O. extra virgin olive oil. In addition, a further issue of this project aims at warranting their protection of quality, their authenticity and their healthy properties, creating a policy of territorial marketing. In the following sections (Sect.4.1 and Sect.4.2), materials and methods of the panel-tests and the description of the *data set* are outlined.

4.1. Panel tests: materials and methods

Pastificio Toscano, located in S. Quirico d'Orcia, produces "pici" using organic semolina from *durum wheat* grown in Val d'Orcia, a geographical area with an ancient tradition for best quality food products. The company has an integrated quality system of certification (organic, ISO:9001, ISO:14001) with ethic SA:8000 certification. Founded in 2003, it uses a traditional production process: the bronze drawing and a 24 hours drying.

The Tuscany or certified P.D.O. extra virgin olive oil is produced by the company Olivicoltori Toscani Associati (OTA), a cooperative of Tuscan olive growers, with 23,000 olive companies associated. The product considered in this research was an extra virgin olive oil made in 2007.

Pork "cinta senese" products were from "Podere Bioamiata". Cinta senese is a breed of free range pigs characterized by a variously pigmented (black down and a white belt all over the chest and the front legs) skin, subject to a specific P.G.I. protocol. Podere Bioamiata is located at above 700 a.s.l., between the Val d'Orcia and Monte Amiata.

For each product, a specific panel-test evaluation form, with sensory descriptors, such as visual taste and olfactory, were optimized. More precisely, sensory descriptors of the pork products are considered according to the production protocol.

4.2. The data

Conjoint and choice based designs are carried out during several survey-steps during the period July 2007-March 2008; for each single product, an ad-hoc questionnaire is supplied to the respondents together with the panel-test. More precisely, for the extra virgin olive oil a choice-experiment is planned, while conjoint analysis is carried out for pasta V.O. and ”salumi”. Subsequently, the combined study of choice and conjoint exercises are matched in order to evaluate a global consumer set of quality local food products, as explained in Sect.4.

Note that, in our case study, the price is directly involved only in the oil survey. In the other surveys, related to pasta V.O. and ”salumi”, we include only a qualitative evaluation of the price (high, low). This decision has been related to the impossibility of creating just one price for the consumer-set of the local food products, given that each product is separately evaluated through: a panel-test plus a questionnaire plus a choice or conjoint experiment.

As regards each single product, in every questionnaire a specific section was inserted in order to measure the consumer’s preferences according to the organic, the certification criteria and the territorial certification. We define ”territorial certification” as the combination of several features, i.e. packaging, place of purchase, kind of retailer, connection with traditional products and environmental preservation.

In addition, information related to the baseline variables of respondents, such as gender, age, social-economic situation and the number of family components, are collected.

Furthermore, each questionnaire has three different parts: two of these are in common, the other one, which is dedicated to the choice or conjoint step, differs from survey to survey, as mentioned above. The first part of the questionnaire encloses questions related to the social and demographic status. The second part aims at evaluating the respondent’s interest to the specific product; in particular, questions about the usage and the knowledge of this kind of foods are included: ”How often are you buying typical food?”, ”In your opinion, is it important the territorial certification?”, ”Is it important organic certification for you?”, ”Are you interested in preserving typical food?”. In addition, a four-question-box is also included in this second part: each respondent has to give an evaluation, on a five points metric scale, with respect to some product’s characteristics. The evaluation is given according to the importance that each characteristic has during the pre-purchasing step. The features here con-

sidered are: production in a specific geographical area, process/product certification, purchasing place, price. We finally compute an indicator, called ”involving”, defined as the average of this 4 judgements.
People participating to the surveys were volunteers selected randomly, without any rule. In the oil survey, 102 people were involved and the location of the meeting was the OTA’s site; in the pasta survey, 64 individuals were interviewed during a meeting in the Pastificio Toscano site, while 91 individuals participated to the ham and ”capocollo” surveys and they were interviewed during several panel-tests. The questionnaire has been filled by each volunteer jointly with a panel-test. Finally 257 questionnaires were filled.

4.3. Matching and bootstrap procedures

A statistical matching procedure suggested by Rubin (1979) is carried out by including all the *data sets* collected during the taste occasions, as mentioned in Sect.4.2. The applied method is based on the identification of a set of common variables present in all the *data sets*, in order to match the individuals with respect to all the variables. For example, by considering two *data sets* with one or more common variables, the aim is to obtain a unique *data set* containing all the variables of the two original archives (common and non common). The matching procedure is then performed through a linear model; therefore, we must define which is the dependent variable and which are the independent ones, among all the common variables. In our case study, the common variables are: the indicator mentioned above as ”involving”, used as the dependent variable; while social and demographic characteristics are used as independent variables.

The matching step allows us to obtain a sample of 47 matched individuals. Subsequently, the bootstrap is carried out in order to replicate the *data set* obtained through the matching procedure; the final *data set* consists of 184 respondents who have expressed their preferences for all the four food products and their propensity to buy territorial food products of high quality and specific origin. Therefore, the global *data set* is composed of 4224 observations.

The MULTTEST and MDC procedures of the SAS (Statistical Analysis System) for the bootstrap and for the application of the discrete choice models, respectively, are applied.

4.4. The variables

The variables involved in the application (Sect.5) are: the attributes present in the alternatives, such as organic, P.D.O. and territorial certification; gender and age. Dummy variables are built in order to measure the associations between the products, the attributes and the respondents' characteristics. In Table 1 a description is summarized; all the attributes are binaries.

5. Model results

The results related to the applied statistical models are shown in Tables 2-11. In each Table, the estimates with the standard errors (s.e.) and the *p*-values are shown. In addition, specific ratios of the coefficients are computed (Tab.12); diagnostic measures, in particular the AIC and the McFadden LRI, are inserted at the bottom of the tables, for each model. Note that, when estimating the HEV models, the scale parameter θ_j of formula (2) is denoted as "scale-p." in the corresponding Tables 5, 6, 9, 10.

As shown in Section 2, three different discrete choice models are carried out in this case study. The framework of the analysis is the following:

1. Four general models are estimated in order to evaluate the organic and the P.D.O. certification criteria and the product's features, i.e. the territorial certification (as described in Sect. 4.2) for all the four food products (consumer-set), together with the baseline variables, such as gender and age of respondents. The response variable for these four models (Tables 2-5) is the expressed preference by discriminating between two alternatives ($J = 2$ in each choice-set) for all the products.

Table 1. Attributes' and dummies' description

Acronym	meaning
organic	organic
PDO	P.D.O.: Protected Designation of Origin
e. v. o. oil	extra virgin olive oil
capo	capocollo
ham	ham
pasta	pici
tc	territorial certification
age-PDO	1 if age $\leq$ 35 and P.D.O.=1; 0 otherwise
age-org	1 if age $\leq$ 35 and org=1; 0 otherwise
age-tc	1 if age $\leq$ 35 and tc=1; 0 otherwise
fe-PDO	1 if female and P.D.O.=1; 0 otherwise
fe-org	1 if female and org=1; 0 otherwise
fe-tc	1 if female and tc=1; 0 otherwise
oil-PDO	1 if oil and P.D.O.=1; 0 otherwise
oil-org	1 if oil and org=1; 0 otherwise
capo-PDO	1 if capo and P.D.O.=1; 0 otherwise
capo-org	1 if capo and org=1; 0 otherwise
ham-PDO	1 if ham and P.D.O.=1; 0 otherwise
ham-org	1 if ham and org=1; 0 otherwise
pasta-PDO	1 if pasta and P.D.O.=1; 0 otherwise

2. a specific model for each product is estimated and the results are shown in Tables 6-11; in this case, for each estimated model, the response variable is the binary choice related to the supplied choice-sets for that product.

3. within point no.*2.*, a further analysis is carried out in order to compare different discrete choice models, such as Mixed Logit and HEV model, for the same food product (Tables 8-11).

In general, all the models achieve satisfactory results; for each model applied, the diagnostic measures are considered. In the heteroschedastic extreme value model, the correlation matrix of the estimates is evaluated

in order to measure the relation of the scale parameter with respect to the estimated coefficients. By considering the mixed logit, the diagnostic measures related to the mixing term and the simulations setting are checked.
The general conditional logit shown in Table 2 could be named ”product’s characteristics and baseline variables”. The results do not reveal heteroschedasticity or respondents heterogeneity and a conditional logit is applied. A general equilibrium is obtained between the organic and the certification criteria (Tab.12); furthermore, a product is preferred 3.7 times if it has the organic and/or the certification features; this last consideration is confirmed also by evaluating the choice probability for the organic (p=0.57) or the P.D.O. feature (p=0.57) and even when the choice probability is computed for a female and the P.D.O. characteristic (p=0.27).

Table 2. Products’ characteristics and baseline variables-CL model

Coefficient	estimate	s.e.	p-value	exp(coef)
organic	1.314	0.204	0.0001	3.720
PDO	1.315	0.062	0.0001	3.726
age-org	–0.298	0.118	0.0115	0.742
fe-org	–1.324	0.186	0.0001	0.266
McFadden LRI=0.230			AIC=1994	

A connected model is the second one (Tab.3), where the organic and the P.D.O. criteria are evaluated for each product with gender and age. The significant variables reveal a not great preference toward ”pici” (the corresponding coefficient is not significant). A positive judgement is related to ”capocollo” and the organic e. v. o. oil. Negative coefficients are obtained for the ham. Furthermore, a very positive preference for the P.D.O. certification is obtained when the respondent is a female or when the respondent is young (less then 35 years); in fact, the odd ratios are 4.14 and 1.86, respectively. This result is further confirmed by the computation of the choice probability: for example, for the e. v. o. oil, the probability is equal to 0.60 if the P.D.O. is present in the alternative jointly with a young and a female respondent.

The last general model applied (Tab.4) is related to the territorial cer-

Table 3. P.D.O. and organic certifications-CL model

Coefficient	estimate	s.e.	p-value	exp(coef)
oil-org	0.214	0.122	0.0797	1.238
oil-PDO	–0.663	0.155	0.0001	0.515
capo-org	–0.318	0.121	0.0085	1.375
ham-org	–0.293	0.936	0.0017	0.746
ham-PDO	–0.716	0.127	0.0001	0.488
age-PDO	0.620	0.107	0.0001	1.859
fe-PDO	1.420	0.099	0.0001	4.136
McFadden LRI=0.255		AIC=1717		

tification of the food products. There is a marked preference toward the typical feature if the respondent is young and/or female. However, undoubtedly, it is confirmed the opposite result for age and organic, because the coefficient is negative. Nevertheless, in general, there is a high coefficient for the organic feature and the choice probability when only this attribute is present is p=0.54, greater than the P.D.O. probability, equal to 0.27. The greater value of the coefficient for the organic attribute in this model with respect to the same parameter of the first model (Tab.12) allows us to conclude that the general preference versus the organic characteristic becomes more relevant when the territorial certification is present.

Table 4. Products' criteria and the territorial certification-CL model

Coefficient	estimate	s.e.	p-value	exp(coef)
organic	3.115	0.372	0.0001	22.538
PDO	1.985	0.113	0.0001	7.285
age-org	–0.692	0.192	0.0003	0.500
fe-org	–2.677	0.359	0.0001	0.069
age-tc	0.404	0.181	0.0252	1.499
fe-tc	0.832	0.149	0.0001	2.299
McFadden LRI=0.442		AIC=921		

The fourth applied model (Tab.5) is the heteroschedastic extreme value model (HEV model), where the specific P.D.O. criterium is investigated in relation with the specific products. The coefficients are all positive, except for the P.D.O. preference evaluated with the age of the respondent. Therefore, there is a general attention toward the P.D.O. characteristic, above all when the products are the e. v. o. oil and the ”capocollo”, which is the most preferred product when the P.D.O. criterium is considered. This result is evident by considering the odd ratios and the the ratios of the coefficients; for example, (Tab.5), the odd ratio for an e. v. o. oil and P.D.O. is about 2.97 and the odd ratio for a ”capocollo” P.D.O. is about 5.57.

Table 5. The P.D.O. through a heteroschedastic extreme-value model

Coefficient	estimate	s.e.	p-value	exp(coef)
oil-PDO	1.088	0.246	0.0001	2.969
capo-PDO	1.716	0.311	0.0001	5.567
ham-PDO	0.799	0.203	0.0001	2.224
pasta-PDO	0.752	0.228	0.0010	2.121
age-PDO	–0.433	0.138	0.0017	0.648
fe-PDO	0.259	0.170	0.1252	1.296
scale-p.	0.926	0.153	0.0001	-
McFadden LRI= n.a.			AIC=2077	

The results obtained through the previous models are further confirmed by the following results, where each product is separately evaluated.

In Table 6, the discrete choice model for ”pici” is illustrated. In this case, the best fitted model is carried out through a HEV model, where the coefficient for ”pici” is negative. We remark that the heteroschedasticity in this case is very relevant, jointly with the respondents’ variables. Surely, some problems related to the available data have a direct effect on the ”pasta” coefficient, which is negatively influenced by the age and the P.D.O. criterium. The e. v. o. oil (Tab.7) does not present heteroschedasticity or respondents heterogeneity, therefore a conditional logit is applied. This result confirms a general preference of the respondent toward

Table 6. The pasta Val d'Orcia-HEV model

Coefficient	estimate	s.e.	p-value	exp(coef)
pasta-PDO	–0.171	0.116	0.1400	0.843
fe-PDO	1.025	0.061	0.0001	2.788
scale-p.	1.664	0.281	0.0001	-
McFadden LRI= n.a.		AIC=2024		

the organic oil and, in addition, a clear decision when the protocol of the product exists, above all when the respondent is female and/or young. In fact, the choice probability for an organic e. v. o. oil is equal to 0.22, while the choice probability for a P.D.O. e. v. o. oil is 0.14. This last result notably increases when we consider the choice probability for a P.D.O. e. v. o. oil when the respondent is a young female (p=0.52). This last result is due to the estimate of the coefficient ”fe-PDO” (odd ratio = 3.14). A further analysis on this product, by inserting in the conditional logit model the price together with P.D.O. and organic features, has not given a significant value for price.

Table 7. The extra-virgin olive oil-CL model

Coefficient	estimate	s.e.	p-value	exp(coef)
oil-PDO	–0.311	0.139	0.0248	0.732
oil-org	0.212	0.121	0.0793	1.237
age-PDO	0.440	0.101	0.0001	1.553
fe-PDO	1.142	0.083	0.0001	3.136
McFadden LRI=0.234		AIC=1759		

The last four tables are related to ”salumi”of ”cinta senese”: ham and ”capocollo” (Sect. 4.1). As regards these two food products, the results are also interesting by a theoretical point of view. The first model (Tab.8) is a simple conditional logit which analyzes both products. The results confirm a general attitude of female versus the P.D.O. criterium, with choice probability versus ”salumi” and P.D.O. equal to 0.40, and a specific preference versus ”capocollo” with the organic and P.D.O. characteristics. As to the ham, this product presents a negative coefficient when associated with the P.D.O. criterium and also when compared with the same coefficient of ”capocollo” (ratio ≈ -1.20). This last result is further

Table 8. The conditional logit for ”salumi”

Coefficient	estimate	s.e.	p-value	exp(coef)
capo-PDO	0.443	0.157	0.0048	1.557
capo-org	0.340	0.126	0.0070	1.405
ham-PDO	–0.368	0.125	0.0033	0.692
age-PDO	0.430	0.101	0.0001	1.538
fe-PDO	1.101	0.104	0.0001	3.006
McFadden LRI=0.246		AIC=1735		

confirmed in the specific ham model (Tab.9), where the HEV model is applied. The computed heteroschedasticity highlights a negative respondent’s preference versus the organic and P.D.O. criteria when the product is the ham, even though the P.D.O. characteristic is just a little preferred. The scale parameter is positively correlated with the ”ham-PDO” coefficient (0.69) and inversely correlated with the coefficient related to the association between the gender and P.D.O. (-0.73).

Table 9. The heteroschedastic extreme-value model for ham

Coefficient	estimate	s.e.	p-value	exp(coef)
ham-PDO	–0.269	0.135	0.0462	0.764
ham-org	–0.425	0.096	0.0001	0.654
age-PDO	0.491	0.082	0.0001	1.633
fe-PDO	1.013	0.104	0.0001	2.753
fe-org	0.222	0.064	0.0005	1.248
scale-p.	1.744	0.447	0.0001	-
McFadden LRI= n.a.		AIC=1728		

Two models for ”capocollo” are carried out: the mixed logit (Tab.11) and the HEV model (Tab.10). The application of both models confirms the presence of heteroschedasticity and respondents heterogeneity for this product. Undoubtedly, both models capture the variability due to the respondents and evaluated through the ”age-PDO” coefficient, which verifies the association between age and P.D.O. In addition, in the HEV model, the scale parameter measures the variability of the P.D.O. certification by decreasing the coefficients linked to this last one; the correlation between the scale parameter and the ”capo-PDO” coefficient is -0.64, while lower correlations are obtained with the ”age-PDO” coefficient (0.27) and with the coefficient of the association between the gender and P.D.O. (-0.22).

Table 10. The heteroschedastic extreme-value model for ”capocollo”

Coefficient	estimate	s.e.	p-value	exp(coef)
capo-PDO	0.381	0.156	0.0164	1.463
capo-org	0.337	0.105	0.0014	1.400
age-PDO	0.421	0.089	0.0001	1.523
fe-PDO	0.878	0.078	0.0001	2.407
scale-p.	1.530	0.317	0.0001	-
McFadden LRI= n.a.			AIC=1739	

In the mixed logit (Tab.11), the Normal distribution is assumed with the estimation of two coefficients for mean and standard deviation; in this case, the mixing term is uniquely composed by the ”age-PDO” coefficient. The dispersion parameter is highly significant and there is a further confirmation of the respondent’s preference versus a ”capocollo” with the P.D.O. characteristic.

Table 11. The mixed-logit for ”capocollo”

Coefficient	estimate	s.e.	p-value	exp(coef)
capo-PDO	0.699	0.182	0.0001	2.013
capo-org	0.440	0.164	0.0074	1.553
age-PDO-M	1.308	0.574	0.0226	3.700
age-PDO-S	2.507	0.915	0.0061	-
fe-PDO	1.048	0.102	0.0001	2.853
McFadden LRI=0.245			AIC=1735	

Table 12. Ratios of the coefficients

Model	Ratio	value
Model of Table (2)	org/PDO	0.999
Model of Table (3)	oil-org/oil-PDO	–0.323
	ham-org/ham-PDO	0.409
Model of Table (4)	bio/PDO	1.569
	eta-org/eta-tc	–1.710
	fe-org/fe-tc	–3.217
Model of Table (5)	e. v. o. oil/capo	0.633
	e. v. o. oil/ham	1.361
	e. v. o. oil/pasta	1.448
	capo/ham	2.147
	capo/pasta	2.283
	ham/pasta	1.063
Model of Table (7)	oil-PDO/oil-org	–1.468
Model of Table (8)	capo-org/capo-PDO	1.302
	capo-PDO/ham-PDO	–1.204
Model of Table (9)	ham-PDO/ham-org	0.634
	ham-org/ham-PDO	1.577
	fe-PDO/fe-org	4.568
	fe-org/fe-PDO	0.219
Model of Table (10)	capo-PDO/capo-org	1.130
	capo-org/capo-PDO	0.885
Model of Table (11)	capo-PDO/capo-org	1.589
	capo-org/capo-PDO	0.630

5.1. Discussion

Statistical analysis reported in Section 5 allows us to draw some implications related to the general features as well as the single products.

We must point out that our aim is to establish the main characteristics for a consumer-set of quality food products with respect to the consumer's preference.

By considering the four general estimated models (Tab. 2-5), an equilibrium between organic and P.D.O. certification emerges (Tab.2); however, (Tab.4), when these two features are evaluated jointly with the consumer's variables and the "territorial certification" (tc), the organic attribute acquires more importance. The explanation of this result could be due to the indirect effect of "tc" on the organic certification, if we consider that "tc" is built through local and product characteristics (Sect. 4.2), and the organic certification is naturally included into the territory. In addition, this result can be confirmed by the positive relevance of "tc" with gender and age.

On the other hand, organic certification clearly prevails over P.D.O. as regards the e. v. o. oil, when considering the result reported in Table 3, while the same can not be highlighted for the "salumi".

The result of Table 3 obtained for the extra v.o. oil is confirmed in the specific estimated model (Tab.7), where organic oil is preferred. The controversial result on "salumi" could be explained by considering the following two remarks:

1. the lack of general knowledge of the "capocollo" product; this implies that the preference versus "capocollo" is mainly influenced by a high variability on the respondents, as confirmed through the Mixed Logit (Tab. 10);

2. the preference versus a P.D.O. ham in the general model of P.D.O. certification (Tab. 5) is slightly confirmed in the HEV model of Table 9, in comparison with the "ham-org" coefficient.

Nevertheless we must also remark that: i) females give always their positive preference versus P.D.O. certification, and ii) in general, an organic product is less expensive than a P.D.O. product.

Therefore, when we concentrate our attention to the P.D.O. characteristic (Tab. 5), positive results are obtained, except for the coefficient "age-PDO", which confirms that a non-young consumer prefers the organic feature. At the same time, when we compare, for each single product, the organic and P.D.O. attributes, the organic one is preferred by the consumer if this feature is deep-routed on the territory, as for the extra v.o. oil.

A last remark is related to the pasta V.O. This product does not achieve good results; the positive coefficient for a P.D.O. pasta (Tab. 5) is not confirmed in the specific HEV model (Tab. 6). This result can be surely explained through the following observations:

1. a generic consumer does not associate pasta with a certified food product;

2. pasta is a staple product and the price of the certified pasta ”pici” is very expensive relative to the corresponding commercial product;

3. at the moment, there is a total lack of a quality protocol.

6. Final remarks

Companies involved in the project could create a network in order to enhance the quality of traced products originating from V.O., through the identification of specific markers and the optimization of panel-test questionnaires. This network could help companies both in promoting and in spreading their products in order to induce consumers to prefer products from V.O. The purpose is to bring local products close to the market, addressing both sensory analysis and quality of the products themselves.

The joint study of the four food products of the V.O. area has achieved good and satisfactory results, both by considering the application of discrete choice models and their theoretical comparison and the results about the specific food products investigated during the four survey steps. Furthermore, the food products are studied through discrete choice models by applying three general models for all the products jointly, and through a single model for each product. Generally speaking, the certification and organic criteria received the positive preference by the respondents and, in addition, the differentiations related to the gender and to the age are highlighted. Furthermore, the heteroschedaticity and the heterogeneity of the respondents is verified through the heteroschedastic extreme value model and the mixed logit model respectively, in order to measure the real preference of the respondent versus the certification criteria and the variability due to the presence of large variances for the error terms.

Surely, we obtain some indications for the future decisions with respect to the certification criteria and the organic quality system for each product, evaluating the respondents' characteristics as well. Furthermore, the panel-test data must be studied jointly with the consumer preferences. All these elements will be analyzed in the following by joining the panel-test data on sensorial variables and the judgments of consumers.

Regarding the panel test, questionnaires and descriptors will be used to implement technical specifications both for the characterization and the definition of the company quality requirements and for a verification of the actual quality expectations of consumers.

References

Berni R., Rivello R. (2009), Choices and conjoint analysis: critical aspects and recent developments, in: Bini M., Monari P., Piccolo D., Salmaso L. (eds.), *Statistical Methods and Models for the Evaluation of Educational Services and Products' Quality*, Series Contribution to Statistics, Springer-Verlag, 119–137.

Bhat C.R. (1995), A heteroschedstic extreme value model of intercity travel mode choice, *Transportation Research: part B: Methodological*, 29, 471–483.

Fotopoulos C., Krystallis A. (2001), Are quality labels a real marketing advantage? A conjoint application on Greek PDO protected olive oil, *Journal of International Food and Agribusiness Marketing*, 12, 1–22.

Haab T.C., Hicks R.L. (1997), Accounting for choice set endogeneity in random utility models of recreation demand, *Journal of Environmental Economics and Management*, 34, 127–147.

Hensher D.A., Greene W.H. (2003), The Mixed Logit model: the state of practice, *Transportation*, 30, 133–176.

Lenk P.J., DeSarbo W.S., Green P.E., Young M.R. (1996), Hierarchical Bayes conjoint analysis: recovery of part heterogeneity from reduced experimental designs, *Marketing Science*, 15, 173–191.

McFadden D., Train K. (2000), Mixed MNL for discrete response, *Journal of Applied Econometrics*, 15, 447–450.

Netzer O., Toubia O., Bradlow E. T., Dahan E., Evgeniu T., Feinberg F. M., Feit E. M., Hui S. K., Johnson J., Liechty J. C., Orlin J. B., Rao V. R. (2008), Beyond conjoint analysis: advances in preference measurement, *Marketing Letters*, 19, 337–354.

Orth U.R., Firbasova Z. (2003), The role of consumer ethnocentrism in food product evaluation, *Agribusiness*, 19, 137–153.

Rubin D.B. (1979), Using multivariate matched sampling and regression adjustment to control bias in observational studies, *Journal of the American Statistical Association*, 74, 318–328.

Scarpa R., Thiene M., Train K. (2007), Utility in WTP space: a tool to address confounding random scale effects in destination choice to the Alps, *Working Paper in Economics*, Department of Economics, University of Waikato, New Zealand, 15, 1–22.

Stefani G., Romano D., Cavicchi A. (2006), Consumer expectations, liking and willingness to pay for speciality foods: Do sensory characteristics tell the whole story?, *Food Quality and Preference*, 17, 53–62.

Train K. E. (1998), Recreation demand models with taste differences over people, *Land Economics*, 74, 230–239.

van del Lans I.A., van Ittersum K., De Cicco A., Loseby M. (2001), The role of region of origin and EU certificates of origin in consumer evaluation of food products, *European Review of Agricultural Economics*, 28, 451–477.

van Ittersum K., Candel M.J.J.M., Meulenberg M.T.G. (2003), The influence of the image of a product's region of origin on product evaluation, *Journal of Business Research*, 56, 215–226.