Food Quality and Preference 40 (2015) 230–239

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

The use of quantile regression in consumer studies

Cristina Davino [a], Rosaria Romano [b,*], Tormod Næs [c,d]

[a] University of Macerata, Department of Political Sciences, Communications and International Relations, 62100 Macerata, Italy
[b] University of Calabria, Department of Economics, Statistics and Finance, 87036 Arcavacata di Rende, Cosenza, Italy
[c] Nofima Mat AS, Osloveien 1, NO-1430 Ås, Norway
[d] University of Copenhagen, Department of Food Science, Rolighedsvej 30, DK-1958 Frederiksberg, Denmark

ARTICLE INFO

Article history:
Received 31 March 2014
Received in revised form 10 October 2014
Accepted 17 October 2014
Available online 27 October 2014

Keywords:
Quantile regression
Least squares regression
Purchase intent pattern
Questionnaire data

ABSTRACT

The main objective of this paper is to describe and discuss the use of quantile regression in consumer studies. The situation analyzed is one of relating segments of consumers obtained based on their acceptance pattern to additional consumer characteristics, including attitudes, habits and demographics variables. The paper shows how the conditional quantiles of a distribution can provide additional insight that is not provided by standard regression approaches. This type of information can be important for understanding how for instance a consumer characteristic may influence disagreement in liking, which can be equally important as the predicted average liking. The advantages of the proposed methodology will be illustrated by data from a consumer test on iced coffee.

Introduction

In consumer studies such as preference mapping (Carroll, 1972; McEwan, 1996) and conjoint analysis (Green & Srinivasan, 1978; Gustafsson, Hermann, & Huber, 2007) a major challenge is to understand individual differences in acceptance and how these relate to product characteristics. In many cases one is also interested in understanding how these differences in acceptance relate to consumer characteristics such as age, gender and habits. This type of information is very important for the planning of for instance appropriate marketing strategies. This can be done in several ways as discussed in for instance (Næs, Brockhoff, & Tomic, 2010). One possibility is to incorporate consumer characteristics directly into an ANOVA model together with the conjoint factors or sensory attributes (Næs, Lengard, Johansen, & Hersleth, 2010). Another possibility is to use a type of two-step procedure, based on an initial analysis of the liking pattern and a subsequent analysis of the relations between the individual differences and consumer characteristics (Endrizzi, Menichelli, Johansen, Olsen, & Næs, 2011; Romano, Davino, & Næs, 2014). Alternatively, one can use procedures allowing for a simultaneous analysis of product hedonic scores, product descriptors and consumer characteristics (Martens et al., 2005; Vigneau, Charles, & Chen, 2013; Vinzi, Guinot, & Squillacciotti, 2007). The three different types of information may also be related to each other using structural equations modeling (Guinot, Latreille, & Tenenhaus, 2001; Menichelli, Hersleth, Almøy, & Næs, 2014).

In all cases that we know of, the focus is on how the consumer characteristics relate to the average liking of the products. The liking scores may be individually related to the consumer characteristics or synthesized by a data compression method in a reduced number of components or simply by averaging over individual scores and then related to the consumer characteristics. In each of these cases, the main goal is to measure how the mean of the response variables (liking) changes depending on the values of the explanatory variables (consumer characteristics). This is a very useful approach, but it overlooks possible information that may lie in other aspects of the distribution. It might for instance be that the distribution of liking is much broader, i.e. it has a larger variance for certain values of the consumer characteristics, for instance it may have a larger variance for men than for women. If not taken into account, valuable aspects of the data are simply neglected or overlooked.

This paper proposes the use of quantile regression (QR) (Koenker & Basset, 1978) for this purpose, i.e. for estimating the conditional quantiles of liking as functions of the consumer characteristics. QR can be considered as a complementary method to classical regression approaches because it is able to supplement the average conditional estimates of the dependent variable with information about its whole distribution. The main motivation of the paper is to propose a statistical method that is appropriate to

* Corresponding author at: University of Calabria, Department of Economics, Statistics and Finance, Cubo 1/C, 87036 Arcavacata di Rende, Cosenza, Italy. Tel.: +39 0984 492448; fax: +39 0984 492421.

E-mail addresses: cdavino@unimc.it (C. Davino), rosaria.romano@unical.it (R. Romano), tormod.naes@nofima.no (T. Næs).

http://dx.doi.org/10.1016/j.foodqual.2014.10.003
0950-3293/

the analysis of consumer studies. Here, the standard approach to the analysis of relations between different data sets is based on the use of data compression methods like the PLSR and alike which provides loadings, scores and regression coefficients: important structures and relationships between variables (both *X* and *Y* variables) can be revealed by looking at the loadings or by looking at the regression coefficients which are combinations of the different loadings, while the scores can be used to get information about how the different individuals relate to the different variables. The use of QR in consumer studies is innovative in that regard. Still the purpose is what would always be the purpose in a regression analysis: to estimate the regression coefficients, but QR also permit to look at the variation around the regression line instead of the regression line itself. The method we propose here is one which models the percentiles of the variation around the regression line as functions of the independent variables. In this way information is obtained about for which constellations of the independent variables the variability is large and small. This is information which is impossible to get from a regression equation or looking at scores and loadings.

In particular, the present work proposes the use of QR as part of the data analysis process initially proposed by Endrizzi et al. (2011) and later by Romano et al. (2014). In the first phase of this process, consumer preferences are put in relation with the characteristics of the products using a particular ANOVA model (Endrizzi et al., 2011). The residuals of this ANOVA model, which correspond to the consumer preferences adjusted for the scale effect, are analyzed by a Principal Component Analysis (PCA). The significant components of the PCA are then put in relation with the additional consumer characteristics to evaluate possible relationships between external characteristics and consumer preferences. The paper proposes the use of QR in this last phase of the data analysis process as a complementary approach to those generally used. The advantages of the proposed methodology will be illustrated by data from a consumer test on iced coffee.

Quantile regression

It is well known that, for a given quantile θ, $100\theta\%$ of the observations are smaller than θ and $100(1-\theta)\%$ are larger. This interpretation of quantiles allows introducing quantiles as a solution of a minimization problem (Fox & Rubin, 1964). The θth quantile of a generic $\boldsymbol{y}$ variable is simply obtained minimizing the weighted absolute sum of deviations:

$$\begin{aligned} Q(\theta) &= \arg\min_{\tau} \sum_{i:y_i>\tau} \theta|y_i-\tau| + \sum_{i:y_i\leqslant\tau}(1-\theta)|y_i-\tau| \\ &= \arg\min_{\tau}\sum_i \rho_\theta(y_i-\tau) \end{aligned} \quad (1)$$

where i ($i = 1, \ldots, n$) is the generic statistical unit and ρ_θ is the so-called check function:

$$\begin{cases} \rho_\theta(u) = \theta u & \text{if} \quad u > 0 \\ \rho_\theta(u) = (\theta-1)u & \text{if} \quad u \leqslant 0 \end{cases}$$

As u is the argument of the check function, in Eq. (1) it represents the deviation between each value of the $\boldsymbol{y}$ variable and the identified quantile value.

Eq. (1) represents an asymmetric loss function because a weight θ or $1-\theta$ is assigned, respectively to positive and negative deviations; namely to points greater than or less than the quantile. A more elaborate description of how to obtain this minimization criterion is given by Koenker (2005) and Davino, Furno, and Vistocco (2013).

QR, as introduced by Koenker and Basset (1978), may be considered a generalization of Eq. (1) to the regression case. Let us denote by $Q_\theta(\hat{\mathbf{y}}|\mathbf{X}) = \mathbf{X}\hat{\beta}(\theta)$ the generic conditional quantile function for the θth quantile where $0 < \theta < 1$, $\boldsymbol{y}$ is the response variable and $\mathbf{X}$ a set of predictor variables. Similarly to the unconditional quantile minimization (Eq. (1)), the conditional quantile estimator is:

$$\hat{\beta}(\theta) = \arg\min_{\beta(\theta)}\sum_{i=1}^{n}\rho_\theta(y_i - \mathbf{x}_i'\beta_i) \quad (2)$$

where ρ_θ is the previously defined check function which weights positive and negative residuals asymmetrically. Although different functional form can be used, the paper restricts attention to linear regression models. No parametric distribution assumptions are required for the error distribution. For more information about quantiles and QR we refer to Appendix.

Application data set

The data used in this paper come from a study on consumer preferences for iced coffee (Asioli, Næs, Granli, & Almli Lengard, 2014). The study was conducted on 12 different coffees and involved 100 consumers, who have evaluated the products only through an image on the screen of the PC, without tasting the products. The products varied according to the following four conjoint factors: calories (60 kcal or 90 kcal per 100 ml), origin (Norway o Italy), price (17, 23 or 29 Norwegian crones), and coffee type (latte or espresso, corresponding to mild and strong). A fractional factorial design of the twelve products was conducted (Table 1). In the original paper by Asioli et al. (2014) it is stated that, on the basis of knowledge about the products and the complexity of the task, a full factorial design with a large number of possible factors and levels combinations (2 * 2 * 2 * 3 = 24) was not considered appropriate for the study. The *probability of buying* was evaluated on a 9-point scale, with 1 = "not very likely", and 9 = "very likely".

The questionnaire used to collect the data also included a series of questions on additional consumer characteristics related to demographics, habits and attitudes. For illustration purposes, the present work focuses only on two groups of variables: *importance of attributes* (i.e. the consumers' assessment of how important they consider the attributes) and *neophobia* (Table 2). The *importance of attributes* was evaluated on a scale from 1 to 5, with 1 meaning "not important at all" and 5 "very important". Consumers had to evaluate their agreement with each of the "food-neophobia" statements on a 1-to-7 point scale, with 1 meaning "completely disagree" and 7 "completely agree". It is worth noticing that the last neo-phobia question is reversed compared to first two questions. This aspect will be properly considered in the interpretation of the results.

Table 1
Iced coffee profiles obtained by means of the factorial design.

Product	Coffee type	Calories (kcal per 100 ml)	Origin	Price (Norwegian crones)
1	Espresso	90	Italy	29
2	Latte	90	Norway	23
3	Latte	60	Norway	23
4	Espresso	60	Norway	17
5	Latte	90	Norway	29
6	Espresso	60	Norway	29
7	Espresso	90	Norway	17
8	Latte	90	Italy	17
9	Latte	60	Italy	29
10	Espresso	90	Italy	23
11	Latte	60	Italy	17
12	Espresso	60	Italy	23