Advances in Data Analysis and Classification
https://doi.org/10.1007/s11634-020-00410-x

REGULAR ARTICLE

On the use of quantile regression to deal with heterogeneity: the case of multi-block data

Cristina Davino[1] · Rosaria Romano[1] · Domenico Vistocco[2]

Received: 18 July 2019 / Revised: 2 July 2020 / Accepted: 8 July 2020

Abstract
The aim of the paper is to propose a quantile regression based strategy to assess heterogeneity in a multi-block type data structure. Specifically, the paper deals with a particular data structure where several blocks of variables are observed on the same units and a structure of relations is assumed between the different blocks. The idea is that quantile regression complements the results of the least squares regression by evaluating the impact of regressors on the entire distribution of the dependent variable, and not only exclusively on the expected value. By taking advantage of this, the proposed approach analyses the relationship among a dependent variable block and a set of regressors blocks but highlighting possible similarities among the statistical units. An empirical analysis is provided in the consumer analysis framework with the aim to cluster groups of consumers according to the similarities in the dependence structure among their overall liking and the liking for different drivers.

Keywords Quantile regression · Group dependence structure · Individual differences · Consumer analysis

Mathematics Subject Classification 62G08 · 62P20 · 91B42

1 Introduction

In recent years the growing availability of data of various types and often collected from different sources requires changes in the classical statistical methods. Ad hoc

✉ Cristina Davino
cristina.davino@unina.it

Rosaria Romano
rosaroma@unina.it

Domenico Vistocco
domenico.vistocco@unina.it

[1] Department of Economics and Statistics, University of Naples Federico II, Naples, Italy

[2] Department of Political Science, University of Naples Federico II, Naples, Italy

Published online: 19 July 2020

techniques must be tailored for managing data arranged in different types of structures. In complex domains it is very common that the observed variables are grouped into homogeneous blocks measuring partial aspects of the phenomenon under investigation. Such data is usually labelled as multi-block data and the statistical tecniques used for their analysis are called multi-block methods (Smilde et al. 2000). Important examples of multi-block data comes from several application fields (marketing, sociology, econometrics, sensory analysis, spectroscopy, ecology, image analysis, etc.). The basic requirement of multi-block methods is that all blocks have one dimension (mode) in common. The way in which the blocks are connected to each other gives rise to different data structure. Different sets of variables can be measured on the same units. Examples of this structure are several: many sets of indicators to describe different aspects of a complex concept, *quality of life*, for instance (Davino and Romano 2014); different components (i.e, teaching, research, internationalisation) to define a university ranking (Romano and Davino 2016); different dimensions that affect the quality of teaching in high schools (motivation, emotions, strategies, teaching) (Romano and Palumbo 2013). Another possible structure consists of a set of two-way matrices of the same units and variables (three-way data). Three-way data is widely used in sensory analysis when the scores of the judges are not averaged so that the three sources of variation are products, attributes and judges (Romano et al. 2008, 2015; Bro et al. 2008). This three-way data structure can also be related to a dependent data set (Romano et al. 2011). Finally, another possible structure comes out when blocks of different dimensions are connected through a common dimension. Examples in consumer studies concern the link of product attributes, liking scores and consumers' characteristics (Martens et al. 2005; Romano et al. 2014; Davino et al. 2015).

Different types of approaches can be used for the analysis of these multi-block data structures, the choice being based on the type of relationship between the various blocks (Höskuldsson 2008; Cariou et al. 2018). An exploratory multi-block approach (Hanafi and Kiers 2006) has to be employed in case no specific causal relationship is assumed among the different blocks. A supervised approach (Westerhuis et al. 1998), is instead used whereas a structure of relationships is assumed between the blocks. Here, the blocks of predictor variables are generally called input blocks, while the block of dependent variables is named output block. If there is a chain of relationships among the blocks, the data follows the typical structure of the structural equations models (Jöreskog and Wold 1982; Bollen 2014).

The present paper deals with multi-block data where all blocks of variables are observed on the same units and both input blocks and a single output block are available. Therefore the approach is supervised since the aim is to explore the dependence structure between the input blocks and the output one. Consumer analysis is considered as a field of application, since data is generally collected into multi-block structures (Næs et al. 2011). Here, the output block generally corresponds to the liking scores given by a sample of consumers on a predefined set of products. The input blocks can concern both other sensory variables, defined drivers of liking (or specific liking), and other additional variables on consumers (demographic, habits, attitudes). The relationships between the input blocks and the output block can be investigated through different strategies, each corresponding to a different way of arranging the data into blocks (Næs et al. 2010). The simplest strategy is to transform each block into a single vector by stacking the corresponding columns and estimating a multiple linear regression model. This strategy does not take into account heterogeneity among consumers, which is a further source of complexity of data coming from consumer studies. To this end, some approaches propose to estimate separate regression models for each consumer to be aggregated a posteriori, by a simple arithmetic average or by clustering procedures, to highlight segments of homogeneous consumers with respect to the liking model (Menichelli et al. 2013; Asioli et al. 2016).

This paper proposes a multi-step procedure to deal with heterogeneity in multi-block data. It exploits Quantile Regression (QR) (Koenker and Basset 1978; Davino et al. 2013; Furno and Vistocco 2018) to evaluate the effect of the regressors on the entire distribution of the dependent variable. The idea is to complement the classical approach based on the least squares regression (LSR) to focus beyond the conditional mean. In the case of consumer data this complementary information can be crucial given the typical asymmetric distributions of liking scores. QR has already been used in consumer studies: for estimating the conditional quantiles of liking when segments of consumers obtained according to their acceptance pattern are related to additional consumer characteristics (Davino et al. 2015); for assessing heterogeneity across product similarities (Davino et al. 2018). The proposed strategy has significant implications so that the results can be used to adopt appropriate marketing strategies. In particular, the information obtained concerns the identification of consumer groups that have similar liking models, that is, for each group it is possible to identify the effect of each driver of liking on the overall liking. In addition to the different liking models, the detailed analysis of each group in terms of liking for the individual products also allows for useful information for the product development.

This paper extends a QR multi-step procedure used to assess heterogeneity (Davino and Vistocco 2018) to multi-block data. The main idea is to explore if and how the effects of the drivers on the overall liking differ for groups of consumers at the lower and higher levels of liking. The strategy consists of three main steps. The first step aims to identify the best model for each consumer, based on the quantile that best represents each consumer. In the second step, consumers segments are identified according to similarities in the dependence structure, using cluster analysis. In the final step, a different model is estimated for each group of consumers identified in the previous step (a group can also consist of a single consumer). The proposed procedure is tested on data from consumer study. The basic aim of the proposal is to learn knowledge from a complex data structure, where information gathered in several blocks is combined and analysed in the different steps. The use of QR allows to deal with heterogeneity, a typical source of variation in many fields of applications. Note that this is one of the main strengths of QR as compared to the classical LSR. Such an aspect is exploited in using QR once selecting different models for each consumer according to his/her position (quantile) in the dependent variable distribution. In addition, the QR is used in the final step of the procedure when identifying the representative quantiles of the groups obtained from the cluster procedure. The estimation of different QR models on predefined quantiles using all the observations of the samples allows statistical comparisons between the groups both with respect to the entire model and to individual coefficients.

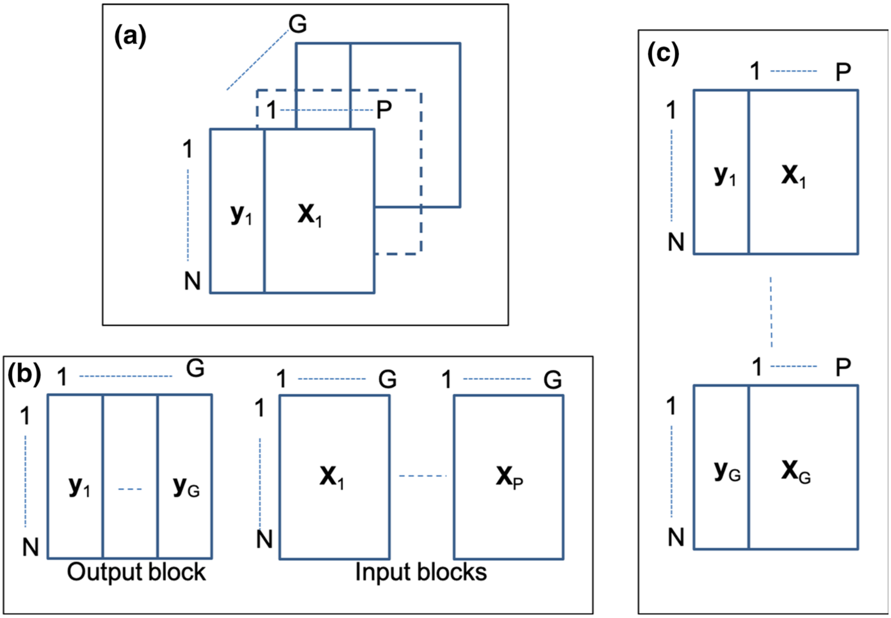

Fig. 1 Description of different multi-block data-structures

The paper is structured as follows. In Sect. 2 the main notation is introduced and the quantile regression based strategy is described. Data used for testing the procedure are described in Sect. 3, while the corresponding results are included in Sect. 4. Finally, some concluding remarks and directions of future avenues of research are described in Sect. 5.

2 Methodology

2.1 Main notation

Let us consider a particular multi-block data arranged as three-way data table, where an array $\mathbf{Z}$ is partitioned in G blocks: $\mathbf{Z} = [\mathbf{Z}_1, \ldots, \mathbf{Z}_G]$. In this paper, each block $\mathbf{Z}_g$ $(g = 1, \ldots, G)$ has dimensions $N \times (P+1)$ because it is a column partitioned matrix composed by the vector $(\mathbf{y}_g)$ for the response variable (a single element of the output block as defined in Sect. 1) and a data matrix $(\mathbf{X}_g)$ for the regressors. The dimension G can represent any possible stratification variable, even time. In the empirical analysis provided in Sects. 3 and 4, the N rows represent a set of products, the P+1 variables are the liking attributes with the dependent variable (the overall liking), while G is the set of consumers. Figure 1 shows three different points of view to represent and analyse such a kind of multi-block data. For the rest of the paper we will refer to the structure represented in Fig. 1c, where the G blocks obtained from the stratification variable are stacked. However, the proposed approach will not consider a single model on all stacked blocks simultaneously. A multistep strategy is proposed in which the response variable and corresponding predictors are related to each other for each individual block.

The aim of the paper is to model and cluster the data simultaneously. That means analysing the relationship among the dependent variable and the set of regressors but highlighting possible similarities in the G levels of the stratification variable. It is a matter of fact that if two units show a very similar dependence structure, they can be considered belonging to the same group.

2.2 Quantile regression based strategy

The strategy proposed in this paper consists of three main steps. The first step consists of identifying the best model for each level g of the stratification variable, based on the quantile of the response variable that best represents each level. Subsequently, the G levels of the stratification variable are grouped according to similarities in the dependence structures. Finally, a different model is estimated for each group identified in the previous step.

The procedure has been introduced by Davino and Vistocco (2018) and applied in consumer science to handle products effects by Davino et al. (2018). In the present paper, the approach is adapted in the case of multi-block data with a large number of blocks. As discussed in Sect. 1, the aim of this paper is to consider an additional source of complexity in the data given by heterogeneity. Since this involves a great number of blocks, the strategy proposed in Davino and Vistocco (2018) is here combined with clustering techniques.

The main strength point of the procedure is represented by the exploitation of QR in the whole process of analysis. QR allows to estimate the whole distribution of the conditional quantiles of the response variable thus replacing the classical estimate of a single value (conditional mean) with estimates of several values (conditional quantiles). A typical QR model is formulated as:

$$Q_\theta(\hat{\mathbf{y}}|\mathbf{X}) = \mathbf{X}\hat{\beta}(\theta), \tag{1}$$

where $Q_\theta(.|.)$ is the conditional quantile function for the θ-th conditional quantile with $0 < \theta < 1$. Each $\hat{\beta}_p(\theta)$ coefficient represents the rate of change in the θ-th conditional quantile of the dependent variable per unit change in the value of the p-th regressor ($p = 1, \ldots, P$), holding the others constant. Although it is theoretically possible to estimate an infinite number of quantiles, a finite number is numerically distinct, the so-called quantile process. Also for QR, several are the functional forms that can be considered. The paper will refer to linear regression models. The interested reader may refer to the reference literature for methodological details (Koenker and Basset 1978; Davino et al. 2013; Furno and Vistocco 2018).

The approach to model and cluster the G levels is structured in the three steps detailed below.

(1) *Identification of the best model for each level*

In the first step, a representative quantile θ_g^{best} is identified for each level g. It will be named from now on as the best quantile. In particular, computing the empirical cumulative distribution function $F(\cdot)$ on the overall $\mathbf{y}$ variable, the best quantile representative of each block g will be obtained as:

$$\theta_g^{best} = \sum_{y \in \mathbf{y_g}} \frac{F(y)}{N}, \tag{2}$$

where N denotes the number of units of the generic block g. By computing the empirical cumulative distribution function on $\mathbf{y}$, we refer to the percentile rank of each observation, i.e. the location of each y_i $(i = 1, \ldots, N \times G)$. For a discussion on the use of the percentile ranks and the choice of the proper location index to summarise them, see (Davino and Vistocco 2018).

(2) *Identification of the group dependence structure*
QR is then carried out on data arranged as in Fig. 1c (for each single block), using the representative quantiles, that is, the G quantiles θ_g^{best} previously identified. Each model provides a set of coefficients, one for each level g and for each regressor: $\hat{\beta}_p\left(\theta_g^{best}\right)$ $(p = 1, \ldots, P)$. Such coefficients can be arranged into a matrix $\hat{\mathbf{B}}\left(\theta^{best}\right)_{[G\times(P+1)]}$ where the additional column refers to the intercept.

The aim of the second step is to identify if there are similar dependence structures among the G levels. For this purpose, a hierarchical cluster analysis (CA) is performed on the $\hat{\mathbf{B}}\left(\theta^{best}\right)$ matrix and a partition of G in K groups is identified $(k = 1, \ldots, K)$. Each group will be then characterised by a different quantile deriving from an average of the θ^{best} associated to the units assigned to the group $(\bar{\theta}^{best})$.

(3) *Estimation of the group dependence structure*
In the final step, QR is carried out again on data arranged as in Fig. 1c using the representative quantiles, that is, the K quantiles assigned to the K groups in the previous step. Each of the K estimated models provides a set of coefficients, one for each regressor; differences among the coefficients highlight differences in the group dependence structure. It is worth of noticing that coefficients can be compared because all of them are estimated on the whole sample $(N \times G)$. A testing procedure is implemented to evaluate the significance of the differences among the coefficients related to each cluster, exploiting the classical inferential tools available in the QR framework. Two models estimated at two different quantiles can be compared using a joint tests on all slope parameters or separate tests on each of the slope parameter. The hypothesis of interest is that the slope coefficients of two models are identical and the test statistic is a variant of the Wald test described in Koenker and Bassett (1982).
Let us consider the case of the comparison among the coefficients related to the p–th regressor and estimated at two different quantiles, θ_k^{best} and $\theta_{k'}^{best}$. The null is $H_0 : \beta_p(\theta_k^{best}) = \beta_p(\theta_{k'}^{best})$, and the test statistic is:

$$T = \frac{\left[\hat{\beta}_p(\theta_k^{best}) - \hat{\beta}_p(\theta_{k'}^{best})\right]^2}{\hat{var}\left[\hat{\beta}_p(\theta_k^{best}) - \hat{\beta}_p(\theta_{k'}^{best})\right]}, \tag{3}$$

where $p = 1, \ldots, P$ and $k, k^{'} \in [1, K]$. Under the null hypothesis, the test statistic has an approximate χ^2 distribution with one degree of freedom.

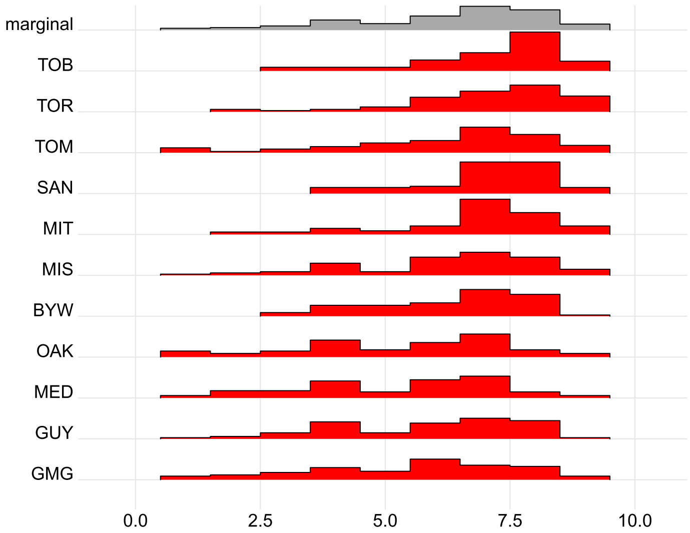

Fig. 2 Distributions for the overall liking: marginal (top) and product specific (from bottom to second to last)

Such a test statistic can be exploited for coefficients pairwise comparisons. An extension of it is used as global test on all the slopes. The standard errors, used to evaluate the statistical significance of the coefficients, can be estimated using resampling methods (Parzen et al. 1994).

3 Data description

The empirical analysis is based on data from a consumer testing on 11 tortilla chips, in which 73 consumers expressed their overall liking for each product on a 9-point hedonic scale (from 1 $=$ dislike extremely to 9 $=$ like extremely) (Meullenet et al. 2008). Furthermore, consumers themselves have provided a judgment on some drivers (appearance, flavor, texture) using the same 9-point hedonic scale. Considering the notation introduced in Sect. 2, the structure of the tortilla dataset is made by $N = 11$ products, $P + 1 = 4$ liking variables and $G = 73$ consumers (levels).

Figure 2 shows the distribution of the overall liking scores on the whole sample of consumers (marginal) and for each single product. All products show a left skewed distribution, even if some of them present a higher variability and a less pronounced skewness (MIS, OAK, MED, GUY, GMG).

A multivariate analysis of consumers' liking is carried out using a principal component analysis (PCA) on the output block ($products \times consumers$) (data arranged as in Fig. 1b). Results demonstrate that consumers show preferences for different products. In Fig. 3 consumer vectors are mainly concentrated in the positive verse of the first dimension, even if there are a few consumers also lying along the second dimension. Individual differences among consumers for the likings of specific prod-

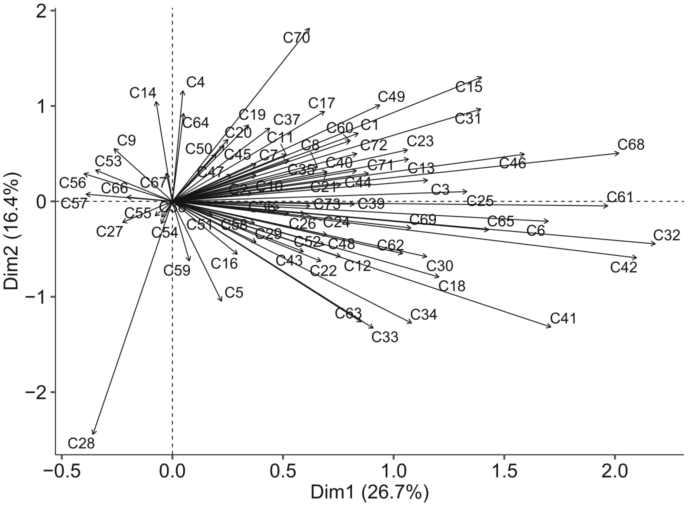

Fig. 3 Loading plot by principal component analysis on the overall liking

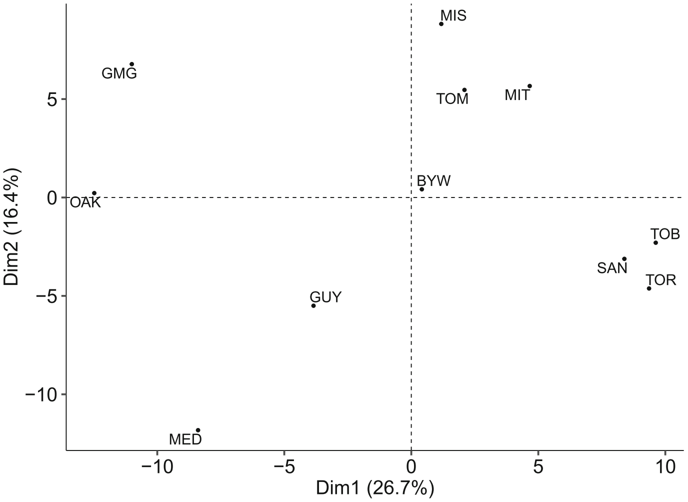

Fig. 4 Score plot by principal component analysis on the overall liking

ucts can be further understood from Fig. 4, which represents the products arranged on all four quadrants. Specifically, the most liked products are SAN, TOB and TOR, along the main PCA dimension, followed by MIS, MIT and TOM on the second dimension.

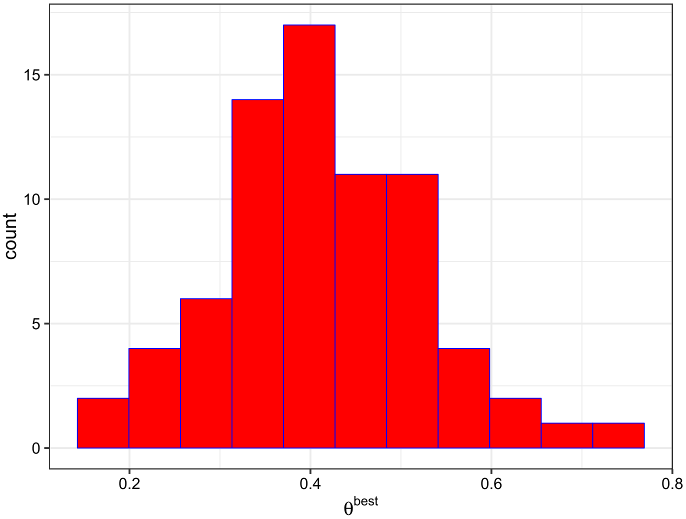

Fig. 5 Histogram of the rank percentiles

4 Results

Heterogeneity among consumers described through the PCA in Sect. 3 can be further investigated linking the overall liking of the consumers to their specific likings (appearance, flavor, texture) by following the QR based strategy described in Sect. 2.2.

In the first step, a representative quantile is identified for each consumer through the average rank percentile of the overall liking she/he has expressed on the considered set of products. The histogram in Fig. 5 shows the distribution of the rank percentiles on the sample of consumers. It is worth noting that there exist a variability on the θ^{best} and thus an heterogeneity in the liking.

QR is then carried out on data arranged as in Fig. 1c using the representative quantiles, that is, the G quantiles θ_g^{best} assigned to the each consumer. Each model provides a set of coefficients, one for each driver, that can be arranged in a matrix *consumers* × *coefficients*. The information gathered into such a matrix is crucial for highlighting the individual differences/similarities among consumers in the way they weight the drivers linked to the overall linking. To this end, the second step consists of identifying consumers' segments by a CA on the different dependence structures estimated by the QR on the G quantiles θ_g^{best}.

On the tortilla dataset, the elbow rule identifies as best partition the one with $k = 3$ groups of consumers, homogenous with respect to the liking models. Table 1 describes each cluster through the following information: size, minimum, maximum and average of the θ_g^{best} of the consumers assigned to the cluster. The last four columns of the table describe the average values of the original variables. Note that for the rest of the paper the $\bar{\theta}^{best}$ will be considered as the quantile representative of each group. Results show that each cluster is characterised by a different position in the ranking of the overall liking. Specifically, the first cluster corresponds to the consumers' segment with lower

Table 1 Description of clusters exploiting summaries of θ^{best} and original variables (average)

Size	θ_{min}	θ_{max}	$\bar{\theta}^{best}$	Overall	Appearance	Flavor	Texture
$n_1 = 6$	019	0.24	0.22	4.89	5.38	4.85	5.44
$n_2 = 34$	0.30	0.41	0.35	5.88	6.32	5.77	6.31
$n_3 = 33$	0.42	0.76	0.51	6.94	7.13	6.71	7.02

Table 2 QR coefficients for the three representative quantiles of the three clusters

	$\bar{\theta}_1^{best} = 0.22$	$\bar{\theta}_2^{best} = 0.35$	$\bar{\theta}_3^{best} = 0.51$
(Intercept)	− 1.24	− 0.48	0.00
Appearance	0.23	0.16	0.17
Flavor	0.70	0.77	0.67
Texture	0.18	0.12	0.17

θ^{best}, i.e. to consumers scoring products lower than the others. The last two clusters, which are the most interesting because of their size, behave in the opposite way. Note that the third cluster show a wider range of the θ^{best} as highlighted by the minimum and maximum vales in the Table 1. This emphasises a higher degree of heterogeneity in this cluster as compared to the others. Another relevant information from Table 1 comes from the analysis of the average values both on liking and drivers. On one hand, the first cluster has a very small size, with a low degree of liking that can hardly be modified. On the other hand, clusters 2 and 3 show higher overall liking values than can be further improved by acting on the specific likings, in particular on the flavor that reveals the lowest averages inside each cluster.

In the final step, QR is carried out again on data arranged as in Fig. 1c using the representative quantiles assigned to each cluster ($\bar{\theta}^{best}$). Coefficients in Table 2 are all significant for $\alpha \leq 0.05$ but intercept and texture coefficient in group 2, which are significant for $\alpha \leq 0.10$. Moreover the intercept in group 3 is not significant at all.

Combining the information in Tables 1 and 2 it is possible to argue that flavor is the most interesting driver to improve the overall liking of the less satisfied consumers (cluster 1 and 2). If on one hand, flavor has the highest QR coefficients, on the other hand consumers score this driver lower than the others, with averages close to the threshold of sufficiency in a scale ranging for 1 to 9.

Results in Table 3 complements the characterisation of the groups by testing their difference in terms of both the whole model (joint test) and the specific coefficients. The classical Bonferroni correction (Shaffer 1995) has been applied. The p-values included in the table show that the three models for the three clusters are significantly different. Focusing on differences among cluster 2 and 3, the p-values emphasise that even if the size of the flavor and the texture coefficients in the two clusters is quite similar, they can be considered different from an inferential perspective.

Starting from the results of the testing procedure showing that flavor and texture are the most discriminating predictors for clusters 2 and 3, a deepen analysis of differences among products in each cluster would be useful in a product development perspective. At this aim Fig. 6 visualizes the averages of the variables (each single panel) in the

Table 3 *p*-values derived from pairwise comparisons on the whole model and on single coefficient

	Joint test	Appearance	Flavor	Texture
Cluster 1 versus cluster 2	0.018	0.024	0.042	0.141
Cluster 1 versus cluster 3	0.009	0.111	0.540	2.295
Cluster 2 versus cluster 3	0.000	2.598	0.000	0.000

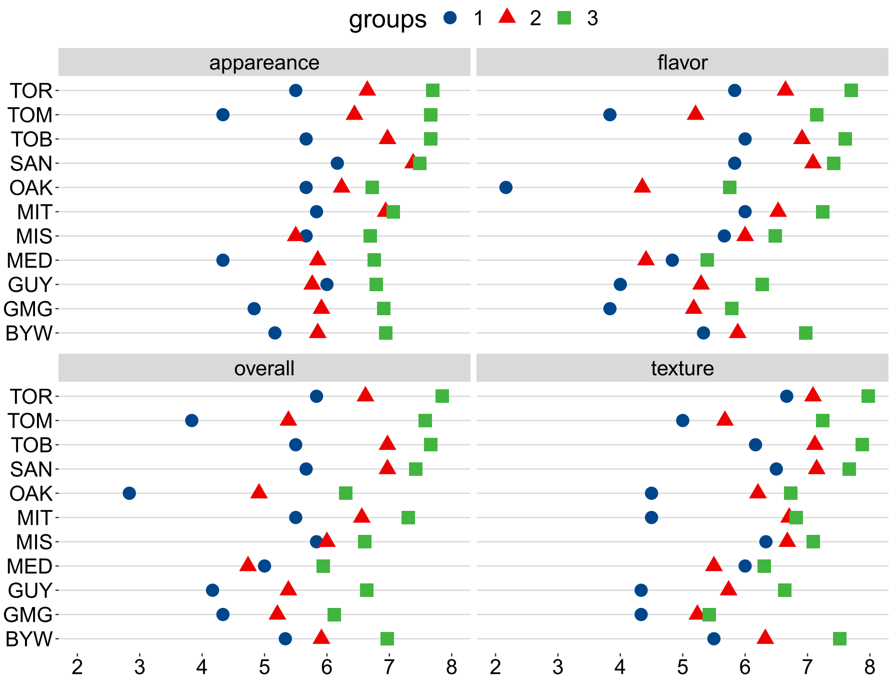

Fig. 6 Description of the clusters according to the original variables partitioned by products

three clusters (different colors and symbols), partitioned by products (labels on the left side). The most relevant information from the figure is that the differences between these two clusters are evident for the flavor and the texture, where it is possible to highlight, among the most liked products, those presenting the wider range between the averages in the two clusters. Specifically the liking of TOR, TOM and BYW that present lower averages in cluster 2 will be more affected by an improvement in the flavor. Such information in terms of the liking for products inside each cluster can be used to suggest appropriate decisions both for marketing and product development departments.

5 Concluding remarks and further developmnents

The paper has shown how to treat an additional source of complexity given by heterogeneity in fields of applications where data follows multi-block structures. Focus has been on one specific data structure, within the supervised framework, where blocks of variables (a single output block and several input blocks) are observed on the same statistical units. A QR based strategy has been proposed as a multi-step procedure able to model and cluster units according to similar dependent structures. The proposal originates from alternative approaches that combine the estimation of separate dependent relations for each single units with a clustering on the model results. For instance in conjoint studies (Gustafsson et al. 2003), where a preference model for each consumer is estimated and then results from the estimated models are synthesised by simple averages or clustering procedures. The strategy proposed only in some aspects can be traced back to classic approaches. For example, the logic of obtaining different models for individual consumers and then synthesizing them is in common. The peculiarity of the proposal is already in the selection of the model, which is based on the selection of the quantile that best represents each consumer. This aspect makes the proposal complementary and not alternative to the classic approaches, which are limited to the study of the effects on average. Any comparison with classical methods would lead to the classical conclusions deriving from a comparison between QR and LSR: the two models provides similar results when the homoscedasticity assumption is satisfied. This means that the different blocks (consumers) present the same dependence structure linking the overall to the specific liking.

The innovative contribution of the proposed approach consists of the following aspects:

- The use of QR in alternative to the classical LSR to model the whole distribution of the dependent variable.
- The identification of different models for each unit obtained by defining the quantile that best represent the position of the unit in the distribution of the dependent variable.
- The estimation of the model characterizing each cluster obtained on all the units and not only on the ones belonging to the cluster. This is a relevant aspect since the estimation of the models using all units allows comparisons among the clusters both for the whole models and for the specific coefficients.
- The description of each cluster according to the corresponding specific quantile, which provides information on the location of the response conditional distribution mainly affected by the units of the cluster.

Further developments are in the direction of the selection of the best clustering partition (Bruzzese and Vistocco 2015; Tibshirani and Walther 2001) and on the assessment of results in terms of stability. The second aspect is even more relevant in application fields where the number of statistical units (products) is low, like in consumer analysis.

Acknowledgements Open access funding provided by Università degli Studi di Napoli Federico II within the CRUI-CARE Agreement.

References

Asioli D, Næs T, Øvrum A, Almli VL (2016) Comparison of rating-based and choice-based conjoint analysis models. A case study based on preferences for iced coffee in Norway. Food Qual Prefer 48:174–184

Bollen KA (2014) Structural equations with latent variables, vol 210. Wiley, New York

Bro R, Qannari EM, Kiers HA, Næs T, Frøst MB (2008) Multi- way models for sensory profiling data. J Chemom 22(1):36–45

Bruzzese D, Vistocco D (2015) DESPOTA: DEndrogram slicing through a permutation test approach. J Classif 32(2):285–304

Cariou V, Qannari EM, Rutledge DN, Vigneau E (2018) ComDim: from multiblock data analysis to path modeling. Food Qual Prefer 67:27–34

Davino C, Romano R (2014) Assessment of composite indicators using the ANOVA model combined with multivariate methods. Soc Indic Res 119(2):627–646

Davino C, Vistocco D (2018) Handling heterogeneity among units in quantile regression. Investigating the impact of students' features on university outcome. Stat Interface 11(3):541–556

Davino C, Furno M, Vistocco D (2013) Quantile regression. Theory and applications. Wiley, series in probability and statistics. Wiley, UK

Davino C, Romano R, Næs T (2015) The use of quantile regression in consumer studies. Food Qual Prefer 40(A):230–239

Davino C, Romano R, Vistocco D (2018) Modelling drivers of consumer liking handling consumer and product effects. Ital J Appl Stat 30:359–372

Furno M, Vistocco D (2018) Quantile regression. Estimation and simulation., vol 216. Wiley, New York

Gustafsson A, Herrmann A, Huber F (2003) Conjoint measurement: methods and applications. Springer, Berlin

Hanafi M, Kiers HA (2006) Analysis of K sets of data, with differential emphasis on agreement between and within sets. Comput Stat Data Anal 51(3):1491–1508

Höskuldsson A (2008) Multiblock and path modelling procedures. J Chemom 22:571–579

Jöreskog KG, Wold HO (1982) Systems under indirect observation: causality, structure, prediction, vol 139. North Holland, Amsterdam

Koenker R, Basset GW (1978) Regression quantiles. Econometrica 46(1):33–50

Koenker R, Bassett G (1982) Tests of linear hypotheses and L1 estimation. Econometrica 50(6):1577–1584

Martens H, Anderssen E, Flatberg A, Gidskehaug LH, Høy M, Westad F, Martens M (2005) Regression of a data matrix on descriptors of both its rows and of its columns via latent variables: L-PLSR. Comput Stat Data Anal 48(1):103–123

Menichelli E, Kraggerud H, Olsen NV, Næs T (2013) Analysing relations between specific and total liking scores. Food Qual Prefer 28(2):429–440

Meullenet JF, Xiong R, Findlay CJ (2008) Multivariate and probabilistic analyses of sensory science problems, vol 25. Wiley, New York

Næs T, Lengard V, Johansen SB, Hersleth M (2010) Alternative methods for combining design variables and consumer preference with information about attitudes and demographics in conjoint analysis. Food Qual Prefer 21(4):368–378

Næs T, Brockhoff PB, Tomic O (2011) Statistics for sensory and consumer science. Wiley, New York

Parzen MI, Wei LJ, Ying Z (1994) A resampling method based on pivotal estimating functions. Biometrika 81(2):341–350

Romano R, Davino C (2016) Assessing scientific research activity evaluation models using multivariate analysis. Stat Interface 9(3):303–313

Romano R, Palumbo F (2013) Partial possibilistic regression path modeling for subjective measurement. J Methodol Appl Stat 15:177–190

Romano R, Brockhoff PB, Hersleth M, Tomic O, Næs T (2008) Correcting for different use of the scale and the need for further analysis of individual differences in sensory analysis. Food Qual Prefer 19(2):197–209

Romano R, Vestergaard JS, Kompany-Zareh M, Bredie WL (2011) Monitoring panel performance within and between sensory experiments by multi-way analysis, in Classification and Multivariate Analysis for Complex Data Structures, 335–342. Springer, Berlin

Romano R, Davino C, Næs T (2014) Classification trees in consumer studies for combining both product attributes and consumer preferences with additional consumer characteristics. Food Qual Prefer 33:27–36

Romano R, Næs T, Brockhoff PB (2015) Combining analysis of variance and three-way factor analysis methods for studying additive and multiplicative effects in sensory panel data. J Chemom 29(1):29–37

Shaffer JP (1995) Multiple hypothesis testing. Ann Rev Psychol 46:561–584

Smilde AK, Westerhuis JA, Boque R (2000) Multiway multiblock component and covariates regression models. J Chemom 14(3):301–331

Tibshirani R, Walther G, Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. J R Stat Soc Ser B (Stat Methodol) 63(2):411–423

Westerhuis JA, Kourti T, MacGregor JF (1998) Analysis of multiblock and hierarchical PCA and PLS models. J Chemom 12:301–321

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.