Assessing heterogeneity in consumer analysis across product similarities and within consumers differences

Cristina Davino
Federico II Univ. of Naples

Rosaria Romano
University of Calabria

Domenico Vistocco
University of Cassino

Table of contents

A case study on tortilla chips

Case study product set

Eleven commercially available toasted white corn tortilla chip products [1].

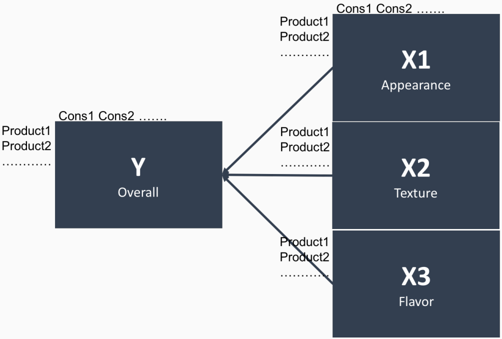
ProductID	Producer	Shape	Salt content (%)	Fat content (%)
BYW	Fleming Companies, Inc.	Triangle	4	12
GMG	Green Mountain Gringo	Strip	5	13
GUY	Guy's Snack Foods	Round	3	9
MED	Medallion Food Corporation	Triangle	2	11
MIS	Mission Food Corporation	Strip	4	10
MIT	Mission Food Corporation	Triangle	4	10
OAK	Oak Creek Farms	Round	2	11
SAN	Frito-Lay	Triangle	5	8
TOB	Frito-Lay	Round	5	12
TOM	Tom's Foods Inc.	Triangle	5	10
TOR	Frito-Lay	Triangle	3	9

Consumer Testing

Each consumer was asked to evaluate the **appearance**, **flavor**, **texture**, and **overall impression** of each sample on a 9-point hedonic scale with 9 being "like extremely" and 1 being "dislike extremely".

[1] Meullenet, J. F., Xiong, R., Findlay, C. J. (2008). Multivariate and probabilistic analyses of sensory science problems (Vol. 25). John Wiley & Sons.

Case Study Objective



Analysing relations between specific and overall liking:

- on an **global level** (full panel);
- on a **product basis** to detect products or groups for which the relation between the overall liking and its drivers is most different;
- on an **individual** level to detect consumer's segments.

The toolkit: quantile regression for handling heterogeneity

Quantile Regression

Quantile Regression[2] (QR) can be considered an extension of conditional mean models to the whole conditional distribution of the response variable:

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

- $Q_\theta(.|.)$ is the conditional quantile function for the θ-th quantile.
- θ is a given conditional quantile, with $0 < \theta < 1$.

Estimates in QR follows the classical interpretation:

Each $\hat{\beta}(\theta)$ coefficient represents the rate of change in the θ-th conditional quantile of the dependent variable per unit change in the value of the j-th regressor ($j = 1, \ldots, p$), holding the others constant.[3]

[2] Koenker, R., Bassett Jr, G. (1978). Regression quantiles. *Econometrica*, 46, 33-50.
[3] Davino, C., Furno, M., Vistocco, D. (2013). *Quantile regression: theory and applications*. Wiley.

Quantile Regression in consumer study

Previous studies

QR has been used in consumer studies for estimating the conditional quantiles of liking in terms of the consumer characteristics [4]

Current study

- QR to relate specific liking attributes to overall liking - **Global Model**
- QR-based strategy to detect **product effect** in such a relation:[5]
    1. Identification of the best model for each product.
    2. Estimation of the group dependence structure.
    3. Test of the differences among the groups.

[4] Davino, C., Romano, R., Naes, T. (2015). The use of quantile regression in consumer studies. *Food Quality and Preference*, 40, 230-239.

[5] Davino, C., Vistocco, M. (2017). Handling heterogeneity among units in Quantile Regression. *Statistics and its interface*, forthcoming.

Data structure for the case study

Data structure:

n: number of units $\rightarrow$ consumers

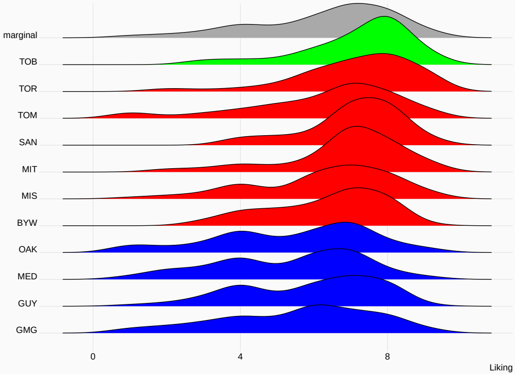
p: number of regressors $\rightarrow$ drivers

G: number of groups $\rightarrow$ products

- $\mathbf{X}_{[n \times p]}$
  - ${}_gX_{ij}$ $(i = 1, \dots, n; j = 1, \dots, p; g = 1, \dots, G)$
- $\mathbf{y}_{[n]}$
  - ${}_gy_i$ $(i = 1, \dots, n; g = 1, \dots, G)$
- n_g: number of units in level g

Main results

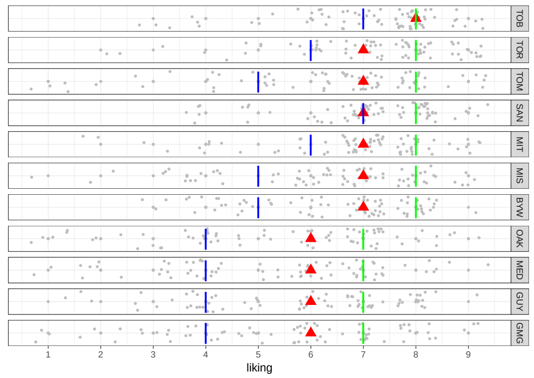
Overall liking distribution



Liking score density plot illustrates satisfaction profile:

- The overall liking (full panel) shows a strong left skewness, i.e. very high overall liking scores.
- TOB, TOR, and SAN (from the same producer Frito-Lay) appear to have the highest overall liking scores.
- GMG and OAK appear to be the less liked products.

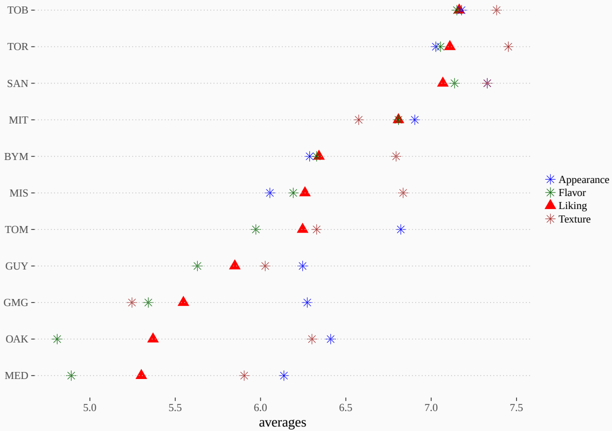
Overall liking descriptive statistics



Liking score descriptive statistics illustrate satisfaction profile:

- TOB, TOR, SAN and MIT completely satisfy all consumers, with approximately 75% of ratings scoring higher than 6.
- GMG, GUY, MED and OAK do not satisfy all consumers, with approximately 25% of ratings scoring less than 4.

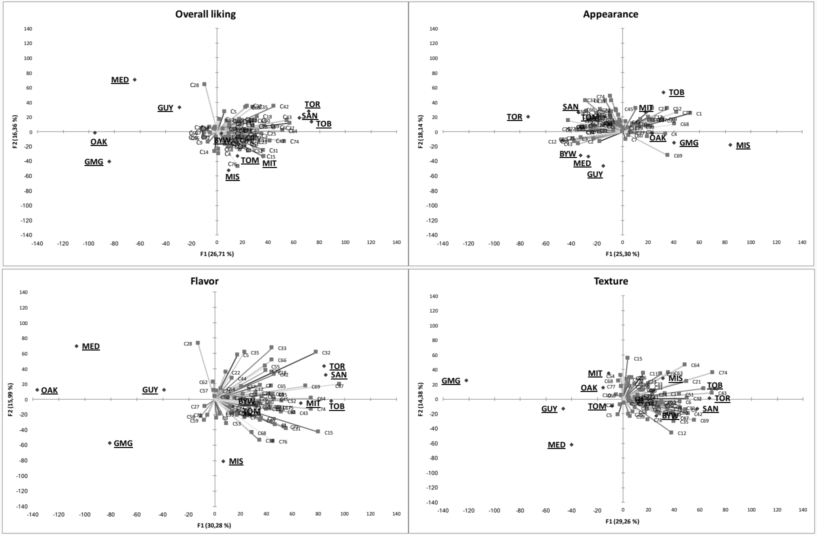
Overall and liking drivers: average values



Differences in the liking profile for each liking attribute

Products that are less appreciated (lower overall liking), such as GUY, MED, and OAK, have lower flavor values than the other two drivers (appearance and texture).

Overall and liking drivers: individual differences



Quantile Regression in consumer study

Current study

- QR to relate specific liking attributes to overall liking - **Global Model**
- QR-based strategy to detect **product effect** in such a relation:[6]
  1. Identification of the best model for each product.
  2. Estimation of the group dependence structure.
  3. Test of the differences among the groups.

[6] Davino, C., Vistocco, M. (2017). Handling heterogeneity among units in Quantile Regression. *Statistics and its interface*, forthcoming.

Relating specific attributes to overall liking - Global Model

	OLS	$\theta = 0.25$	$\theta = 0.50$	$\theta = 0.75$
Appearance	0.21	0.21	0.17	0.16
Flavor	0.64	0.71	0.67	0.58
Texture	0.15	0.15	0.17	0.16

All coefficients are significant at $\alpha = 0.10$.

The QR standard errors are estimated by *xy*-pair bootstrap.

Global model results:

- *Flavor* is the main driver of *overall liking*, followed by *appearance* and *texture*.
- The impact of *flavor* decreases in the higher part of the distribution ($\theta \geq 0.50$), i.e. the less satisfied consumers are more influenced by this attribute.
- *Appearance* slightly outperforms *texture*, but only in the lowest quantile ($\theta = 0.25$).

Quantile Regression in consumer study

Current study

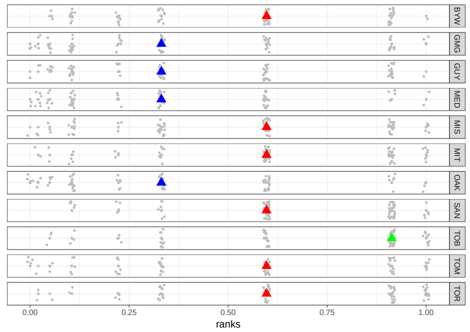
- QR to relate specific liking attributes to overall liking - **Global Model**
- QR-based strategy to detect **product effect** in such a relation:[7]
  1. Identification of the best model for each product.
  2. Estimation of the group dependence structure.
  3. Test of the differences among the groups.

[7] Davino, C., Vistocco, M. (2017). Handling heterogeneity among units in Quantile Regression. *Statistics and its interface*, forthcoming.

Relating specific attributes to overall liking - Product effect

1. Find the best model for each product: the product reference quantile

Median by product of the rank percentiles of each statistical unit with respect to the response variable: ${}_g\theta^{best}$ $g = 1, \ldots, G$



Best model for each product

$\theta = 0.33$: MED, OAK, GMG, GUY

$\theta = 0.60$: MIS, BYW TOM, MIT, SAN, TOR

$\theta = 0.91$: TOB

Relating specific attributes to overall liking - Product effect

2. **Estimation of the products' group dependent structure**

QR is carried out on the whole sample using the θ_g^{best} quantiles, namely estimating the 3 best models: $Q_\theta(\hat{\mathbf{y}}|\mathbf{X}) = \mathbf{X}\hat{\beta}({}_g\theta^{best})$

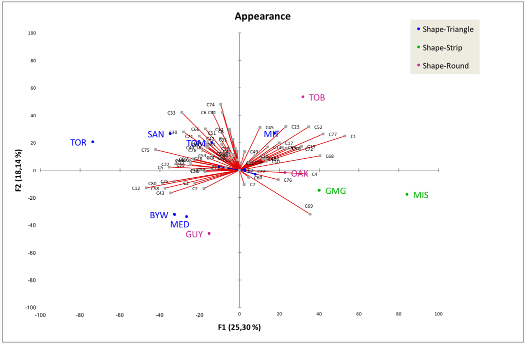
	$\theta = 0.33$ (MED, OAK, GMG, GUY)	$\theta = 0.60$ (MIS, BYW TOM, MIT, SAN, TOR)	$\theta = 0.91$ (TOB)
Appearance	0.16	0.16	0.19
Flavor	0.76	0.65	0.44
Texture	0.12	0.14	0.12

All coefficients are significant at $\alpha = 0.10$.

- *Flavor* has the highest impact on *overall liking* for all 3 models but in different size: it is higher for the first two groups of products.
- The *appearance* has the second highest effect, which becomes higher for the TOB product.

Relating specific attributes to overall liking - Product effect

	θ = 0.33 (MED, OAK, GMG, GUY)	θ = 0.60 (MIS, BYW TOM, MIT, SAN, TOR)	θ = 0.91 (TOB)
Appearance	0.16	0.16	0.19
Flavor	0.76	0.65	0.44
Texture	0.12	0.14	0.12



Relating QR best model to individual differences

- TOB's great performance could be further improved with respect to appearance.
- A remarkable segment of consumers expresses great appreciation for triangular products.

Relating specific attributes to overall liking - Product effect

3. **Testing differences among groups** [8]

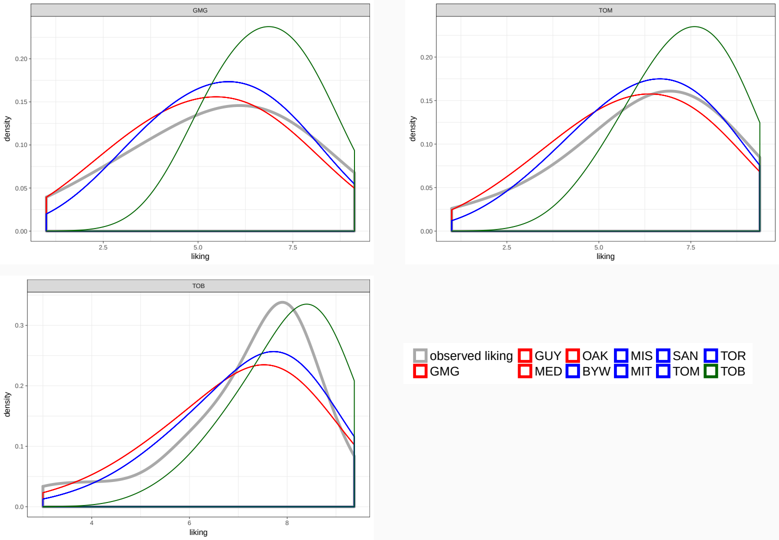
- Separate testing on each slope coefficient.
- Testing if all the slope coefficients of the groups are identical.

	Appearance	Flavor	Texture	joint test
I vs II	0.964	0.000	0.398	0.00
I vs III	0.373	0.000	0.925	0.00
II vs III	0.263	0.000	0.821	0.00

P-values derived from testing differences on each slope coefficient and on the whole model.

[8] Koenker R.W., Basset G. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, 50(1).

Comparisons among models in terms of predictions



Conclusion

Conclusions

Recap:

- Estimation of the group dependence structure
- Friendly interpretation of the final results
- Inferential tools for testing differences among groups

Future avenues:

- Extension to consumers' segmentation
- Management of multigroup structures

THANK YOU FOR YOUR ATTENTION!

Cristina Davino
Federico II Univ. of Naples
cristina.davino@unina.it

<u>Rosaria Romano</u>
University of Calabria
rosaria.romano@unical.it

Domenico Vistocco
University of Cassino
vistocco@unicas.it

STAT FOOD

Heteroschedasticity test

$$Q_{\theta_i}(\hat{\mathbf{y}}|\mathbf{x}) = \hat{\beta}_0(\theta_i) + \hat{\beta}_1(\theta_i)\mathbf{x}$$

$$Q_{\theta_j}(\hat{\mathbf{y}}|\mathbf{x}) = \hat{\beta}_0(\theta_j) + \hat{\beta}_1(\theta_j)\mathbf{x}$$

$$H_0 : \beta_1(\theta_i) = \beta_1(\theta_j)$$

Test Statistic:

$$T = \frac{\left[\hat{\beta}_1(\theta_i) - \hat{\beta}_1(\theta_j)\right]^2}{var\left[\hat{\beta}_1(\theta_i) - \hat{\beta}_1(\theta_j)\right]} \sim \chi^2_{1gdl}$$

where $var\left[\hat{\beta}_1(\theta_i) - \hat{\beta}_1(\theta_j)\right] =$
$var\left[\hat{\beta}_1\left(\theta_i\right)\right] + var\left[\hat{\beta}_1\left(\theta_j\right)\right] - 2cov\left[\hat{\beta}_1\left(\theta_i\right)\hat{\beta}_1\left(\theta_j\right)\right]$

A possible solution to estimate $var\left[\hat{\beta}_1(\theta_i) - \hat{\beta}_1(\theta_j)\right]$: bootstrap