Ordinal Lorenz Regression with application in Customer Satisfaction Surveys

Emanuela Raffinetti

Abstract Two recent proposals, the ordinal Gini measure and the Lorenz zonoid-based approach, originally employed in other fields, are here discussed and jointly illustrated to assess the customers satisfaction. The idea is focused on extending the Gini measure construction when the underlying variable describes the customer satisfaction degree. Moreover, since in the univariate case the Gini measure corresponds to the Lorenz zonoid, dependence relations between an ordinal response variable and a set of generic covariates can be analyzed through appropriate measures based on Lorenz zonoids tool. By properly combining these techniques, a novel linear regression model called "Ordinal Lorenz Regression" (OLREG) has been proposed.

Key words: Customer satisfaction, ranks, ordinal Gini measure, Lorenz zonoids.

1 Statistical approaches in Customer Satisfaction Surveys

Competitors prospering in the new global economy recognize that measuring customer satisfaction is a basic key.

In literature, many approaches providing measures able to assess the satisfaction degree among customers have been proposed. Some of them make use of statistical models to estimate the relationship between the latent variable and the manifest variables through structured equation models obtained via the Partial Least Square method (PLS), as illustrated for instance in [8]. In order to evaluate the consumer satisfaction and to make customers questionnaire calibrated, a recent development has been proposed in [3], where two alternative methods are adopted. These methods are respectively Rasch Models (RM) (see e.g. [1]) and Nonlinear Principal Components Analysis (NLPCA) (see [2]). A different contribution (Covariates in the

Emanuela Raffinetti
Department of Economics, Business and Statistics, Università degli Studi di Milano, Via Conservatorio 7, Milano, Italy, e-mail: emanuela.raffinetti@unimi.it

mixture of Uniform and shifted Binomial distributions - CUB models) has been discussed in [5] to deal directly with the probability distribution of a choice among a set of ordered alternatives and to express the stated evaluation via the subjects' covariates. In such a context, an interesting issue concerns the analysis of dependence relations among an ordinal variable and a set of generic variables.

In this paper we propose a new linear regression model characterized by an ordinal response variable, denoting the customers satisfaction degree, and a set of explanatory variables representing a service or a product evaluated items. Our proposal will be called "*Ordinal Lorenz Regression*" (OLREG), being the result of a combination between two recent research contributions illustrated in [4] and [7].

2 A novel approach: the Ordinal Lorenz Regression

As well known, the transformation process of ordered categories into numerical values represents an open issue, due to the subjectivity related to the different measure scales arbitrary choice. A possible solution to this problem has been developed in [4]. In an ordinal context [4] proposes to transform each ordered category by assigning rank 1 to the smallest observed one and rank $(r_{k-1} + n_{k-1})$ to the largest one, where r_{k-1} and n_{k-1} correspond to the rank and to the absolute frequency associated to the $(k-1)-th$ ordered category. Through the above procedure, one generates a variable with integer values expressed in terms of ranks. The new defined variable nature allows to apply the classical Gini measure with the role of variability measure.

Let us consider a variable Y describing the customer satisfaction degree. The Gini measure, here called "*ordinal Gini measure*", can be defined as:

$$G(r) = 1 - \sum_{r=1}^{k} (Q(r-1) + Q(r))(F(r) - F(r-1)), \tag{1}$$

where k is the total number of ordered categories, $F(r) = \sum_{j=1}^{i} n_j / \sum_{j=1}^{k} n_j$ and $Q(r) = \sum_{j=1}^{i} r_j n_j / \sum_{j=1}^{k} r_j n_j$, for $i = 1, \ldots, k$.

In many cases, the customer satisfaction degree depends on specific factors concerning the main features of a service or a product. The attention is focused on formalizing suitable dependence measures to establish which of the evaluated items can effectively impact on the customers satisfaction degree.

Our proposed contribution is based on a novel linear regression model, called "*Ordinal Lorenz Regression*" (OLREG), since obtained by a combination between the ordinal Gini measure and the Lorenz zonoid approach discussed in [4] and [7].

Given p explanatory variables, we aim at defining the response variable variability "explained" by each $p-th$ evaluated service or product item. In this direction, a suitable tool is recognized in the Lorenz zonoid which corresponds to the Gini measure in the univariate context (see e.g. [6]). In an ordinal framework, by denoting with Y the customer satisfaction degree, it results that $LZ(Y) = G(r)$, where

$LZ(Y)$ represents the Y variable Lorenz zonoid. The following step, consisting in detecting the product or service items that affect the customers satisfaction degree, is characterized by the application of a forward selection procedure in a multiple linear regression model. Each introduced covariate relative contribution is obtained according to the following measure:

$$RGI_{Y,X_{i+1}|X_1,\ldots,X_i} = \frac{LZ(\hat{Y}_{X_1,\ldots,X_{i+1}}) - LZ(\hat{Y}_{X_1,\ldots,X_i})}{LZ(Y) - LZ(\hat{Y}_{X_1,\ldots,X_i})}, \tag{2}$$

for $i = 1,\ldots,p-1$, under the condition that $LZ(X_i) = 0$ for $i = 0$, where $LZ(\hat{Y}_{X_1,\ldots,X_{i+1}})$ and $LZ(\hat{Y}_{X_1,\ldots,X_i})$ are the Lorenz zonoids of the Y linear estimated values obtained by considering respectively covariates $X_1,\ldots,X_{i+1}$ and $X_1,\ldots,X_i$. *RGI*[1] is the acronym of "*Relative Gini Index*". At each step the covariate with the highest *RGI* value is included into the model. The total number of covariates impacting on the customer satisfaction degree is set through the "*RGI*-scree test" (for more details, see [7]). The "*RGI*-scree test" is built on the "*RGI* scree-plot" which presents similarities with the classical scree-plot used in factor analysis. More precisely, the place where the smooth decreases and the *RGI* values appear to level off to the right of the plot defines the "*RGI*-scree". The total number of covariates to be taken into account corresponds to the number of *RGI* values preceding the scree (see Fig. 1).

3 Application: evaluating a restaurant services

This Section is devoted to check the opinion of people visiting a restaurant in the United Kingdom. The complete dataset has been proposed for the Olympic Games in Customer Satisfaction organized by University of Turin in 2009. The dataset sample analyzed here involves 338 individuals. Specifically, the survey has been based on a questionnaire where each visitor has been asked to give a score for expressing his/her overall satisfaction towards the general service provided by the restaurant. The answers are based on a four-levels ordered scale ranging from "completely unsatisfied" to "completely satisfied".

In this survey, 6 items have been investigated: *Restaurant Cleanliness*, *Drinks Offer*, *Restaurant Rooms*, *Menu Contents*, *Staff Behavior* and *Food Quality*. More in detail, all the variables denoting the above considered factors have discrete nature, in fact they assume integer values running from 1 to 10.

In Table 1, we show the *RGI* measures associated to the analyzed factors. In order to establish which items affect the customer satisfaction degree, let us focus the attention on Fig. 1. Fig. 1 shows that the covariate preceding the so called "scree" is X_1, implying that the items which mainly impact on the customers satisfaction degree are *Cleanliness*, *Drinks Offer*, *Restaurant Rooms* and *Menu Contents*.

---

[1] The *RGI* measure assumes the same role of the partial correlation coefficient in a classical multiple linear regression model.

Table 1 Results in terms of *RGI* measures

Covariate	Introduction ordering	Evaluated item	*RGI* values
X_6	1	*Restaurant Cleanliness*	$RGI_{Y\|X_6} = 0.583664$
X_4	2	*Drinks Offer*	$RGI_{Y,X_4\|X_6} = 0.251918$
X_5	3	*Restaurant Rooms*	$RGI_{Y,X_5\|X_6,X_4} = 0.080669$
X_1	4	*Menu Contents*	$RGI_{Y,X_1\|X_6,X_4,X_5} = 0.015298$
X_3	5	*Staff Behavior*	$RGI_{Y,X_3\|X_6,X_4,X_5,X_1} = 0.009130$
X_2	6	*Food Quality*	$RGI_{Y,X_2\|X_6,X_4,X_5,X_1,X_3} = 0.000016$

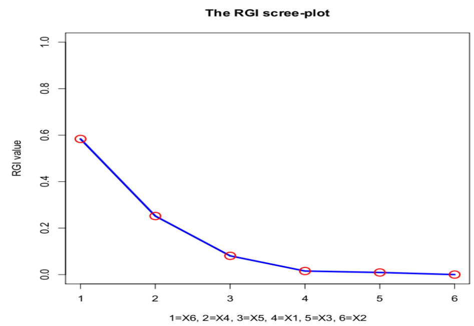

Fig. 1 *The RGI plot in customer satisfaction survey*

Acknowledgements The author acknowledges financial support from the European Social Fund Grant and Prof. Pier Alda Ferrari for her helpful suggestions.

References

1. De Battisti, F., Nicolini, G., Salini, S.: Rasch Model in Customer Satisfaction Survey Data, Quality Technology & Quantitative Management, **7**(1), 15-34 (2010)
2. Ferrari, P.A., Manzi, G.: Nonlinear Principal Components Analysis as a Tool for the Evaluation of Customer Satisfaction. Quality Technology & Quantitative Management, **7**(2), 117-132 (2010)
3. Ferrari, P.A, Salini, S.: Coplementary use of Rasch Models and Nonlinear Principal Components Analysis in the Assessment of the Opinion of Europeans About Utilities, Journal of Classification, **28**, 53–69 (2011)
4. Giudici, P., Raffinetti, E.: A Gini concentration quality measure for ordinal variables. Department Paper No. 1/2011 available on http://www-5.unipv.it/webdesed/wp/2011.1.pdf (2011)
5. Iannario, M., Piccolo, D.: A New Statistical Model for the Analysis of Customer Satisfaction. Quality Technology & Quantitative Management, **7**(2), 149–168 (2010)
6. Koshevoy, G., Mosler, K.: The Lorenz Zonoids of a Multivariate Distribution. Journal of the American Statistical Association, **91**(434), 873–882 (1996)
7. Raffinetti, E., Giudici, P.: Model selection based on Lorenz zonoids. Book of Papers ASMDA 2011 International Conference, ETS Eds., ISBN 97888467-3045-9, 1145–1152 (2011)
8. Tenenhaus, M., Vinzi, E. V., Chatelin, Y. M., Lauro, C.: PLS path modeling. Computational Statistics and Data Analysis, **48**(1), 159–205 (2005)