The anchoring vignette approach to measure customer satisfaction

The anchoring vignette approach to measure customer satisfaction

L'approccio delle anchoring vignettes per la misura della soddisfazione del cliente

Omar Paccagnella, Mariangela Guidolin, Giada Derboni, Teresa Bago d'Uva

Abstract Customer satisfaction may be defined as a self-evaluation on a product or service consumption experience. Being a self-evaluation, it may be affected by Differential Item Functioning, that is a different way in interpreting and answering questions depending on individual characteristics of respondents. Such heterogeneity among customers makes difficult the comparison among their answers. The anchoring vignette approach is a promising solution to this problem. In this paper anchoring vignettes and their model solution are applied to a customer satisfaction survey on a smartphone purchase. The main research findings concern both the results of the empirical application and the discussion of the vignette key assumptions.
Abstract *La soddisfazione del cliente può essere definita come un'autovalutazione sull'esperienza di consumo di un prodotto o servizio. Per sua natura essa può essere affetta da Differential Item Functioning, cioè un diverso modo di interpretare e rispondere alle domande a causa di un insieme di fattori individuali. Questa eterogeneità rende difficile il confronto tra consumatori differenti. L'approccio delle Anchoring Vignettes è una promettente soluzione al problema. In questo lavoro, tale approccio e il corrispondente modello vengono applicati alla valutazione della soddisfazione dopo l'acquisto di uno smartphone. I risultati principali concernono sia l'applicazione empirica sia la discussione delle assunzioni relative a tale approccio.*

Key words: Anchoring Vignettes, Customer Satisfaction, Chopit Model, Differential Item Functioning, Response Scales

---

Omar Paccagnella
University of Padua, Via Battisti 241, Padova e-mail: omar.paccagnella@unipd.it

Mariangela Guidolin
University of Padua, Via Battisti 241, Padova e-mail: guidolin@stat.unipd.it

Giada Derboni
University of Padua, Via Battisti 241, Padova e-mail: giada.derboni@studenti.unipd.it

Teresa Bago d'Uva
Erasmus University Rotterdam, 3000 DR Rotterdam, e-mail: marreirosbagoduva@ese.eur.nl

1 Introduction

Customer satisfaction is one of the most widely studied concepts in marketing: indeed, it may be argued that satisfying customers is one of the major goals of a firm, [7]. The ability to know how customers define satisfaction and to realise a post-consumption segmentation on this basis is a primary need of managers. However, despite ample research produced in this field, marketing scholars have not yet developed a unique definition of customer satisfaction. [5] gave a relevant contribution to the understanding of the mechanisms underlying satisfaction: in particular, he theorized that customers form their expectations about the product (or service) before consumption, experience the performance of it, and compare this performance with formed expectations. Satisfaction is based on this comparison: specifically, when performance exceeds expectations, there is a *positive disconfirmation*, and satisfaction arises. On the contrary, expectations larger than performance, give rise to *negative disconfirmation*, that is to dissatisfaction. This model, generally indicated as the *discrepancy paradigm*, still represents a literature cornerstone.
Customer satisfaction may be generally measured through surveys (for a review, see [7]) and the typical instruments adopted in this context are measurement scales [2]. One of the critical points of surveys refers to respondents' heterogeneity: indeed, respondents may differ for cultural, social and demographic aspects, which may entail a different way in interpreting questions and in using response scales. Such problem has been termed in the literature as Differential Item Functioning (DIF). The presence of DIF may systematically bias the measurement of variables and the relative assessment. As customer satisfaction is based on self-reporting, the DIF may represent an issue. In this paper we face the problem of DIF in customer satisfaction measurement through the *Anchoring Vignette* approach. An introduction of this methodology in this literature has been provided by [6]. The anchoring vignette methodology is an original and innovative type of questionnaire introduced by [4] in the field of social sciences. Vignettes may help to identify systematic differences in the use of response scales within groups or market segments and produce a DIF-free measurement. The paper is organized as follows. In section 2 we introduce the vignette methodology, highlighting its key assumptions, and specify the related statistical model. In section 3 we present an empirical application concerning the satisfaction after a smartphone purchase. Section 4 is left for some final remarks.

2 The Anchoring Vignette approach: definition and key assumptions

Self-assessments provided by respondents are typically affected by DIF. Anchoring vignettes are additional questions contained in the survey that may be employed to obtain a DIF-free measurement of the variable of interest. Each vignette describes a stylised situation, and respondents are asked to evaluate it, by using the same scale

adopted for self-evaluations. These additional answers provide the anchor to estimate the individual DIF and therefore produce DIF-free evaluations, that may be compared across individuals pertaining to different categories or market segments. The validity of the vignette approach relies on two key assumptions concerning measurement: *response consistency* and *vignette equivalence*. The assumption of *response consistency* states that each individual applies the same response scale for both self-evaluation and vignette evaluation. The *vignette equivalence* assumes that all respondents perceive in the same way and on the same one-dimensional scale the level of the variable represented in any vignette. Testing the validity of these assumptions is an open issue in the literature, because no formal tests (without assuming auxiliary assumptions) have been developed so far. [8] and [1] are currently the best solutions available for testing response consistency and vignette equivalence assumptions respectively.

The parametric approach dealing with vignette data is called *chopit* (Compound Hierarchical Ordinal Probit) model. It basically consists of a joint modelling of self-assessed and vignette answers by means of an ordered probit modelling approach. Let Y_i^* be the (unobserved) *perceived* own level of the concept of interest for respondent i ($i = 1,\ldots,n$). We assume it is the result of a linear specification

$$\begin{aligned} Y_i^* &= X_i\beta + \varepsilon_i \\ \varepsilon_i &\sim N(0,1) \end{aligned} \tag{1}$$

where X_i are exogenous variables, β is the vector of coefficients to be estimated (without constant for identification) and ε_i is an independent and identically distributed error term. Respondent i turns the continuous unobserved level into a reported category Y_i (recorded as an ordered variable), by means of a model with individual-specific thresholds τ_i^k

$$Y_i = k \;\; if \;\; \tau_i^{k-1} \leq Y_i^* < \tau_i^k$$

where $-\infty = \tau_i^0 < \tau_i^1 < \ldots < \tau_i^K = \infty$. Thresholds are modelled as a function of some exogenous variable V_i (which may overlap X_i) and a vector of parameters γ:

$$\begin{aligned} \tau_i^1 &= \gamma^1 V_i \\ \tau_i^k &= \tau_i^{k-1} + exp\left(\gamma^k V_i\right) \;\; k = 2,\cdots,K-1 \end{aligned} \tag{2}$$

where the exponential assumption guarantees that thresholds increase with k. Since the information provided by the self-assessments does not allow to identify the parameter vectors β and γ separately, the answers to the vignettes are exploited to overcome this problem. Let Z_{ij}^* be the (unobserved) *perceived* level of the concept of interest described in vignette j ($j = 1,\ldots,J$) for respondent i. According to the *vignette equivalence* assumption, the true level of the variable described in each vignette is perceived in the same way by all respondents. Hence, each vignette equation is defined as a function of a vignette-specific intercept plus an independent and

identically distributed error term (independent of ε_i, X_i and V_i):

$$\begin{aligned} Z^*_{ij} &= \theta_j + u_{ij} \\ u_{ij} &\sim N(0, \sigma^2_u) \end{aligned} \tag{3}$$

As before, respondent i turns the continuous unobserved level into a reported category Z_{ij}, by means of a threshold model with individual-specific thresholds τ^k_i

$$Z_{ij} = k \;\; if \;\; \tau^{k-1}_i \leq Z^*_{ij} < \tau^k_i$$

According to the *response consistency* assumption, the thresholds τ^k_i are the same as the self-assessment equation. As a consequence, self-assessed and vignette questions are asked on the same scale and this allows to identify threshold and vignette dummy parameters (up to scale and location normalisation) from the vignettes' equation alone and β parameters from the self-assessment equation alone. In order to control for individual unobserved heterogeneity, [3] extend the standard version of the *chopit* model including a random individual effect in the thresholds' equation.

3 Empirical application: customer satisfaction in smartphone purchase

In this section we present the main results of the empirical application, concerning the satisfaction after a smartphone purchase. Data come from the LISS panel, specifically from the VECS project (Vignette Evaluation of Customer Satisfaction), and have been collected by CentERdata. The LISS panel is a representative sample of Dutch individuals taking part to monthly web surveys. The first wave of the VECS project was realized in November 2011 and the application here presented refers to the third wave collected in March 2013.
The main results of the application according to the extended *Chopit* model and a standard ordered probit are illustrated for comparative purposes. Both models analyse ordered responses, but the ordered probit settles the same thresholds for all the individuals, while the *Chopit* estimates different thresholds according to individual characteristics. Satisfaction is evaluated by a 5-point Likert scale, from "Very satisfied" to "Very dissatisfied". The explanatory variables included in the models are of two types: some describe respondents' demographic and socio-economic conditions (*Gender, Age, Household size, Partner, Dwelling, Urban, Work, Degree*) while others refer to the smartphone purchase experience (*Group A, Wave 3, Group Wave 3*). In particular, the variable *Group A* refers to the fact that respondents have been divided into two groups (A and B) in order to test a possible difference in changing the order between self-assessment and vignette questions (in *Group A* self-evaluations are asked before the vignettes). The variable *Wave 3* takes value 1 if the respondent purchased the smartphone between May 2012 and March 2013 and 0 for smart-

phones purchased before May 2012. The variable *Group Wave 3* describes the interaction between the two previous variables.
Table 1 shows parameter estimates in the following order: the standard order probit with related thresholds, estimates of the self-evaluation equation (*Chopit*) and related threshold parameters. The estimated coefficients allow to determine the effect of each explanatory variable on the dependent variable: in the presence of a positive sign of an estimate, the larger the variable, the lower the probability of being satisfied, and *vice versa* when a negative sign appears. A formal test conducted on threshold coefficients confirms that these are different and depend on individual characteristics: this suggests that the *Chopit* model is a more suitable choice than the standard ordered probit. Focusing on specific results of the *Chopit* model we may see that the variables that most affect the self-assessment are: *Work*, *Group A*, *Wave 3* and the interaction variable *Group Wave 3*. Thus, persons with a job are more satisfied than those without a job, as well as persons that bought the smartphone more recently. Moreover, individuals pertaining to Group A and Wave 3 experience a higher level of satisfaction. The analysis of threshold coefficients suggests some interesting insights. On the one hand, the variable *Age* is significant in the first and fourth threshold equations, with a negative and positive sign respectively: so as age increases, the first threshold decreases and the fourth one increases, which would indicate that older persons are more likely to use intermediate categories, rather than extreme ones. On the other hand, *Age* is not significant in the self-reported equation, differently from the ordered probit findings and this allows to highlight how a *Chopit* model may better explain the mechanisms behind the development of an own judgement. The variable *Household size* is significant in the first and second threshold equation, with negative and positive sign respectively: so, as the number of households increases, the first threshold gets lower and the second higher. In other words, respondents living in larger households are more likely to assess themselves satisfied, while singles are more likely to be "very satisfied". We may notice that the variable *Work* exerts a significant effect on the third threshold equation, with negative intensity, as in the self-evaluation equation.
The validity of the response consistency and vignette equivalence assumptions are tested according to the solutions proposed by [8] and [1] respectively. In particular, response consistency is tested exploited a comparison between the expectations before buying the product and its perceived performance after the purchase for each respondent (such information is collected in the LISS panel together with the vignette answers). The validity of both assumptions cannot be rejected.

4 Final remarks

In this paper, we have applied the anchoring vignettes approach and investigated its validity, in the measurement of customer satisfaction with reference to a smartphone purchasing. Using a sample of Dutch respondents, we can state that this approach

Table 1 Ordered probit and *Chopit* models: determinants of satisfaction. (*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1)

Variable	Ordered Probit Model	Chopit model				
		Self-assessment	Thresholds			
			γ_1	γ_2	γ_3	γ_4
Gender	-0.069	-0.138	-0.071	0.026	-0.025	-0.227
Age	0.071***	0.027	-0.053***	0.009	0.006	0.256*
Household size	0.056	0.023	-0.058**	0.051***	-0.004	-0.102
Partner	-0.191*	-0.056	0.178**	-0.090	0.038	0.648*
Dwelling	-0.154	-0.171	-0.017	0.028	-0.215	-0.126
Urban	-0.025	0.057	0.096*	-0.048	0.067	0.303
Work	-0.172**	-0.260***	-0.070	0.011	-0.306	-0.656
Degree	-0.033	-0.041	-0.046	0.088*	-0.090	0.324
Group A	0.181*	0.244**	0.077	-0.099*	0.000	0.789**
Wave3	-0.051	-0.280**	-0.248***	0.011	0.170	0.027
Group Wave3	-0.376**	-0.403**	-0.028	0.078	-0.017	-0.613
Constant	–	–	-0.002	0.181	-0.129	-0.961
τ_1	-0.101	–	–	–	–	–
τ_2	1.290	–	–	–	–	–
τ_3	1.790	–	–	–	–	–
τ_4	2.662	–	–	–	–	–

may help with the DIF correction in the field of customer satisfaction, revealing aspects which are hidden in a standard ordered probit solution.

References

1. T. Bago d'Uva, M. Lindeboom, O. O'Donnell, and E. Van Doorslaer. Slipping anchor? testing the vignettes approach to identification and correction of reporting heterogeneity. *J Hum Resour*, (46):875–906, 2011.
2. Y. Kao, L. Huang, and M. Yang. Effects of experiential elements on experiential satisfaction and loyalty intentions: a case study of the super basketball league in taiwan. *International Journal of Revenue Management*, (1):79–96, 2007.
3. A. Kapteyn, J. Smith, and A. van Soest. Vignettes and self-reports of work disability in the united states and the netherlands. *Am Econ Rev*, (97):461–473, 2007.
4. G. King, C. Murray, J.A. Salomon, and A. Tandon. Enhancing the validity and cross-cultural comparability of measurement in survey research. *Am Polit Sci Rev*, (98):191–207, 2004.
5. R. L. Oliver. *Satisfaction: A Behavioural Perspective on the Consumer*. The McGraw-Hill Companies, New York, 1997.
6. O. Paccagnella. A new tool for measuring customer satisfaction: the anchoring vignette approach. *Statistica Applicata- Italian Journal of Applied Statistics*, (23):425–442, 2015.
7. R.A. Peterson and W.R. Wilson. Measuring customer satisfaction: Fact and artifact. *J Acad Market Sci*, (20):61–71, 1992.
8. A. van Soest, L. Delaney, C. Harmon, A. Kapteyn, and J. Smith. Validating the use of anchoring vignettes for the correction of response scale differences in subjective questions. *J Roy Stat Soc A Sta*, (174):575–595, 2011.