Intern. J. of Research in Marketing 27 (2010) 293–307

Contents lists available at ScienceDirect

Intern. J. of Research in Marketing

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

Estimating aggregate consumer preferences from online product reviews

Reinhold Decker [a,*], Michael Trusov [b,1]

[a] Bielefeld University, Department of Business Administration and Economics, Universitätsstraße 25, 33615 Bielefeld, Germany
[b] University of Maryland, Robert H. Smith School of Business, 3454 Van Munching Hall, College Park, MD 20742-1815, USA

ARTICLE INFO

Article history:
First received in 10, September 2009 and was under review for 4 and 1/2 months

Area Editor: Jacob Goldenberg

Keywords:
Consumer behavior
Electronic commerce and internet marketing
Market analysis and response
Preference analysis
Product management
User generated content

ABSTRACT

Today, consumer reviews are available on the Internet for a large number of product categories. The pros and cons expressed in this way uncover individually perceived strengths and weaknesses of the respective products, whereas the usually assigned product ratings represent their overall valuation. The key question at this point is how to turn the available plentitude of individual consumer opinions into aggregate consumer preferences, which can be used, for example, in product development or improvement processes.

To solve this problem, an econometric framework is presented that can be applied to the mentioned type of data after having prepared it using natural language processing techniques. The suggested methodology enables the estimation of parameters, which allow inferences on the relative effect of product attributes and brand names on the overall evaluation of the products. Specifically, we discuss options for taking opinion heterogeneity into account. Both the practicability and the benefits of the suggested approach are demonstrated using product review data from the mobile phone market. This paper demonstrates that the review-based results compare very favorably with consumer preferences obtained through conjoint analysis techniques.

1. Introduction

The basic relevance of consumer preferences, e.g., in connection with new product development processes, is widely confirmed in marketing research and practice. Typically, consumer preferences are estimated by means of conjoint analysis using online or paper-and-pencil surveys. However, this type of preference elicitation can easily become expensive in terms of time and money. Moreover, the quality of the data resulting from consumer surveys directly depends on the willingness of the respondents to participate in the particular study and the length or complexity of the questionnaire (De Leeuw & de Heer, 2002; Groves, 2006). Thus, it is worthwhile to consider alternative possibilities for eliciting aggregate consumer preferences.

Promising data sources that can be used for the measurement task at hand are online forums, particularly in the form of online shops with additional feedback options. In their recent review, "Beyond Conjoint Analysis: Advances in Preference Measurement," Netzer et al. (2008) explicitly stimulate the use of external sources of data, such as product reviews on the Internet, to measure consumer preferences.

1.1. Product reviews as a source of preference data

The non-negligible role of online communities in marketing, and particularly in the context of new product development, has already been discussed by Balasubramanian and Mahajan (2001) as well as Nikolaus and Piller (2003). Schweidel, Rindfleisch, O'Hern, and Antia (2010, p. 27) found that user-generated content in the form of reports, requests, and revisions related to open source software brings both rewards and risks and, therefore, needs "to be cultivated carefully."

Empirical findings also support the notion that online consumer reviews can be a good proxy for overall word-of-mouth[2] (Zhu & Zhang, 2010), which, in turn, can have a strong influence on the decision-making processes of other potential buyers, who search the Internet for product information (Chevalier & Mayzlin, 2006; Urban, 2005). Zhu and Zhang (2010) found that 24% of Internet users access online reviews prior to paying for a service delivered offline. Thus, the conclusion by Hu, Liu, and Zhang (2008, p. 201) that "online product reviews provided by consumers who previously purchased products have become a major information source for consumers and marketers regarding product quality" is intuitively understandable. In line with this idea, Sher and Lee (2009, p. 137) stated that "online consumer reviews provide a trusted source of product information for consumers." Lee (2007), by referring to user-oriented new product

* Corresponding author. Tel.: +49 521 106 6913; fax: +49 521 106 6456.
E-mail addresses: rdecker@wiwi.uni-bielefeld.de (R. Decker), mtrusov@rhsmith.umd.edu (M. Trusov).
[1] Tel.: +1 301 405 5878; fax: +1 301 405 0146.

[2] Recent Internet-related discussions of how word-of-mouth influences consumers' actual behaviors can be found in De Bruyn and Lilien (2008) and Trusov, Bucklin, and Pauwels (2009).

0167-8116/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijresmar.2010.09.001

design, additionally highlighted that online product reviews published on the Internet enable marketers and manufacturers to draw closer to their customers.

The analysis of freely expressed customer opinions is a promising alternative to conventional survey techniques used in preference elicitation studies. Because the reviewers or respondents have not been requested to communicate their opinions but are doing so voluntarily, a high level of authenticity can be expected. By writing product reviews, consumers intentionally invest time and energy into sharing their opinions and, indirectly, providing benefits to others (Ghose & Ipeirotis, 2007). Awad and Zhang (2007, p. 153) found that “the online ratings on a website are significantly correlated with online purchases.” Similar results are also available from Li and Hitt (2008). Dellarocas, Zhang, and Awad (2007) demonstrated that the average valence assigned to online reviews provides a valid basis for predicting future movie sales and that the early volume of online reviews can be used as a proxy of early sales. In a similar context, Chintagunta, Gopinath, and Venkataraman (2010) found that the valence and not the volume of online user reviews is the main driver of box office performance in the movie market. Related findings from considering online consumer reviews and sales of music albums are reported by Dhar and Chang (2009) as well. Furthermore, Zhu and Zhang (2010) demonstrated that online reviews are more influential when consumers have relatively greater Internet experience. Moreover, because the reviewers independently decide what and when they write, trends and topics that are currently relevant for larger consumer communities can be uncovered (Gamon, Aue, Corston-Oliver, & Ringger, 2005).

Commercial websites like Epinions.com, Cnet.com or Ciao.com do not only provide access to a multitude of products but also offer easy-to-use facilities to share opinions and experiences by writing reviews in a more or less formalized way. Today, such reviews are available for virtually each product category in the consumer market, ranging from mobile phones and books to cars and package tours. Online product reviews typically consist of three elements, namely, the pros and cons that explicitly refer to the perceived strengths and weaknesses of a product, the associated product ratings (e.g., in the form of stars) and the formless comments and remarks (full text). Sometimes, online forums also provide the possibility to indicate whether or not one would recommend the respective product. In a recent study based on online restaurant, mobile phone and digital camera reviews downloaded from the Epinions.com website, Branavan, Chen, Eisenstein, and Barzilay (2009) investigated and confirmed the link between the textual descriptions of the individual opinions and the corresponding lists of key phrases (pros and cons) summarizing the reviews. Among others, they conclude that “pros/cons key phrases are an appealing source of annotations for online review texts” (p. 575). If structured summarizations of consumer opinions in the form of pros and cons are not available, then the positive or negative orientation of a review can be determined by means of sentiment classification[3] (Dave, Lawrence, & Pennock, 2003; Popescu & Etzioni, 2005; Cui, Mittal, & Datar, 2006; Wilson, Wiebe, & Hoffmann, 2009).

Because it can be assumed that the contents of the pros and cons are usually referring to those attributes of a product that have been noteworthy for the reviewer due to some positive or negative experiences, it seems convenient to focus on these data.[4] Furthermore, the individual product ratings, in a way, represent the quantitative crystallization of the qualitative terms, phrases or statements noted in the pros and cons. In a broader sense, the ratings can be assumed to be positively correlated with the unknown individual preferences. By focusing on recommender systems, Cheung, Kwok, Law, and Tsui (2003) argued that customer preferences can be predicted by analyzing the relationship between product ratings and the corresponding product attributes. Based on an empirical study of the usefulness and impact of product reviews, Ghose and Ipeirotis (2007) suggested that customer opinions provide a meaningful basis for identifying those product attributes that are important for marketing purposes. In the same sense, Lee and Bradlow (2007, p. 3) used customer reviews to generate attributes and attribute levels for preference measurement by means of conjoint analysis, using “the language of the consumer rather than that of product designers and manufacturers.” Their methodology is based on user-authored pro/con summaries in digital camera reviews from the Epinions.com website.

1.2. Alternative approaches to preference measurement

Today, a large variety of methods for preference measurement are available. With respect to the data used, these methods can be roughly divided into two groups: survey data-based approaches and behavioral data-based approaches.

Survey data-based approaches typically rely on interviewing respondents, either by means of traditional paper-and-pencil or online questionnaires. In addition to self-explicated methods, which are based on the direct elicitation of preferences for individual product characteristics, conjoint analysis (e.g., Sawtooth's Adaptive Conjoint Analysis (ACA) or choice-based conjoint analysis (CBC))[5] has turned out to be the most widespread class of methods in this area, particularly for identifying and evaluating new product concepts (Sattler & Hensel-Börner, 2003; Jiao, Simpson, & Siddique, 2007; Giesen, Mueller, Schuberth, Wang, & Zolliker, 2007).

In contrast to survey data-based approaches, behavioral data-based preference measurement does not have such a long history and has a significantly smaller number of implementations. An early and often cited contribution to this field was made by Fader and Hardie (1996), who applied a discrete choice model to point of sale scanner data in a conjoint-like manner in order to estimate consumer preferences for selected product attributes. To the best of our knowledge, preference measurement based on online consumer review data using a negative binomial regression approach is the original contribution of our research. The paper by Archak, Ghose, and Ipeirotis (2007) differs in that it suggests a hedonic regression approach for analyzing the strength and polarity of consumer review opinions and does not consider opinion heterogeneity in the below-mentioned sense. The basic potential of using an econometric framework for detecting the weight customers put on different product features is also emphasized in the outlook on further applications in a paper by Ghose, Ipeirotis, and Sundararajan (2007) on the extraction of a quantitative interpretation of opinions — the “economic value of text.”

Therefore, the remainder of the paper is organized as follows. Section 2 describes the data pre-processing steps and the final database. In Section 3, an econometric framework is presented, which we then apply to a data set from the German mobile phone market in Section 4. Finally, Section 5 concludes the paper by outlining managerial implications and future research directions.

[3] A public tool for determining the polarity (negative or positive) of given texts called SentiWordNet 3.0 is available at http://www.sentiwordnet.isti.cnr.it/.

[4] The pros and cons can be seen as the condensed associations regarding the reviewed products and, insofar, the essence of the partly lengthy full text comments. This interpretation is also supported by the fact that some authors (see, e.g., Lee & Bradlow, 2007) explicitly use the term “pro/con review summaries” in the present context. Lui, Hu, and Cheng (2005, p. 343), by introducing an analysis system for comparing consumer opinions on different products, even state that “analyzing short sentence segments (phrases) in Pros and Cons produce more accurate results” then compared to the full text comments.

[5] Some academic studies (e.g., Hauser & Toubia, 2005) have identified the ACA (Sawtooth Software, 2007) as one of the most popular conjoint methods used by business practitioners today. Because it also allows to consider larger sets of attributes (which is important in the context of this study), we used ACA for cross-validation in the empirical part of the paper.

2. Data pre-processing and attribute extraction

Even though the basic usefulness of pro/con data from online product reviews for qualitative marketing research seems plausible for various reasons, its practical usage for preference measurement is not that obvious. The main problem stems from the challenges associated with information extraction from the reviews written in natural language. To quantify the effects of the pro/con data using, for example, statistical modeling, textual information needs to be transformed into some appropriate form.

Table 1 presents a typical example of the pros and cons section of five reviews from the mobile phone product category, with 5 stars being the maximum/best rating. Although intuitively the information contained in the pros and cons columns must have a relation with the star rating given in the right column, it is not obvious how this relation can be established and quantified.

The first key challenge in this process is to accurately identify product attributes, which are referred to in the review. The problem of identifying product attributes from consumer reviews has been studied extensively in the recent past in the data mining and natural language processing communities (see, e.g., Archak, Ghose, and Ipeirotis (2007), Abulaish, Jahiruddin, Doja, and Ahmad (2009), and Wei, Chen, Yang, and Yang (2010) for recent discussions of alternative techniques). One option is the search for statistical patterns, i.e., words (particularly nouns and adjectives) and phrases that appear frequently in the reviews (Dave, Lawrence, & Pennock, 2003). Hu and Liu (2004) apply an association rule approach to identify frequent *n*-grams in product reviews, which are then used as potential attributes. This approach is also suitable for eliminating redundant words and phrases. A more recent paper on the automatic extraction of product features from online product reviews, which applies a maximum entropy approach, was published by Somprasertsri and Lalitrojwong (2008). Feldman, Goldenberg, and Netzer (2010, p. 12) suggested an information extraction approach called CARE (*C*onditional random fields *A*ssisted *R*elation *E*xtraction), which combines “an automatic name entity recognition process” with “manually crafted rules to define relationships between entities.” Thus, it is less astonishing that a systematic comparison of two state-of-the-art algorithms for extracting features from product reviews on different consumer goods (including mobile phones) by Ferreira, Jakob, and Gurevych (2008, p. 151), among others, resulted in the conclusion that “the choice of algorithm to use depends on the targeted dataset.”

Starting from the remarkable spectrum of techniques available today and according to corresponding suggestions published in the relevant literature, the following procedure proves feasible in the present context:

1. Review-wise partitioning of the pros and cons into individual words and phrases.
2. Elimination of those words and phrases that neither point to explicit nor implicit product attributes[6] (e.g., the phrase “nothing else” in Table 1).
3. Aggregation of redundant words and phrases (e.g., reduction of “superb camera” and “good camera” in the pro category to “camera”).
4. Transformation of implicit candidate attributes (e.g., “expensive”) into explicit ones (“price”).
5. Mergence of synonyms (e.g., pooling of “accumulator” and “battery” to “battery”).
6. Elimination of those candidate attributes that are less frequent (e.g., with frequency $<1\%$).
7. Binary coding of the pro/con summaries using the available set of attributes.

Table 1
Typical pros and cons of mobile phones.

Product[a]	Pros	Cons	Rating
A	“Superb camera, decent MP3 player and FM Radio”	“Low internal memory, no GPS”	****
B	“Good features and applications, good looking phone.”	“Touch screen too sensitive, poor battery life, poor flash facility on camera”	***
C	“Good camera, lots of features”	“Expensive. Nothing else.”	*****
D	“Size and weight. Internet enabled Java applications.”	“Accumulator, external screen.”	****
E	“Looks fabulous, great touch screen feature”	“Low battery life, terrible reception”	**

[a] In order to protect data privacy, the original product names have been anonymized.

Table 2
Structure of nominally coded product reviews.

Review	Product	Attribute					Overall rating
		$l=1$	$l=2$	$\cdots$	$l=L-1$	$l=L$	
$k=1$	E	pro	mv	$\cdots$	con	mv	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$
$k=K$	C	con	pro	$\cdots$	mv	mv	5

Because of the broad literature on feature extraction from product reviews and because data pre-processing (including attribute identification) is not the focus of this paper, we abstain from a further discussion of the relevant techniques and refer the reader to the above-mentioned papers and the included literature reviews instead.

The procedure described above results in a data set structured according to Table 2. Here, the labels “pro” and “con” indicate whether the respective attribute l $(l=1,\ldots,L)$ occurs in the pro or con section of review $k(k=1,\ldots,K)$. Otherwise, attribute l is coded as a missing value (“mv”). The suggested data model can easily be extended if further information, such as consumer recommendations (“Yes” or “No”), is available.

Depending on the required depth of analysis, the variable “product” can be split into subcategories brand and model (e.g., when considering technical products such as mobile phones or digital cameras). In our empirical example we explicitly make use of the brand as an additional variable for explaining the overall evaluation of a product. “Nokia E90 Communicator”, for example, can be split into brand name “Nokia” and model “E90 Communicator.”

3. Econometric preference analysis using product review data

In the following subsections, we outline a simple methodology for determining the effects attributes have on the products' overall evaluation, e.g., in view of future product improvement or new product development. The basic idea behind this approach is that each product review (consumer opinion) can be represented by a combination of positively (the “pros”) and negatively (the “cons”) valued characteristics, completed by an overall evaluation of the product. We consider three formulations of the model: homogenous preference model, heterogeneous model with a discrete distribution of preferences (latent class), and heterogeneous model with a continuous distribution of preferences.

[6] In the present context, two types of product attributes are distinguished. Explicit attributes may occur as nouns (e.g., “screen”) or nominal combinations (e.g., “touch screen”), whereas implicit attributes are expressed as adjectives (e.g., “expensive”).

3.1. Modeling homogeneous preferences

Let $k=1,\ldots,K$ be the subscript used to identify the individual reviews, $l=1,\ldots,L$ denote the subscript of the functional (product) attributes,[7] and $h\in\{1,2\}$ be the respective level or value of attribute l. Using this notation, the nominally coded opinion data (see Table 2) can be converted into binary data, with

$$x_{klh}=\begin{cases}1, & \text{if functional attribute } l \text{ (e.g. "camera") takes level } h \\ & \text{(e.g. bproQ) in review k} \\ 0, & \text{otherwise}\end{cases} \tag{1a}$$

and $x_{kl1}+x_{kl2}\leq 1\,\forall\, k,l$.[8] To explicitly account for brand name effects, we further introduce a second set of variables

$$\tilde{x}_{km}=\begin{cases}1, & \text{if the product discussed in review } k \text{ belongs to brand } m \\ 0, & \text{otherwise}\end{cases} \tag{1b}$$

with $m=1,2,\ldots,M=$ number of brands considered and $\sum_{m=1}^{M}\tilde{x}_{km}=1\forall k$.

In many product categories, at least the leading brands usually offer several variants or models of the product concerned. Even though the individual variants may differ according to the functional attributes, the evaluation of the products as a whole will also be influenced by the image of the respective supplier captured by the brand name. The above operationalization explicitly accounts for this effect.

Referring to the popular approach of conjoint analysis, the observed evaluation (rating)[9] y_k of the product being discussed in review k is defined as a function of the relevant features. If y is regarded as the realization of a random variable Y, then the following Poisson regression (PR) approach with conditional mean $\lambda>0$ can be specified[10]:

$$\text{Prob}^{PR}(Y=y)$$
$$=\frac{\lambda^y}{y!}\exp(-\lambda)\quad \text{with } \lambda=\exp\left(\alpha+\sum_{l=1}^{L}(\beta_{l1}x_{l1}+\beta_{l2}x_{l2})+\sum_{m=1}^{M}\delta_m\tilde{x}_m\right) \tag{2}$$

The above equation determines the probability of observing product evaluation y given a certain pro/con summary coded by means of explanatory variables x and $\tilde{x}$.

The Poisson assumption can be motivated, among other things, by the fact (Archak, Ghose, & Ipeirotis, 2007, p. 63) that products that repeatedly get bad reviews "tend to disappear from the market quickly, and do not get many further negative evaluations." Accordingly, extremely low ratings will rather be the exception in the long term, and therefore, a normal distribution as required for OLS regression cannot be expected to be inevitable. Furthermore, the exponential function ensures positive responses for all values of the explanatory variables. However, Poisson regression requires that the data do not suffer from overdispersion, i.e., at least approximate equality of expectation and variance.

The unknown parameters $\Phi^{PR}=(\alpha,\beta_{11},\ldots,\beta_{L2},\delta_1,\ldots,\delta_M)$ of the above model can be estimated by maximizing the following log-likelihood function:

$$LL\left(\Phi^{PR}\right)=\ln\left(\prod_{k=1}^{K}\frac{\lambda^{y_k}}{y_k!}\exp(-\lambda)\right)=\sum_{k=1}^{K}y_k\ln(\lambda)-\lambda-\ln(y_k!) \tag{3}$$

Accordingly, the fit of the PR model can be evaluated using the Akaike information criterion (AIC) and common significance measures.

Because of the dichotomous definition of the explanatory variables, the parameter estimates $\hat{\beta}_{11},\ldots,\hat{\beta}_{L2}$ and particularly $\hat{\delta}_1,\ldots,\hat{\delta}_M$ can be interpreted in a similar way to those resulting from conjoint analysis. $\hat{\beta}_{21}$, for example, quantifies the estimated relationship regarding strength and direction between the occurrence of attribute 2 in the pro section and the probability of observing (weighted) product rating y. In a similar way, $\hat{\delta}_2$ represents the power or value implicitly assigned to brand 2 by the consumers (after the effects of the functional attributes have been taken into account) and therewith quantifies the estimated influence of the second brand's name on product evaluation. Previously, Degeratu, Rangaswamy, and Wu (2000) explicitly considered the effect of the brand name on choice behavior as the "intangible value" of a product, whereas the "tangible value" is attributed to the levels of functional product attributes.

3.2. Modeling heterogeneous preferences

In consumer goods industries, many product categories are characterized by a large number of varieties, mostly offered to serve heterogeneous needs and preferences. Therefore, heterogeneity has been a basic issue in consumer behavior modeling for decades. In this subsection, we discuss how the basic PR model can be extended to account for possible heterogeneity patterns. One option is negative binomial regression, and the other is latent class Poisson regression.

3.2.1. The negative binomial regression approach

Because a certain amount of variation in the observed data can be attributed to unknown sources (apart from variables x and $\tilde{x}$), the use of a suitable statistical distribution for representing this "unobserved" heterogeneity appears promising. The negative binomial distribution used below results from the homogeneous Poisson distribution by assuming that the conditional mean λ follows a Gamma distribution. A detailed substantiation of the negative binomial distribution for modeling heterogeneity in consumer behavior was previously provided by Ehrenberg (1959) as well as Wagner and Taudes (1987), for instance. In addition to this behavior-oriented motivation, there is also a more technical one. According to Janes et al. (2006), a negative binomial regression does better than a Poisson regression if the data are overdispersed because it allows the variance of the dependent variable to exceed the conditional mean.

Therefore, let ε be a random variable that accounts for the "unobserved" heterogeneity characterizing the data set to be analyzed and that is not correlated with x and $\tilde{x}$ but satisfies $E(\varepsilon)=0$. We can then replace the conditional mean λ with a random variable Λ, defined as

$$\Lambda=\lambda\cdot\exp(\varepsilon)=\lambda\cdot\tilde{\varepsilon}\quad \text{with } E(\Lambda)=\lambda \tag{4}$$

The probability of observing a certain value y can then be computed using the following mixture distribution:

$$\begin{aligned}\text{Prob}(Y=y)&=\int_0^\infty\frac{(\lambda\tilde{\varepsilon})^y}{y!}\exp(-\lambda\tilde{\varepsilon})\cdot\frac{\gamma_1^{\gamma_2}\tilde{\varepsilon}^{\gamma_2-1}}{\Gamma(\gamma_2)}\exp(-\gamma_1\tilde{\varepsilon})d\tilde{\varepsilon}\\ &=\frac{\gamma_1^{\gamma_2}\lambda^y}{\Gamma(\gamma_2)y!}\cdot\int_0^\infty\tilde{\varepsilon}^{\,y+\gamma_2-1}\exp(-(\lambda+\gamma_1)\tilde{\varepsilon})d\tilde{\varepsilon}\\ &=\binom{y+\gamma_2-1}{y}\cdot\left(\frac{\lambda}{\lambda+\gamma_1}\right)^y\cdot\left(\frac{\gamma_1}{\lambda+\gamma_1}\right)^{\gamma_2}\end{aligned} \tag{5}$$

---

[7] In product configuration design, a functional attribute is a feature used to specify a unique variant of a product family (Erens, 1996). For instance, the Samsung SGH-U600 mobile phone is equipped with an FM player, whereas the Samsung SGH-F480 is equipped with an MP3/MPEG-4 player. That is, both models, albeit belonging to the same brand, are distinguishable on the basis of the functional attribute "equipment/functionality" (see Section 4). Analogous differentiations are possible across brands of course.

[8] If attribute l is not referred to in review k, then $x_{kl1}+x_{kl2}=0$ holds.

[9] By weighting the star rating with a possibly existing consumer recommendation ("Yes" or "No") the significance of the product evaluation can be increased. A simple way of doing this weighting is $[\mathit{rating}\cdot\exp(1)]$ if "Yes" and $[\mathit{rating}\cdot\exp(-1)]$ if "No".

[10] Please note that for identification purposes in the model with an intercept, we need to set δ_M to zero.

with $\gamma_1, \gamma_2 > 0$. Finally, assuming the scale parameter γ_1 and the shape parameter γ_2 to be identical, for analytical convenience (Long & Freese, 2006), leads to a common form of the negative binomial regression (NBR) approach:

$$\text{Prob}^{\text{NBR}}(Y = y) = \binom{y + \gamma - 1}{y} \cdot \frac{\lambda^y \gamma^\gamma}{(\lambda + \gamma)^{y + \gamma}} \quad \text{with } \gamma > 0 \tag{6}$$

The inverse of the Gamma parameter γ is sometimes called the dispersion parameter because an increase of $1/\gamma$ increases the conditional variance of Y (Kleiber & Zeileis, 2008). The unknown parameters $\Phi^{NBR} = (\gamma, \alpha, \beta_{11}, \ldots, \beta_{L2}, \delta_1, \ldots, \delta_M)$ can be estimated by the maximum likelihood method.

3.2.2. The latent class Poisson regression approach

In our second approach, we assume the observations to be drawn from a finite mixture of Poisson distributions. Heterogeneity is explicitly considered by postulating that each review belongs to one class i ($i = 1, \ldots, I \geq 2$), which is a priori unknown (Wedel & DeSarbo, 1995; Wedel, DeSarbo, Bult, & Ramaswamy, 1993; Leisch, 2004). The point-masses of the discrete mixing distributions are denoted by π_i, with $\sum_{i=1}^{I} \pi_i = 1$ and $0 < \pi_i < 1$ for all i. Furthermore, the coefficients of the explanatory variables x and $\tilde{x}$ are assumed to vary across the given classes. Starting once again from a discrete mixture of I Poisson distributions of the individual evaluations y, the following latent class Poisson regression (LCPR) model can be specified:

$$\begin{aligned}&\text{Prob}^{\text{LCPR}}(Y = y)\\ &= \sum_{i=1}^{I} \pi_i \frac{\lambda_i^y}{y!} \exp(-\lambda_i) \quad \text{with } \lambda_i = \exp\left(\alpha_i + \sum_{l=1}^{L} (\beta_{il1} x_{l1} + \beta_{il2} x_{l2}) + \sum_{m=1}^{M} \delta_{im} \tilde{x}_m\right)\end{aligned} \tag{7}$$

Here, $\text{Prob}^{\text{LCPR}}(Y = y)$ denotes the probability of observing evaluation y in a review that belongs to one of the assumed classes i. To estimate the $I(2 + 2L + M)$ unknown parameters $\Phi^{LCPR} = (\pi_1, \ldots, \pi_I, \alpha_1, \ldots, \alpha_I, \beta_{111}, \ldots, \beta_{IL2}, \delta_{11}, \ldots, \delta_{IM})$, the following log likelihood function can be maximized:

$$LL\left(\Phi^{LCPR}\right) = \sum_{k=1}^{K} \ln\left(\sum_{i=1}^{I} \pi_i \frac{\lambda_i^{y_k}}{y_k!} \exp(-\lambda_i)\right) \tag{8}$$

3.3. Measures of the relative strength of effect

The above models can be used to estimate the relative strength of the effects resulting from the pro/con attributes and the brand names on the overall evaluation of the respective products. For demonstration purposes the following considerations are limited to the non-clustering case, but analogous specifications are also possible for the LCPR model.

Based on the specification of the conditional mean λ (Eq. (2)), a simple indicator

$$\phi^{Attribute}(x_l) = \left(\exp\left(\hat{\beta}_{l1}\right) + \exp\left(\hat{\beta}_{l2}\right) - 2\right) \times 100\% \tag{9}$$

can be defined for each pro/con attribute l.[11] Eq. (9), which we call the “backlog of impact,” measures the average sensitivity of the product evaluation to variations of the quality of functional attributes. Accordingly, for $\hat{\beta}_{l1}$, representing the estimated effect of observing attribute l in the pro section, and $\hat{\beta}_{l2}$, analogously denoting the con effect, the above indicator shows how much more (percent-wise) a confirmation of attribute l would increase the overall evaluation of the product compared to a decrease due to rejection. A negative $\phi^{Attribute}(x_l)$ value indicates that an impairment with respect to attribute x_l has a higher influence on product evaluation than an improvement of the same magnitude. Referring to the popular Kano model, an attribute with negative backlog of impact can be called a basic or must-be feature (Matzler & Hinterhuber, 1998). If the product does not fulfill customers' requirements regarding this attribute, then this will cause an overall dissatisfaction reflected in a lower rating. However, because the customers may take these requirements for granted, their fulfillment will not increase their satisfaction in a significant way. The reverse relationship is also possible and has been documented as excitement features in the literature.

A similar measure can also be defined for the brand variables. Here, the relative strength of effect between two brands m_1 and m_2

$$\phi^{Brand}\left(\tilde{x}_{m_1}, \tilde{x}_{m_2}\right) = \left(\exp\left(\hat{\delta}_{m_1}\right) - \exp\left(\hat{\delta}_{m_2}\right)\right) \times 100\% \tag{10}$$

allows for tentative inferences about the relative brand value or power from the consumer's perspective. For example, $\phi^{Brand}(\tilde{x}_{m_1}, \tilde{x}_{m_2}) = 25$ would indicate that the contribution of m_1's brand name to the overall product evaluation is 25% higher than the one of m_2's brand name. If the ϕ values of the competing brands exhibit significant differences, then the respective product category (or at least opinion formation) is characterized by a noticeable sensitivity to brand names.

4. Empirical study

To demonstrate the benefits of the proposed methodology for preference measurement, a data set containing more than 20,000 online product reviews covering four leading brands in the German mobile phone market are considered in the following. Each review contains both a pro/con summary and a product rating, often accompanied by a partially comprehensive comment written as continuous full text and a “Yes” or “No” recommendation. According to Branavan, Chen, Eisenstein, and Barzilay (2009) the pro/con phrases do not only summarize the continuous textual description of the opinion but are also representative of the semantic properties of the latter. In a recent study on mining opinions from comparative sentences in consumer reviews, Ganapathibhotla and Liu (2008, p. 245) explicitly motivate the use of pro/con summaries instead of continuous texts because “the brief information in pros and cons contains the essential information related to opinions …” and “… the user opinion on the product feature is clear.” Therefore, continuous text is ignored in this study as well. Including the information made available by the full texts may further improve the explanatory power of the subsequent analyses but also involves a more comprehensive data preparation procedure and is therefore left to future exploration.

We decided to consider mobile phones in the empirical study because this is a comparatively new product category that still shows an interesting feature development. This, in turn, ensures the existence of a sufficient number of online reviews presenting meaningful contents. Furthermore, because mobile phones can be seen as high-involvement products and because the level of involvement also determines how detailed the reviews are (Grønhaug, 1977), it can be assumed that a broad spectrum of aspects is considered, which enlarges the space of candidate attributes that can be extracted from the reviews.

Before being able to apply the above methodology to these data, the pros and cons must be pre-processed according to the procedure outlined in Section 2. In doing so, a total of 23 attributes were extracted (see Table 3), each featuring a relative frequency of at least 1%. All attributes, though to a different degree, can be assumed to play a role in word-of-mouth communication. These data are used to calibrate the models (PR, NBR and LCPR) outlined in Section 3.

Although the development of a text mining algorithm was not the objective of this study, we have performed an independent experiment to test the performance of the adopted classification algorithm.

[11] For derivations, please, refer to the Appendix B.

Table 3
Extracted functional attributes and associated relative frequencies [in %].

Attribute	Freq.	Attribute	Freq.	Attribute	Freq.
Size/weight	33.5 (34.8)	Menu navigation	8.9 (10.0)	Interfaces	4.1 (4.4)
Appearance ("look")	23.2 (23.5)	Camera	7.8 (10.2)	Robustness/ stability	3.9 (5.8)
Equipment/ functionality[a]	21.7 (22.9)	Memory	7.2 (8.4)	Manufacturing quality	3.7 (7.6)
Battery	16.8 (20.1)	Other functions	6.7 (7.8)	Software	3.2 (6.3)
Price	16.0 (17.1)	Mobile Internet	6.7 (8.2)	Radio technology	3.0 (3.4)
Operation/handling	15.5 (18.4)	Keypad/ touch screen	6.2 (7.3)	Sound	2.5 (3.3)
Display	12.6 (14.2)	Messaging	5.8 (3.9)	Reliability	1.0 (1.9)
Multimedia	9.0 (11.0)	Reception/ voice quality	5.8 (8.3)		

[a] Functionality is referred to in the meaning of "variety of functions."

For this experiment, we randomly drew 7 non-overlapping subsamples from the original set of reviews, containing 600 reviews each. Then, a group of 7 independent test persons with an average experience with mobile phone usage (ownership and use for more than 12 months) were asked to identify relevant attributes in the selected pro/con summaries. Within a four-hour time period, one human coder, on average, was able to process 476 reviews. Although the resulting list of attributes identified across the reviews by human coders had some discrepancies with the list generated according to the procedure described above, the agreement between the two lists is very high. In Table 3, the corresponding relative frequencies of the 23 attributes identified by the coders are given in brackets.[12] The apparent similarity of the two frequency profiles, characterized by high (about 0.99) correlation, confirms the robustness of the presented classification.

4.1. Parameter estimation and fit

Applying the models to the available product review data ($N=20{,}419$) leads to the parameter estimates displayed in Table 4 (values rounded to improve readability). A comparatively large number of attributes prove to be significant for each model, at least at the 10% level (two-tailed p-values). As to be expected, the Akaike information criterion decreases when switching from the PR model ($AIC=124159$) to the NBR model ($AIC=122775$). Obviously, the implicit consideration of unobserved heterogeneity seems to be appropriate in the present case. The usefulness of applying the NBR model also finds its expression in a significant dispersion value[13] of 1.395. Thus, it is also less surprising that the computation of a multi-cluster solution by means of the LCPR model further improves the fit measure.

The two-cluster solution exhibits a remarkably lower $AIC=103069$, which, at first glance, suggests the use of a multi-cluster solution. The two classes are characterized by significantly different sizes. Cluster 1 comprises 88.3% of the observations, cluster 2 only about 11.7%. However, only the intercept (-0.1545) proves significant at level 0.1 in cluster 2. This asymmetry also occurs when the number of classes is increased. In the three-cluster solution (with $AIC=102971$), the main cluster once again represents 88.3% of the observations, whereas in the 4-cluster case (with a negligibly worse $AIC=103107$) the respective value slightly declines to 86.3%. Because of this quite uniform structure of the LCPR outputs, only the parameter vector of the main cluster of the two-cluster solution is displayed in Table 4.[14]

A first look at the different parameter vectors reveals some noticeable aspects. First, the intercepts are highly significant in each case and are of the same magnitude. For all three models, the coefficients of the brand variables show the same relations. Obviously, the included brands are characterized by remarkable differences regarding their reputation in the given competitive environment. Brand 2 clearly dominates consumer perception. Furthermore, most of the product attributes, starting from the attribute "battery" and ending with the attribute "reliability," feature plausible signs, i.e., they are negative for the con and positive for the pro level.[15] Only three attributes, namely "messaging," "mobile Internet," and "multimedia," seemingly do not have a significant effect. This finding, together with the fact that the NBR model seems to be a good compromise between the basic homogeneous PR model and the explicit clustering approach in terms of LCPR, triggered the re-calibration of the former one by excluding the above triple of attributes. The results are given in Table 5 (values rounded).

As to be expected, the estimates for the reduced NBR model (with $AIC=122767$) are very similar to the original one. Interestingly, brand 3 now also features a significant coefficient, which is in line with the corresponding results of the full PR and LCPR models. Furthermore, for all product attributes, at least one of the levels is statistically significant; thus, it is worth maintaining all attributes in the following analyses. Only the weak significance of the negative coefficient of the pro level of the price might be astonishing at first glance but can be explained by the special definition of the functional attributes. Its occurrence in the pro category might be due to the fact that the respective mobile phones are rated as low-priced or even cheap. A cheap mobile phone, however, must not necessarily imply high satisfaction (i.e., high product ratings), particularly if cheapness is achieved at the expense of manufacturing quality and reliability, for instance. This supposition is supported by the fact that if only pro/con summaries with a bad or medium rating (1 to 3 stars) are considered, then 67.7% of all mentions of the term "price" occur in the pro category (versus 32.3% in the con category).[16]

The estimated parameters $\hat{\beta}_{1,1},\ldots,\hat{\beta}_{20,2}$ and $\hat{\delta}_1,\ldots,\hat{\delta}_4$ have the following meaning: the contribution of brand 2's name to the overall evaluation of a mobile phone relative to the base brand 4 equals $\exp\left(\hat{\delta}_2\right) = \exp(0.036) = 1.037$ and thus is the strongest of all brand names. The smallest value, 0.795, results for brand 1. Because of the exponential structure of the conditional mean λ, the contribution for the base brand is set to 1 (as $\exp(0)=1$). Hence, for all other brands, a value smaller than 1 implies a relatively negative effect, whereas a value larger than 1 points to a positive contribution.

Analogously, the con level of the attribute "reliability," with value $\exp\left(\hat{\beta}_{20,2}\right) = 0.416$, has the strongest effect of all functional attributes on the dependent variable. Deterioration with respect to this feature, for instance, as a consequence of the use of cheaper and therefore potentially less durable components, may significantly impair the evaluation of the respective mobile phones. The second strongest effect (with $\exp\left(\hat{\beta}_{19,2}\right) = 0.659$) comes from the con level of the attribute "manufacturing quality." In contrast to these effects, the corresponding positive effects of the pro levels amount to $\exp\left(\hat{\beta}_{20,1}\right) = 1.070$ and $\exp\left(\hat{\beta}_{19,1}\right) = 1.077$, respectively. In conclusion, the average strength of the positive effect of the pro

[12] In addition to the 23 attributes mentioned in Table 3, the following 8 attributes (with absolute frequencies given in brackets) were identified by the coders: "antenna" (31), "vibration/alarm" (30), "technology/age" (27), "cover" (23), "service" (20), "accessories" (17), "reaction/charging time" (17), and "scope of delivery" (16). Because the relative frequencies of these attributes were below 1%, following the same threshold rule as discussed in Section 2, we dropped these from further analysis.

[13] Dispersion values significantly larger than 1 explicitly support the use of a NBR approach (see Cameron and Trivedi, 1990; Kleiber and Zeileis, 2008 for details).

[14] For $I \geq 5$, parameter estimation was impossible with the current version of the R package on a 3.5 GB RAM computer due to memory capacity problems.

[15] A similar concordance of the parameters of a PR, NBR and LCPR (3 clusters) model, applied to a data set resulting from a direct mailing activity in the Netherlands, was also reported by Wedel, DeSarbo, Bult, and Ramaswamy (1993).

[16] The specific role of price information in connection with product ratings in online reviews is discussed in a recent paper by Kuksov and Xie (2010).

Table 4
Parameter estimates of the three models.

	PR model			NBR model			LCPR model (class 1)		
Variable	Estimate	Std. error	*p* value	Estimate	Std. error	*p* value	Estimate	Std. error	*p* value
Intercept	2.429	0.008	***	2.428	0.011	***	2.545	0.008	***
Brand 1	−0.225	0.010	***	−0.228	0.012	***	−0.069	0.010	***
Brand 2	0.036	0.080	***	0.036	0.010	***	0.015	0.008	.
Brand 3	−0.022	0.011	*	−0.022	0.014	0.11	−0.020	0.011	.
Brand 4	-	-	-	-	-	-	-	-	-
Pro_Battery	0.041	0.008	***	0.042	0.010	***	0.013	0.008	.
Con_Battery	−0.069	0.008	***	−0.072	0.010	***	−0.043	0.008	***
Pro_Equipment/Functionality	0.071	0.005	***	0.074	0.007	***	0.018	0.005	***
Con_Equipment/Functionality	−0.182	0.012	***	−0.189	0.014	***	−0.075	0.012	***
Pro_Operation/Handling	0.066	0.006	***	0.069	0.008	***	0.012	0.006	*
Con_Operation/Handling	−0.244	0.018	***	−0.247	0.022	***	−0.080	0.018	***
Pro_Display	0.024	0.007	**	0.025	0.010	**	0.006	0.008	0.45
Con_Display	−0.111	0.012	***	−0.115	0.015	***	−0.049	0.012	***
Pro_Reception/VoiceQuality	0.015	0.011	0.16	0.015	0.014	0.27	0.002	0.011	0.88
Con_Reception/VoiceQuality	−0.270	0.019	***	−0.275	0.023	***	−0.086	0.019	***
Pro_Size/Weight	0.050	0.005	***	0.054	0.007	***	0.013	0.005	*
Con_Size/Weight	−0.062	0.006	***	−0.064	0.008	***	−0.037	0.006	***
Pro_Camera	0.075	0.010	***	0.077	0.013	***	0.024	0.010	*
Con_Camera	−0.078	0.013	***	−0.079	0.016	***	−0.061	0.013	***
Pro_Sound	0.051	0.015	***	0.054	0.019	**	0.013	0.015	0.40
Con_Sound	−0.065	0.031	*	−0.066	0.040	.	−0.024	0.031	0.44
Pro_MenuNavigation	0.017	0.008	*	0.019	0.011	.	−0.001	0.008	0.93
Con_MenuNavigation	−0.228	0.018	***	−0.235	0.022	***	−0.059	0.019	**
Pro_Messaging	0.001	0.013	0.94	0.001	0.017	0.95	−0.001	0.013	0.92
Con_Messaging	−0.013	0.013	0.29	−0.016	0.016	0.32	−0.015	0.013	0.23
Pro_MobileInternet	0.007	0.012	0.56	0.006	0.015	0.69	0.007	0.012	0.58
Con_MobileInternet	0.017	0.012	0.14	0.016	0.015	0.28	0.005	0.012	0.66
Pro_RadioTechnology	0.033	0.016	*	0.036	0.021	.	0.020	0.016	0.22
Con_RadioTechnology	−0.005	0.022	0.81	−0.004	0.028	0.90	−0.015	0.022	0.49
Pro_Multimedia	0.014	0.009	0.11	0.016	0.011	0.17	−0.006	0.009	0.50
Con_Multimedia	0.005	0.013	0.68	0.006	0.016	0.73	−0.028	0.013	*
Pro_Appearance	0.013	0.005	*	0.013	0.007	.	0.008	0.005	0.14
Con_Appearance	−0.116	0.016	***	−0.120	0.020	***	−0.050	0.016	**
Pro_Price	−0.020	0.008	*	−0.020	0.011	.	−0.028	0.009	**
Con_Price	0.009	0.007	0.21	0.008	0.010	0.40	0.005	0.007	0.49
Pro_Robustness/Stability	0.015	0.011	0.17	0.017	0.015	0.25	−0.021	0.011	.
Con_Robustness/Stability	−0.170	0.038	***	−0.173	0.048	***	−0.074	0.039	.
Pro_Interfaces	0.030	0.015	*	0.032	0.019	.	0.012	0.015	0.43
Con_Interfaces	−0.039	0.015	**	−0.039	0.019	*	−0.029	0.015	*
Pro_Software	0.014	0.027	0.60	0.015	0.035	0.67	0.024	0.027	0.37
Con_Software	−0.268	0.015	***	−0.272	0.019	***	−0.084	0.015	***
Pro_OtherFunctions	0.033	0.011	**	0.035	0.014	*	0.011	0.011	0.30
Con_OtherFunctions	−0.072	0.014	***	−0.074	0.018	***	−0.026	0.014	.
Pro_Memory	0.059	0.011	***	0.062	0.014	***	0.014	0.011	0.18
Con_Memory	−0.016	0.012	0.19	−0.015	0.016	0.33	−0.027	0.012	*
Pro_ Keypad/TouchScreen	−0.022	0.018	0.22	−0.019	0.023	0.42	−0.012	0.018	0.52
Con_ Keypad/TouchScreen	−0.031	0.010	**	−0.031	0.013	*	−0.032	0.010	**
Pro_ManufacturingQuality	0.072	0.013	***	0.074	0.017	***	0.021	0.013	0.11
Con_ManufacturingQuality	−0.411	0.024	***	−0.418	0.029	***	−0.117	0.025	***
Pro_Reliability	0.065	0.022	**	0.069	0.028	*	0.015	0.022	0.49
Con_Reliability	−0.869	0.119	***	−0.875	0.133	***	−0.115	0.136	0.40
AIC	124159			122775			103069		

Significance levels (two-tailed *p* values): ***=0.001 / **=0.01 / *=0.05 / .=0.1.
The grey-colored attributes are excluded from further consideration because of insufficient significance.

levels is less distinct than the negative effect of the con levels ($mean\left(\exp\left(\hat{\beta}_{.1}\right)\right) = 1.039$ versus $mean\left(\exp\left(\hat{\beta}_{.2}\right)\right) = 0.860$). This structure can be regarded as evidence of the generally high expectations of the consumers concerning mobile phones and the consequently low willingness to accept any quality reductions in the future. In Subsection 4.3, we are going to deepen these aspects.

Furthermore, the distinctiveness of the con category points to a certain propensity in the mobile phone market considered for word-of-mouth communication in the case of dissatisfaction. In an empirical study on a population's propensity to engage in post-purchase online word-of-mouth, using movie data, Dellarocas and Narayan (2006) also found evidence for concluding that extreme (dis)satisfaction correlates with a higher propensity to discuss a product online. However, because both studies do not establish causality, further research is needed to prove or disprove this hypothesis. The cross-validation outlined in Subsection 4.4 also addresses the plausibility of the measured effects by referring to ACA part-worth utilities.

4.2. Predictive accuracy of the NBR approach

The focus of the above discussions was on explanation. However, because models and theories, despite their explanatory power, may be “subject to potential predictive falsification” (Shugan, 2009, p. 992), we investigated the predictive power and robustness of the suggested framework by carrying out cross-validation simulations. Therefore, the whole data set was repeatedly divided into a calibration set

Table 5
Parameter estimates of the reduced NBR model.

Variable	Estimate	Std. error	*p* value
Intercept	2.429	0.011	***
Brand 1	−0.229	0.012	***
Brand 2	0.036	0.010	***
Brand 3	−0.023	0.014	.
Brand 4	–	–	–

Variable	Estimate	Std. error	*p* value	Variable	Estimate	Std. error	*p* value
Pro_Battery	0.043	0.010	***	Pro_Appearance	0.013	0.007	.
Con_Battery	−0.072	0.010	***	Con_Appearance	−0.120	0.020	***
Pro_Equipment/Functionality	0.074	0.007	***	Pro_Price	−0.020	0.011	.
Con_Equipment/Functionality	−0.188	0.014	***	Con_Price	0.008	0.010	0.39
Pro_Operation/Handling	0.068	0.008	***	Pro_Robustness/Stability	0.016	0.015	0.26
Con_Operation/Handling	−0.247	0.022	***	Con_Robustness/Stability	−0.173	0.048	***
Pro_Display	0.026	0.010	**	Pro_Interfaces	0.035	0.019	.
Con_Display	−0.115	0.015	***	Con_Interfaces	−0.039	0.019	*
Pro_Reception/VoiceQuality	0.015	0.014	0.27	Pro_Software	0.015	0.035	0.67
Con_Reception/VoiceQuality	−0.276	0.023	***	Con_Software	−0.272	0.019	***
Pro_Size/Weight	0.054	0.007	***	Pro_OtherFunctions	0.037	0.014	**
Con_Size/Weight	−0.064	0.008	***	Con_OtherFunctions	−0.074	0.018	***
Pro_Camera	0.080	0.013	***	Pro_Memory	0.064	0.014	***
Con_Camera	−0.078	0.016	***	Con_Memory	−0.018	0.015	0.23
Pro_Sound	0.053	0.019	**	Pro_ Keypad/TouchScreen	−0.017	0.023	0.46
Con_Sound	−0.066	0.040	.	Con_ Keypad/TouchScreen	−0.031	0.013	*
Pro_MenuNavigation	0.019	0.011	.	Pro_ManufacturingQuality	0.074	0.017	***
Con_MenuNavigation	−0.235	0.022	***	Con_ManufacturingQuality	−0.418	0.029	***
Pro_RadioTechnology	0.037	0.021	.	Pro_Reliability	0.068	0.028	*
Con_RadioTechnology	−0.003	0.028	0.93	Con_Reliability	−0.877	0.133	***

$AIC=122767$; significance levels (two-tailed p values): $^{***}=0.001/^{**}=0.01/^{*}=0.05/.=0.1$.

(containing 19,419 observations/reviews used for parameter estimation) and a prediction set (containing 1000 observations/reviews used for computing holdout hit rates). 20 random splits of the original data set lead to the results displayed in Table 6. For the underlying five-point rating scale, the random hit rate equals 1/5 ≡ 20%. However, since the distribution of ratings in our sample is asymmetric across star categories (with shares of 51%, 34%, 10%, 4% and 1% for 5-, 4-, 3-, 2- and 1-star ratings, respectively) we also calculated the proportional chance criterion (38.9% hit rate) as another benchmark metric (Morrison, 1969; Hair, Anderson, Tatham, & Black, 1998). Across 20 random splits, the average hit rate of the proposed NBR model is 51.5%, which is about 2.5 times higher than the hit rate by chance. Compared to the proportional chance criterion, the proposed model offers a relative improvement of 32.4%. Furthermore, all Pearson correlations r (average value = 0.38) between the observed and estimated holdout values are highly significant (two-tailed $p<0.01$). Both the hit rates and the correlations are rather robust across splits.

Furthermore, if we transform the five-point scale into three categories (low, medium, and high valuation), the corresponding holdout hit rates vary between 73.4% and 77.6% with an average value of 75.4% (random hit rate = 33.3%). This relation equals an improvement factor of 2.3 and also means a clear outperformance of the proportional chance criterion (54.99% hit rate). All in all, the available hit rates are in accordance with common standards (see, e.g., Moore (2004) for corresponding values in different conjoint validation studies). Although the ability to predict the dependent variable in a satisfactory way is not sufficient to substantiate the overall adequacy of the suggested NBR approach, this at least provides a certain level of confidence (Shugan, 2009).

4.3. Relative strength of effects and implications

To gain further insights into the explanatory power of the suggested approach, we computed the relative strength of the effects resulting from the different brands and the functional attributes on the overall perception of mobile phones. Applying Eqs. (9) and (10) to the NBR parameters at hand leads to the results displayed in Tables 7 and 8. In both cases, concrete managerial implications can be drawn, for instance, regarding the targeted improvement of already existing products or the customer-oriented development of new products.

The value 24.10% in the first numerical column of Table 7, for instance, indicates that the relative strength of the effect of brand 2's name on the overall evaluation of a mobile phone is about 24% higher than the one of brand 1. The name of brand 2 clearly influences the perception of the products concerned in a positive way and therefore seems suitable for well-directed image campaigns and brand image transfers in new product introduction processes. The values of suppliers 3 and 4 are quite similar, which point to a comparable brand power. A similar interpretation of the regression coefficients of binary-coded brand variables estimated on the basis of grocery store scanner data can be found in Martínez-Garmendia (2010). This paper measures relative brand equity by exponentially transforming the dummy coefficients and relating them to a baseline brand. The resulting values are interpreted as relative preference ranks, making them comparable to the above measures. The basic suitability of

Table 6
Holdout hit rates in terms of overall evaluations.

No. of split	Hit rate [%] *r*	No. of split	Hit rate [%] *r*	No. of split	Hit rate [%] *r*	No. of split	Hit rate [%] *r*
1	50.4; 0.46	6	51.7; 0.40	11	50.3; 0.41	16	50.0; 0.33
2	50.1; 0.39	7	53.6; 0.39	12	53.9; 0.37	17	51.4; 0.32
3	49.7; 0.39	8	54.8; 0.36	13	52.8; 0.40	18	52.0; 0.39
4	50.1; 0.37	9	51.8; 0.42	14	52.1; 0.34	19	50.6; 0.41
5	51.1; 0.37	10	49.8; 0.38	15	53.1; 0.31	20	51.6; 0.43

Table 7
Percentage backlog of impact (brands).

$\phi^{Brand}(.)$	Brand 1	Brand 2	Brand 3	Brand 4
Brand 1	0.00	−24.10	−18.23	−20.45
Brand 2	24.10	0.00	5.87	3.64
Brand 3	18.23	−5.87	0.00	−2.22
Brand 4	20.45	−3.64	2.22	0.00

Table 8
Percentage backlog of impact (functional attributes).

Attribute	Backlog
Battery	−2.6
Equipment/functionality	−9.5
Operation/handling	−14.8
Display	−8.2
Reception/voice quality	−22.6
Size/weight	−0.7
Camera	0.8
Sound	−0.9
Menu navigation	−19.0
Radio technology	3.5
Appearance	−9.9
Price	−1.1
Robustness/stability	−14.2
Interfaces	−0.3
Software	−22.3
Other functions	−3.4
Memory	4.8
Keypad/touch screen	−4.8
Manufacturing quality	−26.5
Reliability	−51.3

regression-based approaches for brand value estimation can also be inferred from a comprehensive review of brand measurement studies by Kartono and Rao (2008), but no example of using online product reviews is highlighted there.

Taking into account recent discussions in the relevant literature, a brand's relative strength can be considered as an indicator of its relative value. Accordingly, the presented methodology enables a tentative verification of the (relative) brand values repeatedly reported by commercial management consultancies and market research institutes. In addition, it particularly allows for the assessment of smaller brands, which are typically not the subject of such standard analyses. In contrast to the elicitation of brand values from household panel data (see, e.g., Kamakura and Russell, 1993), online opinion data of the type at hand enables virtually up-to-date computations of the interesting figures. A two-tailed *t*-test based on the percentage effects rejects the null hypothesis that the true mean is equal to 0%, which points to a noticeable sensitivity of the considered mobile phone market to brand names.

The interpretation of the values depicted in Table 8, although computed in a similar way as for the brands, is somewhat different because of the pro/con definition of the attributes. Apart from "camera," "radio technology" and "memory," all attributes have negative backlog values, which means that in all of these cases, an impairment of the attribute's perception would influence the evaluation more than an improvement of the same magnitude would. This conclusion particularly applies to the attribute "reliability." The present value of −51.3% emphasizes both the existing sensibility to this product feature and the obviously existing high standard in this regard. This finding is also supported by the fact that the respective attribute only rarely (see Table 3 and Fig. 1) occurs in a product review at all. Products that do not fulfill the requirements concerned may cause extreme dissatisfaction and risk being the subject of repeated negative (online) word-of-mouth. Due to the increasing tendency to use the Internet for comparing and purchasing products (Hu, Liu, & Zhang, 2008), the economic consequences of unfavorable word-of-mouth communications/reviews can be considerable.[17] A value close to 0 (as with the "interfaces") indicates that product modifications in both directions would result in effects of the same strength.

In Fig. 1, the estimated percentage backlog of impact is plotted against the observed relative frequency of the attributes in the pro/con summaries. The higher the negative backlog of impact, the more important it is that modifications regarding the attributes concerned are executed with care. The dashed lines mark the corresponding "centers" (17.3 and −23.3) between the respective extremes. Accordingly, four areas can be identified. With a view to further improvement or new development of a mobile phone, the areas can be interpreted as follows.

The attributes located in area I, particularly those in the lower left corner, are not that predominant in online word-of-mouth communications. Obviously, they are not really crucial for the evaluation of a mobile phone and, hence, modifications of these attributes would probably be noticed only en passant. The comparatively weak effect of the price in the lower center can be explained, among other things, by the fact that in the German market a high percentage of mobile phones are sold on a contract with a mobile network operator, making the actual price of the device itself an attribute that can hardly be evaluated by the customers and often does not matter because of the virtual bundling of device and network. The comparatively low transparency of the predominant pricing policies seemingly implies low elasticity regarding the original manufacturer's price.

The attributes "equipment/functionality," "appearance" and "size/weight," located in area IV, are obviously popular topics of online reviews but with rather modest effect on the final evaluation of a mobile phone. Taking into account that mobile phones nowadays are often multifunctional devices that are not only used for phone calls (primary usage) but also, for example, for gaming, entertainment, mobile banking (secondary usage), the available result regarding the first attribute appears plausible. In a very recent study focusing on the Korean mobile phone market Lee, Lee, and Cho (2009) could empirically demonstrate that multifunctional mobile phones have the potential to drive replacement purchases and broaden the mobile phone market. However, the more functions a device offers, the more important the ease of "operation/handling" becomes for obvious reasons. This fact probably also motivates the attribute's closeness to the position of "equipment/functionality."[18] Although both "appearance" and "size/weight" are comparatively often discussed in online product reviews, their negative impact in case of dissatisfaction is rather moderate and low, respectively. A plausible justification of this finding might be the fact that both aspects are easy to control in purchase decisions. Therefore, a "wrong" decision is probably first attributed to oneself and not to the product.

Finally, both the reliability and the manufacturing quality of the devices located in area II are crucial attributes in so far as the current technological progress in the mobile phone industry ensures a high standard per se. Thus, trouble in this respect is rather rare but typically results from real experience with the product. If customers meet a problem of this type, then they are all the more willing to express their displeasure in appropriate reviews.

The emptiness of area III once again supports our beliefs about the generally high quality standards provided by the leading mobile phone brands. Serious shortcomings, which repeatedly attract the attention of larger numbers of customers, do not exist (see also the remark for motivating the Poisson assumption in Subsection 3.1).

4.4. Cross-validation of the NBR results by means of ACA

To further check the plausibility and validity of the NBR results, we additionally conducted an ACA study comprising $N = 195$ respondents.[19] The sample was provided by a professional market research

[17] From a recent empirical study on customer channel migration, Ansari, Mela, and Neslin (2008) conjecture, among other things, that migration to the Internet lowers switching costs, and makes it easier to compare competing products.

[18] A study by UK-based industry analyst Freeform Dynamics published in 2009 and comprising 1500 respondents from Europe and the USA recently ranks "ease of use" as the most important buying criterion (out of a given set of criteria) for mobile phones (http://www.3g.co.uk/).

[19] ACA was used as a benchmark because "it is the industry and academic standard for within-respondent adaptive question design …" and "… appears to predict reasonably well in many situations" (Toubia, Simester, Hauser, & Dahan, 2003, p. 283).

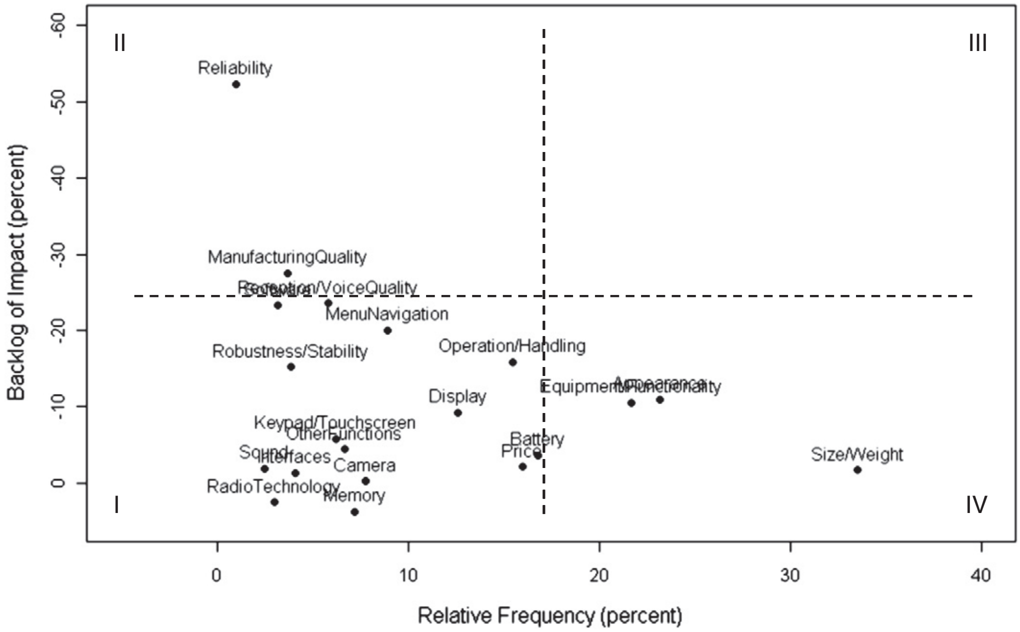

Fig. 1. NBR-based relevance map for the functional attributes.

institute and is representative with respect to sex, age, education and profession of the respondents. We deliberately did not limit our survey to the users of online reviews in order to avoid sample biasing. The attribute levels have been generated predominantly from the available continuous review texts and the pro/con summaries ("language of the consumer"). To account for possible number-of-level effects (Verlegh, Schifferstein, & Wittink, 2002), we limited the number of levels for each attribute to three, except for the brand, which features four levels. In total, 15 functional attributes have three levels, and five of them have two levels. Details are given in Appendix A.

The elicited aggregate part-worth utilities are depicted in Fig. 2. The available values have been generated by averaging the individual part-worth utilities over all respondents. The black bars represent the attribute levels mentioned first in the table depicted in Appendix A, while the grey ones refer to the levels mentioned last. For example, in the case of attribute "battery," the black bar (level: "8 to 12 days standby") equals −0.2158, the empty bar (level: "13 to 17 days standby") corresponds to −0.0244, and the grey bar (level: "18 and more days standby") equals 0.1834.

A closer look at the bar chart reveals both already known and new findings. The attributes "reliability," "manufacturing quality," "robustness/stability," "reception/voice quality," and "operation/handling" are characterized by comparatively large ranges (maximum minus minimum part-worth utilities) indicating high importance. The large range with respect to "reliability," for example, is remarkable insofar as we avoided inappropriate levels with this and other monotonous attributes. In the case of "reliability," we only used non-negative levels ("very high", "above average", and "moderate") because it can be assumed that negative levels such as "very low" or "low" would have stretched the ranges artificially. Other attributes such as "keypad/touch screen," "memory," "other functions," "radio technology," and "size/weight" are rated less important. Both observations are in line with the NBR results given in Tables 5 and 8. The ranges of the corresponding NBR parameters are included in Fig. 2 to facilitate comparisons.[20] The significant correlation between both sets of values equals 0.466 (one-tailed $p<0.05$).

[20] Please note that, due to the different measurement bases for ACA part-worth utilities and NBR model parameters, the comparison of these values is only meaningful in a relative sense within the columns.

However, there are also some interesting differences compared to the NBR estimates. The price, for example, achieves a somewhat higher importance in the ACA study than in the NBR study which, among other things, supposedly results from its explicit address in the first case. The low importance resulting from the NBR approach was already motivated in connection with the interpretation of Fig. 1. The increased ACA importance of the attribute "interfaces" might result from the rapidly rising importance of connecting mobile phones to different devices such as, computers and car's hands-free communication facilities. An interesting finding also evolves from the attribute "equipment/functionality," which features a comparatively large NBR parameter range ($\text{pro}^{\text{NBR}}=0.074$ vs. $\text{con}^{\text{NBR}}=-0.188$). The range of the corresponding ACA part-worth utilities, in contrast, is very small (see Fig. 2). The reason for this seeming discrepancy becomes obvious if we look at the individual part-worth utilities depicted in Fig. 3.

Both levels of this attribute are characterized by a distinct variance across the sample, which leads to values close to 0 for both attribute levels when aggregating the data. Additional cluster analyses of the ACA sample by means of Ward and *k*-means explicitly unfold the cause of the detected divergence. The most adequate cluster solution uncovers a clear separation of the respondents: one group nearly exclusively votes for "limitation to telephony functions," whereas the remaining respondents clearly vote for "usability as an all-in-one device." Both groups are of almost the same size. Furthermore, the importance value of "equipment/functionality" in group 2 is the second highest of all attributes considered. In contrast to this, both "reliability" and "manufacturing quality" as well as "reception/voice quality," exhibit nearly equal importances.

We speculate that another possible source of such discrepancies (see, for example, attribute "sound") is the "active" approach to measuring attribute importance in an ACA setting. In other words, questioning about the attribute creates an awareness of the attribute, which may not necessarily occur to this extent when discussing the product in an online review without this type of reminder. The problem (aided recall versus unaided recall) is well-known from advertising research but definitely needs further research in this new field. Regardless of this lack of research, a combination of both approaches obviously can help to obtain a more objective idea of what really matters concerning the interesting consumer preferences.

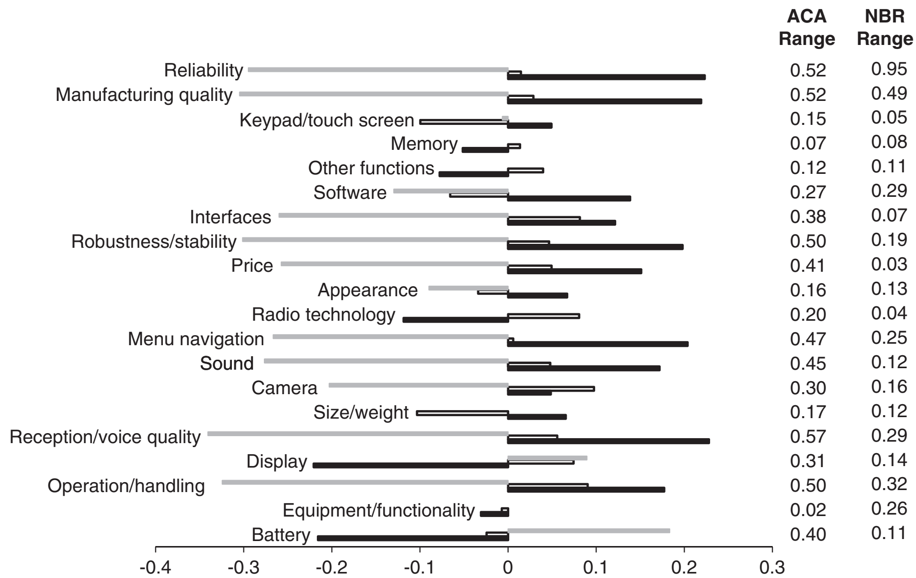

Fig. 2. ACA-based part-worth utilities (illustrated by bars) and corresponding ACA part-worth utility ranges and NBR parameter ranges.

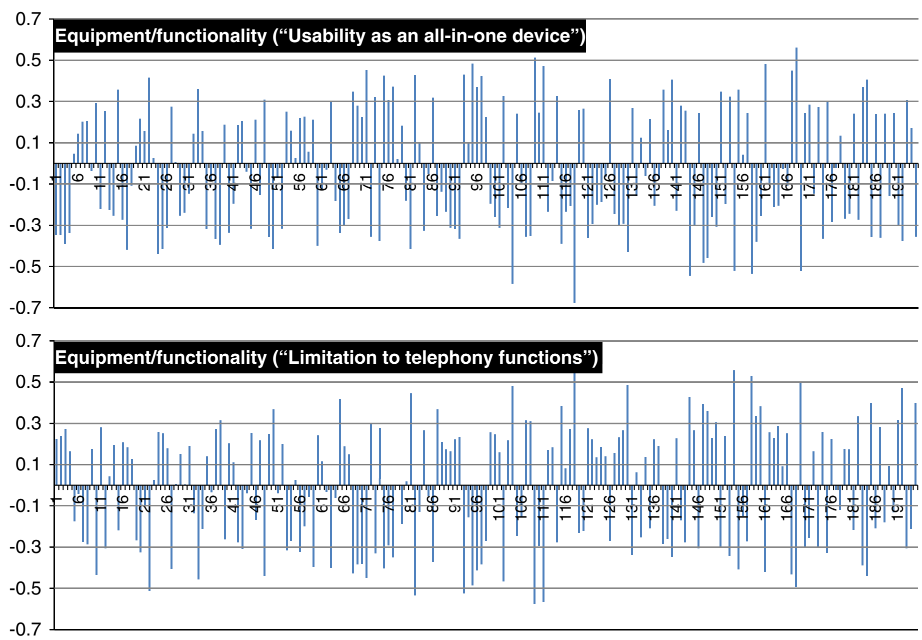

Fig. 3. Individual part-worth utilities for attribute “equipment/functionality”.

The high consistency between the NBR and ACA estimates also finds its expression in the correlation (see Table 9) between the brand parameters resulting from the two methods.

To further evaluate the ACA results, we calculated the coefficients of determination (R^2) between a ranking of the holdout sample evaluations and the ranking predicted by ACA for five mobile phones/concepts presented to each respondent in the final stage of the survey (see Green, 1984 for a similar approach). The resulting cross-sample mean value of 0.757 (with min. = 0.42, max. = 1.00, and standard deviation = 0.15) points to a strong performance (see, e.g., Huber, Wittink, Fiedler, and Miller, 1993; Toubia et al., 2003; Scholz, Meißner, and Decker, 2009 and particularly Moore, 2004). Summing up, we conclude that the survey-based ACA results strongly support the results obtained from the NBR analysis of online reviews and the pro/con patterns inferred from the NBR model align well with the part-worth utilities inferred through traditional conjoint analysis.

Table 9
Comparison of brand parameters.

Brand	NBR	ACA
2	0.036	0.067
4	0.0	0.019
3	−0.023	−0.034
1	−0.229	−0.128

Correlation = 0.950 (one-tailed $p<0.05$).

5. Conclusion and implications

In this paper, we presented an econometric framework that can be applied to turn the plentitude of individual consumer opinions made available by online product reviews into aggregate consumer preference data. After preparing the pro/con summaries that typically accompany full-text reviews by means of natural language processing techniques, the considered approaches, and in particular the negative binomial regression (NBR) model, enable the estimation of meaningful parameters. These parameters allow for inferences on the relative effect of functional attributes and brand names on product evaluations and purchase decisions.

The practicability and benefits of our methodology is demonstrated using online product reviews from a national mobile phone market. Our analyses reveal that the NBR approach can be a promising compromise between simplicity of application and the need of adequately considering opinion heterogeneity. Most of the parameter estimates are statistically significant and highly plausible in size and sign, which, in turn, enable the computation of measures that further quantify the effect the brand name or a certain product attribute has on the overall evaluation of a mobile phone. Furthermore, we demonstrate that the plausibility of the NBR results is supported by an additional ACA study using the concerning attributes. This evaluation also identifies benefits that can result when combining both methods to reach a more reliable estimation of the preferences existing in a market of interest.

The results at hand point to different directions of impact that can be followed when considering further improvements to existing mobile phone offers or developing new devices that meet current user needs. The suggested methodology proves useful not only in collecting information about consumer preferences but also promises to be helpful in reputation analysis. In addition, the elicited relative impacts of attributes can be used in online shop optimizations, particularly, by presenting those attributes in the product navigation menu that prove important in word-of-mouth communication and presumably affect purchase decisions.

However, not least because of the special type of data used, the suggested approach is not free of limitations. Our study is based on the assumption that the reviews have been written by consumers and not by professionals who might be led by commercial or even intentionally manipulative interests. The latter is currently researched as opinion spam, which, according to Jindal and Liu (2008) and Lui (2010), can be categorized as untruthful reviews (e.g., to damage the image of a product), reviews on brands only (not referring to a particular product) and non-reviews (e.g., advertisements). The authors also introduce and empirically verify techniques for detecting this type of spam using reviews from Amazon.com. Opinion spam detection is a crucial and largely unresolved problem, leaving space for intensive future research. Promising suggestions in a methodical respect can be obtained, among others, from the growing literature on fraud detection in data mining (Bolton & Hand, 2002). However, because more and more Internet opinion forums ask their users to evaluate the usefulness of the reviews, the benefits to self-interested parties from manipulating product reviews can be assumed to decrease with an increase in the number of such votes.[21] Further factors limiting the attractiveness of manipulating online reviews are discussed by Dellarocas (2006).

Another issue is self-selection, which can bias the representativeness of the available reviews. Li and Hitt (2008), for instance, explore the presence and implications of self-selection biases using online book reviews. Depending on the investigated product category, the willingness of individual consumer segments to participate in online discussions of products might be different. Particularly, if the category of interest is characterized by a high degree of innovativeness and high new product introduction rates, then the early adopters can be overrepresented. However, it is precisely these lead users who may provide interesting information regarding the optimization of subsequent variants of the basic new products. To increase the chance of adoption, the opinions of lead users can be useful both in the pre-launch stage of product development and in connection with product modification (Schreier, Oberhauser, & Pruegel, 2007).

According to the above limitations, future research should be devoted to the development of powerful filters for detecting fake reviews and to the further automation of the time-consuming data pre-processing and attribute extraction steps (see, e.g., Wei, Chen, Yang, and Yang (2010) and the literature cited there). An exciting issue might be the comparison of consumer preferences elicited from online product reviews written by consumers of different countries. Empirical findings of this type might help to manage the balancing act between product individualization according to national preference patterns and standardization in favor of measure efficiency in international marketing. In a very recent paper, Koh, Hu, and Clemons (2010), among others, address the question of how culture influences raters' behavior when writing reviews and how cultural differences are manifested in differences among ratings. Particularly, if different languages must be considered, the restriction to pro/con summaries can involve a significant facilitation in comparison to the analysis of full text reviews. However, feature extraction approaches accounting for multiple languages are still very scarce. An interesting study on the use of sentiment analysis methodologies for the classification of Web forum opinions in multiple languages (English and Arabic) was recently published by Abbasi, Chen, and Salem (2008), however, without building a bridge to the marketing context at hand.

Another promising topic is reviewer profiling. An increasing number of opinion portals not only collect the mentioned review data but also ask the reviewers to additionally provide characteristics of themselves. Moe and Schweidel (2010) empirically demonstrate that product opinions can be affected substantially by the composition of the considered consumer base and identify four different types of review posters. The in-depth investigation of the relationships between poster types and the relevance of certain attributes might open new insights into online review based market segmentation.

A more strategic issue is exploratory market structure analysis. By combining text mining techniques with a network analysis framework, Feldman, Goldenberg, and Netzer (2010) were able to transfer consumer messages posted on a car manufacturer's forum into perceptual maps of the respective market, among other things. Although they examined product attributes as well, they did not relate them to product ratings as done here. Instead, the co-occurrences of products (e.g., car models) and terms (e.g., adjectives) in the forum from which the data stem are considered. Preference patterns inferred from the NBR approach can be used to enrich such maps with respect to attribute importances.

By adding a temporal dimension to the presented work, it would be interesting to investigate how managerial actions (e.g., promotions, price changes, and new product introductions) and competition influence consumer product ratings and, in particular, the perception of attribute importance. Moreover, our framework can be used to gain a more in-depth understanding of advertising mechanisms. Although there are numerous studies that consider such outcomes of advertising as increased product sales and improved brand recognition or recall, to the best of our knowledge, the impact of advertising on product attribute valuation that leads to a subsequent change in product market performance has not been studied yet. By pairing online review data with market performance and advertising data, the proposed approach can help to shed light on this question.

Last but not least, the application of alternative, more sophisticated estimation approaches, such as Hierarchical Bayes (HB), may further improve the results to be achieved from review data. Moore (2004), for example, demonstrates in a traditional conjoint setting that an HB rating-based model may provide higher hit rates than both an OLS rating-based model and latent segment models.

[21] Epinions.com, for example, explicitly motivates its visitors to “rate others' reviews” and even selects “Top Reviewers” in each category (accessed on April 2010).

Acknowledgments

The authors would like to thank two anonymous IJRM referees as well as the Area Editor, Jacob Goldenberg, and the Editor of this journal, Marnik G. Dekimpe, for detailed comments and useful suggestions which led to a significant improvement of the article. Furthermore, we sincerely thank Dr. Kumiko Gnibba-Yukawa, Prudentia Marketing Research Ltd., for providing the data used in this study and very fruitful discussions with respect to the analysis of online reviews.

Appendix A

Attribute	Attribute levels
Brand[a]	Brand 1–Brand 2–Brand 3–Brand 4
Battery	8 to 12 days standby–13 to 17 days standby–18 and more days standby
Equipment/functionality	Usability as an all-in-one device (e.g., telephony, organizer functions, camera, and games) – limitation to telephony functions
Operation/handling	Very simple (no practice required)–simple (only minor practice required)–sophisticated (practice required)
Display	Small (<2.3 in.)–medium (2.3 to 3.7 in.)–large (>3.7 in.)
Reception/voice quality	Very good–above average–moderate
Size/weight	Usual size and weight – rather small and light
Camera	Suitability for photographing/filming at a professional level–suitability for normal/simple snapshots–no camera
Sound	Very good–above average–moderate
Menu navigation	Menu navigation for all functions–menu navigation for telephony functions only–no menu navigation
Radio technology	Availability of one common network (e.g., GSM, GPRS or UMTS)–availability of all common networks
Appearance	Elegant/sterling–youthful/modern–simple/unimposing
Price	Rather low–average–rather high
Robustness/stability	Very high–above average–moderate
Interfaces	USB or Bluetooth–all common interfaces (e.g., USB, Bluetooth, and WLAN)–no interface
Software	Organizer functions (incl. email and SMS)–organizer functions plus office applications (e.g., text processing and spreadsheet) – no additional application software
Other functions	Many secondary functions (e.g., torch and thermometer)–rather few secondary functions
Memory	Medium-sized (2 to 8 GB)–large (more than 8 GB)
Keypad/touch screen	Data input via keypad–data input via touch screen–data input via keypad and touch screen
Manufacturing quality	Very good–above average–moderate
Reliability	Very high–above average–moderate

[a]The true brand names have been used in the survey but are made anonymous here because of data privacy concerns.

Appendix B

Calculating a backlog of impact indicator for functional attributes

Eq. (2) defines the conditional mean λ as:

$$\lambda = \exp\left(\alpha + \sum_{l=1}^{L}\left(\beta_{lPRO}x_{lPRO} + \beta_{lCON}x_{lCON}\right) + \sum_{m=1}^{M}\delta_m\tilde{x}_m\right).$$

After regrouping with respect to attribute j we get:

$$\begin{aligned}\lambda &= \exp\left(\beta_{jPRO}x_{jPRO} + \beta_{jCON}x_{jCON} + \alpha + \sum_{l=1, l\neq j}^{L}\left(\beta_{lPRO}x_{lPRO} + \beta_{lCON}x_{lCON}\right) + \sum_{m=1}^{M}\delta_m\tilde{x}_m\right)\\ &= \exp\left(\beta_{jPRO}x_{jPRO} + \beta_{jCON}x_{jCON}\right)\cdot\exp\left(\alpha + \sum_{l=1, l\neq j}^{L}\left(\beta_{lPRO}x_{lPRO} + \beta_{lCON}x_{lCON}\right) + \sum_{m=1}^{M}\delta_m\tilde{x}_m\right).\end{aligned}$$

Let us call the second factor $\exp(\alpha + \ldots)$ a *base*, which can be interpreted as an average product evaluation when attribute j is not listed in any of the two sections.

Average evaluations when attribute j is listed in the pro and con section respectively are:

$$\lambda_{PRO} = \exp\left(\beta_{jPRO}x_{jPRO}\right)\cdot base \text{ and } \lambda_{CON} = \exp\left(\beta_{jCON}x_{jCON}\right)\cdot base$$

or, given a dichotomous nature of x:

$$\lambda_{PRO} = \exp\left(\beta_{jPRO}\right)\cdot base \text{ and } \lambda_{CON} = \exp\left(\beta_{jCON}\right)\cdot base.$$

Hence, the percentage increase (%GAIN) and decrease (%LOSS) from the base for attribute j are respectively:

$$\lambda_{\%GAIN} = \left(\frac{\exp\left(\beta_{jPRO}\right)\cdot base - base}{base}\right)\times 100\% = \left(\exp\left(\beta_{jPRO}\right) - 1\right)\times 100\%$$

and

$$\lambda_{\%LOSS} = \left(\frac{base - \exp\left(\beta_{jCON}\right)\cdot base}{base}\right)\times 100\% = \left(1 - \exp\left(\beta_{jCON}\right)\right)\times 100\%.$$

Note that both are defined as non-negative numbers for $\beta_{jPRO}>0$ and $\beta_{jCON}<0$.

Finally, we define the backlog of impact indicator as the difference between gain and loss:

$$\begin{aligned}\text{backlog} &= \left(\exp\left(\hat{\beta}_{jPRO}\right) - 1\right)\times 100\% - \left(1 - \exp\left(\hat{\beta}_{jCON}\right)\right)\times 100\%,\\ &= \left(\exp\left(\hat{\beta}_{jPRO}\right) + \exp\left(\hat{\beta}_{jCON}\right) - 2\right)\times 100\%\left(=: \phi^{Attribute}\left(x_j\right)\right).\end{aligned}$$

Fig. A1 illustrates the basic idea behind the backlog of impact indicator.

Calculating a backlog of impact indicator for brands

The backlog of impact indicator for brands is a comparative (to the base brand) measure and is calculated analogously to the backlog of impact indicator for functional attributes:

$$\begin{aligned}\lambda_{BRAND\_m_j} &= \left(\frac{\exp\left(\delta_{m_j}\right)\cdot base\_brand - base\_brand}{base\_brand}\right)\times 100\%\\ &= \left(\exp\left(\delta_{m_j}\right) - 1\right)\times 100\%,\end{aligned}$$

where *base_brand* is defined as:

$$base\_brand = \exp\left(\alpha + \sum_{l=1}^{L}\left(\beta_{lPRO}x_{lPRO} + \beta_{lCON}x_{lCON}\right)\right).$$

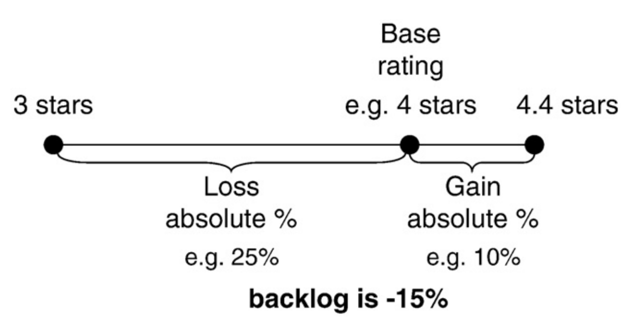

Fig. A1. Backlog of impact indicator illustrated.

Comparative measures for Table 7 are calculated by taking differences for corresponding $\lambda_{BRAND\_m_j}$:

$$\phi^{Brand}\left(\tilde{x}_{m_1}, \tilde{x}_{m_2}\right) = \left(\exp\left(\hat{\delta}_{m_1}\right) - \exp\left(\hat{\delta}_{m_2}\right)\right) \times 100\%.$$

□

References

Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Transactions on Information Systems*, *26*(3), 12—34.

Abulaish, M., Jahiruddin, Doja, M. N., & Ahmad, T. (2009). Feature and opinion mining for customer review summarization. *Pattern Recognition and Machine Intelligence — Lecture Notes in Computer Science*, *Vol. 5909*. (pp. 219—224).

Ansari, A., Mela, C. F., & Neslin, S. A. (2008). Customer channel migration. *Journal of Marketing Research*, *45*(1), 60—76.

Archak, N., Ghose, A., & Ipeirotis, P. G. (2007). Show me the Money! Deriving the pricing power of product features by mining consumer reviews. *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose* (pp. 56—65).

Awad, N. F., & Zhang, J. (2007). Stay out of my forum! Evaluating firm involvement in online ratings communities. *System Sciences — HICSS 2007 — 40th Annual Hawaii International Conference* (pp. 153—158).

Balasubramanian, S., & Mahajan, V. (2001). The economic leverage of the virtual community. *International Journal of Electronic Commerce*, *5*(3), 103—138.

Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science*, *17*(3), 235—249.

Branavan, S. R. K., Chen, H., Eisenstein, J., & Barzilay, R. (2009). Learning document-level semantic properties from free-text annotations. *Journal of Artificial Intelligence Research*, *34*, 569—603.

Cameron, A. C., & Trivedi, P. K. (1990). Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics*, *46*(3), 347—364.

Cheung, K. -W., Kwok, J. T., Law, M. H., & Tsui, K. -C. (2003). Mining customer product ratings for personalized marketing. *Decision Support Systems*, *35*, 231—243.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*(3), 345—354.

Chintagunta, P. K., Gopinath, S., and Venkataraman, S. (2010). The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. Forthcoming in *Marketing Science*. doi:10.1287/mksc.1100.0572

Cui, H., Mittal, V., & Datar, M. (2006). Comparative experiments on sentiment classification for online product reviews. *American Association for Artificial Intelligence Proceedings 2006* (pp. 1265—1270). AAAI Press.

Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. *Proceedings of the 12th International World Wide Web Conference, Budapest* (pp. 519—528).

De Bruyn, A., & Lilien, G. L. (2008). A multi-stage model of word-of-mouth influence through viral marketing. *International Journal of Research in Marketing*, *25*(3), 151—163.

De Leeuw, E., & de Heer, W. (2002). Trends in household survey nonresponse: A longitudinal international perspective. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 41—54). New York: Wiley.

Degeratu, A. M., Rangaswamy, A., & Wu, J. (2000). Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes. *International Journal of Research in Marketing*, *17*(1), 55—78.

Dellarocas, C. (2006). Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management Science*, *52*(10), 1577—1593.

Dellarocas, C., & Narayan, R. (2006). A statistical measure of a population's propensity to engage in post-purchase online word-of-mouth. *Statistical Science*, *21*(2), 277—285.

Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, *21*(4), 23—45.

Dhar, V., & Chang, E. A. (2009). Does chatter matter? The impact of user-generated content on music sales. *Journal of Interactive Marketing*, *23*(4), 300—307.

Ehrenberg, A. S. C. (1959). The pattern of consumer purchases. *Applied Statistics*, *8*(1), 26—41.

Erens, F.-J. (1996). The synthesis of variety developing product families. *Doctoral Thesis*, Eindhoven University of Technology, Universitaire Drukkerij TU Eindhoven.

Fader, P. S., & Hardie, B. G. S. (1996). Modeling consumer choice among SKUs. *Journal of Marketing Research*, *33*(4), 442—452.

Feldman, R., Goldenberg, J., and Netzer, O. (2010). Mine your own business: Market structure surveillance through text mining. working paper, available at Wharton Interactive Media Initiative (http://www.whartoninteractive.com).

Ferreira, L., Jakob, N., & Gurevych, I. (2008). A comparative study of feature extraction algorithms in customer reviews. *Proceedings of the IEEE International Conference on Semantic Computing, Santa Clara* (pp. 144—151).

Gamon, M., Aue, A., Corston-Oliver, S., & Ringger, E. (2005). Pulse: Mining customer opinions from free text. *Advances in intelligent data analysis*, *vol. VI*. (pp. 121—132). Berlin: Springer.

Ganapathibhotla, M., & Liu, B. (2008). Mining opinions in comparative sentences. *Proceedings of the 22nd International Conference on Computational Linguistics, Manchester* (pp. 241—248).

Ghose, A., & Ipeirotis, P. G. (2007). Designing novel review ranking systems: Predicting the usefulness and impact of reviews. *Proceedings of the 9th International Conference on Electronic Commerce, Minneapolis* (pp. 303—310).

Ghose, A., Ipeirotis, P. G., & Sundararajan, A. (2007). Opinion mining using econometrics: A case study on reputation systems. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague* (pp. 416—423).

Giesen, J., Mueller, K., Schuberth, E., Wang, L., & Zolliker, P. (2007). Conjoint analysis to measure the perceived quality in volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, *13*(6), 1664—1671.

Green, P. E. (1984). Hybrid models for conjoint analysis: An expository review. *Journal of Marketing Research*, *21*(2), 155—169.

Grønhaug, K. (1977). Exploring consumer complaining behaviour: A model and some empirical results. *Advances in Consumer Research*, *4*, 159—165.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*(5), 646—675.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis.* London: Prentice Hall.

Hauser, J. R., & Toubia, O. (2005). The impact of utility balance and endogeneity in conjoint analysis. *Marketing Science*, *24*(3), 498—507.

Hu, M., & Liu, B. (2004). Mining opinion features in customer reviews. *Proceedings of the 19th National Conference on Artificial Intelligence, San Jose* (pp. 755—760).

Hu, N., Liu, L., & Zhang, J. J. (2008). Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. *Information Technology and Management*, *9*(3), 201—214.

Huber, J., Wittink, D. R., Fiedler, J. A., & Miller, R. (1993). The effectiveness of alternative preference elicitation procedures in predicting choice. *Journal of Marketing Research*, *30*(1), 105—114.

Janes, A., Scotto, M., Pedrycz, W., Russo, B., Stefanovic, M., & Succi, G. (2006). Identification of defect-prone classes in telecommunication software systems using design metrics. *Information Sciences*, *176*(24), 3711—3734.

Jiao, J., Simpson, T. W., & Siddique, Z. (2007). Product family design and platform-based product development: A state-of-the-art review. *Journal of Intelligent Manufacturing*, *18*, 5—29.

Jindal, N., & Liu, B. (2008). Opinion spam and analysis. *Proceedings of the International Conference on Web Search and Web Data Mining, Palo Alto* (pp. 219—230).

Kamakura, W. A., & Russell, G. J. (1993). Measuring brand value with scanner data. *International Conference on Web Search and Web Data Mining*, *10*(1), 9—22.

Kartono, B., & Rao, V. R. (2008). *Brand equity measurement: A comparative review and normative guide. Johnson School Research Paper Series No. 24-09* available at Social Science Research Network (http://www.ssrn.com).

Kleiber, C., & Zeileis, A. (2008). *Applied econometrics with R*. New York: Springer.

Koh, N. S., Hu, N., & Clemons, E. K. (2010). Do online reviews reflect a product's true perceived quality? An investigation of online movie reviews across cultures. *Electronic Commerce Research and Applications*, *9*(5), 374—385.

Kuksov, D., and Xie, Y. (2010). Pricing, Frills, and Customer Ratings. Forthcoming in *Marketing Science*. doi:10.1287/mksc.1100.0571

Lee, T. Y. (2007). Needs-based analysis of online customer reviews. *Proceedings of the 9th International Conference on Electronic Commerce, Minneapolis* (pp. 311—318).

Lee, T. Y., and Bradlow, E. T. (2007). Automatic construction of conjoint attributes and levels from online customer reviews. Working paper, The Wharton School, University of Pennsylvania.

Lee, M., Lee, J., & Cho, Y. (2009). How a convergence product affects related markets: The case of the mobile phone. *ETRI Journal*, *31*(2), 215—224.

Leisch, F. (2004). FlexMix: A general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software*, *11*(8), 1—18.

Li, X., & Hitt, L. M. (2008). Self selection and information role of online product reviews. *Information Systems Research*, *19*(4), 456—474.

Long, J. S., & Freese, J. (2006). Regression models for categorical dependent variables using stata. College Station: Stata Press.

Lui, B. (2010). Sentiment analysis and subjectivity. In N. Indurkhya, & F. J. Damerau (Eds.), *Handbook of natural language processing*. Boca Raton: Chapman & Hall/CRC Press.

Lui, B., Hu, M., & Cheng, J. (2005). Opinion observer: Analyzing and comparing opinions on the web. *Proceedings of the 14th International Conference on World Wide Web* (pp. 342—351). New York: ACM.

Martínez-Garmendia, J. (2010). Application of hedonic price modeling to consumer packaged goods using store scanner data. *Journal of Business Research*, *63*(7), 690—696.

Matzler, K., & Hinterhuber, H. H. (1998). How to make product development projects more successful by integrating Kano's model of customer satisfaction into quality function deployment. *Technovation*, *18*(1), 25—38.

Moe, W. W., and Schweidel, D. A. (2010). Online product opinions: Incidence, evaluation and evolution. Working paper, available at Wharton Interactive Media Initiative (http://www.whartoninteractive.com).

Moore, W. L. (2004). A cross-validity comparison of rating-based and choice-based conjoint analysis models. *International Journal of Research in Marketing*, *21*(3), 299—312.

Morrison, D. (1969). On the interpretation of discriminant analysis. *Journal of Marketing Research*, *6*(2), 156—163.

Netzer, O., Toubia, O., Bradlow, E. T., Dahan, E., Evgeniou, T., Feinberg, F. M., Feit, E. M., Hui, S. K., Johnson, J., Liechty, J. C., Orlin, J. B., & Rao, V. R. (2008). Beyond conjoint analysis: Advances in preference measurement. *Marketing Letters*, *19*, 337—354.

Nikolaus, F., & Piller, F. T. (2003). Key research issues in user interaction with user toolkits in a mass customisation system. *International Journal of Technology Management*, *26*(5–6), 578—599.

Popescu, A. -M., & Etzioni, O. (2005). Extracting product features and opinions from reviews. *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Morristown* (pp. 339−346).
Sattler, H., & Hensel-Börner, S. (2003). A comparison of conjoint measurement with self-explicated approaches. In A. Gustafsson, A. Herrmann, & F. Huber (Eds.), *Conjoint measurement: Methods and applications* (pp. 147−159). Berlin: Springer.
Sawtooth Software. (2007). The ACA/Web v6.0 technical paper. Sawtooth Software Inc.
Scholz, S. W., Meißner, M., & Decker, R. (2009). Measuring consumer preferences for complex products: A compositional approach based on paired comparisons. *Journal of Marketing Research*, *47*(4), 685−698.
Schreier, M., Oberhauser, S., & Pruegel, R. (2007). Lead users and the adoption and diffusion of new products: Insights from two extreme sports communities. *Marketing Letters*, *18*, 15−30.
Schweidel, D. A., Rindfleisch, A., O'Hern, M. S., and Antia, K. D. (2010). The impact of user-generated content on product innovation. Working paper, available at Wharton Interactive Media Initiative (http://www.whartoninteractive.com).
Sher, P. J., & Lee, S. -H. (2009). Consumer skepticism and online reviews: An elaboration likelihood model perspective. *Social Behavior and Personality*, *37*(1), 137−144.
Shugan, S. M. (2009). Relevancy is robust prediction, not alleged realism. *Marketing Science*, *28*(5), 991−998.
Somprasertsri, G., & Lalitrojwong, P. (2008). Automatic product feature extraction from online product reviews using maximum entropy with lexical and syntactic features. *IEEE International Conference on Information Reuse and Integration, Las Vegas* (pp. 250−255).
Toubia, O., Simester, D. I., Hauser, J. R., & Dahan, E. (2003). Fast polyhedral adaptive conjoint estimation. *Marketing Science*, *22*(3), 273−303.
Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. *Journal of Marketing*, *73*(5), 90−102.
Urban, G. L. (2005). Customer advocacy: A new era in marketing? *Journal of Public Policy & Marketing*, *24*(1), 155−159.
Verlegh, P. W. J., Schifferstein, H. N. J., & Wittink, D. R. (2002). Range and number-of-levels effects in derived and stated attribute importances. *Marketing Letters*, *13*(1), 41−52.
Wagner, U., & Taudes, A. (1987). Stochastic models of consumer behaviour. *European Journal of Operational Research*, *29*(1), 1−23.
Wedel, M., & DeSarbo, W. S. (1995). A mixture likelihood approach for generalized linear models. *Journal of Classification*, *12*, 21−55.
Wedel, M., DeSarbo, W. S., Bult, J. R., & Ramaswamy, V. (1993). A latent class regression model for heterogeneous count data. *Journal of Applied Econometrics*, *8*, 397−411.
Wei, C. -P., Chen, Y. -M., Yang, C. -S., & Yang, C. C. (2010). Understanding what concerns consumers: A semantic approach to product feature extraction from consumer reviews. *Information Systems and E-Business Management*, *8*(2), 149−167.
Wilson, T., Wiebe, J., & Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, *35*(3), 399−433.
Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, *74*(2), 133−148.