Incorporating Heterogeneity Into Count Data Models Applied to Marketing

Delane Botelho, EBAPE-FGV, Brazil
Pedro Jesus Fernandez, EBAPE-FGV, Brazil

In numerous empirical investigations into consumer behavior, the focus of the analysis is centered on explaining a limited dependent variable. Such limitation takes several forms, namely as a variable with limited continuous distribution to the left or to the right. Other limitations include variables with a finite number of values, such as those that take non-negative discrete values. Those data are usually referred to as “count data” and occur frequently in different marketing contexts, because they describe the number of times an event is observed. Examples include the number of purchases in a product category or the number of clients who visit a given store within a specified period of time. In marketing, this type of data has two basic characteristics: 1) excess of zeros (zero-inflation), more than expected in any Poisson distribution and; 2) heterogeneity among observations (buyers or consumers). Such characteristics may represent real problems if we use traditional models (such as the Poisson Regression model) to treat these data. This article deals with these two problems and incorporates some unusual reflection on count data modeling in marketing: 1) models which do not take zero-inflation into account will have poor fit; and, 2) the inclusion of unobserved heterogeneity may significantly improve the model’s goodness of fit, depending on the data (which would support the idea that independent variables are not always crucial to explain the dependent variable); and; 3) in many cases the heterogeneity of consumers may be sufficient (or even necessary) to explain the behavior of the dependent (count) variable.

[to cite]:

[url]:
http://www.acrwebsite.org/volumes/1000038/la/v1/LA-01

[copyright notice]:
This work is copyrighted by The Association for Consumer Research. For permission to copy or use this work in whole or in part, please contact the Copyright Clearance Center at http://www.copyright.com/.
Incorporating Heterogeneity into Count Data Models Applied to Marketing
Delane Botelho, EBAPE-FGV, Brazil
Pedro Jesus Fernandez, EBAPE-FGV, Brazil

EXTENDED ABSTRACT

In numerous empirical investigations into consumer behavior, the focus of the analysis is centered on explaining a limited dependent variable. Such limitation takes several forms, namely as a variable with limited continuous distribution to the left or to the right. Other limitations include variables with a finite number of values, such as those that take non-negative discrete values. Those data are usually referred to as “count data” and occur frequently in different marketing contexts, because they describe the number of times an event is observed. Examples include the number of purchases in a product category or the number of clients who visit a given store within a specified period of time. In marketing, this type of data has two basic characteristics: 1) excess of zeros (zero-inflation), more than expected in any Poisson distribution and; 2) heterogeneity among observations (buyers or consumers). Such characteristics may represent real problems if we use traditional models (such as the Poisson Regression model) to treat these data. This article deals with these two problems and incorporates some unusual reflection on count data modeling in marketing: 1) models which do not take zero-inflation into account will have poor fit; and, 2) the inclusion of unobserved heterogeneity may significantly improve the model’s goodness of fit, depending on the data (which would support the idea that independent variables are not always crucial to explain the dependent variable); and; 3) in many cases the heterogeneity of consumers may be sufficient (or even necessary) to explain the behavior of the dependent (count) variable.

The problem of zero-inflation may be resolved, for example, by modifying the Poisson distribution and introducing Zero-Inflated Poisson (ZIP), distribution, combining zero-inflated distribution (to represent purchases not made during the period under analysis) with Poisson distribution. Explanatory variables may also be included, generating a Zero-Inflated Poisson Regression model. The other problem is the low homogeneity among buyers or consumers. Economists attempt to resolve this issue by using econometric models referring to the “average consumer,” which may be not attractive for marketing scholars, as they are interested in understanding consumer behavior on a more disaggregated level. Also, heterogeneity among individuals is usually taken into account by incorporating explanatory variables, such as demographic variables. Nonetheless, this heterogeneity may contain an unobserved component, and the inclusion of such variables may not be sufficient or may even be unnecessary, when more robust models may be obtained by considering individual differences (as in the case of this article). This means that the observed events (e.g., buying, consuming) are formed by individual components that when combined are capable of explaining more complex phenomena. In other words, explanatory variables, which are highly useful in aggregated behavior models, may lose their importance when heterogeneity among consumers is incorporated.

Estimates of unobserved heterogeneity may be achieved by using Bayesian methods, which have appeared increasingly in the marketing literature and cover a variety of problems and data sources. This increase is based on the notion that buyers differ in their preferences and that firms should therefore take this into consideration when optimizing their marketing actions. All the models presented here were estimated with Bayesian methods.

Two datasets were analyzed: one of which was obtained from Wooldridge (2000) (cigarette consumption-dataset 1), and the other was collected in a Brazilian supermarket chain (purchase of cocoa-dataset 2). From the results obtained and discussed, the combination of unobserved heterogeneity and the explicit acknowledgement of zero-inflation resulted in the best model (that which presents the best fit) for dataset 1, namely the ZINBD (Zero-Inflated Negative Binomial Distribution) model. Depending on the data, the inclusion of unobserved heterogeneity can substantially improve the quality of the adjustment of the models. The consideration of heterogeneity (intrinsic differences among buyers/consumers) in the two datasets may be more important than the inclusion of explanatory variables. Indeed for both datasets, the NBD (Negative Binomial Distribution) model without the inclusion of such variables (NBD distribution model) shows better fit than the model that incorporates them (NBD regression model). This would indicate to marketing researchers that the inclusion of explanatory variables may not always be the best option for the explanation of the response variable. Hence, it is useful to identify explanatory variables that are more adequate. Explanatory variables are important because they normally involve variables of the marketing mix and therefore serve as instruments of action for marketing practitioners to alter the response variable, which is often related to demand (as was the case with the two datasets analyzed in this article). So, the inclusion of variables that do not enhance the fit of a purely stochastic model, such as NBD, should be queried by researchers. The results also indicate that the NBD or ZINBD distribution models (in case there is zero-inflation) could be used as a benchmark to be attained by competing models.

Future research should analyze the existence of a finite number of segments of different buyers/consumers. This could be done by using other heterogeneity distributions, such as a mixture of normal distributions. Concentration on a finite (although unknown) number of segments, or non-parametric methods, such as Dirichlet Process Prior, could be attempted in the Latin America arena.

References

**CHANNEL BLURRING AND CONSUMER EXPECTATIONS: EVIDENCE IN CHILE**

Constanza C. Bianchi, Universidad Adolfo Ibáñez, Chile
Joaquin Mena, Universidad Adolfo Ibáñez, Chile

**EXTENDED ABSTRACT**

Retail firms from all over the world are struggling to satisfy customers by providing more than just goods and services, but an ideal shopping experience. The shopping experience incorporates several elements such as low prices, product quality, information, assortment, convenience, entertainment and services, among others (Arnold et al., 1983).

Nevertheless, it is difficult to construct a single shopping experience because distinctions between retail formats are beggning to blur and overlap. The increased competition in retailing has led to a situation of channel blurring, where retailers have broadened their assortment and consumers can find groceries and electronics in supermarkets, as well as in department stores, drug stores, or home improvement stores. Due to this aggressive level of competition, where consumers can buy the same products in any of the four formats, it is important to understand what are the most important attributes for consumers when shopping at these retail formats.

Studies on store choice and shopper experience have generally been tested in one specific retail format (e.g., Sirohi et al., 1998). A few studies have found that the importance assigned to a store attribute may vary depending on the reason for buying the product (e.g., Green & Krieger, 1995; Thelen & Woodside, 1997; Van Kenhove, et al., 1999). Overall, the literature suggests that consumers may associate certain store formats with specific expectations and needs. However, it is not clear if these consumer expectations, and therefore saliencies, vary across retail formats.

Chile has one of the most concentrated and developed retailing industry in Latin America. The Chilean retailing industry has consistently reacted towards foreign threats, resulting in a local industry with increased level of consolidation and concentration. An important trend in Chilean retailing industry during the last decade has been the increase in channel blurring. For example, grocery retailers sell appliances, electronics, clothes, books, stationary products, gardening and outdoor products. Department stores sell books, food, and provide hairdressers, and tourist services in their stores. Drug store retailers offer food, drinks and gifts in their stores. Finally home improvement retailers sell kitchen appliances, gifts and garden products.

This investigation attempts to identify consumer expectations regarding the relevant store attributes that lead consumers to have an ideal shopping experience for the different retail formats (grocery retailers, department store retailers, drug store retailers, and home improvement retailers centers). Consumers may have different expectations of an ideal shopping experience for different retail formats. This suggests that consumers may assign different saliencies to store attributes depending on what retail format they are shopping. Can one retail format fulfill all of the attributes required by consumers? The literature review suggests that consumers still associate certain store formats with specific expectations and needs. Thus, the main research question of this is the following: *Do consumer store attribute saliency vary for different retail formats?*

To assess the research question, this study utilized a mix of qualitative and quantitative methodologies, such as in depth interviews, observation, and surveys. These methodologies helped identify customer expectations regarding the salient attributes that lead to an ideal shopping experience for the different retail formats. The study was held in two stages, with a total duration of 8 months (July 2003–February 2004).

Twenty interviews were held with consumers of different sex and ages. The interview data identified twelve store attributes that were salient for an ideal shopping experience for consumers when shopping at the different retail formats (drug stores, grocery stores, department stores and home improvement stores).

Based on the qualitative data, a questionnaire was developed, pre-tested and applied personally to a population of approximately 700 graduate students of a Chilean University. This questionnaire asked respondents to evaluate the importance of each variable mentioned, when shopping at each of the different retail formats mentioned in this study. The 12 variables included in the survey where statistically tested in a One-way ANOVA test with a significance level of 95%, in order to verify if there existed differences between the medias. At