Journal of Retailing 92 (4, 2016) 470–485

Journal of Retailing

A Model for Inferring Market Preferences from Online Retail Product Information Matrices

Timothy J. Gilbride[a,*], Imran S. Currim[b], Ofer Mintz[c], S. Siddarth[d]

[a] *Mendoza College of Business, University of Notre Dame, Notre Dame, IN 46556, United States*
[b] *Paul Merage School of Business, University of California, Irvine, CA 92697, United States*
[c] *E. J. Ourso College of Business, Louisiana State University, Baton Rouge, LA 70803, United States*
[d] *Marshall School of Business, University of Southern California, Los Angeles, CA 90089, United States*

Available online 16 August 2016

Abstract

This research extends information display board methods, currently employed to study information processing patterns in laboratory settings, to a field based setting that also yields managerially useful estimates of market preferences. A new model is proposed based on statistical, behavioral, and economic theories, which integrates three decisions consumers must make in this context: which product-attribute to inspect next, when to stop processing, and which, if any, product to purchase. Several theoretical options are considered on how to model product attribute selection and how to treat uninspected attributes. The modeling options are empirically tested employing datasets collected at a popular e-tailer's website, while customers were making product evaluation and purchase decisions. Subsequent to identifying the best model, we show how the resulting attribute preference estimates can be managerially employed to improve customer targeting of abandoned shopping carts for follow up communications aimed at improving sales conversions.

Keywords: Choice models; Information processing; E-commerce; Sequential search; Expected utility

A large number of retailer websites (i.e., Apple, Nikon, Dell, Ford, Best Buy, etc.) organize or have a feature that allows customers to organize product-attribute information at the point-of-purchase in the matrix form reminiscent of information display boards (IDBs) widely employed in lab-based information processing research (see Fig. 1). In this research, we update the IDB methodology to an online format and propose and estimate an econometric model that infers attribute importance weights from consumers making durable goods purchases at a vertically integrated e-tailer's website. Product information matrices require shoppers to make a sequence of decisions: (i) which product-attribute to inspect next, (ii) when to stop processing information, and (iii) which, if any, product to purchase. Our model-based, multi-disciplinary approach incorporates statistical, behavioral, and economic theories to estimate attribute importance weights using data from this sequence of three decisions.

Determining the relative importance of product attributes is one of the canonical problems in marketing management. However, as Feit, Beltramo, and Feinberg (2010) discuss, both market and lab based techniques to measure preferences have their shortcomings. Specifically, while market data comes from consumers making actual decisions, it often lacks the variability in attribute levels needed to estimate attribute preference weights. On the other hand, survey and lab based methods such as conjoint analysis provide the necessary variation in product attributes, but produce inconsistencies between predictions and actual market outcomes, suggesting that respondents may not make "hypothetical survey choices exactly as they make purchase decisions" (pp. 785–786). Our research attempts to overcome these shortcomings by showing how IDB data collected by e-tailers outside the lab and with relatively little attribute level variation, can still yield meaningful market preferences for attributes provided consumer decisions are modeled in an integrated manner. We anticipate that the proposed method will be of most interest

* Corresponding author.
E-mail addresses: tgilbrid@nd.edu (T.J. Gilbride), iscurrim@uci.edu (I.S. Currim), omintz@lsu.edu (O. Mintz), siddarth@usc.edu (S. Siddarth).

http://dx.doi.org/10.1016/j.jretai.2016.07.002
0022-4359/

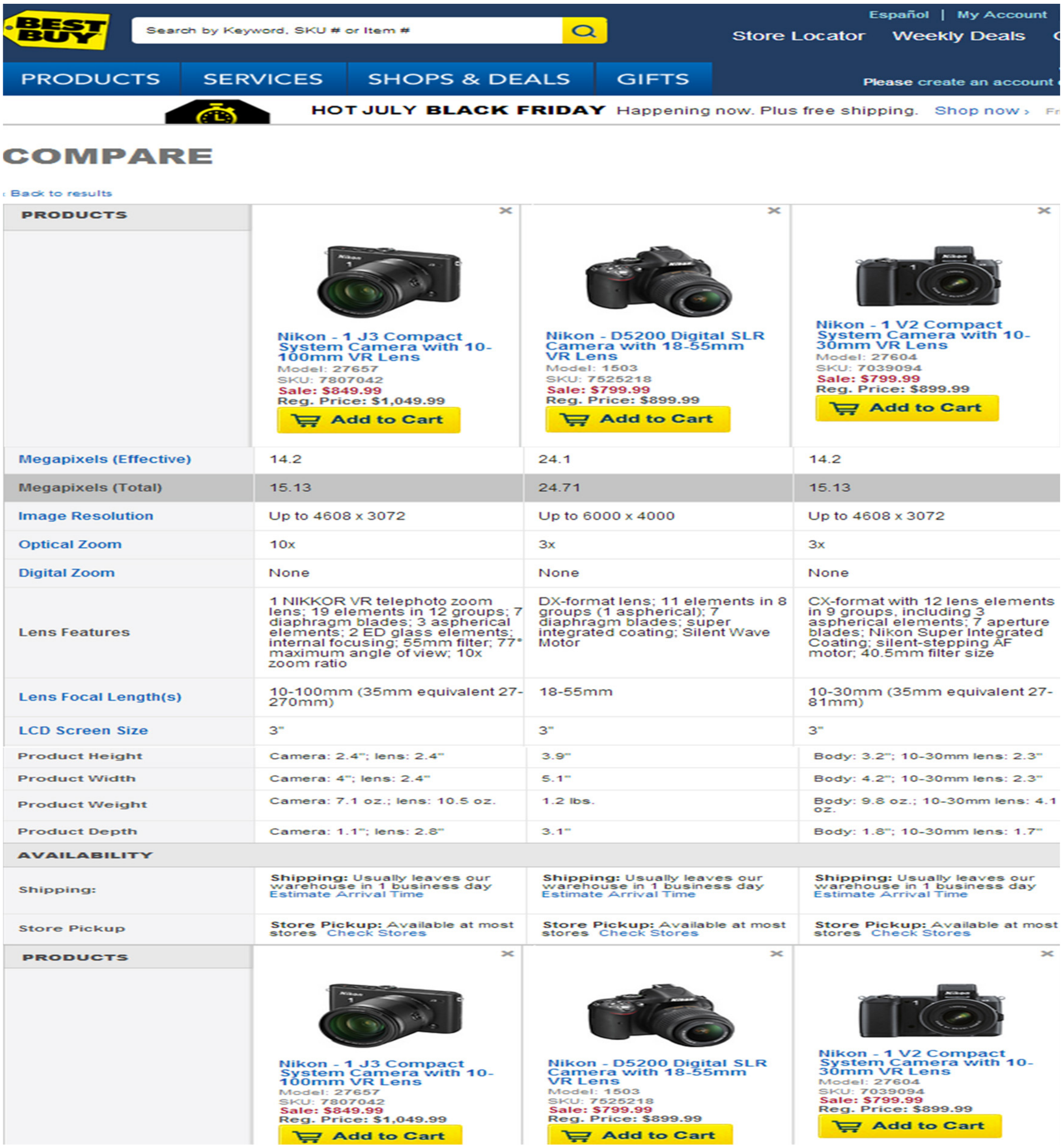

Fig. 1. Online shopping example at Best Buy.

to e-tailers vertically integrated with manufacturers. However, the ease of implementing the online IDB make it attractive for other e-tailers to collect this information to collaborate with their suppliers on product design, pricing, and advertising.

IDBs have been extensively used to study how consumers process product attribute information (e.g., see Bettman, Luce, and Payne 1998; Payne, Bettman, and Johnson 1993 for reviews) but have not been used to infer product attribute weights. Yang, Toubia, and De Jong (2015) and Meißner, Musalem, and Huber (2016) have shown how using eye-tracking technology and modeling the information processing of subjects can be used to improve holdout sample predictions in a laboratory based conjoint study. However, they rely on specialized equipment, experimental methods, and do not test their model's predictions with actual market outcomes. The current research uses data from shoppers who made a single visit to an e-tailer's website and models their information processing and choices on an online IDB. The proposed method is not necessarily a substitute for other lab based studies, but it does leverage data that can uniquely be collected at the point-of-purchase by e-tailers.

Consumers can learn about products and their attribute-levels from a variety of sources through both active and passive learning over a period of time. We assume that the cumulative effect of all past learning and search can be summarized at a point in time by a consumer's current preferences for particular attributes and current expected level and range of attribute values in the market. So for instance, someone shopping for a car would know the relative importance (to them) of horsepower, fuel economy, price, etc. and might expect that the fuel economy for subcompact cars to range from 30 to 35 miles per gallon, say. We do not model the consumer's learning process over time or the source of their preferences; we assume that the data to model this would