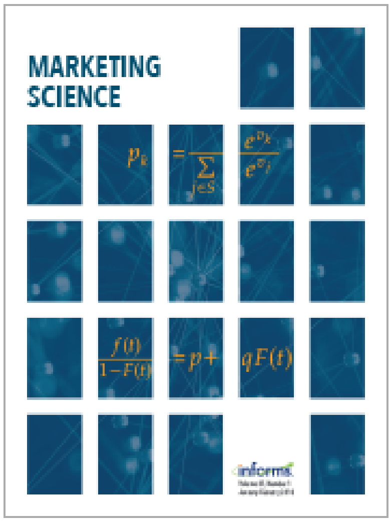


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Database Submission—The ISMS Durable Goods Data Sets

Jian Ni, Scott A. Neslin, Baohong Sun,

To cite this article:
Jian Ni, Scott A. Neslin, Baohong Sun, (2012) Database Submission—The ISMS Durable Goods Data Sets. Marketing Science 31(6):1008-1013. https://doi.org/10.1287/mksc.1120.0726

Marketing Science
Vol. 31, No. 6, November–December 2012, pp. 1008–1013
ISSN 0732-2399 (print) | ISSN 1526-548X (online)

http://dx.doi.org/10.1287/mksc.1120.0726

Database Submission

The ISMS Durable Goods Data Sets

Jian Ni
Carey Business School, Johns Hopkins University, Baltimore, Maryland 21202, jni@jhu.edu

Scott A. Neslin
Tuck School of Business at Dartmouth, Dartmouth College, Hanover, New Hampshire 03755, scott.neslin@dartmouth.edu

Baohong Sun
Cheung Kong Graduate School of Business, New York, New York 10019, bhsun@ckgsb.edu.cn

This paper describes two new data sets available to academic researchers (at http://www.informs.org/Community/ISMS). The first is a panel data set containing the transactions of 19,936 households made over the period from December 1998 to November 2004 at a major U.S. consumer electronics retailer. There are a total of 173,262 transactions, including purchases and returns of products as well as extended warranties. There are 16 product categories and 292 subcategories, ranging from big-ticket items such as televisions to small-ticket items such as CDs and batteries. The second data set features a field experiment for a Christmas promotion that took place in December 2003 in the form of a direct mailing sent to a randomly selected group of households at the end of November 2003. We describe the data and the potential research issues that can be studied using these two durable goods data sets.

Key words: retailer; durable goods; panel data; product adoption; holiday promotion; sales forecasting
History: Received: November 21, 2011; accepted: June 4, 2012; Preyas Desai served as the editor-in-chief and Bart Bronnenberg served as associate editor for this article. Published online in *Articles in Advance* August 9, 2012.

1. Introduction

Durable goods play a crucial role in the economy. In 2008, personal consumption expenditures on durables exceeded $1.1 trillion (Federal Reserve Bank of St. Louis 2009). Compared with fast-moving consumer packaged goods (CPG), consumer decisions for durable goods are much more sophisticated, dynamic, and deliberative, and they raise numerous research questions for microeconomic and marketing analysis. A thorough understanding of consumer decisions with respect to durables will help develop and test both economic and consumer behavior theories, and it will have important implications for managerial decisions.

In recent decades, a rich analytical literature in both marketing and economics has examined the competitive behavior of firms that sell durable goods. However, empirical research investigating consumer decisions about durable goods is sparse in marketing. Examining *Marketing Science* and the *Journal of Marketing Research*, more than 400 papers have been published regarding consumer purchase behavior of CPG products using IRI and ACNielsen data sets over the past three decades. This contrasts with 36 papers focused on durable goods, among which 28 used aggregate sales and only 8 used individual consumer purchase history (consumer panel data).

The purpose of this paper is to address this disparity by introducing two distinct databases to the research community. Administered by the INFORMS Society for Marketing Science (ISMS), these databases are called ISMS Durable Goods Dataset 1 and ISMS Durable Goods Dataset 2 (see §4). To the best of our knowledge, the ISMS Durable Goods data sets are the most comprehensive customer-level transaction data available to researchers. We make these data sets publicly available with the wish to facilitate researchers in marketing, economics, psychology, and other fields to conduct research that helps in understanding consumer purchase decisions about durable goods.

Both databases are provided by an anonymous major U.S. consumer electronics retailer. ISMS Durable Goods Dataset 1 is panel data: it contains the complete transaction records of a large set of customers encompassing most of the retailer's stores over a six-year period. ISMS Durable Goods Dataset 2 is also at the customer level, but it is cross-sectional and features the results of a direct-mail promotion field experiment. It contains a host of variables calculated before the promotion, an indicator of whether the customer was offered the promotion, and dollar purchases made by the customer during the promotion period. In what follows, we describe both databases and suggest a few illustrative research topics. More time is spent on Dataset 1 because it can be used to investigate a broader set of topics. However, ISMS Durable Goods Dataset 2 is also quite promising, and we discuss it as well. Detailed documentation of the variables in each database is available at http://www.informs.org/Community/ISMS/ISMS-Research-Datasets.

2. ISMS Durable Goods Dataset 1

2.1. Data Description

The first data set consists of all transaction records of 19,936 randomly selected households from the same national retail chain during a six-year period (from December 1998 to November 2004). During this period, the 19,936 households made 173,262 transactions of durable goods and/or services from 1,176 of the focal retailer's stores located throughout the United States.[1] There are 16 product categories (e.g., PC hardware), 292 subcategories (e.g., desktop computers), and 22,210 specific products (models), ranging from big-ticket items such as televisions, cameras, and personal digital assistants (PDAs) to accessories and small-ticket items such as CDs and batteries. Brand information for each model is also available in the data set.

There are six types of transactions: product purchase, product return, extended service contract purchase, extended service contract return, sales discount, and miscellaneous. Each record includes detailed information about each transaction type made by a particular customer on a particular date and, depending on the transaction type, information such as brand purchased or returned, service contracts purchased or returned, product category, price paid or refunded, length of coverage of the service contract, and time and location of the transaction. Finally, each record contains customer-level demographic information such as income, gender, family size, and age.

Table 1 shows frequency counts of the six transaction types across all categories (panel A) as well as for the top 10 categories (panel B).[2] There are thousands of transactions per category. Wireless and DVS have the highest product return rates, at 15.2% and 17.5%, respectively, whereas music has the lowest return rate, at 5.4%. PC hardware has the highest purchase rate of extended service contracts, at 21.6%. Interestingly, consumers who purchase an extended service contract for the mobile category are more likely to return that contract.

Table 2 provides descriptive statistics for representative subcategories such as digital cameras, camcorders, TVs, VCRs desktops, and notebooks. Again, there are thousands of purchases in these subcategories, suggesting that research can be conducted at the subcategory level. The number of brands per subcategory ranges from 11 to 67, a rich but manageable number to analyze. Taking digital cameras as an example, 1,953 customers made 2,524 purchases of 20 brands during the six years observed at an average price of $311.95, and consumers purchased, on average, $71.27 in extended service plans offered by the retailer. Among these transactions, 22% are associated with the purchase of an extended service contract (ESC), 11.9% of the products were returned, 3.4% were purchased online, and 24.8% were bought during the holiday season (Thanksgiving through Christmas). Big-ticket items such as digital cameras, camcorders, desktops, and notebooks have about 20% of their purchases associated with an extended service contract purchase, but those categories also have a relatively high product return rate. However, the TV subcategory stands with the highest ratio of service contract price to product price.

Table 3 illustrates "market basket" statistics for the "PC hardware" category. There are six subcategories within PC hardware: scanners, PDAs, notebooks, desktops, printers, and monitors. A total of 5,266 customers bought at least one of these products. Table 3 shows the combinations that were bought most frequently over the observation period (though not necessarily on the same purchase occasion). With six subcategories, there are $2^6 = 128$ possible combinations, or market baskets. However, Table 3 shows that 12 baskets account for more than 90% of the baskets actually purchased. The most common basket, not surprisingly, is purchasing a desktop, printer, and monitor (1,287 customers). However, 908 customers purchased just a printer, 390 purchased just a monitor, and 331 purchased just a desktop. A natural question is, what distinguishes the single-item basket customers from the multiple-basket customers?

Table 4 provides purchase statistics at the customer level. The data can be aggregated per trip or examined at their most micro level—the transaction. Note that the standard deviations are at least as high as the means for these variables, so there is ample variation and, obviously, some "heavy users." For example, the average amount spent per trip is $317 with a standard deviation of $497; the average amount per transaction is $109 with a standard deviation of $295. This suggests that it will be practical to segment the heavy users from the light users and examine differences in the categories purchased, for example.

Table 5 provides descriptive statistics of customer demographics. Again, we see ample variation in age, children, and income, and we note a roughly equal split between male and female head-of-household.

[1] For each randomly selected household, we have the complete transaction record. Thus, whether the data are evenly distributed across years depends on the purchase frequency of each individual customer.

[2] We report the top 10 major categories instead of all 16 because there are sparse observations in the other 6 categories.

Table 1 Number of Observations of Different Types of Transactions (ISMS Durable Goods Dataset 1)

Panel A: Counts across all categories

Transaction type	All categories
Product purchases	139,580
Product returns	14,724
Extended service contract purchases	15,033
Extended service contract returns	2,437
Product purchases with discount	1,452
Miscellaneous transactions	36
Total transactions	173,262

Panel B: Counts separately for major categories

Transaction type	Audio	DVS[a]	Mobile	Imaging	Music
Product purchases	17,876	3,082	18,098	10,300	14,008
Product returns	2,102	540	1,834	1,104	759
Extended service contract purchases	2,168	524	2,439	1,266	328
Extended service contract returns	313	110	394	126	24
Product purchases with discount	183	31	111	141	60
Miscellaneous transactions	0	0	0	0	0
Total transactions	22,642	4,287	22,876	12,937	15,179

Transaction type	$P*S*T$[b]	PC hardware	TV	Video hardware	Wireless
Product purchases	22,889	12,601	13,116	9,762	5,393
Product returns	2,796	1,494	1,449	1,132	818
Extended service contract purchases	1,101	2,716	2,219	1,405	490
Extended service contract returns	194	356	251	196	121
Product purchases with discount	128	307	259	80	54
Miscellaneous transactions	0	0	0	0	0
Total transactions	27,108	17,474	17,294	12,575	6,876

[a]DVS includes direct TV system/accessories, satellite dishes, digital video recorders, etc.
[b]$P*S*T$ stands for products such as computer accessories, cables, computer media, etc. See the documentation online for detailed info about category and subcategory specifications.

Table 2 Descriptive Statistics for Subcategories (ISMS Durable Goods Dataset 1)

Category =	Imaging	Imaging	Imaging	Television	Video hardware
Subcategory =	Digital camera	Camcorder	Camcorder accessories/batteries	TV	Digital video
No. of customers	1,953	1,604	1,020	7,040	3,489
No. of transactions	2,524	3,058	1,456	17,294	6,554
No. of brands/models	20	11	25	67	35
Average price paid for product ($)	311.95	475.58	57.01	370.04	165.32
Average price paid for ESC ($)	71.27	101.17	NA	147.13	35.15
% of product purchases with ESC	22.1	22.9	0	12.8	15.4
% of product purchases returned	11.9	8.4	10.4	8.4	10.7
% of product purchases made online	3.4	0.33	0.76	0.39	1.62

Category =	Video hardware	Music	PC hardware	PC hardware	Wireless
Subcategory =	VCR home	Music	Desktop	Notebook	Wireless
No. of customers	1,963	3,902	2,443	911	2,721
No. of transactions	3,029	15,179	5,074	2,320	6,876
No. of brands/models	21	NA	14	12	NA
Average price paid for product ($)	105.65	39.65	738.68	1,221.70	91.16
Average price paid for ESC ($)	29.75	23.73	223.11	278.32	52.85
% of product purchases with ESC	11.6	2.2	22.6	22.6	7.1
% of product purchases returned	8.2	5.0	8.1	9.8	11.9
% of product purchases made online	0.83	3.3	0.26	0.34	0.33

Examination of the data reveals that 98% of the regular prices paid on the same day are the same across stores. This suggests that pricing decisions are made at the retailer's headquarters and are relatively uniform across stores. To construct a price time series, one can use the *Transaction_Type* and *Unit_Price* variables (see the documentation at http://www.informs.org/Community/ISMS/ISMS-Research-Datasets). *Transaction_Type* reveals whether the transaction was a product purchase and also identifies possible

Table 3 Market Basket Statistics for the PC Hardware Category (ISMS Durable Goods Dataset 1)

Scanner	PDA	Notebook	Desktop	Printer	Monitor	No.	%	Cumulative %
			✓	✓	✓	1,287	24.4	24.4
				✓		908	17.2	51.7
		✓				501	9.5	51.2
	✓					394	7.5	58.7
					✓	390	7.4	66.1
			✓			331	6.3	72.4
			✓		✓	264	5.0	77.4
✓						232	4.4	81.8
✓			✓	✓	✓	174	3.3	85.1
		✓		✓		114	2.2	87.3
		✓	✓	✓	✓	93	1.8	89.0
			✓	✓		92	1.7	90.8

Note. Shown are the most popular combinations purchased over the period of observation, representing 4,780 of the 5,266 customers who purchased PC hardware.

Table 4 Descriptive Statistics for Household Purchases (ISMS Durable Goods Dataset 1)

Variable	Mean	Standard deviation
Average expenditure per shopping trip ($)	317.02	497.52
Average expenditure per transaction ($)	109.23	295.50
Average no. of items purchased per trip	2.39	3.79
Average no. of items purchased per transaction	0.82	1.38
Average no. of purchase trips per household	2.99	2.78
Average no. of purchase transactions per household	8.69	11.67
Average price paid per item ($)	108.91	295.42
Average amount spent on holiday shopping[a] ($)	189.77	352.71
No. of purchases of gift cards	1,487	NA
No. of online purchases	2,550	NA
No. of different items purchased	22,210	NA

[a] *Holiday shopping* refers to purchases made on Black Friday through Christmas.

promotional price discounts. The *Unit_Price* variable states the price paid or the amount of the promotional price discount.

For example, in Figures 1 and 2, we plot the price and sales trends of Apex digital video camera models 763370 and 749912. To prepare these figures, we aggregated across customers to count the number of units sold for the product and the average price paid, per week. Although it is possible to construct weekly price and purchase series for many product models, this is not possible for infrequently purchased models. Users will find it easier to construct price and sales time series at the brand level rather than model level (e.g., Sony digital video cameras rather than each Sony model) and at the monthly level rather than weekly level. Or researchers may be able to construct a price series using data collected from other sources, such as NPD.com.

2.2. Illustrative Research Issues for ISMS Durable Goods Dataset 1

Given that Dataset 1 is panel data at the customer level over time, researchers can study rich purchase decisions such as product adoption (e.g., digital video cameras), brand choice (e.g., which brand of digital video camera to purchase), add-ons (e.g., televisions and television stands), multicategory choice (e.g., television equipment, audio equipment), and adoption of online purchase channel. Data can also be used to examine issues related to "Black Friday," the Friday after Thanksgiving, which has become the much-heralded "kickoff" to the Christmas shopping season. In light of the substantial economic impact of Black Friday, this shopping event deserves closer scrutiny than has been reported in the academic literature and popular press.

The data also contain consumer purchase history on ESCs and on product returns. Contributing more than half of the total profits of major electronics retailers, ESCs have become a major profit engine for consumer

Table 5 Descriptive Statistics for Household Demographics (ISMS Durable Goods Dataset 1)

Variable	Observations[a]	Mean	Standard deviation	Minimum	Maximum
Age_Head	16,384	48.42	15.17	18	99
Gender_Head	16,566	0.64[b]	0.48	0	1
Children_Presence	8,451	0.66	0.47	0	1
No. of children	19,936	0.40	0.80	0	6
Income	16,811	5.68[c]	2.36	1	9

[a] Demographic information is missing for some customers.
[b] Male is coded as 1, and female is coded as 0.
[c] Income level of the household is coded on a 1–9 scale, where a larger number indicates higher income.

electronics retailers such as Best Buy (Berner 2004). It is imperative to understand why and how consumers purchase ESCs and how to improve marketing mix decisions. Similarly, allowing consumers to return products encourages them to try a product for which they have huge uncertainty. However, consumers can also abuse the return policy. When too many products are returned, retailers bear significant cost. It is therefore important to understand the fundamental driving force behind product returns and how product returns affect future purchasing. For example, how do product returns affect customers' product adoption and their subsequent purchases? What types of products are more likely to be returned?

Retailers provide a platform where manufacturers compete to maximize their long-term profits. Many competitive marketing mix strategies such as new product introductions, (product line) pricing, and advertising can be studied. For example, when Sony lowers the prices on its digital cameras, how does Cannon adjust its prices to defend its position? Static or dynamic models can be developed to understand the nature of interactions among firms and test theories of competition; such models can also be used for policy analysis by simulating behavior under a variety of market environments. Given its panel structure,

Figure 1 **Weekly Unit Sales Volume and Price Trend for Apex Digital Video Camera Model 763370 (ISMS Durable Goods Dataset 1)**

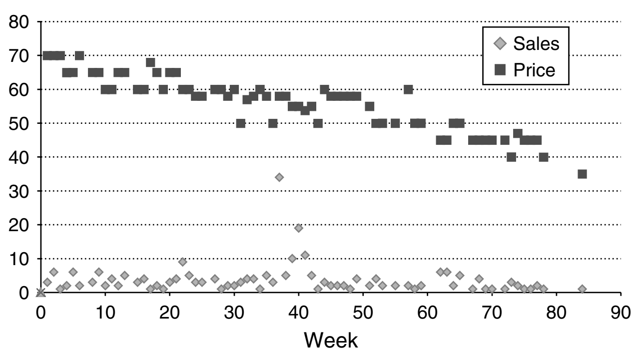
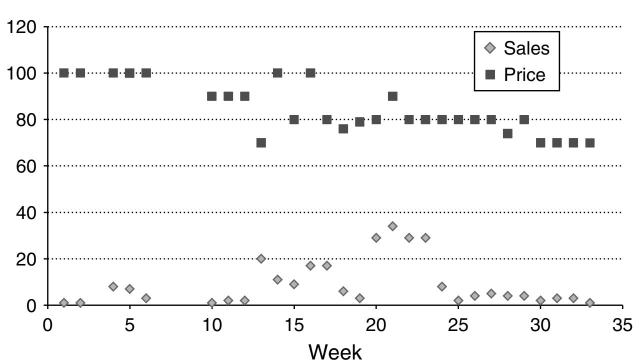


Figure 2 **Weekly Unit Sales Volume and Price Trend for Apex Digital Video Camera Model 749912 (ISMS Durable Goods Dataset 1)**



Dataset 1 affords the opportunity to study manufacturer competition using customer-level data.

2.3. Limitations of ISMS Durable Goods Dataset 1

Its six-year observation window and complete household-level transaction records of a large number of electronic durable goods make this data set unique. However, it is also subject to limitations. First, it does not contain information on non-price-related promotions such as TV advertising, catalogs, store displays, etc.[3] Second, it is from a single retailer. Without information on consumer purchases from other retail outlets, studies on some consumers decisions such as product adoption, upgrade, and replacement could be subject to bias. Third, the data set includes the price paid but does not provide a "store environment" file; i.e., it does not include prices for all available models at each point in time. One can create a price time series as described earlier by aggregating across the 1,176 stores. However, for low-sales items, there may not be enough observations to construct a satisfactory time series.[4] We caution researchers to be aware of these limitations when making assumptions, formulating models, and drawing conclusions.

3. ISMS Durable Goods Dataset 2

3.1. Data Description

ISMS Durable Goods Dataset 2 features a field experiment for a Christmas promotion run by the same chain as that featured in Dataset 1. The field experiment took place in December 2003 in the form of a direct-mail campaign offered at the end of November 2003 to randomly selected households. The promotional offer was as follows: Households received $10 off if they purchased during the promotional time period (December 4–December 15). If they did purchase, they would get 10% off on a subsequent purchase, good through the end of December. Roughly half of the 176,961 households in the database (promotion group) received the Christmas holiday promotional mailer; the other half (control group) did not. When running the experiment, the retailer did not know a priori the customer attributes that would lead to the most incremental sales. Thus, the assignment of promotion in the field experiment was truly random. In addition to receipt of and response to the promotion, the data

[3] Users will notice a steep increase in January 2001 and January 2002 sales for several products. We could not confirm with certainty, but this may be due to special promotions that were run during these months.

[4] It is not uncommon in marketing and economics to collect price information from alternative sources.

Table 6 Descriptive Statistics of ISMS Durable Goods Dataset 2

Variable	Control group	Promotion group
Sample size	88,625	88,336
Average sales amount during December promotion period (across both purchase and non-purchase occasions) ($)	9.73 (104.14)[a]	12.42 (120.85)
No. of transactions in previous 12 months	1.73 (3.44)	1.74 (3.59)
No. of large-ticket item purchases in previous 12 months	0.17 (0.51)	0.17 (0.50)
No. of small-ticket item purchases in previous 12 months	0.96 (3.57)	0.99 (4.10)
No. of ESCs purchased in previous 12 months	0.16 (0.54)	0.16 (0.54)

Note. Standard deviations are reported in parentheses.

[a]The average is lower than the *average expenditure per shopping trip* in Dataset 1 (see Table 4) because many households who received the promotion in Dataset 2 did not make purchases.

contain approximately 150 descriptors of all customers, covering purchase history prior to the promotion, response to previous promotions, purchase of warranties/extended service, product returns, etc.

Table 6 provides statistics for a few key variables characterizing the reaction to the Christmas promotion for the promotion and control groups. Mean sales for those receiving the promotional offer is $12.42; for those not receiving the offer, it is $9.73.[5] The other statistics show that the experimental and control groups were equally matched on variables available before the promotion. This, of course, is as it should be, given that the promotion was distributed randomly.

3.2. Research Opportunities and Limitations

Dataset 2 provides the opportunity to analyze the results of a promotional field test, plus other topics such as the "direct-mail-deal–prone" consumer. Its most attractive attribute is the field test; hence the most obvious use of the database is to analyze the results of the field test—both predicting the results and understanding them. For example, what are the best methodologies for predicting customer-level incremental sales? What variables best predict incremental sales for this promotion? In addition, because the data contain responses to previous promotions—all delivered via direct mail—one could profile the direct-mail-deal–prone customer (see Chapter 3 in Blattberg and Neslin 1990 for a summary of research on the deal-prone consumer in the context of CPG promotions).

One limitation is that the data are not longitudinal like those in Dataset 1. The recency, frequency, monetary value (RFM) variables, for example, precede the field experiment, but the data do not contain week-by-week purchase histories of each customer. Although Datasets 1 and 2 are offered by the same company, the data cannot be linked—the customers' IDs in the two data sets do not have a one-to-one mapping.

[5] The average is a lot lower than for Dataset 1 (see Table 4) because a lot of households who received the promotion did not purchase. Among the purchasers, the average would be much higher.

4. Process for Obtaining the Databases

The databases are provided by the ISMS (http://www.informs.org/Community/ISMS). The distribution of the data is controlled to ensure their use is consistent with ISMS's mission of enabling the development, dissemination, and implementation of research based on marketing science approaches. Database documentation, purchase agreement, open forums, working paper collections, and downloading instructions are available at http://www.informs.org/Community/ISMS/ISMS-Research-Datasets.

5. Summary

In sum, ISMS makes available to the marketing science community two durable goods data sets—a panel database and a field test database. We hope these databases will spark a major acceleration in research on durable goods that will benefit the practice of marketing for many years to come.

References

Berner J (2004) The warranty windfall. *BusinessWeek* (December 20) http://www.businessweek.com/magazine/content/04_51/b3913110_mz020.htm.

Blattberg RC, Neslin SA (1990) *Sales Promotion: Concepts, Methods, and Strategies* (Prentice Hall, Englewood Cliffs, NJ).

Federal Reserve Bank of St. Louis (2009) Personal consumption expenditures: Durable goods. Technical report, Federal Reserve Bank of St. Louis, St. Louis.

The data set described in this paper is maintained by the INFORMS Society for Marketing Science (ISMS) and available through http://mktsci.pubs.informs.org. Any fees charged by the ISMS for the distribution of the data set will be used for the continual maintenance and updating of the data. Funding support is available for students who cannot afford the fee (contact the President of the ISMS in order to apply). Please see http://www.informs.org/Community/ISMS/ISMS-Research-Datasets for further details.