Predictive CRM In A B2B Context: Evaluating Cross-Selling Opportunities Based On Likely Response To The Offer

Lynd Bacon[1], Peter Lenk[2], C. Michael Troutman[3], Ashwin Sridhar[4]
[1]Loma Buena Associates, P.O. Box 620960, Woodside CA 94062 USA
[2]Stephen M. Ross School of Business, University of Michigan, 701 Tappan St., Ann Arbor MI USA
[3]CMT Marketing Analytics, 90 Sycamore Ridge, Honeoye Falls NY 14472 USA
[4]ZLINQ Solutions, 39270 Paseo Padre Pkwy #500, Fremont CA 94538 USA

Abstract
Most companies struggle with the issue of what products to offer their existing customers. An important CRM goal is offering customer what they are most likely to buy, not necessarily what they they ask for. The key is to offer customers what they are must likely to buy, given that an offer or recommendation is made. We describe the development of a B2B CRM system that provides tactical sales support through the use of sensor (GIS) and account management data. It deploys hierarchical Bayes models of purchase behavior that condition on sales interventions, and that incorporate customer covariates to account for heterogeneity. We compare our procedure to the company's current practices, and provide an contextual computing IT systems perspective.

Key Words: CRM, Hierarchical Bayes, sales support, cross-selling, recommendation systems, decision support systems

1. Background

Making effective recommendations to customers is at the core of effective customer relationship management (CRM) programs. Knowing what cross- and up-selling offers to make based on customer data is accomplished in a variety of ways. The procedures include collaborative filtering, market basket analysis using associative rule induction or other methods, and supervised learning techniques like support vector machines. Hierarchical Bayes models have been used for on line applications, and they have been described by Bodapati (2008), Ansari and Kolhi(1999), and others.

The work we describe here is part of a project in progress for a company that sells products directly to its customers using sales representatives who visit customers assigned to them. Due to the proprietary nature of the data we can't reveal the company, the specific nature of its products, or the nature of the business of its customers. We'll describe it here as Company Z, a supplier to commercial car wash facilities located in the continental United States. Our work addresses one of Company Z's sales management goals, which is the improve the effectiveness of product recommendations made by its sales representatives. Company Z has about 30,000 customers, approximately 125 sales representatives, and 150 products clustered in two main lines reflecting what they are used for. Z's products carry a few main brands, and also nested subbrands. Some of the products are substitutes for one another, while others may be used together in complementary ways.

Company Z aspires to create an information system that combines various sources of data and that produces tactical sales support information that is published to its sales force using channels that include mobile computing devices like smart phones. At the core of the system is a database that accumulates data on customers, their environments, and their behaviors. It also stores information about the behaviors of sales representatives. The data include purchase transaction records, product descriptions, customer characteristics like location and size, and GPS-based measures of visits to customers by sales personnel that include date and duration. Data on the specific content of sales visits, including specific products recommended to the customer, are planned additions.

Company Z's system also has at its core a model component whose predictions are the basis for recommendations by sales personnel to customers. The specification and development of this model is the focus of this paper. Upon pondering this application it struck us that it is different from the kinds of recommendation applications we've seen elsewhere in a couple of important respects. First, published examples in marketing are business-to-customer applications, and not business-to-business (B2B) applications. Because of the B2B context, we have a lot more data on the customers than is typically the case in B2C applications, or at least for the on line ones. Second, because the suggested recommendations needn't be generated in real time but can be done on a weekly basis, we have relative luxury in regard to the computational intensity required to produce them. Finally, due to the nature of Company Z's market, its customer base is unlikely to grow by even an order of magnitude due to the nature of the demand its customers face.

2. A Periodic Review Inventory Model

To further define the modeling problem, we operationalized the ordering decision process by assuming that customers order quantities of products at particular times based on a review of what they currently have in inventory, and their expectations regarding what they should be holding. This "periodic review inventory model" (PRIM) assumes that a customer orders a quantity Q of a product at time t when inventory on hand of the product falls below level r, as follows:

$$r_{k-1} + Q_{k-1} - \left[D(t) - D(t_{k-1})\right] < r_k \qquad (1)$$

In Equation 1., above, D(t) and D(t_{k-1}) are the cumulative demand up to and including time t, and the cumulative demand up to and including the last order point k-1, respectively. Q_{k-1} is the quantity of the last order placed at t_{k-1}. The LHS is the amount of the product on hand, i.e., the inventory up to and including the last order. When this amount falls short of the customer's unobserved requirement at order point k, r_k, new order is made. The r_k are latent quantities: they aren't observed.

To make our PRIM a probability model, we assume that differences between successive r_k are i.i.d. with a logistic CDF. Then we can express the probability p of and order at time t, $t > t_{k-1}$, as:

$$p(\text{order at time } t \geq t_{k-1}) = p\left(\left[r_k - r_{k-1}\right] < \left[D(t) - D(t_{k-1}) - Q_{k-1}\right]\right) .$$

Assuming the logistic CDF we have:

$$p(\text{order at time } t \geq t_{k-1}) = \frac{\exp\left(\left[D(t) - D(t_{k-1}) - Q_{k-1}\right]/\sigma\right)}{1 + \exp\left(\left[D(t) - D(t_{k-1}) - Q_{k-1}\right]/\sigma\right)} \qquad (2)$$

where **σ** is a scale factor to be estimated along with other model parameters that will be described below.

We define the order time distribution conditional on the time of the last order and the quantity of product ordered when the last order was made as:

$$P(T=t | t_{k-1}, Q_{k-1}) = P(\text{Order at t}) \prod_{s=t_{k-1}}^{t-1} P(\text{No Order at s}) \qquad (3)$$

or

$$P(T=t | t_{k-1}, Q_{k-1}) =$$

$$\exp\left[\left[D(t) - D(t_{k-1}) - Q_{k-1}\right]/\sigma\right] \times \prod_{s=t}^{t-1} \left(1 + \exp\left[\left[D(s) - D(t_{k-1} - Q_{k-1})/\sigma\right]\right]\right)^{-1} \qquad (4)$$

Equations 3 and 4 define what is essentially a kind of hazard function that is approximated by a sequence of binary events. Sueyoshi(1995) has described a class of such models that are nested within proportional hazard models. Following Sueyoshi, we approximate a Bernoulli process with a continuous time distribution $f_k(t)$ as follows:

$$f_k(t) = [1 + \exp[-Q_{k-1}/\sigma]][d(t)/\sigma] \times$$

$$\frac{\exp[[D(t) - D(t_{k-1}) - Q_{k-1}]/\sigma]}{[1 + \exp[[D(t) - D(t_{k-1}) - Q_{k-1}]/\sigma]]^2} \quad \text{for } t \geqslant t_{k-1}, \qquad (5)$$

$$d(t) = \frac{dD(t)}{dt} \geqslant 0$$

In Equation 5, d(t) is the weekly demand for a product. The first term on the RHS serves to prevent the predicted time of an order k from preceding the time of the preceding order, k-1. Figure 1 gives some example order time distributions that Equation 5 describes. For illustration we have assumed that cumulative demand, D(t), is a linear function of time, and that the scale factor σ = 1.0. Three curves are shown, corresponding to order quantities Q of 2, 4, and 8. You can see from the Figure how the distributions are left truncated at zero due to the specification of Equation 5.

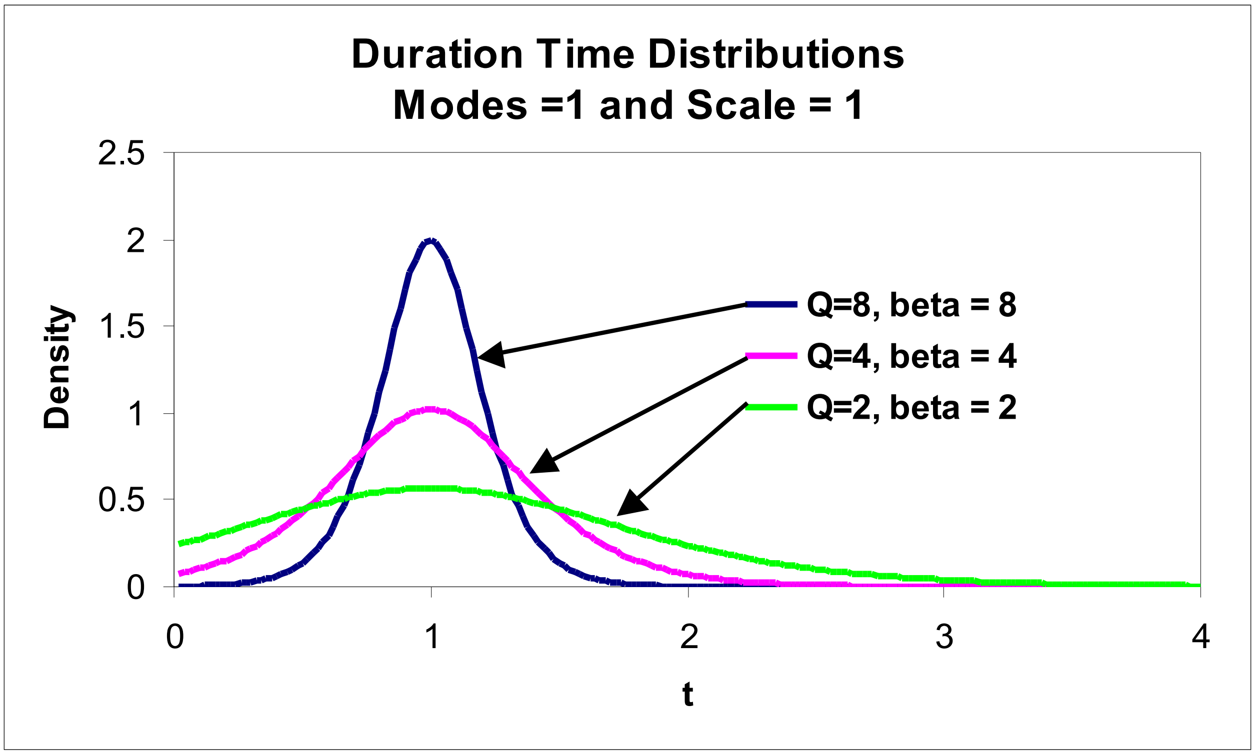

Figure 1. Example order time distributions assuming that cumulative demand, D(t), is linear in time, D(t) = beta*t. Q is the quantity of product units ordered.

3. The Impact Of Sales Visits

One of the main goals for Company Z's system is to predict the impact of specific product recommendations on customer order behavior so that sales representatives can have more impact during their customer visits. At the time of this writing, Company Z was still implementing procedures for capturing reliable data on recommendations made. The data available for modeling included GPS data indicating when sales representatives visit customers and how long they visited them. So as a next step we extended our PRIM to estimate the impacts of visits on demand for each product.

Company Z provided us with a test data set consisting of the product ordering and sales visit histories of 460 of its customers who are located in three different regions of the continental United States. The data include 20,505 distinct orders of 41 different products. The products bore two main brands, and six sub-brands. They represent the purchasing behavior of the 460 customers from the time period spanning January 2007 to December 2007, inclusive. The sales visit data are left truncated at 10 minutes due to the implementation details of Z's automobile-mounted GPS systems. The frequency distribution of visit durations is positively skewed with a mode of approximately 20 minutes. The distribution of product unit quantities is "clumpy," reflecting factors that include product packaging and ad hoc deal making.

Despite these features of the purchase order data, and in the interest of simplicity, we chose to proceed by extending PRIM to be a model of demand as a function of the occurrence and duration of sales visits, customer characteristics, and product attributes. Our extension to PRIM has two components, one expressing demand for specific products by individual customers, and the second representing the effect of sales visits. We specified the first component of demand in terms of deviations from a region-specific "base" demand:

$$D_{i,j}(t) = T(x_j{}'a_i) D_A(t) \quad (6)$$

In Equation 6, $D_{i,j}(t)$ is the demand for product j by customer i at time t. $D_A(t)$ is the base demand in region A, represented by a linear spline function with weekly knots, as illustrated in Figure 2. The linear-exponential function is exponential at values less than b, and is linear for values greater than b. Daily demand d(t) is constant between the knots. $T(x_{j'} a_i)$ adjusts the base demand. x_j is a vector of product attributes, and a_i is a set of parameters for customer i. Figure 2 illustrates type type of linear-exponential function we used for the splines.

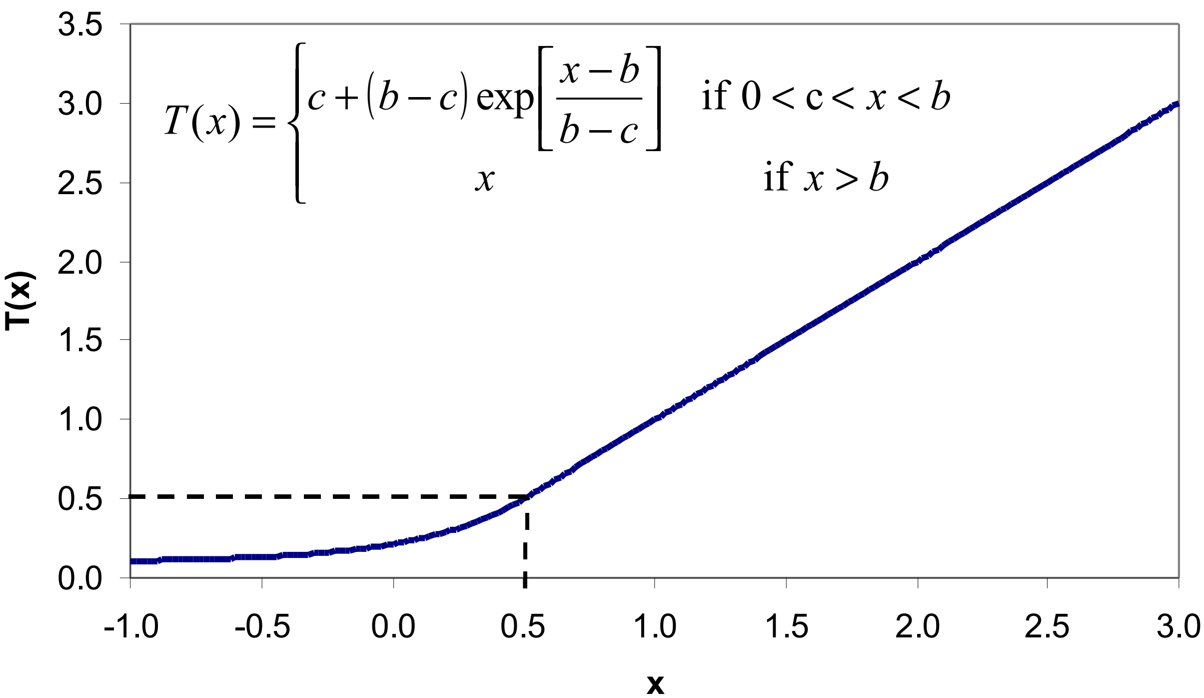

Figure 2. Linear exponential function of the type used for demand splines.

The second component of our PRIM extension captures the impact of sales visits on demand. It is based on two assumptions: (1) the impact of a visit is like a shock or impulse that then decreases over time, and (2) the impacts of visits are cumulative over time. We model the shock, or instantaneous, effect of a single visit to customer i as:

$$v_i(t) = \sum_{u=1}^{U_i} V_{i,u}\left(t - s_{i,u} + 1\right)^{-1} \chi\left(t - s_{i,j}\right), \quad \chi(t) = \begin{Bmatrix} 0 & if & t<0 \\ t & if & t \geq 0 \end{Bmatrix} \qquad (7)$$

and their cumulate effect by integrating over $v_i(t)$:

$$V_i(t) = \int_0^t v_i(s)\,ds \;=\; \sum_{u=1}^{U_i} V_{i,u} \ln\left(t - s_{i,u} + 1\right) \chi\left(t - s_{i,u}\right) \qquad (8)$$

Figure 3 gives an example of how the $v_i(t)$ combine into $V_i(t)$. $V_i(t)$ is the component that will be included in our demand model, below. The top graph in Figure 3 illustrates a sequence of shocks based on Equation 7. The bottom graph shows the sum of the shocks in the top graph.

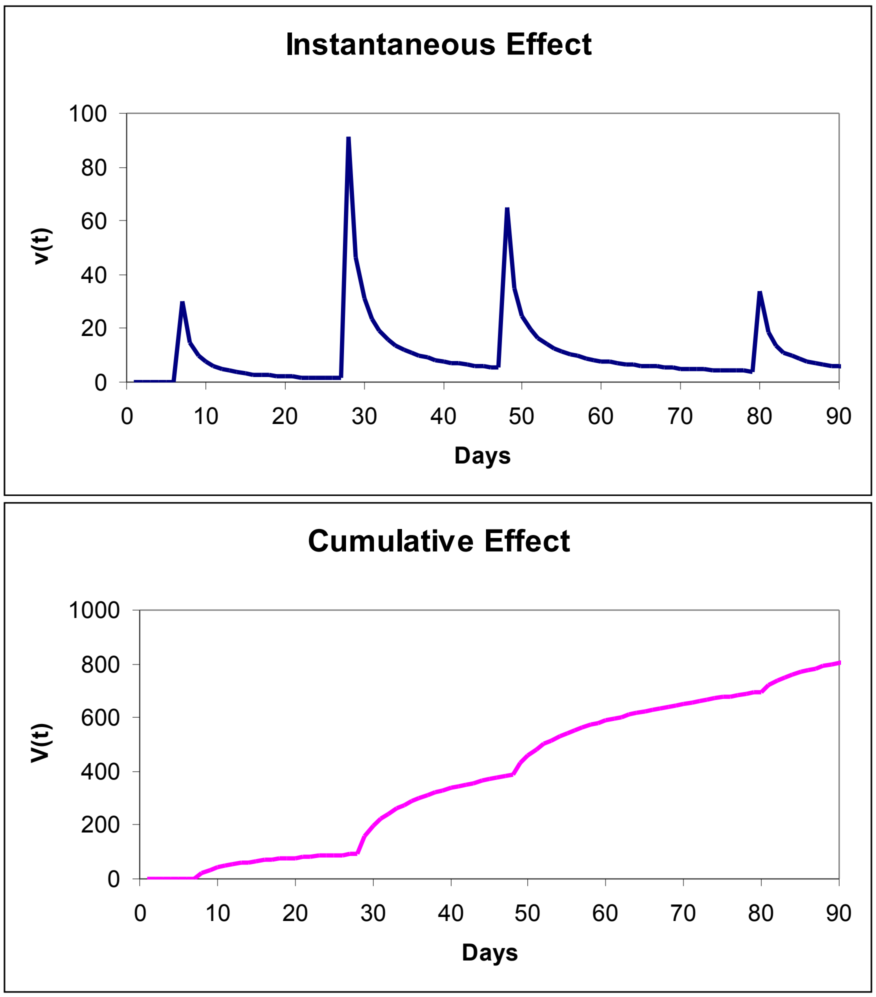

Figure 3. Sales visit functions. Top: The effect of individual sales visits deteriorates over time. Bottom: The effects of individual visits are cumulative.

Analogous to the way we adjusted base demand in Equation 6, we moderate the effect of visits on demand with $T(x_j' \beta_i)$, where x_j is a vector of attributes for product j, and β_i is a vector of coefficients for customer i. Putting all the above together, we can express the order time distribution for a particular product and customer as follows:

$$f(t) = \left[1+\exp\left(-Q_{i,k|j}/\sigma_i\right)\right]\frac{\psi(t)}{\sigma_i}\frac{\exp\left[\left[\Psi(t)-Q_{i,j|k}\right]/\sigma_i\right]}{\left(1+\exp\left[\left[\Psi(t)-Q_{i,j|k}\right]/\sigma_i\right]\right)^2} \quad \text{for } t > t_{i,k-1|j} \qquad (9)$$

where:

$$\Psi(t) = T\left(x_j{}'\alpha_i\right)\left[D_A(t)-D_A\left(t_{i,k-1|j}\right)\right]+T\left(x_j\beta_i\right)\left[V_i(t)-V_i\left(t_{i,k-1|j}\right)\right]$$

and:

$$\psi(t)=T\left(x_j{}'\alpha_i\right)d_A(t)+T\left(x_j{}'\beta_i\right)v_i(t)$$

In Equation 9, $Q_{i,k|j}$ is the quantity of project j ordered by customer i at order point k, and $t_{i,k|j}$ is the time from our arbitrary temporal origin of that order. The other elements of Equation 8 have been defined above.

To complete the model, we assume multivariate normality. Specifically, we assume:

$$\sigma_i \sim T\left(\xi_i\right),$$

$$\xi_i \sim N_1\left(Z_i{}'\theta_\xi, \tau^2\right) \text{ where } Z_i \text{ are customer-specific attributes,}$$

$$\alpha_i \sim N_p\left(\Theta_\alpha Z_i, \Lambda_\alpha\right), \text{ and}$$

$$\beta_i \sim N_p\left(\Theta_\beta Z_i, \Lambda_\beta\right)$$

For identification purposes, we impose the following constraint on the Λ_a matrix:

$$\Lambda_\alpha[1,1] = 1$$

4. Model Estimation And Results

We have estimated this model using MCMC procedures and the test data provided by Company Z. The MCMC chain had 100,000 iterations of which the first 50,000 were burn-in and the last 50,000 iterations were used for estimation where the chain was thinned by a factor of 5. Our results thus far indicate considerable heterogeneity in the impacts of visits on demand for products and customers. Figures 4 and 5 illustrate this. Figure 4 shows the distribution of the estimated posterior means of customers' base demand multipliers defined in Equation 6 as a box plot for each of the 41 products. In each box plot, the box is the interquartile range, the middle bar is the median, and the whiskers are the extreme values. The heterogeneity across customers is obvious. It's also interesting to note that products 1 through 32 share the same main brand, while products 33 through 41 share the other main brand. A qualitative difference between the brands is that the first brand, representing products 1-32, is not in a category of products closely related to the demand for basic car wash services. The second brand, on the other hand, represents inputs that are critical to delivering basic car wash services.

Figure 5 illustrates analogous results for the visit multiplier defined in Equation 8. It's apparent from this figure that the effects of visits on purchases of any particular product vary substantially across customers.

5. Future Development

Our next steps in developing our PRIM methodology include the following. Company Z will soon make available data about specific recommendations made to particular customers on individual sales visits. We'll extend PRIM to take advantage of this data so we can help Z achieve its goal of increasing the effectiveness of recommendations. We'll also be extending PRIM so as to explain variation in quantities ordered. A quantity model is one likely place to express the effect of product-specific recommendations, in fact.

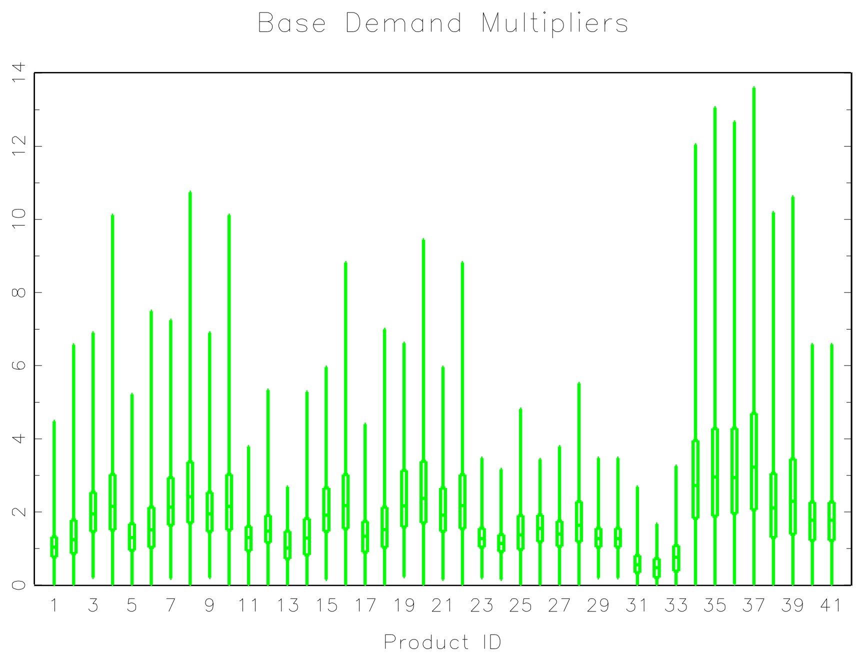

Figure 4. Distributions of mean base demand multipliers defined in Equation 6 for customers by product. Each box is the interquartile range, and the bar in the middle of each box is the median. The whiskers are the extrema.

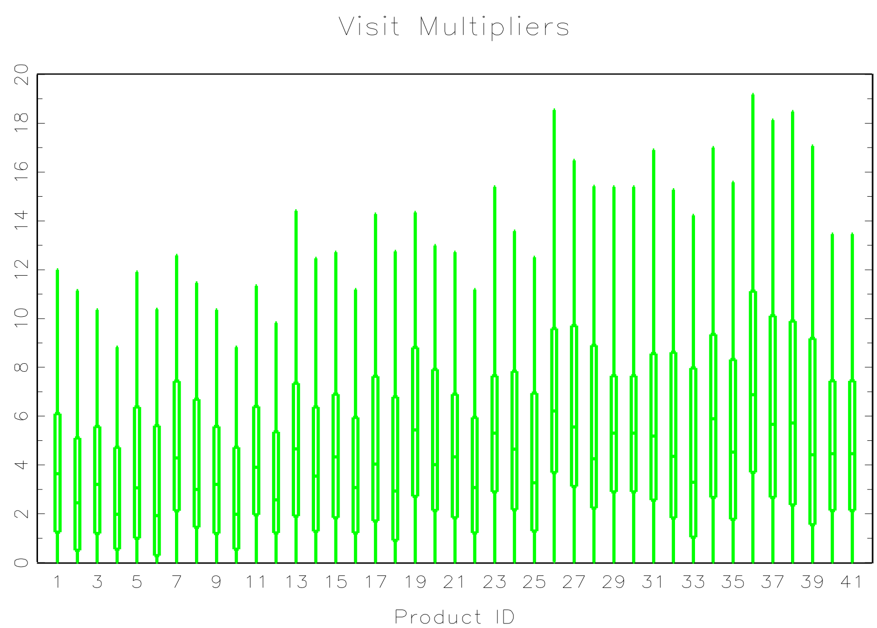

Figure 5. Distributions of visit multipliers. See Figure 4 for details.

6. Acknowledgments

The authors wish to thank Company Z, their anonymous source of data for this project.

7. References

Ansari, A. & Kohli, R. (1999) “Internet recommendation systems.” *Journal of Marketing Research,* XXXVII (Aug.), 363-375.

Bodipati, A. (2008) “Recommendation systems with purchase data.” *Journal of Marketing Research,* Vol. XLV (Feb.), 77-93.

Sueyoshi, G. T. (1995) “A Class of Binary Response Models for Grouped Duration Data,” *Journal of Applied Econometrics*, Vol. 10, No. 4. (Oct. - Dec., 1995), pp. 411-431.