MMRC

DISCUSSION PAPER SERIES

No. 259

One-to-one demand forecasting model and its application for inventory management

Sotaro Katsumata

Graduate School of Economics,the University of Tokyo

Masato Itohisa

Graduate School of Economics,the University of Tokyo

April 2009

東京大学ものづくり経営研究センター

Manufacturing Management Research Center (MMRC)

Discussion papers are in draft form distributed for purposes of comment and discussion. Contact the author for permission when reproducing or citing any part of this paper. Copyright is held by the author.

http://merc.e.u-tokyo.ac.jp/mmrc/dp/index.html

One-to-one demand forecasting model and its application for inventory management

Sotaro KATSUMATA
(Graduate School of Economics, the University of Tokyo)

Masato ITOHISA
(Graduate School of Economics, the University of Tokyo)

Abstract:
In this study, we propose a demand forecasting model for inventory management and apply it to low-involvement and repeat-purchase products, particularly to households' demand for alcoholic beverages. We constructed a model involving household heterogeneity, commonality and time-trend effects. The results revealed that the proposed model maintained a high forecasting performance for long periods. This model can estimate household-level demand and potential tendencies of product consumption. We also discuss the wide range of applications of the model.

Key word: Inventory Management, Marketing Demand Forecasting Model

One-to-one demand forecasting model and its application for inventory management

Sotaro KATSUMATA
(Graduate School of Economics, the University of Tokyo)

Masato ITOHISA
(Graduate School of Economics, the University of Tokyo)

1 Introduction

Demand forecasting is one of the most important issues in the field of supply chain management (Willemain, Smart and Schwarz, 2004). Understanding the demand for final consumption is a fundamental aspect of production planning, including inventory management for each stage of the supply chain, which involves dealers but also OEM (original equipment manufacturers), suppliers and material handlers. In addition, the profit of each company depends on the accuracy of demand forecasting. Although it is almost impossible to predict demand with perfect accuracy, the longer the period, the more difficult it is to forecast demand.

Two streams of research have attempted to solve the above problem. One stream of research aims to develop as accurate a model as possible by utilizing detailed purchase data, for instance, POS (point of sales) data and data obtained through the use of advanced mathematics (e.g. Brown, 1960; Berry and Linoff, 1997; 2000). The other stream of research is based on the BTO (build to order; Holweg and Pil, 2004) approach, which involves initiating production after an order is taken from a customer, as a result of which, there is no need to accurately predict uncertain demand. However, BTO cannot be applied to all products. For example, BTO is a realistic solution for European luxury-car makers but for low-involvement products like food and beverages, if customers do not find the products on store shelves, they go to other shops. Therefore, increasing the accuracy of demand forecasting for low-involvement products is regarded as the first solution. Regardless of the preciseness of a model, it remains a fact that predicting the demand with accuracy is a difficult task because the variance in the demand of the market produces the inventory cost.

Our research will focus on low-involvement products and develop a forecasting model on which the one-to-one marketing (Peppers and Rogers, 1993) method is applied to accumulate individual demands. The models developed in the past view the market as a whole, whereas the new model directs attention to the commonality and heterogeneity of individuals and proposes two views: raising the accuracy of demand forecasting and efficiently promoting the sale of unsold goods to individuals. For instance, providing discount offers through direct mail to those who have latent purchasing power might be an efficient method to sell unsold goods.

In this paper, we begin by conducting a literature review of the demand forecasting model and of one-to-one marketing. We then proceed to construct the one-to-one demand forecasting model and test it using validation data. After that, we discuss the limitations and expansions of the model. Finally, we present the conclusion along with the directions for future research.

2 Literature Review

2.1 Demand Forecasting

Since early times, companies have recognized the importance of demand forecasting. This has led to the development of numerous forecasting methods. In the earliest years, firms believed that demand volume could be predicted from past sales records. Na?ve forecasting, the moving average method and exponential average are commonly used primitive forecasting methods (e.g. Brown, 1959). Na?ve forecasting uses the nearest sales as the predicted value. The moving average and exponential average methods use the data of an appropriate past period with some assigned weighting value. After that, many quantitative models were developed and proposed; some examples of these models are the time-series analysis methods, of which the ARMA (autoregressive moving average) model is well known, and the state-space model (e.g. Hamilton, 1994).

In addition to the above-mentioned methods, regression models that considered market structure and customer characteristics were also developed. These methods are actively researched in the field of marketing. Efforts to achieve predictive accuracy gave rise to the concept of market segmentation. Many segmentation methods were developed to analyse the market segments, for example, mixture regression models (e.g. DeSarbo and Cron, 1988; Kamakura and Russel, 1989; Wedel and Kamakura, 1998). Mixture models divide the whole market into heterogeneous segments, based on which the market aggregate demand can be derived from the sum of each heterogeneous segment, which is a primitive method of one to one demand forecasting. In recent years, data mining techniques like the neural network (Berry and Linoff, 1997; 2000) and database marketing methods (Blattberg, Kim and Neslin, 2008) propose a way to deal with vast amounts of data. In demand forecasting, the volume of data is increasing and the structure of the forecasting models is more complicated.

2.2 One-to-One Forecasting Model

Recently, these market segments have become more fractionizing and are being developed into one-to-one marketing (Rogers and Peppers, 1993) and CRM (Customer Relationship Management; e.g. Reinartz et al., 2004), which assume customer heterogeneity and aim to capture individual demand. The validity of these marketing activities is confirmed on the basis of two reasons. First, the progress of information systems enables us to accumulate enormous quantities of customer information. Second, the introduction of the Bayesian inference allows us to estimate complex models (Rossi, Allenby and McCulloch, 2005). The one-to-one forecasting model is often expressed by the hierarchical Bayes model (e.g. Rossi, McCulloch and Allenby, 1996; Katsumata, 2008). The hierarchical structure is also applied to other objects. Ansari, Essegaier and Kohli (2000) constructed a model that considers customer and product heterogeneities and applied it to movie rating data. Ansari and Mela (2003) and Danaher, Mullarkry and Essegaier (2006) also used this multi-heterogeneous model.

Although one-to-one models have been actively studied in the field of marketing, few researches have applied this model to long-term demand forecasting. In this paper, we construct a model that contains customer heterogeneity and commonality, and then apply it to long-term forecasting. After the estimation, we discuss a wide range of applications for inventory management.

3 Data Overview and Model Construction

Based on the discussion above, in this section, we show the construction and the application of the demand forecasting model. In particular, we forecast the usage of alcoholic beverages by each household. At the same time, we use the ordinary method of forecasting for the purpose of comparing the forecasting of demand by the two methods.

3.1 Data and Target Products

We use the food usage record of households obtained from the 2008 Data Analysis Competition hosted by the Joint Association Study Group of Management Science. These records were collected over a period of one year.

In this research, the analysis and forecast focus on alcoholic beverages. We chose this product category for the following three reasons. First, since the fermentation of alcohol takes a long time, this category of products has PLT (production lead time). Therefore, alcoholic beverages belong to a product category that needs high accuracy in demand forecasting because manufacturers often require a production plan considering PLT. Second, this product category has some flexibility regarding production volume. Therefore, the price of alcohol is stable compared to the prices of vegetables and other raw foods. Third, the products under this category have an expiration date. Even though some alcoholic products such as wine and spirits do not have an expiration date, ordinary brewed beverages have a short shelf life. Since these products with an expiration date face the risk of being discarded, firms have to adjust the production and volume of stock. These reasons show that there is a high requirement for demand forecasting methods in firms producing and selling alcoholic beverages. In the following section, we construct a model to forecast the usage of alcoholic beverages by each household.

3.2 Model Construction

3.2.1 Divide Multiple Decision-Making

In household h, if a product j is used on 20 out of 100 days, we can estimate the usage probability of product j as $20/100 = 0.2$. However, this observed probability contains the following two factors:

- Eating meals at home but not using product j
- Not eating meals at home (eating out)

Here, let us define y_{ht} and z_{htj} as follows:

- $z_{htj} = 1$ if household h uses product j at time t and 0 otherwise.
- $y_{ht} = 1$ if household h makes a meal at time t and 0 otherwise.

Then, we can define the observed food usage probability as a joint probability of z_{htj} and y_{ht}

$$\text{Observed product usage probability} = P(z_{htj} = 1, y_{ht} = 1) \tag{1}$$

When we try to estimate $P(y_{ht} = 1, z_{htj} = 1)$ directly, we need to consider two factors. It is reasonable to estimate each factor. We decompose this probability into two parts. In the above situation, if household h ate out 20 out of 100 days, the observed eat-in probability $P(y_{ht} = 1) = 80/100 = 0.8$, and the conditional food usage probability is $P(z_{htj} = 1|y_{ht} = 1) = 20/80 = 0.25$. These probabilities have the following relations:

$$P(z_{htj} = 1, y_{ht} = 1) = P(y_{ht} = 1)P(z_{htj} = 1|y_{ht} = 1) \tag{2}$$

This means that the observed product usage probability can be decomposed into the multiplication of eat-in probability and conditional food usage probability. However, we cannot observe $P(z_{htj} = 1|y_{ht} = 0)$; therefore, in this paper, we assume z_{htj} and y_{ht} to be independent. Then, we obtain that the observed usage probability of product j is the simple multiplication of eat-in probability and unconditional usage probability of product j.

Next, to estimate the demand volume, we have to estimate the number of attendance at the dinner table. We can interpret this variable as an outcome of the collective decision-making of the family. Here, we formulate the number of attendance at the dinner table. Let M_h be the size of household h and λ_{ht} be the number of attendance at time t in household h. When we observe M_h attendance above, we treat it as $M_h + 1$. Then, we can define the range of λ_{ht} as follows:

$$\lambda_{ht} \in \{1, 2, \cdots, M_h, M_h + 1\} \tag{3}$$

We can estimate the number of attendance by using the discrete ordered choice model with $M_h + 1$ numbers. Let $P(\lambda_{ht} = m)$ be the probability of attendance of m individuals; then, we can calculate the expected number of attendance as

$$\bar{\lambda}_{ht} = \frac{1}{M_h + 1} \sum_{m=1}^{M_h+1} P(\lambda_{ht} = m) \tag{4}$$

Assuming that all decision-making is independent, the demand volume d_{htj} can be expressed by multiplying three factors: the expected number of attendance, eat-in probability and unconditional usage probability of product j.

$$d_{htj} = \bar{\lambda}_{ht} P(y_{ht} = 1)P(z_{htj} = 1) \tag{5}$$

Assuming that the number of households is H, we can express the whole market demand as follows:

$$D_{tj} = \sum_{h=1}^{H} d_{htj} \tag{6}$$

3.2.2 Define Multiple Decision-Making

Based on the settings mentioned above, we can construct a model using a probit-type model. The number of attendance is estimated by using the ordered probit model, while the eat-in probability and product usage probabilities are estimated by using the binomial probit model. Let us bundle these observations as u_{ht},

$$u_{ht} = (y_{ht}, \lambda_{ht}, z'_{ht})' \tag{7}$$

where, $z_{ht} = (z_{ht1}, \cdots, z_{htJ})'$. Let the number of products be J; then, the size of u_{ht} becomes $J + 2$. Then, we set $J^* = J + 2$. In the probit model, Abler and Chib (1993) introduced the latent variable applied by data augmentation (Tanner and Wong, 1987). In this paper, we use latent variables in the same manner as that used in Albert and Chib (1993).

We introduce latent variable for observed vector . When $u^*_{ht} = 1$ and $u_{ht2} \cdots u_{htJ^*}$ are observable, the corresponding relation is as follows:

$$\begin{array}{l} u^*_{htj} > 0 \text{ if } j = 1 \\ c_{hm} \leq u^*_{htj} < c_{h,m+1}, \text{ if } u_{hj2} = m, j = 2 \\ u^*_{htj} \left\{ \begin{array}{l} > 0 \text{ if } u_{htj} = 1 \\ \leq 0 \text{ if } u_{htj} = 0 \end{array} \right. , j = 3, \cdots, J^* \end{array} \tag{8}$$

where cs are the threshold parameters, and $c_{h1} = -\infty$、$c_{h,M_h+1} = 0$、$c_{h,M_h+2} = \infty$ is fixed (Koop, 2003). If $u_{ht1} = 0$, since we cannot observe production usage and the number of attendance, we complement the missing variables stochastically using data augmentation.

$$\left\{ \begin{array}{l} u^*_{htj} \leq 0, j = 1 \\ u_{htj} \sim \mathcal{N}(\mu_{htj}, \sigma^2_{htj}), j = 2, 3, \cdots, J^* \end{array} \right. \quad \text{if } u_{ht1} = 0 \tag{9}$$

If eat-in probability, product usage probabilities and the number of attendance are independent, we can compute each probability using latent variable u^*_{ht} and the probability density function of normal distribution $\phi(\cdot)$.

$$\text{Estimated eat-in probability} : P(u_{ht1} = 1) = \int_{-\infty}^{u^*_{ht1}} \phi(\mu) d\mu \tag{10}$$

$$\text{Estimated usage probability of } j\text{-th product} : P(u_{ht,j+2} = 1) = \int_{-\infty}^{u^*_{ht,j+2}} \phi(\mu) d\mu \tag{11}$$

$$\text{Estimated number of attendance} : \bar{\lambda}_{ht} = \frac{1}{M_h + 1} \sum_{m=1}^{M_h+1} m \int_{c_{hm}}^{c_{h,m+1}} \phi(\mu) d\mu \tag{12}$$

3.2.3 Construction of Proposed Model

In this section, we build the effect factors into the model. At first, we consider household specific factors. Let x_{ht} be the K-dimension explanatory variables of household $h = 1, \cdots, H$ at $t = 1, \cdots, T$. We can obtain the following hierarchical model.

$$\begin{array}{l} u^*_{ht} = x_{ht} B_h + \varepsilon_{ht}, \ \varepsilon_{ht} \sim \mathcal{N}_{J^*}(0, I_{J^*}) \\ \beta_{hk} = w_{hk} \Delta_k + \zeta_{hk}, \ \zeta_{hk} \sim \mathcal{N}_{J^*}(0, \Theta_k) \end{array} \tag{13}$$

where $B_h = (\beta_{h1}, \cdots, \beta_{hK})$. In this paper, we assume that each factor is independent. Let the covariance matrix be the identity matrix I_{J^*}. Again, considering the time-trend factor, we can estimate u^*_{ht} as a state space model.

$$\begin{aligned} u^*_{ht} &= F\gamma_t + \varepsilon_t,\ \varepsilon_t \sim \mathcal{N}_{J^*}(0, I_{J^*}) \\ \gamma_t &= G\gamma_{t-1} + \nu_t,\ \nu_t \sim \mathcal{N}_{J^*}(0, \Phi) \end{aligned} \tag{14}$$

Combining the structures above, we construct a model considering household heterogeneities and time trend. In this paper, we assume the state space term to be a local-level model, $F = G = I_{J^*}$.

$$\begin{aligned} u^*_{ht} &= x_{ht}B_h + \gamma_t + \varepsilon_{ht},\ \varepsilon_{ht} \sim \mathcal{N}_{J^*}(0, I_{J^*}) \\ \beta_{hk} &= w_{hk}\Delta_k + \zeta_{hk},\ \zeta_{hk} \sim \mathcal{N}_{J^*}(0, \Theta_k) \\ \gamma_t &= \gamma_{t-1} + \nu_t,\ \nu_t \sim \mathcal{N}_{J^*}(0, \Phi) \end{aligned} \tag{15}$$

Since this model has a hierarchical structure, we estimate the parameters using the MCMC (Markov Chain Monte Carlo) method. The detailed procedures of estimation are described in the Appendix.

3.2.4 Explanatory Variables

In this paper, we use the following explanatory variables: x_{ht} and w_{hk},

$$x_{ht} = \begin{pmatrix} \text{Intercept} \\ \text{Wifeś Holiday (D)} \\ \text{Husbandś Holiday (D)} \end{pmatrix}, \tag{16}$$

$$w_{hk} = \begin{pmatrix} \text{Intercept} \\ \text{Income} \\ \text{Full Time Housewife (D)} \\ \text{Wife has a collage education (D)} \\ \text{Husband has a collage education (D)} \\ \text{log(Age of housewife)} \\ \text{Age difference between husband and wife} \\ \text{Number of elderly persons} \end{pmatrix} \tag{17}$$

where (D) denotes dummy variables and x_{ht} is a dummy variable vector. In household h, if time t is a holiday of the wife or husband, the corresponding element becomes 1.

3.3 Target Period

In this paper, although we have data for one year, to estimate the annual trend, we divide the data into two parts and treat it as data for two years. To be more precise, there are 194 households in the data. We select 100 households to estimate parameter γ, and estimate the residual 94 households using the estimated γ as the given parameter. In addition, in these 94 households, we estimate the one-to-one parameter B_h using data for the first half of the year and, for forecasting, we use the data for the latter half of the year as validation data. Therefore, we accomplish the following:

1. The full year estimation of 100 households

2. The estimation of the first half of the year for 94 households using γ of the full year estimation as the given parameter

3. The forecasting of the latter half of the year of 94 households

Figure 1 provides the details of these calculations.

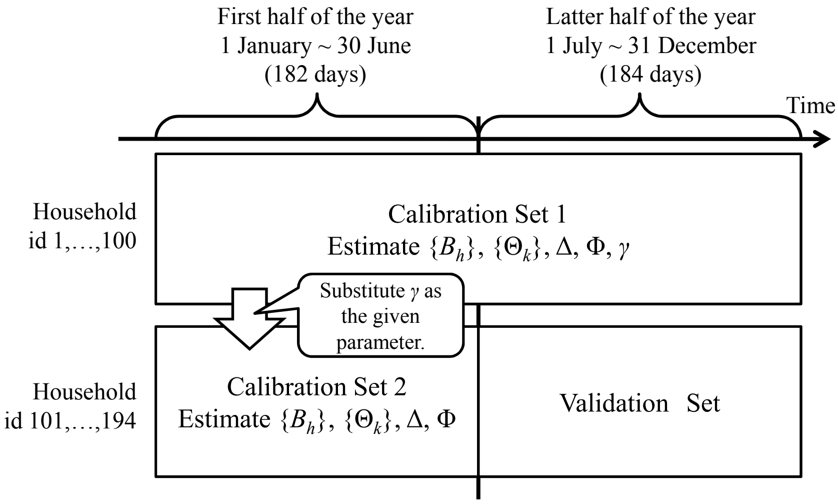

Figure 1: Calibration and Validation Period

3.4 Comparison Models

To compare the forecasting performance of the proposed model, we use the following two models.

Simple Trend Model

As mentioned above, there are many cases wherein time-series analysis methods are used as a demand forecasting model. Therefore, we use a simple time-trend model as the comparison model. This model estimates the aggregate demand.

At time t, we will estimate the aggregated market demand $D_t = (D_{t1}, \cdots, D_{tJ})'$:

$$\begin{aligned} D_t &= \alpha_t + \varepsilon_t,\ \varepsilon_t \sim \mathcal{N}_J(0, I_J) \\ \alpha_t &= \alpha_{t-1} + \nu_t,\ \nu_t \sim \mathcal{N}_J(0, \Psi) \end{aligned} \tag{18}$$

In this model, we use only Calibration Set 1 of Figure 1 for full year estimation. After estimation, we forecast the demand of the validation set using trend parameter α.

$$\tilde{D}_t = \frac{94}{100}\alpha_t \tag{19}$$

Naïve Forecasting

As a primitive method, we use naïve forecasting as our second comparison model. We consider the aggregate demand of the target set at 30 June (one day prior to the calibration period) as the future demand.

4 Result

4.1 Trend and Regression Coefficients

Figure 2 shows the estimated time trend. For beer and sake, we plot the observed aggregate-demand movement on a thin line.

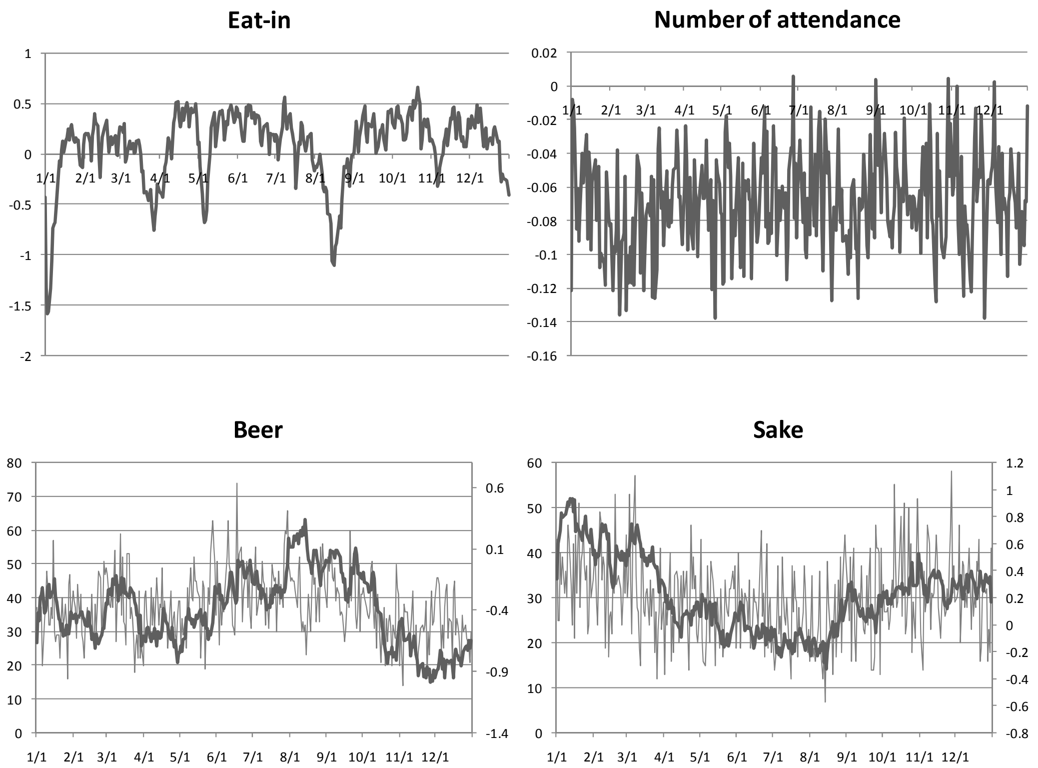

Figure 2: Posterior Mean of Trend Parameter γ

The eat-in trend shows that there are some periods in which demand declines. The first such period is during the New Year holidays. We also observe a decline in demand in early May, which is when schools are closed and people go on trips. Another period during which a decline in demand is observed is mid-August (Bon period), which is when families take days off and visit their hometowns. Compared to eat-in probabilities, there is no obvious trend with regard to the number of attendance. With regard to the trends for beer and sake, beer consumption rises in summer, while sake consumption falls in the same season. Considering all these trends, we find that this model divides the demand factors of usage. For example, although the observed beer demand falls around 15 August, the estimated beer trend rises in this season. This decline in demand can be explained by the eat-in trend.

Hereafter, we discuss the estimation result of commonality parameter Δs. Δs indicates the relation between eat-in probability, number of attendance and product usage probability, and demographic variables. Since we set x_{ht} = (intercept, wife's holiday, husband's holiday), we can obtain Δ for each variable. Table 1 shows the posterior mean of Δs.

Δ_1(Intercept)	Eat-in	# of attendance	Beer	Sake
Intercept	-1.02	-0.60	-2.28	-12.26 *
Income	0.09	-0.02	0.23	-0.07
Full Time Housewife	-0.50	0.09	1.36 *	1.95 *
Wife has a collage education	-0.31	0.01	-0.51	-0.15
Husband has a collage education	0.13	0.00	0.50	0.06
Age of housewife	0.60	0.11	-0.67	2.04
Age difference between husband and wife	0.00	0.00	0.04	0.05
Number of elderly persons	-0.06	0.00	0.39	0.08
Δ_2(Wife's Holiday)	Eat-in	# of attendance	Beer	Sake
Intercept	-0.39	-0.53	-2.12	-10.99 *
Income	-0.04	-0.01	0.38 *	0.22
Full Time Housewife	0.36	0.02	-0.09	-0.16
Wife has a collage education	0.14	-0.01	-1.33 *	-1.01
Husband has a collage education	0.02	0.01	0.30	0.11
Age of housewife	0.21	0.11	-0.12	2.36
Age difference between husband and wife	0.04	0.00	-0.01	-0.01
Number of elderly persons	0.07	-0.01	0.29	-0.07
Δ_3(Husband's Holiday)	Eat-in	# of attendance	Beer	Sake
Intercept	-2.60	-0.41	2.06	-10.35 *
Income	-0.10	0.00	0.35 *	0.06
Full Time Housewife	-0.37	0.05	0.97 *	1.57 *
Wife has a collage education	0.16	-0.01	-0.98	-0.55
Husband has a collage education	-0.23	0.02	1.13 *	0.76
Age of housewife	0.84 *	0.08	-1.41	1.79
Age difference between husband and wife	0.00	0.00	0.05	0.06
Number of elderly persons	0.25	-0.02	0.15	-0.07

Table 1: Posterior Mean of Commonality Parameter Δs

Note that if a parameter is significant at 95%, we add '*' to the right side of the figure. It can be seen that households with full-time housewives tend to consume both beer and sake more than other households. Furthermore, this tendency becomes stronger when their husbands take holidays. In households with older housewives, the eat-in probability increases when husbands take holidays. We can infer the household tendency of product usage from the relation between holiday and demographic information. This result can be applied to marketing activity.

4.2 Predictive Performance Comparison

As shown in Figure 1, we predict the product demand of 94 households for the latter half of the year.

We obtain the demand volume directory from the MCMC samples; let $\hat{d}_{htj}^{(n)}$ be the estimated demand calculated from the n-th sample of N samples obtained by the following procedures.

First, from the following integration, we define eat-in probability $P(u_{ht1}^{(n)} = 1)$, usage probability of the j-th product $P(u_{h,j+2}^{(n)} = 1)$ and the mean number of attendance $\lambda_{ht}^{(n)}$:

$$P(u_{ht}^{(n)} = 1) = \int_{-\infty}^{(u_{ht1}^*)^{(n)}} \phi(\mu) d\mu \tag{20}$$

$$P(u_{ht,j+2}^{(n)} = 1) = \int_{-\infty}^{(u_{ht,j+2}^*)^{(n)}} \phi(\mu) d\mu \tag{21}$$

$$\lambda_{ht}^{(n)} = \frac{1}{M_h + 1} \sum_{m=1}^{M_h+1} m \int_{c_{hm}^{(n)}}^{c_{h,m+1}^{(n)}} \phi(\mu) d\mu \tag{22}$$

From the values provided above, we obtain the estimated demand.

$$\hat{d}_{htj}^{(n)} = \lambda_{ht}^{(n)} P(u_{ht1}^{(n)} = 1) P(u_{ht,j+2}^{(n)} = 1) \tag{23}$$

The estimated demand of product j of household h at time t is derived from the mean of the samples.

$$\hat{d}_{htj} = \frac{1}{N} \sum_{n=1}^{N} \hat{d}_{htj}^{(n)} \tag{24}$$

By summing this individual demand for all H households, we can obtain the estimated demand at time t.

$$\hat{D}_{htj} = \frac{1}{H} \sum_{h=1}^{H} \hat{d}_{htj} \tag{25}$$

We also use comparison models (simple trend model and naïve forecasting) for forecasting. Figure 3 shows the cumulative volume of the prediction results of each model and the realized demand. It seems that the predictive performances of the proposed model and the simple trend model are superior to that of na?ve forecasting. In the case of beer demand, the proposed model fits the realized demand for the next four months, after which the performance of the simple trend model is better than that of the proposed model. In the case of sake demand, among the three models, the proposed model has the highest performance throughout the period. Particularly for the next two months, the proposed model has a fairly good forecasting performance. The predictive ability of the proposed model is satisfactory, especially for recent periods. In practice, many manufacturers and retailers routinely make updates on a short term basis through the use of the rolling horizon-type prediction adjustment. In these circumstances, the proposed model is more feasible.

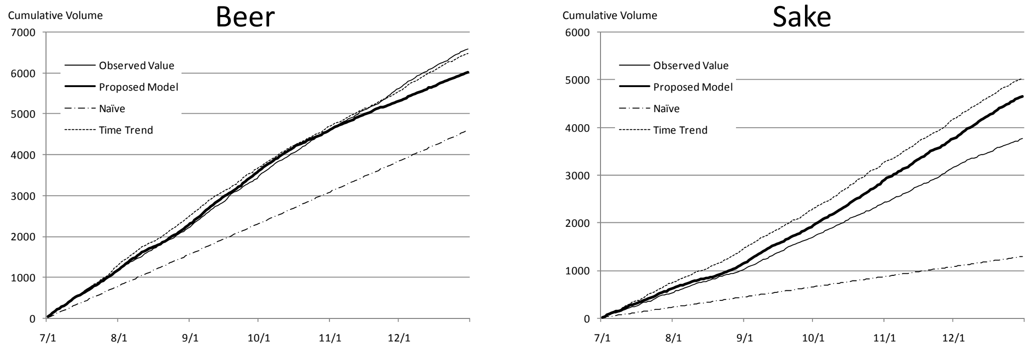

Figure 3: Prediction Performance

5 Discussion

5.1 Applicable Scope of the Model

In this section, we discuss the applicable scope of the proposed model. As mentioned above, the proposed model can be applied to low-involvement and repeat purchase goods. Further, we assume that these products can be purchased through more than one channel. Therefore, we can apply this model to almost all commodity goods.

Firms dealing in these products face severe competition in customer acquisition. They try various expedients to acquire and enclose customers to the greatest extent possible. These firms have to sell their products not only by forecasting passive demand but also by actively selling their products. Although many researchers aim to improve forecasting performance of demand forecasting methods, it is impossible for firms to eliminate unsold goods that face uncertain demand. Even if the distribution, mean and variance of demand are completely known to us, the dead stock volume would increase when we aim to raise the safety inventory level to decrease opportunity loss. Previously, to solve this unsold goods problem, firms would adjust the subsequent order quality. However, in the following discussion, we will discuss how to actively adjust demand through sales promotion.

5.2 Targeting

If household-level demand volume could be predicted, it would help implement marketing activities such as sales promotion. Hereafter, we introduce a method of customer discrimination that targets and that is applicable to validation data.

We use joint-usage probability $\hat{q}_{htj}$ to rank customers. This is obtained by multiplying the unconditional product-usage probability and eat-in probability. We exclude the number of attendance because it is reasonable to assume that when housewives are contacted, they cannot adjust the number of attendance.

$$\hat{q}_{htj} = \frac{1}{N}\sum_{n=1}^{N} P(u_{ht}^{(n)} = 1)P(u_{ht,j+2}^{(n)} = 1) \tag{26}$$

In this research, we set the threshold as 0.5. If q_{itj} is greater than or equal to 0.5, we determine that household h will purchase the jth product. We use hit rate to measure performance. Hit rate indicates the proportion of households that match prediction and observation. This indicator ranges from 0 to 1. Since aggregate models such as the simple trend model and naïve forecasting cannot make predictions at the household level, we provide only the result of the proposed model. We calculate household scores from 1 June

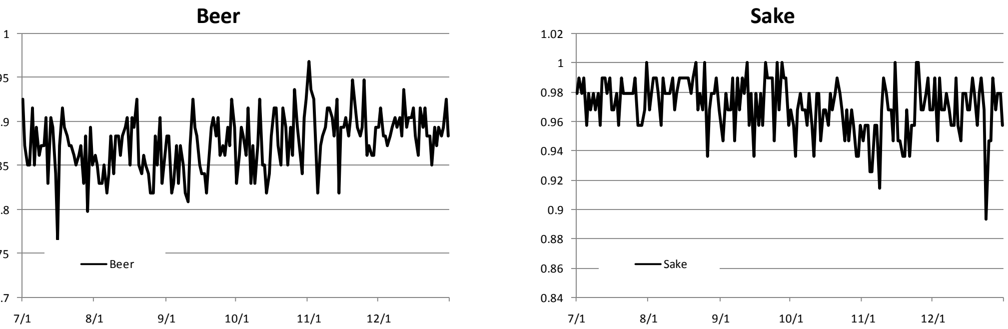

Figure 4: Hit Rate

until 6 months after. Figure 4 shows the hit rate of each day. The proposed model can be applied as a customer determination method as this figure shows that the proposed model maintains a high performance throughout the period. This implies that a household that has a high estimated usage probability actually uses the products. Based on this result, we can implement sales promotion for customers who have high probability of using the product but do not actually use the product. Additionally, the proposed model can be used to realize customer retention and improve customer value.

5.3 PED (Perfect Enclosed Demand)

As mentioned above, the proposed model estimates the market demand from the accumulated demand of all households. Although we determined that the proposed model maintains a high predictive performance even for demand forecasting at the market-aggregate level, this model can also predict the demand at the household level. Furthermore, since this model divides household eat-in probability and product usage probabilities, we can estimate the demand when a household eats dinner on all the days. We term this indicator as *PED* (*perfect enclosed demand*). PED can be derived from estimate demand restricted as $P(u_{ht}^{(n)} = 1) = 1$; therefore,

$$\hat{d}_{htj}^{PED} = \frac{1}{N}\sum_{n=1}^{N} \lambda_{ht}^{(n)} P(u_{ht,j+2}^{(n)} = 1) \tag{27}$$

Figure 5 depicts the estimated demand at the household level, observed demand and PED. The upper two figures show the findings for beer, while the lower two figures show the findings for sake. The estimated demand at the household level can predict observed demand with high accuracy. PED increases dramatically in households with a low eat-in probability, while it increases only marginally in households with a high eat-in probability.

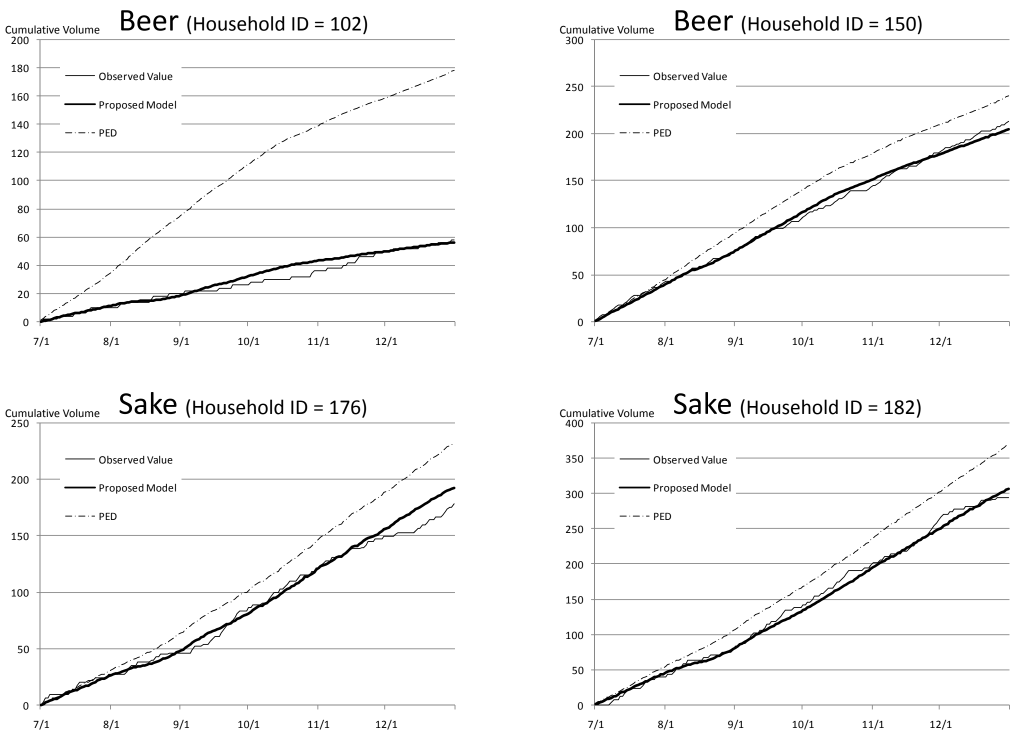

Figure 5: Estimated Demand, Observed Demand, PED

This implies that we can improve demand until PED is attained. We can use PED as the basic information for customer share (Verhoef, 2003) and customer LTV (lifetime value; Blattberg, Getz and Thomas, 2001; Blattberg, Kim and Neslin, 2008; Abe, 2008).

6 Conclusion

This research proposes a model that is a hierarchical Bayes model containing one-to-one prediction and a time-trend factor. This model estimates market demand from accumulated demand at the household level and maintains high predictive accuracy. We have shown that this model has an adequate ability to be used as a demand forecasting model. From this model, we can estimate household-level usage probability, which is also highly accurate. Further, we can apply the model not only to passive inventory control but also to basic information concerning sales promotion such as product recommendation. Furthermore, since decision-making is divided into various factors, PED and the degree of enclosure for each household can be estimated. This information will serve as meaningful information that can be used by firms implementing CRM programmes.

With regard to the application of the model, we raise the following two issues. The first issue is that the model can be expanded and we can propose more relaxed models. In this research, we assume independence between the number of attendance and product usage probability, although it is possible to assume that these factors are correlated. While it is necessary to consider the specific covariance structure to maintain the independence of eat-in probability, we can estimate this matrix using the M-H algorithm proposed by Manchanda, Ansari and Gupta (1999). In this research, we fix the state-space term as the simple trend model (local-level model). We can expand other forms of the state-space model that can express the time-series behaviour more appropriately.

The second issue is the application of this model to other products. This model is applicable to many firms and stores that record POS data. If the firms and stores have additional data such as visit records of other stores, they can estimate PED for every customer of electronics retail stores and supermarkets. In the future, we need to validate a model that considers long-term economic conditions. In this paper, the proposed model uses one-year records. However, it is better to use two-year records and test the predictive accuracy of the model. Additionally, it is necessary to validate the robustness of the model using other products and data.

Acknowledgement
The authors would like to thank Yanyan Jiang and Mark Tan for advice and feedback. Also, Joint Association Study Group of Management Science for providing data, and the University of Tokyo Manufacturing Management Research Center (MMRC) for their research support.

A Appendix

A.1 Detailed Formulation

The proposed model can be expressed as

$$\begin{array}{l} u_{ht}^{*} = x_{ht}B_h + \gamma_t + \varepsilon_{ht},\ \varepsilon_{ht} \sim \mathcal{N}_{J^*}(0, I_{J^*}) \\ \beta_{hk} = w_{hk}\Delta_k + \zeta_{hk},\ \zeta_{hk} \sim \mathcal{N}_{J^*}(0, \Theta_k) \\ \gamma_t = \gamma_{t-1} + \nu_t,\ \nu_t \sim \mathcal{N}_{J^*}(0, \Phi) \end{array} \tag{28}$$

In this paper, household demographic information w_{hk} can be updated within the period $t = 1, \cdots, T$. To consider the above condition, we redefine the model. First, we explain the case of household h. If household h updates its demographic information at time t^*, there are two demographic variables covering time $1, \cdots, t^* - 1$ and time $t^*, \cdots, T$. Let the demographic variables be $w_{h_1k}, w_{h_2k}, k = 1, \cdots, K$ and the range of these variables be $\mathscr{T}_{h_1} and \mathscr{T}_{h_2}$. We can say that $\mathscr{T}_{h_1} = \{1, \cdots, t^* - 1\}, \mathscr{T}_{h_2} = \{t^*, \cdots, T\}$. Let a set of household information variables collected from h be $\mathscr{C}_h = \{h_1, h_2\}$. When we aggregate these sets, we get $\bigcup_{i\in\mathscr{C}_h} \mathscr{T}_i = \{1, \cdots, T\}$. Therefore, we can cover the entire period for mutually exclusive and collectively exhaustive situations.

Hereafter, we will provide the formulation in general. Let any household information be $i = 1, \cdots, L$ and the range of information be $\mathscr{T}_i$. Suppose that the information is collected from household h and let the set of household information of h be $\mathscr{C}_h$; then, $i \in \mathscr{C}_h$. When all information collected from h is aggregated, the range of time equals $\{1, \cdots, T\}$. Therefore, $\bigcup_{i\in\mathscr{C}_h} \mathscr{T}_i = \{1, 2, \cdots, T\}$. Considering these discussions, the proposed model can be redefined as

$$\begin{array}{l} u_{it}^{*} = B_i x_{it} + \gamma_t + \varepsilon_{it},\ \varepsilon_{it} \sim \mathcal{N}_{J^*}(0, I_{J^*}), t \in \mathscr{T}_i \\ \beta_{ik} = \Delta_k w_{ik} + \zeta_{ik},\ \zeta_{ik} \sim \mathcal{N}_{J^*}(0, \Theta_k) \\ \gamma_t = \gamma_{t-1} + \nu_t,\ \nu_t \sim \mathcal{N}_{J^*}(0, \Phi) \end{array} \tag{29}$$

Let the set of household information that contains time t be $\mathscr{I}_t$, where the element count of $\mathscr{I}_t$, $n(\mathscr{I}_t) = H, \forall t \in \{1, \cdots, T\}$.

A.2 Prior Distributions

We set prior distributions as follows:

$$\beta_{ik} \sim \mathcal{N}_{J^*}(\Delta_k w_{ik}, \Theta_k),\ k = 1, \cdots, K \tag{30}$$
$$\Delta_k \sim \mathcal{N}_{J^* \times P}(\Delta_0, \Psi_0),\ k = 1, \cdots, K \tag{31}$$
$$\Theta_k \sim \mathcal{W}(\nu_0, S_0),\ k = 1, \cdots, K \tag{32}$$
$$\gamma_0 \sim \mathcal{N}_{J^*}(\gamma_0, \Phi_0) \tag{33}$$
$$\gamma_t \sim \mathcal{N}_{J^*}(\gamma_{t-1}, \Phi) \tag{34}$$
$$\Phi \sim \mathcal{W}(p_0, P_0) \tag{35}$$

where, $\Delta_0 = O_{J^* \times P},\ \Psi_0 = 1000 I_{J^*},\ \nu_0 = P,\ S_0 = 1000 I_{J^*},\ \gamma_0 = 0,\ \Phi_0 = 1000 I_{J^*},\ p_0 = J^*,\ P_0 = 1000 I_{J^*}$.

A.3 Posterior Distributions

In this section, we show posterior distributions. Detailed explanations of the MCMC method can be found in Gamerman (1998) and Rossi, Allenby and McCulloch (2005). Rowe (2002) discusses the generation of random variables from a matrix distribution. Geweke (1991) shows an efficient generation method of truncated normal random variables.

- Latent variable u

  We draw samples for each element,

  - $j = 1$
  $$\begin{cases} u^*_{it1} \sim \mathcal{TN}_{(0,\infty)}(B_{i1\cdot}x_{it} + \gamma_{t1}, 1), \text{ if } u_{it1} = 1 \\ u^*_{it1} \sim \mathcal{TN}_{(-\infty,0]}(B_{i1\cdot}x_{it} + \gamma_{t1}, 1), \text{ if } u_{it1} = 0 \end{cases} \tag{36}$$
  - $j = 2$
  $$\begin{cases} u^*_{it2} \sim \mathcal{TN}_{(c_{i,m},c_{im+1})}(B_{i2\cdot}x_{it} + \gamma_{t2}, 1), \text{ if } u_{it2} = m \wedge u_{it1} = 1 \\ u^*_{it2} \sim \mathcal{N}(B_{i2\cdot}x_{it} + \gamma_{t2}, 1), \text{ if } u_{it1} = 0 \end{cases} \tag{37}$$
  - $j = 3, \cdots, J^*$
  $$\begin{cases} u^*_{itj} \sim \mathcal{TN}_{(0,\infty)}(B_{ij\cdot}x_{it} + \gamma_{tj}, 1), \text{ if } u_{itj} = 1 \wedge u_{it1} = 1 \\ u^*_{itj} \sim \mathcal{TN}_{(-\infty,0]}(B_{ij\cdot}x_{it} + \gamma_{tj}, 1), \text{ if } u_{itj} = 0 \wedge u_{it1} = 1 \\ u^*_{itj} \sim \mathcal{N}(B_{ij\cdot}x_{it} + \gamma_{tj}, 1), \text{ if } u_{it1} = 0 \end{cases} \tag{38}$$

- Threshold parameter c

  For household information i, c_{i1j}, c_{iM_ij} and $c_{i,M_i+1,j}$ of c_{imj}, $j = 1, \cdots, J^*$ are fixed as $c_{i1j} = -\infty, c_{iM_ij} = 0$ and $c_{i,M_i+1,j} = \infty$. Then, we draw samples for $m = 2, \cdots, M_i - 1$. An estimate algorithm of the ordered probit model is found in Koop (2003).
  $$c_{im} \sim \mathcal{U}(\bar{c}_{i,m-1}, \bar{c}_{i,m+1}) \tag{39}$$

- Parameter $B_i = (\beta_{i1}, \cdots, \beta_{iK})$
  $$\beta_{ik} \sim \mathcal{N}(M, V) \tag{40}$$
  where
  $$\begin{aligned} V &= \left(\textstyle\sum_{t\in\mathcal{T}_i} x'_{itk}x_{itk} + \Theta_k^{-1}\right)^{-1} \\ M &= V\left(\textstyle\sum_{t\in\mathcal{T}_i} x'_{it}(u^*_{it} - B_{i,-k}x_{it,-k} - \gamma_t) + \Theta_k^{-1}\Delta_k w_{ik}\right) \end{aligned} \tag{41}$$
  $B_{i,-k}$ is a matrix that eliminates the k-th element from B_i. $x_{it,-k}$ is a vector that eliminates the k-th element.

- Parameter Δ
  $$\Delta_k \sim \mathcal{N}(M, V, \Theta_k) \tag{42}$$
  where
  $$\begin{aligned} V &= (W'_kW_k + \Theta_{k0})^{-1} \\ M &= V(W'_kB_k + \Theta_{k0}\Delta_0) \end{aligned} \tag{43}$$

- Parameter Θ

$$\Theta_k \sim \mathcal{IW}(v_1, V_1) \tag{44}$$

where

$$\begin{aligned} v_1 &= v_0 + L \\ V_1 &= \left(\sum_{i=1}^{L} (\beta_{ik} - \Delta_k w_{ik})' (\beta_{ik} - \Delta_k w_{ik}) + S_0\right)^{-1} \end{aligned} \tag{45}$$

- Parameter γ

We reconstruct the model for γ_t:

$$u^*_{it} - B_i x_{it} = \gamma_t + \varepsilon_{it},\ \varepsilon_{it} \sim \mathcal{N}_{J^*}(0, I_{J^*}) \tag{46}$$

We sum the above expression for household information $\mathscr{I}_t$, which contains time t.

$$\Longleftrightarrow \sum_{i\in\mathscr{I}_t} (u^*_{it} - B_i x_{it}) = H\gamma_t + \sum_{i\in\mathscr{I}_t} \varepsilon_{it},\ \sum_{i\in\mathscr{I}} \varepsilon_{it} \sim \mathcal{N}_{J^*}(0, H^2 I_{J^*}) \tag{47}$$

Dividing both sides by H, we get

$$\frac{1}{H}\sum_{i\in\mathscr{I}_t} (u^*_{it} - B_i x_{it}) = \gamma_t + \frac{1}{H}\sum_{i\in\mathscr{I}_t} \varepsilon_{it},\ \frac{1}{H}\sum_{i\in\mathscr{I}} \varepsilon_{it} \sim \mathcal{N}_{J^*}(0, \frac{1}{H^2} H^2 I_{J^*}) \tag{48}$$

Let $\tilde{u}^*_{it} = \frac{1}{H}\sum_{i\in\mathscr{I}_t} (u^*_{it} - B_i x_{it})$, $\tilde{\varepsilon}_{it} = \frac{1}{H}\sum_{i\in\mathscr{I}_t} \varepsilon_{it}$; then, we obtain the following state space model:

$$\begin{cases} \tilde{u}^*_t = \gamma_t + \tilde{\varepsilon}_t,\ \tilde{\varepsilon}_t \sim \mathcal{N}_{J^*}(0, I_{J^*}) \\ \gamma_t = \gamma_{t-1} + \nu_t,\ \nu_t \sim \mathcal{N}_{J^*}(0, \Phi) \end{cases} \tag{49}$$

From this model, we sequentially draw γ_t, using the Kalman filter and smoothing.

- Parameter Φ

$$\Phi \sim \mathcal{IW}(v_1, V_1) \tag{50}$$

where

$$\begin{aligned} v_1 &= p_0 + T \\ V_1 &= \sum_{t=2}^{T} (\gamma_t - \gamma_{t-1})'(\gamma_t - \gamma_{t-1}) + P_0^{-1} \end{aligned} \tag{51}$$

References

Abe, M. (in press). Counting your customers one by one: A hierarchical Bayes extension to the Pareto/NBD model, *Marketing science*.

Albert, J. H. & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data, *Journal of American Statistical Association*, **88**, 669-679.

Ansari, A. & Mela, C. F. (2003). E-Customization, *Journal of Marketing Research*, **40**,131-145.

Berry, M. J. A. & Linoff, G. S. (1997). *Data mining techniques: For marketing, sales, and customer relationship management*, New York: Wiley.

Berry, M. J. A. & Linoff, G. S. (2000). *Mastering data mining: Art and science of customer relationship management*, New York: Wiley.

Blattberg, R. C., Getz, G. & Thomas, J. (2001) *Customer equity: Building and managing relationships as valuable assets*, Boston: Harvard Business School Press.

Blattberg, R. C., Kim, B. & Neslin, S. A. (2008). *Database marketing: Analyzing and managing customers*, Springer.

Brown, R. G. (1959). *Statistical forecasting for inventory control*, New York: McGraw-Hill.

Danaher, P. J., Mullarkey, G. W. & Essegaier, S. (2006). Factors affecting web site visit duration: a cross-domain analysis, *Journal of Marketing Research*, **43**, 182-194.

DeSarbo, W. S. & Cron W. L. (1988). A maximum likelihood methodology for clusterwise linear regression, *Journal of Classification*, **5**, 249-282.

Hamilton, J. D. (1994). *Time series analysis*, Princeton University Press.

Holweg, M. and Pil, F. K. (2004). *The second century: Reconnecting customer and value chain through build-to-order*, MIT Press.

Kamakura, W. R., & Russel, G. J. (1989). A probabilistic choice model for market segmentation and elasticity structure, *Journal of Marketing Research*, **26**, 379-90.

Katsumata, S. (2008). Estimating individual preferences in the music CD market, *Annals of Business Administrative Science*, **7**, 19-32.

Koop, G. (2003). *Bayesian econometrics*, Chichester: Wiley.

Manchanda, P., Ansari, A. & Gupta, S. (1999). The shopping basket: a model for multicategory purchase incidence decisionsters,*Marketing science*, **18**, 95-114.

Peppers, D & Rogers, M. (1993). *The one to one future: Building relationships one customer at a time*, New York: Currency Doubleday.

Reinartz, W., Krafft, M., & Hoyer, W. D. (2004). The customer relationship management process: Its measurement and impact on performance, *Journal of Marketing Research*, **41**, 293-305.

Rossi, P. E., Allenby, G. & McCulloch, R. (2005). *Bayesian statistics and marketing*, Wiley.

Tanner, M. A. & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation, *Journal of American Statistical Association*, **82**, 528-540.

Verhoef, P. C. (2003). Understanding the effect of customer relationship management efforts on customer retention and customer share development, *Journal of Marketing*, **67**, 30-45.

Wedel, M. & Kamakura W. (1999). *Market segmentation: Conceptual and methodological foundations: 2nd Ed.*, Springer.

Willemain, T. R., Smart, C. N. & Schwarz, H. F. (2004). *A new approach to forecasting intermittent demand for service parts inventories*, *International Journal of Forecasting*, **20**, 375-387.