Technological Forecasting & Social Change xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach

Hakyeon Lee [a], Sang Gook Kim [b], Hyun-woo Park [b], Pilsung Kang [a,*]

[a] Department of Industrial and Systems Engineering, Seoul National University of Science and Technology, 172 Gongreung 2-dong, Nowon-gu, Seoul 139-746, South Korea
[b] Korea Institute of Science and Technology Information (KISTI), 66 Hoegi-ro, Dongdaemun-gu, Seoul 130-741, South Korea

ARTICLE INFO

Article history:
Received 23 April 2013
Received in revised form 12 August 2013
Accepted 17 August 2013
Available online xxxx

Keywords:
Pre-launch forecasting
Bass model
Multivariate linear regression
Machine learning
Ensemble

ABSTRACT

This study proposes a novel approach to the pre-launch forecasting of new product demand based on the Bass model and statistical and machine learning algorithms. The Bass model is used to explain the diffusion process of products while statistical and machine learning algorithms are employed to predict two Bass model parameters prior to launch. Initially, two types of databases (DBs) are constructed: a product attribute DB and a product diffusion DB. Taking the former as inputs and the latter as outputs, single prediction models are developed using six regression algorithms, on the basis of which an ensemble prediction model is constructed in order to enhance predictive power. The experimental validation shows that most single prediction models outperform the conventional analogical method and that the ensemble model improves prediction accuracy further. Based on the developed models, an illustrative example of 3D TV is provided.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Increasing market uncertainty is making forecasting new product demand more difficult than ever, while shorter product lifecycles are forcing managers to produce new product demand forecasts more frequently. Correct forecasts are a key determinant of the success of new products, but accurate forecasting can be challenging. Forecasting demand for a mature-stage product is unproblematic since enough historical sales data are available. However, forecasting the future sales of a new product that has a short history or no history at all is complicated. Indeed, as Bass et al. [1] note, "the most important forecast is the forecast prior to launch," as it drives important pre-launch decisions such as capital equipment purchases and capacity planning [2]. Overestimating demand can result in excess inventories, while underestimation may incur significant opportunity costs and reduce market share.

Despite its importance, however, little progress has been made in pre-launch forecasting in the literature. Product demand forecasting has mainly been examined in the context of innovation diffusion. The diffusion of an innovation is "the process by which an innovation is communicated through certain channels over time among the members of a social system" [3]. Since its introduction to marketing studies in the 1960s, a variety of models have been developed to empirically model the diffusion of innovations [4]. The main thread of the diffusion models has been based on the pioneering work of Bass [5]. The assumption underlying the Bass model is that an innovation is spread through two types of communication channels: mass media (external influence) and word-of-mouth (internal influence). Over the past 40 years, the Bass model has thus enjoyed a number of applications because it has a relatively high explanatory power despite its simple structure [6].

Pre-launch forecasting also has been investigated mainly based on the Bass model. Previous studies of pre-launch forecasting with the Bass model have struggled against how to estimate the model parameters of a new product when little or no historical sales data are available. Several methods

* Corresponding author. Tel.: +82 2 970 7279; fax: +82 2 974 2849.
E-mail address: pskang@seoultech.ac.kr (P. Kang).

0040-1625/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.techfore.2013.08.020

Please cite this article as: H. Lee, et al., Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach, Technol. Forecast. Soc. Change (2013), http://dx.doi.org/10.1016/j.techfore.2013.08.020

have been developed, which can be categorized into three types: Bayesian approaches, subjective approaches, and analogical approaches. The Bayesian approach starts with pre-launch forecasts and updates them as additional data become available. Various methods for updating the parameter estimates or the forecasts have been proposed [7–10]. However, this approach centers on how to update forecasts after launch and still calls for initial pre-launch forecasts to be updated as data become available. The subjective approach produces parameter estimates through an algebraic estimation procedure on the basis of managerial judgments of tangible information such as the time and level of peak sales [11] and the sum of the coefficients of the external and internal influences [12]. The drawback of the subjective approach is that obtaining accurate judgments is as difficult as estimating accurate parameters [13]. Finally, the analogical approach, called "guessing by analogy," has prevailed in the literature. It assumes that a new product will have a diffusion pattern similar to those of its analogous products over time [1]. Under this approach, the parameter estimates of a new product are obtained by taking a weighted sum of the parameters of analogous products, with the weights derived by establishing similarities between the new product and several analogous products [14].

Although the analogical approach has been applied widely [1,15–19], it has two main limitations. First, there are no clear guidelines for how to select benchmarks even though the estimated parameters are highly dependent on the analogous products under consideration. Second, the similarities are established by expert judgments that are naturally subjective in nature. A promising solution to these problems is using the historical empirical relationship between the parameters and attributes of analogous products. Rogers [3] emphasizes that the attributes of an innovation are important variables in explaining its rate of adoption. Once this relationship is identified, the parameters of a new product can be estimated by knowing its characteristics [20]. Although identifying the relationship between diffusion parameters and product attributes can serve a reliable basis for pre-launch forecasting of a new product, however, little research has been carried out thus far in this direction and no systematic approach has yet been developed.

The tenet of this study is that a statistical and machine learning-based approach can overcome the limitations of the conventional analogical approach to pre-launch forecasting. The goal of statistical and machine learning is to discover intrinsic, sometimes unanticipated, relationships between variables with the help of high computational power [21,22]. A typical procedure of an inductive statistical and machine learning approach is as follows. The first step is to set up the model structure by defining a learning task, configuring input–output variables, seeking appropriate algorithms, and selecting proper performance criteria. The second step is to collect sufficient real-world examples, which are then divided into training and test data sets. In the third step, the employed learning algorithms are optimized on the basis of the training data set. Finally, the best model is identified on the basis of the test data set, using the predetermined performance criteria. In this study, the learning task is defined as predicting the parameters of the Bass model prior to launch, while input and output variables are configured as product attributes and diffusion characteristics, respectively. Various regression algorithms, such as multilayer linear regression, support vector regression, and Gaussian process regression, as well as an ensemble model of these, are used in building the prediction models. Finally, their performances are evaluated using the mean absolute error (MAE) and the root mean squared error (RMSE).

Pre-launch product demand forecasting on the basis of statistical and machine learning algorithms has several advantages over conventional analogical methods. First, a reliable relationship between the attributes and diffusion characteristics of existing products can be found, which in turn enables new product demand forecasts to be based solely upon these attributes without any human manipulation. In other words, analogous products are automatically selected and their contributions to forecasting systematically determined by the prediction model; although such selection and determination processes are not easily understood by humans, they are mathematically sound and analytically tractable. Therefore, forecasting is no longer dependent on the subjective judgments of human experts, but becomes an objective outcome obtained by the combination of learning algorithms and product data. Second, since statistical and machine learning algorithms are designed for interpolation as well as extrapolation, forecasting accuracy can be improved. The parameter values predicted by conventional analogical methods are bounded by the current maximum and minimum estimates of reference products. If a new type of diffusion style occurred, whose parameter values were far from the current boundary, conventional analogical methods would not properly reflect this eventuality. Statistical and machine learning-based approaches, by contrast, would scent the change from the inside of the product and digest it into prediction. In light of the foregoing, this study proposes a new approach to the pre-launch forecasting of new product demand, which utilizes the Bass diffusion model and statistical and machine learning-based regression algorithms. In addition, we also boost prediction accuracy by constructing an ensemble of individual prediction models.

The remainder of this paper is organized as follows. Section 2 reviews previous studies of pre-launch forecasting with the Bass model. Section 3 demonstrates the proposed framework including the product and diffusion database (DB) design, single prediction model development, and ensemble model construction. Section 4 validates the single and ensemble prediction models and provides an illustrative case study on the basis of the best single and ensemble models. Finally, the conclusion and limitations of this paper are presented alongside future research directions in Section 5.

2. Bass model and pre-launch forecasting

The Bass model assumes that a technological innovation is spread by two types of influences: externally by the mass media and internally by word-of-mouth. It can be derived from a hazard function that represents the probability that an adoption occurs at time t given that it has not yet occurred:

$$h(t) = \frac{f(t)}{1-F(t)} = p + qF(t) \quad (1)$$

where $f(t)$ is the density function of time to adoption, $F(t)$ is the cumulative proportion of adopters at time t, and p and q

Please cite this article as: H. Lee, et al., Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach, Technol. Forecast. Soc. Change (2013), http://dx.doi.org/10.1016/j.techfore.2013.08.020

are the coefficients of external and internal influences, respectively. Eq. (1) can thus be rewritten as

$$f(t) = \frac{dF(t)}{dt} = [p + qF(t)][1 - F(t)]. \quad (2)$$

Because the number of adopters at time t, $n(t)$, can be obtained by multiplying $f(t)$ by the potential market size, m, rearranging Eq. (2) yields

$$n(t) = \frac{dN(t)}{dt} = \left[p + \frac{q}{m}N(t)\right][m - N(t)]. \quad (3)$$

The first term, $p[m - N(t)]$, represents adoptions owing to external influences, while the second term, $\frac{q}{m}N(t) \times [m - N(t)]$, stands for adoptions owing to internal influences. In a pure innovation scenario where only external influences exist (i.e., $q = 0$), the Bass model reduces to an exponential function. If p is zero, however, it is equivalent to a logistic model, assumed to be driven by only imitative processes, namely a pure imitation situation. Solving Eq. (2) produces the following closed form solution:

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}}. \quad (4)$$

The cumulative number of adopters at time t, $N(t)$, can then be written as

$$N(t) = M\left[\frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}}\right]. \quad (5)$$

The three parameters of the Bass model (m, p, and q) can usually be estimated through conventional estimation procedures such as ordinary least square (OLS) [5], maximum likelihood estimation (MLE) [23], and nonlinear least squares (NLS) [24]. However, these methods can be applied only when enough sales data are available. Previous studies demonstrate that stable and robust parameter estimates for the Bass model can be obtained only if the data under consideration include the peak of the noncumulative adoption curve [24,25]. When insufficient data are available, such as early in the product life cycle or prior to launch, these conventional estimation methods cannot produce reliable parameter estimates. Nevertheless, as discussed earlier, sales forecasting is even more crucial when little or no data are available. Estimates of the diffusion parameters early in the diffusion process or prior to launch are extremely valuable for managerial decision-making such as capital equipment purchases, production planning, and marketing strategy [26].

Time-varying estimation procedures that rely on Bayesian updating have been introduced by several researchers to cope with estimating when little or no data are available. These procedures start with forecasts prior to launch and update those forecasts whenever an additional record becomes available. Lilien et al. [7] propose a Bayesian approach whereby new product sales prior to market entry are predicted by considering the forecasts of a previously introduced similar product and then updated once data are available using Bayesian regression. Rao and Yamada [8] expand the work of Lilien et al. [7] to show that incorporating priors and the use of perceptual data can improve forecasting performance. Lenk and Rao [9] further suggest the Hierarchical Bayes procedure that explicitly considers between-product and within-product variations in establishing initial estimates for the new product. An adaptive procedure combined with Bayesian updating, called the augmented Kalman filter with continuous state and discrete observations, is also proposed by Xie et al. [10]. While the aforementioned approaches employ prior information derived from the diffusion of previously introduced products deemed most similar, Sultan et al. [27] develop initial parameter estimates by conducting a meta-analysis of 213 sets of parameters from 15 articles. Lee et al. [28] utilize consumer reservation price data in order to construct prior distributions. Regardless of the vehicle used for obtaining initial forecasts, however, time-varying estimation approaches mainly focus on updating after launch, which still requires initial pre-launch forecasts to be updated as data become available.

Several approaches to producing pre-launch forecasts have been proposed in previous studies and can be classified into two types: subjective approaches and analogical approaches. In terms of subjective approaches, Mahajan and Sharma [11] propose an algebraic estimation procedure that requires three pieces of information: 1) market potential (m), 2) time of peak sales (t^*), and 3) level of peak sales (n^*). Once these values have been obtained, p and q can be inferred. However, estimating both the time and level of peak sales is another difficult task. In addition, as Bass [29] notes, these are the key outputs intended to be forecast from the observed data using the diffusion model; therefore, if one could guess such items, there would be no need to estimate the diffusion curve. A similar procedure suggested by Lawrence and Lawton [12] also involves obtaining three pieces of managerial information: 1) market potential (m), 2) number of adoptions in the first period ($n(1)$), and 3) an estimate of the sum of the coefficients p and q ($p + q$). A similar problem occurs here; estimating $p + q$ is also difficult. Although general guidelines suggest a value of 0.5 for consumer goods and 0.66 for commercial goods, such generalizations fail to mirror the idiosyncratic characteristics of particular products. In fact, Lawrence and Lawton [12] suggest that prior sales histories of similar products may produce better prior parameter estimates.

In this spirit, the analogical approach has been widely employed for pre-launch forecasting. This approach assumes that a new product will behave as analogous products do. As mentioned before, several previous studies using the Bayesian approach utilize the diffusion information of exsiting similar products as a priori [7–10]. A more systematic approach on the basis of consumer choice theory is proposed by Thomas [14]. In this approach, the parameters of a new product are estimated by taking a weighted sum of the parameters of analogous products. These weights can be obtained by establishing the similarities between the new product and several analogous products in terms of five dimensions: environmental situation, market structure, buyer behavior, marketing mix strategy, and innovation characteristics. This "guessing by analogy" approach has often been adopted for the pre-launch forecasting of various types of products and services [1,16–19]. However, there are no clear standards for the selection of benchmark products. Analogous products are often chosen simply because

Please cite this article as: H. Lee, et al., Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach, Technol. Forecast. Soc. Change (2013), http://dx.doi.org/10.1016/j.techfore.2013.08.020

they are recent or easily recalled [30]. Although Ilonen et al. [31] employ the self-organizing map for automatically identifying analogies, their approach is centered on selecting analogous countries rather than analogous products. Bayus [15] suggests a grouping procedure based on hierarchical clustering with factor analysis to generate priors on the basis of various products, but the parameter estimates obtained in this way still highly depend on what analogous products are selected. Moreover, similarities are usually identified by expert judgments or consumer opinions, which are inherently subjective.

Using a historical relationship between the parameters and attributes of analogous products can be a promising solution to this problem. Nonetheless, few studies have been conducted in this direction. Sultan et al. [27] propose a meta-analysis model utilizing ANOVAs with four types of attributes. However, these attributes are related to the characteristics of the research itself such as model specification, estimation method, and data reuse, while product-related attributes only include type of innovation and geographic effect. Similarly, Gatignon et al. [32] develop a cross-national econometric model of innovation diffusion, but its included variables are associated with country-level patterns of social communication such as cosmopolitanism, mobility, and sex roles; again, no product-related attributes are considered. Finally, although Srivastava et al. [33] suggest a multi-attribute model for forecasting the adoption of investment alternatives that includes five attributes of investment products. Due to the high correlation between these attributes, however, only two attributes, information costs and likelihood of loss of principal, are employed to forecast the acceptance of a potential investment alternative. Further, the attributes considered in this study are too industry-specific to be generalized to other products and services. In summary, although establishing the historical relationship between diffusion parameters and product attributes can be considered to be promising for pre-launch new product forecasting, no systematic approach has thus far been proposed in the literature.

3. Methodology

3.1. Overall procedure

The present study proposes a statistical and machine learning approach to the pre-launch forecasting of new product demand on the basis of the Bass model. Fig. 1 depicts the overall procedure of the proposed approach. Two types of DBs are required to develop the prediction models: a product diffusion DB and a product attribute DB. The product diffusion DB includes the Bass diffusion parameters of existing products obtained by applying the NLS technique to historical sales data on each product. To construct the product attribute DB, various attributes that can explain the diffusion characteristics of a product need to be selected and defined. The values of the attributes of each product are measured through expert judgments. By taking the product attribute DB as inputs and the product diffusion DB as targets, single prediction models are developed to identify the relationship between diffusion parameters and product attributes. Six statistical and machine learning-based regression algorithms are utilized. To improve prediction performance, ensemble prediction models combining these single models are also constructed. Comparing prediction performance among the developed models produces the best-fit models that will be utilized for the case study. The performance comparison with the results of the simple analogical approach is also provided to uphold the validity of the proposed approach. Finally, the working of the developed models is provided with the help of a simple illustrative example of 3D TV.

3.2. Data

As many products as possible for which previous sales data are available should be collated in the DBs in order to improve the reliability of the prediction models. One of the major data sources was the CE Historical Data provided by the Customer Electronics Association. This data set contains US sales data on more than 60 categories of electronic products from their market introductions to the present. In total, 21 US products whose historical sales data were presented in recent papers and reports were included in the DBs. In addition, Korean sales data on over 40 products were gathered from relevant associations. Since small numbers of observations may produce unreliable estimates, all products with fewer than 12 sales records were then excluded from the data set, which resulted in 87 products remaining.

In the next step, the diffusion parameters of each of these 87 products were estimated by fitting their historical data to the Bass model in order to construct the product diffusion DB. Although the Bass model is specified by three parameters, the potential market size (m), the coefficient of external influence (p), and the coefficient of internal influence (q), the product diffusion DB only includes p and q. Most previous studies of pre-launch forecasting center on estimating the two parameters of communication effects, p and q, while separately estimating m from market research [1,9,10,14,15,18]. Similarly, Lawrence and Lawton [12] and Mahajan and Sharma [11] also utilize the potential market size as inputs for the subjective algebraic procedures rather than as outputs of the procedure, because the market size is likely to be affected by marketing efforts and various environmental factors rather than by product characteristics [34]. Consequently, our prediction model also only includes estimates of p and q. These two parameters of the 87 products were estimated using the NLS procedure, with average values of p and q found to be 0.0087 and 0.3273, respectively. The parameter estimates included in the product diffusion DB were then employed as the target variables in the prediction models.

The next step was to construct the product attribute DB for the 87 products. Firstly, product attribute variables affecting the diffusion patterns need to be figured out. An extensive literature review of the drivers and determinants of innovation diffusion and a series of discussions with senior marketing managers in practice led us to derive 17 attribute variables for constructing the product attribute DB. Table 1 presents these 17 variables with their abbreviations and measurement scales. These variables can be grouped into four categories: industry, market, technology, and use. Four variables (IC, TCS, TIE, and DN) were valued on nominal scales, while the other variables were measured on a five-point Likert scale from very low (1) to very high (5). The values of the attribute variables for each of the 87 products were then measured by industry experts including marketing managers and engineers and used as input variables in the prediction models.

Please cite this article as: H. Lee, et al., Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach, Technol. Forecast. Soc. Change (2013), http://dx.doi.org/10.1016/j.techfore.2013.08.020

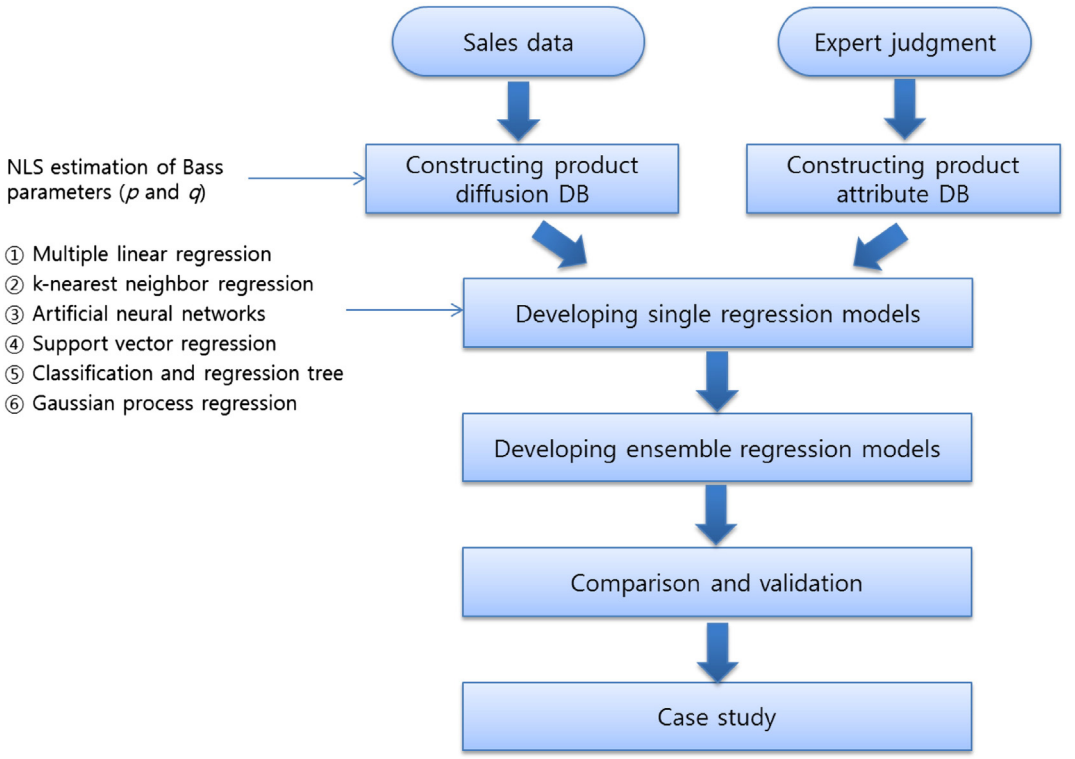

Fig. 1. Overall procedure.

3.3. Experimental design

3.3.1. Data preprocessing and variable selection

As explained earlier, some products were removed owing to the limited number of historical sales records. Another issue in the modeling process was the existence of outliers whose estimates were noticeably different. Fig. 2 shows the estimated Bass model parameters of the 87 products. This figure clearly shows that certain products have significantly large p or q values compared with others. Since such outliers would degrade prediction accuracy, seven products whose estimates were beyond two standard deviations from the mean were also removed. Consequently, 80 products were finally used to build the prediction models, with the average estimates of p and q decreasing to 0.0063 and 0.2783, respectively. The parameter estimates of each of these 80 products are presented in Appendix A.

Because most of the regression algorithms employed in this study can only handle numerical variables, the four nominal variables (IC, TCS, TIE, and DN) were transformed into binary variables using the 1-of-C coding method. In this method, C binary dummy variables are created for a nominal variable with C categories; for each dummy variable, 1 is assigned if the original value falls in the same category, with 0 otherwise (see Fig. 3). Once the variable transformation had been completed, the number of variables increased from 17 to 24.

Some of these 24 candidate input variables were dispensable because they had little effect on the prediction. Thus, using a stepwise linear regression, we selected only the crucial variables before training the prediction models. This stepwise selection process began with the single most relevant input variable, and the following two procedures were then conducted alternately until all significant variables had been identified: (1) among the candidate variables, the one that most improves accuracy is added to the selected variable set, from which (2) the one that is most irrelevant to improve accuracy is removed.

3.3.2. Regression algorithms

Six statistical and machine learning-based regression algorithms were employed to build single prediction models, namely multivariate linear regression (MLR), k-nearest neighbor regression (k-NN), artificial neural network (ANN), support vector regression (SVR), classification and regression tree (CART), and Gaussian process regression (GPR).

MLR [35] fits the functional relationship between multiple input variables and the target variable of the given data in the form of a linear equation. Let y_i denote the target value (p or q) of the i^{th} product, while x_{ij} denotes the j^{th} input variable of the i^{th} product. Then, the MLR equation of d predictors with n training instances can be written as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_d x_{id}, \quad \text{for} \quad i = 1, 2, \cdots, n. \tag{6}$$

Table 1
Product attribute variables.

Dimension	Variable	Abbreviation	Scale
Industry	Industry classification	IC	Nominal (E: electrical, G: electronic, M: mechanical)
	Types of competitive structure	TCS	Nominal (P: perfect competition, M: monopoly, O: oligopoly, N: monopolistic competition)
	Necessity of complementary goods	NCG	Five-point Likert
	Availability of substitute goods	ASG	Five-point Likert
Market	Level of price	LP	Five-point Likert
	Types of income elasticity	TIE	Nominal (N: Necessities, L: Luxuries)
	Number of potential customers	NPC	Five-point Likert
	Number of suppliers	NS	Five-point Likert
	Diversity of distribution channels	DDC	Five-point Likert
Technology	Degree of newness	DN	Nominal (R: radical, I: incremental)
	Rate of technological change	RTC	Five-point Likert
	Easiness of imitation	EI	Five-point Likert
	Degree of functional variety	DFV	Five-point Likert
Use	Necessity of learning	NL	Five-point Likert
	Frequency of use	FU	Five-point Likert
	Duration of use	DU	Five-point Likert
	Necessity of repurchase	NR	Five-point Likert

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta},\quad \mathbf{y}=\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},\qquad \mathbf{X}=\begin{pmatrix} 1 & x_{11} & \cdots & x_{1d} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nd} \end{pmatrix},\quad \boldsymbol{\beta}=\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_d \end{pmatrix}. \tag{7}$$

The regression coefficients $\boldsymbol{\beta}$ can be obtained by minimizing the squared residual error between the target ($\mathbf{y}$) and predictions ($\hat{\mathbf{y}}$), as shown in Eq. (8), using the ordinary least squares method:

$$E=\frac{1}{2}\sum_{i=1}^{n} e_i^2=\frac{1}{2}(\mathbf{y}-\hat{\mathbf{y}})^{\mathrm{T}}(\mathbf{y}-\hat{\mathbf{y}})=\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}).$$
$$\frac{\partial E}{\partial \boldsymbol{\beta}}=\mathbf{X}^{\mathrm{T}}\mathbf{y}-\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta}=0,\quad \boldsymbol{\beta}=\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}. \tag{8}$$

k-NN [36] is the most popular case-based reasoning algorithm. Since it does not require a separate training procedure, it has been employed in various domains where rapid and frequent model updates are required [37–39]. k-NN predicts a new instance on the basis of the similarity to its neighbors. Once a test instance $\mathbf{x}_t$ is given, k-NN first searches the k most similar instances in the reference data set using a certain distance metric and allocates weights to them under the principle that the greater the similarity, the greater is the weight. Then, the prediction is made according to the weighted average of the target values and assigned weights of the selected neighbors as follows:

$$\hat{y}_t=\sum_{j\in \mathrm{NN}(\mathbf{x}_t)} w_j y_j, \tag{9}$$

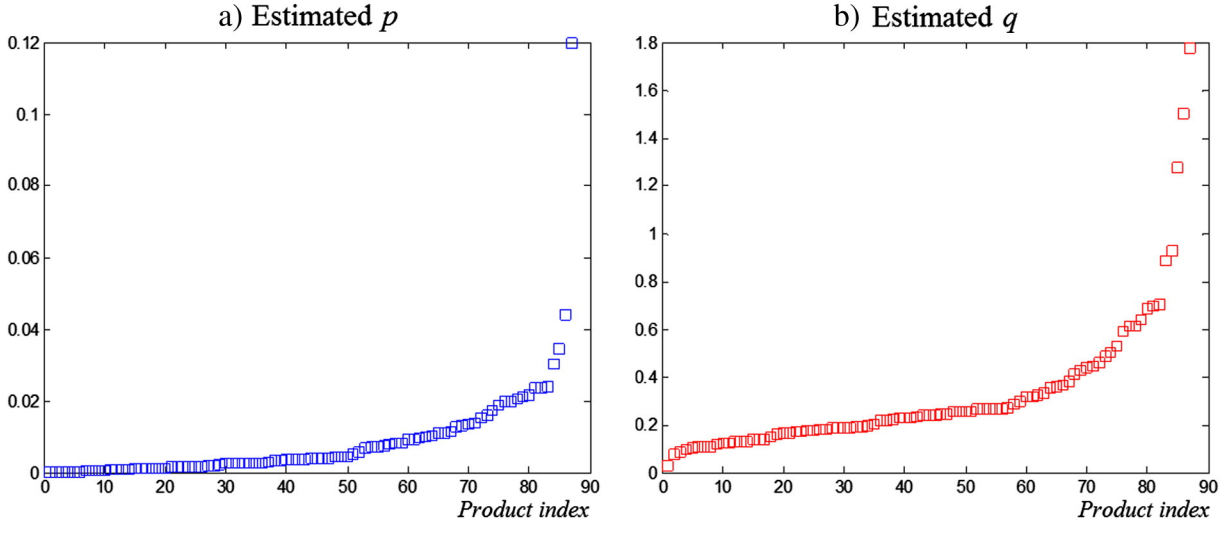

Fig. 2. The estimated Bass model parameters (x-axis: product index, y-axis: estimated value).

Please cite this article as: H. Lee, et al., Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach, Technol. Forecast. Soc. Change (2013), http://dx.doi.org/10.1016/j.techfore.2013.08.020

Product	Original variable
	IC
A	G
B	E
C	M

Dummy variable after 1-of-C coding		
IC_G	IC_E	IC_M
1	0	0
0	1	0
0	0	1

Fig. 3. An example of 1-of-C coding.

where $NN(\mathbf{x}_t)$ and w_j denote the index set of the k-nearest neighbors of $\mathbf{x}_t$ and the assigned weight to the jth nearest neighbor, respectively. In the k-NN regression, two user-specific parameters must be declared prior to the prediction: the number of nearest neighbors (k) and weight allocation method.

ANN [21] is one of the most widely used nonparametric regression algorithms in many application domains owing to its ability to capture nonlinear relationships between the input and output variables [40–42]. A three-layer feed-forward neural network is used in our experiments. In ANN, the target is expressed as a combination of input values, activation functions, and weights as follows:

$$y_i = \sum_{q=1}^{h} w_q^{(2)} g\left(\sum_{p=1}^{d} w_{qp}^{(1)} x_{ip}\right), \quad for \quad i = 1, 2, \cdots, n, \tag{10}$$

where $w_q^{(2)}$, $w_{qp}^{(1)}$, and $g(\cdot)$ are the weight connected from the qth hidden node to the output node, the weight connected from the p^{th} input node to the q^{th} hidden node, and the activation function, respectively. Training ANN is equivalent to finding the optimal weights that minimize the objective loss function, as shown in Eq. (11):

$$\text{Loss function } L = \frac{1}{2}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2. \tag{11}$$

Since there exists no explicit solution for ANN, combinatorial optimization algorithms such as the gradient descent algorithm [43], Newton's method [44], and Levenberg–Marquardt algorithm [45] are used to optimize the weights.

Contrary to ANN, which is based on the empirical risk minimization (ERM), SVR [46] is based on the structural risk minimization (SRM) principle. SVR fits the regression equation $\hat{y} = \mathbf{w}^T\mathbf{b} + b$ with the constraints of including training instances as many as possible in the ε -tube as follows:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n}(\zeta_i + \zeta_i^*), \\ s.t \quad & y_i - \mathbf{w}^T\mathbf{x}_i - b \le \varepsilon + \zeta_i, \quad \zeta_i \ge 0, \\ & \mathbf{w}^T\mathbf{x}_i + b - y_i \le \varepsilon + \zeta_i^*, \quad \zeta_i^* \ge 0. \end{aligned} \tag{12}$$

C in the Eq. (12) controls the trade-off between the flatness and the error of the training samples outside the ε-tube. Lagrangian formulation is derived by eliminating the constraints using slack variables as follows:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}(\zeta_i + \zeta_i^*) - \sum_{i=1}^{n}\alpha_i\left(\varepsilon + \zeta_i - y_i + \mathbf{w}^T\mathbf{x}_i + b\right) \\ & - \sum_{i=1}^{n}\eta_i\zeta_i - \sum_{i=1}^{n}\alpha_i^*\left(\varepsilon + \zeta_i + y_i - \mathbf{w}^T\mathbf{x}_i - b\right) - \sum_{i=1}^{n}\eta_i\zeta_i^*, \\ s.t. \quad & \alpha_i, \alpha_i^*, \zeta_i, \zeta_i^* \ge 0. \end{aligned} \tag{13}$$

By taking the derivatives of the primal variables, the optimal conditions for the above Lagrangian are obtained; in turn, Wolfe's dual problem is derived by replacing the conditions in the primal problem:

$$\begin{aligned} \max \quad & -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_i - \alpha_i^*)\mathbf{x}_i^T\mathbf{x}_j - \varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) \\ s.t. \quad & \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0, \quad 0 \le \alpha_i, \alpha_i^* \le 1. \end{aligned} \tag{14}$$

SVR enables nonlinear fitting using a mapping function $\phi(\mathbf{x})$ that transforms data from a low dimensional input space to a high dimensional feature space. Since only inner products between input vectors are required during optimization, a kernel trick $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)\cdot\phi(\mathbf{x}_j)$ is employed to compute the inner product in the feature space without an explicit mapping function.

CART [47] is another approach to nonlinear regression in that the entire input space is recursively partitioned into small chunks until the members can be fit by a simple model. Beginning with all instances in a single parent node, the splitting criterion, which maximizes the information gain usually measured by the decrease in the sum of squared errors between before and after the splitting, is determined. Then, the training instances are assigned to one of the child nodes in accordance with their variable status and the splitting criterion. Again, each child node becomes a new parent node and the same partitioning procedure is recursively conducted until the full tree is constructed, i.e., there is no more information gain after splitting.

Another key procedure of training CART is pruning. Since the full tree constructed by CART divides the input space into such a large number of tiny segments, it is usually exposed to an over-fitting risk; it would degrade the generalization

ability if the model memorized all the characteristics of data, some of which are even unnecessary, e.g., noises or outliers. The role of pruning is to merge adjacent leaf nodes if the prediction accuracy is improved or at least unchanged on the validation data set that is not used in the tree construction. Once pruning has been completed, CART predicts a new instance by averaging the target values in the leaf node to which the new instance belongs.

GPR [48] begins with the Bayesian approach to MLR and extends the expressiveness by adopting kernel tricks. In GPR, the target y is expressed as a linear combination of the inputs with a Gaussian noise as follows:

$$y = f(\mathbf{x}) + \varepsilon, \quad f(\mathbf{x}) = \mathbf{x}^T\mathbf{w}, \tag{15}$$

assuming that the noise follows an independent, identically distributed (i.i.d.) Gaussian distribution with zero mean and variance σ^2,

$$\varepsilon \sim N\left(0, \sigma^2\right). \tag{16}$$

The likelihood, which is the probability density of the given data and parameters, can be directly obtained as

$$p(\mathbf{y}|\mathbf{X},\mathbf{w}) = \prod_{i=1}^{n} p(y_i|\mathbf{x}_i,\mathbf{w}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\left(y_i - \mathbf{x}_i^T\mathbf{w}\right)^2}{2\sigma^2}\right) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}|\mathbf{y}-\mathbf{X}^T\mathbf{w}|^2\right) = \mathcal{N}\left(\mathbf{X}^T\mathbf{w}, \sigma^2 I\right) \tag{17}$$

As a prior distribution over the weights, a zero mean Gaussian with covariance matrix $\mathbf{\Sigma}_p$ is generally used

$$\mathbf{w} \sim N\left(0, \Sigma_p\right). \tag{18}$$

Inference in GPR is based on the posterior distribution over the weights by applying the Bayes' theorem,

$$p(\mathbf{w}|\mathbf{y},\mathbf{X}) = \frac{p(\mathbf{y}|\mathbf{X},\mathbf{w})p(\mathbf{w})}{p(\mathbf{y}|\mathbf{X})} \propto p(\mathbf{y}|\mathbf{X},\mathbf{w})p(\mathbf{w}) \sim \mathcal{N}\left(\frac{1}{\sigma^2}\mathbf{A}^{-1}\mathbf{X}\mathbf{y}, \mathbf{A}^{-1}\right), \tag{19}$$

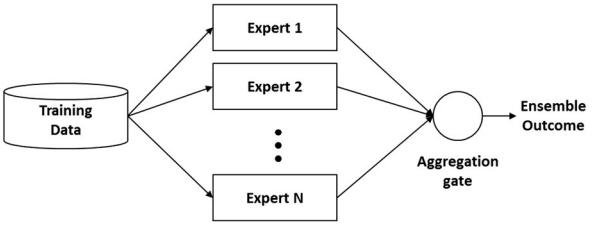

Fig. 4. The structure of an ensemble of individual forecasting models.

where $\mathbf{A} = \left(\frac{1}{\sigma^2}\mathbf{X}\mathbf{X}^T + \mathbf{\Sigma}_p^{-1}\right)$. The prediction distribution of f_t at a test instance $\mathbf{x}_t$ is then obtained by averaging the output of all possible linear models with regard to the Gaussian posterior

$$p(f_t|\mathbf{x}_t,\mathbf{X},\mathbf{y}) = \int f(\mathbf{x}_t|\mathbf{w})P(\mathbf{w}|\mathbf{X},\mathbf{y})d\mathbf{w} = \mathcal{N}\left(\frac{1}{\sigma^2}\mathbf{x}_t^T\mathbf{A}^{-1}\mathbf{X}\mathbf{y},\ \mathbf{x}_t^T\mathbf{A}^{-1}\mathbf{x}_t\right). \tag{20}$$

As in SVR, GPR can fit a nonlinear relationship by introducing a mapping function $\phi(\mathbf{x})$ to project the data from a low dimensional space to a higher dimensional feature space and using kernel tricks to compute inner products in the feature space without an explicit form of $\phi(\mathbf{x})$.

3.3.3. Ensemble model

After training the single prediction models, ensemble prediction models [21,49] are constructed in order to enhance the predictive power. Fig. 4 shows a general structure of an ensemble model. A number of variations are possible according to the diversity of algorithms or parameters; experts can consist of different learning algorithms, an identical learning algorithm with different parameters, or a combination of the both. The prediction outcome of an ensemble is formulated by aggregating the output of every expert. In our experiments, the best ensemble model was identified among all possible combinations of regression algorithms and subsequently used in a comparative validation analysis as well as in the case study.

3.3.4. Validation method and performance measures

As a benchmark method to verify our proposed prediction models, a conventional analogical prediction model was also constructed. In the analogical model, the two parameters of the Bass model can be predicted as follows:

$$p = \sum_{j=1}^{m}\sum_{i=1}^{n} w_i x_{ij} p_j \tag{21}$$

$$q = \sum_{j=1}^{m}\sum_{i=1}^{n} w_i x_{ij} q_j \tag{22}$$

where m is the number of analogous products, n is the number of dimensions for evaluating similarities, w_i is the importance weight of dimension i, and x_{ij} is the similarity weight between the product being evaluated and product j in terms of dimension i. Further, p_j and q_j are the coefficients of the external and internal influences of product j, respectively. We employed the same four dimensions as those used for defining the product attributes: industry, market, technology, and use. The importance weights of each dimension were first determined using the pairwise comparison method [50]. For each of the 80 products, three to five similar products were selected, and their similarity weights were assigned by the experts so that the sum of the weights of the analogous products for each dimension equaled 1.

Please cite this article as: H. Lee, et al., Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach, Technol. Forecast. Soc. Change (2013), http://dx.doi.org/10.1016/j.techfore.2013.08.020

The leave-one-out validation method was adopted to verify the prediction models. If there is a large amount of data, k-fold cross-validation with $k = 5$ or 10 is generally used; the entire data set is divided into k subgroups, and every subgroup is set aside in turn for validation and the model is trained on the basis of the remaining $(k - 1)$ subgroups. The prediction outcomes of the subgroups are then integrated to measure the overall performance of the model. If there is a relatively small amount of data, by contrast, the leave-one-out validation is more appropriate, as it increases k up to the total number of instances in order to secure as many training instances as possible. Because using 80 products is considered to be insufficient, the leave-one-out validation is a better choice in our experiment.

The prediction models were then evaluated in terms of the mean absolute error (MAE) and the root mean squared error (RMSE),

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \quad \mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{23}$$

where y is the target value (estimated p or q of the Bass diffusion model) and $\hat{y}$ is the predicted value derived by the model. The lower the performance measures, the better the prediction model.

4. Results

4.1. Variable selection

The input variables selected by the stepwise linear regression are presented in Table 2. Among the 24 candidate input variables, only seven were identified as being crucial for the prediction of p and q at a significance level of 0.05. Three input variables (DDC, DN_R, and EI) are positively related to p, while the other four (NCG, ASG, LP, and NR) are negatively related. By contrast, four input variables (IC_O, NCG, DDC, and DN_I) are positively related to q, while the other three (RTC, EI, and NL) are negatively related. It was also found that three input variables (NCG, DDC, and EI) are necessary for predicting both p and q. Note that only DDC has positive coefficients for both targets; more diversified distribution channels thus give rise to greater internal and external influence on product demand. However, NCG and EI result in the reversed sign of coefficients if the target changes; a higher necessity of complementary goods therefore interrupts the external but promotes the internal influence, while a higher ease of imitation promotes the external but interrupts the internal influence. In addition, the demand creation led by the mass media is encouraged by a higher DN_R but discouraged by a higher ASG, LP, or NR, while the demand creation led by word-of-mouth is encouraged by a higher IC_O or DN_I, but discouraged by a higher RTC or NL.

Table 3
The prediction performances of each regression algorithm for p and q (The bold face numbers denote the lowest error among the algorithms).

Algorithm	p		q	
	MAE	RMSE	MAE	RMSE
MLR	**0.0026**	**0.0031**	**0.0688**	**0.0898**
k-NN	0.0029	0.0041	0.0709	0.1018
ANN	0.0036	0.0045	0.1428	0.1813
SVR	0.0027	0.0036	0.0961	0.1301
CART	0.0054	0.0079	0.1230	0.1723
GPR	**0.0026**	0.0034	0.0790	0.1053
Analogy	0.0053	0.0073	0.1337	0.1907

4.2. Forecasting performances: single models

The performances of each regression algorithm in predicting p and q are summarized in Table 3. Note that because there are few differences between these performance measures, we focus on the MAE in the presented analysis. MLR is found to be the most outstanding of the single prediction models, followed by SVR and GPR in the case of p and by k-NN and GPR in the case of q. Its MAE for p, 0.0026, is the lowest among all algorithms and it is less than 42% of the average of the estimates. For the prediction of q, MLR is even more accurate (MAE = 0.0688) and no greater than 25% of the average of the estimates. It is worth noting that ANN and CART provide inferior prediction performances; the MAE of CART for p and MAE of ANN for q are almost twice as large as those of MLR. In contrast to SVR and GPR, however, they are both based on the empirical risk minimization scheme, under which principle good

Table 2
The selected input variables for p and q by stepwise linear regression ($\alpha = 0.05$, the variables with bold face are selected for both p and q).

p				q			
Variable	Coefficient	t-statistic	p-value	Variable	Coefficient	t-statistic	p-value
NCG	−0.000786	−2.20	0.031	NCG	0.036543	3.82	<0.001
DDC	0.001988	5.40	<0.001	DDC	0.037047	3.09	0.003
EI	0.000911	2.71	0.008	EI	−0.04629	−5.09	<0.001
ASG	−0.000992	−2.79	0.007	IC_O	0.071268	2.85	0.006
LP	−0.000934	−2.51	0.014	DN_I	0.077644	3.19	0.002
DN_R	0.002237	2.65	0.010	RTC	−0.02975	−2.86	0.005
NR	−0.001546	−4.05	<0.001	NL	−0.02753	−2.86	0.006

Please cite this article as: H. Lee, et al., Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach, Technol. Forecast. Soc. Change (2013), http://dx.doi.org/10.1016/j.techfore.2013.08.020

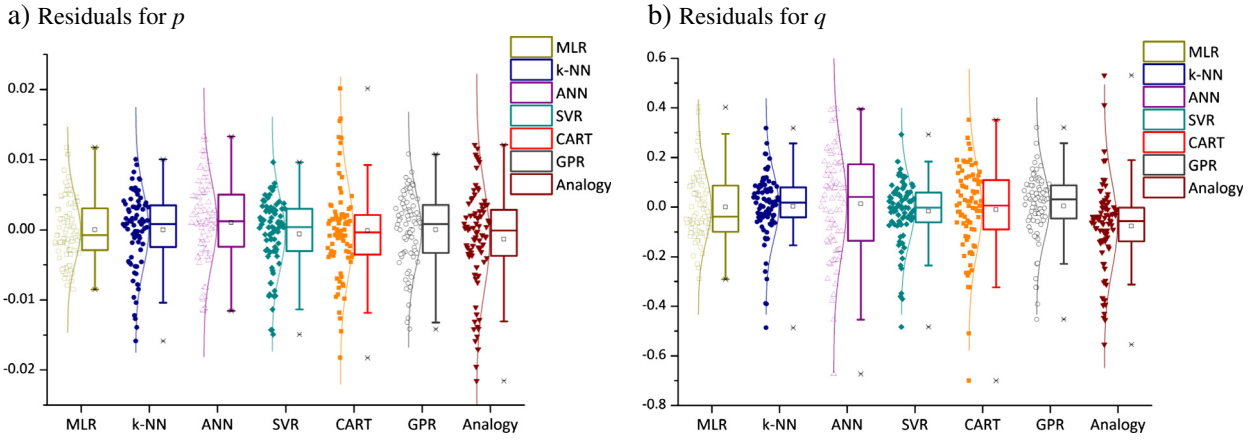

Fig. 5. The distribution and the box plot of the prediction residuals for each single prediction model.

generalization is guaranteed only when a sufficient number of training instances are provided. This requirement is not likely to be met by having 80 products only, and therefore, ANN and CART result in lower prediction accuracies.

Turning to the comparison of prediction performance between the statistical and machine learning-based algorithms and the conventional analogical method, the former are superior to the latter with only a few exceptions, namely CART for p and ANN for q. Setting these exceptions aside, the analogical method is at least 47% worse than the others at predicting p and at least 8.7% worse at predicting q. Compared with the best single algorithm, the MAE of the analogical method is almost twice that of MLR for both p and q. In other words, in the best-case scenario, the new prediction methodology proposed in this study has a twice as strong predictive power compared with the conventional analogical method.

Fig. 5 shows the distribution and box plot of the residuals $(y-\hat{y})$ produced by each single prediction model. A good prediction model should meet the following two qualifications: (1) the average of the residuals should be as close to zero as possible and (2) the dispersion of the residuals should be as narrow as possible. Given these conditions, we find that the conventional analogical method is the poorest model because not only does it have the widest residual dispersion but also its average is below zero for both p and q. As previously noted, ANN and CART are inferior to the other statistical and machine learning-based models since their dispersions are almost as wide as that of the analogical method. This confirms that MLR produces the most desirable outcomes in that its prediction residuals are most narrowly distributed with an average value close to zero. However, the box plot implies that k-NN, SVR, and GPR seem better than MLR at predicting q, because their inter-quantile ranges are smaller than that of MLR. However, their MAEs and RMSEs are spoiled by a few products, for which they fail to make proper predictions, in contrast to MLR.

4.3. Forecasting performances: ensemble model

Of the 57 possible combinations of regression algorithms for constructing ensemble models, the union of MLR and GPR was found to be most accurate. Table 4 compares the prediction performances of the best ensemble model with those of the analogical method and MLR. Constructing an ensemble model resulted in the prediction errors deceasing greatly for both p and q; almost 90% of the MAE and RMSE of the analogical method disappeared in both cases. Compared with the best single model (i.e., MLR), the ensemble prediction model still provided a surprising improvement; over 75% of prediction errors were reduced for both p and q regardless of the performance measures.

Fig. 6 depicts the estimated Bass model parameters (target, x-axis) and prediction outcomes (y-axis) derived by the three models: the analogical method, the best single model (i.e., MLR), and the best ensemble model (i.e., MLR and GPR). The straight line in the figures represents the ideal cases where predicted outcomes equal to their actual targets. Thus, the closer the points approach the line, the better is the prediction model. In this respect, the analogical method is the

Table 4
The prediction performance of the ensemble model compared to the analogical method and the MLR (The bold face numbers denote the lowest error among the algorithms).

Algorithm	p		q	
	MAE	RMSE	MAE	RMSE
MLR	0.0026	0.0031	0.0688	0.0898
Ensemble	**0.0006**	**0.0007**	**0.0180**	**0.0223**
Analogy	0.0053	0.0073	0.1337	0.1907

Please cite this article as: H. Lee, et al., Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach, Technol. Forecast. Soc. Change (2013), http://dx.doi.org/10.1016/j.techfore.2013.08.020

most inferior of the three models because the points in Fig. 6(a) and (b) seem to be almost randomly distributed. In addition, its prediction coverage is much narrower than that of the actual targets; its predicted p values are generally located within 0 and 0.01, although their target values are in the range of 0 to 0.03. Moreover, its predicted q values are mostly located within 0 and 0.3 even though their target values are in the range of 0 to 0.7. It seems as though the randomness and narrow coverage of these predicted outcomes can be resolved using MLR (Fig. 6(c) and (d)). However, two other issues arise from the MLR results: (1) MLR tends to over-predict the estimates beyond a certain extent ($p > 0.02$ and $q > 0.05$) and (2) the prediction outcomes of some products are negative. Fig. 6(e) and (f) shows that the ensemble model can overcome the problems caused by MLR: there are no over-predicted values and negative outcomes appear in only two products when predicting p.

The experimental results in this study can be summarized as follows. First, the product attributes configured herein are shown to be valid, as they lead the regression algorithms to accurate prediction results. Second, statistical and machine learning-based regression algorithms result in a higher predictive power compared with the conventional analogical model. Among the regression algorithms, the best prediction model was MLR; not only because its prediction error rate was the lowest, but also because its residual errors were most compactly distributed. Lastly, the ensemble model significantly enhanced the prediction accuracy of the single regression algorithms. In

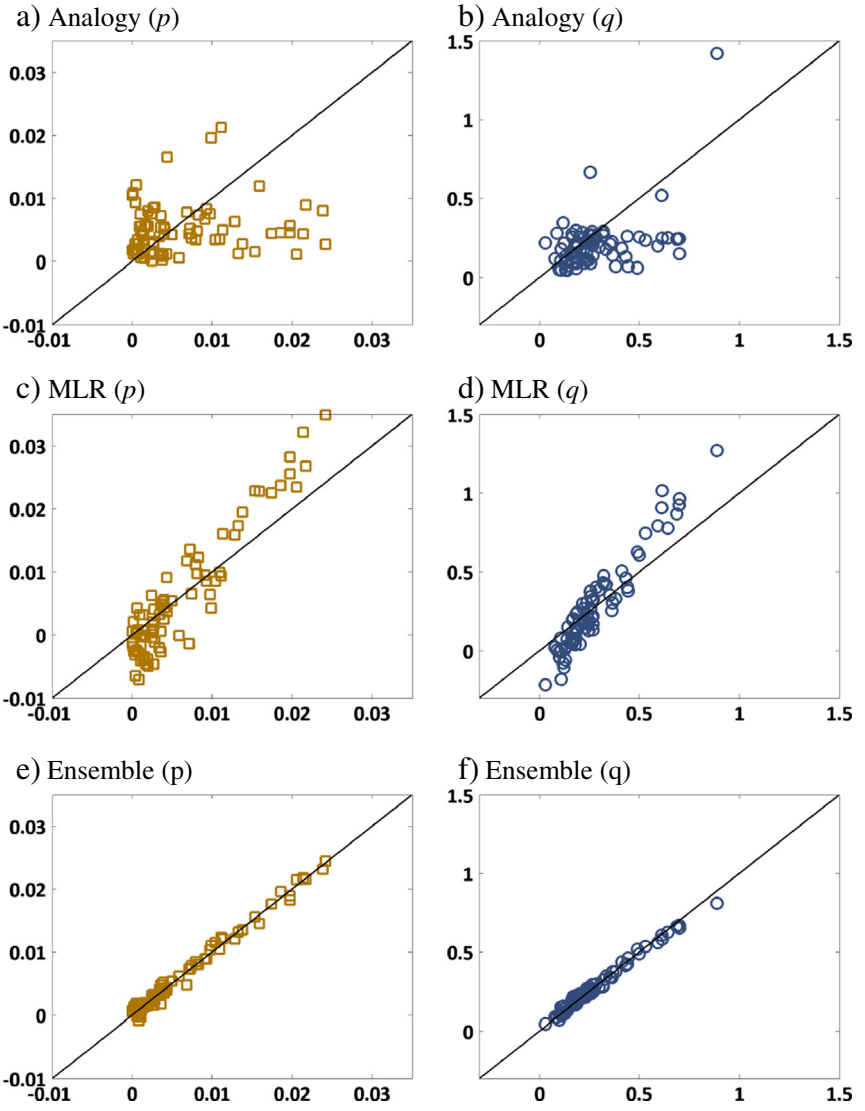

Fig. 6. The predicted against target values of the analogical method, the best single model (MLR), and the ensemble model.

Please cite this article as: H. Lee, et al., Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach, Technol. Forecast. Soc. Change (2013), http://dx.doi.org/10.1016/j.techfore.2013.08.020

addition, it resolved the practical issues caused by the other approaches, such as low correlation between the actual and predicted estimates, over-prediction beyond a certain extent, and negative outcome generation.

5. Illustrative example: demand forecasting of 3D TV

For the purpose of illustration, demand for 3D TV in the North American market is forecasted in this section. 3D TV is a next-generation display that conveys depth perception to the viewer by employing techniques such as stereoscopic display, multi-view display, 2D-plus-depth, or any other form of 3D display. As interest in 3D, which started with the movie *Avatar*, has exploded, major TV manufacturers such as Samsung, LG, Sony, and Panasonic have aimed to steal market share by launching various types of 3D TVs. Although Samsung launched the first commercial 3D TV in 2008, 2010 was considered to be the breakthrough year of the technology, as over six million 3D TVs were sold in 2010 [51]. Because only three years of data points are available at this stage, reliable parameter estimates of the Bass model cannot be produced using the NLS estimation. We therefore apply the developed prediction models to forecast 3D TV demand.

First, the attribute values of 3D TV for the seven significant variables were measured by expert judgments. These experts included engineers and marketing managers of one of the leading display manufacturers as well as professors majoring in electronic engineering, particularly display devices. The measured values of 3D TV are presented in Table 5. The best single regression model (MLR) and best ensemble model (MLR and GPR) were then employed to predict the parameters of 3D TV.

Inputting the obtained attribute values of 3D TV into the two models produced two sets of Bass diffusion parameters, as shown in Table 6. The estimates obtained from the MLR model were found to be slightly higher than those from the ensemble model for both p and q. Once p and q had been obtained, important points in the product lifecycle, such as takeoff (T1) and peak (T^*), could then also be straightforwardly predicted [52]. Times to takeoff and peak in the MLR model were found to slightly precede those in the ensemble model.

Next, potential market size (m) was estimated in order to forecast annual demand. This illustrative example limits its scope to the North American market. Although previous studies have surveyed consumers' purchase intentions to estimate the potential market size, this study rather uses the total number of households in North America as a proxy for the potential market size because this example case aims to illustrate how to apply the developed prediction models to estimate the parameters of the two types of communication effects.

Demand for 3D TV in North America for 15 years (from 2010 to 2024) was then forecasted by combining the total number of the North American households in 2010 (128 million) by the two predicted parameters. Fig. 7(a) and (b) depicts the annual and cumulative demand patterns derived from the ensemble model. Comparing the forecasts with real sales data in first years may uphold the validity of the prediction models. 3D TV shipments in the North American market totaled 8.4 million units in 2012, from 4.5 million units in 2011. The forecast cumulative sales in the ensemble model are 4.2 million in 2011 and 8.5 million in 2012, which are very similar to reals sales, although the annual sales in 2010 are somewhat overestimated.

6. Conclusions

This study proposed a statistical and machine learning algorithm-based approach to pre-launch product demand forecasting on the basis of the Bass model. Taking the product attribute DB as inputs and product diffusion DB as outputs, single prediction models were developed using the six regression algorithms, on the basis of which an ensemble prediction model was constructed to enhance predictive power. It was shown that most single prediction models outperformed the conventional analogical method and that the ensemble model improved prediction accuracy further. An illustrative example of 3D TV was also provided to demonstrate how the developed models could be used in practice.

This study contributes to the field of pre-launch forecasting by proposing a new approach that utilizes statistical and machine learning-based regression algorithms. Despite the importance of the pre-launch forecasting of new products, conventional approaches such as subjective and analogical methods fail to produce objective estimates of diffusion parameters. However, adopting statistical and machine learning-based regression algorithms can reliably portray the relationship between the attributes and diffusion characteristics of existing products, which, in turn, enables forecasting

Table 5
Attribute values of 3D TV.

Variable	Value
Industry classification (IC)	I
Necessity of complementary goods (NCG)	4
Availability of substitute goods (ASG)	2
Level of price (LP)	2
Diversity of distribution channels (DDC)	4
Degree of newness (DN)	I
Rate of technological change	2
Easiness of imitation (EI)	2
Necessity of learning (NL)	2
Necessity of repurchase (NR)	3

Table 6
Predicted parameters of 3D TV.

Model	Parameter estimates		Life cycle	
	p	q	Takeoff	Peak
MLR	0.0087	0.5732	4.9 (Late in 2014)	7.2 (Early in 2017)
Ensemble	0.0091	0.5525	5.0 (Early in 2015)	7.3 (Early in 2017)

Please cite this article as: H. Lee, et al., Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach, Technol. Forecast. Soc. Change (2013), http://dx.doi.org/10.1016/j.techfore.2013.08.020

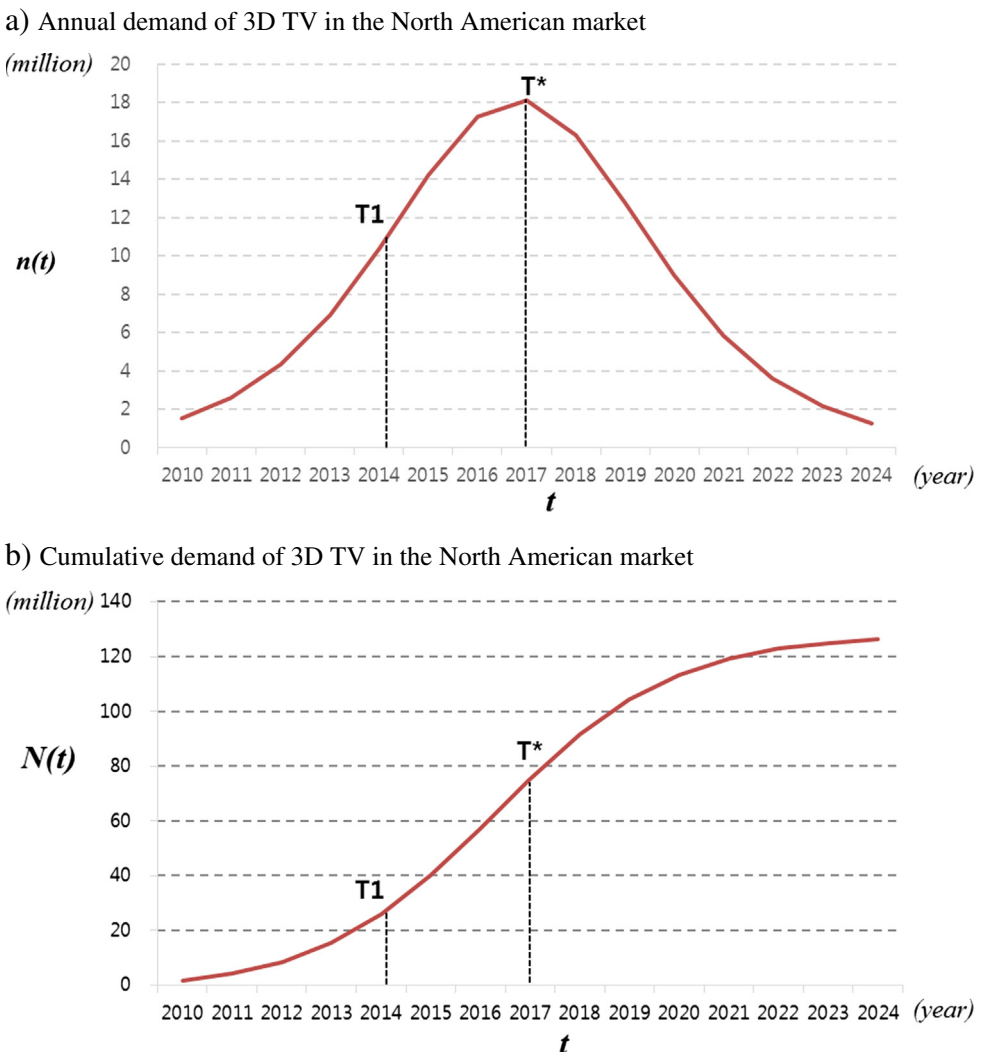

Fig. 7. Demand forecasts of 3D TV in the North American market from the ensemble model.

new product demand solely on the basis of product attributes (i.e., without human manipulation) and fosters effective pre-launch decision-making.

However, the prerequisite for benefiting from the proposed approach is maintaining a relevant and sufficient DB of existing products. Although the primary purpose of this study was to propose a new approach, the data used for constructing the prediction models herein may not be enough to take advantage of the statistical and machine learning algorithms. In addition, as we accumulated as many products as possible in the product DB, the nations and industries into which the products were diffused varied, implying that the validity of the parameter estimates may be questionable.

Moreover, the performance of the proposed approach relies heavily on the product DB used to construct the prediction models. The more homogeneous products included in this product DB, the higher is forecasting accuracy. The product DB should also be regularly updated. Also, the variables selected as product attributes are by no means exhaustive or fixed. Although this paper employed 17 variables, they may not be sufficient to explain the idiosyncratic characteristics of the diffusion of various products. The significant variables can also vary depending on the country or industry context. Future exploratory studies may help identify the context-specific factors that affect the diffusion of products.

Acknowledgments

This work was supported by the Korea Institute of Science and Technology Information (KISTI) and the National Research Foundation of Korea (NRF) grants funded by the Korea government (MEST) (No. 2011-0012759 and 2011-0021893).

Appendix A. The Bass model parameters estimated by the NLS and predicted by the MLR, ensemble, and analogy

Product	p				q			
	NLS	MLR	Ensemble	Analogy	NLS	MLR	Ensemble	Analogy
PC printers	0.000007	−0.001303	0.000686	0.001749	0.173776	0.108571	0.170087	0.131227
Optical cable	0.000027	0.000547	0.000448	0.010440	0.611847	0.907479	0.606509	0.516968
Portable and transportable navigation	0.000114	−0.001895	0.000784	0.010841	0.887896	1.269737	0.809062	1.419575
Press machine	0.000189	0.002099	0.000781	0.001110	0.443570	0.381084	0.461186	0.068896
Automobile	0.000232	−0.002607	0.001240	0.001880	0.218231	0.268223	0.212583	0.115398
Cellular phones	0.000344	−0.003137	0.001410	0.002719	0.267173	0.169772	0.254270	0.134534
Typewriter	0.000409	−0.006499	0.001476	0.009295	0.488950	0.626315	0.517113	0.059778
DBS satellite	0.000504	−0.002351	0.001054	0.012096	0.253918	0.374968	0.246854	0.665767
Internal-combustion engine	0.000535	0.000255	0.001818	0.001076	0.107675	0.082482	0.115944	0.102486
Mobile telephone	0.000627	0.004263	−0.000144	0.002708	0.431015	0.462085	0.415481	0.127981
Cellular phone registration	0.000728	0.000772	0.000627	0.003380	0.700340	0.965365	0.653765	0.146509
Elevator	0.000760	−0.002713	0.000390	0.001053	0.227349	0.173853	0.223963	0.090518
Voice recorder	0.000854	−0.007097	−0.000815	0.002800	0.245686	0.168662	0.263859	0.108042
Digital TV sets & monitors	0.000967	0.003219	0.001373	0.005509	0.500333	0.607216	0.487432	0.252469
Portable MP3 players	0.001064	−0.000071	0.000146	0.007446	0.641945	0.777833	0.624977	0.248487
Dishwashers	0.001072	−0.004083	0.001253	0.000605	0.141863	0.150857	0.131015	0.073823
Telephone	0.001105	−0.001669	−0.000292	0.004705	0.266719	0.133493	0.292473	0.235467
Color TV	0.001163	−0.000091	0.000668	0.005498	0.268695	0.327285	0.249017	0.265833
Plasma DTV	0.001324	−0.000342	0.001189	0.004391	0.593112	0.792773	0.559591	0.197693
Domain registration	0.001368	0.003230	0.001705	0.003413	0.411505	0.506795	0.436337	0.182340
Modems/fax modems	0.001444	−0.003999	0.001252	0.004012	0.178130	0.040046	0.200587	0.207144
Electric fan	0.001511	−0.000289	0.001301	0.001717	0.184559	0.123643	0.182164	0.089883
Washing machine	0.001511	0.000714	0.001722	0.000429	0.184559	0.236248	0.222586	0.056745
Food disposers	0.001584	−0.003468	0.001945	0.001339	0.108291	0.006477	0.099216	0.047727
Digital cameras	0.001629	−0.000039	0.001815	0.005770	0.440367	0.403553	0.422049	0.258400
Analog color TV	0.001719	−0.004620	0.001604	0.005466	0.162791	0.115477	0.198342	0.271943
Compact audio systems	0.001974	−0.003861	0.001454	0.007952	0.229504	0.178163	0.240487	0.219916
Total CD players	0.001996	−0.004940	0.001404	0.005514	0.238066	0.230657	0.251809	0.216656
Air conditioners	0.002452	0.006253	0.002589	0.001058	0.128405	0.017830	0.109323	0.129191
VCR decks with stereo	0.002496	−0.000429	0.003139	0.007533	0.297619	0.385133	0.275455	0.189849
Electronic copier	0.002498	0.000672	0.002770	0.000005	0.187467	0.242066	0.175729	0.121643
Dehumidifiers	0.002582	0.002038	0.002972	0.001259	0.138786	0.073962	0.124480	0.045287
Vending machine	0.002653	0.000863	0.003208	0.008398	0.317252	0.430895	0.284182	0.281935
Record player	0.002683	−0.004661	0.002052	0.002731	0.241972	0.203099	0.260761	0.108182
Cordless telephones	0.002694	0.001564	0.001547	0.003105	0.203386	0.042198	0.199541	0.165024
Clothes dryers	0.002750	0.004099	0.002247	0.001246	0.101752	−0.041971	0.105615	0.048250
VCR decks	0.002831	−0.001168	0.002363	0.008520	0.164984	0.197098	0.158947	0.252340
Key phone	0.003406	−0.002080	0.003095	0.003895	0.029691	−0.214969	0.043558	0.216050
Electric rice cooker	0.003449	0.003296	0.004149	0.001190	0.098339	−0.006931	0.071583	0.048523
Portable CD equipment	0.003593	−0.002656	0.001768	0.007193	0.264427	0.318750	0.291797	0.286239
Black-and-white television	0.003599	0.000551	0.004717	0.005358	0.321698	0.475157	0.325070	0.262775
Bed coverings	0.003612	0.004822	0.003640	0.001053	0.128234	0.010993	0.119144	0.090518
Electric cultivator	0.003760	0.004883	0.004183	0.000116	0.230795	0.145375	0.212446	0.109116
Crane for construction	0.003911	0.005385	0.005165	0.000738	0.168819	0.204768	0.217062	0.096372
Analog projection TV	0.004006	0.005599	0.003467	0.005334	0.256251	0.344262	0.272023	0.266551
Aftermarket PC monitors	0.004006	0.002498	0.004135	0.004267	0.256251	0.267027	0.240701	0.213241
Video tape recorder & player	0.004076	0.004364	0.003950	0.005118	0.318898	0.427691	0.279698	0.290792
Lawn mowers	0.004344	0.009071	0.003879	0.001118	0.129472	−0.057270	0.139290	0.046032
Refrigerator	0.004395	0.003688	0.004310	0.016532	0.121771	−0.108044	0.138332	0.214533
Analog TV/VCR combinations	0.004979	0.005422	0.005353	0.004222	0.353099	0.354612	0.341510	0.208771
Lathe	0.005895	−0.000093	0.006135	0.000540	0.185121	0.124767	0.211764	0.094741
Freezers	0.006867	0.011703	0.004715	0.007772	0.077488	0.022107	0.089495	0.116123
Camcorders	0.007141	−0.001363	0.007285	0.004300	0.119277	−0.077661	0.154889	0.344024
Digital projection sets & monitors	0.007229	0.013551	0.007307	0.005148	0.687376	0.868696	0.662500	0.241678
Telephone answering devices	0.007422	0.006548	0.007815	0.003654	0.189819	0.246770	0.178793	0.206442
Home theater-in-a-box	0.007972	0.011119	0.008432	0.003404	0.332291	0.419107	0.345450	0.174776
Analog color TV with stereo	0.008169	0.009690	0.008003	0.004754	0.199138	0.188172	0.233555	0.251856
Gas range	0.008272	0.012255	0.007980	0.007311	0.237224	0.306208	0.266662	0.152235
Microwave oven	0.009138	0.009446	0.008936	0.006617	0.190294	0.135113	0.185489	0.189779
Aftermarket remote controls	0.009307	0.008507	0.008875	0.008278	0.224803	0.204132	0.226278	0.182739
DVD players/recorders	0.009722	0.006413	0.010305	0.007518	0.363046	0.255407	0.335606	0.224463
Analog handheld LCD color TV	0.009892	0.004231	0.011012	0.019576	0.173174	0.065502	0.185151	0.147190
Video camera	0.010421	0.008484	0.011464	0.003426	0.530827	0.745438	0.535237	0.234278
Fax machines	0.010969	0.009894	0.010448	0.003441	0.266902	0.217588	0.268064	0.201818
Blank Videocassettes	0.011130	0.009372	0.012286	0.021251	0.137842	0.030209	0.125984	0.106521
Monochrome TV	0.011322	0.015997	0.012090	0.004912	0.086825	0.005411	0.088787	0.276325
CDP	0.012830	0.015818	0.012043	0.006269	0.183301	0.079264	0.176501	0.294352
Rack audio systems	0.013253	0.017287	0.013161	0.001184	0.366526	0.304232	0.372504	0.137702
Corded telephones	0.013786	0.019458	0.013523	0.002727	0.109153	−0.181639	0.149943	0.176069

Please cite this article as: H. Lee, et al., Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach, Technol. Forecast. Soc. Change (2013), http://dx.doi.org/10.1016/j.techfore.2013.08.020

Appendix A (*continued*)

Product	*p*				*q*			
	NLS	MLR	Ensemble	Analogy	NLS	MLR	Ensemble	Analogy
Family radio devices	0.015335	0.022861	0.015506	0.001505	0.380731	0.332642	0.376496	0.068652
Personal computers	0.015950	0.022743	0.014494	0.011918	0.179772	0.078699	0.164870	0.127937
Portable tape and radio/tape players	0.017427	0.022506	0.017595	0.004357	0.232762	0.184680	0.258116	0.217054
LCD TV (Digital and analog)	0.018607	0.023682	0.019608	0.004492	0.245743	0.126912	0.269804	0.267157
Videocassette players	0.019732	0.025518	0.018246	0.004515	0.285609	0.402980	0.296578	0.247694
MP3	0.019738	0.028225	0.018932	0.005578	0.697928	0.925170	0.670885	0.242889
Personal word processors	0.020585	0.023438	0.021476	0.001072	0.215886	0.298904	0.205950	0.282171
LCD monitor	0.021403	0.032172	0.021779	0.004330	0.612865	1.015626	0.581345	0.245977
Analog handheld LCD monochrome TV	0.021752	0.026762	0.021560	0.008903	0.163545	0.046244	0.180689	0.155856
Electronic calculator	0.023836	0.035584	0.023161	0.007975	0.255873	0.378878	0.233736	0.089886
Facsimile	0.024187	0.034919	0.024426	0.002648	0.264186	0.212156	0.253331	0.201981

References

[1] F.M. Bass, K. Gordon, T.L. Ferguson, M.L. Githens, DIRECTV: forecasting diffusion of a new technology prior to product launch, Interfaces 31 (2001) S82–S93.
[2] E. Ozkaya, Demand management in global supply chains, (Doctoal thesis) Georgia Institute of Technology, Atlanta, 2008.
[3] E.M. Rogers, Diffusion of Innovations, The Free Press, New York, 1995.
[4] N. Meade, T. Islam, Modelling and forecasting the diffusion of innovation—a 25-year review, Int. J. Forecast. 22 (2006) 519–545.
[5] F.M. Bass, A new product growth model for consumer durables, Manag. Sci. 15 (1969) 215–227.
[6] T. Turk, P. Trkman, Bass model estimates for broadband diffusion in European countries, Technol. Forecast. Soc. Change 79 (2012) 85–96.
[7] G.L. Lilien, A.G. Rao, S. Kalish, Bayesian estimation and control of detailing effort in a repeat purchase diffusion environment, Manag. Sci. 27 (1981) 493–506.
[8] A.G. Rao, M. Yamada, Forecasting with a repeat purchase diffusion model, Manag. Sci. 34 (1988) 734–752.
[9] P.J. Lenk, A.G. Rao, New models from old: forecasting product adoption by hierarchical Bayes procedures, Mark. Sci. 9 (1990) 42–53.
[10] J. Xie, X.M. Song, M. Sirbu, Q. Wang, Kalman filter estimation of new product diffusion models, J. Mark. Res. (1997) 378–393.
[11] V. Mahajan, S. Sharma, A simple algebraic estimation procedure for innovation diffusion models of new product acceptance, Technol. Forecast. Soc. Change 30 (1986) 331–345.
[12] K.D. Lawrence, W.H. Lawton, Applications of diffusion models: some empirical results, in: V. Mahajan, R.C. Cardozo (Eds.), New Product Forecasting, Lexington Books, Lexington, 1981, pp. 529–541.
[13] T. Kim, J. Hong, H. Koo, Forecasting diffusion of innovative technology at pre-launch, Ind. Manage. Data Syst. 113 (2013) 800–816.
[14] R.J. Thomas, Estimating market growth for new products: an analogical diffusion model approach, J. Prod. Innov. Manag. 2 (1985) 45–55.
[15] B.L. Bayus, High-definition television: assessing demand forecasts for a next generation consumer durable, Manag. Sci. 39 (1993) 1319–1333.
[16] J.M. Choffray, G.L. Lilien, A decision-support system for evaluating sales prospects and launch strategies for new products, Ind. Market. Manag. 15 (1986) 75–85.
[17] J. Lee, P. Boatwright, W.A. Kamakura, A Bayesian model for prelaunch sales forecasting of recorded music, Manag. Sci. 49 (2003) 179–196.
[18] H. Seol, G. Park, H. Lee, B. Yoon, Demand forecasting for new media services with consideration of competitive relationships using the competitive Bass model and the theory of the niche, Technol. Forecast. Soc. Change 79 (2012) 1217–1228.
[19] P. Goodwin, K. Dyussekeneva, S. Meeran, The use of analogies in forecasting the annual sales of new electronics products, IMA J. Manag. Math. 24 (2012) 407–422.
[20] V. Mahajan, E. Muller, F.M. Bass, New product diffusion models in marketing: a review and directions for research, J, Marketing (1990) 1–26.
[21] C.M. Bishop, Pattern Recognition and Machine Learning, Springer-Verlag, New York, 2006.
[22] T.M. Mitchell, Machine Learning, McGraw-Hill, Inc., Singapore, 1997.
[23] D.C. Schmittlein, V. Mahajan, Maximum likelihood estimation for an innovation diffusion model of new product acceptance, Mark. Sci. 1 (1982) 57–78.
[24] V. Srinivasan, C.H. Mason, Technical Note—Nonlinear least squares estimation of new product diffusion models, Mark. Sci. 5 (1986) 169–178.
[25] R.M. Heeler, T.P. Hustad, Problems in predicting new product growth for consumer durables, Manag. Sci. 26 (1980) 1007–1020.
[26] W.P. Putsis, V. Srinfvasan, Estimation techniques for macro diffusion models, in: V. Mahajan, E. Muller, Y. Wind (Eds.), New-product Diffusion Models, Springer, New York, 2000, pp. 263–291.
[27] F. Sultan, J.U. Farley, D.R. Lehmann, A meta-analysis of applications of diffusion models, J. Mark. Res. (1990) 70–77.
[28] J. Lee, C.-Y. Lee, K.S. Lee, Forecasting demand for a newly introduced product using reservation price data and Bayesian updating, Technol. Forecast. Soc. Change 79 (2012) 1280–1291.
[29] F.M. Bass, The adoption of a marketing model: comments and observations, in: V. Mahajan, Y. Wind (Eds.), Innovation Diffusion Models of New Product Acceptance Ballinger, Cambridge, 1986.
[30] W.Y. Lee, P. Goodwin, R. Fildes, K. Nikolopoulos, M. Lawrence, Providing support for the use of analogies in demand forecasting tasks, Int. J. Forecast. 23 (2007) 377–390.
[31] J. Ilonen, J.-K. Kamarainen, K. Puumalainen, S. Sundqvist, H. Kälviäinen, Toward automatic forecasts for diffusion of innovations, Technol. Forecast. Soc. Change 73 (2006) 182–198.
[32] H. Gatignon, J. Eliashberg, T.S. Robertson, Modeling multinational diffusion patterns: an efficient methodology, Mark. Sci. 8 (1989) 231–247.
[33] R.K. Srivastava, V. Mahajan, S.N. Ramaswami, J. Cherian, A multi-attribute diffusion model for forecasting the adoption of investment alternatives for consumers, Technol. Forecast. Soc. Change 28 (1985) 325–333.
[34] V. Mahajan, R.A. Peterson, Innovation diffusion in a dynamic potential adopter population, Manag. Sci. 24 (1978) 1589–1597.
[35] S.M. Ross, Introduction to Probability and Statistics for Engineers and Scientists, Elsevier Science, London, 2009.
[36] T. Hastie, R.J. Tibshirani, J.J.H. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, New York, 2009.
[37] G.S. Iwerks, H. Samet, K. Smith, Continuous k-nearest neighbor queries for continuously moving points with updates, Proceedings of the 29th International Conference on Very Large Data Bases (VLDB'03), VLDB Endowment, Berlin, Germany, 2003, pp. 512–523.
[38] Y. Xiaohui, K.Q. Pu, N. Koudas, Monitoring k-nearest neighbor queries over moving objects, Proceedings of the 21st International Conference on Data Engineering (ICDE 2005), Tokyo, Japan, 2005, pp. 631–642.
[39] K. Mouratidis, M.L. Yiu, D. Papadias, N. Mamoulis, Continuous nearest neighbor monitoring in road networks, Proceedings of the 32nd International Conference on Very Large Data Bases, VLDB Endowment, Seoul, Korea, 2006, pp. 43–54.
[40] W.G. Baxt, Use of an artificial neural network for the diagnosis of myocardial infarction, Ann. Intern. Med. 115 (1991) 843–848.
[41] D.C. Park, M.A. El-Sharkawi, R.J. Marks II, L.E. Atlas, M.J. Damborg, Electric load forecasting using an artificial neural network, IEEE Trans. Power Syst. 6 (1991) 442–449.
[42] A. Heiat, Comparison of artificial neural network and regression models for estimating software development effort, Inform. Software Tech. 44 (2002) 911–922.
[43] J. Snyman, Practical mathematical optimization: an introduction to basic optimization theory and classical and new gradient-based algorithms, Springer, New York, NY, 2005.
[44] T.J. Ypma, Historical development of the Newton–Raphson method, SIAM Rev. 37 (1995) 531–551.
[45] D.W. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, SIAM J. Appl. Math. 11 (1963) 431–441.
[46] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (2004) 199–222.
[47] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Wadsworth and Brooks, Monterey, 1984.
[48] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, The MIT Press, London, 2005.
[49] R. Avnimelech, N. Intrator, Boosted mixture of experts: an ensemble learning scheme, Neural. Comput. 11 (1999) 483–497.
[50] T.L. Saaty, How to make a decision: the analytic hierarchy process, Eur. J. Oper. Res. 48 (1990) 9–26.

Please cite this article as: H. Lee, et al., Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach, Technol. Forecast. Soc. Change (2013), http://dx.doi.org/10.1016/j.techfore.2013.08.020

ARTICLE IN PRESS

[51] Displaybank, 3D TV Industry Trend and Market Forecast, Speical Report May 2010, Gyeonggi-do, Korea, 2010.
[52] V. Mahajan, E. Muller, R.K. Srivastava, Determination of adopter categories by using innovation diffusion models, J. Mark. Res. 27 (1990) 37–50.

Hakyeon Lee is an assistant professor in the Dept. of Industrial and Information Systems Engineering, Seoul National University of Science and Technology. He received the degrees of B.S. and Ph.D. from Seoul National University. His research interests are in technological forecasting and innovation diffusion. He has authored several published papers in leading journals of technology and innovation management, including *Technological Forecasting and Social Change*, *Research Policy*, *Technology Analysis & Strategic Management*, *Journal of Engineering and Technology Management*, *and Technology in Society*.

Sang Gook Kim is a senior researcher of the Industry Information Analysis Center, Korea Institute of Science and Technology Information. He received the degree of BS in Industrial Engineering at Seoul National University and the Ph.D. in Operations Research at Florida Institute of Technology. His recent research interests are in technology valuation, product lifecycle forecasting, and technology commercialization and main study areas also include queuing models based on stochastic process and optimization theories. He has authored a few papers on related topics in leading journals such as *Nonlinear Analysis: Theory*, *Methods & Applications*, *Journal of Supply Chain and Operations Management*, *and Asian Journal of Innovation and Policy*.

Hyun-woo Park is a technology economist at the Korea Institute of Science and Technology Information (KISTI). He worked with San Francisco State University as a Visiting Fellow (1996–1997) and University of California, Santa Cruz as a Research Scholar (2008–2009). He received the B.S., M.S, and Ph.D. in International Business (1991) from Hong-Ik University, and Ph.D. in Science and Technology Studies (2007) from Korea University. He published many articles and books in the research fields including technology valuation, innovation management, technology commercialization, and scientometrics. He has authored a number of papers on related topics in leading journals such as *Scientometrics*, *Asian Journal of Technology Innovation*, and *Journal of Supply Chain and Operations Management*.

Pilsung Kang is an assistant professor in the Dept. of Industrial and Information Systems Engineering, Seoul National University of Science and Technology. He received B.S. and Ph.D. in Industrial Engineering at Seoul National University. His research interests include instance-based learning, learning kernel machines, novelty detection, learning algorithms in class imbalance, and network analysis. His research also includes application areas such as keystroke dynamics-based authentication, fault detection in manufacturing process, and technological forecasting. He has published a number of papers on related topics in leading journals such as *Pattern Recognition*, *Intelligent Data Analysis*, and *Neurocomputing*.

Please cite this article as: H. Lee, et al., Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach, Technol. Forecast. Soc. Change (2013), http://dx.doi.org/10.1016/j.techfore.2013.08.020