MEASURING UPSELLING POTENTIAL OF LIFE INSURANCE CUSTOMERS: APPLICATION OF A STOCHASTIC FRONTIER MODEL

Byung-Do Kim
Sun-Ok Kim

BYUNG-DO KIM is Assistant Professor of Marketing at the School of Business Administration, Seoul National University, Korea

SUN-OK KIM is a doctoral candidate in Marketing at the School of Business Administration, Seoul National University, Korea

The authors wish to thank Hyun-Moon Park, Executive Director of Marketing Planning at Samsung Life Insurance Co., for supplying the data used in this study.

ABSTRACT

How much more of a service or product can we potentially sell to a customer? Recognizing the effect of selling inefficiency on upselling potential, this article offers a concept of upselling potential that is different from that currently in use and introduces a methodology for calculating customer-specific upselling potential for life insurance customers. The proposed model was applied to the data of 5,000 life insurance customers. This paper shows that the insurer analyzed could have sold an additional 25% worth of premiums for more than half of its customers. Other uses of the technique are also discussed.

© 1999 John Wiley & Sons, Inc. and
Direct Marketing Educational Foundation, Inc.
CCC 1094-9968/99/040002-08

JOURNAL OF INTERACTIVE MARKETING
VOLUME 13 / NUMBER 4 / AUTUMN 1999

1. INTRODUCTION

The life insurance market is undergoing a major paradigm shift mainly because of stiff competition from other financial institutions and a significant decrease in database-related costs. Insurers are now focusing on customers, rather than products or services. As the industry matures, the cost of acquiring a new customer becomes more expensive. As a result, insurers are putting greater emphasis on retaining their current customers. Carefully analyzing various data on their current customers, insurers attempt to sell more of the same services (*upselling*) and/or sell other types of services (*cross-selling*) to those current customers. The success of life insurers critically depends on the long-term relationship with their customers.

Recognizing the importance of upselling in their overall profit contributions, several insurers have developed somewhat heuristic models to calculate upselling potential for their customers. To our knowledge, there are no published papers that discuss models to measure upselling potential. However, informal conversation with managers suggests that insurers often develop regression models to explain customers' insurance premiums using their demographic variables. In this regression framework, upselling possibilities occur when customers change their demographic status. For example, age changes or special events such as marriage will change the insurance needs of those customers.

The current practice of conceptualizing upselling potential is limited in use, however. We claim that selling (or managerial) inefficiency of the insurer should be included in computing customers' upselling potential as well. Life insurance buyers may not be sure of the core benefits and the quality of service a policy provides since the product is intangible, complex, and abstract (Crosby and Stephens, 1987). Hence, they rely on the advice of salespeople in finding a suitable policy. Insurers (or salespeople) should deliver relevant information to their customers, let them recognize their insurance needs, and motivate them to purchase appropriate services/premiums. When the insurer does not do well in its selling effort, customers do not appreciate the value of the policy and purchase a policy with a premium much lower than the potential premium obtainable. Because of this selling inefficiency on the part of the insurer, customers often do not purchase any insurance services. This inefficiency may be measured and eliminated by additional selling efforts with appropriate sales training. In the life insurance industry, selling inefficiency is mainly the result of factors such as the lack of ability in salespeople, outdated selling methods, inappropriate allocation of marketing expenditures.

The main objective of this paper is to calculate customer-level upselling potential for life insurers. Recognizing current customers as response units, we employ a stochastic frontier model to estimate the maximum premium obtainable from each customer and the (customer-level) inefficiency of the life insurer's selling activity. These estimates allow us to compute upselling scores that indicate how much more of the same product/service we could potentially sell for each current customer. With these upselling scores, insurers can rank-order current customers to maximize their marketing efficiency. The procedure presented here differs from current standard practice, in that customers with large upselling potentials (due to selling inefficiency from the insurer) are recognizes even though they do not change their demographic status.

2. MODEL SPECIFICATION

This model considers customers as response units that respond to the inputs offered by various marketers. Responding to these inputs, customers will decide to purchase a life insurance service to fit their insurance needs. However, given the same type of insurance product/service (e.g., protection insurance), customers might select policies with different levels of coverage (or different premiums) depending on factors such as their insurance needs, the selling efficiency of salespeople, and marketing efforts of insurers.

Our goal is to explain the variability of monthly premiums chosen by current customers using a set of independent variables. We now conceptualize premiums (sold to customers by an insurance firm) as functions of demographics of the policy owners and the insured (e.g., sex, type of profession, and age), marketing activity of the firm and competitors (e.g., advertising, premiums versus benefits, number of salespeople), and macro environment (i.e., GNP). More specifically, we consider a following regression model with n observations or customers.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_K X_{Ki} + \varepsilon_i \quad \varepsilon_i = \nu_i - u_i \tag{1}$$

where Y_i ($i = 1, \ldots, n$) is the dependent variable (or monthly premium) of ith customer, X_{ki} is the kth ($k = 1, \ldots, K$) independent variable of the customer i (proposed to explain the dependent variable), and β is an unknown parameter to be estimated. The point of departure from the classical regression is in the specification of error term εi ($= v_i - u_i$), which consists of two components. The symmetric part v_i is assumed to follow the error structure of standard regression and supposed to capture the effect of statistical noise, measurement error, and random shocks outside the firm's control: $v_i \sim N(0, \sigma_v^2)$. The other part is supposed to capture management inefficiency under the firm's control and we assume that $u_i \sim [N(0, \sigma_u^2)]$. We also assume that u_i is distributed independent of v_i. Note that u_i is always ≥ 0. This one sided assumption of u_i implies that the monthly premium of customer i (Y_i) must lie on or below the customer's frontier ($\beta_0 + \beta_1 X_{1i} + \cdots + \beta_K X_{Ki} + v_i$) since $u_i \geq 0$. Hence, the frontier can be defined to be the maximum premium that the insurer can sell. In our econometric specification, the frontier is the sum of two components: the part we have modeled with a set of independent variables ($\beta_0 + \beta_1 X_{1i} + \cdots + \beta_1 X_{1i}$), and the random component we did not or could not modeled (v_i). The observed premium for customer i is lower or equal to his/her frontier. Any deviation from its frontier (u_i) is the result of factors under the firm's control such as the inefficiency due to salespeople's selling activity and the inappropriate allocation of marketing expenditures.

The specification of frontier (or envelope) functions and its estimations have been a major concern in econometrics and management science for several decades. There have been two research traditions in estimating frontier functions. The first approach estimates the frontier deterministically. This approach, known as data envelopment analysis (DEA), does not allow for stochastic errors and employs mathematical programming to estimate the deterministic frontier. (See Seiford and Thrall [1990] for further details on DEA.) Hence, the calculated frontier may be unstable when the data are contaminated by statistical noise. On the other hand, explicitly considering random errors in the model specification, the stochastic frontier model that we have adopted in equation (1) describes the data more realistically. However, the specification of the frontier is too restrictive in stochastic frontier model (Bauer, 1990).

The stochastic frontier model in equation (1) was first proposed by Aigner, Lovell, and Schmidt (1977). Other researchers have employed different assumptions for u_i distributions, such as exponential, truncated normal, and gamma distributions (Aigner, et al., 1977; Stevenson, 1980; Greene, 1990). We adopt half normal for u_i because of its popularity in stochastic frontier models, leaving more flexible distributions as a future research.

The stochastic frontier model has been applied to various practical problems in management. The concept of *frontier* was originally used to estimate production function where the frontier is defined as maximum outputs obtainable from given inputs. More recently, this concept has been applied to other problems such as estimating profit function or selling function (Aly, Grabowski, Pasurka, and Rangan, 1990; Kamakura, Ratchford, and Agarwal 1988; Mahajan 1991). Here decision-making units are firms that manage the allocation of inputs and produce outputs in their ways.

This study applies the frontier model to the problem of finding the customer potential value. Our application is conceptually different from previous research where the decision-making units are firms. We consider customers as response units who would make purchase decisions based on their insurance needs and the insurers' selling efforts. Here the frontier is the maximum premium/sale obtainable for each customer, which is not observable to researchers but can be estimated by the equation (1). As noted by Bauer (1990), we can regard deviation from the frontier as a measure of the inefficiency once we estimate the frontier. Moreover, the notion of the frontier and the concept of efficiency provide many interesting marketing implications.

Once we estimate the model parameters (βs) in equation (1), we are able to derive the customer- (or observation-) specific estimates of inefficiency u_i. The conditional distribution of u_i given the estimate of the total error ϵ_i has been derived by Jondrow, Lovell, Materov, and Schmidt (1982). We would use the mean of this conditional distribution as the estimates of inefficiency u_i that is given by

$$E(u_i|\varepsilon_i) = \frac{\sigma_u^2\sigma_v^2}{\sigma^2}\left[\frac{\phi(\varepsilon_i\lambda/\sigma)}{1-\Phi(\varepsilon_i\lambda/\sigma)} - \varepsilon_i\lambda/\sigma\right] \quad (2)$$

where $\sigma^2 = \sigma_u{}^2 + \sigma_v{}^2$, $\lambda = \sigma_u/\sigma_v$, $\phi(\cdot)$ is the probability density function of standard normal distribution, and $\Phi(\cdot)$ is its cumulative density function.

As noted above, the frontier (or the maximum premium obtainable) for customer i can be written as $\hat{\beta}'x_i + \hat{v}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \cdots + \hat{\beta}_K X_{Ki} + \hat{v}_i$ which is the premium, with $u_i = 0$ (i.e., there is no inefficiency). Hence, managerial or selling efficiency in percentage term can be measured by $[Y_i/(\hat{\beta}'x_i + \hat{v}_i)] \times 100\%$. Similarly, the percentage inefficiency can be defined as $[\hat{u}_i/(\hat{\beta}'x_i + \hat{v}_i)] \times 100\%$. For example, assume that $Y_i = \$90$, $\hat{\beta}'x_i = \$80$, $\hat{v}_i = \$20$, and $\hat{u}_i = \$10$ for customer i. We say that the monthly (observed) premium of this customer is \$90, which can be broken down into three quantities. $\hat{\beta}'x_i = \$80$ represents the premiums modeled/explained by a researcher, $\hat{v}_i = \$20$ represents the (random) premium gains that are not expected or modeled, and $\hat{u}_i = \$10$ represents the premium loss due to the selling inefficiency. Thus: a customer was willing to pay a monthly premium of \$100, but purchased a policy with a \$90 premium because of the firm's selling inefficiency. In this case, the percentage selling efficiency is 90% (\$90/\$100 × 100%) while the selling inefficiency becomes 10 % (\$10/\$100 × 100%).

3. DATA, ESTIMATION, AND RESULTS

3.1. Data

We applied the stochastic frontier model developed in the previous section to the data kindly supplied by a large life insurance company. The source of the data was 5,000 randomly selected customers who purchased protection insurance from the company from January 1993 to July 1995. The company offers other life insurance products, such as annuity, education, and savings policies; however, we limit our attention to the protection insurance to avoid comparing dissimilar products. For example, a consumer who purchases protection insurance may have insurance needs that are different from those of a consumer who purchases annuity insurance.

The dependent variable of our model is the logarithm of the monthly premium paid. Customers can choose their payment intervals among monthly, quarterly, semiannual, or annual schedules when they contract for protection insurance. The premiums other than monthly ones were adjusted to be comparable to the monthly premiums. Similar to other life insurance companies, this insurer maintains various types customer-specific information (e.g., contract date, premium, type of payment, name of policy owners) collected mainly from the initial contract form. The selection of independent variables was made partly by manager's suggestions and partly by statistical methods such as stepwise regression. The brief descriptions on each of the selected variables are presented in Table 1 below.

Several independent variables included in the final model represent demographic characteristics of the policy owner such as sex, age, job type, and so on. Insurers consider the policy owner's income as an excellent indicator of his

TABLE 1
Brief Descriptions of Independent Variables

Variables	*Description of Variable*
SEX_OWNER	1 if the policy owner is male, 0 if not
SEX_INSURED	1 if the insured is male, 0 if not
SAME	1 if the policy owner and the insured are same person, 0 if not
LOAN	1 if the policy owner has the policy loan, 0 if not
EMPLOYEE	1 if the policy owner is SLI's employee, 0 if not
AGE	Age of insured
AGE_SQ	AGE squared
JOB1[a]	1 if the job type of the policy owner is in the 1st job category, 0 if not
JOB2	1 if the job type of the policy owner is in the 2nd job category, 0 if not
JOB3	1 if the job type of the policy owner is in the 3rd job category, 0 if not
JOB4	1 if the job type of the policy owner is in the 4th job category, 0 if not
JOB5	1 if the job type of the policy owner is in the 5th job category, 0 if not
LENGTH	length of paying premiums (in months)
LENGTH_SQ	LENGTH squared
PAYMENT1	1 if the method of payment is quarterly, 0 if not
PAYMENT2	1 if the method of payment is semi-annual, 0 if not
PAYMENT3	1 if the method of payment is annual, 0 if not
BRANCH1[b]	Dummy variable for 1st segment of branches of making a policy contract.
BRANCH2	Dummy variable for 2nd segment of branches of making a policy contract.
BRANCH3	Dummy variable for 3rd segment of branches of making a policy contract.
BRANCH4	Dummy variable for 4th segment of branches of making a policy contract.
SEASON1	1 if the contract was made during the first quarter, 0 if not
SEASON2	1 if the contract was made during the second quarter, 0 if not
SEASON3	1 if the contract was made during the third quarter, 0 if not
TREND	Linear trend effect
TREND_SQ	TREND squared

[a] The type of customer's job is indexed by 143 different codes in the original data. Instead of creating 142 dummies for each job code, we segment them into 6 classes and use each segment membership as dummies. For example, job category 3 includes scholars, researchers, and some technical jobs.
[b] There are 142 different branch codes in the original data. Similar to the type of jobs, we create 5 branch categories and the corresponding dummies.

or her insurance requirements. Unfortunately, the variable *Income* is not available in our data. However, a couple of proxy variables for income (e.g., policy owner's job type and where he or she lives) are included. We have also included the squared terms of some variables such as age, trend effect, and length of paying premiums since they have shown nonlinear effects. Finally, two independent variables, the job type of the policy owner (JOB) and the location of the branch where the salesperson making a policy contract is assigned (BRANCH), are segmented into five to six clusters before we fit our regression model. For example, the job type of the policy owner is originally indexed by 143 different codes. Instead of creating 142 dummies for each job code (that would require the estimation of too many parameters), we segment them into 6 classes in terms of their mean insurance premiums and use each segment membership as dummies (JOB1 to JOB5). For example, job category 3 (JOB3) includes several job types such as student, researchers, some technical jobs, and so on. Similarly, the original

TABLE 2
Parameter Estimates of OLS and Frontier Model

Variables	*OLS*	*Frontier*
Intercept	10.11 (0.00)[a]	10.53 (0.00)
SEX_OWNER	0.10 (0.00)	0.10 (0.00)
SEX_INSURED	0.09 (0.00)	0.10 (0.00)
SAME	−0.08 (0.00)	−0.08 (0.00)
LOAN	0.05 (0.06)	0.05 (0.06)
EMPLOYEE	−0.37 (0.00)	−0.25 (0.00)
AGE	0.00 (0.84)	−0.00 (0.58)
AGE_SQ	0.00 (0.00)	0.00 (0.00)
JOB1	−0.34 (0.00)	−0.33 (0.00)
JOB2	−0.20 (0.00)	−0.21 (0.00)
JOB3	−0.19 (0.00)	−0.21 (0.00)
JOB4	−0.14 (0.00)	−0.15 (0.00)
JOB5	−0.14 (0.01)	−0.16 (0.00)
LENGTH	0.01 (0.00)	0.01 (0.00)
LENGTH_SQ	−0.00 (0.00)	−0.00 (0.00)
PAYMENT1	−0.38 (0.00)	−0.33 (0.00)
PAYMENT2	−0.22 (0.00)	−0.12 (0.00)
PAYMENT3	−0.04 (0.41)	0.02 (0.48)
BRANCH1	−0.19 (0.00)	−0.18 (0.00)
BRANCH2	−0.11 (0.00)	−0.11 (0.00)
BRANCH3	−0.06 (0.00)	−0.06 (0.00)
BRANCH4	0.05 (0.08)	0.05 (0.05)
SEASON1	0.11 (0.00)	0.11 (0.00)
SEASON2	0.06 (0.00)	0.06 (0.00)
SEASON3	0.02 (0.25)	0.02 (0.21)
TREND	−0.01 (0.05)	−0.01 (0.02)
TREND_SQ	0.00 (0.00)	0.00 (0.00)
Log-likelihood	−2,260	−2,211
σ_u^2		0.19 (0.00)
σ_v^2		0.09 (0.00)
$\lambda\ (=\sigma_u/\sigma_v)$		1.45 (0.00)

[a] The number in parenthesis represents the p-value of the parameter estimates.

data encompass 142 distinct branch codes. We create five branch segments and the corresponding dummies (BRANCH1 to BRANCH4).

3.2. Estimation Results

Table 2 presents the parameter estimates for the ordinary least square (OLS or classical regression) and our stochastic frontier model. Most of the parameter estimates are statistically significant at $p = .01$ for both models. It is important to note that the log-likelihood has been improved from −2,260 for the OLS to −2,211 for our stochastic frontier model. The log-likelihood test shows that this likelihood improvement is statistically significant at $p = .01$. Similarly, the σ_u^2 term in the frontier model is statistically significant at $p = .01$. That is, there exist customer variations in their premiums due to selling inefficiency.

Several parameters are worthwhile to mention. The negative parameter for SAME imply that they select the insurance with lower premiums when the policy owners contract the protection insurance for themselves rather than for their spouse or dependents. The positive AGE_SQ with statistically not significant AGE implies that older policy owners pay the higher premium. This result is frequently found in the analysis of protection insurance data. It is also interesting to note that the payment interval (PAYMENT1 to PAYMENT3) is meaningful to explain the premiums paid.

3.3. Customer Specific Selling Inefficiency

Once we estimate our frontier model, we are able to derive the estimate of selling inefficiency ($\hat{u}_i$) for each customer, using the formula given in equation (2). We calculate these estimates for all 5,000 customers. As mentioned previously, we can compute the maximum attainable premiums (or their frontiers) that are equal to $y_i - \hat{u}_i = \hat{\beta}' x_i + \hat{v}_i$. Hence, selling inefficiency as a percentage of the frontier is represented by $[\hat{u}_i/(\hat{\beta}' x_i + \hat{v}_i)] \times 100\%$, which is actually the upselling potential for customer i. We plot the frequency distribution of this upselling potential for each customer in Figure 1. The mean of this frequency distribution is 28%. It implies that customers on the average have purchased life insurance at a rate 28 %. below their frontiers (or the maximum premiums obtainable). Considering that the insurer analyzed has a reputation in running the company very efficiently, it is surprising to see that more than half of customers have more than 25 %. upselling potential.

Estimating customer-specific upselling poten-

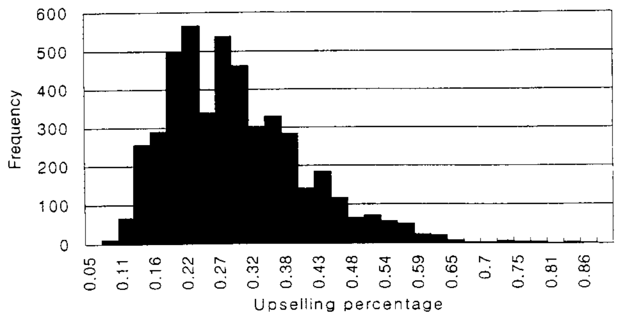

FIGURE 1
Frequency Distribution of Upselling Possibilities

tial allows the insurer to allocate its salespeople (or other marketing expenditures) in more efficient way. The insurer may rank order its customers in terms of their upselling potential and make additional selling efforts accordingly. In addition, identifying the reasons for selling inefficiency (or upselling potential) will help the insurer to find out how to sell more. As mentioned, selling inefficiency comes from various sources, including the lack of capability or motivation among salespeople. Unfortunately, our data do not allow us to pinpoint the sources of the inefficiency. For this purpose, we may need to conduct a survey of current customers and salespeople. More sophisticated data on marketing expenditures may also help. The benefits of identifying sources of inefficiency would be tremendous in developing marketing strategy. For example, if the lack of competence in salespeople turns out to be the major problem, the insurer should put more emphasis on training its salespeople. On the other hand, the lack of motivation of salespeople may imply the need to change the commission structure.

4. CONCLUSIONS

One-to-one marketing is a marketer's dream. The capability of obtaining a vast amount of customer data with low computing costs makes this dream come true. This paper introduces a methodology for calculating the customer-specific upselling potential for life insurance customers. The proposed model was applied to data concerning 5,000 life insurance customers. Employing a stochastic frontier model, we showed that the selling activity of the firm had an average inefficiency of 28%. The customer-level estimate of selling inefficiency (or upselling potential) guides us to select a group of customers with large upselling potential.

As a side benefit, our model can also be used for customer acquisition. Based on their demographic characteristics such as types of professions and age of insured, we can predict the maximum insurance premium that can be sold for each of the potential customers. As a result, we can rank potential customers in terms of their maximum premiums and decide the order of solicitation efforts. Moreover, these predicted values may be used as target premiums, and their differences from realized premiums could be used to evaluate the performance of salespeople.

Our attempts should be considered to be an initial step toward more scientific customer management for life insurers. Several future research directions can be provided. First, more flexible distributions for the inefficiency u can be assumed. For example, a truncated normal adopted by Stevenson (1980) may be an appropriate choice since our assumed half-normal distribution is a nested distribution into the truncated normal. Second, we only considered customers who have purchased protection insurance from the insurer analyzed. A more sophisticated model may be employed to incorporate the behavior of non-purchasers. The authors are currently working on the model that basically combines a tobit model with a stochastic frontier model for this research direction. Third, it may be managerially useful to identify the reasons for inefficiency. As noted, these may include the lack of motivation or the inexperience of sales agents, the bureaucracy of thesales organization, the inefficient allocation of marketing expenditure, and so on.

Most importantly, it is critical for life insurers to quantify the lifetime value of their customers. Marketing expenditures should be justified only to increase the customer lifetime value. The upselling potential is an important component of the customer lifetime value. However, we may also need to know the expected duration of staying in the firm for each customer. In addition, the cross-selling possibility is another important component to quantify the lifetime value. A few researchers have proposed simple methodologies to compute the customer lifetime value (Blattberg, 1998; Jackson, 1989); however, more sophisticated methodology is called for.

REFERENCES

Aigner, D., Lovell, C., and Schmidt, P. (1977). Formulation and Estimation of Stochastic Frontier Production Function Models. Journal of Econometrics, 6, 21–37.

Aly, H., Grabowski, R., Pasurka, C., and Rangan, N. (1990). Technical, Scale, and Allocative Efficiencies in U. S. Banking: An Empirical Investigation. Review of Economics and Statistics, pp. 211–218.

Bauer, P. (1990). Recent Developments in the Econometric Estimation of Frontiers. Journal of Econometrics, 46, 39–56.

Blattberg, R. (1998, January). Managing the Firm Using Lifetime-Customer Value. Chain Store Age, pp. 46–49.

Crosby, L. and Stephens, N. (1987). Effects of Relationship Marketing on Satisfaction, Retention, and Prices in the Life Insurance Industry. Journal of Marketing Research, 24 (4), 404–411.

Greene, W. (1990). A Gamma Distributed Stochastic Frontier Model. Journal of Econometrics, 46, 101–115.

Jackson, D. (1989, May). Determining a Customer's Lifetime Value. Direct Marketing, 24–30.

Jondrow, J., Lovell, C., Materov, I., and Schmidt, P. (1982). On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model. Journal of Econometrics, 19, 233–238.

Kamakura, W., Ratchford, B., and Agarwal, J. (1988, December). Measuring Market Efficiency and Welfare Loss. Journal of Consumer Research, pp. 298–302.

Mahajan, J. (1991). A Data Envelopment Analytic Model for Assessing the Relative Efficiency of the Selling Function. European Journal of Operations Research, 53, 189–205.

Seiford, L. and Thrall, R. (1990),. Recent Developments in DEA. Journal of Econometrics, 46, 7–38.

Stevenson, R. (1980). Likelihood Functions for Generalized Stochastic Frontier Estimation. Journal of Econometrics, 13, 57–66.