INGENIERIA INDUSTRIAL
UNIVERSIDAD DE CHILE

CENTRO DE FINANZAS

Modeling Pricing Strategies Using Game Theory and Support Vector Machines

**Cristián Bravo,
Nicolás Figueroa y
Richard Weber**

El Centro de Finanzas cuenta con el significativo apoyo del Banco de Crédito e Inversiones BCI

fcfm

Modeling Pricing Strategies Using Game Theory and Support Vector Machines[*][†]

Cristián Bravo, Nicolás Figueroa, and Richard Weber
Department of Industrial Engineering
Universidad de Chile
{cbravo,nicolasf,rweber}@dii.uchile.cl

Final version: 24 April 2010

Abstract

Data Mining is a widely used discipline with methods that are heavily supported by statistical theory. Game theory, instead, develops models with solid economical foundations but with low applicability in companies so far. This work attempts to unify both approaches, presenting a model of price competition in the credit industry. Based on game theory and sustained by the robustness of Support Vector Machines to structurally estimate the model, it takes advantage from each approach to provide strong results and useful information. The model consists of a market-level game that determines the marginal cost, demand, and efficiency of the competitors. Demand is estimated using Support Vector Machines, allowing the inclusion of multiple variables and empowering standard economical estimation through the aggregation of client-level models. The model is being applied by one competitor, which created new business opportunities, such as the strategic chance to aggressively cut prices given the acquired market knowledge.

1 Introduction

Among the diverse decisions taken by companies, pricing is one of the most important. Decision makers do not only have a product's or service's price as a tool to affect demand, but also several marketing actions (e.g. mailings or call

[*]The following is a self-archive version of the paper published at ICDM 2010 conference in Berlin, Germany. The original publication is available at www.springerlink.com, in particular http://www.springerlink.com/content/g728865818w8217g/.

[†]Please cite this paper as follows: C. Bravo, N. Figueroa, and R. Weber. Modeling pricing strategies using game theory and support vector machines. In Petra Pertner, editor, Advances in Data Mining. Applications and Theoretical Aspects, Lecture Notes in Computer Science 6171, pages 323-337, 2010.

centers). The final consumer decision is thus influenced by market prices as well as by the stimuli he or she has been subject to.

The dynamics that these elements define can be modeled by game theory [8] which proposes results based on a solid economical background to understand the actions taken by agents when maximizing their benefit in non-cooperative environments. In companies, however, for more than twenty years data mining has been used to retrieve information from corporative databases, being a powerful tool to extract patterns of customer response that are not easily observable.

As of today, these two approaches (i.e. data mining and game theory) have been used to describe similar phenomena, but with limited interaction between each other. This work attempts to combine these approaches thus exploiting both the strong economical background used by game theory to model the relations that define competitive actions, as well as sophisticated data mining models to extract knowledge from the data companies accumulate.

In this model a customer-level, highly detailed demand estimation is introduced, built from Support Vector Machines that can handle a large number of variables from different sources, in contrast with common economics estimations. This demand is used to empower a market-level model based on game theory that details the situation the companies in the market are in, delivering an integrated picture of customers and competitors alike.

This work is structured as follows. Section 2 presents the game theoretic model used for this problem. In Section 4 the demand model is introduced, followed by technical details on Support Vector Machines (SVMs) which is the main technique utilized. The following section presents results obtained for a financial company. Finally, conclusions are drawn in Section 7. Possible future work is outlined in section 8.

2 Competition as a Game

Prior to the definition of game dynamics presented in this work, three definitions are necessary to fully understand the proposed model.

Definition 1. *A* strategy s_j *of a player* j *corresponds to a complete plan of actions, selected from a set of possible actions* S_j *that determines his or her behavior in any stage of the game. The player may, instead of using a fixed action* s_j*, define a probability distribution for the set* S_j *to determine his or her actions, this probability distribution* p_j *is called a* mixed strategy*.*

Definition 2. *Let* p_j *be the strategies for a set of* J *players in a given game. A* Nash Equilibrium *is a vector* $p^* = (p_1^*, \ldots, p_J^*)$ *containing the strategies of the players such that no player has incentives to change his or her particular strategy. If* $S_j(p)$ *is the payout for player* j*, then a Nash Equilibrium is such that*

$$S_j(p^*) = \max_{p_j} S(p_1^*, \ldots, p_j, \ldots, p_J^*) \quad \forall j \in \{1, \ldots, J\}\,. \tag{1}$$

Definition 3. *A* Perfect Sub-Game Equilibrium *is a refinement of the Nash Equilibrium concept where the state is an equilibrium to the game, and also is an equilibrium to all the sub-games that can be constructed from the original one.*

With these concepts at hand we can now define our game. Studies of competition dynamics are usually limited to a game theoretic framework where the players are the companies in the market under analysis. For this particular approach, Nash - Bertrand specification is useful, where players (companies) compete using prices as strategic variables, a reasonable assumption when quantity is flexible when compared to the different levels of demand [6].

In this context, the Nash - Bertrand equilibrium in a one-stage game has only one stable equilibrium: perfect competition, where each player fixes its price according to his marginal cost. However, Friedman [5] argued that when these games are played for a long (infinite) span of time, then every possible configuration of utilities that falls in an "acceptable" or "rational" range will be a perfect sub-game equilibrium.

The previous result, known as *folk theorem*, has an interesting interpretation: from the game theoretic point of view, companies that compete monthly for a fixed (or stable enough) set of customers fall in strategy configurations that will always result in a new equilibrium. If one agent modifies one of its decision variables then, under the assumption of rationality of the players, the new reached state will be a perfect sub-game equilibrium. Then it is useful to look for models that determine, given that one or more conditions are modified, *which* equilibrium will be obtained.

For this theorem to be applicable, the set of strategies must be non-empty. Rotemberg and Saloner[10] define a set of assumptions that are fulfilled in most markets, including the one in this application, that assure the existence of at least one equilibrium.

We will follow the steps of Sudhir *et al* [13] to define the model. Suppose there are J firms in the market with N_t customers in each period t; the firms must fix prices p_{jt}, marketing actions between L available (given by $x_{jt} \in \{0,1\}^L$) and face a cost vector c_{jt}. Under these conditions, Vilcassim, Kadiyali and Chintagunta [16] postulate that the marketing budget does not influence pricing, because it corresponds to a fixed cost. In this work it is considered that marketing strategies are determined *a priori*. This assumption seems realistic since usually marketing budgets and actions are planned at the beginning of each year whereas prices are fixed on a monthly basis.

Each period, the companies maximize the following expression:

$$\max_{p} \; N_t(p_{jt} - c_{jt})S_{jt}(p, x, \chi) \tag{2}$$

Where S_{jt} is the market share which has as inputs the price vector p, all observable marketing actions x, and the observable market heterogeneity given by $\chi \in \mathbb{R}^{I \times n}$ that is intrinsic to each company's customer database and is observable by the players using their respective databases. This assumption means that future utilities are infinitely discounted, being supported by the fact that even though the firm wishes to maximize its future benefits, managers usually prefer short-term goals, implying decisions such as price determination for a certain month, not long-term price fixing. The objective function to be maximized can be represented as a discounted sum, as done e.g. by Dubé and Manchanda [3]. The approach proposed in our paper simplifies the study and is centered on the determination of demand patterns.

To estimate the different cost functions, a matrix of cost factors C_t will be used as input along with a parameter vector λ_j that will be estimated from the firm's data: $c_{jt} = \lambda_j \cdot C_t + \varepsilon_t$ where ε_t is the error that occurs when using this method. The conditions of first order from (2) lead to the price definition for each firm:

$$p_{jt} = c_{jt} - \frac{S_{jt}}{\partial S_{jt}/\partial p_j} \tag{3}$$

The second term on the right hand side of (3), called *Bertrand margin*, must be adjusted by a parameter to allow deviations from the theoretical equilibrium. The parameter that adjust equation (3) will be named κ_j, and corresponds to a numerical measure of how competitive the market is.

This parameter κ_j is of utmost importance, because it indicates the deviation respect to the equilibrium of each company. κ_j, with values between zero and one, indicates exactly how efficient is each company that is being modeled. Values close to one indicate efficiency (near-optimal behavior) and values close to zero represent the lack of it. It allows to identify the companies that are being inefficient and are subject to aggressive behavior from their competitors. To obtain the final model expression, it is necessary to replace the expression for costs and to include κ_j in (3):

$$p_{jt} = C_t \cdot \lambda_j - \kappa_j \frac{S_{jt}}{\partial S_{jt}/\partial p_j} + \varepsilon_t \tag{4}$$

The prior expression is identical to the one presented in previous works, since it represents the classical solution of Bertrand's competition with deviation assuming variable costs. The specification of the market share S_{jt} is where the present work differs from the usual economical modeling, because demand is modeled based on SVMs from disaggregated data at a customer level.

In general, aggregated data is used to model demand, according to the specification of Dubé *et al.* [2], but this approach does not take into account the real drivers for customer decisions, because the aggregated data usually corresponds to variables that indirectly interpret demand. In this work market share is modeled using the direct effects (prices), the indirect effects (marketing strategies) and the customer characteristics, expanding the spectrum used so far and attaining a buying propensity on a case by case basis.

This approach allows to handle a large number of variables in an efficient manner and also permits to construct a demand function with strong statistical support and generalization power, simultaneously providing a high level of detail.

3 Support Vector Machines

Support Vector Machines, the technique used to model demand, is based on the concepts of statistical learning created by Vapnik and Chervonenkis [14] around 1960. Pairs (x, y) are considered, in which an object $x \in X$ is represented by a set of attributes (the "input space" X) and $y \in \{-1, 1\}$ represents the class of the object. If a function $f : X \to Y$ exists that assigns each element in X to its correct class, the error incurred when approximating f can be measured in two ways. The empirical risk R_{emp} is the error accounted when approximating f from a sample set $M \subseteq X$ and the structural risk R is the error incurred in the whole set X. Modelers would like to minimize R, but only observe R_{emp}.

Statistical learning theory establishes bounds for the structural risk based on the empirical risk and a property of the family of functions used to determine classes, called "VC dimension". Defining the margin between a set and a hyperplane as the minimum distance between the hyperplane and the elements of the set, there is a unique hyperplane that maximizes the margin to each of the classes [15]. Furthermore, VC dimension is decreasing as the margin increases, so maximizing the margin and simultaneously minimizing the empirical risk is equivalent to minimizing the structural (real) risk of performing classification using the function f. These types of functions are called "classifiers based on hyperplanes" and possess functional forms given by:

$$f(x) = sgn\left[(w \cdot x) + b\right], \quad w \in R^n,\, b \in R,\, x \in X \tag{5}$$

A support vector machine is then a hyperplane built from a sample $M \subseteq X$ using an efficient algorithm to train it. The first interesting aspect has to do with the properties of the set X: to build the hyperplane one must possess an inner dot product in the space defined by X, and to maximize the margins is necessary that the elements in M are linearly separable. In order to bypass this the "kernel trick" can be used, that starts by considering a function $\Phi : X \to \Phi(X)$ with $\Phi(X)$, the "feature space", a Hilbert space [14] that possesses a defined dot product. In this particular space the separation of the classes can be done linearly, because for any finite - dimension space there is a higher dimensional space such that a non-linear separation in the input space becomes a linear separation in the feature space.

An interesting property of the previous definition is that, in order to construct an SVM, only the value of the dot product in the feature space is necessary, not the explicit form of $\Phi(x)$, so a function $K : X \times X \to \mathbb{R}$ can be used, where $K(x_1, x_2) = \Phi(x_1) \cdot \Phi(x_2)$. Function K is known as a kernel function, hence the "kernel trick". Unfortunately, not all functions that behave as a dot product are kernels, there is an additional property that must be fulfilled, called the Mercer condition. In short, the condition states that the kernel function must be able to represent the dot product for every point in the space X. For a detailed explanation of these conditions, as well as examples of known kernel functions, the reader is referred to [1].

An SVM is then defined when solving the optimization problem that maximizes the margin between the classes, considering the following quadratic optimization problem:

$$\begin{aligned} &\min_{w,b} \frac{1}{2}\left\|w\right\|^2 + \eta \sum_{j=1}^{M} \epsilon_j \\ s.t. \quad & y_j \left[K(w, x_j) + b\right] \geq 1 - \epsilon_j, \quad j = 1, \ldots, m \\ & \epsilon_j \geq 0, \quad j = 1, \ldots, m \end{aligned} \tag{6}$$

The solution of this problem corresponds to hyperplane defined by the normal vector w and the distance to the origin given by b that maximizes margin and minimizes the error incurred. Slack variables ϵ_j account for the fact that not all problems are linearly separable, so the restrictions are relaxed by using these variables to consider an error in the classification. The objective function must account for both effects - minimum error and maximum margin - at the same time, which is done by including a relative weight η, that allows the modeler to balance both goals. The normal vector obtained by this problem is built from a weighted sum of a set of samples from the set M [12], subset named "support vectors" of the hyperplane f, hence the name of the technique. Since the solution of SVMs is composed of a subset of the original set, the solution of the SVM problem is sparse, which also gives numerical advantages over other types of models.

3.1 One-versus-All SVM

SVMs are by definition binary operators. However, some extensions have been developed that allow for multiclass classification, which are of interest for this paper. In particular, One-versus-All (OVA) [9] method will be used, that has been tested as one of the simpler, yet complete, approach to determine multiclass labels.

The main idea of the OVA approach is to train K SVMs, one for each class defined by object j's label $y_j \in \{1, ..., K\}$. In this case, the continuous output of each SVM is used (function (5) without the sign function), that represents the distance to hyperplanes with signs associated to which side of the hyperplane the example is in. The classifier is then given by $f(x) = (f_1(x), \ldots, f_K(x))$ and the class y is determined by the index of the maximum value in vector $f(x)$:

$$y_i = \begin{cases} 1 & i = arg\max_k\{f_k\} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

4 Customer-Level Demand and Aggregated Model

For the data mining twist to model (4), the market share S_{jt} must be defined. We propose to use SVMs for this task, which offers two main advantages:

1. Allows the use of atomic data: Econometric estimations usually employ aggregated data and general indicators as regressors. It is in the interest of both researchers and end-users to exploit the large quantity of data that exist in today's companies' databases. To model this phenomenon at a customer level is therefore of high relevance.

2. Allows to generalize demand: Data mining approaches model demand without the assumptions about the capacity to observe customers' characteristics. Instead, they are based on the patterns that each customer leaves about his or her behavior in the company databases. This allows to consider empiric demands based on the customers' actions (atomic model) and econometric models of the behavior of the firm (aggregated model), hence empowering both approaches.

3. Possesses methodological advantages: SVMs are a powerful mathematical model to approximate almost any type of phenomena. In particular, the problem (6) does not possess local minima, increasing the confidence of the solution.

The use of SVMs instead of Support Vector Regression (SVR) is justified because the demand must be estimated in a per-customer basis, since demand is not modeled by considering a continuous function of aggregated results, but by a set of different customers taking separated decisions.

To develop the final model, consider a database with customer attributes χ over T periods of time and a matrix of marketing actions $X \in \{0,1\}^{L\times T}$ directed to customers. Finally, one must assign labels indicating whether the customer chose the company, its competitors or neither (no-sale) for each one of the periods. Then, each customer can be represented by a vector of attributes consisting of their personal characteristics ($\chi_i \in \mathbb{R}^n$), the prices he or she observed at that particular time ($p = (p_1, \dots, p_J) \in \mathbb{R}_+^J$) and the marketing strategies realized to the set of customers ($x_t \in \{0,1\}^L$): (x_i, p, χ_t). This data plus the labels associated to the firms ($y_i \in \{-1,1\}^{J+1}$) allow to train an SVM in OVA approach (section 3.1) in order to obtain, for each firm J, the predicted amount of customers that chose it in each period t and also the number of customers that do not choose any firm. There are then $J+1$ SVMs that model the tendency to buy (or not to buy) for each customer.

$$f_j(x_i, p, \chi_t) = sgn\left[k\left((w_j^x, w_j^p, w_j^\chi), (x_i, p, \chi_t)\right) + b_j\right], \quad j = \{1, \dots, J\}, i = \{1, \dots, N\} \tag{8}$$

The market share for a given period is the known market share from the previous period ($S_{(j,t-1)}$), adjusted by the new number of customers in period t minus the number of customers that are no longer in the captivity of the company at the end of period $t-1$ (e_{t-1}), plus the fraction of customers that are selected by the SVMs as customers of the company in period t:

$$
\begin{aligned}
S_{j,t}(p,x,\chi) &= \hat{S_{j,(t-1)}} + \frac{\sum_{i\in N_t} f_j(p,x_i,\chi_t)}{N_t} \\
\hat{S_{j,(t-1)}} &= \frac{S_{j,(t-1)} \cdot N_{t-1} - e_{t-1}}{N_t}
\end{aligned} \tag{9}
$$

Equation 9 is a demand function that includes relevant customer-describing variables and also prices and information about strategic actions (e.g. the prices and marketing actions).

It is now necessary to estimate the derivative of (9) $\partial S_j(p,x,\chi)/\partial p_j$, which will be numerically estimated obtaining the number of customers that change their choice due to a price change. In particular, the secant method [4] will be used for the derivation, approximating it through moving the price in a small quantity Δp_j.

$$
\begin{aligned}
\frac{\partial S_j(p,x,\chi)}{\partial p_j} &\approx \frac{1}{2}[\frac{S_j(p^+,x,\chi) - S_j(p,x,\chi)}{\Delta p_j} \\
&\quad + \frac{S_j(p,x,\chi) - S_j(p^-,x,\chi)}{\Delta p_j}] \\
&= \frac{S_j(p^+,x,\chi) - S_j(p^-,x,\chi)}{2\Delta p_j} \\
&= \frac{\sum_{i=1}^{N_t} [f(p^+,x_i,\chi_t) - f(p^-,x_i,\chi_t)]}{2\Delta p_j} \\
\text{with } p^+ &= (p_1,\ldots,p_j+\Delta p_j,\ldots,p_J), \\
p^- &= (p_1,\ldots,p_j-\Delta p_j,\ldots,p_J)
\end{aligned} \tag{10}
$$

Now, replacing (9) and (10) in (4) the final model is obtained:

$$
p_{jt} = \lambda_j \cdot C_t + \kappa_j \; \frac{\hat{S_{j,(t-1)}} + \frac{\sum_{i=1}^{N_t} f(p,\, x_i,\, \chi_t)}{N}}{\frac{\sum_{i=1}^{N_t} [f(p^+,\, x_i,\, \chi_t) - f(p^-,\, x_i,\, \chi_t)]}{2\Delta p_j}} + \epsilon_j \tag{11}
$$

An interesting feature of (11) is, even though the full expression is non-linear, the final estimation is done linearly. To train this model, we propose the following three-step procedure:

1. Construction of integrated database: We need three different kinds of data, an internal database to construct the matrix χ with customer data, information about the competitors, for example whether they have performed a commercial or other visible activity on the customers, stored in matrix X and finally the observed prices must be collected to construct the price matrix P. These prices are usually available to public.

2. Train SVMs on the integrated database: It is necessary to determine the different expected market shares for the competing companies. Expression

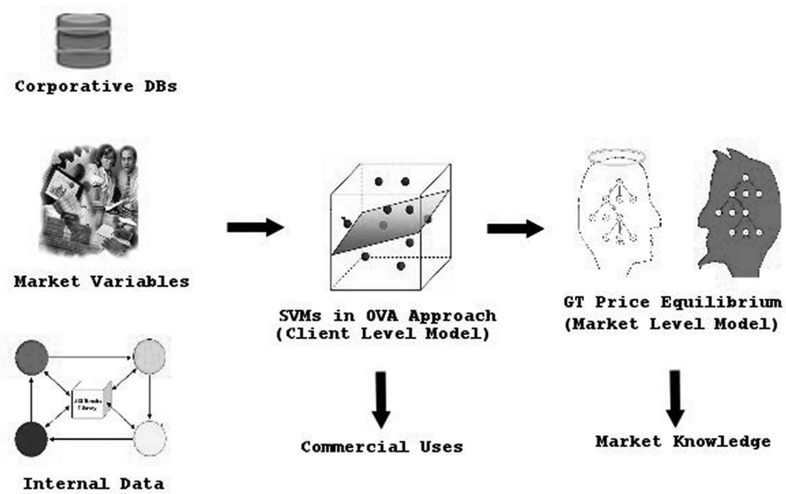

Figure 1: Model diagram.

(9) allows a customer to choose more than one company at the same time, an event that actually can happen in most markets. This case can, however, be extracted from the database or considered as participation anyway, although this is a debatable step.

3. Estimate the parameters in (11): To estimate the resulting linear regression, cost regressors are needed, and can be found through global indicators. In this case, indicators such as the Producer Price Index (PPI) from Chile's National Institute of Statistics and other widely available global indicators were used to model the marginal cost for each company.

To adjust the final model and considering that the resulting problem is linear, least squares or a maximum likelihood method such as the generalized method of moments can be used. The input data are the monthly cost regressors, the observed shares and the estimation of the derivative from equation (10) for each period of time measured. The schematic application of the proposed model is illustrated in Figure 1.

5 Benchmark Model

In order to check the performance of the proposed approached, an artificial neural network (ANN) [11] will be used as a benchmark of the performance when estimating demands. Neural networks are composed of neurons (nodes) that are organized in layers. Each neuron receives as input all the outputs from the previous layer, and applies a specific weight and a transfer function to this input, to then pass this result to the neurons in the next layer. The first layer (input layer) only consists of weights and each neuron is associated to one input variable of the dataset. The final layers is called output layer, and presents the final result in any specified format, while the layers between the input and output layers are called hidden layers. The configuration of the neural network consists of:

- Hidden Layers: The number of hidden layers, and the number of neurons in each hidden layer must be decided. It has been stated [7] that only one hidden layer is necessary to approximate any bounded function, while two are necessary to approximate any unbounded one. In this experiment, since a probability will be estimated, only one hidden layer will be used.
- Output Layer: The output layer must have an output function that presents the results in a logical format for the problem being modeled. In this case, a softmax function must be used, given by:

$$p_i(x) = \frac{exp(\beta_j \cdot x)}{\sum_{j'} exp(\beta_{j'} \cdot x)} \tag{12}$$

  with β_i a vector of output parameters associated to each class j, and x the vector of variables.
- Transfer functions: The functions that determine the input of each layer must be decided. In this case, linear functions were used, representing the final output as a multinomial logistic function, typically used for this type of problems.
- Training parameters: Depending on the software used to train a neural network, several parameters must be defined to determine convergence. The number of epochs (times the neuron is presented with the data) is one common attribute.

In order to obtain the number of epochs and the number of neurons of the hidden layer, a grid-search process was conducted, as described in section 6.

To perform a comparison on the performance to approximate demand, the results of this ANN are aggregated by simply adding the probability that each customer chooses any of the companies, adjusted in the same way as equation 9. This gives the expectation of the demand for the period, as desired.

6 Experimental Results

To apply this model in a real-life situation, data from a well-known local company was available to the authors. The company offers a complete line of financial services, and between them are loans that are directly discounted from the customer's income. This database has some advantages that make it perfect for this particular problem:

- The products placed by any competitor are known: The company in study has access to the other companies that also place products to a given customer, because the discount of the loan's installments is done through the company. This allows knowing in real-time when the customer has chosen the company's competitors and the company in study, fulfilling the most restrictive assumption in the model.

- The market is highly concentrated: 93% of total loans are concentrated in three companies, with the rest offering the remaining 7%, allowing them to behave as an oligopoly and possessing some market influence. The companies with market share of 7% will be referred to as one single company (company O) for simplicity of study.

The market is then formed by four companies (E, A, H and O) that struggle to acquire N customers, each one of them characterized by the variables described in section 2. In particular, 80 variables were studied, characterizing 100,000 different customers over 18 months. The variables came from different sources:

- Internal Databases: possessing demographic data, the income for the customers and the shares for the companies.

- External Databases: Information about prices and cost regressors, which came from Chilean Central Bank, the National Institute of Statistics (INE) and the organism that supervises the companies in the market, called Superintendence of Social Security (SUSESO).

- Generated Variables: Some indicators were built from global income, debt, specific per-company debt and so on, in order to improve the results of the models and to attempt to discover new relationships between the variables.

The variables were selected utilizing a complete study that maximizes the contact of the modeler with the variables and so the extracted knowledge. This process begins by eliminating variables highly concentrated in one single value or with a high rate of missing values, then it continues with a process of univariate feature selection, where variables that possess no univariate discriminating capacity are eliminated. The capacity is obtained from simple and widely known χ^2 and K-S tests, supposing that independence from the objective variable in a univariate way also implies independence from that objective in a multivariate environment. Finally, over-adjusted classification trees are built under the hypothesis that if the variable does not appear in any levels of the tree then it would likely not appear in a different multivariate model. This procedure has been tested previously by our team with good results.

Finally, a set of 20 discriminating variables was achieved, being the input for the data mining models. The available database consists of approximately 100,000 registers and is highly unbalanced, considering that company E has a market share of 50%.

In order to perform the experiments, a search must be performed for optimal parameter setting. 20% of the database was reserved for such task and to ensure the model was able to cover all classes two precautions were applied:

- The samples were artificially balanced using an adjustment parameter that grants a value of one to all the elements that are in the set associated with the class with less cases and a value of $\frac{\text{cases-minor-class}}{\text{cases-mayor-class}}$ to the elements of the class with more registers associated. This is done for each SVM, considering they are in an OVA scheme.

- An *ad-hoc* error function was used to measure performance of the particular parameter configuration, that balances the errors in each class (e_m). This function multiplies the errors for each class, so only solutions that represent all classes are considered. Considering $e_c, c \in \{E, A, H, O, NB\}$ the errors per class, the error function is given by (13). The reader should note that the "No-Buy" class is included ("NB") that consists of all the customers that choose not to buy in a particular period in any company.

$$e_m = \prod_{c \in \{\text{E,A,H,O,NB}\}} e_c \tag{13}$$

- Finally the error is averaged over 3-sample cross-validation, with the error for each parameter set being the average of the errors from equation (13).

With these steps the optimal parameter setting was found and used to train the model in the remaining 80% of the sample.

The elements of this sample were divided in five different subsets to perform cross-validation once again, to reduce the sample error, also keeping additionally 20% off the training for testing. Table 1 displays the results for each company, consisting on the per-class error, which is close to 10% in average, this being a satisfactory measure. The benchmark model performs well below this index, with errors around 15%. Since SVMs allow to fine tune each class performance, the model offers more chances to improve the result obtained, as is reflected in this experiment.

Table 1: Class Error for each company.

Class	Class Error (SVM)	Class Error (ANN)
E	21,53%	22,75%
A	4,37%	13,78%
H	1,89%	10,72%
O	1,91%	8,98%
No Buy	21,61%	43,78%

Table 1 showed the results on an individual customer level, but in this work we require the estimation of aggregated demand, which is not at all common in data mining models, that usually are at an atomic level. The results were aggregated using equations (9) and (10), with results close to an 80% of accuracy (Table 2), which also is highly satisfactory and supports the use of data mining models for aggregated estimations. The benchmark model once again is outperformed by the SVM results, as is expected, since the results from the client level model should be somewhat transferred to the estimation of demand functions, also, the standard deviation of the model is higher, which indicates that the SVM is capable of capturing a wider range of different patterns.

Table 2: Accuracy of aggregated results.

Company	SVM Error	ANN Error
E	16.40% ± 11.50%	43.22% ± 19.38%
A	19.60% ± 5.86%	30.38% ± 17.77%
H	29.00% ± 9.80%	23.99% ± 15.16%
O	19.90% ± 8.38%	26.10% ± 15.53%

The final step is to estimate the game theoretic model from equation (11). The cost factors used consider the cost of life (Consumer Price Index, IPC), the maximum interest rates allowed, and price indexes to producers, salary indexes and others from the sources previously indicated. The dataset consists on weekly data and includes the previous variables, plus the prices (target) charged by each company in that week, and the estimated demands divided by the derivative. The regression was run using the software package SPSS and feature selection was performed using backward and forward selection, conciliating both approaches by keeping the feature combination that performed best in the sample. The results are highly satisfying (table 3) with over 95% of accuracy in average and low standard error, which once again supports the use of this kind of models to predict price changes.

The efficiency coefficient κ_j is of particular interest, because it represents how efficient company j is when fixing its price, a well-known result in game theory. Companies with higher market shares are more efficient, establishing that the most important drivers of price changes are changes in demand and competition. Companies that are less efficient, on the other hand, present smaller values, which indicate that their main drivers to fix prices are their observed costs and their lack of interest (or capacity) to take demand into account. This result is really interesting because it establishes a quantitative measure of the different companies' market position in a given market and goes beyond the results each single approach - data mining and game theory - could provide.

7 Conclusions

The model introduced in this work provides a novel tool to find market equilibria and to determine the expected market share when modifying strategic variables.

Table 3: Results for regression on prices.

Company	R	**Adjusted** R^2	**Std. Error**	κ_j
E	0.973	0.796	0.016	0.684
A	0.919	0.758	0.054	0.757
H	0.961	0.903	0.031	0.096
O	0.994	0.981	0.02	0

Moreover, demand is modeled in terms of directly measurable variables such as price, and in terms of indirect variables such as marketing strategies that the company employs and in terms of the customers' characteristics. This provides a more profound knowledge regarding the customers' attitude towards the different companies. The model offers an integrated view of the elements that define the respective market, integrating the available knowledge, providing a major advantage over the use of a single technique.

In general, the use of models based on successions of games represents an effective alternative to measure the effects of changes in the market's competition conditions. This way, a theoretical limitation (the existence of infinite market equilibriums) is transformed into a useful tool, granting the possibility to determine this new equilibrium in terms of modeling past behavior.

Currently, the so-called "indirect" effects consume a great deal of hours and resources spent in a company, so they cannot be neglected. The connection with data mining allows overcoming this challenge, explaining complex phenomena by obtaining the statistical patterns present in the large quantity of data that companies are storing. This way the reasons that drive a person to prefer a determined company can be studied in detail.

The main limitation of the presented model is the data that needs to be collected, in particular the data referred to competitors' product placement. A workaround to this limitation consists in collecting this data through surveying customers of the company that produces the study's database.

The use of data mining models to estimate aggregated demand is another interesting contribution of this paper. A simple methodology is introduced to aggregate the obtained atomic results that gives very good results. The main reason a researcher would like to utilize this type of demand model is that data mining allows an efficient handling of large quantity of variables, so it is useful when compared to classical demand estimation models that cannot do so.

The model gives useful and applicable results that can be utilized in day-to-day decisions. In particular, the work from this paper was used to design a campaign to acquire competitors' customers, which had a high positive response rate and allowed to increase the market share of company E, a fact that gives even more credibility to the application of such models in companies.

Considering all these elements, the combination of data mining with game theory provides an interesting research field that has received a lot of attention from the community in recent years, and from which a great number of new models are expected. Future studies will generate promising results in all aspects where both a large number of data and interaction between agents are present. An integrated vision that takes into account, at the same time, consumers and companies has been introduced in this paper. This integrated vision allows interpreting the relationships of all the participants and giving a full spectrum of the market.

8 Future Work

Two separate lines of work have been developed from this paper. The first consists of improving the presented model using analytical techniques to avoid the numerical estimations and to improve the model results. The second one, still under development, is to use the techniques here presented to improved credit scoring models, modeling the loan granting process as a game and then applying credit scoring techniques.

Acknowledgments

The first author would like to acknowledge CONICYT for the grants that finance this research, and to C. Mora for her aid in editing this paper. Also, support from the Millennium Science Institute on Complex Engineering Systems (ICM: P-05-004-F, CONICYT: FBO16) (`www.sistemasdeingenieria.cl`) and from the Ph.D. in Engineering Systems (for the first author) is greatly acknowledged.

References

[1] N. Cristiannini and J. Shawe-Taylor. *An Introduction To Support Vector Machines And Other Kernel-Based Methods*. Cambridge University Press, 2003.

[2] J. Dubé, P. Chintagunta, B. Bronnenberg, R. Goettler, A. Petrin, P. Seetharaman, R. Thomadsen, and Y. Zhao. Structural applications of the discrete choice model. *Marketing Letters*, 13(3):207–220, 2002.

[3] J. Dubé and P. Manchanda. Differences in dynamic brand competition across markets: An empirical analysis. *Marketing Science*, 24(1):81–95, 2005.

[4] J. D. Faires and R. Burden. *Numerical Analysis*. Thomson Brooks/Cole, 1986.

[5] J. Friedman. A non-cooperative equilibrium for supergames. *Review of Economic Studies*, 38:112, 1971.

[6] D. Fudenberg and J. Tirole. *Game Theory*. MIT Press, 1991.

[7] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359 – 366, 1989.

[8] J. Nash. Non-cooperative games. *Annals of Mathematics*, 54:286–295, 1951.

[9] Ryan Rifkin and Aldebaro Klautau. In defense of one-vs-all classification. *The Journal of Machine Learning Research*, 5:101 – 141, 2004.

[10] I. Rotemberg and G. Saloner. A super-game theoretic model of business cycles and price wars during booms. *American Economic Review*, 76(3):390–407., 1986.

[11] D. W. Ruck, S. K. Rogers, M. Kabrisky, M. E. Oxley, and B. W. Suter. The multilayer perceptron as an approximation to a bayes optimal discriminant function. *IEEE Transactions on Neural Networks*, 1(4):296298, 1990.

[12] B. Schlkopf. Statistical learning and kernel methods. Technical Report Msr-Tr-23-2000, Microsoft Research, 2000.

[13] K. Sudhir, P. Chintagunta, and V. Kadiyali. Time varying competition. *Marketing Science*, 24(1):96–110, 2005.

[14] V. Vapnik. *Statistical Learning Theory*. John Wiley & Sons Inc., 1998.

[15] V. Vapnik and A Lerner. Pattern recognition using generalized portrait method. *Automatization And Remote Control*, 24:774 780, 1963.

[16] N. Vilcassim, V. Kadiyali, and P Chintagunta. Investigating dynamic multifirm market interaction in price and advertising. *Management Science*, 45(4):499–518, 1999.