Estimating multiplicative competitive interaction model using kernel machine technique

Jooyong Shim[1] · Malsuk Kim[2] · Hyejung Park[3]

[1]Department of Data Science, Institute of Statistical Information, Inje University
[2]Division of Computer Technology, Yeungnam College of Science & Technology
[3]College of Liberal Art, Daegu University

Abstract

We propose a novel way of forecasting the market shares of several brands simultaneously in a multiplicative competitive interaction model, which uses kernel regression technique incorporated with kernel machine technique applied in support vector machines and other machine learning techniques. Traditionally, the estimations of the market share attraction model are performed via a maximum likelihood estimation procedure under the assumption that the data are drawn from a normal distribution. The proposed method is shown to be a good candidate for forecasting method of the market share attraction model when normal distribution is not assumed. We apply the proposed method to forecast the market shares of 4 Korean car brands simultaneously and represent better performances than maximum likelihood estimation procedure.

Keywords: Kernel machine technique, market share attraction model, maximum likelihood estimation, multiplicative competitive interaction model.

1. Introduction

Multiplicative competitive interaction (MCI) model is a popular tool for analyzing market competitive structures (Cooper and Nakanishi, 1988; Fok and Franses, 2004) and is typically applied for simultaneously forecasting the market shares of several brands within a given product category. The model helps to evaluate both the effect of marketing-mix variables on brands' performances and the effect of an individual brand's own efforts while conditioning on competitors' reactions. Kumar (1994) summarizes and extends this important stream of research, and recent research suggests the MCI model produces market share forecasting superior to those from simpler linear or multiplicative models (Kumar and Heath, 1990). A detailed analysis of the model can be found in Fok *et al.* (2002). A common way to solve estimation problems of MCI model is a maximum likelihood (ML) estimation, which is

[1] Adjunct professor, Department of Data Science, Institute of Statistical Information, Inje University, Obang-Dong, Kimhae 621-749, Korea.
[2] Adjunct professor, Division of Computer Technology, Yeungnam College of Science & Technology, Daegu 705-703, Korea.
[3] Corresponding author: Invitation professor, College of Liberal Art, Daegu University, Gyeongbuk 712-714, Korea. E-mail: hyjpark@daegu.ac.kr

optimal in cases the response variables are drawn from a conditional normal distribution. In cases of non-normality, the least squares techniques could lead to mismatched solutions for some densities (Vapnik, 1982). In cases like this, to obtain the improved estimation Gruca and Klemz (1998) applied the neural network instead of least squares estimation procedure under the assumption of independence of error terms. Park (2009) applied the least squares support vector machine (Suykes and Vanderwalle, 1999) for the improved estimation.

The proposed method is based on the kernel machine technique applied in support vector machine (Vapnik, 1995) and other machine learning techniques, which has been applied to the regression problems of various data types (Kim *et al.*, 2008; Shim and Seok, 2008; Seok, 2010; Hwang, 2010a; Hwang, 2010b; Shim, 2011). The proposed method uses a penalty term, which provides the robustness to outliers and the avoidance of overfitting, which is shown to be a good candidate for forecasting the market share attraction model. We apply the proposed method to forecast the market shares of 4 Korean car brands simultaneously and represent better performances than ML estimation procedure.

The rest of paper is organized as follows. In section 2 we briefly review the differential effect MCI model. In section 3, we propose the nonlinear differential effect MCI model based on the kernel machine technique. In Section 4 we perform the numerical studies with real data set. In Section 5 we give the conclusions.

2. Differential effect MCI model

We consider the MCI model of the form,

$$\log A_{ij} = \mu_{0i} + \boldsymbol{x}'_{ij}\boldsymbol{\beta_i} + \epsilon_{ij}, \ j = 1, \cdots T, \ i = 1, \cdots, I, \tag{2.1}$$

where A_{ij} is the attraction of ith brand at time t, $\boldsymbol{x}_{ij}$ is a logarithm of input vector of size $p \times 1$, $\boldsymbol{\beta}_i$ is an unknown parameter vector of size $p \times 1$, error terms at each time $t = 1, \cdots, T$, $(\epsilon_{1t}, \epsilon_{2t}, \cdots, \epsilon_{It})$, are assumed to follow a normal distribution with zero means and covariance vector $\Sigma = \{\sigma_{ij}\}_{i,j=1}^{I}$.

Put $y_{ij} = \log A_{ij} - \log A_{Ij}$ where I-th brand is the reference, and then we have the differential effect MCI model as follows:

$$y_{ij} = \mu_i + \boldsymbol{x}'_{ij}\boldsymbol{\beta}_i - \boldsymbol{x}'_{Ij}\boldsymbol{\beta}_I + e_{ij}, \ j = 1, \cdots T, \ i = 1, \cdots, I - 1, \tag{2.2}$$

where $\mu_i = \mu_{0i} - \mu_{0I}$ and $e_{ij} = \epsilon_{ij} - \epsilon_{Ij} \sim N(0, \sigma_{ii} + \sigma_{II} - 2\sigma_{iI})$. The equation (2.2) can be written in a vector-matrix form as follows:

$$\boldsymbol{y} = U\boldsymbol{\mu} + X\boldsymbol{\beta} - H\boldsymbol{x}_I\boldsymbol{\beta}_I + \boldsymbol{e} \tag{2.3}$$

where $\boldsymbol{y} = (y_{11}, y_{12}, \cdots, y_{1T}, \cdots, y_{I-1,1}, \cdots, y_{I-1,T})'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \cdots, \boldsymbol{\beta}'_{I-1})'$ is a $(I-1)p \times 1$ vector, U is a $(I-1)T \times (I-1)p$ block diagonal matrix of $\mathbf{1}_{I-1}$, which is used for $U\boldsymbol{\mu} = (\mu_1, \cdots, \mu_1, \cdots, \mu_{I-1}, \cdots, \mu_{I-1})'$, $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_1, \cdots, \mu_{I-1}, \cdots, \mu_{I-1})'$, X is a $(I-1)T \times (I-1)p$ block diagonal matrix of $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT})'$ $(i = 1, \cdots, I-1)$ which is a $T \times p$ matrix and used for $X\boldsymbol{\beta} = (\boldsymbol{x}_1\boldsymbol{\beta}_1, \cdots, \boldsymbol{x}_{I-1}\boldsymbol{\beta}_{I-1})'$. X_I is a $(I-1)T \times p$ matrix of $(I-1)$ $\boldsymbol{x}_I$'s, which is used for $X_I\boldsymbol{\beta}_I = H\boldsymbol{x}_I\boldsymbol{\beta}_I = (\boldsymbol{x}_I\boldsymbol{\beta}_I, \cdots, \boldsymbol{x}_I\boldsymbol{\beta}_I)'$, H is a $(I-1)T \times T$ block diagonal matrix of I_T.

The vector of error terms $\boldsymbol{e}$ follows a normal distribution with zero means and covariance vector W, where W is a $(I-1)T \times (I-1)T$ matrix such that

$$W((i-1)T+1:iT,(j-1)T+1:jT) = S_{ij}\mathrm{I}_T, \ i,j = 1,2,\cdots,I-1,$$

where $S = S_1\Sigma S_1'$, $S_1 = [I_{I-1}, \mathbf{0}_{(I-1)\times 1}] - [\mathbf{0}_{I-1\times I-1}, \mathbf{1}_{(I-1)\times 1}]$. From the fact that the error term $\boldsymbol{e}$ follows a normal distribution, we have the negative log likelihood function,

$$L(\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\beta}_I|\boldsymbol{y},\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{y}-U\boldsymbol{\mu}-X\boldsymbol{\beta}+X_I\boldsymbol{\beta}_I)W^{-1}(\boldsymbol{y}-U\boldsymbol{\mu}-X\boldsymbol{\beta}+X_I\boldsymbol{\beta}_I).$$

From the partial differentiation, we can get the followings:

$$\frac{\partial L}{\partial \boldsymbol{\mu}} = 0 \quad \rightarrow \quad U'W^{-1}(\boldsymbol{y}-U\boldsymbol{\mu}-X\boldsymbol{\beta}+X_I\boldsymbol{\beta}_I) = \mathbf{0}$$

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = 0 \quad \rightarrow \quad X'W^{-1}(\boldsymbol{y}-U\boldsymbol{\mu}-X\boldsymbol{\beta}+X_I\boldsymbol{\beta}_I) = \mathbf{0}$$

$$\frac{\partial L}{\partial \boldsymbol{\beta}_I} = 0 \quad \rightarrow \quad X_I'W^{-1}(\boldsymbol{y}-U\boldsymbol{\mu}-X\boldsymbol{\beta}+X_I\boldsymbol{\beta}_I) = \mathbf{0}.$$

The estimates of $(\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\beta}_I,\boldsymbol{\alpha})$ can be obtained from the linear system:

$$\begin{pmatrix} U'W^{-1}U, & U'W^{-1}X, & -U'W^{-1}X_I \\ X'W^{-1}U, & X'W^{-1}X, & -X'W^{-1}X_I \\ X_I'W^{-1}U, & X_I'W^{-1}X, & -X_I'W^{-1}X_I \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\beta} \\ \boldsymbol{\beta}_I \end{pmatrix} = \begin{pmatrix} U'W^{-1} \\ X'W^{-1} \\ X_I'W^{-1} \end{pmatrix}\boldsymbol{y},$$

which leads to the generalized least squares (GLS) estimates,

$$\begin{pmatrix} \widehat{\boldsymbol{\mu}} \\ \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\beta}}_I \end{pmatrix} = \begin{pmatrix} U'W^{-1}U, & U'W^{-1}X, & -U'W^{-1}X_I \\ X'W^{-1}U, & X'W^{-1}X, & -X'W^{-1}X_I \\ X_I'W^{-1}U, & X_I'W^{-1}X, & -X_I'W^{-1}X_I \end{pmatrix}^{-1} \begin{pmatrix} U'W^{-1} \\ X'W^{-1} \\ X_I'W^{-1} \end{pmatrix}\boldsymbol{y}$$

If error terms, e_{it}'s, are assumed to be independent, the ordinary least squares (OLS) estimates are obtained as follows:

$$\begin{pmatrix} \widehat{\boldsymbol{\mu}} \\ \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\beta}}_I \end{pmatrix} = \begin{pmatrix} U'U, & U'X, & -U'X_I \\ X'U, & X'X, & -X'X_I \\ X_I'U, & X_I'X, & -X_I'X_I \end{pmatrix}^{-1} \begin{pmatrix} U' \\ X' \\ X_I' \end{pmatrix}\boldsymbol{y}. \tag{2.4}$$

For given $\boldsymbol{x}_{it}$ and $\boldsymbol{x}_{It}$, the forecasted market shares are obtained as follows:

$$\widehat{M}_{It} = \frac{1}{1+\sum\limits_{i=1}^{I-1}\exp(\widehat{y}_{it})} \text{ and } \widehat{M}_{it} = \frac{\exp(\widehat{y}(\boldsymbol{x}_{it}))}{1+\sum\limits_{i=1}^{I-1}\exp(\widehat{y}_{it})}, \ i = 1,\cdots,I-1, \tag{2.5}$$

where $\widehat{y}_{it} = \widehat{\mu}_i + \boldsymbol{x}_{it}'\widehat{\boldsymbol{\beta}}_i - \boldsymbol{x}_{It}'\widehat{\boldsymbol{\beta}}_I$ is the estimated logarithm of ratio of market share of the brand i to market share of the brand I.

3. Nonlinear differential effect MCI model

In this section we consider the MCI model of the form,

$$\log A_{ij} = \mu_{0i} + \phi(\boldsymbol{x}_{ij})'\boldsymbol{w} + \epsilon_{ij}, \ j = 1, \cdots T, \ i = 1, \cdots, I \tag{3.1}$$

where $\phi(\boldsymbol{x}_{ij})$ is an unknown nonlinear feature mapping function of $\boldsymbol{x}_{ij}$ (a logarithm of input vector) of size $d_f \times 1$ vector, $\boldsymbol{w}$ is an unknown parameter vector of size $d_f \times 1$, error terms at each time $t = 1, \cdots, T$, $(e_{1t}, e_{2t}, \cdots, e_{It})$, are assumed to follow a normal distribution with zero means and covariance vector $\Sigma = \{\sigma_{ij}\}_{i,j=1}^I$.

Put $y_{ij} = \log A_{ij} - \log A_{Ij}$ where I-th brand is the reference, we have the nonlinear differential effect MCI model as follows:

$$y_{ij} = \mu_i + (\phi(\boldsymbol{x}_{ij}) - \phi(\boldsymbol{x}_{Ij}))\boldsymbol{w} + e_{ij}, \ j = 1, \cdots T, \ i = 1, \cdots, I-1, \tag{3.2}$$

where $\mu_i = \mu_{0i} - \mu_{0I}$ and $e_{ij} = \epsilon_{ij} - \epsilon_{Ij} \sim N(0, \sigma_{ii} + \sigma_{II} - 2\sigma_{iI})$.

The equation (3.2) can written as

$$\boldsymbol{y} = U\boldsymbol{\mu} + (\phi(\boldsymbol{x}) - H\phi(\boldsymbol{x}_I))\boldsymbol{w} + \boldsymbol{e}, \tag{3.3}$$

where $\boldsymbol{y} = (y_{11}, y_{12}, \cdots, y_{I-1,T})'$, U is a $(I-1)T \times (I-1)$ block diagonal matrix of $\boldsymbol{1}_{I-1}$, which is used for $U\boldsymbol{\mu} = (\mu_1, \cdots, \mu_1, \cdots, \mu_{I-1}, \cdots, \mu_{I-1})'$, $\phi(\boldsymbol{x})$ is a $(I-1)T \times M$ matrix of $\phi(\boldsymbol{x}_{ij})$'s and $\phi(\boldsymbol{x}_I)$ is a $T \times M$ matrix of $\phi(\boldsymbol{x}_{It})$'s. $\boldsymbol{e}$ follows a normal distribution with zero mean vector and covariance vector W such that $W((i-1)T+1 : iT, (j-1)T+1 : jT) = S_{ij}I_T$, $i, j = 1, 2, \cdots, I-1$. Since we assume that

$$\begin{pmatrix} e_{11} \\ \vdots \\ e_{I-1,1} \end{pmatrix}, \begin{pmatrix} e_{12} \\ \vdots \\ e_{I-1,2} \end{pmatrix}, \cdots, \begin{pmatrix} e_{1T} \\ \vdots \\ e_{I-1,T} \end{pmatrix}$$

have the same covariance matrix such that

$$S = \begin{pmatrix} \sigma_{11}+\sigma_{II}-2\sigma_{1I}, & \sigma_{12}+\sigma_{II}-\sigma_{1I}-\sigma_{I2}, & \cdots & \sigma_{1I-1}+\sigma_{II}-\sigma_{1I}-\sigma_{I,I-1} \\ \sigma_{21}+\sigma_{II}-\sigma_{2I}-\sigma_{I1}, & \sigma_{22}+\sigma_{II}-2\sigma_{2I}, & \cdots & \sigma_{2,I-1}+\sigma_{II}-\sigma_{2I}-\sigma_{I,I-1} \\ & & \vdots & \\ \sigma_{I-1,1}+\sigma_{II}-\sigma_{I-1,I}-\sigma_{I1}, & \sigma_{I-1,2}+\sigma_{II}-\sigma_{I-1,I}-\sigma_{I2}, & \cdots & \sigma_{I-1,I-1}+\sigma_{II}-2\sigma_{I-1,I} \end{pmatrix}.$$

From the fact that the error term $\boldsymbol{e}$ follows a normal distribution, we consider the penalized negative log likelihood function,

$$L(\boldsymbol{\mu}, \boldsymbol{w}|\boldsymbol{y}, \boldsymbol{x}) = \frac{1}{2}(\boldsymbol{y} - U\boldsymbol{\mu} - (\phi(\boldsymbol{x}) - H\phi(\boldsymbol{x}_I))\boldsymbol{w})'W^{-1}(\boldsymbol{y} - U\boldsymbol{\mu} - (\phi(\boldsymbol{x}) - H\phi(\boldsymbol{x}_I))\boldsymbol{w}) + \frac{\lambda}{2}\boldsymbol{w}'\boldsymbol{w} \tag{3.4}$$

where λ is a nonnegative constant which control the trade-off between the goodness-of-fit on the data and $||\boldsymbol{w}||^2$ and H is a $(I-1)T \times T$ matrix consisted of I_T.

Here the feature mapping function $\phi(\cdot) : R^d \to R^{d_f}$ maps the input space to the higher dimensional feature space where the dimension d_f is defined in an implicit way. It is known

that $\phi(\boldsymbol{x}_i)'\phi(\boldsymbol{x}_j) = K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ which is obtained from the application of Mercer's conditions (1909). The representer theorem (Kimeldorf and Wahba, 1971) guarantees the minimizer of the penalized negative log likelihood is consisted of $(\phi(\boldsymbol{x}) - H\phi(\boldsymbol{x}_I))\boldsymbol{w} = K^*\boldsymbol{\alpha}$ for some vector $\boldsymbol{\alpha}$.

Here $K^* = (\phi(\boldsymbol{x}) - H\phi(\boldsymbol{x}_I))(\phi(\boldsymbol{x}) - H\phi(\boldsymbol{x}_I))'$ can be expressed in terms of kernel functions as follows,

$$
\begin{aligned}
K^* &= (\phi(\boldsymbol{x}) - H\phi(\boldsymbol{x}_I))(\phi(\boldsymbol{x}) - H\phi(\boldsymbol{x}_I))' \\
&= \phi(\boldsymbol{x})\phi(\boldsymbol{x})' - 2H\phi(\boldsymbol{x}_I)\phi(\boldsymbol{x})' + H\phi(\boldsymbol{x}_I)\phi(\boldsymbol{x}_I)'H \\
&= K(\boldsymbol{x}, \boldsymbol{x}) - 2HK(\boldsymbol{x}_I, \boldsymbol{x}) + HK(\boldsymbol{x}_I, \boldsymbol{x}_I)H'.
\end{aligned}
\tag{3.5}
$$

Now the problem (3.4) becomes obtaining $(\boldsymbol{\mu}, \boldsymbol{\alpha})$ to minimize

$$
L(\boldsymbol{\mu}, \boldsymbol{\alpha}|\boldsymbol{y}, \boldsymbol{x}) = \frac{1}{2}(\boldsymbol{y} - U\boldsymbol{\mu} - K^*\boldsymbol{\alpha})'W^{-1}(\boldsymbol{y} - U\boldsymbol{\mu} - K^*\boldsymbol{\alpha}) + \frac{\lambda}{2}\boldsymbol{\alpha}'K^*\boldsymbol{\alpha}.
\tag{3.6}
$$

From the optimality conditions, we can get the followings:

$$
\frac{\partial L}{\partial \boldsymbol{\mu}} = 0 \quad \rightarrow \quad U'W^{-1}(\boldsymbol{y} - U\boldsymbol{\mu} - K^*\boldsymbol{\alpha}) = \boldsymbol{0}
$$

$$
\frac{\partial L}{\partial \boldsymbol{\alpha}} = 0 \quad \rightarrow \quad \boldsymbol{y} - U\boldsymbol{\mu} - K^*\boldsymbol{\alpha} - \frac{1}{\lambda}W\boldsymbol{\alpha} = \boldsymbol{0}.
$$

The estimates of $(\boldsymbol{\mu}, \boldsymbol{\alpha})$ can be obtained from the linear system:

$$
\begin{pmatrix} U'W^{-1}U, & U'W^{-1}K^* \\ U, & K^* + \lambda W \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} U'W^{-1} \\ I \end{pmatrix}\boldsymbol{y},
$$

which leads to

$$
\begin{pmatrix} \widehat{\boldsymbol{\mu}} \\ \widehat{\boldsymbol{\alpha}} \end{pmatrix} = \begin{pmatrix} U'W^{-1}U, & U'W^{-1}K^* \\ U, & K^* + \lambda W \end{pmatrix}^{-1} \begin{pmatrix} U'W^{-1} \\ I \end{pmatrix}\boldsymbol{y}.
\tag{3.7}
$$

Since $W = Cov(\boldsymbol{e})$ is not known we cannot find the estimates of $(\boldsymbol{\mu}, \boldsymbol{\alpha})$ directly from (3.7) but iterative method as follows:

(i) With $W = \mathrm{I}_{(I-1)T}$,
(ii) Obtain the estimates of $(\boldsymbol{\mu}, \boldsymbol{\alpha})$ from (3.6).
(iii) With residuals $\boldsymbol{r} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$, find estimate of W,
$\widehat{W}((i-1)T+1 : iT, (j-1)T+1 : jT) = S_{ij}\mathrm{I}_\mathrm{T}, i, j = 1, 2, \cdots, I-1$ where S_{ij} is the (i, j)th element of

$$
\widehat{Cov}(\boldsymbol{e}_{.t}) = \widehat{Cov}\begin{pmatrix} e_{1t} \\ e_{2t} \\ \vdots \\ e_{I-1,t} \end{pmatrix} = \frac{1}{T-1}\sum_{t=1}^{T}\boldsymbol{r}_{.t}\boldsymbol{r}'_{.t} \text{ and } \boldsymbol{r}_{.t} = \begin{pmatrix} y_{1t} - \widehat{y}_{1,t} \\ y_{2t} - \widehat{y}_{2t} \\ \vdots \\ y_{I-1,t} - \widehat{y}_{I-1,t} \end{pmatrix}, t = 1, \cdots, T.
$$

(iv) Iterate (ii) and (iii) until convergence.

For given $\boldsymbol{x}_{it}$ and $\boldsymbol{x}_{It}$, the estimated logarithm of ratio of market shares of the brand i to market share of the brand I is obtained as follows:

$$
\widehat{y}_{it} = \widehat{\mu}_i + (K(\boldsymbol{x}_{it}, \boldsymbol{x}) - 2K(\boldsymbol{x}_{It}, \boldsymbol{x}) + K(\boldsymbol{x}_{It}, \boldsymbol{x}_{It})H')\widehat{\boldsymbol{\alpha}}, \ i = 1, \cdots, I-1.
\tag{3.8}
$$

4. Numerical studies

We perform the estimation of MCI model by using the proposed method, OLS estimation procedure and GLS estimation procedure. Given data include the monthly market shares and prices of 4 car brands - Sonata, SM5, Lotze, Tosca, from January 2006 to July 2009. The response variable is the log-ratio of market shares of 3 car brands based on Sonata as the reference. The input variables include current prices and sales numbers of previous month. We use an expanding window of historical in-sample data to produce one-step ahead forecast for a given out-of-sample month from May 2008 to July 2009. That is, we train the differential effect MCI model 15 times, each time using slightly different, one-month-extended data, to forecast the market shares of incoming months. Thus, to forecast the first out-of-sample month May 2008, we use the first 28 months from January 2006 to April 2008. For each following out-of-sample month we add one month of historical data to the estimation window. In the end, we compute the root mean squared error (RMSE) and the mean absolute error (MAE) per brand per month, where an error is defined as the difference between true brand market share and one-step ahead forecasted market share as follows:

$$RMSE_i = \sqrt{\frac{1}{15}\sum_{t=29}^{43}(y_{it} - \widehat{y}_{it})^2}, \ MAE_i = \frac{1}{15}\sum_{t=29}^{43}|y_{it} - \widehat{y}_{it}|, \ i = 1, \cdots, 4.$$

The performance of forecasting market shares using the proposed method is characterized by hyperparameters - λ and the kernel parameters. For each set of in-sample months, the optimal values of hyperparameters can be chosen by minimizing the average of 5 MSEs as follows:

$$V(\theta) = \frac{1}{5}\sum_{i=1}^{I-1}\sum_{t=T_r-4}^{T_r}(y_{it} - \widehat{y}_{it})^2,$$

where θ is a set of hyperparameters, $\widehat{y}_{it}$ is one-step ahead forecast of y_{it} calculated from data $\{y_{ij}, \boldsymbol{x}_{ij}\}_{i=1 j=1}^{I-1 t-1}$ and $\{y_{ij}, \boldsymbol{x}_{ij}\}_{i=1 j=1}^{I-1 T_r}$ which are data of in-sample months.

In the given data we empirically found that the radial basis function (RBF) kernel provides better performance than the linear kernel. The results of the numerical studies are presented as averages of 15 RMSEs and 15 MAEs for each brand in Table 1. Overall, the proposed method outperforms ML estimation procedures in terms of RMSE and MAE over the 15-month out-of-sample period from May 2008 to July 2009. The average monthly RMSE (MAE) over all brands for the out-of-sample period is equal to 0.0478 (0.0368) for OLS estimation procedure, 0.0438 (0.0337) for GLS estimation procedure and for the proposed method with RBF kernel the RMSE is 0.0373 (0.0291). Figure 1 shows plots of the logarithm of ratio of market share to the market share of Sonata from 2008 May to 2009 July, where dots indicate the observed values, dotted lines are one-step ahead forecasts by OLS estimation procedure, dashed lines are one-step ahead forecasts by GLS estimation procedure, and the solid lines are one-step ahead forecasts by the proposed method. In figure we can see that OLS and GLS estimation procedures show the similar pattern of forecasting but both do not seem to superior to the proposed method.

Table 4.1 Averages of 15 RMSEs and 15 MAEs for one-step ahead forecasts

	RMSE			MAE		
	OLS	GLS	proposed	OLS	GLS	proposed
Sonata	0.0725	0.0669	0.0487	0.0578	0.0538	0.0392
SM5	0.0414	0.0437	0.0341	0.0331	0.0356	0.0265
Lotze	0.0578	0.0463	0.0519	0.0414	0.0320	0.0386
Tosca	0.0193	0.0181	0.0145	0.0149	0.0134	0.0120
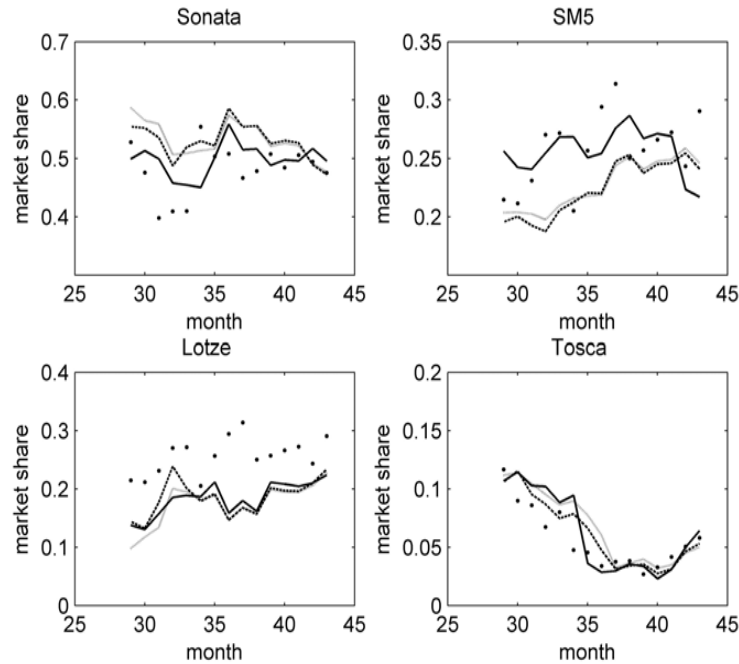
Figure 4.1 Plots of the market shares from 2008 May to 2009 July

5. Concluding Remarks

To forecast the market shares of several brands simultaneously in a differential effect multiplicative competitive interaction model we proposed a kernel machine technique for estimation of a differential effect multiplicative competitive interaction model. Through the example we showed that the proposed method yields the satisfying results.

References

Cooper, L. G. and Nakanishi, M. (1988). *Market share analysis*, Kluwer Academic Publishers, Boston.

Fok. D. and Franses, P. H. (2004). Analyzing the effects of a brand introduction on competitive structure using a market share attraction model. *International Journal of Research in Marketing*, **21**, 159 - 177.

Fok, D., Franses, P. and Paap, R. (2002). Econometric analysis of the market share attraction model. In *Advances in econometrics*, edited by Franses, P. and Montgomery, A., **16**, Elsevier Science, 223-256.

Gruca, T. S. and Klemz, B. R. (1998). Using neural networks to identify competitive market structures from aggregate market response data. *International Journal of Management Science*, **26**, 49-62.

Hwang, C. (2010a). Support vector quantile regression for longitudinal data. *Journal of the Korean Data & Information Science Society*, **21**, 309-316.

Hwang, H. (2010b). Fixed size LS-SVM for multiclassification problems of large data sets. *Journal of the Korean Data & Information Science Society*, **21**, 561-567.

Kimeldorf, G. S. and Wahba, G. (1971). Some results on Tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications*, **33**, 82-95.

Kim, M. S., Park, H. J., Hwang, C. and Shim, J. (2008). Claims reserving via kernel machine. *Journal of the Korean Data & Information Science Society*, **19**, 1419-1427.

Kumar, V. (1994). Forecasting performance of market share models: An assesment, additional Insights, and guidelines. *International Journal of Forecasting*, **10**, 295-312.

Kumar, V. and Heath, T. B. (1990). A comparative study of Market share models using disaggregate data. *International Journal of Forecasting*, **6**, 163-174.

Mercer, J. (1909). Functions of positive and negative and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society A*, 415-446.

Park, H. (2009). Analysis of market share attraction data using LS-SVM. *Journal of the Korean Data & Information Science Society*, **20**, 879-886.

Shim, J. (2011). V ariable selection in the kernel Cox regression. *Journal of the Korean Data & Information Science Society*, **22**, 79 5-801.

Shim, J. and Seok, K. H. (2008). Kernel poisson regression for longitudinal data. *Journal of the Korean Data & Information Science Society*, **19**, 1353-1360.

Seok, K. H. (2010). Semi-supervised classification with LS-SVM formulation. *Journal of the Korean Data & Information Science Society*, **21**, 461-470.

Suykens, K. A. K. and Vanderwalle, J. (1999). Least square support vector machine classifier. *Neural Processing Letters*, **9**, 293-300.

Vapnik, V. N. (1982). *Estimation of dependences based on empirical data*, Springer, Berlin.

Vapnik, V. N. (1995). *The nature of statistical learning theory*, Springer, New York.