

A Modified Diffusion Model Considering Autocorrelated Disturbances: Applications on CT Scanners and FPD TVs

자기상관 오차항을 고려한 수정된 확산모형: CT-스캐너와 FPD TV에의 응용

Cha, Kyoung Cheon(차 경 천)*
Kim, Sang-Hoon(김 상 훈)**

Estimating the Bass diffusion model often creates a time-interval bias, which leads the OLS approach to overestimate sales at early stages and underestimate sales after the peak. Further, a specification error from omitted variables might raise serial correlations among residuals when marketing actions are not incorporated into the diffusion model. Autocorrelated disturbances may yield unbiased but inefficient estimation, and therefore invalid inference results. This phenomenon warrants a modified approach to estimating the Bass diffusion model.

In this paper, the authors propose a modified Bass diffusion model handling autocorrelated disturbances. To validate the new approach, authors applied the method on two different data-sets: CT Scanners in the U.S, and FPD TV sales in Korea. The results showed improved model fit and the validity of the proposed model.

Key words: Bass diffusion model, serially correlated residuals, autocorrelated disturbances

I. Introduction

The Bass diffusion model(Bass 1969) has been well accepted and applied over a large number of new products. Meanwhile, most

research in predicting new product diffusion has employed discrete time-series data. As Putsis(1996) correctly mentioned, a serious time-interval bias may incur in this process. Specifically speaking, the OLS (ordinary least squares) approach to the Bass model may lead

* Assistant Professor, Seoul School of Integrated Sciences & Technologies(kccha@assist.ac.kr)

** Corresponding author, Associate Professor of Marketing, Graduate School of Business, Seoul National University (profkim@snu.ac.kr)

to overestimate sales at early stages and underestimate sales after the peak. It is because the Bass model has consistently assumed that the error terms are temporally uncorrelated, which is rather unlikely in the real world (Mahajan et al. 2000, p266). Despite this, there are very few models that accommodate the serially correlated errors. It is apparent that this is an important point that shouldn't be overlooked when estimating diffusion models using time-series data. Desiraju et al.(2004) pointed out another important issue. They claimed that the "endogeneity" may be a problem for all diffusion studies that employ a linearized estimation equation when serial correlations are present among residuals. Under such circumstances, using OLS to estimate the linearized version of the Bass diffusion model would result in not only biased but also inconsistent estimates. The researchers might turn to a so-called "instrument variables" approach to handle the endogeneity problems.

There is another source of autocorrelated disturbances in the diffusion model. The specification error from omitted variables may raise serial correlations among residuals because, in general, marketing actions are not incorporated into the Bass diffusion model(Lilien et al. 1992, p473). Johnston(1991, p309-310) pointed out that there are three reasons for autocorrelated disturbances and that they are omitted explanatory variables, misspecification of the form of the relationship, and measurement

error in the dependent variable. With them, autocorrelated disturbances may result in unbiased but inefficient estimation and therefore invalid inference procedures. In fact, Bass et al. (1994) proposed a "generalized" diffusion model that incorporates a few marketing variables into the original Bass model. But apparently, it is impossible to take into account all marketing and other variables that affect the diffusion process.

There have been few studies that considered serial correlation in estimating the growth curves (e.g., Mar-Molinero 1980; Meade 1988; Desiraju et al. 2004). Mar-Molinero (1980), for instance, used a logistic curve with first-order autocorrelated error term in predicting the diffusion of tractors in Spain to obtain minimum variance estimates while improving the model fit. Meade(1988) tried and compared various error structures in growth curves and concluded that only Mar-Molinero (1980)'s method was effective in improving the model fit.

With all above being said, it is quite obvious that a better diffusion model is called for that may take care of the possible serial correlation problem among residuals, especially to improve the predictive validity. Srinivasan and Mason (1986) mentioned that it is necessary to develop an econometric procedure to handle autocorrelation problem when the serial correlation is large enough to be statistically significant.

In the subsequent sections, authors derive a modified approach to the Bass diffusion model

in the presence of autocorrelated disturbances, and then apply the proposed model to the datasets of CT Scanners in U.S hospitals and Flat Panel Display (FPD) TV sales in Korea. Finally, managerial implications are discussed.

II. Model Derivation

Let S_t be the adoption at time t , Y_t be the cumulative adoption at time t , N be market potential and u_t a disturbance term at time t . Then, equation (1) is a typical Bass diffusion model. As well known, the parameter p indicates the innovation effect and q the imitation effect.

$$S_t = (p + q \frac{Y_{t-1}}{N})(N - Y_{t-1}) + u_t \quad (1)$$

When u_t in equation (1) is an autoregressive process of order p , $u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \dots + \phi_p u_{t-p} + \varepsilon_t$, where ε_t is a white noise. Using a backshift lag operator that shifts time one period back, such as $Bu_t = u_{t-1}$, $B^2 u_t = u_{t-2}$, \dots , $B^p u_t = u_{t-p}$, we can rewrite it as $(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) u_t = \varepsilon_t$.

Therefore, equation (1) can be arranged as equation (2).

$$S_t = (p + q \frac{Y_{t-1}}{N})(N - Y_{t-1}) + \frac{1}{(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)} \varepsilon_t \quad (2)$$

Multiplying $(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ on both sides of equation (2) leads to a Bass diffusion model with autocorrelated disturbances of order p as shown in equation (3).

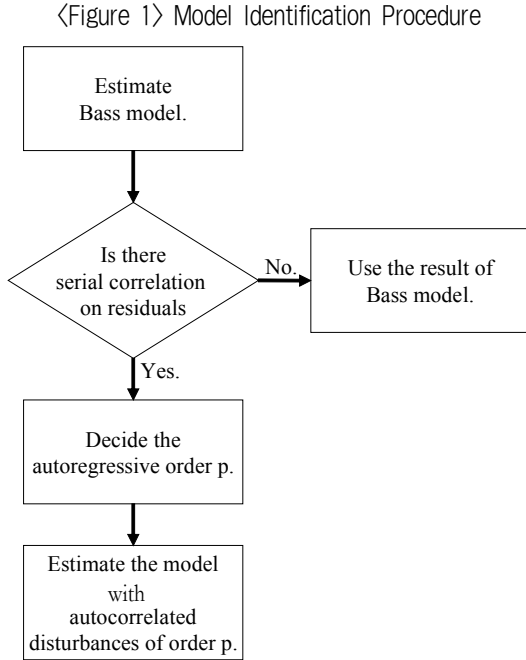
$$S_t = (p + q \frac{Y_{t-1}}{N})(N - Y_{t-1}) + \sum_{i=1}^p \phi_i [S_{t-i} - (p + q \frac{Y_{t-i-1}}{N})(N - Y_{t-i-1})] + \varepsilon_t \quad (3)$$

The second term of the right hand side of equation (3) represents the errors generated in the past time periods. These errors might have come from time-interval bias and/or omitted variables. The modified diffusion model rectify such errors in current time period by the weight parameters, ϕ_i .

III. Model Identification

To calibrate the proposed model, a researcher should first estimate the original Bass model. Then he or she needs to test if there is any significant serial correlation among residuals. Once the existence of serial correlations is confirmed, the autoregressive order (p) has to be specified. The autoregressive order is to be determined by matching the patterns in the sample autocorrelations and partial autocorrelations with those of the known theoretical models (Box et al. 1994, p185). The final step

is to estimate equation (3). <Figure 1> summarizes the model identification procedure.



If residuals have statistically significant serial correlations, the forecasting function of the diffusion model should also change. The forecasting function in the presence of autocorrelated disturbances of order p can be written as below.

$$\hat{S}_t = (p + q \frac{Y_{t-1}}{N})(N - Y_{t-1}) + \sum_{i=1}^p \phi_i [S_{t-i} - (p + q \frac{Y_{t-1-i}}{N})(N - Y_{t-1-i})] \quad (4)$$

IV. Application

To validate the model, authors applied the proposed diffusion model to two datasets: CT Scanners in U.S. hospitals and FPD TV sales in Korea.

The dataset of CT scanner diffusion among U.S. hospitals (1973, 6~1981, 12) is the one which Tragtenberg and Yizhaki (1989) used for their study. The first step of model identification procedure with the dataset confirmed that there exist significant serial correlations in residuals when a Bass model was estimated. <Table 1> shows the estimation results of the Bass model. Durbin-Watson (DW) statistic was used to test for the absence of serial correlation of lag 1. With positive correlations, DW falls into the range from 0 and 2; with negative correlations, it lies between 2 and 4; when the residuals are uncorrelated, DW statistics would be about 2. From <Table 1>, we can conclude that there exists a positive correlation.

<Figure 2> presents the result of residual test including the sample autocorrelation(AC), sample partial autocorrelation(PAC), Q-statistic and

<Table 1> The estimation result of Bass model for CT Scanners

| Product \ parameters | p | q | N | DW Statistic | AIC_c |
|----------------------|----------------|----------------|----------------|--------------|---------|
| CT Scanners | 0.003 (***) | 0.056 (***) | 1,256 (***) | 0.85 | 6.53 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

<Figure 2> The result of residual test for CT Scanners

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob | |
|-----------------|---------------------|----|--------|--------|--------|-------|
| | | 1 | 0.570 | 0.570 | 34.154 | 0.000 |
| | | 2 | 0.465 | 0.207 | 57.062 | 0.000 |
| | | 3 | 0.443 | 0.179 | 78.131 | 0.000 |
| | | 4 | 0.400 | 0.069 | 95.468 | 0.000 |
| | | 5 | 0.229 | -0.158 | 101.22 | 0.000 |
| | | 6 | 0.176 | -0.041 | 104.65 | 0.000 |
| | | 7 | 0.138 | -0.024 | 106.77 | 0.000 |
| | | 8 | 0.130 | 0.059 | 108.68 | 0.000 |
| | | 9 | -0.025 | -0.163 | 108.75 | 0.000 |
| | | 10 | -0.030 | -0.009 | 108.86 | 0.000 |
| | | 11 | 0.091 | 0.210 | 109.02 | 0.000 |
| | | 12 | -0.015 | -0.115 | 109.85 | 0.000 |
| | | 13 | 0.008 | 0.103 | 109.85 | 0.000 |
| | | 14 | -0.080 | -0.219 | 110.62 | 0.000 |
| | | 15 | -0.078 | -0.058 | 111.37 | 0.000 |
| | | 16 | -0.049 | 0.102 | 111.66 | 0.000 |
| | | 17 | -0.164 | -0.207 | 115.01 | 0.000 |
| | | 18 | -0.152 | 0.006 | 117.93 | 0.000 |
| | | 19 | -0.033 | 0.101 | 118.07 | 0.000 |
| | | 20 | -0.088 | -0.003 | 119.07 | 0.000 |
| | | 21 | -0.100 | 0.024 | 120.38 | 0.000 |
| | | 22 | -0.123 | -0.183 | 122.38 | 0.000 |

<Figure 3> The residual test result of proposed model for CT Scanners

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob | |
|-----------------|---------------------|----|--------|--------|--------|-------|
| | | 1 | -0.120 | -0.120 | 1.4979 | 0.221 |
| | | 2 | 0.056 | 0.042 | 1.8251 | 0.402 |
| | | 3 | 0.134 | 0.148 | 3.7365 | 0.291 |
| | | 4 | 0.209 | 0.249 | 8.4282 | 0.077 |
| | | 5 | -0.045 | 0.002 | 8.6486 | 0.124 |
| | | 6 | 0.037 | -0.019 | 8.7997 | 0.185 |
| | | 7 | 0.006 | -0.066 | 8.8033 | 0.267 |
| | | 8 | 0.154 | 0.112 | 11.465 | 0.177 |
| | | 9 | -0.140 | -0.100 | 13.676 | 0.134 |
| | | 10 | -0.123 | -0.191 | 15.401 | 0.118 |
| | | 11 | 0.214 | 0.173 | 20.697 | 0.037 |
| | | 12 | -0.118 | -0.075 | 22.312 | 0.034 |
| | | 13 | 0.093 | 0.186 | 23.341 | 0.038 |
| | | 14 | -0.104 | -0.093 | 24.646 | 0.038 |
| | | 15 | -0.044 | -0.159 | 24.880 | 0.052 |
| | | 16 | 0.105 | 0.133 | 26.241 | 0.051 |
| | | 17 | -0.154 | -0.187 | 29.188 | 0.033 |
| | | 18 | -0.138 | -0.076 | 31.566 | 0.025 |
| | | 19 | 0.143 | 0.054 | 34.151 | 0.018 |
| | | 20 | -0.074 | 0.001 | 34.868 | 0.021 |
| | | 21 | -0.014 | 0.146 | 34.882 | 0.029 |
| | | 22 | -0.058 | -0.101 | 35.331 | 0.036 |

its p-value. As <Figure 2> reconfirms, there exists a serial correlation among residuals. By matching the autocorrelation patterns with those of the known models, authors set the order as 1 and re-estimated the diffusion model via the proposed model with autocorrelated disturbances of order 1. The new estimation results are in <Table 2> and <Figure 3>. In <Table 2>, the estimated parameter for auto-correlation disturbance (ϕ) is also presented. From the results, we can see that the fit has improved and the serial autocorrelations have been handled properly.

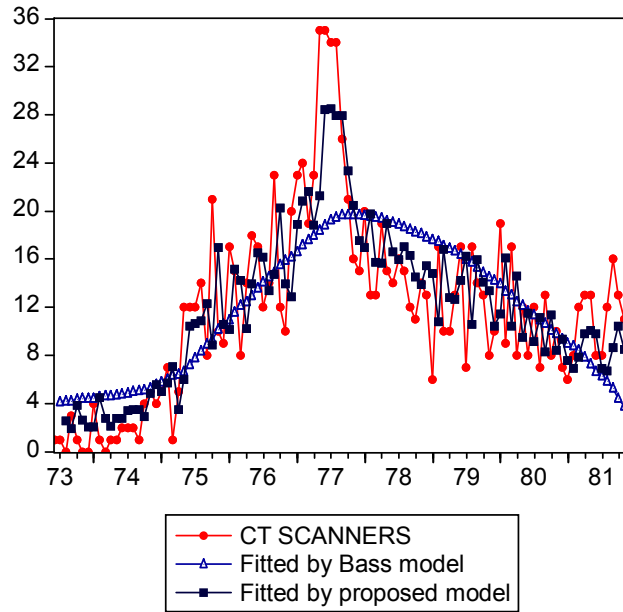
AIC (Akaike's information Criterion) is a useful measure in selecting the best specification among alternative models (Burnham and Anderson 2002). However, authors used the "corrected" AIC (AIC_c) to compare between the Bass and the proposed models because AIC perform poorly when there are too many parameters relative to the size of the sample. Burnham and Anderson (2002) suggested to use AIC_c when the ratio of the number of observations to the number of the estimated parameters is small (say the ratio < 40). When N is the number of observations and K is the

<Table 2> The estimation result of proposed model for CT Scanners

| Product | parameters | p | q | N | ϕ | DW Statistic | AIC_c |
|-------------|------------|---------------|----------------|----------------|----------------|--------------|---------|
| CT Scanners | | 0,004 (**) | 0,056 (***) | 1,250 (***) | 0,572 (***) | 2.2 | 6.34 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

〈Figure 4〉 The fitted graph by Bass model and proposed model



number of estimated parameters, AIC_c is defined as equation (5).

$$AIC_c = AIC + \frac{2K(K+1)}{N-K-1} \quad (5)$$

As shown in 〈Table 1〉 and 〈Table 2〉, the value of AIC_c of the proposed model is smaller

than that of the Bass model ($6.34 < 6.53$), which means that the proposed model is better than the traditional Bass model.¹⁾ 〈Figure 4〉 presents the fitted graph of CT Scanners by both Bass model and the proposed model. It confirms our expectation that the Bass model would overestimate the sales before peak and underestimate them after peak.

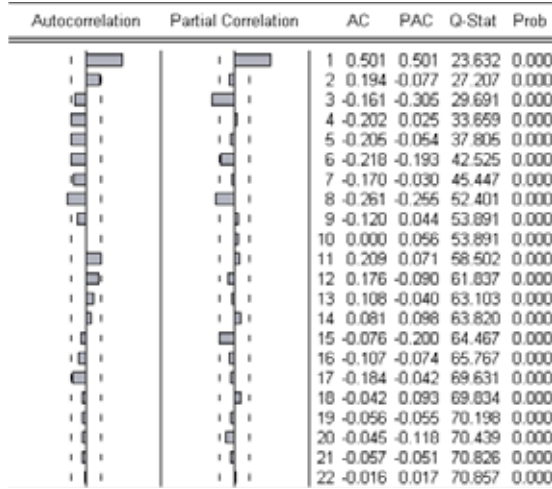
〈Table 3〉 The estimation result of Bass model for FPD TV sales

| Product | parameters | q | N | DW Statistic | AIC_c |
|-----------------|------------|----------------|--------------------|--------------|---------|
| FPD TV in Korea | | 0.100 (***) | 3,826,492 (***) | 0.97 | 20.09 |

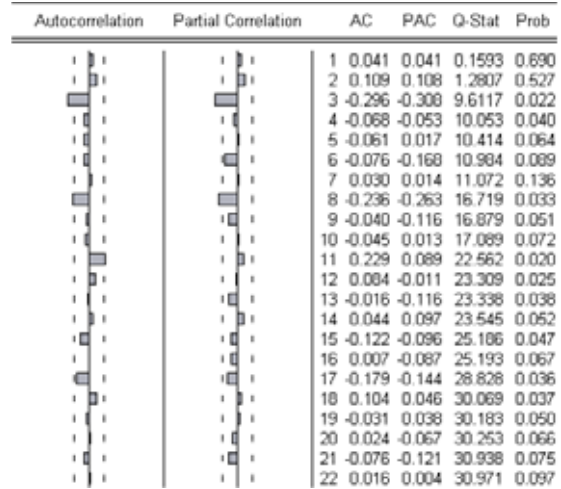
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1) Since AIC is based on information theory, it does not use a traditional “hypothesis testing” paradigm. A so-called “J-test” may be used to test which model performs better between two non-nested models, but it was not conducted in this paper since our priority concern is put on the elimination of correlated error terms.

〈Figure 5〉 The result of residual test for FPD TV sales



〈Figure 6〉 The residual test result of proposed model for FPD TV sales



〈Table 4〉 The estimation result of proposed model for FPD TV sales

| Product | parameters | q | N | ϕ | DW Statistic | AIC_c |
|-----------------|------------|----------------|--------------------|----------------|--------------|---------|
| FPD TV in Korea | | 0.100 (***) | 4,002,656 (***) | 0.518 (***) | 1.90 | 20.01 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The second dataset used for the model validation is about the monthly FPD TV sales in Korea (2000.1~2007.8). As above, 〈Table 3〉 and 〈Figure 5〉 show the estimation results from the Bass model. In this case, authors restricted the innovation effect (parameter p) to be "0" because it was estimated negatively in the preliminary estimation. As shown in 〈Figure 5〉, there exists a serial correlation in residuals. Authors set the order as 1 and estimated the proposed model with autocorrelated disturbances of order 1. The results from the new model are in 〈Table 4〉 and 〈Figure

6〉. Again, after the estimation of proposed model, model fit improved and no serial correlation was identified. AIC_c of the proposed model was also slightly smaller than that of the Bass model (20.01<20.09).

V. Discussion

Whatever the reasons are for the serial correlation among residuals, the OLS approach to the traditional Bass model tends to overe-

<Table 5> The estimation result of Bass model for Bass et al. (1994)

| Product | parameters | <i>p</i> | <i>q</i> | <i>N</i> | DW Statistic | <i>AIC_c</i> |
|----------------------|------------|----------------|----------------|------------------|--------------|------------------------|
| Room Air Conditioner | | 0.020 (**) | 0.389 (***) | 173,134 (***) | 1.59 | 16.95 |
| Color TV | | 0.020 (*) | 0.706 (***) | 35,746 (***) | 1.53 | 20.84 |
| Clothes Dryer | | 0.024 (***) | 0.325 (***) | 15,652 (***) | 1.53 | 16.27 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

estimate and underestimate the sales around the peak. In such circumstances, a new method of estimating the Bass model is warranted, and the authors of the present paper devised a new approach called “a modified Bass model with autocorrelated disturbances.” The proposed model has a simple procedure and proved to improve predictive validity. The proposed model corrects the errors generated on past time periods in current time period with weighting parameters.

Authors applied the proposed model to the datasets of CT Scanners in U.S. hospitals and the FPD TV sales in Korea. The estimation results confirmed the existence of serial correlation among residuals, and therefore possibility of improper statistical inference. When the proposed model was applied, the model fit improved and the serial correlation problem was resolved.

Naturally, there are cases where the proposed model estimation is unnecessary. In fact, when the authors analyzed the diffusion data used by Bass et al. (1994), no meaningful serial correlation was found among residuals. <Table 5> shows the estimation results of a Bass

<Figure 7> The result of residual test for Bass et al. (1994)

(a) Room Air Conditioner

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 0.149 | 0.149 | 0.3383 | 0.561 |
| | | 2 -0.135 | -0.161 | 0.6462 | 0.724 |
| | | 3 -0.153 | -0.110 | 1.0804 | 0.782 |
| | | 4 -0.507 | -0.517 | 6.4849 | 0.166 |
| | | 5 -0.126 | -0.055 | 6.8756 | 0.230 |
| | | 6 -0.022 | -0.277 | 6.8887 | 0.331 |
| | | 7 0.183 | 0.095 | 8.0158 | 0.331 |
| | | 8 0.208 | -0.246 | 9.8250 | 0.278 |
| | | 9 -0.007 | -0.078 | 9.8280 | 0.365 |
| | | 10 -0.036 | -0.276 | 9.9352 | 0.446 |

(b) Color TV

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|----------|--------|--------|-------|
| | | 1 0.187 | 0.187 | 0.4335 | 0.510 |
| | | 2 -0.491 | -0.546 | 3.8481 | 0.146 |
| | | 3 -0.365 | -0.178 | 6.0464 | 0.109 |
| | | 4 -0.016 | -0.237 | 6.0515 | 0.195 |
| | | 5 0.126 | -0.186 | 6.4475 | 0.265 |
| | | 6 0.117 | -0.107 | 6.9011 | 0.330 |
| | | 7 -0.011 | -0.185 | 6.9075 | 0.439 |

(c) Clothes Dryers

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|----------|--------|--------|-------|
| | | 1 0.188 | 0.188 | 0.5383 | 0.463 |
| | | 2 -0.478 | -0.532 | 4.3772 | 0.112 |
| | | 3 -0.355 | -0.171 | 6.7354 | 0.081 |
| | | 4 0.009 | -0.179 | 6.7370 | 0.150 |
| | | 5 -0.003 | -0.366 | 6.7371 | 0.241 |
| | | 6 0.075 | -0.053 | 6.8942 | 0.331 |
| | | 7 0.035 | -0.330 | 6.9362 | 0.436 |
| | | 8 0.058 | -0.079 | 7.0757 | 0.528 |
| | | 9 0.003 | -0.211 | 7.0762 | 0.629 |
| | | 10 0.014 | -0.121 | 7.0915 | 0.717 |

model and <Figure 7> provides the residual test result. This gives one reason why the Bass model worked well enough without marketing variables in the study of Bass et al. (1994). However, whenever the serial correlation is present, the proposed model would be an easy and useful approach to the marketers who are eager for better prediction of new product diffusion.

<논문 접수일: 2008. 12. 12>

<게재 확정일: 2009. 02. 04>

References

- Bass, Frank M. (1969). "A New Product Growth Model for Consumer Durables," *Management Science*, 15, 215-227.
- Bass, Frank M., Trichy V. Krishnan, and Dipak C. Jain. (1994). "Why the Bass Model Fits without Decision Variables." *Marketing Science*, 24 (November), 1568-1578.
- Box, George E.P., Gwilym M. Jenkins and Gregory C. Reinsel. (1994). *Time Series Analysis: Forecasting and Control*, 3rd ed., Prentice-Hall International Inc.
- Burnham, Kenneth P. and David R. Anderson. (2002). *Model Selection and Multimodel Inference*, Springer.
- Desiraju, Ramarao, Harikesh Nair and Pradeep Chintagunta. (2004). "Diffusion of new pharmaceutical drugs in developing and developed nations," *International Journal of Re-*
- search in Marketing*, 21, 341-357.
- Green, William H. (1997). *Econometric Analysis*, 3rd ed., Prentice Hall International Edition.
- Johnston, J. (1991). *Econometric Methods*, 3rd ed., McGraw-Hill International Editions.
- Lilien, Gary L., Philip Kotler, K. Sridhar Moorthy. (1992). *Marketing Models*, Prentice Hall.
- Mahajan, Vijay, Eitan Muller and Yoram Wind. (2000). *New-Product Diffusion Models*, Kluwer Academic Publishers.
- Mar-Molinero, C. (1980). "Tractors in Spain: A Logistic Analysis," *The Journal of the Operational Research Society*, Vol. 31, No. 2, 141-152.
- Meade, Nigel. (1988). "Forecasting with Growth Curves: the effect of error structure," *Journal of Forecasting*, Vol. 7, 235-244.
- Putsis, Jr., William P. (1996). "Temporal Aggregation in Diffusion Models of First-Time Purchase: Does Choice of Frequency Matter?" *Technological Forecasting and Social Change*, 51, 265-279.
- Srinivasan, V. and Charlotte H. Mason. (1986). "Nonlinear Least Squares Estimation of New Product Diffusion Models," *Marketing Science*, Vol. 5, No. 2, 169-178.
- Tragtenberg, Mauel and Shlomo Yizhaki. (1989). "The Diffusion and Innovations: A Methodological Reappraisal," *Journal of Business and Economic Statistics*, Vol. 7, No. 1, 35-47, Jan.

자기상관 오차항을 고려한 수정된 확산모형: CT-스캐너와 FPD TV에의 응용

차 경 천*
김 상 훈**

요 약

시계열 확산 데이터를 활용하여 Bass 확산모형을 최소자승법(OLS)으로 추정하면, 초기에는 과다 추정하고 변곡점을 지나서는 수요를 낮게 추정하는 경향이 있다. 또한 확산모형에서 필요한 변수가 모형에서 빠짐으로 인해 발생하는 설정오류는 잔차의 자기상관을 발생시킬 수 있다. 자기상관이 오차항에 있을 경우, 추정된 모형의 모수들은 불편추정치이나 비효율적 추정치가 된다. 따라서 이러한 문제를 해결하는 확산모형의 개발이 요구된다.

본 연구에서는 자기상관 오차항을 고려한 수정된 확산모형을 제안하였다. 모형의 검증을 위해 미국의 CT-스캐너와 우리나라의 FPD TV 판매량을 제안된 모형에 응용하였다. 분석결과, 제안된 모형이 기존 모형에 비해 적합도와 모형의 주요 추정 통계량에서 우수함을 보였다.

핵심개념: Bass 확산모형, 자기상관 오차

* 서울과학종합대학원 조교수(kccha@assist.ac.kr)

** 서울대학교 경영전문대학원 부교수, 교신저자(profkim@snu.ac.kr)