Journal of Korean
Data & Information Science Society
2007, Vol. 18, No. 4, pp. 995 ~ 1004

Modelling Online Word-of-Mouth Effect on Korean Box-Office Sales Based on Kernel Regression Model[1)]

Siyun Park[2)] · Jin Gyo Kim[3)]

Abstract

In this paper, we analyse online word-of-mouth and Korean box-office sales data based on kernel regression method. To do this, we consider the regression model with mixed-data and apply the least square cross-validation method proposed by Li and Racine (2004) to the model. We found the box-office sales can be explained by volume of online word-of-mouth and the characteristics of the movies.

Keywords : Cross-Validation Method, Kernel Regression, Mixed-Data, Smoothing Parameter Selection

1. 서론

회귀함수의 비모수적 추론에 관한 많은 연구에 있어 자료가 연속임을 가정한다. 그러나, 마케팅, 계량 경제등의 분야에서 관측되는 자료에는 연속형 자료와 함께 많은 범주형 자료가 혼재되어 있는 경우가 빈번하며 이들 중 몇 개의 범주형 설명변수가 연구의 주요한 관심인 경우가 많다. Aitchison과 Aitken(1976)이 이산형 변수의 평활법을 제안한 이후 이산형 변수에 대한 커널 평활법에 관한 많은 연구가 있어 왔으며, 특히, Bierens(1983), Ahmad와 Cerrito(1994)는 이산형과 연속형 자료가 동시에 존재하는 혼합자료(mixed-data)를 설명변수로 갖는 회귀함수에 대한 국소 상수 회귀 추정량(local constant regression estimator)을 제안하였다. 한편, Li and Racine (2004)은 최소 제곱 교차 타당법(least square cross-validated method)에 기초한 국소 선형 회귀 추정량(local linear regression estimator)을 제안하였는데, 회귀 함수를 추정하는데 있어 국소 선형 추정량(local linear estimator)은 경계 수정(boundary correction) 등의 좋은 성질들을 가지고 있어 많은 연구자들의 주된 연구대상이 되어 왔다(Fan and

---

1) 본 연구는 2007년도 2단계 BK21 서울대학교 경영전문사업단에 의하여 지원되었음.
2) BK 계약조교수, 151-742 서울특별시 관악구 신림동 산 56-1 서울대학교 경영전문대학원
Email : siyunpark@snu.ac.kr
3) 부교수, 151-742 서울특별시 관악구 신림동 산 56-1 서울대학교 경영전문대학원

Gijbels(1995)). 본 연구에서는 이들이 제시한 방법론에 기초하여 국내 개봉 영화의 매출에 미치는 온라인 구전효과를 측정하는 실증 연구를 수행한다. 개봉영화의 매출을 온라인상에서 측정된 구전(word-of-mouth), 영화의 특성과 영화 개봉정보로 설명하는 비모수적 회귀모형을 가정하고 커널 평활법에 기초하여 분석한다. 본 논문의 구성은 다음과 같다. 2장에서는 Li and Racine(2004)의 방법론을 간략히 소개하고 3장에서는 이들의 방법을 실제 우리나라 개봉영화 자료에 대하여 적용한 결과에 대해 살펴본다. 4장에서는 자료 분석을 통해 얻은 결론과 추후 연구 과제에 대해 언급한다.

2. 교차 타당법에 기초한 국소 선형 회귀추정량

본 장에서는 혼합형 자료를 설명변수로 갖는 비모수적 회귀모형에서 커널 평활법에 기초하여 회귀함수를 추정하는 Li and Racine(2004)의 방법론을 살펴보겠다. 이를 위하여 다음과 같은 회귀모형을 고려한다.

$$y_j = g(x_j) + \epsilon_j,\ \ j = 1, \cdots, n, \tag{2.1}$$

단, (x_j, y_j)은 서로 독립이고 동일한 분포를 갖으며 $\epsilon_j = y_j - g(x_j)$는 4차 적률까지 존재한다.

먼저, 식(2.1)의 설명변수x_i가 모두 연속형 변수인 경우를 고려하면, x_i를 q차원의 연속형 벡터이고, $E(y_j|x_j) = g(x_j)$ a.s.. X의 support S는 컴팩트 집합일 때 다음을 가정한다.

$\epsilon > 0$에 대해 $\inf_{x \in S} f(x) \geq \epsilon > 0$ 이고, $g(x), f(x)$와 $\sigma^2(x) = E(u_i^2|x_i = x)$ 는 S에서 4차까지 미분가능하며, $g_{ss}(x)$ 를 g의 x_s에 대한 2차 도함수로 두면,

$s = 1, \cdots, q$ 에 대해 $\int g_{ss}(x)^2 f(x)dx > 0$ 를 만족한다.

$w(\cdot): R \rightarrow R$ 은 $\int w(v)v^4 dv < \infty$ 를 만족하는 유계 대칭 확률 밀도함수로서 m-차 까지 미분가능하다. $w^{(s)}(\cdot)$를 $w(\cdot)$의 s차 도함수라 두면, 모든 $s = 1, \cdots, m$에 대해, $\int |w^{(s)}(v)v^s| dv < \infty$ 이다. 여기서, m은 $m > \max\{2 + 4/q, 1 + q/2\}$ 인 양의 정수이다.

이제, 함수 $g(x)$와 $g(x)$의 첫 번째 도함수를 이용하여 $q+1$차원 벡터 $\delta(x)$를 다음과 같이 정의한다.

$$\delta(x) \equiv (g(x), \beta(x)')',\ \text{단},\ \beta(x) \equiv \Delta g(x) \equiv \partial\, g(x)/\partial\, x \text{이다.} \tag{2.2}$$

모형(2.1)의 $g(x_j)$를 x_i에서 첫 번째 차수까지 테일러 시리즈 전개를 하고, y_j를 $(1, (x_j - x_i)')$에 커널 가중 회귀 (kernel weighted regression)시킴으로써 $\delta(x_i)$의 국소

선형 추정량을 고려하면, 이에 대한 leave-one-out 국소 선형 커널 추정량은 다음과 같다.

$$\hat{\delta}_{-i}(x_i) = \begin{pmatrix} \hat{g}_{-i}(x_i) \\ \hat{\beta}_{-i}(x_i) \end{pmatrix} \tag{2.3}$$
$$= \left[\sum_{j \neq i} W_{h,ij} \begin{pmatrix} 1, & (x_j - x_i)' \\ (x_j - x_i), & (x_j - x_i)(x_j - x_i)' \end{pmatrix} \right]^{-1} \sum_{j \neq i} W_{h,ij} \begin{pmatrix} 1 \\ x_j - x_i \end{pmatrix} y_i,$$

단, $W_{h,ij} = \Pi_{s=1}^{q} h_s^{-1} w((x_{js} - x_{is})/h_s)$는 곱커널 함수 (product kernel function)이고 $h_s = h_s(n)$은 x의 s번째 요소와 관련되는 평활모수(smoothing parameter)이다.

다음의 교차 타당 목적함수(cross-validation object function)를 정의하고 이를 최소화 하는 $(h_1,\cdots,h_q)$를 $(\hat{h}_1,\cdots,\hat{h}_q)$라 한다.

$$CV(h_1,\cdots,h_q) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{g}_{-i}(x_i))^2, \tag{2.4}$$

단, $\hat{g}_{-i}(x_i) = e_1{}'\hat{\delta}_{-i}(x_i)$, $e_1{}' = (1,0,\cdots,0)'$ 이다.

위에서 선택된 평활모수 $(\hat{h}_1,\cdots,\hat{h}_q)$에 기초하여 다음과 같은 추정치를 얻는다.

$$\hat{\delta}(x) = \begin{pmatrix} \hat{g}(x) \\ \hat{\beta}(x) \end{pmatrix} = \left[\sum_i W_{\hat{h},ix} \begin{pmatrix} 1, & (x_i - x)' \\ x_i - x, & (x_i - x)(x_i - x)' \end{pmatrix} \right]^{-1} \sum_i W_{\hat{h},ix} \begin{pmatrix} 1 \\ x_i - x \end{pmatrix} y_i,$$

단, $W_{\hat{h},ix} = \Pi_{s=1}^{q} \hat{h}_s^{-1} w((x_{is} - x_s)/\hat{h}_s)$ 이다.

$\hat{h}_1,\cdots,\hat{h}_q$의 수렴과 $\hat{g}(x)$의 점근적 정규성에 대한 자세한 사항은 Li and Racine (2004)의 (A1)-(A3)와 Appendix A를 참고하라.

이제, 모형(2.1)에서 설명변수 x가 연속형과 이산형의 혼합자료인 경우를 고려하면, 연속형인 경우의 커널함수 $w(\cdot)$에 대한 가정에 다음의 가정을 더한다.

$f(x)$를 $X = (X^c, X^d)$의 support인 $S \times S^d$에서 정의된 양의 상수에 의해 유계 되어 있고, $\sigma^2(x) = E(\epsilon_i^2 | X_i = x)$, $\sigma^2(\cdot, x^d)$, $g(\cdot, x^d)$, $f(\cdot, x^d)$는 모든 $x^d \in S^d$에 대해 모두 G_2^4에 속한다. 단, 다음을 만족하는 함수의 클래스 G_μ^α의 정의는 다음과 같다.

$\alpha > 0$ 와 양의 정수 $\mu > 0$에 대해, $m(x^c)$가 μ번 미분가능하고, $m(x^c)$와 이의 μ차까지의 부분도함수 α차 적률까지 유한한 함수들에 의해 유계되어 있으면, $m \in G_\mu^\alpha$ 이

라 한다.

이제 (2.2)의 $g(x)$에 대해 다음을 정의한다.

$$\beta(x) \equiv \Delta g(x) \equiv \partial g(x^c, x^d)/\partial x^c,$$

단, $x=(x^c, x^d)$, 단, $x^c \in R^q$는 연속형 설명변수에 대한 $q\times 1$ 벡터이고, x^d는 이산형 설명변수에 대한 $r\times 1$ 벡터이다.

편의상 연속형 설명변수 x^c에 대해서는 국소 선형 추정량을, 이산형 변수 x^d에 대해서는 국소 상수 추정량을 고려하자. 설명변수가 연속형인 경우와 유사한 방법으로 추정량 $\hat{\delta}(x)$을 구하기 위해 곱커널 함수를 정의하겠다.

먼저, 명목형 이산 설명변수에 대하여 다음의 커널함수를 정의하면 $0 \le \lambda_s \le 1$ 이고, $\lambda_s = 0$ 일 때 커널함수는 지시함수가 되고 $\lambda_s = 1$일 때 상수함수가 된다. 즉, $\lambda_s = 1$이면 x_s^d는 제거(smooth-out)된다.

$$l(x_{is}^d, x_{js}^d) = \begin{cases} 1, & x_{is}^d = x_{js}^d, \\ \lambda_s, & x_{is}^d \neq x_{js}^d. \end{cases}$$

이를 이용하여 이산형 설명변수에 대한 곱커널(product kernel) 함수를 정의한다.

$$L(x_i^d, x_j^d, \lambda) = \Pi_{s=1}^r \lambda_s^{1-I(x_{is}^d = x_{js}^d)}, \quad \text{단, } I(\cdot)\text{는 지시함수이다.}$$

다음의 교차 타당 목적함수를 최소화하는 $(h_1,\cdots,h_q,\lambda_1,\cdots,\lambda_r)$를 $(\hat{h}_1,\cdots \hat{h}_q,\hat{\lambda}_1,\cdots,\hat{\lambda}_r)$로 구한다.

$$CV(h_1,\cdots,h_q,\lambda_1,\cdots,\lambda_r) = \frac{1}{n}\sum_{i=1}^n (y_i - \hat{g}_{-i}(x_i))^2,$$

여기서, $\hat{g}_{-i}(x_i) = e_1'\hat{\delta}_{-i}(x_i)$, $e_1' = (1,0,\cdots,0)'$는 커널함수 $K_{h,ij}$가 다음과 같이 정의 되었을 때 커널 가중회귀에 의해 구해진 leave-one-out 추정량이다.

$$K_{h,ij} = W_{h,ij}L_{\lambda,ij},\ \ W_{h,ij} = \Pi_{s=1}^q h_s^{-1} w((x_{is}^c - x_{js}^c)/h_s),\ \ L_{\lambda,ij} = \Pi_{s=1}^r l(x_{is}^d, x_{js}^d, \lambda_s)$$

위에서 선택되어진 평활모수에 기초하여 얻어진 추정량은 다음과 같다.

$$\hat{\delta}(x) = \begin{pmatrix} \hat{g}(x) \\ \hat{\beta}(x) \end{pmatrix} = \left[\sum_i K_{\hat{h},ix} \begin{pmatrix} 1, & (x_i^c - x^c)' \\ (x_i^c - x^c), & (x_i^c - x^c)(x_i^c - x^c)' \end{pmatrix}\right]^{-1} \sum_i K_{\hat{h},ix} \begin{pmatrix} 1 \\ x_i^c - x^c \end{pmatrix} y_i,$$

단, $K_{\hat{h},ix} = \Pi_{s=1}^{q} \hat{h}_s^{-1} w((x_{is}^c - x_s^c)/\hat{h}_s) \Pi_{s=1}^{r} l(x_{is}^d, x_s^d, \hat{\lambda}_s)$ 이다.

추정량 $\hat{\delta}(x)$와 $(\hat{h}_1, \cdots \hat{h}_q, \hat{\lambda}_1, \cdots, \hat{\lambda}_r)$의 점근적 성질에 대한 가정과 증명은 Li and Racine (2004)의 (B1), (B2)와 Appendix B를 참고하라.

순서형 자료(ordered categorical data)에 대한 커널함수를 다음과 같이 정의하면 λ_s는 $0 \le \lambda_s \le 1$을 만족하며 $\lambda_s = 0$이면 커널함수는 지시함수가 되고, $\lambda_s = 1$일 때 균등 가중치 함수가 된다.

$$l(x_{is}^d x_{js}^d, \lambda_s) = \begin{cases} 1, & x_{is}^d = x_{js}^d, \\ \lambda_s^{|x_{is}^d - x_{js}^d|}, & x_{is}^d \neq x_{js}^d. \end{cases}$$

위의 커널함수에 기초하여 자료가 연속형, 명목형, 순서형의 혼합자료일 경우 위에서 살펴본 방식과 유사하게 회귀함수의 추정치를 구할 수 있다. 모수적, 비모수적 회귀모형의 추론에서 고려하는 설명변수의 수가 많을 경우 해당 모수에 대한 유의성 검정을 수행한 결과를 반영한 후 회귀모형에 대한 추론을 하는 일련의 작업은 번거로울 수 있다. 이제까지 살펴본 Li and Racine(2004)의 방법론은 반응변수에 대한 정보가 없는 설명변수에 대해 큰 평활모수 추정치를 할당함으로서 추론 이전의 단계에서 해당 변수를 제거(smooth-out)하는 성질을 가지고 있다. 이에 대한 자세한 내용은 Hall, Racine, and Li(2004)를 참고하라. 3장에서는 몇 가지 모형에 대해 추정된 평활 모수 값과 Racine, Hart, and Li (2006)의 붓스트랩 검정의 수행결과를 비교하겠다.

3. 국내 개봉영화의 온라인 구전효과

앞서 언급한 바와 같이 본 장에서는 온라인 구전의 함수로서 매출을 설명하는 비모수적 회귀모형을 고려하고 Li and Racine(2004)의 방법론을 적용한다. 사용된 자료는 2004년 12월부터 2007년 2월까지 국내에서 개봉된 420개 영화에 대한 매출, 영화특성, 개봉에 관한 자료와 온라인 구전에 관한 자료이다. 매출은 한국영화진흥위원회에서 개봉기간 전체에 대한 매출을, 온라인 구전변수는 포털사이트 naver.com에서 해당 영화에 대한 네티즌의 평점과 평점부여에 참여한 네티즌의 수를 사용하였다. Godes and Mayzlin(2004)는 온라인 구전의 규모(volume)와 확산(dispersion)의 두 가지 차원을 고려하였으나 본 분석에서는 규모 한 차원만을 고려하되, 규모를 측정하는 네티즌의 수와 더불어 평균평점을 사용한다. 본 장에서 고려하는 모형은 다음과 같다.

$$\log(y_i) = g(x_i^c, x_i^d) + \epsilon_i, \; i = 1, \cdots, 420, \tag{3.1}$$

단, y_i는 i번째 영화에 대한 매출, x_i^c는 i번째 영화에 대한 온라인 구전으로서 모두 연속형 변수라 가정한다. 즉, $x_i^c \equiv (x_{i1}^c, x_{i2}^c)$, 단, x_1^c는 log(네티즌의 수), x_2^c는 네티즌의 평균평점이다. x_i^d는 i번째 영화에 대한 영화특성 정보로서 $x_i^d \equiv (x_1^d, x_2^d)$, 단, x_1^d는 순

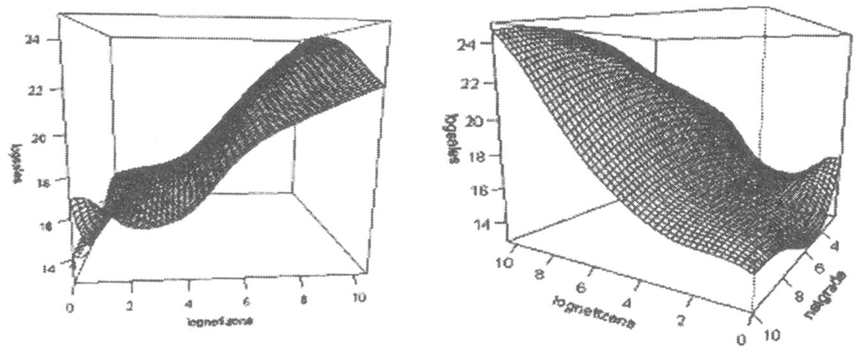
서형 자료로서 영화의 전체개봉월수, x_2^d는 명목형 자료로서 영화가 만들어진 지역을 나타낸다.

다음 (표1)은 2장에서 설명한 Li and Racine(2004)의 방법론을 모형(3.1)에 대한 적용한 결과이다. (표1)의 괄호 안의 숫자들은 전체 420개 자료를 350개의 training 자료와 70개의 evaluation 자료로 분할하여 추정평균제곱오차(predicted mean squared error)를 계산 한 값으로 모형2와 모형3의 오차가 가장 작음을 알 수 있다. 셀 안의 숫자는 평활모수의 추정치들로서 옆의 기호는 Racine, Hart, and Li(2006)의 붓스트랩 유의성 검정에서 붓스트랩 type I 검정을 붓스트랩 반복 200회로 실시한 결과 얻어진 유의확률을 0.01, 0.05, 0.1의 값과 비교한 결과이다. 자세한 사항은 (표1)에 대한 각주를 참고하라. (표1)에서 고려한 각 모형에 대한 설명변수가 모두 유의함을 알 수 있다.

(표1) 각 설명변수에 대한 평활모수 추정치와 predicted MSE

설명변수	모형1	모형2	모형3	모형4	모형5
	(1.15)	(1.01)	(1.01)	(1.13)	(1.03)
log(네티즌수)	1.20***	0.88***	0.72*	0.97***	1.22***
평균평점	1.51***	1.85*	1.90***	1.63***	1.68***
개봉기간(월)		0.72***			
개봉기간1개월이상			0.11***		0.09***
지역		0.07*	0.05	0.09*	

* 유의확률 < 0.1, ** 유의확률 < 0.05, *** 유의확률 < 0.01 ' ' 유의하지 않음



(그림 1) (표1)의 모형1에 대한 추정함수

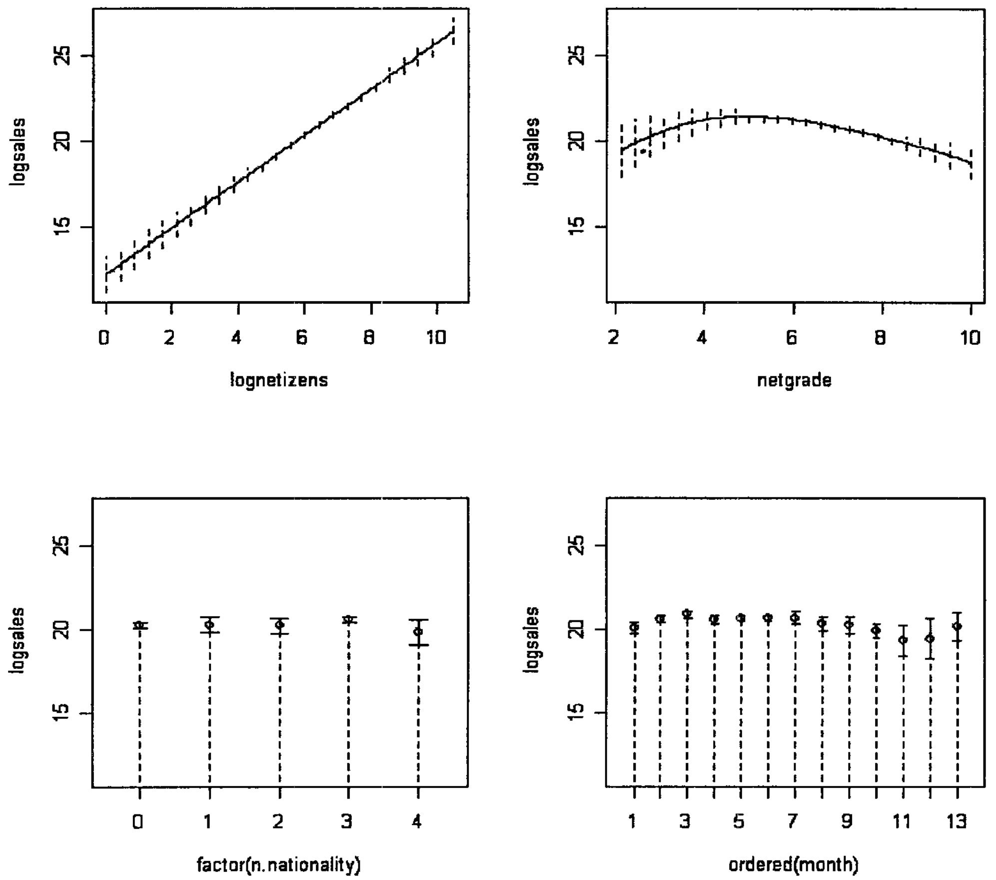
(그림1)은 (표1)의 모형1에서 추정된 함수로서 매출을 구전만으로 설명하여 추정한 함수이다. 함수의 형태를 파악하기 위하여 두 가지 다른 각도에서 함수를 그린 그림을 각각 두 칸으로 제시하였는데 세로축 logsales는 log(매출)을 lognetizens은 log(네티즌수)를 netgrade는 네티즌이 부여한 평균평점을 나타낸다. (그림1)에서 볼 수 있듯이 네티즌의 수가 증가함에 따라 매출이 증가하는 반면 네티즌의 수가 적을 때에는

네티즌의 평균평점이 증가함에 따라 매출이 다소 감소했다가 다시 증가하는 형태임을 알 수 있다. (표1)의 모형2는 모형1에 개봉월수와 영화가 제작된 지역을 추가적으로 고려한 모형이다. 모형2의 추정된 함수에 대하여 다른 설명변수들의 값을 자료의 중간 값으로 고정하여 각각의 변수가 매출에 미치는 영향을 나타낸 편회귀 그림(partial regression leverage plot)이 (그림2)에 제시되어 있다. (그림2)의 왼쪽 아래 칸의 factor(n.nationality)는 영화가 만들어진 지역을 나타내는 명목형 자료로서 1부터 4까지의 5개 범주값은 순서대로 오세아니아 및 동유럽 등의 기타, 한국, 아시아, 북미, 유럽을 나타내며, 오른쪽 아래 칸의 ordered(month)는 개봉 월수(1-12개월, 13은 12개월 이상인 경우)를 나타내는 순서형 자료이고, 선 위의 점들은 붓스트랩 오차 한계(bootstrap error bound)를 나타낸다. (그림2)에서 알 수 있듯이 모형2에서는 다른 변수들의 값들이 고정되었을 때 네티즌의 수와 매출의 관계는 거의 선형적이며 네티즌의 평균평점이 증가함에 따라 매출은 증가하다가 다시 감소하는 형태를 보인다. (표1)의 모형3은 모형1에 개봉기간 한달 이상에 대한 더미변수와 영화가 만들어진 지역을 추가적으로 고려하는 모형이다. (그림3)은 (표1)의 모형3에 대한 편회귀 그림으로 오른 쪽 아래 칸의 factor(onemonth)는 개봉기간 1개월 이상에 대한 더미변수를 나타낸다. (그림2)의 왼쪽 위칸의 그림과 (그림3)의 왼쪽 위칸의 그림을 비교하면 네티즌의 수에 대한 붓스트랩 오차 한계는 모형2에 비하여 모형3에서 큼을 알 수 있다. 결론적으로, 다른 변수들을 고정했을 때 온라인 구전과 매출의 관계는 선형이 아니지만 개봉월수를 순서형 변수로 고려하는 모형에서는 거의 선형관계를 보임을 알 수 있다. 참고로 (표1)에는 제시하지는 않으나 추가적으로 log(개봉 스크린수)를 고려한 모형을 적합한 결과 본 분석의 주요한 관심변수인 네티즌수에 큰 평활모수 추정치를 할당하여 제거(smooth-out)시켜 버리므로 본 분석에서는 제외하였다.

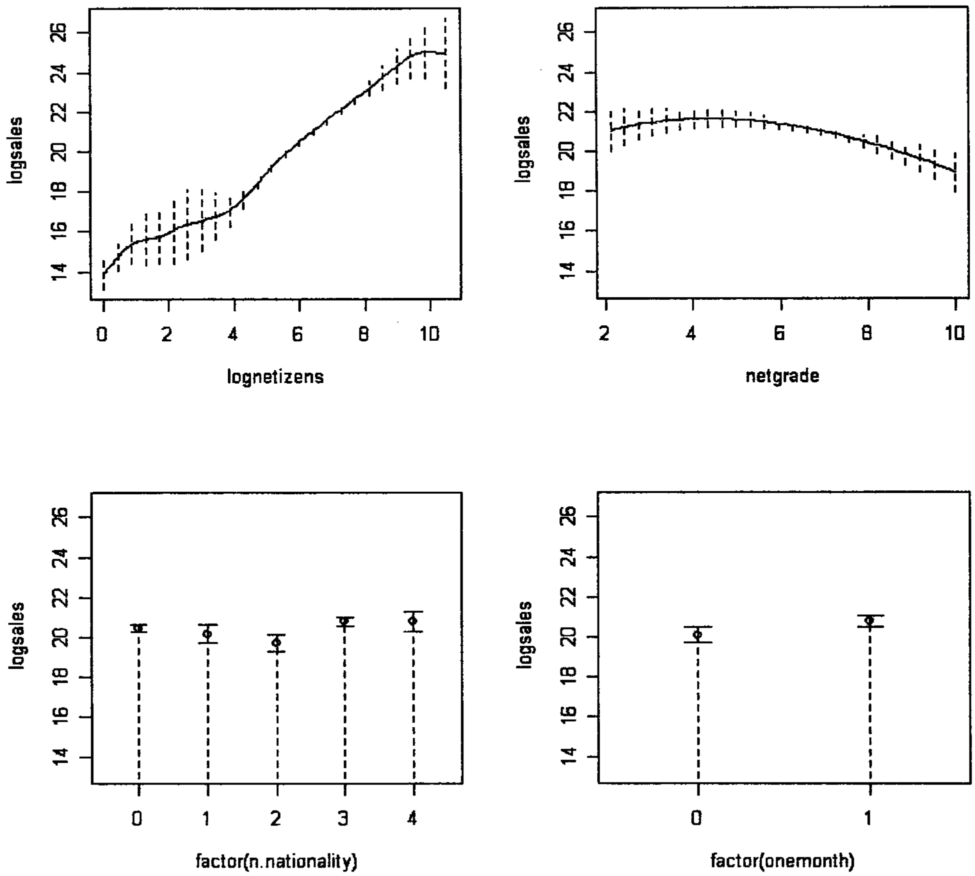
4. 결론 및 향후 연구과제

본 연구에서는 커널 회귀분석 방법론에 기초하여 온라인 구전과 영화특성의 함수로서 영화의 매출을 설명하였다. 온라인 구전의 규모의 측면인 평점부여에 참여한 네티즌의 수와 네티즌의 평균평점과 함께 영화가 제작된 지역과 개봉기간이 매출에 미치는 영향을 함수로 추정하였다. 기존의 구전효과 모형이 모수적 모형에 기초한데 반해 본 분석은 매출에 미치는 구전의 함수형태를 찾는 비모수적 접근법이라는데 의의가 있다 할 수 있다. 그러나, 본 분석에는 몇 가지 한계가 있으며 이는 자료가 전체 개봉기간에 대한 평균이나 합으로 이루어진 통합된 자료라는데 기인한다. 일반적으로 구전의 효과는 시차를 가지고 나타난다. Godes와 Mayzlin(2004)은 미국 드라마의 구전효과 측정에 관한 모수적 모형에서 구전의 시간적 측면을 고려하였는데 주별 구전자료를 이용함으로서 구전의 시간적 측면을 고려하였다. Chevalier와 Mayzlin(2006)은 두 개의 온라인 서점의 구전이 매출에 미치는 효과에 대한 모수적 모형에서 평균평점과 부정적 혹은 긍정적 구전의 영향에 대한 측정을 고려하였는데, 평점의 평균보다는 부정적 평점의 비율이 유의함을 발견하였다. 추후 분석에 있어서 온라인 구전과 매출의 시계열 자료를 이용 한다면 구전의 측정에 있어 시간적 측면을 고려할 수 있을 뿐 아니라 평점의 평균적인 구전효과가 아닌 부정적 혹은 긍정적 구전효과를 분리하여 모형화 할 수 있을 것이다. 또한, 사이트별 혹은 커뮤니티별 평점자료로서 온라인 구

전의 확산에 대한 측정이 가능하다면, 온라인 구전의 규모와 확산의 두 가지 차원에서 매출에 미치는 영향을 측정할 수 있을 것이라 사료된다. 이를 위하여 몇 개의 인터넷 사이트의 이질적인 커뮤니티에서 얻어진 온라인 구전을 개봉 주차에 따라 관측한 자료와 개봉 주차에 따른 매출을 이용하여 비모수적 방법론을 적용한다면 보다 정교한 구전효과 모형의 개발이 가능할 것이다.



(그림 2) (표1)의 모형2 에 대한 부분회귀그림



(그림 3) (표 1)의 모형3에 대한 부분회귀그림

참고문헌

1. Ahmad, I. A., and Cerrito, P. B. (1994). Nonparametric estimation of joint discrete-continuous probability densities with applications, *Journal of Statiatiscal Planning and Inference*, 41, 349-364.
2. Aitchison, J., and Aitken, C. G. G. (1976). Multivariate binary discrimination by the kernel method, Biometrika,. 63(3), 413-420.
3. Bierens, H. J. (1983). Uniform consistency of kernel estimators of a regression function under generalized conditions. *Journal of the American Statistical Association*, 78, 699-707.
4. Chevalier, J., D., and Mayzlin, D. (2006). The effect of word of mouth

on sales: online book review, *Journal of Marketing Research,*. 43(3), 345-354.
5. Fan, J., and Gijbels, I. (1995). *Local Polynomial Modeling and its Applications.* Chapman and Hall.
6. Godes, D., and Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication, *Marketing Science,*. 23(4), 545-560.
7. Hall, P., Racine, J., and Li, Q. (2004). Cross-validation and the estimation of conditional probability densities, *Journal of the American Statistical Associasion,* 99, 1015-1026.
8. Hurvish, C. M., Simonoff, J. S., and Tsai, C. L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion, *Journal of Royal Statistical Society B*, 60, 271-293.
9. Li, Q., and Racine, R. (2004). Cross-validated local linear nonparametric regression, *Statistica Sinica,* 14 485-512.
10. Li, Q., and J.S. Racine. (2007). *Nonparametric Econometrics.* Princeton University Press.
11. Racine, J.S., J. Hart, and Li, Q. (2006). Testing the Significance of Categorical Predictor Variables in Nonparametric Regression Models, *Econometric Reviews,* 25(4), 523-544.
12. Wand, M.P., and M.C. Jones. (1995). *Kernel Smoothing.* Chapman and Hall.

[2007년 10월 접수, 2007년 11월 채택]