

로그노말분포를 이용한 통화시간 분석 (Analysis of Call Duration with Log-Normal Distribution)

박상준*: 동서대학교 상경정보학부

민태기 : 한국통신

한민희 : 한국과학기술원 테크노경영대학원

요약문

통화시간에 대한 할인율 적용 혹은 짧은 통화에 요금을 부과하지 않은 정책 등 마케팅활동의 효과분석 및 마케팅전략을 수립하는데 있어서 통화시간 분석은 매우 중요하다. 기존연구들은 설명변수의 도입이 편리한 비례위험함수(proportional hazard function)을 활용해 왔다. 그러나 이 방법은 추정이 용이하지 않다는 약점을 갖고 있다. 특히 통화행태를 설명하기 위해 다양한 설명변수를 도입하는 경우, 기존에 상용화된 통계패키지로 추정하는데 한계점을 갖고 있다. 본 연구에서는 추정의 용이성에 초점을 맞추어, 실무자들이 손쉽게 통화시간에 대한 마케팅 노력의 효과를 분석할 수 있는 분석방법론을 제시하고, 제시된 분석방법을 이용하여 판촉활동(요금할인)이 통화대기시간과 통화시간에 미치는 영향을 실증분석 하고 있다.

1. 도입

음성통신 서비스 수요는 크게 가입수요와 통화수요로 나뉘어 진다. 통화수요는 통화빈도(호수: message frequency)와 통화시간(도수: total time of calling or call duration)으로 나뉘어 진다. 유선전화시장에서 가입수요는 이미 포화상태에 도달해 있다. 그러므로 유선전화의 수익구조에 있어서는 통화수요가 가입수요보다 더 영향을 준다고 할 수 있다.

통화수요에 관한 연구는 시장수준에서 모형화와 개인수준에서의 모형화로 나누어 생각해 볼 수 있다(한민희 외 1992). 시장전체 자료를 이용한 기존연구로는 호수와 도수에 대한 가격탄력성 측정(Park, Wetzel, and Mitchell 1983), 도시간 장거리 전화의 "point-to-point" 가격탄력성 측정(Parcy 1983) 연구가 있었다. 이들 연구들은 통화시간에 대한 좀 더 정밀한 연구를 위해서 개인수준 자료를 이용한 모형화 및 분석이 강조하고 있다. 개인수준 자료를 이용한 연구로는 가입자특성을 반영한 회귀분석(Infosino 1980), 장거리 전화의 도수와 호수와의 관계 모형화(Veitch 1985), Generalized Gamma 분포를 이용한 시외전화 도수 분석(Cameron and White 1990), 비례위험함수(proportional hazard function)을 이용한 개별통화시간 모형화(한민희 외 1992) 연구 등이 있다.

비례위험함수를 이용한 통화수요 모형화는 일반적인 확률분포뿐 아니라 다양한 분포의 특성을 반영할 수 있고 기타 설명변수를 모형에 도입하는 것이 편리하다는 장점 때문에 많은 실증연구들이 수행되었다(Heckman and Singer 1984, Jain and Vilcassim 1991 등).

그러나 비례위험함수를 이용하여 통화수요를 분석하는데는 추정상 어려움이 존재한다. 일반적으로 이런 유형의 모형화 방법은 최우추정법(Maximum Likelihood Estimation) 방법이 요구되는데, 설명변수가 많을 경우 상용화된 패키지로 분석하는데 한계가 있으며, 추정시간이 오래 걸린다는 단점을 갖고 있다. 본 연구에서는 추정의 용이성에 초점을 맞추어 Lognormal 분포를 이용하여 통화시간 분포의 신축성을 유지하였고 회귀분석의 최대 약점은 통화수요가 음이 될 수 있다는 약점을 극복하였다. 또한 최소자승법(OLS) 모수를 추정하는 추정법을 제시하였다. 최소자승법은 추정이 매우 용이할 뿐 아니라 추정시간이 매우 짧다는 장점을 갖고 있다. 대규모의 과금자료를 감안하면 이 장점은 실무자에게 중요하게 부각되는 사항이다. 또한 본 연구에서는 본 연구가 제시한 모형화방법을 이용하여 통화시간에 판촉활동이 어떤 경로로 구매수량(quantity choice)에 영향을 미치는가 실증분석하고 그 결과를 제시한다. 통화시간에 영향을 줄 수 있는 변수로는 다양한 변수를 생각할 수 있다: 가격,

통화시간대, 요일 등. Neslin, Henderson, and Quelch (1985) 는 어떤 경로를 통하여 구매수량의사결정에 이르는가에 관하여 연구하였다. 본 연구에서는 그들의 연구프레임을 기본으로 하여 통화시간에 영향을 주는 설명변수를 파악하고자 하였다.

$$v = \sum_k b_k x_k \quad (6)$$

특정변수 x_k 가 v 와 양의 관계가 있으면, 즉 모수 b_k 가 양수라면 특정변수 x_k 가 증가할수록 v 가 증가된다. v 의 증가에 따라 통화시간 분포가 어떻게 변화하는지 살펴보면 <그림 1> 과 같다. v 가 감소할수록 왼쪽으로 통화수요가 편이됨을 알 수 있다. 즉 마케팅활동을 하면 v 가 증가하고, 통화시간이 길어진다는 것을 의미한다.

2. 로그-노말 분포를 이용한 통화수요 모형화

음이 아닌 무작위 변수 (random variable) e 가 로그-노말 (Log-Normal) 분포를 따른다고 가정하면, 로그-노말의 p.d.f 는 식 (1) 과 같이 표현될 수 있다.

$$f(e; u, \sigma) = \frac{1}{\sqrt{2\pi\sigma e}} \exp\left(-\frac{[\ln(e)-u]^2}{2\sigma^2}\right), e \geq 0 \quad (1)$$

u 를 영 (zero) 으로 가정하면 식 (1) 은 식 (2) 와 같이 단순한 형태의 로그-노말 함수가 된다.

$$f(e; u, \sigma) = \frac{1}{\sqrt{2\pi\sigma e}} \exp\left(-\frac{[\ln(e)]^2}{2\sigma^2}\right), e \geq 0 \quad (2)$$

통화시간을 나타내는 변수 y 를 다음과 같이 식 (3) 으로 표현한다.

$$y = \exp(v)e \quad (3)$$

식 (3) 에서 v 는 통화시간을 늘리거나 줄이는데 영향을 줄 수 있는 변수의 효과를 나타낸다. 식 (3) 으로 표현된 통화시간의 분포를 유도해 보면 다음과 같다.

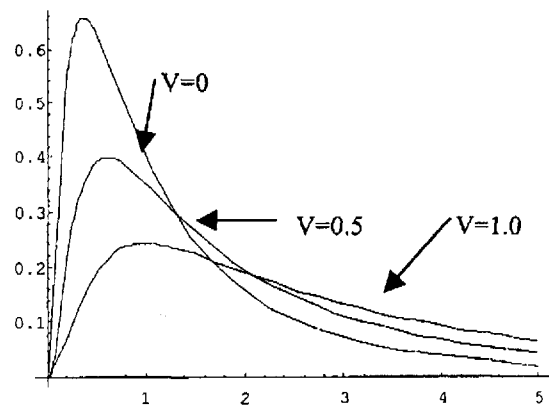
$$\begin{aligned} G(y) &= P(Y \leq y) = P(\exp(v)e \leq y) \\ &= P(e \leq \exp(-v)y) \\ &= \int_0^{\exp(-v)y} f(e)de \end{aligned} \quad (4)$$

결국 통화시간의 p.d.f 는 식 (5) 로 정리될 수 있다.

$$f(y; u, \sigma) = \frac{1}{\sqrt{2\pi\sigma y}} \exp\left(-\frac{[\ln(y)-v]^2}{2\sigma^2}\right) \quad (5)$$

식 (5) 에서 v 를 통화시간에 영향을 주는 변수들의 선형결합 함수로 나타내면 식 (6) 과 같이 표현될 수 있다.

<그림 1> $\sigma=1$ 일때 v 에 따른 통화시간 분포의 변화



모수의 추정을 위해 식 (3)에서 양변에 log 를 취하고, v 에 식 (6) 을 대입하면 식 (7) 이 유도된다.

$$\ln(y) = \sum_k b_k x_k + \ln(e) \quad (7)$$

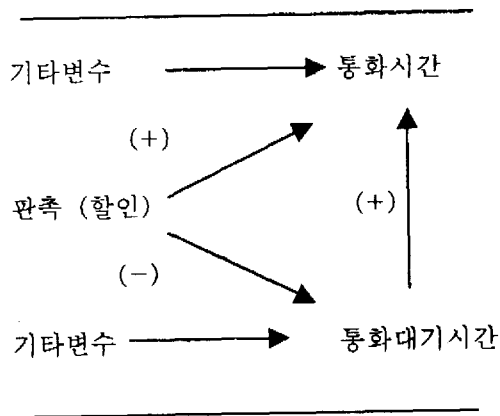
식 (3) 에서 $\ln(e)$ 는 평균이 "0" 이고 분산이 σ^2 인 정규분포 (Normal Distribution) 를 따른다. y 가 모수 (u, σ^2) 를 따르는 로그-노말 분포일 때 $\ln(y)$ 는 평균이 u 이고 표준편차 σ 인 정규분포를 따른다는 것을 잘 알려져 있는 사실이다 (Devore, 1990). 그러므로 식 (7) 을 이용하여 로그-노말 분포의 모수를 추소자승법 (OLS) 으로 추정할 수 있다.

3. 판촉활동 분석

<그림 2> 에서 보는 바와 같이, 판촉활동 (할인) 은 통화시간을 늘리면서, 동시에 통화대기시간을 줄일 것으로 예상된다. 줄어든 통화대기시간

은 구매량을 줄일 것으로 예상된다. 결국 판촉활동은 직접적으로 통화시간(구매량)을 늘리는 효과와 통화 대기시간을 매개로 하여 통화시간(구매량)을 감소시키는 효과를 갖을 것으로 예상하였다.

<그림 2> 판촉(할인)이 통화시간에 미치는 영향 분석



통화상품에 있어서, 통화시간에 영향을 주는 변수는 판촉활동(할인시간대), 대기시간(waiting time)과 더불어 다양한 영향변수를 고려할 수 있고, 또한 대기시간에 대해서도 다양한 영향변수를 고려할 수 있다. 대기시간은 크게 3가지 종류로 구분이 가능하다. 첫째는 가족 구성원에 상관없이 착신번호에 상관없이, 한 전화번호에서 직전 통화가 끝난 후 다음 통화가 시작될 때까지의 시간이다(대기시간 1). 두번째는 가족 구성원에 상관없이 한 착신번호에 통화가 끝난 후, 그 착신번호로 다음 통화가 시작될 때까지의 시간이다(대기시간 2). 마지막으로 한 구성원이 한 착신번호로의 통화가 끝난 후, 그 구성원이 그 착신번호로 다음 통화가 시작될 때까지의 시간이다(대기시간 3). 그러나 세 번째 대기시간은 현실적으로 측정하기 어렵기 때문에 본 연구의 분석대상에서 제외하였다. 기타 변수로는 소득수준, 전화대상(부모, 형제/자매, 등), 요일효과(월/화, 수/목, 금/토, 일), 지역(개봉, 신사, 창동 등) 등의 변수가 사용되었다.

분석결과에서 판촉활동(할인)이 통화시간에 영향을 주며($\alpha = 0.01$), 통화대기시간에는 영향을 주지 않는다($\alpha = 0.1$)는 결과를 얻었다. 그러나 통화대기시간 2의 경우에는 유의수준 $\alpha = 0.15$ 하에서는 유의하게 판촉활동(할인)이 영향을 준다는 결과를 얻었다. 이런 결과는 판촉활동이 통화량을 늘리는 효과와 감소시키는 효과 두 가지 모두 갖고 있다는 의미로 해석되며, 판촉여부, 판촉수준 결

정등에 이에 대한 고려가 필요함을 의미한다.

참고문헌

- [1] 한민희 외 (1992), "통신상품관리모형의 개발" 한국통신보고서
- [2] Cameron, Trudy Ann and Kenneth J. White (1990), "Generalized Gamma Family Regression Models for Long Distance Telephone Call Durations," in DeFontenay et al. (eds.) *Telecommunications Demand Modeling: An Integrated View*, North Holland.
- [3] Heckman, J. and B. Singer (1984), "Relationships Between the Demand for Local Telephone Calls and Household Characteristics," *The Bell System Technical Journal*, 59, pp.931-53.
- [4] Infosino, W. J. (1980), "Relationships Between the Demand for Local Telephone Calls and Household Characteristics," *The Bell System Technical Journal*, 59, pp.271-320.
- [5] Jain D. C. and N. J. Vilcassim (1991), "Investigating Household Purchase Timing Decisions: A conditional Hazard Function Approach," *Marketing Science*, 10, pp. 1-23.
- [6] Neslin S. A, C. Henderson, and J. Quelch (1985), "Consumer Promotions and the Acceleration of Product Purchases," *Marketing Science*, 4 (Winter), pp.147-65.
- [7] Pacey, Patricia L. (1983), "Long Distance Demand: A Point-to-Point Model," *Southern Economic Journal*, Vol. 49, pp. 1094-107.
- [8] Park, Rolla Edward, Bruce M. Wetzel, and Bridger M. Mitchell (1983), "Price Elasticities for Local Telephone Calls," *Econometrica*, 51, pp.1699-730.
- [9] Veitch, J. G. (1990), "Relation of Conversation Time to Message to Message Frequency in Residential Usage," in De Fontenay et al. (eds.), *Telecommunications Demand Modeling: An Integrated View*, North Halland