

통신서비스산업 예측모형 예측력 비교 분석

The Comparative Evaluations of Telecommunications Service Forecasting Models for Forecasting Performance

조상섭(S.S. Jo)

경제분석연구원 선임연구원

정동진(D.J. Jeong)

경제분석연구원 선임연구원

본 연구는 현재 통신서비스 산업에서 가장 많이 사용하고 있는 5개 예측모형(단순 성장 모형, 단순 Logistic 모형, Gompertz 모형, 확장 Bass 모형, 시간 변동 Bass 모형)을 이용한 초고속 인터넷 가입자에 대한 예측력을 비교 평가하는 데 있다. 예측모형의 추정 방법으로 비선형 회귀방정식(nonlinear regression)을 사용하여 추정의 효율성을 높였다. 예측력 비교분석 기준은 (i) 포화점에 대한 타당성 (ii) 모수에 대한 통계적 유의성 (iii) 실제치 대비 예측치에 대한 AAD 기준을 통하여 예측모형의 예측력을 비교 평가하였다. 본 연구에서 실시한 방법론에 따라 다섯 가지 통신서비스 예측모형의 예측력을 분석한 결과 가장 작은 AAD를 나타낸 예측모형은 Log-Logistic 모형으로 나타났으며, 가장 큰 AAD를 나타낸 예측모형은 단순 Logistic 모형으로 나타났다. 또한 AAD 기준에서 보면 일반적으로 많이 사용하고 있는 Gompertz 예측모형과 Bass 모형 중에서는 Gompertz 예측모형이 더 우월한 것으로 나타났다.

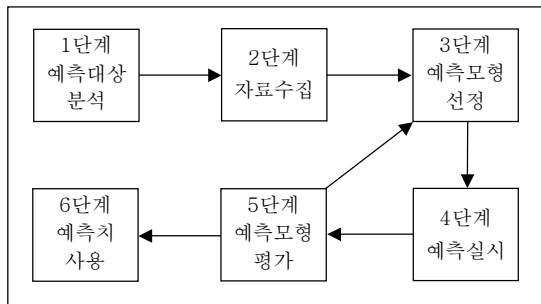
I. 연구 개요

본 연구의 목적은 자주 사용하는 예측모형들이 갖는 초고속 인터넷 가입자에 대한 예측력을 평가하는 데 있다. 통신서비스 예측모형에 대한 예측력 비교 평가는 다음과 같은 세 가지 점에서 매우 중요하다. 첫째, 정보통신산업 여러 방면에서 통신서비스 예측은 통신사업자 및 통신장비제조업자 그리고 통신산업 정책시행자에게 중요한 관심사이다. 따라서 어떤 예측모형에 의한 통신서비스 예측을 실시하였는지에 대한 결과치에 대하여 이해당사자의 이해관계에 영향을 준다. 둘째, 현재 주로 사용되고 있는 다수 예측모형은 통신서비스 예측을 위하여 만들어진 모형이 아니다. 따라서 이에 대한 검증이 연구자의 입장에서는 매우 필요하다. 마지막으로 통신서비스 예측관계자는 미래 통신서비스 예측모형 개발을 위해서 기존의 통신서비스 예측모형을 비교 분석함

으로써 통신환경에 맞는 각 모델의 장점에 대한 비교정보가 필요하다.¹⁾

기존 통신서비스 예측모형에 관한 연구는 활발히 진행되고 있다. 특히 김국희 외[1]는 통신산업환경을 중심으로 일반적인 확산모형에 대하여 자세히 설명하고 있지만, 실제적 자료를 이용한 예측치 제공 또는 예측모형의 비교는 생략하고 있다. 그리고 본 연구와 가장 밀접하게 관련된 Meade et al.[2]의 연구는 통신서비스를 바탕으로 예측모형을 비교 분석한 점에서 가장 포괄적인 모형으로 볼 수 있으나, 비교평가에 있어서 중요한 모수 추정방법에 대한 설명이 빈약하다.

1) 본 연구의 부차적 목적은 2005년까지 초고속 인터넷 가입자를 통신서비스별(ADSL, C-Modem) 및 통신서비스 제공업자별(4개 사업자)로 예측하는 데 있다. 이에 대한 예측자료를 원하는 독자는 저자에게 연락 바람(e-mail: choss@etri.re.kr).



<자료>: Armstrong(2001, p. 363)[3]에서 발취

(그림 1) 예측에 대한 단계별 절차

본 연구의 중요성은 (그림 1)에서 보듯이, 예측모형평가 정보가 다시 예측모형선정에 영향을 주고, 이 선정모형에 의하여 예측치가 결정된다는 데 있다. 따라서 예측모형의 정확한 비교는 미래 예측력을 높이는 데 중요한 역할을 한다.

본 연구에서 실시한 방법론에 따라 다섯 가지 통신서비스 예측모형의 예측력을 분석한 결과 가장 작은 AAD(Average Absolute Difference)를 나타낸 예측모형은 Log-Logistic 모형으로 나타났으며, 가장 큰 AAD를 나타낸 예측모형은 단순 Logistic 모형으로 나타났다. AAD 관점에서 일반적으로 사용하고 있는 Gompertz 예측모형과 Bass 모형 중에서는 Gompertz 예측모형이 더 우월한 것으로 나타났다.

본 연구의 전개 순서는 다음과 같다. 제 II장에서는 본 연구에서 비교 분석하게 될 5개 일반적인 예측모형을 간단하게 기술하고, 예측모형비교에서 매우 중요한 위치를 차지하고 있는 모수추정방법론 중 비선형 추정방법론의 계량적 의의 및 추정기법을 설명하였다. 제 III장에서는 본 연구의 핵심인 예측모형의 예측력을 실제 우리나라 초고속 인터넷 가입자 자료를 통하여 비교 분석하였으며, 비교분석 기준도 제 III장에서 함께 제시하였다. 마지막 장에서는 본 연구를 요약하고, 앞으로의 연구방향 및 한계를 간단하게 서술하였다.

II. 예측모형과 추정기법

통신산업에서 주로 사용되는 예측모형은 간단한

성장 함수 모형에서 비교적 복잡한 NNE(Neural Network Estimation) 기법에 이르기까지 매우 광범위하다. 예측모형 선정기준은 첫째, 통신서비스 가입자를 비교적 간단하고, 빠르게 예측할 수 있는 예측모형으로 국한하였다. 둘째, 현재 통신서비스 예측 실무자들이 가장 많이 사용하는 모형을 대상으로 하였다. 마지막으로 예측비용 즉 시간이 비교적 적게 드는 예측모형을 기준으로 하였다. 따라서 복잡하고 실무적으로 사용하기에 어려운 NNE와 Kalman 필터링을 이용한 상태공간예측모형에 대한 비교는 생략하기로 한다. 또한 비교적 서로 유사한 예측모형들은 예측력 비교에서 제외하였다. 본 연구는 이러한 모형 선정기준에 의하여 다음과 같은 5개 예측모형을 선정하였다.

1. 통신서비스에서 사용되는 예측모형

먼저 가장 쉽고 많이 사용되는 모형으로서 간단한 지수성장함수를 이용한 예측모형이 있다. 이 모형의 특징은 성장곡선상에 변곡점이 없이 지수적으로 성장하는 것을 가정한다는 것이다. 따라서 통신서비스 가입자가 포화점에 가까이 갈수록 그 성장변화율은 일정하게 감소하게 된다. 통신서비스 누적가입자 수 y 대한 지수적 성장함수식은 (1)과 같다.

$$y_t = a - c \exp(-bt) \quad (1)$$

여기서 a 는 통신서비스 가입자의 포화점을 나타내는 모수이고, c 는 성장폭을 조정하는 모수이며, b 는 확산율을 나타내는 모수라고 볼 수 있다.

다음으로 단순 Logistic 성장모형이 있다. 이 예측모형 역시 매우 단순하고 예측계산이 용이하며, 통신서비스 특유의 형태인 S자형 성장패턴을 갖는다는 점에서 실무적으로 많이 사용되고 있다. 그러나 이 모형의 문제점은 변곡점이 성장함수 전체의 1/2 수준에서 결정된다는 단점이 있다. 통신서비스 누적가입자 수 y 대한 Logistic 성장함수식은 (2)와 같다.

$$y_t = \frac{a}{1+c \exp(-bt)} \quad (2)$$

여기에서 사용된 기호는 앞에서 언급한 지수성장 함수와 같은 의미를 갖는다.

세번째는 통신산업에서 예측모형으로 Bass 예측 모형과 함께 가장 대중화된 Gompertz 예측모형이다. 단순 Logistic 모형과 비교한 이 모형의 특징은 포화점 앞에서 변곡점이 발생하며 비대칭적이라는 점이다. 따라서 단순한 Logistic 모형보다는 실무적으로 많이 사용되고 있다. 통신서비스 누적가입자 수 y 대한 Gompertz 예측함수식은 (3)과 같다.

$$y_t = a \exp(-c(\exp(-bt))) \quad (3)$$

다음으로 Bass[4]에 의해서 내구적 소비재의 (durable goods) 특성을 고려한 성장곡선이 도입되면서 통신서비스 예측에서 보편화된 Bass 예측모형이 있다. 이 예측모형의 의미는 단순한 함수적 표현에 의한 예측모형보다 경제적 의미를 부여한 예측모형이라는 데 있다. 즉 창조적인 확산과 모방적인 확산이라는 두 가지 요소에 의하여 모든 소비재확산은 이루어지며, 이를 예측모형적으로 해석한 데 이 모형의 의미가 있다. 통신서비스 누적가입자 수 y 에 대한 Bass 예측함수식은 (4)와 같다.

$$y_t = \frac{a - c_1 \exp(-bt)}{1 + c_2 \exp(-bt)} \quad (4)$$

여기서 c_i ($i=1,2$)는 확산크기를 조정하는 모수이다. 다른 기호의 의미는 앞 예측모형에서 같다.

다음 예측모형으로는 Tanner[5]에 의하여 소개된 Log-Logistic 성장모형이다. 이 예측모형의 특징은 단순 Logistic 모형의 단점인 변곡점이 대칭적인 수준에서 결정된다는 단점을 보완하여 비대칭적인 지점에서 결정되도록 한 예측모형이다. 통신서비스 누적가입자 수 y Log-Logistic 성장함수식은 (5)와 같다.

$$y_t = \frac{a}{1+c \exp(-b(\ln(t)))} \quad (5)$$

Log-Logistic 예측모형과 앞에서 본 다른 예측 모형과의 큰 차이는 시간변수에 대한 로그 변환을 통하여 시간에 따라 변곡점의 발생시점을 비대칭적으로 변환했다는 점이다.

이상 5개 예측모형을 통하여 세 가지 관점에서 예측력을 비교 평가하였다. 첫째, 통신서비스 가입자 포화점에 대한 타당성 기준이다. 둘째, 예측모형의 모수에 대한 통계적 유의성에 관한 기준을 보았다. 마지막으로 AAD를 통하여 실제 5개 가입자에 대한 예측치와 비교함으로써 각 예측모형의 예측력을 평가하였다.²⁾

2. 예측모형의 추정기법

본 절에서는 5개 예측모형에 대한 모수추정방법론을 간단하게 기술하기로 한다. 본 연구에서 사용한 예측모형 추정방법론의 중요성은 다음과 같다. 첫째, 최소자승법(OLS)으로 대표되는 선형회귀추정방법은 효과적인 추정을 하기 위해서는 실제 자료 변화 또는 추정하고자 하는 비선형식의 변환이 필수적이다. 그러나 인위적인 비선형식의 선형형태로의 변환은 문제점이 발생할 수 있다. 둘째, 선형추정방법론은 실제 데이터의 변환을 수반할 수 있다. 따라서 예측의 본질을 흐릴 수 있다(data explain themselves). 따라서 본 연구에서는 비선형모수추정방법론을 사용하여 각 예측모형의 예측력을 비교 분석하였다. 이를 간단히 설명하면 다음과 같다.³⁾

(6)과 같은 앞에서 살펴본 일반적인 함수형태의 통신서비스 예측모형이 있다고 가정하자.

$$y_t = f(x_t, \beta) + e_t \quad (6)$$

여기서 β 는 추정하고자 하는 모수 벡터를 의미하며, e_t 는 잔차항을 의미한다. 이때 추정 잔차를 최소로 하는 기준으로 추정치를 추정하면 (7)과 같다.

2) 통계적으로나 경제적으로나 의미가 없는 단순 성장모형을 제외한 4개 예측모형에 대하여 실시하였다.
3) 비선형 추정방법론에 대한 자세한 소개는 Judge et al.[6]를 참조하기 바라며, 일반적인 추정방법론에 따른 비선형 추정 방법론에 대한 자세한 소개는 Bera et al.[7]를 참조.

$$S(\beta) = \sum_{i=1}^T [y_i - f(x_i, \beta)]^2 \quad (7)$$

모수추정치 기준이 잔차를 최소로 하는 추정치를 구하는 방식이므로 각 β 에 대하여 최소값을 구하는 방식과 같다. 그러나 선형함수 가정과 달리 비선형함수 가정에 의하여, 최소를 위한 1차조건에서 (8)과 같은 비선형방정식이 나타난다. 즉

$$\frac{\partial S}{\partial \beta} = -2 \sum_{i=1}^T [y_i - f(x_i, \beta)] \left(-\frac{df(x_i, \beta)}{d\beta} \right) = 0 \quad (8)$$

(8)을 수치적으로 풀기 위해서는 $f(x, \beta)$ 를 근사적 추정치로 사용하며, 현재 가장 많이 사용하는 방법으로는 Taylor Series 근사치 적용방법으로 풀고 있다. 즉 적절한 초기값 β 벡터에 대하여 컴퓨터에 의존하여 수치적으로 계산하는 과정을 이용한다. 또한 지속적으로 발생하는 미지수 β 는 일정하게 작은 간격으로 연속적 계산을 하여 최후계산 값과 현시점에서 계산된 값과 같아지는 때까지, 즉 수렴(convergence)에 이를 때까지 상기에서 설명한 계산과정을 수행하게 된다.⁴⁾

비선형 추정방법의 추정량은 다음과 같은 특성이 있다. 첫째, 비선형 특성상 선형모형에서 볼 수 있는 최적선형불편추정량(BLUE)의 특성을 지니지 않는다. 둘째, 그럼에도 불구하고 대표본에서는 일치성과 샘플 크기를 크게 하면 \sqrt{T} 속도로 일정 분산을 갖는 정규분포에 수렴하게 된다.

이상에서 본 바와 같이 비선형 추정방법은 장점과 단점을 동시에 가지고 있다. 따라서 실무적인 차원에서는 추정에 드는 비용(시간)을 줄이기 위하여 일반적으로 선형추정방법론을 사용하는 경향이 높다. 그러나 추정하고자 하는 함수 형태와 모수와 모수간의 관계를 정확하게 모르는 불확실성에서는 비선형방법론이 더 큰 장점이 있다.

4) 이런 일련의 과정을 Gauss-Newton Algorithm 과정이라고 한다.

III. 예측력 비교 분석 결과

본 연구는 1999년 6월 초고속 인터넷 상용화 시점에서 2001년 11월까지 가입자 수를 대상으로 5개 통신서비스 예측모형을 평가하였다. 가입자 수에 대한 자료는 정보통신부에서 다운받았으며, 분석 패키지는 Gauss 3.1을 이용하여 계산하였다.⁵⁾

제 II장에서 기술한 5개의 일반적 통신서비스 예측모형에 대한 각 모수 추정치를 <표 1>과 <표 2>에 제시하였다. 먼저 5개 초고속 가입자 예측모형을 비선형추정량으로 추정된 결과는 다음과 같이 요약된다. 첫째, 단순성장 모형에 의한 추정치 중 포화점 확산율에 대한 추정치가 경제적인 측면에서 볼 때 타당하지 못하게 나타났다. 둘째, 다른 네 개의 예측모형은 경제적인 의미에서 별다른 문제점을 발견하지 못했다. 셋째, 추정치간 상관계수는 비교적 높게 나타남으로써 서로 간에 공선성 문제가 있음을 알 수 있다. (그림 2)는 한 예로 Bass 예측모형의 추정치에 의한 예측치와 실제치를 비교한 그림이다.

다음에는 추정된 모수를 사용하여 2001년 7월에서 2001년 11월까지 5개월 실제 가입자 수와 예측 가입자 수에 대한 절대적 평균량(average absolute difference) 기준을 통하여 비교한 것을 <표 3>에 나타냈다.⁶⁾

이 기준을 근거로 각 통신서비스 예측모형에 대한 예측력을 비교 평가하면 다음과 같은 순서이다. 첫째, 가장 작은 ADD를 나타낸 예측모형은 Log-Logistic 모형으로 나타났으며, 가장 큰 AAD를 나타낸 예측모형은 단순 Logistic 예측모형으로 나타났다. 예측모형의 평가기준을 AAD로 본다면, 일반적으로 가장 많이 사용하고 있는 Gompertz 예측모형이 Bass 예측모형보다는 더 우월한 것으로 나타났다.

5) 만일 본 예측모형에 대한 2005년까지 월별 예측치를 원하는 독자는 저자에게 직접 연락하기 바람.

6) 본 연구에서는 지면관계상 ADSL 예측에 대한 평가만 하였으나, 각 예측모형에 대한 케이블 모뎀 예측력도 똑같이 나왔다. 따라서 케이블 모뎀에 대한 결과는 생략하기로 한다. 만일 이에 대한 결과를 원하는 독자는 직접 본 연구자에게 연락 바람.

<표 1> ADSL에 대한 5개 예측모형 Nonlinear 추정 결과

예측모형	포화점(a)	c		확산율(b)	반복횟수	추정시간 상관계수			
단순 성장 모형	-762,976 (-4.223)	-516,908 (-4.260)		-0.0863 (-10.34)	40	1	0.976	-0.954	
						0.976	1	-0.995	
						-0.954	-0.995	1	
단순 Logistic 모형	4,000,677 (32.013)	172.824017 (7.103)		0.275201 (25.429)	14	1	-0.683	-0.860	
						-0.683	1	0.953	
						-0.860	0.953	1	
Gompertz 모형	5,778,915 (28.071)	9.611580 (25.43)		0.117477 (29.69)	13	1	-0.878	-0.968	
						-0.878	1	0.967	
						-0.968	0.967	1	
확장 Bass 모형	4,250,547 (26.25)	9,892,164 (4.881)	107.072 (5.52)	0.24516 (19.22)	20	1	0.321	-0.909	-0.788
						-0.909	1	-0.454	-0.541
						-0.909	-0.454	1	0.969
						-0.788	-0.541	0.969	1
시간 변동 Logistic 모형	5,864,536 (21.144)	69,575.50 (4.66)		3.57854 (36.95)	41	1	-0.864	-0.928	
						-0.864	1	0.989	
						-0.928	0.989	1	

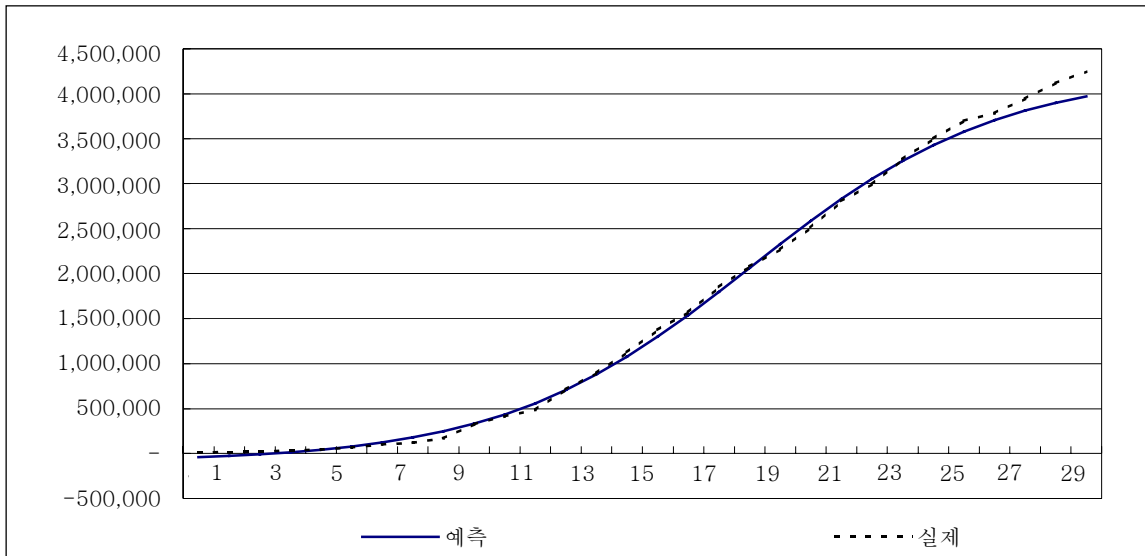
주 1) () 안은 추정치 t값을 나타냄
 2) Bass 모형의 경우 추정해야 할 모수 개수가 네 개이며, 두번째 행은 c1이고 세번째 행은 c2를 나타냄
 3) 상관계수는 순차적으로 모수1과 모수2 그리고 모수3 간의 관계이며, Bass 모형인 경우 마지막 열이 모수4와의 관계임

<표 2> 케이블 모뎀에 대한 5개 예측모형 Nonlinear 추정 결과

예측모형	포화점(a)	c		확산율(b)	반복횟수	모수간 상관계수			
단순 성장 모형	-728,292 (-4.168)	-566,059 (-4.052)		-0.064 (-8.132)	47	1	0.991	-0.976	
						0.991	1	-0.995	
						-0.976	-0.995	1	
단순 Logistic 모형	2,247,312 (20.471)	90.273 (5.291)		0.251 (15.867)	11	1	-0.641	-0.859	
						-0.641	1	0.933	
						-0.859	0.933	1	
Gompertz 모형	3,174,375 (11.525)	7.350557 (12.558)		0.109119 (11.804)	10	1	-0.840	-0.966	
						-0.840	1	0.947	
						-0.966	0.947	1	
확장 Bass 모형	2,515,344 (13.06)	4,861,765 (5.814)	42.66 (3.67)	0.202204 (9.763)	10	1	0.383	-0.909	-0.791
						0.383	1	-0.476	-0.558
						-0.928	-0.476	1	0.985
						-0.791	-0.558	0.985	1
시간 변동 Logistic 모형	3,702,316 (7.098)	9,959.50 (2.671)		2.8999 (14.81)	32	1	-0.834	-0.935	
						-0.834	1	0.975	
						-0.935	0.975	1	

주 1) () 안은 추정치 t값을 나타냄
 2) Bass 모형의 경우 추정해야 할 모수 개수가 네 개이며, 두번째 행은 c1이고 세번째 행은 c2를 나타냄
 3) 상관계수는 순차적으로 모수1과 모수2 그리고 모수3 간의 관계이며, Bass 모형인 경우 마지막 열이 모수4와의 관계임

위의 결과를 종합해 볼 때, 초고속 인터넷 서비스 가입자 예측력은 Log-Logistic 예측모형이 가장 우수하다고 볼 수 있다. 그러나 추정 기법과 실제치와 예측치 비교시점을 고려한다면 Log-Logistic 모형



(그림 2) Bass 모형에 의한 ADSL 예측치와 실제치(1999년 6월~2001년 11월)

<표 3> 4개 통신서비스 예측모형 기준에 의한 ADSL의 평가결과

예측모형	AAD	상대적 순위
단순 Logistic 모형	265,187	4
Gompertz 모형	76,745	2
Bass 모형	164,591	3
Log-Logistic 모형	55,457	1

의 우월성에 대한 평가에는 한계가 있을 수 있다. 즉 한 예로 비선형 추정방법에서 선형 추정방법으로 추정방법을 달리했을 경우 위 예측모형의 결과는 달라질 수 있다는 점이다.

IV. 결론 및 한계

본 연구는 최근에 깊은 관심을 나타내고 있는 5가지 통신서비스 예측모형의 예측력을 추정치의 통계적 유의성과 AAD 관점에서 비교 평가하였다. 5개 예측모형 중에서 단순 지수성장모형의 경우는 포화점이 부의 수치가 나오므로써 다음 단계의 분석에서 제외하였다. 그러나 네 개 예측모형의 경우 통계적 모수 유의성 및 경제적 의미 측면에서 유사하게 평가되었다. AAD를 기준으로 예측모형을 비교했을

경우, Log-Logistic 모형이 가장 작은 수치 즉 가장 예측력이 높은 것으로 나타났다. 실무적 차원에서 가장 많이 사용되고 있는 Gompertz 모형과 Bass 모형과의 비교는 Gompertz 모형이 더 우월한 것으로 나타났다.

그러나 몇 가지 본 연구의 한계점이 존재한다. 첫째, 현재 사용하고 있는 예측모형들은 본 연구에서 분석하고 있는 예측모형보다 훨씬 더 복잡하고 다양하다. 본 연구에서 제외한 NNE와 Kalman 필터를 이용한 상태공간모형의 예측력 평가는 미래 연구방향이 될 것이다. 둘째, 본문 중에서 지적했듯이 추정 기법과 실제치와 예측치 비교시점 등을 고려한다면, 본 연구결과에 대한 유효성은 더욱 낮아질 수 있다는 점이다.

이러한 몇 가지 연구분석의 한계점에도 불구하고, 현재 연구 현장에서는 초고속 인터넷 가입자의 포화점 추정과 이에 따른 통신서비스장비 투자정도 문제가 중요한 화두가 되고 있는 만큼, 이 연구가 앞으로 다른 관련 연구와 통신서비스 산업의 예측모형 적용에 있어 조그마한 기여가 될 것으로 생각한다.

참고 문헌

- [1] 김국희, 김지수, 강희일, 전치형, “확산모형을 이용한 전

- 약품 시장 예측 모형,” 전자통신동향분석, 제15권 제 6호, 2000, pp. 178 – 189.
- [2] N. Meade and T. Islam, “Forecasting with Growth Curves: An Empirical Comparison,” *International Journal of Forecasting*, Vol. 11, 1995, pp. 199 – 215.
- [3] J.S. Armstrong, *Principles of Forecasting*, Kluwer Academic Publishers, 2001.
- [4] F.M. Bass, “A New Product Growth Model for Consumer Durables,” *Management Science*, Vol. 15, 1969, pp. 215 – 227.
- [5] J.C. Tanner, “Long Term Forecasting of Vehicle Ownership and Road Traffic,” *The Journal of the Royal Statistical Society, Series A*, 1978, pp. 14 – 63.
- [6] G. Judge, Hill, Griffiths, Luthephol, and Lee, Introduction of the Theory and Practice of Econometrics, John Wiley and Sons, 1988.
- [7] A.K. Bera and Y. Biliias, “The MM, ME, ML, EL, EF and GMM Approaches to Estimation,” *Journal of Econometrics*, Vol. 107, 2002, pp. 51 – 86.