

성장곡선모형에 의한 인구예측시스템

최종후* · 최봉호** · 양우성*** · 김유진****

이 논문에서는 선형·비선형 성장곡선모형의 종류와 특성을 살펴보고, 이들을 비교·검토하고, 모형선호기준 통계량에 입각하여 추정결과를 비교한다. 또한 최종사용자 환경을 위한 SAS/AF로 구현한 성장곡선모형에 의한 인구예측시스템을 소개한다.

핵심단어 : 성장곡선모형(Growth Curve Model), SAS/AF, 인구예측시스템, Damped-Trend 지수평활법(Exponential Smoothing)

I. 서론

한 기업에서 신상품이 출하되었을 때 판매추세에 대한 예측은 경영자의 입장에서는 가장 중요한 관심거리가 될 수 있다. 그것은 신상품의 출하 이후 가장 많이 팔리는 시기와 그 판매량 등에 대한 예측이 판매전략의 수립에서 결정적인 정보가 될 것이기 때문이다. 시간의 흐름에 따라 얻어지는 누적 수요자료를 설명하기 위하여 흔히 이용되는 성장곡선모형(Growth Curve Model)은 이러한 문제를 해결하는데 유용한 통계적 모형이다.

본 논문은 이렇듯 현실 응용성이 높은 성장곡선 모형을 SAS/AF로 구현함으로써 최종사용자환경(End User Computing)을 제공하고자 하는 데에 그 목적을 두고 있다.

* (339-700) 충남 연기군 조치원읍 고려대학교 정보통계학과 부교수, jchoi@tiger.korea.ac.kr

** (302-701) 대전시 서구 둔산동 920 통계청 국제통계과장, bongho.choi@nso.go.kr

*** Dept. of Environmental & Math. Sciences, Okayama University, Tsushima, Okayama, 700-8530, Japan

**** (339-700) 충남 연기군 조치원읍 고려대학교 대학원 정보통계학과, sasw@naver.com

2장에서는 성장곡선모형을 선형·비선형 성장곡선모형으로 분류하고 이들 모형의 종류와 특성을 살펴보고, 이들을 비교·검토한다. 비교대상이 되는 성장곡선 모형은 선형모형으로 Mansfield-Blackman 모형, Linear Gompertz 모형, Weibull 모형, Bass 모형, NSRL(Nonsymmetric Responding Logistic) 모형, Harvey 모형 등이며, 비선형 모형으로는 Logistic 모형, Probit 모형, Gompertz 모형 등이다. 또한 이들 모형을 이용한 추정결과의 선호를 비교하기 위한 모형선호기준 통계량을 소개한다. 3장에서는 SAS/AF로 구현한 성장곡선모형에 의한 인구예측시스템을 소개한다. 마지막으로 4장에서는 약간의 토의를 덧붙인다.

본 논문에서 예시를 위하여 사용된 자료는 1960년부터 1994년까지의 국내 연도별 총인구수이다.

II. 성장곡선모형

시간의 흐름에 따라 얻어지는 누적 자료를 설명하기 위하여 흔히 이용되는 성장곡선모형(Growth Curve Model)은 선형 모형(Linear Model)과 비선형 모형(Nonlinear Model)으로 분류할 수 있다.

1) 선형 성장곡선모형

선형 모형은 일반적으로 극한값(Upper limiting value) K 를 상정하고 분석해야 하기 때문에 현실 응용면에 그만큼 한계점을 안고 있다. 일반적으로 변화율을 분석 자료로 사용하는 Bass 모형, NSRL 모형, Harvey 모형 등은 비선형 성장곡선 모형보다 정확한 예측을 할 수 있다고 알려져 있다(Young, 1993).

(1) Mansfield-Blackman 모형

$$\ln\left(\frac{Y_t}{K - Y_t}\right) = \beta_0 + \beta_1 t + \varepsilon_t \quad (1.1)$$

위에서, Y_t 는 시간에 따른 누적 관측값,
 β_0, β_1 는 모수,
 K 는 극한값,
 $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$

Mansfield-Blackman 모형은 Logistic 모형의 선형화된 형태로서 Logistic 모형과 마찬가지로 변곡점을 중심으로 좌우대칭의 형태를 띠고 있다. 분석시 극한값 K 가 반드시 알려져 있어야 하며, 선형회귀분석 방법을 통해 모형을 추정한다.

(2) Linear Gompertz 모형

$$\ln\left(-\ln\left[\frac{Y_t}{K}\right]\right) = \beta_0 + \beta_1 t + \varepsilon_t \quad (1.2)$$

Linear Gompertz 모형에서 Y_t 의 분산은 상수가 아닌 함수의 형태이며, 치우친 종모양의 함수이다. 분석시 극한값이 알려져 있어야 하며, 모형의 추정에 선형회귀분석 방법이 사용된다.

(3) Weibull 모형

$$\ln\left(\ln\left[\frac{K}{K - Y_t}\right]\right) = \beta_0 + \beta_1 \ln t + \varepsilon_t \quad (1.3)$$

Weibull 모형은 모수가 가지는 값에 따라 모형이 대칭인지 아닌지가 결정되며, 역시 극한값이 알려져 있어야 분석이 가능하다. 모형의 추정에 선형회귀분석 방법이 사용된다.

(4) Bass 모형

$$y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 (Y_{t-1})^2 + \epsilon_t \quad (1.4)$$

위에서, y_t 는 관측값의 변화율
 $\beta_0, \beta_1, \beta_2$ 는 모수,

Bass 모형은 관측값의 변화율을 사용하는 모형으로 Logistic 모형에서 파생되었다. 지금까지 설명된 모형과는 다르게 모형이 시간에 대한 함수가 아니다. 오차항 ϵ_t 는 상수의 값을 가지고 있으며, 극한값이 알려져 있지 않아도 분석이 가능하다.

(5) NSRL 모형 (Nonsymmetric Responding Logistic Model)

$$\ln y_t = \beta_0 + \beta_1 \ln(Y_{t-1}) + \beta_2 \ln(K - Y_{t-1}) + \epsilon_t \quad (1.5)$$

관측값의 변화율을 사용하는 모형이며, 극한값이 알려져 있어야 분석이 가능하다.

(6) Harvey 모형

$$\ln y_t = \beta_0 + \beta_1 t + \beta_2 \ln(Y_{t-1}) + \epsilon_t \quad (1.6)$$

Logistic, Gompertz, 특별한 경우의 변형된 지수 모형을 포함한다. 변화율을 사용하는 모형 중에는 유일하게 모형에 시간이 변수로 포함되어 있다.

2) 비선형 성장곡선모형

비선형 성장곡선모형으로는 Logistic 모형, Probit 모형, Gompertz 모형 등이 널리 사용된다(Young & Ord, 1990; 최중후 · 황정연, 1997).

Logistic 모형은 박테리아의 증식현상과 같은 개체의 수, 유행 상품의 매출누계 등에 잘 적합되는 모형으로 S자 형의 형태를 갖는다. Gompertz 모형은 Logistic 모형

Logistic 모형과 유사한 S자형 곡선으로 누적 정규분포함수의 형태를 취한다. 이외에도 Von Bertalanffy 모형 등이 있다(Draper & Smith, 1981).

모형이 비선형이기 때문에 얻어진 각종 통계량들은 정확한 분포로부터 생성된 값이 아닌 근사적인 분포로부터 얻어진 값이다. 따라서 비선형 성장곡선 모형에서의 추정 결과에 대한 통계량들은 자료의 수가 적을 때 해석상의 주의가 필요하다. 또한 모형의 추정과정에서 각 모수에 대한 초기값이 필요하지만 이에 대한 일반화된 방법은 없다.

(1) Logistic 모형

$$Y_t = \frac{K}{1 + \exp(\alpha - \beta t)} + \varepsilon_t \quad (2.1)$$

위에서, Y_t 는 시간에 따른 누적 관측값,
 α, β 는 모수
 K 는 극한값
 $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$

곡선의 모양은 변곡점을 중심으로 좌우대칭의 형태를 띠고 있으며 변곡점의 시점은 $t_1 = \ln \beta / k$, 변곡점은 $Y_1 = 1 / 2 K$ 이다. 모형의 추정에 비선형회귀분석 방법이 사용되며, 극한값 K 가 알려져 있지 않아도 분석이 가능하다. 이 모형은 피얼 성장곡선 모형(Pearl Growth Curve Model)이라고도 한다.

(2) Probit 모형

$$Y_t = K\Phi(\alpha + \beta t) + \varepsilon_t \quad (2.2)$$

위에서, Φ 는 누적 정규분포함수

Logistic 모형과 유사한 S자형 곡선으로 누적 정규분포함수의 형태를 취한다.

(3) Gompertz 모형

$$Y_t = K \exp(-\alpha \exp(-\beta t)) + \epsilon_t \quad (2.3)$$

Logistic 모형과 같은 S자 형태를 띠고 있지만 변곡점을 중심으로 좌우대칭의 모양은 아니다. 변곡점의 시점은 Logistic 모형과 같은 $t_i = \ln \beta / k$ 이며, 변곡점은 $Y_i = K/e$ 이다. 극한값이 알려져 있지 않아도 분석이 가능하며, 모형의 추정에 비선형회귀 분석 방법이 사용된다.

3) 선형·비선형 성장곡선 모형의 비교

성장의 유형은 분석대상이 되는 자료의 성격과 그 자료의 성장의 유형에 따라 구별될 필요가 있다. 성장곡선의 종류를 나누는 기준에는 관찰값을 그대로 사용하였는

〈표 1〉 선형·비선형 성장곡선 모형의 비교¹⁾

	Logistic	Probit	Gompertz	Mansfield Blackman	Linear Gompertz	Weibull	Bass	NSRL	Harvey
관측치 or 변화율	관측치	관측치	관측치	관측치	관측치	관측치	변화율	변화율	변화율
극한값이 알려져 있어야 하는가?	×	×	×	○	○	○	×	○	×
추정에 선형회귀 모형이 사용 되었는가?	×	×	×	○	○	○	○	○	○
모형이 대칭인가?	○		×	○	○		○		
추정되는 모수의 갯수	2 or 3	2 or 3	2 or 3	2	2	2	3	3	3

1) 〈표 1〉에서 음영표시된 부분은 저자가 추가한 부분임

될 필요가 있다. 성장곡선의 종류를 나누는 기준에는 관찰값을 그대로 사용하였는가? 혹은 관찰값의 변화율을 사용하였는가? 극한값 K가 알려져 있는가? 모형의 분포가 대칭인가? 추정에 선형회귀 모형이 사용되었는가? 모형이 시간에 관한 함수인가? 등 여러 가지가 있을 수 있다. 각 성장곡선 모형들의 특성을 요약하면 <표 1>과 같다.

극한값이 알려져 있지 않은 경우 Bass 모형과 Harvey 모형이 좋은 결과를 나타내며, 특히 Harvey 모형은 자료의 개수가 많은 데이터에 적합하다. 만약 극한값이 알려져 있지만 자료의 크기가 작은 경우에는 NSRL모형을 사용하는 것을 피해야 한다 (Makridakis 외 1984).

이때 자료에 대하여 적합이 가장 잘된 모형일지라도 가장 정확한 예측을 하는 모형은 아니기 때문에 모형의 선택은 자료의 특성에 따라 선택하는 것이 가장 바람직하다.

4) 모형선호기준 통계량

성장곡선 모형을 이용한 추정결과의 선호를 비교하기 위한 모형선호기준 통계량을 소개한다.

(1) 평균오차(Mean Error)

$$ME = \frac{1}{N} \sum_{t=1}^N (\hat{y}_t - y_t) \quad (4.1)$$

\hat{y}_t = 실제값(actual value),

y_t = 예측값(predicted value)

(2) 평균백분율오차(Mean Percentage Error)

$$MPE = \frac{100}{N} \sum_{t=1}^N (\hat{y}_t - y_t) / y_t \quad (4.2)$$

(3) 평균절대오차(Mean Absolute Error)

$$MAE = \frac{1}{N} \sum_{t=1}^N |\hat{y}_t - y_t| \quad (4.3)$$

(4) 평균절대백분율오차(Mean Absolute Percentage Error)

$$MAPE = \frac{100}{N} \sum_{t=1}^N |(\hat{y}_t - y_t) / y_t| \quad (4.4)$$

(5) 평균제곱오차(Mean Squared Error)

$$MSE = \frac{1}{N} \sum_{t=1}^N (\hat{y}_t - y_t)^2 \quad (4.5)$$

(6) 평방근평균제곱오차(Root Mean Square Error)

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{y}_t - y_t)^2} \quad (4.6)$$

(7) 평방근평균제곱백분율오차(Root Mean Square Percentage Error)

$$RMSPE = 100 \sqrt{\frac{1}{N} \sum_{t=1}^N \{(\hat{y}_t - y_t) / y_t\}^2} \quad (4.7)$$

이상의 통계량들은 그 값이 작을수록 해당 모형을 선호한다. MSE는 \hat{y}_t 가 편의된 추정량이므로 $E\{\hat{y}_t\}$ 를 얼마나 효과적으로 추정하고 있는지를 알아보기 위한 적절한 기준이며, MAPE는 예측값 y_t 와 $E\{y_t\}$ 의 차이를 $E\{y_t\}$ 와의 비율로 환산한 값의 총합으로, y_t 의 단위에 무관하게 서로 다른 모형을 비교하는데 바람직한 기준이 된다.

(8) 결정계수(Corrected Actual R-square)

$$R_c^2 = 1 - \frac{SSE}{CSSA} \quad (4.8)$$

$$\text{위에서, } SSE = \sum_{t=1}^N (\hat{y}_t - y_t)^2$$

$$CSSA = \sum_{t=1}^N (y_t - \bar{y})^2$$

$$\bar{y} = \frac{1}{N} \sum_{t=1}^N y_t$$

결정계수는 그 값이 클수록 해당 모형을 선호한다. 하지만 y_t 들이 서로 독립적이라고 볼 수 없으므로 여기서는 큰 의미를 부여할 수 있는 통계량은 아니다.

(9) 예측의 정확성 통계량(Statistic Measuring the Accuracy of Forecast)

$$U_1 = \frac{\sqrt{MSE}}{\sqrt{\frac{1}{N} \sum_{t=1}^N y_t^2}} \tag{4.9}$$

(10) Theil의 부등식 계수(Inequality Coefficient)

$$U = \frac{\sqrt{MSE}}{\sqrt{\frac{1}{N} \sum_{t=1}^N y_t^2 + \frac{1}{N} \sum_{t=1}^N \hat{y}_t^2}} \tag{4.10}$$

U_1 과 U 의 계수 값은 0에 근접할수록 좋은 모형으로 판정한다. 흔히 Theil의 부등식 계수 U 가 0.55 이하이면 예측이 정확한 것으로 판정한다(Lindberg, 1982; McNess, 1979).

위의 (4.9)과 (4.10)은 모형선택기준 통계량이라기 보다는 최적모형으로 선택된 모형하에서 미래의 예측값을 구했을 때 그 결과의 정확성을 알아보는 기준 통계량이라고 할 수 있다.

Ⅲ. SAS/AF로 구현한 성장곡선모형에 의한 인구예측시스템

앞에서 제시된 성장곡선모형들을 이용하여 인구예측을 일괄적으로 수행하는 인구예측시스템을 소개한다. 이 프로그램은 SAS 시스템의 SAS/AF, SAS/ETS의 프로시저와 함수를 이용하여 구현하였다.

인구예측시스템은 크게 5개의 항목으로 구성되어 있는데 각각 초기화면, 자료의 입력, 모형의 선택 및 인구예측, 추정된 모형의 비교, 도움말 윈도우로 구성되어 있다. 각 과정별로 만들어진 윈도우에서 적절한 옵션을 선택한 후 수행시키면 분석이 일괄적으로 수행되어 결과를 요약 출력한다.

분석에 사용된 자료는 1960년부터 1994년까지의 한국의 연도별 총인구수이다. 자료는 <표 2>와 같다.

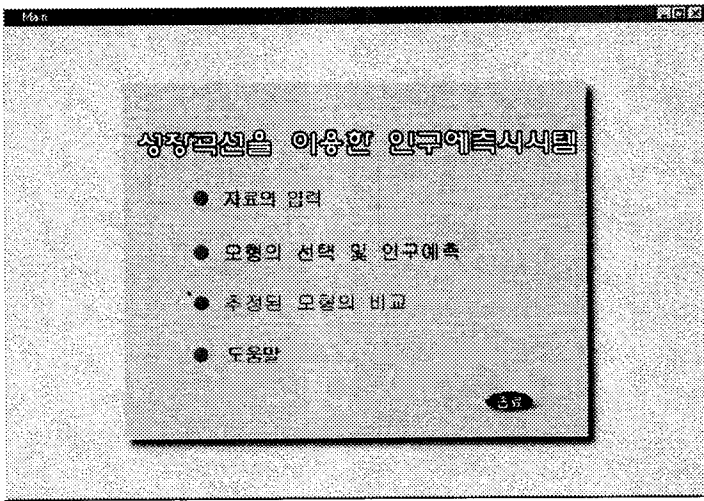
<표 2> 연도별 총인구수 (1960~1994)

연도	인구	연도	인구	연도	인구
1960	25,012,374	1972	33,505,406	1984	40,405,956
1961	25,765,673	1973	34,103,149	1985	40,805,744
1962	26,513,030	1974	34,692,266	1986	41,213,674
1963	27,261,747	1975	35,280,725	1987	41,621,690
1964	27,984,155	1976	35,848,523	1988	42,031,247
1965	28,704,674	1977	36,411,795	1989	42,449,038
1966	29,435,571	1978	36,969,185	1990	42,869,238
1967	30,130,983	1979	37,534,236	1991	43,295,704
1968	30,838,302	1980	38,123,775	1992	43,747,962
1969	31,544,266	1981	38,723,248	1993	44,194,628
1970	32,240,827	1982	39,326,352	1994	44,641,540
1971	32,882,704	1983	39,910,403		

자료출처 : 「장래인구추계」, 1996, 통계청

1) 초기화면

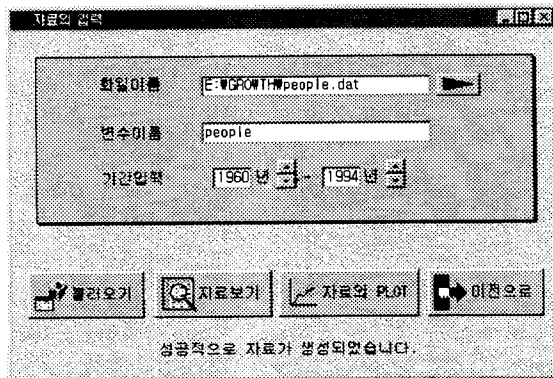
〈그림 1〉은 초기화면으로 위 그림에서 기술되어 있는 순서대로 자료의 분석이 실시된다. 종료 버튼을 누르면 시스템이 종료된다.



〈그림 1〉 초기화면

2) 자료의 입력

〈그림 2〉는 자료의 입력 윈도우로 이 화면에서 선택한 자료를 이용해 이후에 모든



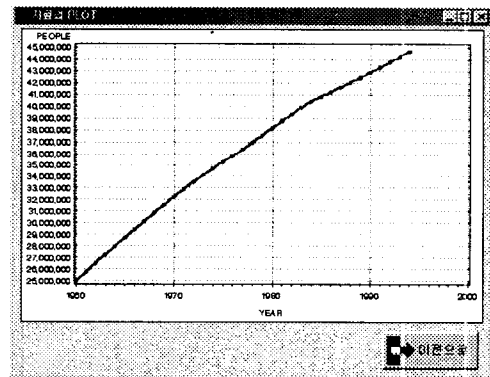
〈그림 2〉 자료의 입력

분석작업이 수행되므로 반드시 데이터를 입력해야 한다. 입력대상이 되는 자료는 ASCII 형식의 데이터이며, 연도별 자료가 분석의 대상이 된다. 불러오기 버튼을 누르면 자료가 SAS 데이터 셋으로 변환되며, 자료가 성공적으로 생성되면 화면 밑 부분에 '성공적으로 자료가 생성되었습니다' 라는 메시지가 출력된다. 만약 자료의 생성에 실패하면 '자료의 생성에 실패하였습니다.' 라는 메시지가 출력된다. 자료보기 버튼을 누르면 입력된 자료를 볼 수가 있으며(〈그림 3〉), 프린터로 출력도 가능하다. 또한 자료의 PLOT 버튼을 누르면 입력된 자료의 연도별 PLOT을 볼 수 있다(〈그림 4〉). 데이터의 입력이 끝나면 이전으로 버튼을 선택해 초기화면으로 돌아갈 수 있다.

3) 모수의 초기값 추정

YEAR	PEOPLE
1960	25,012,374
1961	25,765,673
1962	26,513,030
1963	27,261,747
1964	27,984,155
1965	28,704,674
1966	29,435,571
1967	30,130,983
1968	30,830,302
1969	31,544,266
1970	32,240,827
1971	32,882,704
1972	33,505,406
1973	34,103,149
1974	34,682,266
1975	35,280,725
1976	35,848,529

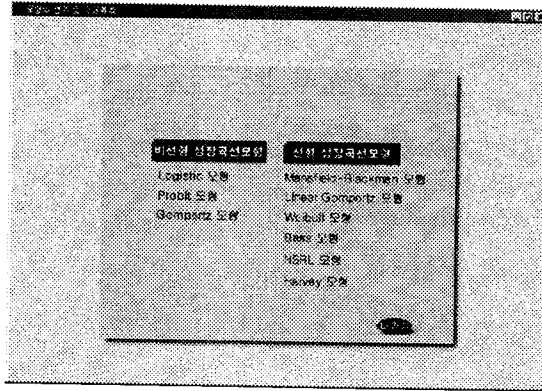
〈그림 3〉 자료 보기



〈그림 4〉 자료의 Plot

각 모형은 적절한 모수의 초기값이 필요하며, 특히 극한값 K 는 초기값에 민감하게 반응하는 경향이 있다. 모수의 초기값을 구하는 일반적인 방법은 없지만 본 시스템에서는 모수의 초기값으로 극한값 K 는 자료의 길이(자료가 수집된 연도의 수)와 마지막 시점의 자료를 이용하여, 자료가 10년 이하의 자료이면 마지막 시점의 자료의 5배, 10~20년의 자료는 마지막 시점 자료의 3배, 30년 이상의 자료는 마지막 시점 자료의 2배를 곱하여 극한값을 추정하는데 있어서 초기값으로 사용하였다. 극한값 이외의 모수의 초기값은 2, 1, 0, -1, -2를 각각 모수의 초기치로 사용하는 격자탐색(Grid search)기법을 사용하였다.

4) 모형의 선택 및 인구예측

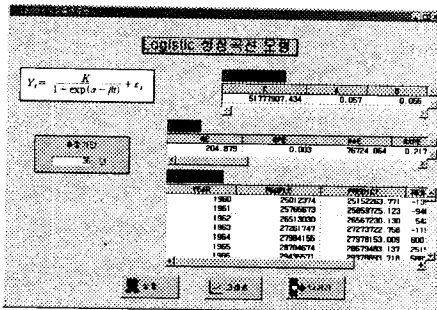


〈그림 5〉 모형의 선택 및 인구예측

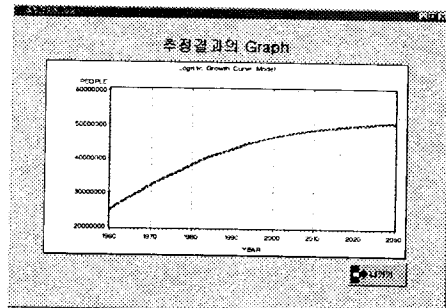
이 윈도우에서는 시스템에서 분석 가능한 모형의 종류를 나열하고 있으며(〈그림 5〉). 모형을 선택하면 각 모형에 대한 분석 윈도우로 이동하게 된다.

모형의 선택화면에서 비선형 성장곡선 모형인 Logistic 모형을 선택하고 추정기간에 분석자가 추정하고 싶은 기간을 입력하면 〈그림 6〉과 같이 모수의 추정결과, 모형 적합성을 판단하는 모형 선호기준 통계량과, 추정된 모형에 입력자료를 적합시킨 결과인 자료에 대한 예측치(Predict Value), 그리고 실제 자료와 예측치의 차이인 잔차(Residual)가 출력된다.

〈그림 6〉의 화면에서 그래프 버튼을 선택하면 입력자료와 예측치를 비교하는 〈그림 7〉의 PLOT이 화면에 출력된다.

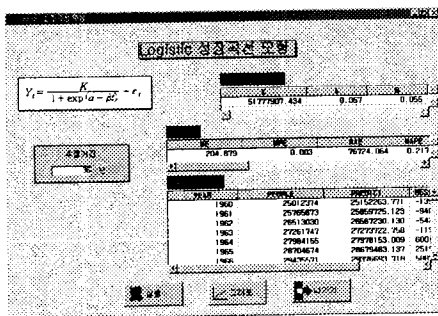


〈그림 6〉 Logistic 성장곡선모형

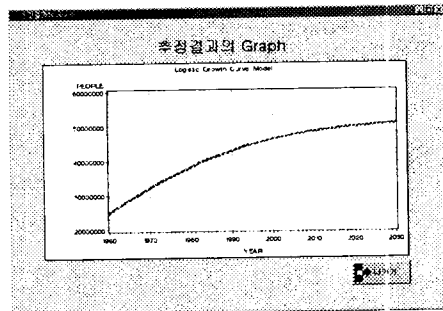


〈그림 7〉 Logistic 모형의 추정결과 Plot

시스템에 적용된 모형 중 극한값 K가 반드시 알려져 있어야 하는 Mansfield-Blackman 모형, Linear Gompertz 모형, Weibull 모형, NSRL 모형은 추정기간을 입력하는 외에 극한값 K를 추가로 입력해야 한다. 하지만 극한값을 안다는 것은 현실적으로 거의 불가능하다. 그러므로 성장곡선모형 중 가장 적합이 잘된 모형에서 추정된 극한값 K를 이용하는 것이 하나의 대안이라고 할 수 있겠다.



<그림 8> Bass 성장곡선모형



<그림 9> Bass 모형의 추정결과 Plot

<그림 9>에서 볼 수 있듯이 선형 성장곡선모형과 비선형 성장곡선모형의 적합도가 크게 차이가 없다는 것을 확인할 수 있다. 또한 본 논문에서 이용된 자료에서는 수요의 변화율을 사용하는 모형인 BASS 모형, NSRL 모형, Harvey 모형은 다른 성장곡선모형에 비해 적합이 잘 되는 것을 확인할 수 있다.

선형성장곡선 모형에서는 신뢰구간을 구하기 위한 수식이 복잡하고 또한 구하기도 어렵다. 본 시스템에서는 이를 해결하기 위해 모의실험을 통한 수치해석적인 방법을 통해 해결하였다.

5) 추정된 모형의 비교

<그림 10>은 시스템에 적용된 모형들의 선택기준이 되는 여러 통계량을 테이블의 형태로 출력해주고 있으며 이를 종합적으로 비교하여 가장 적합한 모형을 선택할 수 있다.

MODEL	ME	MPE	MAE	MAPE
Logistic	204.079	0.003	76724.064	0.217
Probit	188.409	0.003	77260.981	0.218
Gompertz	21.417	0.001	55200.046	0.147
Mansfield	-6671.490	0.016	185548.901	0.555
Linear G	70.077	0.000	55115.619	0.146
Weibull	62224.405	0.361	1486884.740	4.348
Bass	0.000	0.000	29487.992	0.078
NSRL	-1479.720	-0.003	29375.496	0.078
Harvey	-1493.991	-0.003	29461.413	0.078

〈그림 10〉 추정된 모형의 비교

〈표 3〉은 모형선택기준 통계량을 요약한 표이다.

〈표 3〉 모형선택기준 통계량

	RMSE	RMSPE	결정계수	UI	U
Logistic	88713.344	0.253	0.9998	0.0024	0.0012
Probit	89582.973	0.254	0.9998	0.0025	0.0012
Gompertz	69785.887	0.181	0.9999	0.0019	0.0010
Mansfield- Blackman	213684.533	0.698	0.9987	0.0059	0.0029
Linear Gompertz	69791.184	0.181	0.9999	0.0019	0.0010
Weibull	1743855.155	5.471	0.9034	0.0476	0.0238
Bass	37957.115	0.097	1.0000	0.0010	0.0005
NSRL	38243.867	0.098	1.0000	0.0010	0.0005
Harvey	38601.206	0.099	1.0000	0.0011	0.0005

본 논문에 이용된 자료의 적합결과는 비선형 성장곡선모형으로는 Gompertz 모형이, 선형 성장곡선모형으로는 Bass 모형이 가장 자료에 대해 적합이 잘된 것으로 분석되었다. 대체적으로 Mansfield-Blackman모형과 Weibull 모형을 제외한 선형 성장곡선의 적합이 비선형 성장곡선의 경우보다 잘 되었음을 확인할 수 있다.

위에서 적용된 모형들 대부분 2010년을 전후해서 인구의 증가추세가 매우 둔화될

것이며 2030년을 전후로 해서 인구의 증가추세는 거의 멈출 것으로 예상된다. 또한 2011~2013년 사이에 우리나라의 총인구가 5천만을 돌파할 것으로 예측결과 나타났다.

〈표 4〉는 수요예측시스템을 사용한 후 Gompertz 모형과 Bass모형에 의해 예측된 향후 10년간의 누적인구수이다.

6) 도움말

〈표 4〉 Gompertz 모형과 Bass 모형의 예측값

연도	누적 인구	Gompertz 모형을 이용한 예측값	Bass 모형을 이용한 예측값
1994	44,641,540	44532575.944	44592529.684
1995	.	44898538.927	44980845.841
1996	.	45253802.916	45359735.330
1997	.	45598578.211	45729346.671
1998	.	45933078.722	46089847.756
1999	.	46257521.214	46441395.574
2000	.	46572124.602	46784150.942
2001	.	46877109.304	47118275.259
2002	.	47172696.631	47443930.265
2003	.	47459018.241	47761277.813
2004	.	47736565.627	48070479.660
2005	.	48005289.650	48371697.262

인구예측시스템에서는 각 성장곡선모형 설명 도움말, 시스템 설명 도움말, 화면 출력물 도움말의 3가지 도움말을 제공하고 있다. 이 도움말들은 각각 사용자가 시스템을 사용하면서 예측방법이나 예측과정에서 발생하는 여러 가지 결과물에 대한 이해를 돕는 기능을 한다.

IV. 결론 및 토의

지금까지 성장곡선모형을 선형·비선형 성장곡선모형으로 분류하고 이들 모형의 종류와 특성을 비교하고, SAS/AF로 구현한 성장곡선모형에 의한 인구예측시스템의 개발 사례를 소개하였다. 이 시스템에 의해 분석자는 비교적 적은 시간과 노력을 들여서 여러 성장곡선모형에서 추정된 결과를 비교하고 선택할 수가 있다.

그러나 앞서도 언급하였듯이 인구예측시스템에 의하여 적합이 가장 잘된 모형일 지라도 가장 정확한 예측을 하는 모형은 아니기 때문에 모형의 선택에는 다각도의 검토절차가 필요하다. 또한 이 문제에 대한 시계열 모형으로서는 Damped-Trend 지수 평활법(Exponential Smoothing)²⁾을 이용할 수 있다.

이와 같은 성장곡선 모형은 매년 증가되는 컴퓨터 누적 보급대수, 호출기 가입자수, PC 통신망 가입자수 등 여러 분야의 수요예측에 이용될 수 있다.

2) Damped Trend Exponential Smoothing 모형은 ARIMA 모형으로 표현하면 ARIMA(1,1,2) 모형과 동일하다. 즉,

$$(1 - \phi B)(1 - B)Y_t = (1 - \theta_1 B - \theta_2 B^2)\epsilon_t$$

$$\theta_1 = 1 + \phi - \alpha - \alpha\gamma\phi, \quad \theta_2 = (\alpha - 1)\phi$$

으로 표현할 수 있다(SAS Institute, 1995).

참고문헌

- 통계청 (1996), 《장래인구추계》, 통계청
- 최종후, 황정연 (1997), "모형선호 기준에 의한 비선형 성장곡선의 비교," 《한국분류학회지》, 제1권, pp. 25-33.
- Draper N. R. and Smith H. (1981), *Applied Regression Analysis, Second Edition*, John Wiley & Sons Inc..
- Lindberg B. C. (1982), "International Comparison of Growth in Demand for a New Durable Product," *Journal of Marketing Research* 19, 364-371.
- Makidakis, S., Anderson, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J. Parzen, E., and Winkler, R. (1984), *The Forecasting Accuracy of Major Time Series Method*, Willy, Chichester.
- McNess S. K. (1979), "The Forecasting Record for the 1970' s," *New England Economic Review*, September-October.
- SAS Institute (1993), *SAS/AF Software : FRAME Entry Usage and Reference, Version 6, First Edition*, SAS Institute Inc..
- SAS Institute (1995), *SAS/AF Software : FRAME Dictionary, Version 6, First Edition*, SAS Institute Inc..
- SAS Institute (1993), *SAS/ETS User's Guide, Version 6, Second Edition*, SAS Institute Inc..
- SAS Institute (1995), *SAS/ETS Software: Time Series Forecasting System, Version 6, First Edition*, SAS Institute Inc..
- SAS Institute (1994), *SAS Screen Control Language : Reference, Version 6, Second Edition*, SAS Institute Inc..
- Young P. (1993), "Technological Growth Curve : A Competition of Forecasting Models," *Technological Forecasting and Social Chance*, Vol 44, 375-389.

abstract

Population Forecasting System Based on Growth Curve Models

Jong Hoo Choi, Bong Ho Choi, Woo Sung Yang, Eugene Kim

In this paper, we investigate the linear and nonlinear growth curve models, and compare these models based on model selection criteria. And we propose a Population Forecasting System using SAS/AF. It is useful in the behalf of an end-user computing environment.

Key Words : growth curve model, SAS/AF, population forecasting system,
damped-trend exponential smoothing