Journal of the Korean Data & Information Science Society
2013, **24**(2), 321–332

http://dx.doi.org/10.7465/jkdi.2013.24.2.321
한국데이터정보과학회지

초등학교 교원 수 예측을 위한 시계열 회귀모형[†]

류수락[1] · 김종태[2]

[12]대구대학교 전산통계학과

접수 2013년 2월 1일, 수정 2013년 3월 6일, 게재확정 2013년 3월 14일

요 약

본 연구는 지속적인 저출산의 여파로 2020년에는 초등학생 수가 올해 대비 17%, 중고교생은 30%가 감소할 것이라는 예측을 가지고 초등학교 교원 수를 예측하기 위한 방법을 제시하는데 있다. 교육통계연보의 1970년부터 2010년까지의 초등교육 관련 주요 통계 자료를 이용하여 시계열 회귀모형과 시계열 그룹별 회귀모형, 지수평활법 모형을 제시하고, 제시된 모형을 이용하여 향후 10년간의 연도별 초등학교 교원 수를 예측하였다. 모형 예측 결과 시계열 그룹별 회귀 모형이 교원 수 시계열을 가장 잘 설명하는 것으로 나타났으며, 적합한 모형으로 판명되었다. 3가지 분석방법 모형에 따른 예측값에 대한 장단점과 한계를 제시한다.

주요용어: 시계열 그룹별 회귀모형, 시계열 회귀모형, 지수평활법 모형.

1. 서론

2011년 6월 15일 천지일보에서는 “지속적인 저출산의 여파로 2020년에는 초등학생 수가 올해 대비 17%, 중고교생은 30% 감소할 것이라는 예측이 나왔다. 이영 한양대 교수 등 연구진은 지난 6월 초 ‘2012~2020년 중장기 교원 수급 전망 세미나’에서 이와 같이 밝혔다. 연구는 교육과학기술부의 의뢰로 이뤄졌다. 그러나 학생이 줄어도 2020년 교사 1인당 학생 수는 현재 교사 수급상황을 고려할 때 OECD 국가 평균에 크게 못 미칠 것으로 예상돼 교사 수를 늘릴 필요가 있는 것으로 나타났다. 2008년 OECD 국가 평균은 초등 16.4명, 중학교 13.7명, 고등학교 13.5명이다. 2011년 기준 우리나라 국공립 초등학교 교사는 정원이 13만 9448명이므로 2020년 무렵 학생 수가 감소해도 여전히 1인당 학생 수는 OECD 국가보다 약 2명이 많다. 중고교 교사도 현재 정원인 13만 5764명을 그대로 유지할 경우 2020년 1인당 학생 수는 14.25명으로 여전히 OECD국가보다 약 1~2명이 많다. 연구진은 2020년에 1인당 학생 수가 OECD 수준에 미치기 위해서는 초등학교 교사를 매년 4~5천 명, 중고교 교사는 매년 2~5천 명을 새로 선발해야 한다고 분석했다.”라는 기사를 보도했다.

학습활동을 효과적으로 전개하기 위해서는 충분한 교육 시설 뿐만 아니라 학생과 교직원 같은 인적 자원을 효과적으로 활용해야 한다. 그 동안 계속된 교원 증원과 학교, 학급의 새로운 증설로 각급 학교의 교원 1인당 학생 수와 학급당 학생 수가 지속적으로 감소되어 왔으나 아직도 선진국의 수준과는 큰 차이가 있다. 2008년 우리나라 초등학교의 교원 1인당 학생 수는 초등학교 21.3명이다. 이는 선진국인 미국 14.3명, 일본 18.8명, 프랑스 19.9명, 호주 15.8명에 비하여 초등학교의 교원 1인당 학생 수가 과다함을 보여준다.

[†] 이 논문은 2010년도 대구대학교 학술지원비에 의한 연구임.
[1] (712-714) 경상북도 경산시 진량읍 대구대로 201, 대구대학교 전산통계학과, 석사과정.
[2] 교신저자: (712-714) 경상북도 경산시 진량읍 대구대로 201, 대구대학교 전산통계학과, 교수.
E-mail: jtkim@daegu.ac.kr

21세기 학교의 가장 큰 어려움은 학생생활지도 문제가 될 것이라고 교사의 80%가 꼽고 있다. 이들 중 대부분은 학급당 학생 수가 많은 것에 대해 학생생활지도를 어렵게 하는 가장 큰 원인으로 지적하고 있다. 실제로 학급당 학생 수와 교원 1인당 학생 수의 과다는 학교교육과정 운영의 곤란, 학습자의 수준과 능력에 맞는 수업 진행의 어려움에 따른 학습효과의 저하, 학생 생활 지도 곤란, 학생의 심신 발달 저해 등의 문제를 야기하고 있다. 또한 저출산의 여파로 인해 초등학교 학생 수가 줄어들고 있지만 여전히 초등학교 교원 일인당 학생 수는 OECD 국가들의 평균 초등학교 교원 일인당 학생 수에 도달하지 못하고 있는 현실이다.

최근 시계열분석을 통한 교원 수 수요를 예측하는 연구가 시도가 되고 있는 상황이다. Kim (2000)은 대학 졸업생의 수를 예측하였고, Kim (2002)에서 시계열 분석을 이용하여 초등학교 선생님의 수를 예측하였다. Kim (2005a, 2005b)은 대구와 경북지역을 중심으로 고3학생수를 예측하고, 대구와 경북지역의 대학 신입생 수와 비교 분석하였다. Yoon과 Kim (2012)은 비선형 회귀모형을 이용한 학년별 학생수 추계를 하였다.

이에 본 연구에서는 시계열분석을 통해 향후 10년간의 초등학교 교원 수와 교원 일인당 학생 수를 예측해보고 어느 시점부터 OECD 국가의 평균 초등학교 교원 일인당 학생 수에 도달하는지에 대하여 분석하였다. 본 연구는 초등학교 교원양성 정책을 수립하기 위한 기본적인 자료로 활용될 수 있을 것이다.

2. 시계열 회귀분석 모형에 의한 교원 수 예측값

본 연구에서 사용된 연구대상은 1970년부터 2010년도까지 (41년간)의 초등학교 주요 교육통계 자료이며, 통계자료는 교육통계 연보 (교육통계서비스)에서 제공하는 학교 수, 학급 수, 학생 수, 교원 수, 교육예산과 2010년의 물가지수를 100으로 하여 조정한 2010년 불변가격 교육예산규모를 활용하였다. 교원 수를 종속시계열로 하고 학교 수, 학급 수, 학생 수, 불변가격 교육예산 규모 등 4개의 시계열을 독립시계열로 하는 시계열 회귀분석과 다중공선성 (multicollinearity)진단을 같이 하였다. 분석을 한 결과는 다음 Table 2.1과 같다.

Table 2.1 Time series regression analysis and multicollinearity

Variable	DF	Estimate	SE	t-value	p-value	Tolerance	VIF
Constant	1	-17514	19191	-.91	.3673	.	0
School	1	.0024	.0035	.68	.5016	.0273	36.5798
Class	1	.0772	.2356	.33	.7446	.0465	21.5040
Student	1	14.9277	3.1004	4.81	<.0001	.1507	6.6334
Constant Dollars	1	2.05E-06	4.05E-07	5.06	<.0001	.0084	118.9107

Constant Dollars of unit one thousand won

학교 수, 학급 수, 학생 수, 불변가격 교육예산 규모를 독립시계열로 하는 모형을 검토한 결과에 의하면 Table 2.1에서 학교 수와 학급 수는 회귀계수가 유의하지 않으므로 종속시계열인 교원 수를 설명하는데 기여하지 않는다는 것이 나타났다. 그리고 다중공선성 문제를 판단하는 분산팽창인수 (VIF)를 사용하였으며, 일반적으로 분산팽창인수 값이 10이상일 때 심각한 다중공선성 문제가 존재한다고 볼 수 있는데 학교 수와 학급 수, 불변가격 교육예산 규모는 다중공선성 문제을 갖는 것으로 나타났다. 다중공선성 문제를 해결하기 위한 대책으로는 여러 가지 방법이 있지만 가장 일반적으로 사용되는 변수 제거 방법을 사용하였다. 학교 수와 학급 수는 다중공선성이 있고, 종속시계열인 교원 수를 설명하는데 유의하지 않는 것으로 분석 결과를 보인다. 그러므로 이 두 변수를 제거하고 학생 수와 불변가격 교육예산 규모 변수만을 사용하여 다중공선성 문제를 해결하는 방식을 선택하였다.

학교 수와 학급 수의 두 변수를 제거하고 난 후에 통계적으로 유의성을 가지는 학생 수와 불변가격 교육예산 규모 변수만을 독립 시계열로 하는 모형이 최종적으로 선택되었다. 모형의 결정계수 R^2 =

0.9801로서 모형에 대한 높은 설명력을 가졌고, 모형에 사용된 독립 시계열인 학생 수와 불변가격 교육예산 규모의 변수들에서 다중공선성 문제는 나타나지 않았다.

학생 수와 불변가격 교육예산 규모을 독립시계열로 사용하는 모형의 오차항의 자기상관 (autocorrelation)정도에 관하여 Durbin-Watson 검정을 실시한 결과, DW=0.362로 모형의 오차항이 양의 자기상관을 갖는 것으로 나타났다. 이를 해결하기 위하여 시차의 후진제거법 (backstep)을 사용한 Yule-Walker 추정방법을 사용하였다 그 결과 Durbin-Watson 검정 의하여 DW=1.6644의 값이 나타났고, 자기상관이 제거가 되어진 것을 알 수 있다. 이 방법에 의하여 자기상관을 제거한 모형은 $R^2 = 0.9941$의 결정계수를 가졌다. 다음 Table 2.2에는 최종적으로 선택된 시계열 회귀분석 모형의 추정결과가 제시되었다.

Table 2.2 Time series regression model analysis

Variable	DF	Estimate	SE	t-value	p-value
Constant	1	100506	15192	6.62	<.0001
Student	1	.0033	3.0102	3.88	<.0001
Constant Dollars	1	1.58E-06	1.70E-07	9.29	<.0001

Constant Dollars of unit one thousand won

시계열 회귀분석 모형의 추정결과를 이용한 모형식을 정리하면 식 (2.1)과 같다.

$$\begin{aligned} \text{Teachers} &= 100506 + 0.00339 \times \text{Student} + 0.0000015826 \times \text{Constant Dollars} + \epsilon, \\ \epsilon &= 0.860865\epsilon_1 - 0.208651\epsilon_2 \end{aligned} \tag{2.1}$$

위의 절차에 의해서 선택된 식 (2.1)에서 향후 10년 간의 초등학교 교원 수를 예측하기 위해서는 향후 10년 간의 초등학교 학생 수와 불변가격 교육예산 규모가 먼저 예측되어야 한다.

2.1. 향후 10년간의 초등학교 학생 수 예측값

Statistics Korea (2010)에서 2010년도의 인구센서스 결과를 기준으로 발표한 1970년부터 2020년까지의 연령별 장래인구 추계값과 연령별 인구수 추계값으로부터 구해진 초등교육대상 연령 인구수와 1970년부터 2010년까지의 실제 초등학교 취학률을 사용하였다.

Statistics Korea (2010)에서 발표한 연령별 장래 인구 추계값은 7월 1일 0시를 기준으로 한 것이기 때문에 연령별 인구수 추계값은 만 5세 아동 중 7월, 8월, 9월, 10월, 11월, 12월, 1월, 2월에 출생한 아동은 다음 해의 초등학교 입학 대상자가 되며, 11세 아동 중 7월, 8월, 9월, 10월, 11월, 12월, 1월, 2월에 출생한 아동은 다음 해의 중학교 입학 대상자가 된다. 따라서 초등교육대상 연령 인구수는 연령별 인구수 추계값으로부터 다음 식 (2.2)에 의하여 산출되었다.

$$\begin{aligned} \text{초등교육대상 연령 인구수}(N_t) &= 5\text{세인구} \times (8/12) + 6\text{세인구} + \cdots + 10\text{세인구} \\ &\quad + 11\text{세인구} \times (4/12) \end{aligned} \tag{2.2}$$

취학률은 증가와 감소가 반복적으로 일어났으며, 1990년부터 1992년에는 취학률이 100%가 넘는데 이 현상들은 취학률을 산출할 때 '각급학교 학생 수'는 연령에 관계없이 실제로 재학중인 학생을 포함시키기 때문에 취학률이 100% 이상일 경우도 발생한다고 분석된다. 최근 10년간의 평균 취학률은 약 98.41%인 것으로 나타나고 있다.

다음 Table 2.3에는 2011년부터 2020년까지 향후 10년간 초등교육대상 연령 인구수와 초등학교 학생 수 예측값이 수록되었다. 초등교육대상 연령 인구수는 식 (2.2)에 의해서 산출된 값이며, 초등학교 학생 수의 예측값은 향후 10년간 초등교육대상 연령 인구의 최근 10년간의 취학률 98.41%로 가정하여 산출한 값이다.

Table 2.3 Estimated numbers of the elementary school students

Year	Age population	Predicted students
2011	2984422.67	2936970.35
2012	2835811.33	2790721.93
2013	2773474.00	2729375.76
2014	2747487.33	2703802.28
2015	2716583.67	2673389.99
2016	2730216.33	2686805.89
2017	2754205.67	2710413.80
2018	2757211.33	2713371.67
2019	2733903.67	2690434.60
2020	2722755.00	2679463.20

2.2. 향후 10년간의 교육예산 규모 예측

향후 10년간의 교육예산 규모를 예측하기 위하여 1970년부터 2010년까지 41년간의 교육예산 규모의 변동이 검토되었다. 교육예산 규모는 2010년 불변가격으로 1970년의 2,240,809,347천원에서 2010년에는 41,292,619,760천원으로 증가하였는데, 이는 총액 기준으로 40년 동안 39,386,709,652천원이 증가한 것이다. 이를 비율로 계산하면 40년 동안 1,858%가 증가 된 것이고, 이로부터 기하평균에 의하여 연평균 증가율을 산출하면 40년 동안 해마다 전년도 대비하여 평균 7.79%씩 증가한 것으로 나타났다.

한편, 가장 최근 11년간의 교육예산 규모의 변동을 살펴보면 2000년의 30,430,548,939천원에서 2010년의 41,292,619,760천원으로 이는 총액 기준으로 10년 동안 11,196,970,060천원이 증가한 것이다. 이를 비율로 계산하면 10년 동안 137%가 증가 된 것이고, 이로부터 기하평균에 의하여 연평균 증가율을 산출하면 10년 동안 해마다 전년도 대비하여 평균 3.19%씩 증가한 것으로 나타났다. 지난 40년간 우리나라의 경제는 경제 개발 5개년 계획으로 인해 고도성장을 거듭하여 왔으며 이에 따라 교육예산도 큰 비율로 증가하게 되었다. 그러나 최근에는 경제 위기를 경험하면서 1999년에는 IMF 경제위기가 발생하여 교육예산의 감소현상이 발생하기도 했다. 따라서 향후 10년간은 최근 11년간의 교육예산 규모 증가율이 3.19% 증가가 있을 것으로 가정하고 교육예산 규모의 예측값을 산출하였다. Table 2.4에는 이러한 가정에 의하여 예측된 향후 10년간의 교육예산 규모가 수록되었다.

Table 2.4 Estimated educational budget scales

Year	Estimated Education budget
2011	42,955,436,856
2012	44,325,715,292
2013	45,739,705,610
2014	47,198,802,219
2015	48,704,444,009
2016	50,258,115,773
2017	51,861,349,666
2018	53,515,726,721
2019	55,222,878,403
2020	56,984,488,224

unit: one thousand won

Table 2.5는 추정된 위의 식 (2.1)에서 Table 2.3에 있는 초등학교 학생 수의 예측값과 Table 2.4의 교육예산 규모 예측값을 독립변수로 사용하여 향후 10년간 초등학교 교원 수를 예측한 결과이다.

Table 2.5 Estimated numbers of the elementary school teachers

Year	Estimated teachers
2011	178,444
2012	180,116
2013	182,146
2014	184,369
2015	186,648
2016	189,153
2017	191,770
2018	194,398
2019	197,022
2020	199,773

3. 시계열 그룹별 회귀분석 모형에 의한 교원 수 예측값

Figure 3.1은 연도와 교원 수의 시도표를 작성한 그림이다 (Korean Educational Development Institute, 1982-2012). 연도별 교원 수의 변화를 살펴보면 1998년을 기준으로 교원 수에 대한 기울기가 1998년 이전의 변화보다도 급격히 변화하는 것을 알 수 있다. 이러한 그래픽 분석에 따라서 1998년을 기준으로 자료를 두 개의 그룹으로 나눈 후 모형을 다음과 같이 정의한다.

$$\begin{aligned} GR1 &: \text{Teacher} = \beta_{01} + \beta_{11}\text{Year} + \epsilon_{1t}, \quad t = 1970, \cdots, 1998 \\ GR2 &: \text{Teacher} = \beta_{02} + \beta_{12}\text{Year} + \epsilon_{2t}, \quad t = 1999, \cdots, 2010 \end{aligned} \tag{3.1}$$

위 모형을 식 (3.2)과 같이 변형할 수 있다.

$$\begin{aligned} GR1 &: \text{Teacher} = \beta_{01} + \beta_{11}\text{Year} + \epsilon_{1t}, \quad t = 1970, \cdots, 1998 \\ GR2 &: \text{Teacher} = (\beta_{01} + \beta_{01}^*) + (\beta_{11} + \beta_{11}^*)\text{Year} + \epsilon_{2t}, \quad t = 1999, \cdots, 2010 \end{aligned} \tag{3.2}$$

즉, 식 (3.1)에서 $\beta_{02} = \beta_{01} + \beta_{01}^*$이고 $\beta_{12} = \beta_{11} + \beta_{11}^*$로 변형한 것이다. 한편 식 (3.2)를 가변수 D_t를 이용하여 하나의 모형으로 표현하면 식 (3.3)과 같다.

$$\text{Teacher} = \beta_{01} + \beta_{01}^* D_t + \beta_{11}\text{Year} + \beta_{11}^* D_t\text{Year} + \epsilon_t, \quad t = 1970, \cdots, 2010 \tag{3.3}$$

여기에서

$$D_t = \begin{cases} 1, & \text{만약} \quad t \geq 1998 \\ 0, & \text{만약} \quad t < 1998 \end{cases}$$

이다. 식 (3.1), 식 (3.2), 식 (3.3)는 모두 같은 모형임을 쉽게 알 수 있다. 가변수를 이용한 시계열 그룹별 회귀분석을 한 결과는 Table3.1와 같다. 여기서, T는 가변수 D_t를, T∗Year는 $D_t * YEAR$를 의미한다.

Table 3.1 Grouped time series regression analysis

Variable	DF	Estimate	SE	t-value	Approx
Constant	1	-2818006	138796	-20.3	<.0001
T	1	-2869631	581243	-4.94	<.0001
Year	1	1483	69.9345	21.21	<.0001
T∗Year	1	1433	290.665	4.93	<.0001

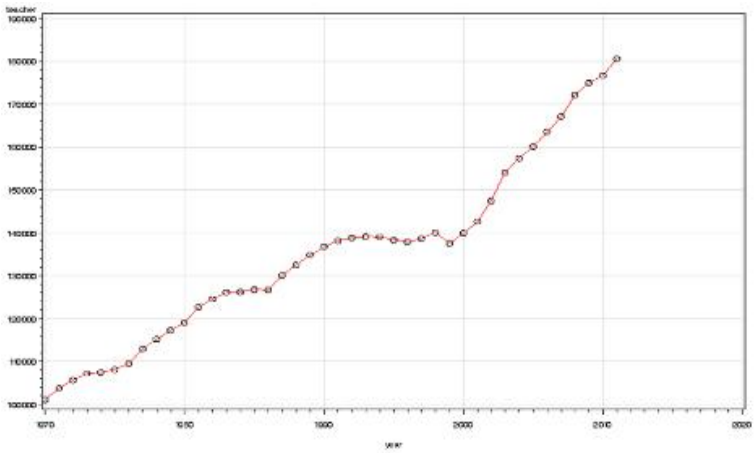

Figure 3.1 Time series chart of the number of teachers

Table 3.1과 같이 독립 시계열 변수들이 모두 유의하게 나타났고, 시차의 후진제거법 (backstep)을 사용한 Yule-Walker 추정값에 의하여 DW=1.3832로 자기상관의 제거가 시도되었으며, 이 방법에 의하여 자기상관을 제거한 모형은 $R^2 = 0.9974$의 결정계수를 가졌다. 따라서 이 연구에서는 이를 교원 수 예측을 위한 최종적인 모형으로 선택하였으며, 시계열 그룹별 회귀분석 모형의 추정결과를 이용한 모형식을 정리하면 식 (3.4)과 같다.

$$\begin{aligned} \text{Teacher} &= -2818006 - 2869631D_t + 1483\text{Year} + 1433D_t \times \text{Year} + \epsilon_t \\ \epsilon &= 0.860865\epsilon_1 - 0.208651\epsilon_2 \end{aligned} \tag{3.4}$$

식 (3.4)을 이용한 향후 10년간 초등학교 교원 수 예측값이 95% 신뢰구간과 함께 Table 3.2에 제시되었다.

Table 3.2 Estimated numbers of the elementary school teachers by grouped time series regression analysis

Year	2011	2012	2013	2014	2015
95% Upper limit	182,288	190,395	189,676	195,723	195,581
Predicted	179,633	182,273	184,875	187,050	189,117
95% Lower limit	176,978	174,152	180,074	178,376	182,654
Year	2016	2017	2018	2019	2020
95% Upper limit	200,485	199,720	206,615	205,197	212,843
Predicted	191,150	193,288	195,649	198,105	200,738
95% Lower limit	181,814	186,856	184,683	191,013	188,632

4. 지수평활법 모형에 의한 교원 수 예측값

지수평활법 모형은 각 방법에 추세와 계절요인의 형태에 따라 다르게 적용되는데 Figure 3.1 연도와 교원 수의 시도표를 살펴보면 시간이 지나면서 시계열의 값이 커지고 있음을 알 수 있어 1차 선형 추세를 갖고 있음을 확인할 수 있으므로 향후 10년간 초등학교 교원 수를 예측하기 위하여 단수지수평활 모형보다는 이중지수평활 모형이 더 적절하다고 판단하였다. 또한 가장 좋은 예측값을 구하는 문제는 전

체 자료 (1970년부터 2010년까지)를 이용한 예측오차의 제곱합을 최소화하는 평활상수 (α)의 결정문제로 귀결된다. 0 <평활상수< 1에 대해 0.1부터 0.9까지 0.1씩 증가시켜 예측오차의 제곱합을 계산하였더니 0.8과 0.9 사이에서 감소하다가 다시 증가하는 패턴을 보이고 있어 0.8과 0.9사이의 평활상수 값을 0.01씩 증가시켜 시뮬레이션을 한 결과가 Table 4.1에 제시되었다.

Table 4.1 The sum of squares of prediction errors for smoothing parameter values

α	SSE	α	SSE
0.1	795,941,650	0.81	97,983,456
0.2	433,080,176	0.82	98,041,231
0.3	249,093,522	0.83	98,156,756
0.4	169,437,376	0.84	98,330,559
0.5	130,983,692	0.85	98,563,322
0.6	111,047,451	0.86	98,855,880
0.7	101,172,652	0.87	99,209,218
0.8	97,983,058	0.88	99,624,472
0.9	100,646,025	0.89	100,102,929

Table 4.1로부터 예측오차의 제곱합이 가장 작은 값을 가지는 평활상수를 0.8로 정하고 이중지수평활법을 수행하여 얻어진 향후 10년간 초등학교 교원 수 예측값이 95% 신뢰구간과 함께 Table 4.2에 제시되었다.

Table 4.2 Estimated numbers of the elementary school teachers by double exponential smoothing method

Year	2011	2012	2013	2014	2015
95% Upper limit	186,606	193,355	200,399	207,543	214,733
Predicted	179,124	181,420	183,715	186,011	188,307
95% Lower limit	171,642	169,484	167,032	164,479	161,880
Year	2016	2017	2018	2019	2020
95% Upper limit	221,947	229,175	236,412	243,656	250,904
Predicted	190,602	192,898	195,193	197,489	199,785
95% Lower limit	159,257	156,621	153,975	151,322	148,666

또한 원자료와 이중지수평활된 값에 대한 시도표 Figure 4.1을 살펴보면 전반적으로 잘 예측이 되고 있음을 알 수 있다.

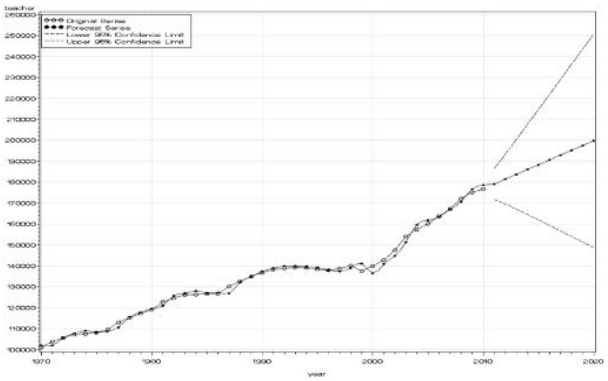

Figure 4.1 Raw data and estimated valus by exponential smoothing method

5. 결론

본 연구에서는 시계열 회귀분석 모형과 시계열 그룹별 회귀분석 모형, 지수평활법 모형을 사용하여 향후 10년간의 초등학교 교원 수가 예측되었다. Table 5.1에는 시계열 회귀분석 모형과 시계열 그룹별 회귀분석 모형, 지수평활법 모형에 의하여 예측된 결과가 함께 수록되어 있다.

Table 5.1 Estimated numbers of the elementary school teachers

Year	Method1	Method2	Method3
2011	178,444	179,633	179,124
2012	180,116	182,273	181,420
2013	182,146	184,875	183,715
2014	184,369	187,050	186,011
2015	186,648	189,117	188,307
2016	189,153	191,150	190,602
2017	191,770	193,288	192,898
2018	194,398	195,649	195,193
2019	197,022	198,105	197,489
2020	199,773	200,738	199,785

method1: time serise regression.
method2: group-specific time series regression analysis.
method3: exponential smoothing.

지수평활법 모형과 시계열 그룹별 회귀모형은 종속시계열 자체의 움직임만을 분석하여 장래 종속시계열의 값을 예측하는 방법이기 때문에 과거의 추세가 미래에도 지속되는 경우에만 예측값의 정확도가 일정수준으로 유지될 수 있게 된다. 초등학교 교원 수를 살펴보면 1970년에는 초등학교 교원 수가 101,095명이었으나 2010년에는 176,754명으로 1.7배 가까이 증가하였으며, 지난 42년간의 기간 중에서 초등학교 교원 수 감소가 발생한 1995년도, 1996년도와 1999연도의 정년단축과 이에 따른 명예퇴직의 영향으로 인하여 교원 수 감소를 제외하고는 대체적으로 지속적인 증가추세를 유지하고 있다. 시계열 회귀모형에서도 학생 수와 불변가격 교육 예산 규모의 시계열이 교원 수 시계열을 잘 설명하는 것으로 나타났는데 학생 수는 점점 감소하는 반면 불변가격 교육 예산은 시간이 지날수록 기하급수적으로 증가하기 때문에 초등학교 교원 수가 증가할 수 밖에 없다. 따라서 시계열 회귀모형과 시계열 그룹별 회귀모형, 지수평활법 모형에 따른 초등학교 교원 수 예측 결과 값만으로 어느 모형이 더 적절한지 판단하기에는 어렵다. 따라서 적절한 모형을 선택하기 위한 방법으로, 첫 번째 방법은 분석방법에 따른 모형 중에서 어느 모형이 좋은지 판단하기 위해서 자주 쓰는 평균오차제곱합 (MSE), 평균절대비율오차 (MAPE), 평균절대오차 (MAE) 통계량을 비교해 가장 작은 값을 가지는 모형을 초등학교 교원 수 예측 모형에 적절하다고 판단하였다.

Table 5.2 Comparisons of MSE, MAPE and MAE

Statistic	Method1	Method2	Method3
MSE	2689375	702080	2300586
Rank	3	1	2
MAPE	0.92	0.45	0.73
Rank	3	1	2
MAE	1228.26	599.13	1152.88
Rank	3	1	2

method1: time serise regression.
method2: group-specific time series regression analysis.
method3: exponential smoothing.

두 번째 방법은 2011년 실제 초등학교 교원 수와 시계열 회귀모형과 시계열 그룹별 회귀모형, 지수평활법 모형에 따른 2011년 초등학교 교원 수 예측값을 비교하여 차이가 가장 작은 모형을 초등학교 교원 수 예측 모형에 적절하다고 판단하였다.

Table 5.3 Difference between actual number and estimated values of the elementary school teachers in 2011

Year	2011	Year	2011	Year	2011
Method1	178,444	Method2	179,633	Method3	179,124
Actual teachers	180,623	Actual teachers	180,623	Actual teachers	180,623
Difference	-2,179	Difference	-990	Difference	-1,499

method1: time serise regression.
method2: group-specific time series regression analysis.
method3: exponential smoothing.

Table 5.2에서 시계열 그룹별 회귀모형이 MSE, MAPE, MAE 값이 가장 작은 값을 가지고 또한 위의 Table5.3에서도 시계열 그룹별 회귀모형이 2011년 실제 초등학교 교원 수와 예측값이 가장 작은 값을 가진다는 것을 보았다. 따라서 이 연구에서는 시계열 회귀모형과 지수평활 모형보다는 시계열 그룹별 회귀분석에 의한 모형이 초등학교 교원 수 시계열을 잘 설명하는 것으로 판단하여 시계열 그룹별 회귀분석 모형이 최종적으로 선택되었다.

$$\begin{aligned} \text{Teacher} &= -2818006 - 2869631D_t + 1483\text{Year} + 1433D_t \times \text{Year} + \epsilon_t \\ \epsilon &= 0.860865\epsilon_1 - 0.208651\epsilon_2 \end{aligned} \tag{5.1}$$

Table5.4에는 Table 2.4에 수록된 초등학교 학생 수 예측값과 최종적인 모형으로 선택된 시계열 그룹별 회귀모형의 향후 10년간 초등학교 교원 수 예측값으로부터 향후 10년간의 교원 일인당 학생 수 예측값이 제시되었다. Table5.4에 의하면 우리나라 초등학교 교원 일인당 학생 수는 2020년에 13.3명이 되는 것으로 예측되었다.

Table 5.4 Estimated student number per elementary school teacher

Year	Estimated student numbers	Estimated teacher numbers	Student number per teacher
2011	2,936,970	179,633	16.3
2012	2,790,722	182,273	15.3
2013	2,729,376	184,875	14.8
2014	2,703,802	187,050	14.5
2015	2,673,390	189,117	14.1
2016	2,686,806	191,150	14.1
2017	2,710,414	193,288	14.0
2018	2,713,372	195,649	13.9
2019	2,690,435	198,105	13.6
2020	2,679,463	200,738	13.3

Student number per teacher: Student number per the elementary school teacher one person.

Table 5.5는 최근 2008년 OECD 주요 국가의 초등학교 교원 일인당 학생 수로 실제 2008년의 우리나라 초등학교 교원 일인당 학생 수는 21.3명으로 OECD 주요 국가인 호주 15.8명, 오스트리아 12.9명, 미국 14.3명, 이탈리아 10.6명보다 많은 숫자이다 (OECD Educational Information Center, 2010). 하지만 OECD 회원국 간의 초등학교 교원 일인당 학생 수 격차는 점점 줄어들고 있다. 그 결과 저출산에 따란 학령인구 변화로 초등학교 교원 일인당 학생 수는 향후 1~3년 이내에 OECD 평균수준에 도달할 것으로 전망된다. 또한 우리나라 초등학교 교원 일인당 학생 수는 1970년부터 2010년까지 연평균 2.72%씩 감소하여 왔는데, 이 연구에서 예측된 교원 수에 따르면 향후 10년간의 교원 일인당 학생 수는 지속적인 학생 수 감소에 힘입어 연평균 2.21%씩 감소될 것으로 예측되었다.

Table 5.5 Student number per elementary school teacher in OECD major countries

OECD countries	Student number per teacher	OECD countries	Student number per teacher
OECD average	16.4	EU19 average	14.6
Australia	15.8	Austria	12.9
Belgium	12.6	USA	14.3
France	19.9	Czechia	18.1
Finland	14.4	Italy	10.6
Hungary	10.6	Spain	13.3

Student number per teacher: Student number per the elementary school teacher one person.

하지만 아래의 Table 5.6와 같이 교원 일인당 학생 수 한국교육개발원과 OECD 교육지표가 서로 다르다는 것을 알 수 있다. OECD 교육지표는 교사 일인당 학생 수의 경우 통계 수집 방법으로 학생을 가르치는 수업담당교원만을 교사로 산출한 결과이고 한국교육개발원의 경우 교장, 교감 등 관리직 교원 및 비교수 전문직을 포함하여 산출하였다. 따라서 2008년 한국교육개발원의 교원 일인당 학생 수가 OECD 교육지표상 교사 일인당 학생 수에 비해 초등학교의 경우 2.8명, 중학교의 경우 1.4명, 고등학교의 경우 1명 정도 적은 것을 볼 수 있다. 이는 전체 비중의 5~10%를 차지할 정도로 큰 수치이다.

Table 5.6 Student number per teacher reported by Korean Educational Development Institute and OECD education indicators (2008)

Division	KEDI	OECD EAG
elementary	21.3	24.1
middle school	18.8	20.2
high school	15.5	16.5

KEDI: Korean Educational Development Institute.
OECD EAG: OECD Education at a glance.

결국 OECD 교육지표에 따른 초등학교 교원 일인당 학생 수로 OECD 주요 국가와 비교하면 훨씬 많은 숫자이다. 우리나라도 OECD 교육지표대로 교원이 아닌 교사 일인당 학생 수로 OCDE 주요 국가랑 비교 하면 한국교육개발원의 교원 일인당 학생 수와 비교 했을 때 보다 더 많은 차이가 날 것이고 OECD 국가들의 평균 교원 수에 도달하는 데 많은 기간이 걸릴 것이다.

종적인 모형인 시계열 그룹별 회귀모형에 의하면 초등학교 교원 수는 향후 10년간 약 2만 4천명정도 증가하게 된다. 본 연구에서는 시계열 회귀모형과 시계열 그룹별 회귀모형, 지수평활법 모형을 사용하여 향후 10년간 초등학교 교원 수를 예측되었다. 지수평활법 모형은 예측 모형의 수립을 위하여 분석한 과거 자료의 추세에 의해서 미래 예측을 수행되는데 지수평활법 모형의 수립을 위하여 사용된 과거 자료 중에서 1999년 등의 자료는 교원의 정년단축과 명예퇴직으로 인하여 교원 수가 비정상적으로 감소한 자료이다. 따라서 이러한 자료는 향후 교원 수에 대한 예측의 정확성을 손상시키고 있을 가능성이 있다. 분석도중 지수평활법은 보통 평활상수가 0.01부터 0.3까지가 지수평활법 모형의 적용이 절적한 것으로 볼 수 있는데, 본 연구에서는 평활상수가 0.8이라는 높은 수치가 나와 지수평활법이 시계열자료 적합에 적절한 모형이 아니라는 결정을 하게 된다. 다음 연구에는 지수평활법 모형이 아닌 ARIMA모형을 이용하여 적절을 모형을 찾는 것이 좋다고 생각된다.

또한 시계열 회귀분석 모형에 의한 예측의 정확성은 교육예산 규모가 이 연구에서 가정하고 있는 바와 같이 실제로 연평균 3.19%씩 증가하는가에 따라서 결정된 것이기 때문에 실제의 미래 교원 수는 교육예산 규모에 결정적으로 의존한다고 볼 수 있다. 그러므로 향후 10년간의 초등학교 교원 수 예측값은 교원양성 정책을 수립하기 위한 정책판단의 기본적인 자료로 활용되는 것이 타당할 것으로 보인다.

참고문헌

Kim, H. C. (2000). Mid and long term forecasts of the number of graduates of universities of education and the number of newly recruited teachers. *The Joumal of Korean Education Administration*, **18**, 133-149.

Kim, H. C. (2002). Forecasting of the number of the elementary school teachers using time-series data analysis : A search for the explanatory variables and the comparison of the results from different forecasting methods. *The Joumal of Korean Education*, **29**, 113-130.

Kim, J. T. (2000a). The forecasting about the numbers of the third graders in a high-school until 2022 year in Daegu. *Journal of the Korean Data & Information Science Society*, **16**, 933-942.

Kim, J. T. (2005b). The forecasting for the numbers of a high-school graduate and the number limit of matriculation in Kyungbook. *Journal of the Korean Data & Information Science Society*, **16**, 969-977.

Korean Educational Development Institute (1970-2010). *Education statistical year book*, Educational Statistics & Information, Seoul.

OECD Educational Information Center. (2010). *2010 OECD education index*, OECD Report, Seoul.

Statistics Korea (2010). *Population projections*, Korean Statistical Information Service, Daejeon.

Yoon, Y. H. and Kim, J. T. (2012). Estimations of the student numbers by nonlinear regression model. *Journal of the Korean Data & Information Science Society*, **23**, 71-77.

Journal of the Korean Data & Information Science Society
2013, **24**(2), 321–332

http://dx.doi.org/10.7465/jkdi.2013.24.2.321
한국데이터정보과학회지

Time series regression model for forecasting the number of elementary school teachers[†]

Soo Rack Ryu[1] · Jong Tae Kim[2]

[12]Department of Computing · Statistics, Daegu University

Received 1 February 2013, revised 6 March 2013, accepted 14 March 2013

Abstract

Because of the continuous low birthrates, the number of the elementary students will decrease by 17% in 2020 compared to 2011. The purpose of this study is to forecast the number of elementary school teachers until 2020. We used the data in education statistical year books from 1970 to 2010. We used the time-series regression model, time series grouped regression model and exponential smoothing model to predict the number of teachers for the next ten years. Consequently time-series grouped regression model is a better model for forecasting the number of elementary school teachers than other models.

Keywords: Exponential smoothing, group-specific time series regression analysis, time serise regression.

[†] This research was supported by the Daegu University Research Grant 2010.
[1] Graduate student, Department Computing · Statistics, Daegu University, Kyoungbook 712-714, Korea.
[2] Corresponding author: Professor, Department Computing · Statistics, Daegu University, Kyoungbook 712-714, Korea. E-mail: jtkim@daegu.ac.kr