

ARIMAX 모형을 적용한 건강보험지출 장기전망

원종욱*·성병찬**·장인수* †

(*한국보건사회연구원, **중앙대학교 응용통계학과)

(2015년 1월 21일 논문접수, 2016년 5월 18일 최종 수정, 2016년 5월 26일 게재확정)

<국문초록>

본 연구는 지금까지 건강보험지출의 장기재정추계 방법론으로 적용된 요인별 예측방법과 조성법의 한계점을 보완하기 위하여 독립변수를 고려한 시계열 방법론인 ARIMAX 모형을 적용하여 건강보험지출을 장기추계하였다. ARIMAX 모형의 다양한 변형을 위하여 자기회귀항 및 독립변수의 지체항을 고려하였으며 또한 로그변환 ARIMAX 모형 및 시계열 회귀모형을 고려하였다.

여러 ARIMAX 모형을 적용한 결과 건강보험지출의 추계결과는 큰 차이를 보였다. 자기회귀항을 고려한 모형에서는 2050년 기준 561.37조 원으로 추계된 반면, 독립변수의 지체항을 고려한 모형의 경우 433.18조 원으로 추계되었다. 로그변환 ARIMAX 모형의 경우에는 430.42조 원으로 추계되었다. 이처럼 모형 간 추계 차이는 조성법이나 OECD의 요인별 예측방법에서도 발생하고 있으며 이는 모형 간의 기본 가정의 차이뿐만 아니라 수십 년 후 미래에 대한 장기 추계로 인하여 불가피하게 발생한다.

본 연구는 건강보험지출의 자기상관적 특성을 고려하는 시계열 분석 방법론 및 모형을 적용하고 다양한 시계열 모형에 따라 건강보험지출의 장기추계 결과를 도출하였다는 점에서 기존 연구와는 차별되며, 더욱 정교한 건강보험지출 장기전망을 위해 확장된 형태의 다변량 ARIMAX 모형과 같은 방법론에 대한 가능성을 제시하고자 한다.

핵심용어: 건강보험지출, 장기전망, ARIMAX 모형, 시계열 모형

† 교신저자: 장인수, 연락처: 044)287-8311, sescis@kihasa.re.kr

주소 : 세종특별자치시 시청대로 370, 세종국책연구단지 3층 한국보건사회연구원

본고는 보건복지부 연구용역 「사회보장 재정추계 기반강화 연구」의 일부 내용을 수정하여 논문화하였음을 밝힘

I. 연구의 목적 및 배경

우리나라의 사회보장지출은 지속적으로 급격하게 증가할 것으로 예상되고 있으며, 특히 가장 큰 비중을 차지하고 있는 건강보험의 경우 제도적 성숙 측면뿐 아니라 재정적 지속가능성을 모두 고려하여야 할 필요성이 존재한다(신화연 외, 2013, p.39). 이에 본 연구에서는 건강보험의 장기재정추계 방법론에 대해 고찰하고, 그동안 많은 연구가 이루어지지 않은 시계열 분석 방법론을 다양하게 적용해 보고자 한다. 또한, 선행연구의 추계결과와 비교하고자 한다.

지금까지 건강보험지출을 장기적으로 추계한 방법론은 조성법과 요인별 예측방법(OECD방법론) 등이 있다. 조성법은 성별 및 연령별 인구와 각 인구그룹의 1인당 평균 급여비를 예측한 후 예측치를 곱하여 총 지출을 계산하는 방법이다. 또한 요인별 예측방법은 소득, 인구, 새로운 의료기술 발달, 정부정책 변화 등 건강보험지출수준에 영향을 미칠 수 있는 요소로 구분하여 건강보험지출을 추계하는 방법이다. 선행연구들이 주로 건강보험지출에 영향을 미치는 (독립)변수들의 예측에 초점을 맞추는 반면, 본 연구에서는 건강보험지출이 시간에 따른 시계열 자료의 특성을 띠고 있다는 점에서, 건강보험지출 자체(종속변수)를 모형화하고 예측하는 시계열 분석 방법을 고려한다. 특히, 독립변수들의 영향을 동시에 고려할 수 있는 자기회귀차분 이동평균모형(ARIMAX)을 적용하고자 한다.

지금까지의 선행연구에서 주로 수행된 조성법이나, 요인별 예측방법의 한계점과 본 연구에서 적용하고자 하는 ARIMAX모형과의 차이점을 정리하면 다음과 같다.

먼저, 과거 실적치 기간 대비 추계기간은 요인별 예측 방법에서 적용되었던 기간과 동일하나, 자료 취급에 있어 이를 시계열로 고려한다는 점을 그 특징으로 제시할 수 있다. 즉, 우리나라 사례에 적용하는 요인별 예측 방법의 경우, 우리나라의 건강보험제도가 안정화된 것이 최근의 일이므로, 분석을 위한 기간이 2000-2010년의 11년으로 비교적 짧다는 단점이 존재하므로, 과거 제도를 구조적으로 반영할 수 있는 모형 설정이 필요하다.

또한, 요인별 예측방법의 경우 투입되는 독립변수가 동태적인 특성을 가지고 있는 점을 고려하지 못하므로 이를 보완하기 위한 방법으로서 ARIMAX 모형의 적용이 요구된다.

ARIMAX 모형의 독립변수는, 다음과 같은 두 가지 측면을 고려하기 위하여, OECD, 통계청, 한국은행 등 공신력 있는 기관에 의해 전망되어 있는 자료를 고려하였다. 첫째는 이 자료들이 정부의 다른 분야의 장단기 재정추계 시 그대로 이용되고 있어 건강보험지출의 장기추계도 이와 일관되게 고려되어야 한다는 점이며, 둘째는 건강보험 예측오차의 원인을 독립변수의 모형화가 아닌 ARIMAX 모형화에만 두기 위함이다. 따라서, 1인당 GDP와 65세 이상 인구의 진료비지출비중 지표를 도출하기 위한 인구 고령화를 전망치는 통계청의 자료를 원용하거나

이를 바탕으로 계산하였다.

마지막으로 요인별 예측 방법에서의 잔차 요인을 정책과 공급자 측면에서 세부적으로 구분하여, 이를 독립변수로 투입하고자 하였다. 즉, 기존 연구에서는 잔차 요인이 인구 및 소득 이외의 다른 요인으로 크게 분류한 것과 다르게, 본 연구에서는 잔차 요인을 정책과 공급자 요인으로 따로 분해하였다는 특징이 존재한다. 먼저 정책적 요인으로는 65세 이상 인구의 의료비 상대지출비중을 고려한 변수를 투입하여 이러한 고령인구의 지출 수준을 어느 정도 수용(충당)할 수 있는지 통제하고자 하였다. 아울러 요인별 예측방법의 잔차이자 공급자 측면의 요인으로는 고가의료장비와 의약품연구개발비로 구분하여 고려하고자 하였다.

II. 선행연구 고찰

시계열을 분석하고 예측할 수 있는 가장 기본적인 모형은 Box와 Jenkins에 의해 고안된 자기회귀와 차분, 이동평균을 고려하는 ARIMA 모형이라고 할 수 있지만, 시계열 자료는 자기 자신의 특성에 의해서만 결정되고 예측될 수 있는 것은 결코 아니다. 이는 Durka et al. (2011)에서 확인할 수 있는 바와 같이 시계열 자료가 법적인 조치나, 정책의 변화, 환경적인 규제 변화 등 여러 외부적 요인에 의해서도 영향을 받을 수 있다는 것과 그 맥락을 같이 하는 것이라 할 수 있다. ARIMA 모형에 추가적으로 외부적 영향을 고려할 수 있는 ARIMAX 모형은 전이함수모형(transfer function model)이나 동태적 회귀모형(dynamic regression model)의 개념과 연결되어 많은 문헌들에서 인용 및 설명되고 있다. 예를 들어, Pankratz (1991)은 ARIMAX 모형을 동태적 회귀모형으로 고려하였으며, Box et al. (2008)은 전이함수모형의 형태로 파악하였다.

ARIMAX 모형은 다양한 분야에서 널리 사용되고 있다. 최근 연구로서 몇 가지 예를 살펴보면, Chaleampong와 Kruangpradit (2013)는 태국의 수출액을 예측하기 위하여 ARIMA 모형과 각 수출국에 대한 경기선행지표를 추가적으로 고려하는 ARIMAX 모형을 비교하였으며, Joyeux et al. (2012)는 관광 분야에서 호주의 여권(passport) 수요 예측을 위하여 ARIMAX 모형이 장기적으로 ARIMA 모형보다 예측력이 더 좋다는 결론을 얻었다. Andrews et al. (2013)은 ARIMA 및 ARIMAX 모형을 이용하여 장기적 장애급여 청구 건 수를 예측하였고, Li와 Chen (2007)은 교통사고 건 수를 예측하기 위하여, ARIMAX 모형을 이용하였으며 이를 통하여 GDP 및 총인구 수, 운전자 수, 총 자동차 대 수 등과 같은 다양한 독립변수들의 영향을 고려하였다. 또한, Alberto et al. (2006)은 항공 수요 예측에서 동일한 모형을 사용하였다.

한편, 국내에서는 유금록(2007)의 연구에서 한국과 일본의 중앙정부 세출예산 과거 실적치에서 세침분포(leptokurtic distribution)가 나타나는 근거를 규명하기 위해 ARIMAX 모형을 적

용한 바 있다. 이론적 근거에 따른 독립변수 투입과 데이터의 안정성 여부를 검증하기 위한 단위근 검정을 바탕으로 한 자료의 차분 여부 결정과 적합의 단계를 거쳐, 중앙정부 세출예산의 퍼센트 변화라는 종속변수에 대하여 독립변수의 효과를 분석하였다. 그러나, 유금록의 연구에는 과거 실적치의 패턴과 영향 관계를 분석하기 위해 ARIMAX 모형을 적용한 것이며, 미래에 대한 예측을 수행하지는 않았다는 점에서 앞서 살펴본 선행연구와는 구별된다.

본 연구는 1990년부터 2010년까지 21년의 과거 자료를 이용하여 표본 길이의 약 2.5배인 2060년까지의 미래 50년을 예측한다. 이는 우리나라를 비롯한 대부분의 외국의 경우에도 장기재정추계기간을 60~100년으로 범으로 명문화하고 있어 불가피한 당면과제이다. 따라서, 본 연구에서의 건강보험 추계결과는 장기추계에 따른 예측오차를 줄이기 위해서 향후 주기적인 독립변수의 예측치 또는 모형 갱신(updating)이 이루어져야 함을 유의하여야 한다.

III. 건강보험지출의 ARIMAX 모형화

1. 연구방법론

본 연구에서 적용하는 모형은 종속변수의 시계열적 특성을 고려하는 대표적인 모형인 Box와 Jenkins의 ARIMA 모형을 근간으로 하고 있다. 자기회귀(autoregressive, AR) 모형은 종속변수의 과거값들로 현재의 시계열을 설명하며, 자기회귀이동평균(autoregressive - moving average, ARMA) 모형은 종속변수의 과거값 뿐만 아니라 백색잡음의 과거값도 포함하여 시계열을 모형화한다. 일반적으로 사회과학 분야에서의 시계열 변수들은 비정상적(nonstationary) 경우가 많으며 이와 같은 경우 변수의 차분을 고려하는 자기회귀차분 이동평균(ARIMA) 모형을 사용한다. 건강보험 장기추계를 위한 ARIMA 모형의 분석결과는 모형 적합성과 변수별 유의성이 낮은 것으로 분석되었으며, 자기회귀나 차분 등의 고려 여부에 따라 추계결과의 편차가 큰 것으로 나타났다. ARIMA(p,d,q)에서 p는 자기회귀의 차수, q는 이동평균의 차수를 나타내며 d는 차분의 차수이다. 다만, ARIMA 모형은 월이나 분기, 반기, 년 등의 비교적 짧은 기간을 예측함에 장점이 있으나, 현재 분석에서 장기간을 추계한다는 점에서는 한계점이 존재한다고 하겠다.

일반적으로 ARIMA 모형은 독립변수를 포함하지 않으므로 전통적인 회귀분석에서의 관심사인 특정 독립변수와의 관계를 알아내기가 어렵다. 또한, 여러 가지의 독립변수들이 결합적으로 부여할 수 있는 우수한 장기 예측력을 가질 수 없게 된다. ARIMAX 모형은 ARIMA 모형에 독립변수까지 고려하는 '외생변수를 포함한 자기회귀이동평균(autoregressive moving average model with exogenous inputs model, ARIMAX(p,d,q,r))' 모형을 말한다. 이 때, p, d

와 q 는 각각 ARIMA 모형의 차수, r 은 독립변수의 지체차수(lagged order)를 각각 의미하며, 아울러 적합모형이 모두 일정한 유의수준에서 유의하지 않을 때는 그 중 설명력이 가장 높은 모형을 적용한다.

2. 분석과정

1) 변수 변환

적절한 시계열 모형을 수립하기 위해서는 변수 변환의 과정이 요구되는데, 이를 위해서는 로그변환이 자주 사용된다. 특히, 거시경제변수들은 시간이 지남에 따라 변동성(분산)이 증가하는 이분산성을 보이므로 이를 안정화하기 위하여 로그변환을 가하는 경우가 일반적이며, 이러한 과정을 바탕으로 모형 수립에서 필요한 가정인 등분산성이나 정규성을 충족한다고 할 수 있다.

시계열 자료를 이용한 (회귀)모형에서 종속변수와 독립변수들이 비정상적일 경우 그 모형은 허구적일 수 있다; 이 경우를 허구적 회귀(spurious regression)라고 부른다. 따라서 비정상 시계열 자료를 이용한 모형은 초기 단계에서 차분(difference)을 행하는 것이 일반적이다. 그 예로서 본 연구의 비교 방법론인 OECD 방법론을 보면 로그 변환 이후 차분을 한 것을 볼 수 있다. 이는 앞서 언급한 허구적 회귀를 방지하기 위한 일종의 장치로 이해된다. 차분 이외에 충분한 추세 제거(detrending)의 과정을 거칠 수도 있으나 본 연구에서는 이러한 결정적(deterministic) 추세향의 영향은 장기적으로 모든 다른 독립변수 항들을 압도할 수 있는 위험이 있으므로 사용하지 않았다.

한편, 종속변수와 독립변수들이 모두 비정상적임에도 불구하고 회귀모형이 가능한 경우는 공적분(cointegration) 관계가 있는 것으로 해석할 수 있다. 이는 다변량 시계열 영역에서는 일반적인 이론이며, 종속변수를 한 개로 고려하는 시계열 회귀모형 수립에서는 적합 후 잔차에 대한 단위근 검정을 통해서 이를 방지할 수 있다. 즉, 비정상 시계열 모형에서 그 적합 모형의 잔차에 단위근이 없는 경우 공적분 관계를 이용한 모형으로 볼 수 있으며, 그 잔차에 단위근이 존재하는 경우 해당 종속변수와 독립변수는 반드시 차분을 거쳐서 모형 적합이 이루어져야 하는 것이다.

2) 변수 선택

다음으로 변수 선택의 단계를 거쳐야 한다. 즉, 이는 시계열 변수들이 독립변수로 고려될 경우, 어떤 과거 시차변수들이 독립변수로 투입될 것인가를 결정하는 단계이다. 이때는 다음과 같은 방법들을 고려할 수 있을 것이다.

먼저, 전이함수모형에서 사용되는 것처럼, 종속변수와 독립변수들간에 교차상관계수(cross correlation coefficients)를 살펴보는 방법을 들 수 있다.

두 번째는 기설정된 독립변수에 대하여 다양한 시차의 독립변수들을 미리 생성한 후, 회귀 분석에서 고려되는 단계적 회귀분석(stepwise regression)을 고려하는 방법이다.

마지막으로는 첫 번째와 두 번째의 방법을 혼합한 방법이다. 이는 자료에 근거한 이론 고찰을 배경으로 한다. 본 방법은 1) 또는 2)의 방법에 의하여 선택된 변수들이 타당한 이론적 해석과 연결되지 않을 수 있다는 단점을 극복하기 위하여 관련 이론과의 조화에 초점을 둔 방법이라고 할 수 있겠다.

3) 시계열 회귀모형 고려

변수선택 이후 다양한 모형의 적합을 시도할 수 있다. 특히 독립변수를 고려할 수 있는 단변량 시계열 모형으로 다음을 고려할 수 있다.

- ARIMAX 모형
- 전이함수모형
- AUTOREG (regression model with autoregressive errors) 모형

위의 세 가지 모형 중에서 가장 일반적인 모형은 전이함수모형으로 볼 수 있으나 그 모형의 가정이 엄격하고 적합 및 해석이 어렵다. 따라서, 본 연구에서는 ARIMAX 모형을 위주로 건강보험지출을 적합 및 해석한다. ARIMAX 모형은 AUTOREG 모형보다 더 일반적인 모형으로 알려져 있다. 그러나, 전이함수모형을 포함한 ARIMAX 모형은 변수선택 결과에 따라 모형 적합이나 분석 이후의 결과에 많은 차이가 발생하므로 해석에 주의하여야 한다.

4) 잔차검정

회귀분석과 마찬가지로 시계열 모형적합에서 잔차분석은 중요한 단계로 이해된다. 이를 통하여 수립된 모형의 적절성 여부를 판단할 수 있기 때문이다. 주로 백색잡음 검토, 잔차의 자기상관함수(autocorrelation coefficient function) 및 부분자기상관함수(partial autocorrelation coefficient function) 검토, 단위근 검정 등에 대한 사항이 검토되어야 할 것이다.

먼저 백색잡음의 검토는 포트만테우(portmanteau) 검정을 통하여 수행할 수 있다. 잔차의 자기상관함수 및 부분자기상관함수를 통하여서는 백색잡음 가정 검토뿐만 아니라 모형의 수정 방향에 대하여 검토할 수 있다. 즉, ARIMA 모형의 차수가 재조정될 수 있는 것이다. 단위근 검정은 차분에 대한 필요성을 검토하는 과정으로서 ADF(Augmented Dickey Fuller) 검정 또는 KPSS(Kwiatkowski-Phillips-Schmidt-Shin) 검정을 통하여 이루어진다. 앞서 언급된 허구적 회귀를 검토한다는 점에서 그 의미를 찾을 수 있다.

5) 최종모형의 선택

앞서 순차적으로 언급된 네 단계를 통하여 적절한 모형을 수립하면 선택 가능한 여러 후보 모형들이 있을 수 있으며, 이 때 각 모형의 분석을 수행하여 다음과 같은 기준에 의하여 최종 모형을 선택할 수 있다.

- 정보량 기준: AIC (Akaike information criterion), SBC (Schwarz bayesian information criterion)
- MSE (mean squared errors), MAE (mean absolute errors)

본 연구에서 적용한 ARIMAX(p,d,q,r) 모형은 다음과 같이 전이함수모형의 특별한 형태로 고려될 수 있다. 편의상, 원자료 y_{0t} 에 대하여 $y_t = (1 - B)^d y_{0t}$ 는 정상성을 만족하는 종속 시계열이라고 가정한다.

$$\phi(B)y_t = \mu + \psi(B)x_t + \theta(B)z_t$$

여기서, $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ 는 시차연산자(B)를 이용하여 종속 시계열의 과거값의 영향을 포함할 수 있으며, $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ 는 백색잡음 z_t 가 서로 상관될 수 있는 유연성 있는 구조를 허용한다. 특히, 외생적 변수인 x_t 가 종속변수인 y_t 와 지체되어 영향력을 가질 수 있도록 그 계수들의 구조가 $\psi(B) = 1 - \psi_1 B - \dots - \psi_r B^r$ 와 같다.

ARIMAX 모형의 단점으로 고려되는 것은 독립변수들의 지체 차수(r)를 결정하는 것으로 볼 수 있으나, 이는 종속 시계열과 독립 시계열 간의 교차상관계수를 통하여 그 후보 차수들을 정할 수 있다. ARIMAX 모형 적합에 대한 전반적인 내용은 SAS/ETS 9.2 (2014) 매뉴얼을 참고할 수 있다. 또한 Hyndman과 Athanasopoulos (2013)은 동적 회귀모형의 관점에서 ARIMAX 모형을 자세하게 설명하고 있다.

3. 실증분석을 위한 자료 구축

본 연구의 종속변수는 건강보험지출이며, 독립변수는 1인당 GDP, 인구가중치를 고려한 진료비 지출비중지표(인구가중치×65세 이상 진료비 지출 비중지표), 의약품연구개발비, 의료고가 장비이다. 독립변수를 다음과 같이 설정한 이유는 앞서 살펴본 요인별 예측방법(OECD 방법론)을 보완하기 위함이다. 즉, 요인별 예측방법은 크게 변수를 인구, 소득, 잔차 요인으로 구분하여 추계하고 있으며, 그중 특히 잔차 요인은 보장성 강화, 제도 변화, 의료기술 발전, 지출효율화 정책 등의 특성을 모두 고려한 것으로 정리할 수 있다. 본 연구에서는 잔차 요인을 더

세분화하여 직접적으로 투입할 수 있는 변수로서 인구가중치를 고려한 진료비 지출비중지표, 의약품연구개발비, 의료고가장비를 투입하였으며, 이것은 각각 보장성 강화, 의료기술 발전 측면을 고려한 변수라고 판단하였기 때문이다. 또한 분석의 세밀성을 도모하기 위해서는 잔차 요인을 세분화하고 이를 구체적이고 직접적으로 고려하여야 할 필요성이 있다고 판단하였다. 특히 인구가중치를 고려한 진료비 지출비중지표는 정책변수로 해석할 수 있는데, 이의 의미는 65세 이상 인구의 진료비 상대지출 비중을 정책이 어느 정도 수용할 수 있는지를 뜻하는 것으로 해석 가능하다. 즉, 인구 고령화에 따라 65세 이상 고령인구의 진료비 급증 수용가능성과 정부의 건강보험지출은 밀접한 관련이 있다고 판단하여 이의 영향력을 파악하고자 한 것이다. <표 1>은 분석을 위해 구축된 변수들의 실제값이다.

<표 1> 분석을 위한 변수별 자료 구축(1990년-2010년)

년도/ 항목	건강보험 지출 (단위: 조 원)	1인당 GDP (단위: 백만 원)	인구 고령화율 (단위: %) (A)	65세 이상 진료비 지출 비중지표 (B)	인구가중 치를 고려한 진료비 지출비중 지표 (AXB)	의약품 연구 개발비 (단위: 억 원)	고가 의료장비 (단위: 대)
1990	2.16	4.46	0.0512	1.607	0.08228	97.9	154
1991	2.49	5.34	0.0523	1.772	0.09268	163.2	184
1992	2.97	6.03	0.0536	1.906	0.10216	203.9	206
1993	3.46	6.76	0.0551	2.119	0.11676	144.9	231
1994	3.97	7.84	0.0569	2.170	0.12347	248.1	240
1995	5.08	9.08	0.0589	2.310	0.13606	288.0	293
1996	6.46	10.12	0.0614	2.345	0.14398	240.7	858
1997	7.80	11.02	0.0637	2.385	0.15192	313.9	1140
1998	8.79	10.83	0.0663	2.509	0.16635	366.1	1285
1999	9.61	11.78	0.0692	2.708	0.18739	673.4	1508
2000	10.74	12.83	0.0722	2.682	0.19364	803.0	1781
2001	14.11	13.76	0.0756	2.778	0.21002	713.2	1964
2002	14.80	15.13	0.0792	3.012	0.23855	976.9	2148
2003	15.97	16.03	0.0829	2.994	0.24820	1125.1	2318
2004	17.33	17.21	0.0867	2.974	0.25785	1710.7	2421
2005	19.98	17.97	0.0907	3.076	0.27899	2711.6	2554
2006	22.82	18.79	0.0949	3.156	0.29950	3350.1	2800
2007	25.89	20.06	0.0993	3.355	0.33315	3851.5	2995
2008	28.27	20.97	0.1032	3.678	0.37957	4337.3	3037
2009	31.19	21.65	0.1069	3.690	0.39446	4599.4	3054
2010	34.93	23.74	0.1103	3.634	0.40083	5193.4	2865

주: 1) 65세 이상 진료비지출비중지표는 65세 이상 인구의 총 급여실적을 인구수로 나눈 평균진료비에 대하여 15세 이상-64세 미만 인구의 평균진료비를 1로 하였을 때의 지표값임
 2) 의약품연구개발비 지출은 각 년도 기업경영분석, 한국은행에서 발췌하여 적용
 3) 의료고가장비 대수는 이상규(2012)에서 발췌하여 적용

각 독립변수는 1인당 GDP와 독립변수 2를 도출하기 위한 인구 고령화율을 제외하고는 모두 추세분석(trend analysis)을 통하여 미래 50년의 자료를 모두 예측하였다. 즉, 인구 고령화율과 1인당 GDP 추세치는 재정추계위원회의 명목 GDP 추세치와 통계청의 장래추계인구를 사용하여 계산하였으며, 65세 이상 인구의 상대진료비지출비중 지표, 의약품연구개발비 및 고가의료장비는 추세의 설명력과 이론적 수용가능성을 고려하여 가장 적절하다고 판단되는 함수식을 적용하였다. 각각 추정된 추세분석 함수식은 <표 2>와 같다.

<표 2> 변수의 장기 예측치 도출을 위한 추세분석 함수식

독립변수명	추세분석 함수식	결정계수(R^2)
독립변수 2 중 65세 이상 진료비지출비중지표	$Y = 1.744e^{0.037x}$	0.9685
독립변수 3: 의약품연구개발비	$Y = 18.513x^2 - 181.02x + 602.75$	0.9693
독립변수 4: 고가의료장비	$Y = 1,211\text{Log}(x) - 996$	0.9502

이를 바탕으로 하여 독립변수를 예측한 결과는 <표 3>과 같다. 특히, 65세 이상 진료비지출 비중 지표는 정책변수를 직접적으로 조절하기 위하여 2010년 3.634에서 2060년 24.123으로 설정하여 예측한 값들이다.

<표 3> 독립변수 예측(2011년-2060년) 결과

년도/항목	1인당 GDP (독립변수1) 단위 : 백 만원	인구 고령화율 (A)	65세이상 진료비지출 비중지표(B)	인구가중치 를고려한 진료비지출 비중지표(A XB) (독립변수2)	의약품 연구개발비 (독립변수3) 단위 : 억 원	고가 의료장비 (독립변수4) 단위 : 대
2011	24.85	0.1136	3.936	0.4471	5,581	2,747
2012	25.45	0.1178	4.084	0.4811	6,233	2,801
2013	26.44	0.1222	4.238	0.5179	6,922	2,853
2014	27.94	0.1266	4.398	0.5568	7,648	2,902
2015	29.76	0.1309	4.564	0.5974	8,411	2,950
2016	31.83	0.1351	4.736	0.6398	9,211	2,995
2017	34.06	0.1396	4.914	0.6861	10,048	3,039
2018	36.41	0.1446	5.100	0.7374	10,923	3,082
2019	38.79	0.1504	5.292	0.7959	11,834	3,123
2020	41.20	0.1572	5.491	0.8632	12,782	3,163
2021	43.67	0.1645	5.698	0.9374	13,767	3,201
2022	46.18	0.1724	5.913	1.0194	14,790	3,238
2023	48.75	0.1807	6.136	1.1088	15,849	3,274
2024	51.38	0.1895	6.367	1.2066	16,945	3,310
2025	54.06	0.1988	6.607	1.3135	18,079	3,344
2026	56.81	0.2083	6.856	1.4282	19,249	3,377

년도/항목	1인당 GDP (독립변수1) 단위 : 백 만원	인구 고령화율 (A)	65세이상 진료비지출 비중지표(B)	인구가중치 를고려한 진료비지출 비중지표(A XB) (독립변수2)	의약품 연구개발비 (독립변수3) 단위 : 억 원	고가 의료장비 (독립변수4) 단위 : 대
2027	59.60	0.2177	7.115	1.5489	20,457	3,409
2028	62.45	0.2266	7.383	1.6730	21,701	3,441
2029	65.33	0.2351	7.661	1.8012	22,983	3,471
2030	68.25	0.2433	7.950	1.9343	24,301	3,501
2031	71.21	0.2512	8.250	2.0723	25,657	3,530
2032	74.23	0.259	8.561	2.2172	27,049	3,559
2033	77.34	0.267	8.883	2.3719	28,479	3,587
2034	80.48	0.2756	9.218	2.5405	29,946	3,614
2035	83.68	0.2843	9.566	2.7195	31,449	3,640
2036	86.97	0.2929	9.926	2.9074	32,990	3,667
2037	90.32	0.3013	10.300	3.1035	34,568	3,692
2038	93.76	0.3093	10.689	3.3060	36,182	3,717
2039	97.32	0.3166	11.092	3.5116	37,834	3,741
2040	101.01	0.323	11.510	3.7176	39,523	3,765
2041	104.86	0.3287	11.943	3.9258	41,249	3,789
2042	108.86	0.334	12.394	4.1395	43,012	3,812
2043	113.02	0.3391	12.861	4.3611	44,812	3,835
2044	117.35	0.3446	13.346	4.5989	46,648	3,857
2045	121.85	0.3507	13.849	4.8567	48,522	3,879
2046	126.50	0.357	14.371	5.1303	50,433	3,900
2047	131.34	0.3628	14.912	5.4102	52,381	3,921
2048	136.41	0.3675	15.474	5.6868	54,366	3,942
2049	141.68	0.3711	16.058	5.9590	56,388	3,962
2050	147.17	0.3739	16.663	6.2302	58,447	3,982
2051	152.86	0.376	17.291	6.5014	60,543	4,002
2052	158.77	0.378	17.943	6.7823	62,677	4,021
2053	164.93	0.3798	18.619	7.0715	64,847	4,040
2054	171.32	0.3817	19.321	7.3747	67,054	4,059
2055	177.90	0.384	20.049	7.6988	69,298	4,078
2056	184.67	0.3869	20.805	8.0494	71,579	4,096
2057	191.66	0.3903	21.589	8.4262	73,898	4,114
2058	198.87	0.3939	22.403	8.8244	76,253	4,132
2059	206.31	0.3975	23.247	9.2407	78,645	4,149
2060	213.96	0.4009	24.123	9.6711	81,074	4,166

IV. ARIMAX 모형 적합 및 분석결과

1. 자기회귀항을 고려한 모형

우선 4개의 독립변수를 모두 투입하되, 종속변수의 자기회귀항 AR(1)을 고려한 ARIMAX(1,0,0,[0,0,0,0]) 모형을 고려하였다. 여기서, [0,0,0,0]은 고려된 4개의 독립변수들이 지체없이 모두 종속변수와 같은 시간에 관측된 시계열 변수임을 나타낸다. 자기회귀항 AR(1)만이 고려된 이유는 <표 4> 및 <표 5>와 같은 자기상관함수와 부분자기상관함수의 형태를 건강보험지출 시계열이 보이기 때문이다. 또한, ADF 단위근 검정을 통하여 건강보험지출 시계열이 정상 시계열임을 확인하였다. 단위근 검정 시 옵션은 디키-풀러의 ρ 타입 검정을 선택하였으며 AR의 최대 차수는 2로 설정하였고, 모형의 형태가 zero mean, single mean 또는 trend에 관계없이 단위근의 귀무가설이 기각되었다.

자기회귀항을 고려한 모형의 추정 결과인 <표 6>을 살펴보면, 예상과 달리 종속변수의 AR(1) 항은 5% 유의수준에서 유의하지 않음을 알 수 있으며, 더불어 의료비상대지출지수 및 의료고가장비 변수 또한 유의한 독립변수가 될 수 없다. 따라서, 정책변수인 의료비상대지출지수를 제외하고, 모형에서 유의하지 않은 독립변수인 AR(1) 및 의료고가장비 변수를 제거하였으며 추가적으로 다중공선성 문제(VIF=54,102,17; 상태지수=21.69)를 해결하기 위하여 분산변동 비율의 가장 큰 값을 차지하고 있는 1인당 GDP 변수를 제거하였다. 최종 모형(축소된 자기회귀모형)의 추정 결과는 <표 7>과 같다.

최종모형의 적합력인 AIC와 SBC는 각각 58.41과 61.54이었으며 다중공선성 문제가 없음을 확인하였다(VIF=11,11; 상태지수=6.56). 참고로, 본 연구에서는 의료비상대지출지수 변수와의 다중공선성을 해결하기 위하여 1인당 GDP 변수를 모형에서 제거하는 방안을 고려하였으나, GDP 변수의 중요성을 제고하기 위하여 종속변수를 GDP 대비 건강보험지출 비율로 정의하는 ARIMAX 모형을 고려할 수도 있다. 포맷도우 검정에 의하여 잔차항의 백색잡음 가정이 만족되었고, ADF 검정 결과 잔차에 단위근이 없음이 확인되어 허구적 회귀의 가능성이 없음을 검토하였다.

<표 4> 건강보험지출의 자기상관함수 분석결과

Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	S.E
0	97.351613	1.00000											*****											0
1	81.653982	0.83875			.								*****											0.218218
2	66.995077	0.68818			.								*****											0.338555
3	52.932732	0.54373			.								*****		.									0.399654
4	39.111353	0.40175			.								*****		.									0.433451
5	26.361007	0.27078			.								*****		.									0.450834

<표 5> 건강보험지출의 부분자기상관함수 분석결과

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	0.83875			.									*****										
2	-0.05171			.	*									.									
3	-0.06750			.	*									.									
4	-0.08412			.	**									.									
5	-0.06298			.	*									.									

<표 6> 자기회귀모형의 적합 결과

	β	s.e	t Value	$P > t $	시차
1인당 GDP	0.5675*	0.3173	1.79	0.0737	0
고령화율*65세 이상 의료비상대지출지수	34.8956	22.6399	1.54	0.1232	0
의약품연구개발비	0.0020***	0.0006	3.12	0.0018	0
의료고가장비	0.3956×10^{-4}	0.0010	0.04	0.9697	0
상수	-3.0923**	1.7455	-2.55	0.0108	0
AR(1)	0.3054	0.2884	1.06	0.2897	0
n			21		
AIC			60.3086		
BIC			66.5757		

주: *,**,***은 각각 10%, 5%, 1%에서 유의함을 뜻함

<표 7> 축소된 자기회귀모형의 적합 결과

	β	s.e	t Value	$P > t $
고령화율*65세 이상 의료비상대지출지수	79.3055***	6.7470	11.75	<.0001
의약품연구개발비	0.0012***	0.0004	3.11	0.0019
상수	-5.2908***	0.9165	-5.77	<.0001

주: ***은 1%에서 유의함을 뜻함

축소된 자기회귀모형을 이용하여 노인의료비 지출비중 시나리오별 향후 50년 동안의 건강보험지출 추계는 <표 8>과 같이 계산되었다. 시나리오는 모두 5가지로 구성되어 있으며, 각각 2060년 65세 이상 진료비지출비중 지표를 5배(5.145), 10배(10.074), 15배(15.176), 20배(20.057), 24배(24.123)로 설정한 것이다. 독립변수의 예측치는 앞서 적용한 바와 같이 추세분석을 바탕으로 한 예측치를 적용하였다.

분석결과 5가지 시나리오에서 2060년 기준 건강보험 지출은 각각 259조 원, 416조 원, 578조 원, 733조 원, 862조 원으로 추계되었다.

<표 8> 자기회귀모형에 따른 노인의료비 지출비중 시나리오별 건강보험지출 장기추계 (2011~2060)

(단위: 조 원)

년도/ 65세(+) 인구 의료비 지출비중	5배 (5.145)	10배 (10.074)	15배 (15.176)	20배 (20.057)	24배 (24.123)
2011	34.55	34.87	34.79	35.11	37.09
2012	36.79	37.66	37.90	38.37	40.60
2013	39.15	40.58	41.13	41.93	44.39
2014	41.64	43.62	44.49	45.60	48.38
2015	44.17	46.71	48.06	49.41	52.50
2016	46.67	49.92	51.75	53.41	56.90
2017	49.38	53.42	55.72	57.78	61.67
2018	52.21	57.13	59.98	62.52	66.72
2019	55.48	61.27	64.76	67.77	72.53
2020	59.04	65.86	70.14	73.71	79.02
2021	62.96	70.89	75.96	80.25	86.20
2022	67.00	76.28	82.23	87.31	93.89
2023	71.33	82.04	89.02	94.89	102.34
2024	75.87	88.16	96.33	103.15	111.47
2025	80.68	94.64	104.16	112.01	121.37
2026	85.63	101.49	112.43	121.47	131.86
2027	90.62	108.54	121.07	131.38	142.96
2028	95.57	115.64	129.83	141.57	154.34
2029	100.49	122.70	138.80	151.96	166.08
2030	105.38	129.97	147.97	162.80	178.26
2031	110.24	137.28	157.35	173.92	190.89
2032	115.14	144.88	167.09	185.56	204.12
2033	120.25	152.68	177.35	197.81	218.19
2034	125.64	161.09	188.37	211.13	233.41
2035	131.15	169.77	199.91	225.13	249.48
2036	136.71	178.75	211.89	239.73	266.22
2037	142.32	187.84	224.16	254.86	283.80
2038	147.82	197.06	236.80	270.50	301.83
2039	153.20	206.17	249.40	286.19	320.21
2040	158.39	215.01	261.88	302.01	338.65

년도/ 65세(+) 인구 의료비 지출비중	5배 (5.145)	10배 (10.074)	15배 (15.176)	20배 (20.057)	24배 (24.123)
2041	163.39	223.82	274.41	317.87	357.29
2042	168.27	232.59	287.07	334.18	376.37
2043	173.20	241.56	300.09	350.93	396.21
2044	178.34	250.98	313.87	368.75	417.37
2045	183.84	261.00	328.57	387.97	440.15
2046	189.38	271.39	344.03	408.19	464.18
2047	194.90	281.81	359.69	428.77	488.80
2048	200.06	291.81	374.93	449.08	513.23
2049	204.95	301.38	389.73	469.04	537.32
2050	209.57	310.60	404.34	488.88	561.37
2051	213.91	319.71	418.76	508.69	585.46
2052	218.39	328.86	433.54	529.19	610.39
2053	222.83	338.06	448.61	550.04	636.01
2054	227.31	347.62	464.28	571.90	662.86
2055	232.08	357.62	480.78	595.06	691.34
2056	237.14	368.47	498.53	620.02	722.09
2057	242.39	379.67	517.19	646.46	754.71
2058	247.86	391.48	536.85	674.28	789.20
2059	253.44	403.49	557.18	703.27	825.24
2060	259.00	415.63	577.89	733.09	862.36

2. 독립변수의 지체항을 고려한 모형

본 모형은 건강보험지출을 모형화 및 예측하기 위하여 ARIMA 자체 또는 독립변수를 종속변수와 동일한 시차로 입력하지 않고, 종속변수와 시차가 있는 독립변수의 지체항들로 구성하는 모형이다. 본 모형에서는 모든 4개의 독립변수를 고려하되, 종속변수와 동일 시차에 한정하지 않았으며 최적의 시차를 단계별 회귀분석(stepwise regression)을 통하여 선택하였다. 그 결과, 1인당 GDP만이 현시점으로 선택되었고 나머지 독립변수들은 모두 1시점 이전의 값이 선택되었다. 모형의 마지막에 표기되어 있는 ARIMAX의 마지막 인수 [0,1,1,1]은 4개의 독립변수들의 지체시차를 나타내고 있다.

모형의 추정결과는 다음의 <표 9>와 같다. 주요 결과를 살펴보면, 자기회귀항을 고려한 모형과 달리 절편을 포함한 모든 독립변수들이 유의수준 10%에서 유의하였으며, 정보량 기준 또한 AIC=41.87와 SBC=46.85로서 자기회귀항을 고려한 모형보다 훨씬 더 좋은 모형 적합력을 보여주고 있다. 잔차에 대한 백색잡음 가정 및 포맨토우 검정은 잘 만족됨을 확인할 수 있었다. 그러나, 상태지수값이 26.14(VIF=81,112,17,25)로 다중공선성에 문제가 있음이 나타났다. 이를 해결할 수 있는 대안으로 분산비율의 값이 가장 큰 값(0.81)을 가지는 1인당 GDP 변수를 제거하였으며 결과된 모형은 상태지수값이 14.41(VIF=48,14,19)로 변하였다. 축소된 모형의 적합 결과는 <표 10>과 같다.

본 모형으로 추계한 각 시나리오별 2011년~2060년까지의 건강보험지출 추계결과는 <표 11>과 같다. 독립변수의 시차를 고려하지 않고 종속변수의 자기회귀항을 고려한 모형에 비하여 건강보험지출 예측이 작은 값으로 나타나고 있다. 이에 대한 주된 원인은, 예측 모형에서 사용하는 독립변수 시차의 차이이며, 더불어 '인구가중치 진료비 지표'에 대한 회귀계수값의 차이(79.31 vs 54.02)로 비롯된 것으로 볼 수 있다.

<표 9> 독립변수의 지체항을 고려한 모형의 적합 결과

	β	s.e	t Value	$P > t $	시차
1인당 GDP	0.4976**	0.2285	2.18	0.0294	0
고령화율*65세 이상 의료비상대지출지수	27.8369*	15.8562	1.76	0.0792	1
의약품연구개발비	0.0022***	0.0004	5.81	<.0001	1
고가의료장비 대 수	0.0013**	0.0007	2.04	0.0411	1
상수	-3.3904***	1.0591	-3.20	0.0014	0
n			21		
AIC			41.8684		
BIC			46.8470		

주 : *,**,***은 각각 10%, 5%, 1%에서 유의함을 뜻함

<표 10> 축소된 모형(독립변수의 지체항을 고려한 모형)의 적합 결과

	β	s.e	t Value	$P > t $	시차
고령화율*65세 이상 의료비상대지출지수	54.0200***	11.4855	4.70	<.0001	1
의약품연구개발비	0.0019***	0.0004	4.86	<.0001	1
고가의료장비 대 수	0.0021***	0.0006	3.30	0.001	1
상수	-2.5829**	1.1021	-2.34	0.0191	0

주 : **,***은 각각 5%, 1%에서 유의함을 뜻함

<표 11> 독립변수의 지체항을 고려한 모형에 의한 노인의료비 지출비중 시나리오별 건강보험지출 추계
(단위: 조 원)

년도/ 65세(+) 인구 의료비 지출비중	5배 (5.145)	10배 (10.074)	15배 (15.176)	20배 (20.057)	24배 (24.123)
2011	34.72	34.72	34.72	34.72	34.72
2012	35.96	36.18	36.12	36.34	37.69
2013	38.27	38.86	39.02	39.35	40.86
2014	40.69	41.66	42.04	42.58	44.26
2015	43.23	44.58	45.18	45.93	47.82
2016	45.84	47.57	48.49	49.41	51.51
2017	48.46	50.68	51.92	53.05	55.43
2018	51.26	54.01	55.58	56.98	59.63
2019	54.17	57.52	59.46	61.19	64.06
2020	57.42	61.37	63.74	65.80	69.04
2021	60.90	65.55	68.46	70.90	74.51
2022	64.66	70.07	73.52	76.44	80.49
2023	68.55	74.87	78.92	82.38	86.86
2024	72.66	79.95	84.71	88.70	93.78
2025	76.95	85.32	90.89	95.53	101.20
2026	81.47	90.97	97.46	102.80	109.18
2027	86.11	96.91	104.36	110.52	117.60
2028	90.81	103.02	111.56	118.58	126.47
2029	95.53	109.20	118.87	126.86	135.56
2030	100.27	115.39	126.36	135.33	144.94
2031	105.01	121.76	134.02	144.12	154.66
2032	109.78	128.20	141.86	153.15	164.71
2033	114.61	134.86	149.99	162.58	175.22
2034	119.61	141.71	158.51	172.44	186.33
2035	124.85	148.99	167.58	183.08	198.26
2036	130.21	156.51	177.04	194.22	210.80
2037	135.63	164.26	186.84	205.80	223.85
2038	141.13	172.13	196.87	217.78	237.50
2039	146.58	180.13	207.19	230.15	251.49

년도/ 65세(+) 인구 의료비 지출비중	5배 (5.145)	10배 (10.074)	15배 (15.176)	20배 (20.057)	24배 (24.123)
2040	152.00	188.08	217.52	242.59	265.76
2041	157.32	195.89	227.81	255.15	280.10
2042	162.55	203.71	238.17	267.78	294.62
2043	167.73	211.54	248.66	280.74	309.48
2044	172.99	219.56	259.42	294.05	324.89
2045	178.42	227.91	270.74	308.13	341.24
2046	184.14	236.70	282.73	323.19	358.73
2047	189.93	245.78	295.27	338.97	377.11
2048	195.73	254.94	307.98	355.03	395.93
2049	201.33	263.83	320.44	370.95	414.65
2050	206.78	272.47	332.65	386.67	433.18
2051	212.09	280.91	344.76	402.35	451.72
2052	217.24	289.31	356.78	418.04	470.33
2053	222.52	297.77	369.08	434.23	489.55
2054	227.82	306.31	381.61	450.71	509.26
2055	233.18	315.13	394.60	467.90	529.86
2056	238.78	324.29	408.19	486.03	551.61
2057	244.60	334.06	422.65	505.41	574.94
2058	250.61	344.12	437.79	525.84	599.58
2059	256.79	354.62	453.64	547.25	625.53
2060	263.09	365.29	469.99	569.49	652.57

3. 로그변환 모형

시계열 분석 방법론을 적용할 경우 일반적으로 거시경제변수의 분산 안정화 변환을 위해 로그변환을 변수에 취하는 것이 바람직하다. 이를 위하여 로그변환 후 건강보험지출의 ARIMAX 모형, 즉 Log-ARIMAX 모형을 시도해 보기로 한다.

우선, 건강보험 지출 변수에 로그변환이 적절한지 검토하기 위하여 Pankratz(1983, pp.256-258) 방법에 따른 SAS의 BOXCOXAR 매크로를 이용하였으며, 로그변환이 원 시계열의 모형화에 비해 RMSE(residual mean square errors)를 35.61%로 감소시킬 수 있음을 확인하였다.

<표 12>와 <표 13>의 로그변환 이후 건강보험지출의 자기상관함수와 부분자기상관함수를 통해서 볼 때, 정상성을 만족하는 AR(1) 모형의 특징을 강하게 가지고 있다. 따라서, 로그변환 이전의 자료에 대한 모형화 순서와 유사하게 최초로 시도하는 모형은 Log-ARIMAX(1,0,0, [0,0,0,0]) 모형이다. <표 14>의 분석결과에서 확인할 수 있는 바와 같이 AR(1) 변수는 유의하지 않다는 결론을 낼 수 있으며 이는 로그변환 이전의 자료와 유사하다.

또한, 로그변환 이전의 모형에서와 동일하게 독립변수의 지체항을 고려한 Log-ARIMAX(0,0,0,[0,1,1,1]) 모형을 적용해 보면 <표 15>와 같은 결과를 얻을 수 있다. 로그변환 이전과 달리 의료비 상대지출지수가 유의하지 않게 나타났으며, 모형의 적합력은

AIC=62.79와 BIC=57.81로 나타났다.

<표 12> 로그변환 건강보험지출의 자기상관함수 분석결과

Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	S.E
0	0.726966	1.00000												*****										0
1	0.622914	0.85687												*****										0.218218
2	0.517608	0.71201												*****										0.342848
3	0.414369	0.57000												*****										0.407218
4	0.311703	0.42877												*****										0.443587
5	0.210621	0.28973												*****										0.462902

<표 13> 로그변환 건강보험지출의 부분자기상관함수 분석결과

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	0.85687													*****									
2	-0.08357									**													
3	-0.07381									*													
4	-0.08758									**													
5	-0.09044									**													

<표 14> Log-ARIMAX(1,0,0,[0,0,0,0]) 모형의 적합 결과

	β	s.e	t Value	$P > t $	시차
1인당 GDP	0.3600*	0.1854	1.94	0.0522	0
고령화율*65세 이상 의료비상대지출지수	0.9604***	0.2277	4.22	<.0001	0
의약품연구개발비 고가의료장비 대 수	0.0253	0.0462	0.55	0.5844	0
상수	1.4968	0.9400	1.59	0.1113	0
AR(1)	-0.1752	0.2864	-0.61	0.5408	
n			21		
AIC			61.7035		
BIC			55.4364		

주 : *, ***은 각각 10%, 1%에서 유의함을 뜻함

로그변환 모형에 의한 예측치는 지수변환(antilog transformation)을 통하여 원 시계열로 재변환(back-transformation)될 수 있다. 그러나, 이러한 예측값은 원 시계열의 중앙값의 예측치이며 원 시계열의 평균값을 과소예측하는 경향이 있음이 알려져 있다. 원 시계열의 평균을 예측하기 위하여서는 재변환시 예측치의 표준오차를 고려하여야 한다(SAS/ETS 9.2; Cowpertwait and Metcalfe, 2009). 이를 반영하여 추계된 2011년~2060년까지의 5가지 시나리오별 예측치는 <표 16>와 같다. 2060년 기준 64세 이상 인구의 의료비지출 상대 비중을 24배로 가정할 경우 건강보험지출 예측치는 696.94조 원으로 도출되었다.

<표 15> Log-ARIMAX(0,0,0,[0,1,1,1]) 모형의 적합 결과

	β	s.e	t Value	$P > t $	시차
1인당 GDP	1.0025***	0.2026	4.95	<.0001	0
고령화율*65세 이상 의료비상대지출지수	0.1360	0.2466	0.55	0.5815	1
의약품연구개발비	0.1238***	0.0460	2.69	0.0071	1
고가의료장비 대 수 상수	0.1533*** -1.8066*	0.0322 1.0631	4.77 -1.7	<.0001 0.0893	1 0
n			21		
AIC			62.7924		
BIC			57.8138		

주 : *,**,***은 각각 10%, 5%, 1%에서 유의함을 뜻함

<표 16> Log-ARIMAX(0,0,0,[0,1,1,1]) 모형에 의한 노인의료비 지출비중 시나리오별 건강보험지출 추계

(단위: 조 원)

년도/ 65세(+) 인구 의료비 지출비중	5배 (5.145)	10배 (10.074)	15배 (15.176)	20배 (20.057)	24배 (24.123)
2011	33.68	33.68	33.68	33.68	33.68
2012	35.52	35.52	35.52	35.52	35.52
2013	36.63	36.68	36.67	36.72	37.01
2014	38.92	39.06	39.09	39.16	39.48
2015	42.04	42.26	42.35	42.47	42.83
2016	45.74	46.06	46.20	46.37	46.78
2017	49.91	50.34	50.56	50.77	51.24
2018	54.43	55.00	55.31	55.58	56.12
2019	59.27	60.02	60.41	60.76	61.37
2020	64.30	65.23	65.74	66.17	66.84
2021	69.57	70.70	71.32	71.84	72.61
2022	75.11	76.47	77.25	77.86	78.72
2023	80.92	82.52	83.45	84.19	85.14
2024	86.97	88.87	89.96	90.83	91.89
2025	93.31	95.52	96.80	97.80	98.99
2026	99.90	102.46	103.96	105.11	106.42
2027	106.80	109.73	111.47	112.78	114.24
2028	113.93	117.27	119.27	120.77	122.36
2029	121.30	125.09	127.37	129.06	130.81
2030	128.85	133.13	135.70	137.61	139.52
2031	136.58	141.35	144.27	146.40	148.48
2032	144.50	149.84	153.09	155.48	157.74
2033	152.65	158.58	162.21	164.86	167.32
2034	161.11	167.69	171.72	174.66	177.33
2035	169.83	177.08	181.54	184.78	187.68
2036	178.85	186.82	191.75	195.32	198.45
2037	188.20	196.94	202.37	206.30	209.68

년도/ 65세(+) 인구 의료비 지출비중	5배 (5.145)	10배 (10.074)	15배 (15.176)	20배 (20.057)	24배 (24.123)
2038	197.83	207.41	213.37	217.66	221.30
2039	207.79	218.25	224.77	229.47	233.40
2040	218.12	229.53	236.66	241.79	246.01
2041	228.85	241.26	249.04	254.62	259.17
2042	239.98	253.45	261.92	268.00	272.88
2043	251.56	266.16	275.38	281.97	287.21
2044	263.62	279.44	289.45	296.61	302.21
2045	276.23	293.35	304.20	311.95	317.96
2046	289.42	307.93	319.67	328.07	334.51
2047	303.27	323.24	335.95	345.04	351.93
2048	317.72	339.28	353.03	362.86	370.23
2049	332.88	356.12	370.98	381.58	389.48
2050	348.59	373.63	389.65	401.08	409.54
2051	364.90	391.80	409.07	421.39	430.42
2052	381.79	410.67	429.27	442.53	452.17
2053	399.27	430.31	450.29	464.55	474.84
2054	417.61	450.87	472.36	487.69	498.66
2055	436.69	472.32	495.41	511.86	523.57
2056	456.44	494.61	519.38	537.02	549.50
2057	476.98	517.80	544.36	563.28	576.57
2058	498.38	542.06	570.49	590.76	604.91
2059	520.66	567.32	597.77	619.48	634.54
2060	543.83	593.67	626.24	649.46	665.49

4. AUTOREG 모형의 적합 탐색

추가적으로 시계열적 특성을 가지고 있는 모든 변수에 대하여 시계열적 처리를 가하지 않고 일반적인 다중회귀모형으로 적합을 시도하였다. 즉, 모든 ARIMAX 모형의 차수를 0으로 두는 회귀모형 ARIMAX(0,0,0,[0,0,0,0])의 적합을 시도하였다. <표 17>의 분석결과에서 확인할 수 있는 바와 같이 1인당 GDP와 고가의료장비가 10%의 유의수준에서 유의하지 않은 것으로 나타났으나, AIC와 BIC 측면에서 모형의 설명력은 적절한 수준을 유지하였다.

<표 17> ARIMAX(0,0,0,[0,0,0,0])의 모형의 적합 결과

	β	s.e	t Value	$P > t $	시차
1인당 GDP	0.4190	0.2906	1.44	0.1495	0
고령화율*65세 이상 의료비상대지출지수	46.1815**	20.4451	2.26	0.0239	0
의약품연구개발비	0.0017***	0.0005	3.31	0.0009	0
고가의료장비 대 수 상수	0.0001	0.0009	0.14	0.8862	0
	-4.71793***	0.8983	-3.23	0.0012	0
n			21		
AIC			59.3790		
BIC			64.6016		

주 : *,**,***은 각각 10%, 5%, 1%에서 유의함을 뜻함

<표 18> ARIMAX(0,0,0,[0,0,0,0])모형 잔차의 자기상관함수 분석결과

Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	S.E
0	0.614777	1.00000											*****											0
1	0.094647	0.15395				.							***	.										0.218218
2	-0.130674	-0.21256				.				****				.										0.223330
3	-0.037057	-0.06028				.				*				.										0.232764
4	-0.114265	-0.18586				.				****				.										0.233506
5	-0.117633	-0.19134				.				****				.										0.240448

<표 19> ARIMAX(0,0,0,[0,0,0,0])모형 잔차의 부분자기상관함수 분석결과

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	0.15395												***	.									
2	-0.24199											****		.									
3	0.01917																						
4	-0.24949											****		.									
5	-0.13891											***		.									

그러나 시계열 모형의 고찰 부분에서 언급된 AUTOREG 모형의 적합은 필요치 않은 것으로 보인다. <표 18>과 <표 19>에서와 같이 회귀모형의 잔차에 대한 자기상관함수(ACF)와 부분자기상관함수(PACF)에서 자기상관 관계를 발견할 수 없기 때문이다. 일반적으로 AUTOREG 모형은 잔차에 자기회귀항을 고려한다는 점에서 종속변수의 자기회귀항을 고려하는 모형과 다른 모형으로 간주된다.

5. 주요 선행연구와의 분석결과 비교

본 연구에서의 건강보험지출 추계결과를 주요 선행연구와 비교하면 <표 20>과 같다. 요인별 예측방법(OECD 방법론)에 따라 추계된 건강보험지출 추계치의 경우 소득탄력도를 0.8/1.0/1.2

로 가정함에 따라 추계결과에 큰 차이를 보이는 것으로 나타났다. 즉, 표에서 확인할 수 있는 바와 같이 OECD 방법론 5, 4, 3 순으로 큰 값으로 추계되었으며, 2050년을 기준으로 815.9조 원, 634.5조 원, 444.8조 원이며 2060년을 기준으로 1,297.9조 원, 865.4조 원, 450.3조 원으로 각각 추계되었다.

OECD방법론 1과 같이 보험급여비증가율 중 잔차증가율이 2050년에 0%로 수렴함으로 가정하였을 경우 2050년을 기준으로 572.43조 원이, OECD방법론 2와 같이 2030년에 0%로 수렴하였을 경우에는 2050년 기준 404.74조 원이 각각 추계되었다. 또한 원종욱 외(2012)에서 확인할 수 있는 바와 같이 조성법의 분석결과는 2050년을 기준으로 511.33조 원으로 추계되었다.

마지막으로 본 연구에서 분석한 ARIMAX 모형의 분석결과는 2060년을 기준으로 자기회귀항을 고려하는 접근에서는 862.36조 원, 독립변수의 지체항을 고려할 경우 652.57조 원이 각각 추계되었으며, 로그변환 ARIMAX 모형을 적용할 경우 697.68조 원으로 추계되었다.

<그림 1>은 <표 20>의 각 방법에 따른 추계 결과를 그림으로 나타낸 것이다. 이를 통해서 볼 때, 본 연구에서 고려된 ARIMAX 모형에 의한 건강보험지출의 장기 추계치는 OECD 방법들 및 조성법에 비하여 중단기 추계는 비교적 낮은 수준으로 예측하고 장기 추계로 갈수록 점점 기울기가 급해지고 있다. 2045년 이후의 장기 추계에 있어서는 ARIMAX 모형에 의한 예측이 다른 방법들에 의한 것들의 평균 수준으로 보여진다.

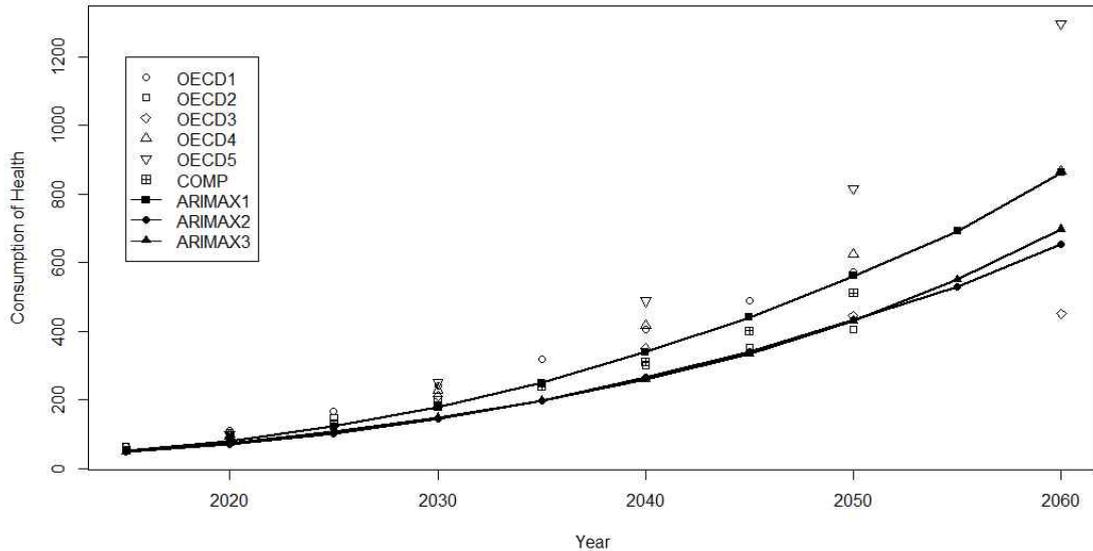
<표 20> 각 모형(OECD방법, 조성법, ARIMAX)에 따른 건강보험지출 추계결과 비교

(단위: 조 원)

	① OECD 방법1	② OECD 방법2	③ OECD 방법3	④ OECD 방법4	⑤ OECD 방법5	⑥ 조성법	⑦ ARIMAX I	⑧ ARIMAX II	⑨ ARIMAX III
2015	64.43	63.25	-	-	-	53.93	52.50	47.82	46.78
2020	109.29	103.34	91.8	94.7	98.2	85.91	79.02	69.04	72.61
2025	164.79	147.33	-	-	-	130.11	121.37	101.20	106.42
2030	239.21	198.65	211.1	228.3	249.1	181.37	178.26	144.94	148.48
2035	317.33	246.24	-	-	-	238.94	249.48	198.26	198.45
2040	406.20	300.30	350.7	416.0	489.2	311.33	338.65	265.76	259.17
2045	490.20	351.33	-	-	-	401.24	440.15	341.24	334.51
2050	572.43	404.74	444.8	624.5	815.9	511.33	561.37	433.18	430.42
2055	-	-	-	-	-	-	691.34	529.86	549.50
2060	-	-	450.3	865.4	1,297.9	-	862.36	652.57	697.68

주: 1) ①의 방법은 원종욱 외(2011)의 분석결과로 보험급여비증가율 중 잔차증가율이 2050년에 0%로 수렴함을 가정
 2) ②의 방법은 원종욱 외(2011)의 분석결과로서, 2030년에 0%로 수렴함을 가정
 3) ③, ④, ⑤는 각각 소득탄력도 0.8, 1.0, 1.2를 가정
 4) ⑥은 원종욱 외(2012)의 분석결과임
 5) ⑦,⑧,⑨의 방법은 본 연구의 분석방법으로서, 2060년을 기준으로 하여 15-64세 인구 대비 65세 이상 노인의 의료비 지출비중이 24배가 됨을 가정

[그림 1] 각 모형에 따른 건강보험지출 추계결과 비교 그림



V. 결론

본 연구에서는 인구 고령화에 따라 급속하게 증가할 것으로 예상되는 사회보험지출 중 가장 큰 비중을 차지하고 있는 건강보험의 장기재정추계 방법론과 추계결과를 알아보고, ARIMAX 모형과 같은 시계열 분석 방법론을 적용한 분석을 수행하였다.

지금까지 건강보험지출의 장기전망을 수행한 선행연구는 주로 조성법이나 요인별 예측방법(OECD 방법론)을 적용하여 추계하였으며, 이에 따른 추계결과는 2050년을 기준으로 404.74조 원에서 815.9조 원까지로 요인별 증가율과 소득탄력도 등 요소별 가정에 따라 큰 차이를 보이는 것으로 나타났다.

본 연구에서는 건강보험 지출이 시계열 자료의 특성을 분명하게 띠고 있음에도 불구하고, 이를 고려하지 못한 한계점을 보완하기 위해 독립변수를 고려하는 자기회귀이동평균모형인 ARIMAX 모형을 바탕으로 실증분석을 수행하고자 하였다. 또한 종속변수는 연간 건강보험지출이며, 독립변수는 요인별 예측방법에서의 인구, 소득, 잔차 요인을 조금 더 세분화하여 1인당 GDP, 인구가중치를 고려한 65세 이상 진료비지출비중지표, 의약품연구개발비, 의료고가장비의 네 개 변수가 투입되었다. 독립변수의 미래치 투입을 위해 1인당 GDP와 65세 이상 진료

비지출비중지표를 도출하기 위한 인구 고령화율은 통계청의 장래인구전망과 재정추계위원회의 명목GDP 예측치를 원용하였고, 나머지 변수는 설명력이 가장 높은 함수식을 적용하여 추세분석(trend analysis)을 바탕으로 예측치를 구하였다. 시계열 모형 적합 과정에서 유의하여야 할 변수 변환이나 변수 선택, 잔차 검정 등의 기본적 과정을 거쳤으며, 종속변수의 자기회귀항 및 독립변수의 지체항을 고려한 모형, 로그변환 ARIMAX 모형을 고려하였고, AUTOREG 모형의 적합 가능성을 탐색하였다.

본 연구에서 적용한 여러 ARIMAX 모형을 적용한 결과 투입변수의 종류 및 자기회귀 과정 등에 따라 추정계수에 차이를 보였으며, 이에 따라 추계결과도 차이를 보였다. 자기회귀항을 고려한 ARIMAX(1,0,0,[0,0,0,0])모형의 경우 2050년 기준 561.37조 원으로 추계된 반면, 독립변수의 지체항을 고려한 모형 ARIMAX(0,0,0,[0,1,1,1])의 경우 433.18조 원으로 추계되었다. 이러한 결과는 앞서 조성법이나 요인별 예측방법을 적용한 추계결과보다 대체로 적은 수치이다. 분석방법론의 차이 및 독립변수를 고려하는 관점으로 인하여, 선행연구인 요인별 예측방법이나 조성법의 분석결과와 차이가 존재하는 이유를 직접적으로 비교하기는 쉽지 않다. 그러나, 요인별 예측방법의 경우 소득탄력도를 0.8, 1.0, 1.2로 적용함에 따라 그 추계치가 크게 차이나며, 조성법은 인구와 급여비 증가율을 각각의 기간별 구간에 대해 일정하게 증가함을 가정하였는데, 이는 일반적인 회귀식의 형태가 아니므로, 이러한 방법론별 특성에 따라 추계치의 차이가 나는 것으로 이해할 수 있다.

앞선 분석결과에서 언급한 바와 같이 로그변환 이후 건강보험지출 시계열의 경우 앞서 확인한 자기상관함수와 부분자기상관함수를 확인할 수 있는 바와 같이 강한 AR(1)의 특징에 따라 차분에 의한 모형 적합은 특별한 정보를 주기가 어려울 것으로 보였으므로 이를 고려하는 것은 추가적인 논의를 요한다고 하겠다. 또한, 본 연구는 정책적 요인으로서의 65세 이상 인구의 의료비 상대지출 비중이 건강보험지출에 미치는 영향을 분석하고자 하였으며 이 독립변수는 다른 독립변수들보다 우선 순위를 두어 변수 제거에 있어 가장 후순위에 두었다.

본 연구는 명문법에 의한 건강보험지출의 장기추계를 목적으로 하고 있어, 불가피하게 표본의 2.5배인 미래 50년을 예측한다. 따라서, 본 연구에 의한 장기추계는 큰 예측오차가 발생할 수 있으며 이를 방지하기 위해서는 독립변수의 예측치 갱신 및 모형 갱신의 과정이 주기적으로 수반되어야 한다.

결론적으로 본 연구는 자료의 시계열적 특성을 고려하고, 모형의 적합 과정을 거쳐 정교한 시계열 계량 방법론을 적용하여 여러 ARIMAX 모형에 따라 추계결과를 도출하였다는 점에서 기존 연구와의 차별성이 존재하며, 더 나아가서는 건강보험지출 장기 전망을 수행하기 위해 본 연구에서 적용한 ARIMAX 모형 및 관련 방법론을 고려할 수 있음을 제시하고자 한다.

참고문헌

- 건강보험공단, 의료보험통계연보, 건강보험통계연보, 각 년도.
- 신화연 · 원종욱 · 허재준 · 서문희 · 강미나 · 이선주 · 김태은, 사회보장 재정추계모형 개발 연구, 보건복지부, 한국보건사회연구원, 2013.
- 신화연 · 원종욱 · 장인수 · 최요한 · 김태은 · 성병찬, 사회보장 재정추계 기반강화 연구, 보건복지부, 한국보건사회연구원, 2014.
- 원종욱 · 신화연 · 윤문구 · 김문길 · 강지원 · 남궁은하, 사회복지 재정추계 모형개발 연구, 한국보건사회연구원, 2011.
- 원종욱 · 신화연 · 임완섭 · 김태은, 사회보장 재정추계 방법론 개발을 위한 기초연구, 한국보건사회연구원, 2012.
- 유금록, “한국과 일본의 중앙정부 세출예산에 있어서 단절균형분석”, 정책분석평가학회지, 17(3), 한국정책분석평가학회, 2007, pp.221-249.
- 이근찬 · 남혜진, 의료장비에 관한 적정 공급체계 연구, 건강보험심사평가원, 2013.
- 이상규, “우리나라 고가 의료장비 현황과 정책대안: 정책의 수직적, 수평적 동기화”, 대한의사협회지, 55(10), 2012, pp.950-958.
- 조신섭·손영숙, SAS/ETS를 이용한 시계열분석, 율곡출판사, 2009.
- Alberto Andreoni·Maria Nadia Postorino, A Multivariate Arima Model to Forecast Air Transport Demand, European transport conferecne; ETC proceedings, 2006.
- Bartosz Przywara., Projecting Future Health Care Expenditure at European Level : Drivers, Methodology and Main Results, European Economy, Economic Papers 417, July 2010.
- Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. *Time series analysis: Forecasting and control*, John Wiley & Sons Inc., New York., 2008.
- Bruce H. Andrews, Matthew D. Dean, Robert Swain, Caroline Cole., Building ARIMA and ARIMAX Models for Predicting Long-Term Disability Benefit Application Rates in the Public / Private Sectors, University of Southern Maine., 2013.
- Chaleampong Kongcharoen·Tapanee Kruangpradit, Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) Model for Thailand Export, ISF 2013 Conference Paper, 2013.

- Durka Peter·Pastorekova Silva, ARIMA vs. ARIMAX - which approach is better to analyze and forecast macroeconomic time series?, Proceeding of 30th International Conference Mathematical Methods in Economics., 2011.
- Economics Department, Economic Policy Committee, Public Spending on Health and Long-Term Care : A New Set of Projection, Working Party No.1 on Macroeconomics and Structural Policy Analysis, ECO/CPE/WPI(2012)23, 2012.
- Hyndman, R.J., Athanasopoulos, G. *Forecasting: principles and practice*. OTexts., 2013.
- Joyeux, R., Milunovich, G., and Rigg, J., "Forecasting Demand for Australian Passports", *Asia Pacific Journal of Tourism Research*, 17(1), 2012, pp. 100-119.
- Li, C. and Chen, J., Traffic Accident Macro Forecast Based on ARIMAX Model, 2009 International Conference on Measuring Technology and Mechatronics Automation., 2009.
- Mark J.Warshawsky, "An Enhanced Macroeconomics Approach to Long-Range Projections of Health Care and Social Security Expenditures as a Share of GDP", *Journal of Policy Modeling*, 21(4), 1997, pp.413-426.
- Pankratz, A. *Forecasting with Univariate Box-Jenkins Models: Concepts and Cases*, New York: John Wiley, 1983.
- Pankratz, A. *Forecasting with dynamic regression models*, Wiley-Interscience., 1991.
- Paul S.P. Cowpertwait, Andrew V. Metcalfe., *Introductory Time Series with R*, Springer, 2009.
- SAS Institute Inc, SAS/ETS 9.2, Cary, NC., 2014.

<Abstract>

An Application of ARIMAX for Predicting Long-Term National Health Insurance Expenditure in Korea

Jongwook Won* · Byeongchan Seong** · Insu Chang*

(*KIHASA, **Department of Applied Statistics, Chung-Ang Univ.)

This study is aimed at applying the ARIMAX model—a time-series forecasting approach that takes into account relevant independent variables—to overcome some of the limitations of such traditional approaches as component method and driver-based simulation in forecasting National Health Insurance expenditure. The versions of the ARIMAX approach that we used in this study include first-order autoregressive model, multiple regression model, log ARIMAX model, and autoreg model.

The results were that the projected expenditures differed widely depending on the forecasting versions used. For example, the National Health Insurance expenditure in 2050 was projected to be anywhere between KRW430.42 trillion and KRW561.37 trillion. Such differences are traceable in large part to the differences in the assumptions each version incorporates. This study presents the possibility of the future application of an extended multivariate ARIMAX model for long-term projection of National Health Insurance expenditure.

Key Words : Health Expenditure, Long-term Projection, ARIMAX model, Time-series

[†] Corresponding author: Insu Chang, Tel: +82-44-287-8311, e-mail: sescis@kihasa.re.kr

Address: Korea Institute for Health & Social Affairs, Building D, 370 Sicheong-daero, Sejong City, 339-007, Korea