

Optimized pricing based on proper estimation of rating factor distribution

Yeong-Hwa Kim^{a,1} · Chul-Hee Jeon^b

^aDepartment of Applied Statistics, Chung-Ang University;

^bDepartment of Statistics, Graduate School of Chung-Ang University

(Received June 29, 2016; Revised August 2, 2016; Accepted August 5, 2016)

Abstract

Auto insurance is an insurance product that requires the proper application of pricing techniques due to intense market competition and the rate regulations of financial authorities. Especially, population change according to aging and rating factor segmentation mainly affect the pricing process. This study suggests a pricing optimization methodology through the proper estimation of age factors. To properly estimate the future distribution of age factor, age change, renewal and conversion of customers are considered as main effects for the optimization of estimation and application. The properness and effectiveness for the suggested method will be proved by a comparison of results applied (one for current distribution and the other for future distribution) at the off-balance process. This study suggests an appropriate risk estimation methodology based on optimization that uses the proper estimation of future distribution to protect from the over or under estimation of risk.

Keywords: auto insurance, future distribution, off-balance

1. 서론

최근 자동차 보험에서 발생하는 손해액이 매년 빠르게 증가함에 따라 자동차 보험 손해율이 90%를 상회하는 수준이고, 보험업계에서 제시하는 자동차 보험 손익분기점 손해율은 77% 선으로 현재 위험 수준에 달한 상황이다. 또한 보험회사들의 치열한 시장경쟁으로 가격경쟁이 심해지고, 금융당국은 물가 안정을 이유로 우회적인 방법을 통한 그림자 규제로 보험회사들은 적정 보험료를 제시하기 어려운 상황이다. 이로 인한 손익 악화로 보험회사들은 적자를 면치 못하고 있다. 이에 따라 최근 금융당국이 ‘보험업 경쟁력 강화 로드맵’을 발표하면서 보험상품의 형태나 가격에 대한 요율 규제를 다소 완화하기로 밝혔으나, 여전히 합리적이지 못한 규제 상황에서 해결방안이 필요한 실정이다. 또한 현재 우리나라는 급격한 출생률 저하와 함께 세계에서 가장 빠른 고령화 속도를 보이고 있다. 이에 따라 고령운전자의 수와 고령운전자에 의한 교통사고 건수가 빠르게 증가하면서 연령그룹별 교통사고 발생 유형과 연령그룹별 자동차 보험 손해를 변화에도 큰 영향을 미치고 있다.

This research was supported by the Chung-Ang University Research Scholarship Grants in 2014.

¹Corresponding author: Department of Applied Statistics, Chung-Ang University, 84 Heuksuk-Ro, Dongjak-Gu, Seoul 06974, Korea. E-mail: gogators@cau.ac.kr

보험개발원에서 제공하는 보도자료에 따르면 2014년 자동차 보험 손해율은 88.3%로 전년대비 1.5% 상승하여 최근 5년 중 가장 높은 손해율을 보이고 있고, 외제차가 전년대비 24.8% 증가하여 최근 5년간 평균 20%에 이르는 높은 증가율을 보이고 있다. 외제차가 증가하면서 수리비, 렌트비 등 외제차 손해율이 지속적으로 증가하고 있다. 외제차 수리비는 1조 1천억 원으로 전년대비 13.9% 증가하였고, 외제차 렌트비는 1,352억 원으로 전년대비 27.8% 증가하여 전체 수리비와 전체 렌트비 상승 요인으로 작용하고 있다. 2014년 자동차 보험 손해액은 1조 1,310억 원으로 누적 적자가 지속적으로 증가하고 있는 추세이다. 이와 같은 상황으로 손익 악화가 이어지면서 보험회사들이 줄줄이 자동차 보험료를 인상한다고 밝혔지만, 이는 교육지책 방안이기 때문에 다른 합리적인 해결방안이 필요한 실정이다.

경찰청 자료에 따르면 우리나라 교통사고 사망률과 보행자 사망률이 OECD 국가 중 가장 높게 나타났다. 특히 보행자 사망률은 37.6%로 OECD 국가 평균 16.5%보다 2배 이상 높게 나타났다. 또한 뉴질랜드(10.7%), 미국(14.1%), 독일(14.4%) 등과 비교하면 더 많은 격차를 보이고 있고, 20년째 OECD 국가 중 가장 높은 보행자 사망률로 불명예를 안고 있다. 최근 우리나라는 교통사고 건수(frequency)가 감소하고 있지만, 고령운전자에 의한 교통사고 건수와 심도(severity)는 오히려 증가하고 있다. 고령운전자에 의한 교통사고 건수가 빠르게 증가하는 원인은 고령화로 인한 고령운전자 수의 증가에 따른 것이고, 심도의 증가는 고령운전자의 교통사고 위험도가 20대를 제외한 다른 연령에 비해 높기 때문이다. 2014년 고령운전자 비율은 7.04%로 지난 2010년 4.9%보다 급격히 늘어났고, 지난 2010년 고령운전자에 의한 교통사고 비율은 5.6%에 불과했지만, 2014년 고령운전자에 의한 교통사고 비율은 9.1%까지 늘어났으며, 사망자 발생 교통사고 비율은 14.5%로 더 높게 나타났다. 고령운전자는 신체기능 저하로 운전 감각이나 반응속도가 둔해지고, 인지 기능과 운동신경이 떨어지면서 위험 요인으로 작용하고 있다. 예를 들어, 고령운전자의 정지 시력은 30대의 80% 수준이고, 원근조절 능력은 10대의 10% 수준이다. 이러한 이유로 인해 최근 일본에서는 운전 면허증을 반납하는 고령운전자에게 다양한 교통관련 혜택을 제공하고 있고, 우리나라 정부도 고령운전자를 대상으로 고령 운전자 안전 운전 대책으로 고령운전자를 위한 맞춤형 정책과 교육 프로그램을 통해 교통안전을 위협하는 요소를 제거하기 위해 노력하고 있다. 이와 같이 고령화 사회로 인한 연령그룹별 교통사고 발생 유형이 변화하면서 인구변화에 대한 관심이 필요하다. 즉, 치열한 시장경쟁과 금융당국의 요율 규제 그리고 요율 세분화로 인한 자동차 보험료 조정 필요하고, 특히 고령화 사회로 인한 인구변화로 고령운전자 수와 고령운전자에 의한 교통사고 건수가 증가하면서 연령그룹별 자동차 보험료 조정이 필요하다. 따라서 자동차 보험료를 결정하는 여러 가지 요율 요소 중 연령에 초점을 두고, 인구분포 변화에 대한 고려가 필요하다. 본 논문은 합리적인 자동차 보험 요율을 활용하지 못하는 상황에서 보험료 인상과 인하 문제가 아닌 인구분포를 통해 가격 최적화 방안을 제시하고자 한다.

본 논문은 총 4장으로 구성하고 있고, 구성은 다음과 같다. 제 2장은 선행연구로 분석에 사용하는 기본 개념에 대해 기술하고, 제 3장은 실증분석으로 실제 보험회사 자료를 사용하여 기존 연령 분포에 대한 결과와 분포 추정 방식으로 미래 연령 분포를 추정하여 얻은 결과를 비교 분석한다. 이를 통해 분포 추정에 대한 효과를 알아보고, 보험산업 전체에 미치는 영향을 알아본다. 제 4장은 결론으로 본 논문의 결과를 요약하여 설명한다.

2. 선행 연구

자동차 보험료 조정은 보험료 수준 조정과 사업비 조정 그리고 요율 요소 위험도 조정으로 구분할 수 있다. 보험료 수준 조정과 사업비 조정은 전체적인 보험료 인상과 인하를 결정하는 문제로 현재 규제사항이기 때문에 현실적으로 적용하기 어려운 부분이고, 보험료 조정 자체는 기본 보험료 인상과 인하 이외에는 불가능하여 다른 합리적인 방안이 필요하다. 따라서 본 논문은 보험료 인상과 인하 문제가 아닌 요

율 요소의 적정 분포를 도출하여 이로 인한 결손을 줄일 수 있는 방안을 제시하고자 한다. 즉, 수치 불균 등 원칙을 통해 현재 요율 조정이 되고 있고, 적용하는 방식에 있어서 추정된 미래 분포를 적용하여 적정 보험료 수준을 유지하도록 유도하고자 한다.

Kim (2013)은 적정 보험료를 빈도와 심도의 추세분석 방식을 통해 산출하는 방안을 제시했다. 이는 자동차 보험료 조정 중 보험료 수준 조정에 관한 내용으로 새로운 추세분석 방식을 제안하였고, 이를 통해 적정 보험료 수준을 보다 합리적으로 예측할 수 있는 방안을 제시했다. 빈도를 frequency, 심도를 severity, 보험료를 premium이라고 할 경우, 빈도와 심도가 서로 독립이라고 가정한다면 다음과 같이 표현할 수 있다.

$$E(\text{premium}) = E(\text{frequency}) \times E(\text{severity}).$$

빈도와 심도가 서로 독립일 경우, 추세 모형을 따로 추정할 수 있다. 교통사고는 추세 요인과 계절 요인이 같이 나타나므로 분기별 지시 함수를 이용하여 계절 요인을 따로 고려하고, 이를 적용한 추세 모형은 다음과 같다.

$$Y = X\beta + \gamma_1 I_1(t_1) + \gamma_2 I_2(t_2) + \gamma_3 I_3(t_3) + \epsilon,$$

$$I_i(t_i) = \begin{cases} 1, & \text{if } t_i = i, i = 1, 2, 3, \\ 0, & \text{otherwise,} \end{cases}$$

여기서 Y 는 추정 변수에 따라 빈도나 심도가 되고, X 는 추세 요인의 모수를 추정하기 위한 설명변수이다. 이와 관련된 연구로 Cook (1970)이 기본적인 추세분석에 대한 연구를 제안하였고, Kim과 Park (2013)은 시계열 모형을 통해 적정 보험료 추정에 적용하는 다양한 모형을 제시했다.

보험료를 산출하는 일반적인 방법으로 GLM 방법이 많이 활용되고 있는데, Kim과 Kim (2009)은 일반 선형모형(generalized linear model; GLM)을 중심으로 오차항에 대한 분포를 다양한 가정을 통해 접근하는 방법을 소개했다. Y 를 손해액이라고 가정하고 X_1, X_2, X_3 를 요율 요소라고 할 때, $g^{-1}(y)$ 를 연결 함수(link function)라고 하면 모형은 다음과 같이 정의할 수 있다.

$$g^{-1}(y) = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + \epsilon.$$

오차항의 분포를 감마분포라고 가정하고, 오차항을 최소화하는 a, b_1, b_2, b_3 를 추정하면 연결 함수가 로그 함수인 경우, 결과는 다음과 같다

$$Y = e^a e^{b_1 X_1} e^{b_2 X_2} e^{b_3 X_3},$$

여기서 e^a 는 손해액의 기본 보험료가 되고, 요율 요소별로 산출된 결과는 곱의 형태로 적용된다.

Choi와 Han (2008)은 손실 함수로 Tweedie's Compound 포아송 분포를 가정하고, 이중 일반선형모형(double generalized linear model; DGLM)을 통해 손해의 평균과 분산을 모델링하여 요율 추정의 정확도를 향상시키는 방안을 제시했다. 국내 선행연구에서는 보험료 수준 조정, 요율 요소 위험도 조정에 대하여 GLM 방법과 베이지안 프레임 평가 방법 등 연구가 많이 이루어졌다. 해외 선행연구에서는 최적화 방안을 통해 적정 보험료를 책정하는 방법을 많이 사용하고 있다. Taylor (1986)는 최적 요율을 산정하는 모형에서 적합한 해를 구하는 방법을 제시하였고, Emms와 Haberman (2005)은 Taylor (1986)가 제시한 방법이 고객의 입장에서 적합하지 않은 해를 생성한다는 것을 보이고, 선형 수요 함수를 가정하여 이를 개선하는 요율 산출 방법을 제시했다.

보험료(premium_{*i*})는 가입 고객수(exposure_{*i*}) × 사고 빈도(frequency_{*i*}) × 사고 심도(severity_{*i*})로 계산할 수 있다. 대부분의 다른 선행연구논문은 자료를 통해 사고 빈도와 사고 심도에 대한 정교한 추정 방

식에 관심을 가지고 있었다. 그러나 요율 세분화로 인한 가격경쟁과 온라인 보험료 비교 사이트 활성화로 보험 가입자가 접할 수 있는 정보가 많아지면서 보험사별 이동이 빈번해지고, 연령 분포의 변동이 심하게 발생하는 현재에는 과거에 비해 가입 고객수에 대한 추정의 필요성이 점차 높아지고 있다. 따라서 가입 고객수($exposure_i$) 또한 분포를 통해 추정된 가입 고객수($\widehat{exposure}_i$)로 대체하여 적용할 필요가 있다. 즉,

$$\begin{aligned} \text{premium}_i &= \text{exposure}_i \times \widehat{\text{frequency}}_i \times \widehat{\text{severity}}_i, \quad i = 1, \dots, k, \\ \text{gross_premium} &= \sum_{i=1}^k \text{premium}_i = \sum_{i=1}^k \text{exposure}_i \times \widehat{\text{frequency}}_i \times \widehat{\text{severity}}_i \end{aligned}$$

에서 $exposure_i$ 를 $\widehat{exposure}_i$ 로 대체하여 다음과 같이 적용할 필요가 있다.

$$\begin{aligned} \text{premium}_i &= \widehat{exposure}_i \times \widehat{\text{frequency}}_i \times \widehat{\text{severity}}_i, \quad i = 1, \dots, k, \\ \text{gross_premium} &= \sum_{i=1}^k \text{premium}_i = \sum_{i=1}^k \widehat{exposure}_i \times \widehat{\text{frequency}}_i \times \widehat{\text{severity}}_i. \end{aligned}$$

진술한 바와 같이 요율 요소 위험도 조정에 대해서는 다수의 선행논문에서 이미 GLM 방법을 통한 제안이 많이 이루어졌고, 이를 통한 적정 보험료 산출 방법이 제시되었다. 이러한 분석 방법에서는 위험도 추정 방식의 선진화와 모델링에 관한 제안을 하고 있다. 이에 반해, 본 논문에서 제시하는 방법은 추정 방식의 선진화가 아닌 고령화 사회로 인한 자동차 보험시장에서 발생하는 인구변화로 이동하는 연령의 미래 분포를 새롭게 추정하여 현재 분포가 아닌 미래 분포를 적용하고자 한다.

미래 분포를 추정하기 위해서 연령이동, 신규 유입으로 자사 신규와 타사 신규 그리고 고객이탈 등 세 가지 방법에 대한 고민이 필요하다. 자동차 보험의 특징 중 하나는 단기보험으로 매년 재가입하는 형태를 가지고 있다. 이에 따라 기존 보험회사에서 재가입을 하는 고객, 새롭게 들어오는 신규 고객 그리고 다른 보험회사로 이동하거나 다른 이유로 이탈하는 고객이 존재한다. 연령이동은 올해 보험 가입 고객수가 내년엔 얼마만큼 재가입하는지에 대한 논의가 필요하다. 이는 연령그룹별 갱신율을 통해 내년 보험 가입 고객수에 대한 갱신수를 예측할 수 있다. 신규 유입의 경우, 자사 신규와 타사 신규로 구분할 수 있다. 자사 신규는 보험 가입이 처음인 신규 고객이고, 타사 신규의 경우는 다른 보험회사에서 유입되는 고객이다. 이는 고객의 가입경력을 통해 알 수 있고, 자료를 통해 연령과 가입경력으로 미래 신규 유입 고객수를 예측하여 파악할 수 있다. 즉, 연령이동, 신규 유입, 고객이탈 방식으로 최적화 방안을 통해 이동하는 고객의 인구변화 분포를 추정하고자 한다.

3. 실증 분석

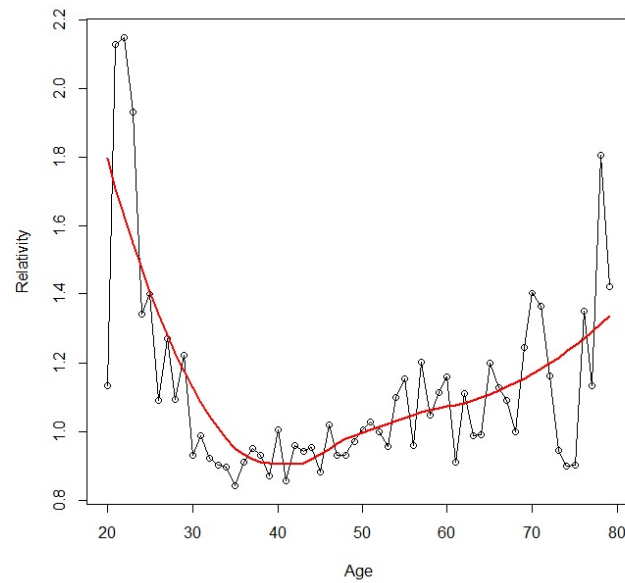
3.1. 분석 개요

실증 분석에서 사용한 자료는 국내 보험회사의 실제 자료를 사용하였고, 자료는 총 932,880개이다. 여러 가지 요율 요소 중 연령그룹 변수를 사용하였고, 연령그룹을 20세부터 79세까지 랜덤 추출하여 1세 단위 연령으로 분류하였다. 편의상 표는 연령그룹별로 정리하였다. 연령그룹별 가입 고객수와 손해액을 통해 계산한 연령그룹 순보험료는 1세 단위 연령으로 분류하여 연령그룹별 가입 고객수와 손해액에 따라 연령그룹별 순보험료를 산출하였다. 순보험료($pure_i$)는 손해액(L_i)을 가입 고객수(E_i)로 나누어 계산할 수 있다.

Table 3.1은 각 연령그룹의 가입 고객수와 손해액을 통해 각 연령그룹의 순보험료를 계산하여 정리한 표이다. 순보험료는 20대가 143,839원으로 가장 높게 나타났고, 70대가 140,946원으로 두 번째로 높게 나

Table 3.1. Age group data

Age	Exposure(E_i)	Loss(L_i)	Pure Premium($pure_i$)
20s	75,196	10,816,086,778	143,839
30s	245,672	25,659,242,953	104,445
40s	306,020	33,123,491,060	108,240
50s	214,402	25,607,253,204	119,436
60s	72,683	9,020,602,502	124,109
70s	18,907	2,664,868,850	140,946
Total	932,880	106,891,545,347	114,582

**Figure 3.1.** Relativity for age.

타났다. 이는 고령운전자가 다른 연령그룹에 비해 가입 고객수가 적음에도 불구하고 높은 순보험료를 보이고 있다. 이를 통해 고령운전자의 교통사고 위험도가 20대를 제외한 다른 연령그룹에 비해 높다고 볼 수 있다. 전체 순보험료 평균(\bar{X})은 전체 손해액($\sum L_i$)을 전체 가입 고객수($\sum E_i$)로 나누어 계산한 결과 114,582원으로 나타났다.

Table 3.1에서 계산한 연령그룹의 순보험료를 통해 각 연령그룹의 상대도(relativity)와 보험료를 산출할 수 있다. 각 연령그룹의 상대도(R_i)는 순보험료($pure_i$)를 전체 순보험료 평균(\bar{X})으로 나눈 것으로 정의되며, i 번째 연령그룹의 보험료($premium_i$)는 해당 연령그룹의 가입 고객수(E_i) \times 전체 순보험료 평균(\bar{X}) \times 해당 연령그룹의 상대도(R_i), 즉 $premium_i = E_i \times \bar{X} \times R_i$ 로 계산할 수 있다.

Figure 3.1은 20세부터 79세까지 1세 단위 연령별 상대도 그래프이다. 전반적으로 20대에서 높은 상대도를 나타내고, 점차 줄어들었다가 50대부터 다시 높아지는 형태를 보이고 있다. 또한 연령그룹별 상대도 격차가 심하게 발생하는 것을 볼 수 있고, 운전 경력이 부족한 20대를 제외하면 연령이 높아질수록 상대도도 같이 높아지는 것을 확인할 수 있다. 순보험료($pure_i$)를 전체 순보험료 평균(\bar{X})으로 나누어 계산한 상대도(R_i)는 연령 이동에 따른 상대도 등락이 많이 발생하여 현실적으로 기존 상대도를 적

Table 3.2. Relativity & estimated relativity for age

Age	Relativity(R_i)	Estimated relativity(\hat{R}_i)
20s	1.475766	1.459611
30s	0.914315	0.986346
40s	0.944975	0.935527
50s	1.055572	1.034168
60s	1.082144	1.106965
70s	1.238707	1.245125

Table 3.3. Exposure, loss, pure premium & premium for age

Age	Exposure(E_i)	Loss(L_i)	Pure premium(pure $_i$)	Premium(premium $_i$)
20s	75,196	10,816,086,778	143,839	11,295,035,083
30s	245,672	25,659,242,953	104,445	27,482,605,953
40s	306,020	33,123,491,060	108,240	32,810,488,915
50s	214,402	25,607,253,204	119,436	25,234,472,021
60s	72,683	9,020,602,502	124,109	9,154,810,648
70s	18,907	2,664,868,850	140,946	2,647,168,948
Total	932,880	106,891,545,347	(mean)114,582	108,624,581,568

용하기 어렵다. 따라서 Figure 3.1의 적합 곡선은 연령그룹별 상대도의 등락을 줄이는 2차형 평활 곡선(smoothing line)으로 상대도를 추정하여 적용하였다.

Table 3.2는 연령그룹별 상대도(relativity)로서 Figure 3.1을 정리한 표이다. 기존 상대도와 새롭게 추정된 상대도에서 20대가 가장 높은 상대도로 나타났고, 70대가 두 번째로 높은 상대도로 나타났다. 적합 곡선을 통해 추정된 새로운 상대도는 연령그룹 상대도 격차가 줄어든 것을 확인할 수 있다. 이는 위험도에 대한 안정성을 제공한다.

본 장에서는 먼저 현재 가입 고객수를 이용하여 기존 보험료와 예상 보험료를 구하고, 이에 따른 수치균등계수(off-balance)를 계산한다. 그리고 분포 추정 방식으로 연령이동, 신규 유입, 고객이탈 등 세 가지 경우에서 각각을 추정하는 방법을 제시하고 이를 이용하여 새로운 미래 가입 고객수 분포를 도출한다. 도출된 미래 가입 고객수를 통해 기존 보험료와 예상 보험료를 산출하여 수치균등계수를 계산하고, 현재 가입 고객수 분포에서의 수치균등계수와 미래 가입 고객수 분포의 수치균등계수를 비교하여 분포 추정에 대한 효과를 비교해보고자 한다.

3.2. 현재 분포 적용

다음 Table 3.3은 Table 3.1의 현재 분포를 적용한 연령그룹 가입 고객수와 손해액, 순보험료를 근거로 보험료를 계산하여 추가 정리한 표이다. Table 3.3에서 보험료(premium $_i$)는 가입 고객수(E_i) × 전체 순보험료 평균(\bar{X}) × 상대도(\hat{R}_i)로 계산할 수 있고, 산출한 보험료 중 가입 고객수가 많은 40대가 32,810,488,915원으로 가장 높게 나타났다. 이때, 예상 보험료(\sum premium $_i$)는 108,624,581,568원이며, 기존 보험료($\sum L_i$)는 106,891,545,347원으로 계산된다. 또한 두 보험료의 차이는 1,733,036,221원으로 나타났다.

요율 조정에서 실제 사용하는 방식으로 기존 보험료와 예상 보험료를 서로 같게 하기 위한 수치균등계수는 기존 보험료($\sum L_i$)를 예상 보험료(\sum premium $_i$)로 나누어 계산할 수 있다. 현재 분포를 적용했을 때, 수치균등계수는 약 0.984로 계산된다. 현재 가입 고객수를 이용한 연령 분포를 준용하여 수치균등

Table 3.4. Exposure, renewal rate & renewal number for age

Age	Exposure(E_i)	Renewal rate(renewal $_i$)	Renewal number(\hat{f}_i)
20s	75,196	0.6135	39,334
30s	245,672	0.7366	170,507
40s	306,020	0.7866	238,321
50s	214,402	0.8112	187,074
60s	72,683	0.8362	69,089
70s	18,907	0.8220	18,111
Total	932,880		722,436

계수를 산출했지만, 인구변화로 인한 연령 분포가 이동한다면 보험료에 대한 수치균등계수가 변화하게 된다. 이에 따라 현재 분포와 미래 분포에서 계산한 수치균등계수의 차이를 알아보고, 그 효과를 살펴보고자 한다.

3.3. 분포 추정 방법

분포 추정 방법을 살펴보면, 먼저 연령이동 인구 추정의 경우, 해당 보험회사의 실제 연령그룹별 갱신율(renewal $_i$)을 반영하여 올해 가입 고객수와 갱신율을 통해 내년 갱신수를 추정하였고, 이를 다음과 같이 정의한다.

$$\hat{f}_{i+1} = E_i \times \text{renewal}_i, \quad i = 1, 2, \dots, 54,$$

$$\hat{f}_i = \frac{1}{5} \left\{ \sum_{i=55}^{59} E_i \times 0.822 \right\}, \quad i = 55, \dots, 59.$$

1세씩 증가하는 내년 연령그룹별 재가입 고객수를 추정할 때, 내년 74세까지는 올해 가입 고객수와 갱신율을 적용하여 계산할 수 있다. 내년 75세부터 79세까지는 올해 74세부터 79세까지의 총 가입 고객수에 동일한 갱신율인 0.822를 적용하였고, 이를 75세부터 79세까지 균등하게 분류하였다. 여기서 i 는 관측치의 순서이므로, 실제 연령은 $i + 20$ 이 된다. 즉, 예를 들어, $i = 54$ 는 74세를 의미한다.

Table 3.4는 해당 보험회사의 실제 갱신율을 적용한 연령그룹 갱신수를 추정하여 정리한 표이다. 연령이 높아질수록 갱신율이 높아지는 것을 볼 수 있고, 40대가 238,321명으로 가장 많은 갱신수로 나타났다. 갱신율은 높지만 가입 고객수가 적은 70대가 가장 적은 갱신수를 나타냈다. 또한 총 갱신수는 722,436명으로 추정되었다.

다음으로 신규 유입 인구 추정의 경우, 올해 보험 가입 고객수와 내년 보험 가입 고객수가 동일하다는 가정과 내년에 신규 고객이 유입되면 그 고객 수만큼 기존의 고객이 이탈되었다는 가정을 하고, 총 신규 유입 고객수($\sum \hat{g}_i$)를 다음과 같이 추정할 수 있다.

$$\sum E_i - \sum \hat{f}_i = \sum \hat{g}_i = \sum \hat{g}_{i1} + \sum \hat{g}_{i2}, \quad i = 1, 2, \dots, 60.$$

올해 보험 가입 고객수와 내년 보험 가입 고객수가 동일하다는 가정을 통해 총 가입 고객수($\sum E_i$)는 총 932,880명으로 연령이동 인구 추정에서 구한 갱신수($\sum \hat{f}_i$) 722,436명을 뺀 나머지 총 신규 유입 고객수($\sum \hat{g}_i$)를 210,444명으로 추정할 수 있다. 또한 총 신규 유입 고객수($\sum \hat{g}_i$)는 자사 신규 고객(new customer)수와 타사 신규 고객(converted customer)수의 추정치인 $\sum \hat{g}_{i1}$ 와 $\sum \hat{g}_{i2}$ 의 합으로 표현할 수 있다. 따라서 연령과 가입경력 변수를 이용하여 총 자사 신규 비율(new customer rate)과 총 타사 신규 비율(conversion rate)을 계산할 수 있다. 즉, 총 자사 신규 비율은 총 자사 신규 고객수($\sum \hat{g}_{i1}$)를 총 가

Table 3.5. New customer, new customer rate, converted customer & conversion rate for age

Age	New customer(g_{i1})	New customer rate($rate_{i1}$)	Converted customer(g_{i2})	Conversion rate($rate_{i2}$)
20s	21,522	0.452885	53,674	0.060624
30s	14,718	0.309709	230,954	0.260859
40s	6,467	0.136084	299,553	0.338341
50s	3,062	0.064433	211,340	0.238706
60s	1,181	0.024852	71,502	0.080761
70s	572	0.012037	18,335	0.020709
Total	47,522	1	885,358	1

Table 3.6. Renewal, estimated new customer & estimated converted customer for age

Age	Renewal(\hat{f}_i)	Estimated new customer(\hat{g}_{i1})	Estimated converted customer(\hat{g}_{i2})
20s	39,334	4,854	12,108
30s	170,507	3,321	52,098
40s	238,321	1,459	67,576
50s	187,074	692	47,674
60s	69,089	266	16,131
70s	18,111	128	4,137
Total	722,436	10,720	199,724

입 고객수($\sum E_i$)로 나누어 계산할 수 있고, 총 타사 신규 비율은 총 타사 신규 고객수($\sum g_{i2}$)를 총 가입 고객수($\sum E_i$)로 나누어 계산할 수 있다. 그 결과, 총 자사 신규 비율은 약 0.051이고, 총 타사 신규 비율은 약 0.949로 계산된다. 즉, 새롭게 회사로 유입되는 고객은 순수한 신규 가입고객 5.1%와 타사로 부터 당사로 전환되어 새롭게 가입되는 고객 94.9%로 구성되어 있다는 것이다. 이를 통해 총 자사 신규 고객수와 총 타사 신규 고객수의 추정치 $\sum \hat{g}_{i1}$ 와 $\sum \hat{g}_{i2}$ 를 구할 수 있는데, 총 자사 신규 고객수의 추정치($\sum \hat{g}_{i1}$)는 총 신규 유입 고객수($\sum g_i$) \times 총 자사 신규 비율(0.051)로 구할 수 있고, 이는 10,720명이다. 총 타사 신규 고객수의 추정치($\sum \hat{g}_{i2}$)는 총 신규 유입 고객수($\sum g_i$) \times 총 타사비율(0.949)로 계산하면 199,724명으로 계산된다.

연령그룹별 신규 유입 고객수를 추정하기 위해 각 연령그룹별 신규 비율을 계산할 수 있다. 자사 신규 비율($rate_{i1}$)은 자사 신규 고객수(g_{i1})을 총 자사 신규 고객수($\sum g_{i1}$)으로 나누어 구할 수 있고, 타사 신규 비율($rate_{i2}$)은 타사 신규 고객수(g_{i2})를 총 타사 신규 고객수($\sum g_{i2}$)로 나누어 구할 수 있다. 각 연령그룹별 자사 신규 비율과 타사 신규 비율은 Table 3.5를 통해 확인할 수 있다. Table 3.5를 살펴 보면, 자사 신규 비율은 20대에서 가장 높게 나타나며, 연령이 많아질수록 자사 신규 비율이 낮아지는 것을 확인할 수 있다. 또한 타사 신규 비율은 40대가 가장 높게 나타난다.

Table 3.6은 연령그룹별 자사 신규 고객수와 타사 신규 고객수를 추정하여 정리한 표이다. 자사비율 0.051과 타사비율 0.949를 총 신규 유입 고객수($\sum \hat{g}_i$)에 적용하면 총 자사 신규 고객수($\sum \hat{g}_{i1}$)는 10,720명이고, 총 타사 신규 고객수($\sum \hat{g}_{i2}$)는 199,724명으로 나타난다. 또한 Table 3.5를 통해 연령그룹 자사 신규 고객수와 타사 신규 고객수를 구할 수 있다. 연령그룹별 자사 신규 고객수(\hat{g}_{i1})은 총 자사 신규 고객수($\sum \hat{g}_{i1}$) \times 자사 신규 비율($rate_{i1}$)로 계산할 수 있고, 타사 신규 고객수(\hat{g}_{i2})는 총 타사 신규 고객수($\sum \hat{g}_{i2}$) \times 타사 신규 비율($rate_{i2}$)로 계산할 수 있다. 추정한 자사 신규 고객수(\hat{g}_{i1})는 20대가 4,854명으로 가장 많으며 타사 신규 고객수(\hat{g}_{i2})는 40대가 67,576명으로 가장 많이 나타났다. 또한 연령이동 인구 추정에서 구한 갱신수(\hat{f}_i), 신규 유입 인구 추정을 통해서 구한 자사 신규 고객수(\hat{g}_{i1})와 타

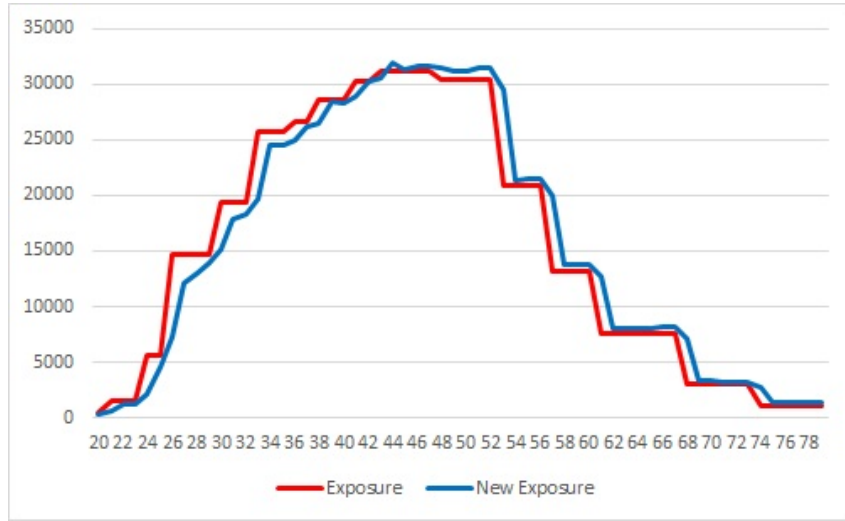


Figure 3.2. Exposure & new exposure for age.

Table 3.7. Estimates of exposure, loss, pure premium & premium for age

Age	Exposure(\hat{E}_i)	Loss(L_i)	Pure premium(\widehat{pure}_i)	Premium($\widehat{premium}_i$)
20s	56,296	10,816,086,778	192,129	8,312,126,498
30s	225,926	25,659,242,953	113,574	25,184,453,680
40s	307,356	33,123,491,060	107,769	32,973,071,178
50s	235,440	25,607,253,204	108,763	27,728,786,891
60s	85,486	9,020,602,502	105,521	10,774,896,251
70s	22,376	2,664,868,850	119,095	3,141,709,724
Total	932,880	106,891,545,347	114,582	108,115,044,222

사 신규 고객수(\hat{g}_{i2})를 합하여 새로운 연령그룹별 가입 고객수(\hat{E}_i)의 미래 분포 추정이 가능하다. 즉, $\hat{E}_i = \hat{f}_i + \hat{g}_{i1} + \hat{g}_{i2}$ 으로 추정된다. 그 결과, 새롭게 추정한 가입 고객수(\hat{E}_i)는 40대가 307,356명으로 가장 많았고, 50대가 235,440명으로 두 번째로 많은 가입 고객수를 차지했다. 이를 기존 연령 분포와 비교해보면 50대 이상의 가입 고객수가 증가한 것을 알 수 있다. 추정한 미래 분포는 기존의 연령 분포와 차이가 있고, 다음 Figure 3.2를 통해 눈으로 쉽게 확인할 수 있다.

Figure 3.2를 보면 연령그룹별로 기존 가입 고객수(E_i)와 연령이동, 신규 유입, 고객이탈 등 세 가지 방법의 미래 추정 방식으로 추정한 인구변화에 따른 연령을 새롭게 정의한 연령그룹별 가입 고객수(\hat{E}_i)의 분포를 비교할 수 있으며, 이를 통해 연령그룹별 가입 고객수가 이동하는 것을 확인할 수 있다. 최근 사회적으로 큰 이슈가 되는 것처럼 우리나라가 빠르게 고령화 사회로 변화하면서 고령운전자의 수도 증가함을 볼 수 있고, 이에 따른 고령운전자의 교통사고 위험도에 대한 연구뿐만 아니라 보험회사들의 손익 악화를 연령분포 추정을 통해 줄일 수 있는 해결 방안 자료로 제시할 수 있을 것이라 판단된다.

3.4. 미래 분포 적용

다음 Table 3.7은 미래 분포를 적용한 연령그룹별 가입 고객수와 손해액, 순보험료 그리고 보험료를 추정하여 정리한 표이다. Table 3.7에서 보험료($\widehat{premium}_i$)는 가입자 수(\hat{E}_i) × 전체 순보험료 평

균(\bar{X}) \times 상대도(\hat{R}_i)로 계산할 수 있고, 가입자 수가 많은 40대가 32,973,071,178원으로 가장 높게 나타났다으며 50대가 27,728,786,891원으로 두 번째로 높게 나타났다. 이때, 예상 보험료($\sum \widehat{\text{premium}}_i$)는 108,115,044,222원이었으며, 기존 보험료($\sum L_i$)는 106,891,545,347원이었다. 기존 보험료($\sum L_i$)를 예상 보험료($\sum \widehat{\text{premium}}_i$)로 나누면 수치균등계수를 계산할 수 있다. 이 데이터에서 현재 분포와 미래 분포를 적용한 수치균등계수는 각각 0.984와 0.989이었고, 두 값의 차이는 0.00463%p로 나타났다.

기존 가입자 수를 통해 현재 연령 분포를 적용하여 실제 받을 총 보험료를 계산하면 105,186,158,619원이고, 미래 연령 분포를 적용한 실제 총 보험료는 105,681,892,367원으로 계산되었다. 따라서 두 보험료의 차이는 495,733,748원으로 나타나 비교적 작은 차이를 보이는 것 같지만, 보험산업 전체 규모를 고려하면 이는 무시할 수 없는 차이이다. 즉, 국내 자동차 보험산업 규모는 약 1조 1300억 원으로, 기존 연령 분포를 적용한 수치균등계수를 통해 계산한 총 보험료는 1,111,920,000,000원으로 산출되고, 미래 분포를 적용한 수치균등계수를 통해 계산되는 총 보험료는 1,117,570,000,000원이므로 두 보험료의 차이는 5,650,000,000원으로 계산된다.

4. 결론

본 논문에서는 여러 가지 요율 요소 중 연령에 대한 분포를 추정하여 현재 분포를 적용한 보험료와 연령이동, 신규 유입, 고객이탈 등 세 가지 방법으로 최적화(optimization)를 통한 분포 추정 방법으로 미래 분포를 적용한 보험료를 비교 분석하였다. 현재 분포를 적용한 수치균등계수와 미래 분포를 적용한 수치균등계수의 차이는 분포를 추정하지 않아 발생하는 결손으로 이는 실제 자료를 사용하여 분석한 결과 현재 분포를 적용한 총 보험료는 105,186,158,619원으로 나타나고, 미래 연령 분포를 적용한 총 보험료는 105,681,892,367원으로 계산된다. 두 보험료의 수치균등계수 차이는 약 0.00463%P 발생하고 액수로는 495,733,748원이다. 이 차이는 현재 보험회사가 그 금액만큼 보험료를 덜 받고 있으며 손해가 발생하여 결손이 생긴다는 의미이다. 따라서 새로운 미래 분포를 추정하여 적용하면 보험사에서 495,733,748원의 결손을 예방할 수 있다. 현재 국내 자동차 보험산업 전체 규모가 약 1조 1300억 원으로 현재 분포를 적용한 총 보험료는 1,111,920,000,000원으로 나타나고, 미래 연령 분포를 적용한 총 보험료는 1,117,570,000,000원으로 계산된다. 여기에 기존 분포를 적용한 수치균등계수와 미래 분포를 적용한 수치균등계수의 총 보험료 차이는 5,650,000,000원이다. 다음과 같이 발생하는 손실을 분포 추정을 통해 보험산업 전체에서 5,650,000,000원만큼의 손해를 줄일 수 있다고 볼 수 있다.

또한 반대의 경우도 나타날 수 있다. 이는 서론에서 언급했듯이 보험료를 인상하는 목적이 아니기 때문에 연령그룹별 가입 고객수를 추정하여 미래에 받을 수 있는 예측 보험료가 어떤 것이 더 적정인지 확인할 수 있으며 보험회사의 초과 이득 또한 발생하지 않도록 한다. 또한 이는 보험료 수준 조정과 거의 동일한 효과를 얻을 수 있다. 요율 요소 위험도 조정에서는 수준 조정을 하지 못하지만, 이러한 분포 추정을 통해서 일정 수준의 효과를 볼 수 있다. 그리고 이를 통해 보험회사가 보험료를 인상하는데 있어서 인상 시점을 연기할 수 있는 요인으로 고려할 수 있다. 분포를 추정하지 않아서 계속해서 발생했던 결손이 고령화 사회로 인한 인구 분포가 점차 이동하면서 교통사고 위험도가 높은 고령운전자들에게 보험료를 더 받아야 했지만, 그렇지 못하고 있기 때문에 계속해서 적자가 발생하고 누적되어 보험회사의 손익 악화로 이어지고 있는 실정이다.

다른 선행논문에서 제시한 것처럼 여러 가지 추정 방식으로 미래 위험도를 판단하는 것도 물론 중요하지만, 위험도의 판단은 현재 자료를 통해서 분석할 수밖에 없고, 미래를 판단하는 수단으로 결국 분포를 추정하여 사용할 수 밖에 없다. 이러한 분포 추정을 통해 적정 보험료를 제시할 수 있고 수치균등계수를 개선하면서 결손되는 부분을 보완할 수 있을 것으로 판단된다.

References

- Choi, W. and Han, S. (2008). Estimating the rate of motor insurance premium by double generalized linear model, *Journal of Korea Insurance Research Institute*, **19**, 37–57.
- Cook, C. (1970). Trend and loss development factors, *Casualty Actuarial Society*, **52**, 1–14.
- Emms, P. and Haberman, S. (2005). Pricing general insurance using optimal control theory, *ASTIN BULLETIN*, **35**, 427–454.
- Kim, M. (2013). A study on trend analysis of severity and frequency for predicting the proper premium - Focusing on auto insurance, *Journal of Korean Academy of Actuarial Science*, **5**, 100–111.
- Kim, M. and Kim, Y. (2009). Various modeling approaches in auto insurance pricing, *Journal of the Korean Data & Information Science Society*, **20**, 77–87.
- Kim, M. and Kim, Y. (2013). A study of Bayesian inference on auto insurance credibility application, *Journal of the Korean Data & Information Science Society*, **24**, 689–699.
- Kim, Y. and Park, W. (2013). Estimating automobile insurance premiums based on time series regression, *Korean Journal of Applied Statistics*, **26**, 237–252.
- Taylor, G. C. (1986). Underwriting strategy in a competitive insurance environment, *Insurance: Mathematics and Economics*, **5**, 59–77.

요율 요소 분포 추정을 통한 가격 최적화 방안 연구

김영화^{a,1} · 전철희^b

^a중앙대학교 응용통계학과, ^b중앙대학교 대학원 통계학과

(2016년 6월 29일 접수, 2016년 8월 2일 수정, 2016년 8월 5일 채택)

요약

자동차 보험은 치열한 시장경쟁과 금융당국의 요율 규제로 가격 조정에 합리적인 적용 방식이 요구되는 대표적인 보험상품이다. 특히 고령화 사회로 인한 인구변화와 가격경쟁으로 인한 요율 세분화는 보험료 결정에 주요한 요인으로 작용하게 된다. 따라서 본 논문에서는 여러 가지 요율 요소 중 연령에 대한 분포 추정을 통해 가격 최적화 방안을 제시하고자 한다. 미래 분포를 추정하는 방식으로 고객의 연령이동, 신규유입, 고객이탈 3가지 요인에 대한 최적화 방안을 적용함으로써 합리적인 연령 분포를 추정하여 가격 조정 과정에 반영하고자 한다. 기존 연령 분포를 적용한 보험료와 미래 연령 분포를 추정하여 적용한 보험료를 수지 불균등 원칙에 적용하여 위험도 추정에 대한 적정성을 비교 분석하는 것이 목적이다. 보험료 인상 또는 인하 문제와는 별개로 인구변화에 대한 적정 분포를 추정함으로써 가입자의 분포 변화로 인한 결손 또는 초과 이득이 발생하지 않는 적정 보험료 추정 방식을 제안하고자 한다.

주요용어: 미래 분포, 수지균등계수, 자동차보험

이 논문은 2014년도 중앙대학교 연구장학기금 지원에 의한 것임.

¹교신저자: (06974) 서울특별시 동작구 흑석로 84, 중앙대학교 경영경제대학 응용통계학과.

E-mail: gogators@cau.ac.kr