보험금융연구 제33권 제3호 (2022. 8) pp. 41-76
DOI : https://doi.org/10.23842/jif.2022.33.3.002

공유된 임의효과를 고려한 다중 위험 보험의 요율에 관한 연구

A Study on the Ratemaking of Multi-peril Insurance via Shared Random Effects

전 옥 현*·정 힘 찬**·최 양 호***·심 현 우****
Okhyeon Jeon·Himchan Jeong·Yang Ho Choi·Hyunoo Shim

하나의 보험 계약이 여러 개의 담보로 구성되어 서로 다른 위험에 의한 사고를 보상하는 보험을 다중 위험 보험이라 한다. 다중 위험 보험은 각 위험에 의한 청구 건수 사이에 의존성이 존재할 수 있다. 의존성이 확인된 경우 보험요율 산정에 있어 위험 간의 상관관계를 반영하는 것이 합리적이다. 본 연구에서는 다중 위험 간의 의존성 존재여부를 확인하고 이를 사후요율결정(posterior ratemaking)에 반영하기 위하여 공유된 임의효과 모델을 적용하였다. 이 모델은 신뢰도 요율을 닫힌 공식(closed formula) 형태로 제공할 수 있어 실무적 적용과 해석이 용이하다는 장점이 있다.

자동차 보험은 다양한 담보로 구성되어 다중 위험 보험의 대표적인 예로 볼 수 있다. 본 연구에서는 제안 모델을 요율산출에 적용하기 위하여 자동차 보험의 대인배상과 대물배상 담보의 실제 청구 데이터를 사용하였다. 해당 데이터를 이용하여 요율을 산정하였고 그 성능을 비교하기 위하여 표본외 검증을 수행한 결과 포아송 모델 등 비교 모델보다 성능이 우수한 것으로 나타났다.

국문 색인어: 다중 위험 보험, 공유된 임의효과, 과대산포, 청구빈도분석, 할인할증제도
한국연구재단 분류 연구분야 코드: B051605, B051608, C030805

* 한양대학교 금융보험학과 석사과정, 메리츠화재(j.okhy@meritz.co.kr), 제1저자
** Simon Fraser University 통계·계리학과 조교수(himchanj@sfu.ca), 공동저자
*** 한양대학교 보험계리학과 부교수(ychoi@hanyang.ac.kr), 공동저자
**** 한양대학교 보험계리학과 부교수(hyunooshim@hanyang.ac.kr), 교신저자
논문 투고일: 2022. 1. 12, 논문 최종 수정일: 2022. 5. 25, 논문 게재 확정일: 2022. 8. 18

Ⅰ. 서론

우리나라는 물론 전 세계적으로 보아도 자동차 수는 지속적으로 증가하고 있다. 그에 따라 자동차 보험은 자연스럽게 대표적이고 보편적인 보험이 되었다. 또한 대부분의 국가에서 자동차 보험은 의무적으로 가입해야 하는 보험 상품이다. 그러므로 자동차 보험을 취급하는 보험회사 입장에서는 자동차 사고에 영향을 미치는 요인을 분석하는 것이 매우 중요하다. 자동차 보험의 계약자가 처음 보험을 가입할 시점에 적용되는 사전적인 보험요율을 결정하는 변수에는 일반적으로 연령, 성별, 직업, 거주 지역, 자동차의 유형과 용도 등이 포함된다. 그 이후에는 운전자의 과거 보험사고 이력 정보를 사후적인 보험요율 변수로 하여 경험요율을 결정한다.

할인할증제도(bonus-malus system)에서 보험료 할증이란 보험사고에 과실 책임이 있는 계약자에게 보험료를 더 높게 받는 것을 말한다. 또한 할인이란 과실이 있는 보험사고 이력이 없는 계약자에게 보험료를 더 적게 받는 혜택을 제공하는 것을 말한다. 예를 들어, 심현우·임형기·최양호(2021)는 실손 의료보험에서 과거 청구 건수에 따른 할인할증제도를 제안하였고, 실제로 2021년 한국의 4세대 실손 의료보험에서 할인할증제도를 제도적으로 도입하였다. 한국을 제외한 모든 국가의 할인할증제도에서는 보고된 보험사고의 횟수 즉, 사고 빈도에 따라 보험료를 할인 또는 할증하지만 보험사고의 크기 즉, 사고 심도는 고려하지 않고 있다. 반면 한국에서는 1989년 할인할증제도를 도입한 이후 사고 빈도뿐만 아니라 사고 심도까지 고려하여 경험요율을 결정하였다.

그러나 2018년 이후부터는 국제적인 할인할증제도를 따라 요율결정에 사고 심도 정보는 제외하고 사고 빈도만을 반영하고 있다. 따라서 자동차 보험 요율을 결정하는데 있어 사고 빈도의 영향이 매우 크고 중요하기 때문에 빈도에 대한 면밀한 분석이 뒷받침 되어야 한다.

자동차 보험의 중요한 특성 중 하나는 다중 위험(multi-peril) 보험이라는 것이다. 우리나라 자동차 보험이 보장하는 담보는 대인배상, 대물배상, 자기신체사고, 자기차량손해로 구분된다. 이러한 여러 담보에서 보장하는 손해가 하나의 보험사고에서 동시에 발생할 수 있다는 특징이 있다. 따라서 각 담보별 손해액은 독립적이라고 보기 어려우므로 담보 간

상관관계를 고려할 필요가 있다.

다양한 담보들 간에 공변량(covariate)이 존재할 때, 담보 유형별로 공변량의 영향이 다르면 각 청구 유형에 대해 서로 다른 회귀계수를 적용해야 한다. 운전 습관과 같이 관찰되지 않은 이질성은 다양한 청구 유형들과 각각 서로 다른 상관관계를 가질 것이다. 예를 들어, 과실 배상책임 담보는 일반적으로 운전자가 어느 정도 통제할 수 있으므로 해당 청구의 발생은 운전 습관과 밀접한 관련이 있다. 그러나 폭우로 인한 침수 피해와 같이 자연재해에 의한 자기차량 손해의 경우 운전자가 거의 통제할 수 없으므로 운전 습관과의 연관성이 매우 낮다. 결국 보험회사는 공변량 정보와 각 보험 계약자의 과거 청구 내역을 활용하여 계약자의 관찰되지 않은 이질성과 보험료를 산출하는 모델을 개발해야 한다.

이에 본 연구에서는 종단면 데이터를 이용하여 관찰되지 않은 이질성을 반영한 공유된 임의효과(shared random effects) 모델을 제안한다. 이 모델은 다중 위험 보험의 특성을 요율에 보다 잘 반영하기 위한 목적으로 각 계약자의 청구이력을 통해 적절한 할인할증요인을 산출하는 모델이다. 이 모델은 기존의 할인할증제도 관점에서 해석이 용이하고 최종적인 요율 산출식도 닫힌 공식(closed formula)으로 도출되어 실무적으로 적용하기에도 편리할 것이다.

우리나라의 자동차 보험은 의무보험으로서 소비자물가지수 품목에 포함될 정도로 보편적이고 중요한 보험 상품이기 때문에 적정 보험요율 산정에 관한 여러 연구가 있었다. 예를 들어 할인할증요율 산출에 대하여 김진한·민재형(1994)은 다중상태모형을, 정중영·김태완(2004)은 기대가치보험료원칙을 이용할 것을 제안하였다. 조재린·이기형·강중철(2012)은 신뢰도 이론을 적용한 자동차 보험요율 산출 결과를 분석하였다. 선행연구에서는 대부분 자동차 보험의 담보간의 청구 건수를 독립으로 가정한 반면 본 연구에서는 담보 간 의존성을 반영하였다는 특징이 있고 보험료 산출식이 기본보험료와 할인할증 요인으로 명확히 구분되어 실무적 적용과 해석이 용이하다는 장점이 있다.

본 연구의 구성은 다음과 같다. 제2장에서는 제안 모델과 관련된 신뢰도이론과 다중 위험 간의 의존성에 대한 선행연구를 살펴본다. 제3장에서는 다중 위험 간의 의존성과 과대산포를 고려한 빈도모형을 적용하여 최종적인 요율산출 모델을 도출한다. 제4장에서는 이 모델을 실제 청구데이터에 적용하여 모델에 적용되는 계수를 결정하여 보험료를 산출한

다. 그리고 그 결과를 다른 모델과 비교하여 모델의 성능을 확인한다. 마지막으로 제5장에서는 본 연구의 결과를 요약한다.

II. 선행 연구

1. 신뢰도 이론

신뢰도 이론은 미래에 발생할 수 있는 보험사고나 그 사고로 인한 손해액 추정 및 보험료 결정 시 과거 청구 이력을 통해 관찰되지 않은 이질성을 도출하는 이론이다. 즉, 과거와 당해 연도의 데이터에 근거하여 그 다음 해의 보험료 산출 방식을 수정 보완하는 접근방법으로, 현재까지 수집된 데이터를 얼마나 믿을 수 있는가에 대한 신뢰도(credibility) 개념이 보험료 책정에 중요한 요소로 사용되고 있다. 신뢰도의 개념은 CAS(미국 손해보험협회)에서 보험료에 현재까지 수집된 데이터를 어떻게 적용할 것인가 하는 문제의 해결책으로 제시되어 보험회사의 가격정책에 적용되기 시작하였다.

Whitney(1918)는 새로운 보험료를 얻기 위하여 현재 사용되고 있는 요율과 과거 데이터를 어떻게 조합하여 사용할 것인가를 연구하여 다음과 같은 기본 개념을 정립하였다.

$$\text{사후적 보험료} = Z \times \text{청구 이력} + (1 - Z) \times \text{사전적 보험료}. \tag{1}$$

여기서 Z를 신뢰도 요인(credibility factor, $0 \le Z \le 1$)이라 한다. 위 식은 청구이력으로부터 얻은 정보 즉, 평균 손해액과 과거 추정치인 사전평균(prior mean)의 가중평균 형태로 볼 수 있다. 이러한 관점에서 베이지안 패러다임과 동일하다고할 수 있다.

이후 신뢰도 이론은 보험 계리 분야에서 주로 연구되었다. Bailey(1950)는 모델분포가 이항분포이고 사전분포가 베타분포인 경우에, 사후적 요율이 이러한 신뢰도의 가중평균이 되는 것을 보였다. Bühlmann(1967)은 서로 독립인 두 추정량의 제곱오차(squared error)를 사용하여 부분신뢰도를 구하는 방법을 제안하였다. Mayerson(1964)과 Jewell(1974)은 베이지안 관점에서 신뢰도 이론을 분석하여 관찰되지 않는 이질성을 반영

하는 방법을 제시하였다. Norberg(1986)는 관찰된 이질성과 관찰되지 않은 이질성은 모두 요율 산출에 고려되어야 한다고 제안하였다. Frees et al.(1999)은 이를 발전시켜 선형 혼합 모델(linear mixed model)을 기반으로 신뢰도 이론과 회귀분석을 통합한 일반적인 체계를 제시하였다. 선형 혼합 모델에서 반응 변수는 관찰된 공변량과 이와 관련된 회귀계수(고정효과)는 물론, 관찰되지 않은 특성(임의효과)의 영향을 받는다. 고정효과와 임의효과를 모두 사용하기 때문에 신뢰도 보험료가 할인할증제도와 밀접한 관련이 있음을 알 수 있다. Boucher and Denuit(2008) 및 Boucher et al.(2009)은 패널 데이터에 zero-inflated Poisson 분포를 사용하여 신뢰도 보험료를 도출하는 방법을 모색하였다. Jeong and Valdez(2020b)는 과거 빈도의 결합분포가 다변량음이항분포(multivariate negative binomial; MVNB)일 때, 미래 건수에 대한 신뢰도 보험료(예측분포의 기댓값)는 다음 공식으로 나타낼 수 있음을 보였다.

$$\text{사후적 보험료} = \frac{r + \text{실제 청구 경험}}{r + \text{기대 청구 경험}} \times \text{사전적 보험료}. \tag{2}$$

여기서 r은 할인할증제도의 평활 계수를 나타내고 r값이 클수록 과거 청구 경험에 대한 가중치는 작아진다.

2. 다중 위험 보험의 의존성

그러나 실무적으로 요율결정에 적용하기 위해서는 여러 담보의 청구 간에 존재할 수 있는 의존성을 고려해야 한다. Frees et al.(2010)은 개별 위험 모델(individual risk model)하에서 다중 위험 보험에 대한 종속성 모델을 제안하였는데, 이는 한 청구 건에 대하여 복합 손실을 청구 유무와 총 청구금액으로 분해하는 모델이다. 여기서 다중 위험 청구 간의 의존성을 모델링하기 위해 Gaussian Copula를 이용한 다변량 이진회귀분석(multivariate binary regression)을 사용하였다. 반면에 Frees et al.(2016)은 총위험모델(collective risk model)하에서 복합 손실을 빈도와 심도 요소로 분해하는 다중 위험 보험의 종속성 모델을 제안하였다. 여기서 다변량 빈도와 심도 각각의 의존성을 반영하기

위해 Copula를 이용하였다. e Silva and de Lourdes Centeno(2017)는 다중 위험의 의존성을 반영하기 위하여 여러 가지 일반화된 이변량 빈도분포를 사용할 것을 제안하였으며, 이를 통해 관찰된 청구 건수에서 과대 산포를 확인하였다. Quan and Valdez(2018)는 다중 위험 청구 금액에 대한 회귀분석에서 다변량 의사결정 트리(multivariate decision tree)를 사용하는 방법을 제안하였다.

그러나 다중 위험 간의 의존성과 종단면적 특성을 모두 포함한 연구는 많지 않다. 예를 들어, Frees(2003)는 다중 위험 계약에 대한 복합 손실의 신뢰도 보험료를 산출하는 방법을 제안하였으나 청구 빈도와 심도의 평균 및 공분산만을 고려한 비모수적 접근방식이었다. 반면 Bermúdez and Karlis(2017)는 두 가지 유형의 청구에 대한 사후적 요율결정을 위해 이변량 Poisson 분포를 적용하였다. 두 연구는 공변량을 고려하지 않은 방식이기 때문에 계약자의 관찰된 이질성과 관찰되지 않은 이질성을 모두 포함하기는 어렵다.

Pechon et al.(2018), Pechon et al.(2019) 및 Pechon et al.(2020)은 상관 로그노말(correlated lognormal) 임의효과를 이용하여 여러 유형의 청구에 대한 신뢰도 보험료 공식을 제시하였다. 이는 종단면하에서 다중 위험 청구 사이의 의존성을 고려하는 포괄적인 체계를 제공하였다. 그러나 모든 계약자 각각에 대한 신뢰도 요인과 빈도의 주변분포를 도출하기 위해서는 다중(6~7중)적분이 필요하다. 일반적으로 보험 계약자 수가 매우 많기 때문에 실무적으로 적용하기에는 부담이 있을 수 있다.

본 연구에서 제안하는 방법론은 공유된 임의효과를 통해 다중 위험 보험에 대하여 계약자의 관찰된 이질성과 관찰되지 않은 이질성을 모두를 고려하고 종단면 분석을 가능하게 한다. 또한, 다중 위험 신뢰도 보험료에 대한 닫힌 공식을 최종적으로 제시한다.

Ⅲ. 연구 방법

1. 다중 위험 보험

일반적으로 다중 위험(multi-peril) 보험이란 필요한 여러 보장을 함께 묶는 보험이다. 다중 위험 보험에 가입하면 피보험자가 여러 가지 유형의 사고에 대해 보장 받을 수 있을

뿐만 아니라 함께 발생하는 손실에 대한 광범위한 보장으로 인해 전체 보험료 비용이 일반적으로 낮아진다. 다중 위험 보험에는 다양한 사례가 있으나 본 연구에서는 자동차 보험을 대상으로 분석하고자 한다. 자동차 보험은 자동차 사고에서 발생할 수 있는 다양한 유형의 경제적 손실을 보상한다. 하나의 자동차 사고로 인해 운전자, 동승자 및 보행자 등 인적 피해는 물론 자동차와 도로 위의 각종 구조물 등에 대한 물적 피해도 발생할 수 있으므로, 이러한 다양한 피해에 대한 보상을 제공하는 자동차 보험은 다중 위험 보험의 대표적 사례이다.

다중 위험 보험의 청구 건수에 대한 일반적인 특징을 표현하면 다음과 같다. 다중 위험 보험에서는 여러 유형의 사고가 발생할 수 있기 때문에 보험회사는 특정 계약자의 각 유형별 청구 건수를 시간의 흐름에 따라 관찰할 수 있다. i번째($i=1, \cdots, I$) 계약자에 대해 t년도($t=1, \cdots, T$)에 관찰 가능한 특성을 x_{it}라 하고, 이 때 발생한 j유형($j=1, \cdots, J$) 사고의 청구 건수를 $N^{(j)}{}_{it}$라고 나타내면 청구빈도에 대한 일반적인 데이터구조는 아래와 같이 나타낼 수 있다.

$$D=\left\{\left(N_{it}^{(1)}, \cdots, N_{it}^{(j)}, \cdots, N_{it}^{(J)}, \mathrm{x}_{it}\right)\Big| i=1, \cdots, I, t=1, \cdots, T\right\}. \tag{3}$$

2. 빈도모형

일반적으로 빈도모형으로는 포아송(Poisson) 분포를 사용하는 경우가 많다. 포아송 분포는 단순하여 사용이 용이한 장점이 있으나 자료가 과대산포(overdispersion)인 경우에는 문제가 발생할 수 있다. 즉, 평균과 분산이 같다고 가정하는 포아송 분포와 달리 경험데이터는 분산이 평균보다 큰 경우에는 포아송 분포가 적합하지 않을 수 있다는 것이다.

이러한 문제를 해결하기 위해 Wedderburn(1974)은 quasi-Poisson 분포를 제안하였다. 빈도 확률변수 N이 quasi-Poisson 분포 $QP(\nu, w)$를 따른다는 것은 확률변수 wN이 포아송 분포 $P(w\nu)$를 따른다는 것과 동치이다.

$$N \sim \ QP(\nu, w) \Leftarrow\Rightarrow wN \sim \ P(w\nu). \tag{4}$$

위 정의에 따르면 quasi-Poisson의 확률분포함수는 다음과 같이 구할 수 있다.

$$\Pr(N=n)=\Pr(wN=wn)=e^{-w\nu}\frac{(w\nu)^{wn}}{(wn)!}. \tag{5}$$

따라서 $E[wN]=wE[N]=w\nu$, $Var[wN]=w^2Var[N]=w\nu$이므로 $E[N]=\nu$, $Var[N]=\nu/w$가 된다. 즉, 분산이 모수 w에 반비례하기 때문에 w에 따라 평균과 분산의 대소 관계가 바뀌어 과대산포 문제를 해결할 수 있다.

Cameron and Trivedi(1990) 등의 연구에서는 아래와 같은 가설검정을 통하여 빈도모델의 과대산포 여부를 검정하는 방법을 제안하였다.

$$H_0 : w=1 \text{ vs. } H_1 : w \neq 1. \tag{6}$$

위 귀무가설은 평균과 분산이 같다($w=1$)는 것이고 이를 검정하기 위하여 회귀분석을 활용하였다. 본 연구에서는 위 연구에서 제안한 방법으로 분석 데이터의 과대산포 여부를 확인한 후 quasi-Poisson 분포를 빈도모형으로 사용하였다.

3. 자동차 보험 할인할증제도

자동차 보험의 할인할증제도는 1910년대 영국인에 의해 처음 도입되었는데 처음에는 NCD(no claim discount, 무청구할인)라는 용어를 사용하였다. 문자 그대로 “우리(보험회사)를 괴롭히지 않는다면 할인을 받는다. 어떤 이유로든 우리에게 접수를 한다면 할인을 받지 못한다”라는 의미였다. 따라서 회사에 보고된 모든 청구는 과실 여부에 관계없이 계약자에게 불이익을 부과한다. 그러나 곧 보험회사들은 이 제도를 폐지하게 되었고 1960년대에 유럽의 보험회사가 할인할증제도를 도입했을 때부터는 과실 사고에만 불이익을 부과하게 되었다.

할인할증제도의 목적은 두 가지로 볼 수 있다. 첫 번째는 보험 계약자와 보험회사 간의 정보 비대칭을 줄여서 보험회사가 각 운전자의 위험 등급에 대해 더 적절하고 공정한 요금을 부과하는 것이다. 예를 들어, 주행거리가 긴 운전자는 사고 위험에 더 많이 노출되기 때문에 더 많은 사고가 발생할 것이다. 그런데 주행거리를 신고하지 않는다면 보험회사는 주행거리를 사전적으로 보험료에 반영할 수 없다. 대신, 긴 주행거리를 사후적으로 요율에

반영하여 위험을 보전한다. 이를 뒷받침하는 가정은 더 많이 운전하거나 위험한 운전 습관을 가진 사람들은 과거에 더 많은 사고를 겪었고 미래에도 많은 사고를 발생시킬 것으로 기대된다는 것이다.

현재 시행 중인 모든 자동차 보험 할인할증제도가 청구 빈도만을 사용한다는 사실은 미래 청구 예측에 있어서도 대부분 청구 빈도에만 적용된다는 것을 의미한다. 청구 심도는 사고위험을 제대로 반영하지 못한다. 그럼에도 우리나라에서 심도를 할인할증제도에 반영했던 이유는 과거 사망사고 등 인적사고가 빈발하여 이를 억제하기 위한 정책적인 목적이었다. 그러나 현재는 물적사고의 비중이 증가하여 이를 유지할 필요성이 없어졌기 때문에 빈도만을 반영한 할인할증제도로 전환하여 사고위험을 보다 정확하게 반영한 요율을 적용할 수 있게 하였다.

자동차 보험 할인할증제도의 두 번째 목적은 도덕적 해이를 통제하는 것이다. 자동차 보험 할인할증제도와 같은 경험요율 제도를 경험한 보험 계약자는 보험금 청구 이력에 따라 보험료가 인상될 수도 있다는 것을 알고 있고 보험금 청구빈도를 줄이기 위한 유인을 가지게 된다. 이러한 과정을 통해 도덕적 해이 문제를 통제할 수 있다.

4. 다중 위험 보험의 임의효과

식 (3)에서 나타나는 청구빈도 데이터구조의 중요한 특징은 종단면성(longitudinality)이다. 즉, 같은 계약자에 대해 시간의 흐름에 따라 청구이력을 반복적으로 관찰한다는 것이다. 따라서 각 계약자의 위험요인에 내재된 관찰되지 않은 이질성을 설명하기 위해서는 빈도모형에 임의효과를 반영하는 것이 필요하다.

관찰 가능한 특성에는 다양한 변수가 포함될 수 있다. 예컨대 연령, 성별, 차량연식, 지역, 용도(개인용, 업무용, 영업용 등) 등이 대표적이다. 이러한 다양한 특성이 각 계약자의 청구빈도를 모두 설명할 수는 없다. 즉, 관찰 가능한 특성이 모두 동일한 계약자들 사이에도 청구이력은 확연히 다를 수 있다는 것이다. 따라서 관찰된 특성이 계약자들이 가진 위험요인의 이질성을 구분하는데 충분하지 않다고 의심할 수 있다. 그런데 관찰하는 특성의 숫자를 늘린다고 해서 모든 이질성을 구분할 수는 없을 것이다. 그 특성이라는 것이 너무 많아 효율성이 떨어질 뿐 아니라 애초에 운전습관과 같이 관찰이 불가능한 특성도 존재하

기 때문이다. 그러므로 관찰되지 않은 이질성을 포함하기 위해서는 임의효과를 빈도모형에 적용할 수 있다. 임의효과 모델은 신뢰도 이론 및 할인할증제도와도 밀접하게 관련되어 있다. Gómez-Déniz and Vázquez-Polo(2005)와 Jeong and Valdez(2020b)를 비롯하여 요율산정 관점에서 임의효과를 적용한 여러 연구가 있었다.

요약하면, 청구빈도는 관찰된 공변량과 회귀계수 모두에 영향을 받고 관찰되지 않은 이질성요인 θ_i에도 영향을 받는 것으로 가정한다. 계약자 i에 있어서 θ_i 정보는 모든 유형의 보장과 연도별 청구 건에 공유된다. 이를 반영하여 앞서 제안한 quasi-Poisson 분포를 표현하면 다음과 같다.

$$N_{it}^{(j)} \big| \mathbf{x}_{it}, e_{it}, \theta_i \overset{indep}{\sim} QP\left(\theta_i \nu_{it}^{(j)}, w^{(j)}\right) \Leftarrow\Rightarrow w^{(j)} N_{it}^{(j)} \big| \mathbf{x}_{it}, e_{it}, \theta_i \overset{indep}{\sim} P\left(\theta_i w^{(j)} \nu_{it}^{(j)}\right). \tag{7}$$

여기서 자료는 풀링하지 않고 패널모형을 기본적으로 사용하며, 모형은 $\boldsymbol{x}_{it}$가 변함에 따라 매년 달라질 수 있는 고정효과(fixed effect)와 한 보험계약자에 대해 동일하게 적용되는 θ_i, 즉 정적확률효과(static random effect)를 동시에 사용한다. 그리고 $\nu_{it}^{(j)} = e_{it} e^{\mathbf{x}_{it}\alpha^{(j)}}$인데 e_{it}는 각 계약자의 t년도의 익스포져(exposure)로 $0 < e_{it} \le 1$이다. $\alpha^{(j)}$는 j번째 보장 유형에 대한 회귀계수, $w^{(j)}$는 j번째 보장 유형에 대한 가중치이다. $\nu_{it}^{(j)}$는 관찰된 공변량과 그 회귀계수로 결정되므로 각 계약자(i)의 유형(j)별로 각 기간(t)동안 위험요인이 내포하고 있는 관찰된 이질성을 설명한다. 반면 임의효과 θ_i는 계약자(i)의 위험요인에 내포된 관찰되지 않은 이질성을 설명한다. 임의효과의 기댓값은 $E[\theta_i] = 1$ 으로 가정하는데 이는 θ_i가 할인할증요인 역할을 하는 승법(multiplicative) 임의효과이기 때문이다. 가법(additive) 임의효과의 경우에는 $E[\theta_i] = 0$으로 가정하는 것이 일반적이므로 승법 임의효과에 대해서는 $E[\theta_i] = 1$로 설정하는 것이 자연스럽다.

위 분포에 의해 청구 건수 $N_{it}^{(j)}$의 평균을 구해보면 다음과 같다. 먼저 quasi-Poisson 분포의 평균을 이용하여 $E\left[N_{it}^{(j)} | \theta_i\right] = \nu_{it}^{(j)} \theta_i$ 임을 알 수 있고 $E[\theta_i] = 1$ 이므로 $E\left[N_{it}^{(j)}\right] = E\left[E\left[N_{it}^{(j)} | \theta_i\right]\right] = E\left[\nu_{it}^{(j)} \theta_i\right] = \nu_{it}^{(j)}$ 이 된다. 그러므로 $\nu_{it}^{(j)}$는 관찰되지 않은 이질성 θ_i에 대한 정보가 없을 때 피보험자 i의 시간 t에서 청구 건수의 사전평균(prior mean)으로 해석할 수 있다.

종단면 분석에 있어서 다중 위험 청구 간에 있을 수 있는 종속성을 포함시키는 방법으로 공유된 임의효과가 유일한 것은 아니다. 대표적인 방법은 Copula를 활용하여 여러 유형 간의 의존성을 반영하는 것으로 Yang and Shi(2019)에서 연구되었다.

관찰되지 않은 이질성이 청구 빈도에 미치는 영향은 사전적으로는 알 수 없다. 따라서 특정한 사전분포(prior distribution)를 가지고 표현해야 한다. 베이지안 분석에서는 임의의 모수(parameter)에 대한 충분한 정보가 있지 않다면 일반적으로 θ에 대한 비정보 사전분포(noninformative prior)를 사용한다. 이러한 분포 중 가장 널리 사용되는 분포가 Jeffreys의 사전분포이다. Jeffreys(1946)에 따르면 Jeffreys의 사전분포는 Fisher 정보행렬(information matrix)의 행렬식(determinant)에 대한 제곱근으로 정의된다. 위 모델에서 θ에 대한 Jeffreys의 사전분포는 $\pi(\theta) = \frac{1}{\sqrt{\theta}}$이 되고 이에 대응되는 사후분포는 다음과 같이 구할 수 있다.

사후분포의 정의는 $\pi(\theta|n)= \frac{\Pr(n|\theta)\pi(\theta)}{\int_0^\infty \Pr(n|\theta)\pi(\theta)d\theta}$ 이므로 분자부분으로 결정된다. 식 (5)에서 구한 quasi-Poisson 확률분포함수를 대입하여 분자부분을 구하고 그에 비례하는 가장 단순한 식을 구해보면 다음과 같다.

$$\Pr(n|\theta)\pi(\theta)= e^{-w\nu\theta}\frac{(w\nu\theta)^{wn}}{(wn)!}\frac{1}{\sqrt{\theta}} \propto \theta^{wn-\frac{1}{2}}e^{-w\nu\theta}. \tag{8}$$

이는 감마분포의 확률분포함수에 비례한다. 즉, $\theta|n \sim \Gamma\left(wn+\frac{1}{2}, \frac{1}{w\nu}\right)$이 된다는 것을 알 수 있다. 그러나 위 사전분포는 식별가능성(identifiability) 문제가 있다. 즉, 분포를 유일하게 결정하는 모수가 존재하지 않는다는 것이다. 따라서 임의효과의 기댓값에 대한 조건인 $E[\theta_i]=1$을 적용하여 이 문제를 해결할 수 있다. 또한 모델분포의 켤레 사전분포(conjugate prior)를 이용하는 것이 유용하므로 이를 만족하는 분포를 찾으면 다음과 같다.

모델분포가 선형지수함수 계통(linear exponential family)에 속할 때 사전분포와 사후분포가 동일한 형태인 켤레 관계가 된다는 성질을 이용한다. 즉, 모델분포를 $\Pr(N=n|\theta)= \frac{p(n)e^{r(\theta)n}}{q(\theta)}$로 표현할 때, 사전분포가 $\pi(\theta)= \frac{[q(\theta)]^{-k}e^{\mu k r(\theta)}r'(\theta)}{c(\mu, k)}$라

면 이에 대응되는 사후분포 역시 동일한 형태의 분포가 된다. quasi-Poisson 분포는 선형 지수함수 계통에 해당되므로 위의 모델분포에 대입하면 다음과 같다.

$p(n)=\dfrac{1}{(wn)!}$, $q(\theta)=e^{w\nu\theta}$, $r(\theta)=w\ln(w\nu\theta)$, $r'(\theta)=\dfrac{w}{\theta}$이므로 이를 대입하여 사전분포를 구하면 $\pi(\theta)\propto\left[e^{w\nu\theta}\right]^{-k}e^{\mu kw\ln(w\nu\theta)}\dfrac{w}{\theta}\propto e^{-kw\nu\theta}\theta^{\mu kw-1}$이 된다. 이는 감마 분포이므로 $\theta\sim\Gamma\left(\mu kw,\dfrac{1}{kw\nu}\right)$ 임을 알 수 있다. 여기에 평균에 대한 조건을 적용하면 $E[\theta]=\dfrac{\mu}{\nu}=1$이므로 $\mu=\nu$이고 따라서 감마분포의 두 모수는 서로 역수관계임을 알 수 있다. 최종적으로 $r=\mu kw=kw\nu$라 하고 결과를 정리하면 다음과 같다.

$$\pi(\theta)\propto\theta^{r-1}e^{-\theta r},\ \text{즉}\ \theta\sim\Gamma\left(r,\frac{1}{r}\right)\text{이므로}\ E[\theta]=1,\ Var[\theta]=\frac{1}{r}. \tag{9}$$

이 결과를 통해 $t,t'=1,\cdots,T$와 $j,k=1,\cdots,J$에 대해 다음 통계량을 산출할 수 있다.

$$\begin{aligned}E\left[N_t^{(j)}\right]&=E\left[E\left[N_t^{(j)}|\theta\right]\right]\\&=E\left[\nu_t^{(j)}\theta\right]\\&=\nu_t^{(j)},\end{aligned} \tag{10}$$

$$\begin{aligned}Var\left[N_t^{(j)}\right]&=E\left[Var\left[N_t^{(j)}|\theta\right]\right]+Var\left[E\left[N_t^{(j)}|\theta\right]\right]\\&=E\left[\frac{\theta\nu_t^{(j)}}{w^{(j)}}\right]+Var\left[\theta\nu_t^{(j)}\right]\\&=\frac{\nu_t^{(j)}}{w^{(j)}}+\frac{\left(\nu_t^{(j)}\right)^2}{r},\end{aligned} \tag{11}$$

$$\begin{aligned}Cov\left(N_t^{(j)},N_{t'}^{(k)}\right)&=E\left[N_t^{(j)}N_{t'}^{(k)}\right]-E\left[N_t^{(j)}\right]E\left[N_{t'}^{(k)}\right]\\&=E\left[E\left[N_t^{(j)}N_{t'}^{(k)}|\theta\right]\right]-E\left[N_t^{(j)}\right]E\left[N_{t'}^{(k)}\right]\\&=E\left[\nu_t^{(j)}\nu_{t'}^{(k)}\theta^2\right]-\nu_t^{(j)}\nu_{t'}^{(k)}\\&=\nu_t^{(j)}\nu_{t'}^{(k)}\left(E\left[\theta^2\right]-1\right)\\&=\nu_t^{(j)}\nu_{t'}^{(k)}\left(Var[\theta]+E[\theta]^2-1\right)\\&=\frac{\nu_t^{(j)}\nu_{t'}^{(k)}}{r},\end{aligned} \tag{12}$$

$$
\begin{aligned}
Corr\left(N_t^{(j)}, N_{t'}^{(k)}\right) &= \frac{Cov\left(N_t^{(j)}, N_{t'}^{(k)}\right)}{\sqrt{Var\left[N_t^{(j)}\right]}\sqrt{Var\left[N_{t'}^{(k)}\right]}} \\
&= \frac{\dfrac{\nu_t^{(j)}\nu_{t'}^{(k)}}{r}}{\sqrt{\dfrac{\nu_t^{(j)}}{w^{(j)}}+\dfrac{\left(\nu_t^{(j)}\right)^2}{r}}\sqrt{\dfrac{\nu_{t'}^{(k)}}{w^{(k)}}+\dfrac{\left(\nu_{t'}^{(k)}\right)^2}{r}}} \\
&= \frac{1}{\sqrt{1+\dfrac{r}{w^{(j)}\nu_t^{(j)}}}\sqrt{1+\dfrac{r}{w^{(k)}\nu_{t'}^{(k)}}}}.
\end{aligned}
\tag{13}
$$

따라서 제안 모델은 시간이 지남에 따라 다중 위험 빈도 사이의 의존성을 포함하고 있다. 예를 들어, $r \to \infty$ 이면 제안 모델은 독립 quasi-Poisson 모델로 축소된다. 또한, $r \to \infty$ 이고 모든 $j = 1, \cdots, J$ 에 대해 $w^{(j)} = 1$ 이면 제안 모델은 독립적인 Poisson 모델로 축소된다.

빈도분석을 통한 요율산출은 빈도변수의 조건부 평균인 $E\left[N_t^{(j)}|\theta\right] = \theta\nu_t^{(j)}$ 로 볼 수 있다. 즉, 보험요율은 $\nu_t^{(j)}$ 와 임의효과 θ 의 곱으로 표현되기 때문에 θ 가 보험료 결정을 위한 할인할증요인의 역할임을 알 수 있다.

하이퍼파라미터(hyperparameter) r 의 값은 사전적 지식이나 적률법(method of moments) 등을 이용하여 결정할 수 있다. 예를 들어, Lemaire(1998)에 따르면, 빈도요율에 대한 할인할증요인의 범위는 일반적으로 54%에서 200%이다. 따라서 제안된 사전분포의 하이퍼파라미터 r 을 선택함에 있어 이러한 사전적 지식을 적용시켜야 한다. 즉, θ 의 95% 최고사후밀도구간(highest posterior density interval, 이하, 'HPDI'라 함)이 (0.54, 2.00)을 포함할 수 있도록 해야 한다. 또한 일치추정량으로 r 을 추정할 수도 있는데 Sutradhar and Jowaheer(2003)에 의하면 $i = 1, \cdots, I$ 에 대해 $\sum_{t \neq t', j \neq k}\left(N_{it}^{(j)} - \nu_{it}^{(j)}\right)\left(N_{it'}^{(k)} - \nu_{it'}^{(k)}\right)$ 의 기댓값은 $\frac{1}{r}\sum_{t \neq t', j \neq k}\nu_{it}^{(j)}\nu_{it'}^{(k)}$ 이고 따라서 r 에 대한 추정량은 다음과 같다.

$$\hat{r} = \frac{\sum_{i=1}^{I} \sum_{t \neq t', j \neq k} \hat{\nu}_{it}^{(j)} \hat{\nu}_{it'}^{(k)}}{\sum_{i=1}^{I} \sum_{t \neq t', j \neq k} \left(N_{it}^{(j)} - \hat{\nu}_{it}^{(j)}\right)\left(N_{it'}^{(k)} - \hat{\nu}_{it'}^{(k)}\right)}. \tag{14}$$

여기서 $\hat{\nu}$는 미리 정해진 평균 구조를 통해 일반화추정방정식(generalized estimating equation; GEE)을 풀어 추정된 일치추정량이다(Liang and Zeger 1986, pp. 15-19). 이 가정에 따르면 다음과 같이 θ에 대한 사후분포를 유도할 수 있다.

사후분포의 정의는 $\pi(\theta|n) = \dfrac{\Pr(n|\theta)\pi(\theta)}{\int_0^{\infty} \Pr(n|\theta)\pi(\theta)d\theta}$ 이므로 모델분포와 사전분포를 대입하면 다음과 같다.

$$\begin{aligned}
\pi\left(\theta|\mathrm{n}_T^{(1)}, \cdots, \mathrm{n}_T^{(J)}\right) &\propto \pi(\theta) \prod_{j=1}^{J} \prod_{t=1}^{T} \Pr\left(w^{(j)} n_t^{(j)} | \theta\right) \\
&\propto \theta^{r-1} e^{-\theta r} \prod_{j=1}^{J} \prod_{t=1}^{T} \theta^{w^{(j)} n_t^{(j)}} e^{-w^{(j)} \nu_t^{(j)} \theta} \\
&= \theta^{\sum_{j=1}^{J} \sum_{t=1}^{T} w^{(j)} n_t^{(j)} + r - 1} e^{-\theta\left(\sum_{j=1}^{J} \sum_{t=1}^{T} w^{(j)} \nu_t^{(j)} + r\right)}.
\end{aligned} \tag{15}$$

따라서 사후분포는 다음의 감마분포임을 알 수 있다. 즉,

$$\theta|\mathrm{n}_T^{(1)}, \cdots, \mathrm{n}_T^{(J)} \sim \Gamma\left(\sum_{j=1}^{J} \sum_{t=1}^{T} \mathrm{w}^{(j)} \mathrm{n}_t^{(j)} + \mathrm{r}, \frac{1}{\sum_{j=1}^{J} \sum_{t=1}^{T} \mathrm{w}^{(j)} \nu_t^{(j)} + \mathrm{r}}\right). \tag{16}$$

또한 $\mathrm{n}_T^{(1)}, \cdots, \mathrm{n}_T^{(J)}$이 주어진 경우 N_{T+1}의 예측 분포를 유도할 수 있다. 이를 위해 $\left(w^{(1)}\mathrm{n}_T^{(1)}, \cdots, w^{(J)}\mathrm{n}_T^{(J)}\right)$와 $w^{(j)} N_{T+1}^{(j)}$가 제안 모델을 따른다고 가정한다. 그리고 표기상 편의를 위해 $r_T = r + \sum_{j=1}^{J} \sum_{t=1}^{T} w^{(j)} n_t^{(j)}$ 이고 $\hat{r}_T = r + \sum_{j=1}^{J} \sum_{t=1}^{T} w^{(j)} \nu_t^{(j)}$라고 하면 식 (15)에서 구한 사후분포는 다음과 같이 표현된다.

$$\pi\left(\theta|\mathrm{n}_T^{(1)},\cdots,\mathrm{n}_T^{(J)}\right) \propto \theta^{\sum_{j=1}^{J}\sum_{t=1}^{T}w^{(j)}n_t^{(j)}+r-1} e^{-\theta\left(\sum_{j=1}^{J}\sum_{t=1}^{T}w^{(j)}\nu_t^{(j)}+r\right)} = \theta^{r_T-1}e^{-\hat{r}_T\theta}. \tag{17}$$

따라서 사후분포 함수는 다음과 같다.

$$\pi\left(\theta|\mathrm{n}_T^{(1)},\cdots,\mathrm{n}_T^{(J)}\right) = \frac{\left(\hat{r}_T\theta\right)^{r_T}e^{-\hat{r}_T\theta}}{\theta\Gamma(r_T)}. \tag{18}$$

마지막으로 과거 데이터 n이 주어졌을 때 x에 대한 예측분포는 $\Pr(x|n)=\int_0^\infty \Pr(x|\theta)\pi(\theta|n)d\theta$으로 산출되므로, 여기에 식 (5), (7)과 (18)을 대입하여 전개하면 다음과 같다.

$$\begin{aligned}
&\Pr\left(w^{(j)}N_{T+1}^{(j)}=w^{(j)}n_{T+1}^{(j)}|\mathrm{n}_T^{(1)},\ \cdots\ ,\mathrm{n}_T^{(J)}\right) \qquad (19)\\
&=\int_0^\infty \Pr\left(w^{(j)}N_{T+1}^{(j)}=w^{(j)}n_{T+1}^{(j)}|\theta\right)\pi\left(\theta|\mathrm{n}_T^{(1)},\cdots,\mathrm{n}_T^{(J)}\right)d\theta\\
&=\int_0^\infty e^{-w^{(j)}\nu_{T+1}^{(j)}\theta}\frac{\left(w^{(j)}\nu_{T+1}^{(j)}\theta\right)^{w^{(j)}n_{T+1}^{(j)}}}{\left(w^{(j)}n_{T+1}^{(j)}\right)!}\frac{\left(\hat{r}_T\theta\right)^{r_T}e^{-\hat{r}_T\theta}}{\theta\Gamma(r_T)}d\theta\\
&=\left(w^{(j)}\nu_{T+1}^{(j)}\right)^{w^{(j)}n_{T+1}^{(j)}}\hat{r}\\
&=\int_0^\infty\left\{\left(w^{(j)}\nu_{T+1}^{(j)}+\hat{r}_T\right)\theta\right\}^{w^{(j)}n_{T+1}^{(j)}+r_T-1}e^{-\left(w^{(j)}\nu_{T+1}^{(j)}+\hat{r}_T\right)\theta}\left(w^{(j)}\nu_{T+1}^{(j)}+\hat{r}_T\right)d\theta\\
&\qquad\times\frac{\left(w^{(j)}\nu_{T+1}^{(j)}\right)^{w^{(j)}n_{T+1}^{(j)}}\hat{r}_T^{r_T}}{\left(w^{(j)}n_{T+1}^{(j)}\right)!\Gamma(r_T)\left(w^{(j)}\nu_{T+1}^{(j)}+\hat{r}_T\right)^{w^{(j)}n_{T+1}^{(j)}+r_T}}\\
&=\frac{\left(w^{(j)}\nu_{T+1}^{(j)}\right)^{w^{(j)}n_{T+1}^{(j)}}\hat{r}_T^{r_T}\Gamma\left(w^{(j)}n_{T+1}^{(j)}+r_T\right)}{\left(w^{(j)}n\right.}
\end{aligned}$$

그러므로 $w^{(j)}N_{T+1}^{(j)}|\mathrm{n}_T^{(1)},\ \cdots\ ,\mathrm{n}_T^{(J)} \sim NB\left(r_T, \frac{w^{(j)}\nu_{T+1}^{(j)}}{\hat{r}_T+w^{(j)}\nu_{T+1}^{(j)}}\right)$이 된다. 예측분포가 음이항분포임을 이용하면 기댓값은 다음과 같다.

$$E\left[w^{(j)}N_{T+1}^{(j)}|\mathrm{n}_T^{(1)}, \cdots, \mathrm{n}_T^{(J)}\right]=\frac{w^{(j)}\nu_{T+1}^{(j)}r_T}{\hat{r}_T+w^{(j)}\nu_{T+1}^{(j)}}\frac{\hat{r}_T+w^{(j)}\nu_{T+1}^{(j)}}{\hat{r}_T}=\frac{w^{(j)}r_T}{\hat{r}_T}\nu_{T+1}^{(j)}. \tag{20}$$

따라서 최종적으로 예측분포의 기댓값은 다음과 같다. 이 결과는 식 (2)와 동일함을 알 수 있다.

$$\begin{aligned} E\left[N_{T+1}^{(j)}|\mathrm{n}_T^{(1)}, \cdots, \mathrm{n}_T^{(J)}\right] &= \frac{r_T}{\hat{r}_T}\nu_{T+1}^{(j)} \\ &= \frac{r+\sum_{j=1}^{J}\sum_{t=1}^{T}w^{(j)}n_t^{(j)}}{r+\sum_{j=1}^{J}\sum_{t=1}^{T}w^{(j)}\nu_t^{(j)}}\nu_{T+1}^{(j)}. \end{aligned} \tag{21}$$

(16)에 의해 사후분포가 감마분포이므로 사후분포의 기댓값은 다음과 같다.

$$E\left[\theta|\mathrm{n}_T^{(1)},\cdots,\mathrm{n}_T^{(J)}\right]=\frac{r+\sum_{j=1}^{J}\sum_{t=1}^{T}w^{(j)}n_t^{(j)}}{r+\sum_{j=1}^{J}\sum_{t=1}^{T}w^{(j)}\nu_t^{(j)}}. \tag{22}$$

이를 식 (21)에서 구한 예측분포의 기댓값에 적용하면 다음과 같이 표현할 수 있다.

$$E\left[N_{T+1}^{(j)}|\mathrm{n}_T^{(1)}, \cdots, \mathrm{n}_T^{(J)}\right] = E\left[N_{T+1}^{(j)}\right]E\left[\theta|\mathrm{n}_T^{(1)},\cdots,\mathrm{n}_T^{(J)}\right]. \tag{23}$$

따라서 과거 T년 동안의 청구 정보가 주어질 때, $N_{T+1}^{(j)}$의 예측 요율은 사전적 보험료 $\nu_{T+1}^{(j)}$와 신뢰도 요인의 합동추정값(pooled estimate)의 곱으로 산출됨을 알 수 있다. 여기서 사전적 요율은 관측 가능한 공변량과 회귀계수에 의해 결정되며 신뢰도 요인의 추정치는 관찰되지 않은 공유된 이질성을 설명한다.

Frangos and Vrontos(2001)는 $J=1$이고 $w^{(1)}=1$인 경우를 연구하였는데 식 (21)은 이를 확장한 형태이다. 예를 들어 $J=2$이고 $N^{(1)}$과 $N^{(2)}$가 각각 대인배상 및 대물배상 담보로 인한 청구 빈도를 나타낸다고 할 때, 관찰되지 않은 이질성의 합동추정값은 아래와 같다.

$$E\left[\theta|\mathrm{n}_{\mathrm{T}}^{(1)}, \mathrm{n}_{\mathrm{T}}^{(2)}\right] = \frac{\mathrm{r} + \mathrm{w}^{(1)}\sum_{\mathrm{t}=1}^{\mathrm{T}}\mathrm{n}_{\mathrm{t}}^{(1)} + \mathrm{w}^{(2)}\sum_{\mathrm{t}=1}^{\mathrm{T}}\mathrm{n}_{\mathrm{t}}^{(2)}}{\mathrm{r} + \mathrm{w}^{(1)}\sum_{\mathrm{t}=1}^{\mathrm{T}}\nu_{\mathrm{t}}^{(1)} + \mathrm{w}^{(2)}\sum_{\mathrm{t}=1}^{\mathrm{T}}\nu_{\mathrm{t}}^{(2)}} = \frac{\text{평활인자} + \text{실제 빈도의 가중합}}{\text{평활인자} + \text{예측 빈도의 가중합}}. \tag{24}$$

따라서 제안 모델을 사용하면 다중 위험에 대해 다양한 가중치 $w^{(j)}$를 적용함으로써 사후적 요율 요인 $E\left[\theta|\mathrm{n}_{\mathrm{T}}^{(1)}, \cdots, \mathrm{n}_{\mathrm{T}}^{(J)}\right]$에 대한 다양한 수준의 기여도를 고려할 수 있다.

식 (24)에서, θ의 사전분포의 하이퍼파라미터 r은 평활 인자의 역할을 한다. 예를 들어, $r \rightarrow \infty$이면 $\pi(\theta|\mathrm{n})$은 1에서 디락 델타 함수(Dirac delta function)로 수렴하며, 이는 $\mathrm{Pr}(\theta=1)=1$을 의미하고 매우 정보적인(very informative) 점-질량(point-mass) 사전분포가 된다. 반면에 더 작은 r을 선택하면 덜 정보적인(less informative) 사전분포를 사용하게 된다.

경험적 베이즈 방법(empirical Bayes method)을 적용하면 $j=1,2,\cdots,J$에 대한 회귀계수 $\alpha^{(j)}$의 추정치는 다음과 같은 결합로그우도(joint loglikelihood)를 최대화하여 구할 수 있다.

$$l\left(\alpha^{(1)}, \cdots, \alpha^{(J)}|\mathrm{n}_{\mathrm{T}}^{(1)}, \cdots, \mathrm{n}_{\mathrm{T}}^{(\mathrm{J})}\right) = \sum_{\mathrm{i}=1}^{\mathrm{I}} \ln \mathrm{Pr}\left(w^{(1)}\mathrm{n}_{\mathrm{iT}}^{(1)}, \cdots, \mathrm{w}^{(\mathrm{J})}\mathrm{n}_{\mathrm{iT}}^{(\mathrm{J})}\right). \tag{25}$$

이 때,

$$\mathrm{Pr}\left(w^{(1)}\mathrm{n}_{\mathrm{iT}}^{(1)}, \cdots, \mathrm{w}^{(\mathrm{J})}\mathrm{n}_{\mathrm{iT}}^{(\mathrm{J})}\right) = \int \prod_{\mathrm{j}=1}^{\mathrm{J}}\prod_{\mathrm{t}=1}^{\mathrm{T}} \mathrm{Pr}\left(w^{(j)}n_{it}^{(j)}|\theta_i\right)\pi(\theta_i)d\theta_i \tag{26}$$

$$\propto \prod_{j=1}^{J}\prod_{t=1}^{T}\left(\frac{w^{(j)}\nu_{it}^{(j)}}{\sum_{j=1}^{J}\sum_{t=1}^{T}w^{(j)}\nu_{it}^{(j)}+r}\right)^{w^{(j)}n_{it}^{(j)}}\left(\frac{r}{\sum_{j=1}^{J}\sum_{t=1}^{T}w^{(j)}\nu_{it}^{(j)}+r}\right)^{r}$$

이고 이는 다변량음이항분포(MVNB)를 따른다.

Ⅳ. 실증분석

한국의 자동차 보험에서 가장 중요한 보장은 대인보상(bodily injury, 이하, ‘BI’라 함)과 대물보상(property damage, 이하, ‘PD’라 함) 담보이다. 자동차 보험은 한 보험계약자가 하나의 보험사고로 인해 여러 담보에서 보험금 청구가 가능하기 때문에 다중 위험(multi-peril) 보험으로 볼 수 있다. 이러한 자동차 보험의 특성을 반영하여 적절한 보험요율 산정 방법을 연구하고자 한다.

본 연구에서는 국내 한 손해보험회사에서 2014년부터 2019년까지 개인용 자동차 보험을 가입한 전체 계약의 데이터를 사용하였다. 1년 미만인 계약을 포함하며, 업무용 및 영업용 차량은 제외한다. 보험개발원 보험통계조회서비스 자료에 따르면 2014년부터 2019년까지 해당 보험회사의 자동차 책임보험의 원수보험료 기준 시장점유율은 5%가 넘는다. 본 연구에서는 1년 평균 55만 건이 넘는 관측치의 자료를 사용하였는데, 기승도·김대환(2009)은 약 50만 건의 1년 자료를 사용하였고 김영화·이현수(2010)는 1년 평균 약 30만 건의 3년 간 자료를 사용한 것과 비교하여 충분한 크기의 표본일 것으로 판단된다. 1년 미만인 데이터에는 개별 계약자에 대한 성별, 보험가입경력, 할인할증등급, 차량크기, 연령, 차량연식, 대인보상 및 대물보상 보험금 청구 건수 정보가 포함되어 있다. 보험개발원 보험통계조회서비스의 전체 모집단의 성별, 연령대별 자동차보험 건수 자료의 분포와 비교하여서도 큰 차이가 없으므로, 분석대상 데이터의 대표성에도 문제가 없을 것으로 판단된다.

일반적으로 자동차 보험료에는 차량의 종류, 배기량, 피보험자의 성별, 연령뿐만 아니라 운전자 연령한정, 범위한정 특약 등의 특약 가입여부, 보험가입경력, 과거 사고발생 실적을 반영한 할인할증요율 등 다양한 요인들이 영향을 미칠 수 있다. 기승도·김대환(2009)은 성별, 연령, 지역, 가입경력, 교통법규위반경력, 할인할증등급을 변수로 고려하였고, 김영화·김미정·김명준(2011)은 성별, 연령, 운전경력, 할인할증등급, 운전자한정 특약가입여부를 변수로 고려하였다. 본 연구는 이를 참고하여 자료에서 얻을 수 있는 변수들을 선택하고 공변량으로 사용하였다. 본 연구의 목적이 다중 위험 보험의 빈도분석에 공유된 임의효과 모델을 적용한 효과를 확인하는 것이므로, 임의효과를 사용함으로써 유의성이 높다고 판단된 변수들을 제외한 나머지 변수와 관찰되지 않는 특성들도 간접적으로 반영될

수 있게 하였다.

2014년부터 2018년까지의 데이터를 훈련자료(training data)로, 2019년 데이터를 시험자료(test data)로 선택하였고 관측 건수는 각각 2,933,233건, 490,940건이다. 시간이 지남에 따라 보험 계약자의 청구를 추적할 수 있는 고유한 식별자가 자료에 포함되어 있기 때문에 본 자료는 패널데이터이며, 계약을 지속한 경우와 지속하지 않은 경우를 포함한 모든 계약이 자료에 포함되기 때문에 특히 불균형 패널데이터이다. 손해보험, 그 중에서도 특히 자동차보험의 자료는 본질적으로 불균형 패널데이터인 경우가 많은데, 본 연구의 방법론은 패널의 관측길이에 관계없이 적용 가능하다는 장점이 있다. 또한 데이터에는 각 계약자의 연도별 계약정보와 BI 및 PD의 청구 정보가 포함되어 있기 때문에 다변수 패널데이터 빈도 모델링을 시도할 수 있다.

〈Table 1〉, 〈Table 2〉는 훈련자료에서 관찰 가능한 계약 특성을 정리한 표이며, 각각 범주형 변수와 양적 변수의 기술통계량에 대해 요약한 결과이다. 범주형 변수로는 성별, 보험가입경력, 할인할증등급, 차량크기를, 양적 변수로는 연령과 차량연식을 공변량으로 선택하였다. 성별(Gender)은 남성(0), 여성(1)으로 구분하였고 보험가입경력(YDE)은 ‘3년 미만’(0), ‘3년 이상 5년 미만’(1), ‘5년 이상 7년 미만’(2), ‘7년 이상’(3)의 4가지 범주로 구분하였다. 할인할증등급(BM)은 자료에 세분화된 등급이 있으나 분석 시 크게 할증(0)과 할인(신규 포함)(1)으로 구분하였다. 마지막으로 차량크기(AutoSize)는 배기량으로 구분하였는데 1,000cc 미만을 경형(0), 1,600cc 미만을 소형(1), 2,000cc 미만을 중형(2), 2,000cc 이상을 대형(3)으로 구분하였다. 양적 변수인 연령(Age)의 평균은 48.9세이며 차량연식(AutoAge)의 평균은 8.2년이다.

〈Table 1〉 Observable policy characteristics (categorical variables)

Categorical variable	Value	Number of policies	Proportion to total
Gender	(0) Male (base)	2,133,111	72.72%
	(1) Female	800,122	27.28%
	Total	2,933,233	100.00%
YDE (years of driving experience)	(0) Under 3 (base)	320,887	10.94%
	(1) 3 to Under 5	202,168	6.89%
	(2) 5 to Under 7	189,949	6.48%
	(3) 7 and over	2,220,229	75.69%
	Total	2,933,233	100.00%
BM (bonus-malus)	(0) Malus (base)	187,400	6.39%
	(1) Bonus	2,745,833	93.61%
	Total	2,933,233	100.00%
AutoSize	(0) Light (base)	658,726	22.46%
	(1) Compact	1,162,854	39.64%
	(2) Medium	793,902	27.07%
	(3) Large	317,751	10.83%
	Total	2,933,233	100.00%

〈Table 2〉 Observable policy characteristics (quantitative variables)

Quantitative variable	Mean	Minimum	Maximum	S. D.
Age	48.92	18.00	100.00	11.17
AutoAge	8.22	0.00	40.00	5.19
BI claim freq.	0.05	0.00	9.00	0.23
PD claim freq.	0.16	0.00	14.00	0.44

BI와 PD가 독립 포아송 분포를 따른다고 할 때, 청구 건수에 대한 포아송 회귀분석의 결과는 〈Table 3〉과 같다. BI 청구 건수 결과를 보면 남성에 비하여 여성의 빈도가 더 높은 것으로 나타났다. 가입경력은 3년 미만에 비하여 ‘3년 이상 5년 미만’, ‘5년 이상 7년 미만’, ‘7년 이상’으로 경력이 길어질수록 청구 건수가 적게 나타났다. 할인할증 등급은 할증에 비하여 할인등급이 청구 건수가 적었다. 연령은 1차항의 계수가 음수이고 2차항의 계수가 양수이므로 연령이 증가함에 따라 청구 건수가 감소하다가 고연령대에서 다시 증

가하는 형태임을 알 수 있다. 이상의 변수들에 대해서는 PD 청구 건수도 이와 동일한 형태를 보였다. 이러한 결과는 유사한 변수로 사고빈도를 분석한 선행연구들의 결과와도 그 경향이 일치한다. 예를 들어, 김동훈·이계연(1998)은 유의미한 변수 중 가입경력이 짧을수록, 할증률이 클수록, 배기량이 클수록, 차량연식이 짧을수록 청구 건수가 증가하는 것으로 분석하였다. 기승도·김대환(2009)은 남자에 비해 여자가, 가입경력이 짧을수록, 할증률이 클수록 청구 건수가 증가하고, 연령이 증가할수록 그 건수가 감소하였다가 장년층에서 다시 증가하는 양상을 보이며, 차종에 따라서 청구 건수가 달라짐을 보여주었다.

반면 차량크기는 BI 청구 건수에서는 중형차량이 청구 건수가 가장 적었으나 PD 청구 건수는 대형차량이 가장 적은 것으로 나타나는 등 차이를 보였다. 또한 차량연식도 BI는 연식이 증가함에 따라 청구 건수가 증가하다가 감소하나 PD는 지속적으로 감소하는 추세를 보였다.

〈Table 3〉 Coefficient estimates of Poisson model

Variables(l)	BI (j=1)		PD (j=2)	
	Estimate ($\alpha_l^{(1)}$)	Standard error	Estimate ($\alpha_l^{(2)}$)	Standard error
(Intercept)	-1.8210***	0.0398	-0.8575***	0.0221
Gender(1)	0.1049***	0.0060	0.1260***	0.0033
YDE(1)	-0.2866***	0.0120	-0.3119***	0.0068
YDE(2)	-0.3619***	0.0126	-0.3745***	0.0071
YDE(3)	-0.4602***	0.0091	-0.4390***	0.0051
BM(1)	-0.5574***	0.0092	-0.3873***	0.0054
AutoSize(1)	-0.0188**	0.0068	0.0287***	0.0038
AutoSize(2)	-0.0915***	0.0076	0.0531***	0.0041
AutoSize(3)	-0.0424***	0.0098	-0.1913***	0.0057
Age	-0.0138***	0.0016	-0.0086***	0.0009
Age^2	0.0002***	0.0000	0.0002***	0.0000
AutoAge	0.0277***	0.0018	-0.0053***	0.0010
$AutoAge^2$	-0.0018***	0.0001	-0.0011***	0.0001

Note: Significance '***' 0.001, '**' 0.01, '*' 0.05
Base: Gender - Male(0)
YDE(years of driving experience) - Under 3(0)
BM(bonus-malus) - Malus(0)
AutoSize - Light(0)

〈Table 4〉 Contingency table for BI and PD claims

Frequency of PD claims	Frequency of BI claims					
	0	1	2	3	4	5+
0	2,493,835	21,719	235	7	0	0
1	271,039	89,816	1,380	21	2	0
2	28,412	15,814	3,164	94	5	0
3	2,996	2,286	815	140	3	2
4	389	451	179	46	9	4
5+	84	132	90	36	14	14

〈Table 5〉 Measures of association between frequencies of BI and PD claims

Item	Pearson correlation	Kendall's τ	Spearman's ρ
Estimate	0.4626	0.4421	0.4457
p-value	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$

식 (7)과 (9)에 명시된 제안 모델은 과대 산포를 갖는 다중 위험 보험의 청구 건수 사이에 연관성이 존재한다고 가정하고 있다. 따라서 단순한 독립 포아송 모델보다 더 복잡한 모델을 적용하는 것이 적절한지 검증하기 위해서는 실제로 그 의존성이 존재하는지 확인해야 한다. 〈Table 4〉는 BI와 PD의 청구에 대한 빈도 테이블을 나타낸다. 이 표를 통해 두 청구 건수 간의 의존성을 추측해 볼 수 있다. 예를 들어 BI 건수가 1건일 때 PD 건수가 1건인 경우의 빈도가 가장 높고 BI 건수가 2건, 3건의 경우도 유사한 형태를 보인다. 〈Table 5〉는 이러한 두 담보 간의 연관성을 계량적으로 확인하기 위해 일반적인 측정 방법인 Pearson correlation, Kendall's τ, Spearman's ρ를 산출한 결과이다. 이를 통해 적어도 공변량의 영향을 제어하지 않을 때에는 BI 및 PD 청구 건수 간에 유의한 양의 상관관계가 있음을 확인할 수 있다.

quasi-Poisson 분포에서 산포 모수를 추정하는 방법을 제시하는 몇 가지 연구가 있다. 예를 들어 McCullagh and Nelder(1989)는 $\hat{\phi}=\frac{1}{m-p}\sum_{k=1}^{m}\frac{\left(n_k-\hat{\nu}_k\right)^2}{\hat{\nu}_k}$을 제안하였고 Cameron and Trivedi(1990)는 $\tilde{\phi}=\frac{1}{m}\sum_{k=1}^{m}\tilde{R}_k$를 제안하였다. 여기서

$\widetilde{R}_k = \dfrac{(n_k - \hat{\nu}_k)^2 - n_k}{\hat{\nu}_k}$ 이고, m은 총 관측 수, p는 추정된 모수의 개수, n_k는 k번째 관측의 청구 건수, $\hat{\nu}_k$는 모델에서 추정된 모수를 가지고 산출한 k번째 관측의 청구 건수 기댓값이다. 위 추정치가 1보다 크다면 과대산포가 존재한다고 볼 수 있다.

〈Table 6〉은 이를 산출한 결과인데 BI, PD 청구 건수의 $\hat{\phi}$, $\tilde{\phi}$ 모두 산포 모수가 1 이상으로 추정되는 것을 확인할 수 있다. 따라서 분석 대상인 청구 데이터에는 과대산포가 존재한다고 볼 수 있다.

〈Table 6〉 Estimates of overdispersion parameters $\hat{\phi}$ and $\tilde{\phi}$ for BI and PD claims

Overdispersion parameter	BI	PD
$\hat{\phi}$	1.0422	1.0807
$\tilde{\phi}$	1.0402	1.1069

그리고 Cameron and Trivedi(1990)는 m이 충분히 크고 $k=1, \cdots, m$에 대해 n_k가 독립일 때, 다음이 성립함을 보였다.

$$H_0 : \phi = 1 \text{하에서 } \hat{Z} = \frac{\tilde{\phi}}{\sqrt{\frac{1}{m}\sum_{k=1}^{m}\left(\widetilde{R}_k - \tilde{\phi}\right)^2}} \simeq N(0,1). \tag{27}$$

주어진 데이터에서 BI 및 PD 청구에 대한 $\hat{Z}$ 값은 각각 21.0456과 66.4850이다. 따라서 두 담보의 청구에 과대 산포가 없다는 귀무가설을 유의하게 기각할 수 있으므로 과대 산포가 존재한다고 볼 수 있다.

위와 같이 데이터에서 과대 산포의 유의성과 위험 간의 의존성을 확인하였으므로 이를 포함하는 제안 모델을 적용하여 요율을 산출한다.

먼저 $w^{(j)}$의 값은 다중 위험의 특성에 대한 사전 지식이나 관측된 데이터의 통계량으로 결정해야 한다. 식 (24)에서 확인할 수 있듯이 신뢰도 요인의 합동 추정치는 예측 빈도의 가중합(weighted sum)에 대한 실제 빈도의 가중합의 비율로 도출된다. 따라서 $w^{(j)}$는 특

정 위험에 대한 계수로 해석될 수 있다. 이 계수는 위험과 관찰되지 않은 이질성 사이의 연관성에 대한 정보를 제공한다.

예를 들면 자동차 보험에서 대물보상 담보와 자기차량손해 담보가 있다. 대물보상 담보는 보험 계약자의 과실이 인정되는 자동차 사고로 인한 손해를 보상하는 반면 자기차량손해 담보는 홍수와 같은 천재지변에 의해 침수가 되는 경우에 대해서도 보상한다. 그러므로 이 두 담보를 분석한다면 자기차량손해 담보에 비해 대물보상 위험이 관찰되지 않은 이질성과 비교적 큰 양의 상관관계가 있다고 볼 수 있다. 이와 관련하여, 신뢰도 계수의 합동 추정치를 산출할 때 자기차량손해 담보의 청구이력보다 대물보상 담보의 과거 청구이력에 더 많은 가중치를 두는 것이 합리적이다. 즉, 대물보상 담보의 $w^{(j)}$를 자기차량손해 담보의 $w^{(j)}$보다 높게 설정한다는 것이다. 따라서 제안된 방법은 실무자가 모델링 과정에서 다중 위험의 특성에 대한 사전 지식을 유연하게 반영할 수 있도록 한다.

사전 지식으로 $w^{(j)}$의 값을 미리 지정할 수 있는 강력한 근거가 부족한 경우에는 과대산포 모수의 추정치를 활용하고 적률법을 사용할 수 있다. 제안 모델에서 $\sum_{t=1}^{T_i}\left(N_{it}^{(j)}-\nu_{it}^{(j)}\right)^2$의 기댓값은 $j=1,\cdots,J$와 $i=1,\cdots,I$에 대해 $\sum_{t=1}^{T_i}\frac{\nu_{it}}{w^{(j)}}+\frac{\left(\nu_{it}^{(j)}\right)^2}{r}$이므로 식 (14)과 유사한 방식으로 $w^{(j)}$의 일치추정량을 구하면 다음과 같다.

$$\hat{w}^{(j)}=\frac{\sum_{i=1}^{I}\sum_{t=1}^{T_i}\hat{\nu}_{it}^{(j)}}{\sum_{i=1}^{I}\sum_{t=1}^{T_i}\left[\left(N_{it}^{(j)}-\hat{\nu}_{it}^{(j)}\right)^2-\frac{\left(\hat{\nu}_{it}^{(j)}\right)^2}{\hat{r}}\right]}. \tag{28}$$

$j=1,2$에 대한 과대 산포 모수 $w^{(j)}$가 사전 지식이나 적률법을 통해 결정되면 식 (25)와 (26)을 이용하여 공유된 임의효과를 포함한 결합 우도(joint likelihood)로부터 $\alpha^{(1)}$와 $\alpha^{(2)}$를 추정할 수 있다. 본 연구의 훈련자료에 적률법을 사용한 결과 r, $w^{(1)}$, $w^{(2)}$가 각각 6.1330, 0.9663, 0.9150으로 추정되었다. 따라서 $w^{(1)} > w^{(2)}$이기 때문에 제안된 사후적 요율결정 체계에서 상대적 기여도는 BI 빈도가 PD 빈도보다 높은 것으로 나타났다.

제안 모델의 회귀계수인 $\alpha^{(j)}$를 추정한 결과는 〈Table 7〉과 같다. 이 결과는 포아송 회

귀분석 결과인 ⟨Table 3⟩과 유사하게 나타났다.

⟨Table 7⟩ Coefficient estimates of proposed model

Variables(l)	BI (j=1)		PD (j=2)	
	Estimate($\alpha_l^{(1)}$)	Standard error	Estimate($\alpha_l^{(2)}$)	Standard error
(Intercept)	-1.8242***	0.0398	-0.8548***	0.0221
Gender(1)	0.1079***	0.0060	0.1252***	0.0033
YDE(1)	-0.2758***	0.0120	-0.3046***	0.0068
YDE(2)	-0.3525***	0.0126	-0.3713***	0.0071
YDE(3)	-0.4565***	0.0091	-0.4349***	0.0051
BM(1)	-0.5476***	0.0092	-0.3776***	0.0054
AutoSize(1)	-0.0182	0.0068	0.0296***	0.0038
AutoSize(2)	-0.0891***	0.0076	0.0525***	0.0041
AutoSize(3)	-0.0348*	0.0098	-0.1887***	0.0057
Age	-0.0139***	0.0016	-0.0086***	0.0009
Age^2	0.0002***	0.0000	0.0002***	0.0000
AutoAge	0.0271***	0.0018	-0.0051***	0.0010
AutoAge^2	-0.0018***	0.0001	-0.0011***	0.0001

Note: Significance '***' 0.001, '**' 0.01, '*' 0.05
Base: Gender - Male(0)
YDE(years of driving experience) - Under 3(0)
BM(bonus-malus) - Malus(0)
AutoSize - Light(0)

모델의 모든 계수가 추정되었으므로 이를 청구 빈도의 예측에 적용해볼 수 있다. 연구방법에서 사후분포의 기댓값을 식 (22)와 같이 도출하였으므로 이를 이용한다. $N^{(j)}{}_{T+1}$에 대한 예측치는 사후적 보험료를 산출한다는 의미이므로 예측분포의 기댓값 즉, $\hat{N}^{(j)}_{T+1} := E\left[N^{(j)}_{T+1}|\mathrm{n}^{(1)}_T, \mathrm{n}^{(2)}_T\right]$로 추정하는 것이 타당하다. 이렇게 산출한 결과는 표본외(out-of-sample) 검증을 통하여 2019년 시험자료의 실제 청구 건수와 비교하여 예측 성능을 확인해볼 수 있다. 성능을 비교하기 위하여 독립 포아송 모델(naive Poisson)과 독립 할인할증(univariate credibility)을 적용한 보험료도 함께 산출하였다.

독립 포아송 모델은 과거 사고이력을 반영하지 않은 방식으로 제안 모델에서 $w^{(1)} = w^{(2)} = 1$이고 $r = \infty$ 인 특별한 경우이다. 독립 할인할증 보험료는 BI, PD의 사

고 이력을 각각 반영한 보험료이다. 제안 모델(multivariate credibility)과 비교하면 $j=1$일 때에는 $w^{(1)}=1$, $w^{(2)}=0$이고 $j=2$일 때에는 $w^{(1)}=0$, $w^{(2)}=1$인 경우이다. 이를 정리하여 나타내면 ⟨Table 8⟩과 같다.

⟨Table 8⟩ Predictive means of three models for out-of-sample validation

Model	Predictive mean
Naive Poisson	$E\left[N_{T+1}^{(j)} \vert \mathrm{n}_{\mathrm{T}}^{(1)}, \mathrm{n}_{\mathrm{T}}^{(2)}\right] = \mathrm{E}\left[\mathrm{N}_{\mathrm{T}+1}^{(\mathrm{j})}\right] = e^{\mathbf{x}_{T+1}\alpha^{(j)}}$
Univariate credibility	$E\left[N_{T+1}^{(j)} \vert \mathrm{n}_{\mathrm{T}}^{(1)}, \mathrm{n}_{\mathrm{T}}^{(2)}\right] = \dfrac{r+\sum_{t=1}^{T} n_t^{(j)}}{r+\sum_{t=1}^{T} \nu_t^{(j)}} \cdot e^{\mathbf{x}_{T+1}\alpha^{(j)}}$
Multivariate credibility	$E\left[N_{T+1}^{(j)} \vert \mathrm{n}_{\mathrm{T}}^{(1)}, \mathrm{n}_{\mathrm{T}}^{(2)}\right] = \dfrac{r+w^{(1)}\sum_{t=1}^{T} n_t^{(1)}+w^{(2)}\sum_{t=1}^{T} n_t^{(2)}}{r+w^{(1)}\sum_{t=1}^{T} \nu_t^{(1)}+w^{(2)}\sum_{t=1}^{T} \nu_t^{(2)}} \cdot e^{\mathbf{x}_{T+1}\alpha^{(j)}}$

미래 청구 빈도에 대한 예측치로서 산출된 각 보험료의 예측력을 평가하기 위해 다음과 같이 정의된 제곱근평균제곱오차(root mean square error, 이하, ‘RMSE’라 함)와 평균절대오차(mean absolute error, 이하, ‘MAE’라 함)를 사용하였다.

$$\mathrm{RMSE} = \sqrt{\frac{1}{2I}\sum_{j=1}^{2}\sum_{i=1}^{I}\left(N_{i,T_i+1}^{(j)} - \hat{N}_{i,T_i+1}^{(j)}\right)^2}, \tag{29}$$

$$\mathrm{MAE} = \frac{1}{2I}\sum_{j=1}^{2}\sum_{i=1}^{I}\left|N_{i,T_i+1}^{(j)} - \hat{N}_{i,T_i+1}^{(j)}\right|. \tag{30}$$

여기서 I는 검증 세트에서 관찰된 보험 계약자 수를 의미한다. 세 가지 모델에 대해 산출한 위 통계량을 ⟨Table 9⟩에 나타내었는데, 세 모형의 RMSE와 MAE가 모두 커서 빈도를 정확히 예측하는 데에는 한계를 가짐을 알 수 있다. RMSE와 MAE는 오차를 나타내는 지표이므로 일반적으로 값이 더 작은 모델이 선호되는데, 제안 모델의 RMSE와 MAE가 모두 가장 작은 값을 보여 세 모델 중에서는 가장 좋은 예측력을 보였다. 좀 더 정확히 모형

들 간 비교를 하기 위하여 BIC(Bayesian information criterion)를 산출한 후 〈Table 9〉에 나타내었는데, 제안모형의 BIC가 가장 낮아서 다른 모형들에 비해 근소하게 더 나은 것으로 보인다.

〈Table 9〉 Out-of-sample validation for frequency prediction

Statistic	Naive Poisson	Univariate credibility	Multivariate credibility
RMSE	0.31496	0.31417	0.31373
MAE	0.16955	0.16814	0.16807
BIC	3,797,314	4,138,017	3,714,699

V. 결론

한 보험계약에서 서로 다른 위험에 의한 사고를 보상하는 다중 위험 보험은 매우 흔한 보험 상품이다. 한 사고로 인하여 복수의 위험으로부터 손해가 발생할 수 있으므로 각 위험으로 인한 청구 건수가 서로 의존성을 가질 수 있다. 또한 보험회사가 여러 해 동안 동일한 보험 계약자를 반복적으로 관찰하는 것은 매우 당연한 일이다. 따라서 보험회사는 계약자의 과거 청구 이력에 대한 종단면 데이터를 자연스럽게 보유하게 되므로 이를 보험요율에 적절히 반영할 수 있어야 할 것이다. 특히 자동차보험과 일반손해보험은 보험기간이 짧고 보험기간이 종료되면 갱신 또는 재가입 형태로 새로운 요율을 적용하는 보험 상품이므로 계약자의 과거 청구 이력을 보험요율에 반영하는 할인할증제도와 신뢰도이론을 활용할 수 있을 것이다.

본 연구에서는 이러한 특성을 고려하여 다중 위험 보험의 각 위험에 내재된 특성을 의미하는 고정효과와 시간의 흐름에 걸쳐 공유된 임의효과를 통합하면서 다중 위험의 청구 건간에 자연스러운 의존 구조를 반영한 요율산출 모델을 제안하였다. 실제 청구 데이터를 확인한 결과 과대산포가 존재하는 것을 확인하여 이를 반영하기 위해 빈도 모델분포로 quasi-Poisson 분포를 채택하였다. 임의효과의 사전분포는 켤레 사전분포의 성질을 이용

하기 위하여 감마분포를 가정하였고, 이로부터 자연스럽게 감마분포인 사후분포를 도출하였다. 최종적으로 예측분포는 음이항 분포임을 유도하였다. 다중 위험 보험의 위험 간 의존성, 관찰된 이질성 및 관찰되지 않은 이질성, 그리고 과대산포 건수 데이터의 분포에 대해서는 다수의 선행연구가 존재하지만 이를 모두 포함하는 연구는 거의 이루어지지 않았다. 본 연구는 이를 모두 반영한 요율산출 모델을 제시하였다는 점에 의의가 있다. 또한, 예측분포가 음이항 분포이므로 제안 모델은 닫힌 형태(closed form)의 다중 위험 보험료 공식을 제공할 수 있고 실무적으로 적용과 해석이 용이하다는 장점이 있다.

본 연구에서는 대인보상 및 대물보상 담보의 청구 건수 데이터를 훈련자료로 사용하여 제안 모델의 계수를 산출하였고 이를 검증하기 위하여 시험자료로 표본외 검증을 수행하였다. 그 결과 청구 빈도 예측에 있어 기존 방법보다 우수한 결과를 보이는 것으로 나타났다. 이 결과는 제안 모델이 보다 적절한 보험요율을 산출할 수 있다는 것을 의미한다. 그러나 이 모델을 실제 우리나라 자동차보험의 보험료 산출에 사용하기에는 한계가 있다. 우리나라의 자동차보험은 의무보험으로서 감독당국에 의한 규제를 많이 받기 때문에 할인할증 제도 역시 정해진 규정에 따라 적용해야 한다. 따라서 제안 모델을 그대로 실무적으로 적용하는 데에는 무리가 있을 수 있으나, 요율의 적정성을 검증하거나 회사 내부적인 분석 목적으로는 유용하게 활용할 수 있을 것이다.

제안 모델에서는 빈도 데이터의 과대산포를 반영하기 위해 quasi-Poisson 분포를 모델 분포로 적용하였다. 이 분포는 과대산포를 적용하면서도 켤레 사전분포를 이용할 수 있고 그 결과 예측 분포가 음이항 분포를 따른다는 해석적인 장점을 가지고 있어 모델 분포로 채택하였다. 최종후·고인미·전수영(2011)에 따르면 과대산포를 고려한 빈도분포에는 음이항(negative binomial), 제로팽창 포아송(zero-inflated Poisson), 제로팽창 음이항(zero-inflated negative binomial) 분포 등 여러 가지가 있고 이에 대한 선행연구도 다수 존재한다. 따라서 빈도 분포로써 quasi-Poisson 분포가 아닌 다른 분포를 본 연구의 제안 모델에 적용하는 것을 시도해볼 수 있다. 또한 본 연구의 분석에서는 계약자 모두에게 동일한 관찰기간 동안 발생한 과거 청구 이력에 대하여 모두 동일한 가중치를 모델에 반영하였는데, 할인할증 요율 산출에 가장 적합한 관찰기간을 계약자집단별로 다르게 선택한 후 그에 따라 서로 다른 기간별 가중치를 모델에 반영하는 방향으로 분석을 발전시킬

수도 있을 것이다.

한편, 본 연구에서는 개인용 자동차보험의 주요 담보인 대인보상과 대물보상, 두 가지 담보만을 분석 대상으로 하였는데, 자동차보험에는 다른 용도와 여러 담보가 존재하므로 후속연구에서는 자동차상해나 자기신체사고와 같은 담보를 추가해서 분석대상 담보를 확장하여 본 연구모형의 적용가능성을 확인해 보는 것을 기대한다. 그리고 일반보험이나 인보험에도 다중 위험 보험이 다양하게 존재하기 때문에, 다른 보험상품으로까지 연구를 확장해보고 본 연구와 비교 및 대조해 보는 것 또한 기대한다.

참고문헌

기승도·김대환 (2009), “일반화선형모형(GLM)을 이용한 자동차보험 요율상대도 산출방법 연구”, **보험연구원 연구보고서,** 2009-05.

(Translated in English) Ki, S., and D., Kim (2009). “Automobile Insurance Ratemaking Using Generalized Linear Model(GLM)”, *Korea Insurance Research Institute*, 2009-05.

김동훈·이계연 (1998), “로짓모형을 이용한 자동차보험 사고확률 추정”, **리스크관리연구**, 제10권 제1호, pp. 75-112.

(Translated in English) Kim, D., and K., Lee (1998). “An Estimation of Automobile Insurance Accident Probability by Logit Model”, *The Journal of Risk Management*, 10(1):75-112.

김영화·김미정·김명준 (2011), “여러가지 신뢰도에 근거한 자동차 보험료 예측”, **응용통계연구 논문집**, 제24권 제2호, pp. 279-292.

(Translated in English) Kim, Y., M., Kim and M., Kim (2011) “Estimating the Automobile Insurance Premium Based on Credibilities”, *The Korean Journal of Applied Statistics*, 24(2):279-292.

김영화·이현수 (2010), “신뢰도에 근거한 자동차 보험 가격 산출 비교”, **한국통계학회 논문집**, 제17권 제5호, pp. 713-724.

(Translated in English) Kim, Y., and H., Lee (2010). “A Comparison Study for the Pricing of Automobile Insurance Premium Based on Credibility”, *Communications for Statistical Applications and Methods,* 17(5):713-724.

김진한·민재형 (1994), “마코프체인을 이용한 자동차보험료의 할인·할증시스템에 관한 연구”, **한국경영과학회 학술대회논문집**, 1994년 제1권, pp. 79-95.

(Translated in English) Kim, J., and J., Min (1994). “An Exploratory Study on the Bonus-Malus System of Automobile Insurance Premium through the Method of Markov Chain", *Proceedings of the Korean Operations and Management Science Society Conference*, 1994, 1:79-95.

보험개발원 (2014), "자동차보험 할인할증제도 개선 정책토론회 개최", 2014. 2. 24.

(Translated in English) Korea Insurance Development Institute (2014). "Policy Discussion Session to Improve the Automobile Insurance Bonus-Malus Systems", 2014. 2. 24.

__________ (2022), 보험통계조회서비스, 2022. 3. 30. (Retrieved from https://incos.kidi.or.kr:5443).

(Translated in English) Korea Insurance Development Institute (2022). Insurance Statistics Consumer Service, 2022. 3. 30. (Retrieved from https://incos.kidi.or.kr:5443).

심현우·임형기·최양호 (2021), "부분적 청구경력 고지 간소화에 따른 건강보험 요율 차등화에 관한 연구", **보험금융연구**, 제32권 제2호, pp. 43-75.

(Translated in English) Shim, H., H., Im and Y., Choi (2021). "A Study on Premium Rate Differentiation of Health Insurance by Simplified Partial Disclosure of Claim History", *Journal of Insurance and Finance*, 32(2):43-75.

정중영·김태완 (2004), "자동차보험의 효율적인 할인·할증제도 분석 연구", **한국자료분석학회지**, 제6권 제6호, pp. 1665-1675.

(Translated in English) Jeong, J., and T., Kim (2004). "An Analysis of Adequate and Efficient Bonus-Malus System in the Korean Automobile Insurance", *Journal of The Korean Data Analysis Society* 6(6):1665-1675.

조재린·이기형·강중철 (2012), "신뢰도를 적용한 자동차보험요율산출 연구", **한국자료분석학회지**, 제14권 제4호, pp. 2267-2276.

(Translated in English) Cho, J., K., Lee and J., Kang (2012). "A Study on the

Pricing of Car Insurance with Credibility", *Journal of The Korean Data Analysis Society*, 14(4):2267-2276.

최종후·고인미·전수영 (2011), "제로팽창 모형을 이용한 보험데이터 분석", **응용통계연구**, 제24권 제3호, pp. 485-494.

(Translated in English) Choi, J., I., Ko and S., Cheon (2011). "A Zero-Inflated Model for Insurance Data", *The Korean Journal of Applied Statistics*, 24(3):485-494.

Bailey, A. (1950). "Credibility procedures: Laplace's generalization of Bayes' rule and the combination of collateral knowledge with observed data", *Proceedings of the Casualty Actuarial Society*, 37:7-23.

Bermúdez, L., and D., Karlis (2017). "A posteriori ratemaking using bivariate Poisson models", *Scandinavian Actuarial Journal*, 2017(2):148-158.

Boucher, J.-P., and M., Denuit (2008). "Credibility premiums for the zero-inflated Poisson model and new hunger for bonus interpretation", *Insurance: Mathematics and Economics*, 42(2):727-735.

Boucher, J.-P., M., Denuit and M., Guillen (2008). "Models of insurance claim counts with time dependence based on generalization of Poisson and negative binomial distributions", *Variance*, 2(1):135-162.

________________________ (2009). "Number of accidents or number of claims? An approach with zero-inflated Poisson models for panel data", *Journal of Risk and Insurance*, 76(4):821-846.

Bühlmann, H. (1967). "Experience rating and credibility", *ASTIN Bulletin: The Journal of the IAA*, 4(3):199-207.

Cameron, A., and P., Trivedi (1990). "Regression-based tests for overdispersion in the Poisson model", *Journal of Econometrics,* 46(3):347-364.

e Silva, J., and M., de Lourdes Centeno (2017). "Ratemaking of dependent risks", *ASTIN Bulletin: The Journal of the IAA*, 47(3):875-894.

Frangos, N., and S., Vrontos (2001). "Design of optimal bonus-malus systems with a frequency and a severity component on an individual basis in automobile insurance", *ASTIN Bulletin: The Journal of the IAA*, 31(1):1-22.

Frees, E. (2003). "Multivariate credibility for aggregate loss models", *North American Actuarial Journal*, 7(1):13-37.

Frees, E., G., Lee and L., Yang (2016). "Multivariate frequency-severity regression models in insurance", *Risks*, 4(1):4.

Frees, E., G., Meyers and A., Cummings (2010). "Dependent multi-peril ratemaking models", *ASTIN Bulletin: The Journal of the IAA*, 40(2):699-726.

Frees, E., V., Young and Y., Luo (1999). "A longitudinal data analysis interpretation of credibility models", *Insurance: Mathematics and Economics*, 24(3):229-247.

Gómez-Déniz, E., and F., Vázquez-Polo (2005). "Modelling uncertainty in insurance bonus-malus premium principles by using a bayesian robustness approach", *Journal of Applied Statistics*, 32(7):771-784.

Jeffreys, H. (1946). "An invariant form for the prior probability in estimation problems", *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007):453-461.

Jeong, H. (2020). "Testing for random effects in compound risk models via Bregman divergence", *ASTIN Bulletin: The Journal of the IAA*, 50(3): 777-798.

Jeong, H., and E., Valdez (2020a). "Bayesian credibility premium with GB2 copulas", *Dependence Modeling*, 8(1):157-171.

______________________ (2020b). "Predictive compound risk models with dependence", *Insurance: Mathematics and Economics*, 94:182-195.

Jewell, W. (1974). "Credible means are exact bayesian for exponential families", *ASTIN Bulletin: The Journal of the IAA*, 8(1):77-90.

Klugman, S., H., Panjer and G., Willmot (2012). *Loss Models: From Data to Decisions*, 4th Edition, Wiley.

Lemaire, J. (1998). "Bonus-malus systems: the european and asian approach to merit-rating", *North American Actuarial Journal*, 2(1):26-38.

Liang, K.-Y., and S., Zeger (1986). "Longitudinal data analysis using generalized linear models", *Biometrika*, 73(1):13-22.

Mayerson, A. (1964). "A bayesian view of credibility", *Proceedings of the Casualty Actuarial Society*, 51:7-23.

McCullagh, P., and J., Nelder (1989). *Generalized linear models,* 2nd Edition, Chapman and Hall.

Norberg, R. (1986). "Hierarchical credibility: analysis of a random effect linear model with nested classification", *Scandinavian Actuarial Journal*, 1986(3-4):204-222.

Pechon, F., M., Denuit and J., Trufin (2019). "Multivariate modelling of multiple guarantees in motor insurance of a household", *European Actuarial Journal*, 9(2):575-602.

_________________________________ (2020). "Home and motor insurance joined at a household level using multivariate credibility", *Annals of Actuarial Science*, 2020:1-33.

Pechon, F., J., Trufin and M., Denuit (2018). "Multivariate modelling of household claim frequencies in motor third-party liability insurance", *ASTIN Bulletin: The Journal of the IAA*, 48(3):969-993.

Quan, Z., and E., Valdez (2018). "Predictive analytics of insurance claims using multivariate decision trees", *Dependence Modeling*, 6(1):377-407.

Sutradhar, B., and V., Jowaheer (2003). "On familial longitudinal poisson mixed models with gamma random effects", *Journal of multivariate analysis*, 87(2):398-412.

Wedderburn, R. (1974). "Quasi-likelihood functions, generalized linear models, and the gauss-newton method", *Biometrika*, 61(3):439-447.

Whitney, A. (1918). "The theory of experience rating", Proceedings of the Casualty Actuarial Society, 4:274-292.

Yang, L., and P., Shi (2019). "Multiperil ratemaking for property insurance using longitudinal data", *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(2):647-668.

Abstract

Multi-peril insurance consists of multiple coverages which compensate for accidents caused by various perils. In multi-peril insurance, there may be dependence among the claims from different perils. When dependence is confirmed, it is reasonable to consider the correlation among perils in ratemaking. In this study, we confirmed such dependence and proposed a shared random effects model to capture and reflect this in posterior ratemaking. The model can provide closed forms of credibility premium, which makes it easy to implement in actuarial practice.

We calibrated the model using the actual claim data of bodily injury and property damage of automobile insurance as an example of multi-peril insurance. Then we used the model for posterior ratemaking and evaluated its performance by out-of-sample validation. The results showed that the performance of the proposed model is superior to that of comparative models such as the naive Poisson model.

※ Key words: Multi-peril Insurance, Shared Random Effects, Overdispersion, Claim Frequency Analysis, Bonus-malus System