Journal of the Korean Data Analysis Society (February 2017)
Vol. 19, No. 1 (B), pp. 239-253

자동차보험 손해 진전 계수 예측력 향상 방식에 관한 연구*

김명준[1], 한소율[2], 김영화[3]

요 약

자동차 사고 발생에 의한 손해 규모는 사고 시점에 결정되는 사항이 아니며 시간 경과에 따라 그 규모가 변하는 특성을 가지고 있다. 이는 치료의 장기화 뿐만 아니라, 물가상승, 정책의 변화 등과 같은 이유로 지급기준이 변경되는 것에 기인한다. 기존의 연구가 사고 추세 예측 방법에 관심이 집중되는 것에 비하여 추세 연구의 근간이 되는 손해 데이터의 추정과 관련된 연구는 미미한 실정이다. 본 연구에서는 이러한 문제점을 보완하여 추세 예측의 근간이 되는 최종 손해 수준을 합리적으로 추정하는 방법을 제안하고자 한다. 자동차 보험에는 물적 담보와 인적 담보가 존재하며, 본 논문에서는 손해 규모의 변동이 상대적으로 크게 발생하는 인적 담보 중 하나인 대인 담보를 분석 대상으로 하며, 시간이 경과함에 따라 변화하는 누적 손해 자료를 사용하여 최종 손해 규모 수준을 적정하게 추정하는 방식 제안을 목적으로 한다. 또한 실증 자료 분석을 통하여 본 논문에서 제안하는 모형과 현재 보험회사 현업에서 사용하는 방식을 비교함으로써 추정 방식의 적합성과 효용성 검정을 실시하여 제시하였다.

주요용어 : 손해액, 손해 진전 계수, 추세 예측, run-off triangle.

1. 서론

자동차보험은 보험 가입이 필수인 의무보험이라는 특징과 1년마다 갱신되는 단기성 소멸 계약이라는 특징을 가지고 있으며, 판매되는 보험 상품 가운데 보험 가입자의 수가 가장 많고 보험료의 변동이 매년 발생하는 보험이다(Kim, 2013). 의무보험인 자동차 보험료 인상은 장기적인 경기침체로 가계의 어려움을 겪고 있는 서민들의 부담을 가중시킬 수밖에 없고, 따라서 자동차보험 가입자는 보험료의 증감에 큰 관심을 가지게 된다. 보험개발원에 따르면 현재 국내 자동차 등록대수는 최근 5년간 매년 40만 대 이상씩 증가하여 연평균 3.3%의 증가율을 보이고 있으며, 2015년 기준 20,990천 대를 기록하고 있다. 또한, 현재 자동차보험 가입자는 2천만 명에 이르고 있지만 자동차 보험료 산정이나 보장에 대한 불만이 끊이질 않고 있어 금융감독원이나 보험사는 계속적으로 대책 마련을 위해 노력하고 있다는 뉴스가 보도되기도 하였다. 손해보험사의 손해율 악화로 인하여 자동차 보험료가 인상되면 자동차보험 가입자들의 불만이 커질 수밖에 없다. 따라서 금융감독원은 물가 억제 정책 등의 기조로 소비자 물가지수에 포함되어 있는 자동차 보험료 인상을 사실상 통제하고 있다. 그 결과 손해보험사는 자동차보험 손해율의 악화로 인하여 경영상의 많은 어려움을 겪고 있으며 다른 보험 분야의 양호한 실적으로 자동차보험의 손해를 메꾸는 불합리한 상황이 지속되고 있다. 손해율은 보험가입 고객으로부터 받은 보험료와 사고발생으로 인하여 지출한 보험금의 비율을 말하는데, 손해보험사의 보험영업 손익의 결정적 요소이며 사업 비율과 함께 손해보험사의

---

*이 논문은 2016년도 중앙대학교 연구년 결과물로 제출됨.
[1]34430 대전광역시 대덕구 한남로 70, 한남대학교 비즈니스통계학과 조교수. E-mail : mkim@hnu.kr
[2]06974 서울시 동작구 흑석로 84, 중앙대학교 대학원 통계학과 석사과정. E-mail : soyul5458@naver.com
[3](교신저자) 06974 서울시 동작구 흑석로 84, 중앙대학교 응용통계학과 교수. E-mail : gogators@cau.ac.kr
[접수 2016년 12월 28일; 수정 2017년 2월 4일, 2017년 2월 17일; 게재확정 2017년 2월 20일]

경영을 분석하고 평가하는 중요한 지표 중 대표적인 것이다. 손해율이 높을수록 손해보험사의 손실이 크다는 것을 의미하는데, 대표적으로 손해보험업계 1위인 삼성화재는 2015년 월별 평균 손해율이 80% 안팎을 기록 중이고 다른 대형 손해보험사들도 월별 손해율이 90%를 넘는다는 뉴스도 있었다. 경영실적 개선을 위해 손해보험사들이 모집 비용과 인건비, 관리비 등 사업비를 줄이는 긴축 경영을 하고 있음에도 불구하고 자동차보험 손해율의 악화는 자동차 보험료 인상을 피할 수 없게 한다.

보험료는 심도(severity)와 빈도(frequency)의 곱으로 나타낼 수 있으며, 여기서 심도란 사고 1건당 피해액을 말하며 빈도란 사건 발생률로 보험기간에 유효한 계약건수 중 사고 건수의 비율을 말한다(Kim, Park, 2013). 자동차 보험료 산정은 빈도와 심도를 어떻게 예측하여 자료에 적합 시킬 것인가에 대해 중점을 둔다. 자료를 적합시키는 모형에 따라 보험료 책정이 크게 다르게 나타나기 때문이다. 과거에 발생한 사고의 특성 파악 및 추세 분석을 통해 미래 사고를 예측하여 보험료를 결정해야 하며, 통계적 기법을 활용하여 이를 수행하게 된다. 현재까지 관련된 연구는 주로 교통사고 발생 건수, 사망률 등에 초점을 두어 진행되어 왔는데, Lee(2010)는 음이항 혼합모형을 이용하여 교통사고 발생 건수에 대한 통계적 모형을 제시하였고, Jeong, Choi(2014)는 사고 건수를 빈도모형으로 분석하였으며, Kang, Heo(2015)는 일반화 포아송 모형을 이용하여 교통사고 모형을 구축하고 상대 위험도를 추정하였으며, Shim, Choi(2016)는 일반화 선형 모형에 근거한 연령-기간 모형을 통해 교통사고 사망률을 예측하였다.

보험료를 산출하는 두 가지 측면으로는 미래 예측, 위험도 구분이 존재한다. 미래 예측이란 과거의 사고 기록을 바탕으로 미래를 예측하는 방법이며, 위험도 구분이란 운전자의 성별, 자동차 크기, 가입 경력, 사고 유무 등 현재 데이터를 사용하여 상대적 비교를 통해 분석하는 방법이다. 후자의 경우 고객별 보험료의 변동이 발생하지만 보험회사 순보험료의 총액은 지급보험금의 총액과 일치되어야 한다. 이를 수지상등의 원칙(the principle of equivalence; off-balance)이라 하며 보험업 감독규정에 명시되어 있다.

이러한 두 가지 측면 가운데 본 논문에서는 미래 예측에 초점을 맞추어 분석을 진행한다. 자동차 사고 발생에 의한 손해액은 한 시점에 결정되는 사항이 아니며 시간이 경과함에 따라 손해액 규모가 변화하는 속성을 가지고 있다. 자동차보험에는 물담보와 인담보가 존재한다. 물적 사고에 해당하는 대물, 차량 담보의 경우에는 시간 경과에 따라 손해액의 변동이 크지 않아 비교적 단기간에 확정되는 경향이 높다. 즉, 시간 경과에 따른 변동이 크지 않으므로 다양한 분석 기법의 적용에 대한 검토의 필요성이 상대적으로 낮다. 그러나 대인 사고나 자기 신체 사고의 경우 피해자 부상 정도에 따라 손해액이 확정되는 시기가 매우 다르게 나타나게 된다. 즉, 경미한 사고의 경우에는 비교적 단기간에 손해액이 확정되지만, 중상의 경우 치료 과정 중 발생하는 추가 치료의 필요성, 후유증 등으로 인한 추가 보험료의 발생으로 손해액의 변화가 크게 나타나 본 논문에서는 인담보 최종 손해액을 추정하는 연구를 진행하고자 한다.

보험료 자율화로 인해 인상되는 보험료에 대해 많은 소비자들이 불만을 토로하고 있다. 소비자와 손해보험사 사이에 합리적이고 적정한 보험료 책정을 위해서는 손해보험사와 소비자 모두의 노력이 필요하다. 손해보험사의 경우, 먼저 사업비를 통한 보험료 조정이 필요하다. 사업비란 회사 운영에 필요한 경비로 인건비 및 물가를 포함한다. 다음으로, 위험도 집단을 세분화하여 적합한 보험상품 개발이 필요하다. 마지막으로 정확한 보험료 예측이 필요하다. 자동차 보험료 예측에 있어 가장 중요한 이슈는 적절한 데이터를 예측 모형의 기반으로 활용하는 것이다. 자동차보험의 경우 물가 상승, 의료비 상승, 자동차보험사 직원들의 보상 실무 경험 증가, 정책의 변화 등과 같은 이

유로 보험금 지급기준이 변경되는 특성을 가지고 있다. 이를 반영하기 위해 최근값 또는 평균값 등을 사용할 경우 형평성 문제가 대두된다. 이 문제를 해결하기 위해 알려져 있는 과거의 정보를 이용하여 미래의 값을 예측하는 방법인 추세 분석이 필요하다. 추세 분석에 대한 연구는 Cook (1970)에 의하여 기본적인 골격이 제시되었고, 그 후 Cummins(1980)와 McCarthy(1998) 등에 의하여 적용 방안들이 발전되어 왔다. 또한 최근 시계열 방식을 적용한 연구가 Kim, Park(2013)에 의해 제안되었고 빈도 심도를 이용한 연구가 Kim(2013)에 의해 제안되었다.

다양한 추세 예측 연구에 비하여 추세 연구의 근간이 되는 적정한 손해액 데이터의 추정과 관련된 연구는 미미한 실정이다. 본 연구에서는 보험금 최종 손해액 수준을 합리적이게 추정하는 방법을 제시하고자 한다. 또한 실제 국내 보험사의 대인담보 손해액 자료를 사용하여 본 논문에서 제안하는 추정 방법과 현업에서 사용하는 산술평균의 적정성을 비교해 보고자 한다.

현재 보험회사에서는 보험금 지급 규모를 추정하기 위해서 손해 진전 계수(the loss development factor; LDF)라는 개념을 사용하고 있다. 따라서 손해 진전 계수를 적절하게 추정하는 것이 추세분석에 사용하는 올바른 데이터를 추정하는 기초 과정이 되며, 본 논문에서는 이에 대한 연구 결과를 제시하고자 한다. 즉, 대인 담보 손해액 실제 자료를 이용하여 손해 진전 계수를 추정하고, 이를 바탕으로 단순 회귀 모형과 다양한 지시함수를 활용하여 최종 손해액을 추정한 후 모형별 해당 결과의 적정성을 분석하고 최적의 방법을 제시하고자 한다. 본 연구의 결과는 현업에서 적절한 추세분석을 위한 데이터 구축에 즉각적인 활용이 가능하다는 점에서 본 연구의 의의를 둘 수 있다. 본 논문의 구성은 다음과 같다.

1장에서는 연구의 배경 및 목적에 대하여 설명하였다. 2장에서는 본 논문에서 보험금 지급 규모를 추정하기 위해 사용한 손해 진전 계수(LDF) 개념을 설명하고 단순 선형 회귀 모형, 절편 지시함수 모형, 절편 기울기 지시함수 모형, 특이점 지시함수 고려 모형을 소개하였다. 3장의 실증자료 분석에서는 국내 S 보험사의 실제 데이터를 사용하여 본 논문에서 제안하는 방법들을 이용하여 분석을 실시하고, 모형별 분석 결과를 제시하였다. 또한 모형별 적합성 검정을 통해 모형 우위 비교함으로써 모형의 효율을 확인하였다. 마지막 4장에서는 3장에서 제안한 여러 방법들의 검정 결과를 바탕으로 가장 효과적인 모형이 어떤 모형인지 정리하고, 현업에서 일률적으로 사용하는 산술평균과 본 연구에서 제안한 방법을 비교하고 정리 요약하여 논문을 마무리하였다.

2. 선행 연구

금융감독원에 따르면 일반적으로 자료 집계 기준의 종류는 사고 연도 기준(the accident year basis; AY), 계약 연도 기준(the underwriting year basis; UY), 회계 연도 기준(the fiscal year basis; FY)으로 분류된다. AY는 해당 연도에 발생한 사고로 인한 손해액은 계약이 체결된 연도나 보험금이 지급된 연도에 관계없이 AY 해당 연도의 손해액으로 집계되는 방식이다. 즉, AY 해당 연도의 사고에 대한 지급 보험금과 자료집계 시점에서 평가한 개별 추산치의 합계가 AY 해당 연도의 발생 손해액이다. UY는 해당 연도에 체결된 모든 보험계약에 대한 손해액을 집계하는 방법이다. UY 기준의 자료집계 방식은 일정 기간 중 계약 체결 건들에 대한 정확한 실적을 집계할 수 있다는 장점이 있으나, 자료 집계까지 상당 기간이 소요된다는 단점을 가지고 있다. FY는 해당 연도에 발생한 회계적 처리 사항은 모두 해당 회계 연도로 귀속시켜 자료를 집계하는 방식이다. FY 해당 연도의 발생 손해액은 특정 사고기간 또는 계약기간과 정확히 대응되지 않는 단점이 있다. 이러한 특성으로 인하여 금융감독원에서는 FY 자료를 분석 자료로 활용하는 것에는 부적합한 면이 존재한다고 지적하고 있으므로, 본 논문의 분석에서는 AY 손해액 데이터를 사용하였다.

2.1. 손해 진전 계수

손해 진전 계수(the loss development factor; LDF)는 최종 손해액을 추정하는 보험업계의 전통적인 방식으로, 현재 시점의 누적 지급 보험금을 직전 시점의 누적 지급 보험금으로 나눈 값으로서 기간이 경과함에 따라 지급 보험금이 어떻게 진전되는가를 나타내는 값이다. 현재 사용되는 LDF로는 산술평균, 기하평균, 절삭평균 등이 있다. 본 논문에서는 LDF를 이용하여 최종 손해액(the ultimate loss)을 추정하고자 한다. 추정된 최종 손해액을 이용하여 보험료 미래 예측분석을 실행하므로 이를 구성하는 데이터가 정확해야 한다.

분석의 첫 단계로 Table 1과 같은 데이터 형식을 Run-off triangle이라 칭하며 대인 손해액 자료를 사용하였다. 손해 진전 삼각형 형태의 자료에서 가장 오래된 사고 발생기간의 경우 자료 집계 최종 시점 이후에 진전되는 손해의 규모가 최종 발생 손해액 총액을 추정하는데 크게 영향을 주지 않을 정도로 작아야 하므로 충분히 긴 경과기간에 걸쳐 자료가 집계되어야 한다(Lee, Lee, 2006). 손해 진전 계수 산출 방법은 Table 1의 손해액 자료를 이용하여 사고 발생 기간별 손해 진전 계수(LDF)를 산출하면 Table 2와 같이 나타나게 된다. 손해 진전 계수(LDF)는 다음과 같이 정의된다.

$$\mathrm{LDF} = [x_{i,j+1}/x_{i,j}]$$

Table 1에 'Run-off triangle'이 예시되어 있으며 삼각형내의 값들은 기준 연도 i에 발생한 보험사고에 대하여 지급 연도 j년 말까지 지급된 누적 손해액을 의미한다. 여기서 기준 연도는 해당 연도에 발생한 사고로 인한 손해액으로 집계되는 시점(AY)을 의미한다(Jang, 1997).

Table 1. Run-off triangle

Accident year(i)	Age in year(j)			
	1	2	3	4
1	22,105	22,329	22,963	23,374
2	22,486	23,557	23,679	
3	30,143	29,240		
4	36,769			

여기서 22,105는 기준 연도 1년에 발생한 보험사고에 대하여 1년 동안에 집계된 누적 손해액을 의미한다. 삼각형의 대각선상의 값들은 각 AY의 실적을 나타낸다. 즉, 동일 대각선상의 값들은 해당 AY까지의 누적 손해액을 의미한다. 따라서 삼각형의 대각선 하단의 공란은 현 시점에서 알 수 없는 미래의 손해액을 나타내게 된다(Jang, 1997).

Table 2. LDF calculation

Accident year(i)	Age in year(j)		
	1-2	2-3	3-4
1	1.010	1.028	1.018
2	1.048	1.005	
3	0.970		

본 논문에 사용된 데이터는 연도 자료가 아닌 분기자료이므로 i는 사고분기, j는 진전분기를 의미한다. 최종 손해액을 산출하기 위해 사고 데이터를 삼각화하여 사용하는 기법에는 지급보험금 진전 추이방법(the paid loss development method; PLDM), 발생 손해액 진전 추이방법(the incurred loss development Method; ILDM), 평균 지급 보험금방법 (the average payment method; APM), 본휴에

터-퍼거슨 방법 (the Bornhuetter-Ferguson method; B-F) 등이 있다(Oh, Lee, 2014). 본 논문에서는 국내 S 보험사의 실제 손해액 자료가 지급보험금 기준이므로 PLDM을 이용하였다.

2.2. 단순 선형회귀 모형(Model 1)

회귀분석은 반응변수가 설명변수들에 의해서 어떻게 설명 또는 예측되는지를 알아보기 위해 적절한 함수식으로 자료를 표현하여 분석하는 통계 방법이다. 본 연구에서 설명변수(X)는 시간, 반응변수(Y)는 추정하고자 하는 분기의 LDF값이다. 본 연구에서 사용된 모형의 식은 다음과 같다.

$$Y=\beta_0+\beta_1 X+\varepsilon$$

여기서 β_0와 β_1은 회귀계수(regression coefficient)라 불리는 상수(constant)이며, ε은 오차(error)이다. Y는 근사적으로 X의 선형함수이며, ε은 $N(0,\sigma^2)$을 따르는 확률변수로 가정한다.

2.3. 절편 지시함수 모형(Model 2)

자동차 사고에는 추세요인과 계절요인(겨울철 폭설 등으로 인한 사고 발생의 급격한 증가 등)이 혼합되어 나타나는 특성이 있기 때문에 분기별 지시함수를 활용하여 계절요인을 별도로 고려하는 모형을 본 논문의 분석방법에 적용하였으며 그 모형은 다음과 같다(Kim, 2013).

$$Y=\beta_0+\beta_1 X+\beta_2 I(t_1)+\beta_3 I(t_2)+\beta_4 I(t_3)$$
$$where\ \ I(t_i)=\begin{cases}1,\ \text{if}\ t_i=i^{th}\ quarter \\ 0,\ \text{if}\ t_i\neq i^{th}\ quarter\end{cases},\ \ i=1,2,3$$

여기서 Y는 추정하고자 하는 분기의 LDF 값이 되는 것이며,X는 시간의 진전에 따라 적용되는 변수로 추세요인에 대한 모수(parameter)를 추정하기 위한 설명변수이다. $I(t_i), i=1,2,3$는 각 해당 분기를 나타내는 지시함수로 1분기의 경우 모형에는 β_2만 남게 되고, 2분기에는 β_3가, 3분기에는 β_4만 포함되는 형태의 모형이며, 4분기에는 모든 지시함수가 없는 모형이 되어 전체 모형의 절편을 통하여 4분기의 효과를 추정하게 된다.

2.4. 절편-기울기 지시함수 모형(Model 3)

절편-기울기 지시함수 모형은 2.3절에서 설명한 절편 지시함수 모형에서 기울기를 추가하여 확장시킨 모형이다. 기울기를 고정하여 LDF 값을 예측하는 모형 보다 기울기를 추가하여 확장시킨 모형이 해당 분기의 LDF 값을 예측하는데 있어 더 효과적인 방식이라고 볼 수 있다. 제안 모형은 다음 식과 같다.

$$Y=\beta_0+\beta_1 X+\beta_2 I(t_1)+\beta_3 I(t_2)+\beta_4 I(t_3)+\beta_5 I(t_1)X+\beta_6 I(t_2)X+\beta_7 I(t_3)X$$
$$where\ \ I(t_i)=\begin{cases}1,\ \text{if}\ t_i=i^{th}\ quarter \\ 0,\ \text{if}\ t_i\neq i^{th}\ quarter\end{cases},\ \ i=1,2,3$$

2.5. 특이점 지시함수 고려 모형(Model 4)

자동차보험 최종 손해액의 예측력을 높이기 위해 모형을 점차 확장하여 분석을 실시하고자 한다. 여기서 특이점을 고려하는 이유는 국가 정책의 변화, 보상기준의 변화, 회사 운영방식의 변화

등으로 급격한 변동이 발생하는 경우 고려할 수 있다는 장점이 있다. 특이점을 고려한 모형은 다음과 같다.

$$Y=\beta_0+\beta_1X+\beta_2I(t_1)+\beta_3I(t_2)+\beta_4I(t_3)+\beta_5I(t_1)X+\beta_6I(t_2)X+\beta_7I(t_3)X+\beta_8S(Z)$$
$$where\ \ I(t_i)=\begin{cases}1,\ \text{if}\ t_i=i^{th}\ quarter\\0,\ \text{if}\ t_i\neq i^{th}\ quarter\end{cases},\ \ i=1,2,3$$
$$S(Z)=\begin{cases}1,\ \text{if}\ y_i>\mu+2\sigma\\0,\ \text{if}\ y_i\leq\mu+2\sigma\end{cases},\ \ i=1,2,...,54$$

여기서 $S(Z)$는 특이점 지시함수이다. 특이점 지시함수 $S(Z)$란 실제 손해액 자료를 이용하여 사고 발생 기간별 손해 진전 계수를 산출한 데이터 y_i 중에서 $\mu+2\sigma$를 초과하는 값을 특이점으로 판단하여 '1'로 지정하고 $\mu+2\sigma$이하인 값을 특이점이 아니라고 판단하여 '0'으로 지정하는 것을 의미한다. 본 연구에서는 특이점을 고려하여 최종 손해액을 예측하고자 한다.

3. 실증 자료 분석

3.1. 분석 개요

자동차보험은 기본적으로 5개의 기본 담보 (대인배상, 대물배상, 자기신체사고, 무보험차량에 의한 손해, 자기차량사고)로 이루어져 있으며, 대인배상은 보상한도에 따라 2개의 담보로 구분된다. 현재 국내에서는 대인배상과 대물배상 모두 담보에 대한 보험 가입이 의무화되어 있어 해당 담보에 대한 손해액 데이터를 사용하여 추세분석을 실시하고자 한다. 일반적으로 대물의 경우 대인에 비해 손해 진전이 작게 나타나므로 분석의 고도화에 대한 필요성이 요구되지 않는다. 반면 대인의 경우는 자동차 보험료 손해액이 확정되는 시기가 매우 다르게 나타나 비교적 변동이 크게 나타나므로 본 연구에서는 대인 담보 손해액 데이터를 사용하여 최종 손해액을 추정하는 연구를 실시하고자 한다.

먼저 최종 손해액 추정을 위해서는 LDF분석을 위한 적절한 데이터 확보가 필요하다. 각 사고 분기별, 경과기간별로 분류하여 손해 진전 삼각형(loss development triangle 또는 run-off triangle) 데이터를 사용하였고 최종 손해액 추정을 위한 기초자료로 사용하였다.

현재 손보사에서 최종 손해액을 추정하는 방법으로는 산술평균과, 기하평균을 이용하여 예측하고 있다. 본 논문에서는 추가적으로 단순 선형 회귀 모형, 절편 지시함수 모형, 절편-기울기 지시함수 모형, 특이점 지시함수 고려 모형 등 네 가지를 제안하고자 한다. 즉, 제안한 네 가지 모형하에서, 최종 손해액 예측력이 높은 방식이 무엇인지 살펴보기로 한다.

3.2. 데이터의 구성

본 연구 분석에 사용한 데이터는 국내 S 보험회사로부터 수집한 실제 자료이며 자료는 총 932,880개로 구성되어 있다. 그중 분석에 필요한 데이터는 대인, 대물 자료의 발생손해액이다. 과거 1993년 1분기부터 2006년 4분기까지 총 56분기의 경과기간 별 누적 손해액 자료를 이용하여 앞에서 제시한 손해 진전 계수(LDF)를 엑셀로 구현해 보았다. Table 3은 AY 기준 누적 손해액 Run-off triangle 형태의 데이터를 이용하여 손해 진전 계수를 산출한 자료이다. 즉, y_1은 1-2분기의 진전계수를 의미한다. 본 논문에서는 손해 진전 계수가 적용된 손해액 데이터를 바탕으로 최종 발생 손해액 추정 분석을 진행한다.

Table 3. LDF of loss amount for AY

AY	Q	y_1	y_2	y_3	y_4	y_5	$\cdots$	y_{48}	y_{49}	y_{50}	y_{51}	y_{52}	y_{53}	y_{54}	y_{55}
1993	1	0.988	0.947	0.954	0.995	0.964	$\cdots$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1993	2	1.008	0.965	0.935	0.990	0.991	$\cdots$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
1993	3	0.940	0.960	0.952	0.983	0.973	$\cdots$	1.000	1.000	1.000	1.000	1.000	1.000		
1993	4	0.940	0.965	0.957	0.980	0.994	$\cdots$	1.000	1.000	1.000	1.000	1.000			
1994	1	0.995	0.946	0.945	0.990	0.981	$\cdots$	1.000	1.000	1.000	1.000				
1994	2	0.986	1.007	0.955	1.000	0.990	$\cdots$	1.000	0.996	1.000					
1994	3	1.009	0.986	0.954	0.975	0.981	$\cdots$	1.000	0.999						
1994	4	1.051	1.048	0.977	1.009	0.974	$\cdots$	1.000							
1995	1	1.010	0.986	0.976	1.000	0.972									
1995	2	0.992	0.964	0.981	0.997	0.976									
1995	3	1.005	0.999	0.984	1.000	0.997									
1995	4	1.012	0.999	0.962	1.035	0.989									
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$								
2004	1	1.002	1.011	1.010	0.994	0.978									
2004	2	1.008	1.024	1.000	1.011	0.989									
2004	3	1.011	1.021	1.024	0.989	0.997									
2004	4	1.001	1.099	1.062	1.042	1.045									
2005	1	1.000	1.018	1.024	1.011	1.000									
2005	2	1.022	1.017	1.025	1.007	1.006									
2005	3	1.015	1.006	1.026	1.015	1.000									
2005	4	1.092	1.120	1.021	1.013										
2006	1	1.031	1.008	1.008											
2006	2	1.000	1.003												
2006	3	1.012													
2006	4														

3.3. 데이터 분석 방법

본 연구는 최종 손해액 예측력 향상 방식에 대한 제안이므로, 2장에서 제안한 모형들을 이용하여 자동차보험 최종 손해액을 잘 예측하였는지 검정하고자 한다. 검정의 방법으로는 Table 3에서 제시한 데이터의 마지막 대각선 한 줄을 삭제한 자료를 이용하여 실제값과 예측값을 비교 검정하였다.

사용한 모형이 유의할 경우 해당 모형을 이용하여 산출한 예측값을 사용하고, 모형이 유의하지 않을 경우 일반적으로 보험회사에서 사용하는 산술평균값을 사용하기로 한다. 실제값과 예측값을 비교하여 예측값이 정확하다면 우리가 궁극적으로 알고 싶은 최종 손해액 예측력 또한 높다고 할 수 있다. y_{31}부터 비교적 손해 진전이 안정화가 되었으므로 모형 예측에서 생략하기로 한다.

3.4. 모형별 분석 결과

실제 누적 손해액 LDF 자료 y가 1부터 55까지 존재하지만 y_{30} 이후의 경우 진전이 안정화되어 있으므로, 본 연구에서는 y_{30}까지 분석을 실시하였다. 이는 Table 3에서 y_{48}시점 이후 손해 진전 계수가 1로 수렴하는 현상이 y_{30} 이후에도 유사하게 발생함을 의미하는 것으로 이해할 수 있으며, 다음 식과 같이 각각의 모형을 적용시켜 5% 유의수준에서 검정을 실시하였다. 모형이 유의할 경우 추정값을 사용하고 그렇지 않을 경우에는 현재 자동차 보험사에서 사용하고 있는 산술평균값을 사

용하여 추정하였다. 다음 네 가지 모형 중 Model 1과 Model 2는 진전시기별로 적용이 가능한 단순 회귀모형과 분기데이터로 구성되어 있는 데이터의 특성을 고려하여 주기적으로 발생하는 현상을 고려하는 절편지시함수 모형을 고려하여 제안하는 모형이다. Model 3은 주기적으로 증가의 폭뿐만 아니라 변동의 수준을 동시에 반영할 수 있는 지시함수 모형이며, 마지막으로 Model 4는 특정시점에서 제도의 변경 등으로 진전계수가 급격히 변동할 개연성이 존재하는 경우를 고려하여 고안된 모형으로, 현업에서는 이를 인지하여 사전에 조정 가능할 수도 있으나, 여기서는 연구에 목적에 맞도록 신뢰 구간의 개념을 활용, 이를 통계적으로 인지하고 적용하는 방식으로 제안하는 모형이다.

1) **단순 선형회귀 모형**(Model 1)

$$\begin{aligned} y_1 &= \beta_{01} + \beta_{11} X_1 \\ y_2 &= \beta_{02} + \beta_{12} X_2 \\ &\vdots \\ y_{30} &= \beta_{0,30} + \beta_{1,30} X_{30} \end{aligned}$$

유의수준 5% 에서 검정한 결과 $y_1, y_2, y_3, y_4, y_5, y_6, y_8, y_9, y_{14}, y_{16}, y_{20}, y_{21}, y_{23}$ 모형이 유의하게 나타났다. 유의한 13개 모형의 경우 단순 선형회귀 모형을 이용하여 산출된 추정값을 사용하였고, 유의하지 않은 모형은 산술평균을 사용하여 분석하였다.

2) **절편 지시함수 모형**(Model 2)

$$\begin{aligned} y_1 &= \beta_{01} + \beta_{11} X_1 + \beta_{21} I(t_1) + \beta_{31} I(t_2) + \beta_{41} I(t_3) \\ y_2 &= \beta_{02} + \beta_{12} X_2 + \beta_{22} I(t_1) + \beta_{32} I(t_2) + \beta_{42} I(t_3) \\ &\vdots \\ y_{30} &= \beta_{0,30} + \beta_{1,30} X_{30} + \beta_{2,30} I(t_1) + \beta_{3,30} I(t_2) + \beta_{4,30} I(t_3) \end{aligned}$$

$$where \;\; I(t_i) = \begin{cases} 1, \text{ if } t_i = i^{th}\, quarter \\ 0, \text{ if } t_i \neq i^{th}\, quarter \end{cases}, \quad i = 1,2,3$$

유의수준 5%에서 검정한 결과 $y_2, y_3, y_4, y_5, y_8, y_9, y_{14}, y_{18}, y_{21}, y_{27}$ 모형이 유의하게 나타났다. 유의한 10개 모형은 분기효과를 고려한 절편 지시함수 모형을 이용하여 추정값을 사용하고 유의하지 않은 모형의 경우 산술평균을 이용하여 분석하였다.

3) **절편-기울기 지시함수 모형**(Model 3)

$$\begin{aligned} y_1 &= \beta_{01} + \beta_{11} X_1 + \beta_{21} I(t_1) + \beta_{31} I(t_2) + \beta_{41} I(t_3) + \beta_{51} I(t_1) X_1 + \beta_{61} I(t_2) X_1 + \beta_{71} I(t_3) X_1 \\ y_2 &= \beta_{02} + \beta_{12} X_2 + \beta_{22} I(t_1) + \beta_{32} I(t_2) + \beta_{42} I(t_3) + \beta_{52} I(t_1) X_2 + \beta_{62} I(t_2) X_2 + \beta_{72} I(t_3) X_2 \\ &\vdots \\ y_{30} &= \beta_{0,30} + \beta_{1,30} X_{30} + \beta_{2,30} I(t_1) + \beta_{3,30} I(t_2) + \beta_{4,30} I(t_3) + \beta_{5,30} I(t_1) X_{30} + \beta_{6,30} I(t_2) X_{30} + \beta_{7,30} I(t_3) X_{30} \end{aligned}$$

$$where \;\; I(t_i) = \begin{cases} 1, \text{ if } t_i = i^{th}\, quarter \\ 0, \text{ if } t_i \neq i^{th}\, quarter \end{cases}, \quad i = 1,2,3$$

절편-기울기 지시함수 모형은 절편 지시함수 모형에서 확장된 모형으로서 5% 유의수준에서 검정한 결과 $y_2, y_3, y_4, y_5, y_8, y_9, y_{14}, y_{21}, y_{27}$ 모형이 유의하게 나타났다. 유의한 9가지 모형은 절편, 기울기 지시함수 모형을 이용하여 추정값을 사용하고 유의하지 않은 모형들은 산술평균을 이용하여 분석을 실시하였다.

4) 특이점 지시함수 고려 모형(Model 4)

특이점 지시함수 $S(Z)$는 각 모형에서 $\mu+2\sigma$를 초과하는 값을 특이점으로 판단하여 '1'로 지정하고 $\mu+2\sigma$ 이하일 경우 특이점이 없다고 판단하여 '0'으로 지정하여 표현한 지시함수이다. 절편 기울기 지시함수 모형에서 특이점 지시함수 $S(Z)$를 추가하여 다음 식을 도출하였다.

$$
\begin{aligned}
y_1 &= \beta_{01}+\beta_{11}X_1+\beta_{21}I(t_1)+\beta_{31}I(t_2)+\beta_{41}I(t_3)+\beta_{51}I(t_1)X_1+\beta_{61}I(t_2)X_1+\beta_{71}I(t_3)X_1+\beta_{8,1}S(Z) \\
y_2 &= \beta_{02}+\beta_{12}X_2+\beta_{22}I(t_1)+\beta_{32}I(t_2)+\beta_{42}I(t_3)+\beta_{52}I(t_1)X_2+\beta_{62}I(t_2)X_2+\beta_{72}I(t_3)X_2+\beta_{8,2}S(Z) \\
&\vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots \\
y_{30} &= \beta_{0,30}+\beta_{1,30}X_{30}+\beta_{2,30}I(t_1)+\beta_{3,30}I(t_2)+\beta_{4,30}I(t_3)+\beta_{5,30}I(t_1)X_{30}+\beta_{6,30}I(t_2)X_{30}+\beta_{7,30}I(t_3)X_{30} \\
&\quad +\beta_{8,30}S(Z)
\end{aligned}
$$

$$
where\ \ I(t_i)=\begin{cases}1, \text{ if } t_i = i^{th}\ quarter \\ 0, \text{ if } t_i \neq i^{th}\ quarter\end{cases},\ \ i=1,2,3
$$

$$
S(Z)=\begin{cases}1, \text{ if } y_i > \mu+2\sigma \\ 0, \text{ if } y_i \le \mu+2\sigma\end{cases},\ i=1,2,\cdots,30
$$

각 모형에서 특이점을 확인한 결과 특이점이 1개 또는 특이점이 없는 경우가 가장 많이 나타났으므로 지시함수 $S(Z)$에 '0'을 대입하여 해당 진전분기의 LDF를 예측하였다. 30개의 모형을 유의수준 5%에서 검정한 결과 $y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11}, y_{12}, y_{13}, y_{14}, y_{17}, y_{18}, y_{19}, y_{20}, y_{21}, y_{22}, y_{24}, y_{25}, y_{27}, y_{29}$ 모형이 유의하게 나타났다. 유의한 24개의 모형은 특이점 지시함수 고려 모형을 이용하여 LDF 값을 추정하고 유의하지 않은 모형은 현업에서 사용하는 산술평균을 이용하여 분석하였다.

5) 제안한 모형을 이용한 예측값

Table 4는 본 연구에서 제안한 모형을 사용하여 만들어진 예측값이다. 음영이 들어간 셀의 경우 해당 모형에서 유의한 결과를 얻어 해당 모형을 사용한 예측한 값이며, 음영이 들어가지 않은 셀은 모형이 유의하지 않아 현재 손해보험사에서 일률적으로 사용하고 있는 산술평균값이다. y_{21}, y_{27}의 절편 기울기 지시함수와 특이점 지시함수 고려 모형의 예측값이 동일하게 나타난 이유는 y_{21}, y_{27}에서 특이점이 검출되지 않아 절편 기울기 지시함수 모형과 특이점 지시함수 고려 모형에서 동일한 예측값이 산출되었다. 본 연구에서 제안한 모형들의 예측값을 사용하여 적합성 검정을 실시한 후 결과 비교에 따라 최적의 방법을 제시하고자 한다.

3.5. 모형별 적합성 검정

손해액 예측에 적합한 모형을 선택하는 과정에서는 실제 데이터와 추정값의 오차를 최소화하는 모형을 찾는 것이 중요하다. 그러나 이 과정에서 각 모형의 상대적 적합성을 고려하지 않고 연구자의 주관적 판단에만 근거하여 모형을 선택한다면 LDF 산출과 적용의 오류로 인하여 보험사는 손해율 악화와 이익 감소로 경영상의 어려움에 빠질 수 있다.

실제로 예측 수행에 있어 요구되는 기법에는 예측 형태, 예측 기간, 자료 유형, 예측 기법 사용의 비용, 요구되는 정확도, 자료의 이용 가능성, 예측 기법의 운용 및 이해의 용이성 등 여러 가지 요소를 종합적으로 고려하여 예측 기법을 선정하여야 한다고 알려져 있다.

본 연구에서는 예측 기법의 정확성을 기준으로 하여 예측 기법 간의 차이를 비교하고자 한다. 이때, 앞서 언급한 기준들은 일반적으로 인정될 수 있는 유일한 측정 기준은 아니므로 항상 일관

된 결과만을 나타내지는 않는다는 점을 유의해야 하므로, 여기서는 각 기준의 보완성과 계산의 용이성을 고려하여 평균제곱근오차(RMSE), 평균절대오차(MAE), 평균절대퍼센트오차(MAPE)를 근거로 각 예측 기법의 차이를 비교 분석하고자 한다.

Table 4. Estimates of LDF

Loss amount	True value	Model 1	Model 2	Model 3	Model 4	Average
y_1	101.236	102.023	100.541	100.541	101.470	100.541
y_2	100.349	104.184	104.133	101.962	100.148	100.853
y_3	100.797	103.585	103.315	103.05	101.472	99.481
y_4	101.300	101.708	103.105	103.776	101.567	100.443
y_5	99.963	100.637	101.215	101.836	100.534	99.274
y_6	100.716	100.583	100.122	100.122	101.114	100.122
y_7	100.521	100.294	100.294	100.294	100.421	100.294
y_8	100.018	101.449	101.876	100.036	100.036	100.051
y_9	100.044	100.369	101.144	101.147	100.044	99.715
y_{10}	99.773	100.076	100.076	100.076	99.851	100.076
y_{11}	100.507	99.911	99.911	99.911	99.972	99.911
y_{12}	99.949	99.907	99.907	99.907	100.107	99.907
y_{13}	99.949	99.960	99.960	99.960	100.938	99.960
y_{14}	99.945	100.102	100.039	100.327	99.232	99.885
y_{15}	100.671	99.910	99.910	99.910	99.910	99.910
y_{16}	100.080	100.903	99.815	99.815	99.815	99.815
y_{17}	100.087	99.905	99.905	99.905	99.581	99.905
y_{18}	99.957	99.959	100.137	99.959	100.394	99.959
y_{19}	100.011	99.931	99.931	99.931	100.112	99.931
y_{20}	100.172	100.093	99.917	99.917	100.293	99.917
y_{21}	99.989	100.149	100.027	99.804	99.804	99.924
y_{22}	99.939	99.970	99.970	99.970	99.926	99.970
y_{23}	99.876	100.136	99.959	99.959	99.959	99.959
y_{24}	99.932	99.994	99.994	99.994	100.043	99.994
y_{25}	99.978	99.993	99.993	99.993	99.836	99.993
y_{26}	100.059	99.986	99.986	99.986	99.986	99.986
y_{27}	100.215	99.925	100.137	99.991	99.991	99.925
y_{28}	100.014	99.985	99.985	99.985	99.985	99.985
y_{29}	100.074	99.996	99.996	99.996	100.018	99.996
y_{30}	99.948	100.004	100.004	100.004	100.004	100.004

1) **평균제곱근오차**(root mean square error: RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2} \quad (y_i : \text{실제값}, f_i : \text{예측값}, n : \text{자료의 수})$$

2) **평균절대오차**(mean absolute error: MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - f_i| \quad (y_i : \text{실제값}, f_i : \text{예측값}, n : \text{자료의 수})$$

3) **평균절대퍼센트오차**(mean absolute percent error: MAPE)

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - f_i|}{y_i} \times 100 \quad (y_i : \text{실제값}, f_i : \text{예측값}, n : \text{자료의 수})$$

이 세 가지 측정치는 예측성과에 대하여 항상 동일한 순서대로 성과를 표시하지 않는다. MAE는 실제값과 예측값 간의 절대적인 차이만을 나타내지만 MAPE는 비율로 나타낸 것으로 실제값의 크기에 영향을 받는다. 그러나 본 논문의 데이터에서는 분모의 값이 1에 가까운 숫자들로 이루어져 있어 MAPE 값과 MAE 값이 유사하게 나타난다. MAPE 값이 MAE 값보다 작은 값으로 이루어져 있으므로 본 논문에서는 MAPE 값을 사용하고자 한다. RMSE는 추정값 또는 모형이 예측한 값과 실제 환경에서 관찰되는 값의 차이를 다룰 때 흔히 사용하는 측도이며 정밀도(precision)를 표현하는데 적합하다. 각각의 차이 값을 잔차(residual)라고도 칭하며, 평균제곱근편차는 잔차들을 하나의 측도로 종합할 때 사용된다. 본 연구에서는 MAPE와 RMSE 값을 이용하여 모형의 우위를 판단하고자 한다.

추정값이 작아지는 것을 방지하기 위해 전체 데이터에 100을 곱하여 단위 조정을 한 후 검정에 사용하였다. 또한 Table 5의 RMSE, MAPE 값들은 모형 적합성 검정에서 모두 30개의 자료를 사용했으므로 위 식에서 n으로 나누는 것을 생략하여 계산하였다. Table 5의 결과는 해당 모형이 5% 유의수준 하에서 유의할 경우, 해당 모형을 이용하여 예측값을 사용하고 유의하지 않을 경우 현업에서 사용하는 산술평균값을 이용하여 적합성 검정을 실시하였다. 전체 모형 기준 적합성 검정 결과 산술평균이 RMSE, MAPE가 모두 가장 작은 값으로 나타났다.

Table 5. Fitting results for the considered models

	Model 1	Model 2	Model 3	Model 4	Average
RMSE	2.044	2.448	1.652	1.753	1.217
MAPE	6.062	6.486	5.052	5.747	3.915

본 논문에서 제안하는 네 가지 모형만을 비교 검정해보면, 평균제곱근 오차인 RMSE를 사용하여 적합성 검정을 실시한 결과, 절편-기울기 지시함수 모형을 이용한 예측값이 실제값에 가장 근사하다는 것을 알 수 있었으며 두 번째로는 특이점 지시함수 고려 모형, 단순 선형회귀 모형, 절편 지시함수 모형 순으로 손해액 예측력이 높다는 것을 알 수 있었다. 또한 예측 정확도를 나타내는 MAPE를 통해 적합성 검정을 실시한 결과, RMSE와 동일한 결과를 얻었다.

Table 6. Fitting results for significant models

	Model 1	Model 2	Model 3	Model 4	Average
Number of significant models	13	10	9	24	30
RMSE	1.433	1.738	1.444	0.382	0.429
MAPE	0.9083	1.2651	1.1190	0.2840	0.2844

Table 5는 유의한 모형과 유의하지 않은 모형을 같이 사용하여 적합성 검정을 한 결과이며, Table 6은 유의한 모형만을 사용하여 적합성 검정을 한 결과이다. 유의한 모형 개수는 30개 중 특이점 지시함수 고려 모형이 24개로 가장 많았고, 단순 선형 회귀 모형은 13개, 절편 지시함수 모형은 10개, 절편-기울기 지시함수 모형은 9개로 나타났다. 유의한 모형 기준 적합성 검정을 위해 RMSE와 MAPE를 사용하였다.

RMSE, MAPE 모두 적합성 우위가 동일하게 나타났으므로 설명의 간략화를 위해 MAPE 적합성

결과를 이용하여 설명하고자 한다. 특이점 지시함수 고려 모형이 0.2840으로 다른 모형들에 비해 가장 작은값으로 나타났다. 즉 실제값과 예측값의 차이가 가장 작게 나타나므로 특이점 지시함수 고려 모형이 현업에서 사용하는 산술평균에 비해 예측 정확도가 높은 모형임을 증명할 수 있다. 그러나 산술평균의 MAPE 값이 0.2844로 특이점 지시함수 고려 모형과 차이가 크지 않다는 것을 알 수 있다. 그러나 실제 자동차 보험 손해액의 단위가 크기 때문에 차이가 작더라도 보험사 입장에서는 실제 손해액 규모가 크게 나타나게 된다.

일반적으로 자동차보험 손해액은 손해 진전이 2년에서 3년 사이까지 변동이 크게 나타나는 특성을 가지고 있다. 따라서 손해액 규모에 대한 추정은 사고 초기의 변동에 초점을 맞추게 된다. 현재 현업에서 활용하고 있는 일률적인 방식보다는 여러 가지 대안 모형을 적용, 활용함으로써 각 회사의 구조에 적합한 모형으로 수렴하는 과정을 거칠 수 있다는 것에 본 논문의 의미를 부여할 수 있다. 이러한 목적에 부합하도록 본 연구에서는 2년 반 즉, y_{10}까지 변동이 클 것으로 간주하여 $y_1 \sim y_{10}$까지 누적 손해액 LDF 데이터를 사용한 적합성 검정 결과를 Table 7에 정리하였다. RMSE 기준 적합성 검정 결과 특이점 지시함수 고려 모형이 0.335로 실제값과 가장 유사한 값을 예측하는 모형임을 증명하였다. 두 번째로는 현업에서 사용하고 있는 산술평균(0.655), 절편-기울기 지시함수 모형(1.757), 단순 선형 회귀 모형(1.799) 마지막으로 절편 지시함수 모형(2.241) 순으로 최종 손해액 예측력이 높다는 것을 알 수 있었다. MAPE기준 적합성 검정 결과 또한 특이점 지시함수 고려 모형이 0.253으로 가장 예측력이 높았으며 산술평균(0.551), 단순 선형 회귀 모형(1.291), 절편-기울기 지시함수 모형(1.547), 마지막으로 절편 지시함수 모형(2.043) 순으로 예측력이 높게 나타났다.

현재 보험사는 일률적으로 산술평균을 사용하여 최종 손해액을 추정하고 있지만 자동차 보험의 경우 국가 정책의 변화, 보상기준의 변화 등이 항상 내포하기 때문에 보험사는 이를 고려하여 최종 손해액 예측에 반영해야 한다. 보험사는 추세에 맞춰 반영할 수 있는 LDF 방식을 현업에서 가지고 있어야 보다 정확한 최종 손해액을 산출할 수 있다

Table 7. Fitting results for $y_1 \sim y_{10}$

	Model 1	Model 2	Model 3	Model 4	Average
Number of significant models	8	6	6	10	10
RMSE	1.799	2.241	1.757	0.335	0.655
MAPE	1.291	2.043	1.547	0.253	0.551

Table 8의 실제값과 예측값의 오차를 보면 y_1부터 y_{11}까지 현업에서 사용하는 산술평균값보다 특이점 지시함수 고려 모형을 이용한 예측값의 오차가 작은 것을 볼 수 있다. 이는 특이점과 변동이 존재하여 본 논문에서 제안한 특이점 지시함수 고려 모형을 사용하여 실제값을 예측하는 방안이 현재 자동차 손해보험사에서 최종 손해액 예측을 위해 일률적으로 사용하는 산술평균 방법보다 실제 손해액을 예측하는데 있어 더 높은 예측력을 나타낸다.

4. 결론

현재까지의 사고율, 손해액 예측과 관련된 연구들은 사고 정보 데이터가 주어졌다는 가정하에 미래에 실현될 사고율 및 손해액의 예측에 관점을 둔 분석방식과 그 결과들을 제안하고 있다. 그러나 미래 예측에 필요한 설명 변수인 손해 관련 데이터가 정확해야만 미래 사고에 대한 예측 결과가 합리적으로 추정 될 수 있다. 따라서 미래 예측의 근간이 되는 최종 손해액 데이터에 대한

정교한 추정이 선결적으로 이루어져야 한다. 현재까지 다양한 추세 예측 연구에 비하여 추세 연구의 근간이 되는 적정한 손해액 데이터의 추정과 관련된 연구는 미미한 실정이다. 본 연구에서는 이러한 문제점을 보완하여 보험금 최종 손해액 수준을 합리적이게 추정할 수 있는 다양한 방법을 제시하고, 그 효용성을 증명하고자 하였다.

Table 8. Difference between estimates and true values

Loss amount	Model 1	Model 2	Model 3	Model 4	Average
y_1	0.787			0.234	0.695
y_2	3.835	3.784	1.613	0.201	0.504
y_3	2.788	2.518	2.253	0.675	1.317
y_4	0.408	1.805	2.476	0.267	0.857
y_5	0.674	1.252	1.873	0.571	0.689
y_6	0.133			0.398	0.594
y_7				0.100	0.227
y_8	1.431	1.858	0.018	0.018	0.033
y_9	0.325	1.100	1.103	0.000	0.328
y_{10}				0.078	0.303
y_{11}				0.535	0.595
y_{12}				0.158	0.042
y_{13}				0.989	0.011
y_{14}	0.157	0.094	0.382	0.713	0.059
y_{15}					0.761
y_{16}	0.823				0.265
y_{17}				0.506	0.181
y_{18}		0.180		0.437	0.002
y_{19}				0.101	0.080
y_{20}	0.079			0.121	0.255
y_{21}	0.160	0.038	0.185	0.185	0.065
y_{22}				0.013	0.031
y_{23}	0.260				0.084
y_{24}				0.111	0.061
y_{25}				0.142	0.015
y_{26}					0.073
y_{27}		0.078	0.224	0.224	0.290
y_{28}					0.029
y_{29}				0.056	0.078
y_{30}					0.056

본 논문에서 사용한 S 보험사의 데이터를 통하여 비교 분석한 결과, 손해 진전이 10-11분기까지 변동 폭이 크게 발생하는 것을 알 수 있었다. 그 결과 본 논문에서 제안하고 있는 특이점 지시함수 고려 모형 방법이 현업에서 사용하는 산술평균 방법의 단점을 보완하면서 예측력이 개선되는 것을 확인할 수 있었다. 결과적으로 y_1부터 y_{10}까지는 본 논문에서 제안하는 특이점 지시함수 고려 모형을 사용하여 LDF 예측값을 산출하고, 그 이후에는 비교적 변동 폭이 줄기 때문에 현업에서 사용하는 산술평균을 이용하여 LDF 예측값을 산출하는 방안을 선별적으로 활용한다면 최종 손해액 예측력을 한층 높일 수 있을 것이다.

각 분기별로 보상에 대한 어떠한 패턴이 나타나지 않거나 일관성 있는 현상이 지속된다면 본 논문에서 제안한 모형 사용에 한계가 발생할 수 있으나, 자동차 보험의 경우 국가 정책의 변화, 보상기준의 변화, 회사 운영방식의 변화 등은 항상 발생할 수 있는 속성을 가지기 때문에 보험사는 이를 고려하여 최종 손해액 예측력을 향상시킬 필요가 있다. 본 논문에서 제시한 다양한 방법들을 고려하여 회사별로 존재하는 보상 데이터의 특징과 보상 환경에 부합하는 최적의 모형을 설계하고 수려해 나가는 과정이 필요한 것이며, 이를 통하여 정확한 손해액을 추정하고 정교한 미래 예측이 가능한 다양한 LDF 방식을 현업에서는 고려할 필요가 있다. 본 논문에서 제안하는 다양한 방식과 이에 대한 효용성 검증이 이러한 목적에서 의미를 둘 수 있다는 사실을 시사하므로, 본 연구에서 제안하는 방식이 현업에서 다양하게 활용되기를 기대한다.

References

Cook, C. (1970). Trend and loss development factors, *Proceedings of the Casualty Actuarial Society*, 57, 1-14.

Cummins, J., Powell, A. (1980). The performance of alternative models for forecasting automobile insurance paid claim costs, *Astin Bulletin,* 11, 91-106.

Jang, D. (1997). A study on the chain ladder method for insurance loss reserving, *The Journal of Risk Management,* 8, 35-48. (in Korean).

Jeong, J., Choi, J. (2014). Poisson regression and negative binomial regression model fit for traffic accidents, *Journal of the Korean Data Analysis Society*, 16(1B), 165-172. (in Korean).

Kang, I., Heo, T. (2015). Analysis of traffic accident data and relative risk estimation using generalized Poisson regression Model, *Journal of the Korean Data Analysis Society*, 17(4B), 1933-1944. (in Korean).

Kim, M. (2013). A study on trend analysis of severity and frequency for predicting the proper premium -focusing on auto insurance-, *The Journal of Actuarial Science,* 5, 3-19. (in Korean).

Kim, Y., Park, W. (2013). Estimating automobile insurance premiums based on time series regression, *The Korean Journal of Applied Statistics*, 26, 237-252. (in Korean).

Lee, C., Lee, K. (2006). A study for the evaluation of loss reserve risk of property-liability insurers, *Korean Insurance Journal*, 74, 199-223. (in Korean).

Lee, J. J. (2010). Motor vehicle crash count data analysis using negative binomial mixture regression models, *Journal of the Korean Data Analysis Society*, 12(3B), 1485-1498. (in Korean).

McCarthy, T. (1998). A frequency based model for excess wind in property ratemaking, *Casualty Society Forum* (Winter), 209-238.

Oh, C., Lee, J. (2014). A study on the rating claims-made insurance policy, *The Journal of Actuarial Science*, 6, 3-35. (in Korean).

Shim, H., Choi, Y. (2016). Age-period analysis on Korea road fatality rates, *Journal of the Korean Data Analysis Society*, 18(5B), 2611-2626. (in Korean).

Research for the Predictability Enhancement of Loss Development Factor in Auto Insurance*

Myung Joon Kim[1], *Soyul Han*[2], *Yeong-Hwa Kim*[3]

Abstract

Loss amount due to car accidents is not finalized at the time of accident and has characteristics of change as time passes away. Not only treatment extension and inflation, policy change affect the payment standard. Previous studies were relatively focused on the accident trend prediction and insignificant interest was given to the proper loss amount estimation which is foundation of the trend prediction. The study compensates the biased interests and proposes the reasonable methodology of ultimate loss estimation. Auto insurance coverages consist of property and bodily damages and this study works with bodily liability coverage which has relatively large fluctuation. Using the cumulative loss amount changed by the time, the precise estimation method is suggested. Also through the real data analysis result, the properness and effectiveness of suggested model will be proved by comparison with currently used method by business field.

Keywords : Loss development factor, loss amount, run-off triangle, trend prediction.

*This research was supported by the Chung-Ang University research grant in 2016.
[1]Assistant Professor, Department of Business Statistics, Hannam University, Hannam-Ro 70, Daeduk-Gu, Daejeon 34430, Korea. E-mail : mkim@hnu.kr
[2]Graduate Student, Department of Statistics, Graduate School, Chung-Ang University, Heuksuk-Ro 84, Dongjak-Gu, Seoul 06974, Korea. E-mail : soyul5458@naver.com
[3](Corresponding Author) Professor, Department of Applied Statistics, Chung-Ang University, Heuksuk-Ro 84, Dongjak-Gu, Seoul 06974, Korea. E-mail : gogators@cau.ac.kr
[Received 28 December 2016; Revised 4 February 2017, 17 February 2017; Accepted 20 February 2017]