Communications of the Korean Statistical Society Vol. 15, No. 4, 2008, pp. 551–562

삼각분할표에서 구조적 변화점 유무에 관한 검정†

이성임[1)]

요 약

보험분야에서 지급준비금(loss reserve)을 추정할 때에는 보험사고의 발생년도와 사건발생 이후의 경과년도에 따라 지급된 보험금을 자료로 사용하게 되는데, 이것은 흔히 삼각분할표(run-off triangular table)의 형태로 주어진다. 이러한 삼각분할표 자료에 대하여 지급준비금 추정에 주로 사용되는 방법으로 사다리법(chain-ladder method)이 있는데, 이것은 사고발생년도부터 보험금이 정산되는 시점까지의 경과기간동안 지급된 누적 보험금의 변화율(진전계수)을 추정함으로써 지급준비금을 추정하는 것이다. 이러한 사다리법은 보험사고의 발생년도에 따른 진전계수의 변화가 없다는 가정을 기본전제로 하고 있다. 그러나 여러 가지 사회 환경적 요인으로 인하여 시간이 지남에 따라 지급보험금의 진전패턴이 달라질 수 있고, 본 논문에서는 사건의 변화에 따른 구조적 변화점 유무를 검정할 수 있는 검정법을 제안하고자 한다. 또한 이를 실제 예제에 적용 고찰해 보고자 한다.

주요용어: 삼각분할표; 사다리법; 구조적 변화.

1. 서론

범주형 자료분석에서 가장 기본적인 자료형태는 $R \times C$ 분할표(contingency table)이다. 다양한 응용 분야에서 이러한 분할표가 기초자료로 주어지는데, 그 중에서 보험 분야의 지급준비금(loss reserve)추정에 사용되는 자료는 그림 1.1과 같은 삼각분할표(run-off triangular table)의 형태를 갖게된다. 지급준비금은 기보고발생준비금과 미보고발생준비금(Incurred But Not Reported: IBNR)으로 크게 나눌 수 있는데, 전자는 사고발생 후 보고된 사고에 대하여 보험금이 지급되지 않은 손해를 보상하기 위하여 적립된 금액을 의미하고, 후자는 보험사고가는 이미 발생했지만 아직 보험회사에 보고되지 않은 계약에 대해 향후 지급될 보험금 추정액을 의미한다 (홍종선과 전흥기, 2006). 그림 1.1에서 자료 $Z_{i,j}$는 i년도에 발생한 사고에 대하여 j년도 경과 후 지급된 보험금을 나타낸다. 단, $j = 1$은 사건 발생 당해년도를 나타낸다. 다시말해, 그림 1.1에서 작성된 자료는 현재를 n년도 기준으로 현재시점까지 지급된 보험금을 나타낸다. 따라서, 그림 1.1에서 $i+j > (n+1)$를 만족하는 $Z_{i,j}$들은 현재시점 n년도 기준으로, 앞으로 지급할 보험금을 나타내며, 이는 곧 자료로부터 예측하고자 하는 지급준비금이 된다. 분석의 주된 관심은 현재까지 지급된 보험금으로부터 앞으로 지급할 지급준비금을 예측하는 것이다.

† 본 연구는 단국대학교 2006–2007년 연구비 지원에 의해 수행되었음.
1) (122-807) 경기도 용인시 수지구 죽전동 126번지, 단국대학교, 조교수. E-mail: silee@dankook.ac.kr

$i \backslash j$	1	2	···	···	···	···	n
1	$Z_{1,1}$	$Z_{1,2}$	···	···	···	···	$Z_{1,n}$
2	$Z_{2,1}$	$Z_{2,2}$	···	···	···	$Z_{2,n-1}$	
⋮	⋮	⋮	⋮	⋮	⋰		
i	$Z_{i,1}$	$Z_{i,2}$	···	$Z_{i,n-i+1}$			
⋮	⋮	⋮	⋰				
⋮	⋮	$Z_{n-1,2}$					
n	$Z_{n,1}$						

그림 1.1: 자료 $Z_{i,j}$는 i년도에 발생한 보험사고에 대하여 j년 경과한 보험사고에 대하여 지급된 보험금을 나타낸 것으로, 현재년도를 n으로 했을 때, 자료는 위와 같은 삼각형 형태로 요약된다. 이 때, $j=1$은 사건발생 당해년도를 가르킨다.

지급준비금의 추정방법은 크게 보험사고유형별로 손해사정자가 지급준비금 산정기준의 세목별 항목에 의거하여 개별적으로 추산하여 적립하는 개별추산법과, 과거에 발생한 손해자료를 이용하여 사고발생기간의 진전에 따른 추세와 사고발생 시기에서 보험금지급까지의 경과기간에 따른 패턴을 추정하여 이러한 과거의 추세와 패턴이 미래에도 계속된다는 가정 하에 미래에 발생할 손해규모를 추정하는 총량적 추산방법으로 나뉜다 (홍종선과 전홍기, 2006). 개별추산방법은 대수의 법칙을 적용시킬 수 없을 정도로 건수가 적고 건당 손해액이 큰 화재보험 등에 주로 적용되는 것으로, 보통은 손해사정 담당자의 경험적인 판단에 의해 지급준비금을 추산하게 된다. 그러나 이 방법은 물가상승률 등에 의해 준비금이 크게 좌우되는 경향이 있고, 동일한 사고에 대해 손해사정자의 주관적인 판단에 따라 서로 다른 준비금 산출이 가능하는 등 객관성이 보장되지 않는다. 총량적 추산방법은 크게 고전적 모형과 확률적 모형으로 나눌 수 있다. 고전적 모형은 경과보험료 중 일정한 비율(예정손해율)의 금액이 장래에 발생될 손해 보험금으로 충당된다는 손해율법(Loss Ratio Method), 손해 발생 이후부터 경과기간동안 사고 건당 평균 지급보험금이 매우 일정한 비율로 진전된다고 가정하여 추산한 평균지급보험금방식(Average Payment Method), 손해가 발생한 시점부터 경과기간동안 손해액이 어떻게 진전되어 가느냐를 총체적으로 분석하여 향후 지급될 것으로 예상되는 지급준비금을 추산하는 사다리법, 손해율법과 사다리법을 합산한 방법으로 예정손해율과 연도별 손해액 진전계수를 동시에 고려하는 보른휴에테르-퍼거슨법(Bornhuetter-Ferguson Method) 등이 있다 (Brown, 1993). 확률적 모형으로는 사다리법을 확장 적용하여 사고발생년도와 경과기간의 효과를 가정한 이원배치 분산분석 모형 (Kremer, 1982) 등이 있다. 이들 모형간의 비교는 Mack과 Venter (2000) 그리고 Verrall과 England (2000) 등을 참조하기로 한다.

삼각분할표로부터 지급준비금을 추정하는 방법들 중 가장 기본적이고 널리 사용되는 방법은 사다리법(chain ladder method)으로 이것은 앞서 말한대로 사고발생년도부터 그 사고로 인한 보험금이 정산되는 시점까지의 누적지급보험금의 변화율(진전계수, development coefficient)을 사용하여 지급준비금을 추정하는 방법이다. 이러한 사다리법은 사고

발생년도에 따른 지급준비금의 변화가 존재하지 않는다는 가정하에 제시된 추정법이므로, 사고발생년도에 따른 진전계수의 급격한 변화의 개연성이 의심된다면 지급준비금 추정 결과에 문제가 있을 수 있다 (Kremer, 1982; Zehnwirth, 1991; Ajne, 1989). 고전적 모형과 확률적 모형 모두 이러한 가정을 기본 바탕으로 하므로, 본 논문에서는 지급준비금을 추정하기 전 이러한 가정을 미리 검토하기 위한 절차로, 보험금 지급년도별 추세에 따른 구조적 변화점의 유무에 대한 검정을 할 수 있도록 통계량을 제안하고 이를 실제 자료에 적용해 보고자 한다.

본 논문의 구성은 다음과 같다. 제 2절에서는 삼각분할표 형태의 자료로부터 지급준비금을 추정하는데 일반적으로 많이 사용되고 있는 추정법으로 사다리법을 소개하기로 한다. 제 3절에서는 변화점 유무에 대한 검정 통계량을 제안하고, 제4절에서는 실제 예제를 통해 검정 통계량을 적용해 보기로 한다.

2. 사다리법

사다리법은 지급준비금 추정에 있어 가장 많이 사용되는 방법으로, 보험 사고의 발생년도로부터 그 사고로 인한 보험금이 정산될 때까지 경과기간에 따른 보험금의 변화율과 사고발생년도에 따른 추세를 고려하여, 과거의 이들 패턴과 추세가 미래에도 계속된다는 가정 하에 미래에 발생할 보험금을 추정하는 것이다 (Verrall, 1994). 즉, 사고가 i년도에 발생하여 그 후로 j년 경과할 때까지 누적하여 지급된 보험금을 $C_{i,j} = \sum_{k=1}^{j} Z_{i,k}$ (단, $1 \leq j < n-i+1$)로 정의하고, i년도에 발생한 사고에 대하여 누적 지급보험금이 $(j-1)$년도에서 j년도로 변화할 때의 변화율을 진전계수(development factor)라 하여 $f_{i,j} = C_{i,j}/C_{i,j-1}$로 나타낸다. 그림 1.1에서 현재시점에서 계산가능한 진전계수들로부터 사고발생 후 $(j-1)$년도에서 j년도로 경과할 때 지급보험금에 대한 평균변화율 F_j가 추정되고, 이들 추정값으로부터 사건 발생년도 i에 대한 최종 지급보험금은 다음과 같이 추정된다.

$$\widehat{C}_{i,n} = \widehat{F}_n \cdots \widehat{F}_{n-i+3} \widehat{F}_{n-i+2} C_{i,n-i+1}.$$

이 때 F_j를 추정하는 방법은 크게 세가지로 산술평균방법, 최근 m년간의 방법 그리고 가중평균방법이 있다. 산술평균방법(arithmetic average method)은 가능한 모든 사고발생년도의 진전계수를 산술평균하는 방법으로 다음과 같이 추정한다.

$$\hat{F}_j = \frac{\sum_{i=1}^{n-j+1} f_{i,j}}{n-j+1}, \quad j = 2, \ldots, n.$$

최근 사고발생 m년간의 산술평균방법(m-year arithmetic average method)은 최근 사고발생 m년간의 진전계수를 산술평균하는 방법으로 다음과 같이 추정한다.

$$\hat{F}_j = \frac{\sum_{i=n-j-(m-2)}^{n-j+1} f_{i,j}}{m}, \quad j = 2, \ldots, n.$$

단, 최근 사고발생이 m년 미만인 경우에는 m년 미만의 자료를 산술평균하여 진전계수를 계산한다. 이 방법은 과거 자료의 신뢰성이 부족할 때 또는 과거 지급보험금이 현재와 비교하여 많이 적거나 혹은 주위환경의 변화가 심하여 급격한 변화의 개연성이 존재하는 경우 등, 주로 최근년도에 얻어진 진전계수의 영향력을 높이고 싶은 경우에 사용된다. 마지막으로, 가중평균방법(volume-weighted average method)은 사고년도별 누적보험금을 가중치로 하여 진전계수들을 평균한 방법으로 다음과같이 추정한다.

$$\hat{F}_j = \sum_{i=1}^{n-j+1} f_{i,j} \times \frac{C_{i,j-1}}{\sum_{i=1}^{n-j+1} C_{i,j-1}} = \frac{\sum_{i=1}^{n-j+1} C_{i,j}}{\sum_{i=1}^{n-j+1} C_{i,j-1}}, \quad j = 2, \ldots, n.$$

이것은 사고년도별 지급보험금의 변화가 커 지급보험금의 크기가 상대적으로 큰 사고년도의 진전계수에 비중을 많이 두고 싶은 경우에 효과적인 추정법이다. Lemaire (1985)와 Brown (1993)는 산술평균방법은 사고발생년도별 진전계수의 추세가 보이지 않고 무작위적으로 변화할 때 사용하는 방법이라고 평가하고, 최근 사고발생 m년간의 산술평균방법은 과거 자료의 신빙성이 부족하거나 과거 지급보험금이 현재와 비교하여 많이 적은 경우, 또는 주위환경의 변화가 심하여 최근년도 진전계수의 영향력을 높이고 싶은 경우에 사용한다고 하였다.

이처럼 지급준비금을 추정할 때, 평균변화율 F_j의 추정법에 따라 지급준비금 추정에 차이가 발생된다. 그런데, 사다리법은 진전계수 $f_{i,j}$에 대해 사고발생년도 i에 대해서 독립이라는 가정을 한 것으로, 사고발생년도 i에 급격한 구조적 변화의 개연성이 존재하는 경우 이를 무시하고 F_j를 추정하게 된다면 추정에 문제가 발생될 수 있다. 예를 들어, 박잉근 (2006년)에서 실제자료를 이용하여 이들 진전계수의 세 가지 산출법에 따라 추정되어진 지급준비금의 평균제곱오차(Mean Squared Error: MSE)를 비교함으로써 그 효율을 평가한 결과를 살펴보면, 최근 사고발생 m년간의 산술평균방법으로 5개월과 10개월을 사용한 결과 최근 5개월의 자료를 사용한 지급준비금의 MSE가 최근 10개월의 자료를 이용한 지급준비금의 MSE보다 작게 나타났음을 보였다.

따라서 사다리법에 근거한 지급준비금 추정 시, 사고발생년도 별 진전계수의 구조적 변화 유무에 대하여 검정할 필요가 있게 된다. 이에 다음 절에서는 검정절차에 대해 소개하기로 한다.

3. 변화점 존재 유무에 대한 검정

사다리법에서 임의의 보험사고 발생 후 $(j-1)$년도에서 j년도로 경과할 때 지급된 보험금의 평균변화율 F_j를 추정 할 때, 일반적으로 진전계수 $f_{i,j}$가 사고 발생년도에 의존하지 않는 동일한 분포를 따른다는 가정을 하고 있다. 하지만 실제 지급보험금의 평균 변화율은 여러 가지 사회적 요인 또는 주변환경의 변화 등으로 시간이 지남에 따라 변화가 있다는 사실은 누구나 주지하고 있고, 이러한 이유로 제 2절에서 언급된 대로 실제 문제에서 지

급준비금을 추정할 경우, 평균 변화율을 추정함에 있어 최근 m년 간의 자료만을 사용할 필요가 있다. 즉 최근 m년 동안에는 진전계수에 대한 변화점이 존재하지 않거나 존재 하여도 미미하다는 가정을 전제로 하고 있다 할 수 있다. 하지만 이러한 접근 또한 m값을 정하는데 있어 특별한 기준이 없이 회사마다 또는 개인 개인의 주관적인 판단에 의존한다는 단점을 지니고 있다. 따라서 이 절에서는 진전계수에 대한 변화점 존재 유무에 대하여 객관적인 판단을 가능하게 하는 검정 통계량를 제안하고, 더 나아가 이 통계량을 이용하여 변화점의 위치를 파악하는 과정에 대하여 이야기 하고자 한다.

먼저 변화점의 존재 유무를 검정하기 위한 통계량을 제안하고자 한다. 진전계수에 대해 최근 $t(> m)$번째 년도에 변화점이 존재한다는 이야기는, 각각의 $j = 2, \ldots, n$에 대하여 $f_{1,j},\, f_{2,j}, \ldots, f_{(t-1),j}$들의 분포가 F_{0j}를 따르고, $f_{t,j},\, f_{(t+1),k}, \ldots, f_{n,j}$들의 분포가 F_{1j}일 때, 어떤 j에 대하여 이 두 분포가 같지 않음을 의미한다. 즉, 변화점 존재 유무 검정을 위한 귀무가설은

$$\mathcal{H}_0^{(t)} = \left\{F_{0j} = F_{1j},\, j = 2, \ldots, n\right\} = \bigcap_{j=2}^{n} \left\{F_{0j} = F_{1j}\right\}$$

으로 표현 될 수 있다. 또, 여기에서 기존 Mack (1993) 등의 연구에서와 같이 두 분포가 동일하다는 가정을 각 분포의 모평균 μ_{0j}와 μ_{1j}가 다르다는 검정하기 쉬운 약화된 가정으로 표현 할 수 있고, 이것은 다시 말해

$$\mathcal{H}_0^{(t)} = \left\{\mu_{0j} = \mu_{1j},\, j = 2, \ldots, n\right\} = \bigcap_{j=2}^{n} \left\{\mu_{0j} = \mu_{1j}\right\}$$

가 된다. 앞서 언급한 것처럼 위 가설은 진전계수들의 변화점 유무에 관한 문제로, 사고발생년도로부터 1차년도에서 2차년도로 경과할 때의 전진계수 $j = 2$에 해당되는 관측값들은 $\left\{f_{t,2},\ t = 1, 2, \ldots, T\right\}$로 주어진다. 이 때, $j = 1$에 해당하는 관측값은 존재하지 않음에 유의한다.

제안된 통계량을 소개하기에 앞서 각 "기본단위 가정" $\mathcal{H}_{0j}^{(t)} = \{F_{0j} = F_{1j}\}$를 검정하는 통계량을 생각하여 보기로 한다. 우선 H_j 각각에 대한 검정은 서로 다른 두 그룹에 대한 모평균의 차에 관한 검정이 되고, 이것은 간단하게 이표본 t검정 통계량

$$z_{t,j} = \left(\bar{f}_{(t-1),j} - \bar{f}_{t,j}\right) \Big/ \sqrt{\frac{s^2_{(t-1),j}}{n_{(t-1),j}} + \frac{s^2_{t,j}}{n_{t,j}}} \tag{3.1}$$

을 사용한다. 식 (3.1)에서 $\bar{f}_{(t-1),j}$와 $s^2_{(t-1),j}$는 $f_{1,j}, f_{2,j}, \ldots, f_{t-1,j}$들의 표본평균과 표본분산, $\bar{f}_{t,j}$와 $s^2_{t,j}$는 $f_{t,j},\, f_{(t+1),j}, \ldots, f_{n,j}$들의 표본평균과 표본분산을 의미한다. 식 (3.1)의 이표본 t검정의 결과로 나온 유의확률을 $P_{t,j}$라 하면

$$P_{t,j} = 2\left\{1 - \Phi\left(\left|z_{t,j}\right|\right)\right\} \tag{3.2}$$

이고, 이들은 변화점이 없다는 가정 아래 서로 독립인 Uniform(0, 1)을 따르는 확률 변수들임이 알려져 있다. 따라서 이들 "기본단위 검정"들의 결과들은 다음의 $\{1, 2, \ldots, T\} \times$

표 3.1: $P_{t,j}$는 변화년도 $t\,(t = 1,\ldots,T)$를 기준으로 각 경과년도 j별 그 전 후의 진전계수의 모평균이 동일하다는 귀무가설을 검정한 결과에 대한 유의확률이다.

t	경 과 구 간					
	1→2	2→3	$\cdots$	$(j-1)\to j$	$\cdots$	$(n-1)\to n$
1	$P_{1,2}$	$P_{1,3}$	$\cdots$	$P_{1,j}$	$\cdots$	$P_{1,n}$
2	$P_{2,2}$	$P_{2,3}$	$\cdots$	$P_{2,j}$	$\cdots$	$P_{2,n}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
t	$P_{t,2}$	$P_{t,3}$	$\cdots$	$P_{t,j}$	$\cdots$	$P_{t,n}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
T	$P_{T,2}$	$P_{T,3}$	$\cdots$	$P_{T,j}$	$\cdots$	$P_{T,n}$

$\{2,3,\ldots,n\}$의 표 3.1로 표현되고, 본 절에서는 이들을 묶어 하나의 통계량으로 표현하는 방법을 살펴 보고자 한다.

위의 기본단위 검정의 결과 $\{P_{t,j},\, t = 1,2\ldots,T,\, j = 2,3,\ldots,n\}$에서 임의의 쌍 $P_{t,j}$와 $P_{s,j'}$는 $j \neq j'$에 대해 서로 독립임을 가정한다. 이 사실을 이용하면 $P_{t(j)}$를 $P_{t,2}, P_{t,3},\ldots,$ $P_{t,n}$들의 j번째 순서 통계량이라 할 때 이들은 $\mathcal{H}_0$의 가정하에서 Uniform$(0,1)$을 따라야 하므로 $P_{t(j)}$는 Uniform$(0,1)$분포의 j번째 순서통계량의 분포인 모수가 $(j,\, n-1-j)$인 Beta분포

$$f_{(j)}(p) = \frac{\Gamma(n-1)}{\Gamma(j)\Gamma(n-1-j)}p^{j-2}(1-p)^{n-2-j}$$

를 따르게 된다. 그리고 이 정보를 이용하여 각각의 기본단위 검정을 하나로 묶어 변화점 존재 $(\mathcal{H}_0)$에 대한 다양한 검정 통계량을 구성 할 수 있게 된다. 아래에서는 이에 대한 두 가지 예를 제시하고 우리의 변화점 검정 문제에서 각 통계량의 의미를 살펴보겠다.

순서화된 유의확률 $P_{t(j)}$가 모수가 $(j,\, n-1-j)$인 Beta분포를 따르므로 $E(P_{t(j)}) = j/(n-1)$임을 알 수 있고 이용하여 $\{P_{t(j)},\, j = 2,3,\ldots,n\}$을 이용하여 전진 계수들이 t년도에 변화점을 갖는다는 가정 $\mathcal{H}_0^{(t)}$를 검정하기 위한 통계량으로 다음의 두 가지 통계량 $T_{1,t}$와 $T_{2,t}$를 고려 할 수 있다.

$$T_{1,t} = \max_{2\le j\le n}\left|P_{t(j)} - \frac{j}{n-1}\right| \quad \text{and} \quad T_{2,t} = \sum_{j=2}^{n}\left|P_{t(j)} - \frac{j}{n-1}\right|.$$

자료에 변화점이 존재한다는 이야기는 $\{H_0^{(t)},\, t = 1,\ldots,T\}$ 중 적어도 하나의 시간에서 진전계수의 분포가 달라짐을 의미하며, 이를 검정하기 위하여 $T_{1,t}$와 $T_{2,t}$의 최대값인 다음의 T_1과 T_2 통계량을 제안한다:

$$\mathbf{T}_1 = \max_{1\le t\le T} T_{1,t} \quad \text{and} \quad \mathbf{T}_2 = \max_{1\le t\le T} T_{2,t}. \tag{3.3}$$

위의 두 통계량에서 보듯이 우리의 "기본단위 검정"의 결과인 $\{p_{t,j},\ t = 1,2,\ldots,T,\ j = 2,3,\ldots,n\}$을 하나로 표현하는 방법은 유일하지 않으며, 실제 문제에 있어서는 각 문제들의 특성을 고려하여 하나를 선택할 것을 제안한다. 실제로 위의 두 통계량을 우리의 전진

계수에서의 변화점 문제에서 고려하게 되면 절대 시간 t 변화가 있다 함은 “모든 경과년도 j”에 있어서 t시점 전후로 변화가 있음, 즉

$$\bigcap_{j=2}^{n}\left\{\mathcal{H}_j^{(t)}\right\}^c$$

을 기대하는 것이고 따라서 몇몇 j들에 민감하게 의존하는 $\mathbf{T}_1$ 보다는 $\{P_{t,j},\ j=2,3,\ldots,n\}$들의 기대오차의 평균을 사용하는 $\mathbf{T}_2$가 보다 적합하다고 사료된다.

이제 변화점이 없다는 귀무가설 $\mathcal{H}_0^{(t)}$하에서, $\mathbf{T}_1$과 $\mathbf{T}_2$ 통계량의 표본분포에 대하여 알아보자. 여기서 $\mathbf{T}_1$ 통계량을 구성하고 있는 $\mathbf{T}_{1,t}$들은 Kolmogorov-Smirov 통계량에 해당되고 $\mathbf{T}_2$를 구성하고 있는 $\mathbf{T}_{2,t}$들은 $P_{t,j}$들의 표본분포와 (0, 1)에서의 균등분포와의 $\mathcal{L}_1$ 거리에 해당되어 일견 이들의 근사 분포가 쉬이 계산 될 듯 하나 $\{T_{1,t},\ t=1,2\ldots,T\}$(또는 $\{T_{2,t},\ t=1,2,\ldots,T\}$)들 사이의 상호 의존성으로 인하여 근사분포의 정확한 계산은 쉽지 않다. 이러한 이유로 이들의 귀무분포의 계산을 위하여 추가적이 컴퓨터 작업이 필요하고 다음의 순열 검정과정을 제안한다.

순열검정과정

(S1) 변화점이 존재하지 않는다는 가정 하에서 모든 $j=2,\ldots,n$에 대하여 $f_{2,j},\ldots,f_{(t-1),j}$가 서로 독립인 같은 분포를 따르므로 $\{1,2,\ldots,T\}$의 임의의 순열 $\pi=(\pi(1),\pi(2),\ldots,\pi(T))$에 대하여 $f_{\pi(1),j},\ f_{\pi(2),j},\ldots,f_{\pi(T),j}$를 얻는다.

(S2) (S1)의 순열된 자료에 대하여 $\mathbf{T}_2$(또는 $\mathbf{T}_1$) 통계량을 계산한다.

(S3) (S1)과 (S2)의 작업을 B번 반복하여 B개의 $\mathbf{T}_2$ (또는 $\mathbf{T}_1$) 통계량 값들을 얻는다.

(S4) 실제 자료로부터 계산된 통계량 값들을 $\mathbf{T}_2^{obs}$(또는 $\mathbf{T}_1^{obs}$)라 할 때 각각의 유의확률은 위에서 구한 순열조합된 표본으로 부터 나온 B개의 $\mathbf{T}_2$(또는 $\mathbf{T}_1$)값들 중 $\mathbf{T}_2^{obs}$(또는 $\mathbf{T}_1^{obs}$) 보다 큰 값들의 비율로 계산된다.

(S5) 마지막으로 위 절차에서 변화점이 존재하는 t의 위치를 찾는 것은, 각각의 t에 대하여 같은 방식으로 순열조합된 표본을 이용하여 $T_{1,t}$(또는 $T_{2,t}$)들을 계산하고 이를 이용하여 $\mathcal{H}_0^{(t)}$에 대한 검정에 대한 유의 확률을 같은 방식으로 계산 하게된다.

4. 실제의 자료분석

이 절에서는 A 보험회사의 건강상품입원비 담보의 실제자료로부터 사고발생년도와 그 후 경과년도(여기서는 경과월별)에 따른 지급보험금이 주어진 삼각분할표로부터, 진전계수에 대한 구조적 변화점 유무를 검정하고자 한다. 먼저 자료로부터 구한 진전계수는 표 4.1과 같다. 위 표 4.1에서 임의의 변화점 t에 대하여 $\mathcal{H}_{0j}=\{\mu_{0j}=\mu_{1j}\}$를 검정할 때, 대표값 μ_{0j}와 μ_{1j}를 추정할 수 있는 적어도 2개 이상의 확률표본을 확보할 수 있도록 $12\le t\le 22$를 설정하였다. 그 결과 식 (3.2)를 계산하면 표 4.2와 같다.

표 4.1: A 보험회사의 건강상품입원비 담보자료로부터 구한 누적보험금의 변화율 (즉, 진전계수 $f_{i,j}$. 이 때, 경과구간은 사건발생후 j월이 $(j+1)$월로 경과한 것을 나타내며, 공란은 현재시점에서는 알 수 없는 미지의 값을 나타냄)

사고년도	경 과 구 간									
	1→2	2→3	3→4	4→5	5→6	6→7	7→8	8→9	9→10	10→11
1	3.654	1.602	1.204	1.059	1.065	1.030	1.044	1.008	1.013	1.007
2	5.458	1.446	1.189	1.088	1.096	1.025	1.023	1.015	1.006	1.002
3	3.964	1.539	1.237	1.080	1.045	1.045	1.029	1.009	1.009	1.002
4	5.454	1.516	1.240	1.123	1.049	1.026	1.032	1.007	1.007	1.004
5	5.782	1.519	1.137	1.105	1.053	1.028	1.024	1.006	1.008	1.003
6	4.636	1.361	1.171	1.078	1.036	1.048	1.021	1.016	1.004	1.006
7	3.786	1.548	1.223	1.063	1.056	1.031	1.021	1.011	1.014	1.005
8	4.667	1.500	1.144	1.088	1.076	1.042	1.035	1.027	1.012	1.003
9	3.908	1.472	1.182	1.076	1.069	1.040	1.037	1.023	1.013	1.011
10	3.974	1.436	1.190	1.069	1.079	1.043	1.020	1.010	1.006	1.004
11	3.714	1.396	1.151	1.083	1.076	1.022	1.037	1.011	1.008	1.004
12	4.909	1.483	1.165	1.112	1.055	1.043	1.023	1.008	1.005	1.004
13	3.909	1.581	1.171	1.044	1.060	1.023	1.018	1.018	1.007	1.010
14	4.370	1.317	1.128	1.094	1.045	1.041	1.036	1.008	1.009	1.014
15	3.919	1.425	1.195	1.066	1.068	1.036	1.025	1.007	1.019	1.004
16	3.662	1.578	1.143	1.062	1.088	1.034	1.031	1.007	1.007	
17	5.648	1.345	1.153	1.073	1.050	1.033	1.035	1.011		
18	3.417	1.468	1.179	1.088	1.045	1.045	1.024			
19	3.836	1.343	1.110	1.055	1.057	1.054				
20	3.508	1.371	1.118	1.094	1.034					
21	3.751	1.311	1.137	1.066						
22	3.344	1.398	1.124							
23	3.792	1.294								
24	3.606									

표 4.2: 표 4.1의 진전계수 자료에 대하여 임의의 변화점 t에 대하여 그 전후 이표본 t검정한 결과의 유의확률 식 (3.2)을 나타냄

변화점(t)	경 과 구 간									
	1→2	2→3	3→4	4→5	5→6	6→7	7→8	8→9	9→10	10→11
12	0.0355	0.0349	0.0011	0.3086	0.9113	0.2703	0.8931	0.0926	0.7999	0.2029
13	0.0526	0.0238	0.0028	0.4453	0.9039	0.4432	0.7421	0.3109	0.6197	0.1071
14	0.0364	0.0014	0.0026	0.3956	0.5428	0.5727	0.9549	0.3195	0.2476	0.0341
15	0.0555	0.0178	0.0015	0.0872	0.2827	0.8271	0.5713	0.7735	0.7271	0.0689
16	0.1163	0.0253	0.0113	0.3743	0.3494	0.3491	0.9817	0.2038	0.9951	0.0840
17	0.0000	0.0000	0.0004	0.1633	0.3609	0.5410	0.5819	0.2262	0.6086	0.0482
18	0.0001	0.0009	0.0017	0.2885	0.6231	0.1368	0.8754	0.2359	0.8834	0.3073
19	0.0000	0.0000	0.0026	0.5431	0.4915	0.2392	0.5166	0.4181	0.9851	0.2185
20	0.0005	0.0000	0.0000	0.6725	0.0092	0.2325	0.9503	0.0255	0.6281	0.1663
21	0.0008	0.0006	0.0000	0.4833	0.0281	0.1546	0.6582	0.0686	0.4503	0.0987
22	0.0038	0.0489	0.0001	0.9527	0.1920	0.0029	0.8273	0.2793	0.4527	0.3980

위 자료의 경우에는 한 개의 경과월에서도 진전계수의 평균값이 다르게 나타나면 다른 것으로 하여, 하나 이상의 경과월에서 서로 다른 패턴을 보여준다면 통계량 T_2가 적절할 것으로 판단된다. 이에, 제 3절에서 제안한 순열검정을 통해 통계량 T_2의 표본분포를 구

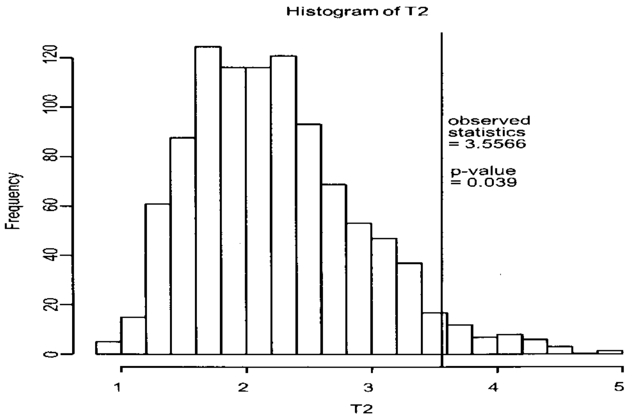

그림 4.1: T_2통계량의 표본분포

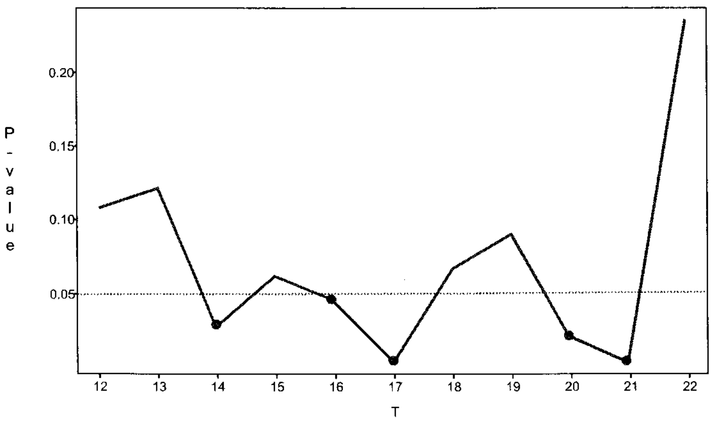

그림 4.2: $T_{2,t}$의 유의확률

하면 다음 그림 4.1과 같고, 통계량 T_2의 관측값은 $T_2 = 3.5566$으로 표본분포로부터 유의확률을 살펴보면 3.9%로 모든 경과년도에서 두 분포가 동일한 것은 아니라고 할 수 있다. 즉, 시간에 따른 진전계수의 변화가 존재 할 수 있다는 것을 의미한다.

이에 변화점 t에 따른 $\mathcal{H}_{0j}^{(t)} = \{\mu_{0j} = \mu_{1j}\}$에 대한 각 검정 결과의 유의확률들을 그림으로 표현하면 다음 그림 4.2와 같다. 그림에서 살펴보면 5개의 변화점 t 전후로 진전계수의 확률분포가 같지 않음을 알 수 있다.

5. 결론

일반적으로 지급준비금은 손해보험회사 부채의 많은 부분을 차지하고 있어 이에 대한

적절한 추정과 평가가 매우 중요하다. 그런데, 지급준비금 추정에 가장 많이 사용되는 사다리법은 서론에 언급한 바와 마찬가지로 보험금 지급년도에 따른 지급보험금의 변화율이 거의 미비하다고 가정하고 있다. 이러한 점을 보완하기 위해 사용된 통계적 모형들의 경우에도 $n \times n$ 삼각분할표에서는 관측 가능한 자료의 수가 $n(n+1)/2$개 이므로, 일반로그선형모형에서처럼 사건의 발생년도와 경과년도 그리고 보험금 지급년도별 추세의 변화를 모두 설명할 수 없다는 것이 잘 알려져 있다. 예를 들어, Barnett와 Zehnwirth (2000)는 이들 추세의 안정성을 가정한 통계적 모형을 연구하였다. 한편 보험료 산출과 관련하여 통계적 분포를 사용하는 것을 보여주는 여러 연구들 (차재형과 이재원 2000; 이정진 등, 1997)이 있다. 본 연구에서는 이러한 통계적 모형에 대한 가정을 미리 검토하기 위한 절차로, 보험금 지급년도별 변화점이 존재하는지 검정할 수 있는 통계량을 제안하고, 이를 실제 자료에 적용해 보았다. 구조적 변화점이 존재할 때, 최근 m년간의 진전계수들로부터 지급보험금에 대한 평균변화율을 구한다면, 변화점을 추정 후 그 이후의 진전계수들로부터 평균변화율을 구하는 것이 타당하다는 객관적 증거를 제시하게 될 것이다.

향후에 지급보험금의 구조적 변화점이 의심되는 경우, 이러한 변화점을 가정하고 지급준비금 추정문제를 연구하는 것이 필요할 것이다.

참고문헌

박잉근 (2006). 사다리법 진전계수에 관한 비교 연구, 석사학위논문, 단국대학교 통계학과.

이정진, 강근석, 이윤오, 김지현, 이창수, 김성철 (1995). 전문가용 한국형 통계패키지 개발연구 I - 생존분석, 베이지안분석, 보험통계를 중심으로, <한국통계학회논문집>, **2**, 434–444.

차재형, 이재원 (2000). 손해액 분포 결정에 따른 보험료 산출, <응용통계연구>, **13**, 247–263.

홍종선, 전홍기 (2006). <신용도 이론>, 자유아카데미.

Ajne, B. (1989). Exponential Run-off Claims Reserving Manual, **2**, Institute of Actuaries.

Barnett, G. and Zehnwirth, B. (2000). Best estimates for reserves, In *Proceedings of the Casualty Actuarial Society*, **87**, 245–303.

Brown, R. L. (1993). Introduction to ratemaking & loss reserving for property & casualty insurance, ACTEX publications, Connecticut.

Kremer, E. (1982). IBNR-claims and the two-way model of ANOVA, *Scandinavian Actuarial Journal*, **1**, 47–55.

Lemaire, J. (1985). *Automobile Insurance Actuarial models*, Kluwer-Nijhoff, Boston.

Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates, *Astin Bulletin*, **23**, 213–225.

Mack, T. and Venter, G. (2000). A comparison of stochastic models that reproduce chain ladder reserve estimates, *Insurance Mathematics and Economics*, **26**, 101–107.

Verrall, R. J. (1994). Statistical methods for the chain ladder technique, *Casualty Actuarial Society Forum*, 393-446.

Verrall, R. J. and England, P. D. (2000). Comments on: "A comparison of stochastic models that reproduce chain ladder reserve estimates" by Mack and Venter, *Insurance Mathematics and Economics*, **26**, 109–111.

Zehnwirth, B. (1991). Interactive Claims Reserving Forecasting System, Version 6.1. Insureware P/L, E. St. Kilda Vic 3 183, Australia.

[2008년 4월 접수, 2008년 6월 채택]

Testing Structural Changes in Triangular Data†

Sungim Lee[1)]

Abstract

The loss reserve is defined as a provision for an insurer's liability for claims or an insurer's estimate of the amount an individual claim will ultimately cost. For the estimation of the loss reserve, the data which make up the claims in general is represented as run-off triangle. The chain ladder method has known as the most representative one in the estimation of loss reserves based on such run-off triangular data. However, this fails to capture change point in trend. In order to test of structural changes of development factors, we will present the test statistics and procedures. A real data analysis will also be provided.

Keywords: Run-off triangle; chain ladder method; structural changes.

† This research was supported by the research fund of Dankook University in 2006–2007.

1) Assistant Professor, Department of Statistics, Dankook University, 126 Jukjeon-Dong, Suji-Gu, Yongin, Gyeonggi-Do 448-160, Korea. E-mail: silee@dankook.ac.kr