

Cox 비례위험모형을 이용한 변액연금 해지율의 추정

김유미¹ · 이항석²

¹보험연구원 · ²성균관대학교 보험계리학과/수학과

접수 2013년 5월 22일, 수정 2013년 6월 17일, 게재확정 2013년 6월 22일

요약

해약율의 추정은 최근 보험제도의 변화 (국제회계기준의 도입에 따른 현금흐름방식의 가격산출체계 시행, 무해약환급금 보험상품의 판매 허용 등)에 따라 보험료의 결정과 손익분석 그리고 리스크 관리 등에 있어서 중요한 요소로 부각되고 있다. 특히, 변액연금은 최저보증옵션으로 인하여 보험계약자의 해약요소가 중요시되고 다른 보험 상품에 비해 복잡하므로 차별성 있는 통계모형의 선택과 분석이 필요하다. 기존의 해약률 연구는 실태분석 또는 회귀분석을 위주로 모형화하는 것에 초점이 맞추어져 있었으나 본 연구에서는 변액연금 계약과 관련된 여러 변수와 최저보증옵션을 반영하기 위하여 생존분석기법 중 하나인 Cox 비례위험모형을 이용하여 해지율을 추정하였다. 변액연금 해지율에 영향을 미치는 주요변수로는 납입방법, 보험료, 보험기간 대비 유지기간, 계약자적립금 대비 최소보증금, 계약자연령이 있으며 본 연구에서는 이에 관하여 분석해보았다.

주요용어: 변액연금, 생존분석, 콕스 비례위험모형, 해지율.

1. 서론

보험 상품의 보험료결정과 손익분석 및 리스크 관리 측면에서 해지율 (해약율)의 추정은 실무적으로 매우 중요하다. 특히 최근의 변화를 살펴보면 2014년 4월에 도입되는 현금흐름방식의 가격산출체계 하에서는 해지율의 모형 선택이 보험료 결정의 핵심요소가 된다. 또한 국제보험회계기준 도입에 따라 보험부채 평가 시 시가로 평가하게 되어 해지율 중심의 리스크 관리가 필요하다. 그리고 2010년 4월부터는 무해약환급금 상품의 개발 및 판매가 허용되어 해지율의 합리적인 모델링이 보험료의 결정과 상품손익분석의 핵심요인이 되었다. 이에 따라 보험 상품 형태별, 계약자속성별 해약추이 및 원인 등을 반영한 정형화된 해지율 모형에 대한 필요성이 높게 요구되고 있다. 특히 보험 상품 중에서도 변액연금의 경우 노후생활자금의 보호 등을 위하여 제도상 적립금의 최저 수준을 보장해주고 있어 변액연금의 해지율의 계산에는 최저보증옵션을 감안해야하므로 다른 보험 상품의 해지율 모형보다 더 많은 설명변수가 필요하게 되어 변액연금상품의 해지율 모형의 설계는 더욱 복잡하다. 예를 들면, 적립금과 최저보증 수준과의 차이로 인하여 계약자의 행동이 변할 수 있으므로 해지율 모형에 적립금과 최저보증과의 차이에 대한 설명변수가 반영될 필요가 있다.

또한 투자형 상품의 경우 고정적인 예정이율을 보장해주는 상품과 달리 펀드수익률에 따라 투자적립금의 수준이 달라지기 때문에 계약자가 예상하는 투자적립금 수준과 실제 적립금의 차이가 계약자 행동에 영향을 주게 된다. 이와 같은 점을 감안하여 본 논문에서는 투자형 보험상품인 변액연금을 분석대상

¹ (150-010) 서울시 영등포구 여의도동 35-4, 보험연구원, 연구원.

² 교신저자, (110-745) 서울시 중로구 명륜동 25-2, 성균관대학교 보험계리학과/수학과, 부교수.

E-mail: hangsuck@skku.edu

으로 한정하였으며 본 연구의 목적은 이러한 계약자의 행동을 반영한 해지율을 추정해 봄으로써 변액연금의 리스크 관리에 도움이 되고자 하는 것이다

생명보험의 해약을 모델링에 관한 연구는 1990년 이후 활발히 진행되기 시작했다. 90년대 초에 Joh (1990)은 유지율의 실태를 계약내용, 계약자의 속성, 계약상황 세 가지 측면에서 분석하였고, Chung (1992)은 일본과 미국의 유지율 비교와 설문조사 등을 통한 단순한 통계기법만을 이용하여 생명보험의 해약 원인을 찾고자 하였다. 이와 같이 우리나라 생명보험의 해약 예측 분석은 자료에 근거한 실태분석이 주를 이루었다. 이 외에도 보험회사의 실효-해약에 관한 분석을 다룬 논문은 Jung과 Ouh (2007), Jung 등 (2011), Kim (2008), Ryu와 Jang (2009), Lee (2009) 등이 있다.

Jung과 Ouh (2007), Jun 등 (2001)은 명목국민총생산, 금리, 실업률 등 거시변수를 이용하여 생명보험의 해지율에 미치는 영향을 그랜저 인과관계검정 및 벡터오차수정모형을 이용하여 규명하였다. Kim (2008)은 로지스틱모형을 이용한 해약모형을 연구하였다. Ryu와 Jang (2009)은 효력·상실 데이터를 기초로 하여 해지율의 실태 및 그 특징을 연령별, 성별, 납입방법별 등으로 분류하여 제시하였으며 이를 바탕으로 거시경제변수에 따라 보험종목별로 효력상실-해지율에 미치는 영향을 보였다. 또한 Hwang과 Lee (2010)은 생명보험상품별 해지율 추정 및 예측모형을 연구하였다. Lee (2009)은 탈퇴원인별 소수 연령분포가 상이하다는 가정 하에 절대탈퇴율에서 다중탈퇴율로의 공식 전환을 유도하였다. 또한 이를 이용하여 계약자적립금과 최소보증금액의 수준에 따라 달라지는 변액연금보험의 해지현상을 반영하기 위해 동적해지율이 적용된 다중탈퇴율의 전환과정을 설명하였다. 동적해지율에 대한 자세한 개요는 Kwon (2011)을 참조하기 바란다.

기존 보험계약의 해지율 연구와 본 연구와의 차별성은 두 가지이다. 첫 번째는 데이터 분석의 접근이 기존 연구에서는 합산된 자료를 이용하여 분석하였지만 본 연구에서는 변액연금의 개별 건별 계약 데이터를 활용하여 분석한다. 두 번째는 기존의 연구가 실태분석 또는 회귀분석에 초점이 맞추어져 있었지만 본 연구에서는 계약 데이터의 관측기간이 91개월이므로 중도절단(censoring)을 포함하고 있기 때문에 생존분석기법 중 하나인 Cox 비례위험모형을 이용하여 추정하는 것이 타당성 있다. 또한 부록에 있는 그래프에서 확인할 수 있듯이 공변수(covariates)별로 유지율(생존율)의 차이가 있음이 나타나기 때문에 이를 반영하여 모형에 적용하는 것이 필요하다.

본 논문은 다음과 같이 구성되어 있다. 1절에서는 본 논문의 연구배경을 밝히고 선행연구 및 해지율 추정의 필요성에 대해 살펴보았다. 2절에서는 생존분석의 개요 및 Cox 비례위험모형에 대하여 설명하였다. 3절에서는 실증분석을 통해 해지율에 영향을 미치는 변수를 분석하고 통계분석의 결과를 제시하였으며, 4절에서는 결론 및 시사점을 도출하였다.

2. Cox 비례위험모형의 이론적 배경

이 절에서는 Cox 비례모형으로 분석하는데 필요한 생존분석의 기본성질을 먼저 논의하고 Cox 비례모형의 기본성질을 살펴본다. 또한 보험의 해지율과 유지율의 측면에서 논의를 추가하여 전개하고자 한다. 생존분석에서 생존시간은 어떤 사건(event)이 일어날 때까지의 시간을 말하며 본 연구에서 사건은 해지(해약)이며 시간은 계약자에 의해서 생명보험계약이 해지될 때까지의 시간을 의미한다. 생존분석과 관련된 자세한 논의는 Song 등 (2007), Kang과 Lee (2009), Park 등 (2012) 및 Fox (2002) 등을 참조하기 바란다.

위험함수(hazard function, $h(t)$)는 시점 t 까지는 해지하지 않았다는 가정 하에 t 시점 바로 직후에 순간적으로 해지할 조건부확률로 정의된다. 이를 식으로 표현하면 다음과 같다.

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \quad (2.1)$$

여기서 확률변수 T 는 생존기간 (보험에서 계약의 유지기간)을 의미하며 연속형 확률변수이다. 여기서 $S(t)$ 는 생존함수이고 $f(t)$ 는 확률밀도함수이다. 생존함수와 위험함수는 다음과 같은 형태로도 나타낼 수 있다.

$$S(t) = e^{-\int_0^t h(u)du} \tag{2.2}$$

본 연구에서 Cox 비례위험모형은 계약자의 해지율을 모형화하는 통계적 모형으로서, 계약자의 탈퇴 (해지)에 작용하는 요인의 영향을 파악하고 시간에 따른 계약자의 유지율을 추정하는데 사용할 수 있다. 비례위험모형의 가장 큰 특징은 공변수의 특성을 표현하는 함수와 기저함수를 분리하여 탈퇴에 영향을 주는 요인들 간의 상대적 효과를 산정하는 것이 가능하다는 것이다. 또한 기저함수의 추정과정에 있어서 계약자의 탈퇴시간에 대한 다양한 계약자의 탈퇴경향을 모형화 할 수 있다는 장점이 있다.

Cox 모형은 t 시점에서의 로그위험함수를 여러 설명변수 (이하 공변수라 일컫는다.)들의 선형식으로 표현한다. 즉, 각 계약자의 설명변수의 다른 값들, 다시 말하면 i 번째 계약자의 설명변수 값이 $x'_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ 일 때, Cox모형은 위험함수를 다음과 같이 표현한다.

$$h_i(t|x_i) = h_0(t) \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \tag{2.3}$$

여기에서 $h_i(t|x_i)$ 는 시간 t , 개별 계약자 i 의 공변수 벡터 x_i 에서의 개별위험함수이며, $h_0(t)$ 는 시간 t 에서의 기저위험함수 (baseline hazard)로 공변량들의 값들이 모두 0일 때의 위험함수의 값이다. 위험함수와 생존함수의 관계는 다음과 같다.

$$S_i(t) = \exp\left(-\int_0^t h_i(s)ds\right) = \exp\left(-\int_0^t h_0(s) \exp(\beta' x_i)ds\right) = [S_0(t)]^{\exp(\beta' x_i)} \tag{2.4}$$

생존함수 $S_0(t)$ 를 추정한 후 최대편우도함수 (maximizing partial likelihood function; LR)로 추정된 회귀계수 β' 을 이용하여 식 (2.4)를 추정할 수 있다.

$$\hat{S}_0(t) = \prod_{j=1}^k \left(1 - \frac{\exp(\hat{\beta}' x_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\hat{\beta}' x_{(j)})}\right)^{\exp(-\hat{\beta}' x_{(j)})} \tag{2.5}$$

여기서 $j = 1, 2, \dots, r, k = 1, 2, \dots, r - 1, R(t_{(j)})$ 는 j 번째 탈퇴시간 ($t_{(j)}$)의 직전까지 탈퇴하지 않은 계약자 집합, 즉 $t_{(j)}$ 에 탈퇴가 일어날 가능성이 있는 계약자들의 집합인 위험집합 (risk set)을 나타낸다. 탈퇴시간의 총 개수가 r 이며, 그 순서가 $t_{(1)} < t_{(2)} < \dots < t_{(r)}$ 일 때 기저생존함수는 각 탈퇴시간에 로그변환로그함수를 이용하여 생존율을 추정하여 계산된다.

Cox의 모형은 공변수의 효과를 모형 안에서 고려함으로써 이 공변수들이 위험률에 비례적으로 변한다는 것을 분석할 수 있다. 즉, 두 관측의 위험률 비가 시간에 대해 일정하다는 것을 의미하므로 공변수의 서로 다른 값에서 $\log(-\log(S(t)))$ 는 t 시점에 관계없이 일정하게 된다. 예를 들어 계약자 i 와 계약자 j 의 위험함수의 관계는 다음과 같다.

$$\begin{aligned} \frac{h_i}{h_j} &= \frac{h_0(t) \exp(\beta_1 x_{i1} + \dots + \beta_p x_{ip})}{h_0(t) \exp(\beta_1 x_{j1} + \dots + \beta_p x_{jp})} \\ &= \exp(\beta_1(x_{i1} - x_{j1}) + \dots + \beta_p(x_{ip} - x_{jp})) \end{aligned} \tag{2.6}$$

기저위험함수 $h_0(t)$ 는 서로 상쇄되어 없어지기 때문에 위험함수는 시간에 대해 일정한 관계가 됨을 알 수 있다.

위에서는 공변수가 연구 시작부터 알려져 있고 그 영향이 변함없는 공변수만을 고려하였다. 그러나 이 효과가 시간에 따라 변해간다면 그 공변수 (x)와 시간 (t)의 교호작용 (즉 xt)이 모형에 추가되어야 하며 시간함수로 표현된 이 교호작용이 삽입된 모형은 비례성 가정이 성립되지 않는다. 따라서 생존시간에 따라 위험률의 비가 변하는 변수에 대한 유의성 검정은 바로 위험함수의 비례성 가정에 대한 검정이 되는 셈이다.

공변수 z 에 대해 비례성 가정이 성립된다면 $z = 0$ 인 군과 $z = 1$ 인 군의 위험함수는 서로 비례적인 관계 즉, 시간에 대해 무관한 상수의 관계가 유지되어야 한다. 이때 위험함수는

$$h(t) = h_0(t) \exp(\beta z) \quad (2.7)$$

으로 표현되며 두 위험군의 위험비 (hazard ratio)는 상수 $\exp(\beta)$ 가 된다. 이 공변수에 대한 비례성 가정이 성립되지 않아 시간에 따라 변하는 변수임이 판명된 경우에는 이를 모형에 추가한다. 즉,

$$h(t) = h_0(t) \exp(\beta z + \gamma t) \quad (2.8)$$

에서 $x = \gamma t$ 로서 공변수와 시간의 교호작용을 나타낸다. 따라서 시간에 따라 변하는 변수를 반영한 위험군의 위험비는 $\exp(\beta + \gamma t)$ 로 시간에 따라 변하므로 비례성 가정이 성립하지 않는다. 이와 같이 시간에 따라 변하는 변수 (time dependent variable)에 대한 유의성검정은 비례성 가정을 검토한 후 층화 모형을 통해 다시 분석한다.

전체 연구기간에 걸쳐 비례성가정에 어긋나는 공변수 일지라도 자료의 일부에서는 비례성 가정이 성립되는 경우가 있으므로 이 공변수는 s 개의 층화군 (strata)으로 나누어 다음과 같이 층화 모형 (stratified model)을 적용한다. 층화 모형의 위험 함수는 다음의 형태로 표현된다.

$$h_j(t; x) = h_{0j}(t) \exp(\beta' x) \quad (j = 1, 2, \dots, s) \quad (2.9)$$

식 (2.9)를 살펴보면 비례성가정에 문제가 없는 공변수 x 의 회귀계수 β' 는 모든 층화군에서 동일하지만, 각 층화군에서의 기저위험함수 $h_{0j}(t)$ 는 서로 다른 형태를 가지도록 허용된 모형이다. 즉, 서로 다른 층화군에서는 비례성 가정이 성립되지 않아도 되는 셈이다. 층화 모형에서 추정하고자 하는 회귀계수 β' 는 각 층화군의 우도비의 모든 층화군에 걸친 합이 최대가 되도록 추정하게 된다.

3. 실증분석

3.1. 실증분석의 개요

생존시간의 최종적인 비례위험모형을 구축하기 위하여, 먼저 모든 공변수에 대한 단일분석을 통하여 생존확률에 영향을 준다고 판단되는 변수를 일차적으로 선정하였다. 선정된 공변수를 대상으로 공변수간의 상호작용을 검토한 후

$$-2\ln LR = -2\ln(L_0/L_1)$$

를 계산하고 각 생존시간 군에 대한 공변수와 그 회귀계수 값을 추정하였다. 각 생존시간 군에 대해 선정된 공변수들의 위험률에 대한 비례성 검토를 실시하여 비례성 가정에 위반될 경우에는 그 공변수를 시간종속형 변수로 추가하여 반영한다. 모형에 선정된 공변수와 그 회귀계수값을 결정한 후 기저생존함수 및 기저위험함수를 추정하면 생존시간에 대한 비례위험모형의 구축이 완료되며, 모형에 포함될 공변수의 선택과 최대우도 추정법을 이용한 회귀계수의 추정은 통계분석용 소프트웨어인 SAS System을 이용하였다.

3.2. 자료의 구성

앞서 소개한 방법론을 이용하여 실제 자료들을 바탕으로 검증해 보았다. 실증분석에 사용된 자료는 국내 생명보험사의 변액연금상품으로 2004년 05월부터 2011년 11월까지 표본으로 선정된 12,412명의 생명보험 계약자 데이터이다. 생명보험 해지에 따른 계약자의 이탈 여부는 기준 월 당시 유지하고 있는 계약이 해약 또는 효력상실 할 경우로 정의한다. 분석의 편의상 이상 값이 관측되는 경우 제거해주었다. Table 3.1은 원본데이터의 구성을 보여주고 있다.

Table 3.1 Source data

	Persistency	Lapse	Total
Frequency	11483	626	12109

Table 3.2는 원본데이터의 기술통계량을 보여준다. 여기서 가입금액과 보험료, 기납입보험료, 계약자 적립금의 단위는 백만원이다. 주요 변수 중 하나인 납입방법은 이산형 변수이기 때문에 Table 3.2에서 생략하였다.

Table 3.2 Descriptive statistics (million)

	Average	Median	Mode	STDEV	Kurtosis	Skewness	Min	Max
Benefit amount	21.159	13.2	12	20.739	113.975	6.926	5	620
Premium	3.367	0.15	0.1	16.226	346.582	14.040	0.1	620
Policyholder age (year)	40.101	40	39	11.051	0.244	0.029	-1	82
Insured age(year)	37.174	37	39	10.289	-0.682	0.011	15	63
Sum of premiums	9.019	4.62	10	17.450	247.154	11.093	0.1	620
Accumulation value	6.515	3.031	0.508	12.639	138.063	8.706	0.075	323.349
Duration/deferred period	0.455	0.343	0.333	0.382	1.249	1.302	0.012	2.048
Accumulation/guarantee	1.470	1.468	1	0.305	-1.115	-0.030	0.499	2.000

Figure 3.1은 Kaplan-Meier method에 의하여 그려진 유지율의 그래프이다. 즉, 공변수를 고려하지 않았을 때의 표본에 포함된 데이터의 생존율이 어떻게 분포되어 있는가를 나타낸 것이다. 처음 100%의 생존율에서 마지막 사건이 발생한 91개월에는 생존율이 82%로 나타났고 91개월 이후 중도절단이 되어 82%로 유지하게 된다.

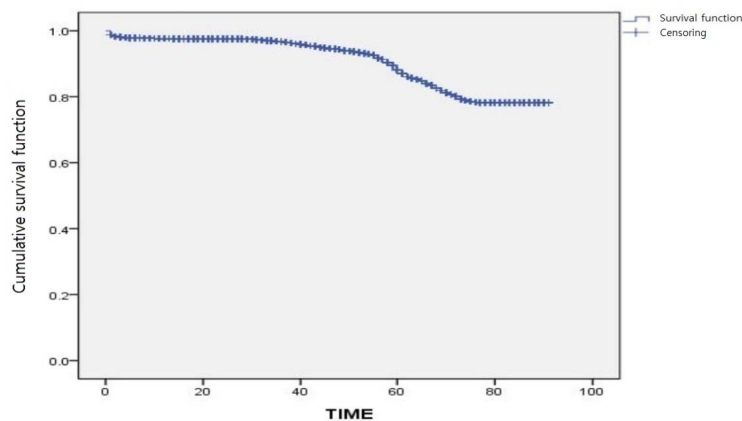


Figure 3.1 Kaplan-Meier survival function

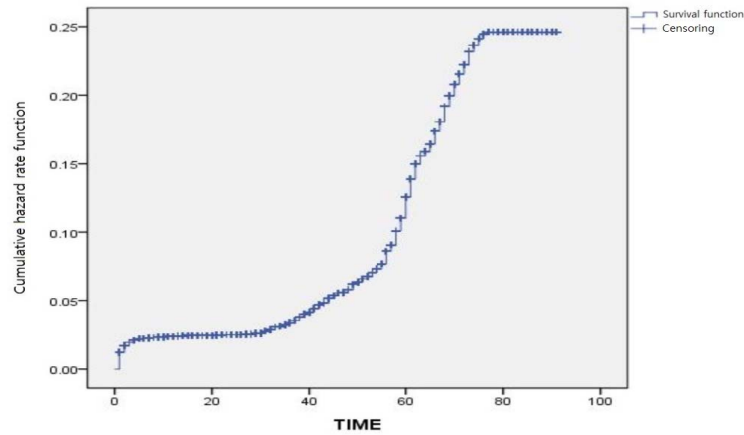


Figure 3.2 Hazard function

비례위험모형의 구축과정은 각 모형에 포함될 공변수를 선택하고 공변수의 회귀계수를 추정한 후 기저생존함수 및 기저위험함수의 추정으로 진행된다. 일반적으로 생존시간에 대해 가장 적합한 비례위험 모형은 해지가 발생할 때까지의 시간에 유의하게 영향을 미치는 공변수를 모두 포함하면서 포함된 공변수의 개수가 가장 적은 모형을 말한다. Table 3.3은 모형의 구축과정에서 일차적으로 고려된 공변수의 정의와 각 계약자에게 부여될 수 있는 공변수를 나타낸다. 우선 특정 기간 사이 (2004.05~2011.11)에 해약 또는 효력상실이 발생한 계약자를 종속변수로 두었으며 해약 또는 효력상실이 발생한 계약자의 경우 '1'의 값을 입력하였고 유지하고 있는 계약자의 경우 '0'의 값을 입력하여 이분형으로 나누었다.

기납입보험료와 계약자적립금은 관행상 계약이 해지된 시점에 산출되기 때문에 유지되고 있는 계약에 대한 가정이 필요하였다. 기납입보험료의 경우 계약시점부터 현시점 (2011.11.15)까지 보험료가 100% 완납되었다고 가정하였다. 변액연금에 경우 계약자적립금과 최소보증금 수준이 해약여부에 영향을 미친다고 볼 수 있기 때문에 계약자적립금 대비 최소보증금의 비율을 공변수로 포함시키고자 한다. 하지만 유지되고 있는 계약에 대해서는 계약자적립금의 수준을 알 수 없음을 지적한 바 있다. 따라서 해약건에 대한 계약자적립금 대비 최소보증금의 수준을 분석한 후 유지건에 대한 수준을 추정하여 가정하였다.

Figure 3.3은 해약건에 대한 계약자적립금 대비 최소보증금을 그래프로 나타낸 것이다. 해약건인 경우 계약자적립금 대비 최소보증금의 비율이 대체적으로 1이하인 것으로 나타났다. 이는 기납입보험료에 비해 계약자적립금이 많이 쌓여 있어 계약자가 해약의 의사를 보인 것으로 해석할 수 있다. 따라서 유지건에 대해서는 계약자적립금 대비 최소보증금의 비율을 1이상의 값으로 하되 난수를 생성하여 결정하였다. 기납입보험료는 계약자적립금 대비 최소보증금의 비율을 이용하여 역으로 산출하였다.

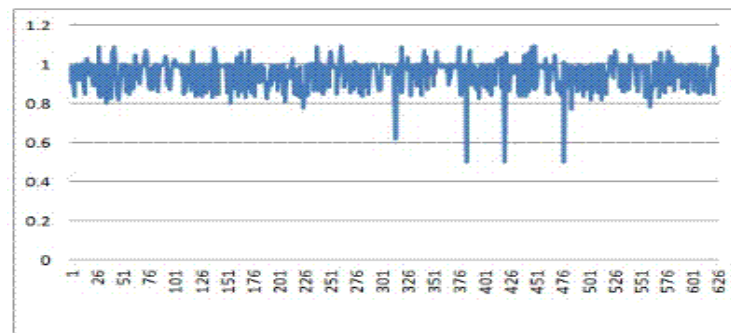


Figure 3.3 The ratio of guaranteed amount to account value

이 외에도 단순히 보험기간, 유지기간, 납입기간이 아닌 보험기간 대비 유지기간의 비율이 생명보험의 해약·효력상실에 영향을 미치는 주요 변수인지 알아보기 위하여 유지기간에 보험기간을 나누어 보험기간 중 유지기간이 몇 %인지를 나타내는 변수를 생성하였다. 실증분석에 앞서 모형의 특성을 잘 설명해 줄 수 있는 공변수를 알아보기 위하여 다음과 같은 기초분석을 실시하였다. 5개의 보험계약관련 변수, 1개의 계약자관련 변수, 1개의 피보험자관련변수 그 외에 새로 생성한 2개의 신계약변수를 채택, 총 9개의 변수들을 1차 생존예측모형의 분석대상변수로 선정하였다.

Table 3.3 Variables used for Cox model

Name	Description	Name of code
Decrement due to lapse	0: no lapse1: lapse	LAPES
Payment frequency	1: monthly2: lump sum	NIBB
Benefit amount	Continuous (million)	GIGA
Premium	Continuous (million)	BHR
Sum of premiums	Continuous (million)	GNIBHR
Accumulation value	Continuous (million)	JGG
Policyholder age	Continuous	KYJAGE
Insured age	Continuous	PBJAGE
Duration/deferred period	Continuous	UZGG.BHGG
Accumulation/guarantee	Continuous (million)	GV_AV

3.3. 공변수의 선정과 공변수 간의 상호작용 검토

탈퇴요인을 보다 잘 설명할 수 있는 모형을 구축하기 위해서는 공변수의 가능한 모든 조합 및 공변수 간의 상호작용 그리고 시간종속 공변수에 대한 검토를 수행하여야 한다. 먼저, 공변수에 대한 단일분석을 통하여 각 변수가 생존율과 연관이 있는지를 살펴보고, 생존율에 영향을 준다고 판단되는 변수를 일차적으로 선정하였다. Table 3.4은 일차적으로 유의한 변수를 선별하기 위해 실시한 단일분석의 결과를 보여주고 있다. 단일분석 결과 계약자적립금을 제외한 공변수들이 유의수준 5%에서 생존율에 영향을 준다고 판명되었다.

Table 3.4 Estimated coefficients of trial model and their statistics

Analysis of maximum likelihood estimates					
Variable	Parameter estimate	Standard error	Chi-square	Pr>chis	Hazardratio
Benefit amount	0.00144*	0.0005049	8.1592	0.0043	1.001
Premium	0.00863*	0.0002913	877.1386	<.0001	1.009
Payment frequency	1.50860*	0.02893	2718.6177	<.0001	4.520
Policyholder age	0.000929*	0.0008969	107.1962	<.0001	1.009
Insured age	0.00695*	0.0009618	52.1410	<.0001	1.007
Sum of premiums	0.00298*	0.0006236	22.8565	<.0001	1.003
Accumulation	-0.000768	0.0009511	0.6526	0.4192	0.999
Duration/deferred period	-11/48479*	0.10868	11166.2984	<.0001	.000
Accumulation/guarantee	0.32559*	0.03050	114.0397	<.0001	1.385

*significant at $\alpha = 0.05$

Table 3.5에서는 최종 선정된 변수들의 계수와 표준오차, Wald 통계량과 유의확률 그리고 상대적 위험도를 나타내는 위험비를 보여준다. 여기에서 회귀계수가 양수이면 계약자의 위험이 증가하는 것이고 음수이면 감소함을 의미한다. 일차적으로 유의하게 선택된 9개의 변수에 대해서 분석을 실시하였으며 변수선택방법은 전진 (forward) 방법을 사용하였다.

Cox 비례위험모형에서 공변수들의 회귀계수는 최대편우도함수를 통해 추정한다. 공변수를 포함한 경우와 포함하지 않은 경우를 살펴보아 설정한 모형이 타당한가를 검정하는 것을 LR (likelihood ratio) 검정이라고 하며 아래 Table 3.5에서 보는 바와 같이 공변수를 포함하지 않은 초기 로그우도값에서 공변수를 포함한 로그우도값을 빼 값인 2830.0518이 LR값을 뜻한다. 이는 유의수준 1%에서 유의한 것으로 나타났으므로 모형에서 공변수들을 포함하는 것이 의미 있는 것으로 판명되었다.

전진방법에 의해 5단계의 선택과정을 거쳐 최종적으로 선택된 변수는 유의수준 5%이내에서 남입방법, 계약자연령, 보험기간 대비 유지기간, 보험료, 계약자적립금 대비 최소보증금이 선정되었다.

Table 3.5 Test statistics and coefficients of final model

Criterion	Model fit statistics	
	Without covariates	With covariates
-2 LOG L	193873.70	191043.65
AIC	193873.70	191055.65
SBC	193873.70	191099.74

Testing global null hypothesis: beta=0		
Test	Chi-square	Pr>chisq
Likelihood ratio	2830.0518	<.0001
Score	4644.8706	<.0001
Wald	3877.2613	<.0001

Variable	Parameter estimate	Standard error	Chi-square	Hazard ratio	95% Hazardratio confidence limits	
					Lower	Upper
					Payment frequency	1.35969
Policyholder age	0.00479	0.00089	28.7622	1.005	1.003	1.007
Duration /deferred period	-0.02998	0.00198	229.408	0.970	0.967	0.974
Premium	0.03724	0.00173	465.959	1.038	1.034	1.041
Accumulation/guarantee	0.20558	0.03121	43.3979	1.228	1.155	1.306

최종모형으로 선택된 공변수를 대상으로 상관분석을 실시해보았다. 상관분석결과 대부분의 공변수들이 양의 상관관계를 보였으며 계약자적립금 대비 최소보증금과 보험기간 대비 유지기간은 음의 상관관계를 보였다. 이는 Table 3.6과 같다. 상관계수가 0이 아니라는 가설은 채택 가능하지만 이는 데이터 개수가 많아서 유의하다.

Table 3.6 Correlations among covariates

	Payment frequency	Premium	Duration/deferred period	Policyholder age
Premium	0.134**			
Duration / deferred period	0.030**	0.248**		
Policyholder age	0.221**	0.058**	0.031**	
Accumulation / guarantee	0.040**	0.016*	-0.024**	0.004*

*significant at $\alpha = 0.05$, **significant at $\alpha = 0.01$

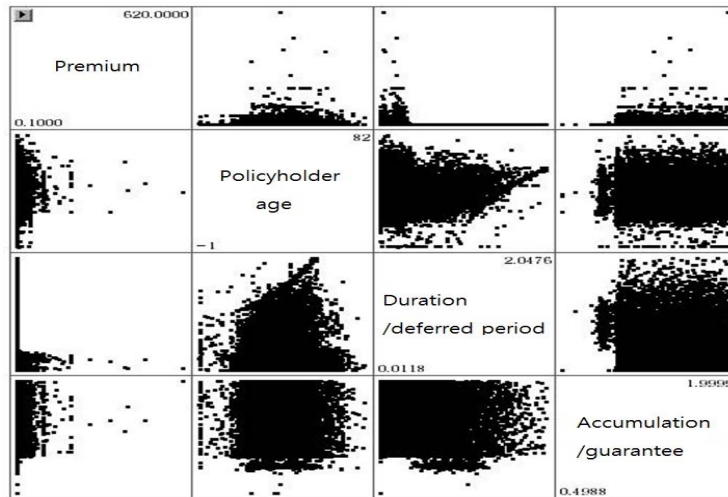


Figure 3.4 Scatter plots among variables

3.4. 공변수의 비례성 가정 검토

비례위험모형에서 채택하고 있는 가장 중요한 가정은 공변수의 위험률에 대한 비례성 가정이다. 즉, 비례위험모형 내의 하나의 공변수가 증가 혹은 감소할 경우 그에 대한 위험률도 공변수의 증감에 비례하여 변화하고, 이러한 경향은 시간에 관계없이 일정하게 유지된다고 하는 것이다. 따라서 만약 모형 구축에 유의한 것으로 판단된 공변수라 할지라도 비례성이 확인되지 않은 공변수의 경우에는 시간에 따라 변화하는 시간중속형 공변수로 모형에서 추가적으로 고려하여야 한다.

비례성 가정을 검토하고자 공변수가 하나일 때 그 공변수에 따라 나누어진 각 부분군에서 Kaplan-Meier 생존율 $S(t)$ 를 구한 후 이를 이용하여 $\ln[-\ln S(t)]$ 를 그래프로 그려보았다. (이하 LLS 그래프라 일컫는다.) LLS 그래프가 직선이든 곡선이든 대체적으로 평행한 패턴을 보이게 되면 그 공변수는 비례성가정에 적합하다고 판단 할 수 있다. Figure 3.5는 공변수별로 LLS의 그래프를 나타낸다. 납입방법의 경우 완전 평행하여 비례성 가정에 완벽하게 부합하였으며 그 외에도 선이 교차할 정도로 비례성가정이 심하게 위배되는 공변수는 없는 것으로 보인다. 따라서 납입방법, 보험료, 보험기간 대비 유지기간, 계약자적립금 대비 최소보유금, 계약자연령에 대해서는 증화분석을 시도할 필요가 없는 것으로 판단된다.

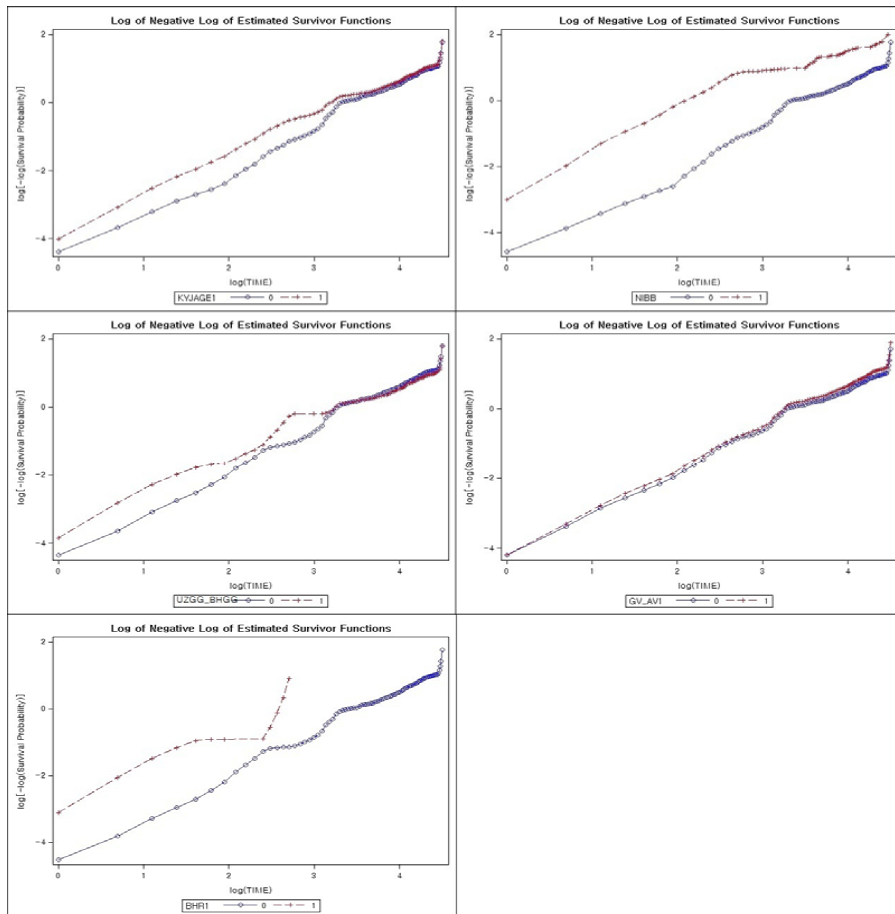


Figure 3.5 Log likelihood ratio by covariates

3.5. 기저함수의 추정

비례위험모형의 기저위험함수는 LLS라고 하는 각 탈퇴시점에 대한 기저생존확률의 추정치의 로그-로그 변환 값 ($\ln[-\ln \hat{S}_0(t)]$)을 적절한 함수형으로 근사시켜서 구한다. LLS를 산정한 결과의 일부는 Table 3.7과 같으며 이를 그래프로 표현하면 Figure 3.6과 같다. 시간과 LLS값은 선형과 유사한 관계를 보여 Weibull 분포를 따른다고 볼 수 있다.

$$LLS(t) = 1.3155 \ln(t) - 4.4883 \tag{3.1}$$

이를 기저생존함수로 변환하면 다음과 같다.

$$\hat{S}_0(t) = \exp(-t^{1.3155} \cdot e^{-4.4883}) \tag{3.2}$$

Table 3.7 Survival function and log likelihood ratio

Payment frequency	Premium	Duration /deferred period	Policy holder age	Accumulation /guarantee	t	S	LLS
0.125	3.636	9.019	40.10	1.469	0	1	.
0.125	3.636	9.019	40.10	1.469	1	0.98862	-4.470
0.125	3.636	9.019	40.10	1.469	2	0.97366	-3.623
0.125	3.636	9.019	40.10	1.469	3	0.95515	-3.081
0.125	3.636	9.019	40.10	1.469	4	0.93703	-2.732

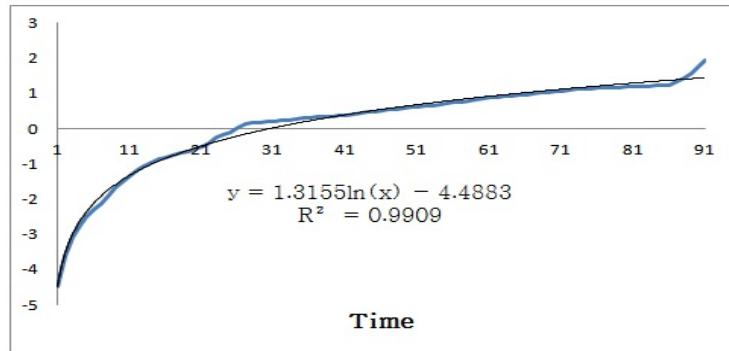


Figure 3.6 Fitted log likelihood ratio of base-line survival function

최종적으로 Table 3.5에서 구한 공변수 및 그 회귀계수와 기저생존함수를 이용하여 생존함수를 표현하면 다음과 같다.

$$\hat{S}_i(t) = \exp(-t^{1.3155} e^{-4.4883}) \exp(1.35969NIBB + 0.03724BHR - 0.02998UZGG_BHGG + 0.20558GV_AV + 0.00479KY_JAGE) \tag{3.3}$$

여기서, NIBB: 납입방법 BHR: 보험료 UZGG_BHGG: 유지기간/보험기간 GV_AV: 계약자적립금 대비 최소보증금 KYJAGE: 계약자연령

3.6. 분석결과의 해석

Wald 추정량은 모형에서 각 공변수가 얼마나 공헌하고 있는가를 나타내며 Wald 추정량이 클수록 모형에서 변수의 공헌도가 크다는 것을 의미한다. 분석결과 중요변수는 납입방법, 보험료, 유지기간/보험기간, 계약자적립금 대비 최소보증금, 계약자연령 순으로 나타났다.

첫째, 납입방법의 경우 변수입력 시 월납의 경우에는 ‘0’, 일시납의 경우에는 ‘1’로 입력하였다. 회귀계수는 1.35969로 양의 관계로 나타났으며 위험비가 3.895로 95% 신뢰구간에서 1을 포함하지 않기 때문에 이 비율은 유의한 것으로 볼 수 있다. 3.895가 의미하는 바는 계약자가 월납보다 일시납을 선택할

경우 위험률이 3.9배 더 높음을 뜻한다. 이는 일시납 계약자의 경우 이미 보험회사에 납입한 보험료가 고액이기 때문에 운용성과에 더욱 민감하게 반응할 수 있으므로 이는 다른 금융상품으로 자산운용방식을 이전할 요인으로 작용할 수 있다.

둘째, 보험료의 경우 회귀계수는 0.03724로 양의 관계를 보였으며 위험비는 1.038로 산출되었다. 이는 보험료 한 단위 (백만원)가 증가하면 위험률이 3.7% 증가함을 의미한다. 또한 95% 신뢰구간에서 1을 포함하지 않기 때문에 이 비율 역시 유의하다고 볼 수 있다. 이는 보험료의 액수가 커질수록 계약자에게는 경제적인 부담으로 작용할 수 있기 때문에 해약요인이 증가하는 것으로 이해할 수 있다.

셋째, 보험기간 대비 유지기간은 위험률과 음의 관계를 보였으며 이는 보험기간 중 유지기간이 차지하는 비율이 증가할수록 위험률이 감소함을 뜻한다. 또한 위험도를 나타내는 위험비 역시 0.970 ($1 - 1.970 = 0.03$)로 산출되었으므로 비율이 한 단위 증가할 때마다 위험률이 0.97배 증가하며, 달리 말해 3%씩 위험이 감소함을 의미한다. 이는 분석 데이터 중 상당수가 계약기간 초기에 해약이 몰려 상대적으로 유지되고 있는 계약을 과소평가한 것으로 보인다. 또한 보험료 납입기간이 종료되었을 경우 혹은 연금지급이 개시되었을 경우 계약자 입장에서 계약을 해지하는 것보다 유지하는 것이 유리하기 때문에 위험률이 어느 정도 감소한 것으로 달리 해석할 수도 있다.

넷째, 계약자적립금 대비 최소보증금의 경우 회귀계수가 양의 값이 나왔으므로 위험률과의 관계가 비례한 것으로 나타났다. 위험비는 1.228로 나타났으며 비율이 한 단위 증가할 때마다 위험률이 1.2배 증가함을 의미하며 95% 신뢰구간에서 1을 포함하지 않기 때문에 유의한 변수로 볼 수 있다. 계약자적립금 대비 최소보증금은 변액연금에서 계약자의 행동에 큰 영향을 주는 것으로 알려져 있다. 투자형 상품인 변액연금은 예정이율 외에 펀드수익률에 따라 계약자적립금 수준이 달라지기 때문에 계약자 입장에서는 자신이 받을 수 있는 최소보증금 수준보다 적립되어 있는 금액이 클 경우 즉, 계약자적립금 대비 최소보증금이 1보다 작을 경우 해약요인이 발생한다는 것을 직관적으로 생각해 볼 수 있다. 그러나 분석결과 비율이 한 단위 증가할 때마다 위험률이 감소할 것이라는 예상과 달리 위험률이 증가하는 것으로 나타났다. 이는 계약자적립금 대비 최소보증금이라는 공변수가 비율이라는 점을 감안하면 분자인 최소보증금과 분모인 계약자적립금 수준의 증감에 따라 달리 해석될 수 있기 때문에 해석 상 오류를 범할 수 있다. 따라서 한 단위 증가할 때마다 위험이 증가하는 것으로 해석하되 그 근거에 대해서는 더욱 신중하게 판단 할 필요가 있다.

마지막으로 계약자연령의 경우 위험률과 양의 상관관계를 보이고 위험비는 1.005이 나왔으며 위험비율의 95% 신뢰구간에 1을 포함하지 않기 때문에 이 비율은 유의한 것으로 나타났다. 즉, 계약자연령이 1세 증가할수록 위험이 1.005배 높아지는 것으로 해석할 수 있다. 비록 위험비가 거의 1에 가까워 위험과는 무관하다고 해석할 수도 있으나 변수선택과정에서 제거되지 않고 유의한 변수로 남아있으며 보험의 특성상 연금의 경우 계약자의 연령이 중요하게 작용할 수 있기 때문에 무시할 수 없는 변수이다.

4. 결론 및 시사점

보험산업의 이미지 제고 및 수익성 강화 차원에서 체계적인 해지율 관리는 그 어느 때보다도 중요해지고 있다. 다른 상품에 비해 투자형 상품에 대한 계약자의 행동예측이 어렵다는 점과 실무에서 변액연금의 대한 해지율의 연구가 필요한 추세에 부합하도록 변액연금상품을 대상으로 국내 보험사의 계약자 데이터 (2004.5~2011.11)를 반영하여 현실적인 해약을 모형을 추정해 보았다.

분석 결과 해지율에 영향을 미치는 변수는 납입방법, 보험료, 보험기간 대비 유지기간, 계약자적립금 대비 최소보증금, 계약자 연령순으로 중요도가 높게 나타났다. 보험기간 대비 유지기간을 제외한 모든 변수들이 양의 관계를 가졌으며 이는 변수들이 한 단위 증가할 때마다 위험률이 증가하는 것으로 해석할 수 있다. 계약자적립금 대비 최소보증금의 수준이 계약자의 행동에 크게 영향을 미친다는 분석이 보고

된 바 본 연구에서는 그 결과가 반대로 나왔으며 공변수가 비율이라는 점을 감안하면 해석 상 오류를 범하지 않도록 주의해야겠다. 또는 최근 몇 차례 금융위기를 겪으면서 투자 형 상품에 대한 불안증으로 인하여 투자형 상품에서 이탈되는 계약이 많았다는 것으로도 해석할 수 있다.

본 연구의 한계와 향후 연구과제는 다음과 같다. 첫째로, 분석에 사용된 해지건에 대한 데이터들의 유지기간이 1년을 채 넘기지 못한 데이터가 상당수 있었다. 이는 특정 보험회사에서만 보이는 특성인지 전반적인 흐름인지에 대한 연구가 선행되지 않아 분석결과에 대한 신뢰성이 떨어질 수 있다. 따라서 차후에는 여러 보험회사의 데이터를 이용하여 패턴을 분석한 후 정교한 예측이 이루어지도록 연구할 필요성이 있다. 둘째로, 본 논문에서는 8개의 기초변수를 대상으로 분석을 실시하였으나 투자 형 상품인 만큼 계약자의 자산 및 신용상태 등 그 외에 고려될 수 있는 변수가 다양하나 데이터의 제약으로 인해 포함하지 못하였다. 따라서 해지율의 예측력을 더욱 높이기 위해서는 해지에 영향을 줄 수 있는 다양한 공변수들이 반영되어야 할 것이며 결과해석에 있어서도 더욱 객관적인 판단과 해석이 이루어져야 할 것이다. 세 번째로, 본 연구에서는 선택된 공변량이 비례성 가정을 만족하여 추가적인 분석이 이루어지지 않았으나 변수가 더욱 다양해진다면 공변량이 시간에 따라 달라지는 모형 (time dependent covariates)이 발견될 수 있으며 이를 적용한 연구가 이루어져야 할 것이다. 이 외에 비모수적 방법이 아닌 모수적 방법의 활용으로도 많은 연구가 이루어져야 할 것이다.

Appendix

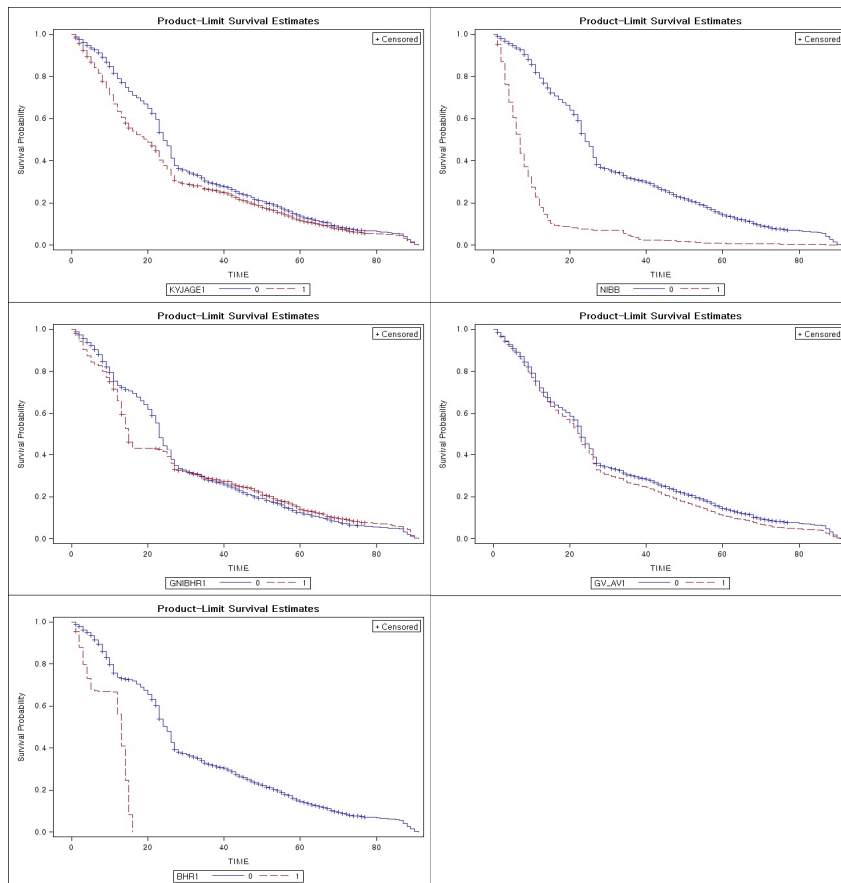


Figure A.1 Survival functions by covariates

References

- Allison, P. D. (1995). *Survival analysis using SAS: A practical guide*, SAS Institute, North Carolina.
- Chung, Y. S. (1992). Suggestions for improving the persistency rates of life insurance contracts. *Korea Insurance Academic Society*, **39**, 241-273.
- Cox, D. R. (1972). Regression models and life tables. *Journal of Royal Statistic Society*, **34**, 187-220.
- Fox, J. (2011). *An R companion to applied regression*, SAGE Publications, California.
- IOAK. (1990). *Improvement method for continuous contract rate using model analysis*, Institute of Actuaries of Korea, Seoul.
- Jung, S. C. and Ouh S. C. (2007). An empirical study on the lapse rate in the life insurance industry. *Korea Insurance Academic Society*, **82**, 155-179.
- Jung, S. C., Ouh, S. C. and Kang, J. C. (2011). An empirical study on the influence of product portfolio and interest rate on the lapse rate in the life insurance industry. *Journal of the Korean Data & Information Science Society*, **22**, 73-80.
- Kang, M. and Lee J. W. (2009). Survival analysis small and medium size construction enterprises using Cox proportional hazards model. *Journal of the Korea Real Estate Analysts Association*, **15**, 44-45.
- Kim, S. N. (2008). *Study on a prediction model for the cancellation of a life insurance contract using logistic and survival analysis*, Master Thesis, Ewha Womans University, Seoul.
- KIRI. (2009). *Analysis on the suspension of effect and the cancellation of life insurance contracts*, Korea Insurance Research Institute, Seoul.
- KIRI. (2010). *Estimation and prediction models for the cancellation rate for each life insurance products*, Korea Insurance Research Institute, Seoul.
- Kuo, W., Tsai, C. and Chen, W. (2003). An empirical study on the lapse rate; The cointegration approach. *Journal of the American Risk and Insurance Association*, **70**, 489-508.
- Kwon, J. Y. (2011). *Study of dynamic lapse rate model*, Master Thesis, SungKyunKwan University, Seoul.
- Lee, H. (2009). Decrement models with an application to variable annuities under fractional age distributions. *Communications for Statistical Applications and Methods*, **16**, 85-102.
- Park, J. K., Oh, K. H. and Kim, M. S. (2012). Survival analysis on the business types of small business using Cox's proportional hazard regression model, *Journal of the Korean Data & Information Science Society*, **23**, 257-269.
- Song, H. H., Jung, G. D. and Lee, W. C. (2007). *The survival analysis*, Chung Moon Gak, Paju.
- Xue, Y. (2010). Interactions between dynamic lapses and interest rates in stochastic modeling. *Product Matters of Society of Actuaries*, **77**, 8-12.

Estimation of lapse rate of variable annuities by using Cox proportional hazard model

Yumi Kim¹ · Hangsuck Lee²

¹Korea Insurance Research Institute

²Department of Actuarial Science/Mathematics, Sungkyunkwan University

Received 22 May 2013, revised 17 June 2013, accepted 22 June 2013

Abstract

The importance of lapse rate is highly increasing due to the introduction of Cash Flow Pricing system, non-refund-of-reserve insurance policy, and IFRS (International Financial Reporting System) to the Korean insurance market. Researches on lapse rate have mainly focused on simple data analysis and regression analysis, etc. However, lapse rate can be analyzed by survival analysis and can be well explained in terms of several covariates with Cox proportional hazard model. Guaranteed minimum benefits embedded in variable annuities require more elegant statistical analysis of lapse rate. Hence, this paper analyzes data of policyholders with variable annuities by using Cox proportional hazard model. The key variables of policy holder that influences the lapse rate are payment method, premium, lapse insured to term insured, reserve-GMXB ratio, and age.

Keywords: Cox proportional hazard model, guaranteed minimum benefit, lapse rate, variable annuities.

¹ Assistant researcher, Korea insurance research institute, Seoul 150-010, Korea.

² Corresponding author: Associate professor, Department of Actuarial Science/Mathematics, Sungkyunkwan University, Seoul 110-745, Korea. E-mail: hangsuck@skku.edu