

비례위험모형에서 비례성 가정에 대한 검증: 도산모형에의 응용

남재우, 김동석, 이회경
KAIST 테크노경영대학원
zeuss@kgsm.kaist.ac.kr

Abstract

The previous quantitative bankruptcy prediction models cannot include time dimension. To overcome this limit, various dynamic models using survival analysis are developed recently. This paper emphasizes that the proportionality assumption must be adapted with caution when the Cox's proportional hazard model is used to explain bankruptcy. It is shown that a non-proportional hazard model including a change point model is a proper alternative, when the proportionality assumption is violated by the change of macroeconomic environment, such as the financial crisis in 1997.

1. 서론

채무제표 변수를 이용한 기존의 정량적 도산 예측모형은 판별분석¹, 확률모형² 등으로 대표되는 이분류(dichotomous classification)모형³으로 분류될 수 있다. 이러한 이분류 모형은 구조적으로 시간 개념을 모형에 효과적으로 포함시킬 수 없다는 문제점이 지적되고 있다. 기업의 도산이란 그 자체로는 특정 시점에 발생하는 하나의 사건이지만, 내부적으로는 오랜 시간에 걸쳐 점진적으로 진행되는 도산과정(failure process)의 결과이므로 도산모형에 있어 시간은 중요한 요소이다. 이런 문제를 극복하기 위하여 기업의 생존기간 자체와 여기에 영향을 미치는 공변량(covariate)으로 구성되는 생존분석 기법을 도산예측 모형에 적용하고자 하는 시도가 계속되고 있다.

대부분의 관련 연구에서는 사용의 편리성에 기 인하여 Cox(1972)의 비례위험모형(proportional hazard model; PH-model)을 적용하고 있다. 그러나 비례위험모형의 기본 가정이라 할 수 있는, 비례적 위험함수의 가정은 대단히 제약적이며 경우에 따라서는 현실적이지 않은 가정일 수도 있다. 그럼에도 불구하고 기존의 많은 연구에서는 이러한 비례성 가정(proportionality assumption)을 적용함으로써 모형의 적합성이 크게 떨어지게 된다.

본 연구에서는 한국의 상장기업을 대상으로 비례위험모형을 추정하고, 추정된 각각의 공변량에 대해 층화모형(stratified model)을 이용하여 비례성 가정을 만족하는지를 검증하였다. 그리고 비례성 가정을 만족하지 못하는 경우에 Liang et. al.(1990)의 변환점을 이용한 비례위험모형(PH-model with change point; PHCP-model)이 적절한 대안이 될 수 있음을 보이고자 한다.

이어지는 2장에서는 본 연구의 분석 모형에 대해 간략히 설명하고, 3장에서는 한국 상장기업의 실제 채무제표 자료를 이용하여 실증분석 하였다. 마지막으로 분석 결과와 향후 연구방향에 대해 4장, 결론에 정리하였다.

2. 분석 모형

생존분석 모형 중 가장 널리 사용되는 것은 식 (1)과 같이 표현되는 Cox(1972)의 비례위험모형(PH-model)이다.

$$h_i(t; x) = h_0(t) \cdot \exp[\gamma'x_i] \quad (1)$$

이 모형은 비례적인 위험함수를 가정하고 있으며, 모형 내에 모수적인 부분($\gamma'x_i$)과 비모수적인 부분($h_0(t)$)을 모두 포함하고 있다. 여기에서 개인의 특성에는 무관하게 시간의 함수로만 정의되는 비모수적 부분을 기저위험함수(baseline hazard function)라 한다. 공변량의 지수함수로 이루어진 모수적 부분의 계수 추정에는 부분우도함수(partial likelihood function)를 이용한 반모수적(semi-parametric) 추정법이 사용된다.

식(1)을 살펴보면 모수적 부분에는 시간 개념이 포함되지 않으며, 기저위험함수에는 개인의 특성이 반영되지 않는다. 이로 인하여 PH-모형은 시간에 따른 개별 위험함수의 변화가 비례적인 경우만을 모형화 하게 되므로, 모형의 적용에 앞서 분석하고자 하는 자료가 비례성 가정을 만족시키는지에 대한 사전 검토가 필수적이라 하겠다.

비례성 가정이 성립하는지 확인하기 위하여 다양한 검증법이 제시되고 있으나 기본 원리에 있어서는 크게 다르지 않다. 다양한 비례성 검증 기법

1. Altman, E. I. (1968)

2. Ohlson, J. S. (1980) and Zmijewski, M. E. (1984)

3. Dimitras et. al. (1996)

들을 두 가지 범주로 대별해 보면 그래프를 이용하는 방법⁴과 검정 통계량에 의한 방법⁵으로 나뉠 수 있다.

본 연구에서는 이들 중 가장 직관적이고 가시적으로 비례성을 확인할 수 있는 층화 모형(stratified model)을 이용하여 어떠한 변수에서 어떤 양상으로 비례성 가정이 위배되는지를 그래프를 이용하는 방법으로 확인하고, 이를 근거로 이러한 불비례성(non-proportionality)을 모형에 포함할 수 있는 변화점을 갖는 PH-모형(PH-model with change point; PHCP-model)을 하나의 대안으로 제시한다.

먼저 층화모형에 대해 간략히 살펴보면, w -개의 층화군(strata)을 갖는 층화모형은 다음과 같은 위험함수로 표현된다.

$$h_i(t; x, s) = h_0^s(t) \cdot \exp[\beta'x_i]; s = 1, \Lambda, w \quad (2)$$

식(2)를 보면, 층화모형은 특정 변수를 제외한 나머지 변수에서는 비례성 가정이 만족된다고 가정하여 w -개의 층화군에서 모두 동일한 계수(β)를 갖는 반면, 개별 층화군마다 상이한 기저위험함수를 가정함으로써 전체적으로 비례적이지 않은 관계를 모형화 하고 있다. 이러한 층화모형에서 계수의 추정 방법은, 추정 과정에서 기저위험함수가 소거되는 PH-모형의 특성상, 일단계로 개별 층화군 내에서 독립적으로 준우도 함수를 구성한 다음에 최종적으로 모든 층화군의 준우도 함수를 곱하는 방식으로 추정된다.

비례성이 깨어지게 되는 여러 양상 중에, 전체 구간에서 어떤 구조적 변화(structural change)에 의해 비례성 가정을 만족하지 않는 것으로 나타나지만, 전체 구간을 몇 개의 소구간으로 나누었을 때 이 소구간 내에서는 비례성 가정이 성립하는 경우가 있다. 이러한 자료의 경우 전체 구간을 몇 개의 적당한 소구간으로 나누어 분석하는 변화점 모형(PHCP-model)을 적용할 수 있다. Liang et. al.(1990)은 하나의 변화점(τ)에 의해 전체 구간이 두 개의 소구간으로 나뉘며, 구간에 따라 위험 요소(공변량)의 영향력이 다르게 나타나는 경우를 모형화하고 이에 대한 추정량을 제시하였다.

$$h_i(t; z, x) = h_0(t) \cdot \exp[(\beta + \theta \cdot I_{\{t \leq \tau\}}) \cdot z_i + \gamma'x_i] \quad (3)$$

식(3)에서 z_i 는 연구기간의 전반부와 후반부에서 그 영향력의 크기가 변하는 공변량이며, x_i 는 전체 구간에 걸쳐 동일한 영향을 미치는 공변량을 의미한다. 이때 구간을 나누는 시점 τ 를 변화점(change point)이라고 하며, 이를 기준으로 z_i 변수에 대한 상대위험도(hazard ratio)⁶가 변하게 된다. 다시 말하면, z_i 변수는 변화점 이전에는 $e^{z_i(\beta+\theta)}$ 의 상대위험

도를 갖는 것으로, 그리고 변화점 이후에는 $e^{z_i\beta}$ 의 상대위험도로 모형화 되는 것이다. Liang et. al.(1990)은 언급하지 않았지만, 이 PHCP-모형은 시간함수와 교호효과 모형(time-by-covariate model)⁷의 특수한 경우로 해석될 수 있다.

3. 실증분석

3.1. 분석 자료

생존분석 기법의 특성상, 분석에 앞서 연구기간(study window)을 명시적으로 정의하여야 한다. 본 연구에서는 1991년부터 2000년까지, 10년의 기간을 연구기간으로 설정하였으며, 분석 기업은 1991년 1월 1일 시점에 한국증권거래소(KSE)에 등록된 상장 기업 중 367개 기업을 선정하였다.

연구기간 중 도산한 기업은 128개 기업 (35%)이며, 연구 종료 시점인 2000년까지 살아남은 나머지 239개(65%) 기업의 수명(life time)은 우측절단(right censored)자료로 처리하였다. 공변량은 도산기업의 경우에는 1991년부터 도산 시점까지, 건전기업의 경우에는 2000년 연구종료 시점까지의 33개 재무비율로 구성되어 있으므로 일종의 불균형 패널(unbalanced panel) 형태이다.

시변공변량 모형(time varying covariate model)의 경우에는 불균형 패널 내의 모든 자료를 그대로 이용할 수 있지만, 정적 모형(static model)에서는 특정 시점의 자료만 이용하든지, 또는 통계적 대표값을 취해야만 한다. 기존의 도산 모형에서는 대부분 '도산 1년 전 자료'와 같은 방식으로 특정 시점을 정의하였으나, 도산 시점을 미리 알 수 없는 예측 모형의 관점에서는 이러한 방법은 사후적 분석의 한계를 갖는다. 본 연구에서는 이러한 논리적 모순을 피하기 위하여 집계(agggregation)에 의한 편익과 비효율성의 문제를 감안하더라도 특정 시점을 정하지 않고 연구 기간 내의 전체 자료에 대한 평균값을 이용하였다.

3.2. 비례위험 모형

모형에 포함될 설명변수(공변량)의 일차 선정을 위하여, 기존의 이분류 모형에서는 전체 기업을 건전과 도산의 두 집단으로 나누어 단순 t-검정을 통해 평균값의 차이를 나타내는 재무비율을 선택하는 방식이 이용되었다. 하지만 처음부터 동일 모집단(identical population)을 가정하는 생존분석에 있어서는, 해당 재무변수의 중앙값(median)⁸을 기준으로 전체 기업을 두 집단으로 나눈 다음 두 집단 간의

4. Hess, K. R. (1995)

5. Ng'andu, N. H. (1997)

6. 상대위험도에 대한 설명은 3.2절 참조.

7. 시간함수와 교호효과(time-by-covariate interaction)를 이용한 비례성 검정방법에 대한 자세한 설명은 Ng'andu, N. H.(1997)로 대신한다.

8. 변수를 두 그룹으로 나누기 위한 기준치로 평균과 중앙값을 모두 고려하였으나, 추정 결과 중앙값을 사용하는 경우 보다 높은 적합도를 보였다. 이는 극단치가 다수 존재하는 상황에서 중앙값이 평균에 비해 극단치의 영향을 적게 받기 때문으로 해석된다.

생존함수(survival function)가 동일한가를 검정하는 방법으로 설명변수를 선정하였다. 두 집단 간의 생존함수의 차이를 통계적으로 검정하기 위하여 로그 순위 검정법(log-rank test)을 이용하였다.

전체 6개 범주, 33개 재무비율에 대해 각각 중앙값 이상과 이하의 두 그룹으로 나누어 두 그룹이 연구기간 내의 모든 시점에 대해 동일한 생존함수를(identical survival function)를 갖고 있는가를 로그 순위 검정한 결과, 전체 33개 변수 중 총 24개 변수가 1% 유의수준 하에서 동일한 생존함수라는 귀무가설을 기각하는 것으로 나타났다.

실제 자료를 이용하여 PH-모형을 추정할 때 직면하는 기술적인 문제로, 동위 자료(tied data)의 문제를 들 수 있다. 부분우도함수(partial likelihood function)를 이용하는 Cox의 비례위험모형에서는 수명 자료 자체가 아니라 그 순서(rank)만을 이용하여 계수가 추정되기 때문에 동위 자료를 어떻게 처리하느냐에 따라 계수의 추정 결과는 크게 달라지게 된다. 본 연구에서는 연 단위로 발표되는 재무제표의 속성을 고려하여, 이산 시간(discrete time)을 가정하여 로그-로지스틱 모형으로 계수를 추정하는 이산 기법(discrete method)을 이용하였다. 비례위험모형에 대한 추정 결과는 다음과 같다.

<표 1> 비례위험모형 추정 결과

변수	변수 설명	계수	상대위험도
X_1	총자본증가율	0.0381 (<0.0001)	1.039
X_2	자기자본구상비율	-0.0406 (<0.0001)	0.960
X_3	현금흐름대총자본비율	-0.1895 (<0.0001)	0.827
X_4	종업원1인당부가가치	-0.0191 (<0.0001)	0.981

주) 괄호 안은 $H_0: \beta_i = 0$ 에 대한 우도비 검정의 p-value.

<표 1>에서 상대위험도(hazard ratio)는 공변량이 한 단위 증가할 때 개별기업의 위험이 얼마나 증가(또는 감소)하는가를 의미한다. 예를 들면 총자본증가율(X_1)이 한 단위 증가하면 그 시점에서 기업의 순간도산확률은 3.9% 증가하는 반면, 현금흐름비율(X_3)이 한 단위 증가하게 되면 기업의 순간 도산확률은 17.3% 감소하게 됨을 의미한다.

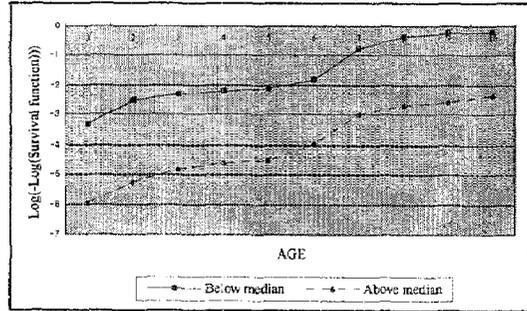
3.3. 비례성 가정에 대한 검정

본 연구에서는 층화 모형(stratified model)을 이용하여 어떤 변수에서 비례성 가정이 위배되는지를 확인하고, 이를 근거로 이러한 형태의 불비례성을 모형에 포함할 수 있는 PHCP-모형을 추정하였다. 이진(binary) 변수가 아닌 연속 변수로 층화(stratification)하는 경우 해당 변수를 적절한 기준에 의해 두 집단으로 나누어야 한다. 본 연구에서는 앞 절에서와 동일한 근거에서 중앙값(median)을 기준으로 분류하였다.

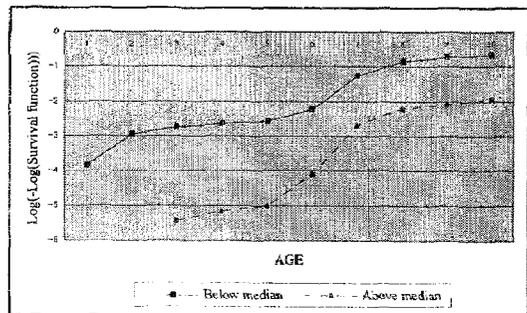
비례성 가정이 성립하는 변수에 대한 층화모형에서는 두 집단의 LLS곡선(log log survival curve)이 시간의 경과에 따라 평행하게 나타나게 된다. 만일

이 그래프가 평행하지 않고 수렴, 또는 발산의 형태를 보이게 되면 비례성 가정이 성립하지 않는 것으로 판별할 수 있다.

<그림 1> LLS곡선: 현금흐름(X_3)에 의한 층화



<그림 2> LLS곡선: 부가가치(X_4)에 의한 층화



층화모형 추정 결과, 총자본증가율(X_1), 총자본증가율(X_1), 현금흐름비율(X_3)의 경우에는 비례성 가정이 성립하는 것으로 판정되었으나, 종업원1인당 부가가치(X_4) 변수의 경우에는 비례성 가정이 성립하지 않은 것으로 나타났다. <그림 1>은 현금흐름비율(X_3) 변수로 층화한 모형에서, 두 집단의 LLS 곡선을 나타내고 있다. 중간 부분에서 IMF의 영향으로 누적위험율이 급격하게 증가하기는 하지만 두 집단 간의 비례관계는 계속 유지되는 것으로 나타났다. 이에 비하여 부가가치(X_4) 변수로 층화한 경우에는 <그림 2>와 같이 비례하지 않는 형태의 LLS 곡선을 나타내고 있음을 확인할 수 있다.

3.4. 변화점을 갖는 위험모형(PHCP-모형)

앞에서 <그림 2>를 살펴보면 1991년부터 5년 동안은 부가가치(X_4) 변수에 대한 두 집단의 LLS 곡선이 일정한 간격을 유지하였으나 1996년을 기점으로 그 비례 양상이 크게 차이를 알 수 있다. 이를 당시 한국의 거시 상황과 비교해 본다면, 이 시기에 한국 경제는 급격한 외환 고갈로 인해 IMF 관리체제에 편입되면서, 1996년 중반부터는 이미 기업의 부도율이 급격히 증가하기 시작하였다. IMF 체제 이후에는, 그 동안 건전기업으로 분류되던 많은 기업들이 부실기업으로 전환되었으며, 살아남은 기업들 역시 이전의 관행과 경영방식에 대한 제질 개선을 적극 추진하게 된다. 이러한 과정에서 개별 공변량이 갖는 영향력의 크기도 변화하게 된 것으로 해석할 수 있다.

이렇듯 소구간에서는 비례적이지만 전체 구간에서는 구조적 변화에 의해 비례성 가정이 깨어지는 경우 앞에서 소개한 PHCP-모형을 적용할 수 있다. 추정에 앞서 사전적으로 정의되어야 하는 변화점은 앞에서의 증화모형의 결과를 사전정보로 하여 $\tau=5$ 로 분석하였으며 그 결과는 다음과 같다.

<표 2> PHCP-모형($\tau=5$) 추정 결과

변수	계수	추정치	상대위험도	$g(t)=\ln(t)$
X_1	γ_1	0.0378 (< 0.0001)	1.039	N.A.
X_2	γ_2	-0.0394 (< 0.0001)	0.961	N.A.
X_3	γ_3	-0.1870 (< 0.0001)	0.829	N.A.
$Z(t \leq \tau)$	$\beta + \theta$	-0.0340 (< 0.0001)	0.967	0.0090 (0.5393)
$Z(t > \tau)$	β	-0.0151 (< 0.0001)	0.985	-0.0272 (0.1942)

주) 괄호 안은 $H_0: \beta_i = 0$ 에 대한 우도비 검정의 p-value.

증화모형에 대한 추정 결과, 연구기간의 전반부에서 종업원1인당부가가치(X_4) 변수의 상대위험도는 0.967로, 후반부에는 0.985로 추정되었다. 앞 절의 PH-모형에서는 부가가치(X_4) 변수에 대한 상대위험도가 $e^{-0.0191} = 0.981$ 로 추정되었으며, 이는 시점에 상관없이 이 변수를 한 단위 증가시키면 기업의 순간위험율은 1.9% 감소하게 됨을 의미한다. 하지만 두 개의 구간으로 나눈 증화모형에서는 β 와 θ 값이 각각 -0.0151, 0.0189로 추정되어 부가가치(X_4) 변수의 상대위험도는 구간에 따라 변하게 된다. 자세히 살펴보면, IMF 체제 편입 이전 기간에는 부가가치(X_4)에 대한 상대위험도가 $e^{-0.0151+0.0189} = 0.967$ 이므로, 이는 부가가치(X_4) 변수를 한 단위 증가시키면 기업의 순간도산율은 3.7% 감소하게 됨을 의미한다. 하지만 연구기간의 후반부에는 상대위험도가 $e^{-0.0151} = 0.985$ 로 변하게 되어 이 구간에서는 부가가치(X_4) 변수 한 단위 증가에 대해 순간도산율이 1.3% 만 감소하게 된다. 이는 특정 경제적 사건을 중심으로, 도산 모형에서 종업원1인당부가가치(X_4) 변수의 영향력이 바뀌고 있음을 의미한다.

이렇게 전체 구간을 두 개의 소구간으로 나누어 분석한 PHCT-모형에서, 개별 소구간에서는 비례성이 성립하는지를 확인하기 위하여, 개별 소구간 내에 시간함수와 교호항을 추가하여 계수의 유의성을 살펴 보았다. 이 교호항 계수가 유의한 것으로 추정되면, 구간 내에 시간에 따라 변하는 불비례성이 존재함을 뜻한다. 시간함수를 로그함수로 가정한 경우에 대한 추정 결과를 <표 2>의 가장 마지막 열에 제시하였다. 여기에서 보면 개별 구간 내에 적용한 교호효과 계수에 대한 p-value가 각각 0.5393 그리고 0.1942로 추정되어 구간 내에서는 비례성 가정이 성립하고 있음을 확인할 수 있다.

4. 결론

본 연구에서는 1991년부터 2000년까지 한국의 상장기업을 대상으로, 생존모형을 이용한 기업도산모형을 구축하였다. Cox의 PH-모형을 적용하여 4개의 재무제표 변수를 공변량으로 하는 도산모형을 구성하였다.

PH-모형 추정 후에, 모형에 포함된 공변량에 대해 비례성 가정이 성립하는지를 증화모형으로 검정하였다. 검정 결과 종업원 1인당 부가가치(X_4) 변수에서 비례성 가정이 위배되는 것을 확인할 수 있었다. 어떤 구조적 변화에 의해 비례성 가정이 어긋나게 되는 경우 적합한 변화점을 갖는 비례위험모형(PHCP-model)을 적용하여 90년대 중반 IMF 관리체제로의 편입을 변환점으로 하는 변환점 모형을 구축하였다. 모형 추정 결과, 변환점 이전과 이후의 소구간에서는 비례성 가정이 만족됨을 확인하였다.

주로 생물학을 중심으로 한 의료 또는 자연과학 분야에서 발전된 생존분석 기법은 동일한 모질단을 가정한다는 측면에서, 사후적 분석이라는 한계를 갖고 있는 기존 도산 모형에 대한 하나의 대안이 될 수 있다. 아직은 도입 단계의 방법론이라, 도산 자료의 특성을 더 정확히 반영할 수 있도록 모형이 수정, 보완되어야 하리라 본다.

< 참고문헌 >

- Altman, E. I., 1968, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *Journal of Finance*, 23, 589-609.
- Cox, D. R., 1972, "Regression models and life-tables," *Journal of the Royal Statistical Society, Series B*, 34, 187-220.
- Dimitras, A. I., S. H. Zanakis, and C. Zopounidis, 1996, "A survey of business failure with an emphasis on prediction methods and industrial applications," *European Journal of Operational Research*, 90, 487-513.
- Hess, R., 1995, "Graphical methods for assessing violations of the proportional hazards assumption in Cox regression," *Statistics in Medicine*, 14, 1707-1723.
- Liang, K., S. Self, and X. Liu, 1990, "The Cox proportional hazards model with change point: An epidemiological application," *Biometrics*, 46, 783-793.
- Ng'andu, N. H., 1997, "An empirical comparison of statistical tests for assessing the proportional hazards assumption of Cox's model," *Statistics in Medicine*, 16, 6, 611-626.
- Ohlson, J. S., 1980, "Financial ratios and probabilistic prediction of bankruptcy," *Journal of Accounting Research*, 18, 1, 109-131.
- Zmijewski, M. E., 1984, "Methodological issues related to the estimation of financial distress prediction models," *Journal of Accounting Research*, 22, Supplement, 59-82.