

상관관계가 존재하는 등급별 동질성 검정방법

홍중선¹ · 이나영²

¹²성균관대학교 통계학과

접수 2012년 11월 5일, 수정 2012년 12월 6일, 게재확정 2012년 12월 28일

요약

신용평가방법에서 등급의 계량화 중 신용등급 변화 검정방법은 등급별로 추정된 예측부도율과 실제부도율과의 동질성을 검정하는 방법으로 한 시점에 대한 이항검정과 카이제곱검정 등이 있고, 여러 시점의 정확성을 검증하는 방법으로 정규성검정, 확장된 신호등검정 등이 있다. 본 연구에서는 현실적인 상황을 고려하여 이런 검정방법들이 상관관계가 존재하는 경우에 등급별 동질성 검정방법을 소개하고 이 방법들을 신용평가 이외에 다양한 분야의 자료에 활용할 수 있음을 알아본다.

주요용어: 로지스틱분석 실패지수, 미숙아출생, 신용등급, 적합성검증, 판별분석.

1. 서론

신 BIS협약에서는 은행의 자체적인 리스크 특성 및 관리수준에 맞는 리스크 측정방법을 선택할 수 있도록 하여, 신용리스크의 경우 표준방법 (standardized approach)과 내부등급법 (internal rating-based approach) 중 선택 적용이 가능하다. 내부등급법을 사용하고자 하는 은행은 적용 최소요건을 충족함을 입증해야하는데 그 최소요건은 주로 신용평가시스템 (credit rating system)에 관한 것으로 리스크의 상대적, 절대적 크기를 측정함에 있어 신뢰성 있는 방법을 사용했는지 여부이고 이는 적합성검증 (validation)을 통해 알 수 있다 (Lim, 2006). 적합성검증이란 신용등급이 리스크를 적절히 차별화하고, 리스크 요소 추정치가 관련 리스크 특성을 적절하게 나타내고 있는지 여부를 평가하는 다양한 절차 및 활동을 의미한다.

바젤위원회에서는 신용평가시스템으로부터 추정된 부도율이 통계적으로 유의한 변별력, 안정성, 예측력을 갖는지 살펴볼 것을 권고하고 있다. 대부분 신용평가시스템이나 리스크요소에 대한 적정성은 통계모형을 기반으로 이루어지는데 특히 적합성검증방법 중 등급의 계량화 검정방법들은 대부분 부도사건 간 독립을 기본가정으로 하기 때문에 현실세계에 그대로 적용하는 것은 이치에 맞지 않는다. 현실세계에서는 어느 정도의 부도사건 간 상관성이 존재하기 때문이다 (Lim, 2005). 이에 적합성 검증방법 가운데 등급의 계량화 방법들에 부도사건 간 상관성을 고려할 수 있는 방법들이 제안되었다. 바젤위원회에서는 상관성을 고려한 이항 검정방법을, So와 Lee (2009b)에서는 상관성을 고려한 정규성 검정방법을, 그리고 Kim과 Ha (2010)는 상관성을 고려한 확장된 신호등 검정방법을 제안하였다.

본 연구에서는 이런 방법들이 기존의 부도사건 간 독립을 가정한 방법들과 비교해 어떤 결과를 도출하는지 살펴볼 것이다. 또한 등급의 계량화 검정방법들을 신용평가에서 뿐만 아니라 등급을 나눌 수 있는 다양한 자료에 적용하여 이러한 방법들을 어떻게 적용할 수 있는지 알아보도록 하고 그 결과에 대해

¹ 교신저자: (110-745) 서울특별시 중로구 성균관로 25-2, 성균관대학교 경제대학 통계학과, 교수.

E-mail: cshong@skku.ac.kr

² (110-745) 서울특별시 중로구 성균관로 25-2, 성균관대학교 통계학과, 석사과정.

살펴보도록 하겠다. 본 논문에서의 2절에서는 기존의 등급의 계량화 검정방법과 상관성을 고려한 등급의 계량화 검정방법들에 대해 알아본다. 3절에서는 2절에서 소개한 방법들을 이용해 실제 신용평가 자료를 이용해 상관성을 고려하지 않은 경우와 고려한 경우 결과에 어떠한 차이가 있는지 알아보도록 한다. 4절에서는 미숙아출산 자료와 국가실패 자료를 이용해 등급을 나누고 예측확률과 실제확률을 구해 2절에서 소개한 방법을 신용평가부분만이 아니라 확장, 응용하여 상관관계의 고려 유무에 따른 결과를 비교해보도록 하겠다. 마지막으로 5절에서는 상관계수가 변환에 따라 검정결과들이 어떻게 변하는지에 대해 전체적으로 살펴보고 향후 남겨진 과제들에 대해서 언급하도록 하겠다.

2. 상관성을 고려한 계량화 검증 방법

등급계량화는 신용평가모형을 통해 산출된 결과를 동일 리스크를 가진 차주끼리 하나의 등급으로 묶어서 그룹화하여 동 등급에 부도율 추정치를 할당하는 일련의 과정을 의미한다. 즉 등급의 계량화 검정이란 동일 신용등급의 차주들이 비슷한 신용위험을 지닌 차주들로 잘 구성되어 있는지 여부를 검정하는 것이다. 일반적으로 등급화는 평가모형의 변별력이 확보되지않는 상태에서는 별다른 의미를 가질 수 없으므로 평가모형의 변별력에 대한 적합성검증이 이루어진 이후에 진행되는 것이다 (Lim, 2006). 등급의 계량화 검정방법에서의 귀무가설과 대립가설은 다음과 같다.

H_0 : 각 등급 내에서 실제부도율은 예측부도율과 같거나 과소 추정되었다.

H_1 : 적어도 하나의 실제부도율은 예측부도율보다 크다.

위의 귀무가설을 바탕으로 BCBS (2005)에서 제안한 이항검정, 정규성검정, 신호등검정, 확장된 신호등검정에 대해 알아보고 기존의 방법에서 상관성을 고려한 방법도 같이 살펴본다.

2.1. 이항검정

이항검정 (binomial test)은 특정시점, 특정등급에서의 검정방법으로 한 시점에서 각각의 등급별로 부도율을 예측했을 때, 이 예측부도율이 실제부도율을 예측하기에 적합한지를 검정하는 것이다. 동일한 등급에서의 차주들의 부도율은 동일하고 각 등급 내에서 발생하는 부도사건이 독립이라는 가정 하에, 해당등급의 부도 차주수 D_n 은 이항분포 $b(n, \hat{P})$ 를 따른다. 여기서 n 은 해당등급의 전체차주의수이고 \hat{P} 는 예측부도율이다. 임계 부도 차주수 d^* 는 정규근사법칙에 의하여 구하고, 유의수준 α 에서 해당등급의 부도 차주수가 임계값 d^* 보다 크거나 같으면 귀무가설을 기각한다.

바젤위원회에서는 이항검정에 상관성을 고려한 방법을 제안하였다. 단일인자모형 (Vasicek, 2002)에 의해

$$R_i = \sqrt{(1 - \rho)\epsilon_i} + \sqrt{\rho}X, \quad (2.1)$$

여기서 R_i 는 i 번째 차주의 자산이며, X 는 모든 차주들이 가지고 있는 요인 (체계적요인)이고, ϵ_i 는 i 번째 차주만 가지는 특별한 요인 (비체계적요인)이다. 그리고 X 와 ϵ_i 는 모두 표준정규분포 $N(0, 1)$ 를 따른다. 자산상관관계 ρ 는 Basel II rule에 의해 0.24보다 클 수 없으며, $cov(X, \epsilon_i) = 0, cov(R_i, R_j) = \rho$ ($i \neq j$)를 가정한다. i 번째 차주의 자산 R_i 가 임계값 r_i 보다 작으면 부도로 간주하며, 확률변수 Y_i 는 지시함수 $I(R_i \leq r_i)$ 로 설정한다.

t모든 차주들이 동일한 부도확률을 가지고 같은 등급 내에 있는 차주들은 동일한 임계점($r_i = r$)을 가진다고 하면 Y_i 는 다음과 같은 성질을 지닌다.

$$E(Y_i | X = x) = P(R_i \leq r) = P(\sqrt{(1-\rho)}\epsilon_i + \sqrt{\rho}x \leq r | X = x) = \Phi\left(\frac{r - \sqrt{\rho}x}{\sqrt{1-\rho}}\right).$$

이때 부도차주수가 임계값 k 보다 작을 확률을 부도비율을 고려하여 정리하면 다음과 같다.

$$P(D_n \leq k) = P\left[\Phi\left(\frac{r - \sqrt{\rho}x}{\sqrt{1-\rho}}\right) \leq \frac{k}{n}\right] = \Phi\left(\frac{\sqrt{1-\rho}\Phi^{-1}\left(\frac{k}{n}\right) - r}{\sqrt{\rho}}\right). \quad (2.2)$$

따라서 $P(D_n \leq k) = \alpha$ 을 만족하는 k 값을 구했을 때 이때의 k 값이 실제부도차수보다 크면 유의수준 α 에서 귀무가설을 기각한다.

2.2. 정규성 검정

정규성 검정 (normal test)은 여러 시점에서 특정등급에 대한 실제부도율과 예측부도율을 비교하는 검정방법이다. 예측부도율이 작은 변동성을 가지고 있고 서로 다른 시점 간에 부도사건은 독립이라는 가정 하에서 총관측기간 n 에서 기간 t ($t = 1, 2, \dots, n$)의 예측부도율 \hat{P}_t 는 독립적이고 평균 P_t 을 가지고 분산 τ^2 을 가진다면, 표준화 통계량 $Z = \sum_{t=1}^n (\hat{P}_t - P_t) / \sqrt{n\tau}$ 는 표준정규분포를 따른다. 여기에서 분산 τ^2 은 편의 (bias)를 고려한 추정치를 사용한다 (Ha, 2009). 따라서 유의수준 α 에서 검정통계량 Z 가 $\Phi^{-1}(\alpha)$ 보다 크면 귀무가설을 기각한다.

여러 시점에서 특정등급에 대한 실제부도율과 예측부도율을 비교하는 검정방법을 확장하여, So와 Lee (2009a)는 다변량 정규분포를 이용해 분산공분산 행렬에 시점 간 상관관계를 반영하여 검정하는 방법을 제안하였다. 시점 ($t = 1, 2, \dots, n$)별로 전체차수 N_t 와 부도 차주수 D_t 및 예측부도율 \hat{P}_t 을 이용하여, 실제부도수와 예측부도수의 차이를 나타내는 $X_t = D_t - N_t\hat{P}_t$ 에 대한 다변량 확률변수 $X = (X_1, X_2, \dots, X_n)'$ 는 평균이 영벡터이고 분산공분산행렬 Σ 의 다변량 정규분포를 근사적으로 따른다고 가정한다. 본 연구에서는 분산공분산행렬을 통해 상관성을 반영할 수 있다. 그러나 실제 분석 대상이 되는 기간 가운데 일정한 시차이상, 즉 T-2이상 시차에 대한 공분산은 추정이 불가능하므로 이를 보완하기 위해 이분산 1차자기상관 (heteroscedasticity AR(1))을 가정한 분산공분산행렬을 사용하였다.

다변량 정규분포를 이용하여 계산한 확률

$$P(X < x) = \int_{-\infty}^x g(y; \Sigma) dy$$

이 유의수준 α 보다 크면 귀무가설을 기각한다 (So와 Lee, 2009a). 본 연구에서는 다중 적분값을 구하기 위하여 수치 해석적 접근방법 가운데 라플라스 근사를 사용하지 않고 R 에서의 pmvnorm 함수를 이용하여 누적확률을 산출하였다.

2.3. 신호등 검정

신호등검정 (traffic light approach)은 바젤위원회가 제안한 방법에서와 마찬가지로 Metron (1974)의 이론을 기반으로 하는 단일인자모형을 이용한다. \hat{P} 에 대한 두 개의 신뢰수준 $1 - \alpha_{high}$ 와 $1 - \alpha_{low}$ 에 대한 임계값 C_{low}, C_{high} 을 각각 구한다 (Tasche, 2003). 임계값을 구하기 위해 부도 차주수 D_n 은 다음과 같다.

$$D_n = \sum_{i=1}^n I[\sqrt{1 - \rho}\epsilon_i + \sqrt{\rho}X \leq r]. \quad (2.3)$$

이때 임계값 $c_{low}c_{high}$ 는 다음과 같이 설정한다.

$$\begin{aligned} c_{low} &= \min[k \mid P(D_n > k) \leq \alpha_{low}] = q(1 - \alpha_{low}, D_n) + 1 \\ c_{high} &= \min[k \mid P(D_n > k) \leq \alpha_{high}] = q(1 - \alpha_{high}, D_n) + 1, \end{aligned}$$

여기서 $q(\alpha, D_n)$ 은 D_n 의 $100 \times \alpha$ 분위수이다 (Gordy, 2002). 관측부도율을 구하기 위하여 Taylor expansion을 이용하고 (Tasche, 2005), Quantile 공식과 Granularity 조정접근방법 (adjustment approach; Gagliardini and Gouriou (2010)와 Martin and Wilde (2002)를 사용하여 구한다. 관측부도율이 c_{low} 보다 작다면 Green으로, c_{low} 보다 크거나 같고 c_{high} 보다 작으면 Yellow로, c_{high} 보다 크거나 같으면 Red로 판단함으로써 신호등식 의사결정 규칙을 설정한다. Green이 많이 나올수록 등급의 계량화가 잘 되어있고, Red가 나타나면 예측부도율과 실제부도율과 차이가 크다는 의미이다.

2.4. 확장된 신호등 검정

확장된 신호등 검정 (extended traffic light approach; ETLA)은 다항분포를 이용하여 일정기간동안 발생한 실제부도율을 통해 특정 등급의 적합성검증 방법이다. 신용등급별 관측부도율이 예측부도율과 떨어져 있는 정도를 Green, Yellow, Orange, Red로 구분한다. 시점 간 부도사건을 독립으로 가정하고 다항분포를 기저분포로 한다. 특정 시점(t)에서의 부도 차주수 D_t 가 같은 등급내의 차주 n_t 들이 동일한 부도확률 P_t 를 가진 이항분포 $D_t \sim b(n_t, P_t)$ 를 따르므로 표준화 확률변수 Z_t 는 표준정규분포에 근사하므로 각 시점별 검정통계량 $\phi(z_t)(t = 1, 2, \dots, n)$ 일 때, 확장된 신호등 검정방법은 다음과 같이 수행한다.

$$\phi(z_t) = \begin{cases} \text{Green,} & z_t \leq \Phi^{-1}(\pi_g) \\ \text{Yellow,} & \Phi^{-1}(\pi_g) < z_t \leq \Phi^{-1}(\pi_g + \pi_y) \\ \text{Orange,} & \Phi^{-1}(\pi_g + \pi_y) < z_t \leq \Phi^{-1}(\pi_g + \pi_y + \pi_o) \\ \text{Red,} & \Phi^{-1}(\pi_g + \pi_y + \pi_o) < z_t \end{cases}$$

여기서 $(\pi_g, \pi_y, \pi_o, \pi_r) = (0.5, 0.3, 0.15, 0.05)$ 을 주로 사용한다 (Blochwitz, 2004). n 년 동안 발생한 신호등 색깔의 횟수를 (k_g, k_y, k_o, k_r) 라고 정의할 수 있다 ($k_g + k_y + k_o + k_r = n$). 각 연도별 부도사건을 독립이라고 가정한 경우 다시점에 걸쳐 측정된 신호등 색깔의 횟수인 벡터는 다항분포를 따른다. 임계구간을 찾기 위해 신호등 색깔에 따라 적당한 가중치를 부여한 $V = p_g k_g + p_y k_y + p_o k_o + p_r k_r$ 식으로부터 분포를 생성한다. 여기에서 가중치 $(p_g, p_y, p_o, p_r) = (1000, 100, 10, 1)$ 또는 $(0.5, 0.3, 0.01, 0.05)$ 등을 사용하여 임계구간을 찾는데, 유의수준 α 에서 $V \leq v_\alpha$ 이면 관측기간동안에 실제부도율이 모두 예측부도율보다 크지않다는 귀무가설을 기각한다. v_α 는 $P(V \leq v_\alpha) \leq \alpha$ 조건을 만족하는 가장 큰 수이다 (자세한 방법은 Blochwitz (2004) 참조).

Kim과 Ha (2010)는 여러 기간의 부도율 정확성검증에서 임계값에 상관관계를 반영한 방법으로 ETLA 방법을 확장하였다. 신호등 검정에서의 부도수가 아닌 부도율을 기준으로 사용한 식은 OeNB (2005)을 참고하였다. 기존의 ETLA 방법에서는 표준정규분포를 Z_t 를 이용하여 각각의 부도확률을 설정하였지만 Kim과 Ha (2010)는 상관성을 고려한 신호등검정방법을 확립하여 부도확률을 구하면서 신호등색을 나타내었다. 기존의 ETLA 방법에서와 마찬가지로 n 년 동안 발생한 신호등 색깔의 횟수를 (k_g, k_y, k_o, k_r) 라고 정의할 수 있으며 $V \leq v_\alpha$ 이면 귀무가설을 기각한다.

3. 신용평가 실증자료분석

3.1. 한 시점 검정

앞에서 소개한 등급의 계량화 검정방법은 시점에 따라 한 시점 또는 여러 시점으로 구분한다. 한 시점에 대해서는 어떤 외국 모회사의 신용자료이며 신용등급을 1에서 10등급으로 나누었고 각 등급에서의 전체 차주수와 예측 부도율, 실제 부도율로 정리한 OeNB (2004)의 자료를 이용하여 분석한다. Table 3.1은 이항검정, 상관관계를 고려한 이항검정, 신호등검정에 대한 결과이다.

Table 3.1 Binomial test and traffic light approach

grade	expected PD	actual PD	binomial ($\rho = 0$)	binomial ($\rho = 0.1$)	traffic ($\rho = 0.1$)
1	0.05	0.00	H_0 not reject	H_0 not reject	Green
2	0.10	0.17	H_0 not reject	H_0 not reject	Green
3	0.25	0.32	H_0 not reject	H_0 not reject	Green
4	0.50	0.45	H_0 not reject	H_0 not reject	Green
5	0.75	1.23	H_0 reject	H_0 not reject	Green
6	1.30	1.99	H_0 reject	H_0 not reject	Green
7	2.00	3.51	H_0 reject	H_0 not reject	Green
8	3.40	4.21	H_0 not reject	H_0 not reject	Green
9	6.00	8.30	H_0 not reject	H_0 not reject	Green
10	10.00	12.84	H_0 not reject	H_0 not reject	Green

Table 3.1을 살펴보면 $\rho = 0$ 일 때는 5, 6, 7등급에서 실제부도율이 임계 부도율보다 작으므로 귀무가설은 기각되며, 예측부도율이 실제부도율에 비해 과소 추정되었다고 판단할 수 있다. 반면 $\rho = 0.1$ 로 주어졌을 경우 모든 등급에서 귀무가설은 기각되지 않는다. 모든 등급에서의 실제부도율과 예측부도율이 거의 차이가 없다는 것을 의미한다. 또한 신호등검정 결과에서는 신호등색이 모든 등급에서 Green으로 나타나며, $\rho = 0.1$ 일 때의 이항검정과 결과와 동일하게 나타난다. 상관관계를 고려하지 않은 이항검정에서는 5, 6, 7등급에 대해 귀무가설이 기각되지만 상관성을 고려한 이항검정의 경우 임계부도율이 증가함에 따라 귀무가설이 기각되지 않는다. 상관성을 고려한 경우에 결과가 달라지는 것을 확인할 수 있다.

3.2. 여러 시점 검정

여러 시점에 대해서는 Lee와 Sim (2006)에서 사용한 5년간 실증 자료를 이용한다. 이 자료는 모은행의 2001년부터 2005년까지의 전체 차주수와 예측부도율, 실제 부도율이 나타나 있으며, 시점 간 부도사건과의 상관관계를 고려한 정규성검정과 확장된 신호등검정의 결과를 Table 3.2에 정리하였다.

Table 3.2 Normal test and ETLA

		$\rho = 0$	$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.15$...	$\rho = 0.55$
Normal test	Cumulative probability	0.0313	0.0367	0.0431	0.0504	...	0.0941
ETLA	2001	Yellow	Yellow	Yellow	Yellow	...	Orange
	2002	Yellow	Yellow	Yellow	Yellow	...	Orange
	2003	Yellow	Yellow	Yellow	Yellow	...	Orange
	2004	Yellow	Yellow	Yellow	Yellow	...	Orange
	2005	Yellow	Yellow	Yellow	Yellow	...	Orange
	V	1.7	1.5	1.5	1.5	...	0.8

정규성검정에서 상관계수의 변화에 따른 누적 확률값을 살펴보면 상관계수가 0.15 이상이면, 누적확률값이 유의수준 α 보다 크므로 귀무가설을 기각한다. 이는 어떤 신용평가모형을 통해 예측한 예측부도율과 실제부도율의 차이가 시점 간 자기상관이 0.15 이상으로 나타난다면 독립을 가정한 정규성검정 ($\rho = 0$)과 결과가 다르게 나타난다는 것을 의미한다.

확장된 신호등 검정결과를 살펴보면 시점 간 독립을 가정한 경우 ($\rho = 0$) 2001에서 2005년 동안의 검정한 결과 Green이 1번, Yellow가 4번 나타나서 $V = 0.5 \times 1 + 0.3 \times 4 + 0.15 \times 0 + 0.05 \times 0 = 1.7$ 이다. $P(V \leq v_\alpha)$ 을 만족하는 v_α 을 구하기 위해 2001년에서 2005년동안 발생할 수 있는 신호등색의 모든 경우의 수를 구한 후, 선형식 V 의 값을 오름차순으로 정렬한 뒤 기저분포인 다항분포를 이용해 확률을 구한다. $P(V \leq v_\alpha)$ 을 만족하는 가장 큰 v_α 는 28번째에 있는 (1, 2, 0, 1)이며 이때의 v_α 값은 1.2이다. 따라서 $V = 1.7 > v_\alpha = 1.2$ 이므로 귀무가설을 기각하지 못한다. 이 결과는 시점 간 독립을 가정한 정규성 검정과 같은 결과를 제공함을 알 수 있다. 그러나 상관계수가 0.55보다 작은 경우에는 V 값이 $v_\alpha = 1.2$ 보다 크거나 같아서 귀무가설을 기각하지 못하고 실제 부도율과 예측 부도율이 거의 차이가 없다는 결론을 내린다. 기존의 ETLA 방법은 상관관계를 고려하지 않은 방법으로 만약 실제 자료에 높은 상관관계가 존재한다면 이를 고려하지 못하고 귀무가설을 채택하게 된다는 뜻이다. 그리고 상관관계를 고려한 정규성검정은 상관계수가 0.15 이상이면 귀무가설이 기각되지만 상관성을 고려한 ETLA 방법은 상관계수가 0.55 이상일 때 귀무가설이 기각된다. 즉 상관계수의 값에 따라 결과가 달라지는 것을 확인할 수 있다.

4. 다양한 분야의 실증자료분석

앞에서 소개한 등급의 계량화 방법은 신용평가에서 사용되었다. 본 연구에서는 다른 분야에서 응용 및 확장한다. 다음의 두 종류의 실증자료를 이용하여 확장해 보았다.

4.1. 미숙아출생 자료

1986년 미국 메사츄세츠주의 베이스테이트 병원에서는 산모 488명을 대상으로 신생아 몸무게 (BWT)가 2.5kg 미만인 미숙아출산에 대한 위험요인들에 관한 연구를 실시하였다. 종속변수는 미숙아를 출산 하였으면 Low는 1, 아니면 0의 값을 가진다. 설명변수로는 네개의 변수이며, 임신기간중 흡연여부에 따라 Smoke=1 or 0, 산모의 인종에 따라 Race=1 (white), 2 (black), 3 (other), 산모의 나이 (Age), 임신직전의 몸무게 (Lwt)이다. 인종은 $d_1 = race1, d_2 = race2$ 인 더미변수를 사용하였다.

Table 4.1 BWT data

grade	number of BWT	actual number of BWT	actual probability of BWT	expected number of BWT	expected probability of BWT
1	20	6	0.3000	2	0.1000
2	48	13	0.2708	7	0.1458
3	61	22	0.3607	15	0.2459
4	69	21	0.3043	14	0.2029
5	73	27	0.3699	14	0.1918
6	65	28	0.4308	19	0.2923
7	40	6	0.1500	16	0.4000
8	31	11	0.3548	6	0.1935
9	37	11	0.2973	13	0.3514
10	44	6	0.1364	19	0.4318

미숙아출생 자료에 대한 로지스틱 회귀분석의 결과를 바탕으로 예측확률을 구하였다. 절단점 (cut-off value)을 설정하고, 예측확률이 절단점보다 크면 미숙아로 예측하여 오류율을 구한다. 절단점값이 0.8일 때 가장 작은 총오류율을 나타내는 것을 발견하였고, 임신직전의 몸무게에 따라 1에서 10등급까지 나누었다. 등급에 따라 실제 미숙아출생율과 예측 미숙아출생률에 대한 결과는 Table 4.1과 같다.

4.1.1. 이항검정

Table 4.1의 자료를 가지고 등급 내에서의 예측 미숙아출생률이 잘 예측되었는지 살펴보기 위하여 우선, 상관계수를 고려한 이항검정을 실시한다. 로지스틱 회귀분석의 결과 나이의 계수가 음의 값을 가지

유의수준 0.05에서 신호등검정의 결과 상관계수를 0.05로 설정한 경우에 1등급과 5등급에서 Yellow, 나머지 등급에서는 Green으로 나타났다. 이는 1, 5등급에서는 예측 미숙아출생률과 실제 미숙아출생률의 차이는 염려할 수준이 아니며, Green으로 나타난 등급에서는 예측이 잘되었다는 의미이다. 그리고 상관계수가 증가할수록 모든 등급에서 Green으로 판단된다. 이는 바젤위원회가 제안한 상관성을 고려한 이항검정의 결과와 동일하다.

t신호등검정의 결과 Table 4.3의 $\rho = 0.05$ 인 경우가 바젤위원회가 제안한 상관성을 고려한 이항검정의 결과 즉, Table 4.2의 $\rho = 0.1$ 인 경우와 일치되는데, 신호등검정이 바젤위원회가 제안한 방법보다 관대하기 때문이다.

4.2. 국가실패지수 자료

미국의 외교전문지 폴린폴리시 (FP)와 평화재단이 2005년부터 매년 공동으로 발표하는 국가실패지수 (FSI) 자료를 이용한다. 전 세계 177개국을 대상으로 인권, 사회, 경제, 정치, 치안 요인 등 12개 항목을 조사하고, 각 항목당 10점이며 최고 120점을 만점으로 국가의 불안한 정도를 측정한다. FP는 이를 바탕으로 177개의 나라를 Critical, In danger, Borderline, Stable, Most stable의 다섯 등급으로 나누어지며 이를 여러 시점에 대한 검정방법을 사용할 수 있다.

2007년부터 2012년까지의 FSI 자료에 대하여 관별분석을 실시하였고, 4.1절에서와 같이 70점을 기준으로 분류를 통해 오분류율이 가장 작게 나타난다 (Table 4.4 참조). 따라서 총점수가 70점 이상이면 실제 위험국가 간주하도록 하여 실제 위험국가율을 구한다. 또한 12개의 항목 중 경제불균등 (uneven development) 항목을 이용해 예측 위험국가율을 구하였다. 경제불균등 항목의 점수가 6.5 이하일 때 위험국가로 예측한 경우에 오분류율이 가장 적게 나타났으며 이에 대한 혼동행렬도 Table 4.4에 같이 표현하였다. 그리고 Table 4.5는 2007년부터 2012년까지의 실제 위험국가율과 예측 위험국가율을 계산하여 정리하였다.

Table 4.4 Confusion matrix

	discriminant analysis		uneven development	
	actual critical	actual stable	actual critical	actual stable
predict critical	113	4	110	18
predict stable	0	60	3	46
sum	113	64	113	64

Table 4.5 FSI data

year	actual number of critical	actual probability of critical	expected number of critical	expected probability of critical
2007	113	0.63842	118	0.66667
2008	110	0.62147	117	0.66102
2009	109	0.61582	117	0.66102
2010	108	0.61017	115	0.64972
2011	108	0.61017	110	0.62147
2012	108	0.61017	102	0.57627

4.2.1. 정규성검정

Table 4.5에 대하여 2.2절에서 논의한 시점 간 독립을 가정한 정규성검정을 실시하면 표준화 검정통계량값은 -1.7993으로 음수값이다. 따라서 임계값이 양수(1.96)이므로 당연히 귀무가설을 기각할 수 없다. 따라서 관측기간 6년동안 실제 위험국가율이 예측 위험국가율보다 크다고 할 수 없다.

그리고 연도별 상관관계를 고려하여 정규성검정을 실시한 결과에서 검정통계량인 누적확률값이 상관 계수에 관계없이 모두 0에 가까운 값을 가지고 있어 귀무가설이 기각되지 않는다. 즉 관측기간 6년 동안 실제 위험국가율이 예측 위험국가율 보다 크다고 할 수 없다.

4.2.2. 확장된 신호등 검정

FSI 자료인 Table 4.5에 대한 ETLA 검정결과 Table 4.6을 통해서 시점 간 독립을 가정했을 때 ($\rho = 0$), 6년 동안 Green이 5번, Orange가 1번 나타나 $V = 0.5 \times 5 + 0.3 \times 0 + 0.15 \times 1 + 0.05 \times 0 = 2.65$ 이다. $P(V \leq v_\alpha) \leq \alpha$ 을 만족하는 가장 큰 v_α 는 47번째에 있는 (0,5,0,1)이며 이때의 v_α 값은 1.55이다. 따라서 $V = 2.65 > v_\alpha = 1.55$ 이므로 귀무가설을 기각하지 못한다. 이 결과는 시점 간 독립을 가정한 정규성검정과 같은 결과를 제공함을 알 수 있다.

Table 4.6 ETLA of FSI data

year	$\rho = 0$	$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.15$	$\rho = 0.2$...	$\rho = 0.55$
2007	Green	Green	Green	Green	Green	...	Green
2008	Green	Green	Green	Green	Green	...	Green
2009	Green	Green	Green	Green	Green	...	Green
2010	Green	Green	Green	Green	Green	...	Green
2011	Green	Green	Green	Green	Green	...	Green
2012	Orange	Yellow	Yellow	Yellow	Yellow	...	Green
V	2.8	2.85	2.85	2.85	2.85	...	3.00

상관계수를 고려한 경우도 모든 검정통계량이 $v_\alpha = 1.55$ 보다 크므로 귀무가설을 기각할 수 없으며, 정규성검정과 일치된 결과이다. 상관계수의 변화에 따른 결과를 자세히 살펴보면 상관계수의 값이 0.05이상이면 Green이 5번 Yellow가 1번 나타나며 2012년도에 신호등색이 상관계수가 0일 때와 다르게 나타나는 것을 알 수 있다. 그리고 상관계수가 0.55이상이 되면 2012년도의 신호등색은 Green으로 나타난다. 즉 상관계수가 커질수록 검정통계량값인 V도 커지면서 귀무가설을 기각할 확률이 줄어든다고 판단할 수 있다.

5. 결론

오늘날 급변하는 금융시장 속에서 정확한 모형을 구축하고 사후검증을 하는 것이 복잡하고 다양해짐에 따라 중요성이 높아지고 있다. 이러한 모형의 적합성검증에서 등급별 예측부도율이 잘 유지되어 있는지를 판단하는 등급의 계량화 검정방법을 연구하였다. 기존의 방법들은 차주간, 시점간 독립을 기본 가정으로 하는데, 현실세계에 연쇄적인 부도사건과 같은 현실적이지 않다는 단점을 가진다. 본 연구에서는 상관성을 고려해주는 방법들을 소개하였고 상관성을 고려해주지 않은 경우와 비교하였다.

등급 내에서 실제부도율이 예측부도율보다 큰 경우가 많으면 상관계수값이 증가할수록 귀무가설을 기각할 확률이 증가하게 되어 모든 등급에서 실제부도율이 예측부도율보다 크다고 판단된다. 반면에 등급 내 예측부도율이 실제부도율보다 큰 경우가 많다면 상관계수의 값이 증가할수록 귀무가설을 채택할 확률이 높아져 모든 등급에서의 예측부도율이 실제부도율보다 크다고 판단한다. 그러므로 너무 큰 상관관계를 고려하면 등급 내에서의 경우의 수가 큰 쪽을 지지할 확률이 높아지게 되어 관대한 결과를 가져오므로 상황에 맞는 적절한 값의 상관관계를 고려해주는 것이 중요하다.

또한 본 논문에서는 이러한 등급의 계량화 검정방법들을 신용평가에서 뿐만 아니라 다양한 자료에 확장하여 응용하였다. 미숙아출생 자료와 FSI 자료에 대하여 등급에 따른 예측율을 구해 등급 안에서의 예측율이 잘 유지되어 있는지를 살펴보았다. 상관관계를 고려한 경우 상관계수의 값이 커질수록 귀무가

설을 기각할 확률이 감소하여 모든 등급에서 예측율이 실제율보다 크다는 결론을 유도하였다. 이에 따라 상관계수의 값을 너무 크게 정하게 되면 관대한 결과를 제공하게 되므로 자료의 상황에 맞는 적절한 상관계수를 설정하는 것이 중요하다는 것을 파악하였다. 본 연구에서는 확장자료의 예측율을 구할 때 한 가지 항목만을 가지고 이용하였지만, 여러 가지 항목들을 종합하여 더 정교한 모형을 구축해서 예측율을 구한 뒤 적용시키는 연구를 향후 과제로 남겨둔다.

참고문헌

- Basel Committee on Banking Supervision. (BCBS) (2005). *Studies on the validation of internal rating systems (revised)*, Working Paper, **14**, Bank for International Settlements, Switzerland.
- Blochwitz, S., Hohl, S., Tasche, D. and Carsten S. W. (2004). *Validating default probabilities on short time series*, Working Paper, Deutsche Bundesbank, Germany.
- Gagliardini, P. and Gouriéroux, C. (2010). *Granularity adjustment for risk measures: Systematic vs unsystematic risks*, Working Paper, CREST, Loughborough University, United Kingdom.
- Gordy, M. B. (2002). A risk factor model foundation for rating based bank capital rules. *Journal of Financial Intermediation*, **22**, 199-232.
- Ha, J. (2009). A study on validating the calibration of the internal rating system. *Journal of the Korean Data Analysis Society*, **11**, 827-837.
- Kim, E. N. and Ha, J. (2010). Study on the consistency methods of calibration considering correlations. *Journal of the Korean Data & Information Science Society*, **21**, 407-417.
- Lee, S. H. and Sim J. H. (2006). Study for quantitative goodness of fit tests based on new BIS criteria. *Risk Review*, Summer, **1**.
- Lim, C. S. (2005). Goodness of fit tests for credit evaluation system based on new BIS criteria. *Risk Review*, Summer, **3**.
- Lim, J. K. (2006). Introduction of goodness of fit tests for credit evaluation system. *Risk Review*, Autumn, **1**.
- Martin, R. and Wilde, T. (2002). Unsystematic credit risk. *RISK Magazine*, **15**, 123-128.
- Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, **29**, 449-470.
- OeNB. (2005). *Rating models and validation, guidelines on credit risk management*, Otto Wagner Platz, Austria.
- So, S. H. and Lee, D. H. (2009a). Risk model validation for multi-period probability of default estimates. *Journal of the Korean Data Analysis Society*, **11**, 839-849.
- So, S. H. and Lee, D. H. (2009b). Credit risk model validation with correlated binomial distribution. *Journal of the Korean Data Analysis Society*, **11**, 327-338.
- Tasche, D. (2003). *A traffic lights approach to PD validation*, Working Paper, Deutsche Bundesbank, Germany.
- Tasche, D. (2005). Calculating credit risk capital charges with the one-factor model. *Journal of RISK*, **7**, 85-101.
- Vasicek, O. A. (2002). The distribution of loan portfolio value. *Journal of RISK*, **15**, 160-162.

Class homogeneous tests with correlation

Chong Sun Hong¹ · Na Young Lee²

¹²Department of Statistics, Sungkyunkwan University

Received 5 November 2012, revised 6 December 2012, accepted 28 December 2012

Abstract

Among class quantitative tests for the credit rating systems, the credit rating tests for calibration are to test the class homogeneous differences between observed and predicted probabilities. For one time period, binomial test and chi-square test are included, and normal test and extended traffic lights test are also contained for several time periods. In this work, we consider real data in which there exists correlation among variables, so that these test methods could be applied to the credit rating systems as well as various kinds of the class data such as BWT data and FSI data.

Keywords: BWT, credit evaluation, discriminant analysis, FSI, logistic analysis, validation.

¹ Corresponding author: Professor, Department of Statistics, Sungkyunkwan University, Seoul 110-745, Korea. Email: cshong@skku.edu

² Graduate student, Department of Statistics, Sungkyunkwan University, Seoul 110-745, Korea.