

상관관계를 반영한 신용등급 계량화 검정기법 연구

김은나¹ · 하정철²

¹²계명대학교 통계학과

접수 2010년 3월 15일, 수정 2010년 4월 20일, 게재확정 2010년 5월 6일

요약

신BIS협약에서는 자본적정성 규제에 있어서 자율성 확대를 위해 내부등급시스템을 허용하였고, 이에 따라 시스템의 적합성 검증의 중요성을 더욱 강조하게 되었다. 적합성검증은 양적·질적 검증으로 나뉘며 이 때 양적 검증은 변별력, 안정성, 등급의 계량화로 구별된다. 본 논문에서는 양적 검증의 등급의 계량화 중 신용등급 변화 검정기법에 대하여 연구하려고 한다. 등급의 계량화 검정은 등급별로 추정된 부도율과 실제 부도율과의 차이를 검정하는 방법으로 한 시점에 대한 검정으로 이항검정과 카이제곱검정, Brier score, 신호등 검정이 있고, 여러 시점의 정확성을 검증하는 방법으로 정규성 검정, 확장된 신호등 검정이 있다. 신용평가시스템의 정확성을 높이고, 현실 상황에 반영이 가능하려면 상관관계를 고려하지 않을 수 없다. 본 논문에서는 부도상관관계를 고려한 확장된 신호등 검정을 제안하고자 한다.

주요용어: 계량화, 상관관계, 신용등급, 적합성 검증, 확장된 신호등 검정.

1. 서론

신BIS협약은 국제결제은행 (BIS) 바젤은행감독위원회에서 기존 BIS협약의 유효성이 저하됨에 따라 이를 대체하기 위해 제정을 추진하여 2004년 6월에 최종 확정된 협약이다. 바젤위원회 회원국은 2006년 말부터 시행하고 비회원국 중에서도 우리나라를 포함하여 많은 나라들이 회원국과 같은 시기 또는 각국의 실정을 고려하여 시행시기를 결정하였다 (금융감독원, 2006). 신BIS협약의 핵심은 리스크에 민감한 자기자본규제 요건을 마련하고 금융회사의 능력에 따라 리스크 관리 방법을 선택하여 적용하는데 있다 (정원길, 2003).

신BIS협약에서는 금융기관의 적정 자기자본 수준에 대한 건전성 점검에 있어서 자체적으로 내부등급 시스템을 허용하게 하고, 내부등급체계에 대한 정기적인 점검을 통하여 적용중인 신용등급체계에 대한 지속적인 모니터링과 일관된 내부등급체계에 대한 확인을 받도록 하였다. 또한 2008년 신BIS협약은 국내 은행에도 도입되었는데 은행의 리스크 특성 및 관리수준에 맞는 리스크 측정방법을 스스로 선택할 수 있도록 함으로써 은행의 자율성 확대를 보장해 주고 있다. 규제목적의 자기자본비율 산출시 은행 자체적으로 도출된 리스크 추정치를 인정하기 위해서는 그 전제조건으로서 신용평가시스템 및 리스크 추정치의 적격성 확보가 우선되어야 하며, 동 적격성을 보장하는 절차가 바로 신용평가시스템에 대한 적합성 검증 (validation)이라 할 수 있다 (Hong과 Choi, 2008; Hong과 Suh, 2008). 적합성 검증은 신용등급이 리스크를 적절히 차별화하고, 리스크요소 추정치가 관련 리스크 특성을 적절하게 나타내고 있는지 여부를 평가하는 다양한 절차 및 활동을 의미한다 (임종건, 2006).

¹ (704-701) 대구광역시 달서구 달구벌대로 2800, 계명대학교 통계학과, 석사과정.

² 교신저자: (704-701) 대구광역시 달서구 달구벌대로 2800, 계명대학교 통계학과, 조교수.

E-mail: jeicy@kmu.ac.kr

적합성 검증은 모든 신용평가시스템에 기술적, 수리적 방법에 의해 수행되어야 하며, 이 검증은 양적 검증과 질적 검증으로 나뉘게 된다. 양적 검증에 대한 주요 기준은 변별력, 안정성, 등급의 계량화가 있다.

신용평가시스템의 변별력 (discriminatory power)은 부도와 정상 차주들을 사전적으로 구별하는 능력을 의미하며 상위 등급의 실제 부도율과 하위 등급의 실제 부도율 간에 차이가 클수록 높은 변별력을 지닌 것으로 간주된다. 변별력을 평가함에 있어 현재 폭넓게 사용되는 ROC, CAP, 또는 지니 (Gini) 계수가 있지만 이 기준에만 단순히 의존해서는 안된다고 알려져 있다 (홍중선과 김지훈, 2009). 신용평가시스템의 안정성 (stability)은 리스크 요인들과 신용도 간의 인과관계를 적절하게 모델화하는 것을 의미하며, 불안정한 시스템은 시간 경과에 따라 예측력이 급격히 감소하는 경향이 있다. 등급의 계량화 (calibration)는 내부등급과 같이 상대적 부도율에 의해 등급이 부여된 차주들에게 PD (예측 부도율: Probability of Defaults)를 할당하는 과정을 의미한다 (이석형과 심재호, 2006; 하정철, 2009). 이 때, 등급의 계량화는 통상 PD와 신용등급의 매핑을 나타내고 있으나, 보다 넓은 의미로 사용된 경우에는 LGD와 EAD 등 추가적인 리스크 측정 요소들의 매핑도 포함한다. 내부등급법 하에서 은행의 소유자기 자본은 각 익스포저에 대한 내부추정치에 의해 결정되고 이는 신용평가시스템으로부터 도출되는 바, 등급의 계량화에 대한 정확성은 신용평가시스템을 평가하는 중요한 기준이 되고 있다 (금융감독원, 2005).

본 논문에서는 신용평가시스템의 양적 검증 중 등급의 계량화에 대한 검증 방법에 대해 알아보고 현실에 더 적합한 검증 방법을 찾아보았다. 2장에서 계량화 검증 방법 중 일반적으로 알려진 여러 방법들에 대해 살펴보고 3장에서 상관관계를 반영한 ETLA 방법에 대해 설명한다. 4장에서는 상관관계를 고려하지 않은 ETLA 방법과 상관관계를 고려한 ETLA 방법의 결과를 비교하고 5장은 결론으로 구성하였다.

2. 계량화 검증 방법

계량화 검증이란 등급 내 예측 부도율과 실제 관측 부도율이 잘 구분되었는지를 판단하는 것으로 추정된 부도율이 적합한 값인지를 판단하는 것이다. 이 때, 평가모형으로부터 PD를 추정하는 방법에는 다양한 방법이 있고 직접적인 방법과 간접적인 방법으로 구분된다.

직접적인 방법으로는 로짓 (logit), 프로빗, 위험모형 등이 있다. 이 방법은 평가 스코어가 직접적으로 차주의 PD로 이용될 수 있고, 스코어가 직접적으로 차주의 PD로 이용될 수 없을 때 간접적인 방법이 이용된다. 간접적인 방법 중 가장 간단한 것은 각 등급별 과거 부도율 자료로부터 PD를 추정하는 것이다. 일반적으로 예측된 PD는 관측되는 부도율과 차이가 있는데 예측 부도율과 관측 부도율에 차이가 무작위로 이루어지는 것인지 아니면 체계적으로 이루어지는 것인지를 여부가 중요한 문제이다. 즉, 그 차이가 무작위가 아닌 구조적인 문제에 의해 발생된다면 예측 부도율과 실제 부도율이 차이가 있다는 것으로 간주되며 등급의 계량화가 적합하게 이루어지지 않았다고 할 수 있다 (이석형과 심재호, 2006).

예측 부도율과 실제 부도율의 차이를 검증하는 방법 즉, 등급별 계량화가 적절하게 이루어졌는지 검증하는 통계적인 방법으로는 이항 검정과 카이제곱 검정, 정규성 검정, Brier Score, 신호등 검정, 확장된 신호등 검정이 있다.

2.1. 이항 검정 (Binomial test)

이항 검정은 특정 시점, 특정등급에서 부도확률 예측의 정확성 검증에 대한 방법이다. 즉, 한 시점에서 각각의 등급별로 부도율을 예측하였을 때, 이 예측 부도율이 실제 부도율을 예측하기에 적합한지를 검증하는 것이다. 기본 가정은 동일한 등급에서 차주들의 부도율은 동일하고 각 등급 내에서 발생하는 부도사건은 서로 독립적이라는 것이다.

H_0 : 각 등급 내에서 실제 부도율 (p)과 예측 부도율 (PD)은 같다.

H_1 : 각 등급 내에서 예측 부도율 (PD)은 실제 부도율 (p)에 비해 과소 추정되었다.

여기서 예측 부도율 (PD)은 등급별로 설정된 목표 부도율이고, 실제 부도율 (p)은 등급에 속한 차주들의 내재적인 부도율을 말한다. 유의수준 α 하에서 만약 해당 등급의 부도 차주의 수 k 가 아래에 산출된 임계값 d^* 보다 크거나 같으면 귀무가설을 기각하게 된다.

$$d^* = \min \{d | P[D_n \geq d] \leq \alpha\} = \min \left\{ d \mid \sum_{i=d}^n \binom{n}{i} PD^i (1-PD)^{n-i} \leq \alpha \right\}$$

D_n : 부도 차주수, n : 해당 등급 차주의 수, PD : 등급의 예측 부도율

이항 검정은 상관관계를 고려하지 않고 부도사건이 서로 독립이라는 가정 하에서 실시된다. 하지만 실제 부도사건들은 독립적으로 발생하지 않으므로, 이를 고려한 모형이 필요하다. 실제로 부도사건 사이에는 약 0.5%에서 3% 정도의 작은 상관계수를 가진다고 알려져 있으므로 상관관계를 반영한 이항 검정을 수행해야 할 것이다.

부도사건 사이의 상관관계를 고려하기 위한 단일요인모형을 가정한 경우 부도 차주수의 확률분포는 이항분포의 정규근사 성질 등을 이용하여 아래의 근사식으로 나타낼 수 있다 (BCBS, 2005).

$$\begin{aligned} P(D_n \leq k) &= \int_{-\infty}^{\infty} \sum_{l=0}^k \binom{n}{l} \Phi \left(\frac{\gamma - \sqrt{\rho}x}{\sqrt{1-\rho}} \right)^l \times \left(1 - \Phi \left(\frac{\gamma - \sqrt{\rho}x}{\sqrt{1-\rho}} \right) \right)^{n-l} \phi(x) dx \\ &\approx \Phi \left(\frac{\sqrt{1-\rho} \Phi^{-1} \left(\frac{k}{n} \right) - \gamma}{\sqrt{\rho}} \right) \end{aligned}$$

단,

D_n : 부도 차주수, $\gamma = \Phi^{-1}(PD)$: 부도 임계점, $\rho = Corr(R_i, R_j)$: 자산상관계수,

n : 해당 등급 차주의 수, Φ : 표준정규분포함수, ϕ : 표준정규 확률밀도함수

2.2. 카이제곱 검정 (Chi-square test)

이항 검정은 특정 시점에서 각각의 등급별에 대한 개별적인 검정인 반면 카이제곱 검정은 서로 다른 등급들을 동시에 검정하는 방법이다. 먼저 0등급부터 k 등급 까지 $k+1$ 개의 등급구간이 존재하고 각각의 구간에서 예측 부도율을 PD_0, \dots, PD_k , 실제 부도율을 p_0, \dots, p_k 라고 할 때 예측 부도율과 실제 부도율을 이용하여 카이제곱 분포에 근사한다는 사실을 이용할 수 있다. 이 때, 각 등급 간 부도사건은 독립적으로 발생한다고 가정한다.

H_0 : $p_0 = PD_0, p_1 = PD_1, \dots, p_k = PD_k$

H_1 : not H_0

$$T_k = \sum_{i=0}^k \frac{(n_i PD_i - d_i)^2}{n_i PD_i (1 - PD_i)} \sim \chi_{k+1}^2$$

n_i : 등급 i 의 차주수, d_i : 등급 i 의 부도 차주수

검정통계량 T_k 는 모든 등급구간에 대해 차주의 수가 충분히 크고 모든 등급구간별 예측 부도율이 정확하다는 가정 하에 카이제곱 분포를 따르게 된다. 따라서 유의수준 α 에서 $T_k > \chi_\alpha^2(k+1)$ 이면 귀무가설을 기각하게 되어 일부 등급에서는 실제 부도율이 예측 부도율과 다르다고 할 수 있다.

2.3. 정규성 검정 (Normal test)

정규성 검정은 연도별로 예측 부도율과 실제 부도율을 비교하는 검정 방법이다. 이항 검정과 카이제곱 검정은 한 시점에 대해서만 검정이 가능하지만 정규성 검정은 여러 시점에 대해서도 비교가 가능하다. 기본 가정은 부도율이 작은 변동성을 가지고 있으며 서로 다른 시점 간에 부도사건은 독립적이라는 것이다.

H_0 : 관측기간 T 년 동안 실제 부도율 p_t 중 그 어느 것도 예측 부도율 PD_t 보다 크지 않다.

$$(p_1 \leq PD_1, p_2 \leq PD_2, \dots, p_T \leq PD_T)$$

H_1 : 관측기간 T 년 동안 적어도 하나의 실제 부도율 p_t 는 예측 부도율 PD_t 보다 크다. (not H_0)

기본 가정이 성립할 경우 중심극한정리에 의해 다음과 같은 통계량을 생각할 수 있다.

$$Z = \frac{\sum_{t=1}^T (\hat{p}_t - PD_t)}{\sqrt{T}\tau} \sim N(0, 1)$$

T : 총 관측기간, PD_t : 기간 t 의 예측 부도율, \hat{p}_t : 기간 t 의 관측 부도율, τ^2 : \hat{p}_t 의 분산추정치

$$\tau^2 = \frac{1}{T-1} \sum_{t=1}^T (\hat{p}_t - PD_t)^2 - \frac{1}{T(T-1)} \left(\sum_{t=1}^T (\hat{p}_t - PD_t) \right)^2$$

유의수준 α 에서 검정통계량 $Z > z_\alpha$ 이면 귀무가설이 기각된다. 즉, 관측기간 T 년 동안 실제 부도율 p_t 중 어느 것은 예측 부도율 PD_t 보다 크다고 할 수 있다.

2.4. Brier score (BS)

Brier Score 방법은 개별 차주의 예측 부도율과 실제 관측된 결과의 편차 제곱을 평균한 평균제곱오차 (MSE; Mean squared error) 개념을 이용한 것이다. 이 검정은 특정 시점에서 전체 등급에 대한 검정 방법이다.

$$BS = \hat{p}(1 - \hat{p}) + \frac{1}{n} \sum_{i=1}^K n_i (PD_i - \hat{p}_i)^2 - \frac{1}{n} \sum_{i=1}^K n_i (\hat{p}_i - \hat{p})^2$$

\hat{p} : 전체 관측 부도율, \hat{p}_i : 각 등급의 관측 부도율, n_i : 차주수, PD_i : 각 등급의 예측 부도율

이 통계량은 등급계량화 적정성 외에도 모형변별력, 불확실성을 나타내는 값이 있어 등급 계량화에 대한 판단이 어렵다. 특히, 첫 번째 불확실성의 값이 상대적으로 크기 때문에 BS 대신에 BS를 불확실성으로 표준화한 BSS (Brier Skill Score)를 이용하기도 한다.

$$BSS = \frac{\hat{p}(1 - \hat{p}) - BS}{\hat{p}(1 - \hat{p})} = 1 - \frac{BS}{\hat{p}(1 - \hat{p})}$$

이상적인 모형일 경우 BSS 값이 1이 될 것이며, 0보다 작은 값이면 등급계량화가 적절히 이루어지지 않았다고 판단할 수 있다.

2.5. 신호등검정 (Traffic lights approach; TLA)

이항 검정에 상관관계를 반영하여 일반화한 것이 신호등 검정 (Traffic lights) 방법이다 (Tasche, 2003). 이 방법에서도 상관관계를 고려한 이항 검정과 마찬가지로 단일요인모형을 이용하여 특정 시점의 각 등급에 대한 검정 방법이다.

신호등 검정 방법에서는 예측 부도율 PD 에 대해 두 개의 신뢰수준 $1 - \alpha_{low}$, $1 - \alpha_{high}$ (ex: 95%, 99.9%)으로 임계값 c_{low} , c_{high} 을 구하여 관측 부도율 \hat{p} 이 c_{low} 보다 작으면 Green으로, c_{low} 보다 크거나 같으면서 c_{high} 보다 작으면 Yellow로, c_{high} 보다 크거나 같으면 Red로 표시함으로써 등급 계량화의 적합성을 검증한다. 임계값 c_{low} , c_{high} 는 다음과 같이 구할 수 있다.

$$c_{low} = \min \{k : P[X > k] \leq \alpha_{low}\} = q(1 - \alpha_{low}, X) + 1$$

$$c_{high} = q(1 - \alpha_{high}, X) + 1$$

X : 실제 관측된 부도 차주수, $q(1 - \alpha, X)$: X 의 $100 \times (1 - \alpha)$ 분위수

관측 부도율을 \hat{p} 이라고 할 때, 관측 부도율을 이용하여 $q(1 - \alpha, X) = nq(1 - \alpha, \hat{p})$ 으로 분위수를 구할 수 있다. Gordy (2002)는 분위수 $q(1 - \alpha, \hat{p})$ 을 근사적으로 얻는 방법을 제안하였고, 부도 차주수가 아닌 부도율을 기준으로 사용한 식은 다음과 같다 (OeNB, 2005).

$$\hat{p} > Q + \frac{(2Q - 1)}{2n} + \frac{1}{2n} \left(\frac{Q(1 - Q)}{\phi \left(\frac{\sqrt{\rho}\Phi^{-1}(\alpha) - t}{\sqrt{1 - \rho}} \right)} \right) \times \left(\frac{\sqrt{\rho}\Phi^{-1}(\alpha) - t}{\sqrt{1 - \rho}} - \sqrt{\frac{1 - \rho}{\rho}} \Phi^{-1}(\alpha) \right)$$

$$\alpha : \text{유의수준}, \rho : \text{자산상관계수}, Q = \Phi \left(\frac{\sqrt{\rho}\Phi^{-1}(1 - \alpha) + t}{\sqrt{1 - \rho}} \right), t = \Phi^{-1}(PD)$$

2.6. 확장된 신호등 검정 (Extended traffic lights approach; ETLA)

확장된 신호등 검정은 Blochwitz 등 (2003)이 시도한 방법으로 신용 등급별 관측 부도율이 예측 부도율과 얼마나 떨어져있는가를 신호등 색깔로 구분하여 여러 시점의 등급 안정성에 대한 검정 방법이다. 부도사건이 차주별로 서로 독립적이고 각 시점별로도 독립적이라고 가정하며, 귀무가설 하에서 신호등 색깔로 정의된 결과의 확률분포는 다항분포를 따른다는 사실을 이용한다.

t 년에 발생한 부도사건이 독립적이고 같은 포트폴리오 안에 속해있는 모든차주가 동일한 부도확률 p_t 를 가진다고 한다면, 부도차주 발생 빈도 D_t 는 이항분포를 따른다. 이항 분포에서 $N_t p_t$ 가 너무 작지 않다면 중심극한정리에 의해 표준정규분포로 근사 시킬 수 있고, 표준화된 부도율 R_t 를 사용 할 수 있다.

$$R_t = \frac{D_t - N_t p_t}{\sqrt{N_t p_t (1 - p_t)}}$$

N_t : 각 시점별 차주수, D_t : 각 시점별 부도차주의 수

계산된 R_t 를 이용하여 각각의 부도확률을 매핑하고, 매핑한 분포함수 $C(x)$ 는 다음과 같이 정의할 수

있다.

$$C(x) = \begin{cases} g, & x \leq \Phi^{-1}(\pi_g), \\ y, & \Phi^{-1}(\pi_g) < x \leq \Phi^{-1}(\pi_g + \pi_y), \\ o, & \Phi^{-1}(\pi_g + \pi_y) < x \leq \Phi^{-1}(\pi_g + \pi_y + \pi_o), \\ r, & \Phi^{-1}(\pi_g + \pi_y + \pi_o) < x \end{cases}$$

$g : green, y : yellow, o : orange, r : red$

확률 $\pi_g + \pi_y + \pi_o + \pi_r = 1$ 이고 Blochwitz 등 (2003)에서 $(\pi_g, \pi_y, \pi_o, \pi_r) = (0.5, 0.3, 0.15, 0.05)$ 으로 사용하고 있다.

T 년 동안 발생한 신호등 색깔의 횟수를 A 로 표시할 때 $A = (k_g, k_y, k_o, k_r)$ 라고 정의한다. 여기서 k_g 는 T 년 동안 green 영역에 발생한 횟수, k_y 는 T 년 동안 yellow 영역에 발생한 횟수 등을 의미한다. 이 때, 각 연도별 부도사건을 독립으로 가정했으므로 벡터 A 는 아래와 같이 다항분포로 근사시킬 수 있다.

$$P[A = (k_g, k_y, k_o, k_r)] = \frac{T!}{k_g!k_y!k_o!k_r!} \pi_g^{k_g} \pi_y^{k_y} \pi_o^{k_o} \pi_r^{k_r}, k_g + k_y + k_o + k_r = T$$

임계구간을 찾기 위해서 관측기간이 9년 이 하인 경우 $V = 1000k_g + 100k_y + 10k_o + k_r$ 로 정의된 통계량을 사용할 수 있다. 가설검정을 위해 유의수준 α 에서 $V \leq \nu_\alpha$ 인 경우 관측기간 T 년 동안에 실제 부도율 어느 것도 예측 부도율을 보다 크지 않다는 귀무가설을 기각하게 된다.

H_0 : 관측기간 t 년 동안에 실제 부도율 어느 것도 예측 부도율을 보다 크지 않다.

$$(p_t \leq PD_t \ t = 1, \dots, T)$$

H_1 : 관측기간 t 년 동안에 적어도 하나의 실제 부도율은 예측 부도율을 보다 크다. (not H_0)

표 2.1 여러 가지계량화 방법 비교

검정 방법	기본 가정	시점	특징
이항 검정	-각 등급 내에서 부도 사건이 독립	한 시점	-특정 시점의 각 등급에 대한 검증방법 -상관관계를 고려할 수 있음
카이제곱 검정	-각 등급 내에서 부도 사건이 독립	한 시점	-특정 시점의 전체 등급에 대한 검증 방법 -상관관계를 고려하지 않음
정규성 검정	-부도율의 변동성이 작음 -서로 다른 연도 간 부도 사건은 독립	여러 시점	-여러 시점의 각 등급에 대한 검증 방법 -상관관계를 고려하지 않음
Brier Score	-각 등급 내에서 부도 사건이 독립	한 시점	-특정 시점의 전체 등급에 대한 검증 방법 -상관관계를 고려하지 않음
신호등 검정	-각 등급 내에서 부도 사건이 상관관계를 가짐	한 시점	-특정 시점의 각 등급에 대한 검증방법 -상관관계를 고려할 수 있음
확장된 신호등검정	-각 등급 내에서 부도 사건이 독립 -서로 다른 연도 간 부도사건은 독립	여러 시점	-여러 시점의 각 등급에 대한 정확성 검증방법 -상관관계를 고려하지 않음

위에서 제시된 여섯 가지의 계량화 방법을 정리하면 표 2.1과 같다.

3. 상관관계를 고려한 ETLA 방법

신용손실에 대한 분포를 추정하는 과정에서 중요한 역할을 하는 요인 중 하나가 상관관계를 고려하는 것이다. 확장된 신호등 검정에는 상관관계를 반영하지 않은 임계값으로 구간을 나누어 검정하고 있으므로, 상관관계를 반영한 신호등검정은 특정 시점의 등급에 대한 검정을 하고 있다. 본 논문에서는 여러 기간의 부도를 정확성 검증에서 임계값에 상관관계를 반영하고자 한다. 검정을 위해 상관관계를 반영한 신호등 검정을 이용하여 확장된 신호등 검정 방법의 임계값을 구하였다. 구간별 비율은 확장된 신호등 검정에서 Blochwitz 등 (2003)이 사용한 $(\pi_g, \pi_y, \pi_o, \pi_r) = (0.50, 0.30, 0.15, 0.05)$ 를 이용하고 관측부도를 \hat{p}_i 을 매핑하고, 매핑한 분포함수 $C^*(x)$ 는 다음과 같다.

$$C^*(x) = \begin{cases} g, & x \leq T(\pi_g), \\ y, & T(\pi_g) < x \leq T(\pi_g + \pi_y), \\ o, & T(\pi_g + \pi_y) < x \leq T(\pi_g + \pi_y + \pi_o), \\ r, & T(\pi_g + \pi_y + \pi_o) < x \end{cases}$$

이 때의 $T(q)$ 는 상관관계를 반영한 임계값으로 신뢰수준 q 를 기준으로 계산된 값이다.

$$T(q) = Q + \frac{1}{2n} \left(2Q - 1 + \frac{Q(1-Q)}{\phi \left(\frac{\sqrt{\rho}\Phi^{-1}(1-q) - t}{\sqrt{1-\rho}} \right)} \left(\frac{\sqrt{\rho}\Phi^{-1}(1-q) - t}{\sqrt{1-\rho}} - \sqrt{\frac{1-\rho}{\rho}} \Phi^{-1}(1-q) \right) \right)$$

n : 차주수, PD: 예측부도율, q : 신뢰수준, ρ : 자산상관계수,

$$t = \Phi^{-1}(PD), Q = \Phi \left(\frac{\sqrt{\rho}\Phi^{-1}(q) + t}{\sqrt{1-\rho}} \right)$$

상관계수를 반영하여 구한 분포함수는 확장된 신호등 검정과 같은 방법으로 임계구간을 구할 수 있다. 따라서 1 ~ T년동안 발생한 신호등 색깔의 횟수를 A로 표시했을 때 $A = (k_g, k_y, k_o, k_r)$ 라고 정의한다. 이 때, 각 연도별 부도사건이 독립이라고 가정하고, 벡터 A는 아래와 같이 다항분포로 근사시킬수 있다.

$$P[A = (k_g, k_y, k_o, k_r)] = \frac{T!}{k_g!k_y!k_o!k_r!} \pi_g^{k_g} \pi_y^{k_y} \pi_o^{k_o} \pi_r^{k_r}$$

$$\text{단, } \pi_g + \pi_y + \pi_o + \pi_r = 1, k_g + k_y + k_o + k_r = T$$

임계구간이나 순서화를 위한 방법은 확장된 신호등 검정의 방법과 동일하며, 다음과 같은 가설검정을 위해 유의수준 α 에서 $V \leq \nu_\alpha$ 인 경우 귀무가설을 기각하게 된다. 이 때 ν_α 는 $P[V \leq \nu] \leq 1 - \alpha$ 를 만족하는 ν 중 가장 큰 수이다.

H_0 : 관측기간 t 년 동안에 실제 부도율 중 어느 것도 예측 부도율 보다 크지 않다.

$$(p_t \leq PD_t \quad t = 1, \dots, T)$$

H_1 : 관측기간 t 년 동안에 적어도 하나의 실제 부도율은 예측 부도율 보다 크다. (not H_0)

4. 검정 결과 비교

분석에 사용된 자료는 이석형과 심재호 (2006)이 정규성 검정과 확장된 신호등 검정에서 사용한 5년간 자료를 동일하게 사용하여 상관관계를 고려한 확장된 신호등 검정과 상관관계를 고려하지 않은 이항검정의 결과를 비교하였다.

표 4.1 정규성 검정과 확장된 신호등 검정

년도	총차주수(n)	실제 부도 차주수(d)	관측 부도율 (d/n)	예측 부도율 (PD _t)	(d/n- PD _t)	R _t	color
2001	2000	52	2.60%	2.48%	0.12%	0.345	Y
2002	2500	84	3.36%	3.19%	0.17%	0.484	Y
2003	3000	70	2.33%	2.28%	0.05%	0.196	Y
2004	2500	79	3.16%	3.06%	0.10%	0.290	Y
2005	3000	77	2.57%	2.71%	-0.14%	-0.483	G
합					0.30%		

표 4.1에서 보듯이 5년간의 자료에 대해서 정규성 검정을 계산하면,

$$Z = \frac{\sum_{t=1}^T (\hat{p}_t - PD_t)}{\sqrt{T}\tau} = \frac{0.003}{\sqrt{5} \times 0.0012} = 1.118 < z_{0.95} (= 1.96)$$

으로 Z-검정 통계량 값 1.118은 임계값 (1.96) 보다 작으므로 귀무가설을 채택하게 되어 관측기간 동안의 실제 부도율은 예측 부도율보다 크지 않다고 할 수 있다.

확장된 신호등 검정의 결과 2005년은 Green이고, 그 외 기간은 Yellow로 표시된다. 아래의 표 4.2는 누적확률과 누적확률 값에 따른 구간의 발생횟수를 나타낸 결과이다.

표 4.2 확장된 신호등 검정 결과

구분	π	누적 π	임계값	발생횟수
Green	0.5	0.5	0	1
Yellow	0.3	0.8	0.842	4
Orange	0.15	0.95	1.645	0
Red	0.05	1		0

따라서 검정통계량을 계산하면, $V = (1000 \times 1) + (100 \times 4) + (10 \times 0) + 0 = 1400$ 으로 유의수준 0.05에서 $\nu_\alpha = 1310$ 이므로 정규성 검정과 마찬가지로 귀무가설을 채택하게 되어, 관측기간 동안에 실제 부도율 어느 것도 예측 부도율 보다 크지 않다고 할 수 있다.

아래의 표 4.3은 정규성 검정과 확장된 신호등 검정에서 사용된 5년간의 자료를 동일하게 사용하여, 상관관계를 반영한 확장된 신호등 검정과 상관관계를 고려하지 않은 이항 검정을 실시한 결과이다. g, y, o 는 각각의 임계값 $T(\pi_g), T(\pi_g + \pi_y), T(\pi_g + \pi_y + \pi_o)$ 를 나타낸다.

표 4.3 상관관계를 반영한 확장된 신호등 검정 (단위: 건, %)

년도	\hat{p}	상관관계를 고려하지 않은 경우					상관관계를 고려한 경우							
							color ρ = 0.01				ρ = 0.1			
		임계값					임계값				임계값			
		g	y	o	color	g	y	o	color	g	y	o	color	
2001	2.60	2.48	2.77	3.05	Y	2.42	3.03	3.71	Y	1.92	3.70	6.46	Y	
2002	3.36	3.19	3.49	3.77	Y	3.12	3.84	4.62	Y	2.53	4.73	8.04	Y	
2003	2.33	2.28	2.51	2.73	Y	2.22	2.77	3.38	Y	1.75	3.40	5.99	Y	
2004	3.16	3.06	3.35	3.63	Y	2.99	3.69	4.45	Y	2.42	4.55	7.75	Y	
2005	2.57	2.71	2.96	3.20	G	2.65	3.27	3.96	G	2.12	4.03	6.97	Y	

전체적인 비교로 표 4.3을 살펴보면, 상관관계를 고려하지 않은 이항 검정보다 상관관계를 고려한 경우의 $T(\pi_g + \pi_y), T(\pi_g + \pi_y + \pi_o)$ 에서 더 큰 임계값을 가지고, $T(\pi_g)$ 의 경우에는 상관관계를 고려할 경우 더 작은 임계값을 가지는 것을 볼 수 있다. 신호등 표시는 상관관계를 고려하지 않은 경우와 상

관관계 0.01을 고려한 확장된 신호등 검정의 결과가 동일하게 나타내고 있으며, 상관계수 0.1을 가지는 경우에는 모두 Yellow로 나타났다. 그러나 상관관계를 반영한 두 개의 결과 모두 누적확률은 0.1524와 0.3344의 값을 가지고 있다. 따라서 귀무가설을 채택하게 되고 실제 부도율이 예측 부도율보다 크지 않다는 결론을 얻을 수 있다.

위의 결과를 가지고 확률을 계산하면, $(k_g, k_y, k_o, k_r) = (1, 4, 0, 0)$ 으로 표시되고, 각 영역에 대한 확률은 $(\pi_g, \pi_y, \pi_o, \pi_r) = (0.5, 0.3, 0.15, 0.05)$ 를 사용한다. 즉 표시된 결과를 해석하기 위해서 앞서 말한 3절의 다항분포를 가정하여 사용하게 되면 $P[A(1, 4, 0, 0)] = (5!/1!4!0!0!) \times \pi_g^1 \pi_y^4 \pi_o^0 \pi_r^0 = 0.0203$ 으로 계산할 수 있고, 발생회수와 확률을 이용하여 계산된 누적확률은 0.3344이다. ETLA의 결과는 누적확률을 기준으로 0.05까지는 Red, 0.2까지는 Orange, 0.5까지는 Yellow, 그 이상은 Green으로 나타내므로, 누적확률 0.3344는 Yellow를 나타낸다고 할 수 있다.

앞에서 설명한 신호등 검정은 상관관계가 높을수록 더 큰 임계값을 가진다고 알려져 있지만, 상관관계를 고려한 ETLA 방법의 임계값을 살펴본 결과 $T(\pi_g)$ 는 상관계수 0.01 반영 했을 때 보다 0.1 반영 했을 때 더 작은 임계값을 가진다는 것을 알 수 있었다. 또한 $T(\pi_g)$ 의 구간비율을 0.5가 아닌 0.67로 정하면 상관계수 0.01과 0.1은 같은 결과를 나타내었고, 0.67이상의 구간에서는 신호등 검정에서 알려진 바와 같이 상관계수가 높을수록 더 큰 임계값을 가지는 결과를 볼 수 있었다.

5. 결론

은행들은 내부등급법 승인을 받기 위하여 신용평가모형의 적합성 검증 체제를 가지고 있어야 한다. 적합한 신용평가 시스템은 은행의 입장에서 내부등급법 승인뿐만 아니라 리스크 관리 측면에서도 가장 중요한 역할을 하고 있다.

가장 많이 사용되고 있는 방법 중 하나인 확장된 신호등 검정은 여러 기간의 부도율 정확성 검증방법에 사용된다. 이 방법은 정상적인 PD 구간과 충분히 많은 익스포저가 존재할 때, 그리고 부도율 관찰횟수가 증가할수록 더 좋은 결과를 갖는다. 그러나 여러 시점의 적합성을 검증하므로 각 시점에 대한 정보의 손실이 있으며, 부도사건이 독립이라는 기본 가정으로 검증하는 방법이므로 상관관계를 고려하지 않아 연쇄적인 부도의 발생과 같은 현실적인 문제를 반영하기 힘들다. 이와 같이 부도사건이 독립적이라는 가정을 갖는 이항검정, 카이제곱, 정규성 검정 또한 이러한 문제를 피하기 어렵다. 따라서, 이러한 문제를 보완하고자 확장된 신호등 검정에 상관관계를 반영하였다. 상관관계를 고려한 방법은 TLA에서 사용한 방법을 이용하여 분석하였다. 분석 결과 상관계수와 구간비율의 결정에 따라 임계값이 달라짐을 볼 수 있었다. 특히 구간비율이 낮은 경우 상관계수가 커짐에 따라 임계값이 낮아지는 것을 확인할 수 있어 Green 보다 Yellow 신호등이 나타날 가능성이 더 높았으며 구간비율이 높은 경우 이 상관계수가 커짐에 따라 임계값이 높아져서 Red 보다 Orange 신호등이 나타날 가능성이 더 높아졌다. 자료의 부족으로 인해 여러 가지 경우를 분석하지 못한 아쉬움이 있었다. 다양한 시뮬레이션을 통해 경우에 따른 변화를 분석하는 것은 다음 논문의 주제로 남겨둔다.

참고문헌

- 금융감독원 (2005). <신용리스크 내부등급법 세부지침 (안)>, 금융감독원, 서울.
- 금융감독원 (2006). <알기 쉬운 신BIS협약>, 금융감독원, 서울.
- 이석형, 심재호 (2006). 신BIS 기준에 따른 양적 적합성검증 방법론 고찰. 리스크리뷰, 2006년 여름호.
- 임종건 (2006). 신용등급 계량화에 대한 적합성검증 방법론, 리스크리뷰, 2006년 가을호.
- 임철순 (2005). 신BIS기준에 따른 신용평가시스템 적합성검증, 리스크리뷰, 2005년 가을호.
- 허정철 (2009). 신용등급 변화감지에 관한 방법 개발 연구. <한국자료분석학회지>, 11, 827-837.

- 홍중선, 김지훈 (2009). 신용평가모형에서 두 분포함수의 동일성 검정을 위한 비모수적인 검정방법. <한국데이터정보과학회지>, **20**, 261-272.
- Basel Committee on Banking Supervision (BCBS) (2005). Studies on the validation of internal rating systems (revised). *Working paper*, **14**.
- Bernd, E. and Robert, R. (2006). *The basel II risk parameters: Estimation, validation, and stress testing*, Springer Verlag.
- Blochwitz, S., Carsten, S. W. and Hohl, S. (2005). *Reconsidering ratings*, Working paper.
- Gordy, M. B. (2002). *A risk factor model foundation for rating based bank capital rules*, Board of Governors of the Federal Reserve System.
- Hong, C. S. and Choi, J. M. (2008). Validation comparison of credit rating models using Box-Cox transformation. *Journal of the Korean data & Information Science Society*, **19**, 789-800.
- Hong, Y. W. and Suh, J. S. (2008). Estimating the credit value at risk of korean property and casualty insurers. *Journal of the Korean data & Information Science Society*, **19**, 1027-1036.
- Jeong, W. K (2003). The return generating process of corporate bonds based on credit ratings. *Journal of the Korean data & Information Science Society*, **14**, 805-815.
- OeNB. (2005). *Rating models and validation*, Guidelines on Credit Risk Management.
- Tasche, D. (2003). *A traffic lights approach to PD validation*, Working paper.

Study on the validation methods of calibration considering correlations

Enn Na Kim¹ · Jeong Cheol Ha²

¹² Department of Statistics, Keimyung University

Received 15 March 2010, revised 20 April 2010, accepted 6 May 2010

Abstract

In Basel II compliance, internal rating systems are allowed for banks to enhance the self control and the validation of the system are getting more important. The validation methods are composed of qualitative test and quantitative test, three basic standards of which are discriminatory power, stability and calibration. The aim of this article is to review the quantitative tests for calibration and find a new method for it. These methods for discrimination between forecasted PD and observed PD include binomial test, chi square test, Brier score, traffic lights approach, normal test and extended traffic lights approach. We introduce a modified extended traffic lights approach considering asset correlations.

Keywords: Calibration, correlation, credit ratings, extended traffic lights approach, validation.

¹ Graduate student, Department of Statistics, Keimyung University, Daegu 704-701, Korea.

² Corresponding author: Assistant professor, Department of Statistics, Keimyung University, Daegu 704-701, Korea. E-mail: jeicy@kmu.ac.kr