

사례기반추론을 이용한 비상장기업 및 신규상장기업의 VaR 추정*

최경덕

한국정보통신
(kdcchoi@kicc.co.kr)

노승중

아주대학교 경영대학
(sjnoh@madang.ajou.ac.kr)

비상장기업이나 신규상장기업의 경우 주식거래량과 거래가격이 없거나 불충분하므로 최근 활발히 연구되고 있는 VaR(Value-at-Risk)를 파악하지 못한다. 본 연구에서는 비상장기업 및 신규상장기업의 미래 가격위험의 척도인 VaR를 추정하는 방법론을 제시하고, 이를 시스템(VAS-CBR)으로 구현하였다. 구체적으로는 사례기반추론(Case Based Reasoning: CBR) 기법을 이용하여 기존의 상장회사들 중에서 신규 및 비상장기업과 유사한 재무적, 비재무적 특성을 갖는 상장기업을 찾아내고, 유사기업의 VaR를 근거로 신규 및 비상장기업의 VaR를 간접적으로 추정하였다. 또한 개발 시스템의 예측력 제고를 위한 운용방안 및 시스템의 예측력을 실험을 통하여 밝혔다.

1. 서론

재무학에서 논하는 효율적인 위험관리를 위해서는 적정 위험수준을 설정하고 선택한 위험 하에서 최대의 기대이익을 얻는 전략이 필요하다. 이를 위해서 특정 시점에서 금융자산과 부채의 노출위험을 측정하는 방법으로 최근 활발히 연구되는 것이 VaR(Value-at-Risk) 시스템이다. VaR란 금리, 주가, 환율 등 기초적 시장가격¹⁾에 대한

미래 분포를 예측하여, 향후 불리한 시장가격 변동이 특정 신뢰구간 내에서 발생하는 경우 나타날 수 있는 포트폴리오 가치의 최대손실규모를 산출하는 기법이다. 금융기관들은 이렇게 계산된 최대손실규모를 분석하여 포트폴리오의 위험노출 정도를 파악하고, 이를 바탕으로 위험에 대한 헤지 전략을 구사하게 된다. 그러나 VaR와 같이 특정분야의 과거 자료를 사용하는 모형의 단점은 과거 자료가 없는 경우 미래를 예측하지 못한다는 것이며, 이는 통합적인 위험관리 시스템 운영상 커다란 문제점이 될 수 있다. 특히, 주식시장의 경우 비상장기업의 상장 신청사례가 과거에 비하여 증가하고 있어 실제적인 VaR의 측정이 취약한 곳이기도 하다. 또한 신규상장기업의 경우 지식획득의 병목현상으로 인하여 상장기업의 가격위험을 제대로 파악한다는 것이 어려운 실정이다.

* 본 논문은 1999년도 한국학술진흥재단의 연구비에 의하여 연구되었음(KRF-99-003-C00322)

1) 기초적 시장가격이란 금리, 주가, 환율등 포트폴리오가치를 결정하는 기본적인 가격변수를 의미한다. 예를 들어, 포트폴리오구성이 전환사채와 콜옵션으로 된 경우 포트폴리오가치는 전환사채와 콜옵션가치의 함수로 나타나지만, 전환사채와 콜옵션의 가치는 다시 금리와 주가의 함수로 표현되므로 이 포트폴리오의 기초적 시장가격 변수는 금리와 주가가 되는 것이다.

VaR의 정의와 측정방법, 유효성 검증에 대해서는 최근 활발한 연구가 있었고, 주로 통계적인 방법을 사용하여 왔다(오세경, 1999; Jorion, 1997; Lopez, 2000; Pritsker, 1997). 또한 기업의 재무정보에 데이터마이닝 기법을 사용하여 판매예측 및 주가수익률 예측 등을 시도한 연구가 있었다(양경식, 1998; 안광열, 2001). 그러나 VaR의 추정에 사례기반추론 기법을 적용한 연구는 아직 없었던 것으로 파악된다. 본 연구에서는 비상장기업 및 신규상장기업의 미래 가격위험과 VaR를 사례기반추론(Case Based Reasoning: CBR) 기법을 이용하여 추정하는 방법론을 제시하고 시스템을 구현하였다. 기존의 상장회사들 중에서 신규 및 비상장기업과 재무적, 비재무적으로 유사한 특성을 갖는 기업을 찾아내고 유사기업의 VaR를 근거로 신규 및 비상장기업의 VaR를 간접적으로 추정하는 접근방법을 사용하였다. VaR 측정방법으로는 델타평가법(delta-normal valuation method)을 사용하였다.

이하 제2장에서는 VaR의 개념을 약술하였고, 제3장에서는 CBR을 이용한 VaR 추정 방법을 소개하였다. 제4장에서는 이를 실제 기업 데이터에 적용하여 신규 및 비상장기업의 VaR를 추정하는 시스템을 구축한 내역을 소개하였고 제5장에서는 이 시스템의 성능을 평가하였으며, 제6장에 결론 및 향후 연구방향을 제시하였다.

2. VaR의 개념

VaR란 목표보유기간, 신뢰수준 및 확률분포를 전제로 할 때, 정상적인 시장조건하에서 발생할 수 있는 포트폴리오의 최대 손실예상액 추정 수치를 말한다. N개의 포지션으로 구성되는 포트폴

리오가 있다고 할 때 t시점에서의 포트폴리오의 가치는 다음과 같다.²⁾

$$S_t = \sum_{j=1}^N S_{j,t} \phi_j(F_{1,t}, F_{2,t}, \dots, F_{m,t})$$

이 때 $S_{j,t}$ 는 t시점에서의 j번째 자산의 투자단 위이고, $F_{1,t}, F_{2,t}, \dots, F_{m,t}$ 는 자산 $S_{j,t}$ 의 가치 변동에 영향을 주는 m개의 기초적 시장가격 또는 기타 위험요인이며, ϕ_j 는 j 번째 자산 한 단위의 시장가치를 평가하는 함수이다. 따라서 확률적으로 정의되는 VaR는 ΔS 를 포트폴리오의 가치변동이라 할 때 다음과 같이 나타낼 수 있다.

$$P[\Delta S > -v] = 1 - \alpha \quad (1)$$

즉, 포트폴리오의 가치변동이 $-v$ 보다 클 확률이 $1 - \alpha$ 라고 할 때 이 값 v 를 신뢰수준 $100 \times (1 - \alpha)\%$ 에서의 VaR라고 정의한다. α 는 유의수준이 된다. 예를 들어, 어떤 포트폴리오의 VaR가 1일 동안 신뢰수준 95%에서 10억 원이라면, 이는 이 포트폴리오를 보유함으로써 향후 1일동안 10억 원 이상의 손실을 보게 될 확률이 5%임을 의미한다. VaR는 기본적으로 포트폴리오 가치의 미래 변동성으로 위험을 측정한다. 포트폴리오 가치의 변동은 이를 구성하는 개별 포지션과 관련된 시장가격변화로부터 가능하기 때문에 사실상 위험은 시장가격들의 미래 변화로부터 나타난다. 따라서 금융상품들의 시장가격의 통계적 특성은 가장 중요한 분석대상이 된다.

2) 통상 N개의 포지션 대신 자산이라는 용어를 사용하기도 한다. 그러나 최근 금융상품은 공매(short sale)등 자산이라고 하기 어려운 것들이 많아 이를 모두 포함하는 포지션이라는 용어를 사용한다.

3. 사례기반추론을 이용한 VaR 추정 시스템 설계

본 연구에서는 1995년도 초를 현재로 가정하여 그 이후 상장된 기업들의 수익률의 VaR를 추정하였다. 여기서 사용한 VaR는 1980년부터 1997년까지의 주가수익률을 토대로 산출한 VaR로서, 주가변동이나 추세 등을 나타내는 지표보다는 기간이나 증시 상황에 덜 민감하기 때문에 보다 더 각 기업의 위험을 잘 반영한다고 볼 수 있다. 사례기반추론에서 사용하는 사례베이스에 저장된 기업(기존 사례)은 1997년 발행된 상장회사정보에 등록된 회사들 중 1995년 이전에 상장된 기업들을 사례로 사용하였다(한빛증권, 1999). 1995년도 및 그 이후 상장된 기업들을 신규사례로 보고, 신규사례의 재무지표 또는 비재무지표 등을 입력변수로 사용하여 사례베이스에서 가장 유사한 속성을 가진 유사사례를 선택한 다음 유사사례의 수익률의 VaR를 신규사례의 수익률의 VaR로 추정하는 방법을 사용한다.

본 연구에서는 1997년 8월 현재 증권거래소에 상장되어있는 774개 상장사의 보통주 주식 데이터를 사용하였다. 신규상장기업은 상장된 지 1년 미만인 회사로서 1995년도에 상장된 기업을, 비상장기업은 그 이후 상장한 기업을 각각 의미한다. VaR의 측정에는 1980.1.1부터 1997.12.31까지의 주식거래 가격과 수익률을 사용하였다. 주식의 가격은 일 단위로 산정하였으며 수익률은 일 단위의 가격을 기준으로 다음과 같은 공식을 사용하여[7] 데이터베이스화 하였다.

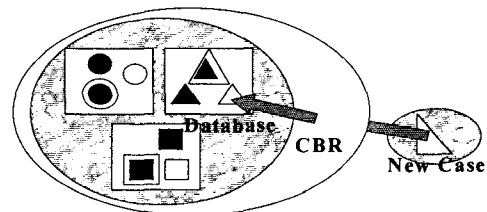
$$r_t = \frac{P_t + D_t - P_{t-1}}{P_{t-1}} \quad (2)$$

여기서 P_t 와 P_{t-1} 은 자산의 현재 및 지난 기의 시장가격, D_t 는 자산으로부터 나온 이자 또는 배당을 의미한다. 단, 여기서는 주식의 가격 위험을 측정하는 것이 주 목적이므로 이자 또는 배당 D_t 는 0으로 간주한다. 위의 식을 통해 나온 주가 수익률의 평균과 분산을 예시하면 <표 1>과 같다.

<표 1> 상장기업의 수익률
(1980.1.1~1997.12.31)

회사명	수익률 평균	수익률 분산
대우정밀	-0.60006444	0.0007305
대원제지	0.0004512	0.0023565
.....

CBR에 사용된 입력변수로는 여러 논문에서 연구된 재무변수 중 중요도가 높은 순으로 고정자산, 자산총계, 자본총계, 매출액, 매출총이익, 영업이익, 당기순이익 등 7가지를 선택하였다(박종역 1995, 이재식 1996, 차봉근 1997). CBR은 잘 알려진 방법론이므로 본 연구에서는 구체적인 설명을 생략하는 대신 참고문헌만을 제시하기로 한다(Aamodt 1996, Watson 1997). CBR을 통해 신규사례와 유사한 사례를 찾는 과정은 <그림 1>과 같다.



<그림 1> CBR을 통한 유사사례 선택과정

CBR을 통하여 특정 신규사례와 비슷한 사례를 검색하기 위한 방법으로 상장기업의 데이터와 신규사례간의 유사도 점수의 합계를 계산하여 신규사례와 비슷한 상장기업을 선택한다. 이를 위해서 첫째, 상장기업의 재무변수(사례의 속성)별로 유사도 값을 계산하는 공식을 이용하여 유사도 값을 계산하고, 각각의 재무변수별로 계산된 유사도 값의 범위에 따라 유사도 점수를 부여하며(<표 2>, <표 3> 참조), 이 유사도 점수에 일정한 가중치를 주어 유사도 점수의 합을 계산하였다(가중치 부여 방법은 제4장 참조).

<표 2> 유사도 점수 매트릭스의 예(재무변수)

재무변수	유사도 값 계산공식	유사도 값	유사도 점수
고정자산	유사도 값 = $\frac{\text{신규및비상장기업의재무변수의값}}{\text{상장기업의재무변수의값}}$	5~10	1
		1~5	5
		1	10
		1~0.5	5
		0.5~0.1	1
		기타	0

<표 3> 유사도 점수 매트릭스의 예(비재무 변수)

비재무변수	유사도 값 계산공식	유사도 값	유사도 점수
명목변수	동일하면 1, 다르면 0	1 (또는 0)	1 (또는 0)
날짜	동년이면 1, 다르면 0	1 (또는 0)	1 (또는 0)

이처럼 재무변수별로 계산된 유사도 점수에 가중치를 곱하여 각 상장기업의 총 유사도 점수를 계산한다. 그 후 총 유사도 점수가 제일 높은 상장기업을 신규사례와 제일 유사한 기업으로 판단한 후, 선택된 상장기업의 수익률의 VaR를 측정하여 신규 및 비상장기업의 수익률 VaR의 예측치로 삼는다. 본 논문에서는 주가 수익율이 정규분포를 이룬다고 가정하여 델타평가법(delta-normal

valuation method)으로 VaR를 측정한다. VaR를 구하는 공식은 다음과 같다(오세경, 1999). 단 현재 사용한 VaR의 공식은 복수의 포트폴리오를 가정한 것이 아니라 단일 포트폴리오를 대상으로 구한 VaR공식이다.

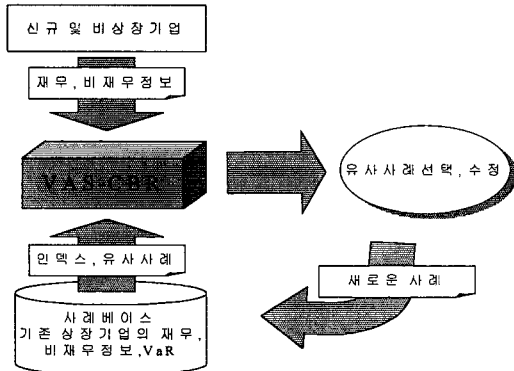
$$VaR = \text{평균} - 1.65 * \text{표준편차} \quad (3)$$

여기서 수익률의 VaR는 신뢰구간 95%범위를 벗어나는 최악의 경우에서 일어나는 수익률을 의미한다. 즉 100번 중 5번 정도 일어날 수 있는 수익률을 위험상황으로 본다. 예를 들어 두 회사의 추가수익률의 평균이 같더라도 VaR가 다르다면 여기서는 수익률의 VaR가 더 작은 회사의 주식이 보다 더 위험한 것으로 본다. 그리고 VaR에 투자할 액수를 곱한 것이 단순 포트폴리오 상에서 위험상황하에서의 추가수익률이라고 볼 수 있다.

4. VAS-CBR 시스템

본 연구에서 Visual Basic 5.0과 ACCESS 97을 이용하여 제 3장에서 설명한 사례기반추론 기법을 적용한 VAS-CBR(VaR Analysis System using Case Based Reasoning)라는 시스템을 구현하였다. VaR를 추정하는 VAS-CBR의 추론과정은 <그림 2>와 같다. 95년을 기준으로 95년도에 상장한 기업이나 96년 이후 상장한 기업을 신규 및 비상장기업으로 간주한다. 그리고 신규 및 비상장기업의 재무지표 및 비재무지표를 입력하면 VAS-CBR은 추론과정을 통해 입력사례와 유사한 상장기업을 사례베이스에서 검색하여 가장 비슷한 유사사례를 제공해준다. 신규 및 비상장기업이 상장된 후 VaR 계산이 가능해지면 그 사례

는 다시 사례베이스에 저장된다.

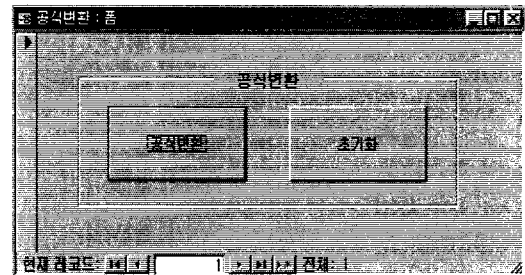


<그림 2> VAS-CBR의 추론과정

구현된 VAS-CBR은 4개의 모듈 구성되어 있다. <그림 3>은 원시데이터를 기초로 Feature Selection을 통해 새로운 가공데이터를 만드는 모듈이다. 본 연구에서는 사례기반추론의 입력데이터로 재무제표에서 사용되는 고정자산, 자산총계, 자본총계, 매출액, 매출총이익, 영업이익, 당기순이익 등을 사용하였으나, 이를 가공하여 통해 수익성, 안전성, 활동성, 성장성 등을 나타내는 지표로 변환시킬 경우 추론의 예측력이 더 높아질 것이라는 판단 하에 원시 및 가공데이터 양자를 다 사용하였다. 가공하기 전의 원시데이터 내용은 <표 4>와 같다. 참고로 <표 4>에서는 특정 사례의 유사도 점수의 합을 구하기 위한 가중치 부여방식을 보여주고 있다. 첫 번째 방식은 재무, 비재무변수 공히 동일한 가중치 10을 준 경우이다. 두 번째는 VaR와 각 재무변수를 다중 회귀분석하여 얻은 표준회귀계수(Beta)³⁾를 가중

3) 이는 회귀계수의 상대적 중요도를 나타낸다. 예를 들어, 독립변수들 간의 단위가 다르다면 회귀계수 그 자체는 큰 의미가 없다. 이 경우 표준화시켜 독립변수들의 상대적 중요도를 알아 볼 수 있는 데 그 지표가 바로 베타(Beta)계수이다.

치로 삼은 경우로서, 비재무변수의 경우 표준회귀계수를 구할 수 없으므로 일률적으로 가중치 1을 부여하였다.



<그림 3> Feature Selection 처리모듈 (예시)

원시데이터를 가공하여 생성한 수익성 지표로는 총자본 이익율, 자기자본 이익률, 매출액 총이익율, 매출액 경상이익율, 매출액 순이익율, 금융비용 부담율 등을 선정했고, 안정성 지표로는 유동비율, 부채비율 등을, 활동성 지표로는 총자본 회전율을 선택했다. 그리고 성장성 지표로는 매출액 증가율, 순이익 증가율, 총자본 증가율 등을 선정했다. 가공 후의 입력데이터 내용은 <표 5>에서 보는 바와 같다. 참고로 <표 5>에서는 특정 사례의 유사도 합을 구하는 데 있어서 <표 4>에서 고려한 두 가지 가중치 부여방식 이외에 재무변수에는 10, 비재무변수에는 1의 가중치를 부여한 경우를 하나 더 고려하고 있다. 이는 설립일, 상장일 등의 비재무변수는 수익률에 별반 영향을 미치지 않을 것이라는 판단에 따른 것이다.

<그림 4>는 신규사례와 기존 사례의 유사도를 측정하기 위해 비재무정보를 입력하는 모듈이다. 신규 및 비상장기업의 경우 비재무 지표로 산업분류, 1/2부, 설립일, 상장일, 계열 등을 입력하고 유사도 측정 단추를 누르면 입력데이터를 기준으로 사례베이스에 저장된 각각의 사례에 대하여

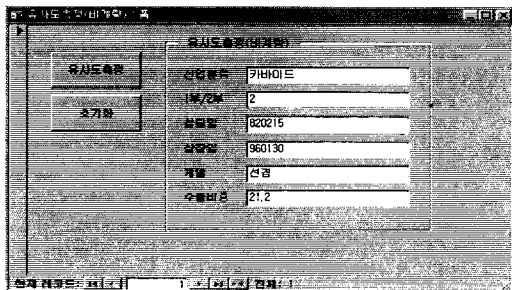
<표 4> 가공전 입력데이터 (예시)

입력변수	Data Type	가중치	가중치	유사도
산업 1부2부	명목변수 명목변수	10 10	1 1	같으면 1 다르면 0
설립일 상장일	날짜 날짜	10 10	1 1	년도가 같으면 1 다르면 0
계열	명목변수	10	1	계열사 유무가 동일하면 1 다르면 0
수출비중 고정자산94 고정자산95 자산총계94 자산총계95 자본총계94 자본총계95 매출액94 매출액95 매출총이익94 매출총이익95 영업이익94 영업이익95 당기순이익94 당기순이익95	계량변수 계량변수 계량변수 계량변수 계량변수 계량변수 계량변수 계량변수 계량변수 계량변수 계량변수 계량변수 계량변수 계량변수 계량변수 계량변수	10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10	0.062 0.303 0.085 0.245 0.025 0.231 0.166 0.055 0.067 0.363 0.009 0.104 0.504 0.033 0.234	동일하면 10, 0.5=< X =< 5이면 5, 0.1=< X =< 이면 1, 이외는 0
			상관계수(R2)*100	상관계수 사용시 설명력은 0.100

<표 5> 가공후 입력데이터 (예시)

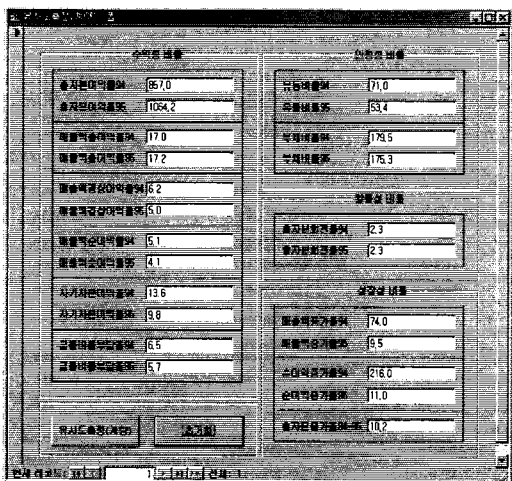
입력변수	입력변수형태	Data Type	가중치	가중치	가중치	유사도	
산업 1부2부		명목변수 명목변수	1 1	10 10	1 1	같으면 1 다르면 0	
설립일 상장일		날짜 날짜	1 1	10 10	1 1	년도가 같으면 1 다르면 0	
계열		명목변수	1	10	1	계열사 유무가 동일하면 1 다르면 0	
수출비중			10	10	0.081	동일하면 10, 0.5=< X =< 5이면 5, 0.1=< X =< 이면 1, 이외는 0	
총자본이익율94 총자본이익율95 매출액총이익율94 매출액총이익율95 매출액경상이익율94 매출액경상이익율95 매출액순이익율94 매출액순이익율95 자기자본이익율94 자기자본이익율95 금융비용부담율94 금융비용부담율95	수익성비율	계량변수	10 10 10 10 10 10 10 10 10 10 10 10	10 10 10 10 10 10 10 10 10 10 10 10	0.016 0.005 0.292 0.312 0.049 0.392 0.095 0.366 0.086 0.094 0.036 0.023		
유동비율94 유동비율95 부채비율94 부채비율95	안정성비율		10 10 10 10	10 10 10 10	0.041 0.005 0.066 0.149		
총자본회전율94 총자본회전율95	활동성비율		10 10	10 10	0.115 0.083		
매출액증가율94 매출액증가율95 순이익증가율94 순이익증가율95 총자본증가율95	성장성비율		10 10 10 10 10	10 10 10 10 10	0.036 0.006 0.008 0.03 0.158		
					상관계수(R2)*100		상관계수 사용시 설명력은 0.065

유사도 점수가 매겨지고 그 결과가 데이터베이스에 저장된다. 산업분류는 각 회사가 속해 있는 산업 군을 의미하며, 계열은 각 회사의 계열사 유무를 파악하여 계열사 유무가 동일하면 점수를 부여하고 다르면 0을 주게 된다. 설립일과 상장일의 경우 동년(同年)인 경우에는 점수를 부여하고 그렇지 않으면 0을 준다.



<그림 4> 비재무정보 처리 모듈 (예시)

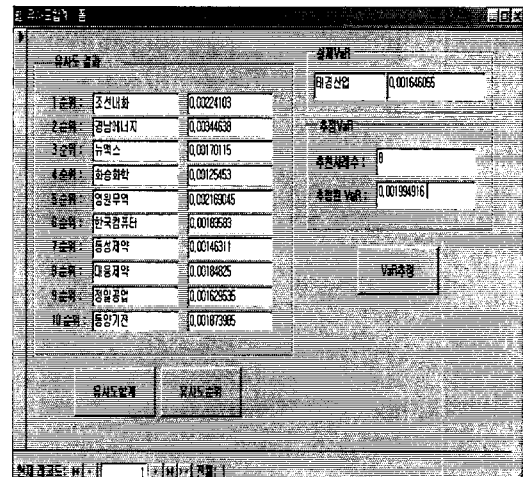
<그림 5>는 재무정보의 유사도를 측정하는 모듈이다. 재무정보의 유사도는 1994년과 1995년의 2년간의 자료를 토대로 측정하였다. 재무정보를



<그림 5> 재무정보 처리 모듈 (예시)

입력하고 유사도 측정 단추를 누르면 사례베이스에 유사도 점수가 매겨지고 점수가 데이터에 저장된다.

<그림 6>은 입력변수와 가장 유사한 사례를 1순위부터 10순위까지 추천해주는 모듈이다. <그림 6>의 화면에서 '유사도합계' 버튼을 누르면 전체 유사도 값이 계산되며 '유사도순위' 버튼을 누르면 1순위부터 10순위까지 VAS-CBR이 추천한 유사 사례를 화면에 표시해준다. 그 다음으로 추천사례수를 입력하고 VaR추정 버튼을 누르면 추천사례의 VaR의 평균이 표시된다.



<그림 6> 유사도 측정 모듈 (예시)

예를 들어 <그림 6>에서 신규사례로 입력된 회사는 태경산업이고 이와 가장 유사한 유사도 점수를 갖는 기존 회사는 조선내화로 나타나고 있다. 실제로 태경산업의 VaR 값인 0.00164605와 가장 가까운 VaR 값을 갖는 기존 회사는 정일공업이지만, 본 논문에서 채택한 유사도 점수 계산 방식으로 산정한 유사도는 조선내화가 가장 가깝다는 의미이다. 여기서 태경산업은 1995년 이후

상장한 회사로서 1995년을 기준으로 본다면 신규 상장기업에 해당하고, 1995년 이후의 실적치를 바탕으로 실제 VaR를 계산한 값이 화면 이상단에 표시되었다. 화면에 나오는 추천사례수는 VAS-CBR이 유사도 순으로 추천한 1순위부터 8순위까지의 회사(기준사례) 개수를 의미하며, 추정된 VaR는 1순위부터 8순위까지의 VaR의 평균을 구한 값이다. 여기서 추천사례수는 사용자가 화면에 입력할 수 있다. 이 예에서 보듯이 태경산업은 우선순위가 가장 높은 조선내화의 VaR보다는 여덟 사례의 VaR를 평균한 값에 더 가깝다는 것을 알 수 있으며, 이러한 이유로 CBR에서는 다수의 기준사례의 값을 평균하여 사용하는 것이 일반적이다(Aamodt, 1996).

5. VAS-CBR 시스템의 평가

본 장에서는 VAS-CBR 모형이 사용하는 원시 데이터의 가공 유무(속성 선택, Feature Selection) 및 유사도 점수에 부여하는 가중치의 변화(Feature Weighting)가 VAS-CBR의 예측력에 주는 영향을 다양한 실험을 통하여 파악함으로써 적절한 실험설계 수준을 정하였다. 또한 이러한 과정을 통하여 얻은 최적의 설계 수준에 기초하여 신규상장기업의 실제 VaR와 시스템이 추정하는 VaR가 신뢰할 만한 유사성을 보이고 있는가의 여부를 실험함으로써 VAS-CBR의 예측력을 평가하였다. 기존의 논문 중에는 사례기반추론을 이용해서 신규 및 비상장기업의 VaR를 추정한 연구가 부재하였기 때문에 방법론의 직접적인 비교는 불가능하다.

적절한 실험설계 수준을 정하기 위해 다음과 같은 실험요소별 수준을 사용하였다. 우선 입력

데이터 수준은 ① 제4장에서 기술한 바와 같은 7개의 주요 재무지표를 가공하지 않고 사용하는 경우('원시'), ② 이를 수익성, 안정성, 활동성, 성장성으로 나누어 가공한 총 12개의 속성변수를 사용한 경우('가공') 등의 두 수준으로 나누었다. 유사도 점수에 부여하는 가중치는 <표 5>에서 본 바와 같이 ① 재무, 비재무변수 공히 동일한 가중치 10을 준 경우('동일'), ② 재무변수에는 10, 비재무변수에는 1의 가중치를 부여한 경우('차별'), ③ 각 재무변수와 VaR의 베타계수를 가중치로 삼은 경우('상관계수') 등 세 수준을 고려하였다. 추천사례 수는 1개에서 10개까지로 실험하였다.

적정 실험설계 수준을 정하는 방법으로는 각 요인의 수준별로 추정 VaR와 실제 VaR의 MSE (Mean Square Error)를 다양한 실험을 통하여 계산하여, 그 중 가장 작은 MSE 값을 주는 수준을 최적 실험수준으로 선택하였다. MSE를 구하는 공식은 다음과 같으며, 100개의 신규사례 데이터를 사용하였다.

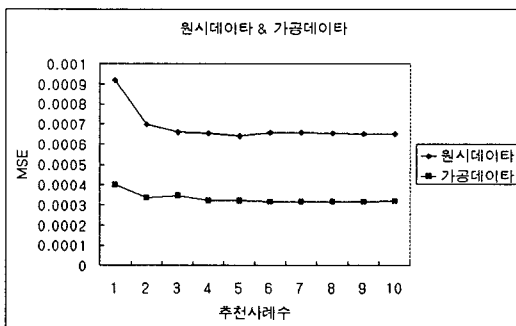
$$MSE = \frac{\sum(A - F)^2}{n} \quad (4)$$

- A : 입력된 기업(신규사례)의 실제 VaR
- F : 추천된 기준사례의 VaR 평균
- n : 실험에 사용한 신규사례 수 (100개)

5.1 Feature Selection 여부에 따른 MSE의 변화

일반적으로 사례기반추론에서는 새로운 모델을 만들고 데이터를 입력하기 전에 Feature Selection과 Feature Weighting을 통해서 데이터를 정제하거나 재구성하는데, 그 이유는 이러한 과정에 따라 예측률, 오차율이 많은 차이가 나타

나기 때문이다(Aarnodt, 1996). 본 연구에서는 일반 재무정보를 통해서 얻을 수 있는 원시데이터와 이를 가공해서 만든 데이터를 각기 적용하여 보다 낮은 오차율을 보이는 입력데이터 수준을 찾았다. <그림 7>은 VaR와 재무지표의 베타계수를 가중치로 택했을 경우의 MSE를 보여주고 있으며, 원시데이터보다는 가공데이터의 MSE가 훨씬 작게 나타나고 있음을 알 수 있다. 또한 가중치를 '동일' 혹은 '차별'로 부여하였을 경우에도 유사한 결과를 얻었다. 따라서 일반적으로 알려진 바와 같이 데이터의 가공, 정제과정이 모델의 예측력과 오차율에 영향을 미친다는 것을 알 수 있으며, VAS-CBR의 경우 가공데이터를 사용함이 바람직한 것으로 판명되었다.

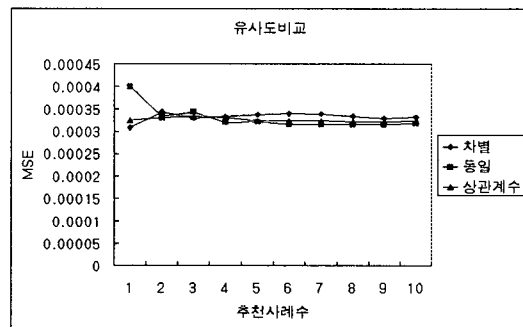


<그림 7> 원시데이터와 가공데이터의 MSE 비교

5.2 Feature Weighting의 변화에 따른 MSE의 변화

유사도 점수에 부여하는 가중치가 예측력에 주는 영향을 분석하기 위하여 <표 5>에서 본 바와 같이 ① 재무, 비재무변수 공히 동일한 가중치 10을 준 경우('동일'), ② 재무변수에는 10, 비재무변수에는 1의 가중치를 부여한 경우('차별'), ③ 각 재무변수와 VaR의 베타계수를 가중치로 삼은 경

우('상관계수') 등 세 수준을 대상으로 MSE를 계산하였으며, 입력데이터로는 가공데이터를 사용하였다. 그 결과 <그림 8>에서 보는 바와 같이 '상관계수'의 경우가 다소 작기는 하나 세 가중치 부여방법이 거의 동일한 수준의 MSE 값을 주는 것으로 판명되었다. 따라서 가중치 부여방법 보다는 Feature Selection을 통한 데이터 가공이 예측력 향상에 커다란 영향을 미친다는 사실을 알 수 있다. 또한 <그림 7> 및 <그림 8>에서 보는 바와 같이 추천사례 수는 2개 이상만 되어도 MSE에 별다른 영향을 주지 않는 것으로 파악되었다.



<그림 8> Feature Weighting의 변화에 따른 MSE의 변화

5.3 VAS-CBR의 예측력

이상의 실험을 통하여 얻은 최적의 설계 수준에 기초하여 신규상장기업의 실제 VaR와 시스템이 추정하는 VaR가 어느 정도의 유사성을 보이고 있는가를 실험함으로써 VAS-CBR의 예측력을 평가하였다. <표 6>은 1995년도 이후 상장한 기업 중 무작위로 고른 10개 기업의 실제 VaR와 각 신규기업별 8개 추천사례의 평균 VaR를 보여주고 있다. 여기서 추천사례들은 유사한 재무 및 비재무지표를 가지고 있고, 따라서 이들의 VaR는

<표 6> VAS-CBR의 예측오차

신규사례 (실제 VaR)	A	B	C	D*	E	F*	G	H	I*	J
	0.00165	0.00590	0.00155	0.05793	0.00000	-0.00052	0.00046	0.00786	-0.00201	0.00546
추천사례 1	0.00224	0.01278	0.00397	0.00206	0.00088	0.00239	0.00786	0.05793	0.00154	0.00096
추천사례 2	0.00345	0.00187	0.00088	0.00046	0.00206	0.00056	0.00067	0.00403	0.00259	-0.00201
추천사례 3	0.00170	0.00383	0.00247	0.00163	0.01547	0.00081	0.05793	0.00046	0.00229	0.00125
추천사례 4	0.00125	0.00317	0.00212	0.01181	0.00397	0.00156	0.00403	0.00067	0.00257	-0.00115
추천사례 5	0.00217	-0.00201	0.00376	0.00403	0.00187	0.00121	0.00139	0.00726	0.00158	0.05903
추천사례 6	0.00184	0.00125	0.00000	0.00786	-0.00201	0.00156	0.00098	0.00061	0.00152	0.00206
추천사례 7	0.00146	0.00314	0.00234	0.00187	0.00321	0.00219	0.00000	0.00300	0.00222	0.00383
추천사례 8	0.00185	0.00546	0.00195	0.00147	0.00348	0.00158	0.00253	0.00295	0.00145	0.01547
평균	0.00199	0.00369	0.00219	0.00390	0.00362	0.00148	0.00942	0.00961	0.00197	0.00993
표준편차	0.00067	0.00428	0.00133	0.00394	0.00515	0.00062	0.01976	0.01966	0.00050	0.02057
구간하한	0.00143	0.00011	0.00108	0.00060	-0.00069	0.00096	-0.00710	-0.00682	0.00156	-0.00727
구간상한	0.00256	0.00726	0.00330	0.00719	0.00792	0.00200	0.02595	0.02605	0.00238	0.02713

어떤 모평균을 중심으로 정규분포한다고 가정하여 이 모평균의 95% 신뢰구간을 구하였다. 그 결과 세 개의 신규기업(D, F, I)을 제외하고는 모두 모평균의 신뢰구간 안에 실제 VaR가 위치하여 있음을 볼 수 있으며, 예외적인 세 기업은 그 VaR의 추정치에 또 다른 변수의 도입이 필요하다는 사실을 암시하고 있다.

6. 결론

비상장기업이나 신규상장기업의 경우 주식거래량과 거래가격이 없거나 불충분하므로 최근 활발히 연구되고 있는 VaR(Value-at-Risk)를 파악하지 못한다. 본 연구에서는 비상장기업 및 신규상장기업의 미래 가격위험의 척도인 VaR를 추정하는 방법론을 제시하고, 이를 시스템(VAS-CBR)으로 구현하였다. 구체적으로는 사례기반추론(Case Based Reasoning: CBR) 기법을 이용하여 기존의 상장회사들 중에서 신규 및 비상장기업과 유사한

재무적, 비재무적 특성을 갖는 상장기업을 찾아내고 유사기업의 VaR를 근거로 신규 및 비상장기업의 VaR를 간접적으로 추정하였다. 또한 개발 시스템의 예측력 제고를 위한 운용방안 및 시스템의 예측력을 실험을 통하여 밝혔다.

CBR을 위한 입력데이터로는 일반 재무정보(원시데이터)보다는 수익성, 안정성, 활동성, 성장성 등을 나타내도록 가공한 재무정보가 VaR추정에 있어서 보다 효과적이며, 유사도에 부여하는 가중치로는 VaR와 각 재무변수를 다중회귀분석하여 얻은 표준회귀계수(Beta)를 사용하는 것이 가장 우수한 예측력을 주었다. 또한 추천사례 수는 2개 이상만 되면 예측력에 별다른 차이를 주지 않는 것으로 나타났다. 본 시스템의 유효성을 판단하기 위해 10개 신규상장기업을 대상으로 실제 VaR와 추정 VaR를 비교한 결과, 그 중 7개 기업의 경우 본 연구에서 제안한 방법에 따른 VaR 추정치가 유사기업의 VaR의 모평균 신뢰구간에 위치하는 바람직한 결과를 보였다. 나머지 3개 기업의 경우 실제 VaR와 추정한 VaR가 상당한 차이를 보인

바, 이는 VaR를 추정함에 있어서 본 연구에서 채택한 변수가 완전한 설명력을 갖지 못한다는 사실을 반증하는 것이며, CBR 기법이 그 예측력에 한계가 있음에 비추어 당연한 결과라 하겠다. 하지만 아직까지는 VaR 측정에 사례기반추론 등의 데이터마이닝 기법을 적용한 사례가 없다는 점에서 본 연구의 의의가 있다고 판단된다.

향후 연구방향으로는 본 연구에서 고려하지 않은 다양한 재무 및 비재무변수를 대상으로 예측력 제고를 위한 실험을 추가로 수행할 필요가 있으며, Feature Selection 시 여타의 변수를 생성하여 결과를 볼 필요가 있다. 또한 각 변수의 유사도에 부여하는 가중치를 다양하게 선정하여 실험할 필요도 있을 것이다. VaR 추정방법으로 사용한 델타평가법은 수익률이 정규분포를 가정하고 있는 바, 시장 수익률의 실제 분포를 사용하여 비모수적 접근을 하는 완전가치평가법(Full Valuation Method)을 사용하는 방안도 고려할 수 있다.

참고문헌

- [1] 박종역, 한인구, "인공신경망을 이용한 한국 종합주가지수 예측," *한국전문가시스템학회 추계 학술대회 논문집*(1995), 359-365.
- [2] *상장회사정보*, 한빛증권, 1999.
- [3] 양경식, 김현수. "판매 및 재무데이터에 관한 데이터 마이닝 기법 적용 연구," *한국전문가시스템학회 추계 학술대회 논문집* (1998), 107.
- [4] 안광열, "인공신경망을 이용한 신규공모주 수익률 예측," *아주대학교 경영정보학과 석사논문*, 2001.
- [5] 오세경, 김진호, 이건호, *위험관리론*, 경문사, 1999.
- [6] 이재식, 한재홍, "사례기반추론을 이용한 중소기업 도산 예측에 있어서의 비재무정보의 활용," *한국전문가시스템학회 추계 학술대회 논문집* (1996), 243-252.
- [7] 차봉근, "유전자 알고리즘을 이용한 인공신경망 구조설계에 관한 연구," *아주대학교 경영정보학과 석사논문*, 1997.
- [8] Aamodt, A. and E. Plaza, *Case-Based Reasoning: Foundational Issues, Methodological Variations, and system Approaches*, Artificial Intelligence Communications (1996), Vol.7, No.1.
- [9] Jorion, P., *Value at Risk: the New Benchmark for Controlling Market Risk* (1997), Chicago: Irwin Publishing.
- [10] Lopez, J. A. and Saidenberg M., "Evaluating Credit Risk Models," *Journal of Banking and Finance* (2000), 24, 151-165.
- [11] Pritsker, M., "Evaluating Value at Risk Methodologies: Accuracy versus Computational Time," (1997), *Journal of Financial Services Research*, 12, 201-242.
- [12] Watson, I., *Applying Case-based Reasoning: Techniques for Enterprise Systems* (1997), Morgan Kaufmann.

Abstract

Estimating VaR(Value-at-Risk) of non-listed and newly listed companies using Case Based Reasoning

Kyoungduck Choi*
Seungjong Noh**

Estimating the Value-at-Risk (VaR) of a non-listed or newly listed company in stock market is impossible due to lack of stock exchange data. This study employs Case-Based Reasoning (CBR) for estimating VaR's of those companies. CBR enables us to identify and select existing companies that have similar financial and non-financial characteristics to the unlisted target company. The VaR's of those selected companies can give estimates of VaR for the target company. We developed a system called VAS-CBR and showed how well the system estimates the VaR's of unlisted companies.

Key words: Case Based Reasoning, Value-at-Risk, Feature Selection

* Korea Information & Communication Co., Ltd.

** Dept. of Business Administration, Ajou University