주성분분석을 이용한 Markowitz 최적 포트폴리오의 연구 : 위험척도로서 표준편차, VaR, ES 이용

<table>
<tr><td colspan="2">신청자 성명</td><td>최 성 훈</td></tr>
<tr><td colspan="2">신청자 소속</td><td>부산대학교 경영학과 박사과정</td></tr>
<tr><td colspan="2">지도교수 성명(소속)</td><td>부산대학교 경영학과 엄철준 교수</td></tr>
<tr><td rowspan="3">신청자
연락처</td><td>주소</td><td>부산 금정구 장전2동 산30 부산대학교
상과대학 금융정보실 309-1호</td></tr>
<tr><td>전화/휴대폰</td><td>051-510-3599 / 019-574-2738</td></tr>
<tr><td>E-mail/팩스</td><td>choi_sunghun@naver.com</td></tr>
</table>

Ⅰ. 서 론

과거부터 자산의 수익률에 영향을 미치는 요인(factor)에 관한 연구는 많이 있었다. 개별 자산의 수익률을 시장(market factor)이라는 하나의 요인으로 관찰하는 단일시장모형[13], 균형상태에서 자산수익률에 영향을 미치는 여러가지 요인을 연구한 모형[7], 이자율의 움직임에 영향을 미치는 공통요인에 대한 연구[9] 등이 있었다. 이들은 자산의 수익률에 영향을 미칠 수 있는 많은 요인들을 모두 고려할 필요가 없이, 소수의 핵심적인 요인들만 관찰함으로써 자산수익률을 설명하고자 하였다.

이러한 연구는 주로 주성분분석(principal component analysis)과 요인분석(factor analysis)과 같은 다변량(multivariate) 통계분석방법을 사용하였다. 이 방법들은 다변량분석에서 차원축소(dimension reduction)에 관한 대표적인 기법들이다. 즉 두 가지 방법 모두, 관찰하고자 하는 자료(data)를 구성하는 요소들을 분석하고, 그 요소간의 관계를 바탕으로 연구자가 관찰하기 쉽도록 자료를 재구성하는 것이다. 이런 차원축소의 방법은 비단 finance 뿐만 아니라 자연과학이나 공학에서도 많이 쓰이는 방법이다.

복잡한 자료를 단순화하여 분석하고자 하는 방법은 여러 개의 기초자산으로 구성된 포트폴리오 관리에 도움이 될 것이다. 예를 들어 효율적 프론티어[10]를 사용하여 포트폴리오에 대한 최적의 투자를 분석할 수 있다. 효율적 프론티어를 도출하는 원리가 포트폴리오를 구성하는 개별자산간의 공분산행렬에 기초하여 지배원리상 최적의 투자를 만드는 기초자산들의 가중치를 계산해 내는 것이기 때문이다. 또한 포트폴리오 위험관리 측면에서도 전통적인 표준편차 대신 새로운 위험측정치를 계산해 낼 때에도 기초자산간의 관계를 단순화하여 분석할 수 있다면 유용한 위험관리 도구가 될 것이다. 본 논문에서는 이 같은 차원축소의 방법을 주식포트폴리오의 최적투자에 적용하고 이를 실증적으로 검증하고자 한다.

본 연구는 서론인 Ⅰ장을 포함해 5개 부분으로 구성되어 있다. Ⅱ장에서는 주성분분석과 고유치분해, 주성분분석을 하는 방법 중의 하나인 특이치분해, 그리고 고유치분해와 특이치분해의 관계에 대해서 설명한다. Ⅲ장에서는 특이치분해를 이용해 주성분분석을 하는 방법과 효율적 포트폴리오를 도출하기 위한 위험

측정치 등을 설명하고, Ⅳ장에서는 이에 대한 실증결과를 제시한다. 마지막으로 Ⅴ장에서는 실증결과가 제시하는 바를 간단히 언급하고 결론을 내린다.

Ⅱ. 연구방법

1. 주성분분석과 특이치분해

1) 주성분분석(principal component analysis)

다변량자료(multivariate data)를 분석할 때, 모든 변량(변수)을 고려하지 않고, 자료의 전체적인 움직임을 잘 설명해 줄 수 있는 소수의 요인들만 뽑아서 사용하면 분석이 간편해질 수 있다. 또한 다차원의 자료를 다루기 쉬운 저차원으로 바꾸게 되면, 계산량이 줄어들 뿐만 아니라, 원래 자료의 복잡성(complexity)때문에 발견할 수 없었던 핵심적인 관계를 알 수 있다. 주성분분석은 이처럼 해석하고자 하는 다차원(multi-dimension)의 데이터를 정보의 손실을 가능한 한 적게 해서 2차원 혹은 3차원 데이터로 축약하는 대표적인 차원축소(dimension reduction) 방법이다.[2]

주성분분석이나 요인분석과 같이 차원을 줄여서 데이터를 분석하고자 할 때 다음과 같은 고유치(eigenvalue)를 계산하게 된다.

$$AX = \lambda X \tag{1}$$

n개의 행과 n개의 열로 이루어진 어떤 정방행렬 (square matrix) A가 위와 같은 식을 만족하도록, 미지의 벡터 X와 상수 λ를 구하는 것을 ‘고유치분해(eigenvalue decomposition)’ 라고 한다.

행렬 A($n \times n$)를 고유치분해하면 n개의 고유치가 계산되고 각각의 고유치에 따라 식 (1)을 만족하는 고유벡터 (eigenvector) X_i (column vector, i=1～n)가 계산된다. 원래 자료인 행렬 A와 고유벡터(X_i)의 선형결합(linear combinations)은 합성변량(principal component)을 만들고, 이를 사용해서 데이터를 분석하고

그 의미를 부여하는 것이 주성분분석이다.

이러한 고유치분해를 하게 되면 고유치에 따른 각각의 고유벡터가 서로 직교(orthogonal)한다는 수학적인 성질 때문에 여러가지 응용분석이 가능하다. 예를 들어 원래 자료와 고유벡터가 곱해져서 만들어진 합성변량은 각각이 서로 독립적인 관계를 유지하고 있는 변수들이므로, 회귀분석에서 다중공선성(multicollinearity)을 해결하는데 사용될 수 있다. 또한 요인분석으로 확장시키면 차익거래이론[12] 처럼 개별주식의 수익률을 설명하는데 사용된다. 그리고 데이터를 저차원(lower dimension)으로 축소시킬 수 있기 때문에 본 논문의 경우와 같이 다양한 주식으로 구성된 포트폴리오의 위험을 분석하는 데에도 사용될 수 있다.

고차원인 자료를 축소를 하려면 전체 변량을 잘 설명할 수 있는 몇 개의 요인을 정해야 한다. 이러한 요인의 수를 정하는 방법으로 누적설명력 기준[1], scree 도표에 의한 방법[3], Kaiser rule[6] 등이 있다.

누적 설명력 기준은 고유치가 각 요인의 분산과 일치하고, 각 요인의 분산의 합이 전체 변량의 분산과 일치한다는 수학적 성질을 이용하는 것이다. 즉 몇 개의 고유치가 전체의 변동성(고유치의 총합)을 얼마만큼 설명하는지 보는 것인데, 예를 들어 n개 보다 작은 k개의 요인을 쓴다면 1~k개까지의 고유치의 합을 전체 고유치의 합으로 나누면 누적기여율이 계산된다. k개의 요인이 전체 변량의 누적기여율 만큼을 설명한다는 것이다. 일반적으로 누적기여율을 70~80% 이상을 많이 쓴다.[1]

scree 도표는 고유치를 내림차순으로 나열하고 그것이 감소하는 모습을 꺾은선 그래프로 보여준다. 종축은 고유치를 나타내고 횡축은 요인의 개수를 나타내는데, 요인의 개수를 정하는 방법은 그래프에서 고유치의 크기가 급격히 작아지는 점까지의 요인을 고려하는 것이다.

Kaiser rule에 따르면 특정요인의 고유치는 그 요인에 적재된 모든 변수의 요인적재값(factor loadings)을 제곱하여 합한 것이므로, 그 요인이 설명할 수 있는 총분산의 크기, 즉 설명력을 의미한다. 그래서 고유치가 '1' 미만인 요인은 한 개의 변수도 설명할 수 없다는 해석이 되므로, 1이상인 것들의 개수를 세어 그 만큼만 요인으로 사용하는 것이다. 이 기준을 사용해서 요인의 수를 결정하려면, 고유치 분해의 대상이 되는 행렬이 공분산행렬이 아닌 상관계수행렬이어야 한다.

2) 특이치분해 (singular value decomposition)

주성분분석을 앞서 말한 고유치분해를 통해서도 할 수 있지만, 특이치분해를 통해서도 할 수 있다.[11] 하지만 분해하고자 하는 행렬이 공분산행렬처럼 정방행렬이어야 하는 고유치분해에 비해, 특이치분해는 어떠한 직사각행렬에 대해서도 적용할 수 있기 때문에 보다 일반적으로 사용될 수 있고, 그것이 가진 여러 가지 수학적인 특성 때문에 유용하다. 특이치분해는 사회과학뿐만 아니라 이미지처리(image processing), 신호처리(signal processing), 생명과학(biotechnology) 같이 분야에서도 사용되고 있다.[8]

특이치(singular value)라는 것은 그 정의상으로는 어떠한 행렬을 역행렬이 존재하지 않게 만드는 숫자이다. 이러한 특이치를 구하는 과정이 곧 행렬을 세 부분으로 나누는 것으로 이어지는데, 나누어진 각각의 행렬이 저마다 중요한 성질을 띠게 된다. 예를 들어 어떤 직사각행렬 A ($T \times N$)는 특이치분해를 이용하면 다음과 같이 행렬을 분해(matrix factorization) 할 수 있다.

$$A = U \bullet S \bullet V^T \tag{2}$$

$$V^T: transpose\, of\, V$$

U는 좌특이벡터(left singular vector)로서 AA^T의 고유벡터이다. S는 특이치가 담긴 대각행렬(diagonal matrix), 그리고 V는 우특이벡터(right singular vector)로서 A^TA의 고유벡터이다. 이것을 계산하는 방법은 다음과 같다.

① A^TA의 고유치를 구하고 근호(root)를 씌우면 특이치가 계산된다.
그것을 내림차순으로 정렬하여 대각행렬(S)로 만든다.
② A^TA의 각 고유치에 따라 고유벡터를 계산해서 V를 계산한다.
③ 위 식의 정의에 따라 U는 $U=A \cdot V \cdot S^{-1}$에 따라서 계산한다.

u와 v를 각각 U와 V의 열벡터로 정의할 때, $u_i^T \cdot u_j=1$, $v_i^T \cdot v_j=1$ (i=j) 그리고 $u_i^T \cdot u_j=0$, $v_i^T \cdot v_j=0$ ($i \neq j$) 을 만족한다. 즉, 각각의 열벡터 원소의 제곱합은 1이

되도록 계산되고 서로 다른 열벡터와는 직교한다. 이러한 성질은 앞서 말한 고유벡터의 성질과 일치하는 것이다.

3) 고유치분해와 특이치분해의 관계

고유치분해와 특이치분해는 수학적으로 관계가 많다. 특히 분해하고자 하는 자료(행렬)가 공분산행렬처럼 대칭(symmetry)행렬일 경우에 고유치와 특이치는 일치한다. 또한 이 경우 고유치분해에 의한 고유벡터(X)와 특이치분해에 의한 좌특이벡터(U) 및 우특이벡터(V) 이 세 가지는 모두 일치한다. 이것을 수학적으로 나타나면 다음과 같다.

예를 들어 어떠한 대칭행렬 B를 다음과 같이 고유치분해하자. 이는 $BX = XD$ 가 되고, $B = XDX^{-1} = XDX^{T}\ (\because X^{-1} = X^{T})$ 로 바꿔 쓸 수 있다. 이 때 D는 고유치가 주대각선에 있는 대각행렬이고, X는 각 고유치에 상응하는 고유벡터이다.

한편 앞서 말한 순서에 따라 B를 특이치분해 하자. 먼저 B^TB의 고유치를 구하고 근호를 씌우면 특이치가 있는 행렬(S)이 계산된다. 그리고 고유치에 따른 고유벡터를 구하면 우특이벡터(V)가 나온다. 계산되지 않은 좌특이벡터(U)는 특이치분해 정의에 의해 계산된다. 그런데 B^TB는 다음과 같이 표현할 수 있다.

$$\begin{aligned} B^TB &= (XDX^T)^T XDX^T \\ &= (XD^TX^T)XDX^T \\ &= XD^TX^TXDX^T \\ &= XD^TDX^T \\ &= XFX^T \end{aligned} \qquad (3)$$

X와 D는 앞서 계산된 고유치분해의 결과이므로 $X^TX=I$ 를 만족하고, $F(=D^TD)$ 역시 대각행렬이다. 그러므로 B^TB는 식 (3)과 같이 고유치분해된다. 한편 특이치분해를 하는 순서에 따라 B^TB의 고유치를 구하기 위해 행렬 F에 근호를 씌우면 특이치가 담기는 행렬 S가 계산된다. B^TB의 고유벡터 X는 우특이벡터 V

가 된다. 나머지 좌특이벡터 U는 추가적인 계산으로 구해진다. B는 $B = U \cdot S \cdot V^T$ 가 아닌 $B = U \cdot D \cdot X^T$ 로서 표현될 수 있다. 한편 U를 계산하기 위해 특이치분해의 정의를 이용하여 U에 대하여 정리하면 $U = B \cdot X \cdot D^{-1} \ = (X \cdot D \cdot X^T) \cdot X \cdot D^{-1} \ = X$ 가 된다. 그러므로 B는 $B = X \cdot D \cdot X^T$ 로서 특이치분해 될 수 있다. 이는 행렬 B를 고유치분해한 결과와 일치한다. 그러므로 대칭행렬일 경우에는 고유치분해와 특이치분해가 같은 기능을 한다. 이는 finance에서 주로 사용되는 대칭행렬인 공분산행렬 및 상관계수행렬에 대해 활용할 수 있음을 보여주는 것이다.

한편 자료가 대칭행렬이 아니라도 자료를 특이치분해 해서 얻어진 특이치나, 원래 자료의 공분산행렬을 고유치분해 해서 얻어진 고유치는 서로 서로 간단한 계산으로 변형시킬 수 있다. 예를 들어 거래일 수가 행(row)에 있고, 기업들의 수익률이 각 열(column)에 있는 주식포트폴리오의 경우, 특이치제곱을 (거래일수 - 1)로 나눠주면 고유치가 계산되고, 그에 따른 고유벡터는 우특이벡터(V)와 일치한다.

이상에서 본 것과 같이 특이치와 고유치는 일대일 대응처럼 밀접한 관계에 있다. 따라서 분석결과 *k*개 고유치가 전체 자료를 설명하는데 필요한 적절한 요인의 수라면, *k*개의 특이치 역시 분석에 필요한 적절한 개수라고 말할 수 있을 것이다. 그러므로 본 논문의 분석 자료인 주식포트폴리오에 대해 특이치분해한 결과를 사용할 때, 고유치를 활용한 요인의 수 결정방식을 사용한다.

Ⅲ. 실증방법

1. 특정요인이 반영된 새로운 수익률 생성

주식포트폴리오는 개별주식의 수익률이 각각 하나의 열벡터를 구성하는 행렬로 볼 수 있다. 특이치분해를 사용해서 주식포트폴리오 행렬을 분해하면 포트폴리오의 전체적인 변동성(σ_p)을 잘 설명할 수 있는 몇 개의 주요한 요인(principal components)만 반영된 새로운 행렬을 만들어 낼 수 있다. 이 때 고려하고자 하는 요인의 수는 앞서 말했듯이 주성분분석이나 요인분석에서 주로 사

용하는 방식을 사용한다.

주식포트폴리오를 행렬 A ($T \times N$, T : 거래일 수, N : 개별종목 수)라고 하자. 행렬 A는 다음과 같이 특이치분해 될 수 있다.

$$A = U \bullet S \bullet V^T \qquad (2)$$

앞서도 언급했듯이 특이치분해를 실시하면 행렬 A ($T \times N$)는 U, S, V 세 부분으로 분해된다. U ($T \times T$)는 좌특이벡터(left singular vectors), S ($T \times N$)는 특이치가 주대각선에 있는 대각행렬, V ($N \times N$)는 우특이벡터(right singular vectors)로 분해된다. 행렬 U는 XX^T의 고유벡터이고, V는 X^TX 의 고유벡터이다. 행렬 S의 대각선에 있는 특이치는 행렬 X^TX의 고유치에 제곱근을 씌운 값이다.

특이치분해 결과 나온 행렬 S의 주대각선에 있는 특이치들은 고유치와 1:1 대응관계에 있으므로, 특이치들 중에서 몇 개를 고르는 것은 그 개수만큼의 전체 자료의 변동성을 설명하는 몇 개의 요인을 골라서 분석한다는 것과 같다. 고유치가 큰 요인이 전체변량(주식포트폴리오)의 변동성을 잘 설명한다. 마찬가지로 특이치도 제일 큰 몇 개(the first largest)의 값이 그러할 것이므로, 중요한 소수의 특이치만 남기고 나머지는 모두 0으로 만든다.

다시 말해서 주성분분석이나 요인분석에서 쓰이는 방법으로 분석에서 사용할 요인의 개수를 정한 뒤, 그 개수만큼만 대각행렬 S에 있는 특이치들을 남겨두고 나머지를 0으로 만들어서, 다시 식 (2)에 의해 행렬을 계산하면 소수의 주성분만 반영된 새로운 행렬(A^*)이 나오게 된다.[1)]

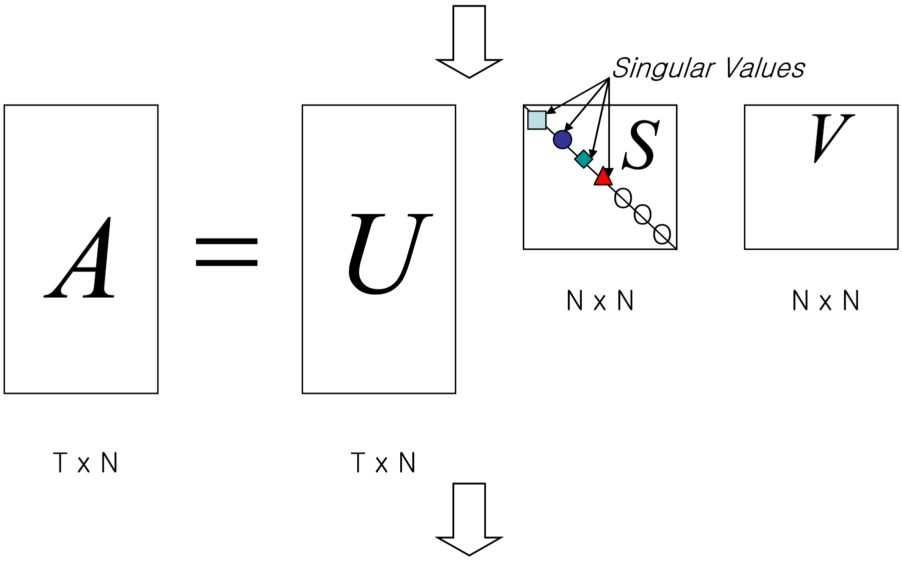
그림 1은 이 같은 과정을 설명하고 있다. S^*는 몇 개의 특이치만 선택하고 나머지 특이치는 0으로 만든 대각행렬이고, A^*는 이러한 방법으로 새롭게 만들어진 주식포트폴리오 행렬이다. 이런 방법으로 포트폴리오를 구성하는 개별주식수익률 데이터를 새로 만들어 내고, 이를 바탕으로 개별기업간의 관계를 보기 위해 공분산행렬을 계산한다

---

1) 이를 rank approximation to the original matrix 라고 부른다. 이는 주성분분석을 실시하는 방법 중의 하나이다.
www.mathworks.com/res/cleve "Professor SVD" column in October 2006

[그림 1] 특이치분해를 사용한 새로운 행렬 생성

$$A = U \cdot S \cdot V^T$$



$$A^* = U \cdot S^* \cdot V^T$$

주 : A ($T \times N$)의 수익률 행렬을 예로 들면, T는 관측치 수(거래일 수)를 나타내고, N은 기업의 수를 의미한다. $T \times N$행렬을 위의 그림과 같이 특이치분해를 하고, 대각행렬 S의 주대각선에 고려하고자 하는 요인의 개수(K개)만큼만 남겨두고 나머지는 모두 0으로 바꾼다. 이 과정을 거치면 대각행렬 S가 S^*로 바뀌고 소수의 요인만 반영된 새로운 A^*(수익률행렬)가 계산된다.

2. 위험측정치

전통적인 Markowitz(1952)에 방법에 따르면 포트폴리오의 위험을 분산(표준편차)을 통해서 관찰하고 있다. 하지만 여기서는 위험관리측면에서 활용도가 높은 VaR와 ES를 위험측정치로 사용하여 효율적 포트폴리오를 도출하고자 한다. VaR와 ES를 사용하는 이유는 기존의 많은 연구들이 자산수익률의 분포가 정규

분포를 따르지 않음을 밝힌 것과, 표준편차와 VaR가 가진 문제점을 극복하기 위해 ES와 같은 새로운 위험측정치를 시도한 것을 고려하기 위해서다.[4,5] 본 논문에서는 위험측정치로서 표준편차와 VaR, ES를 모두 사용한다. VaR와 ES는 수익률이 정규분포를 따른다고 가정하고 다음 식으로 각각 계산한다.

$$\begin{aligned} VaR_p &= -\mu_p + \sigma_p z_\alpha \\ &= -w_i\mu_i + \sqrt{w_i \Sigma w_i^T}\, z_\alpha \end{aligned} \tag{4}$$

$$\begin{aligned} ES &= \frac{1}{1-\alpha}\int_\alpha^1 VaR_p dp \\ &= -w_i\mu_i + \sqrt{w_i \Sigma w_i^T}\,\frac{\Phi(z_\alpha)}{1-\alpha} \end{aligned} \tag{5}$$

주 : μ_p는 포트폴리오의 기대수익률, σ_p는 포트폴리오의 표준편차, z_α는 신뢰수준 α에 해당하는 표준정규분포의 quantile을 의미한다. w_i는 개별자산에 대한 가중치행렬, μ_i는 개별자산의 기대수익률 행렬, Σ는 개별자산의 분산-공분산행렬을 의미한다. $\phi(z_\alpha)$는 수익률의 분포에서 z_α를 넘어서는 영역의 면적을 말한다.

3. Markowitz의 효율적 포트폴리오

주성분분석을 사용해서 포트폴리오의 투자를 최적화시키기 위해 효율적 포트폴리오를 도출하고자 한다. 이에 필요한 기대수익률과 공분산행렬은 앞에서 정의하고 계산되었다. 그리고 효율적 포트폴리오 도출에 쓰이는 위험의 기준 즉, 세 가지 위험측정치 (표준편차, VaR, ES)도 앞에서 정의했다. 앞으로는 Markowitz(1952)의 효율적 포트폴리오를 이용해서 투자를 최적화 시키고 이를 비교할 것이다.[10]

$$
\begin{aligned}
&Min\ \sigma(\omega)=\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}\omega_i\omega_j\sigma_{ij}}\\
&\qquad s.t\ \ E(R_p)=\sum\omega_i E(R_i)\\
&\qquad\qquad \sum_{i=1}^{N}\omega_i=1\\
&\qquad\qquad \omega_i\geq 0\quad \text{for } i=1,....,N
\end{aligned}
\tag{6}
$$

주 : w_i는 개별자산에 대한 가중치행렬, Σ는 개별자산의 분산-공분산행렬, $E(R_i)$는 개별주식의 기대수익률이다.

식 (6)은 Markowitz(1952)의 포트폴리오 최적화과정을 나타낸다. 최적화과정을 수행할 때 위험측정치별로 최소화하는 대상이 달라지므로 그에 따른 포트폴리오 최적화는 다음과 같이 각각 변형된다.

$$
Min\ \sigma(\omega)=\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}\omega_i\omega_j\sigma_{ij}} \tag{7}
$$

$$
Min\ \ VaR(\omega)=\sqrt{\omega\Sigma\omega^T}z_\alpha-\omega\mu \tag{8}
$$

$$
Min\ \ ES(\omega)=\sqrt{\omega\Sigma\omega^T}\frac{\phi(z_\alpha)}{1-\alpha}-\omega\mu \tag{9}
$$

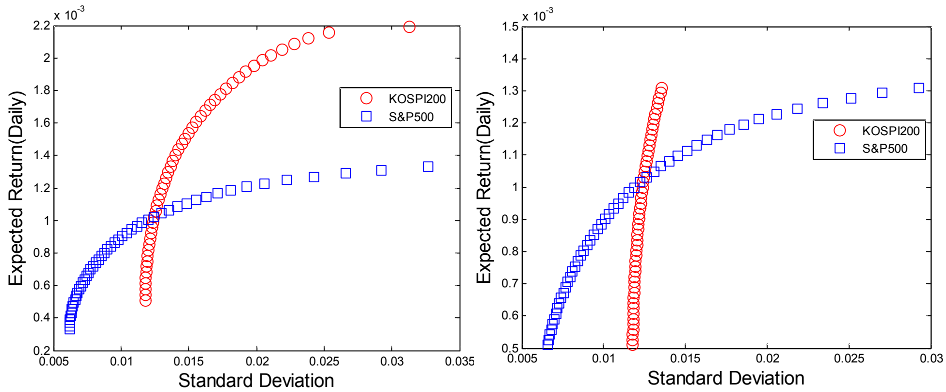
주 : μ_p는 포트폴리오의 기대수익률, σ_p는 포트폴리오의 표준편차, z_α는 신뢰수준 α에 해당하는 표준정규분포의 quantile을 의미한다. w_i는 개별자산에 대한 가중치행렬, μ_i는 개별자산의 기대수익률 행렬, Σ는 개별자산의 분산-공분산행렬을 의미한다. $\phi(z_\alpha)$는 수익률의 분포에서 z_α를 넘어서는 영역의 면적을 말한다.

식 (6)은 Mean-Variance 공간, 식 (7)은 Mean-VaR 공간, 식 (8)은 Mean-ES 공간에서의 효율적 포트폴리오를 찾기 위한 최적화 과정을 의미하며 각각의 제약조건은 Markowitz(1952)을 따른다.

한편 2개 이상의 효율적 포트폴리오를 그려서 비교할 경우에는 효율적 포트폴

리오의 기대수익률의 범위가 같도록 해야 한다. 이는 그림이 가질 수 있는 일종의 착시효과를 통제하고 분석결과 관찰을 용이하게 하기 위해서 이다. 그림 2는 KOSPI200과 S&P500을 구성하는 각각의 개별종목 164개와 343개의 자료를 가지고서 Markowitz(1952)에 의해 도출한 Mean-Variance 공간에서의 효율적 포트폴리오이다.

[그림 2] 기대수익률 조건이 다른 효율적 포트폴리오



왼쪽 그림을 보면 원(KOSPI200, red circle)으로 연결된 효율적 포트폴리오와 사각형(S&P500, blue square)으로 연결된 효율적 포트폴리오 사이에 뚜렷한 우위를 쉽게 판단 할 수 없다. 위험이 작은 구간에서는 S&P500의 효율적 포트폴리오가 같은 기대수익률 상에서 적은 위험을 갖고 있으므로 더 우월하고, 위험이 큰 구간에서는 KOSPI200의 효율적 포트폴리오가 같은 위험에서 더 높은 기대수익률을 보이고 있으므로 더 우월하다.

하지만, 오른쪽 그림을 보면 두 효율적 포트폴리오를 비교적 명료하게 비교할 수 있다. 지배원리 기준에서 평가한다고 할 때, 기대수익률 부분은 같으므로, 위험만 관찰하면 되는 것이다. 본 논문에서는 이처럼 효율적 포트폴리오를 비교할 경우에는 모두 기대수익률을 같게 만들어서 비교할 것이다.

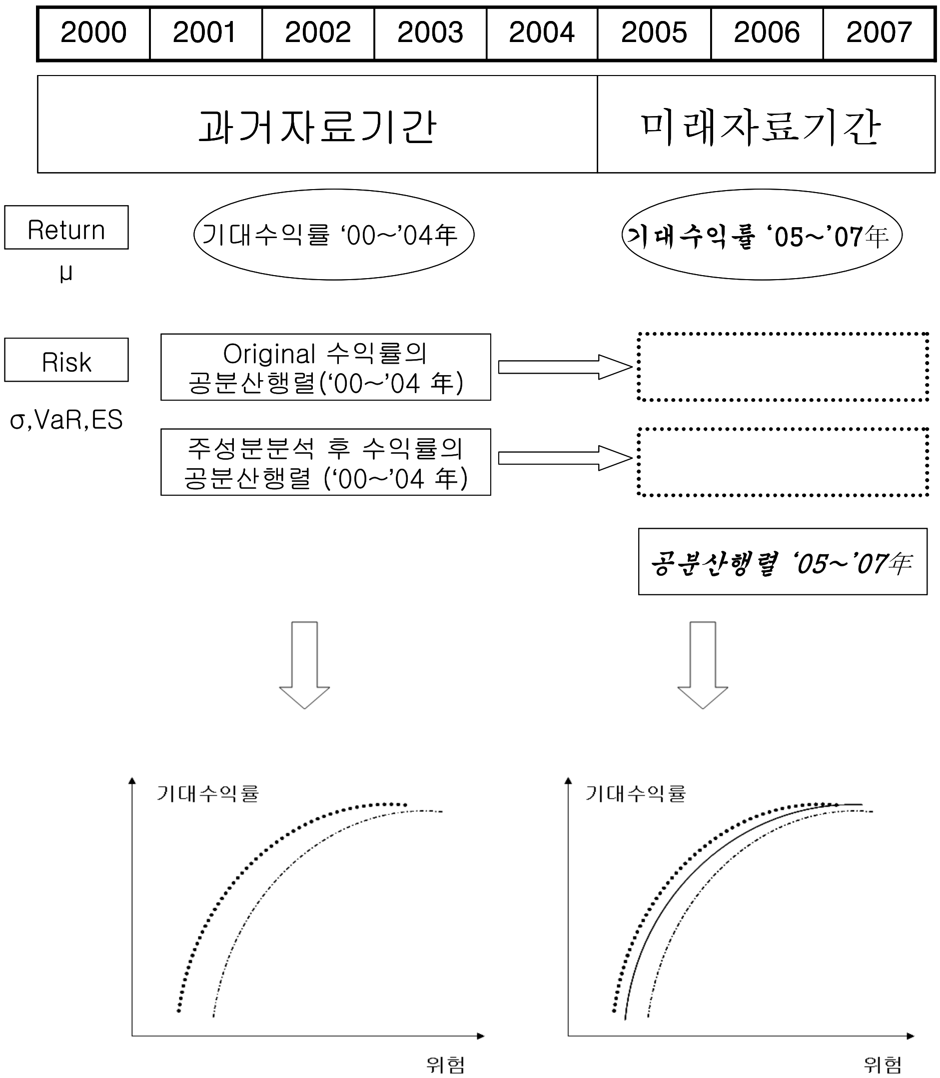
4. 데이터 및 기간 설정

실증분석에 사용된 자료는 KOSPI200의 164개 기업의 일별(daily) 종가이다. 자료의 기간은 2000년부터 2007년까지 총 8년간의 데이터를, 수익률은 로그 차분하여 사용하였다.

본 논문의 목적이 주성분분석으로 소수의 몇 개의 요인만 반영된 주식포트폴리오 수익률의 공분산행렬이 개별기업간의 관계를 더 안정적이고 정확하게 관측한다는 것이다. 그러므로 분석기간을 ① 공분산행렬을 계산하는데 사용되는 과거자료기간(2000년~2004년)과 ② 계산된 공분산행렬 테스트하는 미래자료기간(2005년~2007년)으로 나눈다. 과거자료기간에서는 3가지 위험측정치별로 2개의 효율적 포트폴리오를 유도한다. 하나는 원래 수익률행렬을 이용해서 계산한 공분산행렬을 이용한 것이고, 나머지 하나는 주성분분석을 사용해서 만든 수익률행렬의 공분산을 이용한 것이다.

한편 미래자료기간에서는 3가지 위험측정치별로 각각 3개의 효율적 포트폴리오를 유도한다. 첫번째는 과거의 원래 수익률행렬을 이용해서 계산한 공분산행렬을 이용한 것이고, 두번째는 과거자료를 주성분분석을 사용해서 만든 수익률행렬의 공분산을 이용한 것이다. 세번째는 미래기간에서의 원래 수익률자료의 공분산행렬을 이용한 것이다. 기대수익률은 해당기간의 주식포트폴리오에 아무런 조작을 가하지 않고 계산한 수치를 이용하였다. 그림 3은 실증분석의 구조에 대해서 설명하고 있다.

그림 3 기간의 설정 및 실증분석의 구조

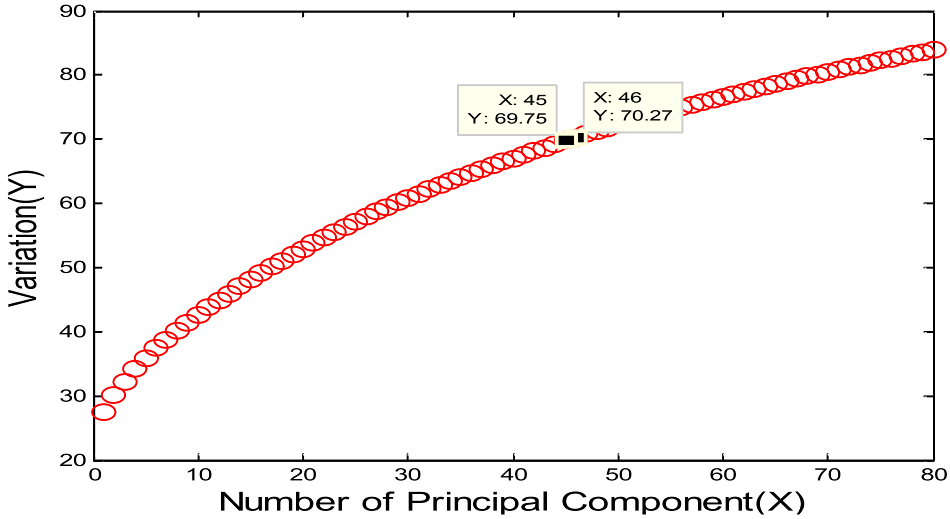


Ⅳ. 실증결과

1. 분석에 필요한 요인의 수 결정

Ⅱ장 1절에서 언급하였듯이 분석에 필요한 요인의 수를 결정하는 방법에는 3가지 방법이 있다. 각각의 방법으로 분석에 필요한 요인의 수를 결정하고, 효율적 프론티어를 유도하였다. 하지만 결과는 모두 유사하므로 여기서는 누적설명력 기준에 의한 결과만 제시하고, 나머지 기준에 의한 결과는 부록에 그림만 제시한다.[2)]

[그림 4] 누적설명력 기준에 의한 주성분 개수의 결정



주 : 고유치(X축)를 크기순으로 나열하고, 누적설명비율(Y축)이 70% 이상에 도달할 때의 고유치의 개수를 분석에 필요한 요인의 개수로 결정한다.

[자료] KOSPI200에 포함된 164개 기업의 일별 종가 수익률 (2000~2004년)

그림 4는 설명에 필요한 요인의 개수를 한 개씩 늘려감에 따라 그것에 의해 설명되는 주식포트폴리오 변동성의 비율을 나타낸 것이다. 대략 46개 정도의 요인을 고려하면 전체변량 변동성의 70%를 설명할 수 있는 것으로 파악할 수 있다. 그러므로 본 논문에서는 46개의 특이치만 사용하고 나머지 118개는 모두 0으로 만들어 버림으로써 새로운 주식포트폴리오 수익률 행렬을 만든다. 이렇게 계산된 자료에 따라 과거자료기간에서의 공분산행렬을 계산하고 이것을 미래자료기간의 효율적 포트폴리오 도출에 사용한다.

2) 누적설명력 기준으로는 46개, scree test에 의해서는 6개, Kaiser rule에 의해서는 33개의 주성분이 선택되었다.

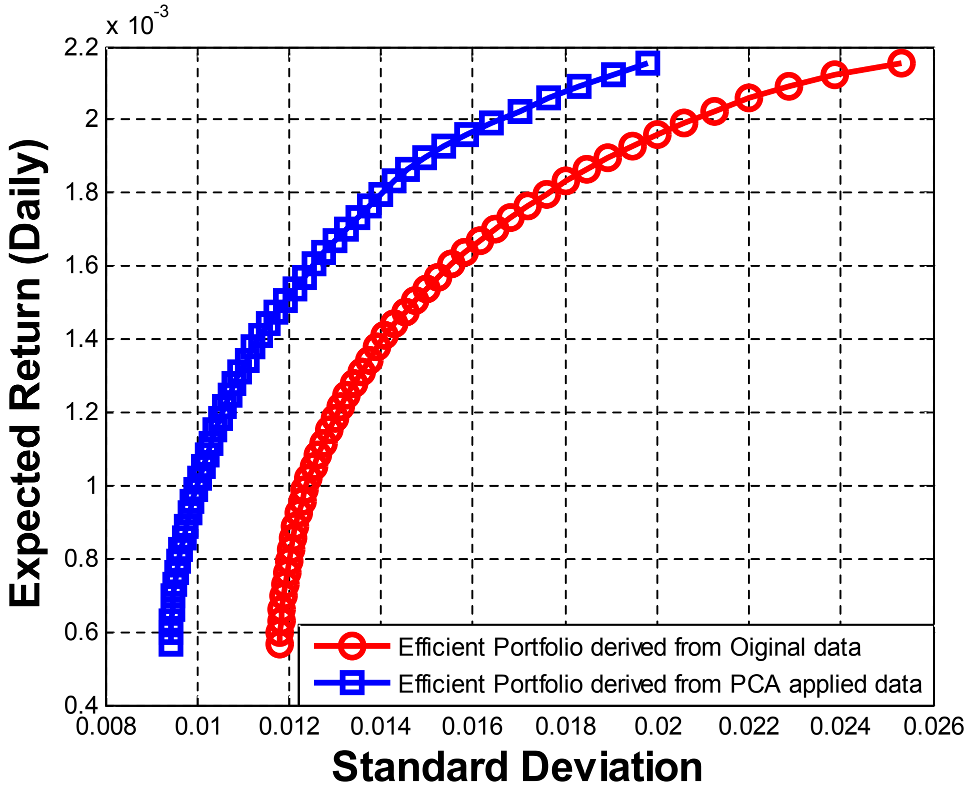
2. 과거자료기간 분석

그림 5, 그림 6, 그림 7에 각각에 효율적 포트폴리오가 2개씩 그려져 있다. 원(○)으로 그려진 것은 아무런 조작을 가하지 않은 원래의 수익률 데이터만 사용해서 도출한 효율적 포트폴리오이다. 그리고 사각형(□)으로 연결된 것은 주성분분석을 사용하여 포트폴리오의 대략 70% 정도의 변동성을 설명할 수 있을 만큼의 요인만을 골라서, 그 요인만 반영된 수익률로 포트폴리오를 구성했을 경우의 효율적 포트폴리오이다.

분석 결과 주성분분석을 거친 수익률 자료를 갖고서 그린 효율적 포트폴리오(□)가 원래의 수익률을 이용해서 그린 효율적 포트폴리오(○) 보다 지배원리상 우월하였다. 즉, 같은 기대수익률일 때 보다 낮은 위험(표준편차, VaR, ES)을, 같은 위험일 때 높은 기대수익률을 보여준다. 이는 나머지 118개의 특이치를 0으로 만들면서 그 요인이 설명하는 위험만큼을 줄인 것에 기인한다. 이러한 결과는 포트폴리오를 구성하는 개별주식간의 관계 분석에 주성분분석을 적용함으로써 기대수익률은 고정시킨 채 위험만 낮출 수 있음을 보여준 것이다.

한편, 횡축에 있는 위험측정치 값들을 비교해보면, 표준편차를 사용한 그림 5보다 VaR와 ES를 사용한 그림 7과 그림 8의 횡축 값이 더 큰 것을 볼 수 있다. 이는 VaR, ES가 포트폴리오의 표준편차만 고려한 것이 아니라 기대수익률까지를 고려한 위험측정치이므로 그 값이 표준편차 보다 큰 것이다. 식 (4)와 (5)에서 이를 알 수 있다. 다시 말해서 VaR, ES를 사용한다는 것이 포트폴리오의 위험을 증가시킨다는 것이 아니며, 해당 위험측정치의 정의에 따라서 수치가 크게 측정된 것이다.

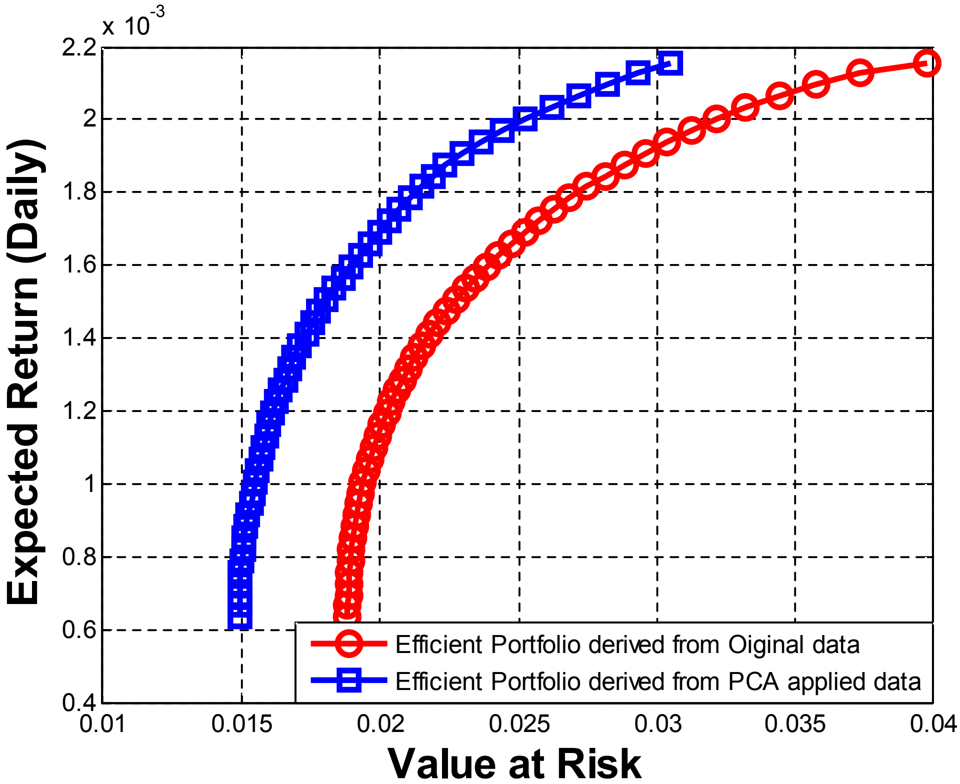
[그림 5] (과거자료기간) Mean-Variance 공간의 효율적 포트폴리오



주 : 주성분분석이 적용된 수익률 자료를 갖고서 그린 효율적 포트폴리오(□)가 원래의 수익률을 이용해서 그린 효율적 포트폴리오(○) 보다 지배원리상으로 더 우월하다.

[자료] KOSPI200에 포함된 164개 기업의 일별 종가 수익률 (2000～2004년)

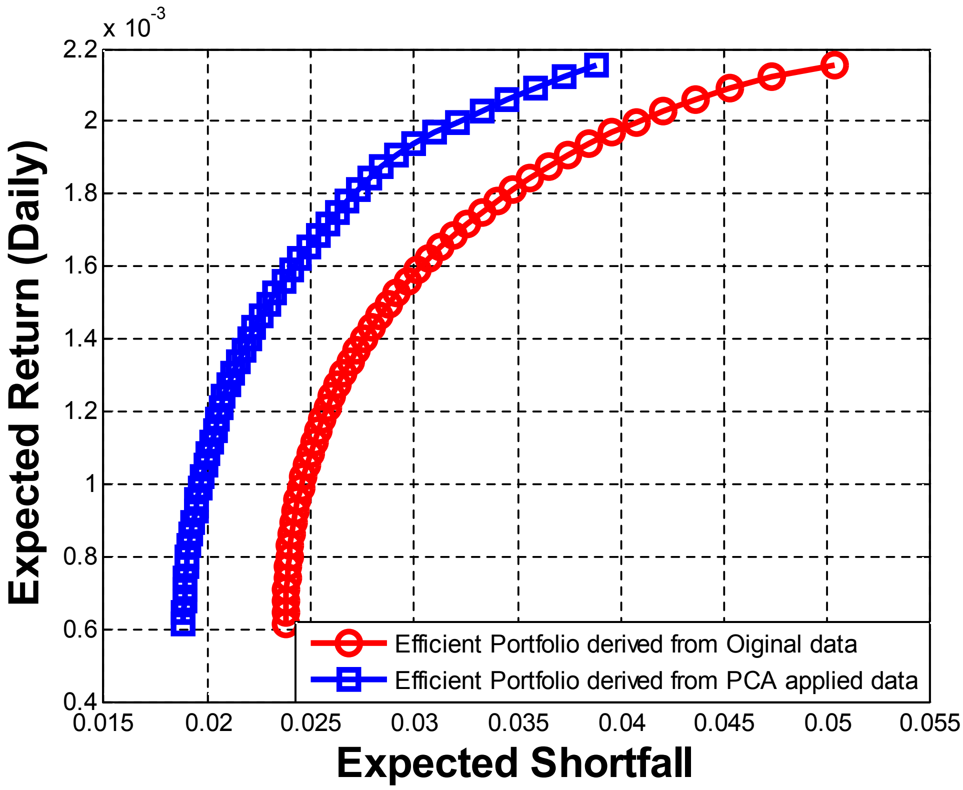
[그림 6]> (과거자료기간) Mean-VaR 공간의 효율적 포트폴리오



주 : 주성분분석이 적용된 수익률 자료를 갖고서 그린 효율적 포트폴리오(□)가 원래의 수익률을 이용해서 그린 효율적 포트폴리오(○) 보다 지배원리상으로 더 우월하다.

[자료] KOSPI200에 포함된 164개 기업의 일별 종가 수익률 (2000～2004년)

그림 7 (과거자료기간) Mean-ES 공간의 효율적 포트폴리오



주 : 주성분분석이 적용된 수익률 자료를 갖고서 그린 효율적 포트폴리오(□)가 원래의 수익률을 이용해서 그린 효율적 포트폴리오(○) 보다 지배원리상으로 더 우월하다.

[자료] KOSPI200에 포함된 164개 기업의 일별 종가 수익률 (2000~2004년)

3. 미래자료기간 분석

미래자료기간(2005년~2007년)에서는 일종의 테스트기간으로서, 과거자료기간(2000년~2004년)에서 포착된 개별기업간의 관계(공분산행렬)가 어느 정도 안정성, 신뢰성을 갖는가를 알아본다. 이를 검증하기 위해 과거자료기간에서 계산한 공분산행렬(Σ)을 이용하고 미래자료기간에서의 주어진 자료로 기대수익률을 계산하여 효율적 포트폴리오를 도출한다.

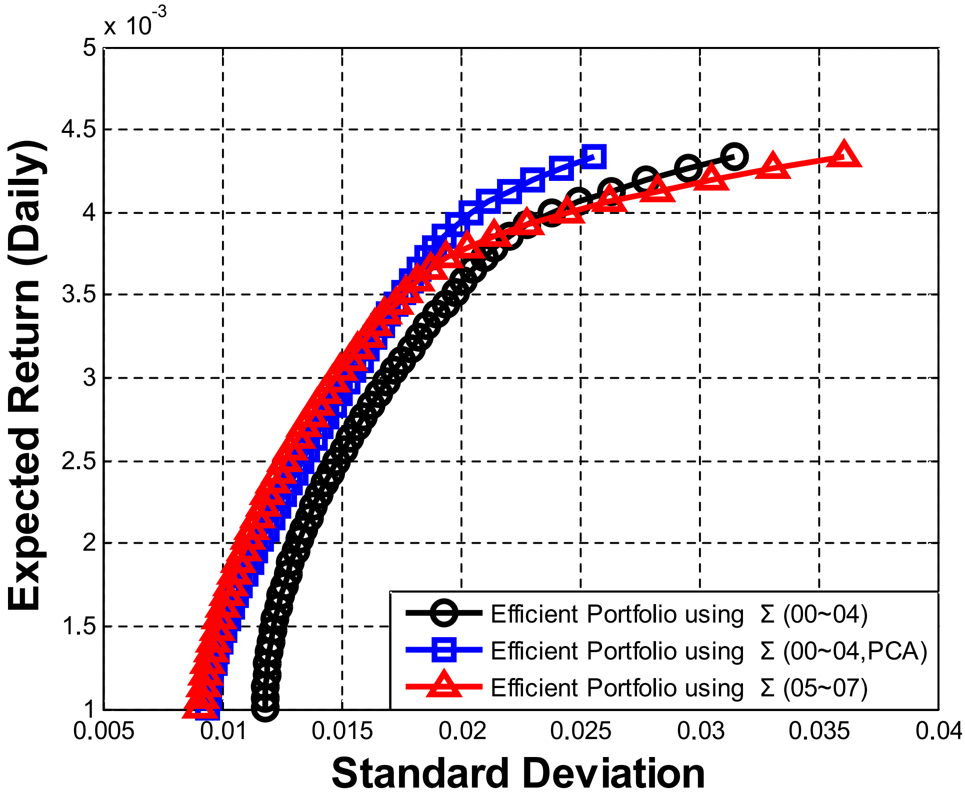
미래자료기간에서 유도하는 3개의 효율적 포트폴리오는 모두 2005년에서 2007년까지의 실제자료의 기대수익률을 사용한다. 하지만 공분산행렬은 다음의 3가지 공분산행렬을 사용한다.

① 과거자료기간에서 주성분분석을 적용하지 않은 수익률로 계산한 공분산행렬
② 과거자료기간에서 주성분분석을 적용한 수익률로 계산한 공분산행렬
③ 미래자료기간에서 주성분분석을 적용하지 않은 수익률로 계산한 공분산행렬

①, ②번 효율적 포트폴리오 중에서 어느 것이 ③번의 효율적 포트폴리오에 근접한지 살펴볼 것이다.

그림 8, 그림 9, 그림 10 에서는 과거자료기간에서의 분석한 방법과 마찬가지로 위험측정치별(표준편차, VaR, ES)로 효율적 포트폴리오를 나타내고 있다. 과거자료기간의 원래 수익률을 바탕으로 한 공분산행렬(Σ)을 미래자료기간에 적용시켰을 경우에는 개별주식간의 관계가 빠짐없이 반영된 것이기 때문에 시간이 변함에 따라 지배원리상 열등한 효율적 포트폴리오(○)를 도출하였다. 하지만, 주성분분석을 사용해서 그린 효율적 포트폴리오(□)는 개별 종목간의 관계를 정확히 관찰했기 때문에 미래자료기간에서 실제 효율적 포트폴리오(△)에 더 근접한 것을 알 수 있다. 결론적으로 주성분분석으로 만들어진 공분산행렬(Σ)의 안정성을 확인한 것이다.

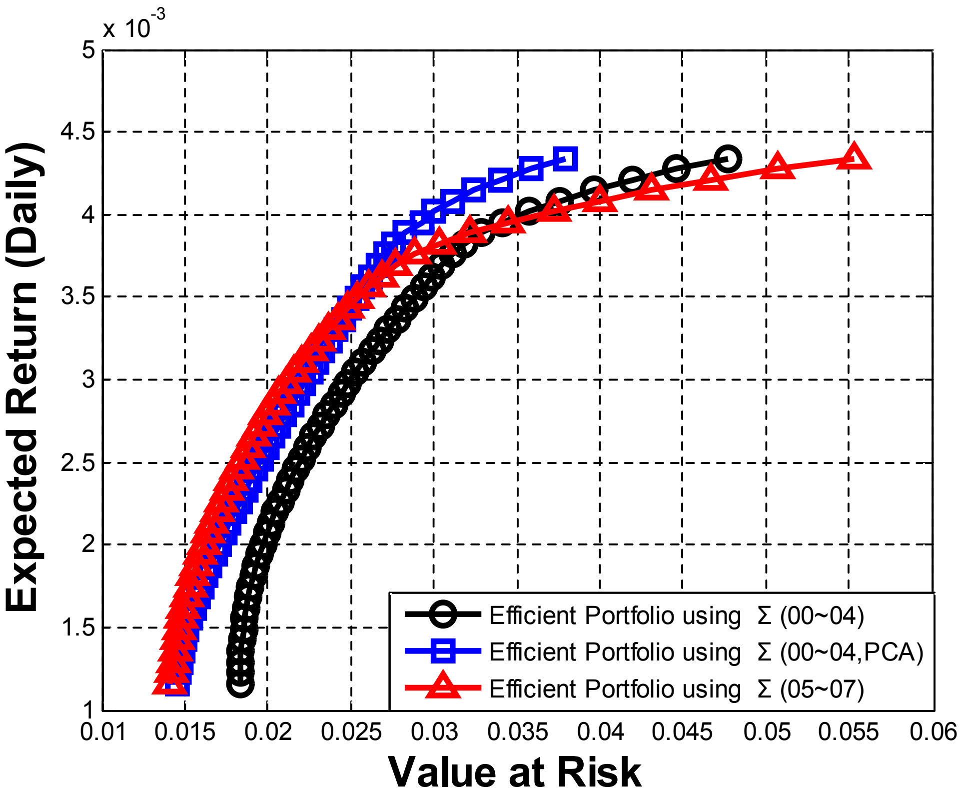
[그림 8] (미래자료기간) Mean-Variance 공간의 효율적 포트폴리오



주 : 과거자료기간에서 주성분분석이 적용되지 않은 수익률을 바탕으로 만든 공분산행렬(Σ, 00~04)의 효율적 포트폴리오(○)와 주성분분석이 적용된 수익률을 바탕으로 한 공분산행렬(Σ, 00~04, PCA)로 만든 효율적 포트폴리오(□) 중에서, 실제 기간의 수익률을 바탕으로 만든 공분산행렬(Σ, 05~07)의 효율적 포트폴리오(△)와 가까운 것은 주성분분석이 적용된 효율적 포트폴리오(□)이다. 한편 3개의 효율적 포트폴리오를 도출하는데 사용된 기대수익률은 모두 미래자료기간(05~07년)의 기대수익률이다.

[자료] KOSPI200에 포함된 164개 기업의 일별 종가 수익률 (00~04년,05~07년)

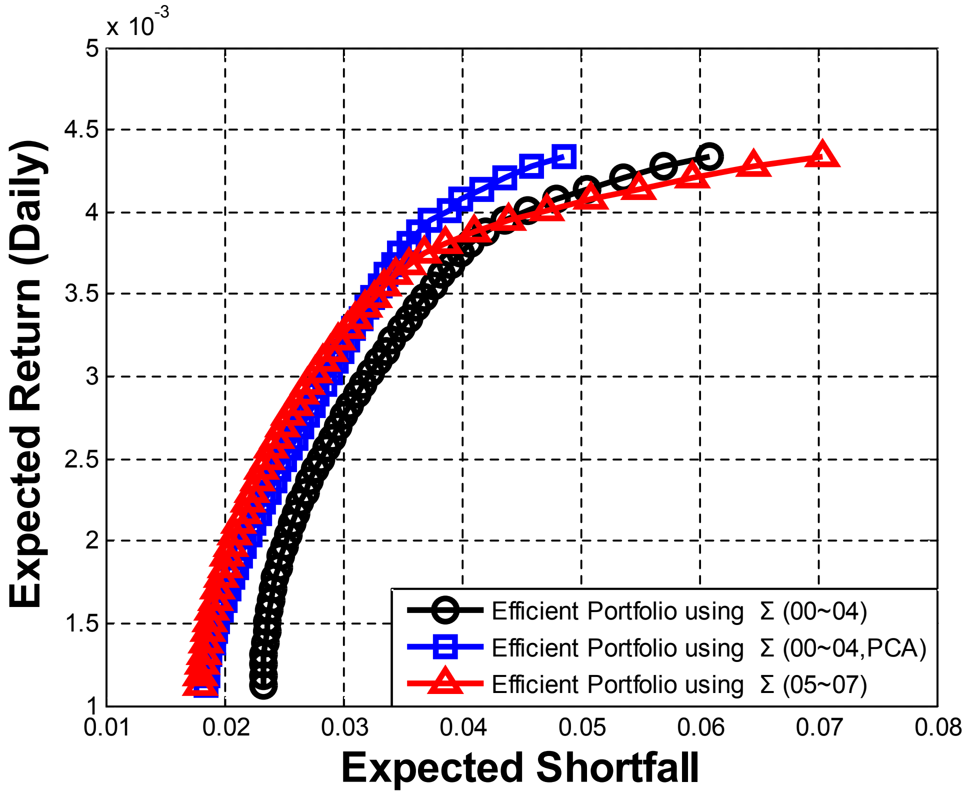
[그림 10] (미래자료기간) Mean-VaR 공간의 효율적 포트폴리오



주 : 과거자료기간에서 주성분분석이 적용되지 않은 수익률을 바탕으로 만든 공분산행렬(Σ, 00～04)의 효율적 포트폴리오(○)와 주성분분석이 적용된 수익률을 바탕으로 한 공분산행렬(Σ, 00～04, PCA)로 만든 효율적 포트폴리오(□) 중에서, 실제 기간의 수익률을 바탕으로 만든 공분산행렬(Σ, 05～07)의 효율적 포트폴리오(△)와 가까운 것은 주성분분석이 적용된 효율적 포트폴리오(□)이다. 한편 3개의 효율적 포트폴리오를 도출하는데 사용된 기대수익률은 모두 미래자료기간(05～07년)의 기대수익률이다.

[자료] KOSPI200에 포함된 164개 기업의 일별 종가 수익률 (00～04년,05～07년)

<그림 11> (미래자료기간) Mean-ES 공간의 효율적 포트폴리오



주 : 과거자료기간에서 주성분분석이 적용되지 않은 수익률을 바탕으로 만든 공분산행렬(Σ, 00～04)의 효율적 포트폴리오(○)와 주성분분석이 적용된 수익률을 바탕으로 한 공분산행렬(Σ, 00～04, PCA)로 만든 효율적 포트폴리오(□) 중에서, 실제 기간의 수익률을 바탕으로 만든 공분산행렬(Σ, 05～07)의 효율적 포트폴리오(△)와 가까운 것은 주성분분석이 적용된 효율적 포트폴리오(□)이다. 한편 3개의 효율적 포트폴리오를 도출하는데 사용된 기대수익률은 모두 미래자료기간(05～07년)의 기대수익률이다.

[자료] KOSPI200에 포함된 164개 기업의 일별 종가 수익률 (00～04년,05～07년)

Ⅴ. 결 론

본 논문에서는 주식포트폴리오 위험관리에 대한 주성분분석의 유용성을 효율적 포트포리오로써 검증하였다.

먼저, 우리나라의 주식시장에 상장된 기업의 주가자료를 이용하여 포트폴리오를 구성하고, 개별종목들 사이의 관계를 알기 위해, 주성분분석의 방법 중의 한 가지인 특이치분해로 몇 개의 요인만 반영된 새로운 주식포트폴리오 수익률을 계산하였다. 효율적 포트폴리오를 도출하는데 필요한 기대수익률 행렬과 공분산행렬 중에서 주성분분석이 조정하는 것은 공분산행렬이므로 기대수익률 행렬은 각 시기의 실제 기대수익률을 사용하며 공분산행렬은 분석의 목적에 따라 바꿔가면서 분석하였다.

표준편차뿐만 아니라 포트폴리오 VaR와 ES를 계산함으로써, Mean-Variance 공간뿐만이 아니라 Mean-VaR와 Mean-ES공간에서 효율적 포트폴리오를 유도하였다. 이는 표준편차 대신 VaR와 ES를 위험측정치로 사용함으로써 포트폴리오의 위험에 대한 노출 정도를 직관적으로 알 수 있는 실무적인 유용성 때문이다. 자료의 기간은 개별 종목간의 관계를 담고 있는 과거의 공분산행렬이 미래에도 적용될 수 있는지 알기 위해 과거자료기간과 미래자료기간으로 나누어 설정하였다.

과거자료기간에서는 주성분분석이 적용된 수익률로 계산한 공분산행렬과 그렇지 않은 공분산행렬, 그리고 아무런 조작을 거치지 않은 기대수익률로서 효율적 프론티어를 위험측정치별로 도출하였다. 분석결과 주성분분석이 적용된 수익률을 사용해서 도출된 효율적 포트폴리오가 원래 수익률 자료를 바탕으로 한 효율적 포트폴리오보다 우월함을 보여주었다. 이는 주성분분석이 포트폴리오의 변동성을 설명하는 요인의 핵심만 반영하고 나머지 부분은 버림으로써 전체적으로 포트폴리오의 위험을 줄여주었기 때문이다. 위험이 줄어든 수익률 자료를 갖고서 공분산행렬을 계산하고 효율적 포트폴리오를 도출했기 때문에 지배원리상에서 보아 같은 기대수익률에 보다 낮은 위험을 보여주었다.

미래자료기간에서는 미래자료기간의 자료로 계산한 기대수익률과 과거자료기간의 공분산행렬(주성분분석 적용된 수익률, 원래 수익률), 그리고 미래자료기간의 실제 자료로 계산한 기대수익률을 사용하여 위험측정치별로 효율적 포트폴

리오를 도출하였다. 미래기간의 실제 효율적 포트폴리오에 근접한 것은 주성분분석이 적용된 수익률로 만들어진 공분산 행렬이었다. 주성분분석 결과로 만들어진 수익률이 개별기업간의 관계를 보다 안정적으로 표현하기 때문에 시간이 흘러 시장상황이 변했어도 포트폴리오의 변동성을 알 수 있는 핵심적인 부분은 공분산행렬이 담고 있는 것이다. 이 같은 결과는 효율적 포트폴리오를 도출하는데 필요한 위험측정치 표준편차, VaR, ES 에 관계없이 동일하였다.

결론적으로 주성분분석을 활용하여 기대수익률을 유지한 채 위험만 낮출 수 있다는 것을 알 수 있었고, 포트폴리오에 편입된 종목들 간의 관계를 안정적으로 파악할 수 있었다. 이를 활용한 보다 심화적인 방법이 포트폴리오의 위험관리와 최적화에 적용될 수 있을 것으로 예상한다.

참고문헌

[1] 김기영 · 전명식,「SAS 주성분분석」, 고려대학교 통계연구소 통계분석강의 총서7, 自由아카데미, 1997

[2] 노형진, 「SPSS에 의한 다변량 데이터의 통계분석」, 효산, 2007

[3] Cattell R. B., “The scree test for the number of factors”. *Multiv. Behav. Res.*, 1, 1966, pp.245-276

[4] Dowd K., *Measuring Market Risk*, Second ed., John Wiley & Sons, Ltd., 2005

[5] Jorion P., *VALUE AT RISK*, Third ed., Mc Graw Hill, 2007

[6] Kaiser H. F., "The application of electronic computers to factors analysis." Educ. Psychol. Meas., 20, 1960, pp.141-151

[7] King B. F., "Market and Industry Factors in Stock Price Behavior", *Journal of Business*, 1996, pp.139-190

[8] Leon S. J., *“Linear Algebra with application"*, Sixth ed., Prentice Hall, 2002

[9] Litterman, R., Scheinkman, J., "Common Factors Affecting Bond Returns", Journal of Fixed Income, June 1991, pp 54-61

[10] Markowitz H., "Portfolio Selection," *The Journal of Finance* 7(1), 1952, pp.77-91

[11] Martinez W. L., Martinez A. R., *"Computational Statistics Handbook with MATLAB“*, Second ed., 2002, Chapman & Hall/CRC

[12] Ross S. A., "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* 13 (December 1976), pp 341-360

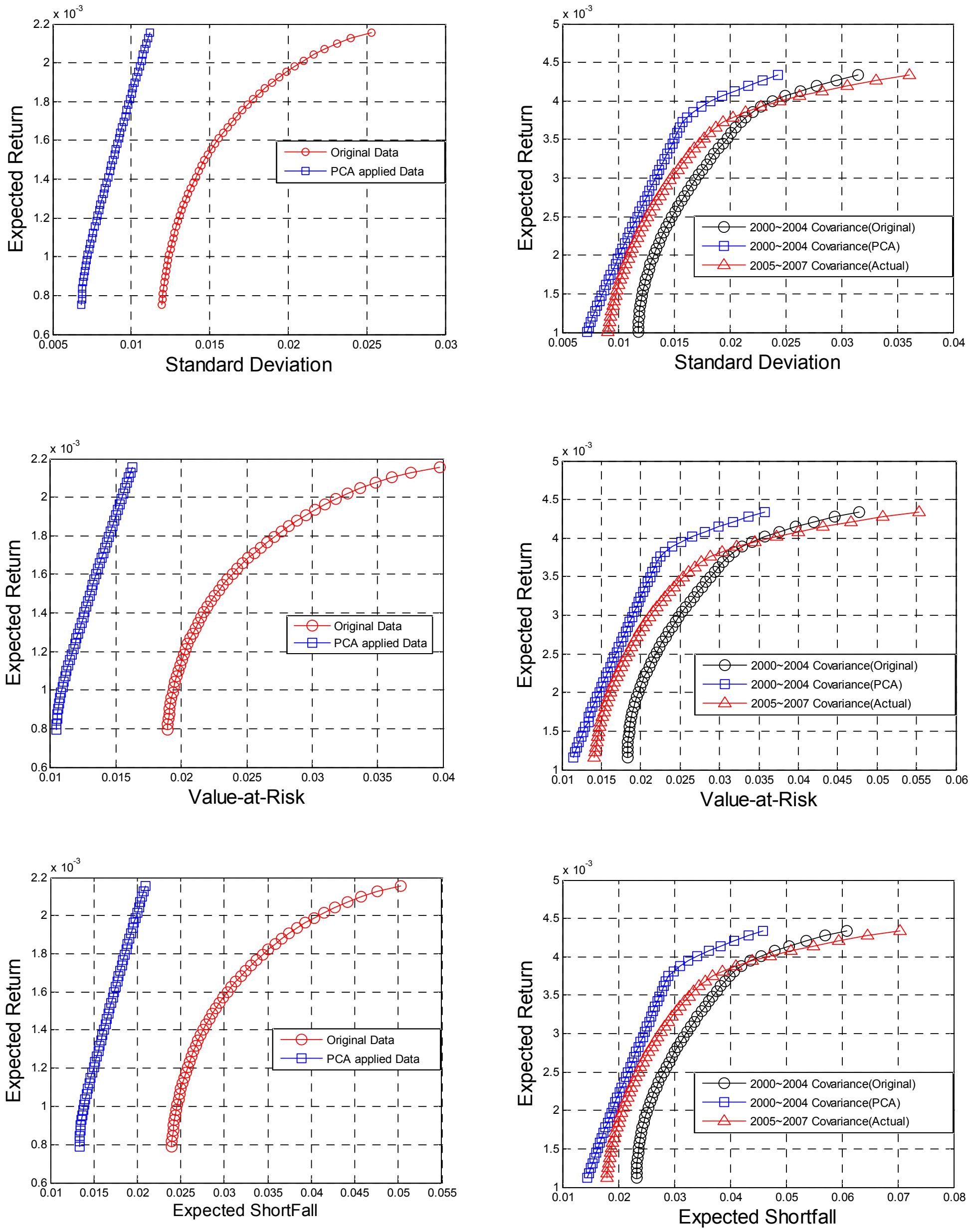
[13] Sharpe, W.F., 1963. A simplified model for portfolio analysis. *Management Science* (January), pp 277 - .293.

[14] Singh M. K., "Value at Risk Using Principal Components Analysis", *Journal of Portfolio Management,* 1997, pp.101-112

[15] Trzpiot G. and Ganczarek A., "Value at Risk Using Principal

Components Analysis on the Polish Power Exchange" ,*From Data and Information Analysis To Knowledge Organization*, 2005, pp.550-557

Scree 도표를 이용한 분석 결과



Kaiser rule에 의한 분석결과