

자산 수익률에 대한 경험적 확률분포를 이용한 포트폴리오 운용 방안

빈기범*

〈요 약〉

자산수익률을 수리적, 통계적으로 모형화하기 위해 정규분포가 종종 사용된다. 각종 확률분포 중 정규분포가 가장 친숙한 탓이다. 하지만, 두터운 꼬리로 잘 알려져 있듯, 정규분포는 자산수익률 또는 증권수익률의 분포로 활용하기에는 꼬리가 과도하게 얇고 이에 자산 또는 증권 가격 변동의 위험을 과소평가한다. 또한, n 개의 자산으로 구성되는 포트폴리오의 수익률이 정규분포를 따르려면, n 개 자산의 확률분포는 n -변량 정규분포여야 하는데, 이러한 다변량 정규분포에 대한 이해가 쉬운 일도 아니다.

본고에서는 관찰한 수익률 데이터를 있는 그대로를 그 확률분포로 보는 방법을 제안한다. 통계학적으로는 비모수적 방법론이다. 그러한 확률분포는 특별한 이름은 없지만 히스토그램으로 그 분포를 편향 없이 나타낼 수 있고, 히스토그램이 곧 그 변수의 확률밀도함수가 된다. 여러 자산 수익률에 대해서도 결합적으로 관찰된 수익률 데이터를 그대로 결합확률분포로 보면 된다. 물론, 3개 이상 변수에 대해서는 그 히스토그램이 시각적으로 그려지지 않는다는.

본고에서는 4개 자산으로 구성되는 포트폴리오를 운용함에 있어 4개 자산에 대한 최적 가중치를 정하는 문제를 다룬다. 이때 4개 자산에 대해서 과거에 관찰된 수익률 데이터를 있는 그대로 결합확률분포로 활용한다. 매 의사결정 시점에 적용하는 최적 가중치 선정의 기준은 표준편차 최소화, 평균수익률 최대화, VaR 최소화, 샤프비율 최대화 등 다양할 수 있다. 이들 여러 가지 기준에 대해서 포트폴리오의 운용을 시뮬레이션하여 결과를 비교한다.

주제어 : 확률분포, 정규분포, 다변량 정규분포, 히스토그램, 포트폴리오 운용

무위험 자산·증권을 제외하면, 어느 자산이나 현재 시점으로부터 미래 특정 시점까지의 수익률은 확률변수(random variable)이다. 사전적으로는 어떤 수익률이 실현될지 알 수

논문접수일 : 2022년 11월 30일 논문수정일 : 2022년 12월 12일 논문게재확정일 : 2022년 12월 15일
본고는 김누리·빈기범(2021)의 일부 내용을 참고하고 준용하였다.

* 명지대학교 경제학과 교수, E-mail: bink1@mju.ac.kr

없고 미래 그 시점까지 가 보아야 수익률이 사후적으로 하나의 값으로 실현된다. 이러한 수익률을 확률변수로 보는 것은 타당하며, 통상 정규분포를 여기에 가져다 쓰는 경우가 많다. 금융경제학에서도 자산 수익률은 종종 정규분포를 이용하여 모형화된다. 기하적 브라운 운동(geometric Brownian motion)이 그 대표적인 사례이다. 정규분포를 전제할 때 자산 또는 포트폴리오 수익률에 대한 이해가 쉬울 수 있고, 변동성, VaR, ES(expected shortfall) 등의 계산도 빠르고 편리할 수 있다. 또한, 무한히 많은 확률분포 중에서 정규분포가 정규라는 용어가 주는 뉘앙스 탓도 있지만 가장 많이 접하는 친근한 확률분포이다. 평균 주변에서 실현될 확률이 높고 거기서 멀어질수록 대칭적으로 확률이 감소하는 정규분포의 특성은 자산 수익률 묘사에도 부합될 수 있다.

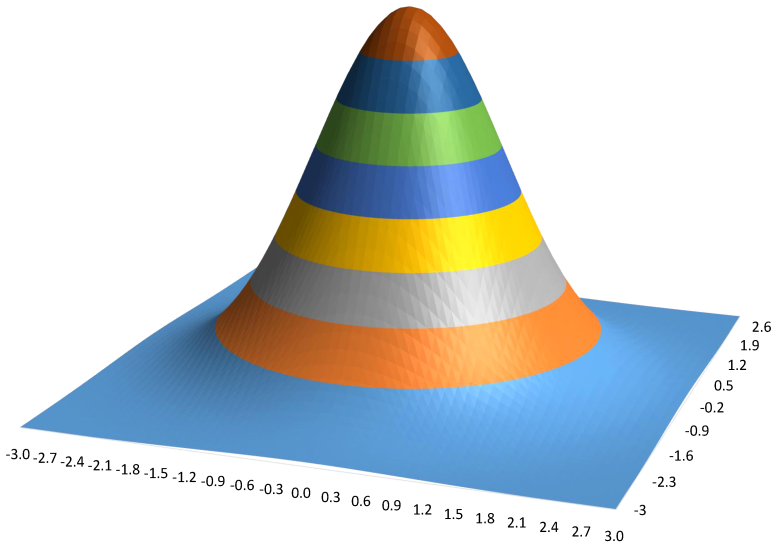
하지만, 자산 수익률을 정규분포로 모형화하는 것은 여러 가지로 단점이 많다. 첫째로, 잘 알려진 대로 자산 수익률 분포의 꼬리는 두껍기 마련인데(fat tail), 정규분포의 꼬리는 극단적으로 얇다. 실제로 정규분포의 꼬리는 통상적으로 이해되는 것 이상으로 얇다. 예를 들어, 정규분포에서는 평균으로부터 3 표준편차 바깥의 확률면적이 0.3%에 불과하다. 즉, 3/1000이다. 그러한 극단적으로 높거나 낮은 수익률은 폭등 또는 폭락을 의미하는데, 정규분포는 그러한 폭등·폭락이 1,000 거래일 중 3일 정도 발생함을 시사한다. 1,000 거래일이면 4년이다. 주식의 경우라면 폭등·폭락이 4년에 겨우 3번 정도의 빈도로 발생하지 않는다. 1년에도 여러 차례 발생한다. 결국 정규분포는 자산 수익률의 꼬리 위험을 크게 과소평가한다. 그 결과 VaR이나 ES도 과소 추정된다.

둘째, 자산 수익률은 그 자산이 무한책임증권이 아닌 한 최악의 수익률은 -100%이다. 이는 자산 구입에 들어간 비용 전액을 모두 날리는 경우이다. -100% 그 이하의 수익률이 나올 수는 없다. 반면, 정규분포는 $-\infty$ 에서부터 ∞ 까지 모든 실수가 가능하다. 다만, 수익률을 로그 수익률로 계산할 경우에는 -100% 이하의 수익률도 가능하다.

셋째, 정규분포는 자주 쓰이지만, 결코 쉽지 않은 확률분포이다. 일단 연속확률분포의 개념부터 쉽지 않다. 확률밀도함수로 그 분포를 묘사하는데 확률밀도가 확률은 아니다. 밀도함수의 식도 간단하지 않다. 미적분을 써야 여러 가지 계산이 가능하다. $-\infty$ 에서부터 ∞ 까지 모든 실수가 가능한 분포이므로 우리가 현실에서 그러한 분포를 갖는 모집단이든 표본이든 관찰할 수 있는 경우는 발생할 수 없다. 연속확률분포는 단지 우리 사고 체계 속의 이론적인 분포일 뿐이다. 나아가, n 개 자산으로 구성된 포트폴리오 수익률이 정규분포를 따르려면, n 개 자산의 수익률은 n -변량 정규분포(n -variate normal distribution)를 따라야 한다. 우리에게 잘 알려져 쉽다고 느끼는 정규분포는 일변량 정규분포일 뿐이다. 포트폴리오 수익률은 일변량이므로 이를 일변량 정규분포로 전제하는 것인데,

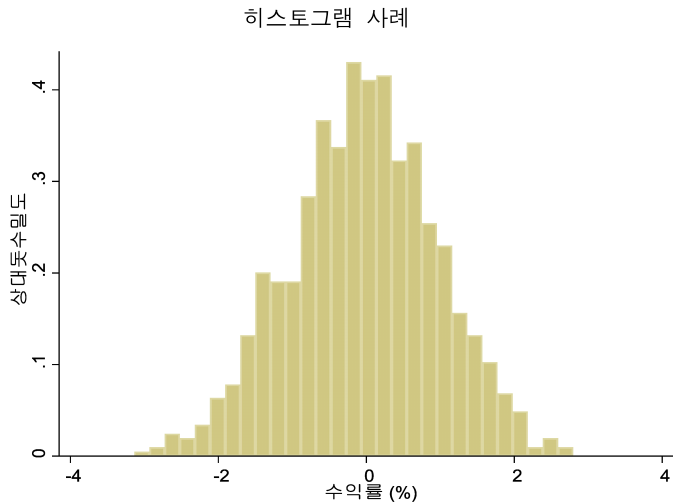
그 이면에는 포트폴리오를 구성하는 n 개 자산 수익률의 결합분포(joint distribution)가 n -변량 정규분포임을 가정하고 있는 것이다. 이 경우, 많은 계산에서 n 중 적분(n -tuple integral)까지 해야 한다.

이변량 정규분포의 밀도함수 그래프 사례



정리하면, 우리가 수익률의 확률분포로 흔히 사용해 오던 정규분포는 쉽지 않은 확률분포일 뿐만 아니라 수익률의 양태를 제대로 묘사하기 어려운 확률분포이다.

그렇다면, 정규분포가 아니며 어떠한 확률분포를 사용해야 하는가? 정규분포든 다른 이름을 가진 분포든 어떤 특정 분포를 사용하려 애쓸 필요가 없다. 우리가 관찰한 자료 그대로 경험적 분포를 특별한 가정이나 전제를 붙이지 말고 최대한 있는 그대로 사용하면 된다. 이를 통계학에서는 흔히 비모수적(non-parametric) 접근법이라 한다. 예를 들어, 수치 50개를 관찰했을 때 그 평균과 분산을 계산한 후, 그 평균과 분산을 갖는 정규분포라면서 애써 정규분포를 끌고 들어올 필요가 없다. 50개 수치의 분포를 (경험적) 히스토그램으로 묘사할 수 있고 그 자체가 아무런 분포나 모형에 대한 가정이 없는 경험적 확률분포를 나타낸다. 실은 히스토그램의 높이가 확률밀도를 나타낸다. 만일, 50개 수치가 양과 음으로 걸쳐 있는 손실(loss)의 수치라면, 이 중 95-백분위수(95-th percentile)가 계산될 수 있고 이것이 95% 신뢰수준 VaR이 된다. 정규분포 가정 없어도 필요한 통계량은 얼마든지 계산할 수 있다.



일변량의 차원을 넘는 다변량에 대해서도 마찬가지다. 예를 들어, 이변량으로서 관찰된 수치들이 있다면 그 분포는 평면 위에서 사각 막대들로 그려지는 히스토그램으로 묘사할 수 있다. 여기에 특별히 이변량 확률분포로서 이름을 갖는 분포를 갖다 붙일 필요가 없다. 그 히스토그램은 그저 이름이 없는 분포이며, 확률분포가 꼭 이름을 가져야 하는 것도 아니다! 한편, 삼변량 이상 넘어가면 우리가 머릿속으로 그려지는 히스토그램에 의존할 수밖에 없다. 시각적으로 편하게 볼 수 있는 방식은 불가할 것이다.

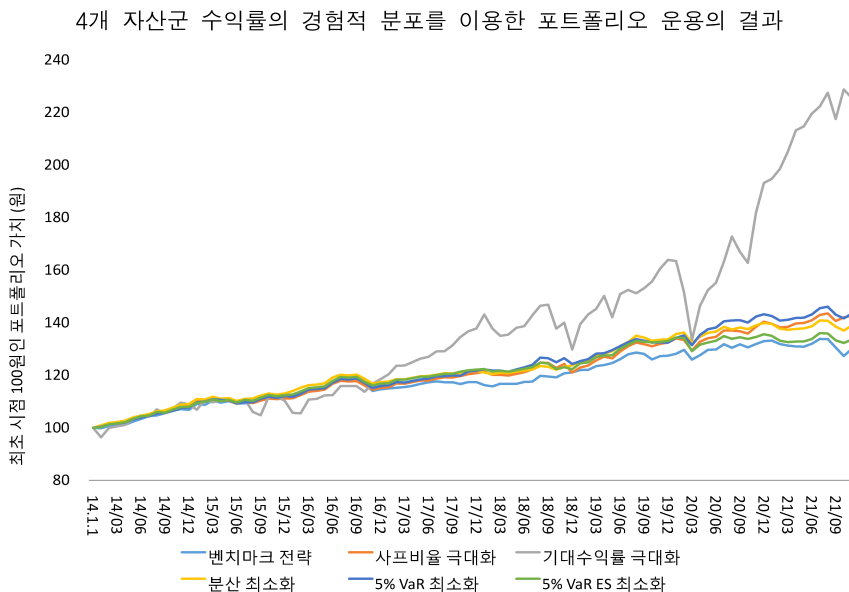
KOSPI200, KIS국채종합총수익지수, S&P500(Hedged to KRW), Barclays Global Aggregate Index(Hedged to KRW)의 4가지 자산군으로 구성되는 포트폴리오를 고려해 보자. 4개 자산군에 대해서 한 달 동안 유지할 가중치를 지금 결정해야 한다고 하자. 추구하는 목표는 다양할 수 있다. 예를 들어, 기대수익률 극대화 전략, 분산 최소화 전략, 샤프 비율 극대화 전략 등이다. 샤프 비율 극대화 전략을 쓴다고 하자. 그렇다면, 어떤 가중치 하에서 포트폴리오 샤프 비율이 최대화될 것인지 고민해야 한다.

이를 위해, 4개 자산군의 과거 50개월 실현치를 가져오자. 4개 수치 세트가 50개 있는 것이다. 이 자체로 4변량 경험적 분포를 가진 것이다. 4개 자산에 대한 가중치 $(\omega_1, \omega_2, \omega_3, \omega_4)$ 의 대안은 무한히 많다. 우리가 무한을 다룰 수는 없다. 그래서, $(\omega_1, \omega_2, \omega_3, \omega_4)$ 에 대한 1,000개의 세트를 만들어 놓자. 각 세트는 다음과 같이 무작위적으로 만들어진다. u_1, u_2, u_3, u_4 를

구간 $[0,1]$ 에서의 균등분포에서 추출하고, $\omega_i = \frac{u_i}{u_1 + u_2 + u_3 + u_4}$ 로 계산하면 된다. 그러면,

$\omega_1 + \omega_2 + \omega_3 + \omega_4 = 1$ 이 된다.

1,000개의 $(\omega_1, \omega_2, \omega_3, \omega_4)$ 중 첫 번째를 4개 수치로 구성된 50개 수익률에 적용하자. 그러면 50개의 포트폴리오 수익률이 나온다. 이것이 포트폴리오 일변량 수익률의 경험적 분포가 된다. 이로부터 평균, 표준편차 등을 계산하여 샤프비율을 계산한다. 동일한 방식을 반복하면 총 1,000가지의 가중치 세트마다 포트폴리오 수익률의 샤프비율이 나올 것이다. 그중 가장 높은 샤프비율을 갖는 가중치를 선정하고, 그걸 오늘 포트폴리오 가중치로 적용하면 된다. 이처럼, 다변량 결합분포의 경우에도 특별한 분포 가정 없이 있는 그대로 경험적 분포를 사용하면 된다. 소위 역사적 수익률(historical return)을 그대로 사용하는 것이다. 앞서 설명한 이런 과정을 역사적 모의실험(historical simulation)이라 부를 수 있다.



상기와 같은 포트폴리오 운용 방식을 2014년 1월 초부터 2021년 10월말까지 총 95개월에 대해서 1개월씩 시간에 따라 이동하면서 모의실험하고 그 투자성과를 산출하고자 한다. 추구하는 전략이 다를 수 있어, 샤프비율 극대화, 평균수익률 극대화, 분산 최소화, 5% VaR 최소화, 5% VaR ES 최소화 각각의 성과를 비교하고자 한다. 이 모의실험에서는 절대 미래를 엿보지 않는다는 전제가 반드시 지켜진다. 초기 100원으로 시작된 포트폴리오의 가치의 변화는 위의 그림과 같다. 평균수익률 극대화 전략의 성과가 두드러지게 높다. 다만, 변동성은 심하다. 다른 전략들의 성과는 서로 유사하다. 한편, 여러 전략 중 사후적으로

샤프비율이 가장 높은 전략은 5% VaR 최소화 전략이다. 샤프비율 극대화 전략이 그 자리를 차지하지 못했다.

정규분포는 우리가 많이 들어서 친근할 뿐 그렇게 쉽게 이해가 되는 분포도 아니다. 게다가, 자산 수익률을 묘사하는데 부적합하다. 그러한 정규분포 사용을 고수할 일이 아니다. 확률분포의 이름이 있어야만 확률분포가 되는 것이 아니다. 이름 없는 분포가 얼마든지 있다. 관찰한 수익률 자료 자체를 그대로 경험적 분포로 받아들여도 필요한 여러 가지 계산이 가능하고 오히려 더 쉽다. 또한, 포트폴리오 운용 전략을 수립하는데도 유용하게 사용될 수 있다.

참 고 문 헌

- 김누리, 빈기범, “중기적 관점에서의 전략적 자산배분수행 방안에 관한 연구”, 국유재산관리기금 제출, 2021.
- 강병진, 김솔, 빈기범, 윤선중, 이상호, “금융기관의 위험관리”, 시그마프레스, 2013.