

GPD 기반의 유전자 알고리즘을 이용한 포트폴리오 최적화[†]

김현돈¹ · 김현태²

^{1,2}연세대학교 응용통계학과

접수 2015년 10월 30일, 수정 2015년 11월 23일, 게재확정 2015년 11월 24일

요약

최적의 포트폴리오를 선택하기 위한 연구는 평균-분산모형을 시작으로 다양하게 진행되어 왔다. 과거에는 위험자산의 확률분포가 정규분포를 따른다고 가정하여, 투자자가 보유한 위험자산의 분산이 최소화되고 기대수익률이 최대가 되도록 포트폴리오를 구성하도록 하였다. 그러나 실제 위험자산의 분포에는 극단적인 사건들이 많이 발생하기 때문에 정규분포보다 훨씬 꼬리부분이 두꺼우며, 또한 왼쪽꼬리와 오른쪽꼬리가 대칭적이지도 않은 것으로 밝혀졌다. 이에 본 논문에서는 위험자산의 확률분포를 극단치 이론에서 널리 사용되는 일반화 파레토분포 (GPD)로 모형화하였고 체계적인 위험의 추정을 위하여 VaR를 이용하는 한편, 최적의 포트폴리오의 탐색을 위해서는 유전자 알고리즘을 사용하였다. 제안 방법의 적정성을 확인하기 위해 국내 증시에서 최적 포트폴리오를 탐색해 보았으며, 그 결과 GPD로 투자자산의 위험을 추정하였을 때 가장 좋은 결과를 얻을 수 있었다.

주요용어: 극단치 이론, 유전자 알고리즘, 최대예상손실액 (VaR), 일반화 파레토분포, 포트폴리오 최적화.

1. 서론

포트폴리오 최적화란 일련의 기준에 따라 투자자산을 조합하여 더 나은 투자효율을 얻는 것을 의미한다. 일반적으로 투자자들은 위험 (Risk)을 최소화하면서 수익률을 극대화하고자 한다. 이를 위해 투자자들은 투자자산과 시장에 대한 분석을 수행하고 투자대상의 위험과 기대수익을 평가한 후 자신의 효율을 극대화하는 포트폴리오를 선정한다. 즉, 투자자산에 대한 모델링 방법과 위험 평가방법이 포트폴리오의 투자효율에 중대한 영향을 미친다.

Markowitz (1952)에 의해 평균-분산모형 (Mean-Variance model)이 제안된 이후, 포트폴리오 선택 및 최적화에 관한 연구가 다양하게 진행되어 왔다. 평균-분산모형의 경우, 위험자산의 확률분포가 정규분포를 따른다고 가정하여, 투자자가 보유한 위험자산의 분산이 최소화되고 기대수익률이 최대가 되도록 포트폴리오를 구성하도록 하였다. 그러나 실제 위험자산의 분포에는 극단적인 사건들이 많이 발생하기 때문에 정규분포보다 훨씬 꼬리부분이 두꺼우며, 또한 왼쪽꼬리와 오른쪽꼬리가 대칭적이지도 않은 것으로 밝혀졌다. 즉, 정규분포에 근거한 위험측정은 과소평가하는 경향이 있고 왼쪽꼬리와 오른쪽꼬리의 특성을 반영하지 못하기 때문에 적절한 위험관리모형으로 부적합한 것이다. 이에 정규분포 보다 꼬리가 두꺼운 분포를 이용하여 포트폴리오 선택을 최적화 하는 연구들이 진행되었는데, Longin (2000)은 극단치 이론을 이용하여 최적의 포트폴리오를 얻고자 하였다.

[†] 본 논문은 한국연구재단의 지원을 받아 수행된 것임 (NRF 2015R1A1A1A05027336).

¹ (03722) 서울특별시 서대문구 연세로 50, 연세대학교 응용통계학과, 석사.

² 교신저자: (03722) 서울특별시 서대문구 연세로 50, 연세대학교 응용통계학과, 교수.

E-mail: jhtkim@yonsei.ac.kr

한편, 평균-분산모형에서 위험을 분산으로 측정된 것과 달리, VaR (Value at Risk)를 위험 측정 수단으로 이용한 평균-VaR모형도 Alexander와 Baptista (2002) 등에 의해 제안되었다. 즉, 일정 신뢰수준 하에서 보유한 투자자산에 대해 발생할 수 있는 최대손실액을 의미하는 VaR를 이용하여 위험을 체계적으로 측정하고자 한 것이다. 또한, Anione과 Tettamanzi (1993)를 시작으로 Shoaf와 Foster (1998), Oh 등 (2006), Byun 등 (2009), Chung과 Oh (2014)는 유전자 알고리즘을 이용한 포트폴리오 최적화 연구를 수행하였다.

최근에는 위의 방법들을 함께 사용한 연구들도 진행되었는데, Lin과 Ko (2009)는 유전자 알고리즘과 극단치 이론을 이용하여 포트폴리오를 최적화하는 방법을 제안하였다. Lin과 Ko의 방법은 VaR 값을 초과한 손실 발생의 빈도가 낮은 포트폴리오를 최적으로 선택하도록 하였다. 그러나, 위험 관리에만 초점을 두고 수익에 대해서는 고려를 하지 않았다는 점에서 현실에 적용하기에는 무리가 따른다. 반면, Ranković 등 (2014)은 유전자 알고리즘과 분포 가정 없는 5% historical VaR 값을 이용하여 최적의 포트폴리오 선택 방법을 제안하였다. Ranković 등은 수익과 위험을 모두 고려하여 포트폴리오 최적화를 수행하였지만, historical VaR 값을 이용함에 따라 과거 보다 큰 위험이 발생할 경우에 대한 대비가 충분하지 못하였다.

본 논문에서는 다수의 위험자산 중, 위험을 최소화 하면서 적정 수익을 낼 수 있는 방법을 제안한다. 최적의 포트폴리오를 구성하기 위해서 유전자 알고리즘 (Genetic Alogrithm)을 이용하였고, VaR의 추정을 위해 투자자산을 두꺼운 꼬리를 가지는 GPD (Generalized Pareto Distribution)에 따라 모델링하여 그 결과를 정규분포, t분포와 비교·분석해 보았다.

본 논문의 구성은 다음과 같다. 2절에서는 GPD에서 VaR를 추정하는 방법을 살펴보고 포트폴리오 최적화를 위해 이용된 유전자 알고리즘에 대해서도 설명하였다. 3절은 GPD 기반 유전자 알고리즘을 이용한 포트폴리오 최적화 방법에 대해서 자세히 설명하고, 4절에서는 제안 방법을 국내 주식시장의 KRX100 종목들에 적용한 결과를 분석·설명하였다. 마지막으로 5절에서는 결론을 도출하고 향후 추가 연구방향을 제시하였다.

2. 관련 방법론

본 논문에서는 포트폴리오의 위험을 평가하기 위해서 투자자산을 GPD로 모델링 한 후 VaR를 추정하였고, 최적의 포트폴리오를 탐색하기 위해서 유전자 알고리즘을 이용한다. 이 절에서는 관련 내용을 간략히 소개하고자 한다.

2.1. GPD에서의 VaR 추정

자료의 극단 값 추정에 이용되는 극단치이론은 극단값을 어떻게 정의 하느냐에 따라 크게 GEV (Generalized Extreme Value) 접근법과 POT (Peak Over Threshold) 접근법으로 나뉜다. GEV 접근법 혹은 구간최대법 (Block Maxima)은 자료가 시간에 따라 관측이 될 때 관측시간을 같은 크기의 세부구간들로 나눈 후 각 구간의 최대값들을 극단치로 정의하게 된다. 이 경우 극단치들은 GEV 분포에 수렴하게 됨이 알려져 있으나 이 방법은 상대적으로 많은 관측치를 버려야 하며 자료가 시간에 의존하지 않는 경우 사용하기가 쉽지 않은 단점이 있다. 이와 달리 POT 접근법은 특정 임계치를 초과하는 값들을 극단치로 정의하여 GPD로 모형화한다. 즉, 어떤 확률변수 X 의 분포에 대해 충분히 큰 임계치 $u > 0$ 를 적절히 선택하면 $X - u | X > u$ 가 점근적으로 GPD를 따른다는 성질을 이용하는 것이다. 이 방법은 버려지는 관측치의 수가 상대적으로 작고 자료가 시간에 의존하지 않아도 문제가 없어 GEV 접근법에 비해 상대적인 장점이 있고, 이러한 이유로 극단치이론의 응용에는 대부분 POT 접근법이 사용되고 있다. 본 논문에서도 포트폴리오 손실의 극단값을 모델링하기 위해 POT 접근법을 이용하여 GPD의 VaR를 위험으로 추정할 것이다.

POT 접근법을 좀 더 자세히 설명해 보자. 먼저, iid (independent identically distributed)인 X_1, X_2, \dots, X_n 의 분포가 F_X 라고 할 때, X_1, X_2, \dots, X_n 중 특정 임계치 (u)를 초과하는 초과잔차 $X - u | X > u$ 를 y 로 둔다면 y 의 확률 분포는 X 에 대한 조건부 확률을 이용하여 아래와 같이 정의할 수 있다.

$$F_u(y) = P(X - u \leq y | X > u) = \frac{F(u+y) - F(u)}{1 - F(u)}, \quad (0 \leq y < \infty) \quad (2.1)$$

Balkema와 de Haan (1974) 및 Pickands (1975)에 의하면 임계값 (u)이 무한히 증가하면 초과잔차의 분포 $F_u(y)$ 는 다음과 같은 GPD 분포를 따른다고 한다.

$$G_{\xi, \beta}(y) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\beta}\right)^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - \exp\left(-\frac{y}{\beta}\right) & \text{if } \xi = 0 \end{cases} \quad (2.2)$$

여기서 β 는 척도모수 (scale parameter)이며, ξ 는 분포의 형태를 결정짓는 형태모수 (shape parameter)이다. 만약 $\xi > 0$ 이면 Type I 파레토분포, $\xi < 0$ 이면 Type II 파레토분포, $\xi = 0$ 이면 지수분포 형태를 가진다 (Oh, 2005). GPD 분포의 모수들은 최우추정법 (Maximum Likelihood Estimation)에 의해 추정하게 되고, 추정된 모수 $\hat{\beta}$ 와 $\hat{\xi}$ 를 통해 신뢰수준 100 $\alpha\%$ 가 주어질 경우 VaR는 다음과 같이 산출한다.

$$VaR_\alpha = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{n}{N_u} (1 - \alpha) \right)^{-\hat{\xi}} - 1 \right) \quad (2.3)$$

여기서 n 은 총 관측치의 수, N_u 는 임계값을 초과한 관측치의 수를 의미한다. 분포적인 관점에서는 식 (2.2)에 위치모수를 넣어 세 개의 모수를 가지는 GPD를 정의할 수도 있으나 초과잔차의 모형에는 사용되지 않아 고려하지 않는다.

한편 모수들을 추정하기 전에, 임계값을 결정하여야 하는데 이를 결정하는 방법으로 표본 평균초과함수 (mean excess function)를 이용하여 선형함수의 형태가 시작되는 지점을 선택하는 방법과 표본에 로그를 취하여 표본 평균초과함수의 역수와 유사한 Hill추정량을 활용한 Hill plot 방법 (Embrechts 등, 1997), 정규분포의 백분위수에 기반하여 표준편차의 1.65배를 가정하는 방법 (Davison과 Smith, 1990) 등이 있다. 본 논문에는 그래프 등을 통해 사람이 판단해야하는 기존 방법과 달리 자동으로 임계치 u 값을 구하기 위해 유전자 알고리즘을 이용하였다.

2.2. 유전자 알고리즘

1975년 Holland에 의해 제안된 유전자 알고리즘은 자연 진화의 법칙인 적자생존 (survival of fittest)과 자연도태 (natural selection)의 원리를 이용하여 주어진 문제 공간에서 최적의 해를 탐색하는 알고리즘이다. 모집단을 구성하는 각 개체는 비트 스트링 형태의 염색체 (chromosome)로 표현되고 적합도 함수를 이용하여 각 개체의 최적 해 근사 정도를 평가한다. 한 세대의 집단은 교차 (crossover), 돌연변이 (mutation), 선택 (selection)의 유전자 연산 (genetic operator)을 수행하여 적합도가 높은 개체들로 부터 새로운 세대의 모집단을 재생산한다. 일반적으로 유전자 알고리즘은 아래와 같은 단계를 통해 최적 해를 탐색한다 (Eberhart 등, 1996).

Table 2.1 Procedure of genetic algorithm

- | |
|---|
| ① Initialize the population of chromosomes |
| ② Calculate the fitness for each individual in the population using fitness function |
| ③ Reproduce individuals to form a new population according to each individual's fitness (selection) |
| ④ Perform crossover and mutation on the population |
| ⑤ Go to step ② until some condition is satisfied |

각 세대의 우수 개체들은 형태 그대로 보존되거나 교차와 돌연변이 연산으로 다음세대를 재생산하게 되는데, 이는 다윈의 진화론을 반영하고 있다. 유전자 알고리즘은 전통적인 최적화 알고리즘과는 달리 도함수를 이용하지 않는다는 특징을 가지고 있으면서도 순회 판매원 문제 (traveling salesman problem), 발전소 최적화 (optimal use of power plants), 그래프 분화 문제 (graph partitioning problem)와 같은 복잡한 문제에 대한 좋은 (best) 해를 찾을 수 있음이 알려져 있다 (Chambers, 1995; Golberg, 1989). 한편, Holland는 Schema 정리를 통해 유전자 알고리즘의 최적 해로의 수렴성을 증명하였다 (Bridges와 Goldberg, 1987).

3. GPD기반 유전자 알고리즘을 이용한 포트폴리오 최적화 방법

3.1. 개요

본 논문에서 제안하는 포트폴리오 최적화방법을 단계별로 설명하면 아래의 Table 3.1과 같다. 최적의 포트폴리오를 선택하기 위한 수익과 위험의 측정은 각각 누적등락률과 VaR를 이용하였다. 누적등락률은 매수 후 보유 (buy and hold) 전략을 따른다고 가정하였으므로 최초 매수시점부터의 등락률 누적값을 이용한다. 유전자 알고리즘 관련 검색체 초기화 및 최적화 함수에 대한 내용은 다음절에 자세히 설명한다.

Table 3.1 Algorithm of portfolio optimization

① Initialize the population :
Randomly initialize 1000 individual's chromosomes (reference section 3.2)
② Iterate the below procedure for individuals of each generation until $iter_{max}$ times
②-1 : Calculate average accumulated rate of return (R_i) of enrollment stock item for each individual (portfolio)
②-2 : Calculate average VaR (V_i) of enrollment stock item for each individual (portfolio)
②-3 : Calculate the fitness for individual i using $R_i VaR_{\alpha, i}$ and target return ($Target_R$) (reference section 3.3)
$Fitness_i = \max(R_i - Target_R, 0) - V_i$
②-4 : Using the $Fitness_i$ of ②-3, perform the selection, crossover and mutation operation on the population to reproduce new generation which mainly contains better fitness individuals of previous generation
③ If step ② is finished, change the number of enrollment stock item and repeat step ①~②
④ change the target return ($Target_R$), and repeat step ①~③

여기서 $iter_{max}$ 는 반복해야할 세대수 즉, 진화회수를 의미한다. 95% VaR는 각 분포별·종목별로 구한 값이다. GPD의 $VaR_{0.95}$ 는 3.1절에서 제안한 방법으로 산출한 값을 사용하였다. 세 분포의 결과를 비교·분석하기 위해서 Table 3.1의 ①~④ 단계를 분포별로 각각 실행한다

3.2. 검색체의 구조

3.2.1. 정규분포, t분포의 경우

먼저, 유전자 알고리즘을 적용하기 위해서는 검색체의 구조를 정의하여야 한다. N 개 주식 종목 중 포트폴리오 편입여부를 검색체로 나타낸다면, 한 종목의 편입여부를 나타내기 위해서 1 비트가 필요하다 (즉, 값이 0이면 불 포함, 1이면 포함). 그러므로 N 개 종목 편입여부를 나타내는 검색체는 N 비트의 비트 스트링이 되고 위의 Figure 3.1과 같다.

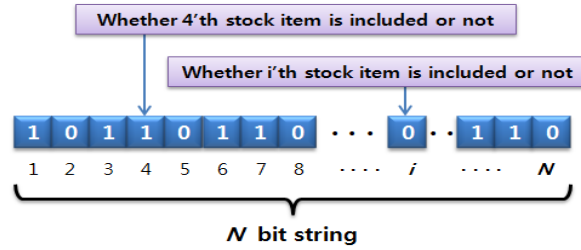


Figure 3.1 Chromosome structure (t and normal distributions cases)

3.2.2. GPD의 경우

극단값을 GPD로 모델링하려면, 임계값을 미리 결정해야 한다. 그래서 GPD의 경우 2단계로 유전자 알고리즘을 적용하였다. 즉, 1단계에서 GPD의 임계값 u 를 탐색하였고 2단계에서 최적 포트폴리오 조합을 탐색하였다.

1단계인 GPD의 임계값 u 를 탐색을 위해 임계값의 범위에 따른 탐색체의 길이를 구하여야 한다. 본 논문에서는 임계값의 범위를 $[0, U_{max}]$ 라고 하고 그 값은 0부터 0.5씩 증가한다고 가정하였다. 그러므로 한 종목 당 $2 \times U_{max}$ 만큼의 정보를 표현할 수 있는 비트 스트링을 탐색체 길이로 지정하여야 한다. 탐색체를 구성하는데 필요한 비트 스트링의 길이 l 은 아래와 같이 나타 낼 수 있다.

$$l = \lceil \log_2(2 \times U_{max}) \rceil \times N \tag{3.1}$$

여기서 N 은 포트폴리오에 편입될 수 있는 종목의 수를, $\lceil \cdot \rceil$ 는 올림 (ceiling) 함수를 나타낸다. 탐색체 구조는 위의 Figure3.2의 1단계와 같다. 편의상 l 은 4장 실험에서 사용된 5로 표시하였다. (4장의 실험 데이터를 살펴본 결과, U_{max} 가 50이었고 이에 따라 l 이 5로 결정되었다.) 2단계 포트폴리오 선택을 위한 탐색체 구조는 3.2.1절의 정규분포, t분포의 경우와 동일하다. 그러므로 GPD를 위한 2단계 유전자 알고리즘의 탐색체 구조는 Figure3.2와 같이 나타낼 수 있다.

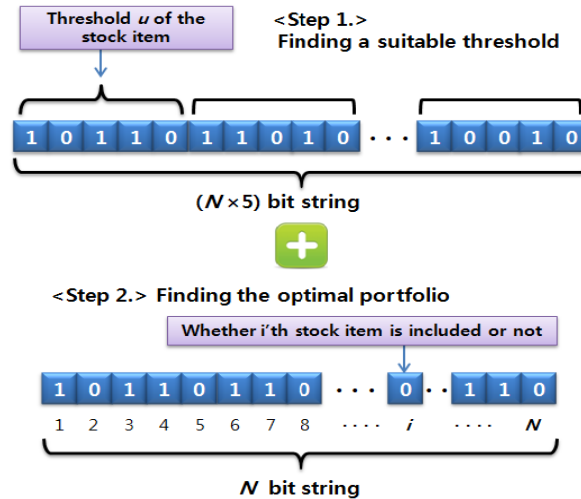


Figure 3.2 Chromosome structure (GPD case)

3.3. 적합도 함수

적합도 함수 Fitness는 높은 수익률을 달성하면서 위험을 최소화하도록 구성하여야 한다. 즉, 수익률을 나타내는 포트폴리오의 누적등락률 평균 R 과 위험을 나타내는 VaR_α 의 평균 V 는 다음의 관계가 있다 (Ranković 등, 2014).

$$Fitness = \lambda R - (1 - \lambda)V \quad (3.2)$$

여기서 λ 는 $[0,1]$ 인 실수이고, R 은 2003년 1월 1일 구매 후 보유한 것으로 가정하였을 때의 누적등락률이다. V 는 페널티의 개념으로 사용되었다. 위 식에서와 같이 Fitness는 λ 값에 따라 무수히 많은 해를 가진다. 유전자 알고리즘의 적합도 함수는 하나의 목적을 최적화하는 해를 찾을 수 있으나, 본 논문에서는 높은 수익률을 달성하면서 위험을 최소화 하도록 하는 두 가지 목적 (bi-objective)의 해를 찾자 한다. 그러므로 우리는 하나의 해를 찾을 수 없고 λ 를 변경해 가면서 해의 집합을 찾아야 하는 것이다.

λ 에 $[0,1]$ 사이의 값을 대입하여 여러 해를 구할 수 있으나, λ 가 수익률과 위험사이에 어떤 의미를 가지는지는 직관적으로 이해할 수 없다. 그래서 Ranković 등 (2014)는 누적수익률 R 을 20등분하여 R 값을 변화시키면서 해를 구하였다. 본 논문에서도 위의 적합도 함수를 아래와 같이 직관적으로 이해하기 쉬운 형태로 변경하였다.

$$Fitness = \max(R - Target_R, 0) - V \quad (3.3)$$

여기서 $Target_R$ 은 목표수익률을 의미한다. 즉, 목표 수익률 이상을 달성하면서 위험이 최소인 해를 탐색하겠다는 것이다. 뿐만 아니라, λ 대신 $Target_R$ 을 나열하면서 직관적인 해를 찾겠다는 것이다. 4장 실험에서 자세히 설명하겠으나, 본 논문에서는 $Target_R$ 을 전체 종목의 수익률 중 50% 분위값 (quantile), 65% 분위값, 80% 분위값으로 변경하면서 결과를 도출하였다. 식 (3.3)의 적합도 함수는 GPD의 임계값을 찾는 적합도 함수로 사용하였다.

유전자 알고리즘을 반복해서 수행하면 조건이 일치하는 하나의 종목을 해로 수렴한다. 즉, 복수의 종목을 해로 지정하지 않는다. 물론 진화를 반복하는 세대 수를 작게 지정한다면 중간단계의 복수 종목이 해로 나타나긴 하지만 이해가 얼마만큼의 수렴이 이루어 졌는지 알 수가 없다. 그러므로 분산 투자를 유지할 수 있도록 포트폴리오에 선택된 종목의 수 " Sel "을 지정하도록 다시 한번 적합도 함수를 변경하였다. Statman (1987), Marsili et al. (1998)을 비롯한 많은 학자들이 분산투자를 위한 적정 투자자산 수에 대한 다수 연구가 진행하여왔으나 특정 종목 수나 기준 등이 명확하게 도출되지 않았다. 그래서, 본 논문에서는 Sel 값을 변경해 가면서 실험 결과를 비교·분석하였다. 변경된 적합도 함수는 아래와 같다.

$$Fitness = \{\max(R - Target_R, 0) - V\} \times \max(Sel + 1 - N_{sel}, 1) \quad (3.4)$$

여기서 N_{sel} 은 해당 개체에서 선택한 포트폴리오의 수를 의미하고 종목수가 Sel 보다 작을 때 페널티를 적용하기 위해서 $\max(Sel + 1 - N_{sel}, 1)$ 값을 적합도 값에 곱해주도록 하였다.

4. 국내 주식시장에서의 포트폴리오 최적화 실험

4.1. 실험 환경

본 논문에서 제안하는 포트폴리오 최적화 방법을 국내 주식시장에 적용해 보았다. 우리나라 주식시장 종목 중 KRX100에 편입되어 있고 2003년 1월 1일 현재 상장되어 있는 종목 66개를 실험 대상으로 하였다. KRX100은 한국거래소 (KRX)가 유가증권 (KOSPI), 코스닥 (KOSDAQ) 전체 중 우량 종목 100개를 선정한 것으로 두 시장 시가 총액 70% 이상을 차지한다. 한편, 주가 변동폭 제한이 있는 국내 증시 상황을 고려하여 주별 가격에 대해서 실험하였는데, 주별 가격을 아래와 같이 등락률 데이터로 변경하여 이용하였다.

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{4.1}$$

여기서 R_t 는 t 시점 수익률을 의미하고 P_t 는 t 시점 가격을 의미한다.

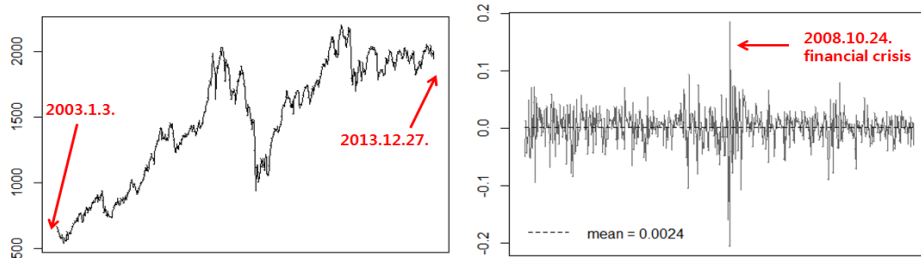


Figure 4.1 Weekly KOSPI and return rate for learning interval

2003년 1월 3일부터 2013년 12월까지의 자료는 학습 데이터로 활용하였고, 2014년 1월 3일 이후의 자료는 테스트 데이터로 활용하였다. 테스트는 주별 종합주가지수가 최고인 지점 2개, 최저 지수인 지점 2개, 총 4개의 시점에서 선택된 포트폴리오 수익률을 비교하는 방식으로 진행하였다. 그 이유는 최고점에서 수익률 관리의 적정성을 확인할 수 있고 최저점에서는 위험관리의 적정성을 확인할 수 있기 때문이다. 한편, Figure 4.1, Figure 4.2에서 보는 바와 같이 등락률의 평균이 학습구간은 양수인 반면 테스트 구간은 음수로 테스트 구간에서는 주가가 하락추세임을 알 수 있다. 이 추세가 결과에 미치는 영향은 결론에서 설명하도록 한다.

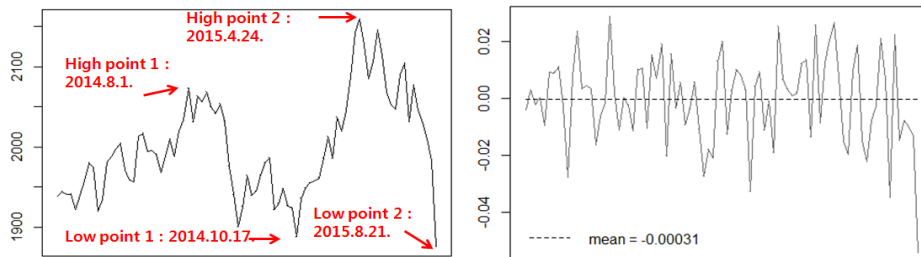


Figure 4.2 Weekly KOSPI and return rate for test interval

앞서 기술한 바대로 실험은 정규분포, t분포, GPD에 대해서 각각 실행하였고, 각 분포별 VaR_α 추정과 임계값 u 의 탐색을 제외한 나머지 실험 환경은 모두 동일하게 구성하였다. 신뢰수준 α 는 0.95로 지정하였고 각 분포에 대해서 목표수익률 $Target_R$ 을 변화시키면서 3번씩 실험을 수행하였다. 또한, 편입종목을 6, 10, 14개로 제한하면서도 실험을 수행하였다. $Target_R$ 은 관측치 데이터의 전체 종목 수익률 중 50% 분위값, 65% 분위값, 80% 분위값으로 변경하면서 결과를 도출하였다. 평균에 해당하는 50%이하의 수익률을 보이는 포트폴리오는 높은 수익률이라는 목표를 달성할 수 없기 때문에 실험대상에서 제외하였다.

4.2. 실험 결과

4.2.1. GPD 임계값 탐색 결과

앞서 기술한 바와 같이 GPD의 $VaR_{0.95}$ 계산을 위해 먼저 임계값을 탐색하는 단계를 실행하였다.

염색체 구조, 적합도 함수 등은 3.2, 3.3절에서 설명한 바와 같고 진화시 5%의 엘리트 개체를 그대로 보존하도록 하였다. 교차연산 확률과 돌연변이 확률은 각각 일반적으로 많이 사용하는 80%와 10% (Scrucca, 2014)로 하였다. (교차 및 돌연변이 확률을 변화시켜 실험을 진행하여도 결과는 동일하였고 80%, 10%일 때 가장 빠른 수렴이 이루어 졌다.) 실험결과 교차 및 돌연변이 확률을 달리해도 결과에는 차이가 나지 않았다. 모집단 내 개체 (종목별 임계값) 수는 3000개로 지정하였고, 최대 반복 회수인 세대수는 150으로 지정하였다.

유전자 알고리즘을 통해 탐색한 임계값 u 는 아래와 같다.

Table 4.1 Thresholds determined by genetic algorithm for each stock

item ID	u	item ID	u	item ID	u	item ID	u
1	3.5	18	5.5	35	2.5	52	3
2	8	19	6	36	8.5	53	7
3	5.5	20	9.5	37	1.5	54	3.5
4	8.5	21	14	38	3.5	55	5.5
5	6	22	2.5	39	3	56	6
6	7.5	23	6	40	7	57	7
7	8.5	24	7	41	5	58	9
8	3.5	25	10.5	42	2.5	59	3.5
9	2	26	3.5	43	9	60	2
10	3.5	27	8	44	3	61	7
11	3	28	7.5	45	6	62	10
12	2	29	4.5	46	1	63	11.5
13	8.5	30	3	47	4	64	1.5
14	8.5	31	3.5	48	8	65	5.5
15	6.5	32	3.5	49	5	66	5
16	9	33	4.5	50	1.5		
17	1	34	5.5	51	1.5		

또한, 세 분포의 $VaR_{0.95}$ 는 아래와 같다.

Table 4.2 95% VaR for each distribution

item ID	normal	t	GPD	itemID	normal	t	GPD
1	8.084	7.959	7.858	34	8.371	8.202	7.546
2	9.862	9.029	7.572	35	8.8	8.428	8.45
3	9.105	8.868	7.426	36	8.501	8.465	8.951
4	11.519	9.806	8.687	37	6.864	6.606	6.736
5	8.275	7.958	7.296	38	10.16	9.281	9.552
6	8.844	8.506	7.63	39	8.788	8.774	8.225
7	8.484	8.222	8.53	40	7.28	7.147	6.783
8	9.41	8.802	8.61	41	8.236	7.628	6.902
9	5.226	5.202	4.923	42	7.986	7.806	8.032
10	9.379	8.998	8.721	43	8.796	7.862	7.35
11	7.607	7.459	7.626	44	25.13	11.853	9.43
12	8.338	8.158	7.539	45	10.308	9.98	8.959
13	8.685	8.314	8.353	46	7.217	7.14	7.824
14	9.529	8.876	9.447	47	8.41	8.32	8.005
15	10.006	9.7	9.649	48	6.501	6.413	6.498
16	8.169	7.472	7.453	49	10.946	10.631	9.811
17	7.979	7.671	7.271	50	7.359	7.273	8.138
18	10.691	9.641	9.181	51	8.631	8.321	8.431
19	6.256	6.042	5.866	52	8.3	8.072	7.251
20	12.132	11.247	10.614	53	25.274	8.487	7.646
21	11.133	10.647	11.856	54	8.412	7.926	6.738
22	12.808	12.467	11.663	55	10.881	10.742	9.717
23	8.968	8.533	8.227	56	6.861	6.581	6.023
24	10.175	10.071	9.386	57	8.65	8.545	8.265
25	10.152	9.197	8.98	58	8.425	8.295	8.504
26	11.318	10.743	11.144	59	11.435	10.564	10.398
27	10.847	10.675	9.984	60	10.71	9.304	8.682
28	10.7	10.153	9.43	61	7.82	7.744	7.306
29	6.049	5.452	4.74	62	8.949	8.807	8.247
30	11.187	10.751	10.266	63	10.367	9.798	8.962
31	11.047	10.657	10.065	64	7.849	7.673	7.662
32	7.489	7.064	6.704	65	9.906	9.618	8.828
33	10.938	10.675	10.807	66	8.753	8.217	7.846

한편, 신뢰수준이 99% 이상일 때는 GPD의 VaR값이 두 분포보다 높은 값을 가지는데, 이는 GPD가 꼬리를 좀더 두껍게 과대 추정하기 때문이다. 반면, 95%의 경우, 위의 Table 4.2와 같이 종목별로 GPD의 VaR값이 두 분포보다 작은 경우도 발생하였다.

4.2.2. 포트폴리오 최적화 결과

4.2.1의 분포별 $VaR_{0.95}$ 값을 이용하여 3장의 포트폴리오 최적화 방법을 국내 증시 데이터에 적용해 보았다. 탐색체 구조, 적합도 함수 등은 3.2, 3.3절에서 설명한 바와 같고 진화시 5%의 엘리트 개체를 그대로 보존하도록 하였다. 교차연산 확률과 돌연변이 확률은 각각 일반적으로 많이 사용하는 80%와 10% (Scrucca, 2014)로 하였다. 모집단 내 개체 (포트폴리오) 수는 1000개로 지정하였고, 최대 반복 회수인 세대수는 100으로 지정하였다. 실험은 포트폴리오 $Target_R$ 과 Sel 을 각각 3단계씩 변화시키면서 진행하였다. 학습 데이터에서의 누적등락률과 실제 수익률은 아래 Table 4.3과 같다. 적합도 평가시 수익 척도는 누적등락률이었고, $Target_R$ 도 누적등락률의 분위값을 사용하였다. 그러나 제안방법의 효율을 확인하기 위해서는 포트폴리오의 누적등락률이 아니라 실제로 얼마나 수익을 냈는지를 확인하여야 한다. 즉, 학습 종료 시점 및 테스트 시점에서 실제로 해당 주식이 얼마나 수익을 창출하였느냐를 측정하여야 하므로 실험 결과는 실제 수익률을 확인한다.

Table 4.3 Accumulated rate of return and return for different quantiles

Quantile	50%	65%	80%
Accumulated rate of return	25.46	29.32	33.21
Return	42.18	64.95	93.61

목표 수익과 편입 포트폴리오 수에 따른 실험의 구성은 Table 4.4와 같다.

Table 4.4 Various experiment combinations for analyses

Exp. ID	condition		Exp.ID	condition	
	target return ($Target_R$)	num. of items (Sel)		target return ($Target_R$)	num. of items (Sel)
1	50%	6	6	65%	14
2	50%	10	7	80%	6
3	50%	14	8	80%	10
4	65%	6	9	80%	14
5	65%	10			

각 실험 1~9를 실시한 결과 실험별로 최종적으로 편입된 종목들은 Table 4.5와 같았다. 분포마다 다르게 추정된 Table 4.2의 $VaR_{0.95}$ 값을 이용하여 유전자 알고리즘 최적해를 탐색한 결과 분포별로 서로 다른 종목들이 포트폴리오에 편입된 것을 확인할 수 있었다. 마찬가지로 목표 수익률과 편입 종목수를 변화시킨 결과, 각 실험별로도 유전자 알고리즘의 최적해인 포트폴리오 편입 종목이 서로 다름을 알 수 있었다.

Table 4.6은 학습 및 테스트 시점의 분포 별 포트폴리오의 평균 수익률이다. 앞서 기술한 바와 같이 테스트 시점은 2개의 고점과 2개의 저점으로 구성되어 있다. 여기서 고점1은 2014년 8월 1일이고 고점2는 2015년 4월 24일이다. 저점1은 2014년 10월 17일이며, 저점2는 2015년 8월 21일이다. 한편 학습 종료 시점은 2013년 12월 27일이다. 테스트 수익률은 2014년 1월부터의 연환산 수익률이며, 학습 수익률은 2003년 1월 1일부터 계산된 연환산 수익률이다. 실험별로 가장 수익률이 높은 시점은 볼드로 표시하였다.

Table 4.5 Portfolios selected from each experiment

Exp. ID	dist.	Stock ID
1	normal	11, 47, 51, 57, 61, 64
	t	11, 47, 51, 61, 64, 66
	GPD	3, 11, 47, 53, 61, 66
2	normal	3, 11, 35, 47, 51, 57, 61, 62, 64, 66
	t	11, 35, 47, 51, 53, 57, 61, 62, 64, 66
	GPD	3, 11, 47, 51, 53, 57, 61, 62, 64, 66
3	normal	3, 10, 11, 14, 15, 35, 47, 51, 57, 61, 62, 64, 65, 66
	t	3, 10, 11, 14, 25, 35, 47, 51, 53, 57, 61, 62, 64, 66
	GPD	3, 4, 10, 11, 35, 47, 51, 53, 57, 60, 61, 62, 64, 66
4	normal	3, 11, 15, 47, 61, 66
	t	3, 11, 47, 53, 61, 66
	GPD	3, 11, 47, 53, 61, 66
5	normal	3, 10, 11, 15, 25, 38, 47, 61, 65, 66
	t	3, 11, 25, 38, 47, 53, 60, 61, 65, 66
	GPD	3, 4, 11, 45, 47, 53, 60, 61, 65, 66
6	normal	3, 10, 11, 15, 18, 25, 28, 38, 45, 47, 61, 63, 65, 66
	t	3, 10, 11, 15, 18, 25, 38, 47, 53, 60, 61, 63, 65, 66
	GPD	3, 4, 10, 11, 18, 25, 45, 47, 53, 60, 61, 63, 65, 66
7	normal	18, 25, 47, 63, 65, 66
	t	18, 25, 47, 53, 60, 65
	GPD	25, 47, 53, 60, 63, 66
8	normal	18, 25, 28, 33, 47, 49, 60, 63, 65, 66
	t	18, 25, 28, 47, 49, 53, 60, 63, 65, 66
	GPD	18, 25, 28, 44, 47, 53, 60, 63, 65, 66
9	normal	11, 18, 21, 25, 28, 30, 33, 45, 47, 49, 60, 63, 65, 66
	t	18, 21, 25, 28, 30, 33, 45, 47, 49, 53, 60, 63, 65, 66
	GPD	4, 18, 25, 28, 30, 44, 45, 47, 49, 53, 60, 63, 65, 66

Table 4.6 Returns of the portfolios

Exp. ID	dist.	learninginterval	high point 1	low point 1	high point 2	low point 2
1	normal	93.53	21.54	-2.33	16.12	0.82
	t	100.76	30.77	0.24	25.62	7.35
	GPD	106.89	38.43	21.24	34.36	12.63
2	normal	89.98	35.19	15.78	23.44	7.23
	t	87.34	31.62	4.78	15.6	1.66
	GPD	89.95	35.02	13.86	22.96	7.31
3	normal	84.29	24.41	7.17	14.77	2.28
	t	87.9	27.6	9.04	15.07	0.51
	GPD	86.44	28.57	11.43	15.52	3.74
4	normal	106.14	29.54	17.08	32.11	12.67
	t	106.89	38.43	21.24	34.36	12.63
	GPD	106.89	38.43	21.24	34.36	12.63
5	normal	101.92	12.62	3.17	12.83	-1.24
	t	106.23	16.61	6.63	14.33	2.38
	GPD	99.56	21.86	9.72	19.35	6.91
6	normal	110.2	7.75	-0.04	9.44	-1.48
	t	103.05	10.02	1.12	8.65	-0.36
	GPD	104.12	13.75	3.6	11.97	1.47
7	normal	124.57	12.82	-6.9	8.5	-2.02
	t	112.11	1.73	-6.07	-0.74	-1.45
	GPD	107.27	22.55	2.66	11.87	-0.27
8	normal	121.33	17.59	6.98	20.1	11.71
	t	120.21	24.4	12.3	18.26	10.58
	GPD	112.22	28.87	21.6	35.19	14.35
9	normal	121.35	13.76	4.84	21.41	11.23
	t	117.11	17.77	5.32	17.52	8.55
	GPD	106.02	27.23	20.39	33.15	13.85

목표 수익률과 편입 종목수에 따른 영향을 확인하기 위해 실험 1~9를 분포별로 다시 정리한 Table은 아래와 같다 (Table 4.7~Table 4.9). 당연한 결과겠지만, 학습시점의 수익률은 $Target_R$ 을 80%분위로 잡았을 때가 세 분포 모두 높은 값을 보였고 정규분포가 학습 시점의 수익률은 전반적으로 높게 나타났고 GPD가 가장 낮게 나타났다.

Table 4.7 Experiment results (normal distribution case)

quantile	num. of items	learning interval	high point 1 (A)	low point 1 (B)	diff. of A and B	high point 2 (C)	low point 2 (D)	diff. of C and D
50%	6	93.53	21.54	-2.33	23.87	16.12	0.82	15.3
	10	89.98	35.19	15.78	19.41	23.44	7.23	16.21
	14	84.29	24.41	7.17	17.24	14.77	2.28	12.49
	avg.	89.27	27.05	6.87	20.17	18.11	3.44	14.67
65%	6	106.14	29.54	17.08	12.46	32.11	12.67	19.44
	10	101.92	12.62	3.17	9.45	12.83	-1.24	14.07
	14	110.2	7.75	-0.04	7.79	9.44	-1.48	10.92
	avg.	106.09	16.64	6.74	9.9	18.13	3.32	14.81
80%	6	124.57	12.82	-6.9	19.72	8.5	-2.02	10.52
	10	121.33	17.59	6.98	10.61	20.1	11.71	8.39
	14	121.35	13.76	4.84	8.92	21.41	11.23	10.18
	avg.	122.42	14.72	1.64	13.08	16.67	6.97	9.7
overall avg.		105.92	19.47	5.08	14.39	17.64	4.58	13.06

Table 4.8 Experiment results (t-distribution case)

quantile	num. of items	learning interval	high point 1 (A)	low point 1 (B)	diff. of A and B	high point 2 (C)	low point 2 (D)	diff. of C and D
50%	6	100.76	30.77	0.24	30.53	25.62	7.35	18.27
	10	87.34	31.62	4.78	26.84	15.6	1.66	13.94
	14	87.9	27.6	9.04	18.56	15.07	0.51	14.56
	avg.	92	30	4.69	25.31	18.76	3.17	15.59
65%	6	106.89	38.43	21.24	17.19	34.36	12.63	21.73
	10	106.23	16.61	6.63	9.98	14.33	2.38	11.95
	14	103.05	10.02	1.12	8.9	8.65	-0.36	9.01
	avg.	105.39	21.69	9.66	12.02	19.11	4.88	14.23
80%	6	112.11	1.73	-6.07	7.8	-0.74	-1.45	0.71
	10	120.21	24.4	12.3	12.1	18.26	10.58	7.68
	14	117.11	17.77	5.32	12.45	17.52	8.55	8.97
	avg.	116.48	14.63	3.85	10.78	11.68	5.89	5.79
overall avg.		104.62	22.11	6.07	16.04	16.52	4.65	11.87

반면, 테스트 시점에서의 수익률은 $Target_R$ 과 관계가 없는 것으로 나타났으며, 오히려 50%분위를 대상으로 하였을 때의 수익률이 가장 높게 나타났다. 이는 선택 대상 종목이 많아진 데 기인한 것이다. 즉, 학습 당시에는 수익률이 높지 않았으나 테스트 시점에서 수익률이 높게 나왔고 Var_α 가 다른 종목에 비해 낮았던 것이다. 전반적으로 GPD가 거의 모든 테스트 시점에서 높은 결과를 보였다. t분포는 정규분포보다 근소하게 좋은 결과를 보였으나 정규분포보다 적합하다고 말하기는 힘들었다. 한편 등락 폭은 포트폴리오 편입종목수가 많을수록 낮게 나타나서 분산투자의 위험 회피 성향을 잘 보여 주었다. 또한 위험이 극대화 되는 저점에서의 수익률이 GPD와 t분포가 정규분포보다 높은 것으로 나타나 두꺼운 꼬리분포가 위험 관리에 더 적합함을 보여주었다.

Table 4.9 Experiment results (GPD case)

quantile	num. of items	learning interval	high point 1 (A)	low point 1 (B)	diff. of A and B	high point 2 (C)	low point 2 (D)	diff. of C and D
50%	6	106.89	38.43	21.24	17.19	34.36	12.63	21.73
	10	89.95	35.02	13.86	21.16	22.96	7.31	15.65
	14	86.44	28.57	11.43	17.14	15.52	3.74	11.78
	avg.	94.43	34.01	15.51	18.5	24.28	7.89	16.39
65%	6	106.89	38.43	21.24	17.19	34.36	12.63	21.73
	10	99.56	21.86	9.72	12.14	19.35	6.91	12.44
	14	104.12	13.75	3.6	10.15	11.97	1.47	10.5
	avg.	103.52	24.68	11.52	13.16	21.89	7	14.89
80%	6	117.99	22.55	2.66	19.89	11.87	-0.27	12.14
	10	112.22	28.87	21.6	7.27	35.19	14.35	20.84
	14	106.02	27.23	20.39	6.84	33.15	13.85	19.3
	avg.	112.08	26.22	14.88	11.33	26.74	9.31	17.43
overall avg.		103.34	28.3	13.97	14.33	24.3	8.07	16.24

Table 4.10은 포트폴리오의 VaR값 초과 정도를 나타낸다. 세 분포 모두 신뢰수준 대비 적절한 초과율을 보여 주고 있으며, 목표수익률 분위가 높아질수록 초과정도가 낮음을 알 수 있다. 한편, Figure 4.3은 포트폴리오 선택과 관련한 세대별 적합도 값의 보여주는 그림이다. 모든 실험에서 유사한 결과를 보여준다.

Table 4.10 Incidence rate exceeding 95% VaR

quantile	num. of items	normal dist.		t dist.		GPD	
		learning	test	learning	test	learning	test
50%	6	4.21	3.45	4.44	4.02	4.62	3.64
	10	3.92	3.22	4.3	3.45	4.62	3.56
	14	3.96	2.46	4.38	2.79	4.68	3.12
65%	6	3.92	2.68	4.07	3.07	4.62	3.64
	10	3.85	1.84	4.08	2.53	4.69	2.64
	14	3.68	1.4	4.14	1.89	4.7	1.97
80%	6	3.34	1.72	3.77	0.96	4.65	2.49
	10	3.45	1.61	3.9	1.61	4.7	1.84
	14	3.65	1.56	4.04	1.48	4.73	1.89

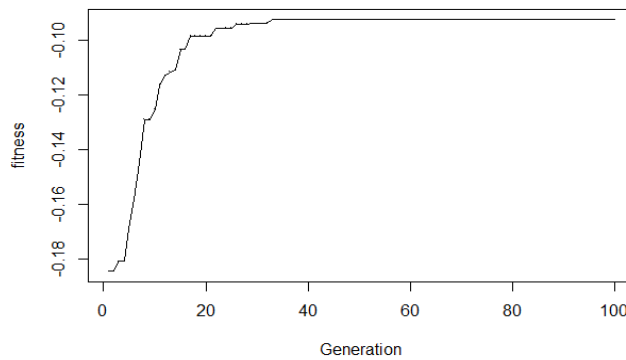


Figure 4.3 Variation of fitness value (in case of GPD with $Target_R = 80\%$ and $Sel = 14$)

Table 4.11은 실험대상 종목표이다.

Table 4.11 List of stocks used in experiments

ItemID	Market	Code	Name	ItemID	Market	Code	Name
1	KOSPI	138930	BNK금융지주	34	KOSPI	006400	삼성SDI
2	KOSPI	001040	CJ	35	KOSPI	000830	삼성물산
3	KOSPI	000120	CJ대한통운	36	KOSPI	009150	삼성전기
4	KOSDAQ	035760	CJ오쇼핑	37	KOSPI	005930	삼성전자
5	KOSPI	139130	DGB금융지주	38	KOSPI	010140	삼성중공업
6	KOSDAQ	028150	GS홈쇼핑	39	KOSPI	016360	삼성증권
7	KOSPI	105560	KB금융	40	KOSPI	000810	삼성화재
8	KOSPI	002380	KCC	41	KOSPI	004170	신세계
9	KOSPI	033780	KT&G	42	KOSPI	055550	신한지주
10	KOSPI	003550	LG	43	KOSPI	002790	아모레G
11	KOSPI	051900	LG생활건강	44	KOSPI	010780	아이에스동서
12	KOSPI	032640	LG유플러스	45	KOSDAQ	056190	에스에프에이
13	KOSPI	066570	LG전자	46	KOSPI	012750	에스원
14	KOSPI	051910	LG화학	47	KOSPI	001800	오리온
15	KOSPI	002550	KB손해보험	48	KOSPI	000100	유한양행
16	KOSPI	005490	POSCO	49	KOSDAQ	039030	이오테크닉스
17	KOSPI	010950	S-Oil	50	KOSPI	030000	제일기획
18	KOSPI	003600	SK	51	KOSPI	021240	코웨이
19	KOSPI	017670	SK텔레콤	52	KOSDAQ	034230	파라다이스
20	KOSPI	000660	SK하이닉스	53	KOSPI	086790	하나금융지주
21	KOSPI	010130	고려아연	54	KOSPI	036460	한국가스공사
22	KOSPI	011780	금호석유	55	KOSPI	071050	한국금융지주
23	KOSPI	000270	기아차	56	KOSPI	015760	한국전력
24	KOSDAQ	035720	다음카카오	57	KOSPI	018880	한온시스템
25	KOSPI	047050	대우인터내셔널	58	KOSPI	009240	한샘
26	KOSPI	042660	대우조선해양	59	KOSPI	000720	현대건설
27	KOSPI	006800	대우증권	60	KOSPI	005440	현대그린푸드
28	KOSPI	005830	동부화재	61	KOSPI	012330	현대모비스
29	KOSDAQ	026960	동서	62	KOSPI	069960	현대백화점
30	KOSPI	000150	두산	63	KOSPI	004020	현대제철
31	KOSPI	034020	두산중공업	64	KOSPI	005380	현대차
32	KOSPI	005300	롯데철성	65	KOSPI	001450	현대해상
33	KOSPI	011170	롯데케미칼	66	KOSPI	008770	호텔신라

5. 결론

본 논문에서는 위험자산의 확률분포를 GPD로 추정하고 유전자 알고리즘을 이용해 최적의 포트폴리오를 탐색하는 방법을 제안하였다. 먼저, 포트폴리오의 적합성을 판단하기 위하여 수익률 척도 및 위험 척도를 결정하였는데, 일정기간 동안의 등락률의 합인 누적등락률을 수익률 척도로 하였고, 특정 신뢰수준 하에서 발생할 수 있는 포트폴리오에 최대손실액을 의미하는 VaR를 위험의 척도로 정의 하였다. VaR는 정규분포, t분포, GPD로 위험자산의 확률분포를 추정하여 95% 신뢰수준 하에서 각각 산출하였다. 한편, GPD의 경우, 임계값의 결정이 필요한데, 본 논문에서는 유전자 알고리즘을 이용하여 임계값을 결정하고 그 임계값을 이용하여 최적의 포트폴리오를 탐색하도록 2단계의 구조의 방법을 제안하였다.

제안한 방법의 적정성을 위하여 국내 증시에서 최적의 포트폴리오를 탐색해 보았고, 그 결과 테스트 시점에서는 GPD를 이용하였을 때가 가장 높은 수익률을 보이는 등 적합함을 알 수 있었다. 테스트 시점은 위험 보다 수익률이 더 강조되는 고점과 수익률 보다는 위험의 추정이 더 강조되는 저점에서 각각

진행하였다. 실험 전 예상으로는 드물게 발생하는 극단적인 사건들을 잘 다룰 수 있는 극단치 이론의 특성상 증시 저점에서는 GPD가 가장 좋은 결과를 보일 것이고 고점에서는 정규분포와 t분포가 더 높은 수익률을 보일 것으로 생각되었다. 그러나 저점뿐만 아니라, 고점에서도 GPD를 이용하였을 때가 더 높은 수익률을 보였다. 이는 실험대상을 KRX100 우량주로 한정된 것과 Figure 4.1, 4.2에서 보듯 테스트 구간에서의 주가가 전반적으로 하락하는 추세라는 점에 기인한 것으로 보인다. 우량주는 일반적으로 주식시장을 선도하는 수익률을 보이면서 위험률도 더 낮은 특성이 있다. 이에 GPD가 우량주의 위험을 더 잘 예측하였기 때문에 고점에서도 더 높은 수익률을 보일 수 있었던 또 다른 이유로 추정된다. 한편, 학습 데이터에 대해서는 정규분포가 가장 적합한 것으로 나타났고 t분포도 GPD 보다 높게 나타났다. 이는 학습 데이터에 전체 데이터 중 등락률이 가장 높았던 2008년 금융위기가 포함되어 이미 주식시장의 위험을 적절하게 반영하였기 때문으로 판단된다. 즉, 위험이 충분히 반영되었을 때는 정규분포를 통해 과거자료를 모델링하는 것이 효과적일 수 있다는 것이다. 이는 여타의 연구에서 정규분포가 아직 많이 사용되는 이유가 될 것이다.

본 연구에서는 GPD가 예상보다 만족할 만한 결과를 보였으나, 좀 더 실제에 가까운 시뮬레이션을 위해 향후 보완해야할 사항들도 있다. 먼저 수익률 산출방법을 매수 후 보유 전략 외 다른 전략을 고안해서 적용할 필요가 있으며, 포트폴리오의 편입과 관련하여서는 편입 여부 0과 1이 아닌 비율형태로 확장할 필요도 있을 것이다. 한편, 위험뿐만 아니라 수익률 척도도 분포로부터 추정할 값을 이용할 수 있을 것이다. 또한, 우량 종목에서만 아니라 비우량 종목의 경우의 고점에서도, 그리고 상승시점의 주식시장에서도 여전히 GPD가 우수한 성능을 보이는 지 확인해 볼 필요가 있을 것이다.

References

- Alexander, G. J. and Baptista, A. M. (2002). Economic implications of using a mean-VaR model for portfolio selection: a comparison with mean-variance analysis. *Journal of Economic Dynamics and Control*, **26**, 1159-1198
- Anione, S., Loraschi, A. and Tettamanzi, A. (1993). A genetic approach to portfolio selection. *Neural Network World*, **6**, 597-604.
- Balkema, A. A. and De Haan, L. (1974). Residual life time at great age. *The Annals of probability*, **2**, 792-804.
- Bridges, C. L. and Goldberg, D. E. (1987). An analysis of reproduction and crossover in a binary-coded genetic algorithm. *Grefenstette*, **878**, 9-13.
- Byun, H. W., Song, C. W., Han, S. K., Lee, T. K. and Oh, K. J. (2009). Using genetic algorithm to optimize rough set strategy in KOSPI200 futures market. *Journal of the Korean Data & Information Science Society*, **20**, 1049-1060.
- Chambers, L. D. (1995). *Practical Handbook of Genetic Algorithms*, CRC Press, Florida.
- Chung, S. H. and Oh, K. J. (2014). Using genetic algorithm to optimize rough set strategy in KOSPI200 futures market. *Journal of the Korean Data & Information Science Society*, **25**, 281-292.
- Davison, A. and R. Smith. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society*, **52**, 393-442.
- Eberhart, R., Simpson, P. and Dobbins, R. (1996). *Computational intelligence PC tools*, Academic Press Professional, San Diego.
- Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997). *Modelling extremal events for Insurance and Finance*, Springer, New York.
- Golberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Addison wesley, Boston.
- Lin, P. C. and Ko, P. C. (2009). Portfolio value-at-risk forecasting with GA-based extreme value theory. *Expert Systems with Applications*, **36**, 2503-2512.
- Longin, F. M. (2000). From value at risk to stress testing: The extreme value approach. *Journal of Banking and Finance*, **24**, 1097-1130.
- Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance*, **7**, 77-91.

- Marsili, M., Maslov, S. and Zhang, Y. C. (1998). Dynamical optimization theory of a diversified portfolio. *Physica A: Statistical Mechanics and its Applications*, **253**, 403-418.
- Oh, K. J., Kim, T. Y., Min, S. H. and Lee, H. Y. (2006). Portfolio algorithm based on portfolio beta using genetic algorithm. *Expert Systems with Applications*, **30**, 527-534.
- Oh, S. K. (2005). Oh, S. K., Extreme Value Theory and Value at Risk focusing on GPD Models. *Journal of Money and Finance*, **19**, 72-114.
- Pickands III, J. (1975). Statistical inference using extreme order statistics. *the Annals of Statistics*, **3**, 119-131.
- Rankovi, V., Drenovak, M., Stojanovi, B., Kalini, Z. and Arsovski, Z. (2014). The mean-Value at Risk static portfolio optimization using genetic algorithm. *Computer Science and Information Systems*, **11**, 89-109.
- Scrucca, L. (2014). GA: a package for genetic algorithms in R. *Journal of Statistical Software*, **53**, 1-37.
- Shoaf, J. and Foster, J. (1998). The efficient set GA for stock portfolios. In *Proceedings of the 1998 IEEE international conference on computational intelligence*, 354-359, IEEE Service Center, New Jersey.
- Statman, M. (1987). How many stocks make a diversified portfolio?. *Journal of Financial and Quantitative Analysis*, **22**, 353-363.

Finding optimal portfolio based on genetic algorithm with generalized Pareto distribution[†]

Hyundon Kim¹ · Hyun Tae Kim²

^{1,2}Department of Applied Statistics, Yonsei University

Received 30 October 2015, revised 23 November 2015, accepted 24 November 2015

Abstract

Since the Markowitz's mean-variance framework for portfolio analysis, the topic of portfolio optimization has been an important topic in finance. Traditional approaches focus on maximizing the expected return of the portfolio while minimizing its variance, assuming that risky asset returns are normally distributed. The normality assumption however has widely been criticized as actual stock price distributions exhibit much heavier tails as well as asymmetry. To this extent, in this paper we employ the genetic algorithm to find the optimal portfolio under the Value-at-Risk (VaR) constraint, where the tail of risky assets are modeled with the generalized Pareto distribution (GPD), the standard distribution for exceedances in extreme value theory. An empirical study using Korean stock prices shows that the performance of the proposed method is efficient and better than alternative methods.

Keywords: Extreme value theory, genetic algorithm, GPD, portfolio optimization, VaR.

[†] This research was supported by Basic Research Program through the National Research Foundation of Korea (NRF 2015R1A1A1A05027336).

¹ Master, Department of Applied Statistics, Yonsei University, Seoul 03722, Korea.

² Corresponding author: Professor, Department of Applied Statistics, Yonsei University, Seoul 03722, Korea. E-mail: jhtkim@yonsei.ac.kr