

이론통계학2

## Project #5. Optimal Portfolio 계산

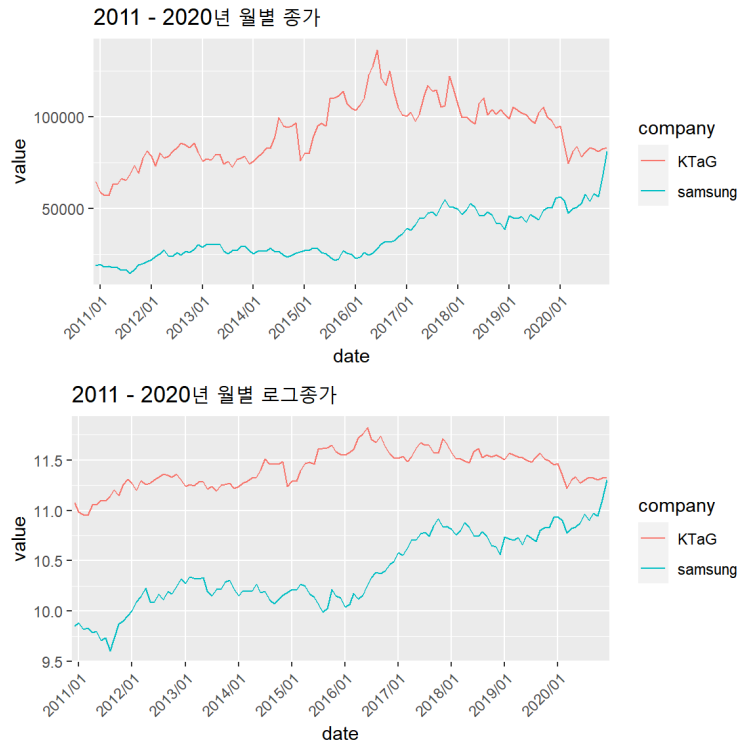
발표일 : 2021. 10. 13

1조

202STG26	박지윤
202STG27	이수현
212STG04	김이현
212STG12	박윤정

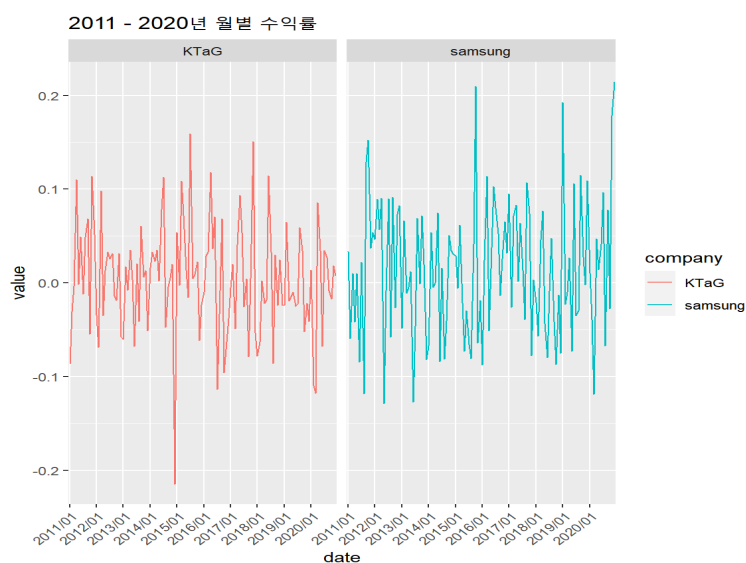
## Part 1 : 주식투자 Portfolio 만들기

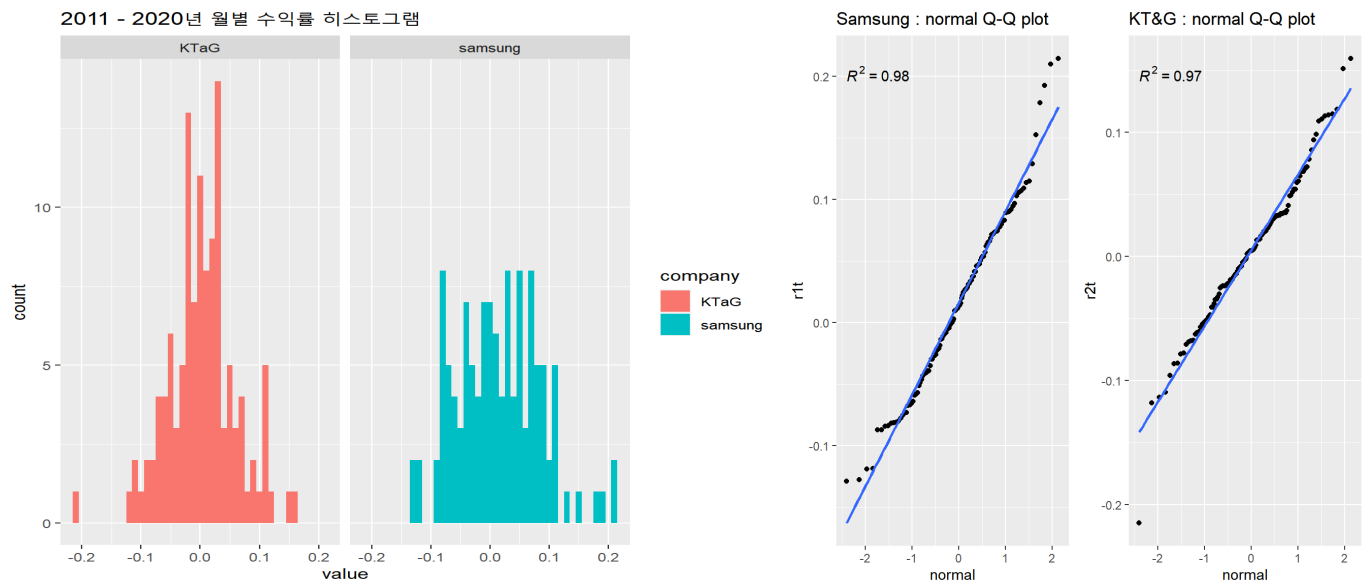
1. 각 종목의 2011-2020년 월별 종가 및 로그종가에 대한 시계열 도표를 그리시오.



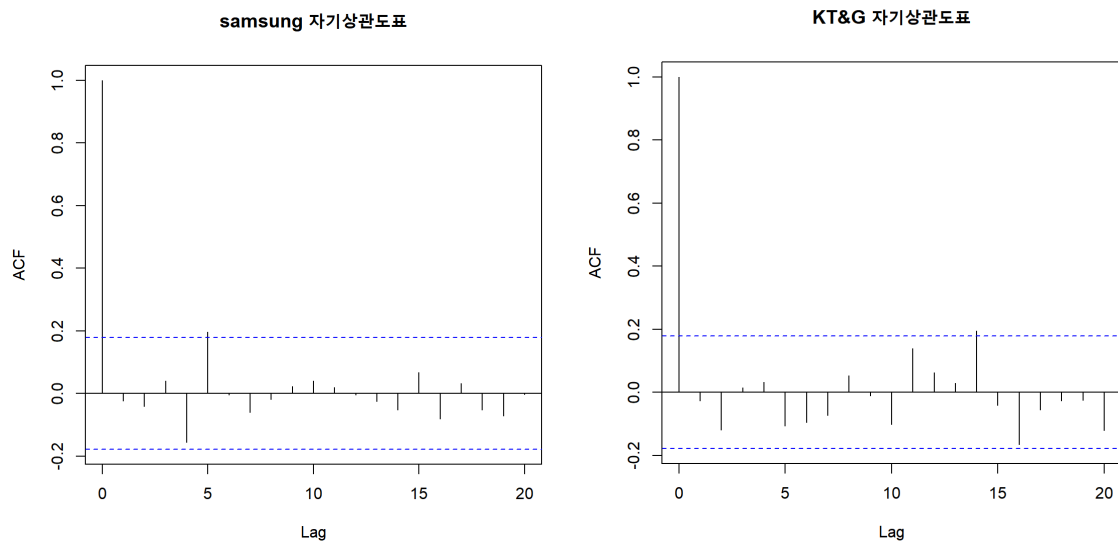
2011-2020년 월별 종가, 월별 로그 종가의 시계열 도표이다. KT&G가 삼성전자보다 높은 값을 가진다. 삼성전자는 2018년 5월 4일 액면분할을 하기 전에는 1주에 265만원대였으나, 액면 분할 이후의 가격대로 조정된 데이터를 활용하였다. (출처 : Yahoo Finance) 로그 종가 그래프를 보면, 10년간 삼성전자의 주가는 꾸준히 상승했으나, KT&G의 10년간 큰 차이가 없었음을 알 수 있다.

2. 개별 종목의 2011-2020년 월별 수익률  $\{r_{it}\}$  자료의 시계열도표, Histogram, Q-Q plot, 자기상관도표를 각각 그려 보고 정규성(normality) 및 독립성 가정을 검토해 보시오.





KT&G의 경우 히스토그램에서 어느정도 종모양의 형태를 띠고 있으므로 정규성을 만족할 것임을 짐작할 수 있었다. Q-Q plot의  $R^2$ 값 또한 0.97로 높았고 점들이 비교적 직선 위에 존재하므로 정규성을 만족한다고 판단했다. 삼성전자의 경우 히스토그램에서는 정규분포를 만족하는지 확신할 수 없었으나 Q-Q plot에서는  $R^2=0.98$ 로 높은  $R^2$ 값을 가졌다. 또한 점들이 대체로 직선에 분포하기 때문에 정규성 가정을 만족한다고 결론 내릴 수 있다.



Lag=5, Lag=14 근처에서 점선을 벗어나기는 하나 각각의 경우를 제외하고는 점선 안에  $\rho_i$  값들이 존재하기 때문에 독립성 가정을 만족한다고 판단했다.

3. 학습자료 (2011-2015) 및 검정자료(2016-2020)에 대해서  $\{r_{it}\}$ 의 평균, 표준편차, 상관계수 ( $\hat{u}_i, \hat{\sigma}_i, \hat{\rho}_i$ ) 및 공분산 행렬  $\Sigma = (\sigma_{ij})$ 을 각각 구하고 그 차이를 비교해 보시오.

	Data	$\hat{u}$	$\hat{\sigma}$	$\hat{\rho}$
Samsung	Train	0.007	0.07	-0.148
	Test	0.022	0.0723	
KT&G	Train	0.0098	0.059	0.121
	Test	-0.002	0.058	

2011-2015년 데이터를 학습자료로, 2016-2020년 데이터를 검정자료로 사용했을 때의 삼성전자와 KT&G의 월별 수익률의

평균, 표준편차, 상관계수이다. 두 경우 모두 학습자료와 검정자료의 표준편차에는 큰 차이가 없지만 평균은 차이를 보인다. KT&G의 경우 train data의 평균은 양수, test data의 평균은 음수인데 이는 시계열 도표에서 확인할 수 있듯이 주가가 2016년 전에는 계속 증가하다가 후로는 감소하기 때문이다.

[학습자료의 공분산 행렬]

	Samsung	KT&G
Samsung	0.0049	-0.0002
KT&G	-0.0002	0.0035

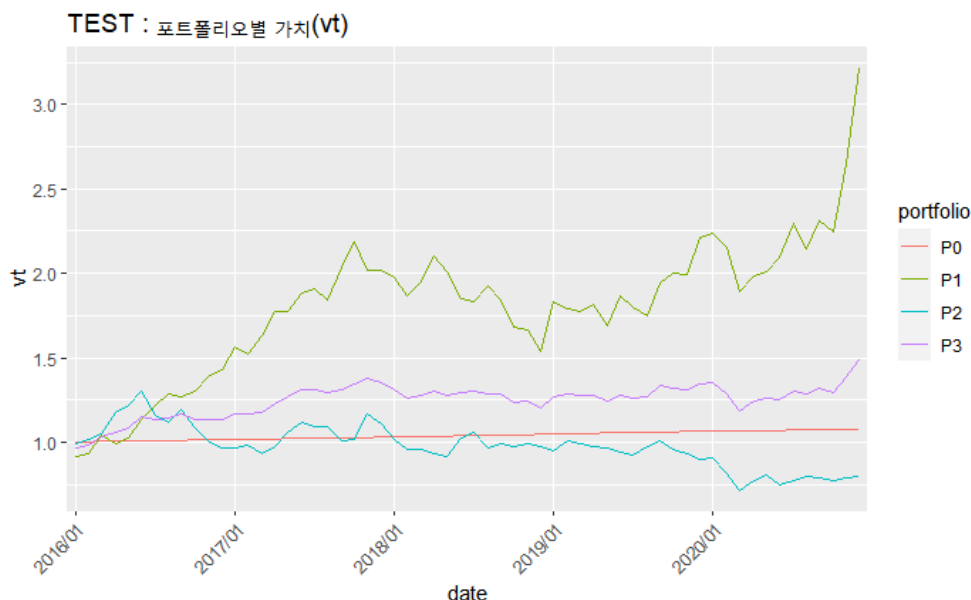
[검정자료의 공분산 행렬]

	Samsung	KT&G
Samsung	0.0053	-0.0003
KT&G	-0.0003	0.0034

삼성전자와 KT&G간의 학습자료와 검정자료의 공분산 행렬이다. 검정자료와 학습자료에서의 분산의 차이는 거의 없는 것으로 보인다. 두 종목 간의 공분산은 모두 음수이며, 검정자료에서의 절대값이 더 크다. 또한 분산의 경우 삼성전자가 KT&G보다 더 크다.

4. 2016년 1월초에 총  $V_0=1$  (예를 들어 1천만원)을 정기예금, 삼성전자, KT&G 주식에 각각  $(\theta_0, \theta_1, \theta_2)$  비율로 분산투자했을 때 미래  $t$  개월 후 각 Portfolio의 가치(Value)  $\{V_t\}$ 의 시계열 도표를 동일한 그래프에 서로 적절히 겹쳐서 그려보고 각 Portfolio의 수익률의 장단점을 상호 비교 설명하시오.

$$\begin{aligned} P_0 : (\theta_0, \theta_1, \theta_2) &= (1, 0, 0) \\ P_1 : (\theta_0, \theta_1, \theta_2) &= (0, 1, 0) \\ P_2 : (\theta_0, \theta_1, \theta_2) &= (0, 0, 1) \\ P_3 : (\theta_0, \theta_1, \theta_2) &= (1/3, 1/3, 1/3) \end{aligned}$$

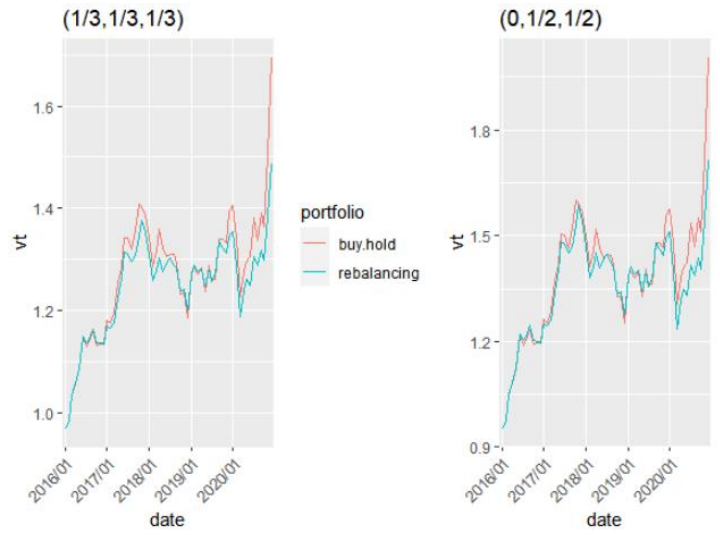


주어진 비율별로 분산 투자했을 때의  $t$ 개월 후의 포트폴리오 가치 그래프이다. 가장 수익률이 높을 것으로 기대되는 경우는 삼성전자에 전액을 투자하는 P1의 경우이다. 삼성전자의 경우 주가가 계속 상승세를 유지하고 있고, 그래프에서도 무위험 이자율보다 수익이 큼을 확인할 수 있기 때문에 해당 회사에 전액을 투자할 경우 큰 이익을 볼 수 있을 것이라 기대된다. 가장 손해가 큰 경우는 KT&G에 전액을 투자하는 P2의 경우로, 해당 회사의 주가는 2016년 중반부터 감소하기 시작하므로 이 회사에 전액을 투자할 경우는 큰 손해를 볼 것임이 예상된다. 삼성전자와 KT&G, 정기예금에 각각 1/3씩 투자하는 P3의 경우 또한 무위험 이자율보다 수익이 큰 시나리오이기 때문에 적절한 이익을 볼 수 있을 것이라 예상된다.

5. 아래 두 Portfolio 에 대해서 a) buy & hold 와 b) rebalancing 경우의 수익률,  $V_t^*$ ,  $V_t$ 의 시계열도표를 서로 겹쳐서 그려보고 그 차이를 설명하시오.

$$P3 : (\theta_0, \theta_1, \theta_2) = (1/3, 1/3, 1/3)$$

$$P4 : (\theta_0, \theta_1, \theta_2) = (0, 1/2, 1/2)$$



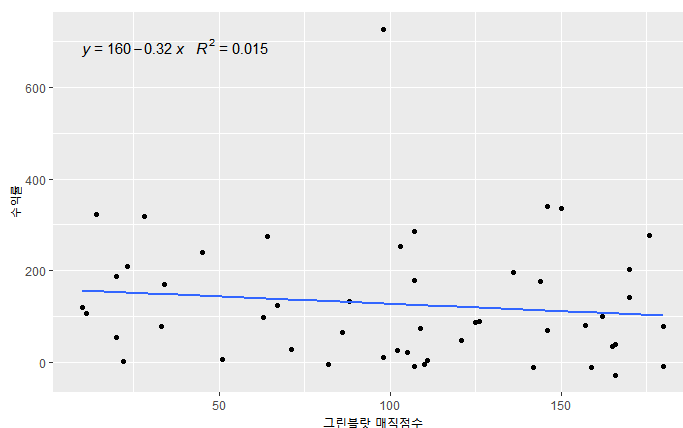
빨간색 그래프는 처음 투자 비율을 그대로 방치한 no rebalancing(buy and hold), 파란색 그래프는 매월 말 포트폴리오를 재조정하는 rebalancing의 경우로 구한 수익률 그래프이다. 두 경우 모두, buy and hold 방법이 rebalancing보다 조금 더 수익률이 큰 것을 확인할 수 있다. 이는 2020년도부터 차이가 조금씩 커져 2020년 말에는 수익률이 비교적 큰 차이를 보인다.

또한 portfolio3과 4를 비교해보면, 정기예금을 제외하고, 삼성전자와 KT&G에 50%씩 투자한 4번의 수익률 상승폭이 좀 큰 것을 확인할 수 있다.

## Part 2 : 재무변수를 이용한 주가수익률 예측

1. 수익률과 그린블랏의 매직점수 (ROA순위+PER순위)의 산점도를 그리고 최적 예측방정식을 구하시오. 또한 두 변수의 상관계수 및 순위상관계수를 구하고 그 의미를 설명하시오.

최적 예측 방정식	상관계수	순위상관계수
$y = 160 - 0.32x$	-0.124	-0.167

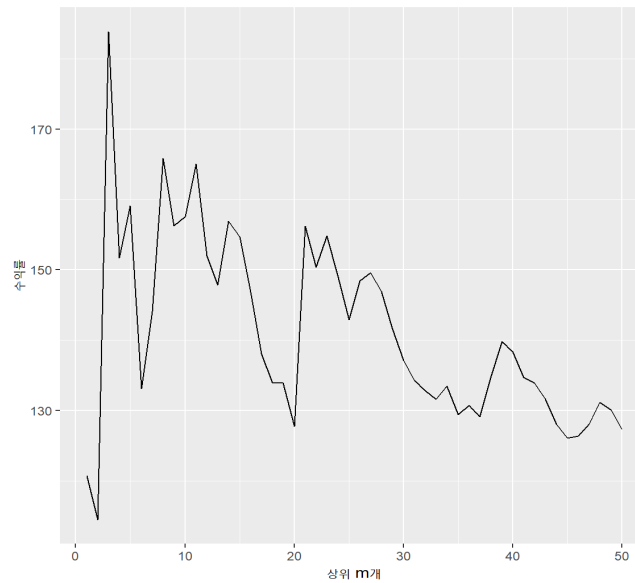


수익률과 그린블랏 매직점수의 최적 예측 방정식은

$$\text{수익률} = 160 - 0.32 * \text{그린블랏 매직점수}$$

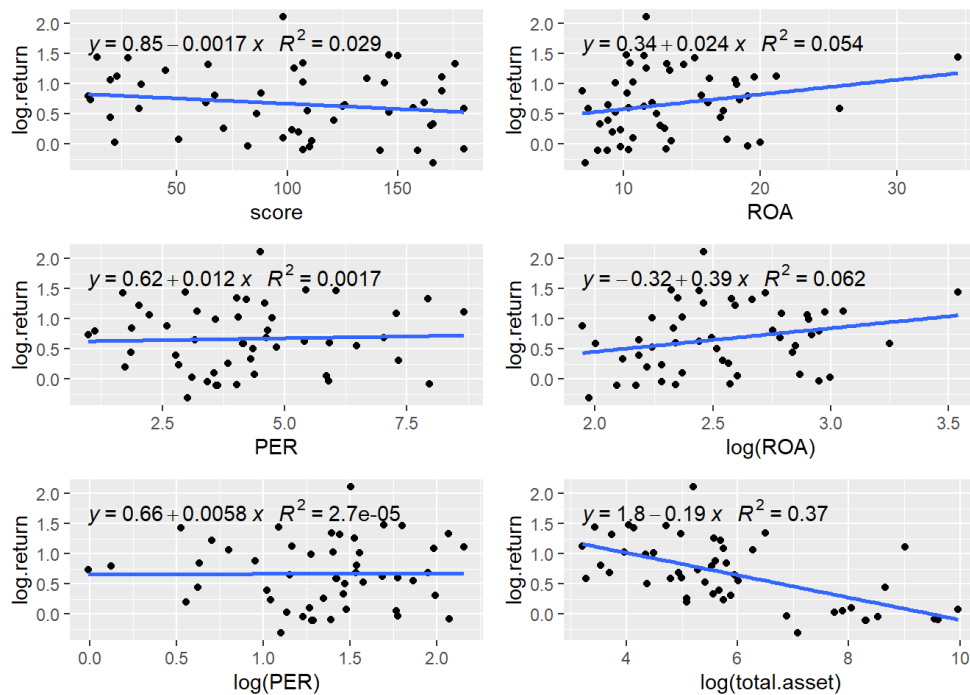
로 추정되었다. 기울기와 상관계수가 모두 음수이므로 매직점수가 작을수록 수익률이 큰 것을 알 수 있다.

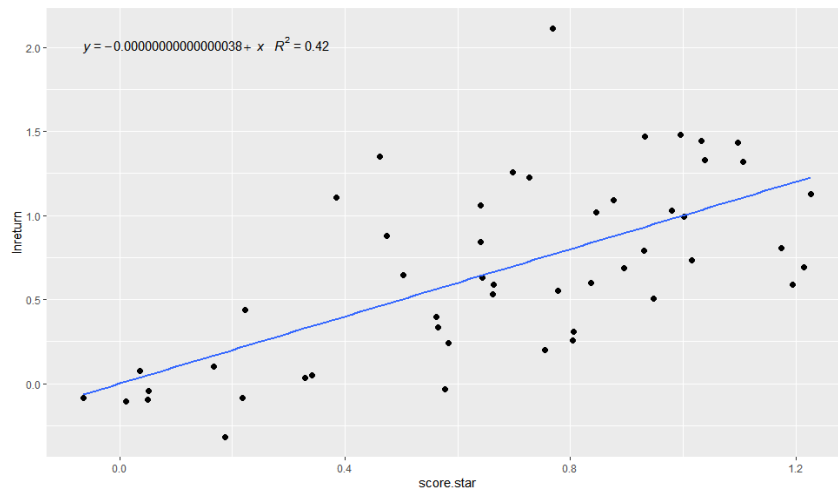
2. 매직점수 상위  $m$  개의 종목에  $1/m$ 씩 분산투자한 포트폴리오의 수익률을  $m=1,2, \dots, 50$  일 때 각각 구하여 이를 그래프로 나타내고 그 의미를 설명하시오.



$m$ 이 커질수록 수익률이 감소하는 경향이 있다. 하지만 경향성에서 튀는 점들도 존재하는데 이는 매직점수는 크지만 수익률이 더 큰 경우들이 존재하기 때문이다. 또한 모든 비율을  $1/m$ 으로 설정했기 때문에, 더 적절한 분산 투자 비율을 찾는 것이 중요하다는 것을 알 수 있다.

3. 재무변수 (PER, ROA, Total Asset)들을 적절히 변환하여 로그수익률 ( $\ln(1+r)$ )을 예측 하는 최적 회귀모형을 찾고 이를 a) 에서 구한 결과와 서로 비교하여 새 예측공식의 의미를 설명하고 이들 결과의 의미를 주식시장의 효율적 시장가설과 연관하여 해석하시오.





$$Score^* = E(\ln(1+r_i)) = 1.83 + 0.184*PER - 0.608*\ln(PER) - 0.191*\ln(TA)$$

PER, ROA, Total Asset 변수의 다양한 변환을 통해 최적 예측 방정식을 도출했다.  $R^2$ 가 가장 큰 경우는  $Score^*$ 변수를 사용했을 때였으며 이때 사용한  $Score^*$ 는 위와 같이 정의된다. 각 종목에서 파생한 다양한 예측 변수들을 이용해 가장 최적화된 모형을 구하고 이를 통해 구한 로그 수익률의 예측값이다. 이때 도출된 최적 예측 방정식은

$$\ln(1+r) = Score^*$$

이다. 주식 시장의 효율적 시장 가설이란, 가격은 “상품에 대해 얻을 수 있는 모든 정보 (All available information)를 빠르게 반영하며 따라서 그 정보들을 이용하여 장기적으로 시장 수익률을 넘을 수 없다”는 가설이다. 즉, 현재 활용 가능한 모든 정보는 이미 현재 주가에 반영되어, 주가를 예측할 수 없다고 본다. 앞서 구한 최적 예측 방정식은 유용한 정보를 최대한 활용한  $Score^*$  변수를 이용해 구한 방정식으로,  $score^*$ 값이 로그수익률과 동일하다는 것으로 보아, 효율적 시장 가설을 뒷받침해준다고 볼 수 있다.

4. 최적예측모형에서 구한 수익률의 추정값(fits) 순위 상위 m 개의 종목에 1/m씩 분산투자한 포트폴리오의 수익률을  $m=1,2, \dots, 50$  일 때 각각 구하여 이를 b)에서 구한 그래프와 겹쳐서 그리고 그 의미를 설명하시오.



두 경우 모두 m이 커질수록 수익률이 감소하는 경향을 보이나, B의 예측모형에서 구한 수익률은 d에 비해서 비교적으로 평이하게 감소하였다. 또한 대체로 d의 수익률이 b의 수익률보다 높은 경향성을 보이는데, 이는 최적 예측 모형에서 구한 수익률의 순위로 선택한 종목들이기 때문에 그런 것으로 해석할 수 있다.

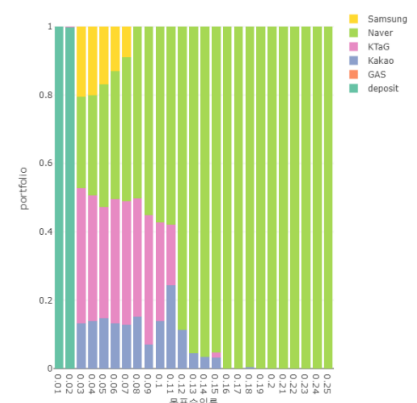
## Part 3 : Optimal Portfolio 찾는 Application 개발 및 최적 포트폴리오 의미 해석

1. Optimal Portfolio Table을 만들고 표의 의미를 설명해 보시오.

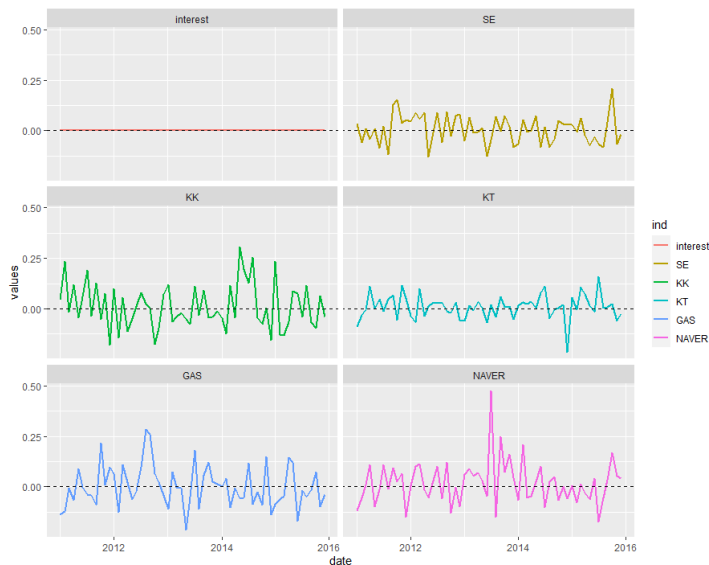
목표	예금	삼성전자	KT&G	KAKAO	NAVER	GAS	MEAN	SD	Sharpe
0.01	1.00	0.00	0.00	0.00	0.00	0.00	0.027	0.002	2.75
0.02	1.00	0.00	0.00	0.00	0.00	0.00	0.027	0.002	1.17
0.03	0.00	0.20	0.40	0.13	0.27	0.00	0.147	0.164	0.21
0.04	0.00	0.20	0.37	0.14	0.29	0.00	0.151	0.169	0.19
0.05	0.00	0.17	0.33	0.15	0.36	0.00	0.159	0.183	0.17
0.06	0.00	0.13	0.36	0.13	0.37	0.00	0.162	0.187	0.16
0.07	0.00	0.09	0.36	0.13	0.42	0.00	0.169	0.200	0.15
0.08	0.00	0.00	0.35	0.15	0.50	0.00	0.182	0.227	0.13
0.09	0.00	0.00	0.38	0.07	0.55	0.00	0.185	0.234	0.12
0.1	0.00	0.00	0.29	0.14	0.57	0.00	0.190	0.244	0.11
0.11	0.00	0.00	0.18	0.24	0.58	0.00	0.194	0.258	0.10
0.12	0.00	0.00	0.00	0.11	0.89	0.00	0.226	0.340	0.10
0.13	0.00	0.00	0.00	0.04	0.96	0.00	0.232	0.358	0.09
0.14	0.00	0.00	0.00	0.04	0.97	0.00	0.233	0.361	0.08
0.15	0.00	0.00	0.01	0.03	0.95	0.00	0.231	0.357	0.07
0.16	0.00	0.00	0.00	0.00	1.00	0.00	0.236	0.371	0.07
0.17	0.00	0.00	0.00	0.00	1.00	0.00	0.236	0.371	0.06
0.18	0.00	0.00	0.00	0.00	1.00	0.00	0.235	0.370	0.05
0.19	0.00	0.00	0.00	0.00	1.00	0.00	0.236	0.371	0.05
0.2	0.00	0.00	0.00	0.00	1.00	0.00	0.236	0.371	0.04
0.21	0.00	0.00	0.00	0.00	1.00	0.00	0.236	0.371	0.03
0.22	0.00	0.00	0.00	0.00	1.00	0.00	0.236	0.371	0.03
0.23	0.00	0.00	0.00	0.00	1.00	0.00	0.236	0.371	0.02
0.24	0.00	0.00	0.00	0.00	1.00	0.00	0.236	0.371	0.02
0.25	0.00	0.00	0.00	0.00	1.00	0.00	0.236	0.371	0.01

목표 수익률이 낮으면 정기에금에 전부 투자하는 것이 가장 적절하나, 목표 수익률이 높아질수록 정기에금보다는 다른 회사의 주식에 분배해 투자하는 것이 적절함을 알 수 있다. 또한 아주 높은 수익률을 원한다면 네이버에 전액을 투자하는 것이 적절한 것으로 보인다.

네이버(주)는 2002년 10월 코스닥시장에 NHN(주)로 상장하였으며, 이후 2008년 11월 코스피시장으로 이전 상장하였고, 기업 인적분할에 따라 2013년 8월 29일 코스피시장에 네이버(주)로 변경 상장하면서 크게 성장하였다.



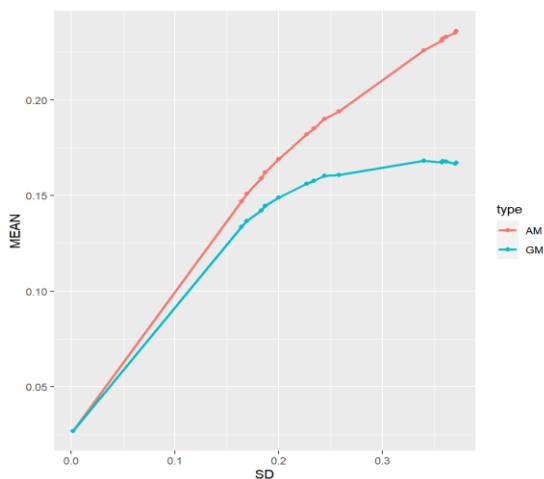




왼쪽 그래프는 학습데이터의 각 기업들의 return 값을 나타낸 시계열 도표이다. 다른 기업들은 비교적 변화가 평이한 것에 비해서 네이버만이 2013년에 0.5까지 증가한 것을 확인할 수 있다. 이는 2013년 분할 상장 이슈가 크게 작용한 것으로 보이며, 이 때문에 위 portfolio table에서도 목표 수익률이 커질수록 네이버의 비중이 점점 커지는 것이라고 볼 수 있다.

또한 목표 수익률이 증가할수록 예측수익률의 표준편차 (SD)가 증가하는 것을 확인할 수 있는데, 이는 높은 수익률을 기대할수록 높은 위험을 감수해야 한다는 즉, High Risk, High Return을 보여준다고 해석할 수 있다.

2. 목표수익률이  $i^* = 0.01, \dots, 0.25$  일 때  $(\hat{\sigma}_p, \hat{\mu}_p)$ 의 산점도 (Efficient Frontier) 그리고 그래프의 의미를 해석해 보시오.



마코비츠 모델은 다양한 위험자산 구성으로 리스크를 최소화하는 포트폴리오 이론이다. 위 마코비츠 차트에서 보면 알 수 있듯이 SD가 커질수록, 즉 다양한 종목에 투자를 할수록 기대 수익률의 평균이 증가한다. 상관성이 낮은 자산들을 동시 보유함으로써 다양한 시장 상황에 흔들리지 않는 수익을 창출할 수 있는 것이다. 하지만 그래프에서 꺾이는 지점이 존재하므로, 무조건적으로 많은 종목에 투자하는 것보다는 수익률을 최대화 할 수 있는 SD를 찾는 것이 중요함을 알 수 있다.

3. 최적 포트폴리오 학습자료에서의 기대수익률과 검정자료에서의 실제수익률을 서로 비교해보고 차이가 있을 경우 그 이유 및 의미를 설명해 보시오. ( $i^* = 0.05$ )

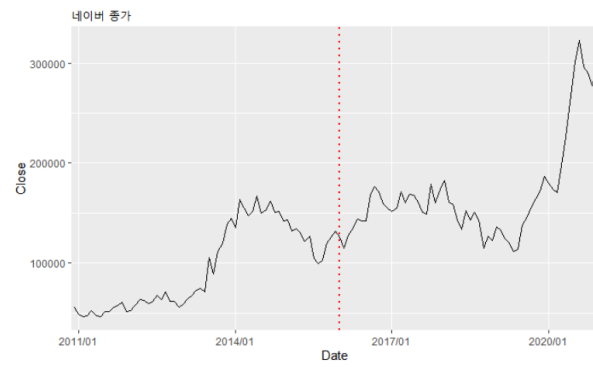
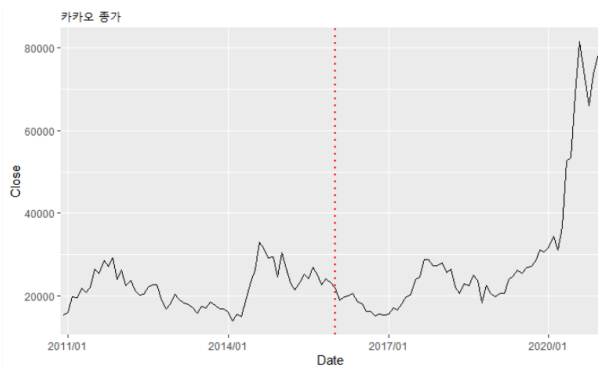
	Mean	Sd
Annualized Expected	0.159	0.183
Annualized Realized	0.156	0.179

연 기대수익률과 연 실제수익률의 평균 및 표준편차가 비슷함을 알 수 있다.

4. 검정자료 (2016.01-2020.12)에서 2016년 1월초 초기자산  $V_0$ 를 최적 포트폴리오에 투자했을 때  $t$  개월 후 최적 포트폴리오의 가치( $vt^*$ )의 시계열 도표와 학습자료 (2011.01-2015.12)에서 구한 투자한지  $t$  개월 후의 최적 포트폴리오의 가치( $vt$ )를 겹쳐서 그려보고 두 시계열 도표의 차이가 있는 경우 그 이유를 설명해 보시오. ( $i^* = 0.05$ )

$t=25$ 와  $t=55$  사이에서 두 시계열 도표의 차이가 커진다. 학습자료의 경우 이 구간에서 수익률이 크게 증가하나 검정자료의 경우는 감소했다가 다시 증가하는 경향을 보인다.

카카오, 네이버 등의 종가 그래프를 확인해보면, 포트폴리오의 가치  $vt$ 의 그래프의 추이가 종가 그래프와 비슷한 것을 알 수 있다.



## Part 4 : R-Shiny

