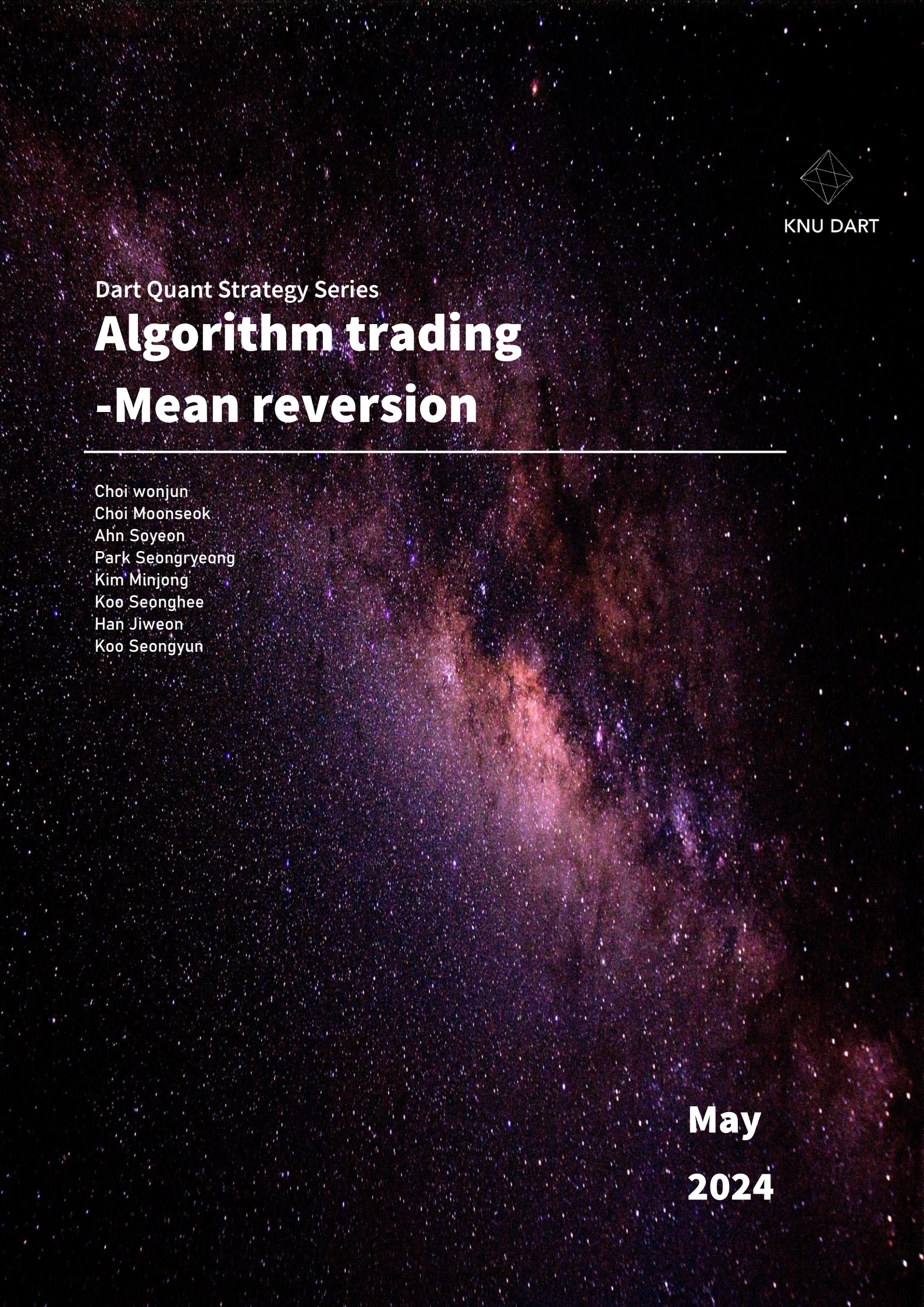


KNU DART

Dart Quant Strategy Series

Algorithm trading -Mean reversion

Choi wonjun
Choi Moonseok
Ahn Soyeon
Park Seongryeong
Kim Minjong
Koo Seonghee
Han Jiweon
Koo Seongyun

May

2024

Contents

Chapter 1. Statistical test for mean reversion

1. Introduction : What is Mean-reversion? 1
2. Time series data 2
3. Random work in Stock market 3
4. Hurst exponent for Mean Reversion 4
5. ADF test for Mean Reversion 5
6. Half – Life of Mean Reversion 6
7. Application of Mean Reversion Strategy in Actual Stock Market 7

DART 2024 First Semester Algorithm trading

Project Manager: Choi wonjun
+82 010-4368-1210
7373wj@naver.com

[Github]
https://github.com/orgs/DART-KNU/teams/algorithm-trading-24fs

Contents

Chapter 2. Practical application of mean reversion

8. Bolinger Band 1
9. Engle & Granger Cointegration Test 2
10. Spread trade 3
11. VAR 4
12. Johansen Test 5

DART 2024 First Semester Algorithm trading

Project Manager: Choi wonjun
+82 010-4368-1210
7373wj@naver.com

[Github]
https://github.com/orgs/DART-KNU/teams/algorithm-trading-24fs

Chapter 1. Statistical test for mean reversion

1. introduction : What is mean-reversion

알고리즘 트레이딩이란 컴퓨터 프로그래밍을 이용하여 일정한 알고리즘에 따라서 증권, 파생상품, 외한등 유동성 자산을 자동으로 매매하는 방식을 말한다. 알고리즘 트레이딩에서 사용하는 알고리즘은 크게 모멘텀과 평균회귀 라는 두 가지 중요한 개념이 존재한다. 모멘텀은 그동안 상승하던 자산의 가격이 더 오를 가능성이 높고, 그 반대의 경우에서도 성립한다는 성질을 말하며, 평균회귀는 자산 가격이나 수익률이 장기적인 평균이나 평균값으로 되돌아가려는 경향을 의미한다.

금융경제학에서 여러 학술 문헌들은 시장과 정보의 성격이 가격에 영향을 어떻게 미치는지 철저히 분석해 왔다. 금융시장, 시장 참여자와 그들의 합리성 수준, 참여자들이 어떻게 상호작용 하는지 등에 관해 여러 가정을 도출해 내었으며, 이는 시장 가격이 이용가능한 모든 정보를 반영한다는 효율적 시장 가설(Efficient market hypothesis)에 까지 이르렀다. 효율적 시장 가설에 따르면, 주가는 랜덤워크와 구별될 수 없는 형태로 진화한다는 것이며, 이에 따르면 가격과 공개적으로 이용 가능한 데이터를 활용해 알파를 획득할 수 있는 패턴을 찾는 투자전략이 통계적 관점에서 투자자가 공감할 수 있는 투자전략으로 이어질 수 없음을 암시한다. 하지만 최근 20 년간 금융 경제학 연구에서는 무작위 행보에서 벗어나는 현상(즉, 효율적 시장가설에서 벗어나는 전략)을 더 자세히 탐구하게 되었고, 이러한 이론은 투자 전략, 자본 적정성, 그리고 옵션의 가격 결정 및 헷지에까지 활용되어 왔다.

무작위 행보에서 벗어나는 이론들은 평균회귀 혹은 평균 이탈 모델을 포함하는 더 넓은 범위의 모델들이 존재한다. 그 중 우리는 평균회귀 모형에 대해 중점적으로 조사해 보고자 한다. 일부 평균회귀 모델은 생명보험 회사처럼 장기적인 투자 위험을 부담하는 기관들의 입장에서 랜덤워크 모델 보다 더 적은 자본으로 수익을 얻을 수 있다고 제안되어 왔다. 이는 평균회귀 모델에 대한 새로운 가능성을 유발했다.

평균회귀의 의미는 상황에 따라 달라질 수 있으며, 가장 넓은 정의는 자산 가격이 최대치(최소치)에 도달한 후 하락(상승)할 경향이 있다는 것이다. 이는 시장의 역사적 최고치를 검토한 후 시장이 이후에 하락했는지를 살펴보는 간단한 테스트를 통해 경험적으로 확인할 수 있다. 보다 정확하게 평균회귀를 정의하고 실증적으로 적용하기 위해서는, 경험적으로 검증하기 보다 통계적 정의로 나아가는 것이 유익할 수 있다. 이는 확률 과정의 여러 검정을 통해서 더 잘 구별할 수 있다.

평균회귀의 기본적인 식은 자산의 수익률이 음의 자기 상관관계를 가지는 경우, 즉 한 기간 동안의 평균 이하 수익률이 이후 기간에 "보상적"으로 평균 이상 수익률을 이어갈 가능성이 높은 경우로 정의된다. 이것의 대표적인 예로는 미국 달러/캐나다 달러의 환 거래에 존재하고 있다. 두 유사한 국가의 통화정책이 유사하기 때문에 그 나라의 통화 가격은 장기적인 평균치로 회귀하는 경향을 보이는 경우가 있다. 이 외에도 우리는 일상생활에서 다양한 평균회귀 경향을 경험하고 있다. 달러의 경우 1300 원, 엔화의 경우 100 엔당 1000 원 휘발유 리터당 1800 원등 예전부터 내려오는 정설이 바뀌지 않고 현대에까지 내려오고 있다는 사실은 시장에서 평균회귀가 존재함을 암시하고 있다.

따라서 본 레포트에서는 우리가 다루고 있는 내용은 다음과 같다. 먼저, 시계열 데이터의 속성과, 주식의 랜덤워크 가설에 대해 알아 본 다음, 시계열 데이터가 이전 차수의 값에 영향을 받는지 확인하는 ADF 검정 방식과 분산이 시간 간격에 비례하여 어떻게 증가하는지 관측할 수 있는 Hurst exponent 검정 방식을 통해 평균회귀를 검정한다. 그 후, 시계열 방정식의 연속적 모델을 정리한 OU process 기반의 식에서부터 평균회귀의 반감기 속도를 계산한다. 이 과정을 통해 얻어진 값으로, 평균회귀가 현대 증권시장에서도 먹히는지 분석해 나갈 것이다.

하지만 주식시장은 다양한 요인으로 증가하는 경향이 있다. 이러한 경향은 주식 시장의 특성을 반영하며, 평균회귀만으로는 시장의 다양한 움직임을 완전히 설명하기 어렵다. 따라서, 평균회귀의 한계를 인식하고 이를 보완하기 위한 다양한 방법을 모색하는 것 또한 중요하다. 이러한 맥락에서, 현대 금융 경제학은 평균회귀의 개념을 깊이 이해하고 이를 최적화하기 위한 전략을 개발해 왔다. 평균회귀 전략이 시간이 지남에 따라 어떻게 발전해 왔는지 살펴보는 것은 투자자들에게 중요한 통찰을 제공할 수 있다. 따라서, 우리는 Chapter2 에서 평균회귀 전략이 현대적으로 어떻게 발전해 왔는지 또한 소개할 예정이다. 우리는 그것들 중 이동평균선을 기준으로 회귀할 것이라는 아이디어인 볼린저밴드와 두개의 유사한 주식이 있다면 그 주식들은 서로 비슷한 방향성으로 회귀할 것이라는 pair-trading 에 대해서 자세히 알아볼 예정이다.

2. Time series data

시계열 데이터란 시간의 흐름에 따라 순서대로 관측되는 자료를 뜻 한다. 시계열 데이터를 표현하는 기본 가정은 시간 순서에 따라 관측된 데이터로, 이때 관측시간이 변수로써 시간에 대한 관측자료의 관계를 표현한다. 이러한 데이터는 주식 가격, 날씨 변화, 경제 지표 등 다양한 분야에서 사용된다. 주로 우리는 이러한 시계열 데이터를 분석하여 과거의 패턴을 파악하고 미래를 예측하는 데 활용한다. 우리가 다루는 데이터 또한 시간의 흐름에 따른 주식의 가격 데이터를 살펴본다. 따라서 시계열 데이터의 기본적 특성 및 이후 우리가 사용할 용어에 대해서 정의할 필요가 있다.

2.1 Unit root and stationarity

단위근(unit root) 검정은 시계열 데이터가 정상성(stationarity)을 가지는지를 판단하는 방법 중 하나이다. 여기서 단위근은 시계열 데이터가 시간 경과에 따라 그 자체의 과거 값에 의존하는 정도가 1 인 경우를 뜻하며, 정상성은 시간에 관계없이 시계열 데이터의 특성이 일정하게 유지되는 것을 뜻한다. 단위근이 존재한다면 시계열 데이터는 정상성을 가지지 않으며, 이를 보정하여 정상성을 만족시킬 필요가 있다. 이해를 돕기 위해 간단한 시계열 모형을 관찰해보자.

$$y_t = \rho y_{t-1} + \epsilon_t \tag{2.1}$$

위 (1) 모형에서 ϵ_t는 확률적 오차항으로 평균이 0이고 분산이 σ^2으로 일정하며 자기상관이 없는 것으로 가정한다. 흔히 이를 백색잡음 오차항(white noise error term)이라고 한다. 위 식에서 차분연산자를 이용하여, 즉 위 식의 양변에서 y_{t-1}을 빼주면 다음과 같다.

$$\Delta y_t = (\rho - 1) y_{t-1} + \epsilon_t = \delta y_{t-1} + \epsilon_t \tag{2.2}$$

여기서 Δ는 차분연산자로 $\Delta y_t = y_t - y_{t-1}$이고 $\delta = 1 - \rho$이다.

단위근 검정은 귀무가설 $H_0 : \delta = 0$을 검정하는 것으로 $H_0 : \rho = 1$의 검정과 동일하다. 대립가설은 $H_1 : \delta < 0$ 또는 $H_1 : \rho < 1$이다. 대립가설로부터 알 수 있는 바와 같이 단위근 검정은 단측검정(one-tail test)이다. 따라서 δ 의 추정값이 0 보다 크거나 같으면 그 시계열은 단위근을 갖는다고 할 수 있다.

2.2 ARIMA(Autoregressive Integrated Moving Average)

ARIMA(Autoregressive Integrated Moving Average) 모델은 시계열 데이터 분석에 널리 사용되는 모델 중 하나이다. ARIMA 모델은 자기회귀(AR), 누적차분(I), 이동평균(MA) 요소를 결합한 모델로, 시계열 데이터의 추세와 계절성을 고려하여 데이터를 모델링하고 예측한다.

- 자기회귀(AR; Autoregressive): 과거의 값이 현재 값에 영향을 주는 모델
- 누적차분(I; Integrated): 차분을 통해 시계열 데이터의 정상성을 보정하는 단계
- 이동평균(MA; Moving Average): 과거의 충격이 현재 값에 영향을 주는 모델

ARIMA 모델은 이러한 요소들을 조합하여 시계열 데이터를 예측하는 데 활용된다. 모델의 성능은 주어진 데이터에 따라 다르며, 적절한 모수 선택과 모델 평가가 중요하다. 이러한 점들을 조합하면 ARIMA 모델의 기본 식은 다음과 같이 나온다.

$$\Delta^{d} Y_t = \alpha + \phi_1 \Delta^{d} Y_{t-1} + \phi_2 \Delta^{d} Y_{t-2} + \cdots + \phi_p \Delta^{d} Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \qquad (2.3)$$

우리는 위 방정식을 통해 ϕ_1를 유추할 수 있다. 우리는 평균회귀를 검증하기 위해 직전 값만 살펴볼 예정이므로, 이동평균은 존재하지 않으므로 q 는 0 이되며, 마찬가지로 직전 값만 확인하므로 차분의 계수 (d)는 0 이 된다. 이전 값에 영향을 받는지 궁금하므로 자기회귀의 부분 차수 p 는 1 이 된다. 따라서 식을 정리하면 앞서 보았던 (2.1) 모형과 같이 AR(1)모형이 나타난다.

$$\Delta y_t = (\rho - 1) y_{t-1} + \epsilon_t$$

이제 이 방정식을 통해, 우리는 δ의 값을 유도할 수 있으며, 이 값에 따라 통계적으로 데이터가 정상성을 띄는지, 그리고 평균회귀적 기질이 있는지 유추할 수 있다. 자세한 내용은 Chapter 1. 5 장 ADF 검정 부분에서 추가로 다룰 예정이다.

3. Random Walk in Stock Market

주가가 시계열성을 띠는 데이터라면, 과거의 데이터를 이용해 미래의 움직임을 예측할 수 있을 것이란 생각이 들 수 있다. 하지만 주가 데이터가 랜덤워크를 따른다고 본다면, 미래를 예측하는 것은 사실상 불가능하다. 랜덤워크 가설이란 술 취한 사람이 비틀비틀 걸어가는 모습을 묘사한 말로, 현재의 주가는 과거의 주가나 그 추세에 영향을 받지 않고 독립적으로 움직인다는 가설이다. 따라서 매 시점의 주가는 상호 독립적이고, 무작위적(random)으로 움직이기에 과거의 주가 데이터를 바탕으로 미래의 주가를 예측하는 것은 불가능하다.

이 가설은 1900 년 프랑스 수학자 루이 바슐리에(Louis Bachelier)의 〈투기이론, Theory of Speculation〉에서 제시되었다. 루이 바슐리에는 금융시장의 가격변동을 브라운 운동(Brownian Motion)으로 모형화 하였는데, 액체나 기체 안에 떠 있는 작은 입자의 불규칙한 운동을 가리키는 브라운 운동처럼 주가의 움직임도 불규칙함을 보인다는 것이다. 이후 1950 년대 중반 미국의 경제학자인 폴 새뮤얼슨(Paul Samuelson)이 바슐리에의 이론을 수정해 기하 브라운 운동(Geometric Brownian Motion)을 정립하였으며, 주식 가격의 움직임을 설명하는 식을 정리하였다.

이어서는 랜덤워크 가설과 기하 브라운 운동을 수학적으로 이해해보고자 한다. 랜덤워크의 경우 먼저 동전을 던지는 상황을 가정해보자. 동전의 앞면이 나오면 +1, 뒷면이 나오면 -1 이라고 하고 수식으로 표현해 보면 아래와 같다.

$$X_j = \begin{cases} 1 & (W_j = Head) \\ -1 & (W_j = Tail) \end{cases}$$

$$M_k = \sum_{j=1}^{k} X_j \quad (k = 1,2,3 \cdots), \qquad M_0 = 0$$

1 번째부터 j 번째까지의 X_j를 더하면 j 번째의 M_k값이 되며, 이 M_k 값을 Symmetric Random Walk 라고 한다.

$$E(M_k) = 0, \quad Var(X_j) = 1$$

$k_0 < k_1 < k_2 \cdots < k_m$일 때

$$M_{k_{i+1}} - M_{k_i} = \sum_{j=k_i+1}^{k_{i+1}} X_j$$

$$Var(M_{k_{i+1}} - M_{k_i}) = \sum_{j=k_i+1}^{k_{i+1}} var(X_j) = \sum_{j=k_i+1}^{k_{i+1}} k_{i+1} - k_i$$

랜덤워크도 확률과정이므로 평균과 분산의 특성을 가지며, 평균은 +1 과 -1 의 조합으로 만들어졌기에 0 이 될 것이고, 그래프의 한 단위 분산은 위나 아래로 1 씩 움직이므로 1 이 나올 것이다. 이어서 두 단위 분산은 2 가 될 것이고, 따라서 k 단위의 분산은 k 가 됨을 알 수 있다.

이번에는 동전을 던지는 횟수 사이 구간을 더 나누어, 실험 사이에 또 여러 번의 동전 던지기를 수행한다고 가정하자. 처음 실험을 1,2,3 … t 번째로, 그 사이의 실험을 1,2,3 … n 이라고 순서를 매기면 (3.1)식으로 표현된다. n 은 t 사이의 실험의 횟수이며, t 를 시간으로 n 을 분으로 비유하여 생각한다면 이해가 쉽다.

$$W^{(n)}(t) = M_{nt} \quad (3.1)$$

$$W^{(n)}(t) = \frac{1}{\sqrt{n}} M_{nt} \quad (3.2)$$

루트 n 으로 나눠준 (3.2)식은 Scale 을 조정해준 것이고, Scale 만 조정하였을 뿐 다른 특성에는 영향을 미치지 않는다. 해당 식을 Scaled Random Walk 라고 부르며, n 이 무한히 커질 때를 브라운 운동(Brownian Motion) 혹은 위너 과정(Wiener Process)라고 부른다.

$$E\left(W^{(n)}(t_j) - W^{(n)}(t_i)\right) = 0$$

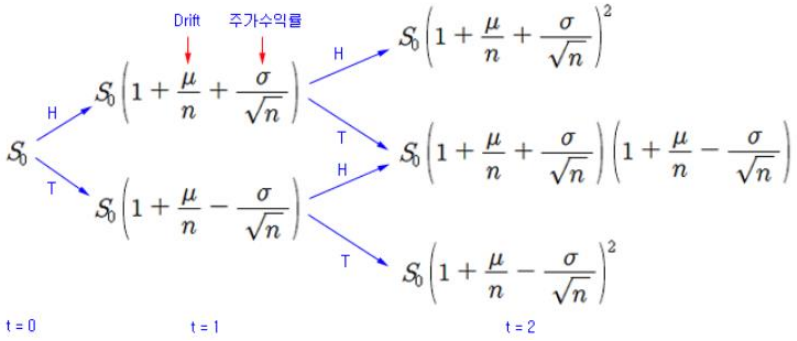
$$Var\left(W^{(n)}(t_j) - W^{(n)}(t_i)\right) = t_j - t_i = \Delta t$$

$$\Delta W^{(n)}(t) \sim N(0, (\sqrt{\Delta t})^2) \quad (3.3)$$

위너 과정도 랜덤워크 과정과 특성이 동일하기에 평균과 분산은 다음과 같다. 위너 과정의 평균은 0 이고, 분산은 Δt가 된다. 위너 과정의 변화량은 평균이 0 이고, 분산이 Δt인 정규분포를 따르게 되며 (3.3)과 같이 나타낼 수 있다.

주가는 무위험 수익률이나 인플레이션과 같은 상승효과의 영향을 받는다. 따라서 무작위성의 위너 과정만으로는

주식의 모형을 만들 수 없기에, 무위험 수익률만큼 Drift 하는 추세에다가 주가의 최근 변동성을 반영한 위너 과정의 효과를 추가하여 모형을 만들어 준다. Drift 를 추가하여 주가가 상승하거나 하락하는지에 관계없이 매 기마다 Drift 만큼은 오르게 해주고, 주가수익률은 수익률의 표준편차로 설정한다. 그런 다음 아래의 그림과 같이, 주가의 흐름을 상승 혹은 하락으로 나눠 나타내면 시간이 경과함에 따라 동전 던지기와 게임과 동일한 형태로 전파된다.



이 전파 과정에 대한 일반항을 만들어 정리하여 주가 모형을 도출해 보겠다.

$$S_t = S_0(1+\frac{\mu}{n}+\frac{\sigma}{\sqrt{n}})^H(1+\frac{\mu}{n}-\frac{\sigma}{\sqrt{n}})^T \quad (3.4)$$

$$nt = H + T$$

$$M_{nt} = H - T,\ \ M_{nt} = M$$

일반항 S_t를 식(3.4)로 나타낼 수 있으며, H 는 주가가 상승하는 경우의 횟수, T 는 주가가 하락하는 경우의 횟수를 나타내어, nt 는 전체 시행의 횟수가 된다. M_{nt}는 랜덤워크의 정의식을 나타내고, 두 식을 이용하여 H 와 T 를 식(3.5)와 같이 나타낼 수 있다.

$$H = \frac{1}{2}(nt+M) \qquad T = \frac{1}{2}(nt-M)$$

$$S_t = S_0(1+\frac{\mu}{n}+\frac{\sigma}{\sqrt{n}})^{\frac{1}{2}(nt+M)}(1+\frac{\mu}{n}-\frac{\sigma}{\sqrt{n}})^{\frac{1}{2}(nt-M)} \quad (3.5)$$

정리된 일반항의 양변에 자연로그를 취하면 아래와 같은 식이 되고, 여기서 로그항을 쉽게 풀기위해 테일러 급수를 적용한다.

$$\ln S_t = \ln S_0 + \frac{1}{2}(nt+M)\ln(1+\frac{\mu}{n}+\frac{\sigma}{\sqrt{n}}) + \frac{1}{2}(nt-M)\ln(1+\frac{\mu}{n}-\frac{\sigma}{\sqrt{n}})$$

$$\ln\left(1+\frac{\mu}{n}+\frac{\sigma}{\sqrt{n}}\right) \approx \left(\frac{\mu}{n}+\frac{\sigma}{\sqrt{n}}\right) - \frac{1}{2}\left(\frac{\mu}{n}+\frac{\sigma}{\sqrt{n}}\right)^2 = \frac{\mu}{n}+\frac{\sigma}{\sqrt{n}} - \frac{\mu^2}{2n^2} - \frac{\mu\sigma}{n\sqrt{n}} - \frac{\sigma^2}{2n} \quad (3.6)$$

$$\ln\left(1+\frac{\mu}{n}-\frac{\sigma}{\sqrt{n}}\right) \approx \left(\frac{\mu}{n}-\frac{\sigma}{\sqrt{n}}\right) - \frac{1}{2}\left(\frac{\mu}{n}-\frac{\sigma}{\sqrt{n}}\right)^2 = \frac{\mu}{n}-\frac{\sigma}{\sqrt{n}} - \frac{\mu^2}{2n^2} + \frac{\mu\sigma}{n\sqrt{n}} - \frac{\sigma^2}{2n} \quad (3.7)$$

식(3.6)과 식(3.7)은 로그 항에 대한 테일러 급수 식이며, 정리된 식을 다시 로그를 취한 일반항에 대입하여 정리해주면 아래의 식이 된다.

$$\ln S_t = \ln S_0 + \mu t - \frac{\sigma^2}{2}t + \frac{M}{\sqrt{n}}\sigma - \frac{\mu^2 t}{2n} - \frac{\mu\sigma}{n}\cdot\frac{M}{\sqrt{n}} \quad (3.8)$$

$$n \to \infty: \frac{M}{\sqrt{n}} = W_t\,, \qquad \frac{\mu^2 t}{2n} = 0, \qquad \frac{\mu\sigma}{n} = 0$$

$$\ln S_t = \ln S_0 + \mu t - \frac{\sigma^2}{2}t + \sigma W_t$$

$$S_t = S_0 e^{\sigma W_t + (\mu - \frac{\sigma^2}{2})t} \quad (3.9)$$

식(3.8)에서 n 을 무한 번 수행하면, 랜덤워크 과정이 브라운운동(위너 과정)이 되고, 마지막 2 개 항은 0 이 된다. 그런 다음 양변에 지수를 취해 정리해주면, 식(3.9)가 바로 기하 브라운 운동 식이 된다. 기하 브라운 운동 식은 위너과정(W_t)을 포함하고 있기 때문에, 결정론적인 식이 아닌 확률론적인 식이 된다.

기하 브라운 운동 모형은 일반적으로 금융 공학에서 파생 상품의 가격 변화를 예측하는 모형으로도 활용된다. 대표적인 예시로 블랙숄즈 모형을 통한 옵션 가격 산출 등이 있다. 이 모형을 활용할 경우 주식의 평균 수익률과 변동성(표준편차)에 관한 데이터만 있으면 미래 가격 분포를 추정할 수 있기 때문에 용이하게 사용된다.

주가의 랜덤워크 및 기하 브라운 운동에 대해 알아보았다면, 이제 본격적으로 평균회귀성을 찾아볼 차례이다. 주가나 환율 등이 평균회귀한다고 하면 두가지 검정을 만족해야 하는데, 기존의 방식에 따르면 크게 Hurst 지수 검정과 ADF 검정이 있다. Hurst 지수 검정을 통해 데이터가 추세를 보이는가에 대한 확인을 하고, ADF 검정은 데이터가 정상성을 띄는지에 대한 검정을 하는 것이다. 다음 챕터를 통해 두 검정의 방식에 대해 자세히 다뤄보겠다.

4. Hurst exponent

기하브라운 운동을 통해 주가의 랜덤워크에 따른 가설을 검증해 보았다. 이를 활용하여 허스트 지수 검정을 통해 주가가 랜덤워크인지 혹은 랜덤워크에서 벗어나는지 확인할 수 있다. 허스트 지수는 금융 시계열 데이터가 순수한 랜덤워크에서 얼마나 벗어나는지를 측정하는 방법을 제공한다. 허스트 지수가 0.5 초과인 경우는 시계열이 추세강화 성질을 가지며, 시계열이 현재 추세를 유지할 가능성이 높으며, 과거의 추세가 미래에도 지속될 것임을 암시한다. 허스트 지수가 0.5 미만인 경우는 시계열이 평균회귀하는 경향을 보이며, 시계열이 장기 평균값으로 돌아 가려는 성질을 가지고 있으며, 시장의 과매수나 과매도 상태일때, 이를 교정하려는 힘이 작동한다는 것을 의미한다. 허스트 지수가 0.5 인 경우는 시계열이 순수한 랜덤워크를 따르며, 이는 앞서 설명한 바와 같이 과거의 움직임이 미래의 움직임과 독립적이다 라는 것을 의미한다.

허스트 지수 검정을 유도하는 가장 오래되고 직관적인 방식은 Rescaled range(R/S) 검정이다. 허스트 지수를 추정하기 위해 시간 간격 n 에 의존하는 R/S 의 로그 회귀계수 값을 사용한다. 구체적인 방식은 다음과 같다.

1. Calculate the mean

$$m = \frac{1}{n}\sum_{i=1}^{n} X_i$$

2. Create a mean-adjusted series

$$Y_t = X_t - m \text{ for } t = 1,2,\dots,n.$$

3. Calculate the cumulative deviate series

$$Z_t = \sum_{i=1}^{t} Y_i \text{ for } t = 1,2,\dots,n.$$

4. Compute the Range R(n)

$$R(n) = \max(Z_1, Z_2, \dots, Z_n) - \min(Z_1, Z_2, \dots, Z_n).$$

5. Compute the standard deviation S(n)

$$S(n) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2}.$$

6. The estimated statistic of R(n)/S(n)

$$E\left[\frac{R(n)}{S(n)}\right] = Cn^H \text{ as } n \to \infty,$$

7. Taking the natural logarithm of both sides

$$H \cdot \ln(n) + C = \ln\left(\frac{R(n)}{S(n)}\right) \tag{4.1}$$

여기서 각 변수의 설명은 다음과 같다.

- *R*(*n*)은 처음 *n* 개의 평균에서 벗어난 누적 편차의 범위입니다.
- *S*(*n*)은 처음 *n* 개의 표준 편차의 합(시리즈)입니다.
- E[*x*]는 기대값을 나타냅니다.
- *n* 은 관찰 기간의 시간 범위입니다 (시계열 데이터 포인트의 수).
- *C* 는 상수입니다.

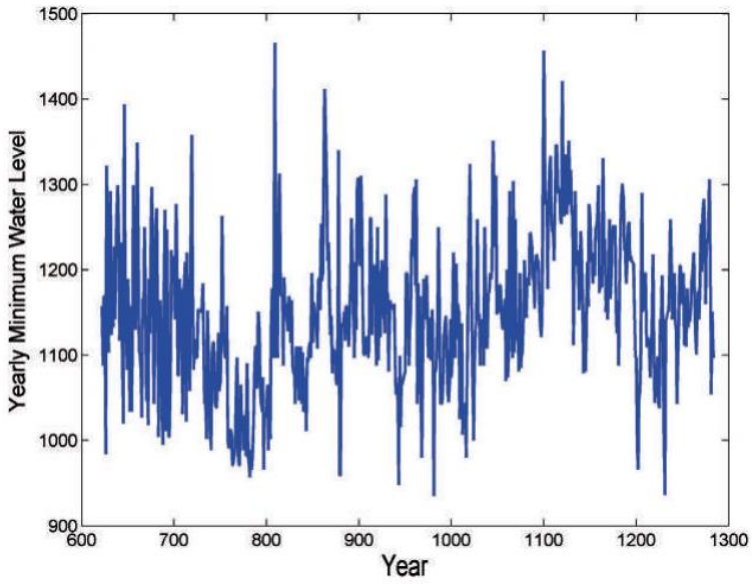
Hurst 지수 검정의 식은 결국 R/S 검정을 거쳐 다음과 같은 결과가 나오게 된다. 위 식에서 우리는 허스트 지수 H 는 ln(R(n)/S(n))를 종속변수로 ln(n)을 독립변수로 선정하여 선형회귀를 돌린 계수값을 의미한다는 것을 알 수 있다. 이것은 결국 허스트 지수가 시간의 크기(n)에 따라 R(n)/S(n)이 변하는 정도를 나타나게 된다.

이제 위 식을 자세히 살펴보자. 앞서 설명한 봐와 같이 브라운 운동에서 시간의 크기 n 에서의 분산은 시간에 따라서 선형적으로 증가한다. 브라운 운동을 따른다면, 시간간격 n 에 따른 변동성 R(n)은 시간 간격의 표준편차 범위 내에 존재할 가능성이 크므로 결국 R(n)은 시간의 제곱근에 비례하여 증가한다. S(n) 또한 결국 n 의 시간 범위에 대한 표준편차를 뜻하므로 시간의 제곱근에 비례하여 증가한다. 따라서 주가가 랜덤워크를 따른다면, 따라서 R(n)/S(n)이 시간 간격의 제곱근이 비례하여 증가한다는 뜻이 된다.

따라서 H 가 0.5 라는 것은 R(n)/S(n)이 시간간격 n 의 제곱근 비례해서 증가한다는 것을 의미한다. 이것은 앞서 설명했듯이 주가의 랜덤워크를 뜻하게 되며 따라서, 'H 가 0.5 일 때 우리는 랜덤워크적 경향이 있다.' 라고 판단할 수 있다. H 가 0.5 보다 큰 경우에서는 변동성의 범위가 시간의 크기보다 지수적으로 빠르게 증가하는 것이 되며 이는 모멘텀적 경향이 있음을 나타낸다. 우리는 반대로 평균회귀 하는 데이터를 찾아야 하므로 H 가 0.5 보다 작은 시계열 데이터를 찾아 평균회귀 데이터 검증을 실시할 예정이다.

5. ADF test for mean - reversion

앞서 설명한 바와 같이 시계열에서의 정상성은 일반적으로 시간이 지나도 평균, 분산 및 자기 상관 구조가 변하지 않는것을 의미한다. 즉 시계열에 추세, 주기 또는 시간에 따라 변하는 체계적인 패턴이 없음을 의미한다.



정상시계열인 경우

평균회귀는 가격 시계열이 일시적으로 평균에서 벗어난 후에 평균값으로 다시 돌아가려는 경향을 의미하는데 가격 시계열이 평균 회귀를 나타내면 일정한 수준을 중심으로 진동하며 이 수준에서의 이탈은 일시적이며 다시 평균으로 돌아간다는 것이다.

즉 정상성을 띄는 시계열은 일정한 수준을 중심으로 진동하며 다시 평균으로 회귀할 것이다. 우리는 ADF 검정을 통해 주가의 정상성을 판단하여 주가가 평균회귀할 것인지 알아볼 수 있다.

ADF 검정의 귀무가설과 대립가설은 다음과 같다.

H_0 : 시계열은 단위근을 포함한다. ($\lambda \neq 0$, 정상성을 가지지 않는다.)

H_1 : 시계열은 단위근을 포함하지 않는다. ($\lambda = 0$, 정상성을 가진다.)

ADF 검정의 식은 다음과 같다.

$$\Delta Y(t) = \lambda Y(t-1) + \mu + \beta t + \alpha_1 \Delta Y(t-1) + \cdots + \alpha_k \Delta Y(t-k) + \varepsilon_t$$

위 식은 다음과 같이 구성되어 있다.

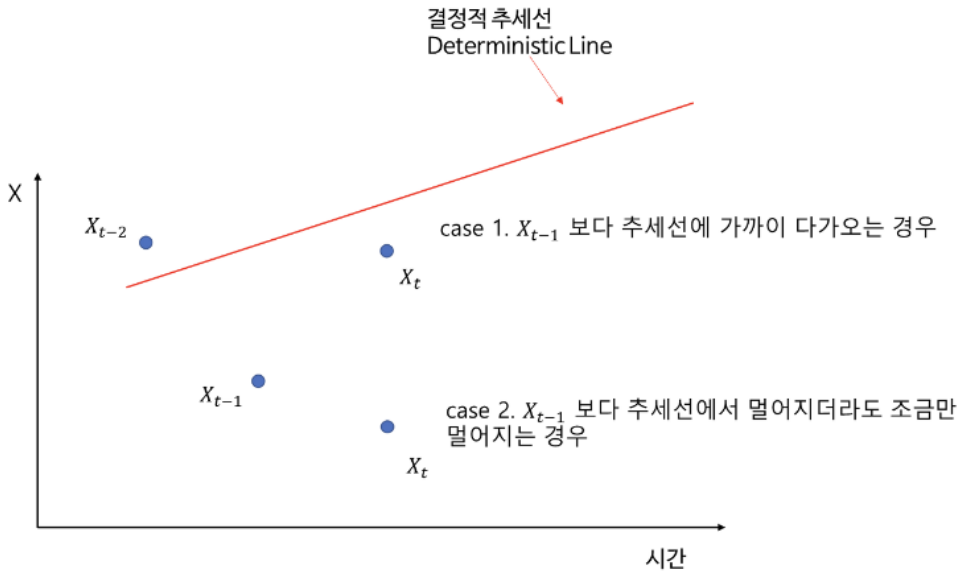
- $\Delta y(t) \equiv y(t) - y(t-1)$는 현재 시점 t 에서의 가격 변화

- λ 는 회귀계수
- μ 는 평균,
- βt 는 추세
- α1Δy(t−1)+…+αkΔy(t−k)는 이전의 가격 변화
- ε_t는 오차 항

이때 주식시장의 일간 변동성은 상당히 높으며 추세에 비해 훨씬 큰 영향을 미치므로 βt 를 0 으로 가정하여 모형을 간단하게 유지하고 변동성에 집중하게 된다. 결국 이는 앞서 설명했던 ARIMA 와 유사한 식이 나온다.

$$\Delta Y(t) = \lambda Y(t-1) + \mu + \varepsilon_t$$

이때 회귀계수 λ = 0 이 아니라면 Δy(t)가 y(t − 1)에 영향을 받으며, λ > 0 이면 추세가 존재하여 주가가 평균회귀하지 않고, λ < 0 이면 주가가 평균회귀할 가능성이 있다. 이때 회귀 계수 λ 를 표준 오차 $SE(\lambda)$로 나눈 $\lambda/SE(\lambda)$를 검정통계량으로 사용하여 임계값보다 낮으면 주가가 정상성을 띄며 평균회귀할 것으로 본다.



λ < 0 인 경우

사진 출처 : https://zephyrus1111.tistory.com/169

6. Half – Life of Mean Reversion

앞서 ADF 검정과 HURST 지수 검정에서 시계열의 정상성과 가격 시리즈가 평균회귀를 하는지에 대한 유무를 알아보았다. 다만, 평균회귀의 전략을 실제 투자전략에 대해서 사용하기 위해서는 거래에 대한 최적의 보유기간을 구해야 한다. 여기서, 우리는 보유 기간을 평균회귀의 반감기로 정의한다. 반감기가 의미하는 바는 초기 y 값(t=0 일 때의 y 값)과 평균 사이 거리가 절반이 되는 시점을 의미한다. 우리가

반감기를 평균회귀의 보유기간으로 삼는 이유는 평균회귀의 초기 단계에서 가장 빠른 속도로 변화가 일어나므로, 이 기간 동안의 투자는 더 높은 효율성(즉, 시간 대비 수익)을 제공하기 때문이다. 이때의 반감기는 OU process 라는 평균회귀 확률 과정 방정식으로부터 계산할 수 있다.

우리는 가격 시리즈가 평균회귀를 한다는 가정을 하므로, 앞서 ADF 검정에서 언급한 회귀 계수 λ는 음수라는 것을 알 수 있다. 이를 고려하여 이산 시계열 방정식을 연속 시계열 방정식으로 표현해주면 OU process 의 모델 하에서 다음과 같이 정의된다.

$$dy_t = \theta(\mu - y_t)dt + \sigma dW_t \qquad (6.1)$$

y_t = t 시점에서의 가격, μ 는 평균, θ 는 -λ 로 가격이 장기평균으로 회귀하는 속도 계수를 의미한다. 식 (6.1)은 확률미적분학에서 평균회귀 확률 과정인 OU process(Ornstein-Uhlenbeck process)의 공식이다. 이 미분 방정식을 풀어보기 전에, 식 (6.1)이 직관적 해석이 필요하다. 먼저 RHS 의 첫 항$\theta(\mu - y_t)dt$ 을 관찰해 보면, t 시점에서의 가격 y_t가 평균 μ로부터 얼마나 떨어져 있는가(즉, 편차를 의미한다.)를 의미한다. 또한 회귀계수 θ는 회귀의 속도를 의미한다. 따라서 편차가 클수록, 회귀계수의 절댓값이 클수록 더욱 빨리 평균으로 회귀한다 라는 것을 알 수 있다.

위에서 OU process 공식의 직관적인 해석을 하였으므로, 이제 미분 방정식을 풀어 보겠다. 먼저, 양 변을 dt 로 나누어 주면 다음과 같다.

$$\frac{dy_t}{dt} = -\lambda(\mu - y_t) + \sigma\frac{dW_t}{dt} \qquad (6.2)$$

y_t항을 LHS 로 이항 시켜준다.

$$\frac{dy_t}{dt} - \lambda y_t = -\lambda\mu + \sigma\frac{dW_t}{dt} \qquad (6.3)$$

이를 풀어주기 위해 양 변에 적분인자 $e^{\int_0^t -\lambda ds}$를 곱해준다.

$$e^{\int_0^t -\lambda ds} = e^{-\lambda t} \qquad (6.4)$$

$$e^{-\lambda t}\frac{dy_t}{dt} - e^{-\lambda t}\lambda y_t = -e^{-\lambda t}\lambda\mu + e^{-\lambda t}\sigma\frac{dW_t}{dt} \qquad (6.5)$$

LHS 는 곱의 미분 형태로 다음과 같이 표현할 수 있다.

$$e^{-\lambda t}\frac{dy_t}{dt} - e^{-\lambda t}\lambda y_t = d\left(e^{-\lambda t}y_t\right) = dX_t \qquad (6.6)$$

이를 통해 (6.5)식은 다음과 같은 형태로 나타낼 수 있다.

$$\int_0^t d\left(e^{-\lambda s}y_s\right) = -\mu\int_0^t \lambda e^{-\lambda s}ds + \sigma\int_0^s e^{-\lambda s}\,dW_s \qquad (6.7)$$

$$X_t = X_0 + \mu\left(e^{-\lambda t} - 1\right) + \sigma\int_0^t e^{-\lambda s}dW_s \qquad (6.8)$$

$X_t = e^{-\lambda t}y_t$이므로 양 변을 $e^{-\lambda t}$ 로 나누어 주자. 또한, 앞 장에서 언급한 바와 같이 W_s(Winner process)가 포함되어 있는 두번째 항은 평균이 장기적으로 0 으로 수렴하므로 다음과 같이 적을 수 있다.

$$y_t = y_0 e^{\lambda t} + \mu\left(1 - e^{\lambda t}\right) \qquad (6.9)$$

λ의 부호는 음수이므로 장기적으로 $e^{\lambda t}$는 결국 0 으로 수렴하게 된다. 따라서, 식(6.9)는 $t = 0$ 에서 y_0 이고 t 가

증가할수록 평균 μ으로 감소하여 수렴하게 되는 그래프를 그리게 된다. 이때 y_t의 반감기, 즉 , $t_{0.5}$는 $-\frac{ln2}{\lambda}$가 되믕ㄹ 관찰할 수 있다.

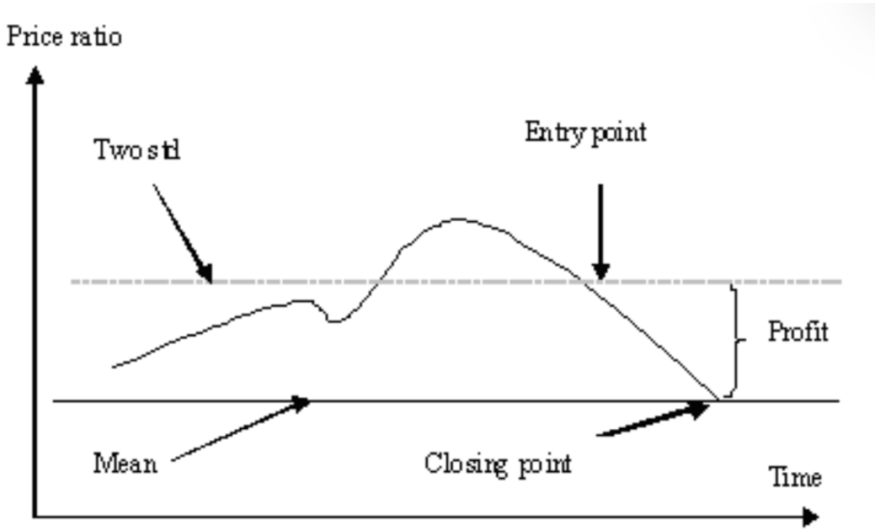
$$y_{t_{0.5}} = y_{-\frac{ln2}{\lambda}} = \frac{y_0}{2} + \frac{\mu}{2} = \frac{y_0 + \mu}{2} \qquad (6.10)$$

반감기가 의미하는 바는 초기 y 값(t=0 일 때의 y 값)과 평균 사이 거리가 절반이 되는 시점을 의미한다. 앞에서 언급한 바와 같이 이 반감기는 거래에 대한 최적의 보유기간을 의미한다. 이제 평균회귀를 활용하는 성공적인 현대 주식에서의 평균회귀 전략에 대해 알아보자.

7. Application of Mean Reversion Strategy in Actual Stock Market

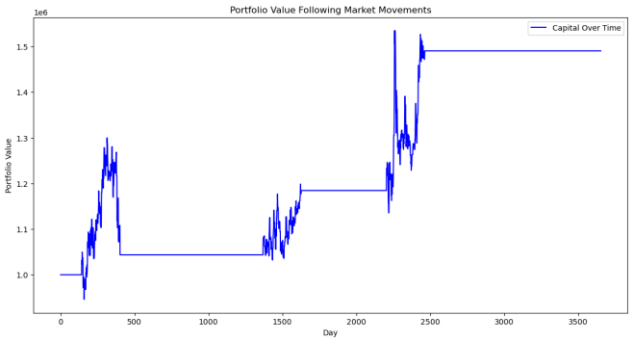
앞서 평균회귀 전략을 통해 평균회귀의 가정 및 기본 개념에 대해서 알게 되었다. 따라서, 우리는 현재 한국시장에서 거래되고 있는 증권, 원자재등 다양한 자산 상품군에 있어 실제 평균회귀의 전략이 먹히는지 백테스트를 통해 알아보고자 한다. 다만, 앞서 우리가 설명했던 통계적 검정들은 적어도 90%의 확신을 요구하는 매우 까다로운 검정들이다. 하지만, 대부분의 금융 투자전략들처럼 훨씬 더 낮은 확신으로도 수익을 내기에는 충분하다. 따라서, 우리는 적절한 실제 통계적 가정의 검증치보다 실증적으로 받아들여지거나, 개인 경험에 따라 추론되는 가설에 조금 더 초점을 두어 투자전략을 내세웠다. 기본 적인 투자 전략은 상품군이 통계적으로 어느정도 평균회귀 한다는 가정하에 다음과 같은 전략에 따라 움직인다.

- 주가가 연속으로 2 표준편차 대역에 도달하였을 때, 포지션을 연다.
- 주가가 평균으로 회귀하였을 경우 포지션을 종료한다.
- 주식의 최대 보유 기간은 OU process 가정 하의 반감기로 설정한다..
- 포지션이 종료되기 전까지 새로운 포지션을 잡을 수 없다.



1. 기업은행 주식 (김민종)

은행, 금융지주는 예금, 투자를 통한 안정적인 수익 구조를 가지고 있고 많은 배당을 주어 투자자에게 일반적으로 안전한 투자 자산으로 인식되어 장기적으로 주가 가격의 변동이 크지 않다. 따라서 평균 회귀 전략에 적합할 것으로 기대해 기업은행 종목을 선택하였다.

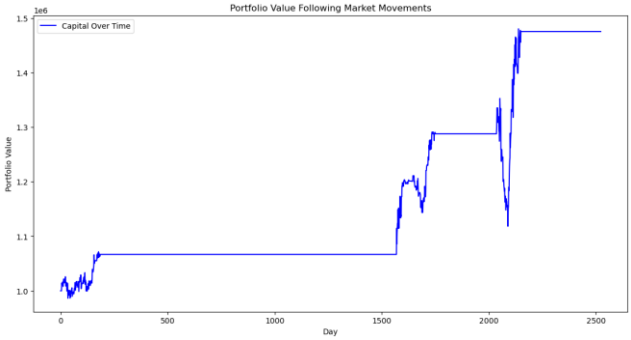


각 세부 통계량의 결과는 다음과 같다.

	ADF-Statistic	p-value	Λ	반감기	Hurst-exponent	CAGR/MDD
기업은행	> 10%	0.34408	0.9973	259.4557	0.5703	4.07%/20%

2-1. 휘발유 (박성령)

휘발유 데이터의 경우 통화 말고 금속이나 에너지 같은 원자재 데이터 중으로 평균회귀 할 법한 데이터를 고민하다, 기름 가격 또한 일반적인 리터 당 원가격이 떠오른다 생각하여 정하게 되었습니다. 휘발유와 경유 중에서, 경유는 러시아/우크라이나 전쟁이나, 경유에 대한 규제와 수요 증가 같은 최근 이슈로 인한 잡음이 다소 있다 생각하여 휘발유로 결정하였습니다.

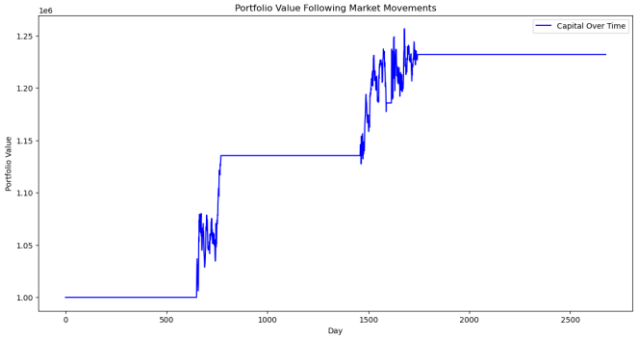


각 세부 통계량의 결과는 다음과 같다.

	ADF-Statistic	p-value	Λ	반감기	Hurst-exponent	CAGR/MDD
휘발유	5% ~ 10%	0.31074	0.9946	129.8416	0.5053	5.79%/17%

2-2. 엔화(박성령)

보통 코스피나 S&P500 과 같은 지수는 수십년에 걸쳐 미미하더라도 성장의 추세가 있다 생각하였습니다. 그러나 환율과 같은 경우, 대게 1 달러는 1100 원, 100 엔은 1000 원이라는 어느 정도의 통념이 존재하기에 평균회귀의 가능성이 비교적 높다고 판단하였습니다. 여러 국가의 통화 중에서도 특히 엔화는 대체적인 기준이 있기에, 그 기준보다 낮아질 때 "엔저", 높을 때 "엔고"와 같은 말도 붙는다고 생각하여 엔화의 평균회귀성 검정을 진행하였습니다.



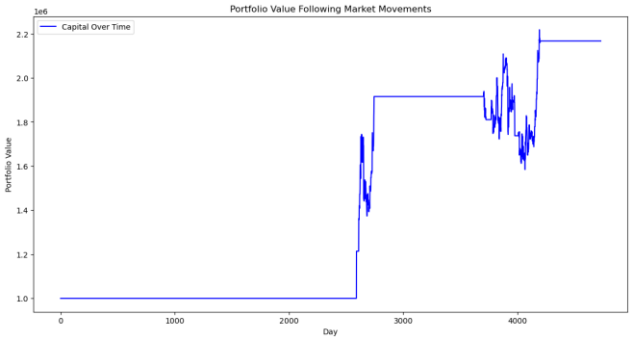
각 세부 통계량의 결과는 다음과 같다.

	ADF-Statistic	p-value	∧	반감기	Hurst-exponent	CAGR/MDD
엔화	5% ~ 10%	0.00176	0.9961	180.2137	0.6285	2.9%/4.8%

3. 우리기술(구성윤)

시장에서는 결국 무위험 수익률이 어떻게든 존재하므로, 무위험 수익률의 영향을 가장 덜 받을 것 같은 종목을 선정하였습니다. 따라서 시장의 기대 인플레이션 영향을 가장 덜 받는, 시장의 변화와 무관한 업종인 기술 관련 업종중에서 선택하였습니다.

이렇게 선별된 기업들 중에, MDD 를 줄이기 위해 상대적으로 거래량이 작고 변동성이 작은 종목을 선정하였습니다.

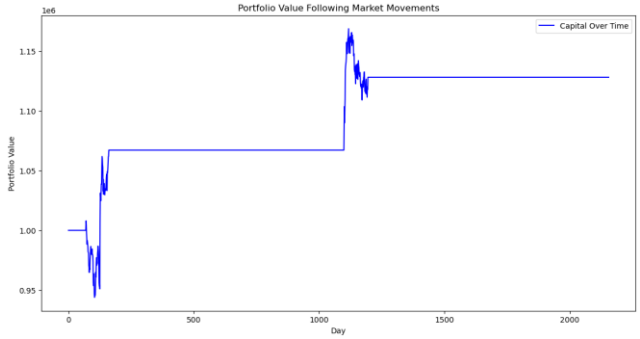


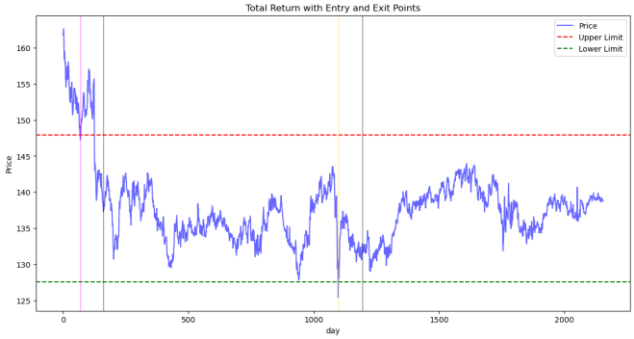
각 세부 통계량의 결과는 다음과 같다.

	ADF-Statistic	p-value	∧	반감기	Hurst-exponent	CAGR/MDD
우리기술	1% ~ 5%	0.05933	0.9963	190.9022	0.5158	6%,24%

4. GBP/XBF(한지원)

영국과 프랑스는 오래 교류한 이웃나라로 인구, GDP, 국방비 등 서로 매우 비슷한 특징들을 보인다. 둘의 환율 또한 비슷하게 수렴할 것이라 생각하여 둘의 환율 데이터를 사용해 백테스트를 진행해보았다.



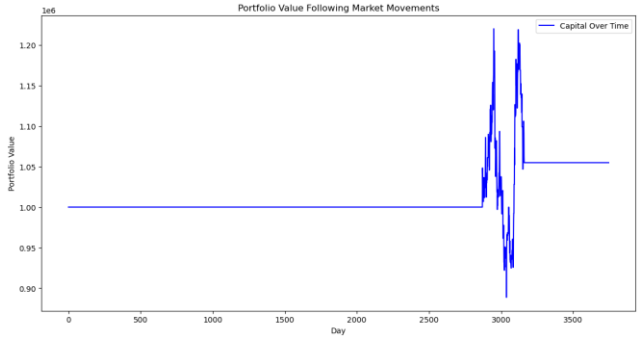


각 세부 통계량의 결과는 다음과 같다.

	ADF-Statistic	p-value	∧	반감기	Hurst-exponent	CAGR/MDD
GBP/XBF	1%	0.000013	0.9927	95.19	0.4916	2%/6.3%

5. 롯데정밀화학(구성희)

정밀화학분야가 오랜 시간동안 큰 변화가 없으며, 원재자에 크게 변동이 없었기 때문에 위 주식이 평균회귀할 것이라 생각했다. 각 세부 통계량의 결과는 다음과 같다. 또한, 정밀화학 분야 중에서 시가총액이 가장 커서 변동성이 가장 적을 거 같아 선정했다.



	ADF-Statistic	p-value	∧	반감기	Hurst-exponent	CAGR/MDD
롯데정밀화학	>10%	0.225	0.997	291.87	0.563	0.5%/27%

결과를 종합해 보면 다음과 같다. 아래의 결과에서 확인 할 수 있듯이 수익률이 꾸준하게 발생하고 있으며, 10 년이상 이라는 장기간동안 MDD 가 비교적 크지 않다는 점에서 평균회귀는 특정 종목에 한해서 시도해볼 만한 전략임을 확인할 수 있었다. 그러나, ADF 통계량의 p-value 의 통계적 유의미성 까지 도달하기 힘들며, 또한 Hurst 지수 또한 우리가 원하는 0.5 미만에 해당하는 값을 가지기는 힘들다. 라는 것 또한 관측할 수 있었다. 거래 횟수 또한 평균적으로 10 년이 넘는 기간동안 2 번~3 번 내지 정도 밖에 발생하지 않은 것으로 보아 이는 실전에서 투입하기에는 다소 무리가 있는 전략임 또한 관측할 수 있었다.

	ADF-Statistic	p-value	λ	반감기	Hurst-exponent	CAGR
기업은행	> 10%	0.34408	0.9973	259.4557	0.5703	4.07%/20%
엔화	5% ~ 10%	0.31074	0.9946	129.8416	0.5053	2.9%/4.8%
휘발유	5% ~ 10%	0.00176	0.9961	180.2137	0.6285	5.79%/17%
우리기술	1% ~ 5%	0.05933	0.9963	190.9022	0.5158	6%,24%
GBP/XBF	1%	0.000013	0.9927	95.19	0.4916	2%/6.3%
롯데정밀화학	>10%	0.225	0.997	291.87	0.563	0.5%/27%

종합적으로, 평균회귀는 특정 종목에 한해서 나름 괜찮은 전략일 수는 있지만, 통계적으로 유의미한 값을 보장해주기는 힘들다는 점. 그리고, 거래빈도가 너무 낮아 CAGR 이 낮게 나와 실전에서 써먹긴 힘들다는 점이 단점으로 존재한다. 따라서, 우리는 평균회귀가 어떻게 현대적으로 발전해왔는지 chapter2 Practical application of mean reversion 에서 분석해 볼 예정이다.

Chapter 2. Practical application of mean reversion

앞서 평균회귀의 통계적 검정 방식과 한계점에 대해 알아 보았다. 현대 금융경제학에서는 이에 대한 한계점을 인식하고 다양한 방식으로 평균회귀의 응용에 대해 연구해 왔다. 예를들어 주가의 이동평균선을 기준으로 회귀할 것이다. 라는 볼린저밴드 이론부터, 공적분 관계에 있는 두 주식은 두주식의 회귀선으로 회귀 할 것이다. 라는 pair trading 까지 다양한 종류의 전략이 연구되어 왔다. 지금부터는 발전된 전략들에 대해서 알아보고, 현대 주식시장에서도 여전히 유용한 전략인지를 검증해볼 예정이다.

8.Bollinger band

볼린저밴드란 1980 년대 초반 미국의 재무분석가인 존 볼린저가 개발한 주가 기술적 분석 도구로, 주가의 변동에 따라 상하밴드의 폭이 같이 움직이게하여 주가의 움직임을 밴드 내에서 판단하고자 고안된 주가지표를 말한다. 기존 지표들이 적절한 매매시기를 알려주지 못한다는 단점을 보완하기 위해 볼린저밴드는 가격변동띠를 탄력적으로 변화시켰다.

볼린저밴드는 상한선과 하한선을 경계로 등락을 거듭하는 경향이 있다는 것을 기본 전제로 한다. 또한 주가의 약 95%가 볼린저밴드 내에서 수렴과 발산을 반복하며 형성된다. 밴드의 폭이 이전보다 상대적으로 크거나 줄어들 경우, 과매수 또는 과매도 상태로 해석할 수 있다.

볼린저밴드 계산 방법

볼린저밴드는 이동평균선(중심선)과 이를 중심으로 한 상단 밴드(상한선)와 하단 밴드(하한선), 총 세 개의 선으로 구성되어 있다. 이 세 개의 선은 대부분의 유가증권가격의 움직임을 포착할 수 있도록 설계되었다. 이동평균선을 추세중심선으로 사용하며 상하한 변동폭은 추세중심선의 표준편차로 계산한다. 따라서 가격변동성 분석과 추세 분석을 동시에 수행할 수 있다.

각 밴드는 다음과 같이 계산한다.

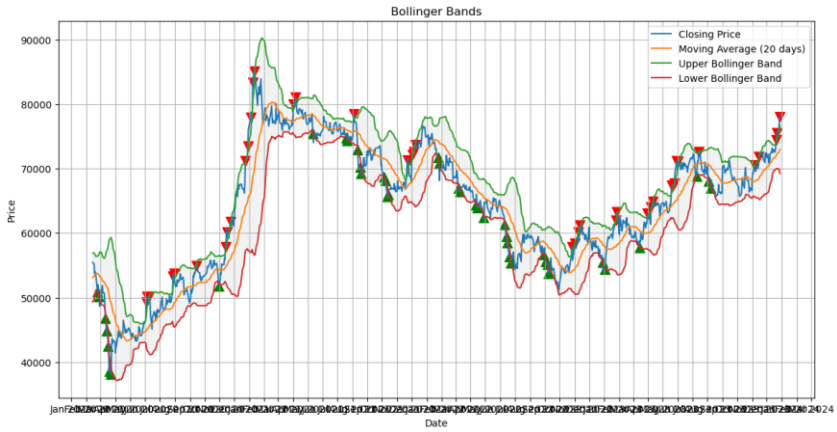
- 중간 밴드 : n일간의 단순 이동 평균 SMA(n), (SMA: Simple Moving Average)
- 상단 밴드 : SMA(n) + 표준 편차 * 2
- 하단 밴드 : SMA(n) – 표준 편차 * 2

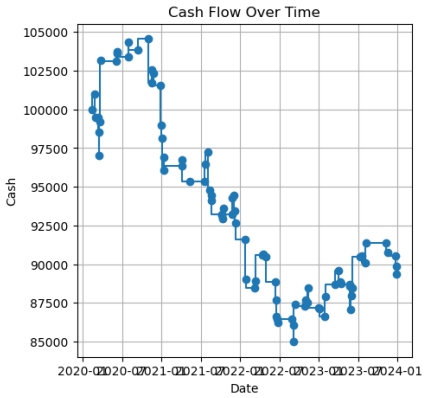
볼린저밴드를 활용한 매매 전략

앞서 볼린저밴드는 상한선과 하한선을 경계로 등락을 거듭하는 경향이 있다는 것을 기본 전제로 한다 을

밝혔다. 따라서 주가가 하한선에 닿으면 다시 상승할 확률이 높으므로 매수하고, 상한선에 닿으면 하락할 확률이 높으므로 매도한다. 주가의 약 95%가 볼린저밴드 내에서 수렴과 발산을 반복하며 형성되지만 5%의 확률로 선을 뚫고 움직일 가능성을 염두해야 한다. 주가가 상한선을 뚫고 상승한다면 이를 과매수라고 판단하고 매도하고, 하한선을 뚫고 하락한다면 이를 과매도로 보고 매수한다.

우리는 우리나라 kospi 의 대표 주식인 삼성전자를 기준으로 잡아 볼린저 밴드 매매를 실험해보았다. 세부 매매 방법은 앞서 말한 것 처럼 주가가 2 표준편차 위에 존재할 경우 과매수 상태라 판단하여 매도하고, 2 표준편차 아래에 존재할 경우 과매도 상태라 판단하여 매수하였다. 보유기간은 스프레드가 강한 정상성을 띄고 있어 앞서 언급한 스프레드의 ouprocess 하의 반감기의 두배로 잡았으며, 20%이상 손실이 발생할 경우 손절하는 것을 원칙으로 삼았다. 테스트 결과는 다음과 같다.





백테스트 결과 CAGR 이 -2.87%, MDD 가 18.71%가 나왔다. 아쉽게도 볼린저밴드의 경우 강한 추세를 띄고 있는 경우 평균으로 회귀하기전에 보유기간이 종료되거나 이동평균선이 매수 시점보다 상승 or 하락 하여 손실을 자주 보는 경우가 발생하였다. 데이터가 강한 추세를 띄고 있는 경우 볼린저밴드 만으로는 매매하였을 경우 강한 수익을 보기 힘들다는 것을 알 수 있었다.

9. Engle and Granger Cointegration Test

9-1. 공적분

시계열에서 두 변수 간의 관계를 나타내어주는 통계적 지표로는 공분산(covariance)과 공적분(cointegration)이 있다. 그렇다면, 공분산과 공적분 간의 차이는 무엇일까?

공분산은 변수 하나의 개념에서 적용하던 분산의 개념을 변수 두 개에 대한 개념으로 확장시킨 것이다. 공분산 식은 다음과 같다.

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

식을 들여다보았을 때, 공분산은 X, Y 편차의 곱이라는 것을 알 수 있다. 즉, 두 변수 간의 직접적인 관계를 의미한다. 반면에, 공적분은 두 변수 사이의 장기적인 균형 관계가 존재하는 지의 유무를 알려주는 지표이다.

위에서 언급한 것처럼, 공적분은 시계열 데이터에서 발생하는 장기적 관계를 분석하는 통계적 개념이다. 일반적으로 시계열 데이터는 추세나 계절성과 같은 특징을 가지는 경우가 있다. 이러한 데이터는 정상성을 가지지 않으므로, 통계적 분석을 할 때 문제가 될 수 있다. 이때 사용되는 개념이 공 적분 개념이다.

공적분은 두 개 이상의 시계열 변수가 장기적으로 관련되어 있다는 것을 나타낸다. 예를 들어, 추세를 띄고 있는 두 개의 비 정상 시계열 데이터 x_t, y_t가 있다고 가정하자. 이 두 시계열을 1 차 차분했을 때 정상 시계열이 된다면, 두 비 정상 시계열의 선형결합으로 이루어진 시계열 $z(t)$는 정상 상태를 따른다. 이때 x_t, y_t는 공 적분 관계에 있다고 한다.

$$If\ x_t \sim I(1), y_t \sim I(1)\ then\ ax_t + by_t \sim I(1) \qquad (2.1)$$

위의 식 (2.1)처럼 일반적으로 적분 차수가 둘 다 1 일 때, 이 둘의 선형 결합의 적분 차수도 $I(1)$인 것이 일반적이다. 그러나 $ax_t + by_t \sim I(0)$ 이 되는 예외적인 경우도 있는데, 이 경우를 두 시계열이 공 적분 되었다고 한다. 이를 일반적으로 표현하면 다음과 같다.

모든 시계열 변수를 모은 $(k \times 1)$벡터 U 가 있을 때, 모든 시계열 변수의 적분 차수가 $I(d)$이고 , b>0 에 대해 선형 결합 $\alpha U \sim I(d-b)$일 때, U 의 성분이 되는 시계열 변수는 공적분 되었다고 한다. 이때 α는 시계열을 공 적분 시키는 '공적분 벡터'이다.

그렇다면 공적분 관계의 유무를 확인 할 수 있는 공적분 검정에는 어떤 것들이 있을까? 대표적으로 앵글 그레인저 공적분 검정, 요한슨 검정이 있다.

9-2. Engle Granger cointegration

앵글 그레인저 공 적분 검정을 위해 제일 먼저 시계열 자료의 적분 차수를 검토한다. 이때 적분 차수는 단위 근 검정을 하였을 때, 안정적이지 않을 때는 1 번 차분을 하고 차분된 시계열을 다시 단위 근 검정을 시행한다. 그리고 차분된 시계열에 단위근이 없다고 검증이 되면 이 시계열의 적분 차수는 $I(1)$이라고 할 수 있다. 그 다음 장기적 (균형) 방정식을 구하고 이 방정식의 잔차를 추출하여 이 잔차에 ADF 검정을 시행한다. 이때, 이 잔차가 정상성이 있다고 확인이 되면 공 적분 관계가 존재한다는 것을 의미한다.

$$y_t = \rho y_{t-1} + \epsilon_t,\ -1 \le \rho \le 1 \qquad (2.2)$$

ϵ_t는 백색 소음으로, 오차 항입니다. 식 (2.2)를 차분하면 다음과 같다.

$$\Delta y_t = (\rho - 1) y_{t-1} + \epsilon_t = \delta y_{t-1} + \epsilon_t \qquad (2.3)$$

만약 단위 근 검정에 따라 δ가 0 보다 작다면, 단위근이 없으므로 위 시계열의 적분 차수는 $I(1)$입니다. 똑같이 단위 근 검정을 통해 적분 차수가 $I(1)$인 것을 확인한 다른 시계열 x_t가 있다고 하자. 이때 이 두 시계열의 선형 결합은 안정적이어야 한다.

$$y_t + \beta x_t = z_t \qquad (2.4)$$

식(2.4)에서 z_t는 안정적이다. β는 공적분 벡터이고 이 공적분 벡터를 알고 있는 경우 바로 ADF 검정을 시행하여 안정성을 검정할 수 있다. 반면에 공적분 벡터값을 모를 경우, OLS(ordinary least square) 회귀분석을 통해 추정된 잔차에 안정성 검정을 시행해야 한다.

$$y_t = c + \beta_1 x_t + \epsilon_t \quad (2.5)$$

위는 OLS 회귀분석을 통해 추정한 장기적 (균형)방정식입니다. 추정된 잔차는 다음과 같다.

$$\epsilon_t = y_t - c - \beta_1 x_t \quad (2.6)$$

이 잔차에 ADF 검정을 시행한다.

$$test\ H_0: \rho = 0\ \ against\ \ \rho \neq 0$$

$$\Delta\epsilon_t = \rho\epsilon_{t-1} + \mu + \pi_t \quad (2.7)$$

식 (2.7)에서 ρ가 0 이라면 ϵ_{t-1}에 영향을 받지 않으므로 정상성을 나타내므로, 이 때 x_t, y_t는 공적분 관계에 있다고 한다.

10. Spread trade

1. 기본 스프레드 매매

기본 스프레드 매매 전략은 두 자산의 가격 차이가 일정 범위에서 벗어났을 때, 이 차이가 평균으로 회귀할 것이라는 가정에 기반한다. 트레이더들은 이러한 스프레드가 역사적 평균으로 회귀할 것이라고 가정하고, 스프레드가 통상적 범위를 벗어나면 해당 자산을 매도하거나 매수하여 이익을 얻으려고 시도한다. 이때 사용하는 수식은 다음과 같다.

$$S_t = p_{1,t} - \beta p_{2,t}$$

여기서 St 는 시간 t 에서의 스프레드, p1,t 와 p2,t 는 각각 시간 t 에서의 두 자산의 가격이다. 계수 β 는 두 자산 가격 사이의 관계를 나타내며, 일반적으로 OLS 회귀분석을 통해 결정된다. 스프레드 St 가 장기 평균 μ 에서 벗어날 때, 트레이더들은 다음과 같이 매매를 진행한다.

$S_t > \mu+\sigma$: $p_{1,t}$ 를 매도하고 $p_{2,t}$ 를 매수합니다.
$S_t < \mu-\sigma$: $p_{1,t}$ 를 매수하고 $p_{2,t}$ 를 매도합니다.

여기서 σ 는 스프레드의 표준 편차로, 통상적 범위를 설정하는 데 사용된다. 이러한 전략은 스프레드가 평균 μ 로 회귀할 것이라는 가정 하에 실행되며, 시장의 변동성에 대응하여 수익을 추구한다.

예를 들어, 특정 통화 쌍의 스프레드를 계산하는 경우, 한 통화가 상대적으로 다른 통화보다 고평가되었을 때, 고평가된 통화를 매도하고 저평가된 통화를 매수함으로써 스프레드가 정상적인 범위로 돌아올 때까지 기다린 후, 반대 포지션을 취하여 이익을 실현한다.

2. 비율 스프레드 매매

기본 스프레드 매매를 한 단계 발전시킨 비율 스프레드는 두 주가의 단순 비율로 만든 스프레드이다. 이 전략에서는 가격 비율이 일정 범위를 벗어났다가 다시 돌아올 것을 예측하여, 비율이 높은 자산을 매도하고 비율이 낮은 자산을 매수한다. 비율이 평균으로 회귀하면 거래를 청산하여 이익을 취한다. 두 주가가 Cointegration 관계에 있으면 비율 스프레드는 추세 성분이 없는 정상성을 갖기 때문에 페어 트레이딩이 가능하다.

$$R_t = \frac{p_{1,t}}{p_{2,t}}$$

여기서 Rt 는 시간 t 에서의 비율 스프레드이다. 이 비율 Rt 이 통계적으로 정상 상태를 나타내며, 일정 범위를 벗어났다가 다시 돌아올 것을 예측하여 다음과 같이 거래한다.

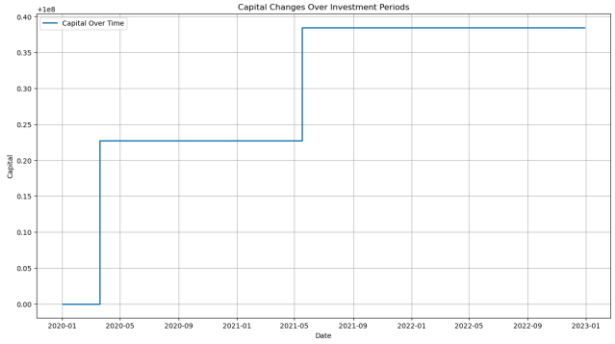
R_t 가 상단 경계를 초과하면, 비율이 높은 자산을 매도하고 비율이 낮은 자산을 매수합니다.
R_t 가 하단 경계를 하회하면, 비율이 높은 자산을 매수하고 비율이 낮은 자산을 매도합니다.

우리는 백테스트에서 로그비율 스프레드를 사용할 예정이다. 로그비율 스프레드를 사용하는 주된 이유는 금융 자료의 비선형성을 효과적으로 다루고 통계적 특성을 개선하기 위해서다. 로그 비율 스프레드를 사용한다면 가격 비율의 변동이 큰 경우에도 스프레드가 과도하게 확장되거나 축소되는 것을 방지할 수 있다. 또한, 로그비율 스프레드는 백분율 변화를 강조하여, 크고 작은 가격 스케일 간의 비교를 용이하게 해준다. 이는 자산 간의 상대적 가격 변화를 분석할 때 중요한데, 예를 들어 두 자산의 가격이 비슷한 경우나 매우 다른 경우에도 일관된 방식으로 비교할 수 있게 도와 준다.

3. 실증 분석

해당 분석에서는 상대적으로 유사하다고 생각되는 동족 산업군에서 페어를 선정하였다. 같은 산업군에 속하기에 유사한 시장 동향과 경쟁 환경에 놓이는 등의 공통 요인이 많기 때문이다. 산업군은 가격의 분산이 크지 않은 증권업으로 그리고 증권업중 대형 증권사 종목인 미래에셋증권과 키움 증권으로 백테스트까지 진행하였다. 백테스트 방식은 가격의 로그비율 스프레드의 2 표준편차를 기준으로 스프레드가 너무 벗어난 경우 과매도된 주식을 매수 과매수된 주식을 매도 하는 방식으로 진행하였으며, 스프레드가 평균값으로 돌아온 경우 포지션을 종료하였다.

분석 결과 아래의 왼쪽 그림과 같은 스프레드를 얻었으며, 노란색 세로선이 진입시점이 된다. 거래 2 일차에 진입하여 8 일간 보유하고 51 일차에 진입하여 23 일간 보유하였다. 오른쪽 그림과 같은 포트폴리오 가치를 보이며 CAGR 은 2.3181%임을 알 수 있다.



자산	ADF-Statistic	p-value	Λ	반감기	CAGR
미래/키움	> 5%	0.315550	0.9677	21.45776	2.3181%

11. VAR - Johanson Cointegration Test

1. VAR(Vector Autoregressive) 모형

VAR(Vector Autoregressive, 벡터 자기회귀) 모형은 Sims(1980)가 고안한 것으로 경제 시계열을 내/외생 변수의 구분없이 적용할 수 있는 다변량 시계열 모형이다. 기존의 ARIMA 모형과 비교하면 ARIMA 모형의 경우 과거 자신의 변동을 설명하는 모형이지만, VAR 은 다른 변수의 변동에 시차까지 고려하여 변동을 설명할 수 있는 모형이다. VAR 모형은 서로 인과관계가 있는 변수들의 현재 관측치를 종속변수로 하고 자신과 여타 변수들의 과거 관측치들을 독립변수로 구성한 n 개의 선형방정식을 통해 시계열을 추정하는 방법이다.

VAR 모형을 추정할 때 사용하는 시계열은 안정성이 보장되지 않은 경우 가성회귀(spurious regression)에 의한 잘못된 결과를 얻을 수 있기에, 단위근 검정과 공적분 검정을 통한 안정성 및 모형의 적합성 검증을 한 후 추정해야 한다. VAR 모형은 아래와 같다.

$$X_t = A(L)X_t + \varepsilon_t = \sum_{k=1}^{\infty} A_k X_{t-k} + \varepsilon_t = \sum_{k=1}^{l} A_k X_{t-k} + \varepsilon_t \quad (X_t \text{는 } n \times 1 \text{ 벡터})$$

VAR 모형을 추정할 때 우도비(likelihood ratio) 검정통계량, AIC(Akaike Information Criterion)와 SC(Schwarz Criterion)를 이용하여 변수의 차수를 적정수준으로 제한하는 것이 필요하다. 원래의 VAR 모형은 시차 k 가 무한대이지만 실제 추정에서는 회귀오차(ε_t)가 평균이 0 이고 분산이 일정한 백색잡음(white noise)에 가까워지는 시점에서 차수를 l로 제한한다.

VAR 모형은 모형내 변수들의 시차관계를 이용하여 예측뿐만 아니라 충격반응 함수를 이용한 파급효과분석도 가능하다는 장점이 있으며, 따라서 거시경제변수에 영향을 미치는 다양한 충격들의 상대적 중요성 분석이나 미래 예측치를 추정하기 위한 도구로서 널리 사용된다.

2. 요한슨 검정(Johanson Cointegration Test)

앞서 보았듯 VAR 모형 추정시에 시계열의 안정성을 검증하지 않을 경우 가성회귀와 같은 문제가 발생하기에 단위근 검정과 공적분 검정이 필요하다고 하였는데, 그 공적분 검정 중에 하나가 요한슨 검정(Johanson Cointegration Test)이다.

요한슨이 1991 년 개발한 공적분 검정법으로, VAR 모형에 대한 가설검정을 통해 적분계열간 안정적인 장기 균형관계가 존재하는지를 점검하는 방법이다. 단위근 검정이 종속변수(ΔY_t)와 설명변수(Y_{t-1})간의 상관관계 존재유무를 나타내는 φ의 유의성을 파악하는 것이라면, 요한슨 검정에서는 두 벡터 ΔY_t와 설명변수 Y_{t-1} 간의 정규 상관계수(ρ)를 분석하여 통계량을 산출한다. 즉 일반적인 VAR(ρ) 모형을 따르는 벡터 시계열 Y_t의 구성변수간 공적분 관계가 존재하면, 요한슨 검정법은 아래와 같이 VAR(ρ) 모형의 변형식을 검정식으로 사용한다.

$$\Delta Y_t = \Pi Y_{t-1} + A_1^* \Delta Y_{t-1} + A_2^* \Delta Y_{t-2} + \cdots + A_p^* \Delta Y_{t-p+1} + \varepsilon_t$$

$$A_j^* = -(A_{j+1} + A_{j+2} + \cdots + A_p)$$

$$\Pi = A_1 + A_2 + \cdots + A_p - I_n$$

여기서 $\Pi = 0$ 여부를 검정하는 것이 바로 요한슨 공적분검정이다. $\Pi = 0$라는 것은 결국 $A_1 + A_2 + \cdots + A_p = I$라는 의미가 되고, I는 단위행렬이 된다. 따라서 다음과 같이 귀무가설과 대립가설을 설정할 수 있다.

$$H_0 : rank(\Pi) \leq \mathrm{r}, \;\; H_1 : rank(\Pi) \geq \mathrm{r} + 1$$

특정한 공적분 차수(rank)를 가정하고 이를 기반으로 검정통계량을 계산하는데, 위 검정시에는 t 나 F 통계량이 아닌 우도비(likelihood ratio:LR) 통계량을 이용한다. 검정 결과 귀무가설을 기각하는 경우, 설정한 공적분 차수보다 높은 차수가 나온것이며, 해당 변수들 간의 공적분 관계가 존재한다는 의미가 된다.

3.1 실증분석 방법론

우선 공적분 관계를 나타낼 만한 페어를 선정한다. 선정 기준은 동종 산업의 종목들로 구성하거나 지주사-계열사 관계 등 밀접한 관련성을 가질 것으로 판단되는 종목들로 3 개씩 페어를 선정하였다. 페어를 선정한 이후 요한슨 검정을 통해 개별 자산 간의 공적분 관계를 검정한다. 공적분 관계를 확인한 후 최적의 헤지 비율을 계산하는데, 요한슨 검정을 통해 구한 eigenvector 를 활용한다. 여기서 Eigenvector (아이젠벡터)는 각각의 고유값(Eigenvalue, 아이젠밸류)에 대응되는 고유공적분 관계에 대한 선형 결합을 나타내며, 시계열 데이터의 선형 관계를 설명해준다. 아이젠벡터를 통해 계산된 헤지 비율을 사용하여 스프레드를 만들고, ADF 검정을 통한 정상성을 확인에 이어 반감기 또한 구해준다. 이후로는 스프레드가 상단 및 하단 임계값을 벗어나는 경우 진입하는 것으로 포지션을 계산하고 백테스트 결과를 확인한다. 앞선 평균회귀 테스트 방식과 동일한 전략의 백테스트를 통해 유효성을 검증해 볼 예정이다. 사용한 데이터의 경우 2023 년 1 월 1 일부터 2024 년

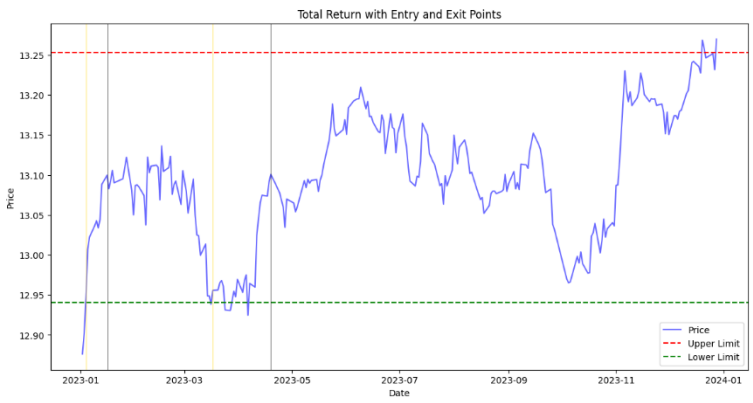
1 월 1 일까지의 데이터를 사용하였다.

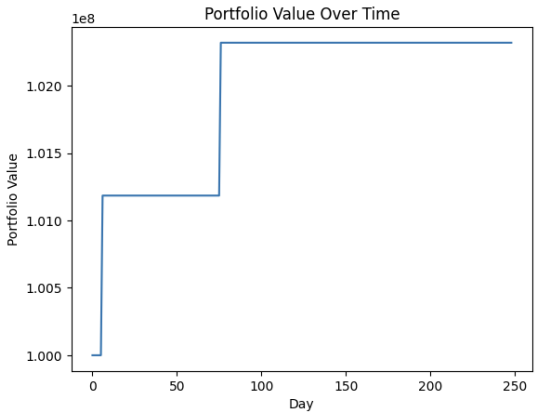
3.2 실증분석

3.2.1 NH 투자증권, 미래에셋증권, 삼성증권 – 동종 산업

해당 분석에서는 동종 산업군에서 종목을 추출하여 페어를 선정하였다. 같은 산업군에 속하기에 유사한 시장 동향과 경쟁 환경에 놓이는 등의 공통 요인이 많기 때문이다. 조선업과 기술분야, 식품산업 등에서 분석을 해보았지만, 그 중 공적분 관계가 잘 보였던 증권사 종목으로 백테스트까지 진행하였다.

분석 결과 아래의 왼쪽 그림과 같은 스프레드를 얻었으며, 노란색 세로선이 진입시점이 된다. 거래 2 일차에 진입하여 8 일간 보유하고 51 일차에 진입하여 23 일간 보유하였다. 오른쪽 그림과 같은 포트폴리오 가치를 보이며 CAGR 은 2.3181%임을 알 수 있다.



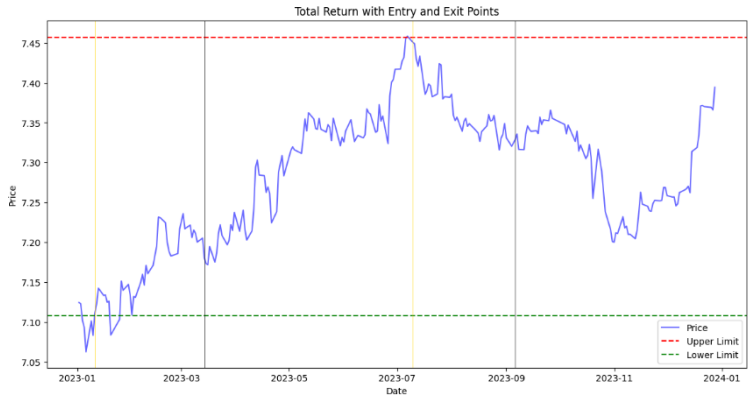


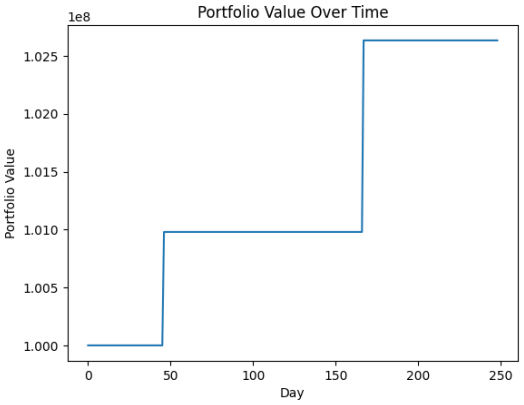
r	ADF-Statistic	p-value	Λ	반감기	CAGR
>= 1	> 10%	0.315550	0.9729	25.6228	2.3181%

3.2.2 현대자동차, 현대건설, 현대인프라코어 – 동일 그룹

다음 분석에서는 동일 그룹에 속한 계열사 간의 공적분 관계가 높을 것이라 예측되어 현대 그룹의 계열사들로 페어를 구성하여 분석을 진행하였다. 삼성, SK, 두산 등의 다양한 그룹의 계열사로 검정을 진행해 보았으나, 현대의 경우 계열사의 수부터, 상장된 경우도 많아 검정에 용이하였으며, 따라서 그 중 공적분 관계가 잘 나타났던 세 종목으로 백테스트를 진행해 보았다.

분석 결과 아래의 왼쪽 그림과 같은 스프레드를 얻었으며, 노란색 세로선이 진입시점이 된다. 거래 7 일차에 진입하여 41 일간 보유하고 128 일차에 진입하여 41 일간 보유하였다. 오른쪽 그림과 같은 포트폴리오 가치를 보이며 CAGR 은 2.6329%임을 알 수 있다.



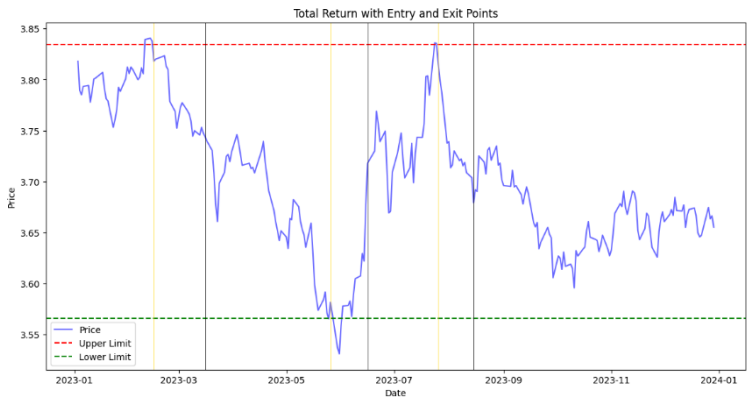


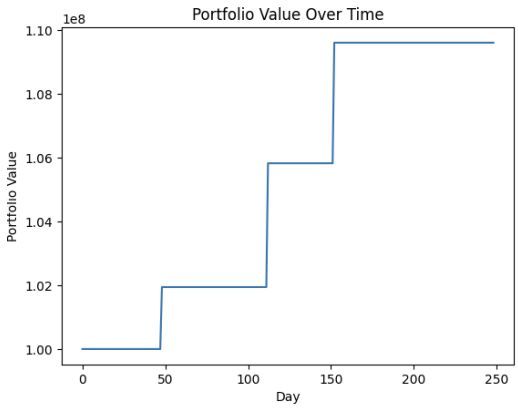
r	ADF-Statistic	p-value	Λ	반감기	CAGR
>= 1	> 10%	0.3986	0.9830	40.6785	2.6329%

3.2.3 콩, 밀, 옥수수 – 원자재

위 세 상품에 대한 검정은 기후변화와 같은 생산 요인이나, 축산업 및 에너지 산업과의 연관성 등 수요 및 공급 변동에 있어 밀접한 관계가 있을 것이라 유추되어 진행하였다.

분석 결과 아래의 왼쪽 그림과 같은 스프레드를 얻었으며, 노란색 세로선이 진입시점이 된다. 거래 30 일차에 진입하여 20 일간 보유, 100 일차에 진입하여 14 일간 보유하고, 거래 140 일차에 진입하여 14 일 보유하는 것으로 총 3 번의 진입시점이 결정되었다. 오른쪽 그림과 같은 시간에 따른 포트폴리오 가치를 보이며 CAGR 은 9.6136%임을 알 수 있다.





r	ADF-Statistic	p-value	Λ	반감기	CAGR
>= 1	> 10%	0.166005	0.9636	19.0450	9.6136 %