시계열 모형 복습 및 파생상품의 가격 예측하기

황정민

1. 시계열 분석 모델(전통적인 모델)

1-1) 시계열의 구성성분

* 시계열의 구성성분 – 추세성분(Trend), 계절성분(Season), 순환성분(Cycle), 불규칙성분(Irregular)
* 추세성분: 시계열의 장기적인 변화 추세로 상승 경향 또는 감소 경향의 상태
* 계절성분: 1년을 단위로 전년과 유사하게 발생하는 단기 변동요인으로 주기에 따라 순환하며 변함.
* 순환성분: 2년~10년 주기를 가지고 순환하는 중기 변동요인
* 불규칙성분: 위의 세 가지 요인으로 설명하고 남은 변동으로 시계열 내에서의 “Noise”.

가법모형: $Z_t = (T_t + S_t + C_t + I_t)$

승법모형: $Z_t = (T_t * S_t * C_t * I_t)$

1-2) 정상성

* 정상성: 시계열의 확률적인 성질들이 시간의 흐름에 따라 불변함을 의미. 시계열이 정상시계열이라면 뚜렷한 추세가 존재하지 않고, 시계열의 진폭이 시간의 흐름에 따라 변화하지 않는다. 정상성에는 강정상성과 약정상성이 있는데 주로 약정상성이 많이 사용된다. 시계열이 다음 3가지 조건을 만족시키면 약정상을 지닌다고 말할 수 있다.

1)모든 t에대해 평균이 일정

2)모든 t에대해 분산이 일정하고, 유한

3)두 시점 사이의 자기공분산은 시점과는 무관하며 시차에만 의존

* 만약 시계열이 비정상 시계열이라면 적절한 변환을 통해 정상시계열로 만들어 줘야한다.

1) 평균이 일정하지 않음(추세가 존재): 차분을 사용(추세성분의 제거, 계절성분을 제거하지는 못한다.)

2) 분산이 일정하지 않음: 로그변환, 제곱근 변환 등을 사용

3) 계절성이 존재하는 존재 하는 경우: 계절차분을 사용

1-3) 시계열 모델

*AR모형(자기회귀모형): 시계열 Z_t를 그 이전 시점의 시계열(Z_{t-1}, Z_{t-2},...)로 회귀시킨 모형으로 p차 자기회귀모형 AR(p)는 다음과 같습니다. 즉, Z_t가 Z_{t-1}, Z_{t-2},..., Z_{t-p} 만의 선형결합, 즉(t-p) 시차 까지만 영향을 받는다고 가정하고 설정된 모형입니다.

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \phi_3 Z_{t-3} + \cdots + \phi_p Z_{t-p} + a_t, \quad (a_t\text{는 } iid\text{인 } white\ noise)$$

*MA 모형(이동평균모형): 시계열 Z_t를 그 이전 시점의 백색잡음과정(White Noise Process)로(a_{t-1}, a_{t-2},...)로 회귀시킨 모형으로 q 차 자기회귀모형 MA(q)는 다음과 같습니다.

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \cdots - \theta_q a_{t-q}$$

*ARMA 모형(자기회귀이동평균모형): AR 모형과 MA 모형을 결합시킨 모형으로, ARMA(p, q)는 다음과 같이 나타낼 수 있습니다.

$$\begin{aligned} Z_t = {} & \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \phi_3 Z_{t-3} + \cdots + \phi_p Z_{t-p} \\ & + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \cdots - \theta_q a_{t-q} \end{aligned}$$

*ARIMIA(누적자기회귀이동평균모형): 원시계열 Z_t를 d 번 차분한 시계열이 ARMA(p, q)를 따를 때, Z_t는 ARIMA(p,d,q)를 따른다고 합니다.

이 밖에도 ARIMA 모형에 계절성분을 고려한 SARIMA 모형, ARIMA 모형에 외생변수(X)를 포함시킨 ARIMAX 모형, ARMA모델을 다변량 시계열 모델에 적합하도록 확장시킨 VARMA 모형 등이 사용되고 있습니다. 또한 대부분의 금융 경제 시계열은 이분산성을 지니는데 이를 반영한 ARCH 모형과 GARCH모형도 금융 시계열 분석에 많이 사용되고 있습니다.

2. 시계열 분석 모델(순환신경망)

2-1) 순차데이터와 순환신경망

*순차데이터: 순서가 의미가 있으며, 순서가 달라질 경우 의미가 손상되는 데이터. 시계열 데이터가 대표적인 순차데이터의 예시이다.
*순환신경망(RNN): MLP에서는 가중치 행렬 W에 과거의 정보를 반영할 수 없었지만, 순차데이터에서는 이를 반영해 줄 필요가 있다는 고민에서 출발한 모형. RNN에서는 MLP와 달리 과거의 정보를 반영할 수 있도록 설정.

2-2) RNN의 다양한 종류

* Vanilla RNN: 가장 기본적인 형태의 RNN이다. 순차 데이터의 길이가 길어질수록 미분값이 너무 커지거나 작아질 수 있어 불안정해지기 쉽다. (기울기 소실/폭발의 문제가 발생)
* LSTM: 중간에 C라는 새로운 cell state를 통해 이전 정보와 현재 정보를 함께 요약하는 RNN이다. 길이가 긴 순차 데이터를 처리하는 데 한계가 있는 Vanilla RNN을 보완할 수 있다.
* GRU: LSTM과 달리 cell state가 없고, output gate를 사용하지 않는 RNN이다. Network의 파라미터가 적다는 장점이 있다.

3. CASE STUDY2: 파생상품의 가격 예측하기

3-1) 배경

파생상품의 가격을 추정하는 방법에는 Black-Sholes Formula를 비롯해 여러 방법들이 있습니다. 그렇지만 각각의 방법들은 현실과 맞지 않는 가정을 기반으로 하고 있거나 가격을 추정하기 위해서 많은 데이터와 자료를 필요로 하는 등의 단점들을 가지고 있습니다. 그렇기 때문에 이번에는 머신러닝을 사용해 파생상품의 가격을 예측해보도록 하겠습니다.

3-2) 옵션과 블랙-숄즈 방정식

* 파생상품: 기초자산의 가격에 따라서 가격이 변동하는 상품. 선물, 선도거래, 옵션 등이 파생상품의 대표적인 예이다.
* 옵션: 기초자산을 정해진 가격에 구매/판매할 수 있는 권리. 구매/판매할 수 있는 가격을 행사가격이라고 한다. 구매할 수 있는 권리를 콜옵션(Call), 판매할 수 있는 권리를 풋옵션(Put)이라고 한다. 옵션을 구매하는 사람(롱포지션)은 옵션을 판매하는 사람(숏포지션)에게 일정한 금액을 지불하는 데 이를 옵션 프리미엄이라고 하고, 이것이 옵션의 가격이 된다.
* 선도거래: 미래의 특정 시점에 기초자산을 현재 정한 가격으로 구매/판매하도록 하는 거래. 롱포지션은 돈을 주고, 기초 자산을 사는 포지션이고, 숏포지션은 돈을 받고, 기초자산을 팔고자 하는 포지션입니다. 그리고 선도거래가 장내화 된 것을 선물거래라고 합니다.
* 옵션과 선도거래의 차이: 선도거래는 롱포지션과 숏포지션에 있는 사람 모두가 권리와 의무를 동시에 지니고 있다. 즉, 미래 특정 시점이 되면 손해를 보더라도 사전에 정한 가격에 거래를 해야 한다. 이와 달리, 옵션은 롱포지션에 있는 사람은 권리만, 숏포지션에 있는 사람은 의무만을 지닌다. 즉, 롱포지션에 있는 사람은 자신에게 이득이 되는 시점에서

옵션을 행사하면 되지만 숏포지션에 있는 사람은 롱포지션에 있는 사람이 옵션을 행사하면 반드시 거래를 해야 한다.

* 예1) A(롱포지션)가 행사가격(X)가 $15인 콜옵션 1개을 B사(숏포지션)로부터 $3(옵션 프리미엄)에 구입했다.
상황1) 만기 시점에 기초자산의 가격이 $10인 경우: A는 권리를 행사하지 않는다. 그러므로 A의 payoff는 -3(옵션 프리미엄)이다. B사의 payoff는 3(옵션프리미엄)
상황2) 만기 시점에 기초자산의 가격이 $16인 경우: A는 권리를 행사한다. A의 payoff는 16-15-3=-2이고, B사의 payoff는 15-16+3=2 이다.
상황3) 만기 시점에 기초자산의 가격이 $30인 경우: A는 권리를 행사한다. A의 payoff는 30-15-3=12이고, B사의 payoff는 15-30+3=-12

옵션의 구매자(롱포지션)은 옵션 프리미엄을 지불하는 대신, 기초자산의 가격 변동에 따른 최대 손실을 제한함과 동시에 기초자산의 가격이 일정 수준보다 높아지면 이득을 볼 수 있게 된다. 옵션의 판매자(숏포지션)은 옵션 프리미엄을 받는 대신 기초자산의 가격이 변동함에 따라 무제한적인 손실이 발생할 수 있는 위험을 감수하게 된다.

* Black-Sholes Formula
블랙숄즈 방정식은 옵션의 가격(의 기댓값)을 산출하는 방정식으로, 주식가격(S), 행사가격(K), 변동성(σ), 만기까지의 기간(T), 무위험 이자율(r) 만을 가지고, 옵션의 가격을 예측할 수 있다. 변동성은 과거의 데이터들로부터 추정하거나(역사적변동성), 현재 옵션의 가격을 사용해 역으로 추정해 볼 수 있다(내재변동성).

$$C = SN(d_1) - Ke^{-rT}N(d_2)$$

$$P = Ke^{-rT}N(-d_2) - SN(-d_1)$$

$$d_1 = \frac{\ln(\frac{S}{K}) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

C : 콜 프리미엄
P : 풋 프리미엄
S : 기초자산 가격
K : 행사가격
r : 무위험 이자율
σ : 기초자산의 변동성
T : 잔존만기
N(x) : 표준정규분포의 누적밀도 함수, 표준정규분포를 따르는 변수가 x보다 작을 확률

다만, 가정들 중 현실과 다른 것들이 존재한다. 몇 가지만 살펴보면
* 무위험이자율과 변동성은 상수이다.
* 주식의 수익률은 정규분포를 따른다(주식의 가격이 로그 정규분포를 따른다)
* 거래 비용이 없다.
* 차익거래가 존재하지 않는다.
* 원하는 만큼 돈을 빌리거나 빌려줄 수 있고, 주식을 사고 팔 수 있다.

3-3) 머신러닝을 사용해 옵션 가격 예측하기
(주피터 노트북 파일 참조)