

시계열 분석

1. 정상성, 차분

1.1 정상성

- 정상성 : 시간의 흐름에 따라 관찰된 값들을 시계열 자료 분석에 기초가 되는 개념
- 정상성 만족하는 조건
 1. 평균이 일정하다
 2. 분산이 시점에 의존하지 않는다
 - :: 3. 공분산은 단지 시차에만 의존하고 실제 어느 지점 t, s 에 의존하지 않는다
- 정상성 여부 확인
 1. ACF와 PACF 그래프로 확인
 - 정상 시계열 : ACF가 빠르게 0으로 떨어진다, PACF는 어느 시점에서 절단점을 가진다(0으로 수렴)
 - 비정상 시계열 : ACF가 느리게 감소
 2. ADF 단위근 검정
 - 귀무가설 : 비정상 시계열이다
 - 대립가설 : 정상 시계열이다
 - p-value 값이 0.05보다 작으면 정상 시계열

* ACF(자기 상관함수)

* PACF(편자기 상관함수) : 간접적으로 영향 주는 부분 버리고 직접적 영향 주는 부분의 상관만 추출

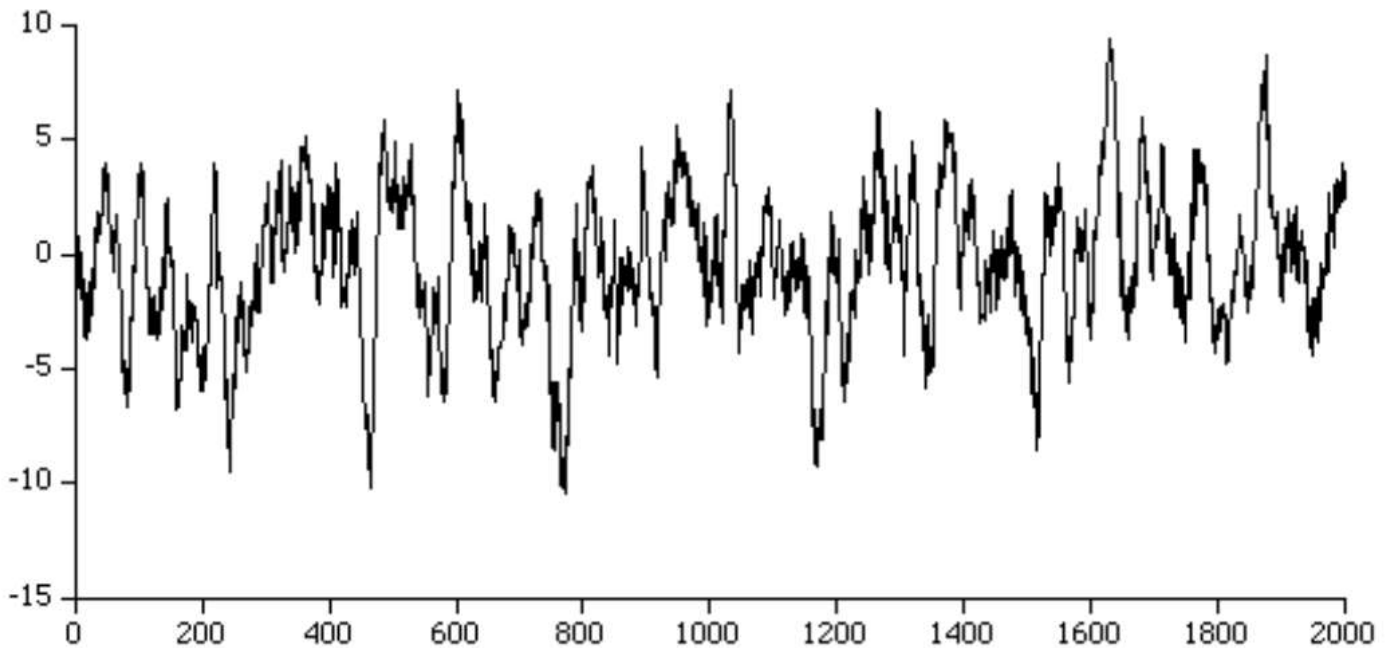
1.2 차분

- 차분
 - 비정상 시계열을 정상 시계열로 가공하는 방법
 - 추세가 보이는데, 평균이 일정하지 않다면 차분
 - 시계열의 수준에서 나타나는 변화(추세나 계절성)를 제거하여 시계열의 평균 변화를 일정하게 만든다
- 변환
 - 로그화 / 박스콕스화
 - 분산이 일정하지 않다면 변환
 - 시계열의 분산 변화를 일정하게 만든다

2. 시계열 모형

2.1 자기회귀 모형 (AR 모형)

- $X(t) = \{a * X(t-1) + c\} + u * e(t), \quad e(t) \sim N(0,1)$



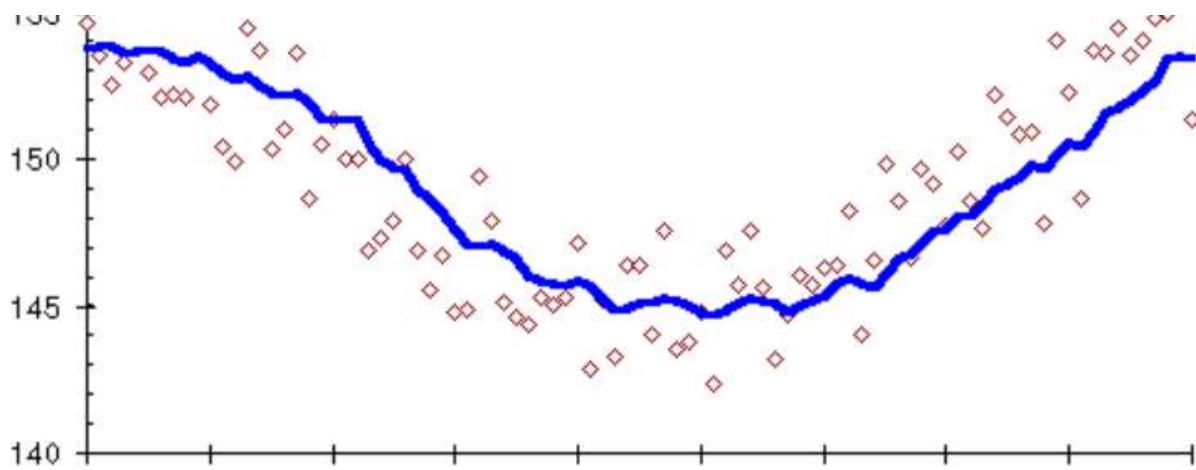
많은 금융정보나 시계열 정보들이 자기 상관의 특징을 잘 가지고 있다. 이 데이터는 평균이 0라는 점 외에도 이전에 양수가 나오면 이후에는 음수가 나올 것이다(반대도 가능)는 일정한 패턴도 예상하게 만든다.

- 정상 시계열 모형
- 이전 시점의 자료가 현재 자료에 영향을 주는 모형
- 종속변수의 이전 값을 이용해서 예측

2.2 이동평균 모형 (MA 모형)

- $X(t) = \{a * e(t-1) + c\} + u * e(t), \quad e(t) \sim N(0,1)$





데이터의 평균값 자체가 시간에 따라 변화하는 경향이 Moving Average이다.

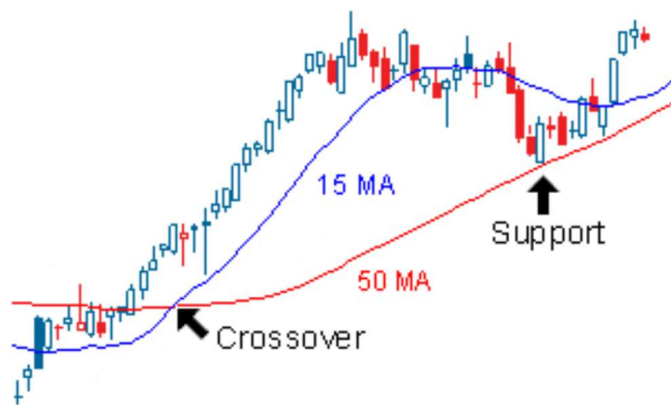


Chart by MetaStock

Copyright © 2005 Investopedia.com

증권가에서 기술적 분석을 할 때 가장 많이 쓰는 것이 MA모형이기도 하다. 예를 들어 최근 50일 평균값보다 최근 15일 이동 평균값이 커지면 주가가 치솟는다, 즉 골든크로스가 발생한다 같은 접근 말이다.

- 정상 시계열 모형
- 시계열의 평균과 이전의 오차를 이용해서 예측
- 과거 잘 알 수 없는 원인이 겹쳐서 현재의 데이터가 생겨났다

2.3 자기회귀이동평균 모형 (ARMA 모형)

- ARMA(1,1) 모형 : $X(t) = \{a \cdot X(t-1)\} + \{b \cdot e(t-1)\} + c + u \cdot e(t)$
- ARMA(2,2) 모형 : $X(t) = \{a_1 \cdot X(t-1)\} + \{a_2 \cdot X(t-2)\} + \{b_1 \cdot e(t-1)\} + \{b_2 \cdot e(t-2)\} + c + u \cdot e(t)$
- 정상 시계열 모형
- AR 모형 + MA 모형 = ARMA(p 자기회귀 차수, q 이동평균 차수)

2.4 자기회귀누적이동평균 모형 (ARIMA 모형)

- 비정상 시계열 모형
- ARIMA(p 자기회귀 차수, d 차분 수, q 이동평균 차수)
- ts() 함수 : 데이터를 시계열 자료로 변환
- diff() 함수 : diff(data, differences=n) 몇번 차분할지
- arima(data, order=c(p,d,q))
- auto.arima(data) 함수를 통해 가장 적절한 arima 모델 찾아준다
- forecast.Arima(data) 함수를 활용해 미래를 예측

2.5 분해 시계열

- 시계열에 영향을 주는 일반적인 요인을 시계열에서 분리해 분석하는 방법
- 회귀 분석적인 방법

1) 추세 요인

그림을 그렸을 때, 형태가 오르거나 또는 내리는 추세 따르는 경우

2) 계절 요인

요일마다, 일년 중 각 월에, 사분기 자료에서 각 분기에 등 고정 주기에 따라 자료가 변화하는 경우

3) 순환 요인

명백한 이유가 없이 알려지지 않은 주기를 가지고 변화하는 경우

4) 불규칙 요인

위 세 가지의 요인으로 설명 할 수 없는 회귀 분석에서 오차에 해당하는 요인