

Week 1

수요예측 - part 1

2021.03.04

개요

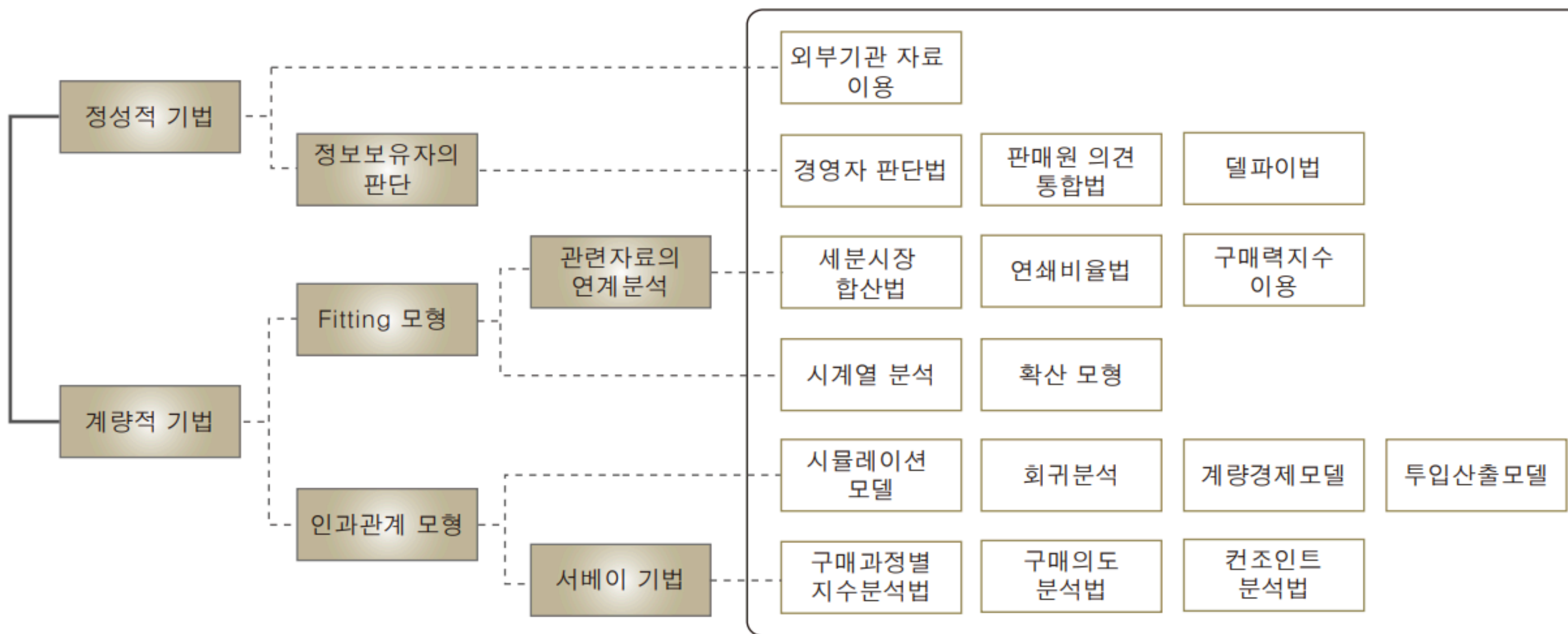
비용과 정확도의 trade-off

- 수요예측 기법
 - 주관적 / 정성적 분석법:
 - 과거 자료가 충분치 않은 경우 주관적인 판단이나 의견에 기초하여 수행하는 예측 방법
 - 소요되는 시간과 비용이 높음. 전문가나 외부 기관으로부터의 정보등을 중요하게 사용
 - 계량적 기법:
 - 동일 제품/서비스의 과거 자료가 존재하는 경우 적용 가능. 신제품의 경우 적용 불가
 - 시계열의 경우 평균에 회귀하거나 설명변수의 변화, 외생 변수의 발생에 민감
 - 단기적인 예측에 주로 활용 가능

개요

비용과 정확도의 trade-off

〈그림 1〉 수요 예측 기법



자료 : 관련자료 종합

개요

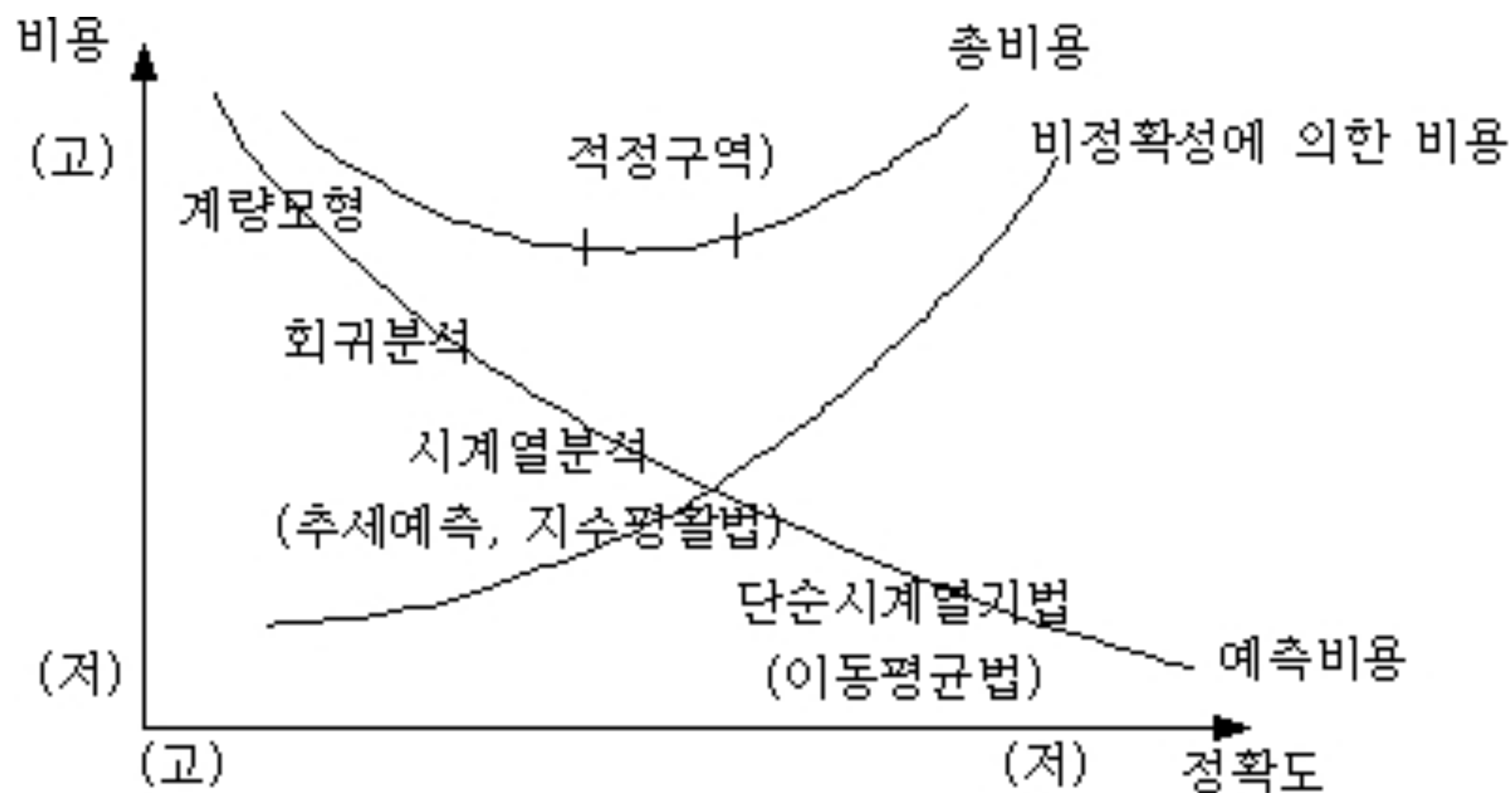
비용과 정확도의 trade-off

〈표 1〉 수요예측 방법론과 그 사용범위

상황 예측기간	동일 서비스의 과거자료 존재	새로운 서비스 혹은 도입 초기의 서비스
단 기	시계열 분석 (Time Series Analysis)	소비자/시장조사 (Consumer - based Approach)
	계량경제모형 (Econometric Model)	주관적 예측 (Subjective Estimation)
장 기	확산 모형 (Diffusion Model)	비교·유추에 의한 방법 (Analogy)

개요

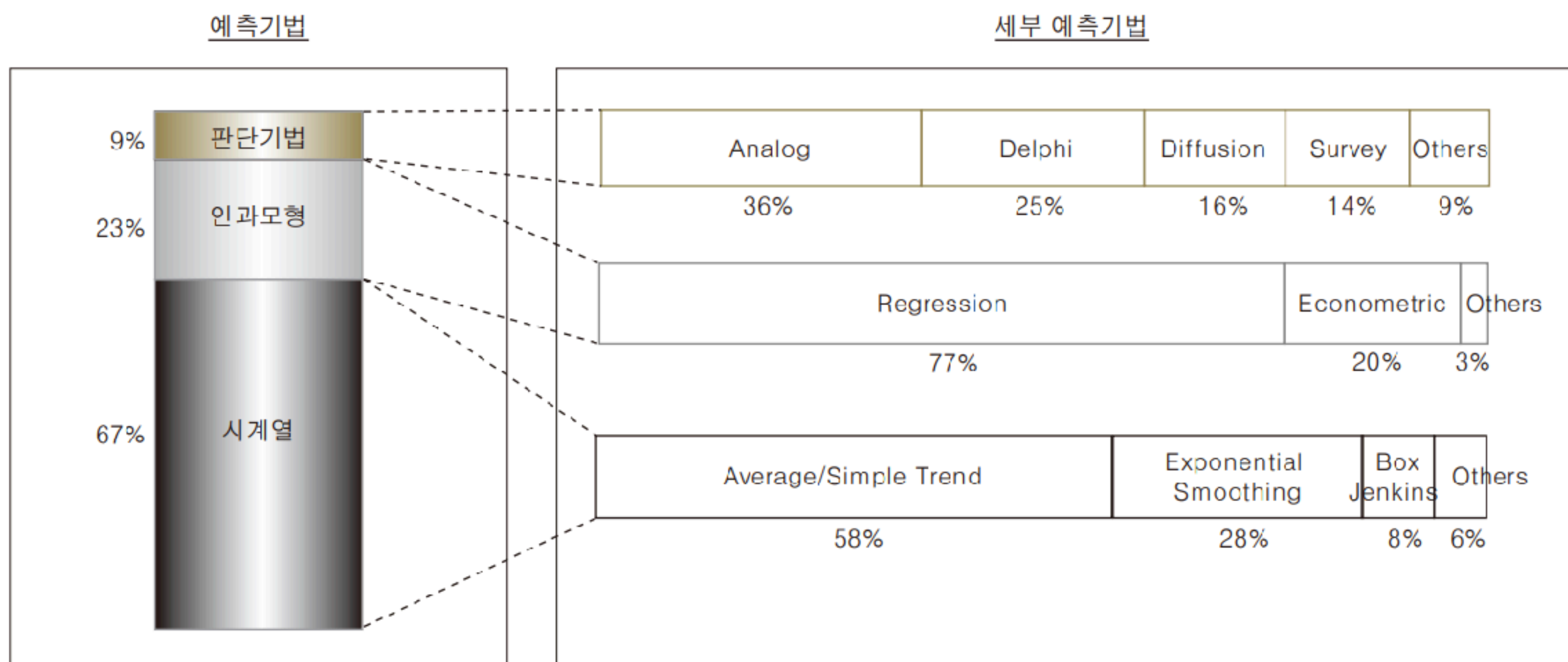
비용과 정확도의 trade-off



개요

비용과 정확도의 trade-off

〈그림 2〉 미국 기업들이 실무에서 사용되는 예측기법



주 : 2004 IBF(Institute of Business Forecast) 주최 컨퍼런스 & 교육 참가자 대상 서베이 결과(N=316개사)

주관적 / 정성적 예측법

중장기 예측에 적용되는 기법

- 델파이법(DelPhi Method): 미래 상황에 대하여 전문가나 담당자로 구성된 위원회를 구성, 개별적 질의를 통해 의견을 수집 후 이를 종합, 분석 및 정리하여, 의견이 일치될때까지 반복
- 시장 / 소비자 조사법: 가장 계량적이고 객관적인 방법으로 소비자로부터 직접 수요에 관한 정보를 얻으려는 방법이며, 시간과 비용이 많이 들지만, 단기 예측 시 비교적 정확한 예측이 가능함
- 경영판단법 (Executive Opinions): 상위층의 경영자들이 모여서 집단적으로 행하는 예측 기법으로 보통 장기 계획이나 신제품 개발을 위해서 사용

주관적 / 정성적 예측법

중장기 예측에 적용되는 기법

- 판매원 의견 통합: 판매원들에게 각자 담당하고 있는 지역의 수요를 예측하도록 하고, 이를 합해서 전체의 수요로 간주하는 예측
 - 경제적 상황 및 경쟁업체 등을 고려해야 함
- 수명주기 유추법 (Life Cycle Analogy): 과거의 자료가 없는 품목 또는 신제품의 수요를 예측하려 할 때 과거의 상황이 미래에도 유사하게 전개된다는 전제하에 제품 수명 주기상의 수요변화를 유추

계량 예측법

단기 예측에 적용되는 기법

- 시계열 모형:
 - 예측기간이 길어질 수록 예측치가 시계열 과정의 평균에 수렴
- 인과형 예측 기법 (Causal forecasting method, 계량경제 모형):
 - 예측하고자 하는 수요와 이에 영향을 미칠 것으로 판단되는 독립변수들 간의 상호관계를 수식화해서 모형으로,
 - 변수들의 불확실성으로 장기적 안정성은 떨어짐
- 확산모형 (Diffusion model):
 - 신제품 도입으로부터 포화상태까지의 과정을 선형적으로 규명하여 장기적 안전성

시계열 분석방법

시계열 분석법

전통적인 방법론

- 이동평균법
- 추세분석법
- 시계열분해법
- 지수평활법

시계열 분석법

시계열의 합성 전략

- 가법 모형: 추세 (T) + 순환변동 (C) + 계절변동 (S) + 우연변동 (I)
- 승법 모형: 추세 (T) * 순환변동 (C) * 계절변동 (S) * 우연변동 (I)

이동 평균법

단순이동평균법

$$f_t = \frac{D_{t-1} + D_{t-2} + \dots + D_{t-n}}{n}$$

가중 이동평균법

$$f_t = w_{t-1}D_{t-1} + w_{t-2}D_{t-2} + \dots + w_{t-n}D_{t-n}, \quad \sum_{i=1}^n w_{t-i} = 1$$

지수평활법

$$F_{t+1} = \alpha D_t + (1 - \alpha)F_t = F_t + \alpha(D_t - F_t)$$

- F_t : 기간 t에서의 시계열 예측치, D_t : 기간 t의 시계열자료 실측
- α : 평활상수로서 $0 \leq \alpha \leq 1$
- **최소의 자료로 단기에측활동에 유용하게 활용할 수 있는 예측기법**
 - 추세, 순환변동, 계절변동이 작용하지 않고 비교적 안정되어 있는 경우 적합
 - 단기에측은 특성상 시계열 요인들이 중요하게 작용하지 않음
 - 시계열 요인이 반영되지 않은 가장 단순한 형태를 단순지수평활법이라고 함
 - 평활상수는 상수로 봄
 - Smoothing 기법의 가장 단순한 형태

적응적 지수평활법

Trigg and Leach (1967)

$$f_t = f_{t-1} + k_t e_{t-1}; \quad e_{t-1} = Z_t - f_{t-1}$$

f_{t-1} : t-1 시점에서 예측한 Z_t 에 대한 예측치

e_{t-1} : 해당 예측오차

$$k_t = \frac{P_t}{Q_t}$$

$$P_t = (1 - \xi)(Z_t - f_{t-1}) + \xi P_{t-1}, \quad 0 < \xi < 1, P_0 \text{ is given}$$

$$Q_t = (1 - \xi)|Z_t - f_{t-1}| + \xi Q_{t-1}, \quad 0 < \xi < 1, Q_0 \text{ is given}$$

P_t : 커질 수도 작아질 수도 있음

Q_t : 계속 커지는 값

구조적인 변화에 빨리 적응할 수 있고, 안정적인 상태에서는 잡음 효과를 제거

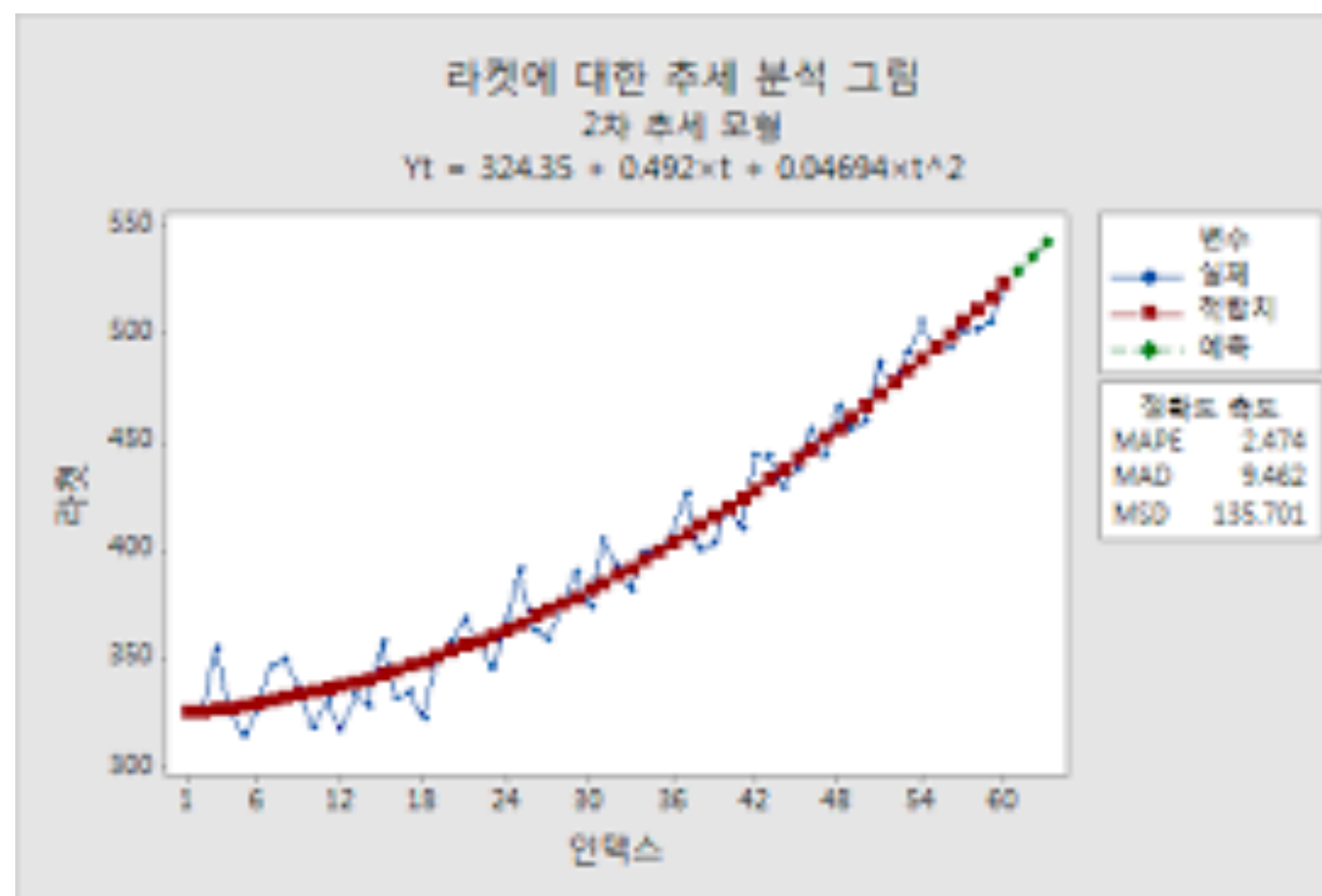
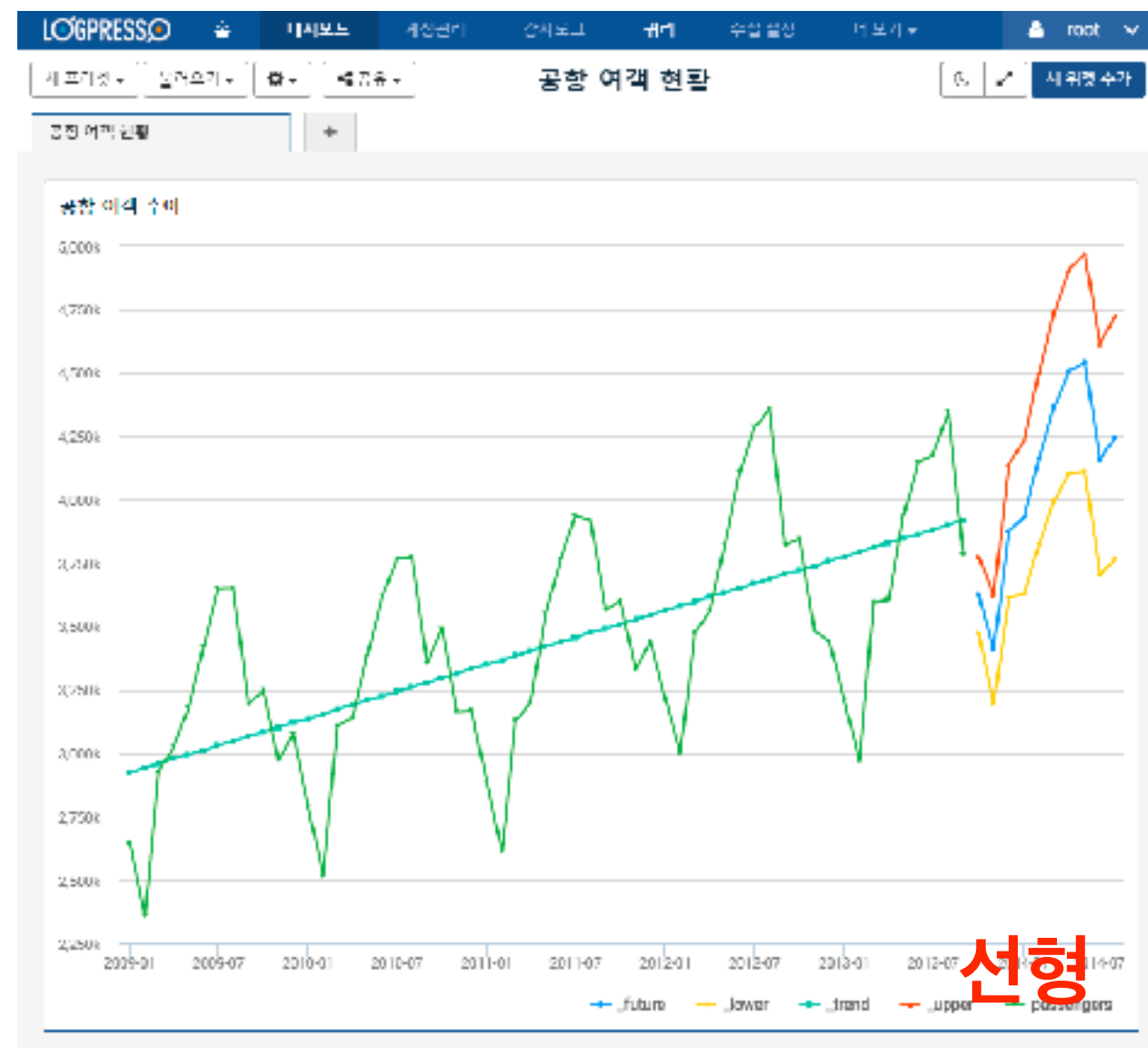
시계열 분석법

시계열 변동 요소

- 추세 변동: 수요의 장기적인 변화 경향
- 순환 변동: 경기변동과 같이, 정치, 경제, 사회적 요인에 의한 변화. 보통 1년 이상의 일정 주기가 없는 사이클 현상
- 계절변동: 어느 제품이나 서비스에 대한 수요가 매년 일정한 패턴으로 변동하는 경우로, 1년 주기로 발생. 월 / 주 주기도 계절변동으로 간주
- 불규칙 변동: 설명될 수 없는 요인 또는 돌발적인 요인에 의하여 일어나는 변화
- 시계열 합성방법
 - 가법 모형: 추세 (T) + 순환변동 (C) + 계절변동 (S) + 우연변동 (I)
 - 승법 모형: 추세 (T) * 순환변동 (C) * 계절변동 (S) * 우연변동 (I)

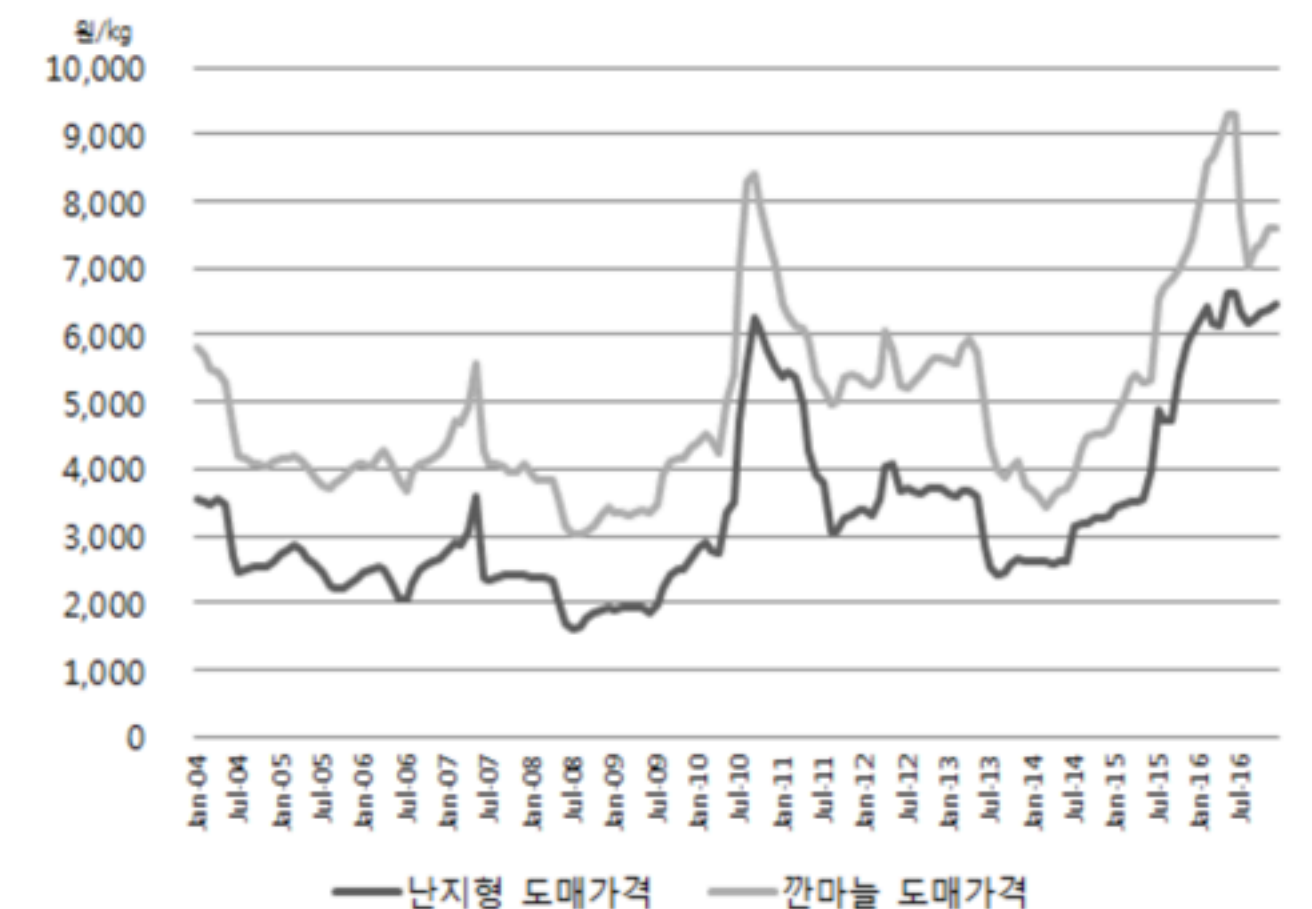
추세분석법

- 오차항의 가정에 대한 고려 없이 회귀분석을 사용
 - 엄밀한 방법이라고 말할 수는 없고, 전체적인 패턴을 볼 수는 있음
 - 시계열 분석에서 추세를 제거하기 위해 활용
 - 패턴에 따라 비선형 회귀 분석을 사용하기도 함
 - 약간 말장난?!



다항함수

그림 2. 마늘 가격 변화 추이



비선형

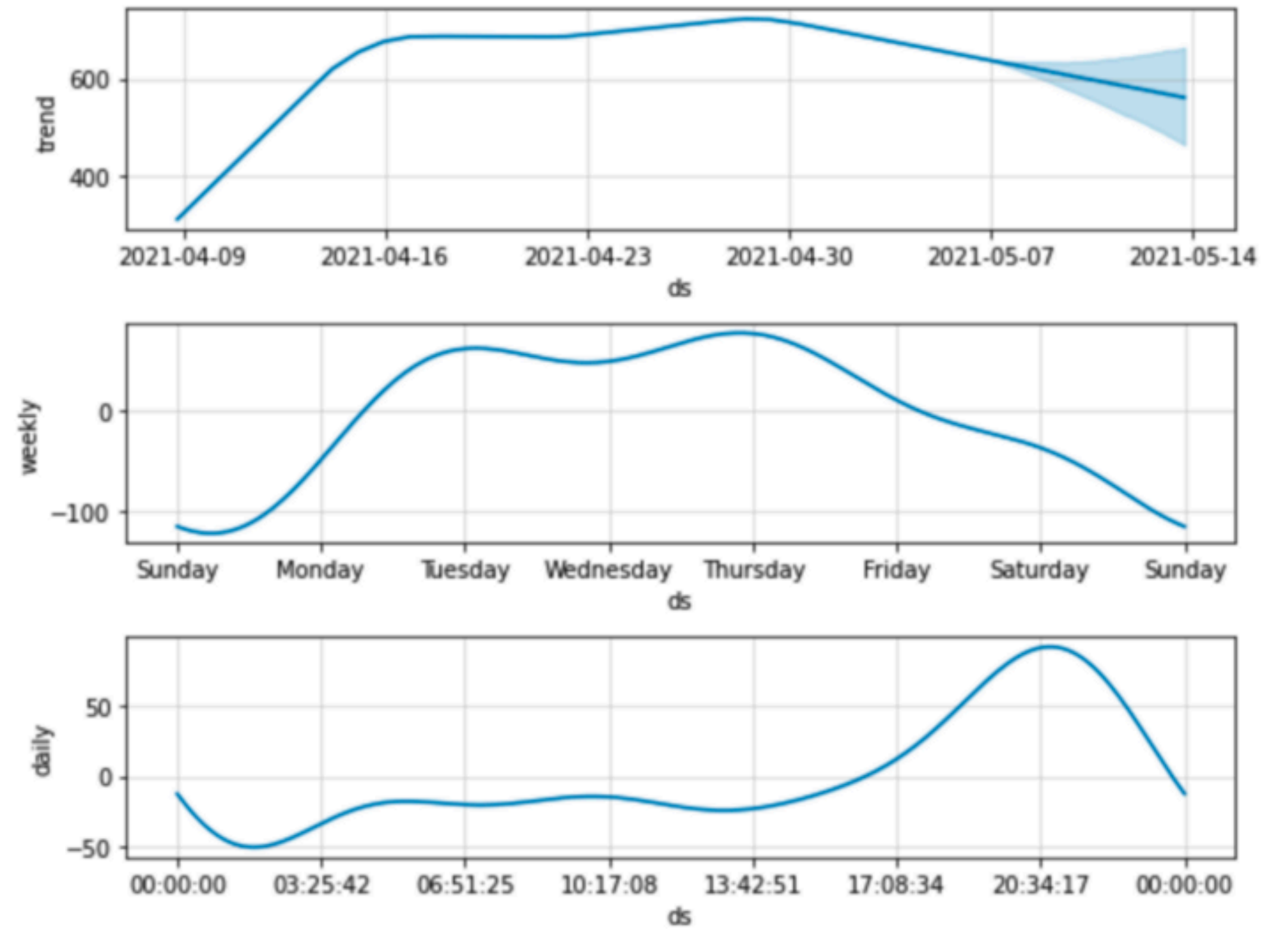
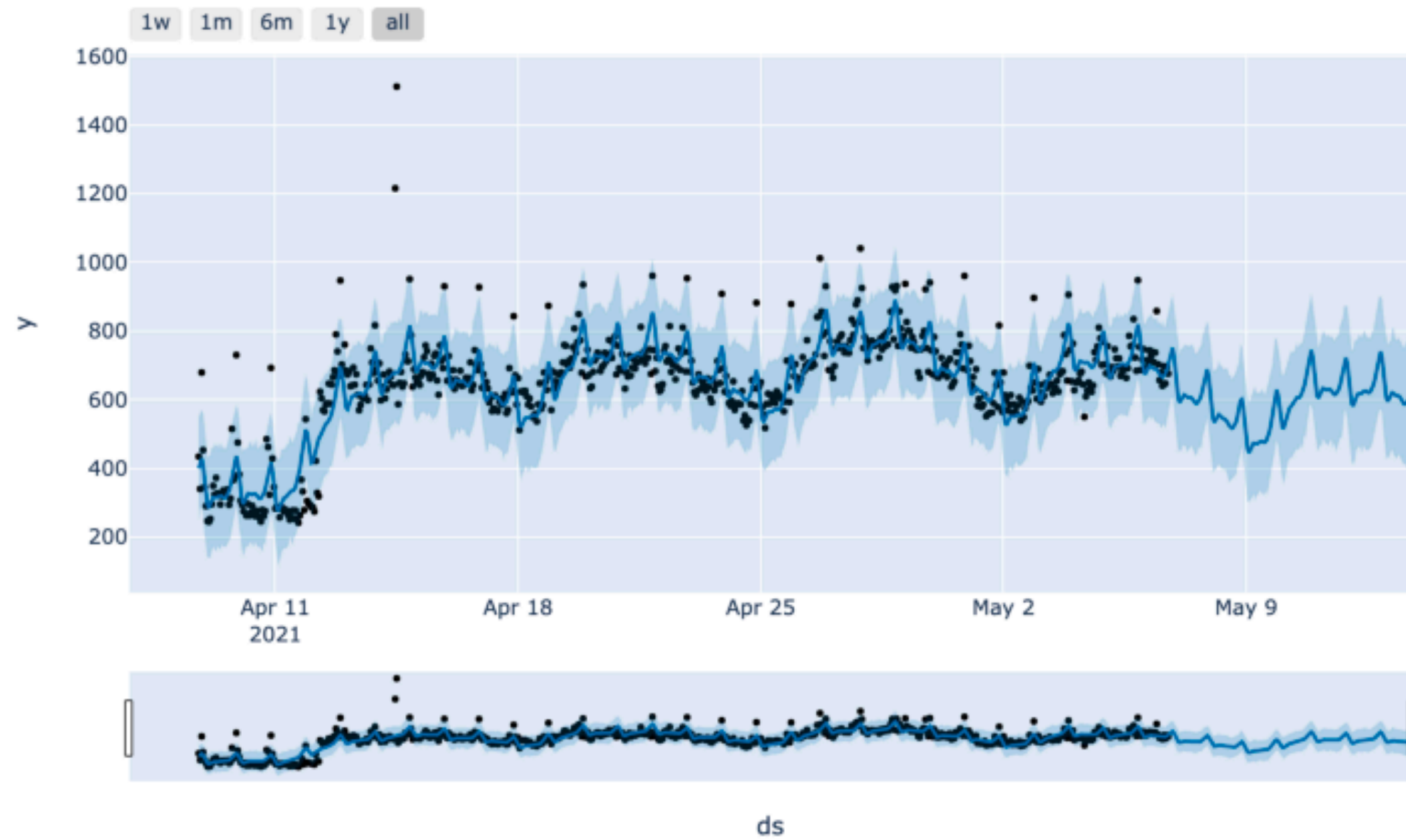
시계열 분석법

분해시계열

- 현업에서 평활법은 수요예측으로 잘 사용하지 않음
 - 실제 충분한 경우도 많으나, 너무 rough하다고 야단맞음!
- 시계열은 추세, 순환, 계절성, 기타 불규칙 변동의 혼합으로 이루어져 있어, 시계열 자료를 형성하고 있는 변동 요소들을 찾아내 시계열 자료를 해당 요소들로 분해 (Decompose)
- 추세 제거에는 회귀분석 방법, 계절성 제거에는 이동평균을 활용
 - Prophet을 사용하면 쉽게 시계열을 분해할 수 있음

시계열 분석법

분해시계열



시계열 분석법

분해시계열

- 독립변수(feature)가 없는 시계열 자료의 경우 현업에서는 가장 많이 사용되는 모형이라고 생각됨
 - 직관적으로 이해하기가 쉽게 설명이 가능
 - 평활법은 별로 얻을 수 있는 정보가 많지 않음
- 시계열(time series)을 횡단면 데이터(cross-sectional)와 구분짓는 가장 큰 특징은 오차항의 상관관계이나
 - 여기에 대한 고려가 없음
 - 경험적으로 이러한 고려가 현실 데이터를 분석하는데 크게 도움이 되지는 않음

시계열 분석법

Box-Jenkins: 오차항의 상관관계를 모델링

- 변수에 관한 정보가 부족하거나, 너무 많은 변수가 영향을 미치고 있는 경우 사용
 - 다른 변수들은 average out 된다고 가정하고 있음
- 역시 시계열의 정상성을 가정하므로, 단기적인 예측에 사용되는 방법임
 - 정상성은 여러 시점의 분포가 시점에 의존하지 않음

$$f(x_0, x_1, x_2, \dots, x_t) = f(x_p, x_{p+1}, \dots, x_{p+t}), \text{Cov}(X_t, X_s) = \text{Cov}(X_{t+h}, X_{s+h}) : \text{강정상성}$$

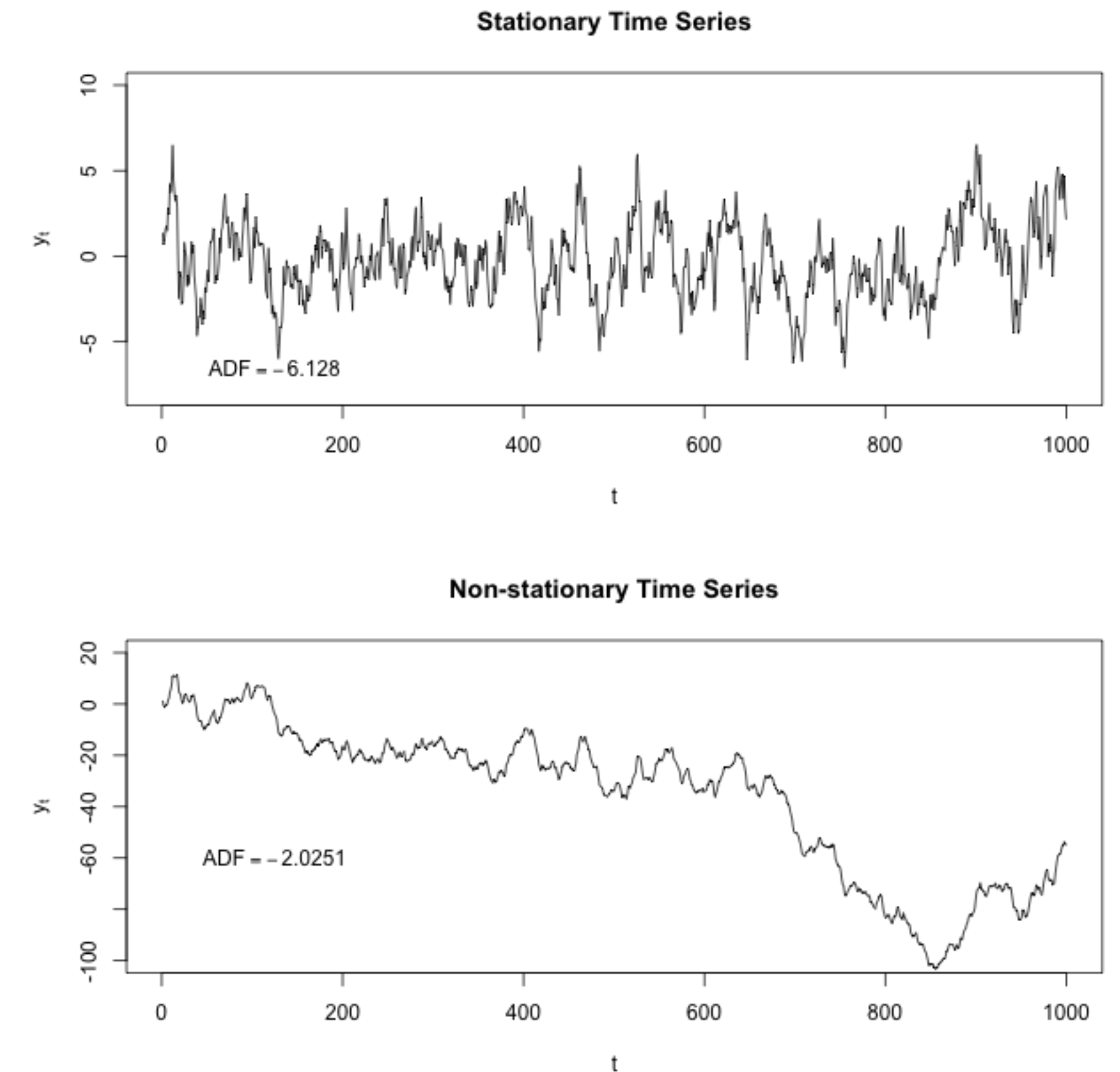
$$E(X_t) = E(X_{t+h}), \text{Cov}(X_t, X_s) = \text{Cov}(X_{t+h}, X_{s+h}) : \text{약정상성}$$

- 추세 없는 안정적인 시계열이라고 이해하면 됨

시계열 분석법

Box-Jenkins: 오차항의 상관관계를 모델링

- 정상 시계열 vs 비정상 시계열
- 현업에서 주로 보는 데이터에서 정상시계열은 없음!
 - 장비에서 나오는 데이터 정도가 정상시계열
 - 마케팅 세일즈 등에서는 거의 없음
- 비정상 시계열은 정상시계열화 해 함
- ACF, PACF를 잘 추정할만큼 긴 시계열도 그렇게 많지는 않음
- 현실에서 Box-Jenkins 방법은 잘 사용하기 어려움



확산모형

확산모형

신제품의 시장도입 초기에 나타난 수요자료를 기초로 상품 성장을 예측

- 생태학에서 세포증식, 전염병 확산 등의 예측을 위해 개발, 상품수명 주기 예측도구로 활용
- 신상품의 누적구매자 수 증가를 S자 형태의 곡선으로 가정
- Bass에 의해 정립: 확산모형을 일반화하여 확산 과정을 다음 요소들의 합으로 정의
 - 시장외적요인: 혁신요인, 기술발전 등
 - 시장내적요인: 모방요인, 구전효과, 경험 효과 등

확산모형

신제품의 시장도입 초기에 나타난 수요자료를 기초로 상품 성장을 예측

- BASS 모형
- 연평균성장률 모형
- Logistic 모형
- Gompertz 확산이론
- Rogers의 혁신확산이론

Bass 확산모형

신제품의 시장도입 초기에 나타난 수요자료를 기초로 상품 성장을 예측

$$n_t = \frac{dN_t}{dt} = p(m - N_t) + \frac{q}{m}N_t(m - N_t), \quad n(0) = pm$$

- m : 잠재구매자
- N_t : 기구매자
- $m - N_t$: t시점의 잠재 수요자수
- p (Innovation Coefficient): 시장 외적 요인 계수,
- q (Imitation Coefficient): 모방계수

Bass 확산모형

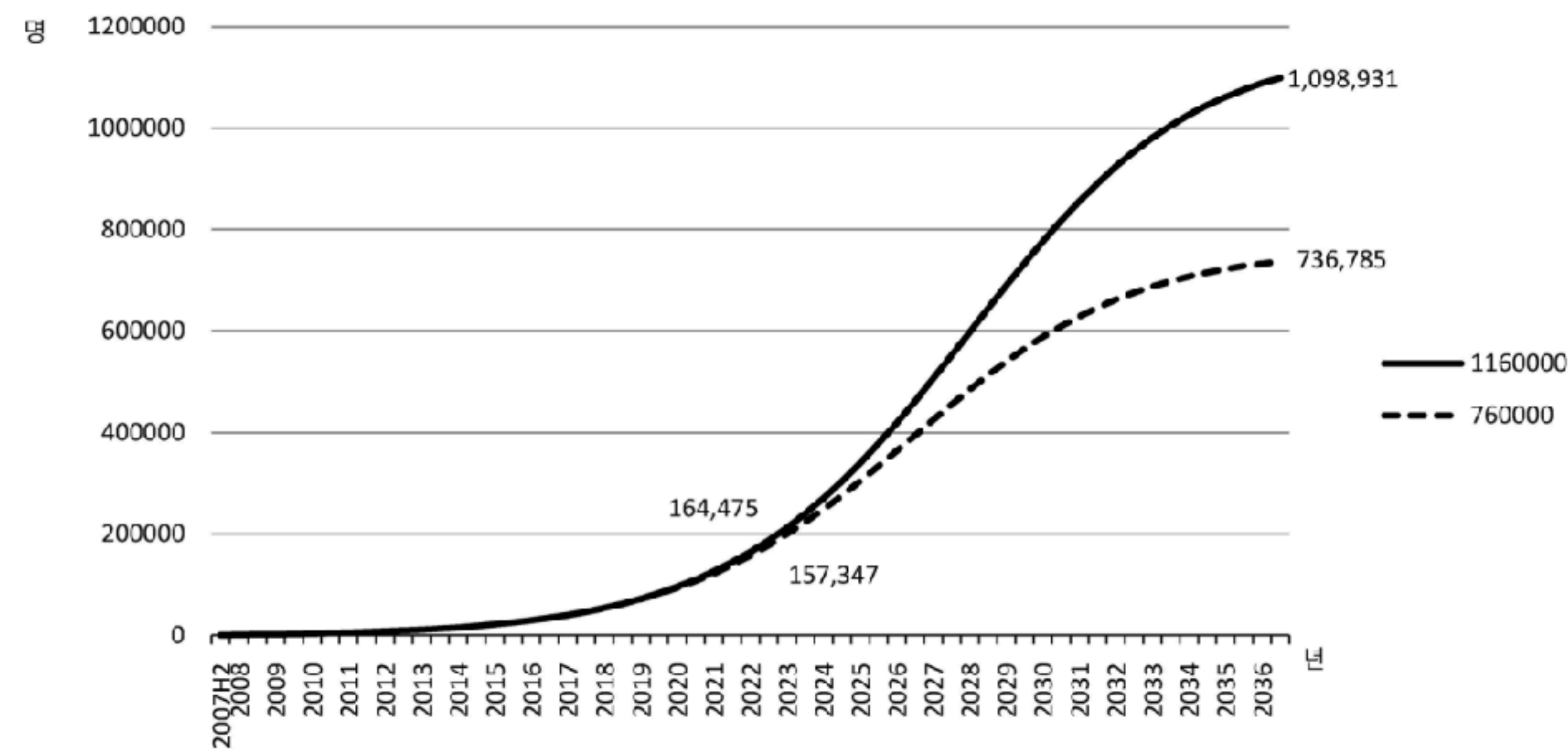
신제품의 시장도입 초기에 나타난 수요자료를 기초로 상품 성장을 예측

$$n_t = \frac{dN_t}{dt} = p(m - N_t) + \frac{q}{m}N_t(m - N_t), \quad n(0) = pm$$

- 잠재 시장의 크기가 달라도 확산형태 계수를 비교함으로써, 확산형태 및 속도 비교 가능
- 추정해야 할 매개변수의 개수가 상대적으로 적어, 매개변수 추정이 용이
- 과거 자료에 전적으로 의존할 필요는 없음
- 매개변수는 여타 신제품 예측기법 (주관적 예측법, 비교/유추에 의한 방법, 소비자조사법) 등으로 추정 가능

Bass 확산모형

신제품의 시장도입 초기에 나타난 수요자료를 기초로 상품 성장을 예측



[그림 1] 20년 후 예측

Logistic 모형

추정이 BASS 모형에 비해 상대적으로 쉬움

$$Z_t = \frac{M}{1 + \exp(a + bt)}$$

- M : 포화상태의 잠재시장 수요
- a, b : 확산속도의 파라미터
- 시장에 대한 자료가 없는 경우, 초기 시장에 대한 설문조사를 통해 포화상태를 추정
- 주어진 포화상태를 이용하여 선형 회귀분석을 통해 시장수요를 추정
- 누적수요량에 대한 자료를 바탕으로 수요 예측으로 신규수요 예측에는 적합하지 않음

Gompertz Growth Curve 모형

정보기술 제품의 성장 모형 추정에 사용되는 대표적인 성장모형

$$Y_t = m \cdot \exp[-\exp(\alpha + \beta t)], t = 1, 2, \dots, T$$

- Y_t : t 시점까지의 누적 수요
- M : 잠재시장의 규모
- α, β : 함수의 위치와 형태를 나타내는 parameter

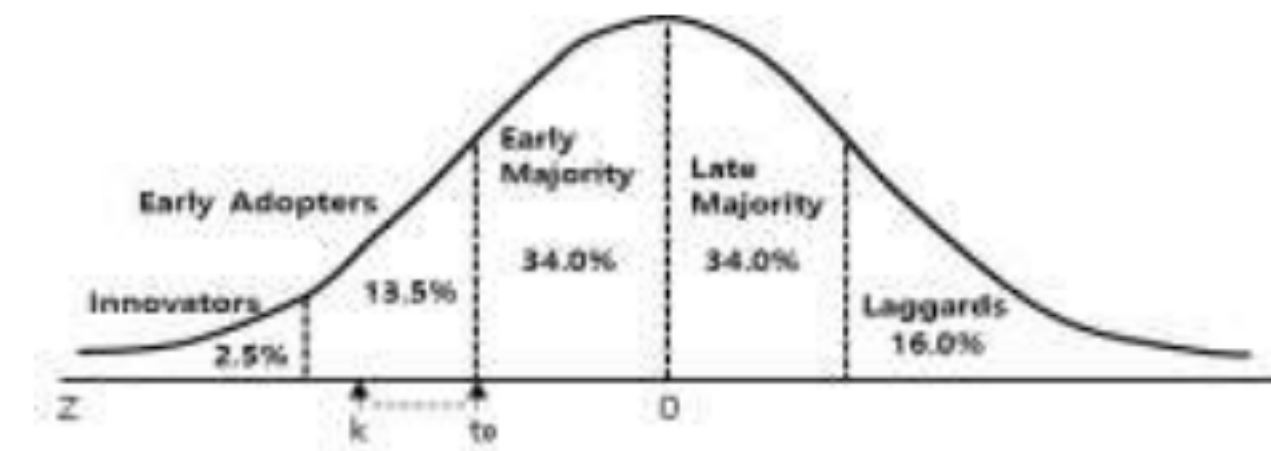


Fig. 2. A conceptual map of factor maturation time in standard normal distribution using Rogers theory

- 최대 성장율은 Y_t 가 약 37%에 이르는 수준에서 발생
- 초기 단계에 상대적으로 빠른 성장률을 보이는 곡선 형태
- 장기적인 예측을 할 경우에는 초기값을 보수적으로 설정해야 하는 한계

Rogers 혁신확산이론

신제품의 시장도입 초기에 나타난 수요자료를 기초로 상품 성장을 예측

- E.M. Rogers가 제안한 혁신의 확산이론은 5가지 분류의 기술수용집단을 분류하여 설명
 - Innovators
 - Early Adopters
 - Early Majority
 - Late Majority
 - Laggards

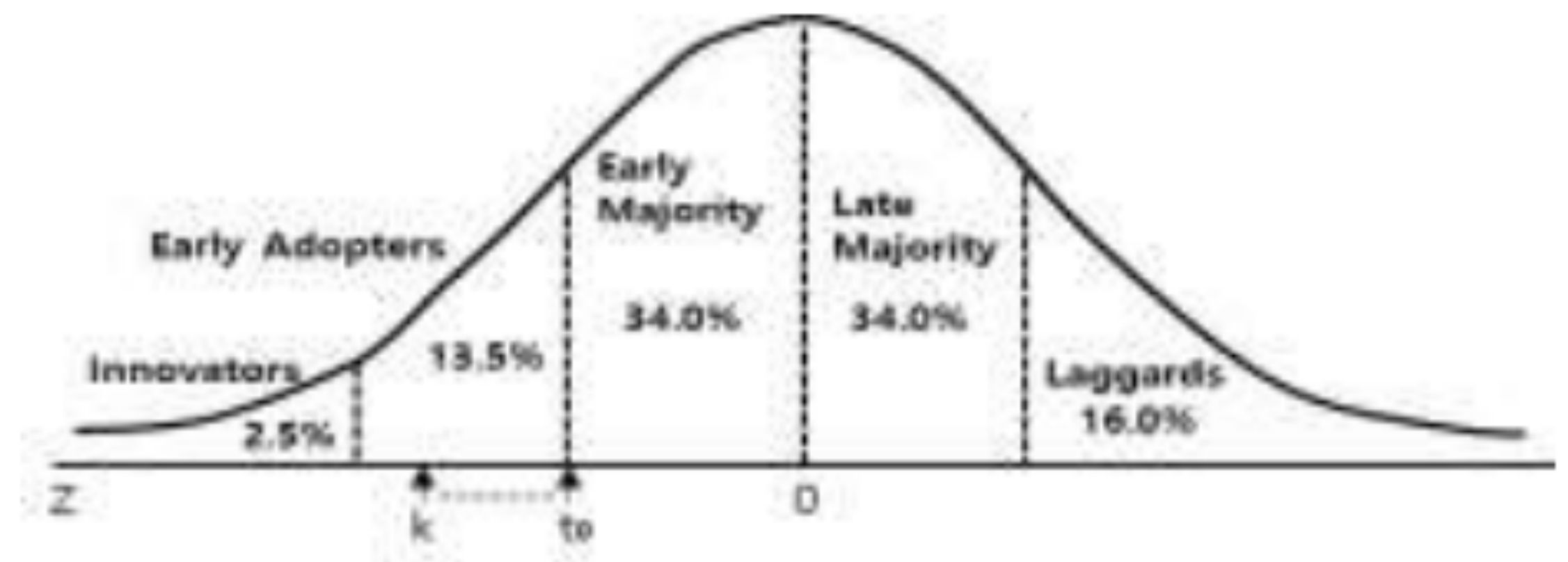


Fig. 2. A conceptual map of factor maturation time in standard normal distribution using Rogers theory

인과형 예측 방법

인과형 예측기법

선형 모형 / ML 모형 계열

- 현업에서 필요한 예측기법은 인과형 예측기법
 - 여러가지 독립변수를 통해 시계열이 설명 될 수 있어야 하고,
 - 현업의 데이터는 보통 정상성을 만족하지 않고
 - 이벤트들이 있기 때문에, Box-Jenkins 모형은 특히 잘 맞지 않음
- 회귀분석, 계량경제 모형, 산업연관분석, 투입산출분석, 수명주기분석, 선도지표법 등
- 다양한 Machine learning 방법들도 사용됨
 - XGBoost
 - Randomforest
 - SVR

인과형 예측기법

Deep Learning 계열

- LSTM, 1D Conv LSTM, CNN LSTM
- ANN, Restricted Boltzman machine
- DeepSD
- Extreme Learning Machine
- deep multi-view spatial-temporal network (DMVST-Net)
- DeepAR
- Gaussian Process Regression
- MCGM(Markov Chain Grey-Markov Model)

Reference

- <https://smalldataguru.com/%EC%88%98%EC%9A%94-%EC%98%88%EC%B8%A1-%EA%B8%B0%EB%B2%95-%EC%8B%9C%EA%B3%84%EC%97%B4-%EB%AA%A8%EB%8D%B8-time-series-model-%EC%86%8C%EA%B0%9C/>
- <https://www.kakaobrain.com/blog/43>
- <https://github.com/microsoft/forecasting>
- <http://contents.kocw.net/KOCW/document/2016/hankyong/kangnamsin1/3.pdf>
- <http://www.koreascience.kr/article/JAKO202022349916778.pdf>
- http://www.sigmapress.co.kr/shop/shop_image/g42415_1405058202.pdf