

제 2 장 예측의 기본 개념

2.1 예측 기법의 종류

1. 예측: 현재와 과거 변수들의 실현된 관측치를 이용하여 그 변수의 미래값을 추정

2. 정성적 방법

- 방법: 전문가의 의견이나 여론조사 과정을 통하여 예측치 추정
- 자료: 과거 자료가 전혀 없거나 미래값 예측에 도움이 되지 않을 경우 사용
- 특징: 통계적 기법을 사용하지 않아 예측의 정확도가 떨어짐

예1) 브레인 스토밍 방법:

- 방법: 리더, 기록자 외에 10명 이내의 참가자(stormer)들이 기존의 관념에 사로잡히지 않고 자유로운 발상으로 아이디어나 의견을 내는 것.
- 6~12명 정도의 사람들이 모여 20분~1시간 가량 문제에 관한 리더의 설명을 듣고, 가능한 많은 대체안을 제시하면 이들은 비판받지 않고 기록→그 후 토의와 분석이 이루어짐.
- 목적: 보다 자유롭고 융통성 있는 사고를 증진하고 구성원들의 창조성을 촉진. 브레인스토밍에서는 어떠한 내용의 발언이라도 그에 대한 비판을 해서 안 되며, 오히려 자유분방하고 엉뚱하기까지 한 의견을 출발점으로 해서 아이디어를 전개시켜 나가도록 함.

예2) Delphi 방법:

- 방법: 한 문제에 대해 여러 전문가들의 독립적인 의견을 우편으로 수집한 다음, 이 의견들을 요약·정리하여 다시 전문가들에게 배부하여 일반적인 합의가 이루어질 때까지 서로의 아이디어에 대해 논평하게 하는 방법
- 장점: 여러 전문가들을 대면회합을 위해 한 장소에 모이게 할 필요 없이 그들의 평가를 이끌어 낼 수 있고, 의사결정과정에서 타인의 영향력을 배제할 수 있음. 의사결정의 범위가 넓거나 장기적인 문제를 해결하는데 유용한 기법
- 단점: 그러나 모든 사람들이 응답을 한 것을 요약·정리하여 다시 우송하는 과정이 합의에 도달하게 될 때까지 계속되므로 소요되는 시간이 길고 응답자에 대한 통제가 힘들. 많은 시간을 요하므로 신속한 의사결정을 필요로 하는 경우에는 사용할 수 없음.

예3) 계층분석 방법(AHP분석):

- 방법: AHP 기법은 Saaty에 의해 개발된 다기준 의사결정모델(multi-criteria decision making model)이며 의사결정 프로세스를 체계적으로 분석하고, 여러 평가항목의 가중치를 쌍대비교(pairwise comparison)에 의하여 단계적으로 도출함으로써 대안들에 대한 합리적 평가를 지원함.

- 장점: AHP 방법론은 복잡한 의사결정 문제를 계층적으로 구조화하여 부분적으로 하나씩 단계적으로 접근하여 최종적으로 종합하는 과정을 거쳐 의사결정을 마무리. 즉, AHP는 계층(hierarchy)의 개념을 통해 의사결정에 필요한 여러 요소들을 계층화시켜 각 요소별, 요소간의 관계를 보다 상세히 논리적으로 보여줌.

3. 정량적 방법

(1) 계량모형 접근법

○ 방법: 경제이론에 기초하여 인과관계를 이용하여 예측치 추정

○ 자료: 인과관계를 추정할 만한 충분한 과거 자료가 있는 경우 사용

○ 특징: 경제 이론뿐만 아니라 통계적 기법이 많이 사용되어 이해가 쉽지 않으나 안정적인 인과관계가 있는 경우 예측의 정확도가 높음

(2) 확정적 시계열 모형(Deterministic Time Series Model)

○ 방법: 시계열 자료의 확률적 변동요인을 제거하고 확정적 추세 요인을 도출하여 예측

○ 자료: 과거자료가 충분하지 않지만 뚜렷한 추세를 나타내는 경우 단기예측 가능

○ 특징: 이해가 쉽고 비용이 적게 드나 확정적 변동요인의 구조변화가 나타나지 않는 한 예측의 정확도는 높음

(3) 확률적 시계열 모형(Stochastic Time Series Model)

○ 방법: 시계열 자료의 과거 패턴을 기초로 ARIMA 종류의 시계열 모형을 추정하여 예측에 사용

○ 자료: 주어진 시계열 자료 자체에 대한 정보는 많은 반면, 다른 변수와의 상관관계가 미약하거나 불분명한 경우 사용

○ 특징: 경제 이론을 도입되지 않는 반면 통계적 기법이 많이 사용되는데 저비용에 전환점 추정 등 변수 움직임을 보는데 유용

2.2 예측 기법의 선정

2.2.1 자료의 성격

- 자료의 특성: 평균 및 분산의 변화, 추세와 계절성, 순환성에 대한 특징 분석
- 정보집합의 규모: 인과관계, 시계열 모형 등 추정에 충분한 정보가 있는 지 고려

2.2.2 소요 비용 vs. 예측의 정확성

- 예측 오차 발생으로 인한 비용 vs. 예측 정확도를 높이는 비용
- 예측 정확도 측정: 예측치가 실제값에 얼마나 가까운지 측정하기 위하여 실제로 실현된 값과 예측치와의 차이를 나타내는 예측 오차 사용

1. 예측 오차에 포함되어 있는 오차

$\varepsilon_t = Y_t - \hat{Y}_t$. Y_t : 실현된 값(실제값), \hat{Y}_t : 예측값, ε_t : 예측 오차

- 편향 오차(bias error): 예측값이 계속적으로 높게 (혹은 낮게) 나타나는 것으로 예측 방법의 구조적 오류가 있을 때 발생
 - 추세나 계절 패턴이 있음에도 이를 무시하고 예측하는 경우에 발생
 - 편향 오차의 존재 유무를 판별하는 방법: 누적평균오차를 이용하여 측정
 - 누적평균오차(CFE; Cumulative Sum of Forecast Error): 각 기간의 실제값과 예측값의 차이를 누적하여 더한 값

예1) 지수평활법:

- 단기예측에 있어서 매우 유용한 기법.
- 가장 최근에 값에 가장 많은 가중치를 두고 자료가 오래될수록 가중치는 지수적으로 감소시키며 예측하는 방법.
- 가중치의 결정을 자동적으로 하는 일종의 가중이동평균법.
- 가중이동평균법은 가중치 결정의 어려움이라는 단점을 가지고 있으므로 이를 해소하기 위해 지수평활법에서는 평활상수(α)를 이용하여 현재에서 과거로 갈수록 더 적은 비중을 주는 방법을 채택. α 의 값이 높을수록 최근의 수요수준에 더 큰 비중이 주어짐.

예) $\hat{Y}_{t+1} = 0.3Y_t + 0.7\hat{Y}_t$ ($\alpha=0.3$)

기간	실제값	예측값	예측오차	CFE
1	10	10(초기예측값 가정)	10-10=0	0
2	22	0.3*10+0.7*10=10	22-10=12	12
3	29	0.3*22+0.7*10=13.2	29-13.2=15.8	27.8
4	39	0.3*29+0.7*13.2=17.9	39-17.9=21.1	48.9

예2) AR(2) 모형 $Y_t = 0.5Y_{t-1} + 0.4Y_{t-2} + \varepsilon_t$, $Y_n = 20, Y_{n-1} = 10$, $\hat{Y}_{n+1} \sim \hat{Y}_{n+4}$ 의 값

$$\hat{Y}_{n+1} = 0.5*20 + 0.4*10 = 14$$

$$\hat{Y}_{n+2} = 0.5*14 + 0.4*20 = 15$$

$$\hat{Y}_{n+3} = 0.5*15 + 0.4*14 = 13.1$$

$$\hat{Y}_{n+4} = 0.5*13.1 + 0.4*15 = 12.55$$

- 확률적 오차(random error): 예측 불가능한 요인으로 인한 오차

2. 예측 오차의 종류

- 평균오차(Mean Error): $ME = \frac{1}{k} \sum_{t=1}^k \varepsilon_t$, k: 예측기간 수

- 실제값과 예측값의 차이의 합을 예측기간 수로 나눔

- 평균절대오차(Mean Absolute Error): $MAE = \frac{1}{k} \sum_{t=1}^k |\varepsilon_t|$

- 실제값과 예측값의 절대값 차이의 합을 예측기간 수로 나눔

- 평균절대값퍼센트오차(Mean Absolute Percent Error): $MAPE = \frac{1}{k} \sum_{t=1}^k \left| \frac{\varepsilon_t}{Y_t} \right|$

- 실제값에 대한 MAE 비율

- 평균자승오차(Mean Squared Error): $MSE = \frac{1}{k} \sum_{t=1}^k \varepsilon_t^2$

- 실제값과 예측값의 차이의 제곱을 합하여 예측기간수로 나눔

- 표준편차 = 평균자승오차의 제곱근

- 평균자승오차제곱근(Root Mean Squared Error): $RMSE = \sqrt{\frac{1}{k} \sum_{t=1}^k \varepsilon_t^2} = \sqrt{MSE}$

예1)

기간	실제값	예측값	오차	평균오차	누적오차	절대오차	평균절대오차	절대값퍼센트오차	자승오차
1	200	225	-25	-25	-25	25	25	25/200=12.5%	625
2	240	220	20	-2.5	-5	20	22.5	20/240=8.3%	400

3	300	285	15	3.33	10	15	20	5%	225
4	270	290	-20	-2.5	-10	20	20	8.7%	400
5	230	250	-20	-6	-30	20	20	8.7%	400
6	260	240	20	-1.66	-10	20	20	7.7%	400
7	210	250	-40	-7.1	-50	40	22.8	19.0%	1600
8	275	240	35	-1.87	-15	35	24.4	12.7%	1225
합						195		81.3%	5275

전체기간 MAE=195/8=24.4

전체기간 MAPE=81.3%/8=10.2%

전체기간 MSE=5275/8=659.4

예2) $Y_t = \varepsilon_t$ 는 평균이 0이므로 예측치는 항상 0

○ Theil의 U 통계치:
$$U = \sqrt{\frac{\sum_{t=1}^k \left(\frac{Y_{t+1} - \hat{Y}_{t+1}}{Y_t} \right)^2}{\sum_{t=1}^k \left(\frac{Y_{t+1} - Y_t}{Y_t} \right)^2}}$$

- 분모의 미래값의 예측값을 현재값을 그대로 사용하는 단순예측치와 비교한 상대적인 예측의 정확도를 나타냄
- U=0→완전예측
- U=1→단순예측치와 같은 수준의 정확도
- U<1→ 단순예측치보다 정확한 예측
- U>1→ 단순예측치보다 부정확한 예측

2.2.3 예측기간(Forecasting Horizon)

- 예측의 목적 및 예측 기간의 길이를 고려하여 적절한 예측 기법 선정
 - 단기 예측: 3개월 미만
 - 중기 예측: 3개월~2년
 - 장기 예측: 2년 이상
- 일반적으로 예측 기간이 길어질수록 예측의 정확도 낮아짐

2.2.4 예측값의 종류

○ 변수 수와 예측 빈도 등을 고려하여 방법 선택

1. 점 예측과 구간 예측

○ 점 예측(Point Forecast): 예측의 목적이 변수의 실제값 추정

○ 구간 예측(Interval Forecast): 대략의 범위를 추정.

예1) 점 예측: 다음 기의 예측치가 1000

예2) 구간 예측: 다음 기의 예측치가 950과 1050 사이에 있을 확률(신뢰도)이 95%

- 예측값 $\hat{Y}_t \pm 2.0 \text{ MAE}$ (or $\pm 1.6 \sqrt{MSE}$) => 확률 89.04

- 예측값 $\hat{Y}_t \pm 2.5 \text{ MAE}$ (or $\pm 2.0 \sqrt{MSE}$) => 확률 95.44

2.3 예측치 추정의 준비 단계 및 평가

2.3.1 예측의 목표 및 대상 설정

○ 예측 목표 설정→대상 변수 설정→예측치 추정의 주기(월별, 분기별) 및 예측 기간의 길이(단기, 중기, 장기) 결정

2.3.2 자료 수집 및 자료의 특징 분석

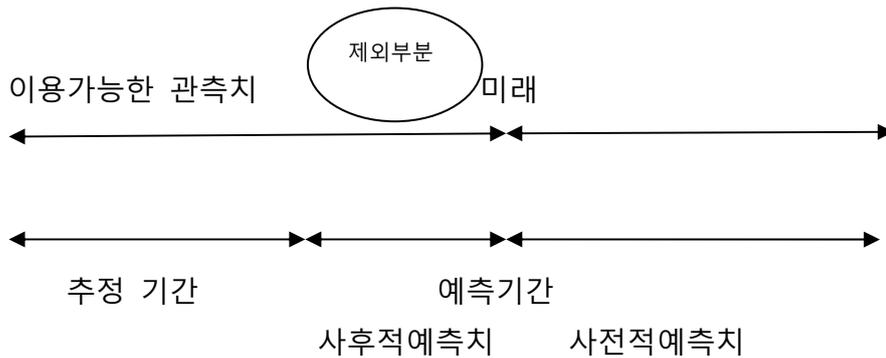
○ 변수에 대한 자료 수집→평균, 분산, 자기상관계수, 다른 변수와의 교차상관계수 등 기술통계량 계산 및 적합한 예측 모형 선정을 위해 전체적인 패턴을 나타내는 그래프 이용

2.3.3 예측 모형의 선정 및 추정

○ 자료의 특성에 맞는 예측 모형 선정→추정→모형의 적합성 검정

○ 미래 예측치의 정확도에 대한 사전 평가 불가능하므로 추정된 예측 모형을 이용해 실험적 예측치 도출

○ 사용할 수 있는 관측치 중 마지막 일부를 제외하고 모형 추정→실제값을 알 수 있는 기간에 대한 사후적 예측치 도출(*Ex post Forecast*)→실제값과 추정치 비교하여 적합성 기준으로 최적 모형 선택



2.3.4 예측치 분석 및 예측 모형 평가

- 최적 모형을 이용하여 사전적 예측치(*Ex ante* Forecast) 도출→이후 시간이 경과한 후의 실제값과 비교→예측 모형에 대한 재평가→필요한 경우 새로운 예측 모형 선택

2.3.5 구조변화에 대한 고려

- 예측치 추정은 기본적으로 과거의 자료가 미래 예측에 유용한 정보를 가지고 있다고 가정
- 가정이 성립하지 않는 경우, 즉 예측치를 추정하고자 하는 변수의 구조변화 (Structural Break)가 나타나서 과거의 움직임과 다른 패턴을 나타낸다면 구조모형을 반영하지 않으면 예측오차가 크게 나타남
- 예) 1997년 아시아 금융위기 당시의 주가 자료. Chow 검정, Quandt-Andrew 검정, Bai-Perron 검정 등으로 확인

EViews 추정 및 예측

- 추정
 - Quick-Estimate equation→GNI C CP
 - t값이나 p-value로 변수로 유의성 확인
 - R²로 모형의 적합도 확인
- 검정
 - 단순회귀에서 상수와 기울기가 모두 0→ 추정 후 View-Coefficient Tests-Wald C(1)=C(2)=0
- 예측
 - equation tool bar에서 Procs/Forecast 를 클릭하면 시계열의 추정치를 구함
 - 예를 들어 위의 quarter.wf1 data를 이용하여 2005년의 국민소득을 예측할

경우 다음과 같은 순서로 하면 됨

- ① workfile window에서 Proc-change workfile range를 선택한 후 1970Q1 2005Q4를 입력
- ② 독립변수(CP)의 값을 먼저 줌(여기서는 1970Q1 2005Q4)
- ③ Quick-Estimate equation을 선택하여 단순회귀모형을 설정하고 추정기간은 1970Q1 2004Q4로 하여 추정.
- ④ Forecast를 선택하고 sample range for forecast에 1970Q1 2005Q4을 입력한다. OK를 클릭하면 국민소득의 예측치는 GNIF라는 계열로 저장
- ⑤ sample을 1970Q1 2005Q4으로 하고 실제치 및 예측치를 비교