성공적 계량분석보고서 작성을 위한 10가지 테크닉

한창호

경제학 박사/콴트글로벌 대표

제4회 콴트글로벌 공개강좌

일시: 2010.09.07~08 16:00~19:00

주제: "애널리스트를 위한 금융데이터분석 워크샵"

주최: 금융공학 온라인교육 전문『**콴트글로벌**』

목차

- 선형회귀분석
  - t-검증
  - F-검증
  - Chow-검증
- 모형 구성 방법
  - 계량모형 구성의 일반적인 절차
  - General to Specific Approach
- 더미변수활용
  - 계절성 모델링1(절편 유)
  - 계절성 모델링2(절편 무)
  - 기울기 더미 변수
- 데이터 변환
  - 로그변환
  - 회귀방정식 의미해석

선형회귀분석

- 선형회귀분석은 모형구성의 출발점
- 비선형모형이 필요한 경우일지라도 데이터에 적절한 변환(로그변환, 차분, 비율)을 가하면 선형모형으로 전환가능

$$Y_t = A X_t^{\beta} e^{u_t}$$

$$\ln Y_t = \ln A + \beta \ln X_t + u_t$$

$$y_t = \alpha + \beta x_t + u_t$$

회귀분석 실행

모형 $y_t = \alpha + \beta x_t + u_t$

회귀직선 $\hat{y}_t = \underset{(SE(\alpha))}{\hat{\alpha}} + \underset{(SE(\beta))}{\hat{\beta}} x_t$

$$\hat{y}_t = \underset{(3.35)}{-59.12} + \underset{(0.0079)}{0.35} x_t$$

잔차 표현 $y_t = \underset{(SE(\alpha))}{\hat{\alpha}} + \underset{(SE(\beta))}{\hat{\beta}} x_t + e_t \qquad e_t = y_t - \hat{y}_t$

t-검증

- t-검증: x_t가 y_t 를 설명할 수 있는가?

$$y_t = \alpha + \beta x_t + u_t$$

- 귀무가설 $H_0 : \beta = 0$
- 검증통계량 $t = \dfrac{\hat{\beta}}{SE(\hat{\beta})}$
- $|t| \geq 2$ 이면 귀무가설 기각 (x_t 가 y_t 설명 가능)

F-검증

- F-검증: 다중가설검증에 사용

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t$$

- 귀무가설 $H_0 : \beta_3 + \beta_4 = 1$

$$H_0 : \beta_2 = 1 \; and \; \beta_3 = -1$$

$$H_0 : \beta_2 = 0 \; and \; \beta_3 = 0 \; and \; \beta_4 = 0$$

F-검증

- 검증통계량 $$F = \frac{RRSS - URSS}{URSS} \times \frac{T-k}{m}$$

$URSS =$ 무제약하 회귀분석에서의 잔차제곱합

$RRSS =$ 제약하 회귀분석에서의 잔차제곱합

$m =$ 제약의 갯수

$T =$ 관측치 갯수

$k =$ 무제약 모형에서의 설명변수 갯수(상수항포함)

- $F \geq F(m, T-k)$ 이면 H_0 기각함

Chow-검증

- 특정 사건을 전후하여 금융시장이나 경제구조 모형에 통계적으로 유의미한 변화가 발생했는지 그 여부에 관심이 있을 경우

- 모형을 구성하고 있는 회귀방정식의 구조변화 여부를 Chow-검증을 이용하여 확인 가능

  - Chow 검증은 F-검증을 응용한 것임

- H_0 : 구조변화가 없다

Chow-검증

- Chow-검증 절차
  - 전체 데이터를 이용하여 회귀분석
  - 데이터를 두 기간으로 나누어 각각 회귀분석
- 검증통계량 $Chow = \dfrac{RSS - (RSS_1 + RSS_2)}{RSS_1 + RSS_2} \times \dfrac{T - 2k}{k}$

  $RSS =$ 전체 데이터를 이용한 회귀분석에서 얻은 잔차제곱합

  $RSS_1 =$ 첫번째 부분 데이터를 이용한 회귀분석에서 얻은 잔차제곱합

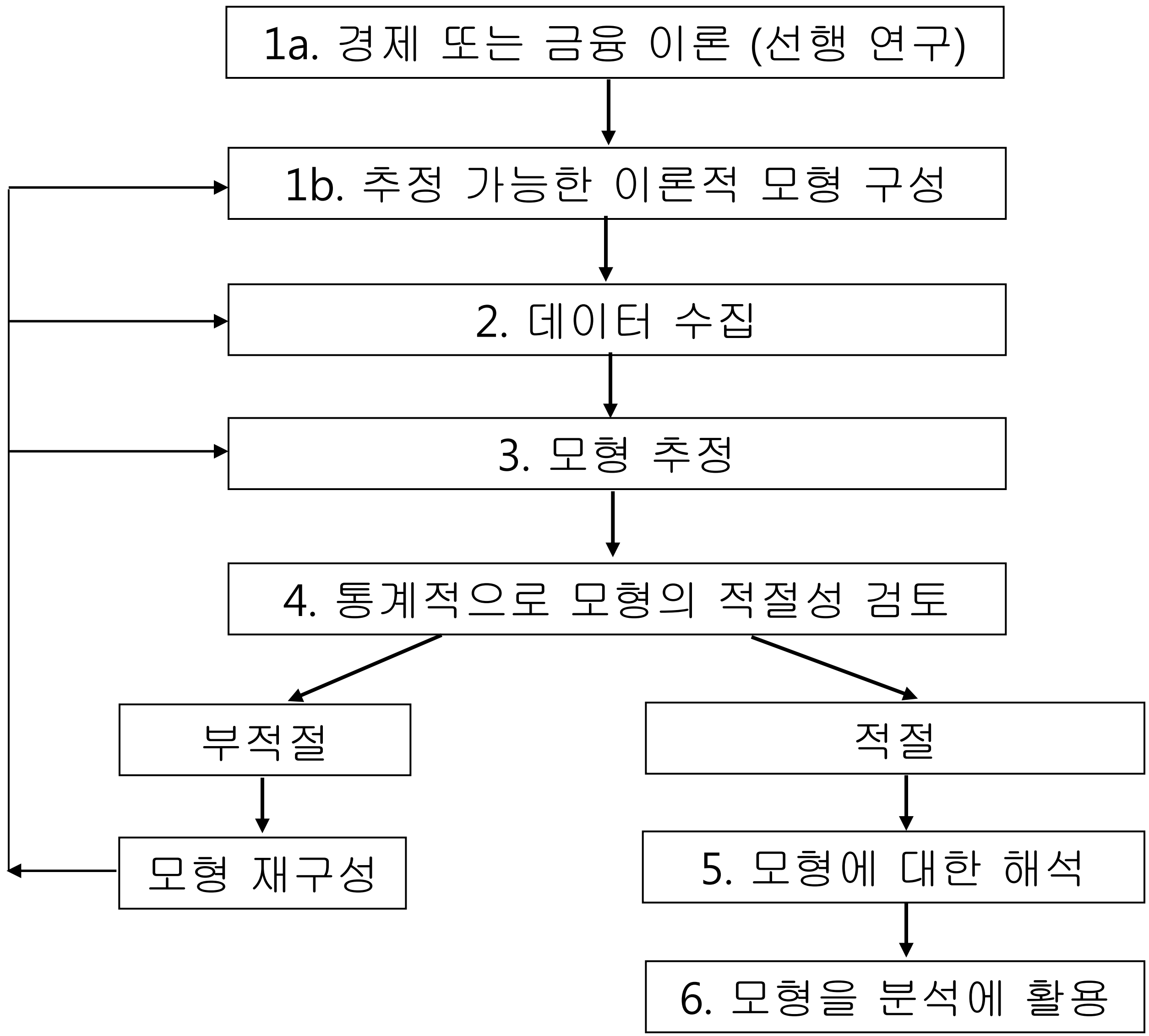
  $RSS_2 =$ 첫번째 부분 데이터를 이용한 회귀분석에서 얻은 잔차제곱합

  $k =$ 설명변수의 갯수(상수항 포함)
- $Chow \geq F(k, T-2k)$ 이면 귀무가설 기각

모형구성방법

- 모형을 구성에는 어떤 정해진 룰은 없음
- 이론에 의해 제시된 모형이 있다면 이를 이용
  - 단, 추정가능한 모형을 구성해야 함
  - 추상적인 개념에 대한 대리변수 선정작업 필요
- 이론적 모형이 없다면 "General to Simple Approach" 사용

계량모형구성 일반적 절차



General to Specific Approach

- 모형구성 방법은 크게 Specific-to-General 과 General-to-Specific approach 로 분류됨
  - Specific-to-General: 단순한 모형에서 출발해서 변수들을 순차적으로 추가
  - General-to-Specific: 규모가 큰 모형에서 출발해서 모형에 제약을 가하고 재배열하여 작은 규모의 최종 모형에 도달

    *Sargan 과 Hendry 가 제안
- 주요 변수를 누락하는 것이 부적절한 변수를 포함하는 것보다 대체적으로 더 심각한 문제

General to Specific Approach

- 모형의 우변에 많은 변수를 지닌 규모가 큰 모형을 구성
  - 이 모형은 금융이론에 의해 도출되어야 함
  - 피설명변수에 영향을 미치는 것으로 생각되는 모든 변수를 포함해야 함
- 회귀분석을 실행, t 값의 절대치가 가장 작은 변수 탈락
- 남은 설명변수를 이용하여 다시 회귀분석 시행
- 모든 변수의 t 값 절대치가 2 이상일 될 때까지 반복 시행

더미변수 활용

- 더미변수(Dummy variables): 특정 조건을 만족할 경우에만 1의 값을 가지고 나머지에서는 모두 0의 값을 가짐

- 異狀置(outlier) 제거에 사용됨
  - 전체 데이터의 흐름과 동떨어진 1회성 또는 극단적인 값을 가지는 데이터 제거

- 계절성 모델링에도 사용됨
  - 주중 요일간 특성, 주말 효과, 1월 효과

계절성 모델링(절편 無)

- 주식수익률이 요일 별로 차이를 보이는가?
- 요일효과모형: $r_t = \gamma_1 D1_1 + \gamma_2 D2_t + \gamma_3 D3_t + \gamma_4 D4_t + \gamma_5 D5_t + u_t$

  r_t : t시점에서의 주식수익률

  $D1_t$: 월요일 더미변수. 월요일이면 1 나머지는 0

  $D2_t$: 화요일 더미변수. 화요일이면 1 나머지는 0

  $D3_t$: 수요일 더미변수. 수요일이면 1 나머지는 0

  $D4_t$: 목요일 더미변수. 목요일이면 1 나머지는 0

  $D5_t$: 금요일 더미변수. 금요일이면 1 나머지는 0

- t-검증을 통해 통계적으로 유의한 $\gamma_i,\ i = 1,2,3,4,5$ 가 있는지 파악

계절성 모델링(절편 有)

- 특정 요일 대비 주식수익률이 차이를 보이는가?
- 월요일 대비 모형: $r_t = \alpha + \gamma_2 D2_t + \gamma_3 D3_t + \gamma_4 D4_t + \gamma_5 D5_t + u_t$
- 회귀분석으로 구한 계수 추정치 $\hat{\gamma}_j, j = 2,3,4,5$ 는 각각 월요일 대비 화, 수, 목, 금요일 주식수익률 차이 의미

  - t-검정을 통해 통계적으로 유의한 차이를 보이는 요일 확인

주의: 상수항 (α) 과 다섯 개 더미 변수를 동시에 모형에 넣으면 안됨

기울기 더미변수

- 절편은 동일한데 분기별로 기울기가 다른가?
- 4분기 대비: $y_t = \alpha + \beta x_t + \gamma_1 D1_t x_t + \gamma_2 D2_t x_t + \gamma_3 D3_t x_t + u_t$

$$1\text{분기 기울기}: \beta + \gamma_1$$
$$2\text{분기 기울기}: \beta + \gamma_2$$
$$3\text{분기 기울기}: \beta + \gamma_3$$
$$4\text{분기 기울기}: \beta$$

$$y_t = \alpha + \gamma_1 D1_t x_t + \gamma_2 D2_t x_t + \gamma_3 D3_t x_t + \gamma_4 D4_t x_t + u_t$$

$\gamma_i, i = 1, 2, 3, 4$: 해당 분기 평균적 기울기

데이터 변환

- 경제, 금융 데이터는 주로 로그변환을 취하여 모델링에 사용하는 경우가 많다.

로그 변환

- 경제, 금융 데이터에 주로 로그변환을 취하는 이유
  - 경제, 금융데이터에 존재하는 비선형적 추세(trend)를 선형적 추세로 전환하므로 모델링이 용이함.

$$Y_t = e^{\delta t}$$

$$\frac{dY_t}{dt} = \delta e^{\delta t} = \delta Y_t$$

$$y_t \equiv \log Y_t = \delta t$$

$$\frac{dy_t}{dt} = \delta$$

로그 변환

- 변수의 변화가 아주 작을 경우, 로그변환을 취한 변수의 1차 차분은 원래 변수의 백분율 변화를 의미함

x 가0 에 가까울 경우 $\ln(1+x) \simeq x$

$$\ln Y_t - \ln Y_{t-1} = \ln \frac{Y_t}{Y_{t-1}}$$

$$= \ln(1 + \frac{Y_t - Y_{t-1}}{Y_{t-1}})$$

$$\cong \frac{Y_t - Y_{t-1}}{Y_{t-1}}$$

선형모형 의미 해석

- Linear $$Y_t = \alpha + \beta X_t + u_t$$

  X_t 가 한 단위증가 -> Y_t 는 β 단위 증가

- Log-linear $$\ln Y_t = \alpha + \beta X_t + u_t$$

  X_t 가 한 단위증가 -> Y_t 는 $100 \times \beta\%$ 증가

- Linear-log $$Y_t = \alpha + \beta \ln X_t + u_t$$

  X_t 가 1% 증가 -> Y_t 는 $0.01 \times \beta$ 단위 증가

- Double log $$\ln Y_t = \alpha + \beta \ln X_t + u_t$$

  X_t 가 1% 증가 -> Y_t 는 $\beta\%$ 증가