Finance Analytics
Chapter5. 회귀모형의 함수 형태

권태연

한국외대 국제금융학과

회귀모형의 함수형태 변환으로 커지는 분산 문제 해결

임금함수 예제: 커지는 분산 문제

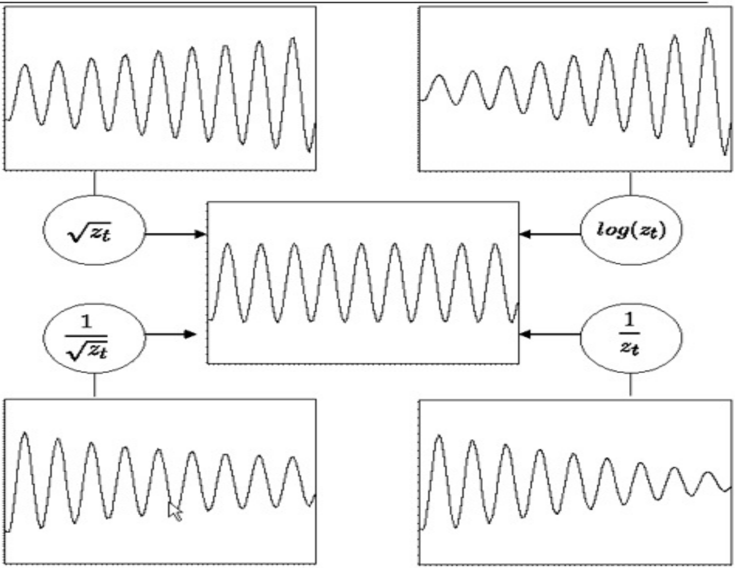
- 임금함수 예제의 잔차플롯에서 확인된 문제는 무엇인가?
  → 등분산가정, $\hat{y}$값이 클수록 분산이 커지는 문제 확인.

  → 해결: 회귀모형의 함수형태 변형, **y**변수 혹은 **x**변수 및 **y** 변수 모두 **log**변환

  → : wage변수를 Log변환 후 log($wage$)와 다른 X변수들간의 선형회귀 모형을 적합하고 문제가 해결되었는제 확인하시오.

  → : wage변수를 Log변환 후 log($wage$)와 다른 X변수들간의 선형회귀 모형을 적합하고 잔차의 정규성 문제를 확인하시오.

적절한 변수 변환을 통한 등분산가정 문제 해결



회귀모형의 함수형태와 해석

회귀모형의 함수 형태

- 1. 로그-로그 모형: x, y변수 모두 로그변환
- 2. 로그-선형 모형: y 변수만 로그변환
- 3. 선형-로그 모형: x 변수만 로그변환

자료의 변환

- 자료 분석적 측면 : 분산 안정화, 등분산 조건 만족하게 하기 위한 변환
- 회귀모형의 해석적 측면 : 탄력성 (elasticity) 혹은 성장 모형등의 의미를 부여할 수 있다.

로그변환 후 회귀계수의 해석 주의사항

✓ 로그변환은 자연로그 변환 $\ln X$을 의미함

✓ $\ln X$ 의 1 증가는 X의 % 증가를 의미함

$\because d\ln X = dX/X$ (몰라도 됨)

$\because e^a \approx (1+a)$ (몰라도 됨)

1. 로그-로그 모형

$$\begin{aligned}\log \hat{y} &= b_1 + b_2 \log x \\ \hat{y} &= exp(b_1)x^{b_2} \\ \hat{y} &= b_1^* x^{b_2}\end{aligned}$$

b_2에 대한 해석:

: $\log x$값이 한 단위 증가하였을 때, $\log y$의 평균 증가분

: x에 대한 y 변수의 탄력성, elasticity으로 해석

: x값이 1% 변화 하였을 때, y값은 b_2 % 변화 하는지를 나타낸다.

: 로그-로그 모형을 상수탄력성 모형이라 하기도 함

2. 로그-선형 모형

$$\log \hat{y} = b_1 + b_2 x$$

b_2에 대한 해석

: x값이 한 단위 증가하였을 때, $\log y$의 평균 증가분

: x에 대한 y 변수의 반-탄력성(semi-elasticity)으로 해석

: x값이 한 단위 증가하였을 때, y값이 $b_2 * 100\%$ 증가

: 로그-선형 모형을 성장 모형으로 부르기도 한다.

3. 선형-로그 모형

$$\hat{y} = b_1 + b_2 \log x$$

b_2에 대한 해석

: $\log x$값이 한 단위 증가하였을 때, y의 평균 증가분

: x값이 100% 증가 하였을 때, y값은 b_2 증가하는지를 나타낸다.

회귀모형의 함수형태와 실습

계량경제 모형 적합 실습

- 앞서 소개된 3가지 계량경제 모형을 각각 자료 "Table2_1", "Table2_5", "Table2_8"을 이용하여 적합 하여 보자.

  1. "Table2_1" : 2005년 미국의 워싱턴 DC를 포함한 50개주, 부가가치 산출물(단위: 천 달러), 노동시간(단위: 천 시간), 제조업 부분의 자본지출액 (단위: 천 달러)
  2. "Table2_5" : 1960년 부터 2007년까지의 미국의 실질 GDP
  3. "Table2_8" : 1995년 미국 867개 가계를 대상으로 조사된 총가계지출 (단위: 달러), 총지출에서 음식료 및 비알코올 음료지출이 차지하는 비율 (단위: %)

- 각 자료 및 변수들의 특징을 파악해 보자.

계량경제 모형 적합 실습

- 자료의 해석적 측면 뿐 아니라 분석적 측면에서도 필요한 변수 변환이었는지 확인하여 보고 그렇지 않다면 다른 형태의 변환을 제시하여 보자.

- 변수 변환 전 모형과 모형을 비교하여 보자.
  a. 등분산 가정 측면에서
  b. 회귀계수의 해석 측면에서
  c. 모형의 적합도 측면에서 : 모형의 적합도를 R-square 혹은 adj-R^2 로 비교하지 못하는 변환은?

자료의 분석적 +해석적 측면을 모두 고려하여 선택된 계량경제 모형

1. 로그-로그 : Cobb-Douglas 생산함수

$$Q_i = B_1 L_i^{B_2} K_i^{B_3}$$

Q: 생산량, L:노동투입량, K:자본

→ 위의 생산함수를 자료를 이용, 선형회귀 모형으로 추정하기 위해서는
위의 함수에 로그를 취하고, 오차항 추가

$$\ln Q_i = B_1 + B_2 \ln L_i + B_3 \ln K_i + u_i$$

- $B_2 + B_3$ 는 규모에 대한 수익 (return to scale)로 해석되며, 투입물의 비율 변화에 따른 산출양의 변화에 대한 정보를 제공
- $B_2 + B_3 = 1$: 규모에 대한 수익불변 (constant return to scale)
- $B_2 + B_3 > 1$: 규모에 대한 수익체증 (constant return to scale)
- $B_2 + B_3 < 1$: 규모에 대한 수익체감 (constant return to scale)

자료의 분석적 +해석적 측면을 모두 고려하여 선택된 계량경제 모형

2. 로그-선형 : 경제 성장 함수, 시간이 흐름에 따라 경제변수가 어떻게 성장하는지에 대한 함수

$$RGDP_t = B_1(1+r)^t$$

$RGDP_t$: (t) 년도의 실질 GDP, r:성장률

$\rightarrow$ 위의 생산함수를 선형회귀모형으로 추정하기 위해서는 위의 함수에 로그를 취하고, 오차항 추가

$$\ln RGDP_t = \ln B_1 + \ln(1+r)t + u$$

$$\ln RGDP_t = B_1 + B_2 t + u$$

자료의 분석적 +해석적 측면을 모두 고려하여 선택된 계량경제 모형

3. 선형-로그 : 엥겔 지출 함수(Engel Expenditure functions):

- 총 지출이 1% 증가할 때, 전체 지출에서 음식 지출이 차지하는 비율이 몇 % (point) 증가하는가?
- 이때, 총 가계지출은 지출액으로 측정되어 있고 전체 지출에서 음식지출이 차지하는 비율은 %로 측정되어 있다.
- 즉 X변수 (총지출)가 1%증가할 때, Y변수(음식지출 비율)가 몇 단위 증가하는 지에 관심

  → 선형-로그 모형

$$SFDHO = B_1 + B_2 \ln(EXPEND) + u$$

역수형 모형

$$Y = B_1 + B_2 \frac{1}{X} + u$$

- 음식지출 비율에 관한 회귀모형을 역수형 모형으로 재 적합 한 후, 앞선 모형과 비교하여 보자.
  a. 등분산 가정 측면에서
  b. 회귀계수의 해석 측면에서
  c. 모형의 적합도 측면에서
    - R^2를 이용할 수 있는가?
    - adj-R^2를 이용할 수 있는가?

      ✓ R^2 혹은 adj-R^2을 이용한 비교는 종속변수가 같을때만 가능함을 주목하자.