제 1 강 계량경제학 기초 Review

Part I. 선형회귀모형 (Linear Regression Model)

I. 계량경제학(Econometrics)이란?

A. 경제적 이론이 설명하는 경제 변수들간의 관계를 경제자료를 바탕으로 통계적으로 추정하고 검정하며 때로는 예측하는 학문

i. 수학적 모형: C=f(Y), f'>0, f''<0 : 경제이론 → 수학적 모형 (예: 엥겔법칙)

ii. 통계모형: 경제자료의 분석을 위해서는 변수들의 관계를 확정적인 관계가 아닌 통계적인 관계로 볼 필요가 있음 :

1. ε 이라는 <u>확률변수</u>가 오차항(error term)으로 포함됨 → C=f(Y)+ ε

iii. 계량경제모형: 자료를 바탕으로 변수들의 관계를 정량적으로 추정하기 위해서는 보다 구체적인 함수형태가 요구되며, 확률변수인 오차항의 분포에 대한 일련의 가정이 필요함

1. 예: $C = \exp(\alpha + \beta Y)$ → $\ln C = \alpha + \beta Y + \varepsilon$,

2. $E(\varepsilon)$=?, $V(\varepsilon)$=?...etc →

3. 종속변수(Dependent Variable): C, 설명변수 or 독립변수: Y, 미지의 모수(unknown parameters) : α, β.

iv. 경제자료 : 변수들에 대한 관측치

1. 통제된 자료(controlled data) vs. 통제되지 않은 자료(uncontrolled data)
   a. 통제된 실험으로부터 산출된 자료 : 주로 자연과학
   b. 통제되지 않는 실험의 결과에서 산출된 자료 또는 관측적 (observational)자료 : 주로 사회과학
2. 관측 형태별: 시계열 자료, 횡단면자료, 패널자료

B. 실증연구의 절차

i. 주장하고 검증 받고자 하는 경제적 이론의 제시 – 변수들간의 수학적 관계로 나타남

ii. 이러한 경제적 이론을 바탕으로 계량경제모형을 제시

iii. 필요한 자료를 습득하고 계량경제모형을 추정, 즉 모수들의 값을 추정

iv. 추정결과로부터 경제적 이론이 제시하는 변수들의 관계에 대한 통계적 검정을 실시 ($\beta \neq 0$, $\beta > 0$, $\beta < 1$),

v. 특히 시계열(패널) 모형의 경우 추정결과로부터 예측 작업을 수행

vi. 결과를 해석하고 평가

II. 단순선형회귀모형 (Simple Linear Regression Model)과 최소제곱추정

A. 단순선형회귀모형의 가정

i. $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$, $t = 1, \ldots, T$

ii. $E(\varepsilon_t) = 0 \iff E(y_t) = \beta_1 + \beta_2 x_t$

iii. $V(\varepsilon_t) = \sigma^2 \iff V(y_t) = \sigma^2$

iv. $Cov(\varepsilon_i, \varepsilon_j) = 0 \iff Cov(y_i, y_j) = 0$, $i \neq j$

v. x 는 <u>확률변수가 아니며</u>, x 는 적어도 두 개의 다른 값을 가져야 한다.

vi. $\varepsilon \sim N(0, \sigma^2)$ (경우에 따라 필요한 가정임)

B. 강 외생성 (Strict Exogeneity)

i. 설명변수 x 가 확률변수가 아니라는 가정

1. 실험실에서의 통제된 자료와 같이 x 를 변화시킴에 있어서, y 에 영향을 미치는 다른 모든 요인들(ε)은 x 의 변화와 무관함을 보증

2. 이 경우 우리는 x 의 변화가 y 에 대해 미치는 영향의 크기를 $\dfrac{\Delta E(y_t \mid x_t)}{\Delta x_t} = \beta_2$ 에서 확인할 수 있음

ii. 설명변수 x 가 확률변수라고 가정해도

1. x 의 변화가 y 의 변화에 영향을 미치는 다른 모든 요인들(ε)과 독립적이라면 동일한 결과를 얻음

2. $E\left(\varepsilon_t \middle| x_1,...,x_T\right)=0$ 를 충족하는 경우 설명변수 x 는 강 외생적(strictly exogenous)라고 함

3. 이는 설명변수 x 가 어떻게 변화하던 무관하게, y 에 영향을 미치는 다른 요인들(ε)이 로 y 에 미치는 평균적 영향은 항상 0 임을 보장함

   a. 따라서 x 의 변화는 ε 의 변화와 독립적임을 의미함

C. 강 외생성 하의 단순선형회귀 모형의 가정 I

i. $y_t=\beta_1+\beta_2 x_t+\varepsilon_t,\ t=1,...,T$

ii. $\left(x_t,y_t\right),\ t=1,...,T$ 는 확률표본(각 쌍은 동일하고 독립적인 분포)

iii. $E\left(\varepsilon_t \middle| x_t\right)=0$ (강 외생성)

iv. $V\left(\varepsilon_t \middle| x_t\right)=\sigma^2$

v. x 는 적어도 두 개의 다른 값을 가져야 한다.

vi. $\varepsilon_t \middle| x_t \sim N\left(0,\sigma^2\right)$

D. 강 외생성 하의 단순선형회귀 모형의 가정 II

i. $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$, $t = 1,...,T$

ii. $E\left(\varepsilon_t \middle| x_1 \ldots x_T\right) = 0$ (강 외생성)

iii. $V\left(\varepsilon_t \middle| x_1 \ldots x_T\right) = \sigma^2$

iv. $Cov\left(\varepsilon_i, \varepsilon_j \mid x_1 \ldots x_T\right) = 0$, $i \neq j$

v. x 는 적어도 두 개의 다른 값을 가져야 한다.

vi. $\varepsilon \sim N\left(0, \sigma^2\right)$

E. 단순회귀모형에서의 모수의 추정

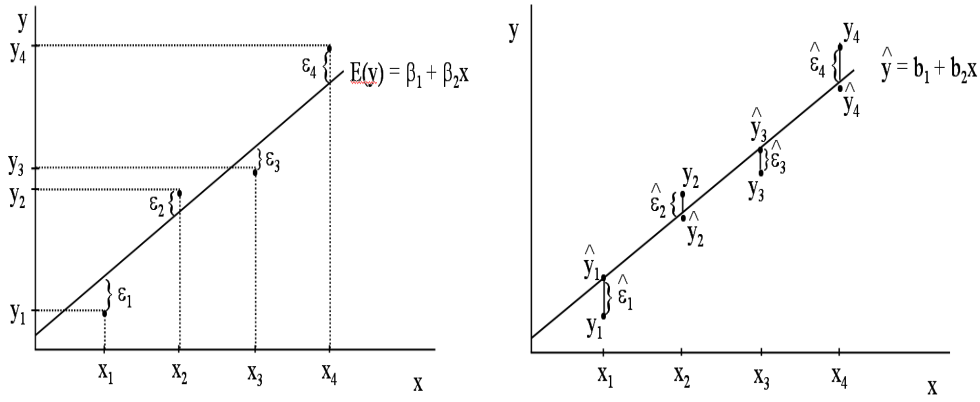
i. 단순회귀모형 $y_t = \beta_1 + \beta_2 x_t + \varepsilon_i$, $t = 1,...,T$ 에서 관측된 x 와 y 의 자료로부터 모수들의 값을 추정

ii. 추정된 값을 b_1, b_2라 하면 $y_t = b_1 + b_2 x_t + \hat{\varepsilon}_t$, $t = 1,...,T$ 로 쓸 수 있음

1. 여기서 $\hat{\varepsilon}_t \equiv y_t - b_1 - b_2 x_t$ 를 잔차(residual)라고 하고, $\hat{y}_t \equiv b_1 + b_2 x_t$ 를 적합값 (fitted values) 또는 예측값(predicted values)이라 함

   a. $\hat{y}_t = b_1 + b_2 x_t$ 의 직선을 적합선(fitted line) 또는 (표본) 회귀선(sample regression line) 이라고 함

<그림> 미지의 실제(true) 회귀선(모집단 회귀선)과 추정된 회귀선



iii. 최소제곱추정 (Least Square Estimation)

1. 잔차의 제곱의 합을 최소로 만들어 주는 모수의 값을 찾는 추정방법

    a. 관측치들의 적합선으로부터의 수직거리의 합을 최소화 시켜 주는 적합선을 찾는 것임

    b. $\min\limits_{b_1, b_2} \sum\limits_t \left(\hat{\varepsilon}_t\right)^2 \equiv \left(y_t - b_1 - b_2 x_t\right)^2 \equiv S\left(b_1, b_2\right)$

        i. $\dfrac{\partial S\left(b_1, b_2\right)}{\partial b_1} = 0 \Rightarrow -2\sum\left(y_t - b_1 - b_2 x_t\right) = 0$

        ii. $\dfrac{\partial S\left(b_1, b_2\right)}{\partial b_2} = 0 \Rightarrow -2\sum x_t\left(y_t - b_1 - b_2 x_t\right) = 0$

c. 정규방정식(normal equations)

i. $\Rightarrow Tb_1 + b_2 \sum x_t = \sum y_t$

ii. $\Rightarrow b_1 \sum x_t + b_2 \sum x_t^2 = \sum x_t y_t$

2. 단순선형회귀모형에 대한 최소제곱추정치(량)

a. $b_2^{LS} = \dfrac{T\sum x_t y_t - \sum x_t \sum y_t}{T\sum x_t^2 - \left(\sum x_t\right)^2} = \dfrac{\sum\left(x_t - \overline{x}\right)\left(y_t - \overline{y}\right)}{\sum\left(x_t - \overline{x}\right)^2}$

b. $b_1^{LS} = \overline{y} - b_2^{LS}\overline{x}$

c. y_t 가 실현된 관측치가 아닌 확률변수일 때, b_1^{LS} , b_2^{LS} 는 추정치가 아닌 추정량이며, 최소제곱(LS) 추정량이라고 함

i. 향후 편의상 LS 추정량을 b_1, b_2로 표기할 것임

III. 최소제곱 추정량의 특성

A. 단순선형회귀모형에 대한 가정 i) – v)가 충족되는 경우 β_1, β_2 에 대한 최소제곱추정량은 다음과 같은 성질을 갖는다.

i. 불편추정량

1. $E\left(b_1\right) = \beta_1$, $E\left(b_2\right) = \beta_2$

ii. 일치추정량

1. $V\left(b_1\right)=\dfrac{\sigma^2\sum x_t^2}{T\sum\left(x_t-\bar{x}\right)^2}\rightarrow 0$, as $T\rightarrow\infty$,

2. $V\left(b_2\right)=\dfrac{\sigma^2}{\sum\left(x_t-\bar{x}\right)^2}\rightarrow 0$ as $T\rightarrow\infty$

iii. 선형추정량

1. $b_1=\sum w_{1t}y_t$, $b_2=\sum w_{2t}y_t$

   a. 즉 추정량이 확률변수의 선형결합으로 표현되는 추정량임

iv. BLUE (Best Linear Unbiased Estimator)

1. 가우스-마코프 정리(Gauss-Markov Theorem): 단순선형회귀모형에 대한 가정 i) –v)하에서 β_1, β_2 에 대한 최소제곱추정량은 모든 선형 불편 추정량 가운데 가장 작은 분산을 갖는다.

   a. → 선형 불편 추정량의 범주내에서 최소제곱 추정량 외의 다른 추정량을 고려할 필요는 없다.

B. 단순선형모형에 대한 가정 i) - vi) 이 충족되는 경우 최소제곱추정량은 다음과 같은 분포를 한다.

i. $b_1^{LS} \sim N\left[\beta_1, \frac{\sigma^2 \sum x_t^2}{T\sum\left(x_t - \bar{x}\right)^2}\right]$, $b_2^{LS} \sim N\left[\beta_2, \frac{\sigma^2}{\sum\left(x_t - \bar{x}\right)^2}\right]$

1. ← 정규분포를 하는 확률변수들의 선형 결합은 정규분포임

   a. cf) 정규분포를 하는 두 확률변수의 공분산이 0 일 때, 두 확률변수는 독립임

ii. vi)을 가정하지 않더라도 충분히 T 가 크면 중심극한정리에 의해 위와 같은 분포로 최소제곱추정량의 분포를 근사할 수 있음

C. 오차항의 분산 σ^2에 대한 추정

i. 잔차를 $\hat{\varepsilon}_t \equiv y_t - b_1 - b_2 x_t$ 오차항에 대한 대용변수로 간주할 수 있으며, 이를 이용하여 오차항의 분산에 대한 추정량을 구축할 수 있음

ii. $\hat{\sigma}^2 \equiv \frac{\sum_1^T \hat{\varepsilon}_t^2}{T-2}$

1. σ^2에 대한 불편 추정량

2. $T-2$로 나누어주는 것은 잔차의 자유도로 나누어 주는 것이며 2 는 추정하고자 하는 모수의 숫자와 같다

iii. 이를 이용해서 최소제곱추정량의 분산에 대한 추정량 역시 도출할 수 있음

1. $\widehat{V(b_1)} = \dfrac{\hat{\sigma}^2 \sum x_t^2}{T\sum\left(x_t - \bar{x}\right)^2}$, $\widehat{V(b_2)} = \dfrac{\hat{\sigma}^2}{\sum\left(x_t - \bar{x}\right)^2}$

a. 여기에 제곱근을 씌운 표준편차에 대한 추정량을 특별히 표준오차(standard error)라 함

IV. 단순회귀모형에서의 추론

A. 구간추정량 (신뢰구간)의 도출

i. 오차항의 분산 σ^2 이 알려진 경우, 가정 vi)하에서

1. $Z = \dfrac{b_2 - \beta_2}{SD\left(b_2\right)} \sim N\left(0,1\right)$ → 이를 이용하여, (1-α) x 100% 구간추정량(신뢰구간)은 다음과 같이 도출됨

a. $P\left(-z_{\frac{\alpha}{2}} < \dfrac{b_2 - \beta_2}{SD\left(b_2\right)} < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$ ⇒

$P\left(b_2 - z_{\frac{\alpha}{2}} \cdot SD\left(b_2\right) < \beta_2 < b_2 + z_{\frac{\alpha}{2}} \cdot SD\left(b_2\right)\right) = 1 - \alpha$ ⇒

$\left(b_2 - z_{\frac{\alpha}{2}} \cdot SD\left(b_2\right), b_2 + z_{\frac{\alpha}{2}} \cdot SD\left(b_2\right)\right)$

ii. 실제로는 σ^2를 알 수 없으므로, 그 불편 추정량을 이용함

1. $t = \frac{b_2 - \beta_2}{SE(b_2)} \sim t_{T-2}$, 단 $SE(b_2) = \sqrt{\frac{\hat{\sigma}^2}{\sum (x_t - \bar{x})^2}}$

a. $V \equiv \frac{(T-2)\hat{\sigma}^2}{\sigma^2} = \frac{\sum_1^T \hat{\varepsilon}_t^2}{\sigma^2} \sim \chi_{T-2}^2$

b. b_1, b_2와 $\hat{\sigma}^2$는 서로 확률적으로 독립

c. $Z \Big/ \sqrt{V \big/ (T-2)} = \frac{b_2 - \beta_2}{SD(b_2)} \Big/ \frac{\hat{\sigma}}{\sigma} = \frac{b_2 - \beta_2}{SE(b_2)} \sim t_{T-2}$

2. $(1-\alpha) \times 100\%$ 구간추정량은 다음과 같이 도출됨

a. $P\left(-t_{\frac{\alpha}{2},T-2} < \frac{b_2 - \beta_2}{SE(b_2)} < t_{\frac{\alpha}{2},T-2}\right) = 0.95 \quad \Rightarrow$

$P\left(b_2 - t_{\frac{\alpha}{2},T-2} SE(b_2) < \beta_2 < b_2 + t_{\frac{\alpha}{2},T-2} SE(b_2)\right) = 1 - \alpha \quad \Rightarrow$

$\left(b_2 - t_{\frac{\alpha}{2},T-2} \cdot SE(b_2), b_2 + t_{\frac{\alpha}{2},T-2} \cdot SE(b_2)\right)$

B. 가설검정

i. 양측검정

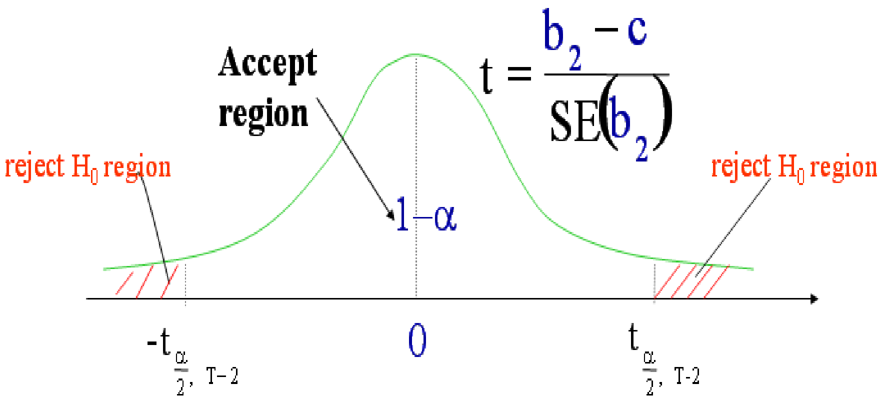
1. 가설: $H_0 : \beta_2 = c$, $H_1 : \beta_2 \neq c$

2. 검정통계량: $t = \dfrac{b_2 - c}{SE(b_2)} \sim t_{T-2}$

3. 유의 수준이 α로 주어졌을 때, 기각역은 $\left[-\infty, -t_{\frac{\alpha}{2},T-2}\right]$, $\left[t_{\frac{\alpha}{2},T-2}, \infty\right]$

즉 $t \geq t_{\frac{\alpha}{2},T-2}$ 이거나 $t \leq -t_{\frac{\alpha}{2},T-2}$ 일 경우, 귀무가설을 기각함

<그림> 양측검정



ii. 단측검정

1. 가설: $H_0 : \beta_2 = c$, $H_1 : \beta_2 > c$ 또는 가설: $H_0 : \beta_2 \le c$, $H_1 : \beta_2 > c$

2. 검정통계량 : $t = \dfrac{b_2 - c}{SE(b_2)} \sim t_{T-2}$

3. 유의 수준이 α로 주어졌을 때, 기각역은, $\left[t_{\alpha,T-2}, \infty\right]$ 즉 $t \ge t_{\alpha,T-2}$ 일 경우, 귀무가설을 기각함

4. 가설: $H_0 : \beta_2 = c$, $H_1 : \beta_2 < c$ 또는 가설: $H_0 : \beta_2 = c$, $H_1 : \beta_2 < c$ 일 경우, 기각역은 $\left[-\infty, -t_{\alpha,T-2}\right]$ 이며, 즉 $t \le -t_{\alpha,T-2}$ 일 때 기각함

iii. P 값 (p-value)

1. p 값은 주어진 검정통계량의 값에 대해 귀무가설을 기각하기 위한 최소크기의 유의 수준임
   a. 양측검정의 경우 : $p - value = 2 \times P\left(t \ge \left|t^*\right|\right)$, 단 t^* 는 검정통계량의 표본값(검정통계치)
   b. 우측단측검정의 경우 : $p - value = P\left(t \ge t^*\right)$
   c. 좌측단측검정의 경우 : $p - value = P\left(t \le t^*\right)$

iv. 유의성 검정(test of significance)

1. 설명변수과 종속변수간에 통계적으로 유의한 관계가 있는가? 에 대한 검정

2. $H_0 : \beta_2 = 0$, $H_1 : \beta_2 \neq 0$

3. $t = \dfrac{b_2}{SE(b_2)}$: 컴퓨터에서 자동적으로 보고함

C. 최소제곱예측

i. 추정결과를 예측에 활용할 수 있음

ii. x_0 이 주어졌을 때, $y_0 = \beta_1 + \beta_2 x_0 + \varepsilon_0$ 에 대한 최소제곱예측(값)은 $\hat{y}_0 = b_1 + b_2 x_0$ 로 주어짐

iii. $f \equiv \hat{y}_0 - y_0$ 를 예측오차라 하며 $V(f) = \sigma^2 \left[1 + \dfrac{1}{T} + \dfrac{(x_0 - \bar{x})^2}{\sum (x_t - \bar{x})^2} \right]$, $E(f) = 0$, 임

1. y_0 에 대한 95% 예측구간은 충분히 큰 관측치나 정규분포의 가정하에서, $\left(\hat{y}_0 - t_{\frac{\alpha}{2}, T-2} \cdot SE(f), \hat{y}_0 + t_{\frac{\alpha}{2}, T-2} \cdot SE(f) \right)$로 주어짐

V. 단순선형회귀 모형의 기타 이슈들

A. 결정계수 (coefficient of determination)

i. 종속변수 y_t 의 변동성을 설명되는 부분과 설명되지 않는 부분으로 분할

1. 종속변수 y_t 의 총변동성 : $\sum_{t=1}^{T}\left(y_t-\overline{y}\right)^2$: Total Sum of Squares (SST or TSS)

2. 설명되는 변동성 : $\sum_{t=1}^{T}\left(\hat{y}_t-\overline{y}\right)^2$: Regression Sum of Squares(SSR or RSS)

3. 설명되지 않는 변동성: $\sum_{t=1}^{T}\left(y_t-\hat{y}_t\right)^2=\sum_{t=1}^{T}\hat{\varepsilon}_t^2$:Error Sum of Squares(SSE or ESS)

ii. SST = SSR + SSE $\Rightarrow R^2 \equiv \dfrac{SSR}{SST}=1-\dfrac{SSE}{SST}$: 결정계수

1. 모형 전체의 설명력을 나타내는 지표

2. $0 \le R^2 \le 1$

3. 단순회귀모형에서의 결정계수는 설명변수와 종속변수간 또는 종속변수와 적합값 간의 표본상관계수의 제곱)과 같음 → 적합도의 척도 (goodness of fitness)

B. 단위의 변경과 추정결과

1. 변수들의 측정단위를 변경하는 것은 결정계수나 t 검정통계량의 값에 영향을 주지 않는다.
2. x 의 단위변경은 x 의 계수에 대한 추정량의 크기에 영향을 줌
   a. $y_t = \beta_1 + (c\beta_2)\left(x_t / c \right) + \varepsilon_t$
3. y 의 단위변경은 모든 계수의 추정량의 크기에 영향을 주며, 오차항 분산의 추정량의 크기에도 영향을 줌
   a. $y_t / c = \beta_1 / c + \beta_2 / c \; x_t + \varepsilon_t / c$

C. 단순선형회귀모형의 여러 가지 형태

i. log-log 모형 : $\ln y_t = \beta_1 + \beta_2 \ln x_t + \varepsilon_t$

   1. β_2 : y 의 x 에 대한 탄력성

ii. 역수 모형 : $y_t = \beta_1 + \beta_2 \dfrac{1}{x_t} + \varepsilon_t$

iii. log-선형, 선형-log: $\ln y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$, $y_t = \beta_1 + \beta_2 \ln x_t + \varepsilon_t$

VI. 다중회귀모형

A. 다중회귀모형에 대한 가정

i. $y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} + \varepsilon_t$, $t = 1,...,T$, K: 추정 모수의 수 (경제적 설명변수의 수 K-1), cf. $x_{t1} \equiv 1$

1. 개별 모수들의 의미: 예컨대, $\beta_2 = \dfrac{\partial E(y_t)}{\partial x_{t2}}$: 다른 설명변수들이 일정할 때, x_{t2} 의 한 단위 변화에 대한 종속변수의 평균값의 변화 $\Rightarrow$ 즉 다른 변수들의 영향력이 통제(control)된 상황에서 첫 번째 설명변수의 종속변수에 대한 영향력을 나타내는 값임

ii. $E(\varepsilon_t) = 0 \;\Leftrightarrow\; E(y_t) = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_K x_{tK}$

iii. $V(\varepsilon_t) = \sigma^2 \;\Leftrightarrow\; V(y_t) = \sigma^2$,

iv. $Cov(\varepsilon_i, \varepsilon_j) = 0 \Leftrightarrow Cov(y_i, y_j) = 0$, $i \neq j$

v. x_{tk} 는 확률변수가 아니며, 다른 설명변수(들)의 정확한 선형함수가 아님 (즉 완전한 공선성(colinearity)가 존재하지 않음)

1. 만일 x_{t2} 와 x_{t3} 간에 정확한 선형관계, 즉 $x_{t2} = a + bx_{t3}$ 가

성립한다면 이들 설명변수의 종속변수에 대한 개별적인 영향력을 추정하는 것은 불가능함 (예: 고정투입생산함수)

vi. $\varepsilon \sim N\left(0, \sigma^2\right)$ (경우에 따라 필요한 가정임)

B. 다중회귀모형(설명변수가 두 개인 경우)에 대한 최소제곱추정

i. $\min_{b_1, b_2, b_3} \sum_t \left(\hat{\varepsilon}_t\right)^2 \equiv \left(y_t - b_1 - b_2 x_{t2} - b_3 x_{t3}\right)^2 \equiv S\left(b_1, b_2, b_3\right)$

ii. 최소제곱추정량의 성질

1. BLUE (Gauss-Markov Theorem)

a. 불편추정량, 선형추정량, (일치추정량)

2. 기타 성질

a. 최소제곱추정에 의한 회귀직선은 $\left(\overline{y}, \overline{x}_2, \overline{x}_3\right)$을 통과함

b. $\sum \hat{\varepsilon}_t = 0$, $\sum \hat{\varepsilon}_t x_{tk} = 0$, $\sum \hat{\varepsilon}_t \hat{y}_t = 0$

iii. 최소제곱추정량의 분산

1. $\operatorname{var}(b_2) = \dfrac{\sigma^2}{\sum \left(x_{t2} - \overline{x}_2\right)^2 (1 - r_{23}^2)}$, $\operatorname{var}(b_3) = \dfrac{\sigma^2}{\sum \left(x_{t3} - \overline{x}_3\right)^2 (1 - r_{23}^2)}$,

단 $r_{23} = \dfrac{\sum \left(x_{t2} - \overline{x}_2\right)\left(x_{t3} - \overline{x}_3\right)}{\sqrt{\sum \left(x_{t2} - \overline{x}_2\right)^2 \sum \left(x_{t3} - \overline{x}_3\right)^2}}$: x_{t2}와 x_{t3}간의 표본상관계수

2. 최소제곱추정량의 분산은 다음과 같을 때 작아짐

a. 오차항의 분산이 작을수록

b. 표본의 크기가 클수록

c. 해당 설명변수의 값이 퍼져 있을수록

d. 설명변수간의 상관관계가 0 에 가까울수록

C. 다중회귀모형(설명변수가 두 개인 경우)에 대한 구간추정 및 가설검정

i. 가정 vi) $y_t \sim N\left[(\beta_1 + \beta_2 x_{t2} + \ldots + \beta_K x_{tK}), \sigma^2\right] \Leftrightarrow e_t \sim N(0, \sigma^2) \Rightarrow$

1. $z = \dfrac{b_k - \beta_k}{\sqrt{\mathrm{var}\left(b_k\right)}} \sim N\left(0, 1\right), \qquad \text{for } k = 1, 2, \ldots, K$

✓ 오차항의 분산을 모를 경우 이를 이용할 수 없음

ii. 오차항의 분산에 대한 불편 추정량

1. $\hat{\sigma}^2 = \dfrac{\Sigma \hat{\varepsilon}_t^2}{T - K}$: 이를 이용하여 최소제곱추정량 b_k 의 표준오차 $\sqrt{\widehat{Var(b_k)}}$ 을 구함

2. $t = \dfrac{b_k - \beta_k}{\sqrt{\widehat{\mathrm{var}}\left(b_k\right)}} \sim t_{(T-K)}$: 이를 이용하여, 개별 모수에 대한 구간추정과 가설검정(양측가설, 단측가설)을 수행함

iii. 이러한 t 검정은 개별 모수 뿐 아니라 개별 모수들의 선형결합에 대한 가설검정에도 이용될 수 있음

1. $H_0 : \sum c_i \beta_i = c_0$, $H_0 : 3\beta_2 - 7\beta_3 = 21$, etc.

   a. 검정통계량: $t = \dfrac{\sum c_i b_i - c_0}{SE\left(\sum c_i b_i\right)} \sim t_{(T-K)}$

iv. F 검정: 모수에 대한 복합가설 검정

1. $H_0 : \beta_2 = \beta_3 = 0$, $H_0 : \sum c_{1i}\beta_i = c_1, \sum c_{2i}\beta_i = c_2, \sum c_{3i}\beta_i = c_3$, etc

2. F-검정의 원리

   a. F-검정은 귀무가설에 의해 부과되는 모수들에 대한 제약하에서의 회귀모형으로부터의 잔차의 제곱의 합과 아무런 제약이 없을 때의 잔차의 제곱의 합을 비교하는 것에 기반함

   b. 귀무가설이 옳다면 두 회귀모형으로부터의 잔차의 차이는 없어야 할 것이며, 두 잔차의 차이가 클수록 귀무가설이 옳지 않은 것으로 판단할 수 있음

3. F-검정통계량의 구축

a. 귀무가설이 참이라는 가정하의 (제약을 부과한) 모형에서의 잔차의 제곱의 합을 SSE_R (Restricted sum of squared errors)

b. 아무런 제약이 부과되지 않은 원래의 모형에서의 잔차의 제곱의 합을 SSE_U (Unrestricted sum of squared errors)

✓ $SSE_R - SSE_U \geq 0$ (Why?)

c. J 를 귀무가설에서의 가설의 수, 즙 부과되는 제약의 수라고 하면

i. 귀무가설하에서 $V_1 = \dfrac{SSE_R - SSE_U}{\sigma^2} \sim \chi^2_{(J)}$

ii. $V_2 = \dfrac{SSE_U}{\sigma^2} \sim \chi^2_{(T-K)}$

iii. V_1 와 V_2 는 확률적으로 독립임

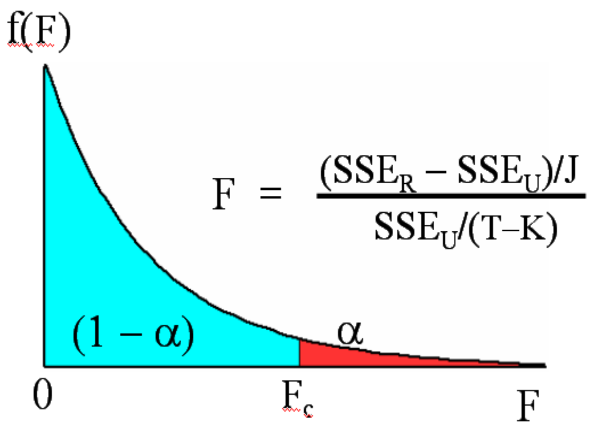
d. 따라서 귀무가설하에서

$$\Rightarrow F = \frac{V_1/J}{V_2/(T-K)} = \frac{(SSE_R - SSE_U)/J}{SSE_U/(T-K)} \sim F_{J,T-K}$$

4. F-검정

a. 대립가설하에서 이렇게 계산되는 F 값은 커질 것임 (F 검정은 항상 기각역이 우측꼬리에 놓이는 검정임)

<그림> F 검정



b. 귀무가설의 가설이 하나일 경우 (즉 제약이 하나일 경우) t 검정과 F 검정의 결과는 같음

✓ $t^2 = F$

c. 모형의 유의성 검정

i. $H_0 : \beta_2 = 0, \beta_3 = 0, \ldots, \beta_K = 0$
H_1 : *at least one* of the β_k is nonzero

ii. $$F = \frac{(SSE_R - SSE_U)/(K-1)}{SSE_U/(T-K)} \sim F_{J,T-K}$$

5. χ^2 검정

a. $V_1 = \frac{SSE_R - SSE_U}{\sigma^2} \sim \chi^2_{(J)}$, under the null →

$\rightarrow \frac{SSE_R - SSE_U}{\hat{\sigma}^2} \simeq \chi^2_{(J)}$, under the null

b. 소표본 분포를 이용한 F 검정이 바람직하나, 정규분포의 가정을 버린다면 둘 다 근사적 분포를 이용.

i. 근사적 분포의 경우에도 F 검정의 소표본 성질이 나으나, 관측치가 커짐에 따라 F 검정은 χ^2 검정으로 수렴함을 보일 수 있음

D. 제한최소제곱추정 (Restricted LS)

i. 예) 어떤 생산과정이 CRS 인 Cobb-Douglas 형태라고 알려져 있음

1. $\log y_t = \beta_1 + \beta_2 \log x_{t2} + \beta_3 \log x_{t3} + \varepsilon_t$ 에서 $\beta_2 + \beta_3 = 1$ 이라는 비표본정보를 이용하여 추정

a. $\log\left(y_t / x_{t3} \right) = \beta_1 + \beta_2 \left(x_{t2} / x_{t3} \right) + \varepsilon_t \;\Rightarrow\; y_t^* = \beta_1 + \beta_2 x_{t2}^* + \varepsilon_t$

ii. 제한최소제곱추정량의 성질

1. 부과한 제약이 옳지 않다면 추정량은 bias 됨 (cost)

2. 분산은 원래 모형에 대한 추정량의 분산에 비해 작아짐 (benefit)

E. 모형의 설정 (Model Specification)

i. 어떠한 모형을 선택할 것인가의 문제

1. 모형에 어떠한 설명변수들을 포함 시킬 것인가?

a. 인과관계 분석: 설명변수의 변화가 종속변수의 평균(체계)적 변화에 미치는 영향의 크기를 알고자 함 (ex 정책 효과 분석)

i. 설명변수에서의 한 단위 변화가 다른 요인들이 일정할 때 종속변수의 평균에 미치는 영향을 끄집어 내어야 함 (외생성이 중요)

b. 예측 모형: 주어진 정보를 토대로 종속변수 값을 예측하고자 함

i. 이 경우 종속변수와 높은 상관이 있는 변수들의 선택이 중요 (외생성 여부보다는 예측력을 높여주는 것이 중요)

2. 포함되는 설명변수들과 종속변수는 어떠한 함수형태로 그 관계를 표현할 것인가?

3. 모형에 대한 적절한 통계적 가정들은?

a. 다중회귀모형에 국한해서 본다면, 1-5의 가정이 적절한가?

ii. 변수의 누락

1. True model: $W_t = \beta_1 + \beta_2 E_t + \beta_3 M_t + \varepsilon_t$ (임금, 경험, 동기부여)

2. Estimate $W_t = \beta_1 + \beta_2 E_t + \varepsilon_t$

   a. 적절한 변수의 누락은 모형에 대한 잘못된 제약부과와 같음

      i. $\beta_3 = 0$ 이라는 제약이 옳지 않음에도 불구하고 이를 부과함

      ii. 추정량의 편향 및 분산의 축소를 낳게 됨

iii. 관련 없는 변수의 포함

1. True model: $W_t = \beta_1 + \beta_2 E_t + \beta_3 M_t + \varepsilon_t$

2. Estimate $W_t = \beta_1 + \beta_2 E_t + \beta_3 M_t + \beta_4 C_t + \varepsilon_t$ (자녀의 수)

   a. 사실 $\beta_4 = 0$

      i. 관련 없는 변수의 포함은 추정량의 불편성에 영향을 주지는 않음

      ii. 원래 모형에 대한 추정량에 비해 분산의 크기가 커짐

iv. 모형 설정 검정

1. Jacque-Bera (JB) Normality test (쟈크베라의 정규성 검정)

   a. 단순선형모형의 가정 1-6 의 적절성을 검정하는 한 방법

   b. 귀무가설과 대립가설 : H_0: X ~ N(μ_X, σ_X^2), H_1: H_0 is not true

      i. 정규분포임을 보이는 목적으로 사용하기에는 부적절한 가설설정임

   c. 검정통계량

      i. $$JB \equiv \frac{T}{6}\left(\hat{s}^2 + \frac{\left(\hat{k}-3\right)^2}{4}\right) \sim \chi^2\left(2\right) : \text{under } H_0$$

      ii. Skewness(비대칭도) – 어떤 확률변수의 분포가 평균을 중심으로 얼마나 비대칭인가를 나타내는 척도 (정규분포의 Skewness = 0) : s ≡ $E[(X-\mu_X)^3]/\sigma_X^3$,

      iii. Kurtosis(첨예도) - 어떤 확률변수의 분포가 얼마나 뾰쪽한가를 나타내는 척도 (정규분포의 Kurtosis=3) : k ≡ $E[(X-\mu_X)^4] / \sigma_X^4$,

d. 기각역

i. JB< $\chi^2_{\alpha,2}$ 이면 유의 수준 α에서 H_0 를 기각할 수 없으며, X 의 분포가 정규분포라고 볼 수 있음 (엄밀하게는 X 의 분포가 정규분포가 아님을 입증하지 못한 것임)

2. RESET(Regression Equation Specification Error Test) 검정

a. 모형 설정에 있어서 변수의 누락과 잘못된 함수형태를 탐지하기 위해 고안됨

b. $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + e_t$ → $\hat{y}_t = b_1 + b_2 x_{t2} + b_3 x_{t3}$

c. 보조적(auxiliary) 회귀

i. $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \gamma_1 \hat{y}_t^2 + e_t$ (1)

ii. $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \gamma_1 \hat{y}_t^2 + \gamma_2 \hat{y}_t^3 + e_t$ (2)

1. 포함되는 적합값의 차수는 더 높아질 수도 있음

d. 가설

i. (1)의 경우: $H_0 : \gamma_1 = 0$, $H_1 : \gamma_1 \neq 0$

ii. (2)의 경우: $H_0 : \gamma_1 = \gamma_2 = 0$ $H_1 : o.w.$

iii. 귀무가설의 기각은 원래 모형이 적절하지 못하며 개선의 여지가 있음을 의미함.

iv. 반면에 귀무가설을 기각하지 못하면 이는 이 검정을 통해 모형설정의 오류를 탐지할 수 없었음을 의미함

F. 공선성(Colinearity)의 문제

i. 설명변수 혹은 독립변수들 상호간에 상관관계가 존재할 수 있음

1. 특히 통제되지 않는 자료의 경우 이러한 상관관계를 통제할 수 없으며, 각 변수의 영향력을 분리해내는데 문제가 발생할 수 있음

ii. 공선성의 효과

1. 완전한 공선성의 경우 최소제곱추정 결과를 얻을 수 없음
2. 높은 공선성은 표준오차를 크게 하고 따라서 신뢰구간을 넓힘
3. 이는 높은 결정계수 값과 유의성 있는 F 값에도 불구하고 개별 t 값은 유의성이 없는 것으로 나타내게 함
4. 몇 몇 관측치 또는 유의성 없는 변수의 삭제 또는 포함에 결과가 매우 민감하게 변함

iii. 공선성의 판단

1. 두 설명변수간의 높은 상관관계 (0.8 또는 0.9 이상?)
2. 한 설명변수를 나머지 설명변수들에 대해 회귀분석시 높은 결정계수의 값 (0.8 이상?)
    a. VIF (variance inflation factor) : 예컨데, x_{tk} 를 나머지 설명변수에 대해 회귀하는 보조적 회귀로부터의 결정계수를 R_k^2 라고 할 때, 해당 변수의 VIF 는 $\frac{1}{1-R_k^2}$ 로 정의됨.
3. 개별 t 값들의 통계적 유의성이 없음에도 유의성 있는 F 값

iv. 공선성의 완화

1. 자료의 추가적 확보
2. 적절한 경제적 또는 통계적 제약의 부과