제 1 강 계량경제학 Review

Part I. 단순회귀모형

I. 계량경제학

A. 계량경제학(Econometrics)이란?

i. 경제적 이론이 설명하는 경제 변수들간의 관계를 경제자료를 바탕으로 통계적으로 추정(estimation)고 검정(test)하는 학문

ii. 거시 소비 함수(Keynse)

1. $C=f(Y), 0<f'<1$: 경제이론
2. $C = \alpha + \beta Y + \varepsilon$: 계량경제모형
   a. 자료를 바탕으로 변수들의 관계를 정량적으로 추정하기 위해서는 보다 구체적인 함수형태가 요구됨
   b. 변수들의 관계를 확정적으로 성립하는 관계가 아닌 통계적으로 성립하는 관계로 봄 : ε 이라는 확률변수가 오차항(error term)으로 포함됨
   c. C: 종속변수(Dependent Variable), Y: 설명변수, 독립변수, α, β, : (알려지지 않음) 모수(unknown parameters)
3. 경제자료 : 변수들에 대한 관측치
   a. 통제된 자료: 통제된 실험으로부터 산출된 자료
   b. 통제되지 않은 자료: 통제되지 않는 실험의 결과에서 산출된 자료
   c. 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

v. 결과를 해석하고 평가

II. 단순선형회귀모형

A. 단순선형회귀모형의 가정

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

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

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 는 확률변수가 아니며, x 는 적어도 두 개의 다른 값을 가져야 한다.

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

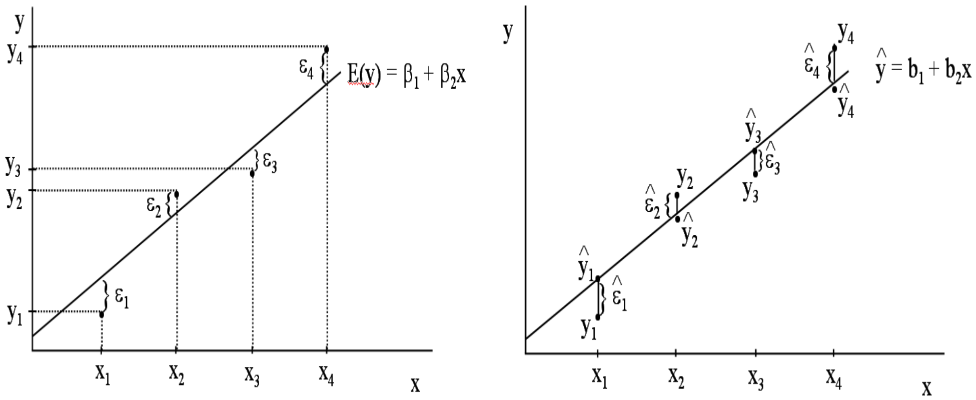
B. 단순회귀모형에서의 모수의 추정

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

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

a. 여기서 $\hat{\varepsilon}_t \equiv y_t - b_1 - b_2 x_t$ 로 정의되면 잔차(residual)라고 하고, $\hat{y}_t \equiv b_1 + b_2 x_t$ 를 적합선(fitted line) 또는 회귀선(regression line) 이라고 함

<그림> 알려지지 않은 실제(true) 회귀선과 추정된 회귀선



ii. 최소제곱추정 (The Least Squared Estimation)

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

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

b. $\min_{b_1, b_2} \sum_t (\hat{\varepsilon}_t)^2 \equiv (y_t - b_1 - b_2 x_t)^2 \equiv S(b_1, b_2)$

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

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

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

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 들이 실현된 관측치가 아닌 확률변수일 때, b_1^{LS}, b_2^{LS} 는 추정치가 아닌 추정량이고 이를 최소제곱추정량이라고 함

III. 최소제곱 추정량의 특성

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

i. 불편추정량

1. $E\left(b_1^{LS}\right) = \beta_1$, $E\left(b_2^{LS}\right) = \beta_2$

ii. 일치추정량

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

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

iii. 선형추정량

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

a. 즉 추정량이 확률표본의 선형결합으로 표현되는 추정량임

iv. BLUE (Best Linear Unbiased Estimator)

1. 가우스-마코프 정리(Gauss-Markov Theorem): 단순선형회구모형에 대한 가정 i) –v)하에서 β_1, β_2 에 대한 최소제곱추정량은 모든 선형 불편 추정량 가운데 가장 작은 분산을 갖는다.
2. 선형 불편 추정량의 범주내에서 최소제곱 추정량외의 다른 추정량을 고려할 필요는 없다.

B. 단순선형모형에 대한 가정 vi) (정규분포에 대한 가정) 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 독립인 두 확률변수의 선형 결합은 정규분포이며, 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^2 에 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^2 에 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. $\widehat{V}\left(b_1\right) = \frac{\hat{\sigma}^2 \sum x_t^2}{T\sum\left(x_t - \bar{x}\right)^2}$, $\widehat{V}\left(b_2\right) = \frac{\hat{\sigma}^2}{\sum\left(x_t - \bar{x}\right)^2}$

      a. 표준편차에 대한 추정량을 특별히 표준오차(standard error)라함

IV. 단순회귀모형에서의 추론

A. 구간추정량 (신뢰구간)의 도출

i. 오차항의 분산 σ^2 이 알려진 경우, 가정 vi)하에서

   1. $Z = \frac{b_2 - \beta_2}{SD\left(b_2\right)} \sim N\left(0,1\right)$
   2. 예컨대 95% 구간추정량은 다음과 같이 도출됨

3. $P\left(-1.96 < Z < 1.96\right) = 0.95 \quad \Rightarrow \quad P\left(-1.96 < \dfrac{b_2 - \beta_2}{SD\left(b_2\right)} < 1.96\right) = 0.95$

$$P\left(-1.96 < \frac{b_2 - \beta_2}{SD\left(b_2\right)} < 1.96\right) = 0.95 \quad \Rightarrow$$

$$P\left(b_2 - 1.96 \cdot SD\left(b_2\right) < \beta_2 < b_2 + 1.96 \cdot SD\left(b_2\right)\right) = 0.95 \quad \Rightarrow$$

$$\left(b_2 - 1.96 \cdot SD\left(b_2\right), b_2 + 1.96 \cdot SD\left(b_2\right)\right)$$

ii. 실제로는 σ^2를 알 수 없으므로, 그 불편 추정량을 이용함

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

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

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

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

2. 예컨대, 95% 구간추정량은 다음과 같이 도출됨

3. $P\left(-t_{0.025,T-2} < t < t_{0.025,T-2}\right) = 0.95 \quad \Rightarrow$

$$P\left(-t_{0.025,T-2} < \frac{b_2 - \beta_2}{SE\left(b_2\right)} < t_{0.025,T-2}\right) = 0.95 \quad \Rightarrow$$

$$\left(b_2 - t_{0.025,T-2} \cdot SE\left(b_2\right), b_2 + t_{0.025,T-2} \cdot SE\left(b_2\right)\right)$$

B. 가설검정

i. 양측검정

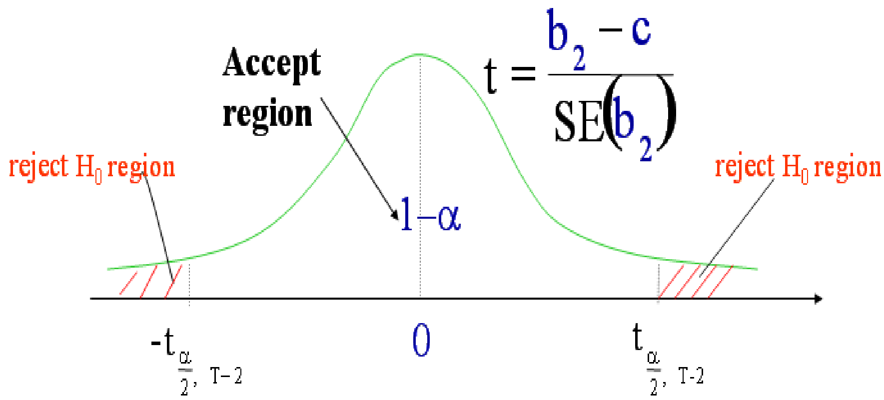
1. 가설: $H_0 : \beta_2 = c$, $H_1 : \beta_2 \neq c$

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

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

iii. P 값 (p-value)

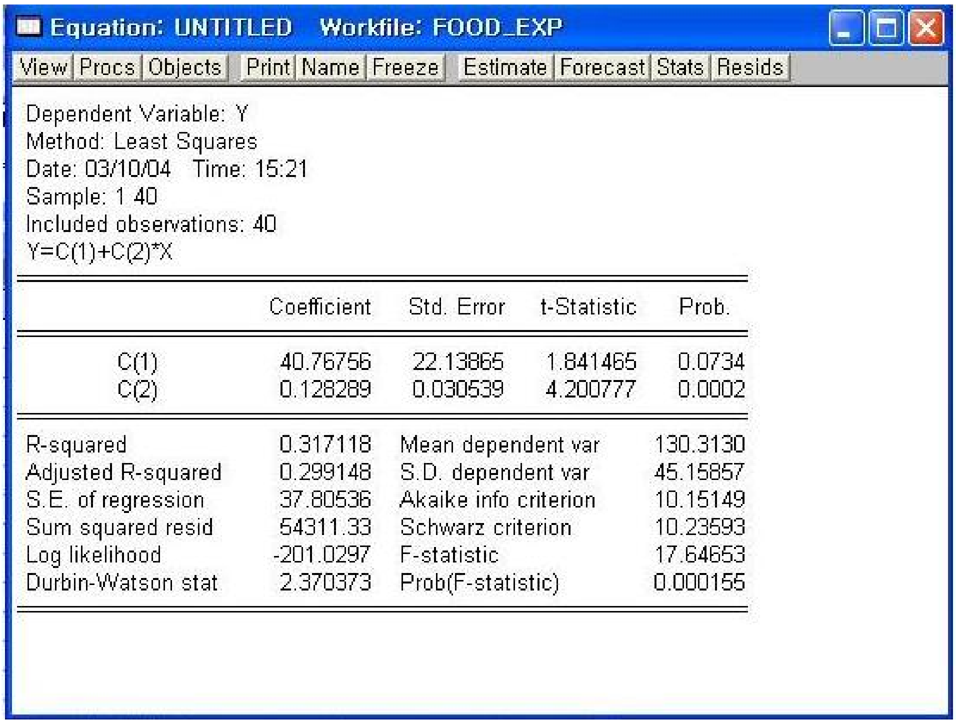
1. p 값은 주어진 검정통계량의 값에 대해 귀무가설을 기각하기 위한 최소크기의 유의 수준임
   a. 양측검정의 경우 : $p - value = 2 \times P\left(t \geq \left|t^*\right|\right)$, 단 t^* 는 검정통계량의 표본값
   b. 우측단측검정의 경우 : $p - value = P\left(t \geq t^*\right)$

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. $H_0 : \beta_2 = 0$, $H_1 : \beta_2 \ne 0$
3. $t = \dfrac{b_2}{SE\left(b_2\right)}$: 컴퓨터에서 자동적으로 보고함

<그림> Eviews 추정 결과 예



Equation: UNTITLED Workfile: FOOD_EXP

View | Procs | Objects | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Y
Method: Least Squares
Date: 03/10/04 Time: 15:21
Sample: 1 40
Included observations: 40
Y=C(1)+C(2)*X

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	40.76756	22.13865	1.841465	0.0734
C(2)	0.128289	0.030539	4.200777	0.0002

R-squared	0.317118	Mean dependent var	130.3130
Adjusted R-squared	0.299148	S.D. dependent var	45.15857
S.E. of regression	37.80536	Akaike info criterion	10.15149
Sum squared resid	54311.33	Schwarz criterion	10.23593
Log likelihood	-201.0297	F-statistic	17.64653
Durbin-Watson stat	2.370373	Prob(F-statistic)	0.000155

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 $V\left(f\right) = \sigma^2\left[1 + \dfrac{1}{T} + \dfrac{\left(x_0 - \bar{x}\right)^2}{\sum\left(x_0 - \bar{x}\right)^2}\right]$, $E\left(f\right) = 0$, 임

1. y_0 에 대한 95% 예측구간은 충분히 큰 관측치나 정규분포의 가정하에

서, $\left(\hat{y}_0 - t_{\frac{\alpha}{2},T-2} \cdot SE(f), \hat{y}_0 + t_{\frac{\alpha}{2},T-2} \cdot SE(f)\right)$로 주어짐

V. 단순선형회귀 모형의 기타 이슈들

A. 결정계수 (coefficient of determination)

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

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

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

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

v. SST = SSR + SSE : 분산분석⇒ $R^2 \equiv \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$: 결정계수

1. 모형 전체의 설명력을 나타내는 지표
2. $0 \le R^2 \le 1$
3. 단순회귀모형에서의 결정계수는 설명변수와 종속변수간의 표본상관계수의 제곱과 같음

B. 단위의 변경과 추정결과

1. 변수들의 측정단위를 변경하는 것은 결정계수는 t 검정통계량의 값에 영향을 주지 않는다.
2. x 의 단위변경은 x 의 계수에 대한 추정량의 크기에만 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 : $y_t = \beta_1 + \beta_2 \frac{1}{x_t} + \varepsilon_t$

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

Part II. 다중회귀모형 및 기본가정의 완화

I. 다중회귀모형

A. 다중회귀모형에 대한 가정

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

- 개별모수들의 의미:
  - ✓ 예컨대, $\beta_2 = \dfrac{\partial E(y_t)}{\partial x_{t2}}$: <u>다른 설명변수들이 일정할 때</u>, x_{t2} 의 한 단위 변화에 대한 종속변수의 평균값의 변화 ⇒ 즉 다른 변수들의 영향력이 통제(control)된 상황에서 첫번째 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, <u>다른 설명변수(들)의 정확한 선형함수가 아님 (즉 완전한 공선성(colinearity)가 존재하지 않음)</u>

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

B. 다중회귀모형(설명변수가 두 개인 경우)에 대한 최소제곱추정

i. $\min\limits_{b_1, b_2, b_3} \sum\limits_t (\hat{\varepsilon}_t)^2 \equiv (y_t - b_1 - b_2 x_{t2} - b_3 x_{t3})^2 \equiv S(b_1, b_2, b_3)$

1. $b_1 = \bar{y} - b_2 \bar{x}_2 - b_3 \bar{x}_3$

2. $b_2 = \dfrac{\left(\sum y_t^* x_{t2}^*\right)\left(\sum x_{t3}^{*2}\right) - \left(\sum y_t^* x_{t3}^*\right)\left(\sum x_{t2}^* x_{t3}^*\right)}{\left(\sum x_{t2}^{*2}\right)\left(\sum x_{t3}^{*2}\right) - \left(\sum x_{t2}^* x_{t3}^*\right)^2}$, 단

   $y_t^* = y_t - \bar{y}$, $x_{t2}^* = x_{t2} - \bar{x}_2$, $x_{t3}^* = x_{t3} - \bar{x}_3$

3. $b_3 = \dfrac{\left(\sum y_t^* x_{t3}^*\right)\left(\sum x_{t2}^{*2}\right) - \left(\sum y_t^* x_{t2}^*\right)\left(\sum x_{t3}^* x_{t2}^*\right)}{\left(\sum x_{t2}^{*2}\right)\left(\sum x_{t3}^{*2}\right) - \left(\sum x_{t2}^* x_{t3}^*\right)^2}$

ii. 최소제곱추정량의 성질

1. 불편추정량
2. 일치추정량
3. 선형추정량
4. BLUE (Gauss-Markov Theorem)
5. 기타 성질

   a. 최소제곱추정에 의한 회귀직선은 $\left(\bar{y}, \bar{x}_2, \bar{x}_3\right)$ 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. $\operatorname{var}(b_2) = \dfrac{\sigma^2}{\sum\left(x_{t2} - \bar{x}_2\right)(1 - r_{23}^2)}$, $\operatorname{var}(b_2) = \dfrac{\sigma^2}{\sum\left(x_{t3} - \bar{x}_3\right)(1 - r_{23}^2)}$, 단

   $r_{23} = \dfrac{\sum\left(x_{t2} - \bar{x}_2\right)\left(x_{t3} - \bar{x}_3\right)}{\sqrt{\sum\left(x_{t2} - \bar{x}_2\right)^2 \sum\left(x_{t3} - \bar{x}_3\right)^2}}$: 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. $z = \dfrac{b_k - \beta_k}{\sqrt{\operatorname{var}\left(b_k\right)}} \sim N\left(0, 1\right),$ for $k = 1, 2, \ldots, K$

   ✓ 오차항의 분산을 모를 경우 이를 이용할 수 없음

2. 오차항의 분산에 대한 불편 추정량

   a. $\hat{\sigma}^2 = \dfrac{\Sigma \hat{\varepsilon}_t^2}{T - K}$

   b. $\dfrac{\left(T - K\right)\hat{\sigma}^2}{\sigma^2}$ □ χ^2_{T-K}

3. 이를 이용하여 최소제곱추정량 b_k 의 표준오차 $\sqrt{\widehat{Var}(b_k)}$ 을 구함

a. $t = \dfrac{b_k - \beta_k}{\sqrt{\hat{\text{var}}\left(b_k\right)}} \sim t_{(T-K)}$

4. 이를 이용하여, 단순회귀모형에서와 마찬가지로 개별모수에 대한 구간추정량과 개별모수에 대한 가설검정(양측가설, 단측가설)을 수행함

ii. 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

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

iii. 두 개 이상의 가설을 동시에 검정하고자 하는 경우 – F 검정을 사용

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

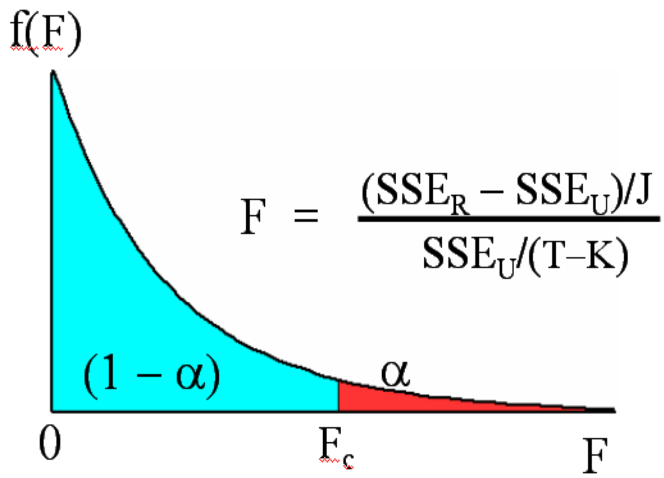
3. F-검정통계량의 구축
   a. 귀무가설이 참이라는 가정하의 (제약을 부과한) 모형에서의 잔차의 제곱의 합을 SSE_R (Restricted sum of squared errors)
   b. 아무런 제약이 부과되지 않은 원래의 모형에서의 잔차의 제곱의 합을 SSE_U (Unrestricted sum of squared errors)
      ✓ $SSE_R - SSE_U \geq 0$ (Why?)
   c. J 를 귀무가설에서의 가설의 수, 즉 부과되는 제약의 수라고 하면
      i. 귀무가설하에서 $V_1 = \dfrac{SSE_R - SSE_U}{\sigma^2} \sim \chi^2_{(J)}$
      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. $F = \dfrac{(SSE_R - SSE_U)/(K-1)}{SSE_U/(T-K)} \sim F_{J,T-K}$

D. 제한최소제곱추정 (Restricted LS)
   i. 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 비표본 정보를 이용하여 추정
         a. $\log\left(y_t / x_{t3}\right) = \beta_1 + \beta_2 \left(x_{t2} / x_{t3}\right) + \varepsilon_t \Rightarrow y_t^* = \beta_1 + \beta_2 x_{t2}^* + \varepsilon_t$
      2. 제한최소제곱추정량의 성질

a. 부과한 제약이 옳지 않다면 추정량은 bias 됨 (cost)

b. 그 분산은 원래의 모형에 대한 추정량의 분산에 비해 작아짐 (benefit)

E. 모형의 설정 (Model Specification)

i. 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 낳게됨

ii. 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

iii. 공선성(Colinearity)의 문제

1. 설명변수 혹인 독립변수들 상호간에 상관관계가 존재할 수 있음

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

2. 공선성의 효과

a. 완전한 공선성의 경우 최소제곱추정 결과를 얻을 수 없음

b. 높은 공선성은 표준오차를 크게 하고 따라서 신뢰구간을 넓힘

c. 이는 높은 결정계수 값과 유의성 있는 F 값에도 불구하고 개별 t 값은 유의성이 없는 것으로 나타내게 함

d. 몇 몇 관측치 또는 유의성 없는 변수의 삭제 또는 포함에 결과가 매우 민감하게 변함

3. 공선성의 판단

a. 두 설명변수간의 높은 상관관계 (0.8 또는 0.9 이상)

b. 한 설명변수를 나머지 설명변수들에 대해 회귀분석시 높은 결정계수의 값 (0.8 이상)

c. 개별 t 값들의 통계적 유의성이 없음에도 유의성 있는 F 값

4. 공선성의 완화

a. 자료의 추가적 확보

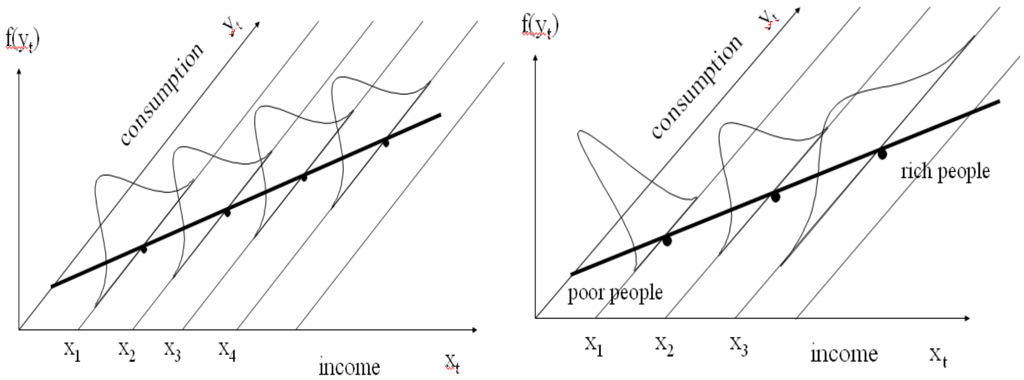
b. 적절한 경제적 또는 통계적 제약의 부과

II. 기본적 가정들의 완화

A. 이분산성 (Heteroskedasticity)

i. 동분산성과 이분산성

<그림> 동분산성과 이분산성



1. $Var(\varepsilon_t) = \sigma_t^2$

2. 이분산성하의 최소제곱추정량

a. 선형추정량, 불편추정량

b. 더 이상 BLUE 가 아님 (유효성이 떨어짐)

c. 통상적 표준오차에 대한 공식은 잘못된 것임 (그에 기반한 신뢰구간이나 가설검정 역시 오류)

i. 단순선형회귀모형에서

1. $V(b_2) = \dfrac{\sigma^2}{\sum (x_t - \bar{x})^2}$: 잘못된 공식

2. $V(b_2) = \dfrac{\sum \sigma_t^2 (x_t - \bar{x})^2}{\left[\sum (x_t - \bar{x})^2\right]^2}$: 정확한 공식

ii. White 의 표준오차

1. $\hat{V}(b_2) = \dfrac{\sum \hat{\varepsilon}_t^2 (x_t - \bar{x})^2}{\left[\sum (x_t - \bar{x})^2\right]^2} \Rightarrow \sqrt{\hat{V}(b_2)}$

2. 대표본하에서 적절

ii. 일반화된 최소제곱추정(Generalized LS Estimation)

1. 두 가지 유형의 이분산성

a. 비례적 이분산성 : $Var(\varepsilon_t) = \sigma_t^2 = \sigma^2 x_t$

b. 분할된 이분산성 :

$$Var(\varepsilon_t) = \sigma_t^2 = \sigma_1.\ \ t = 1, \cdots, T_1 \quad Var(\varepsilon_t) = \sigma_t^2 = \sigma_2, \text{ otherwise}$$

2. 이분산성하의 유효추정(Efficient Estimation)
   a. 비례적 이분산성
      i. $\frac{y_t}{\sqrt{x_t}} = \beta_1 \frac{1}{\sqrt{x_t}} + \beta_2 \frac{x_t}{\sqrt{x_t}} + \frac{\varepsilon_t}{\sqrt{x_t}}$
         1. 즉, $\frac{1}{\sqrt{x_t}}$ 을 가중치로 변수들을 변환
         2. $\Rightarrow y_t^* = \beta_1 x_{t1}^* + \beta_2 x_{t2}^* + \varepsilon_t^* \ \Rightarrow$ 동분산성
      ii. 가중최소제곱 추정(Weighted LS)
         1. 이분산성에 비례적인 변수를 결정
         2. 원래 모형을 그 변수의 제곱근의 값으로 나누어 줌
         3. 변형된 모형에 최소제곱추정을 적용 (상수항 없음)
   b. 분할된 이분산성
      i. $\frac{y_t}{\sigma_1} = \beta_1 \frac{1}{\sigma_1} + \beta_2 \frac{x_t}{\sigma_1} + \frac{\varepsilon_t}{\sigma_1} \qquad t = 1, \ldots, T_1 \ \ \Rightarrow \ \operatorname{var}\left(\frac{\varepsilon_t}{\sigma_1}\right) = 1$
      ii. $\frac{y_t}{\sigma_2} = \beta_1 \frac{1}{\sigma_2} + \beta_2 \frac{x_t}{\sigma_2} + \frac{\varepsilon_t}{\sigma_2} \qquad t = T_1 + 1, \ldots, T \Rightarrow \ \operatorname{var}\left(\frac{\varepsilon_t}{\sigma_2}\right) = 1$
      iii. σ_1^2 와 σ_2^2 를 모름 $\Rightarrow$ $\hat{\sigma}_1^2$ 와 $\hat{\sigma}_2^2$ 를 각각 $t = 1, \ldots, T_1$ 및 $t = T_1 + 1, \ldots, T$ 의 표본으로부터 추정하여 사용
3. 이분산성의 탐지- Goldfeld-Quandt Test
   i. 비례적이분산성의 경우 비례하는 변수를 판별하여 그 변수의 크기대로 자료를 sorting 함
   ii. 자료를 T_1, T_2 의 표본크기를 갖는 두 개의 그룹(비례적 이분산성의 경우 중간 부분은 버림)으로 나누어 양 그룹의 분산의 크기가 같은가에 대한 가설을 검정
      1. $H_0 : \sigma_1^2 = \sigma_2^2, \ H_1 : \sigma_1^2 > \sigma_2^2$
      2. $GQ = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \sim F_{T_1-K, T_2-K}$
      3. 높은 GQ 값에 대해 귀무가설을 기각함

B. 자기상관 (Autocorrelation)

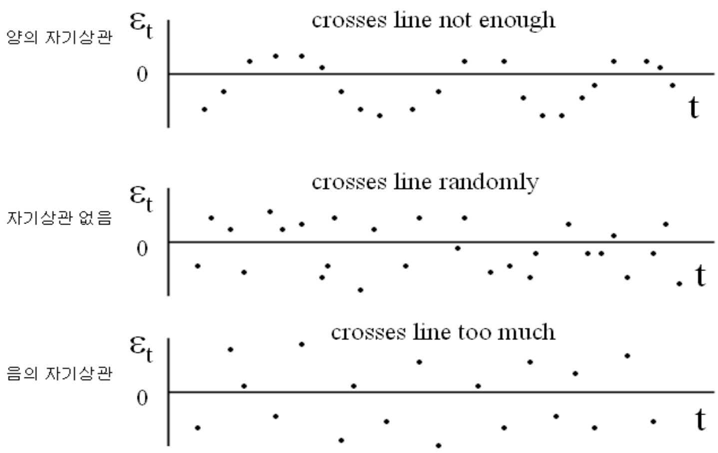
i. 자기상관의 본질

1. 자기상관이란 오차항에 존재하는 체계적인 패턴의 한 형태

a. 양의 자기상관 : 현재의 오차항이 이전의 오차항과 같은 부호를 갖는 경향이 있는 경우

b. 음의 자기상관 : 현재의 오차항이 이전의 오차항과 다른 부호를 갖는 경향이 있는 경우(경제자료에서는 매우 드묾)

<그림> 자기상관



ii. 1 차 자기상관 모형(AR(1))

1. $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$, $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$,

a. $E(v_t) = 0$, $\text{var}(v_t) = \sigma_v^2$, $\text{cov}(v_t, v_s) = 0$, $t \neq s$, $-1 < \rho < 1$

2. $\Rightarrow \varepsilon_t = v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \rho^3 v_{t-3} + \cdots$,

3. $\Rightarrow E(e_t) = 0$, $\text{var}(e_t) = \sigma_e^2 = \dfrac{\sigma_v^2}{1-\rho^2}$, $\text{cov}\left(e_t, e_{t-k}\right) = \sigma_e^2 \rho^k$, $k > 0$

iii. 자기상관 존재시 최소제곱 추정량

1. 불편추정량, 선형추정량 그러나 유효추정량은 아님

2. 통상적 표준오차에 대한 공식은 잘못된 것임 (그에 기반한 신뢰구간이나 가설검정 역시 오류)

iv. 일반화된 최소제곱추정(GLS)

1. 변형

a. $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$, $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$

b. $\Rightarrow y_t = \beta_1 + \beta_2 x_t + \rho\varepsilon_{t-1} + \nu_t$

c. $\Rightarrow\ y_t = \beta_1 + \beta_2 x_t + \rho\left(y_{t-1} - \beta_1 - \beta_2 x_{t-1}\right) + \nu_t$

d. $\Rightarrow\ y_t - \rho y_{t-1} = \beta_1\left(1-\rho\right) + \beta_2\left(x_t - \rho x_{t-1}\right) + \nu_t$

e. $\Rightarrow\ y_t^* = \beta_1 x_{t1}^* + \beta_2 x_{t2}^* + \nu_t,\ t = 2, 3, \cdots T$

2. 두 가지 문제점

a. 변형과정에서 관측치 하나가 줄어듦

i. $y_1 = \beta_1 + x_1\beta_2 + \varepsilon_1$ 를 그대로 추가하는 것은 이분산성의 문제가 발생 ($\mathrm{var}(e_1) = \sigma_e^2 = \sigma_v^2 \big/ (1-\rho^2)$)

ii. $\sqrt{1-\rho^2}\,y_1 = \sqrt{1-\rho^2}\,\beta_1 + \sqrt{1-\rho^2}\,x_1\beta_2 + \sqrt{1-\rho^2}\,\varepsilon_1$ 를 추가

1. $\Rightarrow y_1^* = x_{11}^*\beta_1 + x_{12}^*\beta_2 + \varepsilon_1^*$

b. 자기상관계수(ρ)의 값을 모름

i. $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$: ε_t 를 안다면 회귀분석에 의해 ρ를 추정

ii. $\Rightarrow\ \hat{\varepsilon}_t = y_t - b_1 - b_2 x_t \Rightarrow\ \hat{\varepsilon}_t = \rho\hat{\varepsilon}_{t-1} + v_t$

iii. $\Rightarrow\ \hat{\rho} = \dfrac{\sum_{t=2}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1}}{\sum_{t=2}^{T}\hat{\varepsilon}_{t-1}^2}$

v. 자기상관의 탐지

1. Durbin-Watson Test

a. $H_0 : \rho = 0,\ H_1 : \rho > 0$

b. $d = \dfrac{\sum_{t=2}^{T}\left(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1}\right)^2}{\sum_{t=1}^{T}\hat{\varepsilon}_t^2}$

c. $\Rightarrow\ d \approx 2\left(1-\hat{\rho}\right)\ \Rightarrow\ \hat{\rho} = 0$ 일 때 $d \approx 2$, $\hat{\rho} = 1$ 일 때 $d \approx 0$

d. DW 검정통계량의 분포는 설명변수의 값에 의존

i. 표에서 d_L, d_U 를 구할 수 있으며 이를 이용

2. Lagrange Multiplier Test (LM test)

a. $y_t = \beta_1 + \beta_2 x_t + \rho\varepsilon_{t-1} + \nu_t$

b. $\Rightarrow\ y_t = \beta_1 + \beta_2 x_t + \rho\hat{\varepsilon}_{t-1} + \nu_t$

c. $\Rightarrow$ t 검정 또는 F 검정을 통해 $\hat{\varepsilon}_{t-1}$에 대한 유의성을 검정

d. $\Rightarrow$ LM test 는 대표본에서 유효, 1 차자기상관 뿐 아니라 고차 자기상관에 대한 탐지에도 사용가능

vi. 자기상관존재 시 예측

1. 자기상관이 존재하는 경우 이전의 오차가 다음기의 오차를 예측하는데 도움이 되며, 이 정보를 이용하는 것이 최선의 예측방법임

2. $\hat{y}_{T+1} = \hat{\beta}_1 + \hat{\beta}_2 x_{T+1} + \hat{\rho}\tilde{e}_T$, $\hat{\beta}_1, \hat{\beta}_2$ 는 일반화된 최소제곱추정량,

   $\tilde{e}_T = y_T - \hat{\beta}_1 - \hat{\beta}_2 x_T$

3. h 기 후의 예측: $\hat{y}_{T+h} = \hat{\beta}_1 + \hat{\beta}_2 x_{T+h} + \hat{\rho}^h \tilde{e}_T$

C. 확률변수인 설명변수

i. 새로운 단순선형 모형에 대한 기본가정

1. 기본가정

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

b. (x_t, y_t), $t = 1,\ldots,T$ 는 확률표본(각 쌍은 동일하고 독립적인 분포)

c. $E(\varepsilon|x) = 0$

d. $V(\varepsilon|x) = \sigma^2$

e. x 는 적어도 두 개의 다른 값을 가져야 한다.

f. $\varepsilon|x \sim N(0,\sigma^2)$

2. 이러한 완화된 가정하의 최소제곱추정량

a. a-e 까지의 가정하에서 최소제곱추정량은 BLUE (소표본성질)

b. a-f 까지의 가정하에서 최소제곱추정량은 x 에 조건부 정규분포를 하며 그 분산 역시 통상적인 방법으로 계산되며 그에 근거한 구간추정이나 가설검정 역시 유효함(소표본성질)

c. a-e 까지의 가정하에서 최소제곱추정량은 일치추정량 (대표본성질)

3. 가정 c 을 좀 더 약한 다음의 가정으로 대체

a. 가정 c’: $E(\varepsilon) = 0$, $Cov(x,\varepsilon) = 0$

i. $E(\varepsilon|x) = 0$ 이면 $E(\varepsilon) = 0$, $Cov(x,\varepsilon) = 0$ 이나 그 역은 성립하지 않음

ii. 이 경우 최소제곱추정량의 소표본 성질은 더 이상 성립하지 않음

iii. 최소제곱추정량은 대표본에서 정규분포로 근사할 수 있음

iv. 대표본하에서 통상적인 구간추정이나 가설검정은 유효함

4. 가정 c’ 가 위배될 경우, 특히 설명변수와 오차항간에 상관되어 있을 경우

a. 최소제곱추정량은 일치추정량이 되지 못함

b. 통상적 구간추정이나 가설검정은 유효하지 못함

c. 즉, 설명변수가 확률변수로 간주되는 경우, 설명변수와 오차항간의 상관관계가 최소제곱추정법 적용의 적절함을 판단함에 있어서 결정적임

ii. 설명변수가 오차항과 상관되어 있을 경우

1. 변수오차(Errors in Variables)의 경우: 설명변수가 오차를 가지고 측정될 경우

   a. $y_t = \beta_1 + \beta_2 x_t^* + \varepsilon_t$ (임금, 능력-측정이 어려움)

   b. $\Rightarrow$ $x_t = x_t^* + u_t$ (표준화된 시험 점수: 대용변수(proxy variable))

   * $E(u_t) = 0$, $V(u_t) = \sigma_u^2$, $Cov(u_i, u_j) = 0$, $u \perp \varepsilon$

   c. $\Rightarrow y_t = \beta_1 + \beta_2 (x_t - u_t) + \varepsilon_t = \beta_1 + \beta_2 x_t + \varepsilon_t - \beta_2 u_t$

   $\Rightarrow$ $y_t = \beta_1 + \beta_2 x_t + \eta_t$

   d. $\Rightarrow$ x_t 는 확률변수이며 이 회귀식에 대한 최소제곱추정은 비일치 추정을 낳게됨

   * $\Leftarrow$
   $$\begin{aligned}\mathrm{cov}(x_t, \eta_t) &= E(x_t \eta_t) = E\left[\left(x_t^* + u_t\right)\left(\varepsilon_t - \beta_2 u_t\right)\right] \\ &= E\left(-\beta_2 u_t^2\right) = -\beta_2 \sigma_u^2 \neq 0\end{aligned}$$

2. 연립방정식의 경우: 설명변수가 동시성(Simultaneity)을 가짐

   a. $C_t = \beta_1 + \beta_2 Y_t + \varepsilon_t$ (케인지안 소비함수)

   b. $Y_t = C_t + I_t$ (소득항등식)

   c. $\Rightarrow$ C_t, Y_t 는 모두 내생변수로서 시스템에서 동시에 결정됨

   d. $\Rightarrow$ 소비함수에 대한 최소제곱추정은 비일치 추정을 낳게 됨

   *
   $$\begin{aligned}\mathrm{cov}(Y_t, \varepsilon_t) &= E\left[\left(C_t + I_t\right)\varepsilon_t\right] \\ &= E\left[\left(\beta_1 + \beta_2 Y_t + \varepsilon_t\right)\varepsilon_t\right] \neq 0\end{aligned}$$

iii. 적률방법에 의한 추정(Method of Moment Estimation) : 일치추정량을 제공

2. 수학적 적률과 표본적률을 등치시킴으로써 모수의 추정량을 얻는 방법

   a. $E\left(Y^k\right) = \mu_k$: 확률변수 Y 의 k 번째 (수학적) 적률

   b. $\hat{E}\left(Y^k\right) = \hat{\mu}_k = \sum_{t=1}^{T} y_t^k \Big/ T$: 확률변수 Y 의 k 번째 표본적률

3. 단순회귀모형에서의 적률방법추정

   a. $E(\varepsilon_t) = 0 \quad \Rightarrow \quad E\left(y_t - \beta_1 - \beta_2 x_t\right) = 0$: 적률조건 1

   b. $E(x_t \varepsilon_t) = 0 \quad \Rightarrow \quad E\left[x_t\left(y_t - \beta_1 - \beta_2 x_t\right)\right] = 0$:적률조건 2 (x 가

확률변수가 아니거나, 확률변수라해도 가정 c'를 만족)

c. $\Rightarrow \begin{matrix} \frac{1}{T}\sum\left(y_t - b_1 - b_2 x_t\right) = 0 \\ \frac{1}{T}\sum x_t\left(y_t - b_1 - b_2 x_t\right) = 0 \end{matrix}$ $\Rightarrow$ 최소제곱추정량과 동일한 결과

4. x 가 확률변수이고 오차항과 상관되어 있을 경우

a. $Cov\left(z_t, \varepsilon_t\right) = 0$ 을 만족하는 다른 변수(instrumental variable, 도구변수 또는 수단변수라 부름) z 를 찾음 (도구변수추정법)

b. $E\left(\varepsilon_t\right) = 0 \quad \Rightarrow \quad E\left(y_t - \beta_1 - \beta_2 x_t\right) = 0$: 적률조건 1

c. $E\left(z_t \varepsilon_t\right) = 0 \quad \Rightarrow \quad E\left[z_t\left(y_t - \beta_1 - \beta_2 x_t\right)\right] = 0$:적률조건 2

d. $\Rightarrow \begin{matrix} \frac{1}{T}\sum\left(y_t - b_1 - b_2 x_t\right) = 0 \\ \frac{1}{T}\sum z_t\left(y_t - b_1 - b_2 x_t\right) = 0 \end{matrix}$

e. $\Rightarrow \begin{matrix} \hat{\beta}_2 = \dfrac{\sum z_t \sum y_t - T\sum z_t y_t}{\sum z_t \sum x_t - T\sum z_t x_t} = \dfrac{\sum\left(z_t - \bar{z}\right)\left(y_t - \bar{y}\right)}{\sum\left(z_t - \bar{z}\right)\left(x_t - \bar{x}\right)} \\ \hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x} \end{matrix}$

5. 적률방법추정량의 성질

a. 일치추정량,

b. 대표본하에서 정규분포를 함 $\hat{\beta}_2 \sim N\left(\beta_2, \dfrac{\sigma^2}{\sum\left(x_t - \bar{x}\right)^2 r_{zx}^2}\right)$

* r_{zx}^2 : z 와 x 간의 표본상관의 제곱

* x 와 높은 상관을 가지는 도구변수가 바람직함을 나타냄

c. 오차항의 분산은 다음과 같이 추정함

* $\hat{\sigma}_{IV}^2 = \dfrac{\sum\left(y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t\right)^2}{T-2}$

6. 2 단계최소제곱추정량: 도구 변수가 필요한 것보다 많을 경우

a. z 외에 w 도 도구변수로 역할을 함

b. 다음의 2 단계를 통해 적절한 적률방법추정량을 얻을 수 있음

i. x를 상수항, z, w에 회귀하여 예측치 $\hat{x}$를 얻음

ii. 이 예측치 $\hat{x}$를 x에 대한 도구변수로 사용하여 추정량을 얻음

iii. $\Rightarrow \hat{\beta}_2 = \frac{\sum(\hat{x}_t - \bar{\hat{x}})(y_t - \bar{y})}{\sum(\hat{x}_t - \bar{\hat{x}})(x_t - \bar{x})} = \frac{\sum(\hat{x}_t - \bar{x})(y_t - \bar{y})}{\sum(\hat{x}_t - \bar{x})(x_t - \bar{x})}$

$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$

iv. 여기서 예측치 $\hat{x}$를 설명변수로 사용하여 추정량을 얻을 수도 있음 (2단계 최소제곱추정법)

c. 대표본하에서 $\mathrm{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum(\hat{x}_t - \bar{x})^2}$, $\hat{\sigma}^2_{IV} = \frac{\sum(y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t)^2}{T-2}$

iv. 설명변수와 오차항간의 상관의 탐지 : Hausman Test

1. $H_0 : Cov(x, \varepsilon) = 0$, $H_1 : Cov(x, \varepsilon) \neq 0$

2. 귀무가설하에서 최소제곱추정량과 도구변수추정량 모두 일치추정량

a. 귀무가설하에서 $q = (b_{ols} - \hat{\beta}_{IV}) \to 0$

b. 대립가설하에서 $q = (b_{ols} - \hat{\beta}_{IV}) \to c \neq 0$

c. Davidson and Mackinnon

i. 오차항과의 상관이 의심되는 x_t를 도구변수 z_{t1}와 z_{t2}에 대해 회귀하여 얻은 잔차를 $\hat{v}_t$라 할 때, 이를 원래의 회귀식에 설명변수로 포함시킴

ii. $\Rightarrow$ $y_t = \beta_1 + \beta_2 x_t + \delta \hat{v}_t + \varepsilon_t$: 이 회귀결과로부터 잔차 $\hat{v}_t$에 대한 유의성 검정으로부터 가설검정을 수행

iii. $\Rightarrow$ x와 오차항간에 상관이 있다면 오차항과 상관 없는 도구변수들에 의해 설명되고 남는 부분($\hat{v}_t$)이 오차항과 상관될 것이고 $\delta \neq 0$일 것임