2022/11/25

藪友良

第 8 章の答え

練習問題 1

均一分散は現実には成立しないことが多く、不均一分散が現実的仮定である。

練習問題 2

無作為抽出した横断面データでは、ランダムになっているため、誤差項は相互に無相関となる。

練習問題 3

時系列データでは、誤差項は相互に関係している。これは、何らかのイベントが発生すると、それは現在だけでなく、将来にも影響することが多いためである。

練習問題 4

説明変数X_iと誤差項u_iに相関がないとき、つまり、

$$Cov(X_i, u_i) = 0$$

であれば、説明変数には外生性がある、という。これに対し、説明変数X_iと誤差項u_iに相関があるとき、つまり、

$$Cov(X_i, u_i) \neq 0$$

であれば、説明変数には内生性がある、という。

練習問題 5

説明変数に内生性があると、OLS 推定量は不偏性だけでなく、一致性も持たない。つまり、サンプルサイズが大きくなっても、バイアスは消えない。

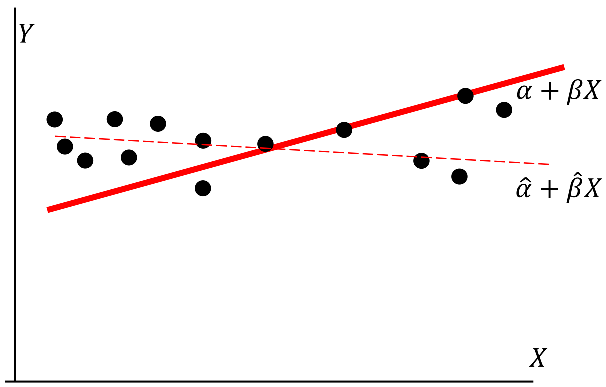
練習問題 6

下図において、実線が真のXとYの関係$(\alpha + \beta X)$を表し、点線が推定された回帰直線$(\hat{\alpha} + \hat{\beta} X)$を表す。ただし、係数$\beta$は正$(\beta > 0)$としており、実線$(\alpha + \beta X)$は右上

がりの関係となる。また、説明変数X_iと誤差項u_iに負の相関がある($Cov(X_i, u_i) < 0$)。

説明変数X_iと誤差項u_iに負の相関があるため、説明変数X_iが小さな値だと、誤差項u_iは大きな値となり、$Y_i = \alpha + \beta X_i + u_i$という関係から実線$(\alpha + \beta X)$の上でデータが観察されやすくなる。逆に、説明変数$X_i$が大きな値だと、誤差項$u_i$は小さな値となり(つまり、負の値)、$Y_i = \alpha + \beta X_i + u_i$という関係から実線$(\alpha + \beta X)$より下でデータが観察されやすくなる。したがって、回帰直線は点線のようになり、OLS推定量$\hat{\beta}$は負のバイアスを持つ。

図　内生性とバイアスの関係



練習問題 7

推定量β^*の確率的表現は、次のようになる。

$$\beta^* = \frac{Y_2 - Y_1}{X_2 - X_1} = \frac{(\alpha + \beta X_2 + u_2) - (\alpha + \beta X_1 + u_1)}{X_2 - X_1}$$

$$= \frac{\beta(X_2 - X_1) + (u_2 - u_1)}{X_2 - X_1} = \beta + \frac{u_2 - u_1}{X_2 - X_1}$$

まず、確率的表現の期待値をとると、

$$E[\beta^*] = E\left[\beta + \frac{u_2 - u_1}{X_2 - X_1}\right] = \beta + \frac{E[u_2] - E[u_1]}{X_2 - X_1} = \beta$$

となり、不偏性を満たすことが確認できる(誤差項は標準的仮定を満たすため、$E[u_1] = E[u_2] = 0$とした)。

次に、推定量の分散は、

$$E\left[\left(\frac{Y_2-Y_1}{X_2-X_1}-\beta\right)^2\right]=E\left[\left(\frac{u_2-u_1}{X_2-X_1}\right)^2\right]=\frac{E[(u_2-u_1)^2]}{(X_2-X_1)^2}$$

$$=\frac{E[u_1^2]+E[u_2^2]-2E[u_1u_2]}{(X_2-X_1)^2}=\frac{2\sigma^2}{(X_2-X_1)^2}$$

となる(誤差項は標準的仮定を満たすため、$E[u_1^2]=E[u_2^2]=\sigma^2$、$E[u_1u_2]=0$とした)。ここで、分散の分母は$(X_2-X_1)^2$であり、$X_1$と$X_2$が互いに離れているほど、分散は小さくなることがわかる。

練習問題 8[1]

(a) 平均 2 乗誤差(MSE)は、次のように分解できる。

$$MSE=E\left[\left(\hat{\theta}-\theta\right)^2\right]$$

$$=E\left[\left(\left(\hat{\theta}-\mathrm{E}[\hat{\theta}]\right)+\left(\mathrm{E}[\hat{\theta}]-\theta\right)\right)^2\right]$$

$$=E\left[\left(\hat{\theta}-\mathrm{E}[\hat{\theta}]\right)^2\right]+\left(\mathrm{E}[\hat{\theta}]-\theta\right)^2+2\left(\mathrm{E}[\hat{\theta}]-\theta\right)E\left[\left(\hat{\theta}-\mathrm{E}[\hat{\theta}]\right)\right]$$

$$=E\left[\left(\hat{\theta}-\mathrm{E}[\hat{\theta}]\right)^2\right]+\left(\mathrm{E}[\hat{\theta}]-\theta\right)^2$$

式展開では、$(\mathrm{E}[\hat{\theta}]-\theta)$は固定した値なので期待値の外に出せること、$E\left[\left(\hat{\theta}-\mathrm{E}[\hat{\theta}]\right)\right]=\mathrm{E}[\hat{\theta}]-\mathrm{E}[\hat{\theta}]=0$であることを用いた。

上式の右辺第 1 項は推定量の分散$V(\hat{\theta})=E\left[\left(\hat{\theta}-\mathrm{E}[\hat{\theta}]\right)^2\right]$となる。右辺第 2 項は、推定量のバイアス$Bias(\hat{\theta})=\mathrm{E}[\hat{\theta}]-\theta$の 2 乗となる。以上から、MSE は次のように表現できる。

$$MSE=E\left[\left(\hat{\theta}-\theta\right)^2\right]=V(\hat{\theta})+Bias(\hat{\theta})^2$$

一般に、「MSE が小さいほど良い推定量である」と判断される。MSE が小さい推定量とは、推定量の分散$V(\hat{\theta})$が小さく、推定量のバイアス$Bias(\hat{\theta})$も小さい推定となる。

MSE の理解を深めるため、下図では、2 種類の推定量を示している。(a)の推定量は不偏性を満たしているが、推定量の分散は大きくなっている。これに対して、(b)の推定量はバイアスはあるが、推定量の分散は小さくなってい

[1] なお、『実践する統計学』の 7 章では、(推定量が不偏性を満たしていれば)推定量が有効であるとは、推定量の分散が小さいことであるとした。本問題では、推定量が不偏性を満たしていない可能性を考慮した有効性の一般的な定義を示している。

る。MSE でみると、(b)のほうが小さくなるため、(b)がより望ましい推定量と判断される。

本書では扱わないが、リッジ推定量といわれる推定量があり、この推定量はバイアスを持っているが、推定量の分散は小さいため、MSE の小さい推定量の 1 つとして知られている。

図：2 つの推定量の比較

(a)バイアスなし、分散は大きい　　**（b）バイアスあり、分散は小さい**

(b) $\hat{\theta}$が不偏推定量であるとしよう。このとき、

$$\mathrm{E}[\hat{\theta}] = \theta$$

であるため、バイアスは

$$Bias(\hat{\theta}) = \mathrm{E}[\hat{\theta}] - \theta = 0$$

となる。このため、MSE は、推定量の分散$V(\hat{\theta})$と一致する。

$$MSE = E\left[\left(\hat{\theta} - \theta\right)^2\right] = V(\hat{\theta})$$

つまり、不偏推定量であれば、分散$V(\hat{\theta})$が小さい推定量が最も望ましいといえる(有効な推定量となる)。ガウス=マルコフの定理では、不偏推定量だけを考えていたため、分散が最小となる OLS 推定量が有効な推定量としていた。