2025/06/24

藪友良

初版以降に新しく追加した練習問題

初版以降に新しく追加した練習問題を掲載しています。購入時期によっては、全ての問題が未掲載かもしれませんし、全て掲載されているかもしれません。なお、これらは少し難易度が高い問題になっていますので、模範解答が理解できたら十分かと思います。

3 章の練習問題

13. ★単回帰モデル$Y_i = \alpha + \beta X_i + u_i$における残差$\hat{u}_i$の性質を考えよう。

(a) $\hat{u}_i = u_i - (\hat{\alpha} - \alpha) - (\hat{\beta} - \beta)X_i$を示せ。

(b) $E[\hat{u}_i^2] = \sigma^2(1 - h_{ii})$を示せ。ここで、$h_{ii}$はレバレッジと呼ばれ、

$$h_{ii} = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

と定義される。なお、h_{ii}は 0 以上 1 以下であり、1 に近いほど、i 番目のデータX_iが他のデータに比べて異なることを意味する(h_{ii}が 1 に近いと外れ値の可能性あり)。

(c) $E[\sum_{i=1}^{n} \hat{u}_i^2] = (n-2)\sigma^2$、$E[s^2] = \sigma^2$を示せ。Hint: $\sum_{i=1}^{n} h_{ii} = 2$

(d) 標準化残差$\bar{u}_i$を$(1 - h_{ii})^{-1/2}\hat{u}_i$と定義する。このとき、$E[\bar{u}_i] = 0$、$E[\bar{u}_i^2] = \sigma^2$を示せ。$\sigma^2$の新たな推定量を$s^2 = \frac{1}{n}\sum_{i=1}^{n} \bar{u}_i^2$としたとき、$s^2$が不偏性を満たすことを示せ。

5 章の練習問題

14. ★(a)～(e)において、多重共線性が生じる理由を述べよ。

(a) 勤続年数が分からないため、勤続年数=年齢－教育年数－6 として計算した。説明変数は、年齢、教育年数、勤続年数とする。

(b) 説明変数は、ダミー変数D_i、その 2 乗D_i^2とする。

(c) 説明変数X_{1i}は、すべて 0 となる変数とする。

(d) 説明変数X_{1i}は、すべて 1 となる変数とする。

(e) 説明変数の数Kはサンプルサイズn以上となる($n \leq K$)。

15. ★ モデルは$Y_i = \alpha + \beta_1 X_i + \beta_2 W_{1i} + \beta_3 W_{2i} + u_i$だが、単回帰モデル$Y_i = \alpha + \beta_1 X_i + u_i^*$を推定したとする(ただし、$u_i^* = \beta_2 W_{1i} + \beta_3 W_{2i} + u_i$)。

(a) OLS 推定量$\hat{\beta}_1$の期待値は、次式になることを示せ。

$$E[\hat{\beta}_1] \ = \ \beta_1 + \beta_2 \ \frac{s_{XW_1}}{s_X^2} + \beta_3 \ \frac{s_{XW_2}}{s_X^2}$$

ただし、s_{XW_1}はXとW_1との標本共分散、s_{XW_2}はXとW_2との標本共分散である。

(b) β_2とs_{XW_1}が同じ符号なら$E[\hat{\beta}_1] \ > \ \beta_1$となるか。

16.　完全な多重共線性とは、ある説明変数X_jがX_j以外の説明変数の線形関数であるといえる。つまり、説明変数はX_0、X_1、…、X_Kとしたとき、任意の定数c_0、c_1、…、c_Kを用いて、次のようになる(ただし、$c_0 = c_1 = \cdots = c_K = 0$ではない)。

$$X_j = \sum_{\ell \neq j} c_\ell X_\ell$$

これが正しいことを示せ。ここで、X_0は常に 1 をとる変数とする(定数項αはαX_0であり、X_0も説明変数の 1 つとみなせる)。なお、藪友良『入門　実践する統計学』(東洋経済新報社、2012 年)では、これを完全な多重共線性の定義としている。

17.　ある新薬は血圧を下げることで病気の症状を緩和させる効果があると仮定しよう。新薬の効果を測るため、被説明変数を患者の症状とし、説明変数を新薬ダミー(新薬を投与したら 1、そうでないと 0)と投与後の血圧とした OLS 推定をしたい。この推定の問題を述べよ。

<u>6 章の練習問題</u>

15. ★　所得Y_iの自然対数である$\ln(Y_i)$を分析したい。

(a) 男性所得の対数の平均は 6.2、女性所得の対数の平均は 5.9 とする。この結果から、男性は女性より所得が平均 30%高いといえるか。Hint: Yの幾何平均は$(Y_1 \times Y_2 \times \ldots \times Y_n)^{1/n}$と定義される[1]。

(b) 被説明変数を所得の対数、説明変数を教育年数としたところ、係数は 0.1 と推定された。この結果から、教育年数が 1 年増えると、所得は平均 10%増えるといえるか。

16. ★　対数線形モデル$\ln(Y_i) = \alpha + \beta X_i + u_i$から$Y_i$の予測を考えたい。ここでパラメータ($\alpha$、$\beta$)は既知とする。

(a) このモデルは$Y_i = e^{\alpha + \beta X_i} e^{u_i}$と表せることを示せ。

(b) 説明変数が非確率変数なら$E[Y_i] = e^{\alpha + \beta X_i} E[e^{u_i}]$となる。このとき、$E[Y_i] = e^{\alpha + \beta X_i}$といえるか。Hint: 仮に$u_i$が$N(0, \sigma^2)$なら$e^{u_i}$は対数正規分布となり、$\sigma^2 > 0$なら$E[e^{u_i}] = e^{\sigma^2/2} > 1$となる[2]。

(c) $E[e^{u_i}]$の推定方法を考えよ。なお、$E[e^{u_i}]$が推定できれば、その結果を用いて$E[Y_i]$も推

[1] 幾何平均は、平均変化率を求めるのに便利な指標である。たとえば、株価の変化率が 70%、-20%、10%としよう。このとき、株価は 1.7 倍、0.8 倍、1.1 倍となるので、幾何平均は$(1.7 \times 0.8 \times 1.1)^{1/3} = 1.14$となる。これに対して、通常の平均は$1.2 (= (1.7 + 0.8 + 1.1)/3)$となる。幾何平均は$1.7 \times 0.8 \times 1.1 = 1.14^3$となり、平均変化率として適切であることがわかる。一方、平均は$1.7 \times 0.8 \times 1.1 < 1.2^3$となり、平均変化率を過大に評価し不適切である。

[2] 対数正規確率変数は、その対数が正規分布する確率変数です。つまり、uが正規分布ならe^uは対数正規分布に従います(e^uの対数は$\ln(e^u) = u$です)。対数正規確率変数は、指数関数ですから 0 より大きな値を取ります。対数正規分布の例としては、体重、所得、資産などがあります。これらは右の裾野が長い分布ですが、対数をとるとベル状の分布になります。

定できる。

17. 貿易における重力モデルは次式とし、誤差項u_{ij}はGDP_i、GDP_j、$Distance_{ij}$とは独立と仮定する[3]。

$$\text{Trade}_{ij} = A\frac{GDP_i^{\beta_1}GDP_j^{\beta_2}}{Distance_{ij}^{\beta_3}} + u_{ij}$$

(a) これは次のように変形できることを示せ。

$$\ln(\text{Trade}_{ij}) = \ln(A) + \beta_1\ln(GDP_i) + \beta_2\ln(GDP_j) - \beta_3\ln(\text{Distance}_{ij}) + \ln(\eta_{ij})$$

ここで、$\ln(\eta_{ij})$は新しい誤差項であり、η_{ij}は次のように定義される。

$$\eta_{ij} = \left(1 + \frac{1}{A}\frac{Distance_{ij}^{\beta_3}}{GDP_i^{\beta_1}GDP_j^{\beta_2}}u_{ij}\right)$$

(b) (a)のモデルでは、説明変数($\ln(GDP_i)$、$\ln(GDP_j)$、$\ln(\text{Distance}_{ij})$)と誤差項$\ln(\eta_{ij})$が相関することを述べよ。これは内生性といわれる現象であり、8.3節で詳しく議論される。

(c) 貿易額は0をとることがあり、とくに小さな国でよく見られる。貿易額のデータを対数対数モデルとして扱う問題を議論せよ。

8章の練習問題

9. 7章例7-2では、$Y_t = \alpha + \beta Y_{t-1} + u_t$を推定した。ここで、$Y_{t-1}$は非確率変数とした仮定は妥当か。

10. ★説明変数は確率変数ではあるが、外生変数としよう(つまり、$Cov(X_i, u_i) = 0$)。このとき、OLS推定量は一致性を満たすが、不偏性は満たさないことを示せ。Hint: nが大きくなると、平均は期待値に収束し、標本分散は分散に収束するとする(『入門 実践する統計学』(巻末参考文献[5])の5.2、5.3節参照)。

9章の練習問題

10. ★ 均一分散が正しい状況を考えよう($E[u_i^2] = \sigma^2$)。ただし、誤差項は互いに独立とする。この問題では、均一分散が正しい状況のもとで、不均一分散に対して頑健な分散の推定量の性質を調べ、推定量を改善する方法を考察する。

(a) 不均一分散に対して頑健な分散の推定量$s_{\hat{\beta}}^2$は、真の分散より小さくなること、つまり、次式が正しいことを示せ。

---

[3] この練習問題は次の論文に基づいて作成しています。Silva, J. M. C. S., and Tenreyro, S. (2006) “The Log of Gravity," *Review of Economics and Statistics*, 88(4), 641-658. 重力モデルの推定に関心がある方は、サポートウェブサイトの追加資料「貿易における重力モデル」と「カウントデータ」を参照してください。

$$E\left[\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2\hat{u}_i^2}{\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right)^2}\right] < \frac{\sigma^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

Hint: 練習問題 3.13(b)から $E[\hat{u}_i^2]=\sigma^2(1-h_{ii})$となる。なお、$h_{ii}$はレバレッジである。

(b) Stata では、不均一分散に対して頑健な分散の推定量として、次の推定量$\hat{V}^{HC1}$を用いている。

$$\hat{V}^{HC1}=\frac{n}{n-2}\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2\hat{u}_i^2}{\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right)^2}$$

この推定量$\hat{V}^{HC1}$の是非を述べよ。Hint: 練習問題 3.13(c)から$\sum_{i=1}^{n}h_{ii}=2$となる(よって、平均は$2/n$)。

(c) 標準化残差$\bar{u}_i=(1-h_{ii})^{-1/2}\hat{u}_i$を用いた次の推定量$\hat{V}^{HC2}$を考えよう。

$$\hat{V}^{HC2}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2\bar{u}_i^2}{\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right)^2}$$

この推定量V^{HC2}の是非を述べよ。Hint: 標準化残差は練習問題 3.13(d)参照

(d) これらの推定量のうち、どれが望ましいといえるか。

11. ★　レバレッジが大きくなるケースを考えよう。回帰モデルは$Y_i=\alpha+\beta D_i+u_i$とし、D_iはダミー変数とする。なお、$\sum_{i=1}^{n}D_i=1$、$E[u_i^2]=\sigma^2$、$E[u_iu_j]=0$とする。$\sum_{i=1}^{n}D_i=1$とは、あるiでだけ$D_i=1$となるが、それ以外はすべて 0 となることを意味する(『入門 実践する統計学』(巻末参考文献[5])12.4.1 節の一時的ダミーに該当する)。

(a) 説明変数の偏差 2 乗和は、次のようになることを示せ。

$$\sum_{i=1}^{n}(D_i-\bar{D})^2=\frac{n-1}{n}$$

(b) $\hat{\beta}$の分散は次のようになることを示せ。

$$V(\hat{\beta})=\sigma^2\frac{n}{n-1}$$

(c) 推定量$\hat{V}^{HC1}$の期待値は次のようになることを示せ。

$$E[\hat{V}^{HC1}]=E\left[\frac{n}{n-2}\frac{\sum_{i=1}^{n}(D_i-\bar{D})^2\hat{u}_i^2}{\left(\sum_{i=1}^{n}(D_i-\bar{D})^2\right)^2}\right]=\frac{V(\hat{\beta})}{n-1}$$

Hint: $D_i=1$なら$h_{ii}=1$、$D_i=0$なら$h_{ii}=1/(n-1)$となる。

10 章の練習問題

10. ★(8)式のもとで$V(u_1)=\sigma^2/(1-\rho^2)$となることを証明せよ。

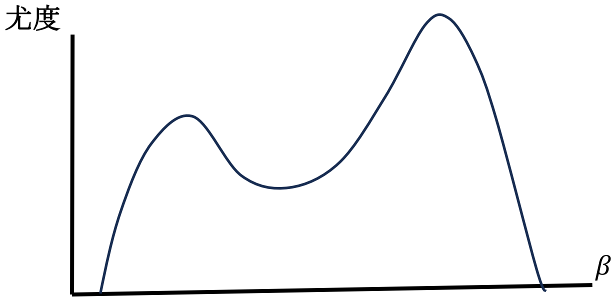
11 章の練習問題

12. ★固定効果モデル$Y_{i,t}=\alpha_i+\beta X_{i,t}+u_{i,t}$を考える。

(a) $Y_{i,t}-\bar{Y}_i = \beta(X_{i,t}-\bar{X}_i)+(u_{i,t}-\bar{u}_i)$を示せ。$\bar{Y}_i$、$\bar{X}_i$、$\bar{u}_i$は時間平均である($\bar{Y}_i=\frac{1}{T}\sum_{s=1}^{T}Y_{i,s}$、$\bar{X}_i=\frac{1}{T}\sum_{s=1}^{T}X_{i,s}$、$\bar{u}_i=\frac{1}{T}\sum_{s=1}^{T}u_{i,s}$)。なお、この式の OLS 推定はダミー変数を用いた固定効果推定と等しい。
(b) 固定効果推定量が一致性を持つには、強外生性の仮定が必要となる。**強外生性**は、全時点t、sに対し$Cov(X_{i,t},\ u_{i,s})=0$を意味する。なぜ強外生性の仮定が必要かを述べよ。Hint: (a)の推定における外生性

<u>12 章の練習問題</u>

11. 尤度関数は、パラメータβにだけ依存しているとする。図では、横軸がβ、縦軸が対応する尤度としている。このとき、ニュートン法で尤度を最大にするβを見つけたい。どのような問題が生じるか述べよ。



<u>13 章の練習問題</u>

11.★ 13.6.2 節では、操作変数の条件②により、$\hat{X}_i=\hat{\gamma}_0+\hat{\gamma}_1Z_i$は$u_i$と無相関とした。しかし、この説明は$n$が小さいとき不正確である。これはなぜか、どうすれば問題に対処できるか。Hint: サンプル分割

12.★　内生性の有無を検証する方法を述べよ。Hint: ハウスマン検定

13.★　パネル分析においてモデル$Y_{i,t}=\alpha+\beta X_{i,t}+\theta W_i+Z_i+u_{i,t}$を考える。個別要因$Z_i$を考慮しながら、係数($\beta$、$\theta$)を推定したい。誤差項は$Z_i+u_{i,t}$である。なお、$X_{i,t}$と$W_i$は$u_{i,t}$と無相関、$Z_i$は$X_{i,t}$と相関するが$W_i$とは無相関とする($W_i$は時間を通じて一定の外生変数、$X_{i,t}$は内生変数)。係数の推定方法を述べよ。Hint: 操作変数

14.★　モデルは$Y_i=\alpha+\beta X_i+u_i$であり、操作変数はZ_iとする。
(a)2SLS 推定量は、次式として表せることを示せ(練習問題 7 参照)。

$$\hat{\beta}_{2SLS}=\beta+\frac{\sum_{i=1}^{n}(Z_i-\bar{Z})u_i}{\sum_{i=1}^{n}(Z_i-\bar{Z})(X_i-\bar{X})}$$

(b)2SLS 推定量は一致性を満たすことを示せ。Hint: nが大きくなると、標本共分散は共分散

に収束する。

15.★異質性を考慮したモデル$Y_i = \alpha + \beta_i X_i + u_i$を考える。係数$\beta_i$は個人$i$によって異なり、確率的に決まるとしよう。

(a) X_iはランダムに決定され、β_i、u_iとは独立とする(X_iは外生変数)。nが大きいとき、OLS 推定量$\hat{\beta}$は$E[\beta_i]$に収束することを示せ。また、$E[\beta_i]$は何を意味するか。Hint: nが大きくなると、標本分散と標本共分散は、それぞれ分散と共分散に収束する。X_iはβ_iとは独立であり、$Cov(\alpha + \beta_i X_i + u_i, X_i) = E[\beta_i]Var(X_i)$

(b) X_iは内生変数であり、$X_i = \gamma_0 + \gamma_i Z_i + e_i$とする($Z_i$は操作変数、$E[\gamma_i] \neq 0$)。反応度$\gamma_i$は個人によって異なり、確率的に決まる。$Z_i$はランダムに決定され、$\beta_i$、$\gamma_i$、$u_i$、$e_i$とは独立とする。$n$が大きくなると、$\hat{\beta}_{2SLS}$は$E[\gamma_i \beta_i]/E[\gamma_i]$に収束することを示せ。また、$E[\gamma_i \beta_i]/E[\gamma_i]$は何を意味するか。Hint: nが大きくなると、$\hat{\beta}_{2SLS}$は$Cov(Z_i, Y_i)/Cov(Z_i, X_i)$に収束する(練習問題 7 参照)。また、$Cov(Z_i, X_i) = E[\gamma_i]Var(Z_i)$

(c) どのようなとき$E[\gamma_i \beta_i]/E[\gamma_i] = E[\beta_i]$は成立するか。

16.★ パネルデータを用いて次のモデルを推定したい。

$$Y_{i,t} = \alpha + \beta Y_{i,t-1} + Z_i + u_{i,t}$$

ただし、誤差項$u_{i,t}$は個別要因Z_iとは独立であり、$u_{i,t}$に系列相関はないとする。

(a) 説明変数$Y_{i,t-1}$は強外生性を満たすか(11 章の練習問題 12 参照)。

(b) $Y_{i,t} - Y_{i,t-1} = \beta(Y_{i,t-1} - Y_{i,t-2}) + (u_{i,t} - u_{i,t-1})$を示せ。

(c) 係数βの推定方法を述べよ。Hint: 操作変数$Y_{i,t-2}$

15 章の練習問題

8. 20 年分の月次データを用いて 5 変量 VAR(定数項あり)を推定したい。季節性を考慮するため、ラグ次数$p = 12$と設定する(12 か月分のラグを使う)。この推定の是非を述べよ。Hint: サンプルサイズとパラメータ数

9.★ 3 変量の構造 VAR を考える。排除制約($b_{10} = c_{10} = c_{20} = 0$)のもとで、誘導型の誤差項は構造ショックを用いて、次のように表現できることを示せ。

$$u_{Yt} = \varepsilon_{Yt}$$
$$u_{Xt} = a_{20}\varepsilon_{Yt} + \varepsilon_{Xt}$$
$$u_{Zt} = a_{30}\varepsilon_{Yt} + b_{30}\varepsilon_{Xt} + \varepsilon_{Zt}$$

なお、これは逆も正しい。つまり、誤差項と構造ショックに上記の関係があれば、3 変量の構造 VAR において排除制約($b_{10} = c_{10} = c_{20} = 0$)が成立する。

10.★ 金融政策の効果を分析したい。四半期データを用いて、3 変量(実質 GDP 成長率、インフレ率、短期金利)の構造 VAR モデルを推定したい。排除制約を用いる場合、3 変数の

順番として何が適当かを理由とともに述べよ。

11.★ 原油の需給ショックが原油価格に与える影響を分析したい[4]。月次データを用いて、3 変量(世界原油生産量の変化率(%)、世界経済活動指数、実質原油価格)の構造 VAR を推定する。排除制約を用いる場合、3 変数の順番として何が適当かを理由とともに述べよ。なお、世界経済活動指数は乾貨物運賃から作成された指標であり、世界経済活動から生じる原油需要を捉えるための変数である[5]。

[4] この練習問題は以下の論文に基づいています。Kilian, L. (2009) “Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market,” *American Economic Review* 99(3), 1053-1069.

[5] 世界経済活動が活発になると海上運賃が上昇することから、世界経済活動指標は経済活動の変動を捉える指標として用いられる。FRED では、世界経済活動指標を Index of Global Real Economic Activity として公開しています(FRED は 1.3.3 節参照)。