第 9 章: 定式化とデータの問題

Jeffrey Wooldridge (2016).
Introductory Econometrics: A Modern Approach
Seventh Edition. Cengage Learning.

2026-03-06

準備

必要なパッケージの読み込み

- wooldridge パッケージの読み込み

```
library(wooldridge)
```

- lmtest パッケージの読み込み

```
library(lmtest)
```

9-1 関数形式の誤設定 (Functional Form Misspecification)

定式化の誤りとは

- 説明変数の非線形性（2 乗項や交差項）を見落としている場合。
- 重要な変数が対数形式であるべきなのに、レベル形式で推定している場合。
- **結果:** 係数の推定値にバイアス（不偏性の欠如）が生じる。

RESET 検定 (Ramsey RESET Test)

- モデルに説明変数の高次項（2 乗、3 乗など）が含まれるべきかを検定する。
- 手順：
  1. 元のモデルを推定し、予測値 $\hat{y}$ を得る。
  2. $\hat{y}^2, \hat{y}^3$ を説明変数に加えた補助回帰を行う。
  3. それらの追加変数の係数がすべて 0 であるという帰無仮説を F 検定する。

RESET 検定の例 (hprice1)

```
data(hprice1)
res <- lm(price ~ lotsize + sqrft + bdrms, data = hprice1)

# lmtest パッケージの resettest を使用
resettest(res, power = 2:3, type = "fitted")
```

```
##
##  RESET test
##
## data:  res
## RESET = 4.6682, df1 = 2, df2 = 82, p-value = 0.01202
```

9-2 観測不可能な変数の代用 (Proxy Variables)

代理変数 (Proxy Variables)

- モデルに含めるべき重要な変数がデータに存在しない場合。
- 例：能力 $(ability)$ が賃金に与える影響を知りたいが、能力は測定不能。
- 代わりに、IQ テストのスコア (IQ) などを**代理変数**として用いる。
- 代理変数が「誤差」を伴っていても、何もしない（脱落変数バイアス）よりはマシな場合が多い。

9-3 ランダム傾きモデル (Models with Random Slopes)

ランダム傾きモデルの考え方

- 個体ごとに傾きが異なるモデル：$y_i = a_i + b_i x_i$
- 平均傾きを $b = E(b_i)$ とおくと、b は平均偏効果（APE）として解釈できる。
- 傾きの異質性は、誤差項に x_i との交差項が入る形（$u_i = c_i + d_i x_i$）で表現できる。

OLS で平均効果を推定できる条件

- a_i と b_i が x_i と平均独立であれば、OLS は平均傾き b を推定できる。
- 一方で、この設定では一般に不均一分散が生じるため、頑健標準誤差の利用が重要。
- 傾きの異質性を観測可能変数で表す場合は、交差項モデルとして実装できる。

9-4 測定誤差 (Measurement Error)

従属変数の測定誤差

- $y^* = y + e$ (y^* が観測値、y が真の値、e が誤差)
- 測定誤差 e が説明変数 x と相関していなければ、OLS 係数は依然として不偏。
- ただし、残差の分散が大きくなり、標準誤差が増大する。

独立変数の測定誤差

- $x^* = x + e$
- **古典的測定誤差 (Classical Errors-in-Variables):** 誤差 e が真の値 x と無相関である場合。
- **減衰バイアス (Attenuation Bias):** 係数 β が 0 の方向に過小推定される。
- $plim(\hat{\beta}_1) = \beta_1 \left(\frac{\sigma_x^2}{\sigma_x^2 + \sigma_e^2} \right)$

9-5 欠損データと非ランダム標本

データの欠損

- **ランダムな欠損:** 推定値にバイアスは生じないが、サンプルサイズが減る。
- **説明変数に基づく非ランダムな抽出:** x の範囲を限定して抽出しても（例：所得が一定以上の人のみ）、$E(u|x) = 0$ が維持されれば OLS は不偏。
- **従属変数に基づく非ランダムな抽出:** y の値によってサンプルが選ばれる場合（例：賃金が高い人のみが回答）、バイアスが生じる。

外れ値 (Outliers)

- 少数の異常なデータが最小二乗法の結果を大きく左右することがある。
- 対数変換を行うことで、外れ値の影響を和らげることができる場合が多い。

まとめ

- RESET 検定は、モデルの非線形性が適切に捉えられているかのチェックに有効。
- 代理変数は、観測不能な変数を近似するために用いられる。
- ランダム傾きモデルでは、平均独立の条件の下で OLS は平均偏効果（APE）を推定できる。
- 説明変数の測定誤差は、係数を 0 に近づけるバイアス（減衰バイアス）を生む。
- サンプルの抽出が y に依存している場合、バイアスが生じるため注意が必要。