2025/07/18

藪友良

ロバスト標準誤差

均一分散は、ガウス=マルコフ定理の前提条件の 1 つであり、数式展開も簡単にしてくれる仮定となります。しかし、均一分散は、現実には成立しない仮定であり、誤差項が不均一分散が現実的仮定になります。本稿では、不均一分散を前提として、不均一分散に頑健な標準誤差(ロバスト標準誤差)を説明します。本稿は、藪友良『入門 実践する計量経済学』(2023 年、東洋経済新報社)の補足資料です。

1. OLS 推定量の分散

ここでは単回帰モデルを考えます。

$$Y_i = \alpha + \beta X_i + u_i$$

横断面データであるため、各変数の右下添字はiとしています(サンプルサイズはn)。ここで、誤差項は不均一分散とします。

$$V(u_i) = E[u_i^2] = \sigma_i^2$$

誤差項の分散は、σ^2ではなく、σ_i^2としていることに注意してください。

このとき、OLS 推定量$\hat{\beta}$の分散$\sigma_{\hat{\beta}}^2$は、均一分散のもとで導出された

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

ではなく、不均一分散を考慮することで得られた次式です(証明は 9 章補足参照)。

$$\sigma_{\hat{\beta}}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sigma_i^2}{\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2}$$

次節では、分散$\sigma_{\hat{\beta}}^2$を推定する方法を説明します。

2. $\boldsymbol{\sigma_{\hat{\beta}}^2}$の推定方法

OLS 推定量$\hat{\beta}$の分散$\sigma_{\hat{\beta}}^2$は、次のようにも表現できます。$q_i = (X_i - \bar{X})^2 \sigma_i^2$と定義すると、

$$\sigma_{\hat{\beta}}^2 = \frac{\sum_{i=1}^{n} q_i}{\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2}$$

$$= \frac{n\bar{q}}{\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2}$$

ただし、$\bar{q}$は平均$\bar{q} = \frac{1}{n}\sum_{i=1}^{n} q_i$と定義しています。

誤差項の分散は$\sigma_i^2 = E[u_i^2]$であるため、分散σ_i^2の推定量は、誤差項の 2 乗であるu_i^2を用いることが自然でしょう。しかし、誤差項u_iは観察できないため、誤差項u_iの推定量である残差$\hat{u}_i$に置き換えることで、q_iの不偏推定量$\hat{q}_i$が得られます。

$$\hat{q}_i = (X_i - \bar{X})^2 \hat{u}_i^2$$

これは分散σ_i^2の推定量として用いることができます。しかし、$\hat{q}_i$は 1 つの観測値から推定

しており、その推定精度はかなり低くなります。

H.ホワイト(Halbert White)は、個々のq_iではなく、その平均$\bar{q}$なら安定的に推定できるとしました。つまり、$\hat{q}_i$は推定誤差が大きい一方、サンプルサイズnが十分大きいなら、平均$\hat{\bar{q}}$は推定誤差が打ち消しあうことで推定誤差が小さくなります[1]。

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

これを先の分散$\sigma_{\hat{\beta}}^2$の式に代入すると、**不均一分散に頑健な分散の推定量**になります(教科書 9.3 節を参照してください)[2]。

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

そして、分散の推定量$s_{\hat{\beta}}^2$の平方根である標準誤差は、**不均一分散に対して頑健な標準誤差**(heteroskedasticity-consistent standard error、HC 標準誤差)、**ロバスト標準誤差**(robust standard error)、開発者の名前にちなんで**ホワイト標準誤差**などと呼ばれます。

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

ロバスト標準誤差は、通常の標準誤差より大きくなることも、小さくなることもあります(小さくなる場合の例として、9 章練習問題 10、11 を参照)。統計ソフトでは、OLS 推定を実行すると通常の標準誤差が算出されてしまいます。しかし、不均一分散が自然な前提であるため、統計ソフトのオプションを変更するなどし、ロバスト標準誤差を算出するようにしてください。たとえば、stata では次のように最後に r を入れるとロバスト標準誤差が表示されます。

```
reg Y X, r
```

[1] 彼の論文は以下を参照してください。White, Halbert. "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica: journal of the Econometric Society* (1980): 817-838. この論文は経済系の論文の中で最も引用された論文の 1 つと言われています。Google Scholar によると、2025 年 7 月 17 日時点で引用関数は 3,6218 回です。ホワイトは 2012 年に亡くなっていますが、ご存命であればノーベル経済学賞を取得する可能性が高い経済学者の 1 人であったといえます。また、彼はコンサルティング会社 Bates White の共同創業者としても有名です。この会社は、経済分析を用いて、さまざまな訴訟案件に専門的知見を提供しています。訴訟案件としては、たとえば、契約紛争、損害賠償額の推定、エネルギー資産をめぐる紛争などがあります。余談ですが、私の知人も PhD 取得後に、この会社で専門的知見を活かした仕事をしています。

[2] $\sqrt{n}\hat{\beta}$の分散は次のようになります。

$$\frac{\bar{q}}{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2}$$

これは次のように推定できます。

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

サンプルサイズnが大きくなると、分子は$\bar{q}$に収束し、$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$は$X$の分散に収束します。