2018/11/21

藪友良

最小 2 乗推定量の意味を理解しよう

ここで重回帰モデルを考えましょう。

$$Y=\beta_0X_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_KX_K+u$$

ただし、X_0 は常に 1 となる変数とします(つまり、β_0 は定数項)。最小 2 乗法では、未知のパラメータ（$\beta_0, \beta_1,\ldots, \beta_K$）は残差 2 乗和を最小にするように決定されます。こうした複雑なプロセスを経て推定量が得られるため、最小 2 乗推定量が何を表しているかという直観的理解が失われがちです。

学部の授業ではあまり扱いませんが、大学院で必ず扱う定理に **Frisch–Waugh–Lovell (FWL)定理**があります。これは**残差回帰**とも呼ばれます。FWL 定理は、最小 2 乗推定量の計算を簡単にしてくれるだけでなく、最小 2 乗推定量に新しい解釈も与えてくれます。また、FWL 定理は、変数間の関係を捉える指標である**偏相関係数**にも深く関連しています。

FWL 定理によれば、X_ℓ の係数 β_ℓ の最小 2 乗推定量$\hat{\beta}_\ell$は、以下 3 段階の手続きを通じて求めることができます。

段階 1: Y を X_ℓ 以外の説明変数で回帰し、残差を e_Y とする。
段階 2: X_ℓ を X_ℓ 以外の説明変数で回帰し、残差を e_X とする。
段階 3: e_Y を e_X で回帰し係数を$\hat{\theta}$とします（このとき、$\hat{\theta}$と$\hat{\beta}_\ell$は一致する）。

段階 1 では、Y を X_ℓ 以外の全ての説明変数で回帰し、その残差を e_Y とするわけですから、e_Y は「Y から X_ℓ 以外の要因を取り除いた変数」となります。同様に、段階 2 では、X_ℓ を X_ℓ 以外の全ての説明変数で回帰し、その残差を e_X としますから、e_X は「X_ℓ から X_ℓ 以外の要因を取り除いた変数」なります。したがって、段階 3 において、e_Y を e_X で回帰することで、（他の変数を一定とした上で）X_ℓ の Y への効果である β_ℓ を推定できるのです。

たとえば、X_1 の係数 β_1 を推定したいとしましょう。段階 1 では、Y を X_1 以外の変数（つまり、X_0、X_2、...、X_K）で回帰して残差 e_Y を求めます（X_0 を含めるとは定数項を含めて推定するということ）。段階 2 では、X_1 を X_1 以外の変数（つまり、X_0、X_2、...、X_K）で回帰して残差 e_X を求めます。段階 3 では、e_Y を e_X で回帰することで、e_X の係数$\hat{\theta}$を推定できます。FWL 定理により、こうして得られた$\hat{\theta}$は、重回帰モデルの最小 2 乗法により得られた$\hat{\beta}_1$と一致します。

[証明] FWL 定理の証明は行列の知識が必要なため、ここでは説明変数が 2 個（K=2)、かつ定数項がないケースにおける証明を考えましょう。モデルは、

$$Y=\beta_1X_1+\beta_2X_2+u$$

とします。ここで FWL 定理を使って β_1 を推定してみましょう。段階 1 は、Y を X_2 で回帰します。これは $Y=\hat{\gamma}_YX_2+e_Y$ となります。ただし、

$$\hat{\gamma}_Y=\sum Y_iX_{2i}\,/\sum X_{2i}^2$$

となり、また残差の性質から、説明変数と残差の積和は 0（$\sum e_{Yi} X_{2i} = 0$）となります（定数項なしの単回帰分析については、12 章練習問題 8 を参照してください）。段階 2 では、X_1 を X_2 で回帰します。これは $X_1 = \hat{\gamma}_X X_2 + e_X$ となります。ただし、

$$\hat{\gamma}_X = \sum X_{1i} X_{2i} / \sum X_{2i}^2$$

となり、また残差の性質から、説明変数と残差の積和は 0（$\sum e_{Xi} X_{2i} = 0$）となります。段階 3 では、e_Y を e_X だけで回帰します。このとき、e_X の係数は

$$\hat{\theta} = \frac{\sum e_{Yi} e_{Xi}}{\sum e_{Xi}^2} = \frac{\sum (Y_i - \hat{\gamma}_Y X_{2i})(X_{1i} - \hat{\gamma}_X X_{2i})}{\sum (X_{1i} - \hat{\gamma}_X X_{2i})(X_{1i} - \hat{\gamma}_X X_{2i})}$$

となります。上式を展開すると、

$$\frac{\sum (Y_i - \hat{\gamma}_Y X_{2i}) X_{1i} - \hat{\gamma}_X \sum (Y_i - \hat{\gamma}_Y X_{2i}) X_{2i}}{\sum (X_{1i} - \hat{\gamma}_X X_{2i}) X_{1i} \quad - \hat{\gamma}_X \sum (X_{1i} - \hat{\gamma}_X X_{2i}) X_{2i}}$$

となります。段階 1 の残差の性質は$\sum (Y_i - \hat{\gamma}_Y X_{2i}) X_{2i} = \sum e_{Yi} X_{2i} = 0$ですし、段階 2 の残差の性質は$\sum (X_{1i} - \hat{\gamma}_X X_{2i}) X_{2i} = \sum e_{Xi} X_{2i} = 0$ですから、これらの関係を代入することで、上式はさらに単純化できます。

$$\frac{\sum (Y_i - \hat{\gamma}_Y X_{2i}) X_{1i}}{\sum (X_{1i} - \hat{\gamma}_X X_{2i}) X_{1i}}$$

次に、通常の最小 2 乗法で β_1 を推定しましょう。残差 2 乗和は

$$\sum (Y_i - \tilde{\beta}_1 X_{1i} - \tilde{\beta}_2 X_{2i})^2$$

となります。ここで$\tilde{\beta}_1$と$\tilde{\beta}_2$で偏微分してから 0 と置くと、以下の正規方程式が得られます。

$$\sum \left(Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}\right) X_{1i} = 0$$
$$\sum \left(Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}\right) X_{2i} = 0$$

上式を満たす$\tilde{\beta}_1$と$\tilde{\beta}_2$は最小 2 乗推定量であるため、^を付けて表しました。2 番目の式を$\hat{\beta}_2$について解くと、

$$\hat{\beta}_2 = \frac{\sum X_{2i} Y_i}{\sum X_{2i}^2} - \hat{\beta}_1 \frac{\sum X_{1i} X_{2i}}{\sum X_{2i}^2} = \hat{\gamma}_Y - \hat{\beta}_1 \hat{\gamma}_X$$

となります（$\hat{\gamma}_Y = \sum Y_i X_{2i} / \sum X_{2i}^2$、$\hat{\gamma}_X = \sum X_{1i} X_{2i} / \sum X_{2i}^2$に注意）。これを正規方程式の 1 番目の式に代入すると、

$$\sum \left(Y_i - \hat{\beta}_1 X_{1i} - (\hat{\gamma}_Y - \hat{\beta}_1 \hat{\gamma}_X) X_{2i}\right) X_{1i} = 0$$

となり、これを$\hat{\beta}_1$について解くと

$$\hat{\beta}_1 = \frac{\sum Y_i X_{1i} - \hat{\gamma}_Y \sum X_{1i} X_{2i}}{\sum X_{1i}^2 - \hat{\gamma}_X \sum X_{1i} X_{2i}} = \frac{\sum (Y_i - \hat{\gamma}_Y X_{2i}) X_{1i}}{\sum (X_{1i} - \hat{\gamma}_X X_{2i}) X_{1i}}$$

となります。これは先に求めた$\hat{\theta}$に他なりません。 [終]

例 1（季節調整）　12.4.3 節では季節性を考慮するため、季節ダミーを導入しました。具体的には、四半期データであれば、季節ダミーは以下と定義されます。

$$Q1_t=\begin{cases}1 & \text{第 1 四半期}\\ 0 & \text{その他}\end{cases}、\quad Q2_t=\begin{cases}1 & \text{第 2 四半期}\\ 0 & \text{その他}\end{cases}、\quad Q3_t=\begin{cases}1 & \text{第 3 四半期}\\ 0 & \text{その他}\end{cases}$$

$Q1_t$ は、t が第 1 四半期であれば 1、そうでないときは 0 となる変数です。このとき、回帰式は以下となります（ただし、X_0 は常に 1 となる変数です）。

$$Y=\alpha X_0+\beta X+\gamma_1 Q1+\gamma_2 Q2+\gamma_3 Q3+u$$

ここで β の最小 2 乗推定量は、FWL 定理によって、次のように求めることができます。まず、Y を X_0、$Q1$、$Q2$、$Q3$ で回帰し、その残差を e_Y します。次に、X を X_0、$Q1$、$Q2$、$Q3$ で回帰し、その残差を e_X とします。そして e_Y を e_X で回帰することで $\hat{\beta}$ が得られるわけです。したがって、季節性が疑われる場合、回帰式に季節ダミーを入れて推定してもよいですし、季節ダミーで回帰して得られた残差（季節調整済み系列）を作成してから、これらの系列を用いて回帰分析をしてもかまいません。

例 2（標本偏相関係数）　標本相関係数(sample partial correlation)は 2 変数(Y、X)間の相関の程度を測るための指標でした。しかし、両変数に影響を与える第 3 の変数(Z)があるとき、単なる Y と X との標本相関係数では、両者の関係を測ることはできません。例えば、火事の被害額 Y と派遣された消防車数 X の関係を調べると、消防車の数 Z が多いほど被害額も多くなっています。しかし、これは派遣する消防車数を増やすと火事の被害額が大きくなることを意味しません。むしろ火事の規模が、火事の被害額と消防車の数に影響を与えているため、両変数には正の相関がみられるのです。正しく両者の関係をみるためには、火事の規模をコントロールする必要があります。標本偏相関係数とは、第 3 の変数 Z の影響をコントロールしたうえで、X と Y との相関を測る指標となります。具体的な計算方法は、FWL 定理と同じとなります。まず、Y を定数項と Z で回帰し、Z の影響を取り除いた残差系列 e_y を求めます。同様に、X を定数項と Z で回帰して、Z の影響を除いた残差系列 e_x を求めます。そして残差系列(e_y、e_x)間の標本相関係数を求めます。これが標本偏相関係数となり、やはり-1 から+1 で定義されます。読者の中には、変数間の関係をより視覚的に確認したいと思っている方もいるでしょう。その場合は、こうして求めた残差系列の散布図を作ることをおすすめします。

練習問題 1:　FWL 定理の修正版を紹介しましょう。

Step 1: X_ℓ を X_ℓ 以外の説明変数で回帰し、残差を e_X とする。
Step 2:Y を e_X で回帰し係数を $\hat{\theta}$ とします（このとき、$\hat{\theta}$ と $\hat{\beta}_\ell$ は一致する）。

つまり、Y については、他の説明変数の影響を除去しないでも、FWL 定理と同じ結果が得られる事がわかります。以下では、この定理が正しいことを証明してください。ただし、証

明では、説明変数が 2 個（K=2）、かつ定数項がないケースを考えなさい。

[答え]　ここで、$\hat{\delta} = \sum e_{Xi} Y_i / \sum e_{Xi}^2$です。このとき、FWL 定理から$\hat{\theta}$と$\hat{\beta}_1$は同じとなるので、ここでは$\hat{\delta}$と$\hat{\theta}$が一致することを示します。

$$\hat{\delta} = \frac{\sum e_{Xi} Y_i}{\sum e_{Xi}^2} = \frac{\sum (X_{1i} - \hat{\gamma}_X X_{2i}) Y_i}{\sum e_{Xi}^2}$$

残差の性質$\sum (X_{1i} - \hat{\gamma}_X X_{2i}) X_{2i} = \sum e_{Xi} X_{2i} = 0$に注意すると、上式は

$$\frac{\sum (X_{1i} - \hat{\gamma}_X X_{2i}) Y_i - \hat{\gamma}_Y \sum (X_{1i} - \hat{\gamma}_X X_{2i}) X_{2i}}{\sum e_{Xi}^2}$$

として書き換えることができます。この式を更に整理すると、以下となります。

$$\frac{\sum (X_{1i} - \hat{\gamma}_X X_{2i})(Y_i - \hat{\gamma}_Y X_{2i})}{\sum e_{Xi}^2} = \frac{\sum e_{Yi} e_{Xi}}{\sum e_{Xi}^2} = \hat{\theta}$$