2010年度 エコノメトリックスII & 上級エコノメトリックスII 第4回講義メモ(訂正版)

2010年10月22日

2.2.3 FWL 定理とその応用

本節では、従来あまり触れられることのなかったFrish=Waugh=Lovell 定理を扱う。この定理は、線形回帰モデルにおいては、結構便利な定理である。

まず、単回帰モデルを例にとって、FWL 定理の意味を見てゆくことにしたい。線形単回帰モデルを

$$\boldsymbol{y} = \mathbf{1}\beta_1 + \boldsymbol{x}\beta_2 + \boldsymbol{u} \tag{2.2.3.a}$$

としたとき、説明変数 (独立変数) ベクトル $\boldsymbol{x}$ を算術平均 $\bar{x}$ からの偏差系列に変換することを考えよう。

$$\boldsymbol{z} \equiv \boldsymbol{x} - \mathbf{1}\bar{x} \qquad where \quad \bar{x} \equiv \frac{1}{n}\mathbf{1}'\boldsymbol{x} = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{x}$$

とすると、(2.2.3.a) 式は

$$\boldsymbol{y} = \mathbf{1}\alpha_1 + \boldsymbol{z}\alpha_2 + \boldsymbol{u} \tag{2.2.3.b}$$

で表され

$$\begin{cases} \beta_2 &= \alpha_2 \\ \beta_1 &= \alpha_1 - \alpha_2\bar{x} \end{cases}$$

になっている。

ゆえに

$$\boldsymbol{z} = (\boldsymbol{I} - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1})\boldsymbol{x} = \boldsymbol{M}_{\mathbf{1}}\boldsymbol{x} \tag{2.2.3.c}$$

であり、また

$$\mathbf{1}'\boldsymbol{z} = \mathbf{1}'\boldsymbol{M}_{\mathbf{1}}\boldsymbol{x} = \mathbf{0}'\boldsymbol{x} = 0$$

が成立する。

$\mathbf{1}'\boldsymbol{z} = 0$ より $\boldsymbol{Z} = (\mathbf{1}, \boldsymbol{z})$ としたとき

$$\begin{aligned} \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y} &= \begin{pmatrix} \mathbf{1}'\mathbf{1} & 0 \\ 0 & \boldsymbol{z}'\boldsymbol{z} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{1}' \\ \boldsymbol{z}' \end{pmatrix} \boldsymbol{y} \\ &= \begin{pmatrix} (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{y} \\ (\boldsymbol{z}'\boldsymbol{z})^{-1}\boldsymbol{z}'\boldsymbol{y} \end{pmatrix} \end{aligned}$$

となり、$\hat{\alpha}_1$ は $\mathbf{1}$ のみに回帰したときの係数、$\hat{\alpha}_2$ は $\boldsymbol{z}$ のみに回帰したときの係数と同一となる。

以上の結果を図を用いて説明しよう。(2.2.3.c) 式より、$\boldsymbol{z}$ は $\boldsymbol{x}$ を $\mathbf{1}$ の直交補空間へ射影したベクトルであることが分かる。図 2.2.3-1 からわかるように、$\mathbf{1}, \boldsymbol{x}$ からなる線形部分空間は、$\mathbf{1}, \boldsymbol{z}$ からなる線形部分空間と一致する。ゆえに、

$$\hat{\boldsymbol{y}} = \mathbf{1}\hat{\beta}_1 + \boldsymbol{x}\hat{\beta}_2 = \mathbf{1}\hat{\alpha}_1 + \boldsymbol{z}\hat{\alpha}_2$$

が成立し、図より $\hat{\beta}_2 = \hat{\alpha}_2$ となることは明らか。

では、次に一般化して考えることにしよう。$\boldsymbol{X}$ および $\boldsymbol{\beta}$ を

$$\begin{aligned}
\boldsymbol{X} &= \Big[\underbrace{\boldsymbol{X}_1}_{n\times k_1} \underbrace{\boldsymbol{X}_2}_{n\times k_2} \Big] \\
\boldsymbol{\beta} &= \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}
\end{aligned}$$

のように分割すると、線形回帰モデルは

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u} \tag{2.2.3.d}$$

のように表現することができる。ここで、一般には $\boldsymbol{X}_1'\boldsymbol{X}_2 \neq \boldsymbol{O}$ であることに注意。 .

いま、

$$\boldsymbol{P}_i \equiv \boldsymbol{P}_{\boldsymbol{X}_i} = \boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i', \quad \boldsymbol{M}_i = \boldsymbol{I} - \boldsymbol{P}_i, \quad (i = 1, 2)$$

とする。このとき、$\boldsymbol{I} = \boldsymbol{P}_1 + \boldsymbol{M}_1$ であるから、(2.2.3.d) 式を

$$\begin{aligned}
\boldsymbol{y} &= \boldsymbol{X}_1\boldsymbol{\beta}_1 + (\boldsymbol{P}_1 + \boldsymbol{M}_1)\boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u} \\
&= \boldsymbol{X}_1(\boldsymbol{\beta}_1 + (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\boldsymbol{\beta}_2) + \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u} \\
&= \boldsymbol{X}_1\boldsymbol{\alpha}_1 + \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\alpha}_2 + \boldsymbol{u} \qquad (1)
\end{aligned}$$

と書きかえることができる。ここで、$\boldsymbol{\alpha}_1 = \boldsymbol{\beta}_1 + (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}'\boldsymbol{X}_2\boldsymbol{\beta}_2$、　$\alpha_2 = \boldsymbol{\beta}_2$ である。

ところで、$\boldsymbol{X}_1'(\boldsymbol{M}_1\boldsymbol{X}_2) = \boldsymbol{O}$ であるから、

$$\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\alpha}_2 + \boldsymbol{v} \tag{2}$$

としたとき、(1) 式と (2) 式における $\boldsymbol{\alpha}_2$ の最小 2 乗推定量 $\hat{\boldsymbol{\alpha}}_2$ は同一になる。

これは以下のようにして確認することができる。

$$\boldsymbol{y} = \boldsymbol{Z}_1\boldsymbol{\delta}_1 + \boldsymbol{Z}_2\boldsymbol{\delta}_2 + \boldsymbol{u} = (\ \boldsymbol{Z}_1\boldsymbol{Z}_2\ ) \begin{pmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \end{pmatrix} + \boldsymbol{u}$$

であるとき、$\boldsymbol{\delta}_i$ の OLS 推定量 $\hat{\boldsymbol{\delta}}_i$, $(i = 1, 2)$ は、

$$\begin{pmatrix} \hat{\boldsymbol{\delta}}_1 \\ \hat{\boldsymbol{\delta}}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{Z}_1'\boldsymbol{Z}_1 & \boldsymbol{Z}_1'\boldsymbol{Z}_2 \\ \boldsymbol{Z}_2'\boldsymbol{Z}_1 & \boldsymbol{Z}_2'\boldsymbol{Z}_2 \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{Z}_1' \\ \boldsymbol{Z}_2' \end{pmatrix} \boldsymbol{y}$$

で表されるが、もし $\boldsymbol{Z}_1'\boldsymbol{Z}_2 = \boldsymbol{O}$ であるならば

$$\begin{pmatrix} \hat{\boldsymbol{\delta}}_1 \\ \hat{\boldsymbol{\delta}}_2 \end{pmatrix} = \begin{pmatrix} (\boldsymbol{Z}_1'\boldsymbol{Z}_1)^{-1}\boldsymbol{Z}_1'\boldsymbol{y} \\ (\boldsymbol{Z}_2'\boldsymbol{Z}_2)^{-1}\boldsymbol{Z}_2'\boldsymbol{y} \end{pmatrix}$$

となるので、この結果を用いればよい。

ただし、残差ベクトルは異なる。なぜなら、(1) 式における残差が

$$\hat{\boldsymbol{u}} = \boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y}$$

で表されるのに対し、(2) 式における残差は、

$$\begin{cases} \boldsymbol{X}^* & \equiv & \boldsymbol{M}_1\boldsymbol{X}_2 \\ \boldsymbol{M}_{\boldsymbol{X}^*} & = & \boldsymbol{I} - \boldsymbol{X}^*(\boldsymbol{X}'^*\boldsymbol{X}^*)^{-1}\boldsymbol{X}'^* \\ & = & \boldsymbol{I} - \boldsymbol{M}_1\boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{M}_1 \end{cases}$$

とおくと、

$$\hat{\boldsymbol{v}} = \boldsymbol{M}_{\boldsymbol{X}^*}\boldsymbol{y}$$

となるからである。

残差ベクトルも同一にするためには、(1) の両辺に左から $\boldsymbol{M}_1$ をかける必要がある。

$$\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{M}_1\boldsymbol{u} \tag{3}$$

これは、次のようにして求めることが出来る。$\boldsymbol{P}_{\boldsymbol{X}} + \boldsymbol{M}_{\boldsymbol{X}} = \boldsymbol{I}$ であるから

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} + \underbrace{\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y}}_{\hat{\boldsymbol{u}}} \\ &= \boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2 + \boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y} \end{aligned}$$

ところで

$$\hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{y}$$

で表されるので

$$\underbrace{\boldsymbol{M}_1}_{n\times n}\underbrace{\boldsymbol{y}}_{n\times 1} = \underbrace{\boldsymbol{M}_1\boldsymbol{X}_1}_{\boldsymbol{O}}\underbrace{\hat{\boldsymbol{\beta}}_1}_{k_1\times 1} + \underbrace{\boldsymbol{M}_1}_{n\times n}\underbrace{\boldsymbol{X}_2}_{n\times k_2}\underbrace{(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{y}_1}_{\hat{\boldsymbol{\beta}}_2} + \underbrace{\boldsymbol{M}_1}_{n\times n}\underbrace{\boldsymbol{M}_{\boldsymbol{X}}}_{n\times n}\underbrace{\boldsymbol{y}}_{n\times 1}$$

を得る。実は $\boldsymbol{M}_1\boldsymbol{M}_{\boldsymbol{X}} = \boldsymbol{M}_{\boldsymbol{X}}$。(別資料を参照のこと。)

定理 2.2.5 Frish=Waugh=Lovell 定理

1. (1) 式における $\boldsymbol{\beta}_2$ の OLSE と (3) 式における $\boldsymbol{\beta}_2$ の OLSE は計算上同一になる。

2. (1) 式における OLS 残差と (3) 式における OLS 残差は計算上同一になる。

FWL 定理の応用

FWL 定理の応用範囲は広いが、固定的季節性やタイムトレンド除去の項を線形回帰モデルの中に含んで推定しても、予め季節性やタイムトレンドを除去してから線形回帰モデルを推定しても、係数の最小 2 乗推定量は一致するという結果はその代表例であろう。(2.2.3.b) 式の $\boldsymbol{X}_1$ を季節性やタイムトレンドを表す説明変数と考えればよい。

もう一つの例が、当てはまりの良さ (goodness of fit) の指標として用いられる決定係数 R^2 である。$\boldsymbol{y}$ は

$$\boldsymbol{y} = \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} + \boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y}$$

と分解できるから、

$$\begin{aligned} \boldsymbol{y}'\boldsymbol{y} &= \boldsymbol{y}'\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y} \\ \parallel \boldsymbol{y} \parallel^2 &= \parallel \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} \parallel^2 + \parallel \boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y} \parallel^2 \\ R^2 &\equiv \frac{ESS}{TSS} = \frac{\parallel \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} \parallel^2}{\parallel \boldsymbol{y} \parallel^2} = \cos^2\theta = 1 - \frac{\parallel \boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y} \parallel^2}{\parallel \boldsymbol{y} \parallel^2} \end{aligned}$$

ここで、θ は $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y}$ と $\boldsymbol{y}$ がなす角である。この決定係数のことを、非中心化 R^2 と呼び、R_u^2 で表す。

ところが、このように定義される R^2 には問題点がある。両辺に定数ベクトル $\mathbf{1}\alpha$ を足しただけで線形関係は同じである線形回帰モデル

$$\boldsymbol{y} + \mathbf{1}\alpha = \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} + \mathbf{1}\alpha + \boldsymbol{M}_{\boldsymbol{X}} \tag{2.2.3.e}$$

では、非中心化決定係数は

$$R_u^2 = \frac{\parallel \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} + \mathbf{1}\alpha \parallel^2}{\parallel \boldsymbol{y} + \mathbf{1}\alpha \parallel^2}$$

となって、α の値によって R_u^2 が変化してしまう。

そこで、FWL 定理を使う。(2.2.3.e) 式の両辺に左側から $\boldsymbol{M}_1$ をかけると、

$$\boldsymbol{M}_1\boldsymbol{y} + \boldsymbol{M}_1\mathbf{1}\alpha = \boldsymbol{M}_1\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} + \boldsymbol{M}_1\mathbf{1}\boldsymbol{\alpha} + \boldsymbol{M}_1\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y}$$

になるが、$\boldsymbol{M}_1\mathbf{1} = \mathbf{0}$ より

$$\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} + \boldsymbol{M}_1\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y} \tag{2.2.3.f}$$

が得られる。(2.2.3.f) から決定係数を求めると、

$$R_c^2 \equiv \frac{\parallel \boldsymbol{M}_1\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} \parallel^2}{\parallel \boldsymbol{M}_1\boldsymbol{y} \parallel^2} = 1 - \frac{\parallel \boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y} \parallel^2}{\parallel \boldsymbol{M}_1\boldsymbol{y} \parallel^2}$$

になり、この R_c^2 が通常用いられる決定係数である。

2.2.4 Influential な観測値とレバレッジ

推定量の頑強性 (robustness) や回帰診断 (regression diagnostics) の観点から見ると、射影行列 $\boldsymbol{P_X}$ は有益な情報をもたらすことが知られている。以下ではその関係についてみることにしよう。

線形単回帰モデルを例にとって見てみよう。図 2.2.4-1 における点 A のようにデータのかたまりから離れた観測データを**外れ値** (outlier) と呼ぶ。点 A を推定に用いるデータから除いたとしても、係数の推定量 $\hat{\beta}_2$ の値に影響をあまり及ぼさないが、点 B のように推定に用いるデータから除くと $\hat{\beta}_2$ の値に大きく影響を及ぼしてしまうものもある。このような観測データ (観測値) を**レバレッジ (てこ) 点** (leverage points) という。

次に、観測データの係数推定量に及ぼす影響を、直観的ではない方法で把握することにしよう。いま線形回帰モデル $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$ において、t 番目の観測データだけ定数項 α だけ異なるような状況を考える。このとき、回帰モデルは

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e}_t\boldsymbol{\alpha} + \boldsymbol{u} \tag{2.2.4.a}$$

に修正される。ここで、$\boldsymbol{e}_t$ は t 番目の要素のみ 1 で、残りは全て 0 である $n \times 1$ ベクトルである。また

$$\boldsymbol{M}_t \equiv \boldsymbol{M}_{\boldsymbol{e}_t} = \boldsymbol{I} - \boldsymbol{e}_t(\boldsymbol{e}_t'\boldsymbol{e}_t)^{-1}\boldsymbol{e}_t'$$

とすると、$\boldsymbol{y}$ から t 番目の観測データを除いて 0 に置き換えたベクトルを $\boldsymbol{M}_t\boldsymbol{y} = \boldsymbol{y} - \boldsymbol{e}_t y_t$ として表現できる。

いま

$$\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{e}_t)\,, \quad \begin{pmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\alpha} \end{pmatrix} = (\boldsymbol{Z}'\boldsymbol{z})^{-1}\boldsymbol{Z}'\boldsymbol{y}$$

とすると、

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{P_Z y} + \boldsymbol{M_Z y} \\ &= \boldsymbol{Z}\begin{pmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\alpha} \end{pmatrix} + \boldsymbol{M_Z y} \\ &= \boldsymbol{X}\tilde{\boldsymbol{\beta}} + \boldsymbol{e}_t\tilde{\boldsymbol{\alpha}} + \boldsymbol{M_Z y} \end{aligned}$$

と表すことができる。上式の両辺に左から $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ をかけると

$$\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e}_t\tilde{\alpha} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_Z y}$$

となるが、$\boldsymbol{X}'\boldsymbol{M_Z} = \boldsymbol{O}$ であることを用いると、

$$\tilde{\beta} - \hat{\beta} = -(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e}_t\tilde{\alpha}$$

のように t 番目の観測データの $\boldsymbol{\beta}$ の推定量に与える影響を $\tilde{\alpha}$ の関数として表すことができる。

FWL 定理を用いると、$\tilde{\alpha}$ を簡単に求めることができる。(2.2.4.a) 式より

$$\boldsymbol{M_X y} = \boldsymbol{M_X e}_t\tilde{\alpha} + residual$$

という回帰モデルの OLSE として $\tilde{\alpha}$ は求まるので、

$$\begin{aligned}\tilde{\alpha} &= \left[(\boldsymbol{e}_t'\boldsymbol{M}_{\boldsymbol{X}})\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{e}_t\right]^{-1}\boldsymbol{e}_t'\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y} \\ &= \frac{\boldsymbol{e}_t'\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y}}{\boldsymbol{e}_t'\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{e}_t} \\ &= \frac{\hat{u}_t}{1-h_t}\end{aligned}$$

が得られる。ここで、h_t は $\boldsymbol{P}_{\boldsymbol{X}}$ の (t,t) 要素である。

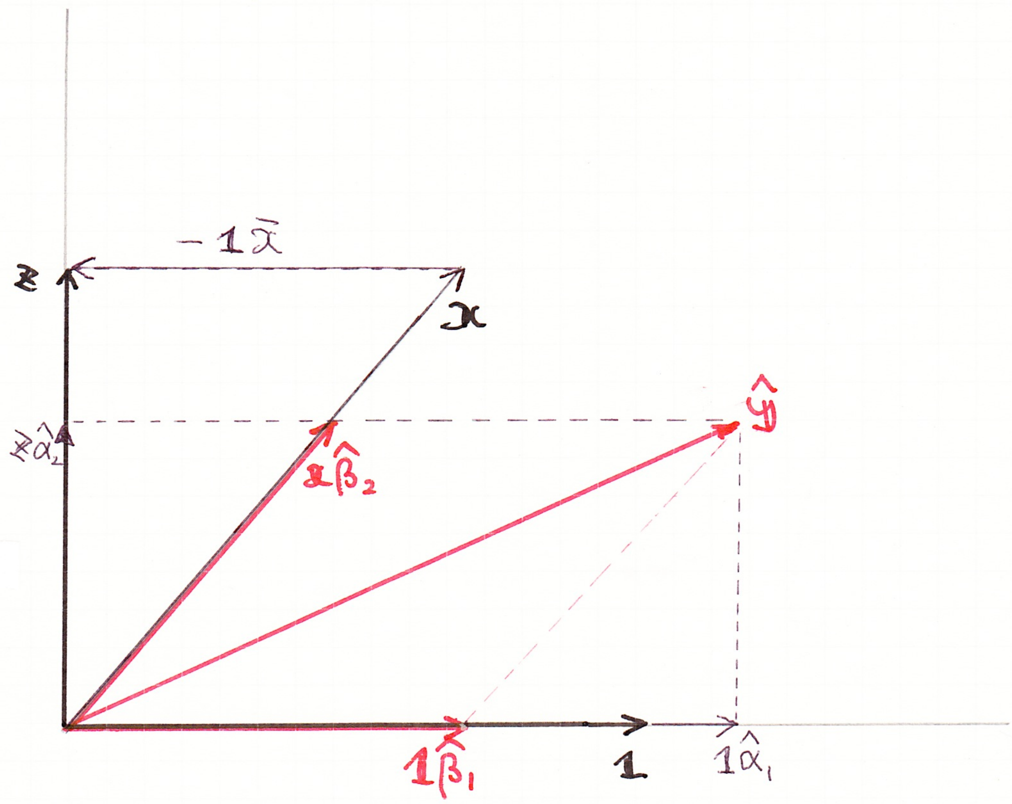
以上の結果より、h_t が大きければ係数推定量に及ぼす影響は大きい (influential) であることがわかる。

実は、

$$\sum_{i=1}^{n} h_t = tr\left(\boldsymbol{P}_{\boldsymbol{X}}\right) = tr\left(\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right)\boldsymbol{X}\right) = tr\left(\boldsymbol{I}_k\right) = k$$

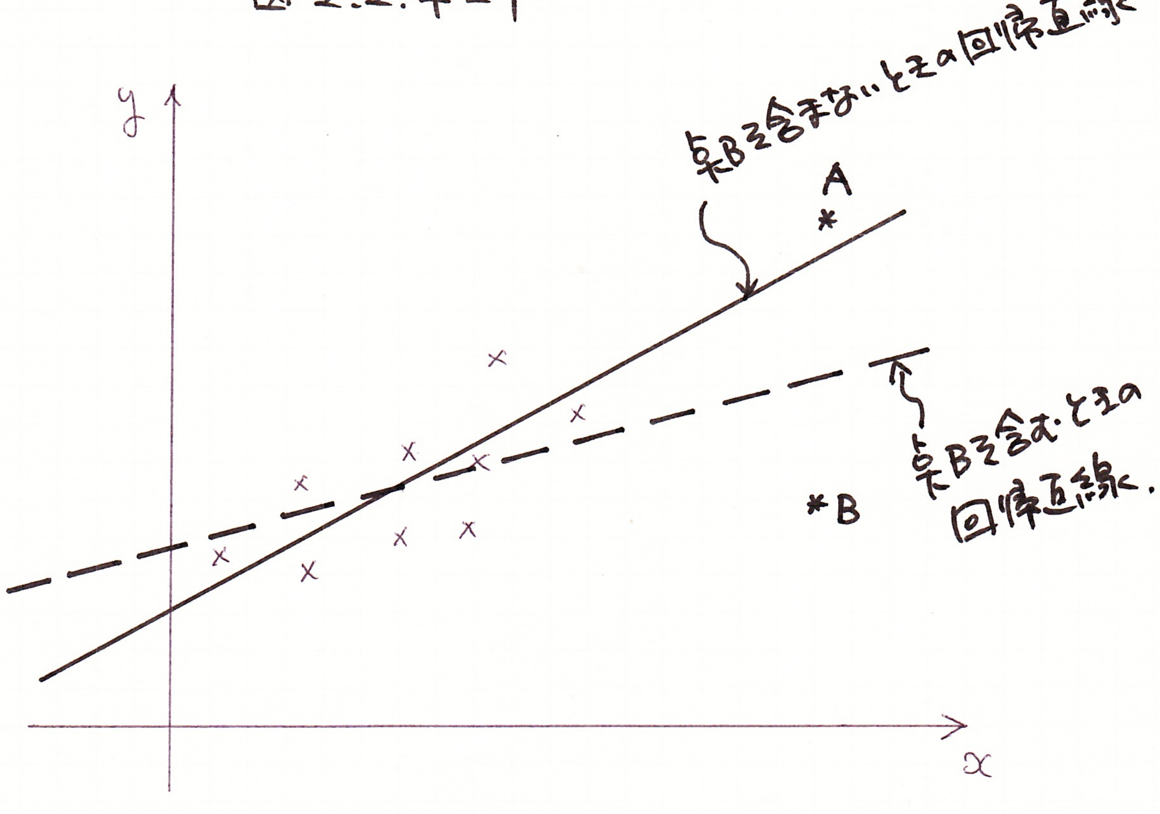
という関係があるので、h_t の算術平均は k/n になっている。

図 2.2.3-1



$-1\bar{x}$
z
x
$\hat{y}$
$z\hat{\alpha}_2$
$x\hat{\beta}_2$
$1\hat{\beta}_1$
1
$1\hat{\alpha}_1$

図 2.2.4-1



点Bを含まないときの回帰直線
A
点Bを含むときの回帰直線.
B
y
x