

2

2010 年度 学部 エコノメトリックス 第 22 回講義メモ

2010 年 12 月 22 日
2011 年 1 月 11 日 (タイトル訂正)
2011 年 1 月 12 日 (訂正版)

§ 4.4 パネルデータの分析

いままで扱ってきた観測データは、クロスセクション（横断面）データのように一時点において観測される様々な個体からなるもの、又は時系列データのように一つの系列が一定時間毎に観測されるもの、のどちらかであった。

しかし、観測される軸として、時間軸と横断（面）軸の二つを共に有するデータも存在する。「上場企業の当期純利益」を例にとって考えてみると、各企業の利益データは、決算期毎に存在する。このような二つの観測軸を持つデータを

- panel data

3

- time series & cross section data
- longitudinal data (biostatistics ではこちらを主に用いる)
- repeated measurement (biostatistics)

のように呼んでいる。

パネルデータのモデル

観測軸が二つあることから、回帰モデルは次のように表現できる。

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + u_{it} \quad (i = 1, \dots, m \text{ and } t = 1, \dots, T)$$

この回帰モデルの一番簡単な推定法は、パネルデータがもつ潜在的な構造を無視し、 u_{it} が i.i.d. である、すなわち

$$E(u_{it}^2) = \sigma^2 \quad \forall i \text{ and } t$$

4

$$E(u_{it}u_{js}) = 0 \quad \forall (i \neq j \text{ or } t \neq s)$$

であるものと仮定して、そのまま OLS を適用するものであり、pooling regression と呼ばれている。

しかし、 u_{it} についてこのような仮定をおくことは、

- (1) 分散不均一性 (heteroscedasticity)
- (2) 時点効果 (ないし個体効果)

の存在を無視している点で問題がある。

もし、この二点が存在するときに OLS 推定を行うと、

- (a) OLSE は効率性を満たさない (inefficient)
- (b) 攪乱項の分散共分散行列の定式化を誤る (misspecified)

5

という問題が起こりうる。

そこで、パネルデータの分析では、攪乱項 u_{it} に関して次のようにモデル化することが行われる。

$$u_{it} = \underbrace{e_t}_{\text{time effect}} + \underbrace{v_i}_{\text{individual effect}} + \underbrace{\varepsilon_{it}}_{\text{iid random dist}} \quad \begin{array}{l} \text{時点効果} \\ \text{個体効果} \\ \text{ランダムエラー} \end{array}$$

これは、分散分析 (ANOVA) の二元配置モデルと同じ構造であることが分かる。(ただし、繰り返し数が 1 であり、交互作用がない点異なる点である。)

諸効果のモデル化の方法

諸効果のモデルとして次の三つが用いられる。

6

(1) fixed effect (固定効果)

e_t や v_i をパラメータとして扱う。

説明変数に **1** (全ての要素が 1 からなるベクトル) を含む場合、パラメータの識別のために $\sum_{t=1}^T e_t = 0$ かつ $\sum_{i=1}^m v_i = 0$ という制約を課す必要がある。

(2) random effect (ランダム効果)

e_t や v_i を確率変数として扱う。

$e_t \perp v_i$ であり、 $e_t \sim (0, \sigma_e^2)$, $v_i \sim (0, \sigma_v^2)$ 。

このモデル化の利点としてパラメータが少数で済むことが挙げられる。

(3) mixed effect (混合効果)

固定効果 と ランダム効果 の mixture。

生物統計、医学統計でよく用いられるモデル。

7

固定効果 (fixed effect) モデルの推定

固定効果モデルの推定は、ダミー変数を用いた OLS 推定により行う。
(以下では、説明の簡略化のために $e_t = 0 \quad \forall t$ と仮定する。)

固定効果モデルは次式で表現できる。

$$\underset{n \times 1}{\mathbf{y}} = \underset{n \times k}{\mathbf{X}} \underset{k \times 1}{\boldsymbol{\beta}} + \underset{n \times m}{\mathbf{D}} \underset{m \times 1}{\boldsymbol{\eta}} + \underset{n \times 1}{\boldsymbol{\epsilon}}$$

$\boldsymbol{\epsilon} \sim \text{iid}(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_n)$

ここで、 $n = mT$ であり、 $\mathbf{y}$, $\mathbf{X}$, $\boldsymbol{\epsilon}$ および $\mathbf{D}$, $\boldsymbol{\eta}$ は以下の形をとるベクトルおよび行列である。

8

$$\mathbf{y} = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1T} \\ y_{21} \\ \vdots \\ y_{mT} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}'_{11} \\ \vdots \\ \mathbf{x}'_{1T} \\ \mathbf{x}'_{21} \\ \vdots \\ \mathbf{x}'_{mT} \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1T} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{mT} \end{pmatrix},$$

9

$$\mathbf{D} = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \ddots & & \vdots \\ \vdots & \mathbf{0} & \mathbf{1} & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{1} \end{pmatrix}, \quad \boldsymbol{\eta} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix}$$

証明は省くが、固定効果モデルにおける $\boldsymbol{\beta}$ の OLS 推定量は

$$\mathbf{M}_D \mathbf{y} = \mathbf{M}_D \mathbf{X} \boldsymbol{\beta} + \text{error} \quad (4.4.1)$$

における $\boldsymbol{\beta}$ の OLS 推定量に一致する。 $(\mathbf{M}_D = \mathbf{I}_n - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}')$ より $\mathbf{M}_D \mathbf{D} = \mathbf{O}$ となるため。よって、 $\boldsymbol{\beta}$ の OLS 推定量とその分散は

$$\mathbf{b}_{LSDV} = (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_D\mathbf{y},$$

$$\text{Var}(\mathbf{b}_{LSDV}) = \sigma_\epsilon^2 (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}$$

10

として得られる。

ここで注意したいのは、 $\mathbf{D}'\mathbf{D} = T\mathbf{I}_m$ であるので、

$$\mathbf{M}_D = \mathbf{I} - \frac{1}{T} \mathbf{D} \mathbf{D}' = \begin{pmatrix} \mathbf{I}_T & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \mathbf{I}_T & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{O} \\ \mathbf{O} & \cdots & \mathbf{O} & \mathbf{I}_T \end{pmatrix} - \frac{1}{T} \begin{pmatrix} \mathbf{1}\mathbf{1}' & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \mathbf{1}\mathbf{1}' & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{O} \\ \mathbf{O} & \cdots & \mathbf{O} & \mathbf{1}\mathbf{1}' \end{pmatrix}.$$

になっている。ゆえに、 $n \times 1$ ベクトルである $\mathbf{M}_D \mathbf{y}$ の $(it, 1)$ 要素は、 $\bar{y}_i \equiv (1/T) \sum_{t=1}^T y_{it}$ で表すと

$$(\mathbf{M}_D \mathbf{y})_{it,1} = (y_{it} - \bar{y}_i)$$

であることが分かる。この性質より、 $\mathbf{b}_{LSDV}$ を 'within estimator' (グループ内変動推定量) と呼ぶ。

11

ランダム効果 (random effect) モデルの推定

次に、 v_i をパラメータではなく $v_i \sim (0, \sigma_v^2)$ である確率変数として考えることにしよう。

このとき、グループ i 毎に T 期間の平均を求めると、回帰モデルは

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T y_{it} &= \left(\frac{1}{T} \sum_{t=1}^T \mathbf{x}_{it} \right)' \boldsymbol{\beta} + \frac{1}{T} \sum_{t=1}^T v_i + \frac{1}{T} \sum_{t=1}^T \varepsilon_{it} \\ &= \left(\frac{1}{T} \sum_{t=1}^T \mathbf{x}_{it} \right)' \boldsymbol{\beta} + v_i + \frac{1}{T} \sum_{t=1}^T \varepsilon_{it} \end{aligned}$$

と書くことができる。この式を行列を用いて表すと、

$$\mathbf{P}_D \mathbf{y} = \mathbf{P}_D \mathbf{X} \boldsymbol{\beta} + \text{error} \quad (4.4.2)$$

12

となり、その OLS 推定量は

$$\mathbf{b}_{BG} = (\mathbf{X}'\mathbf{P}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_D\mathbf{y}$$

で得ることができる。 $\mathbf{P}_D \mathbf{y}$ の $(it, 1)$ 要素は、

$$(\mathbf{P}_D \mathbf{y})_{it,1} = (\bar{y}_i)$$

であることから、 $\mathbf{b}_{BG}$ を 'between estimator' (グループ間変動推定量) と呼ぶ。

さて、分散分析においては

$$\text{全変動} = \text{グループ内変動} + \text{グループ間変動}$$

であった。同様にして pooling regression における $\boldsymbol{\beta}$ の OLS 推定量 $\mathbf{b}$ を変形すると、

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

13

$$\begin{aligned}
 &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{M}_D + \mathbf{P}_D)\mathbf{y} \\
 &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_D\mathbf{X}\mathbf{b}_{LSDV} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_D\mathbf{X}\mathbf{b}_{BG}
 \end{aligned}$$

のように、'within estimator' と 'between estimator' の線形結合として得られる。

$v_i \perp \epsilon_{it}$ であるから、(4.4.2) 式の誤差項の分散共分散行列は、 $\Xi \equiv \sigma_v^2 \mathbf{1}\mathbf{1}' + \frac{1}{T}\sigma_\epsilon^2 \mathbf{I}_T$ とおくと、

$$\begin{pmatrix} \Xi & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \Xi & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{O} \\ \mathbf{O} & \cdots & \mathbf{O} & \Xi \end{pmatrix}$$

で表される。これより $T \rightarrow \infty$ に近づくと、(4.4.2) 式の誤差項の分散共分散における σ_ϵ^2 の寄与は小さくなるが、 σ_v^2 の部分はそのまま残ることが

14

わかる。したがって、漸近効率性を持つ推定量を得るためにも、GLS 推定を行う必要がある。