2000/6/29

第7講

比較統計システム論 講義録(2000 年度)

不完全パネルデータ(Unbalanced Panel Data)
あるいは不完備パネルデータ(Incomplete Panel Data)

$$Y_{it} = \alpha + \beta' X_{ie} + u_i + \varepsilon_{it}$$
$$w_{it} = \varepsilon_{it} + u_i$$
$$w_i = [w_{i1}, w_{i2}, \cdots, w_{it}]'$$

ここで $\Omega = E[w_i w_i']$ とすれば、

$$\Omega = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_\varepsilon^2 + \sigma_u^2 & & \sigma_u^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_\varepsilon^2 + \sigma_u^2 \end{bmatrix} = \sigma_\varepsilon^2 I + \sigma_u^2 ii'$$

ここで i は T×1の単位行列である。誤差共分散行列は次のように表わせる。

$$V = \begin{bmatrix} \Omega & 0 & \cdots & 0 \\ 0 & \Omega & & 0 \\ 0 & 0 & & 0 \\ \vdots & \vdots & \ddots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & \cdots & \cdots & \Omega \end{bmatrix}$$

もし不完全パネルデータを用いるとすれば V 行列の長方行列ではなくなる。

Ωのi番目の要素は

$$\Omega_i^{-1/2} = I_{Ti} - \frac{\theta_i}{T_i} ii'$$

$$\theta_i = 1 - \frac{\sigma_\varepsilon}{\sqrt{T_i \sigma_u^2 + \sigma_\varepsilon^2}}$$

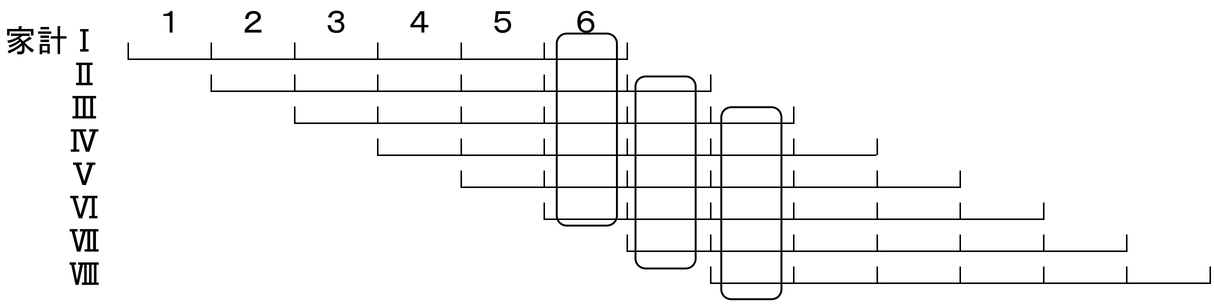
この場合の調整は簡単である。θ_i をi毎に計算してそれを用いて一般化最小 2 乗法(GLS)

を行う必要がある。誤差の分散は次のように表わせる。

$$Var\left[u_i+\frac{\sum_{t=1}^{T_i}\varepsilon_{it}}{T_i}\right]=\sigma_u^2+\frac{\sigma_\varepsilon^2}{T_i}$$

標本(サンプル)が<u>ローテーション</u>（Rotation）して一定割合ずつ変わっていく調査がある。

家計調査　6 ヶ月（毎月 6 分の１（8000／6≈1333）ずつ更新される）



<u>疑似パネル（Psudo Panel）</u>

<u>年齢（cohort）</u>で結び付けた、<u>パネルデータ</u>を作れば、現在のパネルデータより長期のデータが作れる。世代会計、貯蓄、消費分析では有効。

<u>年齢（cohort）</u>
５歳区切りなど cohort の大きさをある程度大きくして、その中に入る標本数をある程度確保することが安定的なパラメータを得る条件である。

その他の属性でグループ分けすることも可能。例えば、地域や男女、教育水準等。

不均一分散問題（Heteroscedasticity）

誤差項間の相関はゼロであっても分散が不均一な場合には、古典的回帰モデルの想定しているケースとは異なる。

例（１）　所得の上昇とともに消費の平均と分散が上昇するような場合、線形回帰モデルを用いる際には注意が必要。

例（２）　集計データ　官庁統計などでは、各市町村、都道府県別の平均値が公表されている。もし個人レベルのデータで古典的回帰モデルが成立しているなら、平均値で表わしたモデルで均一分散性は成立しない。個人レベルのデータで古典的回帰モデルの

仮定が成立し、モデルが次のように書けるとする。

$$y_{ij} = \alpha + \beta x_{ij} + \varepsilon_{ij} \qquad Var(\varepsilon_{ij}) = \sigma^2$$

添字iは個人、jは地域を表わす。各地域の標本数をn_1, n_2, …とし、地域別平均に注目してモデルを書き直せば、j地域のデータは

$$\bar{y}_{ij} = \alpha + \beta \bar{x}_j + \bar{\varepsilon}_j$$

となる。ここで、$\bar{y}_j$, $\bar{x}_j$, $\bar{\varepsilon}_j$ は y, x, ε の地域別平均で、特に$\bar{\varepsilon}_j$ は分散 σ^2 、標本数 n_j 個の平均をあらわす。（集計値の）誤差項$\bar{\varepsilon}_j$ の分散は$Var(\bar{\varepsilon}_j) = \sigma^2 / n_j$ であり、標本数と反比例する。各地域の標本数が異なれば、集計データのモデルでは分散不均一となる。この場合の分散構造は、次のように書ける。

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & 0 \\ 0 & \vdots & \ddots & \vdots \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

不均一分散の検出法

2つのグループで分散が異なるケース

標本の中に2つのグループがあり、グループ内の分散は共通であるがグループ間で異なる疑いがある場合には統計理論に基づいた検定ができる。
グループ1，2の分散を$\sigma^2{}_1$，$\sigma^2{}_2$、標本数をn_1，n_2とする。

グループ1　　$Y_1 = X_1 \beta_1 + \varepsilon_1 \qquad Var(\varepsilon_1) = \sigma_1{}^2 I$

グループ2　　$Y_2 = X_2 \beta_2 + \varepsilon_2 \qquad Var(\varepsilon_2) = \sigma_2{}^2 I$

この枠組みで帰無仮説$\left(H_0 = \sigma_1{}^2 = \sigma_2{}^2\right)$に対して検定を行う。具体的には、分散比率$S_1^2 / S_2{}^2$が自由度$\left(n_1 - k, n_2 - k\right)$の F 分布に従うかどうかを検定する。

White テスト

不均一分散があることはわかっても、それ以上のことが何もわからない場合には、残差の２乗 ε_i^2 を定数項を含めた説明変数間の積からなる項に回帰し注)、決定係数 R^2 を求める。

White はこの回帰の R^2 をn倍したものが漸近的に自由度Ｐ－１のカイ２乗分布に従うことを示した。

注) 説明変数が定数項と W, X, Z なら残差２乗の回帰における右辺は定数項、W, X, Z, W^2, X^2, Z^2, WX, WZ, XZ (P=10)。説明変数が定数項 X, X^2 なら、右辺は定数項 X, X^2, X^3, X^4 (P=5) とする。ここでPは残差２乗の回帰における説明変数の数。

パネルデータ分析の統計的手法

一般に次のような線形モデルを考える。

$$Y_{it} = \alpha_i + \gamma_t + \beta_i X_{it} + u_{it} \qquad i=1, \cdots, N, t=1, \cdots, T \tag{1}$$

（1）傾きが違い、切片も違う （**時系列推計＝β** ts）

$$\alpha_i \neq \alpha_j, \beta_i \neq \beta_j, \gamma_t = 0$$
$$Y_{it} = \alpha_i + \beta_i X_{it} + u_{it} \tag{2}$$

（2）傾きが同じで、切片も同じ （**プーリング推計＝β** pool）

$$\alpha_i = \alpha_j, \beta_i = \beta_j, \gamma_t = 0$$
$$Y_{it} = \alpha + \beta X_{it} + u_{it} \tag{3}$$

（3）傾きが同じで、切片が固定的に違う （**一元配置固定効果推計＝β** of）

$$\alpha_i \neq \alpha_j, \beta_i = \beta_j, \gamma_t = 0$$
$$Y_{it} = \alpha + \beta X_{it} + u_{it} \tag{4}$$

（4）傾きは同じだが、切片は違い、各標本に共通の時間ショックがある

（**二元配置固定効果推計＝β** tf）

$$\alpha_i \neq \alpha_j, \beta_i = \beta_j, \gamma_t \neq 0$$
$$Y_{it} = \alpha_i + \gamma_t + \beta X_{it} + u_{it} \tag{5}$$

（5）ある期間の平均に対して、傾きも切片も同じ

（**ビトウィン推計＝β** betw）

$$\alpha_i = \alpha_j, \beta_i = \beta_j, \gamma_t = 0$$
$$\overline{Y}_{it} = \alpha + \beta \overline{X}_{it} + u_{it} \tag{6}$$

（6）傾きも切片も同じだが、各標本の切片は確率的に変動する

（**一元配置ランダム効果推計＝β** or）

$$Y_{it} = \alpha + \beta X_{it} + u_i + \varepsilon_{it} \tag{7}$$

（7）傾きも切片も同じだが、各標本の切片が確率的に変動し、かつ時間ショックも確率的に

発生する

（二元配置ランダム効果推計＝β_{tr}）

$$Y_{it} = \alpha + \beta X_{it} + u_i + \nu_t + \varepsilon_{it} \qquad (8)$$

これらの推計の間には、次のような関係があることが知られている（β がスカラーの場合にのみ成立）、

ANOVA **定理** ①　$\hat{\beta}_{pool} = \theta\hat{\beta}_{of} + (1-\theta)\hat{\beta}_{betw}$　（9）

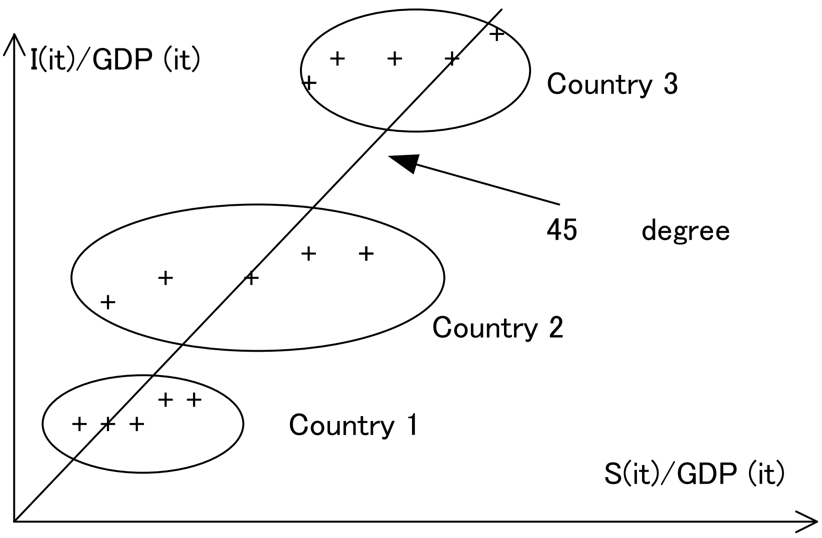
β_{pool} は分散のうち時系列方向の分散の占める比率 θ をウェイトとする β_{of} と β_{betw} の加重平均である。

ANOVA **定理** ②　$\beta_{of} = \sum_{i=1}^{N} \delta_i \beta_{ts}$　（10）

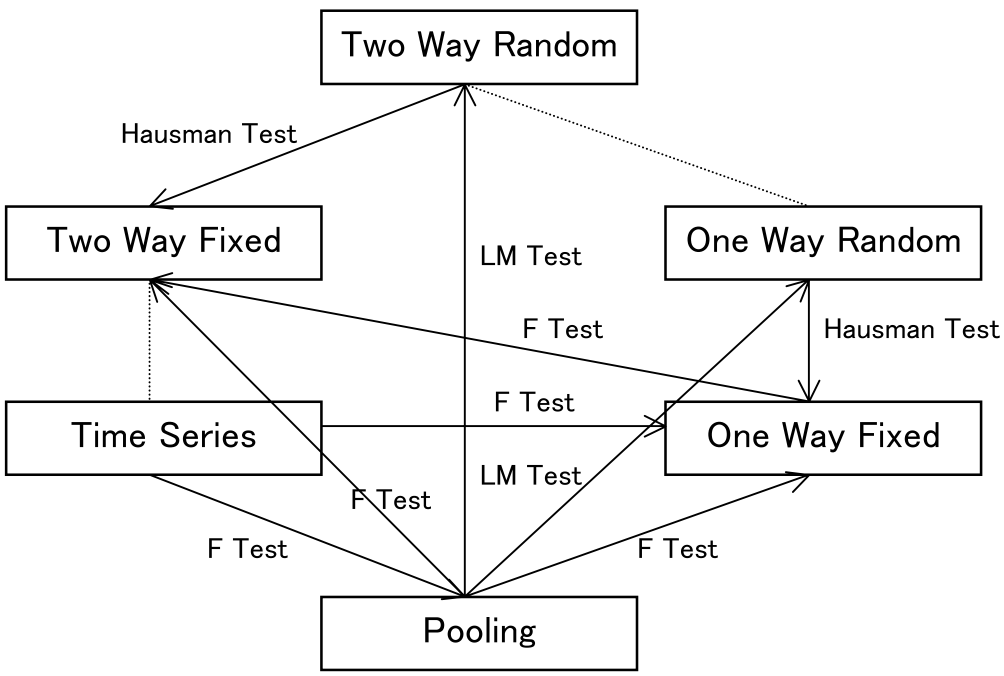
β_{of} は時系列方向の分散のうち経済主体iが占める分散比率 δ_i をウェイトとすると β_{ts} の加重平均である。

Figure 1

Graphical Presentation of the Feldstein-Horioka Paradox



パネルデータ分析におけるモデル選択プロセスの構造



注）矢印は、各検定・診断テストについて、矢印の根元が帰無仮説、矢印の先が対立仮説を表わす。

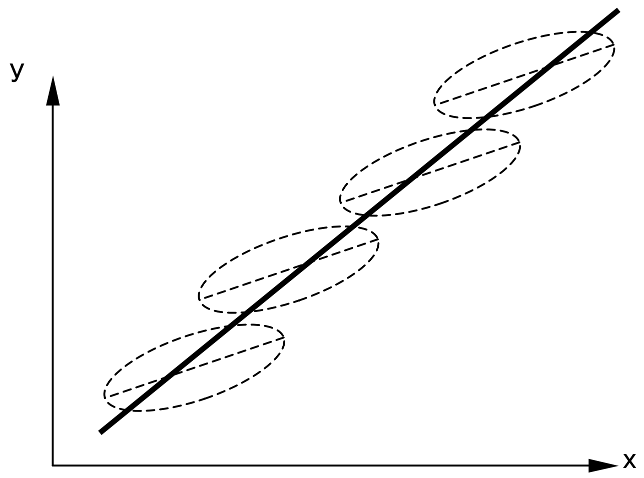


Fig. 1.1

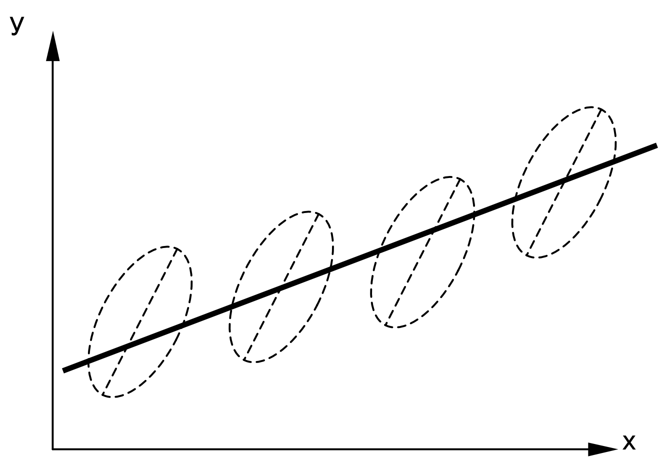


Fig 1.2

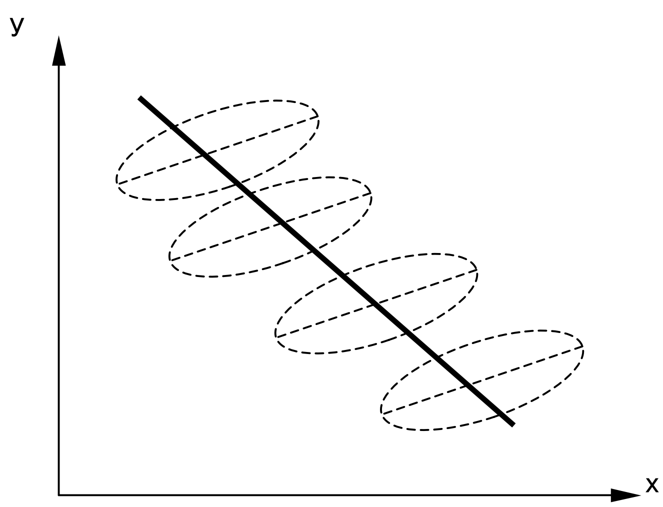


Fig. 1.3

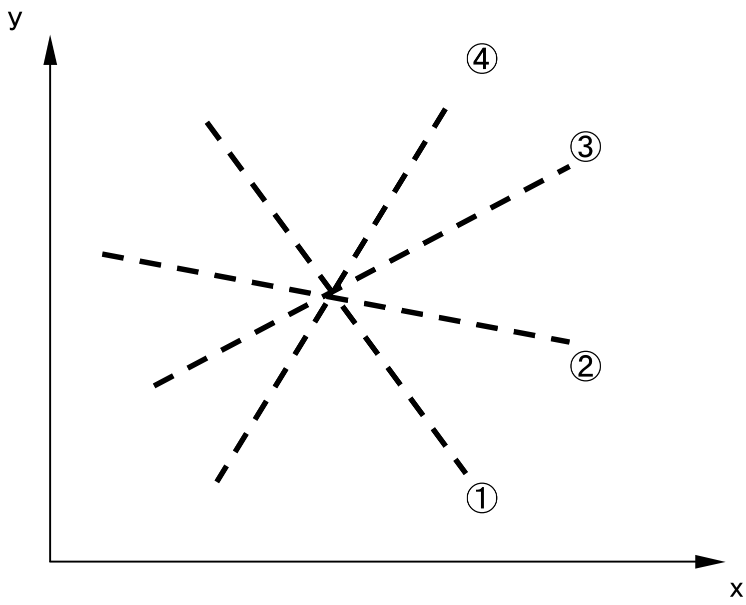


Fig. 1.4

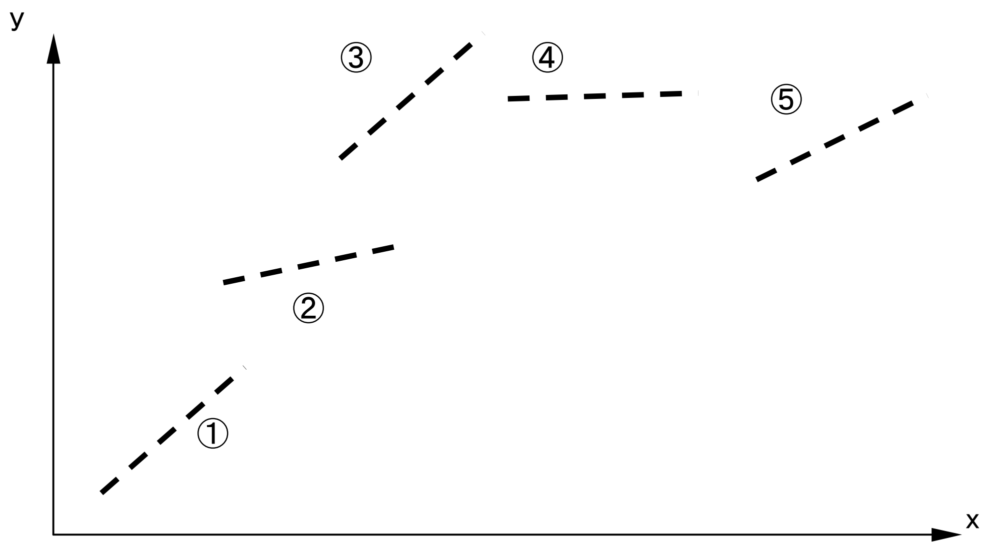


Fig 1.5

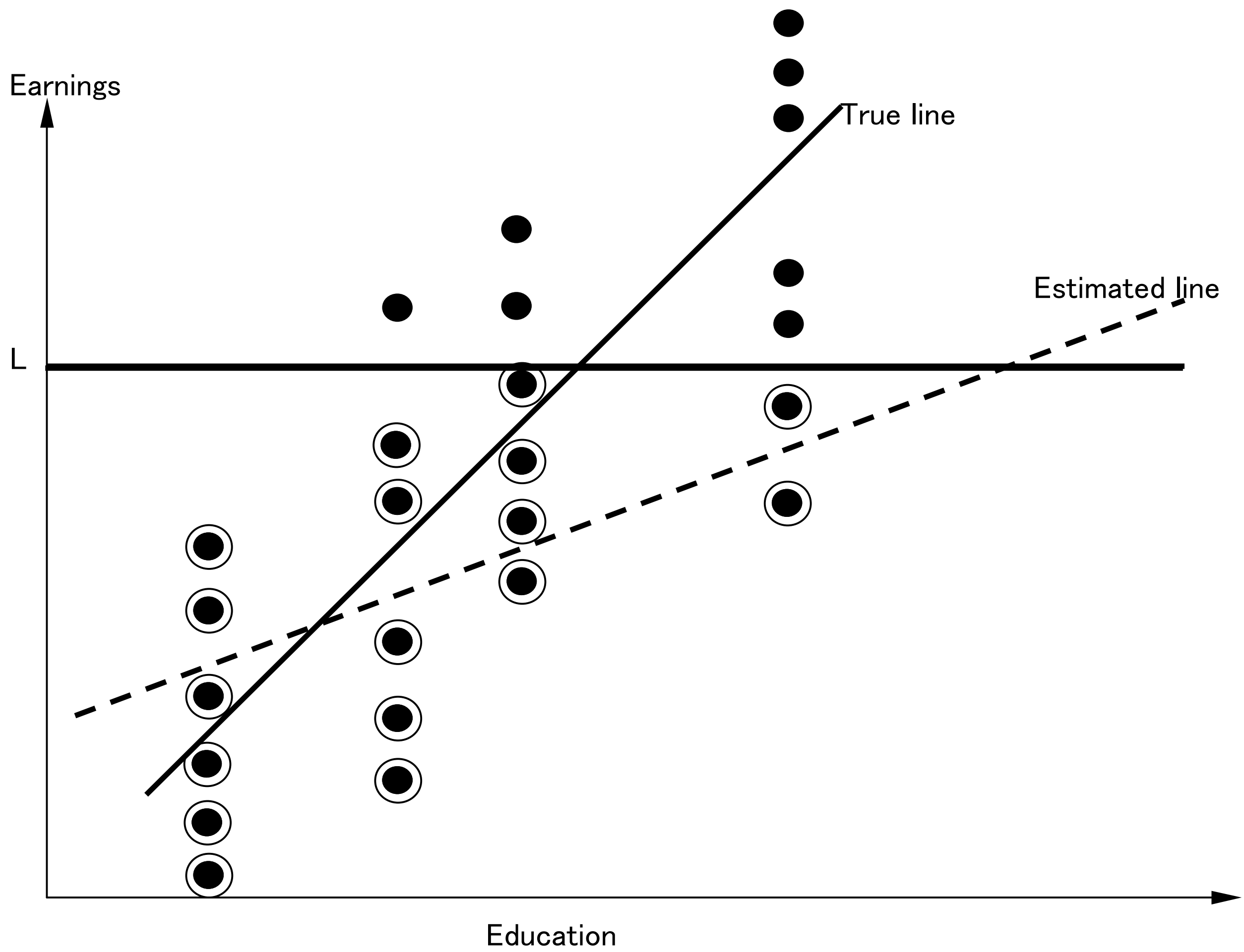
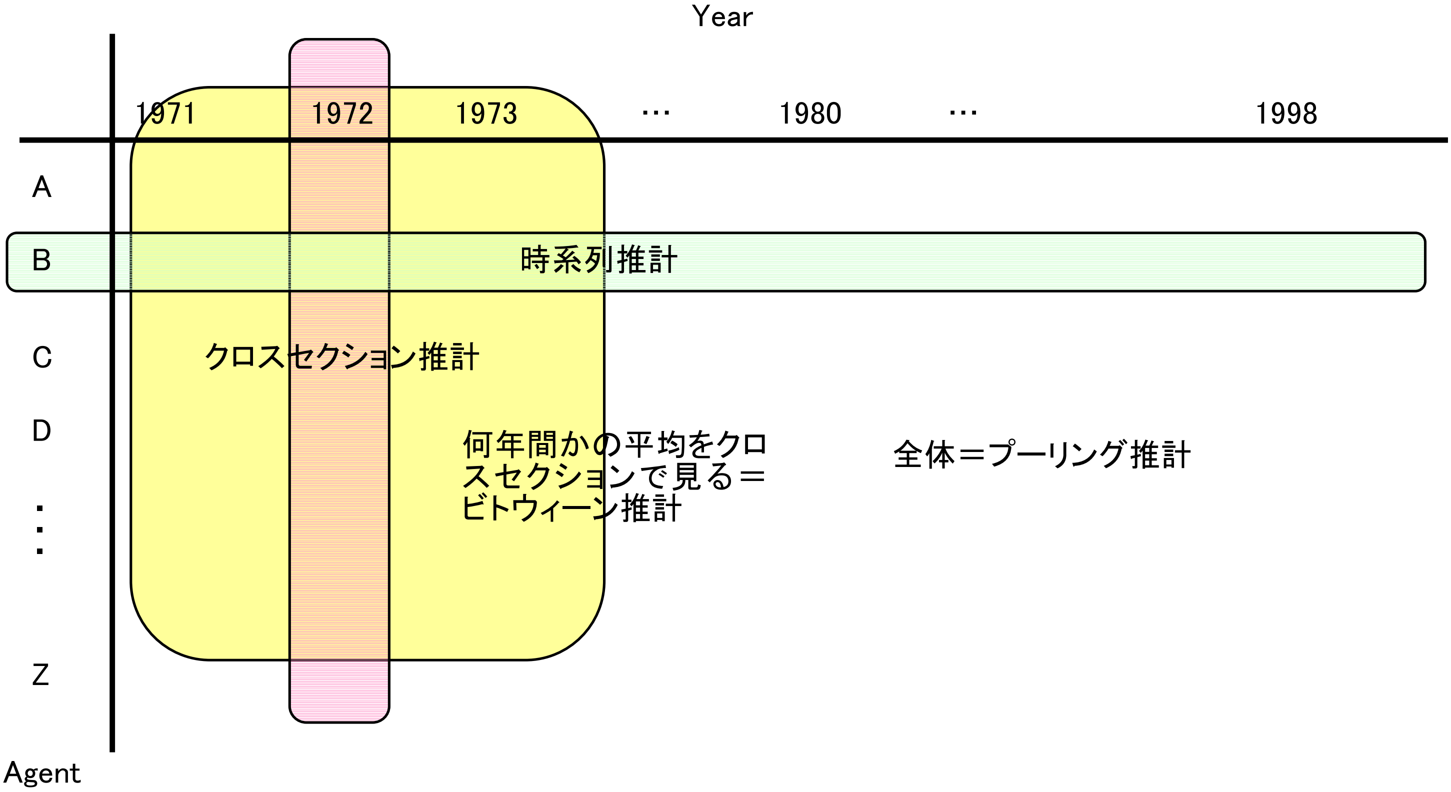


Fig. 1.6

パネルデータの構造


Year
1971
1972
1973
…
1980
…
1998
A
B
時系列推計
C
クロスセクション推計
D
何年間かの平均をクロスセクションで見る＝ビトウィーン推計
全体＝プーリング推計
Z
Agent