パネルデータ分析

$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + a_i + u_{it}$

2. Advanced

Ch.14 パネルデータ分析

1. 固定効果推定
2. ランダム効果モデル
3. 相関ランダム効果アプローチ
4. 他のデータ構造への応用*

14.1 固定効果推定

◆固定効果推定Fixed effects estimation

$$y_{it} = \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + a_i + u_{it},\ i = 1, \ldots, N, t = 1, \ldots, T$$

- a_i:固定効果/潜在的に説明変数と相関

$$\bar{y}_i = \beta_1 \bar{x}_{i1} + \ldots + \beta_k \bar{x}_{ik} + \bar{a}_i + \bar{u}_i$$

- 各個人に対して時間平均で表した式

$$\Rightarrow \boxed{[y_{it} - \bar{y}_i]} = \beta_1 \boxed{[x_{it1} - \bar{x}_{i1}]} + \ldots + \beta_k \boxed{[x_{itk} - \bar{x}_{ik}]} + \boxed{[u_{it} - \bar{u}_i]}$$

- 固定効果を除去: $a_i - \bar{a}_i = 0$

- time-demeaned equationをOLSで推定
  - 固定効果がないためOLS推定でも一致性
  - 横断面ユニット内の時間変化を利用して算出
    - この推定量をwithin estimatorとも呼称

14.1 固定効果推定

◆例14.1:廃棄率への訓練給付の効果

$$scrap_{it} = \beta_1 d88_{it} + \beta_2 d89_{it} + \beta_3 grant_{it} + \beta_4 grant_{it-1} + a_i + u_{it}$$

- 各企業の時間不変特性(ある企業が他の企業より生産性が高い要因等)をa_iによりコントロール
- 重要な点は a_iが他の説明変数と相関する可能性

→ a_iを除去して計算

- 固定効果推定1987-89年

$$\widehat{scrap}^*_{it} = \underset{(.109)}{-\ .080}\ d88^*_{it} \underset{(.133)}{-\ .247}\ d89^*_{it} \boxed{\underset{(.151)}{-\ .252}}\ grant^*_{it} \boxed{\underset{(.210)}{-\ .422}}\ grant^*_{it-1}$$

$$n = 162, R^2 = .201$$

- *元の式から時間平均を除いた変換変数
- 訓練給付は生産性を大幅に上昇(特に遅れて廃棄率低下)

14.1 固定効果推定

◆固定効果推定量について

- 元のモデルで強外生性の仮定が必要
- 変換式のR^2は不適切
- 時間不変変数の効果は推定できない
  - ただし時間不変変数との交差項効果は推定可能(例:教育と時間ダミーとの交差項)
- フルセットの時間ダミーを含む場合、時間変化一定の変数の効果の推定不可能(例:経験年数)
- N個の平均値を使用するため自由度は要修正
  - 修正自由度= *NT-N-k*

14.1 固定効果推定

◆ダミー変数の固定効果解釈

- 固定効果推定量は、元の回帰式の各個人をダミーで表したプールドOLS推定量と等しい

$$y_{it} = a_1 \boxed{ind1_{it}} + a_2 \boxed{ind2_{it}} + \ldots + a_N \boxed{indN_{it}} + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + u_{it}$$

- *indN*:観測値が個人Nであれば1、そうでなければ0

  - 固定効果の推定後、固定効果は次のように推定可能

$$\hat{a}_i = \bar{y}_i - \hat{\beta}_1 \bar{x}_{i1} - \ldots - \hat{\beta}_k \bar{x}_{ik},\ i = 1, \ldots, N$$

- a_i:個人iの推定個人効果

  - ダミーの数が多く、自由度低下というデメリットも

14.1 固定効果推定

- 固定効果or1階階差？
  - T > 2の場合、両手法ともに使用可能
    - T = 2の場合、固定効果と1階階差は同一
    - T > 2の場合、古典的仮定下で固定効果はより効率的
    - 深刻な系列相関（ランダムウォーク等）の場合、1階階差がbetter
    - Tが非常に大きい（Nはそれほど大きくない）場合、パネルは顕著な時系列特性を有し、強依存性等の問題
      - このような場合は、おそらく1階階差がbetter
    - 通常の場合、両方を計算し、頑健性チェックを推奨

14.2 ランダム効果モデル

- ランダム効果モデルRandom effects models

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + a_i + u_{it},\ i = 1, \ldots, N, t = 1, \ldots, T$$

  - 個別効果はランダムと仮定/説明変数と無相関
  - ランダム効果仮定：$Cov(x_{itj}, a_i) = 0, j = 1,2, \ldots, k$
  - 複合誤差項 $a_i + u_{it}$は説明変数と無相関
  - ただし同じ個人 i の観測値に対して系列相関が発生

$$Cov(a_i + u_{it}, a_i + u_{is}) = Cov(a_i, a_i) = \sigma_a^2$$

    - uに系列相関無しの仮定でも系列相関が発生
    - 例：賃金方程式において、ある個人の各期間の誤差項は観察されない同じ能力を含む。したがって、複合誤差項はこの個人に対して期間をまたいで相関

14.2 ランダム効果モデル

- ランダム効果モデルの推定
  - Pooled OLSは一致推定値を算出
    - ランダム効果仮定の下で、説明変数は外生的
    - OLS使用の場合、ユニットiの系列相関のために標準誤差の要修正（→クラスター化標準誤差）
    - また系列相関のため、OLSは非効率的
  - GLS仮定を満たすようにモデルを変換

$$\boxed{[y_{it} - \lambda\bar{y}_i]} = \beta_1 \boxed{[x_{it1} - \lambda\bar{x}_{i1}]} + \ldots + \beta_k \boxed{[x_{itk} - \lambda\bar{x}_{ik}]} + \boxed{[a_i - \lambda\bar{a}_i + u_{it} - \lambda\bar{u}_i]}$$

    - 各変数はQuasi-demeaned dataに変換
    - 誤差項はGLS仮定を満たす

14.2 ランダム効果モデル

- ランダム効果モデルの推定（続き）

with $\boxed{\lambda = 1 - \left[\sigma_u^2 / (\sigma_u^2 + T\sigma_a^2)\right]^{1/2}}$, $0 \le \lambda \le 1$

  - λは不明であるが、推定可能
  - 推定λを用いたFGLSをランダム効果推定と呼称
  - aがuより重要でない場合（λ→0）、FGLSはpooled OLSに近づく
  - aがuより重要な場合（λ→1）、FGLSは固定効果推定に近づく
  - ランダム効果推定は時間不変変数にも有効

14.2 ランダム効果モデル

- 例14.4：パネルデータと賃金方程式

$$\begin{aligned}\widehat{\log}(wage_{it}) = {} & \underset{(.011)}{.092}\ \boxed{educ_{it}} - \underset{(.048)}{.139}\ \boxed{black_{it}} + \underset{(.043)}{.022}\ \boxed{hispan_{it}} \\ & + \underset{(.015)}{.106}\ exper_{it} - \underset{(.0007)}{.0047}\ exper_{it}^2 + \underset{(.017)}{.064}\ married_{it} \\ & + \underset{(.018)}{.106}\ union_{it} + time\ dummies\end{aligned}$$

    - 多くの変数が時間不変のためランダム効果推定
    - ただしこの場合、ランダム効果仮定は現実的か？

- ランダム効果or固定効果？
  - 経済学では、観察されない個別効果は説明変数と相関するため、固定効果がより説得的

14.3 相関ランダム効果アプローチ

- 相関ランダム効果CREアプローチ

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + a_i + u_{it},\ i = 1, \ldots, N, t = 1, \ldots, T$$

  - 相関ランダム効果Correlated Random Effect, CRE

$$a_i = \boxed{\alpha + \gamma_1 \bar{x}_{i1} + \ldots + \gamma_k \bar{x}_{ik}} + r_i$$

    - 個別効果a_iは、説明変数の時間平均に関連する部分と説明変数に無相関な部分r_iに分割

$$\Rightarrow\ y_{it} = \beta_0 + \alpha + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + \gamma_1 \bar{x}_{i1} + \cdots + \gamma_k \bar{x}_{ik} + [r_i + u_{it}]$$

    - このモデルは、無相関ランダム効果r_iと時間平均の追加説明変数を有する通常のランダム効果モデル
    - このモデルでは、説明変数の係数推定値は固定効果推定値と等しくなる
    - 利点：①FE対（普通の）REの検定可能、②時間不変変数の組込み可能

14.4 他のデータ構造への応用

他のデータ構造への応用

- パネルデータ手法は観察されない一定効果を除去すべき場合にも使用可能
- 例：双子の賃金
  - 家庭 i の双子1
    $\log(wage_{i1}) = \beta_0 + \beta_1 educ_{i1} + \ldots + a_i + u_{i1}$
  - 家庭 i の双子2
    $\log(wage_{i2}) = \beta_0 + \beta_1 educ_{i2} + \ldots + a_i + u_{i2}$
    - a：双子の間で共通かつ観察されない遺伝的・家族的特徴

  $\Rightarrow$ $\boxed{\Delta \log(wage_i) = \beta_1 \Delta educ_i + \ldots + \Delta u_i}$
    - 差分式をOLSで推定