11章　パネル分析

藪友良

・パネルデータ

・個別効果

・固定効果モデル

・時間効果

・変量効果モデルとハウスマン検定

パネルデータ

- パネルデータとは

$$(X_{it}\ ,\ Y_{it}\ )$$

--- 横断面データが時間を通じて記録されたデータ

--- iは観測番号、tは時点($i = 1,2, \dots , N,\ t = 1,2, \dots , T$)

--- サンプルサイズは$N\ T$

例) 47都道府県の10年分のデータなら、サンプルサイズは470個

- **パネルデータの長所**

①サンプルサイズが大きいため精度の高い推定が可能

--- 係数の値が小さくても有意な結果が出る

②説明変数の変動が大きく多重共線性の問題を軽減できる

--- 横断面の変動に時間の変動も加わることで、

説明変数間に異なる変動が生じやすい

③各個人や各企業に個別の要因を考慮できる

--- 欠落変数バイアスを除去できる

個別効果

モデル

$$Y_{it} = \alpha + \beta X_{it} + \gamma Z_i + u_{it}$$

- Z_i： 個別効果、i固有の要因

  ①分析者には観察できない

  ②時間を通じて一定

  --- Z_iを考慮しないと推定にバイアスが発生（欠落変数バイアス）

  例) Y_{it}を所得、X_{it}を教育水準、Z_iを能力とする

- $\alpha_i = \alpha + \gamma Z_i$:個別効果(時間を通じて一定)

  $\Rightarrow \quad Y_{it} = \beta X_{it} + \alpha_i + u_{it}$

- 固定効果と変量効果

  --- α_iが説明変数X_{it}と相関したら固定効果(fixed effect)と呼ぶ

  --- α_iが説明変数X_{it}と無相関なら変量効果(random effect)と呼ぶ

  --- 固定効果が現実的な前提となる

プールドOLS

$$Y_{it} = \beta X_{it} + \alpha_i + u_{it}$$

--- 被説明変数をY_{it} 、説明変数をX_{it}としたOLS推定

--- 誤差項$e_{i,t}$は、

$$e_{it} = \alpha_i + u_{it}$$

--- 固定効果なら、欠落変数バイアスが生じる

変量効果なら、欠落変数バイアスは生じない

--- 個体iに着目すれば、誤差項e_{it}は計T個の時系列データ

$$(e_{i1}, e_{i2}, \dots, e_{iT})$$

クラスター(集団)での系列相関を考慮したクラスター構造に対して頑健な標準誤差を用いる(クラスターロバスト標準誤差)

固定効果モデル

- $Y_{it} = \beta X_{it} + \alpha_i + u_{it}$

  --- α_iをダミー変数を用いてコントロールする

- $Y_{it} = \beta X_{it} + \alpha_1 D1_i + \alpha_2 D2_i + \cdots + \alpha_N DN_i + u_{it}$

--- $D1_i$は$i = 1$なら1、$i \neq 1$なら0となるダミー変数

--- $D2_i$は$i = 2$なら1、$i \neq 2$なら0となるダミー変数

--- $i = 1$なら、$D1_i = 1$, $D2_i = D3_i = \cdots = DN_i = 0$となり、

$$Y_{1t} = \beta X_{1t} + \alpha_1 + u_{1t}$$

--- $i = 2$なら、$D2_i = 1$, $D1_i = D3_i = \cdots = DN_i = 0$となり、

$$Y_{2t} = \beta X_{2t} + \alpha_2 + u_{2t}$$

長所：固定効果モデルは、固定効果α_iとX_{it}との相関を許容しており、両者に相関があっても推定量は一致性をもつ（頑健なモデル）

短所：時間を通じて一定の説明変数は、多重共線性のため、含めることはできない

例) 自殺と失業率の関係

- 64カ国の20年間(2000～2019年)のデータ

$$N \times T = 64 \times 20 = 1280$$

- Y_{it}： t年におけるi国の自殺者数(10万人当たり)

  X_{it}： t年におけるi国の失業率(%)

- プールドOLS推定すると、

$$\hat{Y}_{it} = \underset{(1.548)}{11.990} + \underset{(0.150)}{0.142} X_{it}$$

--- 係数βは有意ではない

- 固定効果モデルを推定すると

$$\hat{Y}_{it} = \underset{(0.069)}{0.255} X_{it} + \text{国効果}$$

--- 欠落変数バイアスを取り除くことでX_{it}の係数が正に有意となる

時間効果

$$Y_{it} = \alpha + \beta X_{it} + \theta S_t + u_{it}$$

---S_t : 時間を通じて変動するが、すべてのiに同じ影響を与える

- 時間効果: $\lambda_t = \alpha + S_t$

$$Y_{it} = \lambda_t + \beta X_{it} + u_{it}$$

- 時間固定効果モデル

$$Y_{it} = \beta X_{it} + \lambda_1 d1_t + \lambda_2 d2_t + \cdots + \lambda_T dT_t + u_{it}$$

--- $d1_t$はt = 1なら1、t≠ 1なら0となるダミー変数

--- $d2_t$はt = 2なら1、t≠ 2なら0となるダミー変数

----t = 1なら、$d1_t = 1$, $d2_t = d3_t = \cdots = dT_t = 0$となり、

$$Y_{i1} = \beta X_{i1} + \lambda_1 + u_{i1}$$

----t = 2なら、$d2_t = 1$, $d1_t = d3_t = \cdots = dT_t = 0$となり、

$$Y_{i2} = \beta X_{i2} + \lambda_2 + u_{i2}$$

- 個別効果と時間効果

$$Y_{it} = \alpha_i + \lambda_t + \beta X_{it} + u_{it}$$

- 個別効果と時間効果を考慮した固定効果モデル

$$Y_{it} = \beta X_{it} + \alpha_1 D1_i + \alpha_2 D2_i + \cdots + \alpha_N DN_i$$
$$+\lambda_1 d1_t + \lambda_2 d2_t + \cdots + \lambda_T dT_t + u_{i,t}$$

--- 多重共線性のため、ダミー変数のうち、どれか1つを削除する

$$D1_i + D2_i + \cdots + DN_i = 1$$
$$d1_t + d2_t + \cdots + dT_t = 1$$

例) コロナ感染症と人流との関係

・47都道府県のデータ (2020/1/6～2020/6/28)のデータ

$$N \times T = 47 \times 175 = 8225$$

・ Y_{it}： t日におけるi県のステイホーム指標(2019年に比べて、外出をどれぐらい控えていたかを表している)

--- 0%なら2019年と同程度の外出をしている

--- 100%なら全く外出をしていない

X_{it}： 学校閉鎖ダミー(学校閉鎖中は1となる)

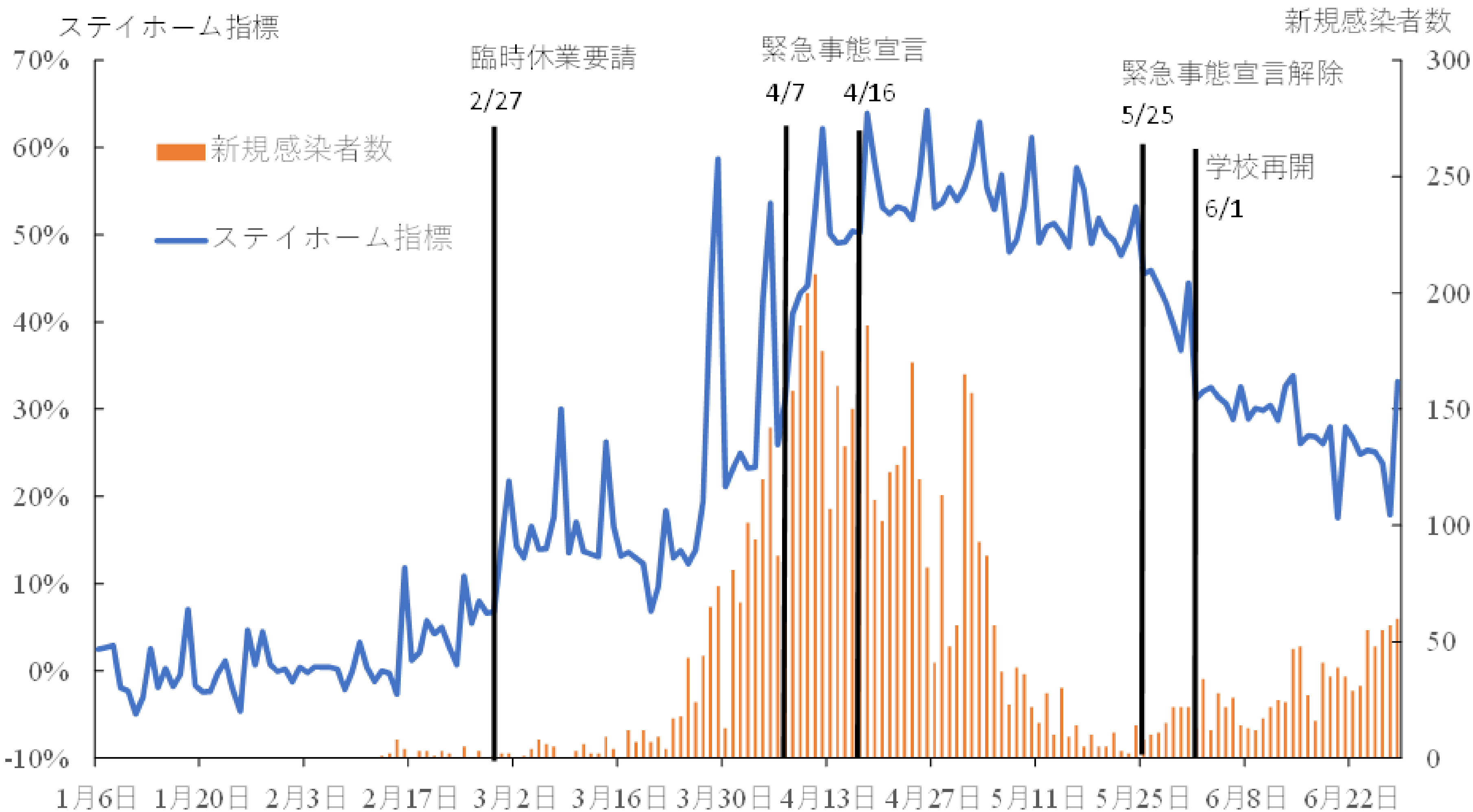
緊急事態宣言ダミー(緊急事態宣言は1となる)

新規感染者数の対数(サインの逆双曲線関数)

$$X^* = \ln(X + \sqrt{X^2 + 1})$$

雨ダミー(県庁所在地に雨が降ったら1となる)

図　11-1　東京都のステイホーム指標と新規感染者数



・ 個別効果と時間効果を考慮した固定効果モデル

$$\hat{Y}_{i,t} = \underset{(1.172)}{8.398}\text{学校閉鎖} + \underset{(0.745)}{8.155}\text{緊急事態} + \underset{(0.560)}{3.255}\text{感染者数} + \underset{(0.376)}{2.307}\text{雨}$$

--- 学校閉鎖すると8.398%だけステイホーム指標があがる

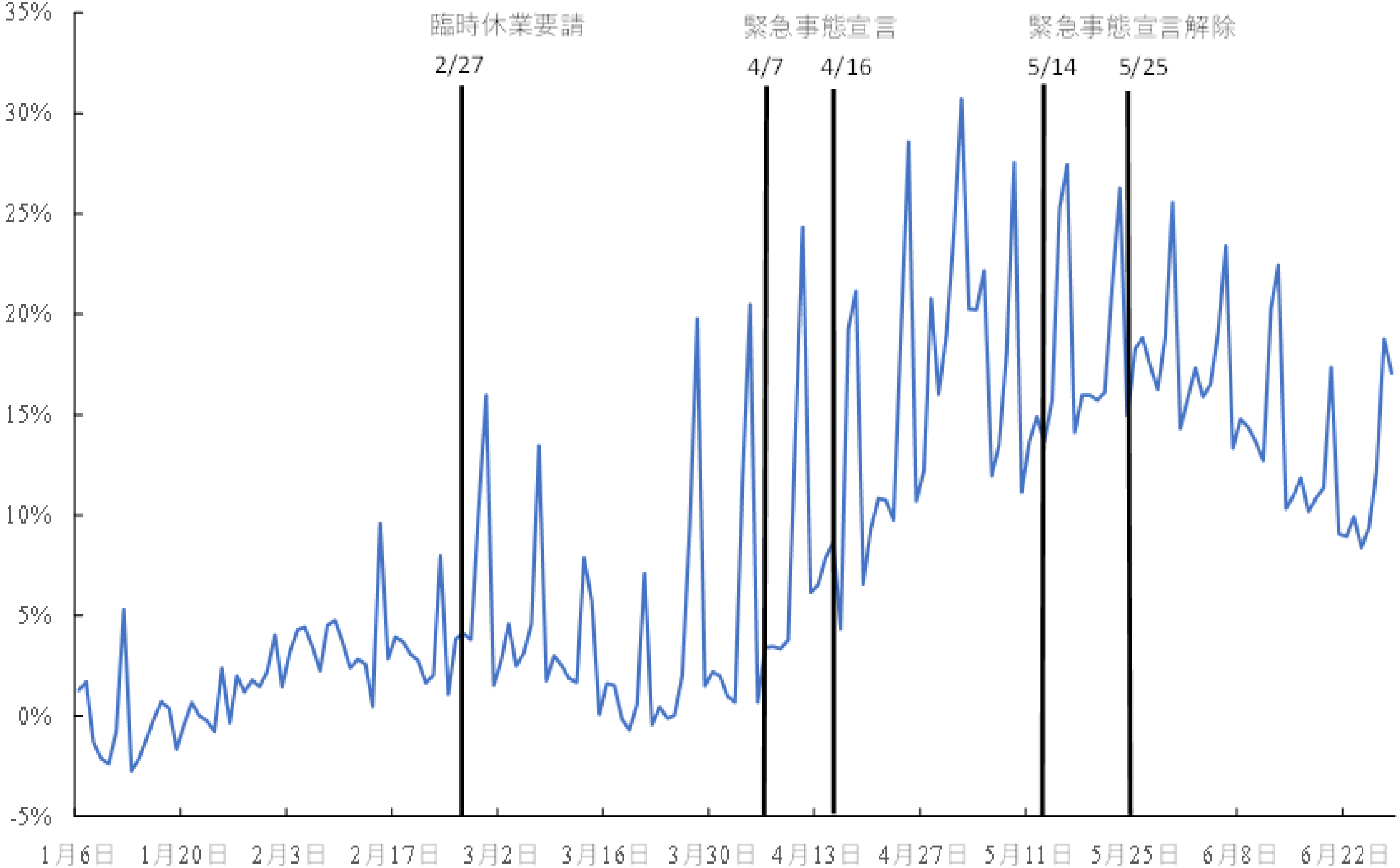
--- 緊急事態宣言中なら8.155%だけステイホーム指標があがる

--- 新規感染者数が1%上がると、$0.03255\% = 3.255 \times 0.01$だけステイホーム指標があがる

--- 緊急事態宣言と学校閉鎖の効果を合わせても約16.5の増加しか説明できない

--- それ以外の要因が重要(人々がニュースなどに反応して自発的に自粛していた)

図 11-2　時間効果の推移



変量効果と
ハウスマン検定

変量効果(Random Effect)モデル

- $Y_{it} = \beta X_{it} + e_{it}$

  変量効果モデルでは、誤差項は

$$e_{it} = \alpha_i + u_{it}$$

- α_iは以下を満たす

$$E[\alpha_i] = \alpha, \mathrm{Var}(\alpha_i) = \sigma_\alpha^2$$

$$\mathrm{cov}(\alpha_i, X_{it}) = 0$$（問題となる仮定）

- プールドOLSと違って、誤差構造を利用してFGLSを行う

- 長所: Nが増えても、推定するパラメータ($\beta, \alpha, \sigma_\alpha^2$)数は不変
- 短所: 変量効果が誤りなら、推定量は一致性をもたない

  一般的に、$\mathrm{cov}(\alpha_i, X_{it}) = 0$が満たされる状況は考え難い

ハウスマン検定

・モデル選択に用いる検定

H_0: 変量効果(説明変数と誤差項$e_{i,t} = \alpha_i + u_{i,t}$に相関なし)

H_1: 固定効果(説明変数と誤差項$e_{i,t} = \alpha_i + u_{i,t}$に相関あり)

・ H_0が正しいもとで、変量効果モデルの推定量 $\hat{\beta}_r$は一致性がある

H_1が正しいもとで、変量効果モデルの推定量$\hat{\beta}_r$は一致性がない

・ 固定効果モデルの推定量$\hat{\beta}_f$は常に一致性がある

・ 固定効果推定量$\hat{\beta}_f$と変量効果推定量$\hat{\beta}_r$を比較し、両者の値が大きく異なるならH_0(変量効果)を棄却する

・ 注意点: H_0(変量効果)の採択はH_0 が正しいことを意味しない

（固定効果モデルを使うことが推奨される）

11章のまとめ

- **パネルデータ**

  --- パネルデータとは、パネルデータの長所

- **個別効果**

  --- 固定効果、変量効果

- **固定効果モデル**

- **時間効果と個別効果を考慮した固定効果モデル**

- **変量効果モデルとハウスマン検定**