2023/07/31

藪友良

第 11 章の答え

練習問題 1

パネルデータは、NはTより大きいことが一般的である。しかし、NよりTの方が大きいケースもある。たとえば、株価の秒単位データを考えよう。秒単位のデータはTが非常に大きいため、TはNよりも大きくなる。

練習問題 2

変量効果は、個別効果が説明変数と無相関である一方、固定効果は、個別効果が説明変数と相関している点で異なる。通常、個別効果と説明変数は相関しているため、変量効果は非現実的である。たとえば、被説明変数Yを賃金、Xを教育年数とすると、個別効果として、生まれつきの能力が挙げられる。能力と教育年数は相関していると考えるのが自然である。

練習問題 3

固定効果モデルは、

$$Y_{i,t} = \beta X_{i,t} + \alpha_1 D1_i + \alpha_2 D2_i + \cdots + \alpha_N DN_i + u_{i,t}$$

となる。ここで、帰無仮説と対立仮説を次のように設定して F 検定をする。

$$H_0: \alpha_1 = \alpha、\ \alpha_2 = \alpha、\ \ldots、\ \alpha_N = \alpha$$

$$H_1: \text{帰無仮説} H_0 \text{は誤りである}$$

仮に帰無仮説H_0が採択されたなら、個別効果はないと判断される。これに対して、帰無仮説H_0が棄却されたなら、個別効果はあると判断される[1]。

なお、7.2 節と同じような仮説にしたいのであれば、上式を

$$\begin{aligned} Y_{i,t} &= \alpha_1 + \beta X_{i,t} + (\alpha_2 - \alpha_1) D2_i + \cdots + (\alpha_N - \alpha_1) DN_i + u_{i,t} \\ &= \alpha_1 + \beta X_{i,t} + \theta_2 D2_i + \cdots + \theta_N DN_i + u_{i,t} \end{aligned}$$

[1] 帰無仮説が棄却されても、プールド OLS に問題があるとはいえない。これは、個別効果が存在しても、説明変数と個別効果が無相関であれば、プールド OLS は一致性を持つためである。ただし、個別効果が存在するなら、個別効果が説明変数と相関する可能性を考慮し、固定効果モデルで推定することが望ましいといえる。

と表現し(ただし、$\theta_2 = \alpha_2 - \alpha_1$、…、$\theta_N = \alpha_N - \alpha_1$と定義した)[2]、仮説は次のように設定すればよい。

$$H_0: \theta_2 = 0、\ \theta_3 = 0、\ \ldots、\ \theta_N = 0$$

H_1: 帰無仮説H_0は誤りである

この仮説検定の結果から何が言えるだろうか。帰無仮説の棄却は、個別効果が存在し、プールド回帰が不適切であることを意味している。しかし、この検定からは、個別効果と説明変数が相関しているかは分からないため、変量効果モデルと個別効果モデルのどちらで推定すべきかわからない。

練習問題 4

2 時点のデータが利用可能であるため、両時点の差をとることで、

$$\begin{aligned} Y_{i,2} - Y_{i,1} &= \left(\alpha + \beta X_{i,2} + \gamma Z_i + u_{i,2}\right) - \left(\alpha + \beta X_{i,1} + \gamma Z_i + u_{i,1}\right) \\ &= \beta(X_{i,2} - X_{i,1}) + (u_{i,2} - u_{i,1}) \end{aligned}$$

となる。ここで、観察できない変数Z_iは式から消えるため、被説明変数を$Y_{i,2} - Y_{i,1}$、説明変数を$X_{i,2} - X_{i,1}$とすれば、通常の OLS で係数βをバイアスなく推定できる。

練習問題 5

固定効果モデルとして、次式を考えよう。

$$Y_{i,t} = \beta X_{i,t} + \theta W_i + \alpha_1 D1_i + \alpha_2 D2_i + \cdots + \alpha_N DN_i + u_{i,t}$$

ただし、W_iは時間を通じて一定の変数であるため、下添字はiだけとなる。また、θはその係数とする。ここで、W_iは、W_1、W_2、…、W_Nという値をとるとしよう(つまり、$i = 1$ならW_iはW_1となり、$i = 2$ならW_iはW_2となる)。

上式には、完全な多重共線性が生じるため、推定できないことを確認しよう。まず、W_iを、ダミー変数($D1_i$、$D2_i$、…、DN_i)を用いて、次のように表記する。

---

[2] 次式が正しいことを確認しよう。

$$\beta X_{i,t} + \alpha_1 D1_i + \alpha_2 D2_i + \cdots + \alpha_N DN_i + u_{i,t} = \alpha_1 + \beta X_{i,t} + (\alpha_2 - \alpha_1) D2_i + \cdots + (\alpha_N - \alpha_1) DN_i + u_{i,t}$$

まず、$i = 1$なら、$D1_i = 1$、$D2_i = D3_i = \cdots = DN_i = 0$なので、$\beta X_{i,t} + \alpha_1 + u_{i,t} = \alpha_1 + \beta X_{i,t} + u_{i,t}$となり、両辺は等しい。次に、$i = 2$なら、$D2_i = 1$、$D1_i = D3_i = \cdots = DN_i = 0$なので、$\beta X_{i,t} + \alpha_2 + u_{i,t} = \alpha_1 + \beta X_{i,t} + (\alpha_2 - \alpha_1) + u_{i,t}$となり、やはり両辺は等しい。他の$i$についても両辺が等しくなることを確認してほしい。

$$W_i = W_1 D1_i + W_2 D2_i + \cdots + W_N DN_i$$

次に、右辺の変数をすべて左辺に移項させると、

$$W_i - W_1 D1_i - W_2 D2_i - \cdots - W_N DN_i = 0$$

となる。ここで、W_1、W_2、…、W_Nは定数なので、$c_0 = 1$、$c_1 = -W_1$、$c_2 = -W_2$、…、$c_N = -W_N$とすると、

$$c_0 W_i + c_1 D1_i + c_2 D2_i + \cdots + c_N DN_i = 0$$

であり、完全な多重共線性が生じていることがわかる。

練習問題 6

時間固定効果モデルとして、次式を考えよう。

$$Y_{i,t} = \beta X_{i,t} + \theta W_t + \lambda_1 d1_t + \lambda_2 d2_t + \cdots + \lambda_T dT_t + u_{i,t}$$

ただし、W_tは個体iに対して同じ影響を与える変数であるため、下添字はtだけとなる。、ここで、W_tは、W_1、W_2、…、W_Tという値をとるとしよう(つまり、$t = 1$ならW_tはW_1となり、$t = 2$ならW_tはW_2となる)。

上式には、完全な多重共線性が生じるため、推定できないことを確認しよう。まず、W_tを、ダミー変数($d1_t$、$d2_t$、…、dT_t)を用いて、次のように表記する。

$$W_t = W_1 d1_t + W_2 d2_t + \cdots + W_T dT_t$$

次に、右辺の変数をすべて左辺に移項させると

$$W_t - W_1 d1_t - W_2 d2_t - \cdots - W_T dT_t = 0$$

となる。ここで、W_1、W_2、…、W_Tは定数であるため、完全な多重共線性が生じていることがわかる。

練習問題 7

固定効果と時間効果を考慮したモデルは、次のモデルとなる。

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

問題 5 から、時間を通じて一定の変数W_iは、

$$W_i = W_1 D1_i + W_2 D2_i + \cdots + W_N DN_i$$

となる。同様に、問題 6 から、個体iに対して同じ影響を与える変数W_tは、

$$W_t = W_1 d1_t + W_2 d2_t + \cdots + W_T dT_t$$

となる。これらに完全な多重共線性が生じていることは明らかであろう。

練習問題 8

どちらも同じモデルを考えている。

$$Y_{i,t} = \alpha + \beta X_{i,t} + e_{i,t}$$

$$e_{i,t} = Z_i + u_{i,t}$$

プールド OLS は、通常の OLS 推定である。これに対して、変量効果モデルによる推定は、誤差項$e_{i,t} = Z_i + u_{i,t}$の系列相関を除くため、被説明変数$Y_{i,t}$と説明変数$X_{i,t}$に複雑な変換を行ったうえで OLS 推定する(つまり、一般化最小 2 乗推定となる)。

変量効果の前提が正しいならば、変量効果推定量は BLUE となるが、プールド OLS は不偏性と一致性はあるが BLUE ではない。固定効果の前提が正しいならば、変量効果推定量とプールド OLS 推定量はどちらもバイアスを持った推定量である。

練習問題 9

ウェブサイトから、再現に必要なデータと STATA の do file をダウンロードできる。STATA 再現コードは以下のとおり。

STATA の再現コード

```
**　データの読み込み
use suicide_data.dta


**　パネルデータ
des
encode country_name, gen(country)
tabulate country, nolabel
sum
xtset country year
```

```
** 11.3.1 節の推定
reg suicide unemployment, vce(cluster country)
display e(r2_a)
reg suicide unemployment i.country, vce(cluster country)
display e(r2_a)

** 11.4.2 節の推定
reg suicide unemployment i.country i.year, vce(cluster country)
display e(r2_a)
```

① データを読み込む。

```
use suicide_data.dta
```

country_name が文字列(str18)であることを des コマンドで確認する。そして、country_name を値に変換したのが country になる。

```
encode country_name, gen(country)
```

そして、パネルデータの i と t を定義する。

```
xtset country year
```

② プールド回帰は、以下のようにする。

```
reg suicide unemployment, vce(cluster country)
```

vce(cluster country)は、クラスターロバスト標準誤差を計算するためのオプションである。また、自由度調整済み決定係数を計算するため、

```
display e(r2_a)
```

としよう。次に、固定効果モデルは、

```
reg suicide unemployment i.country, vce(cluster country)
display e(r2_a)
```

とすればよい[3]。最後に、時間効果を入れるなら、

```
reg suicide unemployment i.country i.year, vce(cluster country)
```

とすればよい。i.year が時間効果にあたる。

---

[3] STATA では、xtreg コマンドがあり、固定効果モデルなら、

```
xtreg suicide unemployment, fe r
```

として推定できる。しかし、このコマンドは標準誤差の計算に問題があることが知られている。https://economics.mit.edu/faculty/acemoglu/data/aj2007

練習問題 10

ウェブサイトから、再現に必要なデータと STATA の do file をダウンロードできる。練習問題 9 と同じようなコードなので、自分で再現してみよう。

練習問題 11

ウェブサイトから、再現に必要なデータと STATA の do file をダウンロードできる。STATA 再現コードは以下のとおり。

STATA の再現コード

```
use stayhome_data.dta


des
encode prefecture, gen(pref)
tabulate pref, nolabel
sum
xtset pref time


gen drain = rain>0
gen linfection = ln(infection+sqrt(infection^2+1))
gen emerg = emerg_start-emerg_end
gen close = close_start-close_end
reg stay close emerg linfection drain i.pref i.time [aw=mobilephones], cluster(pref)
display e(r2_a)
```

① データを読み込む。

```
use stayhome_data.dta
```

prefecture が文字列であることを des コマンドで確認する。そして、prefecture を値に変換したのが pref になる。

```
encode prefecture, gen(pref)
```

そして、パネルデータの i と t を定義する。

```
xtset pref time
```

② データを定義する。雨ダミーは、

```
gen drain = rain>0
```

となる(下記の STATA アドバイス参照)。また、新規感染者数 infection は 0 をとる変数なので、対数を求めることはできない。このため、6 章の補足で学んだ逆双曲線関数を用いる。

```
gen linfection = ln(infection+sqrt(infection^2+1))
```

また、緊急事態宣言ダミーと学校閉鎖ダミーを定義する。

```
gen emerg = emerg_start-emerg_end
gen close = close_start-close_end
```

たとえば、emerg_start は、緊急事態宣言が始まってから 1 となるダミー変数、emerg_end は緊急事態宣言が終わったら 1 となるダミー変数である。よって、緊急事態宣言中に 1 となるダミー変数は、emerg_start-emerg_end となる。

③ 時間効果と固定効果を考慮したモデルによる推定は、次のとおり。ここで、加重最小 2 乗法を用いるため、[aw=mobilephones]としている。なお、mobilephones は携帯電話保有者数。

```
reg stay close emerg linfection drain i.pref i.time [aw=mobilephones], cluster(pref)
```

欠損値がある場合について

この場合、rain に欠損値はないので問題は生じないが、仮に欠損値があるならば、

```
gen drain = rain>0
```

としてしまうと、欠損値も 1 に変換されるという問題が生じる。とくに個票データでは、欠値値のある変数は多く、注意が必要だろう。仮に欠損値があれば、次のようにすれば、欠損していないデータに対してダミー変数を作成できる(「.」は欠損値であることを意味する)。

```
gen drain = (rain > 0) if rain !=.
```

この問題を理解するため、rain に欠損値のある変数 test_rain を作成しよう。

```
gen test_rain = rain
replace test_rain = . if pref == 1
```

ここで、pref=1の場合に、rainが欠損値になる変数がtest_rainとなる。そして、ダミー変数dummyを作成する。

```
gen dummy = test_rain > 0
```

dummyは、test_rainが欠損値であるときにも1になっていることを確認してほしい。これは次のようなコマンドでも確認できる。

```
tabulate dummy if missing(test_rain)
```

この問題を解決するには、次のようにすればよい。

```
gen dummy = (test_rain > 0) if test_rain !=.
```