

xtreg - 線形回帰モデル 【 評価版 】

xtreg コマンドはパネルデータに対して線形回帰モデルをフィットさせる機能を提供します。FE/RE/BE/PA といったいくつかの推定法が用意されています。

Stata 19 では次の機能が新たに加われました。

- Mundlak 回帰を用いた相関変数効果モデル (cre オプション)
- 高次元カテゴリ変数への対応 (absorb オプション)

- | | |
|----------------|-----------|
| 1. パネル用線形回帰モデル | |
| 2. フィット適合度の評価 | |
| 3. xtreg の用例 | Example 1 |
| | Example 2 |
| | Example 3 |
| | Example 4 |
| | Example 5 |
| | Example 6 |
| | Example 7 |
| | Example 8 |

1. パネル用線形回帰モデル

固定効果モデル (fixed-effects models) と変数効果モデル (random-effects models) の機能概要と用例については Baltagi (2013), Wooldridge (2020), Allison (2009) を参照ください。

今、次のようなモデルのフィットについて考えることにします。

$$y_{it} = \alpha + \mathbf{x}_{it}\beta + \nu_i + \epsilon_{it} \quad (1)$$

このモデル式において β は推定対象のパラメータベクトルを、 $\nu_i + \epsilon_{it}$ は誤差項を表します。特徴的なのは ν_i でユニット (パネル) 固有の誤差項を意味します。それはユニット i ごとに異なる値を取りますが、時間 t には依存しないものである点に注意してください。

これに対し ϵ_{it} は通常の誤差項で、

- 平均が 0
- 分散値は均一 (homoskedastic)
- 自身、及び x と ν とは無相関

という通常の特徴を持ちます*1。

推定のために必要となる仮定について話を進める前に (1) 式の変形を行っておきます。 ν_i と ϵ_{it} の特徴がどうであれ、(1) が真であれば、

$$\bar{y}_i = \alpha + \bar{x}_i\beta + \nu_i + \bar{\epsilon}_i \quad (2)$$

も真であると言えます。ただし $\bar{y}_i = \sum_t y_{it}/T_i$, $\bar{x}_i = \sum_t x_{it}/T_i$, $\bar{\epsilon}_i = \sum_t \epsilon_{it}/T_i$ を意味します。(1) 式から (2) 式を引くことによって ν_i を含まない

$$(y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i)\beta + (\epsilon_{it} - \bar{\epsilon}_i) \quad (3)$$

という式を得ることができます。

これら 3 つの数式に基づき β の推定が行われます。最初にまず `xtreg, fe` ですが、これは固定効果推定法 (FE: fixed-effects estimator) と呼ばれ*2、(3) 式に OLS (通常の線形回帰) を適用することによって β を推定します。一方、between estimator (BE) として知られる `xtreg, be` は (2) 式に OLS を適用する形で推定を行います。これに対し変量効果推定法 (RE: random-effects estimator) と呼ばれる `xtreg, re` は BE の推定値と FE の推定値の加重平均を取る形で β の推定を行います。より具体的に言うなら RE は

$$(y_{it} - \theta_i \bar{y}_i) = (1 - \theta_i)\alpha + (x_{it} - \theta_i \bar{x}_i)\beta + \{(1 - \theta_i)\nu_i + (\epsilon_{it} - \theta_i \bar{\epsilon}_i)\} \quad (4)$$

という式を用いて β を推定します。ただし θ_i は $\sigma_\nu^2, \sigma_\epsilon^2, T_i$ の関数です。 $\sigma_\nu^2 = 0$ の場合 — $\theta_i = 0$ (for all i) に対応 — には ν_i は常に 0 となるため、(1) 式は直接 OLS で推定できることとなります。一方、 $\sigma_\epsilon^2 = 0$ の場合 — $\theta_i = 1$ (for all i) に対応 — には ϵ_{it} が 0 となるため、FE 推定法 ((3) 式) によってすべての情報がもたらされることとなります。

評価版では割愛しています。

2. フィット適合度の評価

評価版では割愛しています。

*1 ϵ_{it} をさらに $v_t + w_{it}$ のように分解したモデルを想定することもあります。ただしこの場合には w_{it} が通常の誤差項を表すこととなります。

*2 Within estimator と呼ばれることもあります。

3. xtreg の用例

本 whitepaper では Example データセット `nlswork.dta` を使用します。

```
. use https://www.stata-press.com/data/r19/nlswork.dta *3
```

(National Longitudinal Survey of Young Women, 14-24 years old in 1968)

これは米国における National Longitudinal Survey のデータで、4,711 人の女性労働者に関するデータが 1968 年から 1988 年にわたって追跡調査されています。データセットはパネル変数を `idcode`、時間変数を `year` とする形で `xtset` 済みです。

```
. xtset
```

```
. xtset
Panel variable: idcode (unbalanced)
Time variable: year, 68 to 88, but with gaps
Delta: 1 unit
```

多数の変数が含まれていますが、ここでは分析に使用する変数についてのみその意味を確認しておきます。

```
. describe ln_wage grade age ttl_exp tenure race not_smsa south *4
```

```
. describe ln_wage grade age ttl_exp tenure race not_smsa south
```

Variable name	Storage type	Display format	Value label	Variable label
<code>ln_wage</code>	float	%9.0g		In(wage/GNP deflator)
<code>grade</code>	byte	%8.0g		Current grade completed
<code>age</code>	byte	%8.0g		Age in current year
<code>ttl_exp</code>	float	%9.0g		Total work experience
<code>tenure</code>	float	%9.0g		Job tenure, in years
<code>race</code>	byte	%8.0g	race1b1	Race
<code>not_smsa</code>	byte	%8.0g		1 if not SMSA
<code>south</code>	byte	%8.0g		1 if south

*3 メニュー操作 : File ▸ Example Datasets ▸ Stata 19 manual datasets と操作、Longitudinal-Data/Panel-Data Reference Manual [XT] の `xtreg` の項よりダウンロードする。

*4 メニュー操作 : Data ▸ Describe data ▸ Describe data in memory or in a file

従属変数として使用するのは時給の対数値を表す `ln_wage` です。これに対し回帰変数としては次の7つを使用します。

回帰変数	内容
<code>grade</code>	最終学歴 [0-18]
<code>age</code>	調査時点での年齢 [14-46]
<code>ttl_exp</code>	就業経験年数
<code>tenure</code>	在職年数
<code>race</code>	人種 (1: 白人、2: 黒人、3: その他)
<code>not_smsa</code>	都市部非在住者* ⁵
<code>south</code>	米国南部諸州在住者

なお、以下の用例中では人種に関し `2.race` という指定を行っているので、`race` は2値変数 — 黒人が否か — という形で使用されることになります。また、変数 `age`, `ttl_exp`, `tenure` についてはその2乗項 — 例えば `c.age#c.age` — も回帰変数として含める形でモデルの設定を行います。なお、`2.` とか `c.` といった演算子の意味についてはCS版解説書BR01の因子変数の項をご参照ください。

▷ Example 1: BE モデル

評価版では割愛しています。

▷ Example 2: FE モデル

評価版では割愛しています。

▷ Example 3: FE モデル + `vce(robust)` オプション

評価版では割愛しています。

▷ Example 4: RE モデル (GLS)

評価版では割愛しています。

▷ Example 5: RE モデル (ML)

評価版では割愛しています。

▷ Example 6: PA モデル

評価版では割愛しています。

*⁵ SMSA: standard metropolitan statistical area

▷ Example 7: CRE モデル

Example 2 では FE モデルによる推定を行いました。FE 推定法の場合、`grade` や `race` といった時間的変化のない回帰変数については係数値の推定が行えませんでした。このような変数についてもそれらの効果を推定したい場合には `cre` オプションを指定し、相関変数効果モデル (correlated random-effects model) をフィットさせるのも一法です。

評価版では割愛しています。

▷ Example 8: 高次元カテゴリ変数への対応

Example データセット `nlswork.dta` を使用してフィットを行っている労働市場モデルにおいて、郡 (counties) の効果を加味したいとします。この場合、米国内には約 3,200 の郡が存在しているわけですから、単に

```
. xtreg ... i.county, fe
```

と操作したのでは 3,200 個の指標変数 (ダミー変数) が生成されてしまう形となり、計算時間もさることながら、その出力量は膨大なものとなってしまいます。郡ごとの係数推定値には関心がないとするなら、その効果を吸収する (`absorb`) 形で効率良く推定を行うことが可能です。それが `absorb()` オプションです。このオプションが指定された場合にはパネルごとの効果のみならず郡ごとの効果も吸収された形で推定が行われることになります。

評価版では割愛しています。

■