

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。

- | | |
|----------------|-----------|
| 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 |

1. パネル用線形回帰モデル

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 \mathbf{x} と ν とは無相関

という通常の特性を持ちます*1。

© Copyright Math 工房； 一部 © Copyright StataCorp LP (used with permission)

*1 ϵ_{it} をさらに $\nu_t + w_{it}$ 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。

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} の特性がどうであれ、(1) が真であれば、

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

も真であると言えます。ただし $\bar{y}_i = \sum_t y_{it}/T_i$, $\bar{\mathbf{x}}_i = \sum_t \mathbf{x}_{it}/T_i$, $\bar{\epsilon}_i = \sum_t \epsilon_{it}/T_i$ 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

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

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

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

という式を用いて β を推定します。ただし θ は σ_ν^2 と σ_ϵ^2 の関数です。 $\sigma_\nu^2 = 0$ の場合 — $\theta = 0$ 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 — $\theta = 1$ 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

R^2 は通常の回帰においてよく用いられるフィット適合度 (goodness of fit) の指標です。xtreg の場合、 α と β に対する推定値である $\hat{\alpha}$ と $\hat{\beta}$ が得られたとするなら、(1), (2), (3) 式に関連して適合度の評価を行うことができます。それぞれ予測値に関する方程式を書くと次のようになります。

$$\hat{y}_{it} = \hat{\alpha} + \mathbf{x}_{it}\hat{\beta} \quad (1c)$$

$$\hat{\bar{y}}_i = \hat{\alpha} + \bar{\mathbf{x}}_i\hat{\beta} \quad (2c)$$

$$\hat{\hat{y}}_{it} = (\hat{y}_{it} - \hat{\bar{y}}_i) = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\hat{\beta} \quad (3c)$$

xtreg からはこれら 3 つの方程式に対応する “R-squares” がレポートされてきます。ただし “” が付いている点に注意してください。R-squares は OLS の場合の R^2 値の持つ特質をすべて備えているわけではありません。

評価版では割愛しています。

^{*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 http://www.stata-press.com/data/r17/nlswork.dta *3
```

(National Longitudinal Survey. Young Women 14-26 years of age 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	<code>%9.0g</code>		ln(wage/GNP deflator)
<code>grade</code>	byte	<code>%8.0g</code>		Current grade completed
<code>age</code>	byte	<code>%8.0g</code>		Age in current year
<code>ttl_exp</code>	float	<code>%9.0g</code>		Total work experience
<code>tenure</code>	float	<code>%9.0g</code>		Job tenure, in years
<code>race</code>	byte	<code>%8.0g</code>	<code>racelbl</code>	Race
<code>not_smsa</code>	byte	<code>%8.0g</code>		1 if not SMSA
<code>south</code>	byte	<code>%8.0g</code>		1 if south

*3 メニュー操作 : File ▸ Example Datasets ▸ Stata 17 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となります。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 *mwp-028* 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