

## 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. FE 推定法         | 用例 1 |
| 3. LSDV 推定法       |      |
| 4. RE 推定法         | 用例 2 |
| 5. BE 推定法         | 用例 3 |
| 6. Pooled OLS 推定法 | 用例 4 |
| 7. PA 推定法         | 用例 5 |
| 8. Hausman 検定     |      |
| 9. モデルによる予測       |      |
| 補足 1              |      |

## 1. パネル用線形回帰モデル

xtreg では

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

で与えられるような個体効果モデル (individual effect model) を考察対象とします。その中の誤差項  $\epsilon_{it}$  は通常の OLS (ordinary least squares) の場合と同様、

$$E(\epsilon_{it}) = 0$$

$$E(\epsilon_{it}, \epsilon_{js}) = \begin{cases} \sigma_\epsilon^2 & (i = j \text{ and } t = s) \\ 0 & (\text{otherwise}) \end{cases}$$

という条件を満たすものとします。OLS のモデルと異なるのは個体効果  $\nu_i$  が加わっている点にあります ( $\nu_i$  は時間  $t$  に非依存である点に注意)。この個体効果をどう想定するかによって固定効果モデル (FE: fixed effect model) と変量効果モデル (RE: random effect model) の 2 種類が考えられます。

## (1) 固定効果モデル

このモデルの場合、 $\nu_i$  は推定可能な固定値として扱われます。このため  $N$  の数が大きいパネルデータにおいては推定の自由度が失われることとなります。なお  $\nu_i$  と回帰変数  $x_{it}$  との間には相関があっても構わないという点が変量効果モデルの場合と異なります。

## (2) 変量効果モデル

このモデルの場合には  $\nu_i$  をランダムな確率変数としてとらえ推定を行います。従って  $\nu_i$  と回帰変数  $x_{it}$  との間には相関がないことが前提となります。



個体効果  $\nu_i$  と回帰変数  $x_{it}$  との間に相関がある場合に変量効果モデル用の推定法を実行すると、 $\beta$  の推定値は一致性を持たない (inconsistent) ものとなる点に注意してください。



切片項のみならず傾きについても個体効果を考慮できる混合効果モデルについては [ME] mixed (*mwp-018*) をご参照ください。

パネルデータの解析に際しては最初に固定効果 (FE) モデル、変量効果 (RE) モデルのどちらを前提とするかを判断する必要があります。その上で適切な推定法を用いて解析を進めて行くわけですが、選択できる推定法には次のようなものがあります。

モデル	推定法
固定効果モデル	<ul style="list-style-type: none"> <li>○ FE 推定法 (within estimator)</li> <li>○ LSDV 推定法</li> </ul>
変量効果モデル	<ul style="list-style-type: none"> <li>○ RE 推定法</li> <li>○ BE 推定法 (between estimator)</li> <li>○ Pooled OLS 推定法</li> <li>○ PA 推定法</li> </ul>

## 2. FE 推定法

FE 推定法は mean-differencing という手法により個体効果  $\nu_i$  を除去した上で OLS 推定を実行し、 $\beta$  の推定を行います。すなわち個体効果モデルのモデル式 (1) に対して時間平均を取ると

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

が得られます (ただし  $\bar{y}_i = \sum_t y_{it}/T_i$ ,  $\bar{x}_i = \sum_t x_{it}/T_i$ ,  $\bar{\epsilon}_i = \sum_t \epsilon_{it}/T_i$  とする)。これを (1) 式から引くことにより  $\nu_i$  を含まない形の

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

を得ることができます。この (3) 式に OLS を適用し推定を行うものが FE 推定法です。グループ内平均からの偏差を使って推定を行うことから within estimator と呼ばれることもあります。なお、時間的に変化しない

回帰変数（説明変数）の場合には  $x_{it} - \bar{x}_i$  の値がすべて 0 となってしまうため、係数値の推定は行えない点に注意してください。

FE 推定法の特徴を整理しておく次のようになります。

- $\nu_i$  が  $x_{it}$  と相関を持っている場合であっても  $\beta$  に関する consistent な推定値が得られる。
- 時間的に変化しない回帰変数については係数値の推定が行えない。
- 時間的な変動の小さな回帰変数の場合、推定結果は不正確なものとなる。



個体ごとの  $\nu_i$  の値は predict コマンドを用いることによって算出することができます ([XT] xtreg postestimation (*mwp-167*) 参照)。

## ▷ 用例 1: FE 推定法

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/r15/nlswork.dta *1
```

(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 *2
```

```
. describe ln_wage grade age ttl_exp tenure
+-----+-----+-----+-----+-----+
|          | storage | display | value |          |
| variable | type    | format  | label | variable |
| name     |         |         |      | label    |
+-----+-----+-----+-----+-----+
| ln_wage  | float   | %9.0g   |      | ln(wage/GNP deflator) |
| grade    | byte    | %8.0g   |      | current grade completed |
| age      | byte    | %8.0g   |      | age in current year |
| ttl_exp  | float   | %9.0g   |      | total work experience |
| tenure   | float   | %9.0g   |      | job tenure, in years |
+-----+-----+-----+-----+-----+
```

\*1 メニュー操作 : File ▷ Example Datasets ▷ Stata 15 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。

\*2 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時は給の対数値を表す `ln_wage` です。これに対し回帰変数としては次の4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>t11_exp</code>	就業経験年数
<code>tenure</code>	在職年数

それでは FE 推定法によるフィットを実行してみます。

- Statistics ▸ Longitudinal/panel data ▸ Linear models ▸ Linear regression (FE, RE, PA, BE) と操作
- Model タブ: Dependent variable: `ln_wage`  
Independent variables: `grade age t11_exp tenure`  
Model type: Fixed-effects

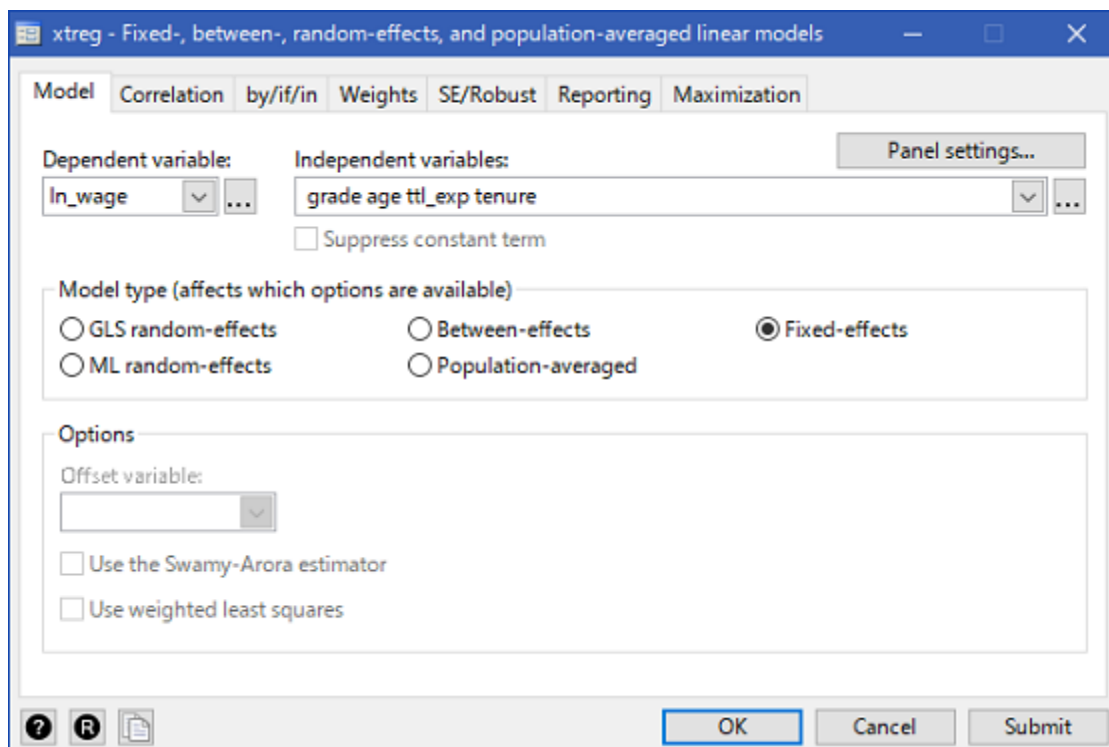


図 1 xtreg ダイアログ – Model タブ

```

. xtreg ln_wage grade age ttl_exp tenure, fe
note: grade omitted because of collinearity

Fixed-effects (within) regression           Number of obs   =    28,099
Group variable: idcode                     Number of groups =     4,697

R-sq:                                       Obs per group:
    within = 0.1443                          min =          1
    between = 0.2745                          avg =          6.0
    overall = 0.1924                          max =          15

corr(u_i, Xb) = 0.1651                       F(3,23399)      =    1315.26
                                           Prob > F        =     0.0000

```

ln_wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
grade	0 (omitted)					
age	-.0030427	.0008644	-3.52	0.000	-.0047369	-.0013484
ttl_exp	.029036	.0014505	20.02	0.000	.026193	.031879
tenure	.0116574	.0009249	12.60	0.000	.0098444	.0134704
_cons	1.547951	.0181798	85.15	0.000	1.512317	1.583584
sigma_u	.3751722					
sigma_e	.29556813					
rho	.61703248 (fraction of variance due to u_i)					

F test that all u\_i=0: F(4696, 23399) = 7.64                      Prob > F = 0.0000

xtreg からの出力において以下の点に注意してください。

- 最終学歴 grade は賃金を規定する重要なファクタと考えられるが、時間的に変化しない変数であるため、その効果は FE 推定法では推定できない。このため  $\nu_i$  とレポートされている情報の中には本来の個体効果の他に grade の効果も含まれている点に注意。
- $\sigma_\nu$  と  $\sigma_\epsilon$  の推定値がそれぞれ sigma\_u, sigma\_e として出力されている。
- $\nu_i$  がすべて 0 であるとする検定に対する p 値は 0.0000 とレポートされており（出力末尾）帰無仮説は棄却される。
- ヘッダ部には  $\text{Corr}(\nu_i, x_{it})$  の推定値が 0.1651 とレポートされている。

ここで  $\nu_i$  の推定値についてチェックしておきます。それには predict コマンドを使用します。

- Statistics ▷ Postestimation ▷ Predictions ▷ Linear predictions, residuals, error components ▷ Launch と操作
- predict ダイアログ: Main タブ: New variable name: u\_fe  
Produce: Fixed or random error component (u\_i)

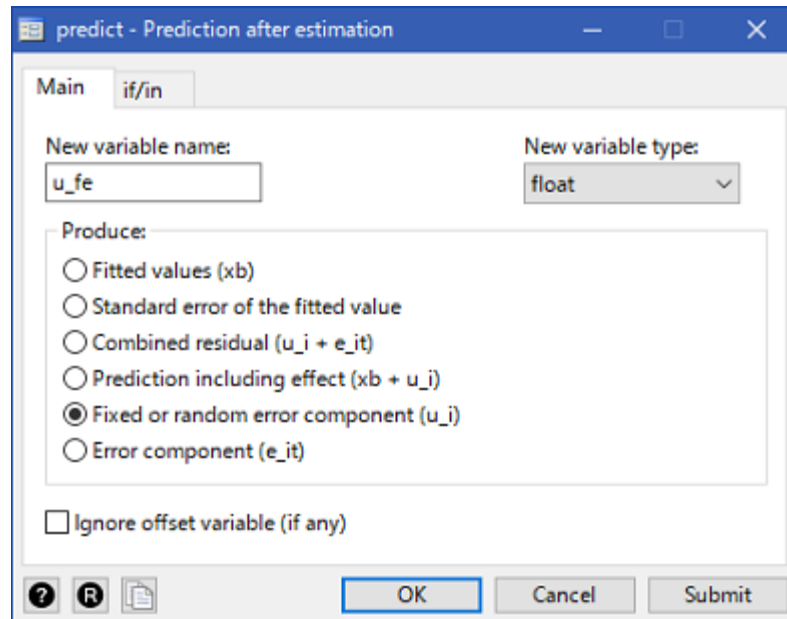


図2 predict ダイアログ – Main タブ

```
. predict u_fe, u
(435 missing values generated)
```

算出された `u_fe` の平均値と標準偏差を `xtsum` コマンドにより計算してみると次のようになります。

```
. xtsum u_fe *3
```

```
. xtsum u_fe
```

Variable		Mean	Std. Dev.	Min	Max	Observations
u_fe	overall	2.67e-10	.3387662	-1.764374	2.337403	N = 28099
	between		.3751722	-1.764374	2.337403	n = 4697
	within		0	2.67e-10	2.67e-10	T-bar = 5.98233

$E(\nu_i) = 0$  である点に注意してください。また `xtreg` 出力中で `sigma_u` としてレポートされている値は、級間変動に対応する標準偏差 0.3751722 であることがわかります。

最後に `xtreg` の推定結果を `est_fe` という名前で退避しておきます。これは後続のセクションにおいて推定結果の比較を行う際に利用します。

```
. estimates store est_fe *4
```

◀

\*3 メニュー操作：Statistics > Longitudinal/panel data > Setup and utilities > Summarize xt data

\*4 メニュー操作：Statistics > Postestimation > Manage estimation results > Save current estimates in memory > Launch

### 3. LSDV 推定法

評価版では割愛しています。

### 4. RE 推定法

評価版では割愛しています。

### 5. BE 推定法

評価版では割愛しています。

### 6. Pooled OLS 推定法

評価版では割愛しています。

### 7. PA 推定法

評価版では割愛しています。

### 8. Hausman 検定

評価版では割愛しています。

### 9. モデルによる予測

評価版では割愛しています。

### 補足 1 – グラフ作成コマンド操作

評価版では割愛しています。

