

xtgee - 一般化推定方程式 【 評価版 】

xtgee はパネルデータを対象に一般化線形モデルのフィットを行います。

1. 一般化線形モデル

2. xtgee の用例

Example 1

Example 2

1. 一般化線形モデル

xtgee はパネルデータを対象にして次のような一般化線形モデル (GLM: generalized linear models) のフィットを行います。

$$g\{E(y_{it})\} = \mathbf{x}_{it}\boldsymbol{\beta}, \quad y \sim F \text{ (パラメータ}\theta_{it}\text{を含む)} \quad (1)$$

ただし $i = 1, \dots, m, t = 1, \dots, n_i$ とします。 $g()$ はリンク関数 (link function) を、 F は分布族 (distributional family) を表します。種々の定義を $g()$ と F に当てはめることにより多様なモデルが構成されることとなります。例えば y_{it} が正規分布に従い、 $g()$ を恒等関数 (identity function) とした場合にはモデル式は

$$E(y_{it}) = \mathbf{x}_{it}\boldsymbol{\beta}, \quad y \sim N() \quad (2)$$

となり、どのような相関構造を前提にするかによって線形回帰モデル、変量効果回帰モデル等が表現されることとなります。

y_{it} がベルヌーイ分布 (二項分布) に従い、 $g()$ をロジット関数とした場合には

$$\text{logit}\{E(y_{it})\} = \mathbf{x}_{it}\boldsymbol{\beta}, \quad y \sim \text{Bernoulli} \quad (3)$$

すなわちロジスティック回帰が誘導されます。また y_{it} の分布をポアソン分布とし、 $g()$ を自然対数とした場合には

$$\ln\{E(y_{it})\} = \mathbf{x}_{it}\boldsymbol{\beta}, \quad y \sim \text{Poisson} \quad (4)$$

という形でポアソン回帰が導かれます。もちろんこれら以外の組合せも想定可能です。



xtgee は一般化推定方程式 (GEE: generalized estimating equations) という手法を用いて GLM のフィットを行います。

評価版では割愛しています。

2. xtgee の用例

以下においては xtgee と他の推定コマンドとの対応関係を具体例によって示します。確かに xtgee は他のコマンドを一般化するものと言えますが、推定手法には違いがあります。従って得られる結果は必ずしも一致するとは限らない点に注意してください。

▷ Example 1

[XT] xtgee の Example 1 には Example データセット nlswork2.dta を用いた用例が紹介されています。

```
. use https://www.stata-press.com/data/r19/nlswork2.dta *1
(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、3,914 人の女性労働者に関するデータが 1968 年から 1978 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 78, but with gaps
Delta: 1 unit
```

Unbalanced と表示されていることから、調査が行われた年度が個人によって異なることが推察できます。データセット中には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 *2
```

```
. describe ln_wage grade age
```

Variable name	Storage type	Display format	Value label	Variable 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

*¹ メニュー操作 : File ▷ Example Datasets ▷ Stata 19 manual datasets と操作、Longitudinal-Data/Panel-Data Reference Manual [XT] の xtgee 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

それぞれの変数の意味は次の通りです。

変数	内容
ln_wage	時給の対数値
grade	最終学歴 [0-18]
age	調査時点での年齢 [14-46]

ここでは ln_wage を grade, age, age² によって説明する線形回帰モデルを考えることにします。最初に regress コマンドによってフィットを行って見ますが、この場合、パネル構造は考慮されずにモデルフィットが行われることになる点に注意してください。すなわち同一の idcode を持った観測データ (observations) であっても独立であることが仮定されるわけです。

```
. regress ln_wage grade age c.age#c.age *3
```

. regress ln_wage grade age c.age#c.age						
Source	SS	df	MS	Number of obs	=	16,085
Model	597.54468	3	199.18156	F(3, 16081)	=	1413.68
Residual	2265.74584	16,081	.14089583	Prob > F	=	0.0000
				R-squared	=	0.2087
				Adj R-squared	=	0.2085
Total	2863.29052	16,084	.178021047	Root MSE	=	.37536
ln_wage	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
grade	.0724483	.0014229	50.91	0.000	.0696592	.0752374
age	.1064874	.0083644	12.73	0.000	.0900922	.1228825
c.age#c.age	-.0016931	.0001655	-10.23	0.000	-.0020174	-.0013688
_cons	-.8681487	.1024896	-8.47	0.000	-1.06904	-.6672577

*3 メニュー操作 : Statistics ▸ Linear models and related ▸ Linear regression なお二乗項の表現のしかたについては *mwp-028* を参照

同じ結果を `xtgee` によって得るためには次のように操作します。キーとなるのは `corr(independent)` という設定です。

- Statistics ▸ Longitudinal/panel data ▸ Generalized estimating equations (GEE)
 - Generalized estimating equations (GEE) と操作
- Model タブ: Dependent variable: `ln_wage`
 Independent variables: `grade age c.age#c.age`
 Family and link choices: (Gaussian, Identity) (デフォルト)

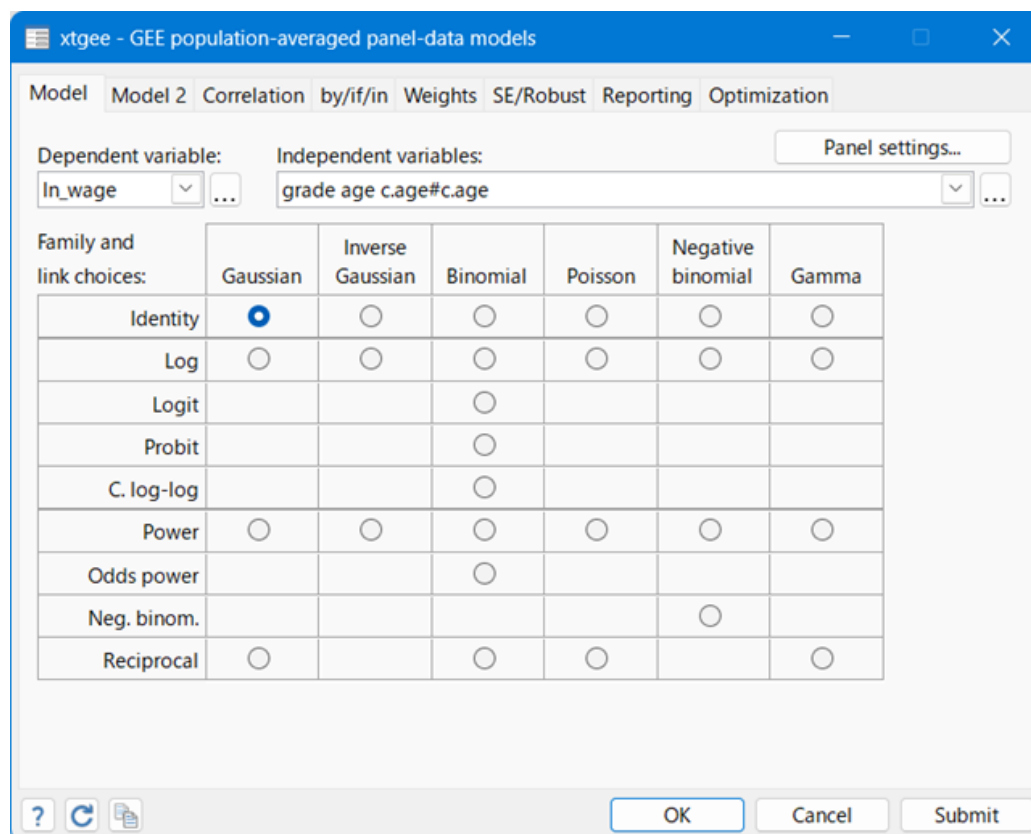


図 1 xtgee ダイアログ- Model タブ

- Correlation タブ: Independent

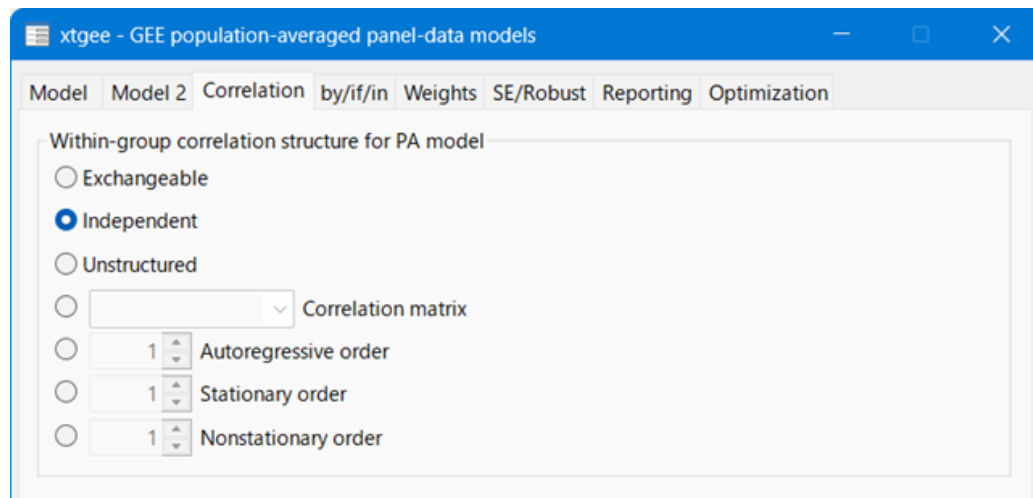


図 2 xtgee ダイアログ- Correlation タブ

- SE/Robust タブ: Scale factors: Use divisor N-P instead of N [nmp]

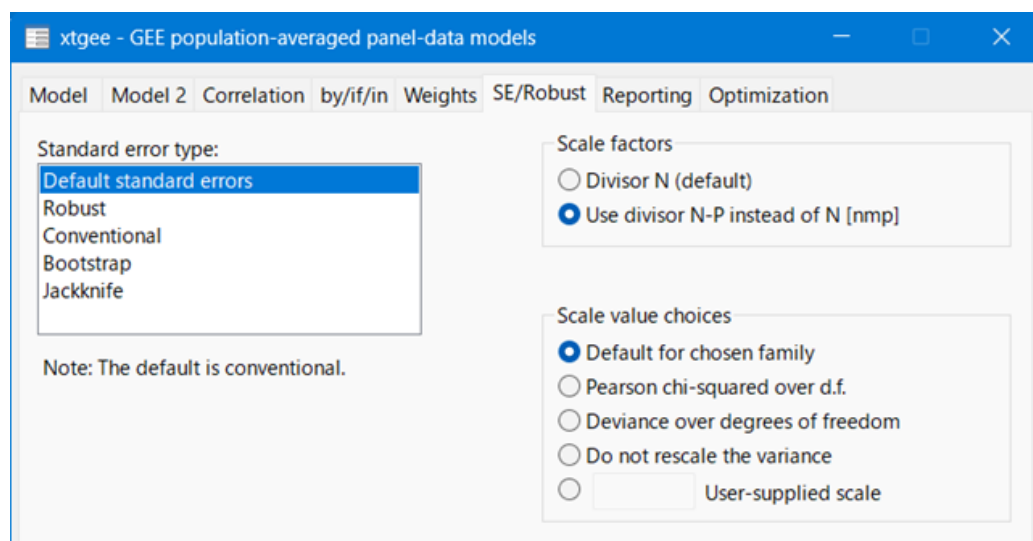


図 3 xtgee ダイアログ- SE/Robust タブ

```
. xtgee ln_wage grade age c.age#c.age, family(gaussian) link(identity) corr(indep
> endent) nmp
```

```
Iteration 1: Tolerance = 8.765e-13
```

```
GEE population-averaged model
Group variable: idcode
Family: Gaussian
Link: Identity
Correlation: independent
```

```
Number of obs   = 16,085
Number of groups = 3,913
Obs per group:
    min = 1
    avg = 4.1
    max = 9
```

```
Wald chi2(3)    = 4241.04
Prob > chi2     = 0.0000
```

```
Scale parameter = .1408958
```

```
Pearson chi2(16081) = 2265.75
Dispersion (Pearson) = .1408958
```

```
Deviance        = 2265.75
Dispersion      = .1408958
```

ln_wage	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
grade	.0724483	.0014229	50.91	0.000	.0696594	.0752372
age	.1064874	.0083644	12.73	0.000	.0900935	.1228812
c.age#c.age	-.0016931	.0001655	-10.23	0.000	-.0020174	-.0013688
_cons	-.8681487	.1024896	-8.47	0.000	-1.069025	-.6672728

推定された係数と標準誤差の値が regress の場合と完全に一致している点を確認できます。なお、nmp というオプションを指定せずスケールファクタをデフォルトの N のまま実行させると、標準誤差の推定値に若干の差が生じてきます。

▷ Example 2

評価版では割愛しています。

