

## 4 非線形パネルデータモデルの推定

中級者向けパネルデータモデルの最終回は非線形モデルの推定に関する Stata の機能を紹介します。Cameron and Trivedi (2010) による“Microeconometrics Using Stata, Revised Edition”, Stata Press の第 18 章 Nonlinear panel models の内容を用いて解説を行います。

### 4.1 非線形パネルデータ分析

今回利用するデータは“Rand Health Insurance Experiment”という社会実験<sup>11</sup>のものです。医療保険の自己負担割合が、受診行動に与える影響を考察する事を目的とし、自己負担割合の異なる保険を無作為割り付けしました。このデータを利用して 2 種類の非線形モデルを推定します。

#### データの要約統計量

モデル推定の前に、データの内容を確認します。

```
. use mus18data.dta,clear
. des dmdu med mdu lcoins ndisease female age lfam child id year
```

variable name	storage type	display format	value label	variable label
dmdu	float	%9.0g		any MD visit = 1 if mdu > 0
med	float	%9.0g		medical exp excl outpatient men
mdu	float	%9.0g		number face-to-face md visits
lcoins	float	%9.0g		log(coinsurance+1)
ndisease	float	%9.0g		count of chronic diseases -- ba
female	float	%9.0g		female
age	float	%9.0g		age that year
lfam	float	%9.0g		log of family size
child	float	%9.0g		child
id	float	%9.0g		person id, leading digit is sit
year	float	%9.0g		study year

最初に、これから推定する 2 種類の非線形モデル (パネルロジットとパネルポアソンモデル) における被説明変数についてチェックしましょう。ロジットモデルの被説明変数 dmdu は当年に受診したことを示すダミー変数です。平均して 69% の人が病院を訪れています。パネルポアソンモデルの被説明変数である mdu は医師と直接面談 (受診) した回数で、平均回数は 2.9 回となっています。

次にモデルで利用する説明変数を見ます。lcoins は自己負担割合に 1 を足した値の対数値です。慢性疾患の数を示す変数として ndisease, その他に 4 つの人口統計学的変数, そして対象者が子供であることを示すダミー変数 child も用意されています。

#### パネルデータの設定

Stata でパネルデータとして分析するために設定を行い、欠損値の情報を確認します。

<sup>11</sup>ランド医療保険実験というキーワードで検索すれば、日本語で詳細な情報を掲載しているウェブサイトを見つけることができます。

```
. xtset id year
```

```

panel variable:  id (unbalanced)

time variable:  year, 1 to 5, but with gaps
                delta: 1 unit

```

この実験は 1971 年から 1986 年にかけて実施され、5 時点で調査を実施しました。

```
. xtodes
```

```

id: 125024, 125025, ..., 632167          n =      5908
year: 1, 2, ..., 5                       T =        5
Delta(year) = 1 unit
Span(year) = 5 periods
(id*year uniquely identifies each observation)
Distribution of T_i:  min    5%   25%   50%   75%   95%   max
                   1      2     3     3     5     5     5

```

Freq.	Percent	Cum.	Pattern
3710	62.80	62.80	111..
1584	26.81	89.61	11111
156	2.64	92.25	1....
147	2.49	94.74	11...
79	1.34	96.07	..1..
66	1.12	97.19	.11..
33	0.56	97.75	..111
33	0.56	98.31	.1111
29	0.49	98.80	...11
71	1.20	100.00	(other patterns)
5908	100.00		XXXXX

この表で“1”とある所は全てのデータが揃っていることを示し，“.”は欠損値であることを示しています。最初に 3 時点では約 90%(全参加者数は 5,908 人) の人から回答得ています。そして、最初の 2 時点で、約 5% の人が調査から脱落しています。

### within と between の変動

パネルデータによるモデル推定を実施する前に within と between における説明変数の変動について考察します。例えば、変数  $x_{it}$  について within は  $x_{it} - \bar{x}_i$  であり、between は  $\bar{x}_i$  です。lcoins, ndisease, female は時間に対して不変な変数ですので、これらの within 変動はゼロです。したがって、それら以外の変数について変動の大きさを調べることにします。

```
. xtsum age lfam child
```

Variable		Mean	Std. Dev.	Min	Max	Observations
age	overall	25.71844	16.76759	0	64.27515	N = 20186
	between		16.97265	0	63.27515	n = 5908
	within		1.086687	23.46844	27.96844	T-bar = 3.41672
lfam	overall	1.248404	.5390681	0	2.639057	N = 20186
	between		.5372082	0	2.639057	n = 5908
	within		.0730824	.3242075	2.44291	T-bar = 3.41672
child	overall	.4014168	.4901972	0	1	N = 20186
	between		.4820984	0	1	n = 5908
	within		.1096116	-.3985832	1.201417	T-bar = 3.41672

これら3つの変数の変動は between 方向のものが within に比べ明らかに大きくなっています。従って, within 推定量を用いる固定効果モデルでは効率的な(標準誤差の小さい)推定量は得られそうにありません。もちろん, 推定値自体も, ランダム効果モデルや PA モデルの推定量とは, 幾分異なる値になることが予想できます。

#### モデルの予想

ランド医療保険実験は医療保険を当該家族に無作為割り付けしていますので, 自己負担割合の変数が内生性を持つことは無いと考えられます。固定効果モデルは  $X_{it}'\beta$  と効果  $u_i$  の間に相関を想定しますが, ここでは無作為化により, この相関を考えないランダム効果モデルや, PA モデルが適切ではないかという予想が成り立ちます。

## 4.2 二項モデル

当年に受診の有無を示す dmdu を被説明変数とするロジットモデルを推定します。最初にプーリングデータにおける単純なロジットモデルを推定し, 次にパネルデータにおけるロジットモデルの推定へと話を展開します。

#### パネルデータにおける被説明変数の考察

ロジットモデルを推定する前にパネルデータとして設定した時の変数 dmdu の分布を確認します。dmdu の変動は between でも within での同程度であることが分かります。

```
. xtsum dmdu
```

Variable		Mean	Std. Dev.	Min	Max	Observations
dmdu	overall	.6875062	.4635214	0	1	N = 20186
	between		.3571059	0	1	n = 5908
	within		.3073307	-.1124938	1.487506	T-bar = 3.41672

次に受診の遷移行列をみると、状態間にある程度の継続性があることが見て取れます。ある年に医師の受診がなかった人のうち、59%は翌年も受診していません。逆にある年に受診した人のうち80%は翌年も受診しています。

```
. xttrans dmdu
```

any MD visit = 1 if mdu > 0	any MD visit = 1 if mdu > 0		Total
	0	1	
0	58.87	41.13	100.00
1	19.73	80.27	100.00
Total	31.81	68.19	100.00

受診における自己相関を見てみると、ラグの回数に関係なく、一定の自己相関のあることが分かります。

```
. corr dmdu l.dmdu l2.dmdu
```

	dmdu	L. dmdu	L2. dmdu
dmdu	1.0000		
L1.	0.3861	1.0000	
L2.	0.3601	0.3807	1.0000

### プーリングデータにおけるロジットモデル

変量効果を考慮せずに、単純なロジットモデルを推定します。ただし、ここでは同一個体ごとに連続でデータを計測していることを考慮して vce オプションを利用します。

```
. logit dmdu lcoins ndisease female age lfam child, vce(cluster id) nolog
```

```
Logistic regression                Number of obs   =    20,186
                                Wald chi2(6)     =    488.18
                                Prob > chi2          =    0.0000
Log pseudolikelihood = -11973.392    Pseudo R2      =    0.0450
                                (Std. Err. adjusted for 5,908 clusters in id)
```

dmdu	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
lcoins	-.1572107	.0109064	-14.41	0.000	-.1785869	-.1358345
ndisease	.050301	.0039657	12.68	0.000	.0425285	.0580735
female	.3091573	.0445772	6.94	0.000	.2217876	.396527
age	.0042689	.0022307	1.91	0.056	-.0001032	.008641
lfam	-.2047573	.0470287	-4.35	0.000	-.2969318	-.1125828
child	.0921709	.0728107	1.27	0.206	-.0505355	.2348773
_cons	.6039409	.1107712	5.45	0.000	.3868333	.8210485

先頭の 4 つの推定値の符号は合理的であると考えられます。家族の構成人数の対数をとった lfam の値が負になっています。一家族の人数が増えると、家族内における医療費の負担を抑制し、それが受診を少なくする方向に影響を与えると理解できます。child の係数は正ですが有意ではありません。例えば、10 才の子供がいる場合、受診確率を計算する  $X'\beta$  の当該変数部の値は  $10 \times 0.0043 + 0.092 = 0.135$ 、一方、31 才の大人の場合は  $31 \times 0.043 + 0 \times 0.092 = 1.333$  でほぼ等しいことが分かります。つまり、他の変数の値が同じであれば、両者が受診確率に与える影響は等しくなります。

限界効果を考察する場合、Stata には margins というコマンドが用意されています。つまり、

```
. margins, dydx(*)
```

```
Average marginal effects          Number of obs   =    20,186
Model VCE      : Robust
Expression    : Pr(dmdu), predict()
dy/dx w.r.t.  : lcoins ndisease female age lfam child
```

	Delta-method				[95% Conf. Interval]	
	dy/dx	Std. Err.	z	P> z		
lcoins	-.0319566	.0021334	-14.98	0.000	-.0361379	-.0277753
ndisease	.0102248	.00078	13.11	0.000	.0086961	.0117535
female	.0628432	.0090029	6.98	0.000	.0451978	.0804886
age	.0008678	.000453	1.92	0.055	-.0000202	.0017557
lfam	-.0416215	.0095198	-4.37	0.000	-.06028	-.022963
child	.0187358	.0147928	1.27	0.205	-.0102576	.0477292

例えば、female が 1 単位増える、つまり、女性であれば受診確率は男性に比べ、0.063 ポイント高いことが分かります。ただし、この 0.063 という数値の計算方法には注意する必要があります。<sup>12</sup>

#### パネルデータにおける PA ロジット推定量

次に logit コマンドのパネルデータ版である xtlogit コマンドによる推定を行います。xtlogit コマンドには pa, re, fe のオプションがあり、re がデフォルトになっています。最初に系列相関を考慮した pa オプションから試してみる事にします。

<sup>12</sup>ここでは i.female としていませんので、female を連続変数として限界効果を計算しています。

```
. xtlogit dmdu lcoins ndisease female age lfam child,pa corr(exch) vce(robust) nolog
```

```
GEE population-averaged model
Group variable:          id      Number of obs   =   20,186
Link:                   logit    Number of groups =    5,908
Family:                 binomial  Obs per group:
Correlation:           exchangeable  min =         1
                                           avg =         3.4
                                           max =         5
                                           Wald chi2(6)  =   521.45
Scale parameter:       1        Prob > chi2     =    0.0000
                               (Std. Err. adjusted for clustering on id)
```

dmdu	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
lcoins	-.1603179	.0107779	-14.87	0.000	-.1814422	-.1391935
ndisease	.0515445	.0038528	13.38	0.000	.0439931	.0590958
female	.2977003	.0438316	6.79	0.000	.211792	.3836086
age	.0045675	.0021001	2.17	0.030	.0004514	.0086836
lfam	-.2044045	.0455004	-4.49	0.000	-.2935837	-.1152254
child	.1184697	.0674367	1.76	0.079	-.0137039	.2506432
_cons	.5776986	.106591	5.42	0.000	.368784	.7866132

プーリングデータの推定結果とよく似た結果になっています。PA 推定の標準誤差が総じて小さくなっていますので、僅かに有効性において優れていることが分かります。先の dmdu の自己相関は時点間の大きさにあまり左右されない、ということが分かりましたので、ここでは自己相関は一定であることを仮定した corr(exch) オプションを利用しました。自己相関の値は次のコマンドで確認できます。

```
. matrix list e(R)
(結果は省略します)
```

## RE ロジット推定量

個体毎の効果を考慮したロジットモデルは次のようになります。

$$\Pr(y_{it} = 1 | \mathbf{x}_{it}, \beta, \alpha_i) = \Lambda(\alpha_i + \mathbf{x}'_{it}\beta) \quad (17)$$

ここで  $\alpha_i$  は固定効果またはランダム効果となります。ランダム効果のあるロジットモデルを次のコマンドで推定します。

```
. xtlogit dmdu lcoins ndisease female age lfam child, re nolog
```

```

Random-effects logistic regression      Number of obs   =   20,186
Group variable: id                    Number of groups =    5,908
Random effects u_i ~ Gaussian         Obs per group:
                                         min =          1
                                         avg =          3.4
                                         max =          5
Integration method: mvaghermite       Integration pts. =    12
Log likelihood = -10878.687           Wald chi2(6)    =   549.76
                                         Prob > chi2     =    0.0000

```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
dmdu						
lcoins	-.2403864	.0162836	-14.76	0.000	-.2723017	-.208471
ndisease	.078151	.0055456	14.09	0.000	.0672819	.0890201
female	.4631005	.0663209	6.98	0.000	.3331138	.5930871
age	.0073441	.0031508	2.33	0.020	.0011687	.0135194
lfam	-.3021841	.0644721	-4.69	0.000	-.4285471	-.175821
child	.1935357	.1002267	1.93	0.053	-.002905	.3899763
_cons	.8629898	.1568968	5.50	0.000	.5554778	1.170502
/lnsig2u	1.225652	.0490898			1.129438	1.321866
sigma_u	1.84564	.045301			1.758953	1.936599
rho	.5087003	.0122687			.4846525	.532708

```
LR test of rho=0: chibar2(01) = 2189.41      Prob >= chibar2 = 0.000
```

PA ロジットの結果と比べると、係数が絶対値で 50% 程度、そして標準誤差も 50% 程度大きくなっています。よって  $z$  値にはあまり変化はありません。PA モデルと RE モデル (FE モデルも) では定式化が異なりますので、係数を比較するということは基本的に意味のないことです。注意してください。

画面下の `sigma_u` からランダム効果  $\alpha_i$  は  $N(0, 1.846^2)$  の分布に従うことが分かります。rho は次式による級内相関です。

$$\rho = \frac{\sigma_{\alpha}^2}{\sigma_{\alpha}^2 + \sigma_{\epsilon}^2}$$

xtlogit では  $\sigma_{\epsilon}^2 = \pi^2/3$  ですので、 $\hat{\rho} = 1.846^2 / (1.846^2 + \pi^2/3) = 0.509$  となります。推定結果の解釈は logit の項目で行ったように margins コマンドを利用します。しかし、ロジットモデルで係数は対数オッズとして解釈できますので、例えば、変数 age は 1 単位増えた時に対数オッズが 0.0073 だけ増えると理解することも可能です。<sup>13</sup>

## FE ロジット推定量

線形モデルでは最小二乗ダミー変数推定量 (LSDV) で  $\beta$  の一致推定量を求めることができましたが、ロジットモデルでは  $T \rightarrow \infty$  でない限り、一致推定量を求めることはできません。そこで、ロジットモデルでは  $\sum_{t=1}^{T_i} y_{it}$  が 1 になるという制約を掛けて推定を行います。これは推定期間中にアウトカム合計が 1 になるという強い制約です。

<sup>13</sup>対数オッズが具体的に何を意味するのか、直感的には分かりにくいですが、説明変数の値に依存せずに評価できる事に利用価値があります。

話を分かり易くするために2時点の状態を考えます。アウトカム  $y_{it}$  は受診するというものでした。ここでは  $y_{i1} + y_{i2} = 1$  ですから1をとるのはどちらか一つの時点です。例えば、

$$\Pr(y_{i1} = 0, y_{i2} = 1 | y_{i1} + y_{i2} = 1) = \frac{\Pr(y_{i1} = 0, y_{i2} = 1)}{\Pr(y_{i1} = 0, y_{i2} = 1) + \Pr(y_{i1} = 1, y_{i2} = 0)} \quad (18)$$

ここで  $y_{i1}$  と  $y_{i2}$  は  $\alpha_i$  と  $\mathbf{x}_{it}$  が与えられた時に独立であると仮定すると、 $\Pr(y_{i1} = 0, y_{i2} = 1) = \Pr(y_{i1} = 0) \times \Pr(y_{i2} = 1)$  となります。ロジットモデルは17式で定義されていますので、

$$\Pr(y_{i1} = 0, y_{i2} = 1) = \frac{1}{1 + \exp(\alpha_i + \mathbf{x}'_{i1}\beta)} \times \frac{\exp(\alpha_i + \mathbf{x}'_{i2}\beta)}{1 + \exp(\alpha_i + \mathbf{x}'_{i2}\beta)}$$

同様に、

$$\Pr(y_{i1} = 1, y_{i2} = 0) = \frac{\exp(\alpha_i + \mathbf{x}'_{i1}\beta)}{1 + \exp(\alpha_i + \mathbf{x}'_{i1}\beta)} \times \frac{1}{1 + \exp(\alpha_i + \mathbf{x}'_{i2}\beta)}$$

この両式を18式に代入して、

$$\begin{aligned} \Pr(y_{i1} = 0, y_{i2} = 1 | y_{i1} + y_{i2} = 1) & \quad (19) \\ &= \exp(\alpha_i + \mathbf{x}'_{i2}\beta) / \{\exp(\alpha_i + \mathbf{x}'_{i1}\beta) + \exp(\alpha_i + \mathbf{x}'_{i2}\beta)\} \\ &= \exp(\mathbf{x}'_{i2}\beta) / \{\exp(\mathbf{x}'_{i1}\beta) + \exp(\mathbf{x}'_{i2}\beta)\} \\ &= \exp\{(\mathbf{x}_{i2} - \mathbf{x}_{i1})' \beta\} / [1 + \exp\{(\mathbf{x}_{i2} - \mathbf{x}_{i1})' \beta\}] \end{aligned}$$

このように計算することで within 推定同様、先に  $\alpha_i$  を削除できます。そして、モデルは説明変数を  $\mathbf{x}_{i2} - \mathbf{x}_{i1}$  とするロジットモデルになります。また、時間に対して不変な変数では、線形モデルのロジットモデル同様、 $\mathbf{x}_{i2} - \mathbf{x}_{i1} = 0$  となり、係数を推定することはできません。

それでは実際に xtlogit コマンドで固定効果モデルを推定します。ただし、このデータは2時点ではありませんので、 $\sum_{t=1}^{T_i} y_{it} = 1, \sum_{t=1}^{T_i} y_{it} = 2, \dots, \sum_{t=1}^{T_i} y_{it} = T - 1$  という制約を掛け、すべての時点で0、または、1となるデータを除いて推定することになります。

```
. xtlogit dmdu lcoins ndisease female age lfam child,fe nolog
```

```
note: multiple positive outcomes within groups encountered.
```

```
note: 3,459 groups (11,161 obs) dropped because of all positive or all negative outcomes.
```

```
note: lcoins omitted because of no within-group variance.
```

```
note: ndisease omitted because of no within-group variance.
```

```
note: female omitted because of no within-group variance.
```

```
Conditional fixed-effects logistic regression   Number of obs   =   9,025
Group variable: id                             Number of groups =   2,449
Obs per group:
min =   2
avg =   3.7
max =   5
LR chi2(3) =   10.74
Prob > chi2 =   0.0132
Log likelihood = -3395.5996
```

dmdu	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lcoins	0	(omitted)				
ndisease	0	(omitted)				
female	0	(omitted)				
age	-.0341815	.0183827	-1.86	0.063	-.070211	.001848
lfam	.478755	.2597327	1.84	0.065	-.0303116	.9878217
child	.270458	.1684974	1.61	0.108	-.0597907	.6007068



推定結果から推定期間中にアウトカムに変動のないデータが 11161 個あり、推定に利用されていないことがわかります<sup>14</sup>。推定に利用している情報量が大幅に減り、標準誤差が大きくなっています。また、RE logit に比べ、係数の値も大きく変化しています。

### パネルロジット推定量の比較

ここまで推定したモデルを一つの表にまとめて表示します。

```
. global xlist lcoins ndisease female age lfam child
. quietly logit dmdx $xlist,vce(cluster id)
. estimates store POOLED
. quietly xtlogit dmdx $xlist,pa corr(exch) vce(robust)
. estimates store PA
. quietly xtlogit dmdx $xlist,re
. estimates store RE
. quietly xtlogit dmdx $xlist,fe
. estimates store FE
. estimates table POOLED PA RE FE,equations(1) se b(%8.4f) stats(N ll) stfmt(%8.0f)
```

Variable	POOLED	PA	RE	FE
#1				
lcoins	-0.1572 0.0109	-0.1603 0.0108	-0.2404 0.0163	(omitted)
ndisease	0.0503 0.0040	0.0515 0.0039	0.0782 0.0055	(omitted)
female	0.3092 0.0446	0.2977 0.0438	0.4631 0.0663	(omitted)
age	0.0043 0.0022	0.0046 0.0021	0.0073 0.0032	-0.0342 0.0184
lfam	-0.2048 0.0470	-0.2044 0.0455	-0.3022 0.0645	0.4788 0.2597
child	0.0922 0.0728	0.1185 0.0674	0.1935 0.1002	0.2705 0.1685
_cons	0.6039 0.1108	0.5777 0.1066	0.8630 0.1569	
lnsig2u				
_cons			1.2257 0.0491	
Statistics				
N	20186	20186	20186	9025
ll	-11973		-10879	-3396

legend: b/se

一覧表を作成するコマンドのオプションとして、equations(1) というものを利用していますが、試にこれを利用しない時の表も作成してみましょう。表はやや、縦長になってしまうはずですが。プールドロジットと PA モデル推定値と標準誤差はほとんど同じです。それに比べると RE ロジットモデルの

<sup>14</sup>利用したデータ (行) 数は全 20,186 行中、9,025 個です。

それは前述のように、絶対値が大きくなっています。FE モデルの結果はこれらに比べ違いが目立ち、時間に対して不変な変数の推定値は得られません。

モデル推定について数式をつかって簡単に違いを整理しましょう。プールドロジットと PA モデルは次式を利用してモデル推定を実行します。

$$\Pr(y_{it} = 1 | x_{it}) = \Lambda(x'_{it}\beta) \quad (20)$$

17 式と比較すると分かりますが、ここでは変量効果は利用しません。また、PA モデルはプールドロジットに比べ、誤差項の相関を考慮します。オプションの `exch` は時点間の距離に関係なく  $\rho_{ts} = \alpha$  としてモデル推定を行います。

### 4.3 カウントデータモデル

ここでは個人の受診回数 (`mdu`) をモデル化するポワソン回帰モデルについて考えます。クロスセクションデータの場合と同じく、パネルデータの場合も標本がポワソン分布に従っていなくても、条件付き平均が適切であれば、一致推定量を得る事が可能です。また、カウントデータの場合、往々にして過分散となってしまうことがあるので、ロバストな標準誤差の計算手法を用いることにします。

前出のように `mdu` の標本分散は  $4.50^2 = 20.25$  であり、これは標本平均 2.86 の約 7 倍となっています。このようなケースではポワソン推定量の標準誤差は結果として過小評価されてしまいます。

`mdu` の変動について考察します。

```
. xtsum mdu
```

Variable		Mean	Std. Dev.	Min	Max	Observations
mdu	overall	2.860696	4.504765	0	77	N = 20186
	between		3.785971	0	63.33333	n = 5908
	within		2.575881	-34.47264	40.0607	T-bar = 3.41672

`between` 変動と `within` 変動の大きさはほぼ同程度ですので、選択されるモデルを予想することはできません。次に `mdu` の時点間での遷移情報を見えます。最初に 4 回以上の訪問実績のある人たちについては、すべて 4 回であるものとする変数 `mdushort` を作成します。

```
. gen mdushort=mdu
. replace mdushort=4 if mdu>=4
. xttrans mdushort
```

mdushort	mdushort					Total
	0	1	2	3	4	
0	58.87	19.61	9.21	4.88	7.42	100.00
1	33.16	24.95	17.58	10.14	14.16	100.00
2	23.55	24.26	17.90	12.10	22.19	100.00
3	17.80	20.74	18.55	12.14	30.77	100.00
4	8.79	11.72	12.32	11.93	55.23	100.00
Total	31.81	19.27	13.73	9.46	25.73	100.00

この遷移行列から受診回数は、比較的、持続性の強い変数であることが分かります。例えば、受診回数 0 の人の半数強は翌年もゼロであることが分かります。逆に 4 回以上の方は、同様に翌年も 4 回以上受診しています。

### プールドポワソン推定量

プールドポワソン推定量は回数  $y_{it}$  が次に示すポワソン分布に従うことを仮定しています。

$$E(y_{it} | \mathbf{x}_{it}) = \exp(\mathbf{x}'_{it}\beta) \quad (21)$$

推定時に気を付けることは、データがポワソン分布に従っていないケースに備えて、ロバストな標準誤差のオプションを利用することです。poisson コマンドによるプールドポワソン推定の結果を次に示します。

```
. poisson mdu lcoins ndisease female age lfam child,vce(cluster id)
```

```
Iteration 0: log pseudolikelihood = -62580.248
Iteration 1: log pseudolikelihood = -62579.401
Iteration 2: log pseudolikelihood = -62579.401
Poisson regression              Number of obs   =    20,186
                                Wald chi2(6)      =    476.93
                                Prob > chi2          =    0.0000
Log pseudolikelihood = -62579.401 Pseudo R2       =    0.0609
                                (Std. Err. adjusted for 5,908 clusters in id)
```

mdu	Robust				
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lcoins	-.0808023	.0080013	-10.10	0.000	-.0964846 -.0651199
ndisease	.0339334	.0026024	13.04	0.000	.0288328 .039034
female	.1717862	.0342551	5.01	0.000	.1046473 .2389251
age	.0040585	.0016891	2.40	0.016	.000748 .0073691
lfam	-.1481981	.0323434	-4.58	0.000	-.21159 -.0848062
child	.1030453	.0506901	2.03	0.042	.0036944 .2023961
_cons	.748789	.0785738	9.53	0.000	.5947872 .9027907

推定結果の標準誤差について考察すると、不均一分散のオプションを使ったケース(省略)よりも 50% も大きく、オプションなしの場合(省略)に比べると 300% も大きくなっています。つまり、poisson コマンドではパネルデータの場合に生じる自己相関と、カウントデータの持つ過分散の情報を適切にコントロールできていないので、標準誤差のオプションを利用して効率的な推定量が得られない事を示しています。

### PA ポワソン推定量

アウトカム  $y_{it}$  は独立であるという仮定を緩めて、相関  $\rho_{ts} = \text{Cor}[\{y_{it} - \exp(\mathbf{x}'_{it}\beta)\}\{y_{is} - \exp(\mathbf{x}'_{is}\beta)\}]$  を許容する PA ポワソンモデルを推定します。このモデルも 21 式の定式化が適切であれば、一致推定量を提供します。

```
. xtpoisson mdu lcoins ndisease female age lfam child,pa corr(unstr) vce(robust)
```

```
Iteration 1: tolerance = .01585489
Iteration 2: tolerance = .00034066
Iteration 3: tolerance = 2.334e-06
Iteration 4: tolerance = 1.939e-08

GEE population-averaged model
Group and time vars:      id year      Number of obs   =   20,186
Link:                     log          Number of groups =    5,908
Family:                   Poisson      Obs per group:
Correlation:              unstructured  min =            1
                                           avg =            3.4
                                           max =            5
                                           Wald chi2(6)    =   508.61
Scale parameter:         1           Prob > chi2      =    0.0000
                                           (Std. Err. adjusted for clustering on id)
```

mdu	Robust		z	P> z	[95% Conf. Interval]	
	Coef.	Std. Err.				
lcoins	-.0804454	.0077782	-10.34	0.000	-.0956904	-.0652004
ndisease	.0346067	.0024238	14.28	0.000	.0298561	.0393573
female	.1585075	.0334407	4.74	0.000	.0929649	.2240502
age	.0030901	.0015356	2.01	0.044	.0000803	.0060999
lfam	-.1406549	.0293672	-4.79	0.000	-.1982135	-.0830962
child	.1013677	.04301	2.36	0.018	.0170696	.1856658
_cons	.7764626	.0717221	10.83	0.000	.6358897	.9170354

corr(unstructured) オプションは  $\rho_{ts}$  が時点間に依存せず、任意の値をとることを可能にします。この推定値はクラスタオプションを利用したブールド推定の結果とほぼ同じです。しかし、標準誤差については、それに比べ 10% ほど小さく、相関を多少上手くモデリングできている事を示しています。

## RE ポワソン推定量

変数効果のあるポワソンモデルは  $y_{it}$  が次のポワソン分布に従うことを仮定するものです。

$$E(y_{it}|\alpha_i, \mathbf{x}_{it}) = \exp(\gamma_i + \mathbf{x}'_{it}\beta) = \alpha_i \exp(\mathbf{x}'_{it}\beta) \quad (22)$$

ここで  $\gamma_i = \ln \alpha_i$  とし、 $\mathbf{x}_{it}$  は定数項を含むものとします。re のデフォルトの推定では  $\alpha_i$  は平均 1、分散  $\eta$  のガンマ分布に従うものとします。一方、対数変換した  $\gamma_i$  が正規分布に従うとするオプション normal も用意されています。また、PA およびブールド推定量は RE モデルの一致推定量であることが分かっています。ランダム効果モデルにおいて、クラスタを考慮したオプションは存在しないので、ブートストラップ法を利用して推定を実行します。ただし、この推定には 5 分程度の時間がかかります。

```
. xtpoisson mdu lcoins ndisease female age lfam child,re
>vce(boot, reps(400) seed(10101) nodots)
```

```
Random-effects Poisson regression      Number of obs   =   20,186
Group variable: id                    Number of groups =    5,908
Random effects u_i ~ Gamma            Obs per group:
                                         min =           1
                                         avg =           3.4
                                         max =           5
                                         Wald chi2(6)    =   553.55
Log likelihood = -43240.556            Prob > chi2     =    0.0000
                                         (Replications based on 5,908 clusters in id)
```

mdu	Observed Coef.	Bootstrap Std. Err.	z	P> z	Normal-based [95% Conf. Interval]	
lcoins	-.0878258	.0083762	-10.49	0.000	-.1042428	-.0714087
ndisease	.0387629	.0022311	17.37	0.000	.03439	.0431359
female	.1667192	.0361028	4.62	0.000	.0959591	.2374793
age	.0019159	.0016321	1.17	0.240	-.001283	.0051148
lfam	-.1351786	.0371143	-3.64	0.000	-.2079212	-.0624359
child	.1082678	.0511981	2.11	0.034	.0079215	.2086142
_cons	.7574177	.0854261	8.87	0.000	.5899856	.9248498
/lnalpha	.0251256	.0270056			-.0278044	.0780556
alpha	1.025444	.0276927			.9725786	1.081183

```
LR test of alpha=0: chibar2(01) = 3.9e+04      Prob >= chibar2 = 0.000
```

推定値はPAモデルに比べ、10%程度大きくなり、標準誤差も同様に大きくなります。つまり、PAモデルではSEを過小評価していることが分かります。

## FE ポワソン推定量

最後に  $x_{it}$  と  $\alpha_i$  の間の相関を許容する FE モデルを推定します.

```
. xtpoisson mdu lcoins ndisease female age lfam child,fe
>vce(boot, reps(400) seed(10101) nodots)
```

```
Conditional fixed-effects Poisson regression   Number of obs   =   17,791
Group variable: id                             Number of groups =    4,977
                                                Obs per group:
                                                min =           2
                                                avg =           3.6
                                                max =           5
                                                Wald chi2(3)    =    4.50
Log likelihood = -24173.211                    Prob > chi2     =    0.2120
                                                (Replications based on 4,977 clusters in id)
```

mdu	Observed Coef.	Bootstrap Std. Err.	z	P> z	Normal-based [95% Conf. Interval]	
age	-.0112009	.0089009	-1.26	0.208	-.0286462	.0062445
lfam	.0877134	.1158023	0.76	0.449	-.139255	.3146817
child	.1059867	.081788	1.30	0.195	-.0543148	.2662882

線形モデルの場合と同じく、ここでも時間に対して不変な変数の推定値は得られません。PA および RE の推定結果と比べると、標準誤差は 2 倍になっており、within 変動だけを利用した結果、効率性が失われていることが分かります。

## パネルポワソン推定量の比較

最後にそれぞれの推定量を比較してみましょう。時間を節約するために、ここでは先に利用したブートストラップオプションは利用しません。

```
. quietly xtpoisson mdu lcoins ndisease female age lfam child,pa corr(unstr)
>vce(robust)
. estimates store PPA_ROB
. quietly xtpoisson mdu lcoins ndisease female age lfam child,re
. estimates store PRE
. quietly xtpoisson mdu lcoins ndisease female age lfam child,re normal
. estimates store PRE_NORM
. quietly xtpoisson mdu lcoins ndisease female age lfam child,fe
. estimates store PFE
. estimates table PPA_ROB PRE PRE_NORM PFE,equations(1) b(%8.4f) se stats(N ll)
>stfmt(%8.0f)
```

Variable	PPA_ROB	PRE	PRE_NORM	PFE
#1				
lcoins	-0.0804 0.0078	-0.0878 0.0069	-0.1145 0.0073	
ndisease	0.0346 0.0024	0.0388 0.0022	0.0409 0.0023	
female	0.1585 0.0334	0.1667 0.0286	0.2084 0.0305	
age	0.0031 0.0015	0.0019 0.0011	0.0027 0.0012	-0.0112 0.0039
lfam	-0.1407 0.0294	-0.1352 0.0260	-0.1443 0.0265	0.0877 0.0555
child	0.1014 0.0430	0.1083 0.0341	0.0737 0.0345	0.1060 0.0438
_cons	0.7765 0.0717	0.7574 0.0618	0.2873 0.0642	
lnalpha				
_cons		0.0251 0.0210		
lnsig2u				
_cons			0.0550 0.0255	
Statistics				
N	20186	20186	20186	17791
ll		-43241	-43227	-24173

legend: b/se

PA と RE の推定結果はほぼ同じです。また、変量効果が正規分布するとした RE モデルも、やや近いものになっています。それらに比べ、FE モデルの推定結果は大きく異なります。

以上  
2016年12月  
株式会社 ライトストーン