Description Remarks and examples Also see

Description

This introduction covers the command cmxtmixlogit. This is the only one of the cm estimation commands that explicitly models panel data. Other cm estimation commands, however, can be used with panel data when run with an appropriate variance estimator, that is, vce(cluster *idvar*), vce(bootstrap, cluster(*idvar*)), or vce(jackknife, cluster(*idvar*)).

cmxtmixlogit fits a mixed logit model to panel choice data. cmxtmixlogit models a sequence of choices rather than a single choice, as commands for cross-sectional data do. As with cmmixlogit, random coefficients can be fit to model the correlation of choices across alternatives, and the property of independence of irrelevant alternatives (IIA) is not assumed. See Overview of CM commands for discrete choices in [CM] Intro 5, and see [CM] Intro 8 if you are not familiar with this assumption.

Remarks and examples

Remarks are presented under the following headings:

Data layout for panel choice data A cmxtmixlogit model Time-series operators Using other cm estimation commands with panel data

Data layout for panel choice data

In panel choice data, decision makers make multiple choices at different times. The data layout is similar to that for cross-sectional data, the difference being that there are repeated cases for each decision maker.

Here is an example of panel choice data. These fictitious data represent individuals' choices of transportation mode at multiple times. We list the data for the first two people:

. use https://www.stata-press.com/data/r19/transport (Transportation choice data) . list if id <= 2, sepby(t)</pre>

	id	t	alt	choice	trcost	trtime	age	income	parttime
1.	1	1	Car	1	4.14	0.13	3.0	3	Full-time
2.	1	1	Public	0	4.74	0.42	3.0	3	Full-time
3.	1	1	Bicycle	0	2.76	0.36	3.0	3	Full-time
4.	1	1	Walk	0	0.92	0.13	3.0	3	Full-time
5.	1	2	Car	1	8.00	0.14	3.2	5	Full-time
6.	1	2	Public	0	3.14	0.12	3.2	5	Full-time
7.	1	2	Bicycle	0	2.56	0.18	3.2	5	Full-time
8.	1	2	Walk	0	0.64	0.39	3.2	5	Full-time
9.	1	3	Car	1	1.76	0.18	3.4	5	Part-time
10.	1	3	Public	0	2.25	0.50	3.4	5	Part-time
11.	1	3	Bicycle	0	0.92	1.05	3.4	5	Part-time
12.	1	3	Walk	0	0.58	0.59	3.4	5	Part-time
13.	2	1	Car	0	4.36	0.23	3.0	2	Full-time
14.	2	1	Public	0	4.43	0.43	3.0	2	Full-time
15.	2	1	Bicycle	0	1.25	1.23	3.0	2	Full-time
16.	2	1	Walk	1	0.89	0.12	3.0	2	Full-time
17.	2	2	Car	0	7.14	0.23	3.2	3	Part-time
18.	2	2	Public	1	1.54	0.12	3.2	3	Part-time
19.	2	2	Bicycle	0	2.75	0.95	3.2	3	Part-time
20.	2	2	Walk	0	0.53	1.64	3.2	3	Part-time
21.	2	3	Car	0	6.69	0.17	3.4	2	Full-time
22.	2	3	Public	1	1.32	0.34	3.4	2	Full-time
23.	2	3	Bicycle	0	0.60	0.49	3.4	2	Full-time
24.	2	3	Walk	0	0.68	0.63	3.4	2	Full-time

Individuals (identified by the variable id) at each of three time points (time variable t) could choose between four modes of transportation (alternatives variable alt) with the one chosen alternative indicated by the binary variable choice. The first person chose to use a car at all three time points. The second person walked at time = 1 and took public transportation at the other two times.

Cost of travel (trcost, measured in \$) and travel time (trtime, measured in hours) are alternativespecific variables. Variables age (measured in decades), income (annual income measured in \$10,000), and parttime (indicating a part-time or full-time job) are case specific. Before we can fit the model, we must cmset the data. For panel data, cmset requires three variables: first, the variable identifying individuals (id), second, the time variable (t), and third, the alternatives variable (alt). (cmxtmixlogit, like cmmixlogit, can fit models without explicitly identified alternatives. In this case, there is no alternatives variable, and the option noalternatives is specified.)

```
. cmset id t alt
note: case identifier _caseid generated from id and t.
note: panel by alternatives identifier _panelaltid generated from id and alt.
Panel data: Panels id and time t
Case ID variable: _caseid
Alternatives variable: alt
Panel by alternatives variable: _panelaltid (strongly balanced)
Time variable: t, 1 to 3
Delta: 1 unit
```

```
Note: Data have been xtset.
```

The notes displayed by cmset say it has created two new variables: _caseid and _panelaltid. Let's list their values for the first two individuals.

	id	id t alt		_caseid	_panelaltid
1.	1	1	Car	1	1
2.	1	2	Car	2	1
3.	1	3	Car	3	1
4.	1	1	Public	1	2
5.	1	2	Public	2	2
6.	1	3	Public	3	2
7.	1	1	Bicycle	1	3
8.	1	2	Bicycle	2	3
9.	1	3	Bicycle	3	3
10.	1	1	Walk	1	4
11.	1	2	Walk	2	4
12.	1	3	Walk	3	4
13.	2	1	Car	4	5
14.	2	2	Car	5	5
15.	2	3	Car	6	5
16.	2	1	Public	4	6
17.	2	2	Public	5	6
18.	2	3	Public	6	6
19.	2	1	Bicycle	4	7
20.	2	2	Bicycle	5	7
21.	2	3	Bicycle	6	7
22.	2	1	Walk	4	8
23.	2	2	Walk	5	8
24.	2	3	Walk	6	8

. list id t alt _caseid _panelaltid if id <= 2, sepby(alt) abbr(11)

_caseid is a variable that identifies cases. For choice model data, remember that a case is a single statistical observation but consists of multiple Stata observations. Each distinct value of panel ID × time represents a single statistical observation, that is, a case. The values of _caseid correspond to the distinct values of panel ID × time, in this example the values of id × t.

_panelaltid is a variable that uniquely identifies the distinct values of panel ID \times alternative. We will explain why this variable is needed when we show an example with time-series operators. But you can skip over the explanation. These new variables make cmxtmixlogit work as you would expect. You need not be concerned about them, just leave them in your dataset.

A cmxtmixlogit model

Continuing with the previous example, we wish to model the effect of travel cost (trcost), travel time (trtime), income, and age on the choice of transportation mode.

We assume that all individuals have the same preferences with respect to travel cost but that preferences with respect to travel time are heterogeneous, and we model these heterogeneous preferences with a random coefficient for trtime by specifying the option random(trtime).

The dependent variable is choice, the binary variable indicating which alternative was chosen. The variable trcost is included following the dependent variable; placing it in this position means that it should have a fixed coefficient. Specifying casevars(age income) includes the case-specific variables age and income in the model with fixed coefficients.

. cmxtmixlogi	t choice trcos	st, random(t	rtime) c	casevars(age inco	ome)			
Fitting fixed	parameter mod	lel:							
Fitting full	model:								
Iteration 0: Iteration 1: Iteration 2: Iteration 3: Iteration 4: Iteration 5:	Log simulated Log simulated Log simulated Log simulated Log simulated Log simulated	l-likelihood l-likelihood l-likelihood l-likelihood l-likelihood l-likelihood	$ \begin{array}{rcl} l &=& -102 \\ l &=& -1014 \\ l &=& -1006 \\ l &=& -1005 \\ l &=& -1005 \\ l &=& -1005 \end{array} $	25.707 (1 1.2513 5.2212 5.9904 5.9899 5.9899	not con	cave)			
Mixed logit c	hoice model		N	lumber of	obs	=	6,000		
			N	lumber of	cases	=	1,500		
Panel Variabi	.e: 1d		N	Number of panels = 5					
Time variable	: t		C	ases per	panel:	min = avg = max =	3 3.0 3		
Alternatives	variable: alt		Α	lts per	case:	min = avg = max =	4 4.0 4		
Integration s Integration p Log simulated	equence: oints: -likelihood =	Hammersley 594 -1005.9899		Wald Prob	chi2(8) > chi2) = =	432.68 0.0000		
choice	Coefficient	Std. err.	Z	P> z	[95]	% conf.	interval]		
alt trcost trtime	8388216 -1.508756	.0438587 .2641554	-19.13 -5.71	0.000	92 -2.0	47829 26492	7528602 9910212		
/Normal sd(trtime)	1.945596	.2594145			1.4	98161	2.526661		
Car	(base alter	rnative)							
Public									
age	.1538915	.0672638	2.29	0.022	.02	20569	.2857261		
income _cons	3815444 5756547	.0347459 .3515763	-10.98 -1.64	0.000 0.102	449 -1.20	96451 64732	3134437 .1134222		
Bicvcle									
age	.20638	.0847655	2.43	0.015	.040	02426	.3725174		
income	5225054	.0463235	-11.28	0.000	61	32978	4317131		
	-1.137393	.4461318	-2.55	0.011	-2.0	11795	2629909		
Walk									
age	.3097417	.1069941	2.89	0.004	.10	00372	.5194463		
LICOME	4183279	.0000042	-13.14	0.000	-1.0	17302	1012018		
_00115	. 1100213		0.10	0.400	1.0	1,002	.0000-100		

The coefficients for trcost and trtime are negative, indicating that as cost and travel time increase, the probability of selecting a method of travel decreases. In the Public, Bicycle, and Walk sections of the output, we see coefficients for the case-specific variables. These are each interpreted relative to the base alternative Car. We can use margins to more easily interpret the results of this model; see [CM] Intro 1 and [CM] margins.

Because we did not specify a distribution in the random() option, we got the default distribution for the random coefficient, which is the normal distribution. Other options for the distribution are available. If we had multiple variables in the random() option, we could optionally specify corrmetric() to pick the form of the correlation modeled. See [CM] cmxtmixlogit for more information on options for random coefficients.

Time-series operators

When you cmset panel data with specified alternatives, your data are automatically xtset. You can type xtset to see the settings:

```
. xtset
Panel variable: _panelaltid (strongly balanced)
Time variable: t, 1 to 3
Delta: 1 unit
```

_panelaltid becomes the "panel" identifier for viewing the data as xt data. This is why cmxtmixlogit needs this variable. It is created so you can use Stata's time-series operators (see [U] 11.4.3.6 Using factor variables with time-series operators) with cmxtmixlogit. For instance, if you want to include lags of alternative-specific variables in your model, the lags must be specific to the alternative, and Stata's time-series lag operator needs to know how to do this.

To illustrate, we add a lag trtime to our earlier model. We also specify correlated for the random coefficients of trtime and its lag so that the distributions of the random coefficients are correlated. Note that because of the additional complexity of this model, it is computationally intensive and may take a few minutes to fit.

```
. cmxtmixlogit choice, random(trtime L.trtime, correlated) casevars(age income)
Fitting fixed parameter model:
Fitting full model:
Iteration 0: Log simulated-likelihood = -726.49438
                                                    (not concave)
Iteration 1:
             Log simulated-likelihood = -725.73356
Iteration 2: Log simulated-likelihood = -724.30029
Iteration 3: Log simulated-likelihood = -720.40177
Iteration 4: Log simulated-likelihood = -720.28639
Iteration 5: Log simulated-likelihood = -720.07741
Iteration 6: Log simulated-likelihood = -720.07434
Iteration 7: Log simulated-likelihood = -720.07411
Iteration 8: Log simulated-likelihood = -720.07411
Refining estimates:
Iteration 0: Log simulated-likelihood = -720.07411
Iteration 1: Log simulated-likelihood = -720.07411
```

Mixed logit cl	hoice model		Number of obs = 4,0 Number of cases = 1,0						
Panel variable	e: id		N	umber of	panels	=	500		
Time variable	: t		C	Cases per panel: min = avg = max =					
Alternatives	variable: alt		A	min = avg = max =	4 4.0 4 82.87				
Integration so Integration po	equence: pints:	Hammersley 625		=					
Log simulated	-likelihood =	-720.07411		Prob > chi2 =					
choice	Coefficient	Std. err.	Z	P> z	[95%	; conf.	interval]		
alt									
trtime									
 L1	-1.02391 - 7797073	.3884411 3843897	-2.64 -2.03	0.008	-1.78	35241 33097	2625792 0263174		
			2.00						
/Normal sd(trtime)	.8594882	.6233604			.207	4386	3.56115		
corr(trtime,	4457922	6071271	0 73	0 463	- 763	10532	961/328		
sd(L.trtime)	1.576005	.4241405	0.10	0.400	.929	99912	2.670768		
Car	(base alter	native)							
Public									
age	.0848749	.0715193	1.19	0.235	055	3005	.2250502		
income	208774	.0336985	-6.20	0.000	274	8219	1427261		
_cons	.1079519	.3923718	0.28	0.783	661	.0826	.8769865		
Bicycle									
age	.2542854	.1066569	2.38	0.017	.045	52418	.4633291		
income	3155109	.0531635	-5.93	0.000	419	7094	2113123		
_cons	462521	.5845974	-0.79	0.429	-1.60	8311	.6832688		
Walk									
age	.5830396	.1878859	3.10	0.002	.2	21479	.9512892		
income	8183397	.1207108	-6.78	0.000	-1.05	54929	5817508		
_cons	.0269189	.9301377	0.03	0.977	-1.79	6118	1.849955		

Including the lag of trtime in this model may not have made much conceptual sense, but we did so for the purpose of showing how to use time-series operators with cmxtmixlogit.

Using other cm estimation commands with panel data

cm estimation commands for cross-sectional data can also be used with panel data. The estimates from these commands have a population-averaged interpretation when used with panel data. The cmsettings tell these cross-sectional cm commands that the data are panel data. In this case, by default, the cm commands report cluster-robust standard errors that account for the within-panel correlation.

Here is what we get if we run a cmclogit model on our previous panel choice data.

. cmclogit ch	oice trcost tr re cmset as pa	time, casev	vars(age	income) efault v	cetvne fo	or nan	el data is
vce(clu	ster id); see	cmclogit.	una one a	orauro v	000390 10	or puir	SI daba ib
Iteration 0: Iteration 1: Iteration 2: Iteration 3: Iteration 4:	Log pseudolik Log pseudolik Log pseudolik Log pseudolik Log pseudolik	elihood = - elihood = - elihood = - elihood = - elihood = -	-1197.990 -1035.481 -1027.634 -1027.622 -1027.622	2 7 6 7 7 7			
Conditional 1	ogit choice mo		Number o	of obs	=	6,000	
Case ID varia	ble: _caseid			Number (of cases	=	1500
Alternatives	variable: alt			Alts per	r case: 1 a 1	nin = avg = nax =	4 4.0 4
Log pseudolik	elihood = -102	7.6227		Wald Prob	chi2(8) > chi2	= =	335.13 0.0000
		(St	td. err.	adjusted	for 500	clust	ers in id)
		Pohuat					
choice	Coefficient	std. err.	z	P> z	[95%	conf.	interval]
alt							
trcost trtime	7667673 6572159	.0464592 .1700226	-16.50 -3.87	0.000	8578 990	3258 0454	6757089 3239778
Car	(base alter	native)					
Public							
age	.1588594	.0784292	2.03	0.043	.0051	1409	.3125779
income	3479798	.0405743	-8.58	0.000	4275	5039	2684557
_cons	8253419	.3651235	-2.26	0.024	-1.540	0971	109713
Bicycle							
age	.2025874	.0867382	2.34	0.020	.0325	5835	.3725912
income	4538989	.0436598	-10.40	0.000	5394	1705	3683273
_cons	-1.505446	.4571108	-3.29	0.001	-2.401	1367	6095252
Walk							
age	.307546	.1077107	2.86	0.004	.0964	1369	.518655
income	7648748	.0616934	-12.40	0.000	8857	7917	6439579
_cons	959179	.5054328	-1.90	0.058	-1.949	9809	.0314511

By default, cmclogit used the variance estimator given by vce(cluster id). If you wish to change the variance estimator, simply set the vce() option to what you want.

Also see

- [CM] Intro 1 Interpretation of choice models
- [CM] Intro 2 Data layout
- [CM] Intro 3 Descriptive statistics
- [CM] Intro 4 Estimation commands
- [CM] **cmclogit** Conditional logit (McFadden's) choice model
- [CM] cmxtmixlogit Panel-data mixed logit choice model

Stata, Stata Press, Mata, NetCourse, and NetCourseNow are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow is a trademark of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2025 StataCorp LLC, College Station, TX, USA. All rights reserved.



For suggested citations, see the FAQ on citing Stata documentation.