

Revisiting Margins

Enrique Pinzón

StataCorp LP

November 13, 2014
Milano

Introduction

- I am going to use today as an excuse to talk about `margins` and how I visualize models and their interpretation
- As many Stata tools, `margins` is powerful yet underutilized.
- `margins` helps us interpret and use the results of our models.

- Review concepts to establish a common ground
 - ▶ Linear cross-sectional models
 - ▶ Nonlinear cross-sectional models
- Present preliminary results of an article I have been working on
 - ▶ Linear panel-data models
 - ▶ Nonlinear panel-data models

Object of Interest

- Interpreting the results from a model depend on the object of interest
- It is rarely the case that the estimated parameters are the object of interest
- We will illustrate this with linear regression

Linear Regression Model

- The model is given by

$$\begin{aligned}y_i &= \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i \\E(\varepsilon_i | x_{i1} \dots x_{ik}) &= 0 \\E(y_i | x_{i1} \dots x_{ik}) &= \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}\end{aligned}$$

- We would like to now how a change in one of the regressors affects $E(y_i | x_{i1} \dots x_{ik})$. For example, for a continuous x_k we have that :

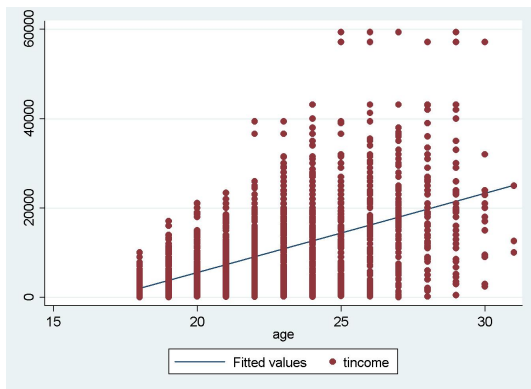
$$\frac{\partial E(y_i | x_{i1} \dots x_{ik})}{\partial x_k} = \beta_k$$

Graphical Representation

$$\text{Income} = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Hispanic} + \beta_3 \text{Male} + \varepsilon$$

$$E(\text{Income} | \text{Age}, \text{Hispanic} = \text{Male} = 1) = \delta + \beta_1 \text{Age}$$

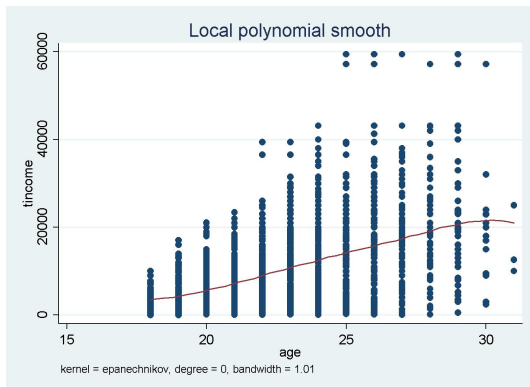
$$\delta \equiv \beta_0 + \beta_2 + \beta_3$$



Graphical Representation

$$\text{Income} = g(\text{Age}, \text{Hispanic}, \text{Male}) + \varepsilon$$

$$E(\text{Income} | \text{Age}, \text{Hispanic} = \text{Male} = 1)$$



Interaction terms

- If our model is of the form

$$Income = \beta_0 + \beta_1 Age + \beta_2 Hispanic + \beta_3 Male + \beta_4 Age^2 + \varepsilon$$

- The marginal effects are of the form

$$\frac{\partial E (Income|Age, Hispanic, Male)}{\partial Age} = \beta_1 + 2\beta_4 Age$$

- `margins` helps us obtain the correct value and standard error for the average marginal effect

```
. regress tincome i.hispanic i.male c.age#c.age, noheader
```

tincome	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
1.hispanic	-13.25757	118.8548	-0.11	0.911	-246.218	219.7029
1.male	2425.08	85.18172	28.47	0.000	2258.12	2592.04
age	908.2874	198.5842	4.57	0.000	519.054	1297.521
c.age#c.age	17.16441	4.26985	4.02	0.000	8.795324	25.5335
_cons	-21689.52	2283.89	-9.50	0.000	-26166.04	-17213

```
. quietly generate margin = _b[age] + 2*_b[c.age#c.age]*age
. summarize margin
```

Variable	Obs	Mean	Std. Dev.	Min	Max
margin	30213	1688.333	100.9031	1526.206	1972.481

```
. margins, dydx(age)
Average marginal effects          Number of obs   =       30213
Model VCE      : OLS
Expression    : Linear prediction, predict()
dy/dx w.r.t.  : age
```

	Delta-method				
	dy/dx	Std. Err.	t	P> t	[95% Conf. Interval]
age	1688.333	15.03678	112.28	0.000	1658.861 1717.806

Probit and Logit Models

- Probit and Logit models are given by:

$$y_i = \begin{cases} 1 & \text{if } y_i^* = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

- The assumptions we make on the distribution of ε_i give us different models.
- The object of interest is $P(y_i | x_{i1} \dots x_{ik})$

$$\frac{\partial P(y_i | x_{i1} \dots x_{ik})}{\partial x_j} = \phi(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}) \beta_j$$

- Average marginal effect

$$\frac{1}{N} \sum_{i=1}^N \phi(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}) \beta_j$$

- Marginal effect at a point, for example \bar{x}

$$\phi(\beta_0 + \beta_1 \bar{x}_1 + \dots + \beta_k \bar{x}_k) \beta_j$$

Probability of Never Being Married

```
. probit nevermarried age i.male i.college, nolog
Probit regression                               Number of obs   =       30213
                                                LR chi2(3)      =       6018.41
                                                Prob > chi2     =       0.0000
Log likelihood = -15229.266                    Pseudo R2      =       0.1650
```

nevermarried	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	-.2302782	.0033003	-69.77	0.000	-.2367467	-.2238097
1.male	.3077872	.0165872	18.56	0.000	.2752768	.3402976
1.college	.354033	.0208055	17.02	0.000	.313255	.3948109
_cons	5.655506	.0757096	74.70	0.000	5.507118	5.803894

```
. estimates store probit
```

Average Marginal Effects

```
. margins, dydx(age)
Average marginal effects          Number of obs   =      30213
Model VCE      : OIM
Expression    : Pr(nevermarried), predict()
dy/dx w.r.t.  : age
```

	Delta-method				
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]
age	-.0650363	.0006976	-93.23	0.000	-.0664036 -.063669

```
. margins, dydx(age) atmeans
Conditional marginal effects      Number of obs   =      30213
Model VCE      : OIM
Expression    : Pr(nevermarried), predict()
dy/dx w.r.t.  : age
at
  at          : age          =      22.72277 (mean)
              : 0.male         =      .4962764 (mean)
              : 1.male         =      .5037236 (mean)
              : 0.college      =      .7830404 (mean)
              : 1.college      =      .2169596 (mean)
```

	Delta-method				
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]
age	-.0741411	.0010415	-71.19	0.000	-.0761823 -.0720998

Logit vs. Probit

- Average marginal effects, marginal effects at a point, and predicted probabilities are similar for logit and probit models.
- The reason is that the logistic and normal distributions are very similar.
- If the objects of interest are the ones mentioned above, the use of one or the other is a fruit of custom.

Logit vs. Probit: Point Estimates

```
. quietly logit nevermarried age i.male i.college, nolog  
. estimates store logit  
. estimates table logit probit
```

Variable	logit	probit
age	-.38456626	-.23027823
male 1	.52122078	.30778721
college 1	.5892684	.35403299
_cons	9.4333826	5.6555062

Logit vs. Probit: Average Marginal Effects

```
. quietly margins, dydx(age) post
. estimates store dydxlog
. estimates restore probit
(results probit are active now)
. quietly margins, dydx(age) post
. estimates store dydxprob
. estimates table dydx*
```

Variable	dydxlog	dydxprob
age	-.06406436	-.06503633

Logit vs. Probit: Marginal Effects at Means

```
. estimates restore logit
(results logit are active now)
. quietly margins, dydx(age) atmeans post
. estimates store atlog
. estimates restore probit
(results probit are active now)
. quietly margins, dydx(age) atmeans post
. estimates store atprob
. estimates table at*
```

Variable	atlog	atprob
age	-.07258275	-.07414107

Logit vs. Probit: Predicted Probabilities

```
. estimates restore logit
(results logit are active now)

. predict prlogit
(option pr assumed; Pr(nevermarried))

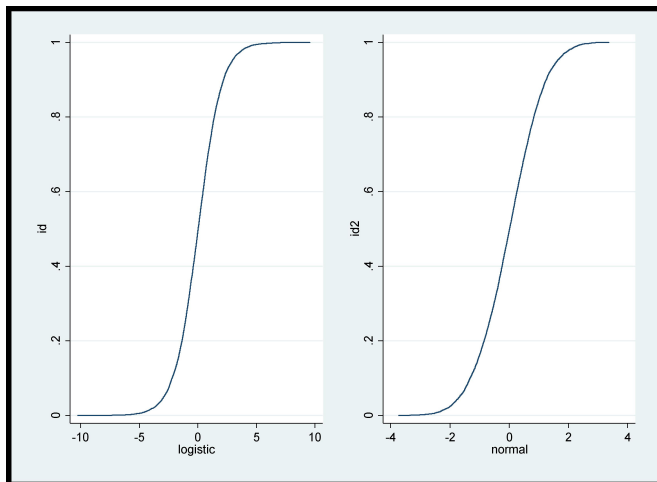
. estimates restore probit
(results probit are active now)

. predict prprobit
(option pr assumed; Pr(nevermarried))

. summarize pr*
```

Variable	Obs	Mean	Std. Dev.	Min	Max
prlogit	30213	.7083044	.2000549	.0766916	.9622073
prprobit	30213	.7102447	.2006532	.0690214	.9739339

Logit vs. Probit: Graphically



Panel Data

- Allows us to model and understand individual heterogeneity
- Follow effects of policy or interventions across time
- Most of our intuition from cross-sectional data (objects of interest) translate to panel-data

Object of Interest: Linear Panel-Data Model

- The linear panel-data model is given by:

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \dots + \beta_k x_{kit} + \alpha_j + \varepsilon_{it}$$
$$E(\varepsilon_{it} | X, \alpha_j) = 0$$

- Random Effects: $E(\alpha_j | x_{1it} \dots x_{kit}) = 0$

$$E(y_{it} | x_{1it} \dots x_{kit}) = \beta_0 + \beta_1 x_{1it} + \dots + \beta_k x_{kit}$$

- Fixed Effects: $E(\alpha_j | x_{1it} \dots x_{kit}) \neq 0$

$$E(y_{it} | x_{1it} \dots x_{kit}) = \beta_0 + \beta_1 x_{1it} + \dots + \beta_k x_{kit} + E(\alpha_j | x_{1it} \dots x_{kit})$$

Fixed-Effects Solution

- Transform variables:

$$\begin{aligned}\bar{y}_i &= \frac{1}{T} \sum_{t=1}^T y_{it} \\ &= \beta_0 + \beta_1 \bar{x}_{1i} + \dots + \beta_k \bar{x}_{ki} + \alpha_i + \bar{\varepsilon}_i\end{aligned}$$

- Generate new variable

$$\begin{aligned}y_{it} - \bar{y}_i &= (\beta_0 - \beta_0) + \beta_1 (x_{1it} - \bar{x}_{1i}) + \dots + \beta_k (x_{kit} - \bar{x}_{ki}) \\ &\quad + (\alpha_i - \alpha_i) + (\varepsilon_{it} - \bar{\varepsilon}_i) \\ \ddot{y}_{it} &= \beta_1 \ddot{x}_{1it} + \dots + \beta_k \ddot{x}_{kit} + \ddot{\varepsilon}_{it} \\ \ddot{y}_{it} &\equiv y_{it} - \bar{y}_i\end{aligned}$$

Object of Interest: Conditional Expectation

- The above procedure gives us a consistent estimator of the parameters $\hat{\beta}_1, \dots, \hat{\beta}_k$
- But:

$$\begin{aligned}\hat{E}(y_{it}|x_{1it} \dots x_{kit}) &= \hat{\beta}_1 x_{1it} + \dots + \hat{\beta}_k x_{kit} \\ &\rightarrow E(y_{it}|x_{1it} \dots x_{kit})\end{aligned}$$

- The reason is that we do not have an estimate of:

$$E(\alpha_i|x_{1it} \dots x_{kit})$$

Object of Interest: Marginal Effects

- Likewise

$$\begin{aligned} \frac{\partial E(y_{it} | \widehat{x_{1it} \dots x_{kit}})}{\partial x_{kit}} &= \hat{\beta}_k \\ \rightarrow &\frac{\partial E(y_{it} | x_{1it} \dots x_{kit})}{\partial x_{kit}} \end{aligned}$$

- The reason is that we do not have an estimate for

$$\frac{\partial E(\alpha_j | x_{1it} \dots x_{kit})}{\partial x_{kit}}$$

Proposed Solution: Mundlak-Chamberlain Approach

- Propose a model for

$$E(\alpha_i | x_{1it} \dots x_{kit})$$

- We do the same in regression were we model $E(y_i | x_{i1} \dots x_{ik})$
- The model is of the form

$$\alpha_i = \theta_1 \bar{x}_{i1} + \dots + \theta_k \bar{x}_{ik} + \nu_i$$

$$E(\nu_i | x_{i1} \dots x_{ik}) = 0$$

$$E(\alpha_i | x_{1it} \dots x_{kit}) = \theta_1 \bar{x}_{i1} + \dots + \theta_k \bar{x}_{ik}$$

- This implies that we are back into the random-effects framework and:

$$E(y_{it} | x_{1it} \dots x_{kit}) = \beta_1 x_{1it} + \dots + \beta_k x_{kit} + \theta_1 \bar{x}_{i1} + \dots + \theta_k \bar{x}_{ik}$$

Fixed-Effects Estimation

```
. xtreg tincome tenure c.age##c.age i.hispanic i.male, fe
note: 1.hispanic omitted because of collinearity
note: 1.male omitted because of collinearity
```

```
Fixed-effects (within) regression      Number of obs      =      30213
Group variable: id                    Number of groups   =      4343
R-sq:  within = 0.4635                 Obs per group: min =      4
      between = 0.2218                 avg =              7.0
      overall  = 0.3644                 max =              9
                                         F(3,25867)        =      7449.50
corr(u_i, Xb) = -0.1075                 Prob > F           =      0.0000
```

tincome	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
tenure	20.00194	.5618164	35.60	0.000	18.90075	21.10313
age	70.32832	159.8937	0.44	0.660	-243.0722	383.7288
c.age#c.age	36.37289	3.476712	10.46	0.000	29.55834	43.18744
1.hispanic	0	(omitted)				
1.male	0	(omitted)				
_cons	-13092.78	1826.164	-7.17	0.000	-16672.17	-9513.4
sigma_u	4941.7061					
sigma_e	5614.1177					
rho	.4365569	(fraction of variance due to u_i)				

```
F test that all u_i=0:      F(4342, 25867) =      5.24      Prob > F = 0.0000
```

Mundlak-Chamberlain Estimates

```
. quietly xtreg tincome tenure c.age##c.age, fe
. estimates store fe
. local muestra = e(sample)
. by id: egen mtenure      = mean(tenure)    if `muestra´
. by id: egen double mage  = mean(age)      if `muestra´
. by id: egen mmale        = mean(male)      if `muestra´
. by id: egen mhispanic    = mean(hispanic) if `muestra´
. by id: egen double mage2 = mean(age^2)   if `muestra´
. quietly xtreg tincome tenure c.age##c.age mtenure mhispanic mmale mage mage2
. estimates store mundlak
```

Mundlak-Chamberlain vs. Fixed Effects

```
. estimates table fe mundlak
```

Variable	fe	mundlak
tenure	20.001943	20.001943
age	70.328319	70.328319
c.age#c.age	36.372886	36.372886
mtenure		16.562624
mhispanic		-259.96981
mmale		2248.8968
mage		3504.7492
mage2		-91.017878
_cons	-13092.784	-47461.511

Testing

```
. estimates restore mundlak
(results mundlak are active now)
. test mtenure mage mmale mhispanic mage2
( 1) mtenure = 0
( 2) mage = 0
( 3) mmale = 0
( 4) mhispanic = 0
( 5) mage2 = 0
      chi2( 5) = 700.54
      Prob > chi2 = 0.0000
```

Objects of Interest

- Conditional expectation

```
. predict xbeta  
(option xb assumed; fitted values)
```

- For the marginal effects

$$\frac{\partial E(y_{it} | x_{1it} \dots x_{kit})}{\partial x_{kit}} = \beta_k + \frac{1}{T} \theta_k$$

- For our example

$$\beta_{age} + (1/10)\beta_{mage} + 2age\beta_{agesq} + (2/10)age\beta_{agesq}$$

Average Marginal Effects

```
. margins, ///
> expression(_b[age] + (1/10)*_b[mage] + _b[c.age#c.age]*2*age + ///
> (1/10)*_b[mage2]*2*age) over(year)
Warning: expression() does not contain predict() or xb().
Predictive margins                                Number of obs   =       30213
Model VCE      : Conventional
Expression     : _b[age] + (1/10)*_b[mage] + _b[c.age#c.age]*2*age + (1/10)*_b[mage2]*2*age
over          : year
```

	Delta-method				
	Margin	Std. Err.	z	P> z	[95% Conf. Interval]
year					
79	1482.805	30.85434	48.06	0.000	1422.332 1543.279
80	1500.187	28.76721	52.15	0.000	1443.804 1556.569
81	1524.764	25.90806	58.85	0.000	1473.985 1575.542
82	1548.672	23.26417	66.57	0.000	1503.075 1594.269
83	1583.816	19.73891	80.24	0.000	1545.128 1622.503
84	1637.464	15.77607	103.79	0.000	1606.543 1668.384
85	1686.49	14.69991	114.73	0.000	1657.678 1715.301
86	1740.62	16.91648	102.89	0.000	1707.464 1773.775
87	1794.102	21.56305	83.20	0.000	1751.839 1836.365
88	1848.186	27.50478	67.20	0.000	1794.277 1902.094

Simulations for Linear Panel Data Models

Table : Averages Marginal Effects

Simulations for `xtreg` with $N = 400$ and $T = 5$

ESTIMATOR	TRUE EFFECTS	ESTIMATED EFFECTS	COVERAGE
MUNDLAK	$x_1 = -.6$	-.602	.952
	$x_2 = .6$.599	.946
	$x_3^* = .36$.363	.939
	$x_4 = .36$.36	.951
R.EFFECTS	$x_1 = -.6$	-.602	.949
	$x_2 = .6$.599	.947
	$x_3^* = .36$.314	.841
	$x_4 = .36$.311	.499
F.EFFECTS	$x_1 = -.6$.	.
	$x_2 = .6$.	.
	$x_3^* = .36$.301	.784
	$x_4 = .36$.299	.32

*The variable x_3 is discrete. Thus, this is a population treatment effect and the estimates are ATEs.

Nonlinear Panel-Data Models: Probit and Logit

- If we are interested in average treatment effects and average effects over the population, population averaged models are best
- Probit panel-data models for fixed-effects are not available in Stata
- Logit fixed-effects models will drop panels that have invariant outcomes
- The main problem is that we cannot get rid of the time-invariant random disturbance
- No numerical integration

Objects of Interest

- The models are given by

$$y_{it} = \begin{cases} 1 & \text{if } y_{it}^* = \beta_0 + \beta_1 x_{1it} + \dots + \beta_k x_{kit} + \alpha_i + \varepsilon_{it} > 0 \\ 0 & \text{otherwise} \end{cases}$$

- The object of interest for the probit fixed-effects model is

$$P(y_{it}|x_{1it}, \dots, x_{ikt}, \alpha_i) = \Phi(\beta_0 + \beta_1 x_{1it} + \dots + \beta_k x_{kit} + \alpha_i)$$

- Using the Mundlak-Chamberlain approach

$$\Phi(\beta_0 + \beta_1 x_{1it} + \dots + \beta_k x_{kit} + \theta_1 \bar{x}_{i1} + \dots + \theta_k \bar{x}_{ik})$$

- The marginal effects for x_{1it} :

$$\phi(\beta_0 + \beta_1 x_{1it} + \dots + \beta_k x_{kit} + \theta_1 \bar{x}_{i1} + \dots + \theta_k \bar{x}_{ik}) \left(\beta_1 + \frac{1}{T} \theta_1 \right)$$

- For a discrete x_{1it}

$$\begin{aligned} & \Phi(\beta_0 + \beta_1 + \dots + \beta_k x_{kit} + \theta_1 \bar{x}_{i1} + \dots + \theta_k \bar{x}_{ik}) \\ - & \Phi(\beta_0 + \dots + \beta_k x_{kit} + \dots + \theta_k \bar{x}_{ik}) \end{aligned}$$

Probit Example

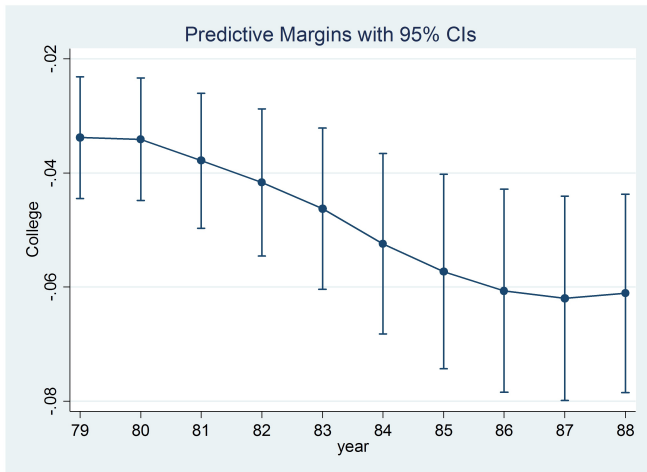
```
. quietly xtprobit nevermarried age i.college mage mmale mcollege, nolog pa
. generate xb = _b[age]*age + _b[mage]*mage + _b[mmale]*mmale + _b[_cons] ///
> + _b[mcollege]*mcollege
. margins, over(year) ///
> expression(normal(xb + _b[1.college]) - ///
> normal(xb))
```

Warning: expression() does not contain predict() or xb().

```
Predictive margins          Number of obs   =       30213
Model VCE      : Conventional
Expression    : normal(xb + _b[1.college]) - normal(xb)
over          : year
```

	Delta-method					
	Margin	Std. Err.	z	P> z	[95% Conf. Interval]	
year						
79	-.0337846	.0054447	-6.21	0.000	-.0444561 -.0231131	
80	-.0341023	.0054792	-6.22	0.000	-.0448412 -.0233633	
81	-.0378459	.0060266	-6.28	0.000	-.0496579 -.0260339	
82	-.0416477	.00657	-6.34	0.000	-.0545246 -.0287707	
83	-.0462502	.0072155	-6.41	0.000	-.0603923 -.032108	
84	-.0524056	.0080612	-6.50	0.000	-.0682052 -.036606	
85	-.0572666	.0086965	-6.59	0.000	-.0743114 -.0402218	
86	-.0606502	.0090759	-6.68	0.000	-.0784386 -.0428617	
87	-.0619659	.0091375	-6.78	0.000	-.079875	-.0440568
88	-.0610928	.0088768	-6.88	0.000	-.078491	-.0436947

Logit vs. Probit: Graphically



Simulations for Linear Panel Data Models

Table : Averages Effects Over Individuals for year 2009:

Simulation Results for `xtprobit` with $N = 400$ and $T = 5$

ESTIMATOR	TRUE EFFECTS	ESTIMATED EFFECTS	COVERAGE
PA MUNDLAK	$x_1 = -.132$	-.132	.951
	$x_2 = .132$.132	.937
	$x_3^* = .067$.069	.939
	$x_4 = .079$.079	.947
R.EFFECTS	$x_1 = -.132$	-.159	.572
	$x_2 = .132$.159	.568
	$x_3^* = .067$.086	.859
	$x_4 = .079$.086	.896
PA	$x_1 = -.132$	-.133	.949
	$x_2 = .132$.132	.933
	$x_3^* = .067$.072	.953
	$x_4 = .079$.071	.827

*The variable x_3 is discrete. Thus, this is a population treatment effect and the estimates are ATEs.

Conclusions

- I presented a new way to obtain marginal effects using a Mundlak-Chamberlain approach
- Along the way we talked about margins and about objects of interest being the guiding principle to use and interpret estimation results
- Thank you