

Estimar efectos parciales con margins

David M. Drukker

Director of Econometrics
Stata

2010 Spanish Stata Users Group meeting
Madrid
September 2010

- Esta charla muestra como usar el comando `margins` para estimar los efectos parciales de una variable
- Hablamos de unos puntos importantes
 - En modelos no lineales, los efectos parciales evaluados en las medias pueden ser muy distintos que las medias de los efectos parciales
 - Los estimadores estándares; como los métodos de máximo verosimilitud (MV), de cuadrados mínimos, y de momentos generalizados; no requieren que la distribución de las covariables sea representativa
 - Estimar la media de un efecto parcial requiere que la distribución de las covariables sea representativa
- También hablamos del uso básico de las variables de factores en Stata

El sintaxis de las variables

- Stata tiene operadores que funcionan en variables de factores
 - un operator unario que especifica indicadores de una variables discreta
 - un operator unario que especifica que una variable es continua
 - un operator binario que especifica interacciones de la variables
 - un operator binario que especifica interacciones factoriales

Datos de ingresos

```
. use earn2b
. summarize age
```

Variable	Obs	Mean	Std. Dev.	Min	Max
age	7373	40.1968	13.22641	15	80

```
. tabulate educ3 hourly
```

educ3	hourly		Total
	nonhourly	hourly	
No high school diplom	192	766	958
HIGH SCHOOL DIPLOMA	616	1,641	2,257
SOME COLLEGE NO DEGRE	472	945	1,417
ASSOCIATE OCCUPATIONA	122	244	366
ASSOCIATE ACADEMIC	110	133	243
BACHELOR'S DEGREE	987	369	1,356
MASTER'S DEGREE	447	89	536
PROFESSIONAL DEGREE	110	18	128
DOCTORATE DEGREE	104	8	112
Total	3,160	4,213	7,373

regress con variables de factores

```
. regress lnearn age c.age#c.age i.educ3 i.hourly
```

Source	SS	df	MS			
Model	1866.97842	11	169.725311	Number of obs = 7352		
Residual	3525.27413	7340	.480282579	F(11, 7340) = 353.39		
				Prob > F = 0.0000		
				R-squared = 0.3462		
				Adj R-squared = 0.3453		
Total	5392.25255	7351	.733540001	Root MSE = .69302		

lnearn	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	.1284447	.0034719	37.00	0.000	.1216388	.1352507
c.age#c.age	-.0013821	.0000405	-34.09	0.000	-.0014615	-.0013026
educ3						
3	.3663099	.0272751	13.43	0.000	.3128428	.419777
4	.3965967	.0293683	13.50	0.000	.3390264	.454167
5	.5247704	.0432303	12.14	0.000	.4400267	.6095141
6	.5574536	.0505165	11.04	0.000	.4584268	.6564805
7	.7062318	.0314011	22.49	0.000	.6446767	.767787
8	.7281191	.0398533	18.27	0.000	.6499951	.8062431
9	.9653706	.0666575	14.48	0.000	.8347028	1.096038
10	.8957075	.0708855	12.64	0.000	.7567514	1.034663
i.hourly	-.2135234	.0186841	-11.43	0.000	-.2501496	-.1768972
_cons	3.373212	.0719173	46.90	0.000	3.232233	3.514191

Use la opción `coeflegend` para ver los nombres

```
. regress lnearn age c.age#c.age i.educ3 i.hourly, coeflegend
```

Source	SS	df	MS	
Model	1866.97842	11	169.725311	Number of obs = 7352
Residual	3525.27413	7340	.480282579	F(11, 7340) = 353.39
Total	5392.25255	7351	.733540001	Prob > F = 0.0000
				R-squared = 0.3462
				Adj R-squared = 0.3453
				Root MSE = .69302

lnearn	Coef.	Legend
age	.1284447	_b[age]
c.age#c.age	-.0013821	_b[c.age#c.age]
educ3		
3	.3663099	_b[3.educ3]
4	.3965967	_b[4.educ3]
5	.5247704	_b[5.educ3]
6	.5574536	_b[6.educ3]
7	.7062318	_b[7.educ3]
8	.7281191	_b[8.educ3]
9	.9653706	_b[9.educ3]
10	.8957075	_b[10.educ3]
i.hourly	-.2135234	_b[1.hourly]
_cons	3.373212	_b[_cons]

El sintaxis de interacciones

```
. regress llearn i.educ3 c.age#c.age c.age##i.hourly, vsquish
```

Source	SS	df	MS			
Model	1873.37108	12	156.114257	Number of obs =	7352	
Residual	3518.88146	7339	.479476967	F(12, 7339) =	325.59	
				Prob > F =	0.0000	
				R-squared =	0.3474	
				Adj R-squared =	0.3464	
Total	5392.25255	7351	.733540001	Root MSE =	.69244	

llearn	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ3						
3	.3672511	.0272535	13.48	0.000	.3138264	.4206757
4	.3954825	.0293452	13.48	0.000	.3379575	.4530076
5	.525017	.043194	12.15	0.000	.4403442	.6096897
6	.5596144	.0504776	11.09	0.000	.4606638	.6585649
7	.7089366	.0313835	22.59	0.000	.6474159	.7704572
8	.7212365	.0398645	18.09	0.000	.6430907	.7993824
9	.9621752	.0666073	14.45	0.000	.8316057	1.092745
10	.8775882	.0709997	12.36	0.000	.7384085	1.016768
c.age#c.age	-.0014171	.0000416	-34.04	0.000	-.0014987	-.0013355
age	.1345898	.0038557	34.91	0.000	.1270315	.1421481
1.hourly	-.0082572	.0592347	-0.14	0.889	-.1243743	.1078599
hourly#c.age						
1	-.0049327	.0013509	-3.65	0.000	-.0075809	-.0022845
_cons	3.178811	.0894314	35.54	0.000	3.0035	3.354122

Un modelo por datos binarios

- El modelo probit por datos binarios es uno de los modelos no lineales más aplicados
- La variable dependiente observada, y_i , toma los valores de 0 and 1 solamente
- Se puede motivar este modelo con una variable continua no observada y_i^*

$$y_i = \begin{cases} 1 & \text{si } y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \epsilon_i > 0 \\ 0 & \text{por lo demás} \end{cases}$$

- Si especificamos que $Pr(y = 1|\mathbf{x}) = F(\mathbf{x}\boldsymbol{\beta})$ es la función de distribución normal de ϵ_i condicional en \mathbf{x} nos da

$$\begin{aligned} Pr(y^* > 0|\mathbf{x}) &= Pr(\epsilon > -\mathbf{x}\boldsymbol{\beta}|\mathbf{x}) \\ &= Pr(\epsilon < \mathbf{x}\boldsymbol{\beta}|\mathbf{x}) \text{ la distribución es simétrica} \\ &= F(\mathbf{x}\boldsymbol{\beta}) \end{aligned}$$

Estimación e inferencia en el modelo probit

- Después de seleccionamos una función de distribución, tenemos un modelo completamente especificado
- El método de máximo verosimilitud es el método de estimación mas frecuente

Datos de accidentes

- Tenemos uno datos simulados acerca de que si cado individuo en grupo ha tenido un accidente de automóvil en el último año
 - `crash` es 1 si una persona ha tenido por los menos un accidente de automóvil en el último año
 - `cvalue` es el valor del automóvil de cado uno
 - `kids` es el número de niños que tiene cada persona
 - `tickets` es el número de multas recibidas en los últimos tres años
 - `male` es una variable que indica si la persona es masculino (1 por masculino)

Un ejemplo de probit

```
. use accidents2
. probit crash tickets traffic i.male, nolog
Probit regression
```

	Number of obs	=	948
	LR chi2(3)	=	720.22
	Prob > chi2	=	0.0000
	Pseudo R2	=	0.8561

```
Log likelihood = -60.522949
```

crash	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
tickets	2.464657	.2768335	8.90	0.000	1.922073	3.00724
traffic	.159089	.0604682	2.63	0.009	.0405735	.2776045
1.male	5.892127	.7758214	7.59	0.000	4.371545	7.412709
_cons	-12.63666	1.529302	-8.26	0.000	-15.63403	-9.639279

Note: 516 failures and 13 successes completely determined.

```
. estimates store probit1
```

Interpretando los parámetros estimados

- El signo del coeficiente da la dirección del efecto pero no es el efecto marginal
- Los coeficientes estimados son consistentes por $\frac{\beta}{\sigma}$, entonces sus valores están en unidades de la desviación estándar de los errores
- Los efectos marginales en el punto $\tilde{\mathbf{x}}$ son

$$\frac{\partial E[y|\mathbf{x}]}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\tilde{\mathbf{x}}} = \frac{\partial F(\mathbf{x}\beta)}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\tilde{\mathbf{x}}} = f(\tilde{\mathbf{x}}\beta)\beta$$
- Los efectos marginales relativos no dependen en \mathbf{x}

$$\frac{\frac{\partial F(\mathbf{x}\beta)}{\partial x_j}}{\frac{\partial F(\mathbf{x}\beta)}{\partial x_k}} = \frac{f(\mathbf{x}\beta)\beta_j}{f(\mathbf{x}\beta)\beta_k} = \frac{\beta_j}{\beta_k}$$

- Use `testnl` para revisar hipótesis acerca de los efectos relativos

```
. testnl _b[1.male]/_b[tickets] = 2
(1)  _b[1.male]/_b[tickets] = 2
      chi2(1) =      8.86
      Prob > chi2 =    0.0029
```

Efectos marginales

- Lo bueno
 - El método de MV nos da todo lo que necesitamos para hacer estimación e inferencia acerca del efecto marginal en el punto \tilde{x}
- Lo malo
 - Tenemos que escoger \tilde{x}
- Use `margins` para estimar efectos en un punto \tilde{x}
- Es comun usar $\tilde{x} = \bar{x}$ cuando las variables x son continuas
- Vean [Long and Freese(2006)] y [Cameron and Trivedi(2009)] por más información acerca de como interpretar los estimatos de los parametros

Efectos marginales en las medias con margins

```
. margins , dydx(tickets traffic) atmeans
```

```
Conditional marginal effects           Number of obs   =           948
Model VCE      : OIM
Expression    : Pr(crash), predict()
dy/dx w.r.t.  : tickets traffic
at            : tickets      =    1.436709 (mean)
               traffic       =    5.201121 (mean)
               0.male        =    .5327004 (mean)
               1.male        =    .4672996 (mean)
```

	Delta-method					[95% Conf. Interval]	
	dy/dx	Std. Err.	z	P> z			
tickets	2.45e-07	8.06e-07	0.30	0.762	-1.34e-06	1.82e-06	
traffic	1.58e-08	5.14e-08	0.31	0.759	-8.49e-08	1.17e-07	

- Los efectos estimados para tickets y traffic son muy peque/ nos

Los cálculos de efectos marginales

```
. estat summarize
```

```
Estimation sample probit          Number of obs =    948
```

Variable	Mean	Std. Dev.	Min	Max
crash	.1624473	.3690553	0	1
tickets	1.436709	1.849456	0	7
traffic	5.201121	2.924058	.005189	9.99823
1.male	.4672996	.4991929	0	1

```
. matrix list r(stats)
```

```
r(stats)[4,4]
```

```

      mean      sd      min      max
crash  .16244726  .36905531      0      1
tickets 1.4367089  1.8494562      0      7
traffic 5.2011207  2.9240582  .00518857  9.9982338
1.male  .46729958  .49919289      0      1
```

```
. matrix r = r(stats)
```

```
. scalar f1 = normalden(_b[tickets]*r[2,1]+_b[traffic]*r[3,1]          ///  
>      +_b[1.male]*r[4,1] + _b[_cons])
```

```
. display f1*_b[tickets]
```

```
2.446e-07
```

```
. display f1*_b[traffic]
```

```
1.579e-08
```

Estimar efectos de variables discretas con margins

```
. margins , dydx(male) atmeans
Conditional marginal effects           Number of obs   =           948
Model VCE      : OIM
Expression    : Pr(crash), predict()
dy/dx w.r.t.  : 1.male
at            : tickets      =    1.436709 (mean)
              : traffic      =    5.201121 (mean)
              : 0.male       =    .5327004 (mean)
              : 1.male       =    .4672996 (mean)
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
1.male	.0087485	.007247	1.21	0.227	-.0054553	.0229523

Note: dy/dx for factor levels is the discrete change from the base level.

Los cálculos de efectos de variables discretas

```
. estat summarize
```

```
Estimation sample probit           Number of obs =    948
```

Variable	Mean	Std. Dev.	Min	Max
crash	.1624473	.3690553	0	1
tickets	1.436709	1.849456	0	7
traffic	5.201121	2.924058	.005189	9.99823
1.male	.4672996	.4991929	0	1

```
. matrix list r(stats)
```

```
r(stats)[4,4]
```

	mean	sd	min	max
crash	.16244726	.36905531	0	1
tickets	1.4367089	1.8494562	0	7
traffic	5.2011207	2.9240582	.00518857	9.9982338
1.male	.46729958	.49919289	0	1

```
. matrix r = r(stats)
```

```
. local xb0 = _b[tickets]*r[2,1]+_b[traffic]*r[3,1] + _b[_cons]
```

```
. display normal(`xb0'+_b[1.male]) - normal(`xb0`)
```

```
.00874852
```

Medias de efectos parciales

- La media de un efecto parcial de x_k is

$$\frac{\beta_k}{N} \sum_{i=1}^N f(\mathbf{x}_i \boldsymbol{\beta})$$

si x_k es continua

- Si x_k es discreta, la media del efecto parcial es la media de los cambios en las probabilidades predecidas

El efecto evaluado en las medias versus la media de un efecto

- Un efecto marginal evaluado en un punto es un estimador del efecto marginal en este punto
 - Se puede interpretar como efecto marginal por una persona, donde el punto describe la persona
 - El estimador es conditional en punto escogido
- La media de un efecto marginal es un estimador de la media del efecto en la población
 - La distribución de las covariables en la muestra tiene que ser representativa de la distribución de las covariables en la población para que el estimador sea consistente
- El efecto marginal evaluado en las medias y la media de los efectos marginales son objetos distintos y pueden tener valores muy distintos
 - Definimos que $g(\mathbf{x}) = \frac{\partial F(\mathbf{x})}{\partial x}$
 - Si $g(\cdot)$ es no lineal
$$g(\bar{\mathbf{x}}) \xrightarrow{p} g(E[\mathbf{x}]) \neq E[g(\mathbf{x})] \xleftarrow{p} N^{-1} \sum_{i=1}^N g(\mathbf{x}_i)$$

La media de los efectos marginales por margins

```
. margins , dydx(tickets traffic)
```

```
Average marginal effects          Number of obs   =          948
Model VCE      : OIM
Expression    : Pr(crash), predict()
dy/dx w.r.t.  : tickets traffic
```

	Delta-method					[95% Conf. Interval]	
	dy/dx	Std. Err.	z	P> z			
tickets	.0857818	.0031049	27.63	0.000	.0796963	.0918672	
traffic	.0055371	.0020469	2.71	0.007	.0015251	.009549	

- Los estimatos de las medias de los efectos son muy distintos de los estimatos de los efectos evaluados en las medias

Los cálculos de las medias de los efectos marginales

```
. predict double xb, xb  
. generate double me_tickets = normalden(xb)*_b[tickets]  
. generate double me_traffic = normalden(xb)*_b[traffic]  
. summarize me_tickets me_traffic if e(sample)
```

Variable	Obs	Mean	Std. Dev.	Min	Max
me_tickets	948	.0857818	.2090093	4.59e-35	.9818822
me_traffic	948	.0055371	.0134912	2.96e-36	.0633787

Las medias de efectos discretos por margins

```
. margins , dydx(male)
```

```
Average marginal effects
```

```
Number of obs = 948
```

```
Model VCE : OIM
```

```
Expression : Pr(crash), predict()
```

```
dy/dx w.r.t. : 1.male
```

	Delta-method				
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]
1.male	.2092058	.0105149	19.90	0.000	.188597 .2298145

Note: dy/dx for factor levels is the discrete change from the base level.

Los cálculos de las medias de los efectos discretos

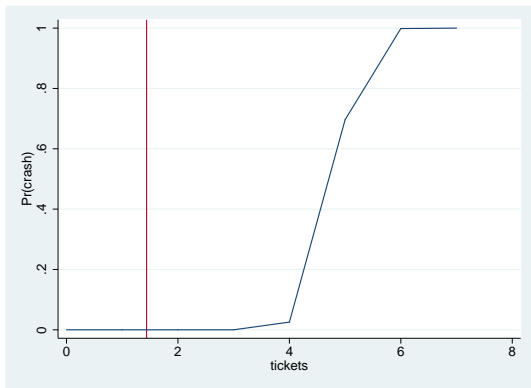
```
. generate double xb0 = _b[tickets]*tickets + _b[traffic]*traffic + _b[_cons]
. generate double de = normal(xb0 + _b[1.male]) - normal(xb0)
. summarize de
```

Variable	Obs	Mean	Std. Dev.	Min	Max
de	948	.2092058	.3605846	7.79e-12	.996267

Tratando tickets como una variable discreta I

```
. estimates restore probit1
(results probit1 are active now)
. preserve
.      replace tickets      = _n-1 in 1/8
(7 real changes made)
.      replace male        = .4672996 in 1/8
(8 real changes made)
.      replace traffic     = 5.2011 in 1/8
(8 real changes made)
.      predict Fhat in 1/8
(option pr assumed; Pr(crash))
(940 missing values generated)
.      graph twoway line Fhat tickets in 1/8, xline(1.4367)
. restore
```


Tratando tickets como una variable discreta II



- La media de tickets es 1.43, y la inclinación de la función de probabilidad es básicamente 0 cuando tickets tiene un valor menos que 3
- Cuando tickets tiene un valor de 3 o más, la inclinación de la función de probabilidad es más que 0

Tratando tickets como una variable discreta III

```
. margins , at(tickets = (0 1 2 3)) post coeflegend
```

```
Predictive margins                                Number of obs   =           948
Model VCE      : OIM
Expression     : Pr(crash), predict()
1._at          : tickets      =           0
2._at          : tickets      =           1
3._at          : tickets      =           2
4._at          : tickets      =           3
```

	Margin	Legend
_at		
1	1.66e-09	_b[1bn._at]
2	.0001208	_b[2._at]
3	.0549183	_b[3._at]
4	.4052946	_b[4._at]

Tratando tickets como una variable discreta IV

```
. lincom _b[2._at] - _b[1bn._at]
( 1) - 1bn._at + 2._at = 0
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
(1)	.0001208	.0001671	0.72	0.470	-.0002067	.0004484

```
. lincom _b[3._at] - _b[2._at]
( 1) - 2._at + 3._at = 0
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
(1)	.0547975	.0177313	3.09	0.002	.0200448	.0895502

```
. lincom _b[4._at] - _b[3._at]
( 1) - 3._at + 4._at = 0
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
(1)	.3503763	.0225727	15.52	0.000	.3061346	.3946179

```
. estimates restore probit1
(results probit1 are active now)
```

Tratando tickets como una variable discreta V

```
. generate double xb_b = _b[_cons] + _b[traffic]*traffic + _b[1.male]*male
. generate double pr0 = normal(xb_b + 0*_b[tickets]) // prob when tickets=0
. generate double pr1 = normal(xb_b + 1*_b[tickets]) // prob when tickets=1
. generate double pr2 = normal(xb_b + 2*_b[tickets]) // prob when tickets=2
. generate double pr3 = normal(xb_b + 3*_b[tickets]) // prob when tickets=3
. generate pe_d01 = pr1-pr0
. sum pe_d01
```

Variable	Obs	Mean	Std. Dev.	Min	Max
pe_d01	948	.0001208	.0003387	2.52e-24	.0031395

```
. generate pe_d12 = pr2-pr1
. sum pe_d12
```

Variable	Obs	Mean	Std. Dev.	Min	Max
pe_d12	948	.0547975	.0794281	1.05e-14	.3911403

```
. generate pe_d23 = pr3-pr2
. sum pe_d23
```

Variable	Obs	Mean	Std. Dev.	Min	Max
pe_d23	948	.3503763	.3749537	1.11e-07	.7821735

Datos faltantes e efectos parciales I

- Estimadores del método MV son consistentes cuando faltan datos por alguna razón aleatoria
(La condición se conoce como “missing at random” en la literatura)
 - “Missing at random” permite que el mecanismo que causa que faltan datos depende en las covariables y un error aleatorio mientras que el error es independiente de lo demás en el modelo
 - También se conoce como “selection on observables”
 - Vean [Cameron and Trivedi(2005)] y [Wooldridge(2002)] por introducciones y resultados formales
 - La distribución de las covariables no tiene que ser representativa a la distribución de las variables en la población para que el método de MV sea consistente

Datos faltantes e efectos parciales II





- Estimar la media en la población de un efecto parcial requiere el supuesto de que la muestra de las covariables sea representativa de la distribución en la población
 - Se conoce como “missing completely at random” en la literatura
 - “Missing completely at random” es un supuesto mucho mas restricto que “missing at random”
 - “Missing completely at random” require que el mecanismo que causa que falten datos es independiente de todo los demás en el modelo y que no puede depender en las covariables
 - A veces, podemos usar pesas para que la muestra de las covariables sea representativa
 - Necesitamos que la muestra de las covariables sea representativa por que necesitamos que

$$N^{-1} \sum_{i=1}^N w_i g(\mathbf{x}_i) \xrightarrow{P} E[g(\mathbf{x})]$$

Datos faltantes e efectos parciales III

- También necesitamos una muestra de las covariables representativa para estimar $E[\mathbf{x}]$
- Si seleccionamos $\tilde{\mathbf{x}}$ en una manera que no depende en la muestra, el método de MV nos da todo para estimar y hacer inferencia acerca de los efectos parciales
 - El método de MV solo requiere “missing at random”

Bibliography

-  Cameron, A. Colin and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and applications*, Cambridge: Cambridge University Press.
-  ———. 2009. *Microeconometrics Using Stata*, College Station, Texas: Stata Press.
-  Long, J. Scott and Jeremy Freese. 2006. *Regression models for categorical dependent variables using Stata*, College Station, Texas: Stata Press.
-  Wooldridge, Jeffrey. 2002. *Econometric Analysis of Cross Section and Panel Data*, Cambridge, Massachusetts: MIT Press.