

# Week 6: Interpreting model results: Marginal and incremental effects

Marcelo Coca Perrailon

University of Colorado  
Anschutz Medical Campus

Health Services Research Methods I  
HSMP 7607  
2020

These slides are part of a forthcoming book to be published by Cambridge University Press. For more information, go to [perrailon.com/PLH](http://perrailon.com/PLH). ©This material is copyrighted. Please see the entire copyright notice on the book's website.

# Outline

- Alternative title for today's class: *it's all about (counterfactual) predictions to interpret model coefficients in the scale of interest*
- Marginal vs incremental effects
- Analytical vs numerical derivatives, one- and two-sided
- Delta-method standard errors
- Replicating margins command output
- Interactions in logistic models
- Testing interactions in logistic models in the probability scale with margins command SEs (delta method)
- GLM models, two-part models

# Why do we need marginal effects?

- In a simple linear model like  $y = \beta_0 + \beta_1 \text{age} + \beta_2 \text{male} + \epsilon$  we can easily interpret the coefficients
- It became more difficult when we had non-linear terms, for example:  
 $y = \beta_0 + \beta_1 \text{age} + \beta_2 \text{age}^2 + \beta_3 \text{male} + \epsilon$
- We saw that using the **analytical derivative** made interpretation a lot easier:  
$$\frac{\partial E[y|\text{age}, \text{male}]}{\partial \text{age}} = \beta_1 + 2\beta_2 \text{age}$$
- No single effect of age, the effect depends on age; that is, one effect at age 20, another at 50, etc. Or you could say that the difference (not effect) in  $E[Y|\mathbf{X}]$  at different ages depends on the value of age (when you change age by a small amount)
- You can find the effect or the difference by plugging in numbers for age in the above expression:  $\frac{\partial E[y|\text{age}, \text{male}]}{\partial \text{age}} \Big|_{\text{age}=a} = \beta_1 + 2\beta_2(a)$

## Why do we need marginal effects?

- With interactions, it was more complicated:

$$y = \beta_0 + \beta_1 \text{age} + \beta_2 \text{male} + \beta_3 \text{male} * \text{age}$$

- But similar in the sense that the effect of age now depends on sex; or the other way around, the effect of sex depends on age
- With simple models, taking the derivative still helps with interpretation
- With a dummy variable, we calculated **incremental effects** (differences or contrasts from 1 vs 0, holding other covariates constant):

$$\frac{\Delta E[y|\text{age}, \text{male}]}{\Delta \text{male}} = \beta_2 + \beta_3 \text{age}$$

- Centering also helps with parameter interpretation:

$$y = \beta_0 + \beta_1(\text{age} - m) + \beta_2 \text{male} + \beta_3 \text{male} * (\text{age} - m)$$

- If  $m$  is average age, then  $\beta_2$  is  $E[Y|X]$  for males versus females of average age. What is  $\beta_1$ ?

# Why do we need marginal effects?

- In the logistic model, things got complicated very quickly:  
$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \text{age} + \beta_2 \text{male}$$
- We could present odds ratios:  $e^{\beta_1}$  and  $e^{\beta_2}$  but we know that **odds ratios** can be **misleading since they tend to be interpreted as relative probabilities**, which in general they are not
- A simple example with no covariates: The probability of death in a control group is 0.40. The probability of death in the treatment group is 0.20
- The **odds ratio** is:  $\frac{\frac{0.2}{1-0.2}}{\frac{0.4}{1-0.4}} = 0.375$ . The treatment reduces the odds of death by a factor of 0.375. Or in reverse, the odds of death are 2.67 higher in the control group  $\left(\frac{1}{0.375}\right)$
- But that's not the **relative risk**, even though most people would forget that they are odds ratios and interpret the odds ratio as a relative risk. The relative risk is  $\frac{0.2}{0.4} = 0.5$ . The **probability** of death is reduced by half in the treatment group

## Why do we need marginal effects?

- Note something else that is important. With odds ratios and relative risks, **we don't have a sense of the magnitude**
- Same example but now the probability of death in the control group is 0.0004 and 0.0002 in the treatment group. The odds ratio is still 0.375 and the relative risk is still 0.5
- The **magnitudes** are of course quite different. A difference of  $(0.0004 - 0.0002)$  0.0002 or 0.02 percent points, which is not a big difference in the scale we care about
- As we will see, marginal effects in non-linear models are a way of presenting model results in the scale of interest, not in the estimation scale. In the case of logit and probit models, we would like to know **differences in probabilities**, which is more informative than odds ratios and relative risks
- The same issue arises in other models like Poisson, negative binomial, probit and so on

## It's about **numerical** derivatives

- The main idea of marginal effects is to use the estimated model to make **predictions changing the value of a covariate** to evaluate its effect on the outcome
- **Incremental** changes are easier to understand: a variable changes from 0 to 1. With continuous variables, we need derivatives
- Our goal will be to take **numerical derivatives** of functions for which analytical derivatives are complicated
- As we saw, in the model  $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$  we can take the derivative with respect to  $X$ :  $\frac{dE[Y_i|X_i]}{dX} = \beta_1 + 2\beta_2 X_i$
- In the logistic model, our estimation scale is the log-odds but we would like to **interpret our model in the probability** scale
- We saw that in the probability scale our model is 
$$p(Y = 1|\mathbf{X}; \beta) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}$$
- **Please note:** In these notes,  $P(X)$  really means  $P(Y = 1|X)$  or  $P(Y|X)$

## It's about **numerical** derivatives

- We know that for a function  $y = e^X$  the derivative with respect to  $X$  is  $\frac{dy}{dX} = e^X$  so the above expression is not going to simplify
- In fact, the effect of  $X_1$  depends on the value of  $X_1$  and the **values of all other covariates**:

$$\frac{\partial p(Y|\mathbf{X};\beta)}{\partial X_1} = \beta_1 \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{(1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)})^2}$$

- We would need to choose values for all the covariates in the model to calculate the effect of  $X_1$
- Please make sure that you absorb this because it's the most important complication of model interpretation when we move away from the world linear additive effects (which in applied research is often): the effect of one variable depends on the value of **all** the other covariates in the model
- Rather than calculating analytical derivatives, we we will approximate the analytical derivative **numerically** using the definition of the derivative to compute the **average effect** of  $X_1$  on  $p$



# Derivatives

- The analytical derivative is a limit:

$$f'(x) \equiv \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

- All the **formulas for the derivative can be derived** using the definition and taking the limit. For example, an easy one for  $f(x) = X^2$

$$f'(x) = \lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h} = \frac{x^2 + 2xh + h^2 - x^2}{x} = \frac{2xh + h^2}{h} = 2x + h = 2x$$

- Numerically, that is, without finding the analytical formula, we could use the definition plugging in a number for  $h$  that is small enough. In that case:

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \approx \frac{f(x+h) - f(x)}{h}$$

- Computationally, it's not trivial to come up with a number  $h$  that is small but "large enough" so that computations are numerically accurate
- For more on choosing  $h$  see Gould, Pitblado, and Poi (2010), Chapter 1 and Greene (2017) Appendix E

## Two-sided derivatives

- The above approximation to the analytical derivative  $f'(x) \approx \frac{f(x+h)-f(x)}{h}$  is not the only way to define a derivative
- A **two-sided derivative** calculates a positive and negative small change  $h$ :
- $f'_2(x) \approx \frac{f(x+h)-f(x-h)}{2h}$
- The same issues about how to choose a small  $h$  remain
- We will calculate marginal effects “by hand” and then we will use the margins command
- We will use **both definitions of the derivative** but **Stata uses the two-sided version**
- Our calculations will be very close but Stata uses an additional iterative procedure that changes the value of  $h$  to achieve numerical accuracy

## Digression: Is it a unit change?

- In a model like  $y = \beta_0 + \beta_1x + \epsilon$  we have said that  $\beta_1$  is the change in  $E[y|\mathbf{X}]$  for a one unit change in  $x$ . And that's fine when the change is constant (linear)
- In the model  $y = \beta_0 + \beta_1x + \beta_2x^2 + \epsilon$ , the **marginal effect/change is no longer for a 1 unit change even though most people would interpret it that way when using marginal effects**. Do it by hand:
- Start with  $x = x^0$ . Then change by one unit to  $x^0 + 1$  and compare the two predictions for  $y$   
$$y^1 - y^0 = \beta_0 + \beta_1(x^0 + 1) + \beta_2(x^0 + 1)^2 - \beta_0 - \beta_1x^0 - \beta_2(x^0)^2$$
- You will find:  $y^1 - y^0 = \beta_1 + 2\beta_2x^0 + \beta_2$ , which is not the same as  $dy/dx = \beta_1 + 2\beta_2x$  evaluated at  $x = x^0$ . There is the extra  $\beta_2$  term lingering there

## Digression: Is it a unit change?

- Easier to get some intuition if we use the definition of the derivative:

$$\lim_{h \rightarrow 0} \frac{\beta_0 + \beta_1(x+h) + \beta_2(x+h)^2 - \beta_0 - \beta_1x - \beta_2x}{h} = \beta_1 + 2\beta_2x + \beta_2h$$

- $\beta_2$  is the coefficient for  $x^2$ , the curvature. The smaller  $\beta_2$  the smaller the curvature and closer to a straight line
- So they are closer when the curve looks more like a line
- **Conclusion:** marginal effects for **continuous** variables are **very small changes** in  $x$ , not unit changes. It's a derivative, which is a **limit** then  $h \rightarrow 0$
- Both are unit (or multiples of units) changes if straight lines

## Simple example of marginal effects

- We are going to first use marginal effects where we **don't quite need** them to illustrate some ideas, in particular how to estimate standard errors. Suppose we estimate this linear/OLS model:

$$bw_i = \beta_0 + \beta_1 smoke_i + \beta_2 mage + \beta_3 smoke_i * mage_i + \epsilon_i$$

- What is the effect of mother's age on birthweight? What is the effect of smoking? We have done this many times:

$$\frac{\partial E[bw|\mathbf{X}]}{\partial mage} = \hat{\beta}_2 + \hat{\beta}_3 smoke$$

$$\frac{\Delta E[bw|\mathbf{X}]}{\Delta smoke} = \hat{\beta}_1 + \hat{\beta}_3 mage$$

- A marginal effect and an incremental effect, respectively
- Note that each of them is a function of the estimated parameters  $\hat{\beta}' = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$  and the data
- In this case, we can just use the coefficients and the formulas above to find marginal and incremental effects
- **Cross-partial derivative:** Note that with two continuous variables the interaction coefficient  $\beta_3$  is a cross-partial derivative:  $\frac{\partial E[Y|\mathbf{X}]}{\partial x_1 \partial x_2} = \beta_3$

# What about standard errors?

- We could also calculate the variance of the marginal effects (the square is the standard errors)
- **The variance is also a function of the parameters**, of course. For the first one above:

$$\text{var}\left(\frac{\partial E[bw|\mathbf{X}]}{\partial \text{mage}}\right) = \text{var}(\hat{\beta}_2) + \text{smoke}^2 \text{var}(\hat{\beta}_3) + 2 * \text{smoke} * \text{cov}(\hat{\beta}_2, \hat{\beta}_3)$$

- I just used basic rules of variances:  $\text{var}(aX) = a^2 \text{var}(x)$  and  $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$
- So the variance for non-smokers is just  $\text{var}(\hat{\beta}_2)$  while the variance for those who smoked is  $\text{var}(\hat{\beta}_2) + \text{var}(\hat{\beta}_3) + 2 * \text{cov}(\hat{\beta}_2, \hat{\beta}_3)$
- We can estimate our model and use the variance-covariance matrix for calculations

# Standard error of marginal effects

- I created interactions by hand because it makes the variance-covariance shorter and easier to display than using the factor variable syntax

```
use https://www.stata-press.com/data/r16/cattaneo2, clear
```

```
gen sm_age = mbsmoke * mage  
qui reg bweight mbsmoke mage sm_age
```

```
matrix list e(V)  
symmetric e(V) [4,4]  
      mbsmoke      mage      sm_age      _cons  
mbsmoke  10733.267  
  mage   71.343306   2.6610229  
sm_age  -403.72658  -2.6610229   15.868305  
  _cons -1997.5365  -71.343306   71.343306   1997.5365
```

```
* var(beta2)  
di e(V) [2,2]  
2.6610229  
* var (beta3)  
di e(V) [3,3]  
15.868305  
* covariance b2,b3  
di e(V) [3,2]  
-2.6610229
```

```
* when smoke is 0 variance of marginal effect is  
di sqrt(e(V) [2,2])  
1.6312642
```

```
* when smoke is 1  
. di sqrt(e(V) [2,2] + e(V) [3,3] + 2*e(V) [3,2])  
3.6341825
```

# Standard error of marginal effects with margins command

- We can match the SEs using the margins command (note that here we need to use the factor variable syntax or you won't get the correct marginal effects)
- But note that Stata tells you that SEs are **Delta Method standard errors**, not the analytical SEs as above, although in this easy case they are exactly the same

```
qui reg bweight i.mbsmoke##c.mage
margins, dydx(mage) at(mbsmoke=(0 1)) vsquish
```

```
Average marginal effects          Number of obs   =       4,642
Model VCE      : OLS
Expression    : Linear prediction, predict()
dy/dx w.r.t.  : mage
1. _at       : mbsmoke      =           0
2. _at       : mbsmoke      =           1
```

```
-----+-----
          |              Delta-method
          |              dy/dx   Std. Err.   t    P>|t|    [95% Conf. Interval]
-----+-----
mage     |
   _at   |
     1   |   11.36258   1.631264    6.97  0.000    8.164523   14.56063
     2   |   -3.950895  3.634182   -1.09  0.277   -11.07562    3.173831
-----+-----
```



# Big picture

- To recap using margins in a simple model like a linear/OLS model with interactions
- We can do two things **analytically**:
  - 1 We can calculate marginal effects
  - 2 We can calculate standard errors
- But in other models both marginal effects and standard errors are more complicated functions of parameters. Look again at the effect of one variable in the probability scale for a logistic model:

$$\frac{\partial p(Y|\mathbf{X};\beta)}{\partial X_1} = \beta_1 \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{(1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)})^2}$$

- Therefore, we will use:
  - 1 **Numerical derivatives** to approximate analytical derivatives
  - 2 The **delta method** to approximate standard error using the reported asymptotic errors (variance-covariance matrix)
- That's why Stata uses the delta method (default). In the example above, the analytical method and delta method approximation are the same
- In essence, **we go from analytical to numerical methods** when working with more complicated marginal or incremental effects

## Aside: Delta method versus unconditional variance-covariance estimation

- I'll focus on the delta method today but it's not the only option
- The `vce(unconditional)` option estimates margins variances using the linearization method which accounts for heteroskedasticity and clustering
- So this is the option to use with survey data or could use it with models estimated with the robust option
- But remember that this is an option for margins, not for the estimated model. In most cases for modeling we pay attention to the model SEs, not margins SEs (with some exceptions we will see later in this class)

# Delta method

- As we saw above, the issue is calculating standard errors of function of coefficients
- The delta method allows us to approximate the mean and variance of a function of random variables using a first-order Taylor approximation (you know that estimated parameters are random variables, right?)
- In this case, we are just interested in the variance and not the mean
- Let's say  $G(\beta)$  is a function of the parameters  $\beta$ , the delta method tell us that
$$\text{var}(G(\beta)) \approx \nabla G(\beta) \text{cov}(\beta) \nabla G(\beta)'$$
- $\nabla G(\beta)$  is the gradient or the vector of partial derivatives of  $G(\beta)$  (the Jacobian), which in this case is the partial derivative of the predictions with respect to coefficients and evaluated at  $\hat{\beta}$ .  $\text{var}(G(\beta))$  in conditional on the data  $\mathbf{X}$

# Delta method

- It looks a bit more complicated than what it is
- Think of it this way: marginal effects are a function of parameters. The delta method gives us an approximation to the variance of that function – we called it  $G(\beta)$
- One needed component in that calculation is the variance-covariance matrix  $cov(\beta)$ , which we obtain after estimating any model
- The second component is the Jacobian. The Jacobian is a matrix of partial derivatives of the margins or marginal effects with respect to each estimated parameter. For  $\frac{\partial E[bw|\mathbf{X}]}{\partial mbsmoke}$  above using margins (not marginal effects), the Jacobian would be a  $2 \times 6$  matrix in Stata because using the factor syntax the reference levels are included
- The Jacobian elements will be functions of coefficients

# Replicate delta method SEs

- We will use Stata's stored Jacobian after margins

```
qui reg bweight i.mbsmoke##c.mage
margins mbsmoke, nofvlabel
Predictive margins                                Number of obs   =       4,642
Model VCE    : OLS
Expression   : Linear prediction, predict()
```

```
-----+-----
```

		Delta-method				[95% Conf. Interval]	
	Margin	Std. Err.	t	P> t			
mbsmoke							
0	3409.435	9.221561	369.72	0.000	3391.356	3427.514	
1	3132.374	19.85928	157.73	0.000	3093.44	3171.308	

```
-----+-----
```

```
* Stored Jacobian from margins
```

```
matrix list r(Jacobian)
```

```
r(Jacobian)[2,6]
```

	0b.	1.		0b.mbsmoke#	1.mbsmoke#	
	mbsmoke	mbsmoke	mage	co.mage	c.mage	_cons
0.mbsmoke	0	0	26.504524	0	0	1
1.mbsmoke	0	1	26.504524	0	26.504524	1

```
* Save it
```

```
matrix J = r(Jacobian)
```

# Replicate delta method SEs

- The relevant calculation is matrix  $V_{rep} = J * e(V) * J'$ , which matches margin's  $r(V)$  (not the same as  $e(V)$  from the reg model)

```
* Stored variance-covariance matrix from reg model
matrix list e(V)
symmetric e(V) [6,6]

      0b.          1.
      mbsmoke     mbsmoke     mage     0b.mbsmoke#     1.mbsmoke#
0b.mbsmoke      0
1.mbsmoke      0     10733.267
  mage      0     71.343306     2.6610229
0b.mbsmoke#
  co.mage      0          0          0          0
1.mbsmoke#
  c.mage      0    -403.72658    -2.6610229      0     15.868305
  _cons      0    -1997.5365    -71.343306      0     71.343306     1997.5365
```

```
* Replicate delta method variance (of margins)
```

```
matrix Vrep = J*e(V)*J'
matrix list Vrep
symmetric Vrep[2,2]

      0.          1.
      mbsmoke     mbsmoke
0.mbsmoke     85.037195
1.mbsmoke    -1.631e-13     394.39086
* match margin output SE for nonsmoker
di sqrt(Vrep[1,1])
9.2215614
```

```
* matches stored variance
```

```
matrix list r(V)
symmetric r(V) [2,2]

      0.          1.
      mbsmoke     mbsmoke
0.mbsmoke     85.037195
1.mbsmoke    -1.631e-13     394.39086
```

# Data

- We will use birth weight data from Wooldridge (bcuse bwght)
- I created an indicator for low birth weight. Very low birth weight is about 50 ounces (1,500 grams) but here I used 100 ounces
- About 15% of children (mean of lw)

```
bwght      birth weight, ounces
faminc     1988 family income, $1000s
motheduc   mother's yrs of educ
cigs       cigs smked per day while preg
```

```
gen      lw = 0
replace lw = 1 if bwght < 100 & bwght != .
```

```
sum lw faminc motheduc cigs
```

Variable	Obs	Mean	Std. Dev.	Min	Max
lw	1,388	.1491354	.3563503	0	1
faminc	1,388	29.02666	18.73928	.5	65
motheduc	1,387	12.93583	2.376728	2	18
cigs	1,388	2.087176	5.972688	0	50

# Logit/logistic model

- We will model low birth weight as a function of cigarettes, mother's education, and family income

$$\log\left(\frac{lw_i}{1-lw_i}\right) = \beta_0 + \beta_1 cigs_i + \beta_2 faminc_i + \beta_3 motheduc_i$$

```
logit lw cigs faminc motheduc, nolog
```

```
Logistic regression                Number of obs   =    1,387
                                   LR chi2(3)         =    24.63
                                   Prob > chi2        =    0.0000
Log likelihood = -572.15891        Pseudo R2      =    0.0211
```

	lw	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	cigs	.0449006	.0104436	4.30	0.000	.0244316	.0653696
	faminc	-.0080855	.004801	-1.68	0.092	-.0174953	.0013243
	motheduc	.0031552	.037153	0.08	0.932	-.0696634	.0759738
	_cons	-1.678173	.4497551	-3.73	0.000	-2.559676	-.7966687



# Odds ratios

- We can show odds ratios: An additional cigarette increases the odds of low birth weight by 4.6%
- Of course, we now know that it does NOT mean that an additional cigarette increases the *relative probability* of low birth weight by 4.6% or that it's even a *difference* in probability, which is what we want
- Recall that the ratio of two numbers gives you percent changes. Percent change is  $\frac{a_2 - a_1}{a_1} = c$  or  $c \times 100$ . Rewrite as:  $\frac{a_2}{a_1} = c + 1$ . So from the ratio of two numbers, subtract 1 and multiply by 100 to get percent change

logit, or

```
Logistic regression                Number of obs   =    1,387
                                   LR chi2(3)         =    24.63
                                   Prob > chi2        =    0.0000
Log likelihood = -572.15891        Pseudo R2      =    0.0211
```

	lw	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
cigs		1.045924	.0109232	4.30	0.000	1.024733	1.067554
faminc		.9919471	.0047623	-1.68	0.092	.9826569	1.001325
motheduc		1.00316	.0372704	0.08	0.932	.9327077	1.078934
_cons		.1867149	.083976	-3.73	0.000	.0773298	.4508283

# Model

- We can also run our trusty linear/OLS model with the caveat that SEs are likely not right (but probably close) and that since low birth probability is (relatively) low we should be more careful
- Now, in the probability scale, an extra cigarette increases the probability of low birth weight by 0.7 percent points. With 10 cigarettes, 7 percent points (linear effects)

```
reg lw cig faminc motheduc, robust
```

```
Linear regression      Number of obs   =    1,387
                      F(3, 1383)             =     6.51
                      Prob > F              =    0.0002
                      R-squared             =    0.0212
                      Root MSE           =    .35304
```

```
-----+-----
```

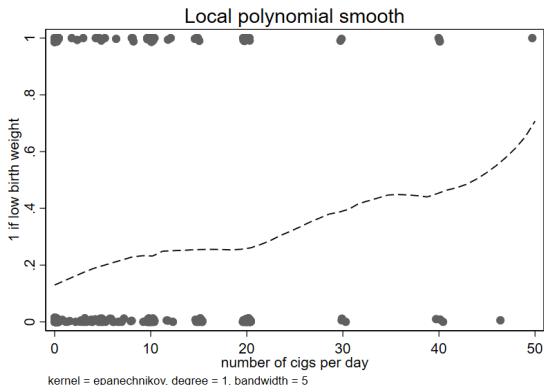
			Robust				
lw		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
cigs		.007757	.0020677	3.75	0.000	.0037009	.0118131
faminc		-.0009345	.0005785	-1.62	0.106	-.0020693	.0002004
motheduc		.0005403	.0042972	0.13	0.900	-.0078895	.00897
_cons		.1531912	.0532648	2.88	0.004	.0487027	.2576797

```
-----+-----
```

# A plot is always helpful

- A plot will help you understand the shape of the relationship of interest but remember that other variables may change the shape. The linearity suggest LPM or logit/probit are going to be similar
- Could use lowess, but Kernel-weighted local polynomial smoothing (lpoly) is another nonparametric option

```
lpoly lw cigs, bw(5) deg(1)  
graph export 1.png, replace
```



# Average Marginal Effect (AME) algorithm

- This is the algorithm to compute the numerical one-sided derivative  $\frac{\partial p(\mathbf{X};\beta)}{\partial x}$  in the probability scale
  - 1 Estimate the model
  - 2 For each observation  $i$ , calculate predictions in the probability scale using values of all the covariates as observed, let's call them  $\hat{Y}_{i0}$
  - 3 Increase  $x$  variable by a "small" amount  $h$  ( $x = x + h$ ) and calculate predictions again, denoted  $\hat{Y}_{i1}$ .  $h$  depends on the units of  $x$ :  
 $h = (|\bar{x}| + 0.001) \times 0.001$ , where  $\bar{x}$  is the mean of  $x$
  - 4 Calculate the difference in the two predictions as a fraction of  $h$  for each observation  $i$ :  $(\hat{Y}_{i1} - \hat{Y}_{i0})/h$
  - 5 The average numerical derivative is  $E\left[\frac{\hat{Y}_{i1} - \hat{Y}_{i0}}{h}\right] \approx \frac{\partial p(Y|\mathbf{X};\beta)}{\partial x}$
- Note that this algorithm follows the definition of the one-sided derivative. Also, **note the connection with predicting counterfactuals in treatment effects**
- Stata calculates two-sided derivatives. Steps 2 and 3 are different. One prediction is for  $x = x - h$  and the other is for  $x = x + h$ , with  $\Delta x = 2h$  in the denominator

# Average Marginal Effect (AME)

- Let's calculate AME for the cigarette variable using the typical formula for the analytical derivative

```
* Compute the initial "small change" h
qui sum cigs
scalar h = (abs(r(mean))+.0001)*.0001
di h
*.00020873
preserve
  qui logit lw cigs faminc motheduc, nolog
  * as is
  predict double lw_0 if e(sample)
  * Change cigs by a bit
  replace cigs = cigs + scalar(h)
  predict lw_1 if e(sample)
  * For each obs
  gen double dydx = (lw_1-lw_0)/scalar(h)
  * Average
  sum dydx
restore
```

Variable	Obs	Mean	Std. Dev.	Min	Max
dydx	1,387	.0055768	.0012444	.0040507	.0113006

- An additional cigarette increases the probability of low birth weight by 0.56 percent points

# Average Marginal Effect (AME)

## ■ Replicate using margins command

```
qui logit lw cigs faminc motheduc, nolog
margins, dydx(cigs)
```

```
Average marginal effects      Number of obs   =       1,387
Model VCE      : OIM
```

Expression : Pr(lw), predict()  
dy/dx w.r.t. : cigs

```
-----+-----
```

		Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]		
cigs	.0055782	.0012814	4.35	0.000	.0030666	.0080898	

```
-----+-----
```

- Very close but Stata uses the two-sided derivative. The formula  $h = (|\bar{x}| + .0001) \times .0001$  is how Stata calculates the initial small change. In this case, .00020873

# Average Marginal Effect (AME) the (almost) Stata way

```
* Two-sided derivative
preserve
qui logit lw cig5 faminc motheduc
* Define small change for cig5
qui sum cig5
scalar h = (abs(r(mean))+0.0001)*0.0001

* Duplicate variable
clonevar cig5_c = cig5

* Small negative change
replace cig5 = cig5_c - scalar(h)
predict double lw_0 if e(sample)
* Small positive change
replace cig5 = cig5_c + scalar(h)
predict double lw_1 if e(sample)

gen double dydx = (lw_1-lw_0)/(2*scalar(h))
sum dydx
restore
sum dydx
```

Variable	Obs	Mean	Std. Dev.	Min	Max
dydx	1,387	.0055771	.001245	.0040404	.0112832

- Remember, Stata has an iterative procedure to ensure numerical accuracy. This is one reason the margins command for some complex models can be slow. Note SEs are not too off, although they are not delta method SEs

# AME for indicator variables (incremental effects)

- Understanding marginal effects it's easier with dummy variables; that's why I have focused on continuous variables
- With dummy variables **we don't have to do a "small" change**. We change from 0 to 1, but the logic is the same, except that we don't need a two-sided change
- Notice the connection with treatment effects prediction of counterfactuals?

```
gen      smoked = 0
replace smoked = 1 if cigs > 0 & cigs ~=.
```

```
* Incremental change
```

```
preserve
```

```
qui logit lw smoked faminc motheduc
```

```
* Nobody smokes
```

```
replace smoked = 0
```

```
predict double lw_0 if e(sample)
```

```
* Everybody smokes
```

```
replace smoked = 1
```

```
predict double lw_1 if e(sample)
```

```
gen double dydx = (lw_1-lw_0)
```

```
sum dydx
```

```
restore
```

Variable	Obs	Mean	Std. Dev.	Min	Max
dydx	1,387	.118284	.0105063	.0935378	.1363816



# AME for indicator variables

- We can of course also use the margins command **with caution**

```
* Wrong
qui logit lw smoked faminc motheduc, nolog
margins, dydx(smoked)
-----
            |           Delta-method
            |           dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
smoked |   .0988076   .0230959    4.28   0.000   .0535405   .1440748
-----
```

```
* Right
qui logit lw i.smoked faminc motheduc, nolog
margins, dydx(smoked)
-----
            |           Delta-method
            |           dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
1.smoked |   .118284   .0322576    3.67   0.000   .0550602   .1815078
-----
```

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

- Even though same margins statement, different results. The first one is **not what we wanted**. We did not use the factor syntax in the first model so Stata didn't go from 0 to 1; instead it used a "small" change
- Smoking increases the probability of low birth weight by almost 12% points (yikes)

## Predicted margins (adjusted means)

- With indicator variables, we can also get what Stata calls predictive **margins** (not marginal effects). Marginal effects are their **difference**. You may know them as adjusted means
- With the post option Stata “posts” results as postestimation results so we can use them in tables or for calculations

```
qui logit lw i.smoked faminc motheduc, nolog
margins smoked, post
```

```
Predictive margins          Number of obs   =       1,387
Model VCE      : OIM
```

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

```
-----+-----
```

		Delta-method			[95% Conf. Interval]	
	Margin	Std. Err.	z	P> z		
smoked						
0	.1305183	.0099014	13.18	0.000	.1111118	.1499248
1	.2488023	.0304311	8.18	0.000	.1891584	.3084461

```
-----+-----
```

```
*ereturn list
```

```
matrix list e(b)
```

```
e(b)[1,2]
```

```
0.      1.
smoked  smoked
y1 .13051827 .24880226
```

```
* marginal effects
```

```
di e(b)[1,2] - e(b)[1,1]
```

```
.11828399
```

## Marginal effects, relative risks, odds ratios - **adjusted**

- Since we have the adjusted probabilities, we can use them to express results as marginal effects (difference/contrasts) ( $P_t - P_c$ ), relative risks ( $\frac{P_t}{P_c}$ ), or odds ratios ( $\frac{\frac{P_t}{1-P_t}}{\frac{P_c}{1-P_c}}$ )

```
. * marginal effects
. di e(b)[1,2] - e(b)[1,1]
.11828399

. * relative risks
. di e(b)[1,2] / e(b)[1,1]
1.9062639

. * odds ratios
. di (e(b)[1,2]/(1-e(b)[1,2])) / (e(b)[1,1]/(1-e(b)[1,1]))
2.2064252
```

- See how relative risks and odds ratios are not the same?
- Please make sure you understand the calculations above. For the nth time: **odds ratios are NOT relative probabilities**. Repeat after me... (you can do the same with probit, of course)

# Marginal effects, relative risks, odds ratios - **unadjusted**

- This would be the “unadjusted” version

```
logit lw i.smoked, nolog or
<... output omitted ...>
```

```
-----+-----
      lw | Odds Ratio   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
    1.smoked |      2.36004   .4228805    4.79   0.000    1.661109    3.353056
      _cons |      .1484375   .0129025   -21.95  0.000    .1251858    .1760079
-----+-----
```

Note: \_cons estimates baseline odds.

```
margins smoked, post
Adjusted predictions      Number of obs      =      1,388
Model VCE      : OIM
Expression      : Pr(lw), predict()
```

```
-----+-----
      |           Delta-method
      |   Margin   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
smoked |
    0 |   .1292517   .0097827   13.21   0.000    .1100779    .1484255
    1 |   .259434   .0301042    8.62   0.000    .2004308    .3184371
-----+-----
```

```
* marginal effects
di e(b)[1,2] - e(b)[1,1]
.13018226
* unadjusted relative risks
di e(b)[1,2] / e(b)[1,1]
2.0071996
* unadjusted odds ratios
di (e(b)[1,2]/(1-e(b)[1,2])) / (e(b)[1,1]/(1-e(b)[1,1]))
2.3600402
```

- Make sure you digest this. It will be in the midterm. Note that Stata's title is the default, but not correct in this case (Adjusted predictions)

# Odds ratios are not relative risks

- Again, odds ratios are not relative risks
- When are they going to be the same?

$$OR = \frac{\frac{P_t}{1-P_t}}{\frac{P_c}{1-P_c}} = \frac{P_t}{P_c} \times \frac{1-P_c}{1-P_t}$$

- One case is if  $P_t \approx P_c$ , although that's not an interesting case
- The more interesting case is  $\frac{1-P_c}{1-P_t} \approx 1$ , which would happen if both  $P_c$  and  $P_t$  are low probability events
- **Probit model:** One advantage of probit models is that unlike logit models, the only meaningful way to understand model results is to use marginal effects, so that would prevent you from presenting odds ratios

## Marginal Effect at the Mean (MEM) and Marginal effects at representative values (MER)

- We have left the values of the covariates **as they were observed** rather than holding them fixed at a certain value
- We can also calculate marginal effects at different value of the covariates, for example at the mean of each covariate or at some other value (MEM and MER, in Stata lingo)
- There is some discussion about which way is better. For example, does it make sense to hold male at 0.6 male? In a sense, yes. We are giving makes the value of the proportion in the sample, 0.6. In another sense, it seems odd
- **Don't waste too much time thinking about this.** When we calculate **marginal effects** (not margins), it doesn't really matter at which value we hold the other covariates constant because we are taking differences in effects
- In general, the difference will be so small that it is better to spend mental resources somewhere else

## Marginal effects at means (MEM)

- Same algorithm except that now we set value for covariates before making predictions to hold them constant – of course, adjusted means will be different

```
qui sum faminc
scalar fincbar = r(mean)
qui sum motheduc
scalar medbar = r(mean)

preserve
  qui sum cigs
  scalar h = (abs(r(mean))+0.0001)*0.0001
  qui logit lw cigs faminc motheduc, nolog
  clonevar cigs_c = cigs

  * Replace with means
  replace faminc = scalar(fincbar)
  replace motheduc = scalar(medbar)

  * Small negative change
  replace cigs = cigs_c - scalar(h)
  predict double lw_0 if e(sample)

  * Small positive change change
  replace cigs = cigs_c + scalar(h)
  predict double lw_1 if e(sample)

  gen double dydx = (lw_1-lw_0)/(2*scalar(h))
  sum dydx
restore
```

- See these slides for more

[https://clas.ucdenver.edu/marcelo-perrailon/sites/default/files/attached-files/week\\_13\\_margins.pdf](https://clas.ucdenver.edu/marcelo-perrailon/sites/default/files/attached-files/week_13_margins.pdf)





# Marginsplot for adjusted means

- Useful to visualize models holding covariates at some value. Handy in DiD models for adjusted trends. Note this is a linear/OLS model

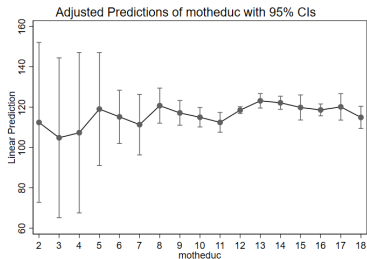
```
margins motheduc, at((mean) faminc parity)
Adjusted predictions      Number of obs   =       1,387
Model VCE      : OLS
Expression      : Linear prediction, predict()
at              : faminc      =    29.04218 (mean)
                 parity      =    1.633021 (mean)
```

```
-----+-----
      |           Delta-method
      |   Margin   Std. Err.   t    P>|t|   [95% Conf. Interval]
-----+-----
      |
motheduc |
      2 |   112.4071   20.19802   5.57  0.000   72.78466   152.0296
      3 |   104.8245   20.18644   5.19  0.000   65.22478   144.4242
```

<... output omitted ...>

```
margins motheduc, at((mean) faminc parity)
```

```
marginsplot
```

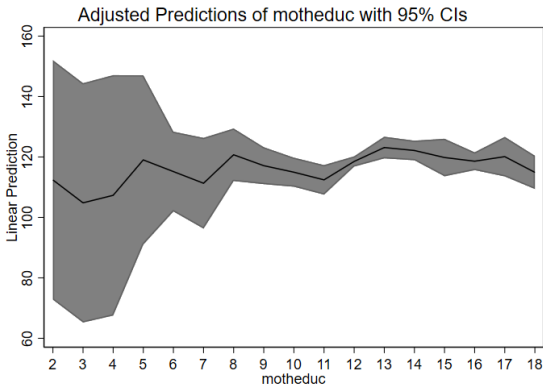


# Marginsplot for adjusted means - lots of options

- Check options for marginsplot. With post options you can get the output to make your own plots too

```
marginsplot, recast(line) recastci(rarea)  
graph export marplot2.png, replace
```

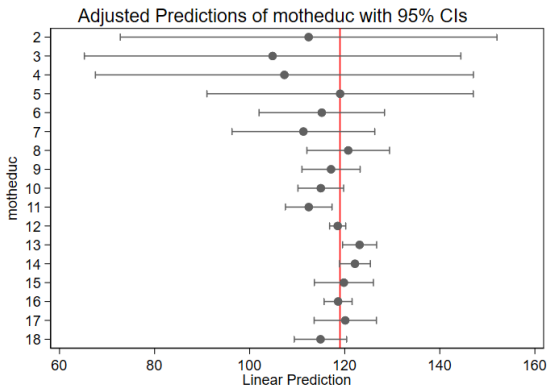
```
* saved margins in e(b)  
*matrix list e(b)
```



# Marginsplot for adjusted means - lots of options

- marginsplot is flexible

```
sum bwght
qui reg bwght i.motheduc faminc parity
margins motheduc, at((mean) faminc parity)
marginsplot, horizontal recast(scatter) xline(119, lcolor(red)) ///
    xscale(range()) yscale(reverse)
graph export mhor.png, replace
```



# Interactions in logistic models

- We estimated the model

$$\log\left(\frac{lw_i}{1-lw_i}\right) = \beta_0 + \beta_1cigs_i + \beta_2faminc_i + \beta_3motheduc_i$$

- We didn't use interactions between cigarettes and income so we have assumed the same effect regardless of income in the **log-odds scale**
- If we add interactions, the margins command is a life saver because effects are hard to interpret otherwise
- In particular, the **magnitude of effects** is hard to make sense when a logistic model has interactions and we use odds ratios for interpretations
- But interactions in logistic (and probit) models have other issues: the **magnitude, sign, and statistical significance of interaction in the log-odds scale may not be relevant to understand the interaction in the probability scale**
- Exploring interactions with marginal effects is better (the old mfx would provide SEs for an interacted variable, but margins with the factor syntax recognizes interaction terms)

# Interactions

- As usual, interactions are easier to understand with two indicator variables. Let's say more than high school (hsp) and male as in previous examples:  
$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 hsp + \beta_2 male + \beta_3 hsp * male$$
- In the log-odds scale, we interpret them exactly in the same way as in the linear model and we can use derivatives and the other tricks we learned (What's  $\beta_1$ ?  $\beta_2$ ?)
- For males:  $\log\left(\frac{p_m}{1-p_m}\right) = \beta_0 + \beta_1 hsp + \beta_2 + \beta_3 hsp$
- For females:  $\log\left(\frac{p_f}{1-p_f}\right) = \beta_0 + \beta_1 hsp$
- Difference males - females:  $\log\left(\frac{p_m}{1-p_m}\right) - \log\left(\frac{p_f}{1-p_f}\right) = \beta_2 + \beta_3 hsp$
- So the effect depends on the level of education
- Note that we can write it in terms of odds:  $\frac{\frac{p_m}{1-p_m}}{\frac{p_f}{1-p_f}} = e^{\beta_2 + \beta_3 hsp}$

# Interactions

- Difference males - females:  $\log\left(\frac{p_m}{1-p_m}\right) - \log\left(\frac{p_f}{1-p_f}\right) = \beta_2 + \beta_3 hsp$
- Difference male - female for educated:  $\log\left(\frac{p_{me}}{1-p_{me}}\right) - \log\left(\frac{p_{fe}}{1-p_{fe}}\right) = \beta_2 + \beta_3$
- Difference male - female for uneducated:  $\log\left(\frac{p_{mu}}{1-p_{mu}}\right) - \log\left(\frac{p_{fu}}{1-p_{fu}}\right) = \beta_2$
- Difference in difference:  
 $\log\left(\frac{p_{me}}{1-p_{me}}\right) - \log\left(\frac{p_{fe}}{1-p_{fe}}\right) - [\log\left(\frac{p_{mu}}{1-p_{mu}}\right) - \log\left(\frac{p_{fu}}{1-p_{fu}}\right)] = \beta_3$
- So same as with linear model. In the log-odds scale, it is a **difference-in-difference**

## Interactions in the odds scale

$$\blacksquare \log\left(\frac{p_{me}}{1-p_{me}}\right) - \log\left(\frac{p_{fe}}{1-p_{fe}}\right) - \left[\log\left(\frac{p_{mu}}{1-p_{mu}}\right) - \log\left(\frac{p_{fu}}{1-p_{fu}}\right)\right] = \beta_3$$

- We can apply the rules of logs and take  $e^{(\ )}$  on both sides:

$$\frac{\frac{p_{me}}{1-p_{me}}}{\frac{p_{fe}}{1-p_{fe}}} / \frac{\frac{p_{mu}}{1-p_{mu}}}{\frac{p_{fu}}{1-p_{fu}}} = e^{\beta_3}$$

- In the odds scale, an interaction is not a difference-in-difference but rather a **ratio of odds ratios**
- The interpretation is... well, good luck with that...
- The interpretation is not that hard, actually. If greater than one, the odds of outcome for educated are higher than for the uneducated (see the numerator—it's the odds ratio for males females for those "educated" or hsp=1. The denominator is same but for those with hsp=0)
- **Interpreting the magnitude in the probability scale is a lost cause.** A lot easier with margins or stratifying the models

# Interactions and marginal effects

- We are going to estimate the model:

$$\text{logit}(lw_i) = \beta_0 + \beta_1 \text{cigs}_i + \beta_2 \text{inc}_i + \beta_3 \text{cigs} * \text{inc}_i$$

- inc is a dummy and equal 1 if income is higher than 40k

```
gen      inc = 0
replace inc = 1 if faminc > 40 & faminc~= .
logit lw c.cigs##i.inc, nolog
Logistic regression                Number of obs   =    1,388
                                   LR chi2(3)         =    22.90
                                   Prob > chi2        =    0.0000
                                   Pseudo R2          =    0.0196

Log likelihood = -573.18336
```

	lw	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	cigs	.0494583	.0104689	4.72	0.000	.0289397	.0699769
	1.inc	-.1003108	.1923845	-0.52	0.602	-.4773774	.2767558
inc#c.cigs							
	1	-.0534926	.0583898	-0.92	0.360	-.1679344	.0609492
	_cons	-1.843	.0954774	-19.30	0.000	-2.030132	-1.655868

- Interaction term not significant, so the model with the interaction doesn't improve fit



# Interactions and marginal effects

- We can also use the margins command to check effects in the probability scale

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

```
Average marginal effects      Number of obs   =   1,388
Model VCE      : OIM

Expression      : Pr(lw), predict()
dy/dx w.r.t.   : cigs 1.inc
```

```
-----
      |           Delta-method
      |           dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
cigs |   .0046941   .0018482     2.54  0.011     .0010717     .0083165
1.inc |  -.0288422   .0225222    -1.28  0.200    -.0729849    .0153006
-----
```

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

- Wait, two effects? The model has three coefficients. **Where is the interaction?**

## Interactions and marginal effects

- This may seem confusing but it's not when you remember how Stata calculates marginal effects. For `cigs`, a continuous variable, it's using the two-sided derivative increasing `cigs` by a little bit and calculating predictions. It's **increasing `cigs` in both the main effect and the interaction**
- Then it takes an average so the marginal effect of `cigs` is the numerical derivative for both `inc=1` and `inc=0` combined
- For the marginal effect of `inc`, it's doing the same going from 0 to 1, averaging over the values of `cigs`, but changing `inc` from 0 to 1 in both the main effect and interaction
- With factor syntax, Stata knows that one variable is a multiplication of the other
- To get the effect of one variable at some values of the other, we have to be more specific with the syntax

# Interactions and marginal effects

- Get the effects at different levels. Below, we compute:  $\frac{\partial P}{\partial \text{cigs}} \Big|_{\text{inc}=0}$  and

$$\frac{\partial P}{\partial \text{cigs}} \Big|_{\text{inc}=1}$$

```
. margins, dydx(cigs) at(inc=(0 1)) vsquish
```

```
Average marginal effects      Number of obs   =       1,388
Model VCE      : OIM
```

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

```
dy/dx w.r.t.    : cigs
```

```
1._at          : inc          =          0
```

```
2._at          : inc          =          1
```

```
-----+-----
```

		Delta-method				
		dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]
cigs						
	_at					
	1	.0062867	.0012881	4.88	0.000	.0037621 .0088113
	2	-.0004394	.0062301	-0.07	0.944	-.0126501 .0117713

```
-----+-----
```

- **Read the output carefully.** at = 1 is for inc = 0. They go in different directions, with the one with higher incomes having a large CIs (why?)

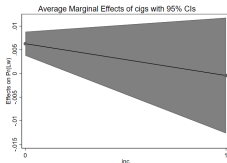
## Is the interaction statistically significant in the probability scale?

- Because the significance of the interaction in the log-odds scale could be different than in the probability scale, we want to check the delta method SEs from the margins command
- For modeling though, the significance in the log-odds scale is telling us about fit. It's the same as using likelihood ratio tests (next class)
- But we should check if the effects are different in the probability scale (magnitude, significant, direction)
- Above, the confidence intervals overlap, so we wouldn't reject the null that the difference is zero using margins
- Marginsplot helps here, but we can also use the `contrast()` option in margins. It performs tests using the delta-method SEs

# Is the interaction statistically significant in the probability scale?

- Check the overlap of delta SEs. Check the magnitudes. Or use the contrast option (handy but not well-known). See Karaka-Mandic, Norton, Dowd (2012) for details. The do file for today's class has code to replicate their example (plus a test using contrast option)

```
margins, dydx(cigs) at(inc=(0 1)) vsquish
marginsplot
graph export inter.png, replace
margins, dydx(cigs) at(inc=(0 1)) contrast(at)
Contrasts of average marginal effects      Number of obs      =      1,388
Model VCE      : OIM
Expression     : Pr(1w), predict()
dy/dx w.r.t.   : cigs
1._at         : inc           =           0
2._at         : inc           =           1
-----+-----
          |      df      chi2    P>chi2
-----+-----
cigs     |
  _at    |      1      1.12    0.2904
-----+-----
```



## So what to do if interaction is not significant with marginal effects?

- There are some deep issues related to statistical model and testing. We should use Wald tests and LR tests to build models, including interactions, but we should present results in the scale of interest and discuss implications of the model. Following Green (2010) (emphasis mine):
  - 1 Build the model based on appropriate statistical procedures and principles. Statistical testing about the model specification is done at this step. Hypothesis tests are about model coefficients and about the structural aspects of the model specification. Partial effects [i.e. marginal effects] are neither coefficients nor elements of the specification of the model. They are **implications** of the specified and estimated model.
  - 2 Once the model is in place, inform the reader with analysis of model implications such as values, predictions, partial effects and interactions. We find that graphical presentations are a very informative adjunct to numerical statistical results for this purpose. **Hypothesis testing need not be done at this point** [that is, using delta errors]. Even where the partial effects are the ultimate target of estimation [as in difference-in-difference models], it seems it would be rare for a model builder to build a structural model by hypothesizing (statistically) about partial effects and/or predictions that would be made by that model.

## Other models

- The logic of marginal and incremental effects extends to any other parametric model
- Calculate the numerical derivative for a small change  $h$  or an incremental 0 to 1 change
- The algorithm is identical, marginal effects by default will show results in the default scale of the predict command
- The scale can be changed with the predict() option in the margins command
- In GLM models with a log link, the marginal effect would be
$$\frac{\partial E[Y|\mathbf{X}]}{\partial x} = \beta_x e^{\mathbf{x}'\beta}$$
- We could have used GLM models for all the logistic examples above

# GLM logit

## ■ Replicate example with GLM

```
qui glm lw c.cigs##i.inc, family(binomial) link(logit)
margins, dydx(*)
```

```
<... output omitted ...>
```

		Delta-method				[95% Conf. Interval]	
		dy/dx	Std. Err.	z	P> z		
cigs		.0046941	.0018482	2.54	0.011	.0010717	.0083165
1.inc		-.0288422	.0225222	-1.28	0.200	-.0729849	.0153006

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

```
margins, dydx(cigs) at(inc=(0 1)) vsquish
```

```
<... output omitted ...>
```

		Delta-method				[95% Conf. Interval]	
		dy/dx	Std. Err.	z	P> z		
cigs							
_at							
1		.0062867	.0012881	4.88	0.000	.0037621	.0088113
2		-.0004394	.0062301	-0.07	0.944	-.0126501	.0117713



## More custom: Two-part models

- Two-part models are often used to estimate cost data with a large proportion of costs that are zero
- The idea is to estimate the model in two parts:
  - 1) Estimate the probability that the cost is greater than zero conditional on covariates:  $P(y_i > 0|X_i)$  (using logit, probit, complementary log-log models)
  - 2) For those observations with non-zero costs, estimate the expected costs conditional on covariates:  $E(y_i|y_i > 0, X_i)$  (using Poisson, linear models, Gamma,  $\log(y)$ , Box-Cox, etc)
- Predictions are obtained combining both parts (multiplication):  
 $P(y_i > 0|X_i) \times E(y_i|y_i > 0, X_i)$

# Two-part models marginal effects

- If you know how to get predictions, you now how to calculate marginal effects
- First, using the user-written command `twopm` and the `margins` command

```
* Get data from example in twopm command
webuse womenwk
replace wage = 0 if wage==.
```

```
* First part logit, second part GLM with Gamma family
qui twopm wage i.married children, firstpart(logit) secondpart(glm, family(gamma) link(log))
margins, dydx(married)
```

```
Average marginal effects          Number of obs   =       2,000
Expression   : twopm combined expected values, predict()
dy/dx w.r.t. : i.married
```

```
-----+-----
          |           Delta-method
          |           dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
    1.married |    6.615493   .5323105   12.43   0.000    5.572183    7.658802
-----+-----
```

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

# Two-part models marginal effects “by hand”

```
gen      nozeroc = 0
replace nozeroc = 1 if wage > 0
preserve
  clonevar marr = married
  * not married
  qui logit nozeroc i.married children
  replace married = 0
  predict double fp0
  replace married = marr
  qui glm wage i.married children if wage > 0, f(gamma) l(log)
  replace married = 0
  predict double c0
  gen chat0 = fp0*c0

  * married
  replace married = marr
  qui logit nozeroc i.married children
  replace married = 1
  predict double fp1
  replace married = marr
  qui glm wage i.married children if wage > 0, f(gamma) l(log)
  replace married = 1
  predict double c1
  gen chat1 = fp1*c1

  gen dif = chat1 - chat0
  sum dif
restore
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
dif	2,000	6.615493	1.660084	2.485561	8.001396

# Subpopulations

- The subpop option is the equivalent of “if” in margins; helpful with survey data but also helpful when we want the margins for some part of the data
- Remember ATET when we discussed treatment effects? We can get ATET after a fully-interacted regression model
- Make sure you understand this. Go back to the class on regression adjustment and propensity scores

```
qui reg bwght i.smoked##(c.fatheduc c.faminc), robust
margins r.smoked, subpop(smoked) vce(unconditional)
<...>
```

```
-----
            |           Unconditional
            | Contrast  Std. Err.   [95% Conf. Interval]
-----+-----
smoked |
(1 vs 0) | -8.648601  1.623139   -11.83314   -5.464057
-----
```

```
teffects ra (bwght fatheduc faminc)(smoked), atet
<...>
```

```
-----
            |           Robust
            | Coef.  Std. Err.   z  P>|z|   [95% Conf. Interval]
-----+-----
ATET
smoked |
(1 vs 0) | -8.648601  1.619049   -5.34  0.000   -11.82188   -5.475324
-----+-----
P0mean
smoked |
0 | 120.1331  .7079497  169.69  0.000   118.7455   121.5206
-----
```

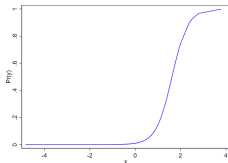
# LPM, logit, and probit redux

- A simulation to understand some features and issues with non-linearity and linear probability models versus logit and probit
- We will simulate 0/1 data in which a linear probability would be a bit problematic, with the outcome  $y$  having low probability (0.56)
- The graph are predictions using the logit model

```
clear
set seed 1234567
set obs 1000
gen x = rnormal(-2/3, 4/3)
gen y = rbinomial(1,invlogit(-4.5 + 3*x))
sum y
```

Variable	Obs	Mean	Std. Dev.	Min	Max
y	999	.0560561	.2301452	0	1

```
qui logit y x
predict y_logit
graph export logitg.png, replace
```



# LPM, logit, and probit redux

- Compare models. Similar average marginal effects but not so close

```
* LPM
qui reg y x
est sto lpm
predict y_lpm
* Logit
qui logit y x
predict y_logit
margins, dydx(x) post
est sto logit_dydx
* Probit
qui logit y x
predict y_probit
margins, dydx(x) post
est sto probit_dydx

. est table lpm logit_dydx probit_dydx, star
```

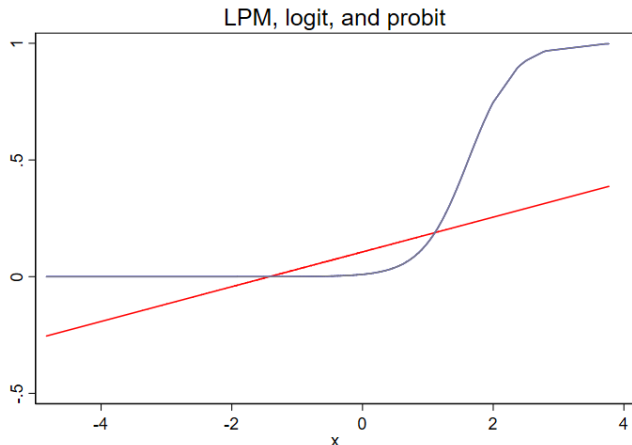
Variable	lpm	logit_dydx	probit_dydx
x	.07450962***	.08365352***	.08365352***
_cons	.10604429***		

legend: \* p<0.05; \*\* p<0.01; \*\*\* p<0.001

# LPM, logit, and probit redux

- Graphs shows the reason. Linear approximation not great. Predictions negative
- Logit and probit predictions identical

```
line y_lpm x, sort color(red) || line y_logit x, color(blue) sort ///  
    || line y_probit x, sort legend(off) title("LPM, logit, and probit")  
graph export lpm_plus.png, replace
```



# LPM, logit, and probit redux

- Marginal effects at different values of  $x$  tell us the story
- In the linear model the slope is the same; it's, well, linear (constant)
- In the logit model, very different at most values, except close to  $x = 0.45$  (see the graph closely). Check with margins

```
margins, dydx(x) at(x=(-4, 0.45, 4)) vsquish  
<...>
```

		Delta-method				[95% Conf. Interval]	
		dy/dx	Std. Err.	t	P> t		
x							
	_at						
	1	.0745096	.0050552	14.74	0.000	.0645896	.0844296
	2	.0745096	.0050552	14.74	0.000	.0645896	.0844296
	3	.0745096	.0050552	14.74	0.000	.0645896	.0844296

```
qui logit y x  
margins, dydx(x) at(x=(-4, 0.35, 4)) vsquish  
<...>
```

		Delta-method				[95% Conf. Interval]	
		dy/dx	Std. Err.	z	P> z		
x							
	_at						
	1	3.15e-07	5.02e-07	0.63	0.531	-6.70e-07	1.30e-06
	2	.0724872	.0165856	4.37	0.000	.0399801	.1049943
	3	.003214	.0026415	1.22	0.224	-.0019633	.0083914



## Topics not covered but helpful to know

- The margins command can do much more
  - 1 You can get elasticities and semielasticities: `eyex()`, `dyex()`, `eydx()`
  - 2 You can use it to make predictions: `margins`, `at(...)`
  - 3 You can use an external dataset to do adjustment through standardization (as in `epi`)
  - 4 You can use the contrast postestimation to test differences after margins to do tests (we covered the `contrast` option)