Estimating and Interpreting Effects for Nonlinear and Nonparametric Models

Enrique Pinzón

September 18, 2018
Objective

- Build a unified framework to ask questions about model estimates
- Learn to apply this unified framework using Stata
A Brief Introduction to Stata and How I Work

- A look at the Stata Interface
- From dialog boxes to do-files
- Loading your data
  - Excel
  - Delimited (comma, tab, or other)
  - ODBC (open data base connectivity)
  - Fred, SAS, Haver
- “Big data”
  - 120,000 variables 20 billion observations (MP)
  - 32,767 variables 2.14 billion observations (SE)
- Stata resources
  https://www.stata.com/links/resources-for-learning-stata/
A Brief Introduction to Stata and How I Work

- A look at the Stata Interface
- From dialog boxes to do-files
- Loading your data
  - Excel
  - Delimited (comma, tab, or other)
  - ODBC (open data base connectivity)
  - Fred, SAS, Haver
- “Big data”
  - 120,000 variables 20 billion observations (MP)
  - 32,767 variables 2.14 billion observations (SE)
- Stata resources
  https://www.stata.com/links/resources-for-learning-stata/
Factor variables

- Distinguish between discrete and continuous variables
- Way to create “dummy-variables”, interactions, and powers
- Works with most Stata commands
Using factor variables

```
.import excel apsa, firstrow
.tabulate d1

<table>
<thead>
<tr>
<th>d1</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2,000</td>
<td>20.00</td>
<td>20.00</td>
</tr>
<tr>
<td>1</td>
<td>2,000</td>
<td>20.00</td>
<td>40.00</td>
</tr>
<tr>
<td>2</td>
<td>2,044</td>
<td>20.44</td>
<td>60.44</td>
</tr>
<tr>
<td>3</td>
<td>2,037</td>
<td>20.37</td>
<td>80.81</td>
</tr>
<tr>
<td>4</td>
<td>1,919</td>
<td>19.19</td>
<td>100.00</td>
</tr>
</tbody>
</table>

.Total | 10,000 | 100.00 |
```

```
.summarize 1.d1

Variable | Obs | Mean  | Std. Dev. | Min | Max |
---------|-----|-------|-----------|-----|-----|
1.d1     | 10,000 | .2     | .40002    | 0   | 1   |
```

```
.summarize i.d1

Variable | Obs | Mean  | Std. Dev. | Min | Max |
---------|-----|-------|-----------|-----|-----|
d1       | 10,000 | .2     | .40002    | 0   | 1   |
    1   | 10,000 | .2044  | .4032827  | 0   | 1   |
    2   | 10,000 | .2037  | .4027686  | 0   | 1   |
    3   | 10,000 | .1919  | .3938145  | 0   | 1   |
    4   | 10,000 | .2037  | .4032827  | 0   | 1   |
```
Using factor variables

```
. import excel apsa, firstrow

. tabulate d1

<table>
<thead>
<tr>
<th>d1</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2,000</td>
<td>20.00</td>
<td>20.00</td>
</tr>
<tr>
<td>1</td>
<td>2,000</td>
<td>20.00</td>
<td>40.00</td>
</tr>
<tr>
<td>2</td>
<td>2,044</td>
<td>20.44</td>
<td>60.44</td>
</tr>
<tr>
<td>3</td>
<td>2,037</td>
<td>20.37</td>
<td>80.81</td>
</tr>
<tr>
<td>4</td>
<td>1,919</td>
<td>19.19</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Total | 10,000 | 100.00 |

. summarize 1.d1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.d1</td>
<td>10,000</td>
<td>.2</td>
<td>.40002</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

. summarize i.d1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>d1</td>
<td>10,000</td>
<td>.2</td>
<td>.40002</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>10,000</td>
<td>.2044</td>
<td>.4032827</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>10,000</td>
<td>.2037</td>
<td>.4027686</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>10,000</td>
<td>.1919</td>
<td>.3938145</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
```
Using factor variables

. import excel apsa, firstrow

. tabulate d1

<table>
<thead>
<tr>
<th>d1</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2,000</td>
<td>20.00</td>
<td>20.00</td>
</tr>
<tr>
<td>1</td>
<td>2,000</td>
<td>20.00</td>
<td>40.00</td>
</tr>
<tr>
<td>2</td>
<td>2,044</td>
<td>20.44</td>
<td>60.44</td>
</tr>
<tr>
<td>3</td>
<td>2,037</td>
<td>20.37</td>
<td>80.81</td>
</tr>
<tr>
<td>4</td>
<td>1,919</td>
<td>19.19</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>10,000</td>
<td>100.00</td>
</tr>
</tbody>
</table>

. summarize 1.d1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.d1</td>
<td>10,000</td>
<td>.2</td>
<td>.40002</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

. summarize i.d1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>d1</td>
<td>10,000</td>
<td>.2</td>
<td>.40002</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>10,000</td>
<td>.2044</td>
<td>.4032827</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>10,000</td>
<td>.2037</td>
<td>.4027686</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>10,000</td>
<td>.1919</td>
<td>.3938145</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>10,000</td>
<td>.1919</td>
<td>.3938145</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Using factor variables

```
.import excel apsa, firstrow
.tabulate d1

<table>
<thead>
<tr>
<th>d1</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2,000</td>
<td>20.00</td>
<td>20.00</td>
</tr>
<tr>
<td>1</td>
<td>2,000</td>
<td>20.00</td>
<td>40.00</td>
</tr>
<tr>
<td>2</td>
<td>2,044</td>
<td>20.44</td>
<td>60.44</td>
</tr>
<tr>
<td>3</td>
<td>2,037</td>
<td>20.37</td>
<td>80.81</td>
</tr>
<tr>
<td>4</td>
<td>1,919</td>
<td>19.19</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>10,000</td>
<td>100.00</td>
</tr>
</tbody>
</table>

.summarize 1.d1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.d1</td>
<td>10,000</td>
<td>.2</td>
<td>.40002</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

.summarize i.d1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>d1</td>
<td>10,000</td>
<td>.2</td>
<td>.40002</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>10,000</td>
<td>.2044</td>
<td>.4032827</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>10,000</td>
<td>.2037</td>
<td>.4027686</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>10,000</td>
<td>.1919</td>
<td>.3938145</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
```
Using factor variables

```
. summarize ibn.d1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>d1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>10,000</td>
<td>.2</td>
<td>.40002</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>10,000</td>
<td>.2</td>
<td>.40002</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>10,000</td>
<td>.2044</td>
<td>.4032827</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>10,000</td>
<td>.2037</td>
<td>.4027686</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>10,000</td>
<td>.1919</td>
<td>.3938145</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
```

```
. summarize ib2.d1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>d1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>10,000</td>
<td>.2</td>
<td>.40002</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>10,000</td>
<td>.2</td>
<td>.40002</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>10,000</td>
<td>.2044</td>
<td>.4032827</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>10,000</td>
<td>.2037</td>
<td>.4027686</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>10,000</td>
<td>.1919</td>
<td>.3938145</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
```
Using factor variables

. summarize ibn.d1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>d1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>10,000</td>
<td>.2</td>
<td>.40002</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>10,000</td>
<td>.2</td>
<td>.40002</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>10,000</td>
<td>.2044</td>
<td>.4032827</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>10,000</td>
<td>.2037</td>
<td>.4027686</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>10,000</td>
<td>.1919</td>
<td>.3938145</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

. summarize ib2.d1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>d1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>10,000</td>
<td>.2</td>
<td>.40002</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>10,000</td>
<td>.2</td>
<td>.40002</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>10,000</td>
<td>.2037</td>
<td>.4027686</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>10,000</td>
<td>.1919</td>
<td>.3938145</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>10,000</td>
<td>.1919</td>
<td>.3938145</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Using factor variables

```
. summarize d1##d2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>d1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>10,000</td>
<td>.2</td>
<td>.40002</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>10,000</td>
<td>.2044</td>
<td>.4032827</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>10,000</td>
<td>.2037</td>
<td>.4027686</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>10,000</td>
<td>.1919</td>
<td>.3938145</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1.d2</td>
<td>10,000</td>
<td>.4986</td>
<td>.500023</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

| d1#d2    |      |       |           |     |     |
| 1 1      | 10,000 | .1009| .3012113  | 0   | 1   |
| 2 1      | 10,000 | .1007| .3009461  | 0   | 1   |
| 3 1      | 10,000 | .1035| .304626   | 0   | 1   |
| 4 1      | 10,000 | .0922| .2893225  | 0   | 1   |
```
Using factor variables

```
. summarize c.x1##c.x1 c.x1#c.x2 c.x1#i.d1, separator(4)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>10,000</td>
<td>.0110258</td>
<td>.9938621</td>
<td>-4.095795</td>
<td>3.714316</td>
</tr>
<tr>
<td>c.x1#c.x1</td>
<td>10,000</td>
<td>.9877847</td>
<td>1.416602</td>
<td>4.18e-09</td>
<td>16.77553</td>
</tr>
<tr>
<td>c.x1#c.x2</td>
<td>10,000</td>
<td>.000208</td>
<td>1.325283</td>
<td>-7.469295</td>
<td>6.45778</td>
</tr>
<tr>
<td>d1#c.x1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>10,000</td>
<td>.0044334</td>
<td>.4516058</td>
<td>-3.021819</td>
<td>3.286315</td>
</tr>
<tr>
<td>2</td>
<td>10,000</td>
<td>.0008424</td>
<td>.4432188</td>
<td>-4.095795</td>
<td>3.178586</td>
</tr>
<tr>
<td>3</td>
<td>10,000</td>
<td>.0025783</td>
<td>.4533505</td>
<td>-3.374062</td>
<td>3.428311</td>
</tr>
<tr>
<td>4</td>
<td>10,000</td>
<td>-.0014739</td>
<td>.4379122</td>
<td>-3.161604</td>
<td>3.714316</td>
</tr>
</tbody>
</table>
```
Models and Quantities of Interest

We usually model an outcome of interest, $Y$, conditional on covariates of interest $X$:

- $E(Y|X) = X \beta$ (regression)
- $E(Y|X) = \exp(X \beta)$ (poisson)
- $E(Y|X) = P(Y|X) = \Phi(X \beta)$ (probit)
- $E(Y|X) = P(Y|X) = [\exp(X \beta)] [1 + \exp(X \beta)]^{-1}$ (logit)
- $E(Y|X) = g(X)$ (nonparametric regression)
Models and Quantities of Interest

We usually model an outcome of interest, \( Y \), conditional on covariates of interest \( X \):

- \( \mathbb{E}(Y|X) = X\beta \) (regression)
- \( \mathbb{E}(Y|X) = \exp(X\beta) \) (poisson)
- \( \mathbb{E}(Y|X) = P(Y|X) = \Phi(X\beta) \) (probit)
- \( \mathbb{E}(Y|X) = P(Y|X) = [\exp(X\beta)][1 + \exp(X\beta)]^{-1} \) (logit)
- \( \mathbb{E}(Y|X) = g(X) \) (nonparametric regression)
Models and Quantities of Interest

We usually model an outcome of interest, $Y$, conditional on covariates of interest $X$:

- $E(Y|X) = X\beta$ (regression)
- $E(Y|X) = \exp(X\beta)$ (poisson)
- $E(Y|X) = P(Y|X) = \Phi(X\beta)$ (probit)
- $E(Y|X) = P(Y|X) = \left[\exp(X\beta)\right] \left[1 + \exp(X\beta)\right]^{-1}$ (logit)
- $E(Y|X) = g(X)$ (nonparametric regression)
We usually model an outcome of interest, $Y$, conditional on covariates of interest $X$:

- $E(Y|X) = X\beta$ (regression)
- $E(Y|X) = \exp(X\beta)$ (poisson)
- $E(Y|X) = P(Y|X) = \Phi(X\beta)$ (probit)
- $E(Y|X) = P(Y|X) = \left[\exp(X\beta)\right]\left[1 + \exp(X\beta)\right]^{-1}$ (logit)
- $E(Y|X) = g(X)$ (nonparametric regression)
Questions

- Population averaged
  - Does a medicaid expansion improve health outcomes?
  - What is the effect of a minimum wage increase on employment?
  - What is the effect on urban violence indicators, during the weekends of moving back the city curfew?

- At a point
  - What is the effect of loosing weight for a 36 year, overweight hispanic man?
  - What is the effect on urban violence indicators, during the weekends of moving back the city curfew, for a large city, in the southwest of the United States?
Questions

Population averaged

- Does a medicaid expansion improve health outcomes?
- What is the effect of a minimum wage increase on employment?
- What is the effect on urban violence indicators, during the weekends of moving back the city curfew?

At a point

- What is the effect of loosing weight for a 36 year, overweight hispanic man?
- What is the effect on urban violence indicators, during the weekends of moving back the city curfew, for a large city, in the southwest of the United States?
Questions

- Population averaged
  - Does a medicaid expansion improve health outcomes?
  - What is the effect of a minimum wage increase on employment?
  - What is the effect on urban violence indicators, during the weekends of moving back the city curfew?

- At a point
  - What is the effect of loosing weight for a 36 year, overweight hispanic man?
  - What is the effect on urban violence indicators, during the weekends of moving back the city curfew, for a large city, in the southwest of the United States?
What are the answers?
A linear model

\[ y = \beta_0 + x_1 \beta_1 + x_2 \beta_2 + x_1^2 \beta_3 + x_2^2 \beta_4 + x_1 x_2 \beta_5 + d_1 \beta_6 + d_2 \beta_7 + d_1 d_2 \beta_8 + x_2 d_1 \beta_9 + \varepsilon \]

- $x_1$ and $x_2$ are continuous, $d_2$ is binary, and $d_1$ has 5 categories.
- There are interactions of continuous and categorical variables
- This is simulated data
A linear model

\[ y = \beta_0 + x_1 \beta_1 + x_2 \beta_2 + x_1^2 \beta_3 + x_2^2 \beta_4 + x_1 x_2 \beta_5 \\
+ d_1 \beta_6 + d_2 \beta_7 + d_1 d_2 \beta_8 + x_2 d_1 \beta_9 + \varepsilon \]

- \( x_1 \) and \( x_2 \) are continuous, \( d_2 \) is binary, and \( d_1 \) has 5 categories.
- There are interactions of continuous and categorical variables
- This is simulated data
Regression results

```
. regress yr c.x1##c.x2 c.x1#c.x1 c.x2#c.x2 i.d1##i.d2 c.x2#i.d1

Source | SS      | df   | MS
------|---------|------|------
Model  | 335278.744 | 18   | 18626.5969
Residual | 479031.227 | 9,981 | 47.9943119
Total  | 814309.971 | 9,999 | 81.439141

Number of obs = 10,000
F(18, 9981) = 388.10
Prob > F = 0.0000
R-squared = 0.4117
Adj R-squared = 0.4107
Root MSE = 6.9278

yr | Coef.   | Std. Err. | t     | P>|t|   | [95% Conf. Interval]
---|---------|-----------|-------|-------|-------------------------
x1 | -1.04884 | .1525255  | -6.88 | 0.000 | -1.347821 - .7498593    
x2 | .4749664 | .4968878  | 0.96  | 0.339 | -.4990339 1.448967     
c.x1#c.x2 | 1.06966 | .1143996  | 9.35  | 0.000 | .8454139 1.293907      
c.x1#c.x1 | -1.061312 | .048992 | -21.66 | 0.000 | -1.157346 - .9652779    
c.x2#c.x2 | 1.177785 | .1673487  | 7.04  | 0.000 | .849748 1.505822       
d1 | 1         | -1.504705 | .5254654 | -2.86 | 0.004 | -2.534723 - .4746865    
   | 2         | -3.727184 | .5272623 | -7.07 | 0.000 | -4.760725 -2.693644     
   | 3         | -6.522121 | .5229072 | -12.47 | 0.000 | -7.547125 -5.497118     
   | 4         | -8.80982 | .5319266 | -16.56 | 0.000 | -9.852503 -7.767136     
   | 1.d2      | 1.615761 | .3099418  | 5.21  | 0.000 | 1.008212 2.223309       
d1#d2 | 1 1  | -3.649372 | .4383277 | -8.33 | 0.000 | -4.508582 -2.790161     
   | 2 1  | -5.994454 | .435919 | -13.75 | 0.000 | -6.848943 -5.139965     
   | 3 1  | -8.457034 | .4364173 | -19.38 | 0.000 | -9.3125 -7.601568       
   | 4 1  | -11.04842 | .4430598 | -24.94 | 0.000 | -11.9169 -10.17993      
d1#c.x2 | 1  | 1.11805 | .3626989 | 3.08  | 0.002 | .4070865 1.829013       
   | 2  | 1.918298 | .3592232 | 5.34  | 0.000 | 1.214149 2.622448       
   | 3  | 3.484255 | .3594559 | 9.69  | 0.000 | 2.779649 4.188861       
   | 4  | 4.260699 | .362315 | 11.76 | 0.000 | 3.550488 4.970909       
```
Effects: $x_2$

Suppose we want to study the marginal effect of $x_2$

$$\frac{\partial E (y|x_1, x_2, d_1, d_2)}{\partial x_2}$$

This is given by

$$\frac{\partial E (y|x_1, x_2, d_1, d_2)}{\partial x_2} = \beta_2 + 2x_2\beta_4 + x_1\beta_5 + d_1\beta_9$$

- I can compute this effect for every individual in my sample and then average to get a population averaged effect.
- I could evaluate this conditional on values of the different covariates, or even values of importance for $x_2$. 


Effects: $x_2$

Suppose we want to study the marginal effect of $x_2$

$$\frac{\partial E (y|x_1, x_2, d_1, d_2)}{\partial x_2}$$

This is given by

$$\frac{\partial E (y|x_1, x_2, d_1, d_2)}{\partial x_2} = \beta_2 + 2x_2\beta_4 + x_1\beta_5 + d_1\beta_9$$

- I can compute this effect for every individual in my sample and then average to get a population averaged effect.
- I could evaluate this conditional on values of the different covariates, or even values of importance for $x_2$. 
Population averaged effect manually

```plaintext
. regress, coeflegend

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 10,000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F(18, 9981) = 388.10</td>
<td>Prob &gt; F = 0.0000</td>
<td>R-squared = 0.4117</td>
<td>Adj R-squared = 0.4107</td>
</tr>
<tr>
<td>Model</td>
<td>335278.744</td>
<td>18</td>
<td>18626.5969</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>479031.227</td>
<td>9,981</td>
<td>47.9943119</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>814309.971</td>
<td>9,999</td>
<td>81.439141</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>yr</th>
<th>Coef.</th>
<th>Legend</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>-1.04884</td>
<td>_b[x1]</td>
</tr>
<tr>
<td>x2</td>
<td>.4749664</td>
<td>_b[x2]</td>
</tr>
<tr>
<td>c.x1#c.x2</td>
<td>1.06966</td>
<td>_b[c.x1#c.x2]</td>
</tr>
<tr>
<td>c.x1#c.x1</td>
<td>-1.061312</td>
<td>_b[c.x1#c.x1]</td>
</tr>
<tr>
<td>c.x2#c.x2</td>
<td>1.177785</td>
<td>_b[c.x2#c.x2]</td>
</tr>
<tr>
<td>d1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-1.504705</td>
<td>_b[1.d1]</td>
</tr>
<tr>
<td>2</td>
<td>-3.727184</td>
<td>_b[2.d1]</td>
</tr>
<tr>
<td>3</td>
<td>-6.522121</td>
<td>_b[3.d1]</td>
</tr>
<tr>
<td>4</td>
<td>-8.80982</td>
<td>_b[4.d1]</td>
</tr>
<tr>
<td>1.d2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.615761</td>
<td>_b[1.d2]</td>
</tr>
<tr>
<td>d1#d2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-3.649372</td>
<td>_b[1.d1#c.x2]</td>
</tr>
<tr>
<td>2</td>
<td>-5.994454</td>
<td>_b[2.d1#c.x2]</td>
</tr>
<tr>
<td>3</td>
<td>-8.457034</td>
<td>_b[3.d1#c.x2]</td>
</tr>
<tr>
<td>4</td>
<td>-11.04842</td>
<td>_b[4.d1#c.x2]</td>
</tr>
</tbody>
</table>
```

September 18, 2018 15/112
Population averaged effect manually

generate double dydx2 = _b[c.x2] + ///
_b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2 + ///
_b[1.d1#c.x2]*1.d1 + _b[2.d1#c.x2]*2.d1 + ///
_b[3.d1#c.x2]*3.d1 + _b[4.d1#c.x2]*4.d1
Population averaged effect manually

```
. list dydx2 in 1/10, sep(0)

<table>
<thead>
<tr>
<th>dydx2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
</tr>
<tr>
<td>2.</td>
</tr>
<tr>
<td>3.</td>
</tr>
<tr>
<td>4.</td>
</tr>
<tr>
<td>5.</td>
</tr>
<tr>
<td>6.</td>
</tr>
<tr>
<td>7.</td>
</tr>
<tr>
<td>8.</td>
</tr>
<tr>
<td>9.</td>
</tr>
<tr>
<td>10.</td>
</tr>
</tbody>
</table>
```

```
. summarize dydx2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>dydx2</td>
<td>10,000</td>
<td>5.4391</td>
<td>2.347479</td>
<td>-2.0755</td>
<td>12.9045</td>
</tr>
</tbody>
</table>
```
Population averaged effect manually

```
. list dydx2 in 1/10, sep(0)

<table>
<thead>
<tr>
<th>dydx2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
</tr>
<tr>
<td>2.</td>
</tr>
<tr>
<td>3.</td>
</tr>
<tr>
<td>4.</td>
</tr>
<tr>
<td>5.</td>
</tr>
<tr>
<td>6.</td>
</tr>
<tr>
<td>7.</td>
</tr>
<tr>
<td>8.</td>
</tr>
<tr>
<td>9.</td>
</tr>
<tr>
<td>10.</td>
</tr>
</tbody>
</table>
```

```
. summarize dydx2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>dydx2</td>
<td>10,000</td>
<td>5.43906</td>
<td>2.347479</td>
<td>-2.075498</td>
<td>12.90448</td>
</tr>
</tbody>
</table>
```
A way to compute effects of interest and their standard errors
Fundamental to construct our unified framework
Consumes factor variable notation
Operates over Stata `predict`, $E(Y|X) = X\hat{\beta}$
margins, dydx(*)

. margins, dydx(x2)
Average marginal effects                       Number of obs  =  10,000
Model VCE     : OLS
Expression    : Linear prediction, predict()
dy/dx w.r.t.  : x2

| Delta-method      | dy/dx | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|-------------------|-------|-----------|-------|-------|----------------------|
| x2                | 5.43906 | .1188069  | 45.78 | 0.000 | 5.206174 5.671945    |

- **Expression**, default prediction \( E(Y|X) = X\beta \)
  - This means you could access other Stata predictions
  - Or any function of the coefficients

- Delta method is the way the standard errors are computed
. margins, expression(_b[c.x2] + ///
> _b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2 + ///
> _b[1.d1#c.x2]*1.d1 + _b[2.d1#c.x2]*2.d1 + ///
> _b[3.d1#c.x2]*3.d1 + _b[4.d1#c.x2]*4.d1)

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

Predictive margins Number of obs = 10,000
Model VCE : OLS
Expression : _b[c.x2] + _b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2 + _b[1.d1#c.x2]*1.d1 +
> _b[2.d1#c.x2]*2.d1 + _b[3.d1#c.x2]*3.d1 + _b[4.d1#c.x2]*4.d1

|          | Delta-method | Margin | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|----------|--------------|--------|-----------|-------|-------|---------------------|
| _cons    |              | 5.43906 | 0.1188069 | 45.78 | 0.000 | 5.206202--5.671917  |
Delta Method and Standard Errors

We get our standard errors from the central limit theorem.

\[ \hat{\beta} - \beta \xrightarrow{d} N(0, V) \]

We can get standard errors for any smooth function \( g() \) of \( \hat{\beta} \) with

\[ g(\hat{\beta}) - g(\beta) \xrightarrow{d} N(0, g'(\beta)' V g'(\beta)) \]
Effect of $x_2$: revisited

\[
\frac{\partial E(y|x_1, x_2, d_1, d_2)}{\partial x_2} = \beta_2 + 2x_2 \beta_4 + x_1 \beta_5 + d_1 \beta_9
\]

- We averaged this function but could evaluate it at different values of the covariates for example:
  - What is the average marginal effect of $x_2$ for different values of $d_1$
  - What is the average marginal effect of $x_2$ for different values of $d_1$ and $x_1$
Effect of $x_2$: revisited

\[ \frac{\partial E (y|x_1, x_2, d_1, d_2)}{\partial x_2} = \beta_2 + 2x_2\beta_4 + x_1\beta_5 + d_1\beta_9 \]

- We averaged this function but could evaluate it at different values of the covariates for example:
  - What is the average marginal effect of $x_2$ for different values of $d_1$?
- Counterfactual: What if everyone in the population had a level of $d_1 = 0$. What if $d_1 = 1$, ...
Effect of $x_2$: revisited

$$\frac{\partial E (y| x_1, x_2, d_1, d_2)}{\partial x_2} = \beta_2 + 2x_2\beta_4 + x_1\beta_5 + d_1\beta_9$$

- We averaged this function but could evaluate it at different values of the covariates for example:
  - What is the average marginal effect of $x_2$ for different values of $d_1$?
- Counterfactual: What if everyone in the population had a level of $d_1 = 0$. What if $d_1 = 1$, ...
Different values of $d_1$ a counterfactual

```plaintext
generate double dydx2 = _b[c.x2] + ///
   _b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2 + ///
   _b[1.d1#c.x2]*1.d1 + _b[2.d1#c.x2]*2.d1 + ///
   _b[3.d1#c.x2]*3.d1 + _b[4.d1#c.x2]*4.d1

generate double dydx2_d10 = _b[c.x2] + ///
   _b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2

generate double dydx2_d11 = _b[c.x2] + ///
   _b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2 + ///
   _b[1.d1#c.x2]

generate double dydx2_d12 = _b[c.x2] + ///
   _b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2 + ///
   _b[2.d1#c.x2]
```

September 18, 2018 24 / 112
Different values of $d_1$ a counterfactual

```plaintext
generate double dydx2 = _b[c.x2] + ///
   _b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2 + ///
   _b[1.d1#c.x2]*1.d1 + _b[2.d1#c.x2]*2.d1 + ///
   _b[3.d1#c.x2]*3.d1 + _b[4.d1#c.x2]*4.d1

generate double dydx2_d10 = _b[c.x2] + ///
   _b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2

generate double dydx2_d11 = _b[c.x2] + ///
   _b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2 + ///
   _b[1.d1#c.x2]

generate double dydx2_d12 = _b[c.x2] + ///
   _b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2 + ///
   _b[2.d1#c.x2]
```
Different values of $d_1$ a counterfactual

```
generate double dydx2 = _b[c.x2] + //_b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2 + //_b[1.d1#c.x2]*1.d1 + _b[2.d1#c.x2]*2.d1 + //_b[3.d1#c.x2]*3.d1 + _b[4.d1#c.x2]*4.d1

generate double dydx2_d10 = _b[c.x2] + //_b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2

generate double dydx2_d11 = _b[c.x2] + //_b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2 + //_b[1.d1#c.x2]

generate double dydx2_d12 = _b[c.x2] + //_b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2 + //_b[2.d1#c.x2]
```
Different values of $d_1$ a counterfactual

generate double dydx2 = _b[c.x2] + ///
_b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2 + ///
_b[1.d1#c.c.x2]*1.d1 + _b[2.d1#c.c.x2]*2.d1 + ///
_b[3.d1#c.c.x2]*3.d1 + _b[4.d1#c.c.x2]*4.d1

generate double dydx2_d10 = _b[c.x2] + ///
_b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2

generate double dydx2_d11 = _b[c.x2] + ///
_b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2 + ///
_b[1.d1#c.c.x2]

generate double dydx2_d12 = _b[c.x2] + ///
_b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2 + ///
_b[2.d1#c.c.x2]
Average marginal effect of $x_2$ at counterfactuals: manually

```
. summarize dydx2_*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>dydx2_d10</td>
<td>10,000</td>
<td>3.295979</td>
<td>1.7597</td>
<td>-2.411066</td>
<td>9.288564</td>
</tr>
<tr>
<td>dydx2_d11</td>
<td>10,000</td>
<td>4.414028</td>
<td>1.7597</td>
<td>-1.293017</td>
<td>10.40661</td>
</tr>
<tr>
<td>dydx2_d12</td>
<td>10,000</td>
<td>5.214277</td>
<td>1.7597</td>
<td>-0.4927681</td>
<td>11.20686</td>
</tr>
<tr>
<td>dydx2_d13</td>
<td>10,000</td>
<td>6.780233</td>
<td>1.7597</td>
<td>1.073188</td>
<td>12.77282</td>
</tr>
<tr>
<td>dydx2_d14</td>
<td>10,000</td>
<td>7.556677</td>
<td>1.7597</td>
<td>1.849632</td>
<td>13.54926</td>
</tr>
</tbody>
</table>
```
Average marginal effect of $x_2$ at counterfactuals:

```
. margins d1, dydx(x2)
Average marginal effects
Model VCE : OLS
Expression : Linear prediction, predict()
dy/dx w.r.t. : x2

| d1 | dy/dx     | Std. Err. | t    | P>|t|   | [95% Conf. Interval] |
|-----|-----------|-----------|------|-------|---------------------|
| 0   | 3.295979  | .2548412  | 12.93| 0.000 | 2.796439 3.795519   |
| 1   | 4.414028  | .2607174  | 16.93| 0.000 | 3.90297 4.925087   |
| 2   | 5.214277  | .2575936  | 20.24| 0.000 | 4.709342 5.719212  |
| 3   | 6.780233  | .2569613  | 26.39| 0.000 | 6.276537 7.283929  |
| 4   | 7.556677  | .2609514  | 28.96| 0.000 | 7.04516 8.068195   |
```
Graphically: `marginsplot`

Average Marginal Effects of x2 with 95% CIs

- Y-axis: Effects on Linear Prediction
- X-axis: d1

The line graph shows the average marginal effects of x2 with 95% confidence intervals.
Thou shalt not be fooled by overlapping confidence intervals

\[ \text{Var} (a - b) = \text{Var} (a) + \text{Var} (b) - 2\text{Cov}(a, b) \]

- You have \( \text{Var} (a) \) and \( \text{Var} (b) \)
- You do not have \( 2\text{Cov}(a, b) \)
Thou shalt not be fooled by overlapping confidence intervals

```
. margins ar.d1, dydx(x2) contrast(nowald)

Contrasts of average marginal effects
Model VCE : OLS
Expression : Linear prediction, predict()
dy/dx w.r.t. : x2

<table>
<thead>
<tr>
<th>Contrast</th>
<th>dy/dx</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>x2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1 vs 0)</td>
<td>1.11805</td>
<td>.3626989</td>
<td>.4070865 1.829013</td>
</tr>
<tr>
<td>(2 vs 1)</td>
<td>.8002487</td>
<td>.3638556</td>
<td>.0870184 1.513479</td>
</tr>
<tr>
<td>(3 vs 2)</td>
<td>1.565956</td>
<td>.3603585</td>
<td>.859581 2.272332</td>
</tr>
<tr>
<td>(4 vs 3)</td>
<td>.7764441</td>
<td>.3634048</td>
<td>.0640974 1.488791</td>
</tr>
</tbody>
</table>
```
Thou shalt not be fooled by overlapping confidence intervals
Effect of $x_2$: revisited

\[ \frac{\partial E(y|x_1, x_2, d_1, d_2)}{\partial x_2} = \beta_2 + 2x_2\beta_4 + x_1\beta_5 + d_1\beta_9 \]

- We averaged this function but could evaluate it at different values of the covariates for example:
  - What is the average marginal effect of $x_2$ for different values of $d_1$ and $x_1$
Effect of $x_2$: revisited

margins d1, dydx(x2) at(x1=(-3(.5)4))
Put on your calculus hat or ask a different question

\[ \frac{\partial E (y|.)}{\partial x_2} \]

- This is our object of interest
- By definition it is the change in \( E (y|.) \) for an infinitesimal change in \( x_2 \)
- Sometimes people talk about this as a unit change in \( x_2 \)
Put on your calculus hat or ask a different question

\[ \frac{\partial E (y|.)}{\partial x_2} \]

- This is our object of interest
- By definition it is the change in \( E (y|.) \) for an infinitesimal change in \( x_2 \)
- Sometimes people talk about this as a unit change in \( x_2 \)
Put on your calculus hat or ask a different question

`. margins, dydx(x2)`

Average marginal effects
Number of obs = 10,000
Model VCE : OLS
Expression : Linear prediction, predict()
dy/dx w.r.t. : x2

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
</tr>
<tr>
<td></td>
<td>Std. Err.</td>
</tr>
<tr>
<td></td>
<td>t</td>
</tr>
<tr>
<td></td>
<td>P&gt;</td>
</tr>
<tr>
<td></td>
<td>[95% Conf. Interval]</td>
</tr>
<tr>
<td>x2</td>
<td>5.43906</td>
</tr>
<tr>
<td></td>
<td>.1188069</td>
</tr>
<tr>
<td></td>
<td>45.78</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>5.206174</td>
</tr>
<tr>
<td></td>
<td>5.671945</td>
</tr>
</tbody>
</table>

`. quietly predict double xb0`
`. quietly replace x2 = x2 + 1`
`. quietly predict double xb1`
`. generate double diff = xb1 - xb0`
`. summarize diff`

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>diff</td>
<td>10,000</td>
<td>6.616845</td>
<td>2.347479</td>
<td>-.8977125</td>
<td>14.08226</td>
</tr>
</tbody>
</table>
Put on your calculus hat or ask a different question

\[
\text{. margins, dydx(x2)}
\]

Average marginal effects
Model VCE : OLS
Expression : Linear prediction, predict()
dy/dx w.r.t. : x2

| Variable | dy/dx   | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|----------|---------|-----------|------|------|----------------------|
| x2       | 5.43906 | 0.1188069 | 45.78| 0.000| 5.206174 5.671945    |

\[
\text{. quietly predict double xb0}
\]
\[
\text{. quietly replace x2 = x2 + 1}
\]
\[
\text{. quietly predict double xb1}
\]
\[
\text{. generate double diff = xb1 - xb0}
\]
\[
\text{. summarize diff}
\]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>diff</td>
<td>10,000</td>
<td>6.616845</td>
<td>2.347479</td>
<td>-.8977125</td>
<td>14.08226</td>
</tr>
</tbody>
</table>
Put on your calculus hat or ask a different question

```stata
. margins, at(x2 = generate(x2)) at(x2=generate(x2+1))
Predictive margins Number of obs = 10,000
Model VCE : OLS
Expression : Linear prediction, predict()
1._at : x2 = x2
2._at : x2 = x2+1

|    | Delta-method | Margin Std. Err. | t | P>|t| | [95% Conf. Interval] |
|----|--------------|------------------|---|------|---------------------|
| _at|              |                  |   |      |                     |
| 1  |              | -.599745         | .0692779 | -8.66 | 0.000 | -.7355437 -.4639463 |
| 2  | 6.0171       | .1909195         | 31.52  | 0.000 | 5.642859 6.39134    |

. margins, at(x2 = generate(x2)) at(x2=generate(x2+1)) contrast(at(r) nowald)
Contrasts of predictive margins
Model VCE : OLS
Expression : Linear prediction, predict()
1._at : x2 = x2
2._at : x2 = x2+1

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th>Contrast Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>_at</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2 vs 1)</td>
<td>6.616845</td>
<td>.1779068</td>
<td>6.268111 6.965578</td>
</tr>
</tbody>
</table>

. summarize diff
Variable | Obs | Mean | Std. Dev. | Min | Max
----------|-----|------|-----------|-----|-----
    diff   | 10,000 | 6.616845 | 2.347479 | -.8977125 | 14.08226
```
. margins, at(x2 = generate(x2)) at(x2 = generate(x2+1))
Predictive margins Number of obs = 10,000
Model VCE : OLS
Expression : Linear prediction, predict()
1._at : x2 = x2
2._at : x2 = x2+1

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Margin</td>
<td>Std. Err.</td>
<td>t</td>
<td>P&gt;</td>
<td>t</td>
</tr>
<tr>
<td>_at</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-.599745</td>
<td>.0692779</td>
<td>-8.66</td>
<td>0.000</td>
<td>-.7355437</td>
</tr>
<tr>
<td>2</td>
<td>6.0171</td>
<td>.1909195</td>
<td>31.52</td>
<td>0.000</td>
<td>5.642859</td>
</tr>
</tbody>
</table>

. margins, at(x2 = generate(x2)) at(x2 = generate(x2+1)) contrast(at(r) nowald)
Contrasts of predictive margins
Model VCE : OLS
Expression : Linear prediction, predict()
1._at : x2 = x2
2._at : x2 = x2+1

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>_at</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2 vs 1)</td>
<td>6.616845</td>
<td>.1779068</td>
<td>6.268111</td>
<td>6.965578</td>
<td></td>
</tr>
</tbody>
</table>

. summarize diff
Variable | Obs | Mean  | Std. Dev. | Min   | Max
----------|-----|-------|-----------|-------|-----
 diff      | 10,000 | 6.616845 | 2.347479 | -.8977125 | 14.08226
Put on your calculus hat or ask a different question

```
. margins, at(x2 = generate(x2)) at(x2=generate(x2+1))
Predictive margins Number of obs = 10,000
Model VCE : OLS
Expression : Linear prediction, predict()
1._at : x2 = x2
2._at : x2 = x2+1

|            | Margin | Std. Err. | t    | P>|t|  | [95% Conf. Interval] |
|------------|--------|-----------|------|------|----------------------|
| _at        |        |           |      |      |                      |
| 1          | -.599745 | .0692779  | -8.66 | 0.000 | -.7355437 -.4639463 |
| 2          | 6.0171  | .1909195  | 31.52 | 0.000 | 5.642859  6.39134   |
```

```
. margins, at(x2 = generate(x2)) at(x2=generate(x2+1)) contrast(at(r) nowald)
Contrasts of predictive margins
Model VCE : OLS
Expression : Linear prediction, predict()
1._at : x2 = x2
2._at : x2 = x2+1

<table>
<thead>
<tr>
<th></th>
<th>Contrast</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>_at</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2 vs 1)</td>
<td>6.616845</td>
<td>.1779068</td>
<td>6.268111  6.965578</td>
</tr>
</tbody>
</table>
```

```
. summarize diff
Variable | Obs | Mean  | Std. Dev. | Min     | Max     
----------|-----|-------|-----------|---------|---------
        diff | 10,000 | 6.616845 | 2.347479 | -0.8977125 | 14.08226 
```

September 18, 2018 35/112
Ask a different question

- Marginal effects have a meaning in some contexts but are misused.
- It is difficult to interpret infinitesimal changes but we do not need to.
- We can ask about meaningful questions by talking in units that mean something to the problem we care about.
A 10 percent increase in $x_2$

```
. margins, at(x2 = generate(x2)) at(x2=generate(x2*1.1)) ///
> contrast(at(r) nowald)
```

Contrasts of predictive margins

Model VCE : OLS
Expression : Linear prediction, predict()
_1._at : $x_2 = x_2$
_2._at : $x_2 = x_2 \times 1.1$

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Delta-method</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2 vs 1)</td>
<td>.7562394</td>
<td>.0178679</td>
<td>.7212147</td>
</tr>
</tbody>
</table>
What we learned

\[ \frac{\partial E (y|x_1, x_2, d_1, d_2)}{\partial x_2} = \beta_2 + 2x_2\beta_4 + x_1\beta_5 + d_1\beta_9 \]

- Population averaged
- Counterfactual values of \(d_1\)
- Counterfactual values for \(d_1\) and \(x_1\)
- Exploring a fourth dimensional surface
What we learned

\[ \frac{\partial E (y|x_1, x_2, d_1, d_2)}{\partial x_2} = \beta_2 + 2x_2\beta_4 + x_1\beta_5 + d_1\beta_9 \]

- Population averaged
- Counterfactual values of \( d_1 \)
- Counterfactual values for \( d_1 \) and \( x_1 \)
- Exploring a fourth dimensional surface
Discrete covariates

\[ E(Y|d = d_1, \ldots) - E(Y|d = d_0, \ldots) \]

\[
\cdots
\]

\[ E(Y|d = d_k, \ldots) - E(Y|d = d_0, \ldots) \]

- The effect is the difference of the object of interest evaluated at the different levels of the discrete covariate relative to a base level.
- It can be interpreted as a treatment effect.
Effect of $d_1$

```
. margins d1
Predictive margins Number of obs = 10,000
Model VCE : OLS
Expression : Linear prediction, predict()

| Margin | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|--------|-----------|-------|------|----------------------|
| d1     |           |       |      |                      |
| 0      | 3.77553   | .1550097 | 24.36 | 0.000 | 3.47168  | 4.079381 |
| 1      | 1.784618  | .1550841 | 11.51 | 0.000 | 1.480622 | 2.088614 |
| 2      | -.6527544 | .1533701 | -4.26 | 0.000 | -.9533906 | -.3521181 |
| 3      | -2.807997 | .1535468 | -18.29 | 0.000 | -3.10898 | -2.507014 |
| 4      | -5.461784 | .1583201 | -34.50 | 0.000 | -5.772123 | -5.151445 |
```

```
. margins r.d1, contrast(nowald)
Contrasts of predictive margins
Model VCE : OLS
Expression : Linear prediction, predict()

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>d1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1 vs 0)</td>
<td>-1.990912</td>
<td>.2193128</td>
</tr>
<tr>
<td>(2 vs 0)</td>
<td>-4.428285</td>
<td>.2180388</td>
</tr>
<tr>
<td>(3 vs 0)</td>
<td>-6.583527</td>
<td>.2182232</td>
</tr>
<tr>
<td>(4 vs 0)</td>
<td>-9.237314</td>
<td>.2215769</td>
</tr>
</tbody>
</table>
```
Effect of $d_1$

```
. margins d1
Predictive margins Number of obs = 10,000
Model VCE : OLS
Expression : Linear prediction, predict()

| Margin | Std. Err. | t   | P>|t|  | [95% Conf. Interval] |
|--------|-----------|-----|------|---------------------|
| d1     |           |     |      |                     |
| 0      | 3.77553   | .1550097 | 24.36| 0.000 | 3.47168 | 4.079381 |
| 1      | 1.784618  | .1550841 | 11.51| 0.000 | 1.480622 | 2.088614 |
| 2      | -.6527544 | .1533701 | -4.26| 0.000 | -.9533906 | -.3521181 |
| 3      | -2.807997 | .1535468 | -18.29| 0.000 | -3.10898 | -2.507014 |
| 4      | -5.461784 | .1583201 | -34.50| 0.000 | -5.772123 | -5.151445 |
```

```
. margins r.d1, contrast(nowald)
Contrasts of predictive margins
Model VCE : OLS
Expression : Linear prediction, predict()

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>d1 (1 vs 0)</td>
<td>-1.990912</td>
<td>.2193128</td>
</tr>
<tr>
<td>(2 vs 0)</td>
<td>-4.428285</td>
<td>.2180388</td>
</tr>
<tr>
<td>(3 vs 0)</td>
<td>-6.583527</td>
<td>.2182232</td>
</tr>
<tr>
<td>(4 vs 0)</td>
<td>-9.237314</td>
<td>.2215769</td>
</tr>
</tbody>
</table>
```
Effect of $d_1$

```
. margins r.d1, contrast(nowald)
Contrasts of predictive margins
Model VCE : OLS
Expression : Linear prediction, predict()

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contrast</td>
<td>Std. Err.</td>
<td>[95% Conf. Interval]</td>
</tr>
<tr>
<td>d1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1 vs 0)</td>
<td>-1.990912</td>
<td>.2193128</td>
<td>-2.420809 -1.561015</td>
</tr>
<tr>
<td>(2 vs 0)</td>
<td>-4.428285</td>
<td>.2180388</td>
<td>-4.855685 -4.000884</td>
</tr>
<tr>
<td>(3 vs 0)</td>
<td>-6.583527</td>
<td>.2182232</td>
<td>-7.011289 -6.155766</td>
</tr>
<tr>
<td>(4 vs 0)</td>
<td>-9.237314</td>
<td>.2215769</td>
<td>-9.671649 -8.802979</td>
</tr>
</tbody>
</table>
```

```
. margins, dydx(d1)
Average marginal effects
Model VCE : OLS
Expression : Linear prediction, predict()
dy/dx w.r.t. : 1.d1 2.d1 3.d1 4.d1

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>t</td>
</tr>
<tr>
<td>d1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-1.990912</td>
<td>.2193128</td>
<td>-9.08</td>
</tr>
<tr>
<td>2</td>
<td>-4.428285</td>
<td>.2180388</td>
<td>-20.31</td>
</tr>
<tr>
<td>3</td>
<td>-6.583527</td>
<td>.2182232</td>
<td>-30.17</td>
</tr>
<tr>
<td>4</td>
<td>-9.237314</td>
<td>.2215769</td>
<td>-41.69</td>
</tr>
</tbody>
</table>
```

Note: dy/dx for factor levels is the discrete change from the base level.
Effect of $d_1$

```
. margins r.d1, contrast(nowald)
Contrasts of predictive margins
Model VCE : OLS
Expression : Linear prediction, predict()

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contrast</td>
<td>Std. Err.</td>
<td>[95% Conf. Interval]</td>
</tr>
<tr>
<td>dl</td>
<td>(1 vs 0)</td>
<td>-1.990912</td>
<td>.2193128</td>
</tr>
<tr>
<td></td>
<td>(2 vs 0)</td>
<td>-4.428285</td>
<td>.2180388</td>
</tr>
<tr>
<td></td>
<td>(3 vs 0)</td>
<td>-6.583527</td>
<td>.2182232</td>
</tr>
<tr>
<td></td>
<td>(4 vs 0)</td>
<td>-9.237314</td>
<td>.2215769</td>
</tr>
</tbody>
</table>
```

```
. margins, dydx(d1)
Average marginal effects
Number of obs = 10,000
Model VCE : OLS
Expression : Linear prediction, predict()
dy/dx w.r.t. : 1.d1 2.d1 3.d1 4.d1

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>t</td>
</tr>
<tr>
<td>dl</td>
<td>1</td>
<td>-1.990912</td>
<td>.2193128</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-4.428285</td>
<td>.2180388</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-6.583527</td>
<td>.2182232</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-9.237314</td>
<td>.2215769</td>
</tr>
</tbody>
</table>
```

Note: dy/dx for factor levels is the discrete change from the base level.
Effect of $d_1$
Effect of $d_1$ for $x_2$ counterfactuals

```
margins, dydx(d1) at(x2=(0(.5)3))
marginsplot, recastci(rarea) ciopts(fcolor(%30))
```
Effect of $d_1$ for $x_2$ and $d_2$ counterfactuals

margins 0.d2, dydx(d1) at(x2=(0(.5)3))
margins 1.d2, dydx(d1) at(x2=(0(.5)3))
marginsplot, recastci(rarea) ciopts(fcolor(%30))
Effect of $x_2$ and $d_1$ or $x_2$ and $x_1$

- We can think about changes of two variables at a time
- This is a bit trickier to interpret and a bit trickier to compute
- \texttt{margins} allows us to solve this problem elegantly
A change in $x_2$ and $d_1$

```
. margins r.d1, dydx(x2) contrast(nowald)

Contrasts of average marginal effects
Model VCE : OLS
Expression : Linear prediction, predict()
dy/dx w.r.t. : x2

<table>
<thead>
<tr>
<th></th>
<th>Contrast Delta-method</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>[95% Conf. Interval]</td>
</tr>
<tr>
<td>x2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1 vs 0)</td>
<td>1.11805</td>
<td>.3626989</td>
<td>.4070865</td>
</tr>
<tr>
<td>(2 vs 0)</td>
<td>1.918298</td>
<td>.3592232</td>
<td>1.214149</td>
</tr>
<tr>
<td>(3 vs 0)</td>
<td>3.484255</td>
<td>.3594559</td>
<td>2.779649</td>
</tr>
<tr>
<td>(4 vs 0)</td>
<td>4.260699</td>
<td>.362315</td>
<td>3.550488</td>
</tr>
</tbody>
</table>
```
A change in $d_1$ and $d_2$

```
. margins r.d1, dydx(d2) contrast(nowald)
Contrasts of average marginal effects
Model VCE : OLS
Expression : Linear prediction, predict()
dy/dx w.r.t. : 1.d2

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Delta-method</th>
<th>dy/dx</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.d2 (base outcome)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.d2</td>
<td>d1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1 vs 0)</td>
<td>-3.649372</td>
<td>.4383277</td>
<td>-4.508582</td>
<td>-2.790161</td>
</tr>
<tr>
<td>(2 vs 0)</td>
<td>-5.994454</td>
<td>.435919</td>
<td>-6.848943</td>
<td>-5.139965</td>
</tr>
<tr>
<td>(3 vs 0)</td>
<td>-8.457034</td>
<td>.4364173</td>
<td>-9.3125</td>
<td>-7.601568</td>
</tr>
<tr>
<td>(4 vs 0)</td>
<td>-11.04842</td>
<td>.4430598</td>
<td>-11.9169</td>
<td>-10.17993</td>
</tr>
</tbody>
</table>
```

Note: dy/dx for factor levels is the discrete change from the base level.
A change in $x_2$ and $x_1$

```
. margins, expression(_b[c.x2] +
> _b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2 +
> _b[1.d1#c.x2]*1.d1 + _b[2.d1#c.x2]*2.d1 +
> _b[3.d1#c.x2]*3.d1 + _b[4.d1#c.x2]*4.d1)

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

Average marginal effects

```
Number of obs = 10,000
Model VCE : OLS
Expression : _b[c.x2] + _b[c.x1#c.x2]*c.x1 + 2*_b[c.x2#c.x2]*c.x2 + _b[1.d1#c.x2]*1.d1 +
> _b[2.d1#c.x2]*2.d1 + _b[3.d1#c.x2]*3.d1 + _b[4.d1#c.x2]*4.d1
```

dy/dx w.r.t. : x1

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
</tr>
<tr>
<td>x1</td>
<td>1.06966</td>
</tr>
</tbody>
</table>
```
Framework

- An object of interest, $E(Y|X)$
- Questions
  - $\frac{\partial E(Y|X)}{\partial x_k}$
  - $E(Y|d = d_{level}) - E(Y|d = d_{base})$
  - Both
  - Second order terms, double derivatives
- Explore the surface
  - Population averaged
  - Effects at fixed values of covariates (counterfactuals)
Framework

- An object of interest, \( E(Y|X) \)
- Questions
  - \( \frac{\partial E(Y|X)}{\partial x_k} \)
  - \( E(Y|d = d_{\text{level}}) - E(Y|d = d_{\text{base}}) \)
  - Both
  - Second order terms, double derivatives

- Explore the surface
  - Population averaged
  - Effects at fixed values of covariates (counterfactuals)
Framework

- An object of interest, $E(Y|X)$
- Questions
  - $\frac{\partial E(Y|X)}{\partial x_k}$
  - $E(Y|d = d_{level}) - E(Y|d = d_{base})$
  - Both
  - Second order terms, double derivatives
- Explore the surface
  - Population averaged
  - Effects at fixed values of covariates (counterfactuals)
Binary outcome models

- The data generating process is given by:

\[ y = \begin{cases} 
1 & \text{if } y^* = x\beta + \epsilon > 0 \\
0 & \text{otherwise} 
\end{cases} \]

- We make an assumption on the distribution of \( \epsilon \), \( f_\epsilon \):
  - Probit: \( \epsilon \) follows a standard normal distribution
  - Logit: \( \epsilon \) follows a standard logistic distribution
  - By construction \( P(y = 1|x) = F(x\beta) \)

- This gives rise to two models:
  1. If \( F(.) \) is the standard normal distribution we have a **Probit**
  2. If \( F(.) \) is the logistic distribution we have a **Logit** model

\[ P(y = 1|x) = E(y|x) \]
The data generating process is given by:

\[ y = \begin{cases} 
1 & \text{if } y^* = x\beta + \varepsilon > 0 \\
0 & \text{otherwise} 
\end{cases} \]

We make an assumption on the distribution of \( \varepsilon \), \( f_\varepsilon \)

- **Probit**: \( \varepsilon \) follows a standard normal distribution
- **Logit**: \( \varepsilon \) follows a standard logistic distribution
- By construction \( P(y = 1|x) = F(x\beta) \)

This gives rise to two models:

1. If \( F(.) \) is the standard normal distribution we have a **Probit**
2. If \( F(.) \) is the logistic distribution we have a **Logit** model

\[ P(y = 1|x) = E(y|x) \]
Binary outcome models

The data generating process is given by:

\[ y = \begin{cases} 
1 & \text{if } y^* = x\beta + \varepsilon > 0 \\
0 & \text{otherwise} 
\end{cases} \]

We make an assumption on the distribution of \( \varepsilon \), \( f_\varepsilon \)

- Probit: \( \varepsilon \) follows a standard normal distribution
- Logit: \( \varepsilon \) follows a standard logistic distribution
- By construction \( P(y = 1|x) = F(x\beta) \)

This gives rise to two models:

1. If \( F(.) \) is the standard normal distribution we have a Probit model
2. If \( F(.) \) is the logistic distribution we have a Logit model

\[ P(y = 1|x) = E(y|x) \]
Effects

The change in the conditional probability due to a change in a covariate is given by

\[
\frac{\partial P(y|x)}{\partial x_k} = \frac{\partial F(x\beta)}{\partial x_k} \beta_k
\]

This implies that:

1. The value of the object of interest depends on \(x\)

2. The \(\beta\) coefficients only tell us the sign of the effect given that \(f(x\beta) > 0\) almost surely

For a categorical variable (factor variables)

\[
F(x\beta|d = d_i) - F(x\beta|d = d_0)
\]
### Coefficient table

```
. probit ypr c.x1##c.x2 i.d1##i.d2 i.d1#c.x1, nolog
Probit regression
Number of obs = 10,000
LR chi2(16) = 2942.75
Prob > chi2 = 0.0000
Log likelihood = -5453.1739 Pseudo R2 = 0.2125

ypr             Coef.  Std. Err.   z  P>|z|     [95% Conf. Interval]
----------------- -------- -------- -------- -------- ------------------------
    x1            -0.3272    0.0424   -7.72  0.000    -0.4102   -0.2442
    x2             0.3105    0.0234  13.26  0.000     0.2647    0.3564
c.x1#c.x2        0.3179    0.0258  12.30  0.000     0.2672    0.3685
     d1            -0.2927    0.0577   -5.08  0.000   -0.4057   -0.1797
     d2             0.2822    0.0575    4.91  0.000     0.1696    0.3949
     1.d2          -0.2747    0.0422   -6.50  0.000   -0.3575   -0.1919
  d1#d2
     1 1             0.2547    0.0818    3.11  0.002     0.0944    0.4150
     2 1             0.6621    0.0840    7.89  0.000     0.4976    0.8266
     3 1             0.8472    0.0901    9.48  0.000     0.6720    1.0222
     4 1             1.2605    0.0999   12.61  0.000     1.0646    1.4564
  d1#c.x1
     1              -0.2747    0.0422   -6.50  0.000   -0.3575   -0.1919
     2              -0.5640    0.0452  -12.47  0.000   -0.6527   -0.4754
     3              -0.9452    0.0512  -18.45  0.000   -1.0456   -0.8448
     4              -1.2206    0.0608  -20.05  0.000   -1.3399  -1.1013
```

Effects of $x_2$

```
. margins, at(x2=generate(x2)) at(x2=generate(x2*1.2))
Predictive margins
Number of obs = 10,000
Model VCE : OIM
Expression : Pr(ypr), predict()
1._at : x2 = x2
2._at : x2 = x2*1.2

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Margin</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
<td></td>
<td>[95% Conf. Interval]</td>
</tr>
<tr>
<td>_at</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.4817093</td>
<td>.0043106</td>
<td>111.75</td>
<td>.000</td>
<td>.4732607</td>
<td>.4901579</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.5039467</td>
<td>.0046489</td>
<td>108.40</td>
<td>.000</td>
<td>.4948349</td>
<td>.5130585</td>
<td></td>
</tr>
</tbody>
</table>
```
Effects of $x_2$ at values of $d_1$ and $d_2$

margins $d_1#d_2$,

at ($x_2=$generate($x_2$)) at ($x_2=$generate($x_2*1.2$))
. quietly logit ypr c.x1##c.x2 i.d1##i.d2 i.d1#c.x1
. quietly margins d1#d2, at(x2=generate(x2)) at(x2=generate(x2*1.2)) post
. estimates store logit
. quietly probit ypr c.x1##c.x2 i.d1##i.d2 i.d1#c.x1
. quietly margins d1#d2, at(x2=generate(x2)) at(x2=generate(x2*1.2)) post
. estimates store probit
## Logit vs. Probit

```
. estimates table probit logit

<table>
<thead>
<tr>
<th>Variable</th>
<th>probit</th>
<th>logit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>_at#d1#d2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 0 0</td>
<td>.53151657</td>
<td>.53140462</td>
</tr>
<tr>
<td>1 0 1</td>
<td>.63756257</td>
<td>.63744731</td>
</tr>
<tr>
<td>1 1 0</td>
<td>.42306578</td>
<td>.42322182</td>
</tr>
<tr>
<td>1 1 1</td>
<td>.62291206</td>
<td>.62262466</td>
</tr>
<tr>
<td>1 2 0</td>
<td>.30922733</td>
<td>.30975991</td>
</tr>
<tr>
<td>1 2 1</td>
<td>.62783902</td>
<td>.62775349</td>
</tr>
<tr>
<td>1 3 0</td>
<td>.26973385</td>
<td>.26845746</td>
</tr>
<tr>
<td>1 3 1</td>
<td>.59004519</td>
<td>.58834989</td>
</tr>
<tr>
<td>1 4 0</td>
<td>.21809081</td>
<td>.21827411</td>
</tr>
<tr>
<td>1 4 1</td>
<td>.5914183</td>
<td>.59140961</td>
</tr>
<tr>
<td>2 0 0</td>
<td>.55723572</td>
<td>.55751404</td>
</tr>
<tr>
<td>2 0 1</td>
<td>.66005549</td>
<td>.65979041</td>
</tr>
<tr>
<td>2 1 0</td>
<td>.4502963</td>
<td>.45117594</td>
</tr>
<tr>
<td>2 1 1</td>
<td>.64854781</td>
<td>.64854287</td>
</tr>
<tr>
<td>2 2 0</td>
<td>.33082849</td>
<td>.33120501</td>
</tr>
<tr>
<td>2 2 1</td>
<td>.65472273</td>
<td>.65506022</td>
</tr>
<tr>
<td>2 3 0</td>
<td>.28400721</td>
<td>.28169093</td>
</tr>
<tr>
<td>2 3 1</td>
<td>.61605961</td>
<td>.61442653</td>
</tr>
<tr>
<td>2 4 0</td>
<td>.22609365</td>
<td>.22538232</td>
</tr>
<tr>
<td>2 4 1</td>
<td>.6154092</td>
<td>.61499622</td>
</tr>
</tbody>
</table>
```
Logit vs. Probit

Predictive Margins of $d_1 \# d_2$ with 95% CIs

- $d_2=0$, $x_2=\text{generate}(x_2)$
- $d_2=0$, $x_2=\text{generate}(x_2 \times 1.2)$
- $d_2=1$, $x_2=\text{generate}(x_2)$
- $d_2=1$, $x_2=\text{generate}(x_2 \times 1.2)$
Fractional models and quasilikelihood (pseudolikelihood)

- Likelihood models assume we know the unobservable and all its moments
- Quasilikelihood models are agnostic about anything but the first moment
- Fractional models use the likelihood of a probit or logit to model outcomes in \([0, 1]\). The unobservable of the probit and logit does not generate values in \((0, 1)\)
- Stata has an implementation for fractional probit and fractional logit models
The model

\[ E(Y|X) = F(X\beta) \]

- \( F(.) \) is a known c.d.f
- No assumptions are made about the distribution of the unobservable
Two fractional model examples

. clear
. set obs 10000
number of observations (_N) was 0, now 10,000
. set seed 111
. generate e = rnormal()
. generate x = rchi2(5)-3
. generate xb = .5*(1 - x)
. generate yp = xb + e > 0
. generate yf = normal(xb + e)

- In both cases $E(Y|X) = \Phi(X\theta)$
- For $yp$, the probit, $\theta = \beta$
- For $yf$, $\theta = \frac{\beta}{\sqrt{1+\sigma^2}}$
Two fractional model examples

```
. clear
. set obs 10000
number of observations (_N) was 0, now 10,000
. set seed 111
. generate e = rnormal()
. generate x = rchi2(5)-3
. generate xb = .5*(1 - x)
. generate yp = xb + e > 0
. generate yf = normal(xb + e)
```

- In both cases $E(Y|X) = \Phi(X\theta)$
- For $y_p$, the probit, $\theta = \beta$
- For $y_f$, $\theta = \frac{\beta}{\sqrt{1+\sigma^2}}$
Two fractional model estimates

```
. quietly fracreg probit yp x
. estimates store probit
. quietly fracreg probit yf x
. estimates store frac
. estimates table probit frac, eq(1)
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>probit</th>
<th>frac</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-.50037834</td>
<td>-.35759981</td>
</tr>
<tr>
<td>x</td>
<td>.48964237</td>
<td>.34998136</td>
</tr>
<tr>
<td>_cons</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

```
. display .5/sqrt(2)
.35355339
```
Fractional regression output

```
.fracreg probit ypr c.x1##c.x2 i.d1##i.d2 i.d1#c.x1
Iteration 0:  log pseudolikelihood = -7021.8384
Iteration 1:  log pseudolikelihood = -5515.9431
Iteration 2:  log pseudolikelihood = -5453.7326
Iteration 3:  log pseudolikelihood = -5453.1743
Iteration 4:  log pseudolikelihood = -5453.1739
Fractional probit regression Number of obs = 10,000
Wald chi2(16) = 1969.26
Prob > chi2 = 0.0000
Log pseudolikelihood = -5453.1739 Pseudo R2 = 0.2125

Robust

|     | Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|-----|-------|-----------|------|------|----------------------|
| ypr |       |           |      |      |                      |
|x1  | -.3271742 | .0421567 | -7.76 | 0.000 | -.4097998 -.2445486 |
|x2  | .3105438  | .0232016 | 13.38 | 0.000 | .2650696 .356018   |
|c.x1#c.x2 | .3178514 | .0254263 | 12.50 | 0.000 | .2680168 .3676859 |
|d1  |       |           |      |      |                      |
|1   | -.2927285 | .0577951 | -5.06 | 0.000 | -.4060049 -.1794521 |
|2   | -.6605838 | .0593091 | -11.14| 0.000 | -.7768275 -.54434  |
|3   | -.9137215 | .0655808 | -13.93| 0.000 | -.1042258 -.7851855|
|4   | -1.276209 | .0720675 | -17.71| 0.000 | -1.417459 -1.134959|
|d1#d2|       |           |      |      |                      |
|1   | .2547359  | .0817911 | 3.11 | 0.002 | .0944284 .4150435 |
|2   | .6621119  | .0839477 | 7.89 | 0.000 | .4975774 .8266464 |
|3   | .8471544  | .0896528 | 9.45 | 0.000 | .6714382 1.022871 |
|4   | 1.260509  | .0999594 |12.61 | 0.000 | 1.064592 1.456425 |
|d1#c.x1|       |           |      |      |                      |
|1   | -.2747025 | .041962  | -6.55| 0.000 | -.3569466 -.1924585|
|2   | -.5640486 | .0447828 | -12.60| 0.000 | -.6518212 -.4762759|
|3   | -.9452172 | .0514524 | -18.37| 0.000 | -.1046062 -.8443723|
|4   | -1.220618 | .0615741 | -19.82| 0.000 | -1.341301 -1.099935|
```

September 18, 2018 62/112
Robust standard errors

- In general, this means we are agnostic about the $E(\varepsilon\varepsilon' | X)$, about the conditional variance.
- The intuition from linear regression (heteroskedasticity) does not extend.
- In nonlinear likelihood-based models like probit and logit this is not the case.
Robust standard errors

- In general, this means we are agnostic about the $E(\varepsilon \varepsilon' | X)$, about the conditional variance
- The intuition from linear regression (heteroskedasticity) does not extend
- In nonlinear likelihood-based models like probit and logit this is not the case
Nonlinear likelihood models and heteroskedasticity

. clear
. set seed 111
. set obs 10000
number of observations (_N) was 0, now 10,000
. generate x = rbeta(2,3)
. generate e1 = rnormal(0, x)
. generate e2 = rnormal(0, 1)
. generate y1 = .5 - .5*x + e1 >0
. generate y2 = .5 - .5*x + e2 >0
Nonlinear likelihood models and heteroskedasticity

```
. probit y1 x, nolog
Probit regression
Number of obs = 10,000
LR chi2(1)    = 1409.02
Prob > chi2   = 0.0000
Log likelihood = -4465.3713 Pseudo R2 = 0.1363

y1      Coef.     Std. Err.     z    P>|z|      [95% Conf. Interval]
--------- -------- ------------- ------ ------ ------------------------
     x   -2.86167   .0812023   -35.24  0.000   -3.020824   -2.702517
  _cons  2.090816   .0415858    50.28  0.000    2.009309    2.172322

. probit y2 x, nolog
Probit regression
Number of obs = 10,000
LR chi2(1)    =  62.36
Prob > chi2   = 0.0000
Log likelihood = -6638.0701 Pseudo R2 = 0.0047

y2      Coef.     Std. Err.     z    P>|z|      [95% Conf. Interval]
--------- -------- ------------- ------ ------ ------------------------
     x   -.5019177   .0636248   -7.89  0.000   -.6266199   -.3772154
    _cons  .4952327   .0290706    17.04  0.000    .4382554    .55221
```
Nonparametric regression

- Nonparametric regression is agnostic
- Unlike parametric estimation, nonparametric regression assumes no functional form for the relationship between outcomes and covariates.
- You do not need to know the functional form to answer important research questions
- You are not subject to problems that arise from misspecification
Nonparametric regression

- Nonparametric regression is agnostic
- Unlike parametric estimation, nonparametric regression assumes no functional form for the relationship between outcomes and covariates.
- You do not need to know the functional form to answer important research questions
- You are not subject to problems that arise from misspecification
Nonparametric regression

- Nonparametric regression is agnostic

- Unlike parametric estimation, nonparametric regression assumes no functional form for the relationship between outcomes and covariates.

- You do not need to know the functional form to answer important research questions

- You are not subject to problems that arise from misspecification
Mean Function

- Some parametric functional form assumptions.
  - regression: $E(Y|X) = X\beta$
  - probit: $E(Y|X) = \Phi(X\beta)$
  - Poisson: $E(Y|X) = \exp(X\beta)$

- The relationship of interest is also a conditional mean:

$$E(y|X) = g(X)$$

- Where the mean function $g(\cdot)$ is unknown
Mean Function

- Some parametric functional form assumptions.
  - regression: $E(Y|X) = X\beta$
  - probit: $E(Y|X) = \Phi(X\beta)$
  - Poisson: $E(Y|X) = \exp(X\beta)$

- The relationship of interest is also a conditional mean:

  $$E(y|X) = g(X)$$

- Where the mean function $g(\cdot)$ is unknown
Mean Function

- Some parametric functional form assumptions.
  - regression: \( E(Y|X) = X\beta \)
  - probit: \( E(Y|X) = \Phi(X\beta) \)
  - Poisson: \( E(Y|X) = \exp(X\beta) \)

- The relationship of interest is also a conditional mean:

\[
E(Y|X) = g(X)
\]

- Where the mean function \( g(\cdot) \) is unknown
Mean Function

- Some parametric functional form assumptions.
  - regression: \( E(Y|X) = X\beta \)
  - probit: \( E(Y|X) = \Phi(X\beta) \)
  - Poisson: \( E(Y|X) = \exp(X\beta) \)

- The relationship of interest is also a conditional mean:
  \[
  E(y|X) = g(X)
  \]

- Where the mean function \( g(\cdot) \) is unknown.
Traditional Approach to Nonparametric Estimation

- A cross section of counties
- citations: Number of monthly drunk driving citations
- fines: The value of fines imposed in a county in thousands of dollars if caught drinking and driving.
Traditional Approach to Nonparametric Estimation

- A cross section of counties
- citations: Number of monthly drunk driving citations
- fines: The value of fines imposed in a county in thousands of dollars if caught drinking and driving.
Implicit Relation

![Graph showing the relationship between number of monthly drunk driving citations and drunk driving fines in thousands of dollars. The graph displays a scatter plot with data points illustrating the correlation between the two variables. The x-axis represents drunk driving fines in thousands of dollars, ranging from 7 to 12, while the y-axis represents the number of monthly drunk driving citations, ranging from 0 to 80. The data points form a trend indicating a negative correlation, where higher fines are associated with fewer citations.](image-url)
Simple linear regression

![Scatter plot showing the relationship between drunk driving fines and some other variable. The plot includes a linear trend line representing the simple linear regression model.](image-url)
Regression with nonlinearities
Poisson regression
Nonparametric Estimation of Mean Function

- lpoly citations fines
Now That We have the Mean Function

- What is the effect on the mean of citations of increasing fines by 10%?
Traditional Approach Gives Us
Additional Variables

- I would like to add controls
  - Whether county has a college town
  - Number of highway patrol units per capita in the county
- With those controls I can ask some new questions
What is the mean of citations if I increase patrols and fines?
How does the mean of citations differ for counties where there is a college town, averaging out the effect of patrols and fines?
What policy has a bigger effect on the mean of citations, an increase in fines, an increase in patrols, or a combination of both?
What We Have Is

Local polynomial smooth

Number of monthly drunk driving citations

Drunk driving fines in thousands of dollars

kernel = epanechnikov, degree = 0, bandwidth = .41
What We Have

- I have a mean function. That makes no functional form assumptions.
- I cannot answer the previous questions.
- My analysis was graphical not statistical
- My analysis is limited to one covariate
- This is true even if I give you the true mean function, $g(X)$
What We Have

- I have a mean function. That makes no functional form assumptions.
- I cannot answer the previous questions.
- My analysis was graphical not statistical
- My analysis is limited to one covariate
- This is true even if I give you the true mean function, $g(X)$
What We Have

- I have a mean function. That makes no functional form assumptions.
- I cannot answer the previous questions.
- My analysis was graphical not statistical
- My analysis is limited to one covariate
- This is true even if I give you the true mean function, $g(X)$
What We Have

- I have a mean function. That makes no functional form assumptions.
- I cannot answer the previous questions.
- My analysis was graphical not statistical
- My analysis is limited to one covariate
- This is true even if I give you the true mean function, $g(X)$
What We Have

- I have a mean function. That makes no functional form assumptions.
- I cannot answer the previous questions.
- My analysis was graphical not statistical
- My analysis is limited to one covariate
- This is true even if I give you the true mean function, $g(X)$
Nonparametric regression: discrete covariates

Mean function for a discrete covariate

- Mean (probability) of low birthweight ($lbweight$) conditional on smoking 1 to 5 cigarettes ($msmoke=1$) during pregnancy

```
.mean lbweight if msmoke==1
Mean estimation

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>lbweight</td>
<td>.1125</td>
<td>.0144375</td>
<td>.0841313 .1408687</td>
</tr>
</tbody>
</table>
```

- `regress lbweight 1.msmoke, noconstant`
- $E(lbweight|msmoke = 1)$, nonparametric estimate
Nonparametric regression: discrete covariates

Mean function for a discrete covariate

- Mean (probability) of low birthweight (lbweight) conditional on smoking 1 to 5 cigarettes (msmoke=1) during pregnancy

```
   . mean lbweight if msmoke==1

Mean estimation Number of obs = 480

                  Mean   Std. Err.     [95% Conf. Interval]
------------------- ------- -------- ------------------------------------------
  lbweight          .1125   .01444     .0841313  .1408687
```

- `regress lbweight 1.msmoke, noconstant`
- \( E(lbweight | msmoke = 1) \), nonparametric estimate
Nonparametric regression: discrete covariates

Mean function for a discrete covariate

- **Mean (probability) of low birthweight** \((lbweight)\) **conditional on smoking 1 to 5 cigarettes** \((msmoke=1)\) **during pregnancy**

\[
\text{. mean lbweight if msmoke==1}
\]

<table>
<thead>
<tr>
<th>Mean estimation</th>
<th>Number of obs</th>
<th>480</th>
</tr>
</thead>
<tbody>
<tr>
<td>lbweight</td>
<td>Mean</td>
<td>0.1125</td>
</tr>
<tr>
<td></td>
<td>Std. Err.</td>
<td>0.0144375</td>
</tr>
<tr>
<td></td>
<td>[95% Conf. Interval]</td>
<td>0.0841313</td>
</tr>
</tbody>
</table>

- **regress lbweight 1.msmoke, noconstant**

\[E(lbweigth|m_{smoke} = 1), \text{ nonparametric estimate}\]
Nonparametric regression: continuous covariates

Conditional mean for a continuous covariate

- low birthweight conditional on log of family income $\text{fincome}$
- $E(lbweight|\text{fincome} = 10.819)$
- Take observations near the value of 10.819 and then take an average
- $|\text{fincome}_i - 10.819| \leq h$
- $h$ is a small number referred to as the bandwidth
Nonparametric regression: continuous covariates

Conditional mean for a continuous covariate

- **low birthweight conditional on log of family income** $\text{fincome}$
- $E(lbweight|\text{fincome} = 10.819)$
- Take observations near the value of 10.819 and then take an average
- $|\text{fincome}_i - 10.819| \leq h$
- $h$ is a small number referred to as the bandwidth
Nonparametric regression: continuous covariates

Conditional mean for a continuous covariate

- low birthweight conditional on log of family income $fincome$
- $E(lbweight|fincome = 10.819)$
- Take observations near the value of 10.819 and then take an average
- $|fincome_i - 10.819| \leq h$
- $h$ is a small number referred to as the bandwidth
Conditional mean for a continuous covariate

- low birthweight conditional on log of family income \( \text{fincome} \)
- \( E(\text{lbweight}|\text{fincome} = 10.819) \)
- Take observations near the value of 10.819 and then take an average
  - \( |\text{fincome}_i - 10.819| \leq h \)
  - \( h \) is a small number referred to as the bandwidth
Nonparametric regression: continuous covariates

Conditional mean for a continuous covariate

- low birthweight conditional on log of family income $fincome$
- $E(lbweight|fincome = 10.819)$
- Take observations near the value of 10.819 and then take an average
- $|fincome_i - 10.819| \leq h$
- $h$ is a small number referred to as the bandwidth
Nonparametric regression: continuous covariates

Conditional mean for a continuous covariate

- low birthweight conditional on log of family income $fincome$
- $E(lbweight|fincome = 10.819)$
- Take observations near the value of 10.819 and then take an average
- $|fincome_i - 10.819| \leq h$
- $h$ is a small number referred to as the bandwidth
Graphical representation
Graphical example
Graphical example continued
Two concepts

1. $h$ !!!!

2. Definition of distance between points, $|x_i - x| \leq h$
Kernel weights

- Epanechnikov
- Gaussian
- Epanechnikov2
- Rectangular (Uniform)
- Triangular
- Biweight
- Triweight
- Cosine
- Parzen
Kernel weights

- Epanechnikov
- Gaussian
- Epanechnikov2
- Rectangular (Uniform)
- Triangular
- Biweight
- Triweight
- Cosine
- Parzen
Discrete bandwidths

- Li–Racine Kernel

\[ k (\cdot) = \begin{cases} 
1 & \text{if } x_i = x \\
h & \text{otherwise}
\end{cases} \]

- Cell mean

\[ k (\cdot) = \begin{cases} 
1 & \text{if } x_i = x \\
0 & \text{otherwise}
\end{cases} \]

Cell mean was used in the example of discrete covariate estimate:

\[ E(\text{lbweigth}|msmoke = 1) \]
Discrete bandwidths

- Li–Racine Kernel

\[ k(\cdot) = \begin{cases} 
1 & \text{if } x_i = x \\
h & \text{otherwise} 
\end{cases} \]

- Cell mean

\[ k(\cdot) = \begin{cases} 
1 & \text{if } x_i = x \\
0 & \text{otherwise} 
\end{cases} \]

- Cell mean was used in the example of discrete covariate estimate

\[ E(lbweigth|msmoke = 1) \]
Selecting The Bandwidth

- A very large bandwidth will give you a biased estimate of the mean function with a small variance.
- A very small bandwidth will give you an estimate with small bias and large variance.
Selecting The Bandwidth

- A very large bandwidth will give you a biased estimate of the mean function with a small variance
- A very small bandwidth will give you an estimate with small bias and large variance
A Large Bandwidth At One Point
A Large Bandwidth At Two Points
No Variance but Huge Bias
A Very Small Bandwidth at a Point
A Very Small Bandwidth at 4 Points
Small Bias Large Variance
Estimation

- Choose bandwidth optimally. Minimize bias–variance trade–off
  - Cross-validation (default)
  - Improved AIC (IMAIC)
- Compute a mean for every point in data (local-constant)
- Compute a regression for every point in data (local linear)
  - Computes constant (mean) and slope (effects)
  - Mean function and derivatives and effects of mean function
  - There is a bandwidth for the mean computation and another for the effects.
- Local-linear regression is the default
Estimation

- Choose bandwidth optimally. Minimize bias–variance trade–off
  - Cross-validation (default)
  - Improved AIC (IMAIC)

- Compute a mean for every point in data (local-constant)

- Compute a regression for every point in data (local linear)
  - Computes constant (mean) and slope (effects)
  - Mean function and derivatives and effects of mean function
  - There is a bandwidth for the mean computation and another for the effects.

- Local-linear regression is the default
Choose bandwidth optimally. Minimize bias–variance trade–off
  ▶ Cross-validation (default)
  ▶ Improved AIC (IMAIC)

Compute a mean for every point in data (local-constant)

Compute a regression for every point in data (local linear)
  ▶ Computes constant (mean) and slope (effects)
  ▶ Mean function and derivatives and effects of mean function
  ▶ There is a bandwidth for the mean computation and another for the effects.

Local-linear regression is the default
Estimation

- Choose bandwidth optimally. Minimize bias–variance trade–off
  - Cross-validation (default)
  - Improved AIC (IMAIC)
- Compute a mean for every point in data (local-constant)
- Compute a regression for every point in data (local linear)
  - Computes constant (mean) and slope (effects)
  - Mean function and derivatives and effects of mean function
  - There is a bandwidth for the mean computation and another for the effects.
- Local-linear regression is the default
Estimation

- Choose bandwidth optimally. Minimize bias–variance trade–off
  - Cross-validation (default)
  - Improved AIC (IMAIC)
- Compute a mean for every point in data (local-constant)
- Compute a regression for every point in data (local linear)
  - Computes constant (mean) and slope (effects)
  - Mean function and derivatives and effects of mean function
  - There is a bandwidth for the mean computation and another for the effects.
- Local-linear regression is the default
Estimation

- Choose bandwidth optimally. Minimize bias–variance trade–off
  - Cross-validation (default)
  - Improved AIC (IMAIC)
- Compute a mean for every point in data (local-constant)
- Compute a regression for every point in data (local linear)
  - Computes constant (mean) and slope (effects)
  - Mean function and derivatives and effects of mean function
  - There is a bandwidth for the mean computation and another for the effects.
- Local-linear regression is the default
Simulated data example for continuous covariate

. clear
. set obs 1000
number of observations (_N) was 0, now 1,000
. set seed 111
. generate x = (rchi2(5)-5)/10
. generate a = int(runiform()*3)
. generate e = rnormal(0, .5)
. generate y = 1 - x -a + 4*x^2*a + e
True model unknown to researchers

quietly regress y (c.x##c.x)##i.a margins a, ///
at(x=generate(x)) at(x=generate(x*1.5))
marginsplot, recastci(rarea) ciopts(fcolor(%30))
npregress Syntax

. npregress kernel y x i.a

- kernel refers to the kind of nonparametric estimation
- By default Stata assumes variables in my model are continuous
- i.a States the variable is categorical
- Interactions between continuous variables and between continuous and discrete variables are implicit
npregress Syntax

. npregress kernel y x i.a

- kernel refers to the kind of nonparametric estimation
- By default Stata assumes variables in my model are continuous
- i.a States the variable is categorical
- Interactions between continuous variables and between continuous and discrete variables are implicit
. npregress kernel y x i.a

- **kernel** refers to the kind of nonparametric estimation
- By default Stata assumes variables in my model are continuous
- **i.a** States the variable is categorical
- Interactions between continuous variables and between continuous and discrete variables are implicit
. npregress kernel y x i.a

- **kernel** refers to the kind of nonparametric estimation
- By default Stata assumes variables in my model are continuous
- **i.a** States the variable is categorical
- Interactions between continuous variables and between continuous and discrete variables are implicit
npregress Syntax

. npregress kernel y x i.a

- **kernel** refers to the kind of nonparametric estimation
- By default Stata assumes variables in my model are continuous
- **i.a** States the variable is categorical
- Interactions between continuous variables and between continuous and discrete variables are implicit
Fitting the model with \texttt{npregress}

\begin{verbatim}
. npregress kernel y x i.a, nolog

Bandwidth

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>0.0616294</td>
<td>0.0891705</td>
</tr>
<tr>
<td>a</td>
<td>0.490625</td>
<td>0.490625</td>
</tr>
</tbody>
</table>

Local-linear regression

<table>
<thead>
<tr>
<th></th>
<th>Number of obs</th>
<th>=</th>
<th>1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous kernel</td>
<td>E(Kernel obs)</td>
<td>=</td>
<td>62</td>
</tr>
<tr>
<td>Discrete kernel</td>
<td>R-squared</td>
<td>=</td>
<td>0.8409</td>
</tr>
<tr>
<td>Bandwidth</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>0.4071379</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>-0.8212713</td>
</tr>
<tr>
<td>a</td>
<td></td>
</tr>
<tr>
<td>(1 vs 0)</td>
<td>-0.5820049</td>
</tr>
<tr>
<td>(2 vs 0)</td>
<td>-1.179375</td>
</tr>
</tbody>
</table>

Note: Effect estimates are averages of derivatives for continuous covariates and averages of contrasts for factor covariates.

Note: You may compute standard errors using \texttt{vce(bootstrap)} or \texttt{reps()}.\end{verbatim}
The same effect

quietly regress y (c.x##c.x)#i.a margins a, ///
   at(x=generate(x)) at(x=generate(x*1.5))
marginsplot, recastci(rarea) ciopts(fcolor(%30))
Longitudinal/Panel Data

- Under large N and fixed asymptotics behaves like cross-sectional models.

- The difficulties arise because of time-invariant unobservables, i.e. $\alpha_i$ in

$$y_{it} = G \left( X_{it} \beta + \alpha_i + \varepsilon_{it} \right)$$

- Our framework still works but we need to be careful with what it means to average over the sample.
Averaging

- Our model gives us:

\[ E(y_{it} | X_{it}, \alpha_i) = g(X_{it} \beta + \alpha_i) \]

- We cannot consistently estimate \( \alpha_i \) so we integrate it out

\[
E_\alpha E(y_{it} | X_{it}, \alpha_i) = E_\alpha g(X_{it} \beta + \alpha_i)
\]

\[
E_\alpha E(y_{it} | X_{it}, \alpha_i) = h(X_{it} \theta)
\]

- Sometimes we know the functional form \( h(.) \). Sometimes we do not.
Averaging

- Our model gives us:

\[ E(y_{it} \| X_{it}, \alpha_i) = g(X_{it}\beta + \alpha_i) \]

- We cannot consistently estimate \( \alpha_i \) so we integrate it out

\[
E_\alpha E(y_{it} \| X_{it}, \alpha_i) = E_\alpha g(X_{it}\beta + \alpha_i)
\]

\[
E_\alpha E(y_{it} \| X_{it}, \alpha_i) = h(X_{it}\theta)
\]

- Sometimes we know the functional form \( h(.) \). Sometimes we do not.
Averaging

- Our model gives us:

  \[ E (y_{it} | X_{it}, \alpha_i) = g (X_{it} \beta + \alpha_i) \]

- We cannot consistently estimate \( \alpha_i \) so we integrate it out

  \[ E_{\alpha} E (y_{it} | X_{it}, \alpha_i) = E_{\alpha} g (X_{it} \beta + \alpha_i) \]

  \[ E_{\alpha} E (y_{it} | X_{it}, \alpha_i) = h (X_{it} \theta) \]

- Sometimes we know the functional form \( h(.) \). Sometimes we do not.
Averaging

- Our model gives us:

\[ E( y_{it} | X_{it}, \alpha_i ) = g( X_{it} \beta + \alpha_i ) \]

- We cannot consistently estimate \( \alpha_i \) so we integrate it out

\[ E_{\alpha} E( y_{it} | X_{it}, \alpha_i ) = E_{\alpha} g( X_{it} \beta + \alpha_i ) \]
\[ E_{\alpha} E( y_{it} | X_{it}, \alpha_i ) = h( X_{it} \theta ) \]

- Sometimes we know the functional form \( h(.) \). Sometimes we do not.
A probit example

```stata
. clear
. set seed 111
. set obs 5000
number of observations (_N) was 0, now 5,000
. generate id = _n
. generate a = rnormal()
. expand 10
(45,000 observations created)
. bysort id: generate year = _n
. generate x = (rchi2(5)-5)/10
. generate b = int(runiform()*3)
. generate e = rnormal()
. generate xb = .5*(-1-x + b - x*b) + a
. generate dydx = normalden(.5*(-1-x + b - x*b)/(sqrt(2)))*((-.5-.5*b)/sqrt(2))
. generate y = xb + e > 0
```
Panel data estimation

```
. xtset id year
    panel variable:  id (strongly balanced)
    time variable:  year, 1 to 10
    delta:  1 unit
. xtprobit y c.x##i.b, nolog
```

Random-effects probit regression

```
Number of obs    =      50,000
Group variable:  id
Number of groups =       5,000
Random effects u_i ~ Gaussian
Obs per group:
    min =        10
    avg =      10.0
    max =        10
Integration method:  mvaghermite
    Integration pts.  =        12
Wald chi2(5)    =    5035.63
Log likelihood  =  -27522.587  Prob > chi2    =     0.0000
```

```
|    | Coef.  | Std. Err. |      z    |     P>|z|   |      [95% Conf. Interval]       |
|----|--------|-----------|-----------|--------|---------------------------------|
| y  |        |           |           |        |                                 |
| x  | -.5212161 | .0393606  | -13.24    |  0.000 | -.5983614  -.4440708           |
| b  |        |           |           |        |                                 |
| 1  |  .4859038  | .0170101  |  28.57    |  0.000 |  .4525647  .519243             |
| 2  |  1.00774   | .0179167  |  56.25    |  0.000 |  .9726241  1.042856            |
| b#c.x|       |           |           |        |                                 |
| 1  | -.5454211  | .0557341  |  -9.79    |  0.000 | -.6546579  -.4361843           |
| 2  |-1.059613   | .0568466  | -18.64    |  0.000 | -1.17103   -.9481958           |
| _cons |   |           |           |        |                                 |
|     | -.506777   | .0187516  | -27.03    |  0.000 | -.5435294  -.4700246           |
| /lnsig2u |     |           |           |        |                                 |
|     | .0004287   | .0298177  |           |        | -.058013   .0588704            |
| sigma_u |       |           |           |        |                                 |
|     | 1.0000214  | .0149121  |           |        | .9714102   1.029873            |
| rho  |       |           |           |        |                                 |
|     | .5001072   | .0074544  |           |        | .4855008   .5147133            |
```

LR test of rho=0:  chibar2(01) = 9819.64  Prob >= chibar2 = 0.000
Effect estimation

.margins, dydx(x) over(year)

Average marginal effects
Number of obs = 50,000
Model VCE : OIM
Expression : Pr(y=1), predict(pr)
dy/dx w.r.t. : x
over : year

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-.2769118</td>
<td>.0058397</td>
<td>-47.42</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>-.2752501</td>
<td>.0058296</td>
<td>-47.22</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>-.2745409</td>
<td>.005857</td>
<td>-46.87</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>-.2769241</td>
<td>.0058773</td>
<td>-47.12</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>-.2764666</td>
<td>.0058452</td>
<td>-47.30</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>-.2731819</td>
<td>.005833</td>
<td>-46.83</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td>-.2725905</td>
<td>.005857</td>
<td>-46.54</td>
<td>0.000</td>
</tr>
<tr>
<td>8</td>
<td>-.271447</td>
<td>.0058275</td>
<td>-46.58</td>
<td>0.000</td>
</tr>
<tr>
<td>9</td>
<td>-.2745909</td>
<td>.0058566</td>
<td>-46.89</td>
<td>0.000</td>
</tr>
<tr>
<td>10</td>
<td>-.2734455</td>
<td>.0058435</td>
<td>-46.79</td>
<td>0.000</td>
</tr>
</tbody>
</table>

.summarize dydx

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>dydx</td>
<td>50,000</td>
<td>-.2609633</td>
<td>.1032875</td>
<td>-.4231422</td>
<td>-.0394023</td>
</tr>
</tbody>
</table>
Effect estimation

Average Marginal Effects of x with 95% CIs

Effects on Pr(Y=1)

year
Effect estimation

Contrasts of Average Marginal Effects of x with 95% CIs
Beware of $p u 0$ or any $\alpha_i = 0$

- The coefficients of population averaged models are useful to compute ATE:

$$ATE = E \left[ F(\mathbf{X}_{it}\delta + \delta_{treat} + \alpha_i) - F(\mathbf{X}_{it}\delta + \alpha_i) \right]$$

$$= E_x \left[ E_\alpha \left[ F(\mathbf{X}_{it}\delta + \delta_{treat} + \alpha_i) \right] \right] - E_x \left[ E_\alpha \left[ F(\mathbf{X}_{it}\delta + \alpha_i) \right] \right]$$

- When we use $\alpha_i = 0$ we get it wrong
- The reason is that $E(g(x)) \neq g(E(x))$ when $g$ is not a linear function:

$$E_x \left[ F(\mathbf{X}_{it}\delta + \delta_{treat} + 0) \right] - E_x \left[ F(\mathbf{X}_{it}\delta + 0) \right] =$$

$$E_x \left[ F(\mathbf{X}_{it}\delta + \delta_{treat} + E(\alpha_i)) \right] - E_x \left[ F(\mathbf{X}_{it}\delta + E(\alpha_i)) \right] \neq$$

$$E_x \left[ E_\alpha \left[ F(\mathbf{X}_{it}\delta + \delta_{treat} + \alpha_i) \right] \right] - E_x \left[ E_\alpha \left[ F(\mathbf{X}_{it}\delta + \alpha_i) \right] \right] = ATE$$
The coefficients of population averaged models are useful to compute ATE:

\[
ATE = E \left[ F(X_{it}\delta + \delta_{treat} + \alpha_i) - F(X_{it}\delta + \alpha_i) \right] \\
= E_x \left[ E_{\alpha} \left[ F(X_{it}\delta + \delta_{treat} + \alpha_i) \right] \right] - E_x \left[ E_{\alpha} \left[ F(X_{it}\delta + \alpha_i) \right] \right]
\]

- When we use \( \alpha_i = 0 \) we get it wrong.
- The reason is that \( E(g(x)) \neq g(E(x)) \) when \( g \) is not a linear function:

\[
E_x \left[ F(X_{it}\delta + \delta_{treat} + 0) \right] - E_x \left[ F(X_{it}\delta + 0) \right] = E_x \left[ F(X_{it}\delta + \delta_{treat} + E(\alpha_i)) \right] - E_x \left[ F(X_{it}\delta + E(\alpha_i)) \right] \neq E_x \left[ E_{\alpha} \left[ F(X_{it}\delta + \delta_{treat} + \alpha_i) \right] \right] - E_x \left[ E_{\alpha} \left[ F(X_{it}\delta + \alpha_i) \right] \right] = ATE
\]
Beware of $\mathbf{p}u0$ or any $\alpha_i = 0$

- The coefficients of population averaged models are useful to compute $\text{ATE}$:

\[
\text{ATE} = \mathbb{E} [F (X_{it}\delta + \delta_{\text{treat}} + \alpha_i) - F (X_{it}\delta + \alpha_i)]
\]

\[
= \mathbb{E}_x [\mathbb{E}_\alpha [F (X_{it}\delta + \delta_{\text{treat}} + \alpha_i)]] - \mathbb{E}_x [\mathbb{E}_\alpha [F (X_{it}\delta + \alpha_i)]]
\]

- When we use $\alpha_i = 0$ we get it wrong
- The reason is that $\mathbb{E}(g(x)) \neq g(\mathbb{E}(x))$ when $g$ is not a linear function:

\[
\mathbb{E}_x [F (X_{it}\delta + \delta_{\text{treat}} + 0)] - \mathbb{E}_x [F (X_{it}\delta + 0)] = \\
\mathbb{E}_x [F (X_{it}\delta + \delta_{\text{treat}} + \mathbb{E}(\alpha_i))] - \mathbb{E}_x [F (X_{it}\delta + \mathbb{E}(\alpha_i))]
\]

\[
\mathbb{E}_x [\mathbb{E}_\alpha [F (X_{it}\delta + \delta_{\text{treat}} + \alpha_i)]] - \mathbb{E}_x [\mathbb{E}_\alpha [F (X_{it}\delta + \alpha_i)]] = \text{ATE}
\]
Concluding Remarks

- Our work is not done after we get the parameters of our model
- After we get the parameters is when our work starts. We can ask interesting questions
- The questions we ask can be placed in a general framework:
  - Define an object of interest $E(y|X)$ or $E(y|X, \alpha)$
  - Explore the multidimensional function
- Use `margins` and `marginsplot`