

# Causal Inference for Complex Observational Data Using Stata

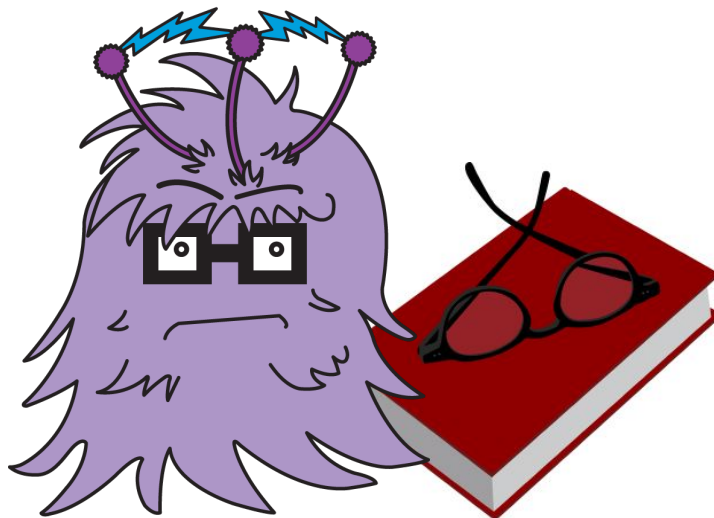
Chuck Huber  
StataCorp  
[chuber@stata.com](mailto:chuber@stata.com)

# ERMs Outline

- Description of the dataset
- Unobserved confounding and endogeneity
- Nonrandom treatment assignment
- Missing not at random (MNAR) and selection bias
- Treatment effects

# The Research Question

- Fictional State University (FSU) has developed a new study-skills program with the goal of improving the grade point averages of their students.



# The Data

```
. use gpa.dta, clear
(Simulated GPA Dataset for ERMs seminars)
```

```
. describe
```

Contains data from gpa.dta

```
obs:      1,000      Simulated GPA Dataset for ERMs seminars
vars:      9         22 Jan 2018 16:06
size:     22,000     (_dta has notes)
```

variable name	storage type	display format	value label	variable label
id	int	%9.0g		Student Identification Number
gpa	float	%9.0g		Final College Grade Point Average
hsgpa	float	%9.0g		High School Grade Point Average
program	byte	%9.0g	YesNo	Student participated in the study skills program?
graduate	byte	%9.0g	YesNo	Did the student graduate college?
income	float	%9.0g		Parent's Income (x \$100,000)
hs_comp	float	%9.0g		High School Competitiveness
roommate	byte	%9.0g	YesNo	Students's roommate is also a student?
scholarship	byte	%9.0g	YesNo	Student received scholarship funds?

Sorted by: id

# The Data

```
. summarize
```

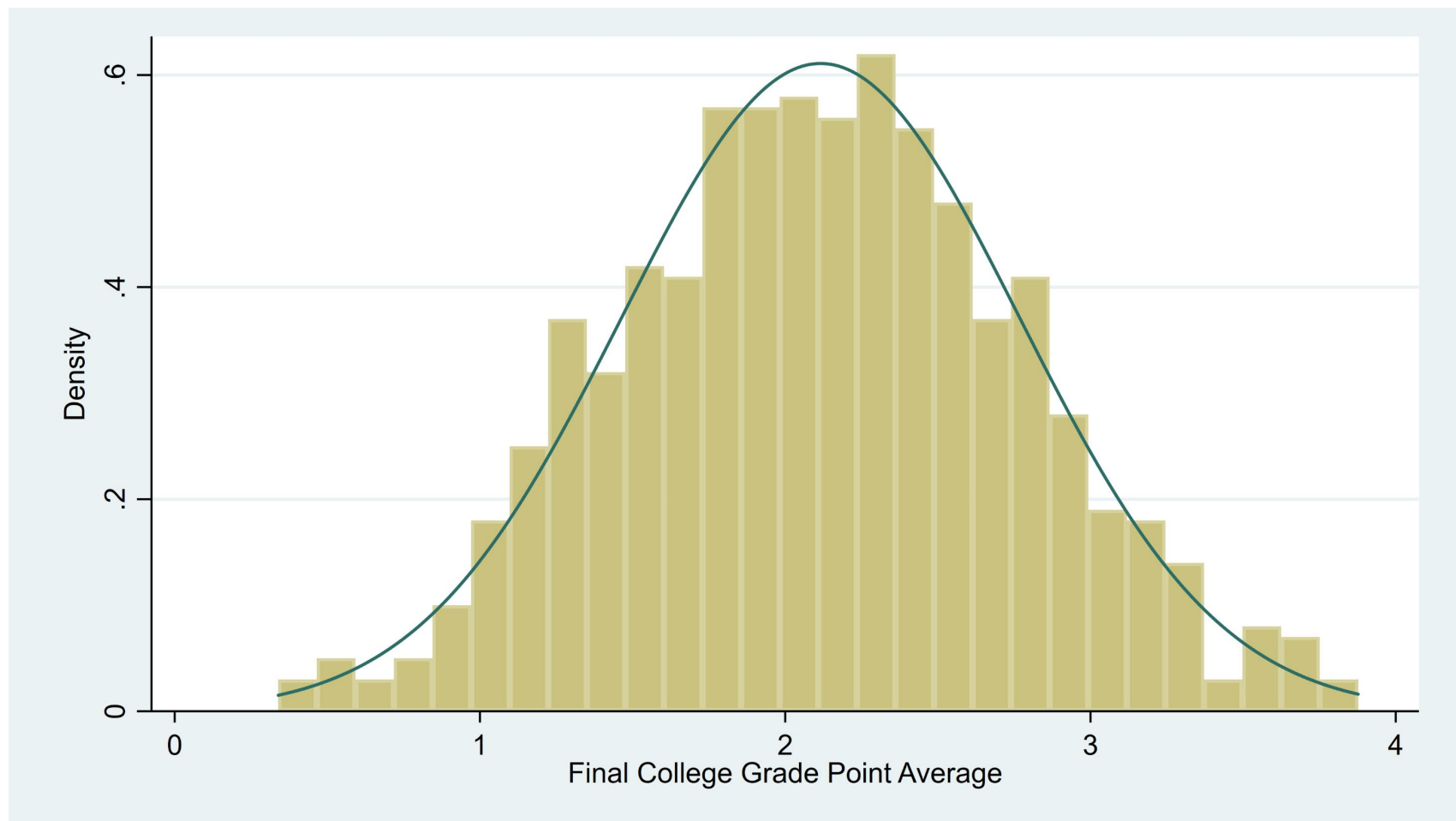
Variable	Obs	Mean	Std. Dev.	Min	Max
id	1,000	500.5	288.8194	1	1000
gpa	792	2.115962	.6529961	.3392706	3.876919
hsgpa	1,000	2.294384	.5714525	.6758502	3.786486
program	1,000	.3	.4584869	0	1
graduate	1,000	.792	.4060799	0	1
income	1,000	.5031867	.2848887	.0004344	.9969745
hs_comp	1,000	.4946027	.286164	.0001878	.9985294
roommate	1,000	.321	.4670944	0	1
scholarship	1,000	.32	.4667096	0	1

# The Data

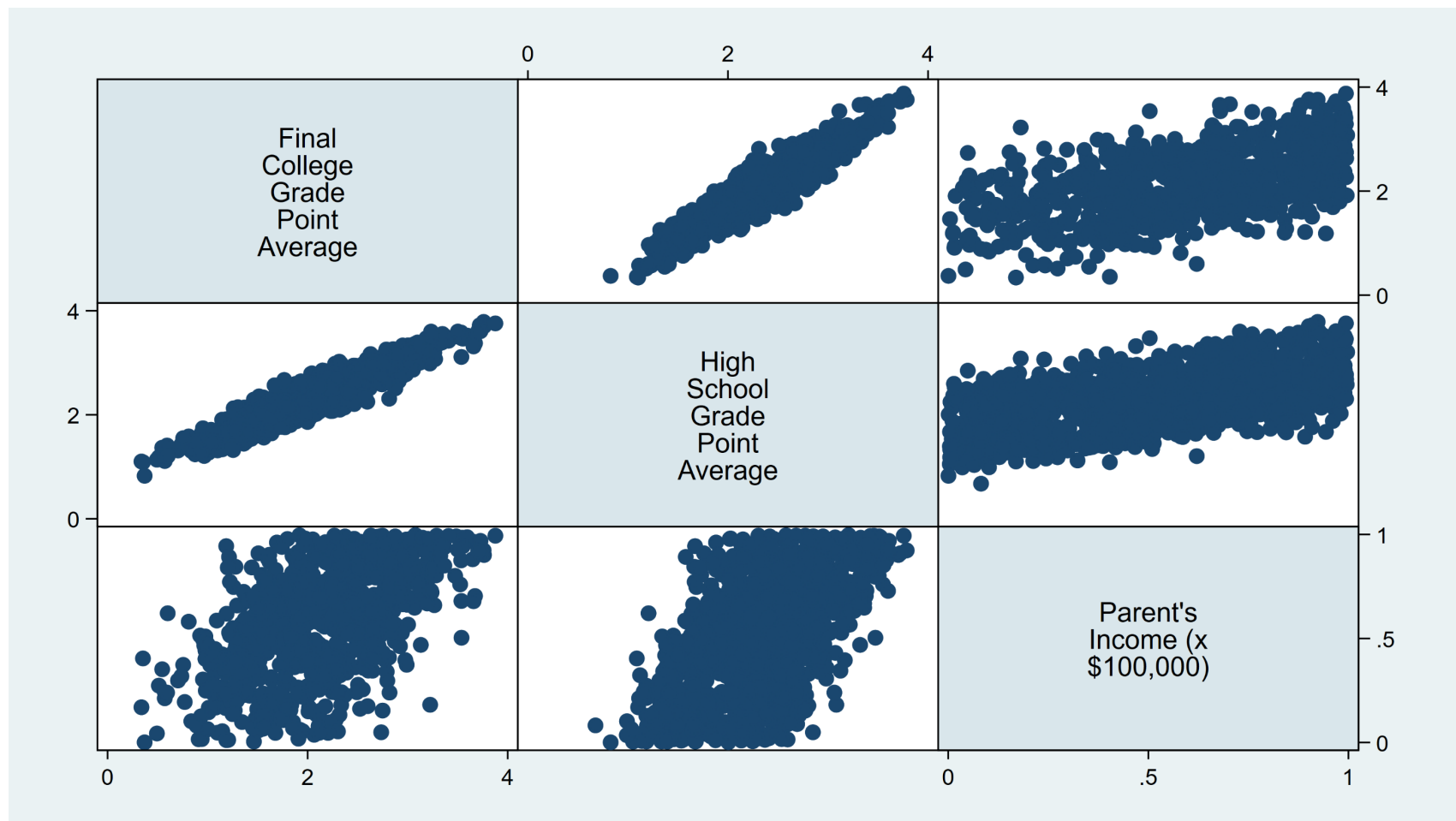
```
. tab graduate
```

Did the student graduate college?	Freq.	Percent	Cum.
No	208	20.80	20.80
Yes	792	79.20	100.00
Total	1,000	100.00	

# The Data

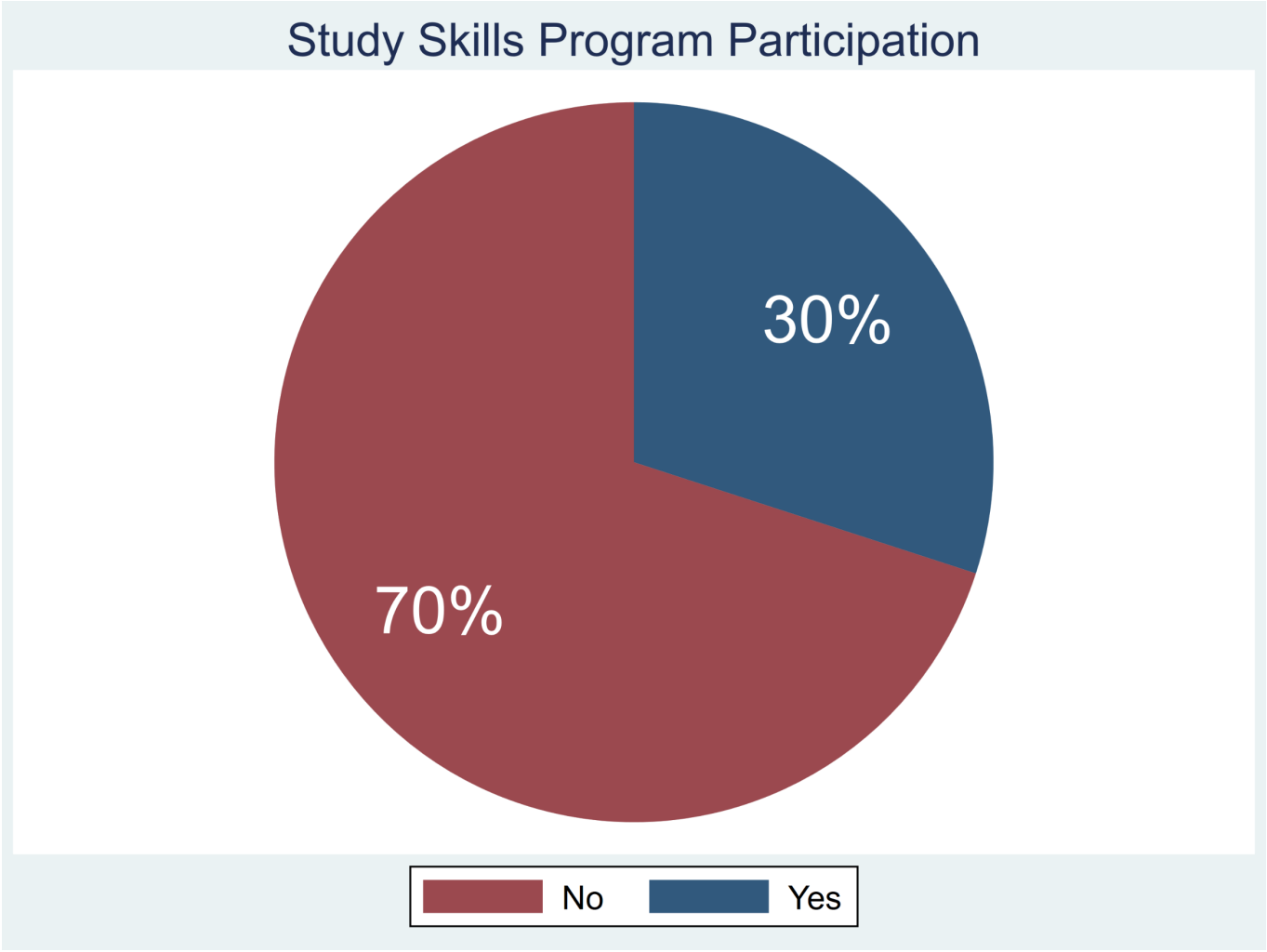


# The Data



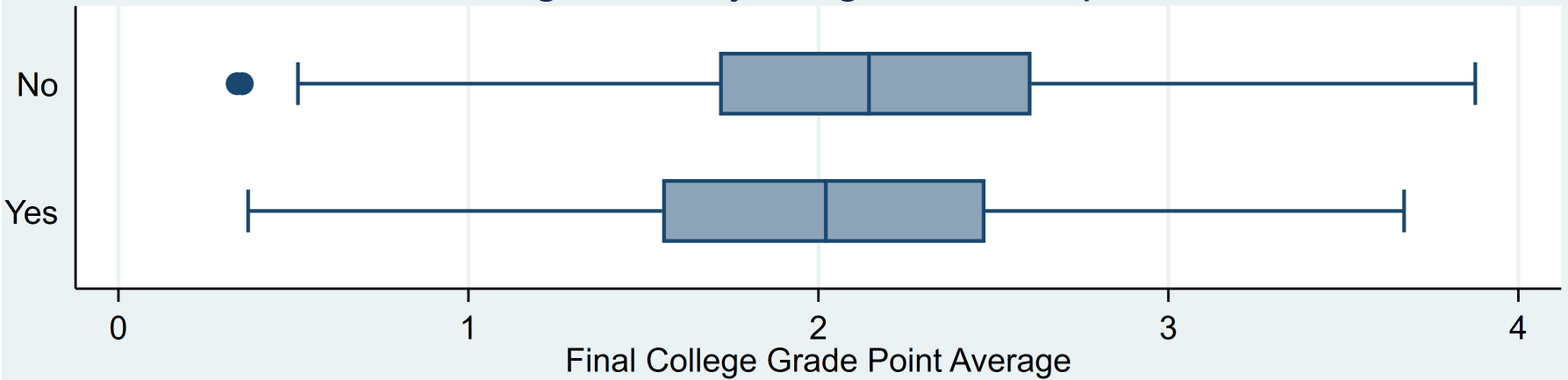


# The Data



# The Data

College GPA by Program Participation



# The Data

```
. regress gpa i.program
```

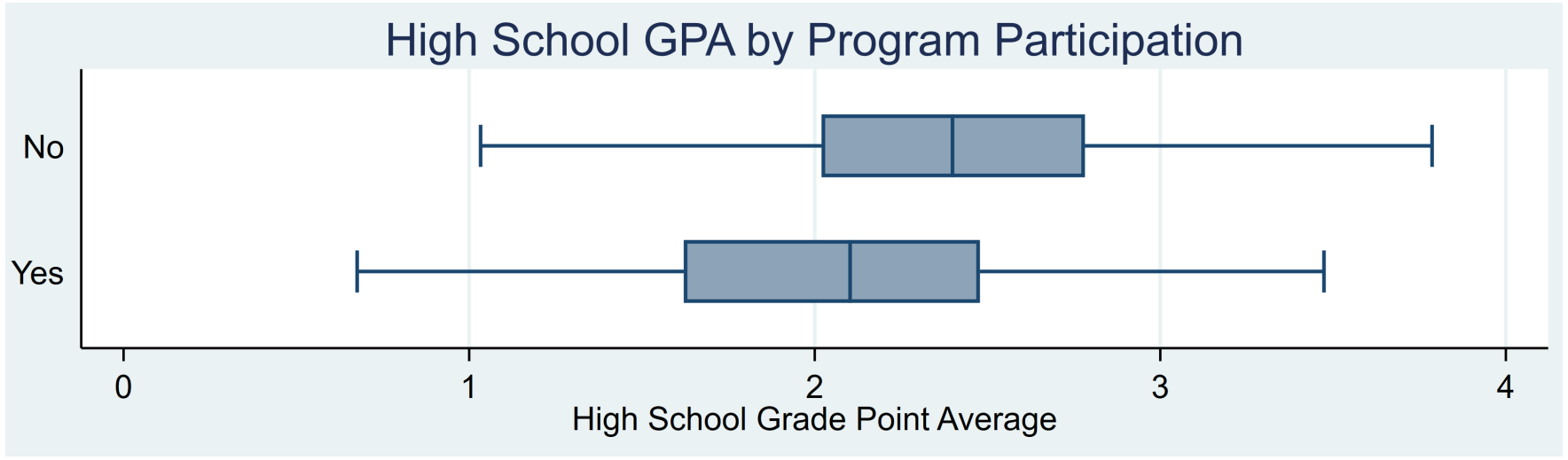
Source	SS	df	MS	Number of obs	=	792
Model	2.43242384	1	2.43242384	F(1, 790)	=	5.74
Residual	334.853048	790	.423864618	Prob > F	=	0.0168
Total	337.285472	791	.426403884	R-squared	=	0.0072
				Adj R-squared	=	0.0060
				Root MSE	=	.65105

gpa	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
program						
Yes	<span style="border: 2px solid blue; padding: 2px;">-.1259343</span>	.05257	-2.40	0.017	-.2291278	-.0227409
_cons	2.149036	.0269406	79.77	0.000	2.096152	2.201919

```
. estimates store univar
```

Students who participated in the program had **lower** GPAs?!?!?

# The Data



# The Data

```
. regress gpa i.program hsgpa
```

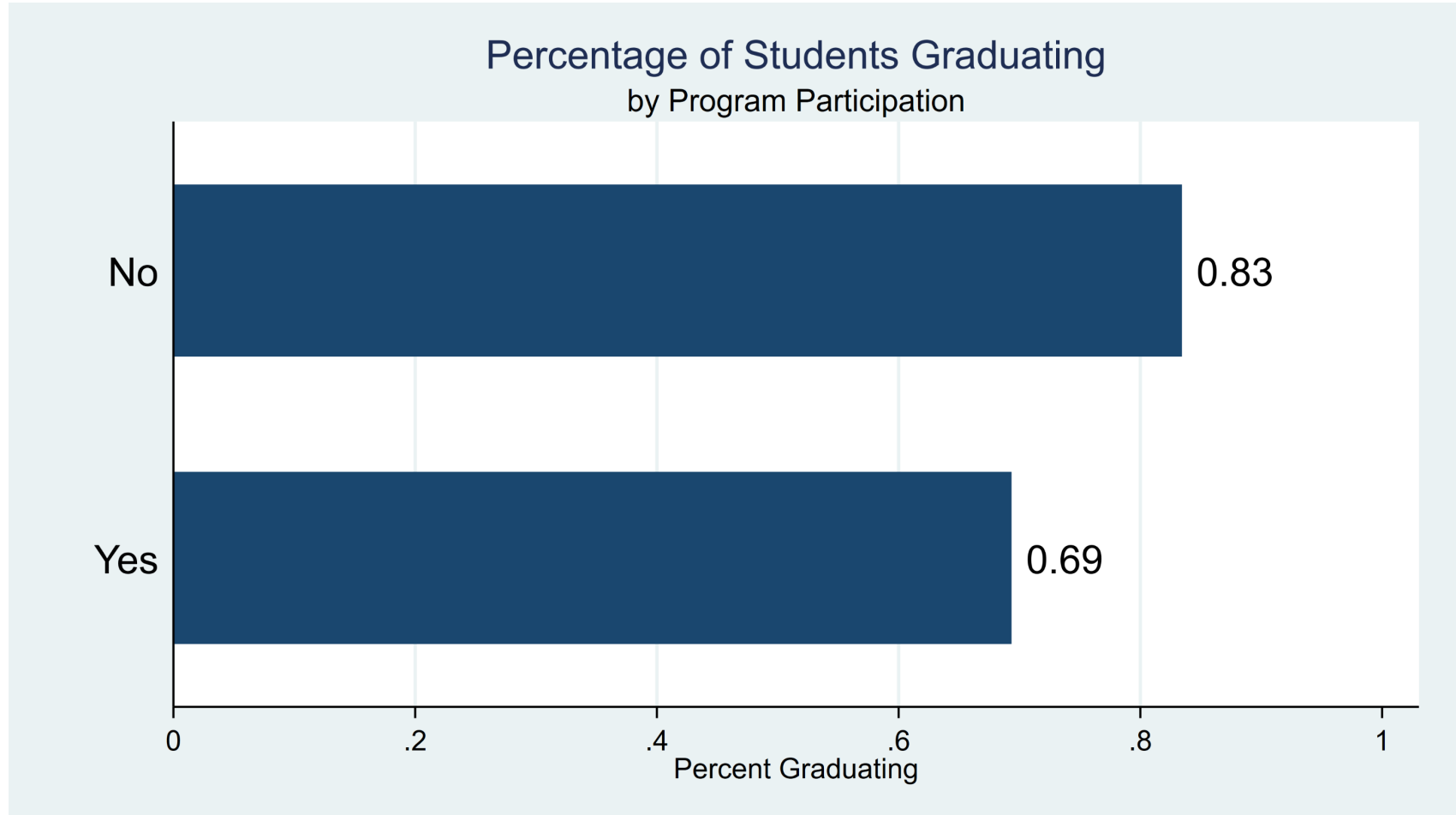
Source	SS	df	MS	Number of obs	=	792
Model	301.753841	2	150.876921	F(2, 789)	=	3350.31
Residual	35.5316304	789	.045033752	Prob > F	=	0.0000
Total	337.285472	791	.426403884	R-squared	=	0.8947
				Adj R-squared	=	0.8944
				Root MSE	=	.21221

gpa	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
program						
Yes	.2002776	.0175963	11.38	0.000	.1657364	.2348187
hsgpa	1.144457	.0140378	81.53	0.000	1.116901	1.172013
_cons	-.6744815	.035729	-18.88	0.000	-.7446166	-.6043464

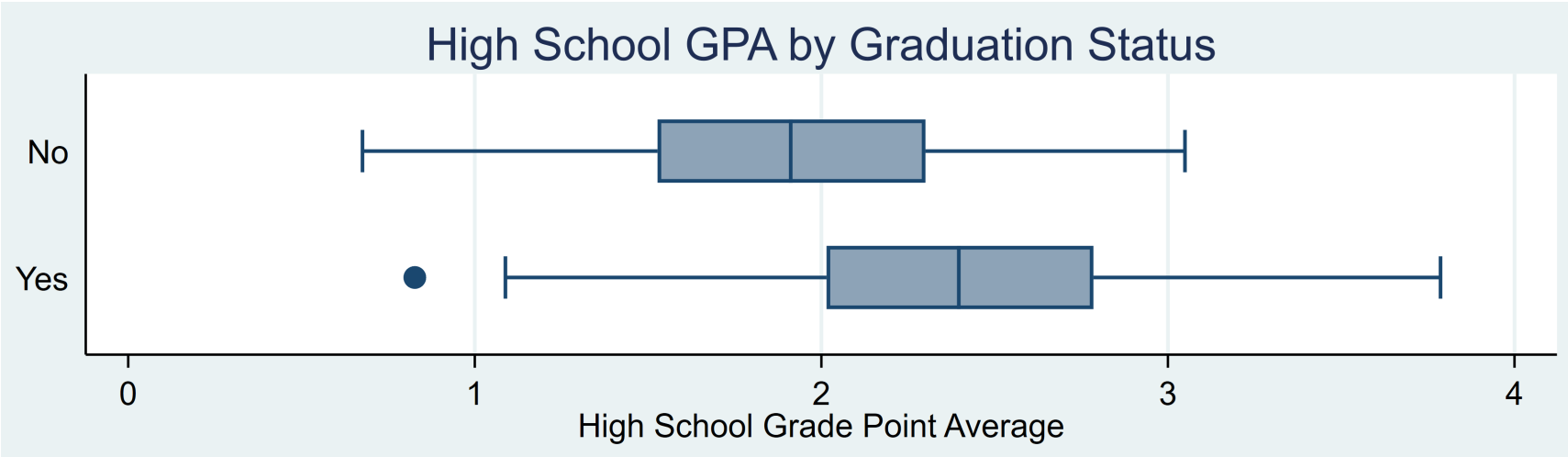
```
. estimates store hsgpa
```

Students who participated in the program had higher GPAs when we account for high school GPA.

# The Data



# The Data



# The Data

What was the effect of the study program on students GPAs?



# Outline

- ✓ • Description of the dataset
- Unobserved confounding and endogeneity
- Nonrandom treatment assignment
- Missing not at random (MNAR) and selection bias
- Treatment effects

# Observed and Unobserved Factors


$y =$  *all factors that influence  $y$*

$y =$  *observed factors* + *unobserved factors*

$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon$


# Endogeneity

“An explanatory variable in a multiple regression model that is correlated with the error term...” (Wooldridge\*, pg 838).

$$y = \beta_0 + \beta_1 x + \beta_2 z + \varepsilon$$

$$\rho_{z\varepsilon} \neq 0$$


# Omitted Variable Bias

$$y = \beta_0 + \beta_1 x + \beta_2 z + \varepsilon$$

  
 $\rho_{xz} \neq 0$

$$y = \beta_0 + \beta_1 x + \varepsilon^* \qquad \varepsilon^* = z + \varepsilon$$

$$y = \beta_0 + \beta_1 x + \varepsilon^*$$

  
 $\rho_{x\varepsilon^*} \neq 0$

# Confounding

“...X and Y are confounded when there is a third variable Z that influences both X and Y...” (Pearl\*, pg 193).

The diagram shows the regression equation  $y = \beta_0 + \beta_1 x + \beta_2 z + \varepsilon$ . The variables  $x$  and  $z$  are written in blue, while  $y$  and  $\varepsilon$  are in black. A curved arrow points from  $z$  to  $x$ , and another curved arrow points from  $z$  to  $y$ , illustrating that  $z$  is a common cause of both  $x$  and  $y$ .

$$y = \beta_0 + \beta_1 x + \beta_2 z + \varepsilon$$

# Unobserved Confounding

$$y = \beta_0 + \beta_1 x + (z + \varepsilon)$$

# Observed and Unobserved Factors

*gpa* = all factors that influence *gpa*

*gpa* = *observed factors* + *unobserved factors*

*gpa* =  $\left[ \begin{array}{l} \text{High school GPA} \\ \text{SAT Scores} \\ \text{Parents Income} \\ \text{Sex} \\ \text{etc...} \end{array} \right] + \left[ \begin{array}{l} \text{Ability} \\ \text{Motivation} \\ \text{Sleep} \\ \text{Support} \\ \text{etc...} \end{array} \right]$

$$gpa = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon$$

# Unobserved Confounding

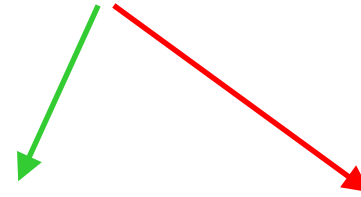
$$gpa = \beta_0 + \beta_1 income + \beta_2 hsgpa + \epsilon_{total}$$

$$gpa = \beta_0 + \beta_1 income + \beta_2 hsgpa + (Ability + \epsilon)$$




# Unobserved Confounding and Endogeneity

$$gpa = \beta_0 + \beta_1 income + \beta_2(hsgpa) + (Ability + \varepsilon_1)$$



$$hsgpa = \pi_0 + \pi_1 hs\_comp + (Ability + \varepsilon_2)^*$$

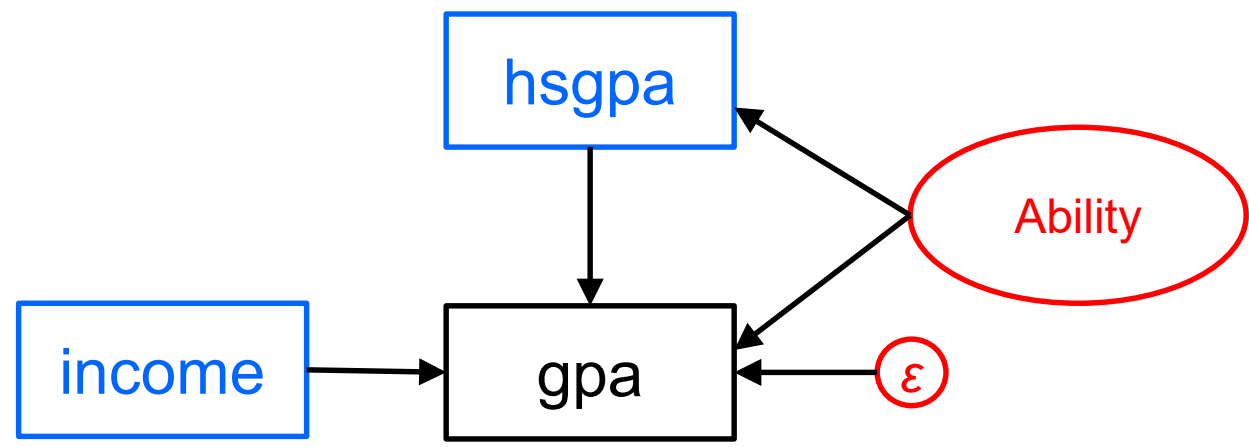
$$gpa = \beta_0 + \beta_1 income + \beta_2(\pi_0 + \pi_1 hs\_comp) + (Ability + \varepsilon_1)^*$$

where  $\rho_{\varepsilon_1^* \varepsilon_2^*} \neq 0$

$$hsgpa = (\text{factors NOT related to Ability}) + (Ability + \text{error})$$

# Unobserved Confounding and Endogeneity

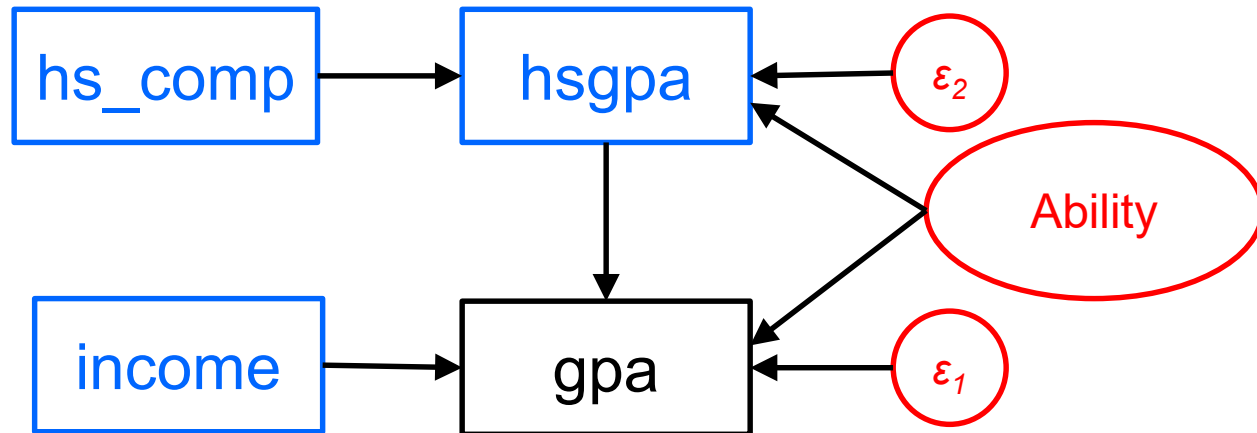
$$gpa = \beta_0 + \beta_1 income + \beta_2 hsgpa + (Ability + \epsilon)$$



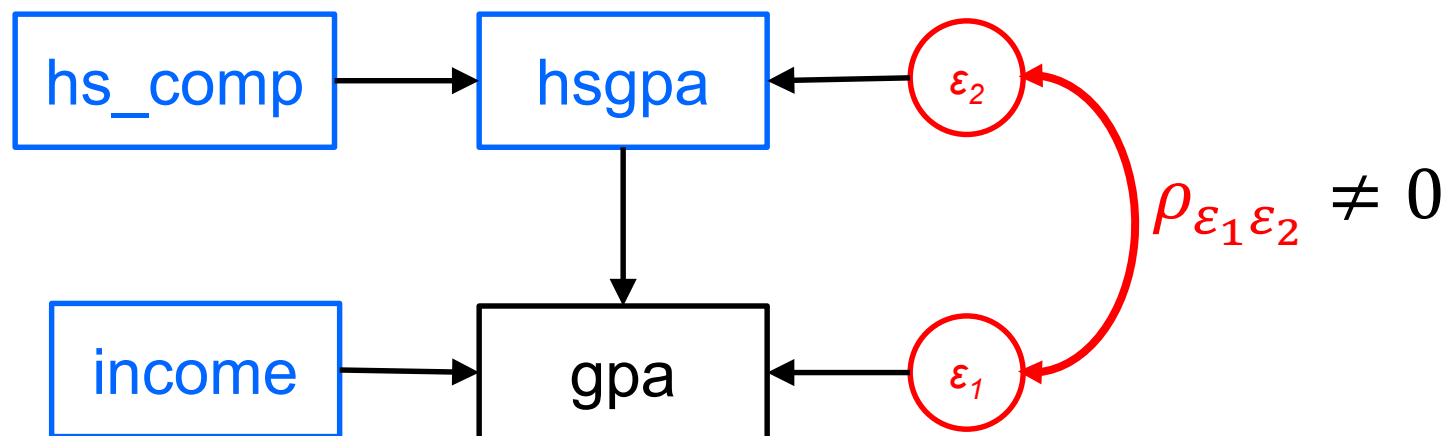
# Unobserved Confounding and Endogeneity

$$hsgpa = \pi_0 + \pi_1 hs\_comp + (Ability + \varepsilon_2)^*$$

$$gpa = \beta_0 + \beta_1 income + \beta_2 (\pi_0 + \pi_1 hs\_comp) + (Ability + \varepsilon_1)^*$$



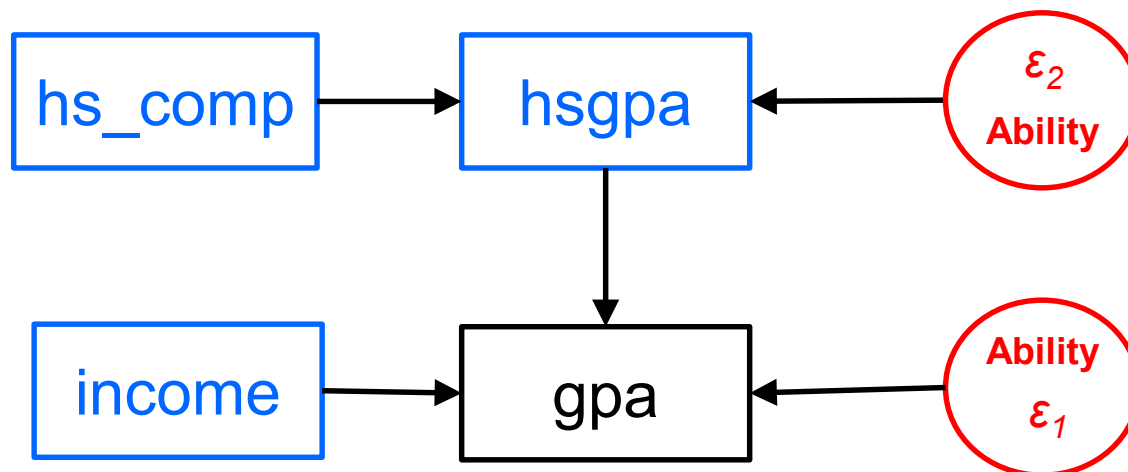
# Unobserved Confounding and Endogeneity



# Unobserved Confounding and Endogeneity

$$hsgpa = \pi_0 + \pi_1 hs\_comp + (Ability + \varepsilon_2)^*$$

$$gpa = \beta_0 + \beta_1 income + \beta_2 (\pi_0 + \pi_1 hs\_comp) + (Ability + \varepsilon_1)^*$$

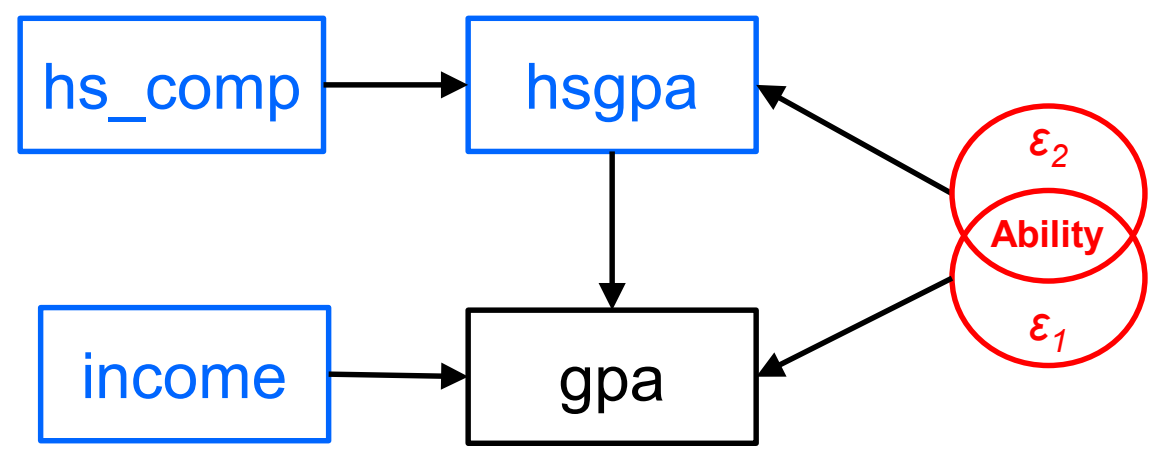


# Unobserved Confounding and Endogeneity

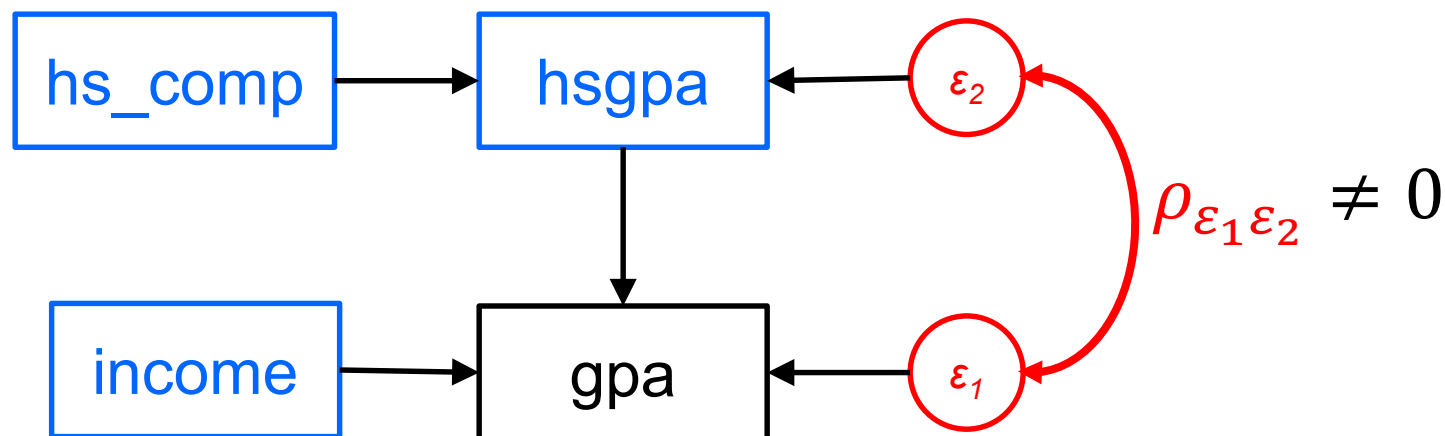
$$hsgpa = \pi_0 + \pi_1 hs\_comp + (Ability + \varepsilon_2)^*$$

$$gpa = \beta_0 + \beta_1 income + \beta_2 (\pi_0 + \pi_1 hs\_comp) + (Ability + \varepsilon_1)^*$$

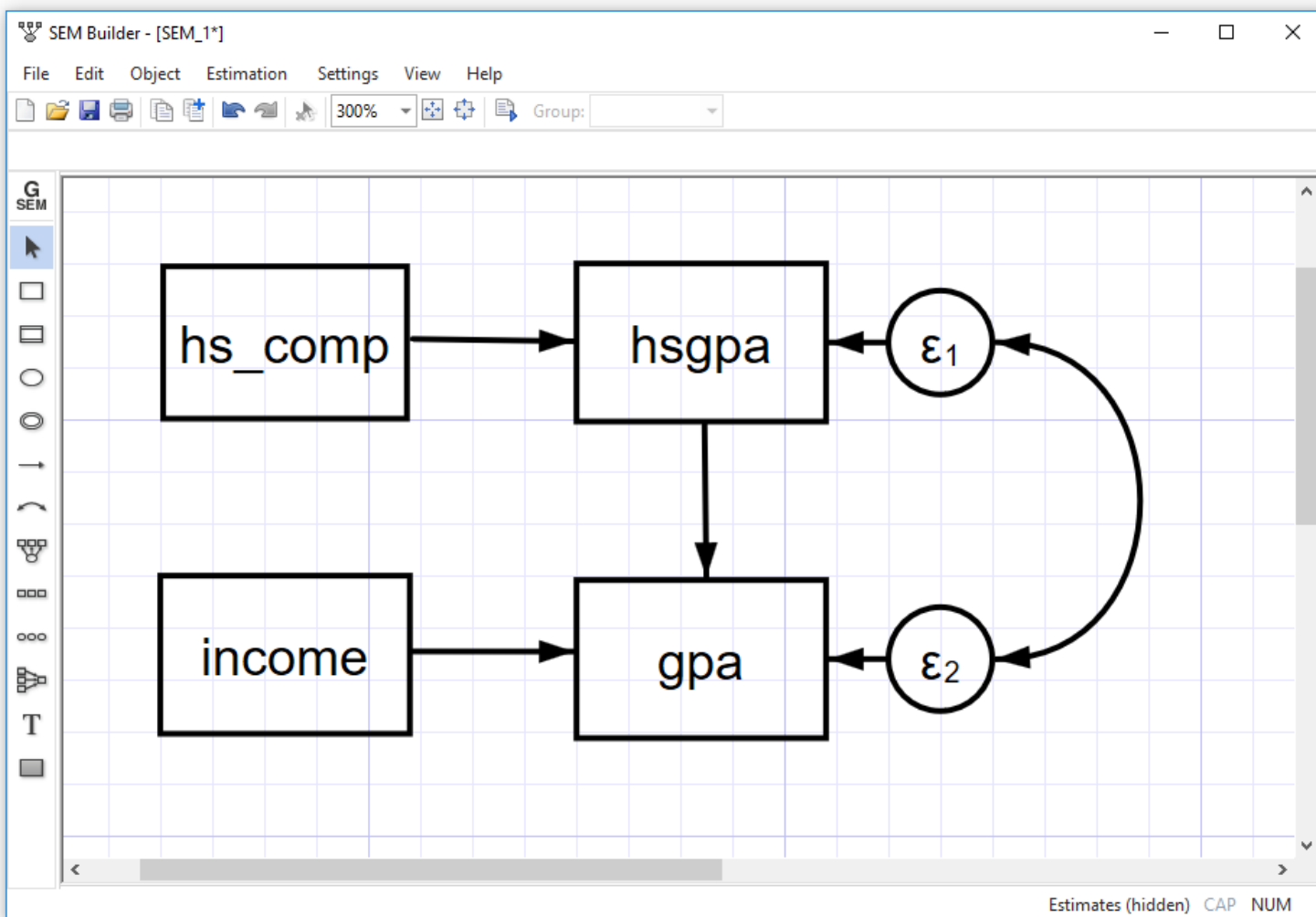
where  $\rho_{\varepsilon_1^* \varepsilon_2^*} \neq 0$



# Unobserved Confounding and Endogeneity

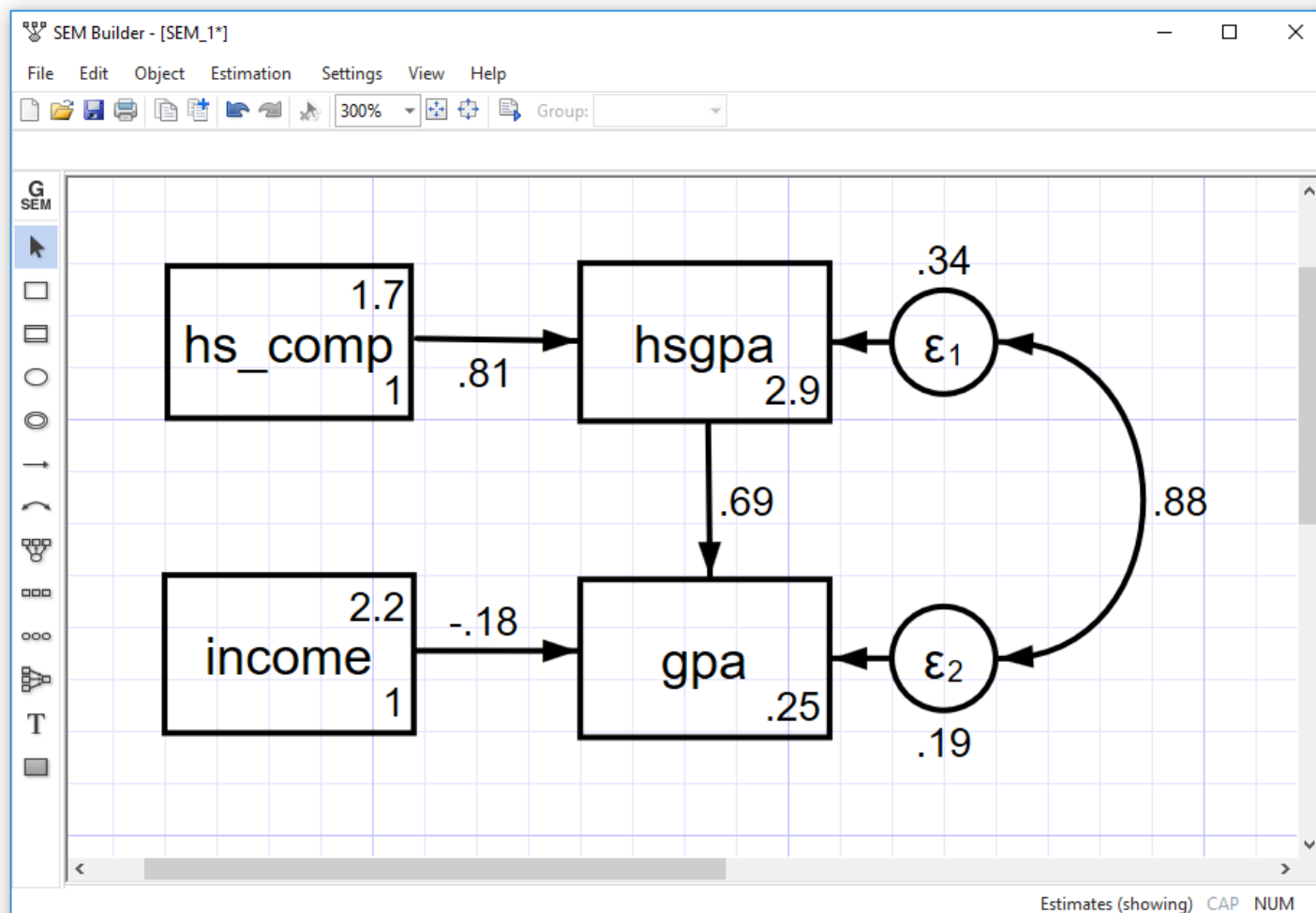


# Unobserved Confounding and Endogeneity





# Unobserved Confounding and Endogeneity



# Unobserved Confounding and Endogeneity

Primary model

```
eregress gpa income, ///  
                endogenous (hsgpa = hs_comp income)
```

Auxillary model

# Unobserved Confounding and Endogeneity

```
. eregress gpa income, ///
> endogenous(hsgpa = hs_comp income) nolog
```

Extended linear regression                      Number of obs        =            792  
Wald chi2(2)                                    =            3951.76  
Log likelihood = 519.11827                      Prob > chi2            =            0.0000

		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
gpa	income	.3702125	.0384969	9.62	0.000	.29476	.4456649
	hsgpa	.9064316	.0193174	46.92	0.000	.8685702	.944293
	_cons	-.2665693	.0397868	-6.70	0.000	-.3445499	-.1885887
hsgpa	hs_comp	1.524814	.0244398	62.39	0.000	1.476913	1.572715
	income	.9898417	.0268549	36.86	0.000	.9372069	1.042476
	_cons	1.072201	.0203047	52.81	0.000	1.032404	1.111997
var(e.gpa)		.0554234	.0030877			.0496902	.061818
var(e.hsgpa)		.0381556	.0019174			.0345767	.0421049
corr(e.hsgpa,e.gpa)		.7503328	.0170348	44.05	0.000	.7149879	.7818521

```
. estimates store endog
```

# Unobserved Confounding and Endogeneity

var (e.gpa)	.0554234	.0030877			.0496902	.061818
var (e.hsgpa)	.0381556	.0019174			.0345767	.0421049
corr (e.hsgpa, e.gpa)	.7503328	.0170348	44.05	0.000	.7149879	.7818521

where  $\rho_{\varepsilon_1^* \varepsilon_2^*} \neq 0$

# Unobserved Confounding and Endogeneity

```
. estimates table univar hsgpa endog, stats(N) equations(1) keep(#1:) b(%9.4f)
```

Variable	univar	hsgpa	endog
program 1	-0.1259	0.2003	
hsgpa		1.1445	0.9064
income			0.3702
_cons	2.1490	-0.6745	-0.2666
N	792	792	792

# Outline

- ✓ • Description of the dataset
- ✓ • Unobserved confounding and endogeneity
  - Nonrandom treatment assignment
  - Missing not at random (MNAR) and selection bias
  - Treatment effects

# Random Treatment Assignment



# Nonrandom Treatment Assignment





# Nonrandom Treatment Assignment

A student's decision to enroll in the study program is based on observed and unobserved factors.

$$P(\text{program} = 1) = \textit{observed factors} + \textit{unobserved factors}$$

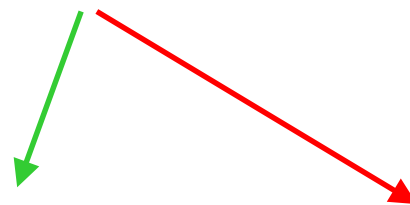
# Unobserved Confounding

$$gpa = \beta_0 + \beta_1 income + \beta_2 program + \varepsilon_{total}$$

$$gpa = \beta_0 + \beta_1 income + \beta_2 program + (Ability + \varepsilon)$$


# Endogenous Treatment

$$gpa = \beta_0 + \beta_1 income + \beta_2(program) + (Ability + \varepsilon_1)$$



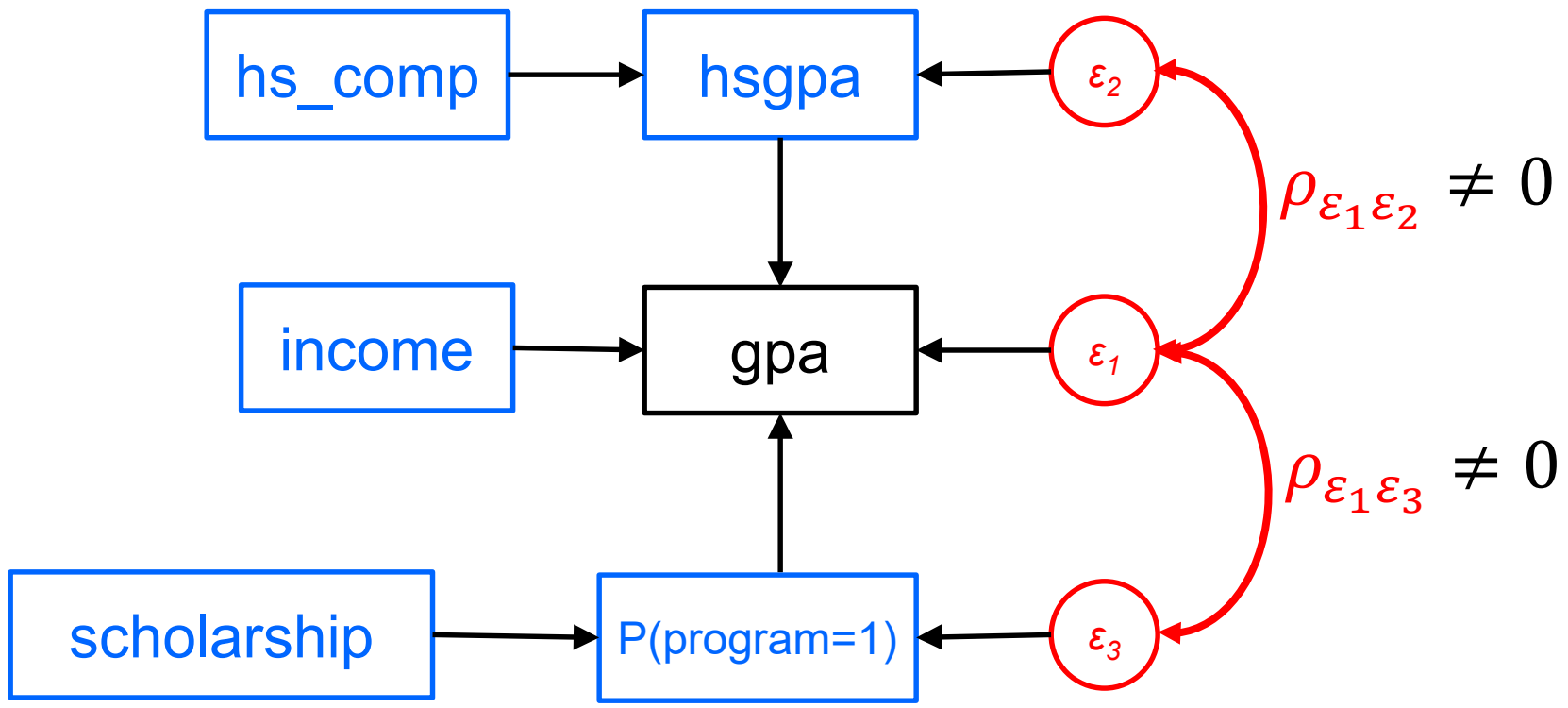
$$P(program = 1) = \pi_0 + \pi_1 scholarship + (Ability + \varepsilon_3)^*$$

$$gpa = \beta_0 + \beta_1 income + \beta_2(\pi_0 + \pi_1 scholarship) + (Ability + \varepsilon_1)^*$$

where  $\rho_{\varepsilon_1^* \varepsilon_3^*} \neq 0$

$$P(program=1) = (\text{factors NOT related to Ability}) + (\text{Ability} + \text{error})$$

# Endogenous Treatment



# Endogenous Treatment

Primary model

```
eregress gpa income,                                     ///  
          endogenous(hsgpa = hs_comp income)           ///  
          entreat(program = income scholarship, nointeract)
```

Auxillary model

# Endogenous Treatment

```
. eregress gpa income,           ///
>   endogenous(hsgpa = hs comp income)  ///
>   entreat(program = income scholarship, nointeract) nolog
```

```
Extended linear regression           Number of obs   =           792
                                     Wald chi2(3)     =           6576.43
Log likelihood = 597.15048           Prob > chi2     =           0.0000
```

		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
gpa	income	.6876358	.0368212	18.67	0.000	.6154676	.7598041
	hsgpa	.9021844	.0150525	59.94	0.000	.872682	.9316869
	program						
	Yes	.3040315	.019458	15.62	0.000	.2658944	.3421685
	_cons	-.5198319	.0352035	-14.77	0.000	-.5888295	-.4508344
program	income	-5.868551	.401813	-14.61	0.000	-6.65609	-5.081012
	scholarship	1.814856	.1636187	11.09	0.000	1.49417	2.135543
	_cons	1.503659	.1629857	9.23	0.000	1.184213	1.823106
hsgpa	hs_comp	1.528458	.0236304	64.68	0.000	1.482144	1.574773
	income	.989619	.0268526	36.85	0.000	.9369889	1.042249
	_cons	1.070543	.0201056	53.25	0.000	1.031136	1.109949
	var(e.gpa)	.0358984	.0020787			.0320469	.0402127
	var(e.hsgpa)	.0381566	.0019175			.0345776	.0421062
	corr(e.program,e.gpa)	.4511304	.0772058	5.84	0.000	.2877691	.5889813
	corr(e.hsgpa,e.gpa)	.8093104	.0134908	59.99	0.000	.7811792	.8341618
	corr(e.hsgpa,e.program)	.480631	.0565509	8.50	0.000	.3624217	.5836218

```
. estimates store entreat
```

# Endogenous Treatment

var (e.gpa)	.0358984	.0020787			.0320469	.0402127
var (e.hsgpa)	.0381566	.0019175			.0345776	.0421062
corr (e.program, e.gpa)	.4511304	.0772058	5.84	0.000	.2877691	.5889813
corr (e.hsgpa, e.gpa)	.8093104	.0134908	59.99	0.000	.7811792	.8341618
corr (e.hsgpa, e.program)	.480631	.0565509	8.50	0.000	.3624217	.5836218

where  $\rho_{\epsilon_1^* \epsilon_3^*} \neq 0$

# Endogenous Treatment

```
. estimates table univar hsgpa endog entreat, stats(N) equations(1) keep(#1:) b(%9.4f)
```

Variable	univar	hsgpa	endog	entreat
program Yes	-0.1259	0.2003		0.3040
hsgpa		1.1445	0.9064	0.9022
income			0.3702	0.6876
_cons	2.1490	-0.6745	-0.2666	-0.5198
N	792	792	792	792



# Outline

- ✓ • Description of the dataset
- ✓ • Unobserved confounding and endogeneity
- ✓ • Nonrandom treatment assignment
  - Missing not at random (MNAR) and selection bias
  - Treatment effects

# No Missingness



# Missing Completely at Random (MCAR)



# Missing at Random (MAR)



# Missing Not at Random (MNAR)



# MNAR and Selection Bias

```
. tab graduate
```

Did the student graduate college?	Freq.	Percent	Cum.
No	208	20.80	20.80
Yes	792	79.20	100.00
Total	1,000	100.00	

# Endogenous Sample Selection

A student's decision to drop out of school is based on observed and unobserved factors.

$$P(\textit{graduate} = 1) = \textit{observed factors} + \textit{unobserved factors}$$

# Endogenous Sample Selection

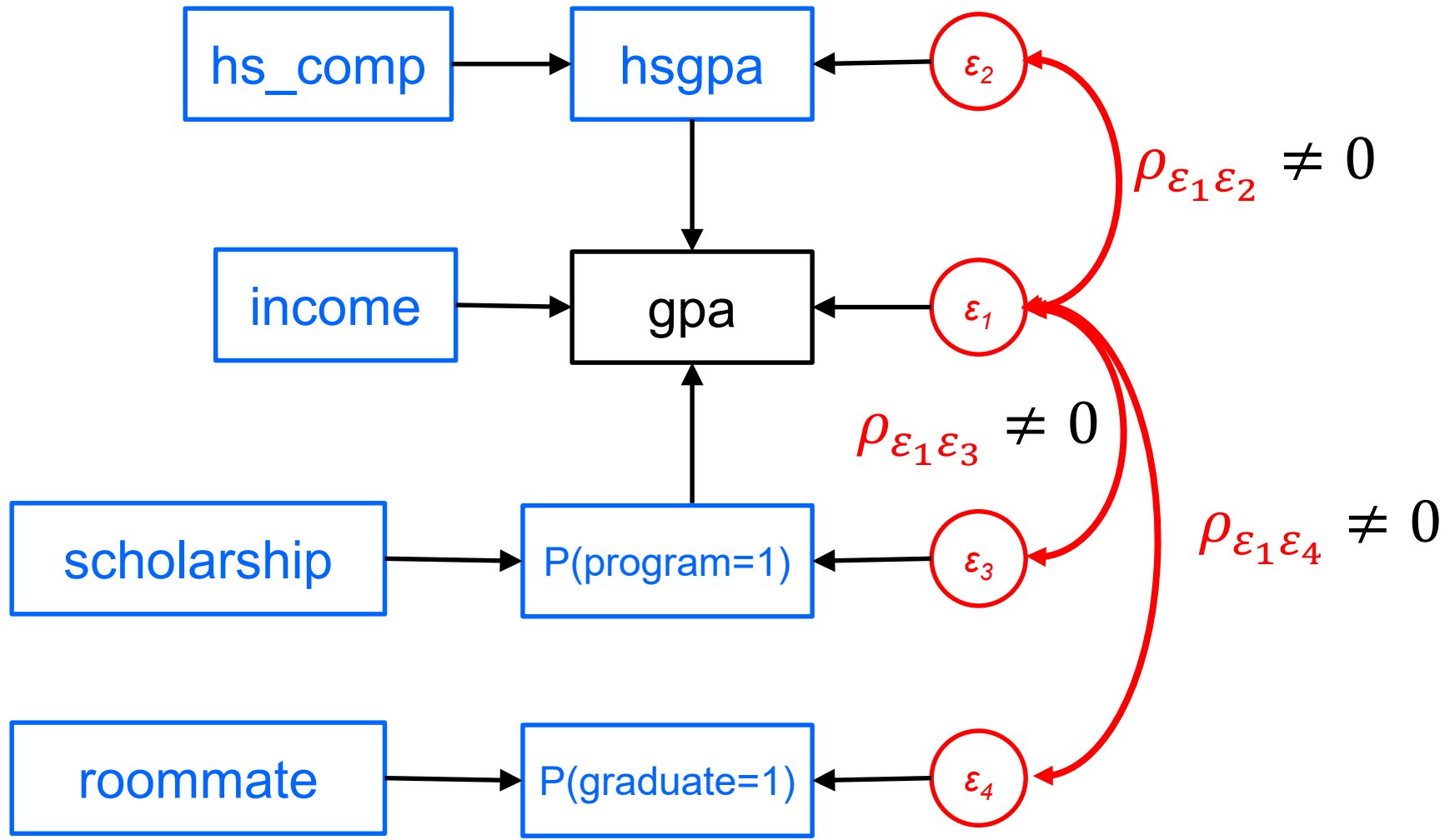
$$gpa = \begin{cases} \beta_0 + \beta_1 income + \beta_2(program) + (Ability + \varepsilon_1)* & \text{if graduate}=1 \\ \text{missing} & \text{if graduate}=0 \end{cases}$$

$$P(\text{graduate} = 1) = \pi_0 + \pi_1 roommate + (Ability + \varepsilon_4)*$$

where  $\rho_{\varepsilon_1^* \varepsilon_4^*} \neq 0$



# Endogenous Sample Selection



# Endogenous Sample Selection

Primary model

```
eregress gpa income,                                     ///  
    endogenous(hsgpa = hs_comp income)                 ///  
    entreat(program = income scholarship, nointeract)  ///  
    select(graduate = income roommate)
```

Auxillary model

# Endogenous Sample Selection

```
. eregress gpa income,                               ///
>   endogenous(hsgpa = hs_comp income)              ///
>   entreat(program = income scholarship, nointeract) ///
>   select(graduate = income roommate) nolog
```

```
Extended linear regression           Number of obs   =   1,000
                                     Selected          =     792
                                     Nonselected       =     208

Log likelihood = 323.23691           Wald chi2(3)    =   8866.38
                                     Prob > chi2      =     0.0000
```

		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
gpa							
	income	.8220509	.0333135	24.68	0.000	.7567576	.8873443
	hsgpa	.8935782	.0136619	65.41	0.000	.8668013	.9203551
	program						
	Yes	.2976643	.0168041	17.71	0.000	.2647288	.3305998
	_cons	-.6071633	.029947	-20.27	0.000	-.6658583	-.5484682
graduate							
	income	4.010154	.2557017	15.68	0.000	3.508988	4.51132
	roommate	1.412072	.1320596	10.69	0.000	1.15324	1.670904
	_cons	-1.053694	.1059937	-9.94	0.000	-1.261438	-.8459504
program							
	income	-4.889741	.2935974	-16.65	0.000	-5.465181	-4.3143
	scholarship	1.791084	.1291875	13.86	0.000	1.537881	2.044287
	_cons	.8297874	.1047466	7.92	0.000	.6244878	1.035087
hsgpa							
	hs_comp	1.512085	.0202588	74.64	0.000	1.472378	1.551791
	income	1.0879	.0221946	49.02	0.000	1.044399	1.1314
	_cons	.9990863	.0161398	61.90	0.000	.9674529	1.03072
	var (e.gpa)	.040487	.0023354			.0361589	.0453332
	var (e.hsgpa)	.0399236	.0017858			.0365726	.0435817
	corr (e.graduate, e.gpa)	.7609452	.0402982	18.88	0.000	.6700487	.8293596
	corr (e.program, e.gpa)	.5402021	.0577087	9.36	0.000	.4175545	.6435181
	corr (e.hsgpa, e.gpa)	.8221551	.0119073	69.05	0.000	.797394	.8441524
	corr (e.program, e.graduate)	.85115	.0432119	19.70	0.000	.7411121	.9166561
	corr (e.hsgpa, e.graduate)	.5633432	.0408602	13.79	0.000	.4780104	.6381415
	corr (e.hsgpa, e.program)	.5265467	.0436265	12.07	0.000	.435811	.6066872

```
. estimates store endsel
```

# Endogenous Sample Selection

var (e.gpa)	.040487	.0023354			.0361589	.0453332
var (e.hsgpa)	.0399236	.0017858			.0365726	.0435817
corr (e.graduate, e.gpa)	.7609452	.0402982	18.88	0.000	.6700487	.8293596
corr (e.program, e.gpa)	.5402021	.0577087	9.36	0.000	.4175545	.6435181
corr (e.hsgpa, e.gpa)	.8221551	.0119073	69.05	0.000	.797394	.8441524
corr (e.program, e.graduate)	.85115	.0432119	19.70	0.000	.7411121	.9166561
corr (e.hsgpa, e.graduate)	.5633432	.0408602	13.79	0.000	.4780104	.6381415
corr (e.hsgpa, e.program)	.5265467	.0436265	12.07	0.000	.435811	.6066872

# Endogenous Sample Selection

```
. estimates table univar hsgpa endog entreat endsel, stats(N) equations(1) keep(#1:) b(%9.4f)
```

Variable	univar	hsgpa	endog	entreat	endsel
program Yes	-0.1259	0.2003		0.3040	0.2977
hsgpa		1.1445	0.9064	0.9022	0.8936
income			0.3702	0.6876	0.8221
_cons	2.1490	-0.6745	-0.2666	-0.5198	-0.6072
N	792	792	792	792	1000

True Model (simulated)

$$\text{gpa} = -0.6 + 0.3 \cdot \text{treatment} + 0.9 \cdot \text{hsgpa} + 0.8 \cdot \text{income}$$

# Outline

- ✓ • Description of the dataset
- ✓ • Unobserved confounding and endogeneity
- ✓ • Nonrandom treatment assignment
- ✓ • Missing not at random (MNAR) and selection bias
- Treatment effects

# ERM Postestimation

- `estat teffects`
- `margins`
- `marginsplot`
- `predict`

# estat teffects

```
. estat teffects
```

```
Predictive margins          Number of obs      =      1,000
Model VCE      : OIM
```

	Delta-method				
	Margin	Std. Err.	z	P> z	[95% Conf. Interval]
ATE					
program	.2976643	.0168041	17.71	0.000	.2647288 .3305998
(Yes vs No)					

Note: Standard errors treat sample covariate values as fixed and not a draw from the population. If your interest is in population rather than sample effects, refit your model using **vce(robust)**.



# estat teffects, atet

```
. estat teffects, atet
```

Predictive margins

Number of obs = 1,000

Subpop. no. obs = 300

Model VCE : OIM

	Delta-method Margin	Std. Err.	z	P> z	[95% Conf. Interval]	
ATET program (Yes vs No)	.2976643	.0168041	17.71	0.000	.2647288	.3305998

# margins

```
. margins i.program, at(hsgpa=(1.5(0.5)4)) predict(fix(hsgpa program)) vsquish
```

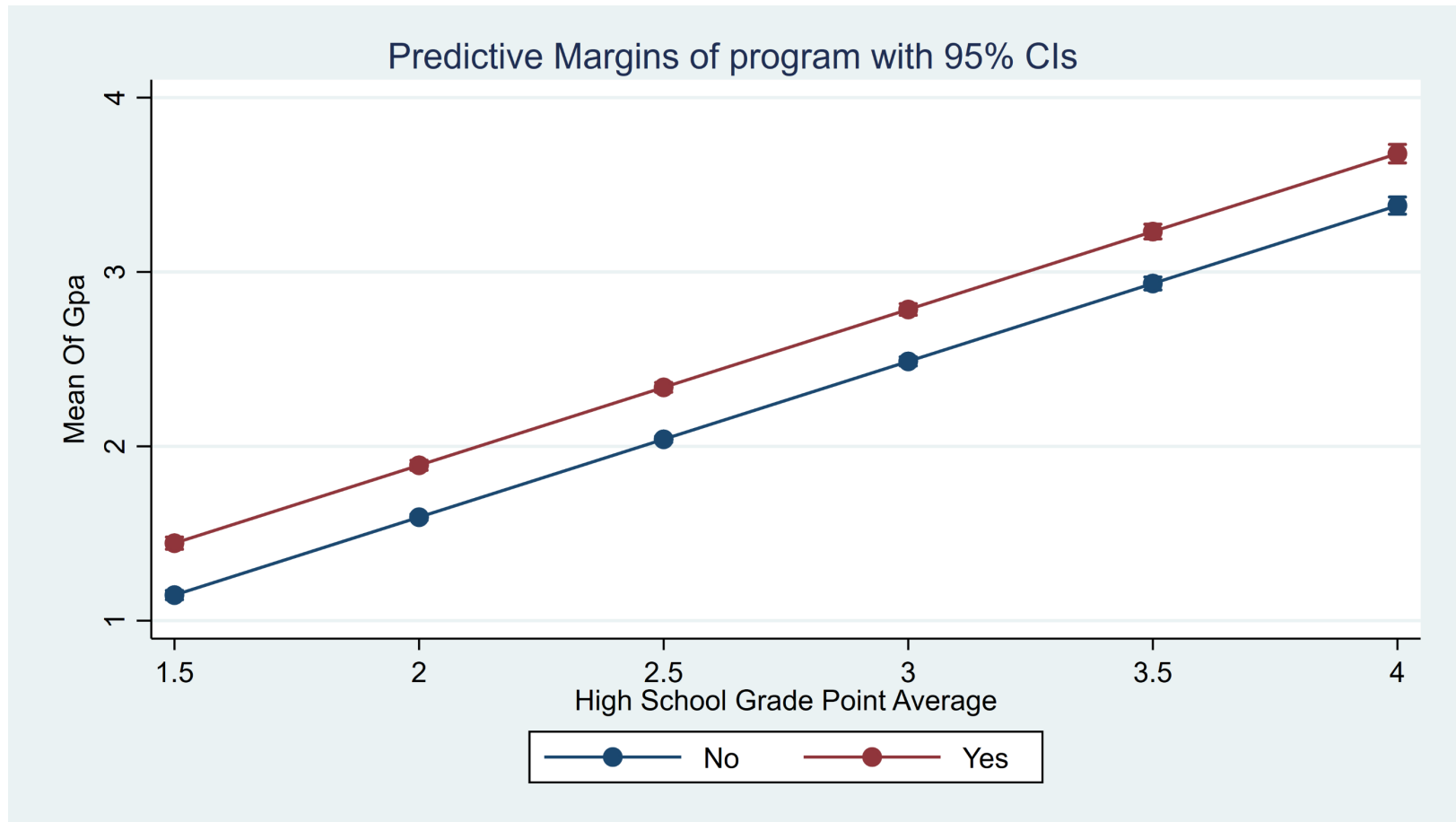
```
Predictive margins                                Number of obs    =           1,000  
Model VCE      : OIM
```

```
Expression   : mean of gpa, predict(fix(hsgpa program))  
1._at        : hsgpa                =           1.5  
2._at        : hsgpa                =           2  
3._at        : hsgpa                =           2.5  
4._at        : hsgpa                =           3  
5._at        : hsgpa                =           3.5  
6._at        : hsgpa                =           4
```

	Margin	Delta-method Std. Err.	z	P> z	[95% Conf. Interval]	
__at#program						
1#No	1.146849	.0134979	84.96	0.000	1.120394	1.173305
1#Yes	1.444513	.0178228	81.05	0.000	1.409581	1.479446
2#No	1.593638	.0092729	171.86	0.000	1.575464	1.611813
2#Yes	1.891303	.0146207	129.36	0.000	1.862646	1.919959
3#No	2.040427	.0091161	223.83	0.000	2.02256	2.058294
3#Yes	2.338092	.0142549	164.02	0.000	2.310153	2.366031
4#No	2.487216	.0131736	188.80	0.000	2.461397	2.513036
4#Yes	2.784881	.0169105	164.68	0.000	2.751737	2.818025
5#No	2.934006	.0189026	155.22	0.000	2.896957	2.971054
5#Yes	3.23167	.0214953	150.34	0.000	3.18954	3.2738
6#No	3.380795	.0251872	134.23	0.000	3.331429	3.430161
6#Yes	3.678459	.0270455	136.01	0.000	3.625451	3.731467

# marginsplot

```
. marginsplot
```



# More ERM's

- **eregress** – continuous outcomes
- **eintreg** – interval outcomes
- **eprobit** – binary outcomes
- **eoprobit** – ordinal outcomes

# ERMs For Panel Data

- **xteregress** – continuous outcomes
- **xteintreg** – interval outcomes
- **xteprobit** – binary outcomes
- **xteoprobit** – ordinal outcomes

# More About ERM

- ERM can include:
  - polynomials of endogenous covariates
  - interactions of endogenous covariates
  - interactions of endogenous with exogenous covariates

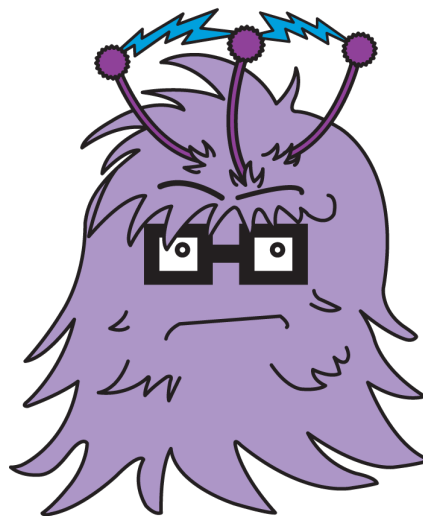
# Cautionary Note

- Nothing about ERMs magically extracts causal relationships.
- As with any regression analysis of observational data, the causal interpretation must be based on a reasonable underlying scientific rationale.

# Thanks for coming!

## Questions?

[chuber@stata.com](mailto:chuber@stata.com)



You can download the slides, datasets, and do-files here:

<https://tinyurl.com/2019CausalInference>