Elliot Tu

1a. Omitting IQ, slope coefficient of B_1* is .0583.

1b. Controlling for IQ, slope coefficient for B_1 is .0389, slope coefficient of B_2 is .005.

1c. Regressing IQ on education, slope coefficient for δ_1 is 3.572.

(For 1a-1c, see **Table 1** below)

1d. Mathematically show the bias affecting education (B_1) caused by omitting IQ (B_2)

$B_1^* = B_1 + B_2\ \delta_1$

$.0583 = .0389 + (.00543)(3.572)$

$.0583 = .0389 + .0194$

$.0583 = .583$

Table 1, Effects of Education and IQ on Wage

VARIABLES	IQ Omitted	IQ Controlled	IQ on Education
educ	0.058***	0.039***	3.572***
	(0.006)	(0.007)	(0.195)
IQ		0.005***	
		(0.001)	
Constant	5.997***	5.708***	53.220***
	(0.083)	(0.098)	(2.672)
Observations	890	890	890
R-squared	0.094	0.122	0.273

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

2a.

Variable	Frequency
1	16
2	65
3	93
4	188
5	228
Total	590

2b. Marriage Happiness = 3.369 - .0484(years married) - .131(male) + .0635(education) + u

Using the OLS is not a good estimator in this scenario because it can produce values outside our actual results (1-5). The only upside is that the OLS gives a rough estimation of what we can expect based on correlation of the variables and the marriage happiness. (See **Table 2** on next page)

2c.There are no differences in the statistical significance of coefficients between both of the two regressions in 2b and 2c. Years married and education are both significant at the 99% level, while gender is not significant at the 90% level or above. One notable difference is in the OLS regression there is a statistically significant constant that isn't present in the ordered probit model. (See **Table 2** below)

Table 2, Effects on the Quality of Marriage

VARIABLES	(1) Standard Regression	(2) Ordered Probit Model
yrsmarr	-0.0484***	-0.0511***
	(0.00796)	(0.00818)
male	-0.131	-0.152
	(0.0951)	(0.0978)
educ	0.0635***	0.0589***
	(0.0214)	(0.0202)
/cut1		-1.563***
		(0.330)
/cut2		-0.679**
		(0.321)
/cut3		-0.103
		(0.321)
/cut4		0.761**
		(0.321)
Constant	3.360***	
	(0.326)	
Observations	590	590
R-squared	0.074	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

2d. For someone in the highest happiness group, one more year of marriage will decrease the probability of being in the highest happiness group by .019 percentage points., one more year of education increases it by .022 percentage points, and being male decreases it by .058 percentage points.

```
Marginal effects after oprobit
      y  = Pr(ratemarr==5) (predict, outcome(5))
         =  .38160049
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
yrsmarr	-.0194714	.00313	-6.22	0.000	-.025603	-.01334	8.1933
male*	-.0579884	.03711	-1.56	0.118	-.130728	.014751	.472881
educ	.0224526	.00771	2.91	0.004	.007346	.037559	16.1407

(*) dy/dx is for discrete change of dummy variable from 0 to 1

3a. House Price = 167591 – 494.5(bedrooms) + .721(lot size) + 45.13(square footage) + u
The coefficients significant at the 95% level or above are square footage and the constant, both of which are significant at the 99% level. (See **Table 3** on next page)

3b. After tabulating the data based on each variable, it appears that there is some censorship. Tabulating lot size and square footage reveals that there are different variables for each house, which we'd expect. Tabulating for bedrooms indicate a concentration of houses have either 2 or 3 bedrooms, which is not indicative of censorship because we'd expect that houses most commonly have from 2-3 bedrooms. However, tabulating price shows that the data for price is capped at 310,000, as almost 40% of the houses show values of exactly 310,000. This likely means the coefficients are downward biased because the price value can't be higher than 310,000, when it probably is several times.

3c. Tobit model. Lot size and square footage are now statistically significant at the 99% level.

Table 3, House Characteristics and Sale Price

VARIABLES	Standard Regression	Tobit Model	Tobit Model
bdrms	-494.5	7,479	
	(5,547)	(8,682)	
lotsize	0.721*	6.102***	
	(0.387)	(2.222)	
sqrft	45.13***	108.6***	
	(8.027)	(18.40)	
var(e.price)			1.829e+09***
			(3.751e+08)
Constant	167,591***	-2,074	
	(18,260)	(42,430)	
Observations	84	84	84
R-squared	0.394		

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

3d. Comparing the differences in the regressions, it appears that the capping of price at $310,000 caused a lot of variance leading to weird values such as an additional bedroom dropping the price of a house by $450 as estimated by the standard regression, and a high constant that would put a house with no bedrooms/square footage/lot size at $167,591.

(For 4a-4f see **Table 4** on next page)

4a. Rooms and area are both statistically significant coefficients at the 95% and 99% levels respectively. An additional room will increase house price by an estimated $5467, while an additional square foot increases the house price by an estimated $34.35.

4b. Adjusting for fixed effects reduced the magnitude of rooms by almost 20% and area a little less. Additionally, rooms standard error increased and is now only significant at the 90% level, while several of the fixed effects of the neighborhoods are almost all statistically significant at the 95% level or higher.

4c. It's likely that the variables we're using (rooms, square footage, and lot size) are correlated with different neighborhoods. Neighborhoods in less affluent suburban/rural areas would likely have smaller houses with less rooms and land size, and neighborhoods in affluent urban areas would likely have higher prices with less land size. Regardless, there is definitely possibility for correlation between the variables and the fixed effects, causing variance that we need to account for.

4d. The neighborhood with the least negative value would be the most expensive, as it drops the house price the least. The order would be 1, 2, 6, 3, 5, and 4 which has the lowest average house price by a large margin.

4e. The inclusion of neighborhood fixed effects lowered all the variable coefficient estimates, with a 20% drop in the magnitude of rooms to a smaller ~9% drop in lot size and square footage. There was also a significant change in the coefficient which went from -15,212 to a positive value of 7,492. This indicates that the constant was absorbing a lot of the effect, causing what was initially a downward-biased estimate.

4f. The house characteristics rooms, land, and area are all now statistically significant at the 95% level. The magnitude of rooms and land have increased while area has decreased. All of their standard errors have also decreased. Additionally, with the exception of neighborhood 3 which is significant at the 95% level, all neighborhood fixed effects are significant at the 99% level, and half of the by-year effects are statistically significant at the 95% level as well. The magnitude of these year effects are all positive, which makes sense, as we would assume house prices tend to increase over time.

Table 4, Effects on House Sale Price

VARIABLES	Standard Regression	Neighborhood Fixed Effect	Year Fixed Effect
rooms	5,467**	4,335*	5,914**
	(2,588)	(2,616)	(2,492)
land	0.0933	0.0831	0.112
	(0.0886)	(0.101)	(0.102)
area	34.35***	32.52***	28.85***
	(4.649)	(4.439)	(4.527)
_Inbh_1		-12,178	-28,380***
		(9,995)	(8,987)
_Inbh_2		-14,127**	-29,054***
		(5,542)	(4,812)
_Inbh_3		-15,395*	-31,317***
		(8,132)	(7,238)
_Inbh_4		-25,472***	-31,913***
		(5,481)	(6,533)
_Inbh_5		-16,124***	-25,629***
		(5,509)	(5,521)
_Inbh_6		-14,576**	-27,015***
		(6,516)	(5,338)
_Inbhyear_1			34,718**
			(15,643)
_Inbhyear_2			40,624***
			(5,479)
_Inbhyear_3			33,776***
			(7,285)
_Inbhyear_4			15,861**
			(6,434)
_Inbhyear_5			28,668***
			(5,615)
_Inbhyear_6			26,771***
			(8,880)
Constant	-15,212	7,492	4,031
	(12,681)	(14,964)	(14,320)
Observations	321	321	321
R-squared	0.409	0.455	0.520

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1