1.

. sum yield					
Variable	Obs	Mean	Std. dev.	Min	Max
yield	14,171	433.6169	395.1982	0	36461.98
. sum rental_in					
Variable	Obs	Mean	Std. dev.	Min	Max
rental_in	14,171	.5031543	4.466769	0	200
. sum rental_ou	t				
Variable	Obs	Mean	Std. dev.	Min	Max
rental_out	14,171	.2164561	1.489389	0	60

- (1) The standard deviation of variable yield is large, which means the output of each households' unit of land is very different.
- (2) On average, land rented in is more than land rented out by the households.
- (3) The difference in land rented in among each household is greater than the difference in land rented out.
- (4) At least one household does not rent in land and at least one household does not rent out land.
- (5) The household who rents in most rents in 200 units of land and the household who rents out most rents out 60 units of land.
- 2. I think  $\beta_1$  will be negative and  $\beta_2$  will be positive.

Because when a household has more land, every unit of land will possibly get less attended and less resources (water, fertilizer, etc.). Thus, the output per unit of land will decrease. However, when a household has less land, every unit of land will possibly get more attended and more resources. Thus, the output per unit of land will increase.

3. Yes, the estimated values agree with my prediction.

Source	SS	df	MS		er of obs	=	14,171
Model	172005.965	1	172005.965		14169) > F	=	0.2940
Residual	2.2129e+09	14,169	156180.516		uared	=	0.0001
RESIGNAL	2121250105	14,105	130100.310		R-squared	=	0.0000
Total	2.2131e+09	14,170	156181.633		MSE	=	395.2
yield	Coefficient	Std. err.	t	P> t	[95% cor	nf.	interval]
rental in	7799978	.7432502	-1.05	0.294	-2.236866	5	.6768703
cons	434.0094	3.340807	129.91	0.000	427.461	1	440.5578
/sum res_y	es_yield_1, re ield_1	siduals					
/predict re	es_yield_1, re ield_1	siduals					
/predict re/ /sum res_y	es_yield_1, re ield_1	siduals	MS		per of obs	=	
/predict ro /sum res_y: reg yield ro Source	es_yield_1, re ield_1 ental_out	df		F(1,	14169)	=	0.09
/predict ro /sum res_y: reg yield ro Source Model	es_yield_1, re ield_1  ental_out  SS  13342.3767	df 1	13342.3767	F(1, Prob	14169) > F		0.09 0.7701
/predict ro /sum res_y: reg yield ro Source	es_yield_1, re ield_1 ental_out	df		F(1, Prob R-sq	14169)	=	0.09 0.7701 0.0000
/predict ro /sum res_y: reg yield ro Source Model	es_yield_1, re ield_1  ental_out  SS  13342.3767	df 1	13342.3767	F(1, Prob R-sq Adj	14169) > F uared	=	0.09 0.7701 0.0000 -0.0001
/predict re /sum res_y: reg yield re Source Model Residual	es_yield_1, re ield_1 ental_out SS 13342.3767 2.2131e+09	df 1 14,169	13342.3767 156191.714 156181.633	F(1, Prob R-sq Adj	14169) > F quared R-squared MSE	= = = =	0.09 0.7701 0.0000 -0.0001 395.21
/predict rows / free yield rows / Source   Model   Residual   Total	es_yield_1, re ield_1  ental_out  SS  13342.3767 2.2131e+09  2.2131e+09	df 1 14,169 14,170	13342.3767 156191.714 156181.633	F(1, Prob R-sq Adj Root	14169) > F quared R-squared MSE	= = = = =	14,171 0.09 0.7701 0.0000 -0.0001 395.21 interval]

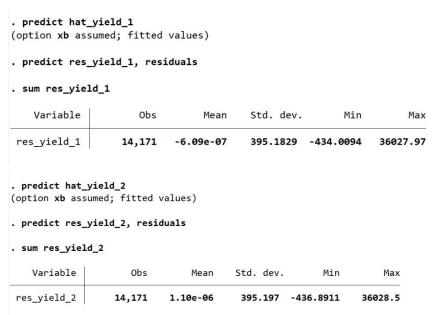
### 4.

 $\beta_1$ -hat: Holding ceteris paribus, when a household rents in one more unit of land, the output per unit of its land will decrease by about 0.7799978 units.

 $\beta_2$ -hat: Holding ceteris paribus, when a household rents out one more unit of land, the output per unit of its land will increase by about 0.6515137 units.

**However**, these relationships are not statistically significant, because their p-values are obviously greater than 0.05. So we may fail to reject the null hypothesis.

5.



We can see that the residuals approximately sum to 0.

### 6.

We can see in sub-question 3 that the  $R^2$  of rental\_in is 0.0001 (about 0.01% is explained) and the  $R^2$  of rental\_out is 0.0000 (about 0.00% is explained).

Source	SS	df		MS	Number	- 1/2	=	14,171
Model	111001 501	-	1110	01 604	F(1, 14		=	0.71
	111081.624	1		81.624	Prob >		=	0.3991
Residual	2.2130e+09	14,169	1561	.84.816	R-squar		=	0.0001
1					Adj R-s			-0.0000
Total	2.2131e+09	14,170	1561	.81.633	Root MS	Ł	=	395.2
yiel	d Coefficient	t Std. e	err.	t	P> t	[95%	conf.	interval]
ental_in_share	.0540097	.06404	127	0.84	0.399	071	5225	.1795418
con								
_cons		3.3327	743	130.03	0.000	426.	8374	439.9026
		3.3327 df	743	<b>130.03</b> MS	Number	of obs	=	14,171
reg yield re	ntal_out_share	df		MS	Number F(1, 14	of obs 169)	=	14,171 24.04
reg yield res	ss 3747795.36	df 1	3747	MS 2795.36	Number F(1, 14 Prob >	of obs 169) F	= =	14,171 24.04 0.0000
reg yield re	ntal_out_share	df	3747	MS	Number F(1, 14 Prob > R-squar	of obs 169) F ed	= = = =	14,171 24.04 0.0000 0.0017
reg yield ren Source  Model Residual	SS 3747795.36 2.2093e+09	df 1 14,169	3747 1559	MS 7795.36 928.149	Number F(1, 14 Prob > R-squar Adj R-s	of obs 169) F ed quared	= = = =	14,171 24.04 0.0000 0.0017 0.0016
reg yield re	ss 3747795.36	df 1	3747 1559	MS 2795.36	Number F(1, 14 Prob > R-squar	of obs 169) F ed quared	= = = =	14,171 24.04 0.0000 0.0017
Source  Model Residual	35 3747795.36 2.2093e+09 2.2131e+09	df 1 14,169 14,170	3747 1559 1561	MS 7795.36 928.149	Number F(1, 14 Prob > R-squar Adj R-s	of obs 169) F ed quared E	= = = =	14,171 24.04 0.0000 0.0017 0.0016
Source Model Residual	SS 3747795.36 2.2093e+09 2.2131e+09	df 1 14,169 14,170 nt Std.	3747 1559 1561 err.	MS 2795.36 228.149 .81.633	Number F(1, 14 Prob > R-squar Adj R-s Root MS	of obs 169) F ed quared E	= = = =	14,171 24.04 0.0000 0.0017 0.0016 394.88

 $\beta_1$ -hat: Holding ceteris paribus, when the proportion of rental\_in to total land area increases by 1% (rental\_in\_share increases by 1 unit), the output per unit of land will increase by about 0.0540097 units.

 $\beta_2$ -hat: Holding ceteris paribus, when the proportion of rental\_out to total land area increases by 1% (rental\_out\_share increases by 1 unit), the output per unit of land will increase by about 0.2350118 units

**However**, the relationship between yield and **rental\_in\_share** may not be statistically significant, because its p-value is obviously greater than 0.05. So **we may fail to reject the null hypothesis**. 8.

reg ln_yield	rental_in_shar	-						
Source	SS	df		MS	Number of	obs	=	14,171
					F(1, 1416	9)	=	1.94
Model	.689144057	1	.689	144057	Prob > F		=	0.1642
Residual	5045.22149	14,169	.356	5074634	R-squared	ı	=	0.0001
				- 0	Adj R-squ	ared	=	0.0001
Total	5045.91063	14,170	.356	5098139	Root MSE		=	.59672
ln_yield	Coefficient	Std. e	err.	t	P> t	[95%	conf.	interval]
ental in share	.0001345	.00009			0.164	000	9055	.0003241
enrar in share			16/	1.39	0.104	000		
_cons	5.940788	.00503	1000	1.39 L180.57	0.000	5.936		5.950651
_cons	227.23.24.40.20.22.22.23.23	.00503	1000		0.000 Number of	5.936 obs	9924	14,171
_cons reg ln_yield Source	5.940788 rental_out_sha	.00503 re	321 1	MS	0.000 Number of F(1, 1416	5.936 obs	= =	14,171 19.59
_cons reg ln_yield Source	5.940788  rental_out_sha  SS  6.96783523	.00503 re	6.96	MS 5783523	Number of F(1, 1416 Prob > F	5.936 obs	= = =	14,171 19.59 0.0000
_cons reg ln_yield Source	5.940788 rental_out_sha	.00503 re	6.96	MS	Number of F(1, 1416 Prob > F R-squared	5.936 obs	= = = =	14,171 19.59 0.0000 0.0014
_cons reg ln_yield Source	5.940788  rental_out_sha  SS  6.96783523	.00503 re	6.96	MS 5783523	Number of F(1, 1416 Prob > F	5.936 obs	= = =	14,171 19.59 0.0000
cons  reg ln_yield  Source  Model Residual	5.940788  rental_out_shalloss  SS  6.96783523 5038.94279 5045.91063	.00503 re df 1 14,169 14,170	6.96 .355	MS 5783523 5631505	Number of F(1, 1416 Prob > F R-squared Adj R-squ	5.936 obs 69)	= = = = = =	14,171 19.59 0.0000 0.0014 0.0013
cons  reg ln_yield  Source  Model Residual  Total	5.940788  rental_out_sha  SS  6.96783523 5038.94279  5045.91063  d Coefficient	.00503 re df 1 14,169 14,170	6.96 .355 .356	MS 5783523 5631505 5098139	Number of F(1, 1416 Prob > F R-squared Adj R-squ Root MSE	5.936 obs (99) I Jared	= = = = = =	14,171 19.59 0.0000 0.0014 0.0013 .59635

 $\beta_1$ -hat: Holding ceteris paribus, when the proportion of rental\_in to total land area increases by 1%,

that is, rental\_in\_share increases by 1 unit, the output per unit of land will increase by about 5.940788%.

 $\beta_2$ -hat: Holding ceteris paribus, when the proportion of rental\_out to total land area increases by 1%, that is, rental\_out\_share increases by 1 unit, the output per unit of land will increase by about 5.938669%.

**However**, the relationship between ln\_yield and **rental\_in\_share** may not be statistically significant, because its p-value is obviously greater than 0.05. So **we may fail to reject the null hypothesis**.

9.

I prefer model (3) in q8, because the log transformation rescales the values. In this way, we can observe the relationship between the percentage change in yield and the percentage change in the rental\_in and rental\_out relatively.

10.

Under this context, the assumptions are:

- (1) The value of rental\_in contains no information about the mean of the unobserved factors.
- (2) The value of rental\_out contains no information about the mean of the unobserved factors.

### Possible reasons:

- (1) In reality, the relationship between yield & rental\_in and yield & rental\_out may not be linear.
- (2) There is endogeneity. When we regress yield to rental\_in or rental\_out, there may be some important factors ignored.

For example, households who are more skilled (we did not include this into our model) may produce higher yields. Meanwhile, they would like to rent (in) more land for more total output. On the other hand, households not that skilled may rent out some land for not being able to handle so much land. Meanwhile, they also produce lower yield.

In this case, the variables rental\_in and rental\_out actually correlate to households' skills, which also influence the yield. In short, the explanatory variables contain some information about the mean of the unobserved factors.

11.

 $\beta_1 < 0, \, \beta_2 > 0$ 

# Mechanism 1:

A household may not have enough energy, resources and willingness to attend to too much land. Thus, they may rent out some land to attend the rest more carefully. On the contrary, when a household rents in some land, every unit of land may get less attended so that the output per unit of land will decrease.

### Mechanism2:

A household may rent out the less productive land for a higher profit rate. In this case, the output per unit of land will increase.

A household may rent in the less productive land from others for higher total output. In this case, the output per unit of land (which is actually an average value), will decrease.

Q2

1.

*yield<sub>i</sub>*: the output per unit of land of village *i* this year.

 $treat_i$ : a boolean variable indicating whether village i received their treatment. 1 for yes and 0 for no

2.

 $yield_i = \beta_0 + \beta_1 treat_i + \mu_i$ 

*yieldi*: the output per unit of land of village *i* this year.

 $treat_i$ : a boolean variable indicating whether village i received their treatment. 1 for yes and 0 for no.

 $\beta_0$ : the intercept indicating the yield (output per unit of land) without treatment (land rental contract law), which may be the average yield of the control group (the remaining 100 villages).  $\beta_1$ : the coefficient of interest, which indicates the effect of the treatment (holding ceteris paribus).

 $\mu_i$ : the error term for village *i* which may influence yield<sub>i</sub>.

3.

- (1) Maybe the richest villages are rich because basically their yields are higher. This issue may increase the  $yield_i$  we observe, which will magnify the treatment effect we observe. In this case,  $\beta_I$  may be too large.
- (2) If we use the average yield of the control group, who failed to be selected to receive the treatment, as  $\beta_0$ , then  $\beta_0$  may be smaller than our ideal assumption. Because the villages in the control group may be poorer and their average yield may be lower than the population (all the 200 villages) before we exert treatment. In this case,  $\beta_I$  may be too small.
- (3) The treatment effect,  $\beta_1$ , may not be applicable for all the villages. It may just be applicable for richer villages. Because the data we acquired is from the richer villages.

## Code

```
clear all
    cap log close
    log using "122090407.log", replace
4
 5
    cd "/Users/30706/Desktop/eco"
7
    use aghousehold, clear
8
   keep if year == 2010
9
10
11 gen yield = d32/d31
   gen rental_in = c10
12
    gen rental_out = c13
13
14
15 sum yield
16 sum rental_in
    sum rental_out
17
19 reg yield rental_in
20 predict hat_yield_1
21 predict res_yield_1, residuals
22 sum res_yield_1
23
24 reg yield rental_out
25 predict hat_yield_2
26 predict res_yield_2, residuals
27 sum res_yield_2
28
   gen rental_in_share = 100*rental_in/d31
29
    gen rental_out_share = 100*rental_out/d31
31
   reg yield rental_in_share
32 reg yield rental_out_share
33
34 gen ln_yield = log(yield+1)
35
   reg ln_yield rental_in_share
   reg ln_yield rental_out_share
37
38 log close
```

## Log

```
name: <unnamed>
    log: C:\Users\30706\Desktop\eco\122090407.log
log type: text
opened on: 6 Oct 2023, 01:45:31
. cd "/Users/30706/Desktop/eco"
C:\Users\30706\Desktop\eco
. use aghousehold, clear
. keep if year == 2010
(0 observations deleted)
. gen yield = d32/d31
. gen rental_in = c10
. gen rental_out = c13
. sum yield
   Variable | Obs Mean
                   Obs Mean Std. dev. Min Max
                                              0 36461.98
                                  395.1982
    yield | 14,171 433.6169
. sum rental_in
  Variable | Obs Mean
                                   Std. dev. Min
                                                          Max
  rental_in | 14,171 .5031543
                                   4.466769
                                                           200
. sum rental_out
                           Mean
                 0bs
   Variable
                                   Std. dev.
                                                 Min
                                                           Max
 rental_out | 14,171 .2164561
                                   1.489389
                                                          60
```

```
. reg yield rental in
   Source | SS df MS
                              Number of obs = 14,171
  1.10
                                           0.2940
                              R-squared =
Adj R-squared =
MSE =
                                           0.0001
                                          0.0000
   Total | 2.2131e+09 14,170 156181.633 Root MSE
                                           395.2
   yield | Coefficient Std. err. t P>|t| [95% conf. interval]
 . predict hat_yield_1
(option xb assumed; fitted values)
. predict res_yield_1, residuals
. sum res_yield_1
 Variable Obs Mean Std. dev. Min
                                        Max
res_yield_1 | 14,171 -6.09e-07 395.1829 -434.0094 36027.97
. reg yield rental_out
                   df MS
                              Number of obs =
   Source
           SS
                                          14,171
  = 395.21
   yield | Coefficient Std. err. t P>|t| [95% conf. interval]
rental_out | .6515137 2.229133 0.29 0.770 -3.717879 5.020907 
_cons | 433.4759 3.354809 129.21 0.000 426.9001 440.0518
```

```
. predict hat_yield_2
(option xb assumed; fitted values)

. predict res_yield_2, residuals

. sum res_yield_2

Variable | Obs Mean Std. dev. Min Max

res_yield_2 | 14,171 1.10e-06 395.197 -436.8911 36028.5

. gen rental_in_share = 100*rental_in/d31

. gen rental_out_share = 100*rental_out/d31

. reg yield rental_in_share

Source | SS df MS Number of obs = 14,171

Model | 111081.624 1 111081.624 Prob > F = 0.3991

Residual | 2.2130e+09 14,169 156184.816 R-squared = 0.0001

Model | 11081.624 1 156181.633 Root MSE = 395.2

yield | Coefficient Std. err. t P>|t | [95% conf. interval]

rental_in_s~e | .0540097 .0640427 0.84 0.399 -.0715225 .1795418

__cons | 433.37 3.332743 130.03 0.000 426.8374 439.9026
```

	ntal_out_share						
Source	SS	df	MS	Number F(1, 14		=	14,171 24.04
Model	3747795.36	1	3747795.36	Prob >	F	=	0.0000
Residual	2.2093e+09	14,169	155928.149	R-squar Adj R-s		=	0.0017 0.0016
Total	2.2131e+09	14,170	156181.633	Root MS		=	394.88
yield	Coefficient	Std. err	. t	P> t	[95% cd	onf.	interval]
rental_out_~e	.2350118	.0479362	4.90	0.000	.141056	95	.3289732
_cons	431.6123	3.342233	129.14	0.000	425.061	11	438.1635
. reg ln_yield   Source	rental_in_sha	re df	MS	Number	of obs	=	14,171
Source	 SS	df		F(1, 14	169)	=======================================	1.94
Source   Model	 SS  .689144057	df 1	.689144057	F(1, 14 Prob >	169) F	=	1.94 0.1642
Source	 SS	df 1	.689144057	F(1, 14 Prob > R-squar	169) F ed	= =	1.94 0.1642 0.0001
Source   	SS .689144057 5045.22149	df 1 14,169	.689144057	F(1, 14 Prob >	169) F red squared	=	1.94 0.1642
Source   Model   Residual   Total	SS .689144057 5045.22149	df 1 14,169 1	.689144057 .356074634 	F(1, 14 Prob > R-squar Adj R-s Root MS	F169) F red squared SE	= = = = = = = = = = = = = = = = = = = =	1.94 0.1642 0.0001 0.0001
Model   Residual   Total	SS .689144057 5045.22149 5045.91063	df 1 14,169 1	.689144057 .356074634 .356098139	F(1, 14 Prob > R-squar Adj R-s Root MS	F169) F red squared SE	= = = = = onf.	1.94 0.1642 0.0001 0.0001 .59672
Source   Model   Residual   Total   In_yield	SS .689144057 5045.22149 5045.91063	df 1 14,169 14,170 Std. err	.356098139	F(1, 14 Prob > R-squar Adj R-s Root MS	:169) F red equared EE 	= = = = = onf.	1.94 0.1642 0.0001 0.0001 .59672 interval]

. reg ln_yield	rental_out_sha	are				
Source			MS			
	6.96783523			F(1, 14169) Prob > F		19.59 0.0000
Residual	5038.94279	14,169	.355631505			
Total	5045.91063	14,170	.356098139	Adj R-squared Root MSE		0.0013 .59635
ln_yield	Coefficient	Std. err.	t	P> t  [95% c	onf.	interval]
				0.000 .00017 0.000 5.9287		
log type: to	:\Users\30706\[		:o\12209040;	7.log		