Q1.

1.

```
. reg yield rental_in_share rental_out_share
```

Source	SS	df	MS		
Model	3856365.5	2	1928182.75	Number of obs	= 14,171
Residual	2.2092e+09	14,168	155931.491	F(2, 14168)	= 12.37
				Prob > F	= 0.0000
				R-squared	= 0.0017
				Adj R-squared	= 0.0016
Total	2.2131e+09	14,170	156181.633	Root MSE	= 394.88

yield	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
rental_in_s~e	.0533957	.0639909	0.83	0.404	-.0720348	.1788262
rental_out_~e	.2349335	.0479369	4.90	0.000	.140971	.3288961
_cons	431.3688	3.35498	128.58	0.000	424.7926	437.945

Yes, it suffers from omitted variable bias.

(1) $R^2 = 0.0017$, indicating the model explains only 0.17% of the variation in the data.

(2) The P-value for **rental_in_share** is 0.404, much greater than 0.05, so we can't reject the null hypothesis.

2. Let's take **health** -- percentage of household population in good health condition – as an example.

   The more villagers in good health condition in a village, the more land they would like to rent in for more output. Because **rental_in_share = rent_in/ total sown area (d31)**, so both the **rent_in and d31** will increase by the same amount. Thus, rental_in_share will increase. So, the covariance between **rental_in_share (X) and health (u)** is positive.

   Therefore, in this example, β_1**-hat** > β_1.

3. To add the possibly omitted variables. I'd like to add variable **health, f9 (Purchase quantity of fertiliser (kg)) and f26 (Purchase value: agricultural diesel (yuan))** to the regression.

```
. reg yield rental_in_share rental_out_share health f9 f26
```

Source	SS	df	MS		
Model	1735683.48	5	347136.697	Number of obs	= 3,949
Residual	182534509	3,943	46293.3068	F(5, 3943)	= 7.50
				Prob > F	= 0.0000
				R-squared	= 0.0094
				Adj R-squared	= 0.0082
Total	184270192	3,948	46674.3141	Root MSE	= 215.16

yield	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
rental_in_s~e	-.1678082	.0615577	-2.73	0.006	-.2884962	-.0471203
rental_out_~e	.294061	.1068383	2.75	0.006	.0845976	.5035244
health	.0353981	.1636156	0.22	0.829	-.285381	.3561772
f9	.0125183	.0022834	5.48	0.000	.0080415	.016995
f26	-.0071274	.0026058	-2.74	0.006	-.0122363	-.0020185
_cons	415.1987	15.07856	27.54	0.000	385.6362	444.7612

4. 100*10%*(-0.1678082) = -1.678082 units of yield.

5.

(1) Regression model:
$yield_i=\beta_0+\beta_1 rental\_in\_share_i+\beta_2 rental\_out\_share_i+\beta_3 health_i+\beta_4 f9_i+\beta_5 f26_i+\beta_6 huzhu\_edu_i+\mu_i$

(2) H0: $\beta_6 = 0$
H1: $\beta_6 \neq 0$

(3) OLS estimation & t-test

```
. reg yield rental_in_share rental_out_share health f9 f26 huzhu_edu
```

Source	SS	df	MS		
Model	1487597.88	6	247932.98	Number of obs	= 3,864
Residual	178325946	3,857	46234.365	F(6, 3857)	= 5.36
				Prob > F	= 0.0000
				R-squared	= 0.0083
				Adj R-squared	= 0.0067
Total	179813544	3,863	46547.6427	Root MSE	= 215.02

yield	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
rental_in_s~e	-.1564082	.0616119	-2.54	0.011	-.2772031	-.0356132
rental_out_~e	.2453433	.1081111	2.27	0.023	.0333829	.4573037
health	.0404872	.1653022	0.24	0.807	-.2836008	.3645752
f9	.0118446	.0022891	5.17	0.000	.0073567	.0163326
f26	-.006786	.0026098	-2.60	0.009	-.0119026	-.0016693
huzhu_edu	-.5648724	1.475569	-0.38	0.702	-3.457842	2.328097
_cons	419.7506	17.37659	24.16	0.000	385.6824	453.8188

(4) We can see that the p-value of huzhu_edu is 0.702, significantly larger than 0.05. Thus, we cannot reject the null hypothesis at the 5% significance level.

6.

(1) Regression model:
$\mathbf{yield_I = \beta_0 + \beta_1 rental\_in\_share_i + \beta_2 rental\_out\_share_i + \beta_3 health_i + \beta_4 f9_i + \beta_5 f26_i + \mu_i}$

(2) H0: $\beta_1 - \beta_2 = 0$
H1: $\beta_1 - \beta_2 \neq 0$

(3) “lincom rental_in_share – rental_out_share” in Stata

```
. reg yield rental_in_share rental_out_share health f9 f26
```

Source	SS	df	MS	Number of obs	=	3,949
				F(5, 3943)	=	7.50
Model	1735683.48	5	347136.697	Prob > F	=	0.0000
Residual	182534509	3,943	46293.3068	R-squared	=	0.0094
				Adj R-squared	=	0.0082
Total	184270192	3,948	46674.3141	Root MSE	=	215.16

yield	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
rental_in_s~e	-.1678082	.0615577	-2.73	0.006	-.2884962	-.0471203
rental_out_~e	.294061	.1068383	2.75	0.006	.0845976	.5035244
health	.0353981	.1636156	0.22	0.829	-.285381	.3561772
f9	.0125183	.0022834	5.48	0.000	.0080415	.016995
f26	-.0071274	.0026058	-2.74	0.006	-.0122363	-.0020185
_cons	415.1987	15.07856	27.54	0.000	385.6362	444.7612

```
. lincom rental_in_share - rental_out_share

 ( 1)  rental_in_share - rental_out_share = 0
```

yield	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
(1)	-.4618692	.1251715	-3.69	0.000	-.7072762	-.2164622

(4) We can see that the p-value is 0.000, obviously smaller than 0.1. Thus, we can reject the null hypothesis. At the 10% significance level, economist B's argument is correct.

7.

Higher. We add more regressors which will influence the yield, so the variation of the data is explained more.

It isn't. The addition of the regressors will make SSE larger and SSR smaller. Thus, R^2 will automatically become larger even if those regressors are not truly related to the dependent variable. Thus, R^2 is not a good measure.

Q2:

1. No and No.

   In this model, u_i contains W_i. Whether located in the coastal region or not determines the probability of receiving treatment. What's more, coastal and inland region have different natural conditions, which will also influence the yield.

   Thus, u_i can influence both X_i and Y_i, and X_i contains some information of u_i. $\mathbf{E(u_i|X_i) \neq 0}$ and the OLS estimator of $\boldsymbol{\beta_1}$ **is biased**.

2.

(1) No and Yes.

   In this model, X_i is a random treatment assignment, it is independent from u_i for u_i does not contain W_i. So, $\mathbf{E(u_i|X_i, W_i)}$ **has nothing to do with** $\mathbf{X_i}$. Also, Y_i and X_i has a linear relationship. Besides, X_i is a random sample. Thus, the OLS estimator of $\boldsymbol{\beta_1}$ **is unbiased**.

(2) Yes.

   W_i represents the geographical location of $village_i$. Whether the village locates at the coastal region or inland region will determine the natural conditions, like weather and soil conditions, which will also influence the yield. However, the model does not add these factors as regressors and they are contained in u_i. Thus, $E(u_i|X_i, W_i)$ has a relationship with W_i.

Q3:

Sample Regression Function: $\hat{Y}_i = \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$

1.

$$\hat{u}_i = Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}$$

$$SSR = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})^2$$

2.

$$\frac{\partial SSR}{\partial \hat{\beta}_1} = \sum_{i=1}^{n} 2(Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})(-X_{1i})$$

$$\frac{\partial SSR}{\partial \hat{\beta}_2} = \sum_{i=1}^{n} 2(Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})(-X_{2i})$$

3.

$$\sum_{i=1}^{n} X_{1i} Y_i - \sum_{i=1}^{n} \hat{\beta}_1 X_{1i}^2 - \sum_{i=1}^{n} \hat{\beta}_2 X_{2i} X_{1i} = 0$$

$\because \sum_{i=1}^{n} X_{1i} X_{2i} = 0$

$$\therefore \hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_{1i} Y_i}{\sum_{i=1}^{n} X_{1i}^2}$$

4.

$$\sum_{i=1}^{n} X_{2i} Y_i - \sum_{i=1}^{n} \hat{\beta}_1 X_{1i} X_{2i} - \sum_{i=1}^{n} \hat{\beta}_2 X_{2i}^2 = 0$$

$$\Rightarrow \hat{\beta}_2 = \frac{\sum_{i=1}^{n} X_{2i} Y_i}{\sum_{i=1}^{n} X_{2i}^2} - \hat{\beta}_1 \frac{\sum_{i=1}^{n} X_{1i} X_{2i}}{\sum_{i=1}^{n} X_{2i}^2}$$

$$\Rightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_{1i} Y_i - \hat{\beta}_2 \sum_{i=1}^{n} X_{2i} X_{1i}}{\sum_{i=1}^{n} X_{1i}^2} = \frac{\sum_{i=1}^{n} X_{1i} Y_i - \sum_{i=1}^{n} X_{2i} X_{1i} \left(\frac{\sum_{i=1}^{n} X_{2i} Y_i}{\sum_{i=1}^{n} X_{2i}^2} - \hat{\beta}_1 \frac{\sum_{i=1}^{n} X_{1i} X_{2i}}{\sum_{i=1}^{n} X_{2i}^2} \right)}{\sum_{i=1}^{n} X_{1i}^2}$$

$$\Rightarrow \hat{\beta}_1 \left[1 - \frac{\left(\sum_{i=1}^{n} X_{2i} X_{1i} \right)^2}{\sum_{i=1}^{n} X_{1i}^2 \cdot \sum_{i=1}^{n} X_{2i}^2} \right] = \frac{\sum_{i=1}^{n} X_{1i} Y_i \cdot \sum_{i=1}^{n} X_{2i}^2 - \sum_{i=1}^{n} X_{2i} X_{1i} \cdot \sum_{i=1}^{n} X_{2i} Y_i}{\sum_{i=1}^{n} X_{1i}^2 \cdot \sum_{i=1}^{n} X_{2i}^2}$$

$$\Rightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_{1i} Y_i \cdot \sum_{i=1}^{n} X_{2i}^2 - \sum_{i=1}^{n} X_{2i} X_{1i} \cdot \sum_{i=1}^{n} X_{2i} Y_i}{\sum_{i=1}^{n} X_{1i}^2 \cdot \sum_{i=1}^{n} X_{2i}^2 - \left(\sum_{i=1}^{n} X_{2i} X_{1i} \right)^2}$$

5.

Sample Regression Function:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$$

For $\hat{\beta}_0$:

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n} Y_i - \hat{\beta}_0 - \hat{\beta}_1 \cdot \frac{1}{n}\sum_{i=1}^{n} X_{1i} - \hat{\beta}_2 \frac{1}{n}\sum_{i=1}^{n} X_{2i} = 0$$

$$\bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2 = 0$$

$$\Rightarrow \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$$

6.

① For $\hat{\beta}_1$:

$$\sum_{i=1}^{n} X_{1i}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) = 0$$

$$\sum_{i=1}^{n} X_{1i}(Y_i - \bar{Y} + \hat{\beta}_1 \bar{X}_1 + \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) = 0$$

$$\sum_{i=1}^{n} X_{1i}(Y_i - \bar{Y}) + \hat{\beta}_1 \sum_{i=1}^{n} X_{1i}(\bar{X}_1 - X_{1i}) + \hat{\beta}_2 \sum_{i=1}^{n} X_{1i}(\bar{X}_2 - X_{2i}) = 0$$

$\because \sum_{i=1}^{n}(\bar{X}_2 - X_{2i}) = 0 \Rightarrow \bar{X}_1 \sum_{i=1}^{n}(\bar{X}_2 - X_{2i}) = 0 \Rightarrow \sum_{i=1}^{n} X_{1i}(\bar{X}_2 - X_{2i}) = \sum_{i=1}^{n}(X_{1i} - \bar{X}_1)(\bar{X}_2 - X_{2i}) = 0$

$\therefore \sum_{i=1}^{n} X_{1i}(Y_i - \bar{Y}) + \hat{\beta}_1 \sum_{i=1}^{n} X_{1i}(\bar{X}_1 - X_{1i}) = 0$

$$\sum_{i=1}^{n} X_{1i}(Y_i - \bar{Y}) - \hat{\beta}_1 \sum_{i=1}^{n} X_{1i}(X_{1i} - \bar{X}_1) = 0$$

Similarly, $\sum_{i=1}^{n} X_{1i}(X_{1i} - \bar{X}_1) = \sum_{i=1}^{n}(X_{1i} - \bar{X}_1)^2$, $\sum_{i=1}^{n} X_{1i}(Y_i - \bar{Y}) = \sum_{i=1}^{n}(X_{1i} - \bar{X}_1)(Y_i - \bar{Y})$

$$\therefore \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_{1i} - \bar{X}_1)(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_{1i} - \bar{X}_1)^2}$$

② They are the same under the condition that $\sum_{i=1}^{n}(X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2) = 0$

Code

```
// Assignment 2
reg yield rental_in_share rental_out_share
reg yield rental_in_share rental_out_share health f9 f26
reg yield rental_in_share rental_out_share health f9 f26 huzhu_edu
reg yield rental_in_share rental_out_share health f9 f26
lincom rental_in_share - rental_out_share
```

Log

```
.
. // Assignment 2
. reg yield rental_in_share rental_out_share

      Source |       SS           df       MS      Number of obs   =    14,171
-------------+----------------------------------   F(2, 14168)     =     12.37
       Model |   3856365.5         2  1928182.75   Prob > F        =    0.0000
    Residual |  2.2092e+09    14,168  155931.491   R-squared       =    0.0017
-------------+----------------------------------   Adj R-squared   =    0.0016
       Total |  2.2131e+09    14,170  156181.633   Root MSE        =    394.88

------------------------------------------------------------------------------
       yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
rental_in_s~e |   .0533957   .0639909     0.83   0.404    -.0720348    .1788262
rental_out_~e |   .2349335   .0479369     4.90   0.000      .140971    .3288961
        _cons |   431.3688    3.35498   128.58   0.000     424.7926     437.945
------------------------------------------------------------------------------

. reg yield rental_in_share rental_out_share health f9 f26

      Source |       SS           df       MS      Number of obs   =     3,949
-------------+----------------------------------   F(5, 3943)      =      7.50
       Model |  1735683.48         5  347136.697   Prob > F        =    0.0000
    Residual |   182534509     3,943  46293.3068   R-squared       =    0.0094
-------------+----------------------------------   Adj R-squared   =    0.0082
       Total |   184270192     3,948  46674.3141   Root MSE        =    215.16

------------------------------------------------------------------------------
       yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
rental_in_s~e |  -.1678082   .0615577    -2.73   0.006    -.2884962   -.0471203
rental_out_~e |    .294061   .1068383     2.75   0.006     .0845976    .5035244
       health |   .0353981   .1636156     0.22   0.829     -.285381    .3561772
           f9 |   .0125183   .0022834     5.48   0.000     .0080415     .016995
          f26 |  -.0071274   .0026058    -2.74   0.006    -.0122363   -.0020185
        _cons |   415.1987   15.07856    27.54   0.000     385.6362    444.7612
------------------------------------------------------------------------------
```

```
. reg yield rental_in_share rental_out_share health f9 f26 huzhu_edu

      Source |       SS           df       MS      Number of obs   =     3,864
-------------+----------------------------------   F(6, 3857)      =      5.36
       Model |  1487597.88         6   247932.98   Prob > F        =    0.0000
    Residual |   178325946     3,857   46234.365   R-squared       =    0.0083
-------------+----------------------------------   Adj R-squared   =    0.0067
       Total |   179813544     3,863  46547.6427   Root MSE        =    215.02

------------------------------------------------------------------------------
       yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
rental_in_s~e |  -.1564082   .0616119    -2.54   0.011    -.2772031   -.0356132
rental_out_~e |   .2453433   .1081111     2.27   0.023     .0333829    .4573037
      health |   .0404872   .1653022     0.24   0.807    -.2836008    .3645752
          f9 |   .0118446   .0022891     5.17   0.000     .0073567    .0163326
         f26 |   -.006786   .0026098    -2.60   0.009    -.0119026   -.0016693
   huzhu_edu |  -.5648724   1.475569    -0.38   0.702    -3.457842    2.328097
       _cons |   419.7506   17.37659    24.16   0.000     385.6824    453.8188
------------------------------------------------------------------------------
```

```
. reg yield rental_in_share rental_out_share health f9 f26

      Source |       SS           df       MS      Number of obs   =     3,949
-------------+----------------------------------   F(5, 3943)      =      7.50
       Model |  1735683.48         5  347136.697   Prob > F        =    0.0000
    Residual |   182534509     3,943  46293.3068   R-squared       =    0.0094
-------------+----------------------------------   Adj R-squared   =    0.0082
       Total |   184270192     3,948  46674.3141   Root MSE        =    215.16

------------------------------------------------------------------------------
       yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
rental_in_s~e |  -.1678082   .0615577    -2.73   0.006    -.2884962   -.0471203
rental_out_~e |    .294061   .1068383     2.75   0.006     .0845976    .5035244
      health |   .0353981   .1636156     0.22   0.829     -.285381    .3561772
          f9 |   .0125183   .0022834     5.48   0.000     .0080415     .016995
         f26 |  -.0071274   .0026058    -2.74   0.006    -.0122363   -.0020185
       _cons |   415.1987   15.07856    27.54   0.000     385.6362    444.7612
------------------------------------------------------------------------------

. lincom rental_in_share - rental_out_share

 ( 1)  rental_in_share - rental_out_share = 0

------------------------------------------------------------------------------
       yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |  -.4618692   .1251715    -3.69   0.000    -.7072762   -.2164622
------------------------------------------------------------------------------

.
. log close
      name:  <unnamed>
       log:  C:\Users\30706\Desktop\122090407.log
  log type:  text
 closed on:  24 Oct 2023, 17:49:30
-------------------------------------------------------------------------------
```