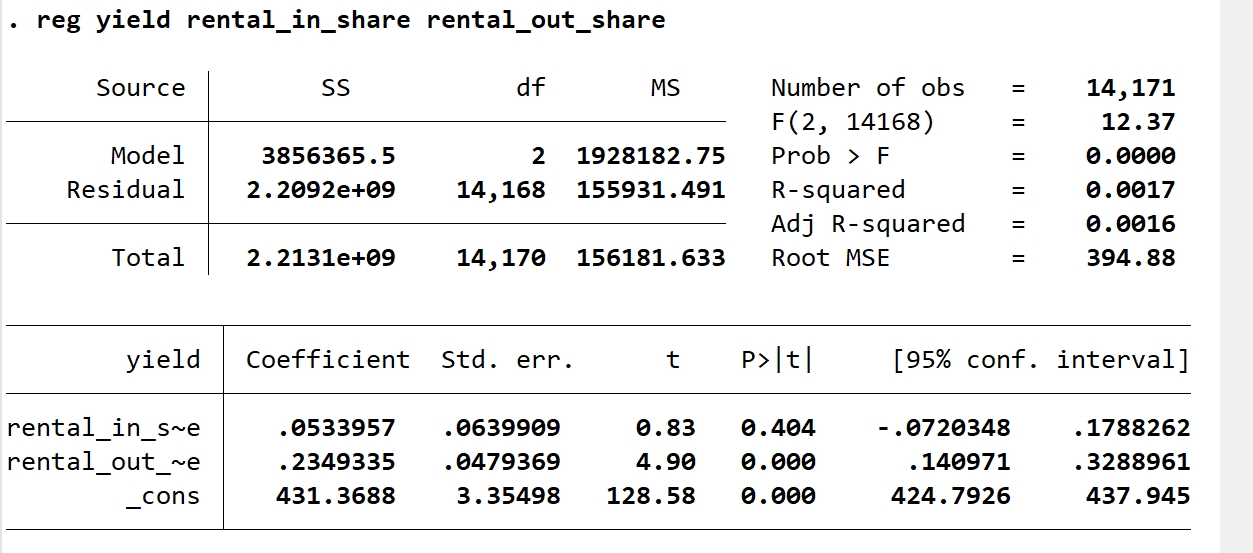
Q1.





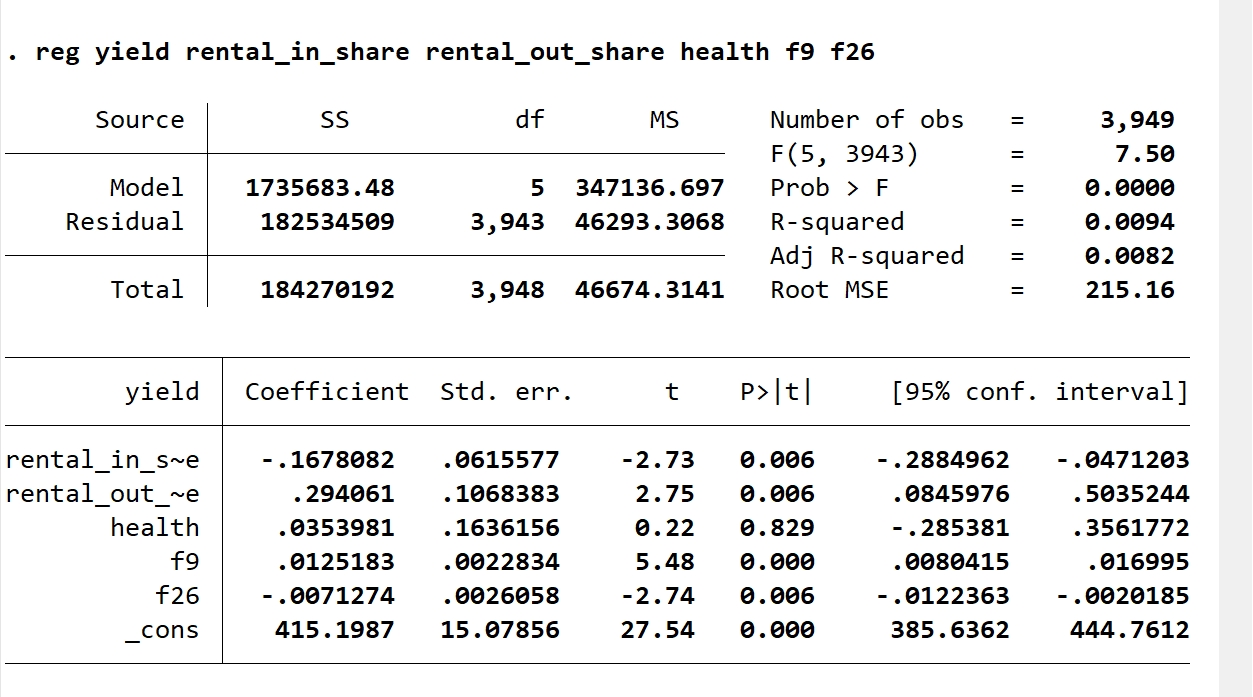
Yes, it suffers from omitted variable bias.

1. R2 = 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.
3. 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, **β1-hat > β1**.

1. 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.



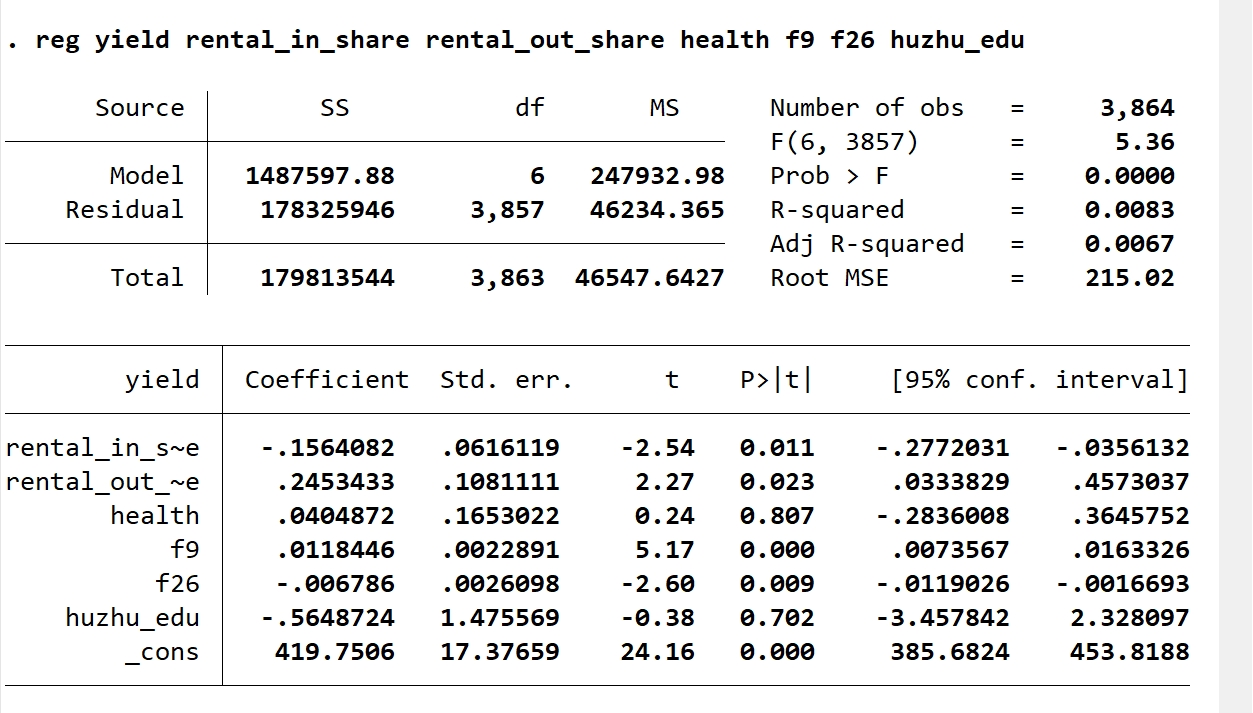
1. 100\*10%\*(-0.1678082) = -1.678082 units of yield.
2. Regression model:

yield­i=β­0+β1rental\_in\_sharei+β2rental\_out\_sharei+β3healthi+β4f9i+β5f26i+β6huzhu\_edui+μi

1. H0: β6 = 0

H1: β6 ≠ 0

1. OLS estimation & t-test



1. 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.
3. Regression model:

**yield­I = β­0 + β1rental\_in\_sharei + β2rental\_out\_sharei + β3healthi + β4f9i + β5f26i + μi**

1. H0: β1 - β2 = 0

H1: β1 - β2 ≠ 0

1. “lincom rental\_in\_share – rental\_out\_share” in Stata

手机屏幕截图

Description automatically generated

1. 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.

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, R2 will automatically become larger even if those regressors are not truly related to the dependent variable. Thus, R2 is not a good measure.

Q2:

1. No and No.

In this model, ui contains Wi. 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, ui can influence both Xi and Yi,and Xi contains some information of ui. **E(ui|Xi) ≠ 0** and the OLS estimator of **β1 is biased**.

1. No and Yes.

In this model, Xi is a random treatment assignment, it is independent from ui for ui does not contain Wi. So, **E(ui|Xi, Wi) has nothing to do with Xi**. Also, Yi and Xi has a linear relationship. Besides, Xi is a random sample. Thus, the OLS estimator of **β1 is unbiased**.

1. Yes.

Wi represents the geographical location of villagei. 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 ui. Thus, E(ui|Xi, Wi) has a relationship with Wi.

Q3: