Model exam

1. Your friend is studying the effect of x1 on y. She has also data for variable x2, and she is considering whether she should also include this variable in her regression model. She has conducted a regression where the dependent variable is y and the independent variables are x1 and x2. Your friend has received the following output:

. regr y x1 x2

```
    Source |    SS       df     MS      Number of obs  =   10,000
-------------+----------------------------------  F(2, 9997)    =   833.53
     Model | 15192.4993      2 7596.24965  Prob > F      =   0.0000
  Residual | 91106.3442   9,997 9.11336843  R-squared     =   0.1429
-------------+----------------------------------  Adj R-squared =   0.1428
     Total | 106298.843   9,999 10.6309474  Root MSE      =   3.0188
```

```
------------------------------------------------------------------------------
         y |   Coef.  Std. Err.    t   P>|t|   [95% Conf. Interval]
-------------+----------------------------------------------------------------
        x1 | .2957783 .0150663   19.63  0.000    .2662453   .3253114
        x2 |  .163956 .0188318    8.71  0.000    .1270419   .2008701
     _cons | .9793178 .0575534   17.02  0.000    .8665015   1.092134
------------------------------------------------------------------------------
```

Your friend has sent you a data set that can be found on the course web page. The data is in an xlsx file with the name "final_exam_2020_OVB.xlsx". She has included the variables x1 (column A) and x2 (column B) but forgot to include y. Your friend would like to know:

1.1. If she has a reason to think there might be an omitted variable bias if she did not include x2 in her regression model?
1.2. How large could the omitted variable bias be if she did not include x2?

Answer both questions and justify your answers. If you can calculate the omitted variable bias, carefully document your calculations.

2. You have conducted a regression and the p-value of your main variable of interest was p = 0.03. Answer the following questions:
2.1. What does this mean in terms of statistical significance?
2.2. Based on this, what can you conclude about the possibility of omitted variable bias? Justify your answer.
2.3. Could [-0.03456, 0.1275] be the 95% level confidence interval of your estimate? Justify your answer.
2.4. Based on the p-value, can you determine what is the standard error of your estimate?