Written Portion:

Question 1

(i) Test 1 is $H_0: B_0 = 0$

T-statistic = (-14.47)/16.27 = -0.889

When alpha = 0.05 and 86 degrees of freedom, our critical t value is about ± 1.988

Because 0.889 < 1.988, we fail to reject the null hypothesis.

Test 2 is $H_0: B_1 = 1$

T statistic = (976 - 1) / 0.049 = 19898

Since t statistic is big, it's likely to be significant, so we would reject the null hypothesis $H_0: B_1 = 1$

(ii) We will calculate the F-statistic:

F-statistic = [(209448.99 - 165644.51) / 2] / [165644.51 / (88 - 2)]

Estimated to be 11.37

With the given hints from homework, we can tell that 11.37 is bigger than all two critical values, so we would reject the joint null hypothesis.

(iii) SSR_full is the sum of squared residuals, and SSR_full = (1-R^2) * SST.

SST = 209,448.99 / (1 – 0.829)

= 1,233,530.53

F-statistic = [(209,466.73 - 209,448.99) / 3] / [209,448.99 / (88 - 5)]

= 0.00234

Since 0.00234 is smaller than 2.15, 2.71, and 4.01, we fail to reject the null hypothesis.

(iv) If the variance of 'price' changes with 'assess,' 'lotsize,' 'sqrft,' or 'bdrms,' it could potentially affect the interpretation of the F-test results from part (iii), making them less reliable. In such situations, it's important to address heteroscedasticity and consider alternative modeling approaches.

Question 2

The problem could occur because of multicollinearity in MLR model. Since two independent variables are highly correlated, it's challenging to isolate their effects on the dependent variable, so each variable may look statistically insignificant.

Computer Problems Output

Wooldridge Computer Exercise C4.2

	rank	salary	cost	LSAT	GPA \
count	156.000000	148.000000	150.000000	150.000000	149.000000
mean	83.750000	38946.714844	12736.240000	158.293333	3.302752
std	50.064169	12188.839844	4012.667528	4.701202	0.199776

min	1.000000	24900.000000	2623.000000	140.000000	2.730000	
25%	40.750000	29950.000000	9663.750000	155.000000	3.200000	
50%	83.500000	34725.000000	12898.500000	158.000000	3.300000	
75%	125.500000	41350.000000	16129.000000	161.000000	3.410000	
max	175.000000	78325.000000	20518.000000	171.000000	3.820000	

	libvol	faculty	age	clsize	north	south \
count	155.000000	152.000000	111.000000	153.000000	156.000000	156.00000
mean	347.070968	70.730263	85.819820	241.830065	0.205128	0.24359
std	188.242214	39.260376	40.891697	113.072275	0.405096	0.43063
min	124.000000	17.000000	3.000000	70.000000	0.000000	0.00000
25%	235.000000	45.000000	63.500000	163.000000	0.000000	0.00000
50%	302.000000	58.000000	85.000000	225.000000	0.000000	0.00000
75%	400.000000	86.250000	112.500000	284.000000	0.000000	0.00000
max	1745.000000	245.000000	206.000000	679.000000	1.000000	1.00000

	east	west	lsalary	studfac	top10	r11_25 \
count	156.000000	156.00000	148.000000	150.000000	156.000000	156.000000
mean	0.307692	0.24359	10.528775	3.706660	0.064103	0.102564
std	0.463025	0.43063	0.276720	1.311862	0.245724	0.304366
min	0.000000	0.00000	10.122623	1.010000	0.000000	0.000000
25%	0.000000	0.00000	10.307280	2.862103	0.000000	0.000000
50%	0.000000	0.00000	10.455215	3.555701	0.000000	0.000000
75%	1.000000	0.00000	10.629826	4.305159	0.000000	0.000000
max	1.000000	1.00000	11.268622	9.342857	1.000000	1.000000

	r26_40	r41_60	llibvol	lcost
count	156.000000	156.000000	155.000000	150.000000
mean	0.083333	0.115385	5.751981	9.392433
std	0.277276	0.320514	0.416548	0.368915
min	0.000000	0.000000	4.820282	7.872074
25%	0.000000	0.000000	5.459577	9.176127
50%	0.000000	0.000000	5.710427	9.464864
75%	0.000000	0.000000	5.991465	9.688369
max	1.000000	1.000000	7.464510	9.929058

(i)

We fail to reject the null hypothesis. The rank does not have a ceteris paribus effect on median starting salary.

(ii)

The p value for LSAT is 9.321670725277372e-07

The p value for GPA is 3.7340930942527556e-06

Both LSAT and GPA are individually significant for explaining salary.

(iii)
There is no need to add clsize and faculty to the equation.

(iv)

There could factors such as funding for faculty and students, as well as research quantity and quality from the school's faculty.
The passing rate for bar exam in the school can also impact the rank.

HW6/HW6.py

```python
import numpy as np
import pandas as pd
import statsmodels.api as sm
from scipy.stats import f


# Include similar code as HW1 for the basic steps
file_location = "/Users/amyliang/Eco441K/HW6/LAWSCH85.DTA"
df= pd.read_stata(file_location)

pd.set_option('display.max_columns', None)

# Wooldridge Computer Exercise C4.2
print("Wooldridge Computer Exercise C4.2")
print(df.describe())

print("(i)")
# Define the dependent variable (y) and the independent variable (X)
y = df['salary']
X = sm.add_constant(df['rank'])  # Add a constant term for the intercept

# Fit the regression model
model = sm.OLS(y, X).fit()

# Perform the hypothesis test on the coefficient for 'rank'
hypothesis = 'rank = 0'
t_test = model.t_test(hypothesis)

# Extract the p-value for 'rank'
p_value = t_test.effect[0]

# Set the significance level (e.g., 0.05)
alpha = 0.05

# Check if the p-value is less than the significance level
if p_value < alpha:
    print(f" We reject the null hypothesis. The rank has a ceteris paribus effect on median starting salary.")
else:
    print(" We fail to reject the null hypothesis. The rank does not have a ceteris paribus effect on median starting salary.")

print()
print("(ii)")
# Remove rows with missing values in 'salary', 'LSAT', and 'GPA' columns
df_clean = df.dropna(subset=['salary', 'LSAT', 'GPA'])


```

```
47
48 # Define the dependent variable (y) and the independent variables (X)
49 y = df_clean['salary']
50 X = df_clean[['LSAT', 'GPA']]
51 X = sm.add_constant(X)  # Add a constant term for the intercept
52
53 # Fit the multiple linear regression model
54 model = sm.OLS(y, X).fit()
55
56 # Perform the hypothesis tests on the coefficients for 'LSAT' and 'GPA'
57 t_test_lsat = model.t_test('LSAT = 0')
58 t_test_gpa = model.t_test('GPA = 0')
59
60 # Extract the p-values for 'LSAT' and 'GPA'
61 p_value_lsat = t_test_lsat.pvalue
62 p_value_gpa = t_test_gpa.pvalue
63 print("The p value for LSAT is", p_value_lsat)
64 print("The p value for GPA is", p_value_gpa)
65
66 # Set the significance level (e.g., 0.05)
67 alpha = 0.05
68
69 # Check if the p-values are less than the significance level
70 if p_value_lsat < alpha and p_value_gpa < alpha:
71     print("Both LSAT and GPA are individually significant for explaining salary.")
72 elif p_value_lsat < alpha or p_value_gpa < alpha:
73     print("At least one of LSAT and GPA is individually significant for explaining
   salary.")
74 else:
75     print("Neither LSAT nor GPA is individually significant for explaining
   salary.")
76
77 print()
78 print("(iii)")
79
80 # Remove rows with missing values in 'salary', 'LSAT', 'GPA', 'clsize', and '
   faculty' columns
81 new_df = df.dropna(subset=['salary', 'LSAT', 'GPA', 'clsize', 'faculty'])
82
83 # Define the dependent variable (y) and the initial independent variables (LSAT
   and GPA)
84 y = new_df['salary']
85 X_initial = new_df[['LSAT', 'GPA']]
86 X_initial = sm.add_constant(X_initial)  # Add a constant term for the intercept
87
88 # Fit the initial multiple linear regression model
89 model_initial = sm.OLS(y, X_initial).fit()
90
91 # Define the additional independent variables (clsize and faculty)
92 X_additional = new_df[['clsize', 'faculty']]
93 X_additional = sm.add_constant(X_additional)  # Add a constant term for the
```

```
intercept

# Fit the extended multiple linear regression model
model_extended = sm.OLS(y, X_additional).fit()

# Perform the F-test to compare the fit of the two models
f_statistic = (model_initial.ssr - model_extended.ssr) / (model_extended.df_model - model_initial.df_model) / model_extended.ssr / (model_extended.nobs - model_extended.df_model)

# Calculate the critical F-value for your desired significance level (e.g., 0.05)
from scipy.stats import f
alpha = 0.05
dfn = model_extended.df_model - model_initial.df_model
dfd = model_extended.nobs - model_extended.df_model
critical_f_value = f.ppf(1 - alpha, dfn, dfd)

# Check if the F-statistic is greater than the critical F-value
if f_statistic > critical_f_value:
    print("The size of the entering class (clsize) or the size of the faculty (faculty) should be added to the equation.")
else:
    print("There is no need to add clsize and faculty to the equation.")

print()
print("(iv)")
ans = """
There could factors such as funding for faculty and students, as well as research quantity and quality from the school's faculty.
The passing rate for bar exam in the school can also impact the rank.
"""
print(ans)
```