Econometrics 322 Lab #5

Extending OLS Regression

Prof. Paczkowski

Enter your Name in the Next Cell

Maanya Tandon (RUID mt958)

Grading Rubric

Score:___: Max(0, 20 - Total Deductions)

Content Area	Deduction	Times Deducted	Check	Comments
Abstract				
Missing	5		[ ]	
Insufficient/Wrong Focus	1		[ ]	
Data Dictionary (Metadata)				
Missing	5		[ ]	
Insufficient/Wrong Form or Wording	1		[ ]	
Graphs				
Missing	5		[ ]	
Missing Title	1		[ ]	
Missing/Wrong Labels	1		[ ]	
Pre-Lab				
Missing	5		[ ]	
Insufficient/Wrong Answer	2 Each		[ ]	
No/Incorrect/Insufficient Model Specification	2		[ ]	
No/Incorrect Statistical Hypothesis Statement	2 Each		[ ]	
Post-Lab				
Missing	5		[ ]	
Insufficient/Wrong Answer	2 Each		[ ]	
Correlations				
Missing	5		[ ]	
Insufficient/Wrong Analysis	2		[ ]	
Missing Graph	2		[ ]	
Estimations				
Missing	5		[ ]	
No or incorrect discussion/interpretation of...				
Hypothesis tests and p-values	2 Each		[ ]	
R^2	2		[ ]	
F-Statistic	2		[ ]	
Multicollinearity/VIF	2		[ ]	
Heteroskedasticity/Test	2		[ ]	
Autocorrelation/Test	2		[ ]	
No/insufficient model selection	2		[ ]	
Elasticities				
Missing	5		[ ]	
Incorrect Interpretation	2		[ ]	
Missing Summary Table	2		[ ]	
Model Portfolio			[ ]	
Missing	5		[ ]	
General Comments:				

Contents

Collaboration Policy

Back to Contents

```
1. Study groups are allowed but I expect students to understand and complete their own assignmen
ts and to hand in one
assignment per student.
2. If you worked in a group, please put the names of your study group in the following table.
3. Just like all other classes at Rutgers, the student Honor Code is taken seriously.

The submitted assignment must be your work.
```

Collaborator(s) Name(s)
name(s) here

Introduction

Back to Contents

Purpose

Back to Contents

The purpose of this lab is to introduce you to the application of the OLS methodology to an actual problem.

At the end of this lab, you will be able to:

1. specify a linear model for a problem;
2. estimate a linear model in Statsmodel;
3. interpret key statistics;
4. identify shortcomings in the proposed linear model;
5. summarize the regression output;
6. estimate elasticities and judge their reasonableness;
7. build a model portfolio.

Problem

Back to Contents

A few markets are key to the health of the economy. Autos and energy are two that most people would quickly cite. The housing market is another. Housing data can be divided into housing permits, housing starts, new home sales, and existing home sales. In addition, housing can be viewed as single family or multifamily.

Existing home sales (as well as new home sales) are dependent on the future, expected state of the economy. For the future state, if there is a concern that the economy will go into a recession soon, people will be less willing to buy a new home for fear of losing their job, so housing sales will be weak or decline. But if the economy looks promising, then housing sales will be strong or increase.

Another factor important for the decision to buy a new home is the current mortgage rate or yield. Usually, people get a fixed rate which means they will be paying that same rate for many years, often 30 years. In essence, the mortgage rate adds to the price of a home.

In this lab, you will estimate a model to explain home sales as a function of the mortgage yield.

Assignment

Back to Contents

Locate annual data for U.S. New Houses Sold and New-home Mortgage Yields. I suggest using the following:

1. *Economic Report of the President: 2013*, "Table B–56. New private housing units started, authorized, and completed and houses sold". Use the "New houses sold" column.
2. *Economic Report of the President: 2013*, "Table B–73. Bond yields and interest rates". Use the "New home mortgage yields" column.

Documentation

Back to Contents

Abstract

Back to Contents

This lab helped me to learn how different markets and variables affect the housing market. The number of houses sold and housing projects started yearly are dependent upon variables like the mortgage yields, unemployment, and others. If the economy is perceived to be strong, the number of houses sold will increase, and if it is perceived as weak the number of houses sold will decrease.

Data Dictionary

Back to Contents

Variable	Values	Source	Mnemonic
New Houses Sold	Thousands	Economic Report of the President, 2013: Table B-56	New_Hous_sold
New Houses Started	Thousands	Economic Report of the President, 2013: Table B-56	rgdp
Mortgage Rate	Percentage	Economic Report of the President, 2013: Table B-73	MORT
Unemployment	Percentage	Economic Report of the President, 2013: Table B-43	UNEMP

Pre-lab Questions

Back to Contents

Before you do any work, please think about the relationship among these macro variables. In particular, think how you would answer the following if called on in class.

What type of data is this and why (i.e., source and domain)?

There are following three datasets used and they have been taken from *Economic Report of the President: 2013* available from https://www.govinfo.gov/app/details/ERP-2013/context:

1. "Table B–56. New private housing units started, authorized, and completed and houses sold". We use the "New houses sold" column.
2. "Table B–73. Bond yields and interest rates". We use the "New home mortgage yields" column.
3. "Table B–43. Civilian unemployment rate by demographic characteristic, 1972–2012". We use the Unemployment Rates for "All civilian workers" column.

We use this data because we want to find which economic variables will affect the annual new houses sold in the United States. We also believe that this is data published by the Government, so it must be accurate to the extent possible.

What is the mortgage yield? Do not tell me that it's the yield on mortgage backed securities. Think about the problem and then answer this question.

Morgage yield, or mortgage rates, is the average fixed rate of interest paid by borrowers on housing loans. The mortage rates we will be using in this lab is for 10-year term mortgages.
For the future state, if there is a concern that the economy will go into a recession soon, people will be less willing to buy a new home for fear of losing their job, so housing sales will be weak or decline. But if the economy looks promising, then housing sales will be strong or increase. Mortgage Rates are an indicator of the future of the economy. The definition in the data source is a follows: "*Effective rate (in the primary market) on conventional mortgages, reflecting fees and charges as well as contract rate and assuming, on the average, repayment at end of 10 years. Rates beginning with January 1973 not strictly comparable with prior rates.*"

How should mortgage yields or rates affect housing? Positively? Negatively? Explain.

Since the morgage rates affect the interest payments by borrowers, it adds to the total cost of ownership of housing. Hence we expect that higher the mortgage rates, higher are the interest payments on housing loans, the lower is the incentive to buy homes. Hence higher mortgage rates should be associated with lower new home sales.

In general, how do you think the housing market has behaved over the past, say, decade? Explain.

Housing sales were at its peak in 2005-2007 when the US economy was in the midst of a real-estate bubble. With the financial crisis of 2008, this bubble had burst, and housing prices and sales went down considerbaly in 2008-2010. Since 2011, the both housing prices and sales have been going up steadily.

Write a tentative <u>specific</u> model. Explain your model.

New houses sold annualy can be realted to mortgage rates using a linear model as follows.
With,
n = Number of observations,
i = i'th observation,
HOUS = New houses sold (in thousands),
MORT = Mortgage rates (in percentage points, e.g. 4.00% as 4.00),
ϵ = residual or disturbances,

${HOUS_i}= \hat{\beta_0} + \hat{\beta_1} MORT_i + \epsilon_i$

i= 1, ..., n

Here, parameter $\hat{\beta_0}$ represents the estimate of intercept, and parameter $\hat{\beta_1}$ represents the estimate of slope in the lienar model. We expect a negative relationship between HOUS and MORT, meaning a negative value for $\hat{\beta_1}$ from regression.

What is a good testable hypothesis? Explain your testable hypothesis.

For the linear model above, I expect a negative relationship between new houses sold (HOUS) and mortgage rates (MORT). This is because I expect that the higher the mortgage rates, the higher the interest payments on housing loans, the lower is the incentive to buy new homes, hence there would be a lower number of new houses sold.

Write the statistical hypothesis to go along with your testable hypothesis. Explain what you wrote.

For the linear model above, I expect the negative relationship between HOUS and MORT, hence the statistical hypotheses for the linear model are:

Null hypothesis: $H_0 = \hat{\beta_1} = 0$, and
Alternative hypothesis: $H_a = \hat{\beta_1} < 0$.

In the Null hypothesis, $\hat{\beta_1} = 0$, means that there is no relationship between new houses sold and mortgage rates. While the alternative hypthesis, $\hat{\beta_1} = 0$, states that there is a negative relationship between new housing sold and morgage rates.

Tasks and Questions

Back to Contents

Load the Pandas, Seaborn, and Statsmodels packages and give them aliases. You will also need the Statsmodels formula API for formulas. Please see Lesson #4 for examples.

In [1]:
```
## Analysis
import pandas as pd
import numpy as np
import statsmodels.api as sm
import statsmodels.formula.api as smf
import statsmodels.stats.weightstats as ztest
from statsmodels.iolib.summary2 import summary_col
## Visualization
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
## sns.set()
sns.set(rc={'figure.figsize':(11.7,8.27)})
```

Import the data. Set the row index to the years.

In [7]:
```
## Read Housing File
data_path = '/Users/maanyatandon/Documents/fall2020/econometrics/Lab5/'
dfh    = pd.read_excel( data_path+'ERP-2013-table56.xls', sheet_name='B56_mod')
dfh.set_index('Year', inplace=True)
dfh.head()
```

Out[7]:

Year	New_hous_start	New_1	New_2	New_5	New_auth_tot	New_auth_1	New_auth_2	New_auth_5	New_comp	New_Hous_Sold
1967.0	1291.6	843.9	71.7	376.1	1141.0	650.6	73.0	417.5	......	487
1968.0	1507.6	899.4	80.7	527.3	1353.4	694.7	84.3	574.4	1319.8	490
1969.0	1466.8	810.6	85.1	571.2	1322.3	624.8	85.2	612.4	1399	448
1970.0	1433.6	812.9	84.9	535.9	1351.5	646.8	88.1	616.7	1418.4	485
1971.0	2052.2	1151.0	120.5	780.9	1924.6	906.1	132.9	885.7	1706.1	656

In [3]:
```
## Read Mortgage Rates File
dfm    = pd.read_excel( data_path+'ERP-2013-table73.xls', sheet_name='B73p1_mod')
dfm.set_index('Year', inplace=True)
dfm.head()
```

Out[3]:

Year	T3M	T6M	T3Y	T10Y	T30Y	Corp_aaa	Corp_baa	Muni	MORT	Prime	Disc_Primary	Disc_Adj	Fed_funds
1941.0	0.103	......	......	......	......	2.77	4.33	2.10	......	1.5	......	1	......
1942.0	0.326	......	......	......	......	2.83	4.28	2.36	......	1.5	......	8 1.00	......
1943.0	0.373	......	......	......	......	2.73	3.91	2.06	......	1.5	......	8 1.00	......
1944.0	0.375	......	......	......	......	2.72	3.61	1.86	......	1.5	......	8 1.00	......
1945.0	0.375	......	......	......	......	2.62	3.29	1.67	......	1.5	......	8 1.00	......

Print the first five (5) records.

In [9]:
```
## Merge the two data sets into final DataFrame df
## df_all = pd.merge( df_SPX, df_AAA,    how = 'inner', left_index = True, right_index = True)
df = pd.merge( dfh, dfm, how = 'inner', left_index = True, right_index=True)
df.head()
```

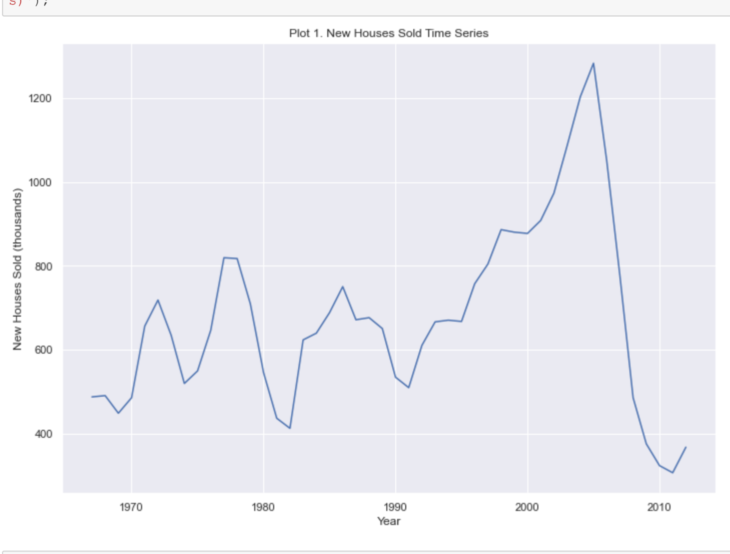
Out[9]:

Year	New_hous_start	New_1	New_2	New_5	New_auth_tot	New_auth_1	New_auth_2	New_auth_5	New_comp	New_Hous_Sold	...
1967.0	1291.6	843.9	71.7	376.1	1141.0	650.6	73.0	417.5	......	487	...
1968.0	1507.6	899.4	80.7	527.3	1353.4	694.7	84.3	574.4	1319.8	490	...
1969.0	1466.8	810.6	85.1	571.2	1322.3	624.8	85.2	612.4	1399	448	...
1970.0	1433.6	812.9	84.9	535.9	1351.5	646.8	88.1	616.7	1418.4	485	...
1971.0	2052.2	1151.0	120.5	780.9	1924.6	906.1	132.9	885.7	1706.1	656	...

5 rows × 23 columns

Graph the data.

In [10]:
```
## Plot 1: 'New Houses Sold' vs time
ax = sns.lineplot( y='New_Hous_Sold', x=dfh.index, data=dfh )
ax.set( title='Plot 1. New Houses Sold Time Series', xlabel= 'Year', ylabel= 'New Houses Sold (thousand
s)');
```



In [11]:
```
## Plot 2: 'New Housing Started' vs time
ax = sns.lineplot( y='New_hous_start', x=dfh.index, data=dfh )
ax.set( title='Plot 2. New Housing Started Time Series', xlabel= 'Year', ylabel= 'New Housing Started
 (thousands)');
```

In [15]:
```
## Plot 3: 'Mortgage Rates' vs time
df['MORT'] = df['MORT'].astype(str).astype(float)
ax = sns.lineplot( y='MORT', x=df.index, data=df )
ax.set( title='Plot 3. Mortgage Rates Time Series', xlabel= 'Year', ylabel= 'Mortgage Rate (%)');
```

In [16]:
```
## Plot 4: New Housing Sold vs Mortgage Rate
ax = sns.relplot( y='New_Hous_Sold', x='MORT', hue=df.index, data=df)
ax.set( title='Plot 4. New Housing Sold vs Mortgage Rate', xlabel= 'Mortgage Rate (%)', ylabel= 'New Ho
uses Sold (thousands)');
```

In [17]:
```
# Plot 5: New Housing Started vs Mortgage Rate
ax = sns.relplot( y='New_hous_start', x='MORT', hue=df.index, data=df)
ax.set( title='Plot 5. New Housing Started vs Mortgage Rate', xlabel= 'Mortgage Rate (%)', ylabel= 'New
 Housing Started (thousands)');
```

When we build a regression model, we say we regress the dependent variable on the independent variable(s). For this lab, you will regress sales on yield. Estimate an OLS model using sales as the dependent variable and yield as the independent variable. Display the summary report.

In [18]:
```
## Step 1: Formula = 'Y ~ X1 + X2 + X3'
formula = 'New_Hous_Sold ~ MORT'

## Step 2: Initialized 'model'
mod = smf.ols(formula, data=df)

## Step 3: Fit
reg01 = mod.fit()

## Step 4: Summarize the fitted model
print(reg01.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          New_Hous_Sold   R-squared:                       0.034
Model:                            OLS   Adj. R-squared:                  0.012
Method:                 Least Squares   F-statistic:                     1.542
Date:                Sun, 01 Nov 2020   Prob (F-statistic):              0.221
Time:                        15:39:58   Log-Likelihood:                -312.24
No. Observations:                  46   AIC:                             628.5
Df Residuals:                      44   BIC:                             632.1
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    810.8574    113.863      7.121      0.000     581.381    1040.333
MORT         -16.3002     13.125     -1.242      0.221     -42.753      10.152
==============================================================================
Omnibus:                        1.542   Durbin-Watson:                   0.222
Prob(Omnibus):                  0.463   Jarque-Bera (JB):                0.836
Skew:                           0.301   Prob(JB):                        0.658
Kurtosis:                       3.272   Cond. No.                         30.9
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Retrieve and display the estimated parameters.

In [19]:
```
print('Estimated Params:\n{}'.format(reg01.params))
print('Number of observations:{}'.format(reg01.nobs))
```

```
Estimated Params:
Intercept    810.857373
MORT         -16.300155
dtype: float64
Number of observations:46.0
```

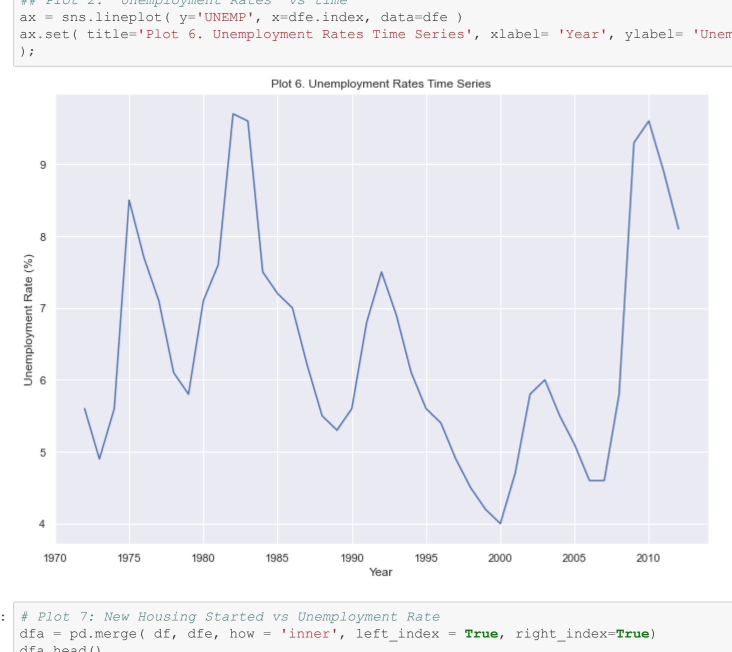
Estimate a yield elasticity.

In [21]:
```
average_mortgage_rates = df['MORT'].mean()
average_new_home_sold  = df['New_Hous_Sold'].mean()
elasticity1 = (average_mortgage_rates / average_new_home_sold) * reg01.params[1]
print('average_mortgage_rates:',average_mortgage_rates)
print('average_new_home_sold:', average_new_home_sold)
print('elasticity1:',elasticity1)
##
elasticity2 = (100.0 / average_new_home_sold) * reg01.params[1]
print('elasticity2:',elasticity2)
```

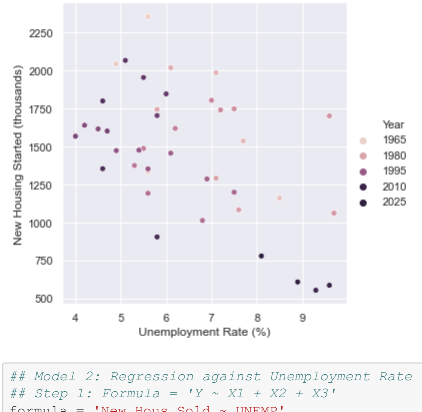
```
average_mortgage_rates: 8.317391304347826
average_new_home_sold: 675.2826086956521
elasticity1: -0.2007674455526804
elasticity2: -2.413827102828881
```

Build a Model Portfolio.

In [22]:
```
## Model2: Read Unemployment Rate File
dfe    = pd.read_excel( data_path+'ERP-2013-table43.xls', sheet_name='B43_mod')
dfe.set_index('Year', inplace=True)
dfe.head()
## Plot 2: 'Unemployment Rates' vs time
ax = sns.lineplot( y='UNEMP', x=dfe.index, data=dfe )
ax.set( title='Plot 6. Unemployment Rates Time Series', xlabel= 'Year', ylabel= 'Unemployment Rate (%)'
);
```



In [23]:
```
# Plot 7: New Housing Started vs Unemployment Rate
dfa = pd.merge( df, dfe, how = 'inner', left_index = True, right_index=True)
dfa.head()
ax = sns.relplot( y='New_hous_start', x='UNEMP', hue=dfa.index, data=dfa)
ax.set( title='Plot 7. New Housing Started vs Unemployment Rate', xlabel= 'Unemployment Rate (%)', ylab
el= 'New Housing Started (thousands)');
```



In [30]:
```
## Model 2: Regression against Unemployment Rate
## Step 1: Formula = 'Y ~ X1 + X2 + X3'
formula = 'New_Hous_Sold ~ UNEMP'

## Step 2: Initialized 'mod'
mod2 = smf.ols(formula, data=dfa)

## Step 3: Fit
reg02 = mod2.fit()

## Step 4: Summarize the fitted model
print(reg02.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          New_Hous_Sold   R-squared:                       0.436
Model:                            OLS   Adj. R-squared:                  0.422
Method:                 Least Squares   F-statistic:                     30.15
Date:                Sun, 01 Nov 2020   Prob (F-statistic):           2.64e-06
Time:                        15:46:43   Log-Likelihood:                -267.93
No. Observations:                  41   AIC:                             539.9
Df Residuals:                      39   BIC:                             543.3
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   1301.5480    113.640     11.453      0.000    1071.689    1531.407
UNEMP        -94.3699     17.188     -5.491      0.000    -129.135     -59.605
==============================================================================
Omnibus:                        7.716   Durbin-Watson:                   0.583
Prob(Omnibus):                  0.021   Jarque-Bera (JB):                6.528
Skew:                           0.909   Prob(JB):                       0.0382
Kurtosis:                       3.720   Cond. No.                         28.8
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

In [27]:
```
## Model 3: Regression against Mortgage Rates and Unemployment Rate
## Step 1: Formula
formula = 'New_Hous_Sold ~ MORT + UNEMP'

## Step 2: Initialized 'mod'
mod3 = smf.ols(formula, data=dfa)

## Step 3: Fit
reg03 = mod3.fit()

## Step 4: Summarize the fitted model
print(reg03.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          New_Hous_Sold   R-squared:                       0.440
Model:                            OLS   Adj. R-squared:                  0.411
Method:                 Least Squares   F-statistic:                     14.94
Date:                Sun, 01 Nov 2020   Prob (F-statistic):           1.63e-05
Time:                        15:46:03   Log-Likelihood:                -267.78
No. Observations:                  41   AIC:                             541.6
Df Residuals:                      38   BIC:                             546.7
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   1334.9164    130.412     10.236      0.000    1070.912    1598.921
MORT          -5.7901     10.771     -0.538      0.594     -27.596      16.015
UNEMP        -91.9772     17.908     -5.136      0.000    -128.231     -55.724
==============================================================================
Omnibus:                        6.117   Durbin-Watson:                   0.558
Prob(Omnibus):                  0.047   Jarque-Bera (JB):                4.932
Skew:                           0.813   Prob(JB):                       0.0849
Kurtosis:                       3.495   Cond. No.                         53.2
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
In [31]: ## create summary table
model_names = [ 'Model ' + str(i) for i in range(1, 6)]
#create a variable to hold the statistics to print; this is a dictionary
info_dict = {'\nn': lambda x: "{0:d}".format(int(x.nobs) ),
             'R2 Adjusted': lambda x: "{:0.3f}".format(x.rsquared_adj ),
             'AIC': lambda x : "{:0.2f}".format(x.aic),
             'F': lambda x: "{0.2f}".format(x.fvalue),
}
#create the portfolio summary table
summary_table = summary_col([reg01, reg02, reg03],
                float_format = '%0.2f',
                model_names = model_names,
                stars = True,
                info_dict = info_dict
)
summary_table.add_title('Summary Table for New Houses Sold Models')
print(summary_table)
```

```
 Summary Table for New Houses Sold Models
==========================================
              Model 1   Model 2    Model 3
------------------------------------------
Intercept   810.86*** 1301.55*** 1334.92***
            (113.86)  (113.64)   (130.41)
MORT        -16.30               -5.79
            (13.13)              (10.77)
R-squared   0.01      0.42       0.41
            0.03      0.44       0.44
UNEMP                 -94.37***  -91.98***
                      (17.19)    (17.91)

n          46        41         41
R2 Adjusted 0.012     0.422      0.411
AIC         628.49    539.87     541.56
F
==========================================
Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01
```

Post-lab Questions

Back to Contents

What is the relationship between home sales and mortgage yields? Is this the relationship you expected? Plot the data as part of your analysis.

The coefficient of regression, β_1, for mortgage rates is -16.3002. This means that for each percentage point increase in mortage rates, annual new home sales in the US decline by 16,300. Since higher the mortgage rates, higher are the interest payments on housing loans; the higher morgage rates add to the total cost of buying the house. So the result is in line with the expectation that with incresing mortgage rates - the incentive to buy homes decreases. Below is the Scatter Plot and Regression Line for New Housing Sold vs Mortgage Rates.

```
In [33]: ax = sns.regplot(x='MORT', y='New_Hous_Sold', color="b", data=df)
ax.set( title='Plot 8: Scatter Plot and Regression Line for New Housing Sold vs Mortgage Rates', xlabel
= 'Mortgage Rate (%)', ylabel= 'New Housing Sold (thousands)');
```

Test your statistical hypothesis about the effect of mortgage rates on new home sales. What do you conclude?

The Null Hypothesis is $\beta_1 = 0$. From the regression, the t-value and p-value associate with this null hypotheis are -1.242 and 0.221 respectively. Since p-value of 0.221 is much higher than the significance level associated with $\alpha = 0.05$, specifically one-side tail test with $\alpha/2 = 0.025$, we cannot reject the Null Hypothesis, $\beta_1 = 0$. This mean that from the data, it cannot be said for sure that the relationship between the new housing sold and mortgage rates is negative.

Interpret the R^{2} from the regression output. What does it say about your model?

The R^2 from regression is 0.034 - a very low number close to zero. This means that SSE is 96.6% (1-0.034) of SST, i.e. the regression is dominated by errors. Only 3.4% of the variation in new houses sold is accounted for by the Morgage Rates. Since, 96.6% of the variation in not accouned for by the model, we cannot be confident of the proposed linear model.

It can be seen from the plots of the regression line above that in the last few years of the data, the observations for new houses sold are way off the regression line. This means, that we need to look for the causes, and maybe use additional explanatory variables for new houses sold.

When we include Unemployment Rates (UNEMP) as an explanatory variable in the regression, the R^2 increases substantially to 0.42.

How can you make the model better? What additional variables can you think of that should be included? Defend your answer.

Most people take housing loans for 15 or 30 years. Hence a 15-year or 30-year term mortgage rates could be used, instead of 10-year rates.

Secondly, if people are unemployed, they would not be able to get the housing loan, no matter how low the mortgage rates are. Hence in the second model, above, we propose a model that includes Unemployment Rates besides Morgage Rates - resulting in an R^2 of 0.440 and Adjusted R^2 of 0.411 - much higher than the R^2 of 0.034 for the original model. In fact, Unemployment Rate alone as an independent variable is a better predictor of new housing sold ($R^2=0.436$) than Mortgage Rates.

Third, the slope of the treasury bill/bond yield-curve tells us the expectation of good economic times vs recessions. Hence another variable that can be considered is the spread between short term treasury yield (say 3-month term) and long term yield (30-year). I expect higher yield curve slope to be associated with higher new home sales.

Fourth, consumer confidence index published by the Conference Board is another indicator of the expectation of future economic conditions. Potentially, it could be explored as an explanatory variable.

Fifth, if house prices increase, then many potential buyers would postpone the decision to buy the house - thus reduced housing sales. Hence year-over- year percentage change in average housing prices could be used as an additional explanatory variable for housing sold.

Interpret the F-statistic from the regression output. What hypothesis does it test? What do you conclude about your model based on the F-statistic?

The regression results in F-statistic of 1.542, while the critical value of F-statistic with 95% confidence with 1 and 44 (=n-2) degrees of is calculated as 4.06 as below. F-statistic calculated from regression is MSR/MSE. For a good model, we ought to have mean of squared errors (MSE) much lower than mean squared from regression (MSR), or calculated F-statisic must be much larger than 1. F-statistic tests the Null Hypothesis that naive model is true, i.e. $\beta_1=0$. Since the calculated regression F-statistic of 1.542 is much lower than the critical F value of 4.06, we cannot reject the Null hypothesis that $\beta_1=0$.

```
In [34]: import scipy.stats
scipy.stats.f.ppf(q=1-0.05, dfn=1, dfd=(reg01.nobs-2))
# scipy.stats.f.cdf(Fcrit, dfn=1, dfd=(reg01.nobs-2))
```

```
Out[34]: 4.061706460119338
```

Are existing home sales elastic or inelastic with respect to the mortgage yield? Does this make sense? Is it what you expect? Defend your answer. Suppose there is a 1\% increase in mortgage rates. How much do home sales change in percentage terms?

Since in this lab, we analyzed 'new' homes sold and did not analyze 'existing' home sales - we will answer for 'new' home sales.

Elasticity of new home sales with respect to mortgage rates is calculated below as:
Elasticity1 = $\eta^H_M = $ (average mortgage rates) / (average new home sold) X $\beta_1 = -0.1974$

This means that when mortgage rates change by 1%, for example from 4% to 4.04%, new houses sold will decrease by about 0.2%. Hence new home sales are inelastic on this measure of elasticity.

However, when elasticity of new home sales with respect to mortgage rates is defined as:
Elasticity2 = $\eta^H_M = $ 100 / (average new home sold) X $\beta_1 = -2.345$

which measure the %-change in new home sales when morgarge rates change by 1 percentage *point*, for example from 4% to 5%. This means that when mortgage rates change by 1 percentage point, for example from 4.00% to 5.00%, new houses sold will decrease by about 2.345%. By this measure, the new home sales are elastic with respect to mortgage yield.

Obviously the two definitions of elasticity give two different answers!

```
In [35]: average_mortgage_rates = dfa['MORT'].mean()
average_new_home_sold  = dfa['New_Hous_Sold'].mean()
elasticity1 = (average_mortgage_rates / average_new_home_sold) * reg01.params[1]
print('average_mortgage_rates:',average_mortgage_rates)
print('average_new_home_sold:', average_new_home_sold)
print('elasticity1:',elasticity1)
##
elasticity2 = (100.0 / average_new_home_sold) * reg01.params[1]
print('elasticity2:',elasticity2)
```

```
average_mortgage_rates: 8.418780487804879
average_new_home_sold: 695.0487804878048
elasticity1: -0.19743567299803524
elasticity2: -2.345181386829517
```

What is the practical significance of your model? In other words, what can this be used for, it for anything at all? How would you answer the question: "*So what?*" Defend your answer.

Because of the low R^2, and also because we could not reject the null hypothesis -- we are not entirely sure if the negative relationship between mortgage rates and new houses sold entirely holds. Hence the model cannot be used as such.

The model establishes the fact that mortgage rates cannot entirely explain the annual new houses sold. Hence, we should explore other explanatory economic variables such as unemployment rates, consumer confidence index, annual average percentage-increase in house prices, and yield curve slope etc. to build a satifactory model.

Secondly, we could combine the new houses sold with existing houses sold, i.e. total number of houses sold annually - which may have a stronger relationship with explanatory variables.

Well done!

Make sure your name is on this notebook at the top and on the file.
Please submit this notebook as a PDF file. Nothing else will be accepted.

```
In [ ]:
```