A4_output

April 4, 2022

Run the cell below:

0.1 Problem 1

0.1.1 (10 x 2 = 20 points)

The Fama-French five-factor model adds two new factors to the three-factor model: the RMW (robust minus weak) profitability factor and the CMA (conservative minus aggressive) investment factor.

1. Download monthly returns from 1970 to 2020 (inclusive) for the Fama-French five factor model. This is the `F-F_Research_Data_5_Factors_2x3` file in the `famafrench` database (the first table in the dictionary returned by pandas_datareader gives you the monthly returns). Call this dataframe `ff5f`. Divide all returns in this dataset by 100. Print the first 5 and last 5 rows in this dataframe.
2. Download monthly adjusted-close prices on AAPL from Yahoo Finance, from 2010 to 2020 (inclusive). Use these prices to calculate monthly returns. Call the returns dataframe `aapl_ret`. Print the first 5 and last 5 rows in this dataframe.
3. Merge the `ff5` and `aapl_ret` dataframes using an inner join on date. Call this merged dataframe `data`. Add a variable in this dataframe called `const` which equals 1 everywhere. Print the first 5 and last 5 rows of `data`.
4. Regress monthly AAPL excess returns from 2010 to 2020 (inclusive) on the Fama-French five factors (Mkt-RF, SMB, HML, RMW, and CMA) and a constant. Print the regression table.

Use the print function to provide answers to the following questions (i.e. your answer should be the argument to the print function). All these answers should be based on the results obtained from the regression above.

5. Is AAPL mispriced with respect to the five-factor model at the 5% significance level? If so, is it overpriced or underpriced? How did you reach your conclusion?
6. Does AAPL have a significant exposure (at the 5% level) to either of the five factors? If so, which ones? How did you reach your conclusion?
7. Using the five-factor model, what percentage of AAPL total risk is systematic?
8. Using the five-factor model, what excess return do we expect AAPL to have if the returns on the market, SMB, HML, RMW, and CMA factors are 0.5% (i.e. 0.005), 0.2%, 0.3%, 0.6% and -0.1% respectively?
9. Use data from 1970 to 2020 to estimate the risk premia (expected excess returns) on the five factors. Print these out.
10. Using the estimated risk premia from point 9 above, calculate the risk premium on AAPL stock assuming the true alpha of AAPL is 0.

Output for part 1:

```
         Mkt-RF     SMB     HML     RMW     CMA      RF
Date
1970-01 -0.0810  0.0313  0.0312 -0.0171  0.0385  0.0060
1970-02  0.0513 -0.0274  0.0393 -0.0232  0.0274  0.0062
1970-03 -0.0106 -0.0240  0.0399 -0.0101  0.0432  0.0057
1970-04 -0.1100 -0.0637  0.0617 -0.0069  0.0626  0.0050
1970-05 -0.0692 -0.0445  0.0332 -0.0125  0.0392  0.0053
…          …      …      …      …      …      …
2020-08  0.0763 -0.0087 -0.0293  0.0427 -0.0130  0.0001
2020-09 -0.0363 -0.0007 -0.0266 -0.0129 -0.0177  0.0001
2020-10 -0.0210  0.0467  0.0419 -0.0093 -0.0073  0.0001
2020-11  0.1247  0.0706  0.0199 -0.0217  0.0132  0.0001
2020-12  0.0463  0.0475 -0.0156 -0.0191 -0.0015  0.0001

[612 rows x 6 columns]
```

Output for part 2:

```
[*********************100%***********************]  1 of 1 completed

             AAPL
Date
2010-02  0.065396
2010-03  0.148470
2010-04  0.111022
2010-05 -0.016125
2010-06 -0.020827
…            …
2020-08  0.214379
2020-09 -0.100908
2020-10 -0.060012
2020-11  0.093606
2020-12  0.116497

[131 rows x 1 columns]
```

Output for part 3:

```
             AAPL  Mkt-RF     SMB     HML     RMW     CMA      RF  const
Date
2010-02  0.065396  0.0340  0.0151  0.0322 -0.0028  0.0140  0.0000      1
2010-03  0.148470  0.0631  0.0185  0.0221 -0.0063  0.0167  0.0001      1
2010-04  0.111022  0.0200  0.0498  0.0289  0.0070  0.0174  0.0001      1
2010-05 -0.016125 -0.0789  0.0004 -0.0244  0.0127 -0.0023  0.0001      1
2010-06 -0.020827 -0.0557 -0.0247 -0.0470 -0.0018 -0.0155  0.0001      1
…            …      …      …      …      …      …      …      …
```

```
 2020-08  0.214379  0.0763 -0.0087 -0.0293  0.0427 -0.0130  0.0001      1
 2020-09 -0.100908 -0.0363 -0.0007 -0.0266 -0.0129 -0.0177  0.0001      1
 2020-10 -0.060012 -0.0210  0.0467  0.0419 -0.0093 -0.0073  0.0001      1
 2020-11  0.093606  0.1247  0.0706  0.0199 -0.0217  0.0132  0.0001      1
 2020-12  0.116497  0.0463  0.0475 -0.0156 -0.0191 -0.0015  0.0001      1

 [131 rows x 8 columns]
```

```
Output for part 4:
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.456
Model:                            OLS   Adj. R-squared:                  0.434
Method:                 Least Squares   F-statistic:                     20.96
Date:                Mon, 04 Apr 2022   Prob (F-statistic):           3.62e-15
Time:                        12:40:22   Log-Likelihood:                 188.04
No. Observations:                 131   AIC:                            -364.1
Df Residuals:                     125   BIC:                            -346.8
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0087      0.006      1.565      0.120      -0.002       0.020
Mkt-RF         1.1960      0.138      8.637      0.000       0.922       1.470
SMB           -0.1328      0.252     -0.526      0.600      -0.632       0.367
HML           -0.3482      0.248     -1.402      0.164      -0.840       0.143
RMW            1.3023      0.362      3.601      0.000       0.587       2.018
CMA           -0.7940      0.421     -1.886      0.062      -1.627       0.039
==============================================================================
Omnibus:                        4.396   Durbin-Watson:                   1.837
Prob(Omnibus):                  0.111   Jarque-Bera (JB):                4.153
Skew:                          -0.287   Prob(JB):                        0.125
Kurtosis:                       3.657   Cond. No.                         87.3
==============================================================================

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

```
Output for part 5:
Question 5 answer: …
```

```
Output for part 6:
Question 6 answer: …
```

```
Output for part 7:
Question 7 answer: …
```

```
Output for part 8:
Conditional prediction =  0.02198378176014695
```

```
Output for part 9:

 Mkt-RF    0.005926
 SMB       0.001631
 HML       0.002530
 RMW       0.002662
 CMA       0.002972
 RF        0.003723
 dtype: float64
```

```
Output for part 10:
Unconditional prediction =  0.007097167517847282
```

0.2 Problem 2

0.2.1 (5 x 8 = 40 points)

For this problem, you will be testing if returns are predictable at the annual level.

Data: 1. Load the CRSP dataset from the "crspm.zip" file and keep only the "permno", "date", and "ret" variables. Drop all rows with any missing values. Call the resulting dataframe "crsp" and print its first two rows. 1. In "crsp", create a new variable "year" that extracts the year from the "date" variable (as an integer, not a period date). Print a table that shows only the min and max of the "year" variable. 2. Calculate annual compounded returns for each firm, each year, and store those in a separate dataframe called "crsp_an". Print the first 5 rows of "crsp_an".

2. Load the Compustat dataset from the "compa.zip" file and keep only the "permno", "datadate", "sich" (industry identifier) and "at" (total assets) variables. Drop all rows with any missing values in this new dataframe. Keep only rows with strictly positive total assets ("at"). Call the resulting dataframe "comp".
    1. In "comp", create a new variable "year" that extracts the year from the "datadate" variable (as an integer, not a period date). Print a table that shows only the min and max of the "year" variable.
    2. In "comp", create a new variable "ag" as the annual percentage growth in total assets. Winsorize it at the 1 and 99 percentiles and call this variable "ag_w". Print a table that shows the mean and standard deviation of "ag" and "ag_w".
    3. In "comp", create a new variable "sic3" that contains the first three digits of the "sich" variable
3. Merge the resulting "comp" dataframe with the "crsp_an" dataframe based on "permno" and "year". Call the merged dataset "data" and print out its shape.
    1. Create a new variable called "future_ret" that, for every firm, every year, stores the annual returns of that firm in the following year. Print the first 5 rows of this dataframe.
    2. Set the index of "data" to be ("permno", "year"). Keep only then "future_ret", "ag_w", and "sic3" variables. Create a new variable called "const" that equals 1 everywhere. Print out the first 2 rows of this dataset.

Analysis:

4. Regress "future_ret", on current winsorized asset growth ("ag_w") and a constant. This should be simple OLS, with no adjustments for fixed effects or corrections to standard errors. Print your results (the regression table)
    1. Use the print function to answer the following question: "Based on these regression results, does current asset growth have statistically significant explanatory power over future returns? How did you reach that conclusion?"
5. Run the same regression as above, only this time add time fixed effects and industry ("sic3") fixed effects, and cluster your standard errors at the firm and year level (ignore the warnings produced in this step, they are harmless). Print your results (the regression table)
    1. Use the print function to answer the following question: "Based on these regression results, does current asset growth have statistically significant explanatory power over future returns? How did you reach that conclusion?"

```
Output for part 1:

    permno        date       ret
1  10000.0  1986-02-28 -0.257143
2  10000.0  1986-03-31  0.365385

Output for part 1.A:

 min    1980
 max    2020
 Name: year, dtype: int64

Output for part 1.B:

                  anret
 permno  year
 10000.0 1986  0.117857
         1987  0.424242
 10001.0 1986  1.217368
         1987  0.898725
         1988  1.163160

Output for part 2:

     permno    datadate    sich      at
 15  10031.0  1988-01-31  5712.0  16.042
 16  10031.0  1989-01-31  5712.0  16.280

Output for part 2.A:

 min    1982
 max    2020
 Name: year, dtype: int64

Output for part 2.B:
```

```
               ag      ag_w
mean     2.706688  0.143621
std   1037.143277  0.451803
```

Output for part 2.C:

```
         permno    datadate    sich      at  year  at_lag1        ag  \
188566  10001.0  1987-06-30  4924.0  11.771  1987      NaN       NaN
188567  10001.0  1988-06-30  4924.0  11.735  1988   11.771 -0.003058
188568  10001.0  1989-06-30  4924.0  18.565  1989   11.735  0.582020
188569  10001.0  1990-06-30  4924.0  18.881  1990   18.565  0.017021
188570  10001.0  1991-06-30  4924.0  19.599  1991   18.881  0.038028

            ag_w sic3
188566       NaN  492
188567 -0.003058  492
188568  0.582020  492
188569  0.017021  492
188570  0.038028  492
```

Output for part 3:

```
(156258, 10)
```

Output for part 3.A:

```
    permno    datadate    sich      at  year  at_lag1        ag      ag_w  \
0  10001.0  1987-06-30  4924.0  11.771  1987      NaN       NaN       NaN
1  10001.0  1988-06-30  4924.0  11.735  1988   11.771 -0.003058 -0.003058
2  10001.0  1989-06-30  4924.0  18.565  1989   11.735  0.582020  0.582020
3  10001.0  1990-06-30  4924.0  18.881  1990   18.565  0.017021  0.017021
4  10001.0  1991-06-30  4924.0  19.599  1991   18.881  0.038028  0.038028

  sic3     anret  future_ret
0  492  0.898725    1.163160
1  492  1.163160    1.687923
2  492  1.687923    0.991279
3  492  0.991279    1.607471
4  492  1.607471    1.012620
```

Output for part 3.B:

```
              future_ret      ag_w sic3  const
permno  year
10001.0 1987    1.163160       NaN  492      1
        1988    1.687923 -0.003058  492      1
```

Output for part 4:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             future_ret   R-squared:                       0.004
Model:                            OLS   Adj. R-squared:                  0.004
Method:                 Least Squares   F-statistic:                     515.0
Date:                Mon, 04 Apr 2022   Prob (F-statistic):          8.59e-114
Time:                        12:40:25   Log-Likelihood:            -1.7249e+05
No. Observations:              125401   AIC:                         3.450e+05
Df Residuals:                  125399   BIC:                         3.450e+05
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.1903      0.003    416.810      0.000       1.185       1.196
ag_w          -0.1374      0.006    -22.695      0.000      -0.149      -0.126
==============================================================================
Omnibus:               259072.226   Durbin-Watson:                   2.071
Prob(Omnibus):              0.000   Jarque-Bera (JB):       5271391516.606
Skew:                      16.824   Prob(JB):                         0.00
Kurtosis:                1006.862   Cond. No.                         2.30
==============================================================================

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

Output for part 4.A:
Question 4.A answer: …

Output for part 5:

```
/home/imb/anaconda3/lib/python3.9/site-
packages/linearmodels/shared/exceptions.py:37: MissingValueWarning:
Inputs contain missing values. Dropping rows with missing observations.
  warnings.warn(missing_value_warning_msg, MissingValueWarning)

                          PanelOLS Estimation Summary
================================================================================
Dep. Variable:             future_ret   R-squared:                        0.0036
Estimator:                   PanelOLS   R-squared (Between):             -0.0149
No. Observations:              125401   R-squared (Within):               0.0072
Date:                Mon, Apr 04 2022   R-squared (Overall):              0.0041
Time:                        12:40:25   Log-likelihood                -1.678e+05
Cov. Estimator:             Clustered
                                        F-statistic:                      445.92
Entities:                       13205   P-value                           0.0000
Avg Obs:                       9.4965   Distribution:                F(1,125091)
Min Obs:                       1.0000
Max Obs:                       66.000   F-statistic (robust):             33.237
```

```
                                        P-value                       0.0000
Time periods:                      38   Distribution:              F(1,125091)
Avg Obs:                       3300.0
Min Obs:                       4.0000
Max Obs:                       5141.0

                             Parameter Estimates
==============================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
------------------------------------------------------------------------------
const          1.1885     0.0031     383.17     0.0000      1.1824      1.1945
ag_w          -0.1254     0.0217    -5.7652     0.0000     -0.1680     -0.0827
==============================================================================

F-test for Poolability: 31.536
P-value: 0.0000
Distribution: F(308,125091)

Included effects: Time, Other Effect (sic3)
Model includes 1 other effect
Other Effect Observations per group (sic3):
Avg Obs: 459.34, Min Obs: 0.0000, Max Obs: 9070.0, Groups: 273
```

```
Output for part 5.A:
Question 5.A answer: …
```

0.3 Problem 3

0.3.1 (10 x 4 = 40 points)

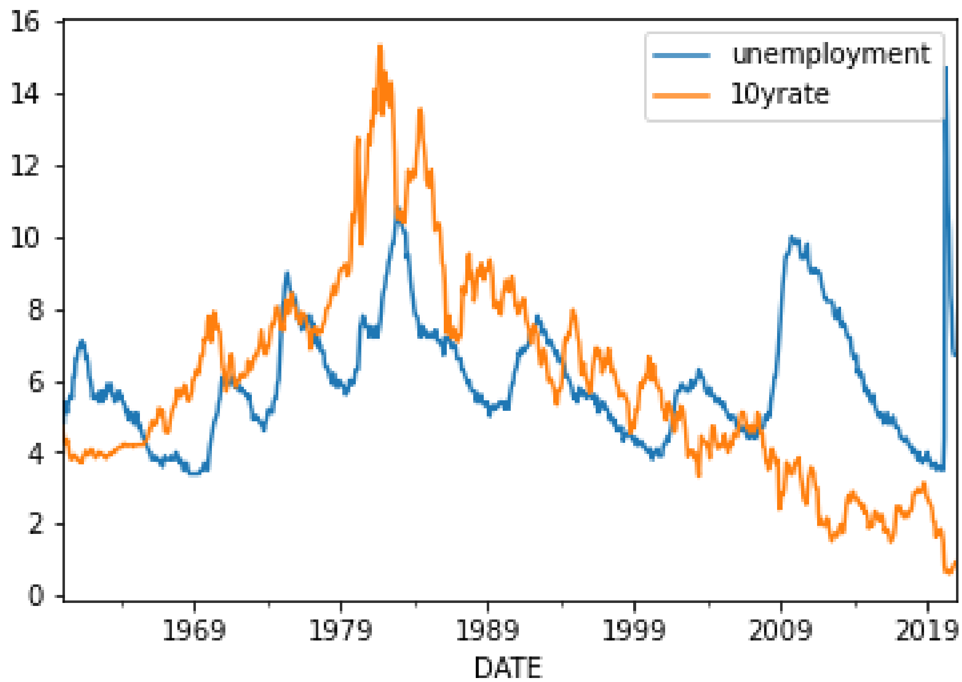
For this problem we will analyze the effect of tresury yields on future unemployment rates.

1. Download data on monthly yields on ten-year treasury bonds (`GS10`) and the unemployment rate (`UNRATE`) from the St Louis Fred from 1960 to 2020 (inclusive). Rename `GS10` to `10yrate` and `UNRATE` to `unemployment`. Plot these two variables on the same plot.
2. Create a variable called `future_unemployment` which gives us the unemployment rate from 12 months into the future. Create a variable called `const` which equals 1 everywhere. Print the first 5 and last 5 rows of this new version of your dataset.
3. Print the correlation between `future_unemployment` and `10yrate`, as well as the 12-month autocorrelations of ‘`future_unemployment` and `10yrate`.
4. Regress `future_unemployment` on `10yrate` and a constant (simple regression with no corrections). Print the regression table.
5. Run the same regression as in point 4 above, but this time use an HAC covariance estimator with maximum 5 lags. Print the regression table.
6. Use an augmented Dickey-Fuller test to test if you can reject the null hypotheses that `future_employment` and `10yrate` are non-stationary at the 99% confidence level.
7. Create a new variable called `unemployment_change` which equals `future_unemployment` minus its value from 12 months prior (i.e. the 12-month difference in `future_unemployment`). Create a new variable called `10yrate_change` which equals `10yrate` minus its value from 12

months prior. Plot these variables on the same graph.

8. Use an augmented Dickey-Fuller test to test if you can reject the null hypotheses that `unemployment_change` and `10yrate_change` are non-stationary at the 99% confidence level.
9. Regress `unemployment_change` on `10y_rate` and a constant. Use an HAC covariance estimator with maximum 5 lags. Print the regression table.
10. Based on the tests you performed above, do you conclude that treasury yields are a statistically significant predictor of future unemployment at the 99% confidence level?

```
Output for part 1:
```




```
Output for part 2:

              unemployment  10yrate  future_unemployment  const
DATE
1960-01-01             5.2     4.72                  6.6      1
1960-02-01             4.8     4.49                  6.9      1
1960-03-01             5.4     4.25                  6.9      1
1960-04-01             5.2     4.28                  7.0      1
1960-05-01             5.1     4.35                  7.1      1
…                        …        …                    …      …
2020-08-01             8.4     0.65                  NaN      1
2020-09-01             7.9     0.68                  NaN      1
2020-10-01             6.9     0.79                  NaN      1
2020-11-01             6.7     0.87                  NaN      1
2020-12-01             6.7     0.93                  NaN      1
```

```
[732 rows x 4 columns]

Output for part 3:
Correlation between 10-year rate and future unemployment =  0.34790650908116644
12-month autocorrelation in future unemployment =  0.7093256351836208
12-month autocorrelation in 10-year rate =  0.9244704162778316

Output for part 4:
                             OLS Regression Results
================================================================================
Dep. Variable:     future_unemployment   R-squared:                       0.121
Model:                             OLS   Adj. R-squared:                  0.120
Method:                  Least Squares   F-statistic:                     98.87
Date:                 Mon, 04 Apr 2022   Prob (F-statistic):           6.50e-22
Time:                         12:40:28   Log-Likelihood:                -1352.3
No. Observations:                  720   AIC:                             2709.
Df Residuals:                      718   BIC:                             2718.
Df Model:                            1
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.7717      0.138     34.660      0.000       4.501       5.042
10yrate        0.2044      0.021      9.944      0.000       0.164       0.245
==============================================================================
Omnibus:                      179.564   Durbin-Watson:                   0.078
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              437.278
Skew:                           1.296   Prob(JB):                     1.11e-95
Kurtosis:                       5.803   Cond. No.                         15.9
==============================================================================

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

Output for part 5:
                             OLS Regression Results
================================================================================
Dep. Variable:     future_unemployment   R-squared:                       0.121
Model:                             OLS   Adj. R-squared:                  0.120
Method:                  Least Squares   F-statistic:                     14.26
Date:                 Mon, 04 Apr 2022   Prob (F-statistic):           0.000172
Time:                         12:40:28   Log-Likelihood:                -1352.3
No. Observations:                  720   AIC:                             2709.
Df Residuals:                      718   BIC:                             2718.
Df Model:                            1
Covariance Type:                   HAC
```

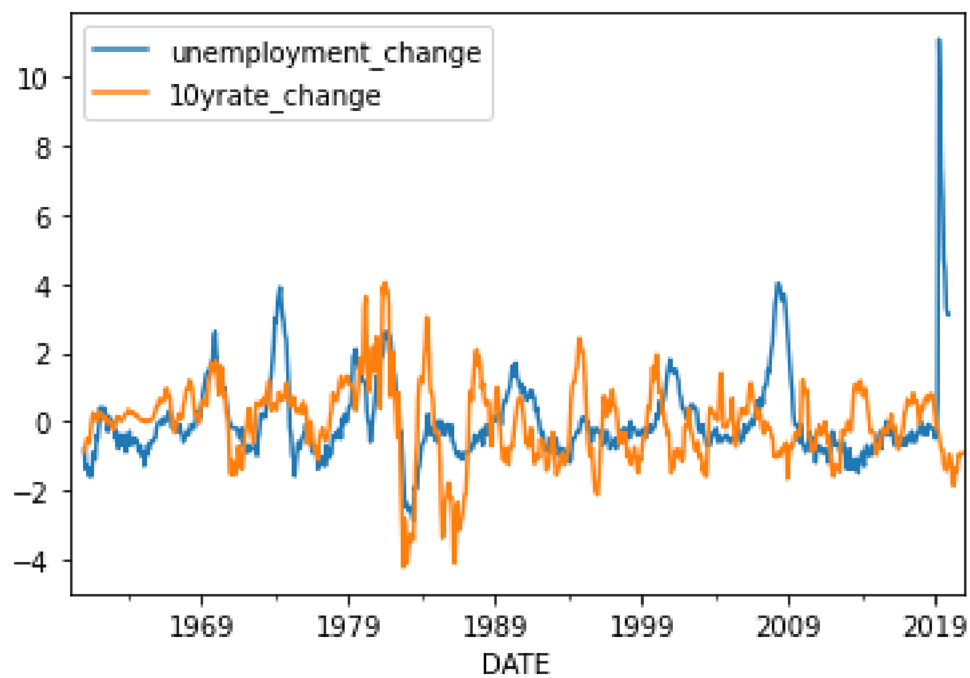
```
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.7717      0.397     12.009      0.000       3.992       5.552
10yrate        0.2044      0.054      3.777      0.000       0.098       0.311
==============================================================================
Omnibus:                      179.564   Durbin-Watson:                   0.078
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              437.278
Skew:                           1.296   Prob(JB):                     1.11e-95
Kurtosis:                       5.803   Cond. No.                         15.9
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC)
using 5 lags and without small sample correction
```

Output for part 6:
P-value on ADF test for future unemloyment = 0.01012335882391747
P-value on ADF test for 10-year rate = 0.7654399454443247

Part 6 answer: …

Output for part 7:



Output for part 8:
P-value on ADF test for unemployment change = 0.00013409498956585427
P-value on ADF test for 10-year rate change = 1.5917835074124478e-06

Part 8 answer: …

Output for part 9:

```
                            OLS Regression Results
===============================================================================
Dep. Variable:     unemployment_change   R-squared:                       0.033
Model:                             OLS   Adj. R-squared:                  0.031
Method:                  Least Squares   F-statistic:                     5.119
Date:                 Mon, 04 Apr 2022   Prob (F-statistic):             0.0240
Time:                         12:40:28   Log-Likelihood:                -1160.6
No. Observations:                  708   AIC:                             2325.
Df Residuals:                      706   BIC:                             2334.
Df Model:                            1
Covariance Type:                   HAC
================================================================================
==
                   coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
--
const            0.0305      0.104      0.292      0.770      -0.174
0.235
10yrate_change   0.2029      0.090      2.263      0.024       0.027
0.379
==============================================================================
Omnibus:                      521.528   Durbin-Watson:                   0.155
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            10067.352
Skew:                           3.113   Prob(JB):                         0.00
Kurtosis:                      20.392   Cond. No.                         1.13
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC)
using 5 lags and without small sample correction
```

Output for part 10:
Part 10 answer: …