AD 685 QUANTITATIVE METHODS FOR FINANC TERM PROJECT

Yuxuan Chen U76697870

1. Data selection

For this project I selected The Coca-Cola Company (KO). It is an American multinational beverage corporation headquartered in Atlanta, Georgia. A quick search on Yahoo Finance shows its Market Cap (intraday) is USD 299.663 billion (retrieved 18 Jun 2025). According to the project instructions, KO falls under the "large-cap" category (market capitalization between $100 billion and $200 billion). Therefore, I have chosen to use the SPDR S&P 500 ETF (SPY) as the benchmark. In addition to the benchmark, I selected SPDR Gold Shares (GLD) as my second explanatory variable (denoted as X2). GLD is an exchange-traded fund (ETF) designed to track the price of gold bullion. As the Coca-Cola Company operates on a global scale and is subject to commodity and macroeconomic fluctuations, the inclusion of gold returns provides a relevant economic signal.

2. Calculate daily earnings and related statistics

Because most of the estimations will refer to returns rather than prices, I choose to use logarithmic return:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

$$where P_{t-1} \ and \ P_t \ are \ the \ starting \ and \ ending \ prices.$$

Log-returns are additive and symmetric, closely approximate simple returns for small moves, and exhibit distributions closer to normality, making them ideal for multi-period quantitative asset analysis. The price data for GLD and SPY are also converted to daily logarithmic yields to maintain consistency with the yield calculation method used by KO.

By using Python, I obtained a series of statistics for the stock returns (see Table 1).

Name	The Coca-Cola Company
Ticker	KO

Mean	0.000193
Median	0.000958
Sample standard deviation	0.015624
Minimum	-0.101728
Maximum	0.062783
Sample skewness	-0.841427
Sample kurtosis	6.966093
Starting date	2020/1/3
Final date	2022/12/30

Table 1 Stock Return Statistics

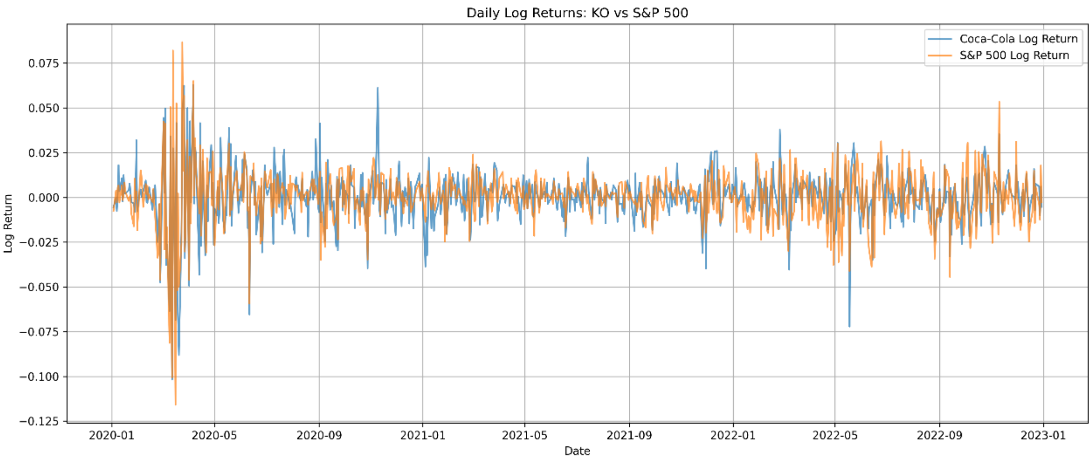

Figure 1: Daily Log Returns KO vs SPY

3. OLS regression and Newey–West estimator

3.1 Ordinary Least Squares Regression

Ordinary Least Squares (OLS) regression is a method for estimating the relationship between a dependent variable and one or more independent variables by minimizing the sum of squared residuals (differences between observed and predicted values).

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

OLS finds the set of coefficients $\beta_0,\ \beta_1, \cdots, \beta_k$ that minimize:

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

3.2 Newey–West estimator

In time-series regressions, residuals often display heteroskedasticity and autocorrelation, violating classical OLS assumptions. This makes standard errors invalid, compromising inference.

The Newey–West estimator corrects this by adjusting the OLS covariance matrix to provide robust standard errors.

The general formula is:

$$\widehat{Var}_{NW}(\widehat{\beta}) = (X'X)^{-1}\hat{S}(X'X)^{-1}$$

where:

$$\hat{S} = \Gamma_0 + \sum_{h=1}^{q} w_h \left(\Gamma_h + {\Gamma_h}^T\right)$$

$$q: maximum\ lag\ order\ (e.g., \boldsymbol{Newey\text{–}West}\ (\boldsymbol{1\ lag})\ uses\ q = 1)$$

$$w_h = 1 - \frac{h}{q+1}: Bartlett\ weights$$

$$\Gamma_h: sample\ lag - hhh\ autocovariance\ of\ residuals$$

For the special case of **Newey–West (1 lag)**:

$$\hat{S}_{lag=1} = \Gamma_0 + 0.5(\Gamma_1 + {\Gamma_1}^T)$$

Therefore, the robust covariance becomes:

$$\widehat{Var}_{NW(1)}(\widehat{\beta}) = (X'X)^{-1}[\Gamma_0 + 0.5(\Gamma_1 + {\Gamma_1}^T)](X'X)^{-1}$$

4. Regression Results

4.1 Separate data

First, I separated all the data into two periods:

- First 2 years. This data will be used to “teach” or establish our model.
- Last year. This data will be used to test our model and to “forecast”.

The results are shown on the table below:

KO-train (2020-2021)	
Mean	0.000147
Median	0.000751
Std Dev	0.016984
Min	-0.101728
Max	0.062783
Skewness	-0.81508
Kurtosis	6.624366

KO-test (2022)	
Mean	0.000286
Median	0.001399
Std Dev	0.012482
Min	-0.072169
Max	0.037942
Skewness	-0.835994
Kurtosis	4.678076

Table 2 KO’s two periods’ statistics

Compared to 2020–2021, KO's log returns in 2022 showed slightly higher average returns, but lower volatility and fewer extreme movements. The return distribution remained slightly left-skewed, but became more stable and closer to normal, as indicated by lower kurtosis. Overall, KO returns in 2022 were calmer and less risky than in the prior two years.

SPY-train (2020-2021)	
Mean	0.000754
Median	0.001648
Std Dev	0.016144
Min	-0.115887
Max	0.086731
Skewness	-1.061486
Kurtosis	13.253851

SPY-test (2022)	
Mean	-0.000863
Median	-0.001815
Std Dev	0.015298
Min	-0.044456
Max	0.053497
Skewness	-0.001245
Kurtosis	0.33286

Table 3 SPY’s two periods’ statistics

In 2022, SPY's average and median log returns turned negative, contrasting with the positive returns in 2020–2021. Volatility stayed at a similar level, but extreme movements became much less frequent, as seen from the sharp drop in kurtosis. The distribution also became more symmetric, with skewness moving closer to zero. Overall, SPY returns in 2022 were weaker but more stable.

GLD-train (2020-2021)	
Mean	0.000341
Median	0.000926
Std Dev	0.010643
Min	-0.05519
Max	0.04739
Skewness	-0.614508
Kurtosis	3.869966

GLD-test (2022)	
Mean	-0.000031
Median	0.000388
Std Dev	0.009619
Min	-0.030162
Max	0.030235
Skewness	0.133872
Kurtosis	0.629863

Table 4 GLD's two periods' statistics

GLD's log returns in 2022 showed slightly lower volatility compared to 2020–2021.

The average return turned slightly negative, though the median remained positive, suggesting minimal directional change. Skewness shifted from left-skewed to mildly right-skewed, and kurtosis decreased notably, indicating fewer extreme price moves. Overall, GLD returns in 2022 were more balanced and less volatile than in the previous two years.

4.2 Stock returns & Intercept

```
=== OLS ===
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.002
Model:                            OLS   Adj. R-squared:                 -0.000
Method:                 Least Squares   F-statistic:                    0.7580
Date:                Thu, 19 Jun 2025   Prob (F-statistic):              0.384
Time:                        20:09:43   Log-Likelihood:                 1339.8
No. Observations:                 504   AIC:                            -2676.
Df Residuals:                     502   BIC:                            -2667.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -2.3085      2.652     -0.871      0.384      -7.518       2.901
x1          3.129e-06   3.59e-06      0.871      0.384   -3.93e-06    1.02e-05
==============================================================================
Omnibus:                      112.389   Durbin-Watson:                   2.092
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              908.295
Skew:                          -0.718   Prob(JB):                    5.84e-198
Kurtosis:                       9.418   Cond. No.                     2.59e+09
==============================================================================
```

Figure 2: KO vs Intercept (OLS)

```
=== Newey-West (lag = 1) ===
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.002
Model:                            OLS   Adj. R-squared:                 -0.000
Method:                 Least Squares   F-statistic:                    0.5620
Date:                Thu, 19 Jun 2025   Prob (F-statistic):              0.454
Time:                        20:09:43   Log-Likelihood:                 1339.8
No. Observations:                 504   AIC:                            -2676.
Df Residuals:                     502   BIC:                            -2667.
Df Model:                           1
Covariance Type:                  HAC
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -2.3085      3.080     -0.750      0.454      -8.359       3.743
x1          3.129e-06   4.17e-06      0.750      0.454   -5.07e-06    1.13e-05
==============================================================================
Omnibus:                      112.389   Durbin-Watson:                   2.092
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              908.295
Skew:                          -0.718   Prob(JB):                    5.84e-198
Kurtosis:                       9.418   Cond. No.                     2.59e+09
==============================================================================
```

Figure 3: KO vs Intercept (Newey-West)

Whether using ordinary OLS or Newey-West, the p-values for Intercept are much greater than 0.05 (0.384 and 0.454) and are not significant.

4.3 Stock returns & Benchmark returns & Intercept

```
=== OLS ===
                            OLS Regression Results
==============================================================================
Dep. Variable:              KO_logRet   R-squared:                       0.571
Model:                            OLS   Adj. R-squared:                  0.571
Method:                 Least Squares   F-statistic:                     669.3
Date:                Thu, 19 Jun 2025   Prob (F-statistic):           2.05e-94
Time:                        19:37:12   Log-Likelihood:                 1552.9
No. Observations:                 504   AIC:                            -3102.
Df Residuals:                     502   BIC:                            -3093.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.0005      0.000     -0.912      0.362      -0.001       0.001
SPY_logRet     0.7953      0.031     25.871      0.000       0.735       0.856
==============================================================================
Omnibus:                       75.347   Durbin-Watson:                   1.873
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              551.441
Skew:                          -0.377   Prob(JB):                    1.80e-120
Kurtosis:                       8.069   Cond. No.                         62.0
==============================================================================
```

Figure 4: KO vs SPY & Intercept (OLS)

```
=== Newey-West (lag = 1) ===
                            OLS Regression Results
==============================================================================
Dep. Variable:              KO_logRet   R-squared:                       0.571
Model:                            OLS   Adj. R-squared:                  0.571
Method:                 Least Squares   F-statistic:                     250.8
Date:                Thu, 19 Jun 2025   Prob (F-statistic):           4.15e-46
Time:                        19:37:12   Log-Likelihood:                 1552.9
No. Observations:                 504   AIC:                            -3102.
Df Residuals:                     502   BIC:                            -3093.
Df Model:                           1
Covariance Type:                  HAC
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.0005      0.001     -0.880      0.379      -0.001       0.001
SPY_logRet     0.7953      0.050     15.837      0.000       0.697       0.894
==============================================================================
Omnibus:                       75.347   Durbin-Watson:                   1.873
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              551.441
Skew:                          -0.377   Prob(JB):                    1.80e-120
Kurtosis:                       8.069   Cond. No.                         62.0
==============================================================================
```

Figure 5: KO vs SPY & Intercept (Newey-West)

In both the OLS and Newey-West regressions, the intercept terms are not statistically significant, with p-values of 0.362 and 0.379, respectively. In contrast, the SPY log return is highly significant in both models, with p-values well below the 0.05 threshold.

4.4 Stock returns & CRSP returns & Intercept

```
=== OLS ===
                            OLS Regression Results
==============================================================================
Dep. Variable:              KO_logRet   R-squared:                       0.518
Model:                            OLS   Adj. R-squared:                  0.517
Method:                 Least Squares   F-statistic:                     538.7
Date:                Thu, 19 Jun 2025   Prob (F-statistic):           1.67e-81
Time:                        19:37:12   Log-Likelihood:                 1523.1
No. Observations:                 504   AIC:                            -3042.
Df Residuals:                     502   BIC:                            -3034.
Df Model:                           1
Covariance Type:            nonrobust
===============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
Intercept     -0.0006      0.001     -1.080      0.281      -0.002       0.000
CRSP_logRet    0.0074      0.000     23.210      0.000       0.007       0.008
==============================================================================
Omnibus:                       93.962   Durbin-Watson:                   1.834
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              785.599
Skew:                          -0.528   Prob(JB):                    2.57e-171
Kurtosis:                       9.024   Cond. No.                         1.67
==============================================================================
```

Figure 6: KO vs CRSP & Intercept (OLS)

```
=== Newey-West (lag = 1) ===
                            OLS Regression Results
==============================================================================
Dep. Variable:              KO_logRet   R-squared:                       0.518
Model:                            OLS   Adj. R-squared:                  0.517
Method:                 Least Squares   F-statistic:                     227.0
Date:                Thu, 19 Jun 2025   Prob (F-statistic):           1.35e-42
Time:                        19:37:13   Log-Likelihood:                 1523.1
No. Observations:                 504   AIC:                            -3042.
Df Residuals:                     502   BIC:                            -3034.
Df Model:                           1
Covariance Type:                  HAC
===============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
Intercept     -0.0006      0.001     -1.033      0.302      -0.002       0.001
CRSP_logRet    0.0074      0.000     15.067      0.000       0.006       0.008
==============================================================================
Omnibus:                       93.962   Durbin-Watson:                   1.834
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              785.599
Skew:                          -0.528   Prob(JB):                    2.57e-171
Kurtosis:                       9.024   Cond. No.                         1.67
==============================================================================
```

Figure 7: KO vs CRSP & Intercept (Newey-West)

In both the OLS and Newey-West regressions, the intercept terms are not statistically significant, with p-values of 0.281 and 0.302, respectively. In contrast, the CRSP log return is highly significant in both models, with p-values far below the 0.05 threshold.

4.5 Stock returns & Benchmark returns & Gold returns & Intercept

```
=== OLS ===
                            OLS Regression Results
==============================================================================
Dep. Variable:              KO_logRet   R-squared:                       0.572
Model:                            OLS   Adj. R-squared:                  0.570
Method:                 Least Squares   F-statistic:                     334.4
Date:                Thu, 19 Jun 2025   Prob (F-statistic):           5.66e-93
Time:                        19:37:13   Log-Likelihood:                 1553.1
No. Observations:                 504   AIC:                            -3100.
Df Residuals:                     501   BIC:                            -3087.
Df Model:                           2
Covariance Type:            nonrobust
===============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
Intercept     -0.0005      0.000     -0.925      0.355      -0.001       0.001
SPY_logRet     0.7926      0.031     25.457      0.000       0.731       0.854
Gold_logRet    0.0267      0.047      0.566      0.571      -0.066       0.120
==============================================================================
Omnibus:                       73.909   Durbin-Watson:                   1.874
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              552.457
Skew:                          -0.349   Prob(JB):                    1.09e-120
Kurtosis:                       8.081   Cond. No.                         96.0
==============================================================================
```

Figure 8: KO vs SPY & Gold & Intercept (OLS)

```
=== Newey-West (lag = 1) ===
                            OLS Regression Results
==============================================================================
Dep. Variable:              KO_logRet   R-squared:                       0.572
Model:                            OLS   Adj. R-squared:                  0.570
Method:                 Least Squares   F-statistic:                     131.4
Date:                Thu, 19 Jun 2025   Prob (F-statistic):           1.35e-46
Time:                        19:37:14   Log-Likelihood:                 1553.1
No. Observations:                 504   AIC:                            -3100.
Df Residuals:                     501   BIC:                            -3087.
Df Model:                           2
Covariance Type:                  HAC
================================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept     -0.0005      0.001     -0.885      0.376      -0.001       0.001
SPY_logRet     0.7926      0.052     15.201      0.000       0.690       0.895
Gold_logRet    0.0267      0.080      0.335      0.738      -0.130       0.184
==============================================================================
Omnibus:                       73.909   Durbin-Watson:                   1.874
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              552.457
Skew:                          -0.349   Prob(JB):                    1.09e-120
Kurtosis:                       8.081   Cond. No.                         96.0
==============================================================================
```

Figure 9: KO vs SPY & Gold & Intercept (Newey-West)

In both the OLS and Newey-West regressions, the intercept terms are not statistically significant, with p-values of 0.355 and 0.376, respectively. The SPY log return remains highly significant across both models, with p-values well below 0.05, confirming its strong predictive power for KO's returns. However, the Gold log return is not statistically significant in either model (p-values of 0.571 and 0.738), suggesting that gold returns do not provide meaningful additional explanatory power for KO's daily return.

4.6 Stock returns & CRSP returns & Gold returns & Intercept

```
=== OLS ===
                            OLS Regression Results
==============================================================================
Dep. Variable:              KO_logRet   R-squared:                       0.518
Model:                            OLS   Adj. R-squared:                  0.516
Method:                 Least Squares   F-statistic:                     269.1
Date:                Thu, 19 Jun 2025   Prob (F-statistic):           4.29e-80
Time:                        19:37:14   Log-Likelihood:                 1523.2
No. Observations:                 504   AIC:                            -3040.
Df Residuals:                     501   BIC:                            -3028.
Df Model:                           2
Covariance Type:            nonrobust
===============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
Intercept     -0.0006      0.001     -1.091      0.276      -0.002       0.000
CRSP_logRet    0.0073      0.000     22.799      0.000       0.007       0.008
Gold_logRet    0.0253      0.050      0.505      0.614      -0.073       0.124
==============================================================================
Omnibus:                       92.115   Durbin-Watson:                   1.835
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              784.589
Skew:                          -0.502   Prob(JB):                    4.25e-171
Kurtosis:                       9.029   Cond. No.                         158.
==============================================================================
```

Figure 10: KO vs CRSP & Gold & Intercept (OLS)

```
=== Newey-West (lag = 1) ===
                            OLS Regression Results
==============================================================================
Dep. Variable:              KO_logRet   R-squared:                       0.518
Model:                            OLS   Adj. R-squared:                  0.516
Method:                 Least Squares   F-statistic:                     117.7
Date:                Thu, 19 Jun 2025   Prob (F-statistic):           1.28e-42
Time:                        19:37:15   Log-Likelihood:                 1523.2
No. Observations:                 504   AIC:                            -3040.
Df Residuals:                     501   BIC:                            -3028.
Df Model:                           2
Covariance Type:                  HAC
===============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
Intercept     -0.0006      0.001     -1.034      0.302      -0.002       0.001
CRSP_logRet    0.0073      0.001     14.432      0.000       0.006       0.008
Gold_logRet    0.0253      0.086      0.296      0.768      -0.143       0.194
==============================================================================
Omnibus:                       92.115   Durbin-Watson:                   1.835
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              784.589
Skew:                          -0.502   Prob(JB):                    4.25e-171
Kurtosis:                       9.029   Cond. No.                         158.
==============================================================================
```

Figure 11: KO vs CRSP & Gold & Intercept (Newey-West)

In both the OLS and Newey-West regressions, the intercept terms are not statistically significant, with p-values of 0.276 and 0.302, respectively. The CRSP log return is highly significant in both models (p-values < 0.001), confirming its strong explanatory power for KO's daily returns. However, the Gold log return is not statistically significant ($p = 0.614$ in OLS, and 0.768 in Newey-West), indicating that gold returns do not significantly contribute to explaining KO's return variation.

In summary, no variable changes its significance status when comparing OLS and Newey-West t-statistics.

5. Results Analysis

Regression of KO vs SPY, GLD and CRSP				
Sample: Jan 01, 2020 - Jan 01, 2025				
	1	2	3	4
Intercept	-0.0005	-0.0006	-0.0005	-0.0006
(OLS s.e.)	0.0000	0.0010	0.0000	0.0010
(NW s.e.)	0.0010	0.0010	0.0010	0.0010
S&P500	0.7953	-	0.7926	-
(OLS s.e.)	0.0310	-	0.0310	-
(NW s.e.)	0.0500	-	0.0520	-
CRSP	-	0.0074	-	0.0073
(OLS s.e.)	-	0.0000	-	0.0000
(NW s.e.)	-	0.0000	-	0.0010
X_2	-	-	0.0267	0.0253
(OLS s.e.)	-	-	0.0470	0.0500
(NW s.e.)	-	-	0.0800	0.0860
R^2 Adjusted	0.571	0.517	0.57	0.516
s.e. Reg	0.01113	0.01181	0.01114	0.01182
NOBS	504	504	504	504

Table 5 4 Estimated Regressions

In this project, I compare two alternative specifications for modeling the daily log returns of The Coca-Cola Company (KO): Model (1), which uses the SPDR S&P 500 ETF (SPY) as the market proxy, and Model (2), which uses CRSP. Although both market return variables are highly statistically significant (p-values < 0.001), I find that Model (1) offers stronger explanatory power with an R-squared of 0.571, compared to 0.517 in Model (2). This suggests that SPY better captures market-related movements in KO returns over the sample period. While I do not observe strong evidence of omitted variable bias in either model, I acknowledge that other macroeconomic factors, such as gold prices or interest rates, could influence KO’s returns. To address this possibility, I estimate additional models that include gold returns as an explanatory variable.

In Models (3) and (4), I incorporate gold returns alongside SPY and CRSP, respectively. In both models, the market return variables remain statistically significant, while the gold return is not (p-values well above 0.05). Model (3) yields the highest R-squared (0.572), which confirms that SPY is a more effective market benchmark in this setting. Based on these results, I select **Model (3)** as the preferred specification for explaining KO's return behavior over the sample period, due to its stronger explanatory power and robustness.