

Consumer Staples Industries Stock Returns Analysis

IEOR 4150 | Professor Antonius Dieker

Nuoya Wang, nw2432; Ming Jin, mj2940; Zihui Zhou, zz2694

Written report

Our project focuses on 10 stocks of the consumer staples industries selected from the list of S&P 500 companies. We apply statistical analysis on the dataset in order to gain business insights into these industries. Firstly, we calculate and plot the histogram of the log-returns of each individual stock (data). Then, we compute confidence intervals for the mean and variance of the data and perform regression analysis of the log-return on time to have a better understanding of companies' business performance from 2018-01-01 to 2019-12-05. Finally, we compare pairs of 10 stocks to figure out if they have similar log-returns and any correlation.

Datasets description

The dataset consists of daily open and close prices of 10 consumer staples companies' ticker stocks from 2018-01-01 to today. Each of the following company's ticker stocks consists of 487 data points, which are drawn from Yahoo Finance.

CPB -- Campbell Soup Company
CL -- Colgate-Palmolive Company
COST -- Costco Wholesale Corporation
SJM -- J.M.Smucker Company
K -- Kellogg Company
PEP -- PepsiCo, Inc.
KO -- The Coca-Cola Company
HSY -- The Hershey Company
KHC -- The Kraft Heinz Company
WMT -- Walmart Inc.

Project's goals

According to the quarterly market updates provided by Fidelity Investments, the global economy is increasing slowly in business cycles in recent years. While most sectors' stocks fluctuate during the past quarter reports, the consumer staples sector remains relatively stable. This phenomenon triggers our interest in researching consumer staples industries' overall performance and provides instructional suggestions on investment.

According to our research, as the consumer staples sector includes food and beverage manufacturers and household goods producers, it is less sensitive to economic cycles. However, stocks in the industries also tend to pay generous dividend yields that grow steadily each year for a long time.

Taking current economic downturns into considerations, we believe that the consumer staples sector is a good choice for investments because of its stable performances and growing dividend. In order to achieve a more solid conclusion, our project conducts several statistical analyses on selected companies' stocks in the consumer staples sector. We also used Jupyter notebook with widgets through Python to develop a user-interactive web app. They may gain deeper insights into a specific stock's performance and make better decisions on which stock to invest for the future through selecting stocks, setting time periods, and comparing with other stocks.

First of all, to make sure that data analysis would be representative, we decided to focus on influential companies in the consumer staples industries -- we selected ten large companies from the S&P 500 consumer staples sector. Then, we perform a thorough analysis of the log-returns (taking the natural logarithm of the divisions of the close price over the open price of the day) of single stocks and pairs of stocks.

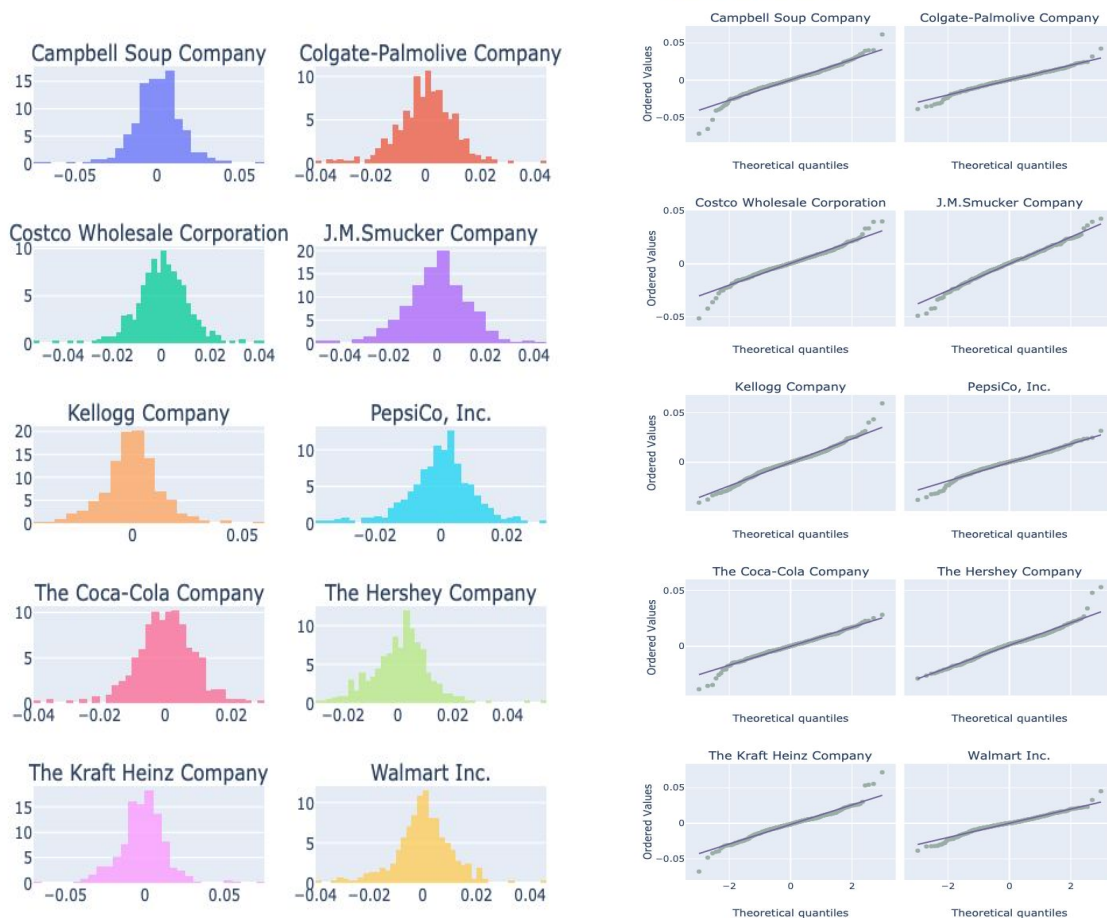
To ensure the accuracy and validity of the analysis, we first graph histograms and normal probability plots to check if the data points of each stock are from a normal distribution. We choose data points randomly from the datasets by setting the number of data points used in all analyses except linear regression on a single stock to guarantee the randomness of data points. Then, we analyze each stock's performance since 2018-01-01 through evaluating means and variances of the stocks' log-returns, and linear-regression modeling.

After analyzing stocks individually, we want to conclude the general trends among the consumer staples industries. To achieve the goal, we first test the equality of pairs of stocks' means to see whether the selected stocks' means are equal. Then, we perform regressions on pairs of stocks' log-returns to figure out if the stocks in the consumer staples industries follow a common pattern.

Data analysis

Single Stock Analyses

First of all, we tested the assumption that all log-returns of a certain company are from a random sample by graphing histograms of log-returns. The histograms illustrate the frequency distributions of log-returns that seem normally distributed. To validate the assertion that those log-returns are from normal distributions, we also plot normal probability plots. The log-return data are plotted against a theoretical normal distribution. Although the head and tail data points deviate from the theoretical straight line, most data points remain on the line and form a linear pattern. Therefore, we may draw the conclusion that the stock log-returns came from a normal distribution.



We analyze the means and variances of the log-returns of each company through their confidence intervals. After proving the normality of the datasets, we may use statistical tests on the sample datasets to determine the confidence intervals for means, μ , and variances, σ^2 . As we only know the sample variance, s^2 , of the log-returns but not the population's variance, σ^2 , of the log-returns' distributions those data points come from, we use t-distribution with $n-1$

degrees of freedom to calculate the confidence interval for μ . Similarly, we use chi-squared distribution to calculate the confidence interval for σ^2 . For a better understanding, we calculate 95% confidence intervals of the means and variance as shown in the following table. From the table, we may notice that each confidence interval of the means, μ , contains 0 and has small a width except The Kraft Heinz Company. The negativity of 95% confidence interval indicates the decreasing trend of this company's stock. The confidence interval of σ^2 is also extremely small and close to zero. Thus, these results suggest that these stocks' do not fluctuate a lot within a day.

company	95% Confidence Interval of μ	95% Confidence Interval of σ^2
Campbell Soup Company	(-0.001011, 0.001554)	(0.000174, 0.000225)
Colgate-Palmolive Company	(-0.000902, 0.00096)	(9.2e-05, 0.000119)
Costco Wholesale Corporation	(-0.000378, 0.001558)	(9.9e-05, 0.000128)
J.M.Smucker Company	(-0.001216, 0.001112)	(0.000143, 0.000186)
Kellogg Company	(-0.001283, 0.000939)	(0.000131, 0.000169)
PepsiCo, Inc.	(-0.001034, 0.00071)	(8e-05, 0.000104)
The Coca-Cola Company	(-0.000741, 0.000841)	(6.6e-05, 8.6e-05)
The Hershey Company	(-3.3e-05, 0.001833)	(9.2e-05, 0.000119)
The Kraft Heinz Company	(-0.002839, -0.00029)	(0.000172, 0.000223)
Walmart Inc.	(-0.000762, 0.001081)	(9e-05, 0.000116)

Then, we perform a linear regression of the log-returns on time. The log-returns on time are computed by taking the natural logarithm of the division of the close price of each day over the open price of 2018-01-02. Observing the table, we notice that besides Campbell Soup Company, Colgate-Palmolive Company, J.M.Smucker Company, and Kellogg Company, R^2 values of the rest six companies are close to one, which indicates stronger linear relations between the stocks' log-returns and time. However, the regression of The Kraft Heinz Company has a negative slope as -0.001608, and R^2 as 0.90259. These indicators suggests a strong negative linear correlation, and a decreasing dividend return for the future. Nevertheless, among the ten companies, five of them show strong positive linearity, we may conclude that the consumer staples industries have a steady increase since 2018-01-01. In this way, we may conclude a potential increasing future log-returns for the consumer staples industries.

	symbol	slope	intercept	R^2	p_value	std_err
	Campbell Soup Company	0.000058	-0.19374	0.01399	8.971976e-03	0.000022
	Colgate-Palmolive Company	0.000051	-0.12593	0.02893	1.625050e-04	0.000013
	Costco Wholesale Corporation	0.000685	-0.03219	0.85836	5.674417e-208	0.000013
	J.M.Smucker Company	-0.000104	-0.07028	0.08055	1.785874e-10	0.000016
	Kellogg Company	-0.000204	-0.00794	0.22824	3.900817e-29	0.000017
	PepsiCo, Inc.	0.000379	-0.14812	0.67272	1.031578e-119	0.000012
	The Coca-Cola Company	0.000326	-0.07539	0.74288	3.812401e-145	0.000009
	The Hershey Company	0.000753	-0.25727	0.77367	1.379619e-158	0.000019
	The Kraft Heinz Company	-0.001608	0.00243	0.90259	2.074820e-247	0.000024
	Walmart Inc.	0.000419	-0.14890	0.63430	5.191415e-108	0.000014

Regression Model of log return of The Kraft Heinz Company vs Time



Two Stock Analyses

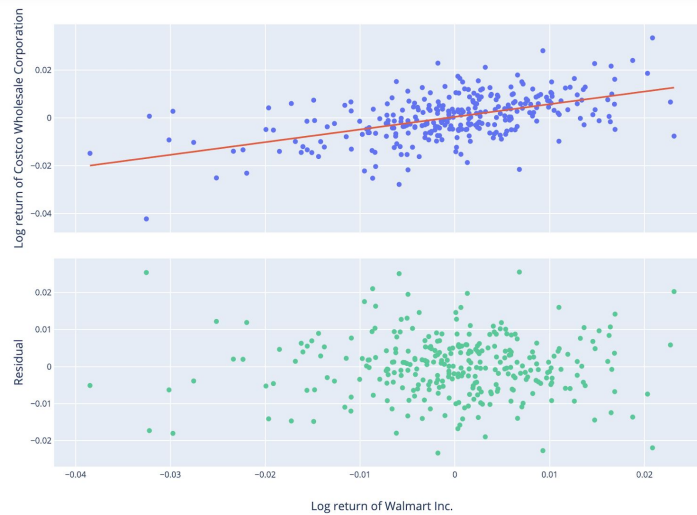
To gain a closer look at the relationships between the ten stocks, we designed the web app to check the equality of the population means and to perform a regression of any two selected stock symbols.

Since both the means, μ , and variances, σ^2 of the log-return distributions were unknown, we used t-test to check the equality of the population means. Here, we will discuss the results of comparing the population mean of the Walmart stock with that of others. We randomly chose 300 samples of each stock and set the confidence level to be 95%. Since the p-values of these nine tests are all greater than 0.05, we accept the null hypothesis that the log-returns of the other nine stocks have the same population mean as the log-returns of Walmart. This conclusion is intuitive because Walmart is a major retailer of the 8 consumer staples brands and a competitor of Costco facing similar challenges and rewards of the industries.

symbol	p_value (& Walmart)
Campbell Soup Company	0.999381
Colgate-Palmolive Company	0.791592
Costco Wholesale Corporation	0.775281
J.M.Smucker Company	0.651538
Kellogg Company	0.671545
PepsiCo, Inc.	0.920010
The Coca-Cola Company	0.913062
The Hershey Company	0.359630
The Kraft Heinz Company	0.147380
Walmart Inc.	0.327049

The web app allows the user to select pairs of stock symbols as either a regressor or a regressand. We tabulate the slopes, intercepts, p-values, and R^2 values of the regressions with the log-returns of the Walmart Inc. stock as the input variable. Since all the p-values are less than α which is 0.05, we can reject the null hypothesis of each regression that there is no correlation between the two variables. The p-values combined with the positive slopes indicate that under the confidence level of 95%, the log-return of Walmart Inc.'s stock is positively correlated to the other nine stocks -- if the log-return of Walmart Inc.'s stock increases, other stocks' log-returns are also likely to increase. However, the linear models do not explain much about the variability of the response variable indicated by the observation that R^2 values of these regressions are below 30%. In other words, the ten stocks' correlations within one day are not obvious. This result can be explained by the fact that the performance of consumer staples products and retailers involves human factors which are hard to predict by the models. Although R^2 values are not large, the residuals are randomly distributed around 0 for all nine plots, similar to the regression graph between Walmart Inc. and Costco Wholesale Corporation depicted below, meaning that the linear models are still appropriate for fitting our data.

	symbol	slope	intercept	R^2	p_value	std_err
	Campbell Soup Company	0.415352	0.00054	0.10677	6.797610e-09	0.069593
	Colgate-Palmolive Company	0.325577	0.00009	0.11928	7.890066e-10	0.051249
	Costco Wholesale Corporation	0.563878	0.00030	0.29531	1.896342e-24	0.050459
	J.M.Smucker Company	0.437779	-0.00066	0.13945	2.319894e-11	0.062998
	Kellogg Company	0.466057	0.00015	0.16573	2.100597e-13	0.060573
	PepsiCo, Inc.	0.407179	-0.00023	0.18407	7.272703e-15	0.049660
	The Coca-Cola Company	0.375594	0.00012	0.20164	2.719297e-16	0.043294
	The Hershey Company	0.321490	0.00055	0.09307	7.005292e-08	0.058135
	The Kraft Heinz Company	0.520755	-0.00128	0.14141	1.640627e-11	0.074333
	Walmart Inc.	1.000000	0.00000	1.00000	0.000000e+00	0.000000



Conclusions

In our analysis of the 10 stocks from the consumer staples industries, we first showed that the log-returns of the stocks are roughly normally distributed by graphing their histograms. To validate this observation, we also include the normal probability plots, whose linearity also indicates the normal distribution of the data points.

After proving the normality, we were able to perform statistical tests to calculate the 95% confidence interval of the population's means and variances of the log-return distributions. We find that these stocks tend to be stable since the confidence intervals of the means contain 0 with small widths and the confidence intervals of the variances are also extremely small and close to zero.

We then conduct the linear regression of the log-returns on time for each company. We notice that three of them (Campbell Soup Company, Colgate-Palmolive Company, and J.M.Smucker

Company) do not possess linearity, two of them (Kellogg Company and The Kraft Heinz Company) possess strong negative correlation, and the rest five of them have strong positive linear correlations. According to our linear regression analysis, we believe that it is better to invest in the five companies which have strong positive linear correlations. Besides, the analysis also supports that dividend yields grow steadily over the years. Thus, we may also conclude that consumer staples industries have relatively positive returns in the stock market.

To get more insights about the relationship between the stocks, we performed t-test to compare the population means of the log-return distributions. The results show that the log-returns of the other nine stocks have almost the same population mean as Walmart. We also discuss the regressions of the log-returns of other stocks over Walmart, which indicate that the log-return of the Walmart stock is positively correlated to the other nine stocks, respectively.

Above all, we may verify the “old wives’ tale” -- consumer staples industries have a steady growing trend in the stock market. Therefore, we believe these industries are suitable for making investments.

Reference

Barchart.com. (2019). *S&P 500 Consumer Staples Sector Charts, Components, Prices - Barchart.com*. [online] Available at: <https://www.barchart.com/stocks/indices/sp-sector/consumer-staples> [Accessed 6 Dec. 2019].

Eresearch.fidelity.com. (2019). *Consumer Staples Snapshot - Fidelity*. [online] Available at: https://eresearch.fidelity.com/eresearch/markets_sectors/sectors/sectors_in_market.jhtml?tab=learn§or=30 [Accessed 6 Dec. 2019].

Medium. (2019). *Detecting stock market crashes with topological data analysis*. [online] Available at: <https://towardsdatascience.com/detecting-stock-market-crashes-with-topological-data-analysis-7d5dd98abe42> [Accessed 6 Dec. 2019].

Research2.fidelity.com. (2019). *Fidelity PDF Report*. [online] Available at: https://research2.fidelity.com/fidelity/research/reports/public/getreport.asp?feedid=2271&doctag=q42019quarterlysectorupdate_92_20191104&versiontag=2019110407520743 [Accessed 6 Dec. 2019].

The Balance. (2019). *What's Inside the Most Popular Stock Index?*. [online] Available at: <https://www.thebalance.com/what-is-the-sector-weighting-of-the-s-and-p-500-4579847> [Accessed 6 Dec. 2019].