**Data Set**

Our data set consists of the daily closing prices of the S&P 500 Index and 9 Exchange-traded Funds (ETFs) every trading day from January 1, 2009 to December 6, 2018. A total of 2499 data points were collected for each stock and the daily closing prices were taken from Yahoo Finance. The ticker symbol and the full name of the ETFs they represent in our data set are:

PBJ - Invesco Dynamic Food & Beverage ETF

XLP - Consumer Staples Select Sector SPDR ETF

PAGG - Invesco Global Agriculture ETF

FUD - UBS ETRACS CMCI Food Total Return ETF

PEJ - Invesco Dynamic Leisure and Entmnt ETF

PBS - Invesco Dynamic Media ETF

BJK - VanEck Vectors Gaming ETF

XLY - Consumer Discret Sel Sect SPDR ETF

IGN - iShares North Amer Tech-Multimd Ntwk ETF

The aforementioned ETFs are drawn from a pool ETFs focusing in food & beverage or media & entertainment.

**Project Goals**

Our project focuses on testing one of the most puzzling investment axioms - the Super Bowl indicator. The axiom states that “when a team from the original NFL (prior to the AFL merger) wins the Super Bowl, the stock market has a good year” (Wallick, 2009). Firstly we hope to test whether that statement is true by identifying the years in which the original NFL teams win and tracking the S&P 500 index. On top of comparing overall market performance, we further hypothesized that impact of an original NFL team win on the stock market is passed through several channels such as sports media and entertainment as well as the food & beverage industry. Therefore we selected ETFs that allocate a predominant weight on the sectors and industries mentioned above.

Despite the fact that the Super Bowl indicator seem extremely strange and nonsensical, if it does predict the market direction with high accuracy, then Super Bowl results each year would present our client with highly useful information in terms of capturing potential investment opportunities or adjusting their portfolio weights. Therefore we hope that conducting statistical tests could lead to

a more informed decision regarding whether or not to increase investment activities along with which ETFs to invest in depending on the winning team of the Super Bowl.

For the S&P 500 index, we first want to test whether the daily log returns are consistent with a normal random sample. Regardless of the results, we will continue with the remaining analysis. Next we assess the S&P 500 index’s performance by computing the mean and variance of the daily log-returns. We will also create approximate confidence intervals for the mean and variance given a 95% confidence level. Finally, we perform a regression of the log-return on time to see if there is any general trends that we could take advantage of. Then we perform the same procedure for all the individual ETFs we selected.

After answering the basic questions, we then ask the question of whether the event of an original NFL team winning Super Bowl correlates with better index or ETF performances. In order to answer this question, we first need to separate years from 2009 to 2018 into those in which an original NFL team wins the Super Bowl (Year = 1) and the rest (Years = 0). Then we calculate the means and variances conditional on Year = 1 and Year = 0. Lastly, we test the equalities of the two means and two variances to see whether there is a statistically significant difference.

In addition to testing equality of population means, we will employ a linear regression of log returns on the Year dummy variable to see whether Year = 1, which corresponds to an original NFL team winning Super Bowl, has a statistically significant effect on log-returns.

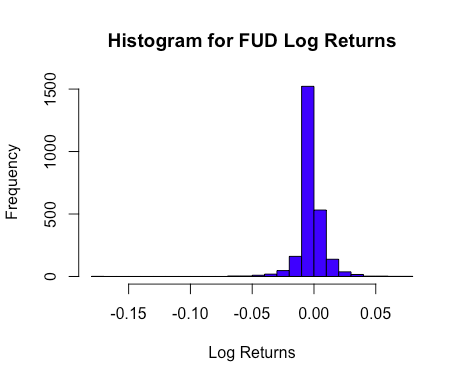
**Analysis**

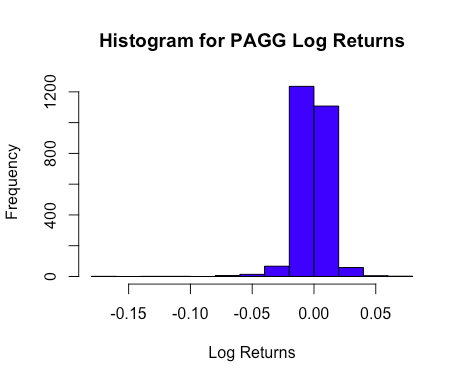
**S&P 500 & Single ETF Analyses**

By looking at the scatterplots of the log-returns of the selected index and ETSs, we detected no apparent patterns that violate randomness. In addition, we performed the runs test the determine non-randomness. The null hypothesis is that the log-returns in the data set are mutually independent, whereas the alternative hypothesis is that they are not and therefore the data isn’t consistent with a random sample.

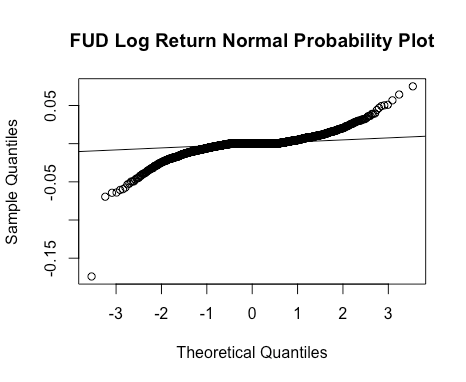
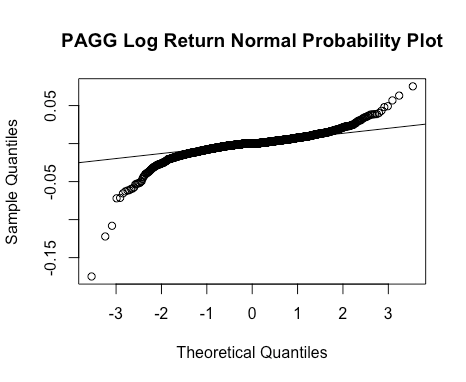
From the test results, we found that only S&P 500 and PAGG had a p-value less than 0.05. Thus we reject the null hypothesis for those two and fail to reject for all the other ETFs. In conclusion, only the log-returns of S&P 500 and PAGG were not consistent with a random sample.

We then examine the assumption that log-returns follow normal distributions. First we draw histograms for the S&P 500 and each individual ETF. The results are mixed. The histograms for IGN, PBJ, PBS, PEJ, XLP, XLY do present frequency distributions that seem approximately normal. However, for S&P 500, BJK, SUD, and PAGG, the log-returns seem to have slightly left-skewed distributions.





In order to ascertain our guesses, we drew normal probability plots, which further confirmed our previous observations. The plots for IGN, PBJ, PBS, PEJ, XLP, XLY mostly follow straight lines. In contrast, those for S&P 500, BJK, SUD, and PAGG, there are relatively large deviations from linearity. The normal probability plots for FUD and PAGG are shown below. Therefore we conclude that the log-returns for IGN, PBJ, PBS, PEJ, XLP, XLY follow came from a normal distribution, whereas those for S&P 500, BJK, SUD, and PAGG do not.

Now that we’ve examined the random sample and normality assumptions, we will continue to construct confidence intervals for log-returns’ mean, 𝜇, and the variance, 𝜎2. Since the population variances are unknown, we utilized t-distribution with n-1 = 2498 degrees of freedom and substituted in the standard deviations to derive the confidence intervals for the means. For the variances, we utilized chi-squared distribution with 2498 degrees of freedom.

**95% Confidence Interval for 𝝁 and 𝝈𝟐**

|  |  |  |
| --- | --- | --- |
| **Ticker Symbol** | **Confidence Interval for 𝝁** | **Confidence Interval for 𝝈𝟐** |
| **S&P 500** | (-3.5e-05, 7.3e-04) | (8.95e-05, 1.00e-04) |
| **BJK** | **(-0.0011, -0.0001)** | (0.00015, 0.00017) |
| **FUD** | **(-0.0010, -0.0001)** | (0.00011, 0.00012) |
| **PAGG** | (-5.2e-04, 7.12e-05) | (0.00011, 0.00012) |
| **PBJ** | (-5.2e-04, 9.0e-05) | (5.78e-05, 6.46e-05) |
| **PBS** | (-0.0005, 0.0003) | (0.00011, 0.00012) |
| **PEJ** | (-0.0003, 0.0005) | (0.00012, 0.00013) |
| **XLP** | (8.1e-05, 6.0e-04) | (4.07e-05, 4.55e-05) |
| **XLY** | **(-0.0001, 0.0006)** | (1.07e-04, 1.20e-04) |
| **IGN** | (-0.0003, 0.0006) | (0.00013, 0.00014) |

In terms of the means, we see that the confidence intervals were small and most of them were centered around zero with the exception of BJK and FUD, whose upper bounds were below zero. This might imply that these two ETFs are not great choices for long-term investments, because if we drew enough amount of large samples of BJK or FUD log-returns, we would get a mean return below zero 95% of the time.

On the other hand, XLY appears to be a relatively better investment with a position mean log-return. In fact, if we did a 90% confidence interval for XLY, the whole interval would be entirely positive. The confidence intervals for the variances were even smaller. This make intuitive sense because the S&P 500 and ETFs were intrinsically less varied than individual stocks.

By converting log-returns and dates into time-series type, we then performed time series regressions of log-returns against time. With no exception, the R^2 values were close to 0 suggesting poor fit for the data with linear regressions. Even though we might not be able to use linear regressions to predict future log-returns for potential investors, several coefficient estimates appear to be positive and highly significant with high t-value. The regression results for BJK, FUD, and PAGG are shown below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | -2.629e-03 | 4.747e-04 | -5.540 | 3.35e-08 \*\*\* |
| tstime\_PAGG | **1.786e-06** | 3.289e-07 | 5.431 | **6.13e-08 \*\*\*** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | -1.800e-03 | 4.254e-04 | -4.231 | 2.42e-05 \*\*\* |
| tstime\_FUD | **9.884e-07** | 2.947e-07 | 3.353 | **0.00081 \*\*\*** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | -2.170e-03 | 5.057e-04 | -4.291 | 1.85e-05 \*\*\* |
| tstime\_BJK | **1.238e-06** | 3.504e-07 | 3.533 | **0.000419 \*\*\*** |

Take BJK as an example, the coefficient estimate of 1.238e-06 that one extra trading day is linked to a 1.238e-06 increase in the value of daily log-return on average. In other words, there is a trivial but upward trend for the three ETFs listed above.

**Testing Equality of Conditional Means**

Firstly, to distinguish years in which the original NFL teams win the Super Bowl from the others, we define the year dummy variable as in the below table:

|  |  |
| --- | --- |
| Year = 1 | 2009, 2011, 2012, 2018 |
| Year = 0 | 2010, 2013, 2014, 2015, 2016, 2017 |

Then for the S&P 500 and each individual ETF, we computed the means of log-returns conditional on Year = 1 or Year = 0. We assume that the two newly generated samples were still consistent with a random sample and came from normal a normal distribution. Then we can start testing if the two conditional means are equal. Specifically, we want to see whether the mean for Year = 1 is greater than that for Year = 0, so it seem reasonable to use the following test:

H0: 𝝁1 > 𝝁0 versus H1: 𝝁1 <= 𝝁0

If we assume that the unknown variances for Year = 1 and Year = 0 are equal (there is no reason to assume otherwise), we can use the pooled estimator of the common variance and a t-distribution with n+m-2 = 2497 degrees of freedom to test the null hypothesis. If we suppose the conditional variances are not only unknown but also not necessarily equal, then we can utilize the Welch two sample t-test considering that the sample size for both populations are large.

The results were as expected mixed. We were able to reject the null hypothesis for PAGG and BJK with extremely low p-values and conclude that log-returns for these two ETFs were not higher in the years in which an original NFL team wins the Super Bowl. We were not able to reject the null hypothesis for any of the other ETFs at a significance level lower than 10%. Notice, however, not being able to reject the null hypothesis does not necessarily lead to the conclusion that 𝝁1 > 𝝁0.

**Regression on Dummy Variable**

Another way to see if whether an original NFL team winning the Super Bowl has any effect on the log-returns is to use a linear regression on the dummy variable. We employed the following regression specification:

log-return = a\*Year + e

whereYear corresponds to the dummy variable defined in the previous section, and e is the residual.

After running the regression for the S&P 500 and ETF log-returns, we found that, surprisingly, all of the coefficient estimates for the dummy variable with the exception of PEJ were negative. Only the estimates for BJK and PAGG were significant with t-values of -2.886 and -3.943, respectively, and both were negative.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| BJK | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | -3.233e-05 | 3.253e-04 | -0.099 | 0.92086 |
| dummy\_year | **-1.493e-03** | 5.174e-04 | **-2.886** | 0.00394 \*\* |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PAGG | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 0.0003620 | 0.0003060 | 1.183 | 0.237 |
| tstime\_BJK | **-0.0019185** | 0.0004866 | **-3.943** | 8.28e-05 \*\*\* |

Take the PAGG regression results above as an example. The coefficient estimate of -0.0019185 implies that, on average, log-returns of PAGG is 0.0019185 lower in years in which an original NFL team wins the Super Bowl than in the other years, and it is significant at 0.01% level! Note that the results for this two

**Conclusion**

From our runs test and normal probability plots, we first conclude that 8 out of 10 log-return samples were consistent with a random sample, but only 6 of them seem to come from a normal distribution. We continued with our analysis nevertheless.

There are week signs of linearity between log returns and time, but BJK, FUD, and PAGG do exhibit small upward trends, which are worth further exploring. Using the confidence intervals for mean and variance of log-returns, we found that longing XLY and shorting BJK and FUD might make sense from a long-term investment perspective. Again we are not sure whether the magnitude of the gain would consistently beat the market average.

Testing the equality of means conditional on NFL original team winning years, we positively concluded for BJK and PAGG that their log returns during those years were not higher. We were not able to reach any definitive conclusions for the S&P 500 and the other ETFs.

Finally, 9 out of 10 coefficient estimates from a linear regression on a single dummy variable were negative. Therefore we can say that, at least at face value, the Super Bowl indicator does not predict the direction of the market based on the 10 index and ETFs we selected. Yet only two estimates were significant, which further confirmed our results from testing the equality of conditional means. As a result, if there is one investment strategy that is truly worth considering based on this report, it is to short BJK and PAGG during years in which an original NFL team wins the Super Bowl.

**References**

[**https://www.mlive.com/business/west-michigan/index.ssf/2009/03/what\_are\_the\_old\_wives\_saying.html**](https://www.mlive.com/business/west-michigan/index.ssf/2009/03/what_are_the_old_wives_saying.html)

<https://www.quora.com/What-are-the-original-eight-NFL-teams>