Pairs Trading Using Cointegration

Jesse Haulk
jesse.haulk@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia

ABSTRACT

By using cointegration, one asset's value can be determined by another asset. Using this information, a pair of assets can be traded against each other to hedge against potential losses. Here, Coca-cola (KO) and PepsiCo (PEP) were chosen as potential trading pairs since they're in the same industry space. The spread was found to be mean reverting, which is an expected behavior for potential pairs. However, it was found that KO and PEP are not cointegrated.

1 INTRODUCTION

Two time series that can be connected together in a linear combination are cointegrated[4]. In finance, cointegration can allow for insights in one asset based on the performance of another asset[3]. Pairs trading tries to take advantage of this by going long in one asset and short the other[2] [1]. Then, as long as the higher asset increases in value faster than the lower asset, or the lower asset decreases in value faster than the higher asset, money has been made. Divergence from the mean of the two assets therefore makes money, while convergence to the mean loses money.

Finding two assets which cointegrate is not a trivial task. While it is possible to create some for loop through assets to find cointegrated assets, the issue becomes p-hacking or the multiple comparisons problem. As such, the pairs chosen should have an economic reasoning that they would be tied together. For example, looking at vertical markets, or supply chain linked assets.

2 METHODOLOGY

The two assets chosen were Coca-cola (KO) and PepsiCo (PEP). Since both assets are competitors in the same market segment, there could likely be some cointegration. From the candle charts in Figure 1 and Figure 2, they do appear to follow the same behavior.

The returns for each asset were then plotted, shown in Figure 3, to get them both on the same scale for better comparison. The assets have similar returns and follow the same general pattern. An important note is one asset is not always higher than the other asset. By plotting the spread, or difference in the returns for KO and PEP, it can be determined if the spread is mean reverting, as shown in Figure 4. As shown, the spread doesn't spend long amounts of sequential time on one side of the mean. If the spread is large, the spread will get smaller on the next day, and vice versa. Figure 5 shows the ACF and PACF plots for the spread and there's strong evidence is the spread is stationary. An augmented Dickey-Fuller

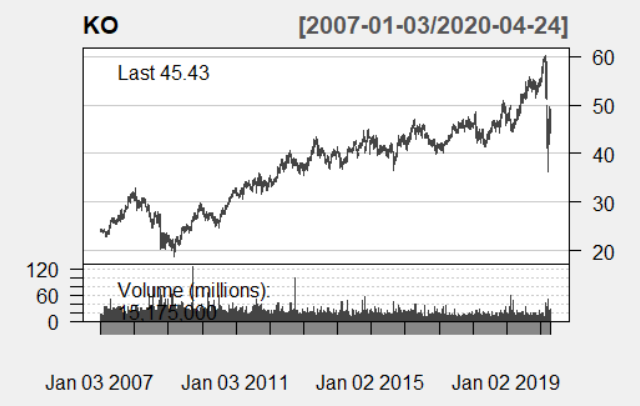

Figure 1: Candle Chart of Coca-cola

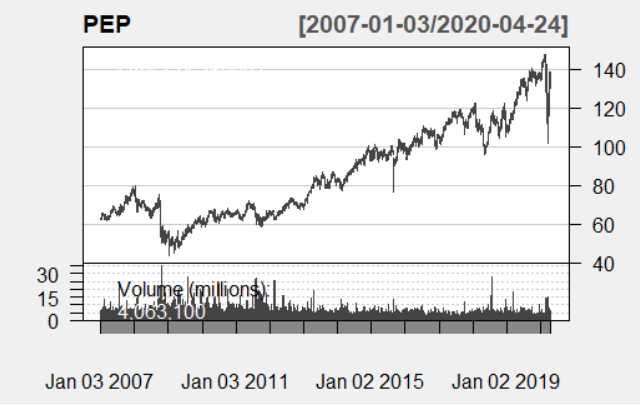

Figure 2: Candle Chart of PepsiCo

Test is more evidence of stationarity with a small p-value. So the spread is likely mean reverting.

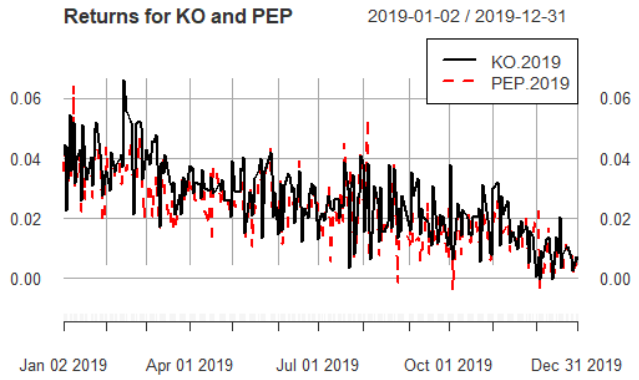

Figure 3: Plot of the Returns for KO and PEP

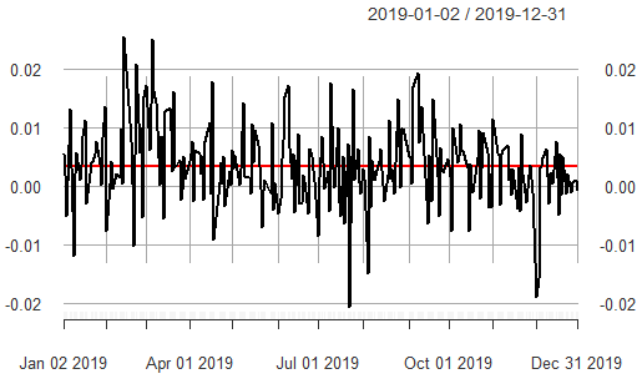

Figure 4: Plot of the Spread with the Mean

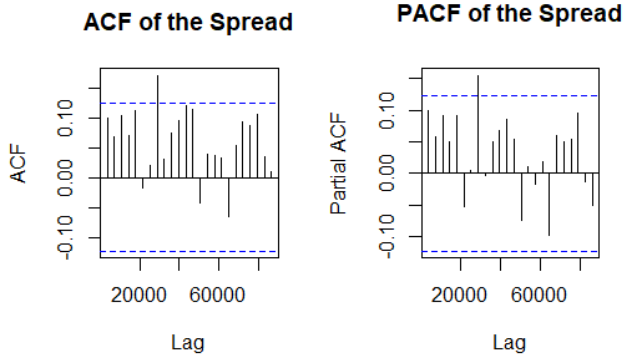

Figure 5: ACF and PACF of the Spread

The data was subset on just the year 2019 returns, and the January through November data was used in the training set and the December data used for the testing set. With the training set, an ARIMA model with seasonality was fit to the spread resulting in an ARIMA(1,0,1), or ARMA(1,1).

From the plot of the residuals, Figure 6, there appears to be constant variance. The ACF of the residuals, Figure 7, shows no autocorrelation. The histogram of the residuals, Figure 8, show that they're unimodal and approximately normally distributed. The Normal Q-Q Plot, Figure 9, shows that the residuals are approximately normal with fat tails. Performing a Box-Ljung test on the residuals is more evidence of no autocorrelation in the residuals.

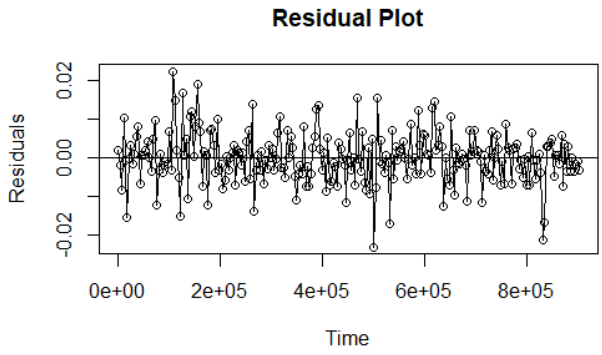

Figure 6: Plot of the Residuals over Time

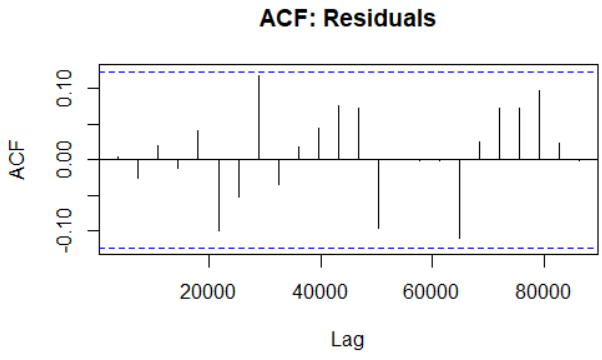

Figure 7: ACF of the Residuals

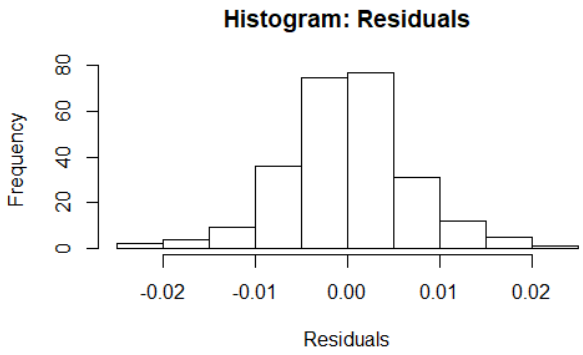

Figure 8: Histogram of the Residuals

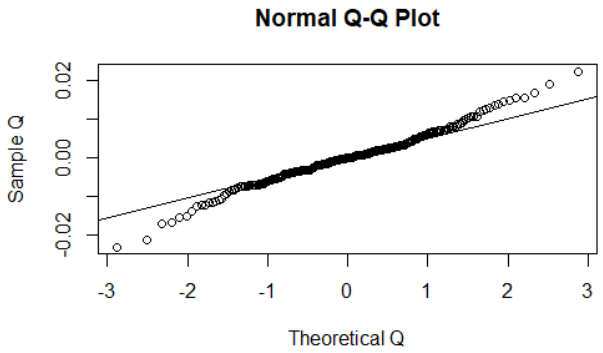

Figure 9: Normal Q-Q Plot of the Residuals

Using the ARMA(1,1) model, a forecast was developed for the December data. As shown in Figure 10, most of the observations in December fall within the 95% prediction window.

Finally, a vector autoregressive model was developed using the 2019 data. The high p-values rejects the hypothesis that KO and PEP are cointegrated.

3 CONCLUSION

The spread was found to be mean reverting and a model for the forecast appears to be appropriate. However, it was determined that KO and PEP are not cointegrated and wouldn't be good candidates for pairs trading. A possible reason for this would be a confounding factor, for example, both KO and PEP are linked to the overall market instead of to each other.

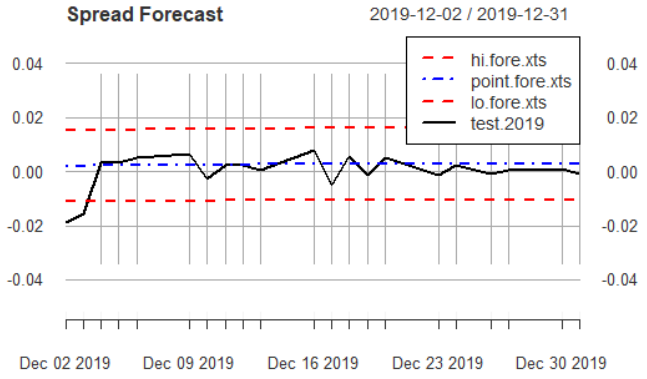

Figure 10: December Forecast of the Spread

REFERENCES

[1] Carol Alexander. 1999. Optimal hedging using cointegration. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 357, 1758 (1999), 2039–2058.

[2] Carol Alexander, Ian Giblin, and Wayne Weddington. 2002. Cointegration and asset allocation: A new active hedge fund strategy. *Research in International Business and Finance* 16, 5 (2002), 65–90.

[3] Robert F. Engle and C. W. J. Granger. 1987. Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica* 55, 2 (1987), 251–276. http://www.jstor.org/stable/1913236

[4] Søren Johansen. 2015. *Time Series: Cointegration* (second edition ed.). Elsevier Ltd.