Gordon Miller

Cointegration of Indices Analysis

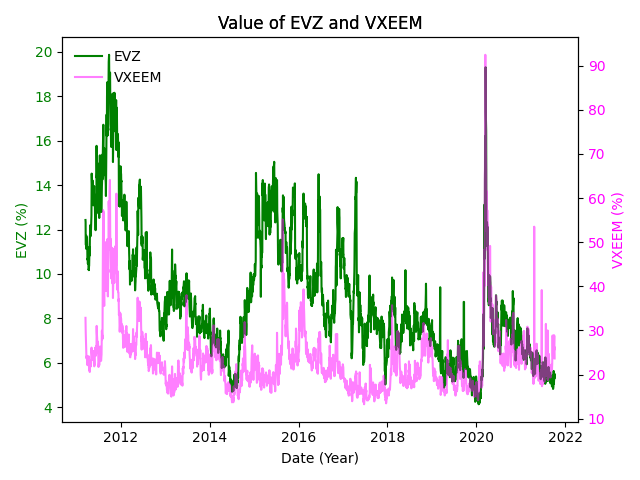
For this assignment I’ve conducted statistical analyses on two volatility indices and attempted to forecast these indices future values. The first of these indices is the CBOE EuroCurrency ETF Volatility Index (EVZ), which tracks the volatility of the EuroCurrency Index and is measured as a percentage or on a scale of 0 to 100 in points. In finance volatility is a measure of dispersion of the returns of a security or index and is used to gauge risk. The second time series used for this assignment is the CBOE Emerging Markets ETF Volatility Index (VXEEM). VXEEM is similar to EVZ in that it tracks volatility of an index. VXEEM tracks the volatility of the MSCI Emerging Markets Index and is also measured in percent or on a scale of 0 to 100 in points. The time period available for EVZ spans from 2007-11-01 to 2021-10-07. The time period available for VXEEM spans from 2011-03-16 to 2021-10-07. Both datasets end on 2021-10-07 since that is the date I downloaded the data. Below is an output from my code that describes the number of observations for both time series, the merged data frame of the time series, and the merged data frame after dropping any NA values:

**EVZ number of observations: 3636**

**VXEEM number of observations: 2757**

**Number of observations for merged dataframe of both series: 2757**

**Number of non-missing observations for both series: 2660**



Above is a line chart that tracks the value of EVZ and VXEEM from 2011-03-16 to 2021-10-07. The left y-axis includes the % values for EVZ and the right y-axis includes the % values for VXEEM. This chart helps us understand both of our time series much better. For instance, we can clearly see that VXEEM’s underlying index seems to be much more volatile than EVZ’s underlying index due to VXEEM’s higher values compared to EVZ. It also seems like these time series may be somewhat correlated to the naked eye. We can also note that these indices’ movements were quite similar during most of 2020, so much so that they overlap.

Chart

Description automatically generated

Above is a plot of 21-day rolling correlations between EVZ and VXEEM overtime. The reason I used 21 days is because that is the number of trading days for each month, and it is helpful to observe the correlations by month. Its clear from the plot that there is a lot of variation in the rolling correlations over time, but what isn’t clear is what the average rolling correlation between the two series is. After applying the max, min, and mean functions on the rolling correlation data I found that the minimum rolling correlation is -0.767, the highest is 0.980 and the average rolling correlation is 0.438. Following this analysis of the rolling correlations we can confidently say there is a positive correlation between the two series, however, there is a fair amount of variation in these correlations over time.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Whole Sample** | | | **Split Sample 1** | | | **Split Sample 2** | | |
|  | EVZ | VXEEM |  | EVZ | VXEEM |  | EVZ | VXEEM |
| count | 2660 | 2660 | count | 1330 | 1330 | count | 1330 | 1330 |
| mean | 8.890598 | 23.44086 | mean | 10.37875 | 25.00971 | mean | 7.402444 | 21.87202 |
| std | 2.85689 | 7.609283 | std | 2.962073 | 7.519202 | std | 1.767306 | 7.374364 |
| min | 4.13 | 13.28 | min | 4.69 | 13.71 | min | 4.13 | 13.28 |
| 25% | 6.85 | 18.6175 | 25% | 8.1725 | 20.3725 | 25% | 6.21 | 17.4875 |
| 50% | 8.19 | 21.965 | 50% | 9.9 | 23.47 | 50% | 7.3 | 20.195 |
| 75% | 10.2825 | 25.82 | 75% | 12.6025 | 26.97 | 75% | 8.19 | 24.1475 |
| max | 19.87 | 92.46 | max | 19.87 | 64.1 | max | 19.31 | 92.46 |

Above is a table of the descriptive statistics for the whole sample of data for EVZ and VXEEM, the 1st split sample and the 2nd split sample. I decided to split my sample in half rather than splitting into 2 thirds and one third because of higher variability in more recent data due to COVID. By splitting my data in half the 2nd split sample has a more accurate estimate of its standard deviation since the higher variability portion of the data isn’t weighted as heavily. When comparing the mean and standard deviation between the two split samples for EVZ and VXEEM there seems to be some fair differences. The mean of split sample 2 for EVZ is 7.40 which is about 25% lower than the mean of 10.38 in sample 1. The standard deviation for EVZ in sample 2 has surprisingly decreased despite COVID taking place in that samples data. On the other hand, VXEEM’s mean doesn’t change too much between the two samples and the standard deviation is surprisingly very similar. However, we can’t come to any conclusions without doing some t tests and f tests. Below are the codes outputs for these tests:

**t-statistic (EVZ) = 31.4688 p-value = 0.0**

**F (EVZ) = 2.8091 p-value = 0.0**

**t-statistic (VXEEM) = 31.4688 p-value = 0.0**

**F (VXEEM) = 1.0397 p-value = 0.2391**

After completing the t test and f test between the two samples it seems that we will reject the null hypothesis that there is no difference between the two means for both data sets. We can also reject the null that there is not difference in variances for EVZ but not for VXEEM. From these results it appears that our data is not stationary, however, this could be due to the increased volatility in split sample 2 due to COVID-19. If retested on data that doesn’t include high volatility periods during 2020 we may come to a different conclusion.

After completing an ADF test we get the output below:

Untransformed ADF test statistic for EVZ = -4.0508 p-value = 0.0012

Untransformed ADF test statistic for VXEEM = -6.3726 p-value = 0.0

Before we can use an autoarima model or test for cointegration we must ensure our data is stationary, so I conducted an ADF test. Surprisingly enough we are going to reject the null hypothesis that the series is a unit root for both series, meaning that the series are in fact stationary. However, I was curious as to how much I could improve these test statistics and it seemed that differencing was the best transformation to use based on the trend of the data found by using decomposition. Afterwards I got the following results from the ADF test:

**First difference transformed ADF test statistic for EVZ = -56.3133 p-value = 0.0**

**First difference transformed ADF test statistic for VXEEM = -63.2902 p-value = 0.0**

**coint t-statistic: -3.4963099974492273 coint p-value: 0.03272390376215259 critical values for t-statistic: [-3.90056355 -3.33842886 -3.04604542]**

Lastly, is the test for cointegration which outputted the values above. We can conclude that EVZ and VXEEM are in fact cointegrated meaning they have a mean reverting relationship that can be statistically exploited for pairs trading. Because these series are correlated in the long term if they ever diverge from this relationship it would give us a trading opportunity to short one and long the other.