Certificate in Quantitative Finance

Final Project

Long\Short Trading Strategy Design & Backtesting

Author:
Boyan Davidov, *CFA, CAIA, FRM*

January 25, 2021

Abstract

The purpose of this report is to study the idea of cointegration and examine trading strategy based on this property of time series. Cointegration describes steady long-term linear dependency of two (or more) different time series, such that they do not diverge from each other without a bound. The basic idea of a cointegration based trading strategy is to exploit a stationary spread by entering a position when the spread deviates from its usual bounds, whereas arriving again in ranges of the spread resp. signals the time to close the position. We have analyzed here a trading opportunity in the pair of volatility ETF and ETN. Before we illustrate the strategy we have introduced in the report the essential statistical concepts such as autocorrelation and mean-reversion, and also demonstrated how to apply different techniques to identify cointegration or to transform time series in order to make them stationary. Last but not least, backtesting and performance evaluation have been discussed, as well as caveats when exploiting supposed arbitrage opportunities.

Contents

1 Introduction **4**
1.1 Dataset and screening pairs . . . 4

2 Random Walks and Stationarity **6**
2.1 Mean-Reversion . . . 7
2.2 Autoregressive (AR) Time-series Model . . . 7

3 Unit Roots and Statistical Tests to detect it **8**

4 Identifying Cointegration **11**
4.1 Engle–Granger Method . . . 11
4.2 Error Correction Model . . . 15
4.3 Johansen Procedure . . . 15

5 Trading Strategy **17**
5.1 Optimization . . . 19
5.2 Performance Evaluation . . . 20
5.2.1 In-Sample Performance . . . 20
5.2.2 Out-of-Sample Performance . . . 21
5.3 Regime Changes . . . 22

6 Conclusion **24**

References **24**

A Appendix **25**
Multiple Regression . . . 26
Autoregression models . . . 26
Dickey-Fuller's test . . . 26
Information Criteria . . . 27
Johansen Test . . . 27
Performance Analytics . . . 27

B Appendix **28**

1. Introduction

Financial markets are in essence highly co-dependent. The classical approach that many investors and practitioners take when studying markets data and projecting future performance of assets is to look for correlation in returns. In contrast to this approach, cointegration based strategies aim to model the co-movements in prices, rates or yields. We say that two time series are cointegrated if in the long-run they do converge to certain bounds due to existing financial or economic link between them. In 1994 Michael Murray[1] made his famous comparison of the concept of cointegration with the tale of the drunk and her dog. In Stochastic Finance "drunkard's walk" illustrates a random walk. So is the walk of the drunkard's dog in the short term. After some time, however, the dog looks for her owner and in the end their paths converge. This humorous example is an intuitive explanation of the idea of a pair of assets that form a stationary series.

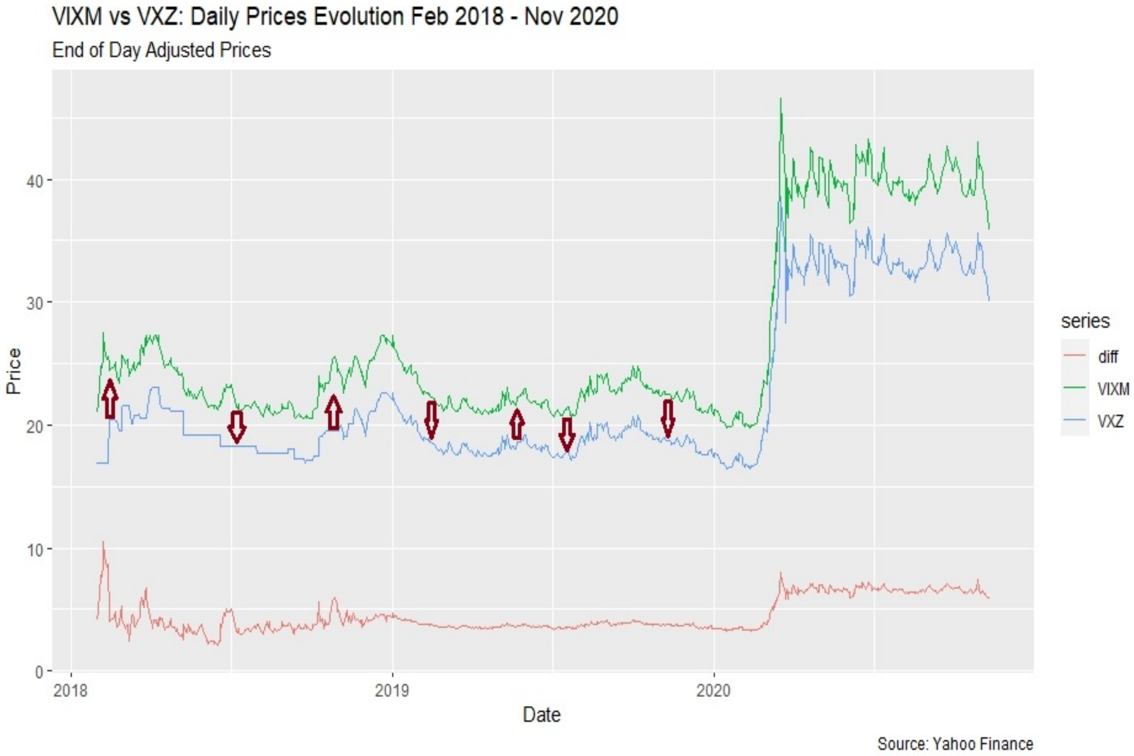
Our objective in this report is to outline the issues one can face when forecasting particular time series and build a trading strategy on the idea of mean-reversion. After this brief introduction we try to follow a roadmap of answering - what (definitions, and concepts), why (technical bits elaborating why does the method work), how (to create and implement the strategy) and finally where (practical bits) to use cointegration. Having this in mind, the outline of the project is as follows:

- In the next section 1.1 we describe the candidates for cointegration and the dataset
- Chapter 2 examines Random Walks and the concept of Stationarity
- Chapter 3 illustrate the statistical techniques to test for Stationarity
- Chapter 4 analysis Cointegration identification and Engle-Granger Method. Section is devoted to the Johansen Procedure and Multiple Cointegration
- Chapter 5 illustrates the Trading Strategy with its Backtesting and Performance Evaluation. Section is dedicated on regime changes and example with impact of the recent Covid crisis
- Chapter 6 concludes and summarizes briefly the findings
- Appendix A describes the mathematical theory behind the methods applied
- Appendix B summarizes the results when testing the strategy in particularly volatile year of 2018 (contrasted to 2020 which was rather "shock" volatility)

1.1 Dataset and screening pairs

In this report, we examine trading strategy on volatility ETF and ETN and more specifically the ProShares VIX Mid-Term Futures ETF (VIXM) and iPath Series B S & P 500® VIX Mid-Term Futures ETN (VXZ). Volatility ETFs provide exposure to the asset class volatility by following particular volatility index. Often referred to as "fear" indicators, they tend to move in the opposite direction of the broad market. Thus, these funds are used primarily by traders looking to capitalize on sharp market downturns. Both ETFs trade in the U.S. markets, VIXM launched on 01/04/11, while VXZ debuted on 01/17/18. VIXM being twice larger in Asset Under Management ($73.41M vs $37.92 AUM in VXZ) is also more frequently traded with Average Daily Volume of $2.92M (vs $0.963M of VXZ). The products are open-ended, quoted and traded in USD, and providing daily liquidity with low transaction fees. Given that they both track the same underlying asset (VIX Mid-Term Futures) it is likely that the respective time series will have a strong cointegrating relationship. As these products offer similar exposure, their prices co-move and this is clearly seen when plotting them. Below is the graph of the two time series from the inception date of VXZ - January 17, 2018 until end of November 2020.

[1] A Drunk and Her Dog: An Illustration of Cointegration and Error Correction



Indeed, even without normalizing prices, the time series follow similar path. This is evident also from the difference (spread) in closing prices, which has stayed almost unchanged in the period January 2019 - January 2020. This can be already used as a hint where to look for cointegration.

Later in implementing our strategy, we follow the standard approach of splitting the time period into in-sample and out of sample periods (i.e. train/test spans). Unfortunately, no clear-cut basis exist that tells how long period to take for examination. This would depend on the data analyzed and more specifically whether there are any regime changes. For instance, as it can be seen from the plot above, there is clear jump in the time series in March 2020. We will discuss this topic in separate chapter later, but one can already assume that the spread features after March 2020 are different than those in period before. In general, choosing longer train/test periods results in more trading signals, however, the link between the time series might have changed and the strategy needs to be re-calibrated. Bearing this in mind, in this report we split the data into one year train period and two months test period (i.e. ratio of 85:15 train/test data).

In addition to cointegration pair candidates VIXM and VXZ, we included also data for iPath Series B S&P 500 VIX Short-Term Futures ETN (VXX) and ProShares VIX Short-Term Futures ETF (VIXY) which offer exposure to short-term volatility. The purpose of this was to check multiple cointegration with the Johansen Procedure. In fact another good candidate for pair-wise cointegration turns out to be these short-term volatility products. The time series data downloaded from Bloomberg is included in separate csv file.

We should mention that considerable time was spent on screening possible cointegration relationships. There are many interesting cases of stocks, for instance the popular Palm/3com Puzzle [2] or Twin Shares Puzzle of the Royal Dutch A & B Shares [3], that logically should be perfect candidates for cointegration. There are also notorious examples of failed arbitrage opportunities due to some fundamental change in the link between considered assets. In general, it was easy to find pairs that are cointegrated, however, it was very difficult to find time series that remain stationary long enough that a solid strategy can be created on them. We have chosen the Volatility ETF/ETNs coincidentally rather with the aim to look for some short-term/mid-term link in volatility. Yet, it turned out interesting to examine related products. In general, the same approach can be utilized for any candidate and more opportunities can be sought time series of bitcoins, commodities, interest rates etc. However, searching for cointegrated ETN and ETF is not always a good idea. Despite being similar products (both are designed to track same underlying asset), ETF and ETN have a key difference. ETN is an unsecured debt note which can be held to maturity or sold at will. So firstly, there is also credit risk involved when trading ETN, secondly (and more important) ETN are less liquid and the strategy depends on possibility to sell in secondary market (large blocks (institutional sizes) ETN can be redeemed only on weekly basis in contrast to ETFs). The VXZ note is sufficiently liquid (with Average Daily Volume of $0.963M), nonetheless, one should be wary of capacity constraints when looking for such trading opportunities.

Before we implement our cointegration based investment strategy, we need to introduce to concepts of stationarity and mean-reversion which is covered in the next section.

[2] A solution to the Palm–3Com spinoff puzzles, Cherkes/Jones/Spatt (2015)
[3] Examples of Dual-listed Companies and Puzzles in Their Pricing, Bedi/Richards, Tennant (2003)

2. Random Walks and Stationarity

Financial analysts often make use of regressions to explain the past and to predict the future of a time series. Unfortunately, a common mistake is that they often omit the examination of how a model can change over time and whether the assumptions of say linear regression model are satisfied. Random walks are such case where one should not use linear regression methods. A random walk is a time series such that the value of the series in one period can be expressed as sum of the value from previous period and unpredictable random error:

$$x_t = x_{t-1} + \varepsilon_t, where\, E(\varepsilon_t) = 0, E(\varepsilon_t) = \sigma^2, Cov(\varepsilon_t, \varepsilon_s) = E(\varepsilon_t, \varepsilon_s) = 0 for\; t \neq s \tag{2.1}$$

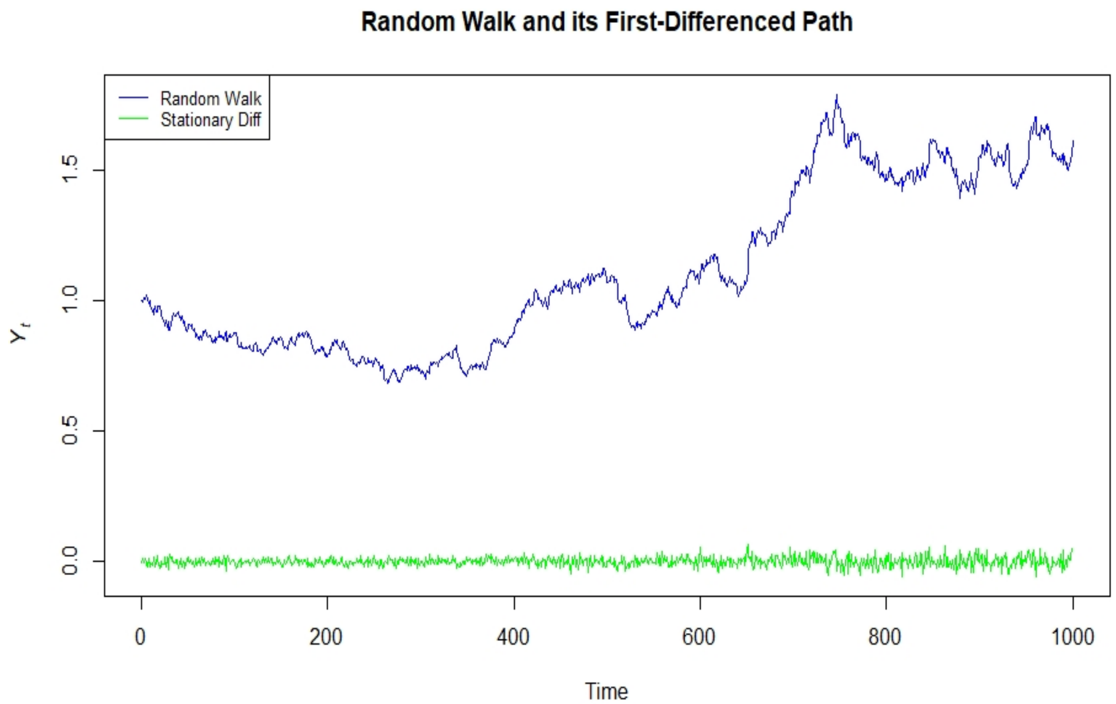
Random pattern can be found in many financial time series. Important characteristics of random walks is that their mean and variance are not constant. We also say that random walk is not a covariance-stationary time series. As such, one can not conduct valid statistical inference. This doesn't mean per se that the data is useless, however, one needs to convert the data to covariance-stationary before claiming that forecasted results have any economic meaning.

A stationary series must satisfy the following principal requirements:

1. $E(y_t) = \mu$, where $|\mu| < \infty, t = 1, 2, ...T$, i.e.finite and constant expected value for all time periods

2. $Cov(y_t, y_{t-s}) = \lambda$,where $|\lambda| < \infty, t = 1, 2, ...T; s = 0, \pm1, \pm2, ... \pm T$,i.e. finite and constant variance for all time periods and covariance for fixed number of periods

Transforming time series model so that it is well specified for linear regression can be simply achieved by differencing it. By subtracting the value of the time series in the first prior period from the current value of the time series we create new time series $y_t = x_t - x_{t-1} = \varepsilon_t$ where the mean is now constant, since $E(\varepsilon_t) = 0$ and $E(\varepsilon_t) = \sigma^2$. In fact, this means that the series has a mean-reverting level of 0. Below we have plotted simulated random walk and its first-differenced transformation.



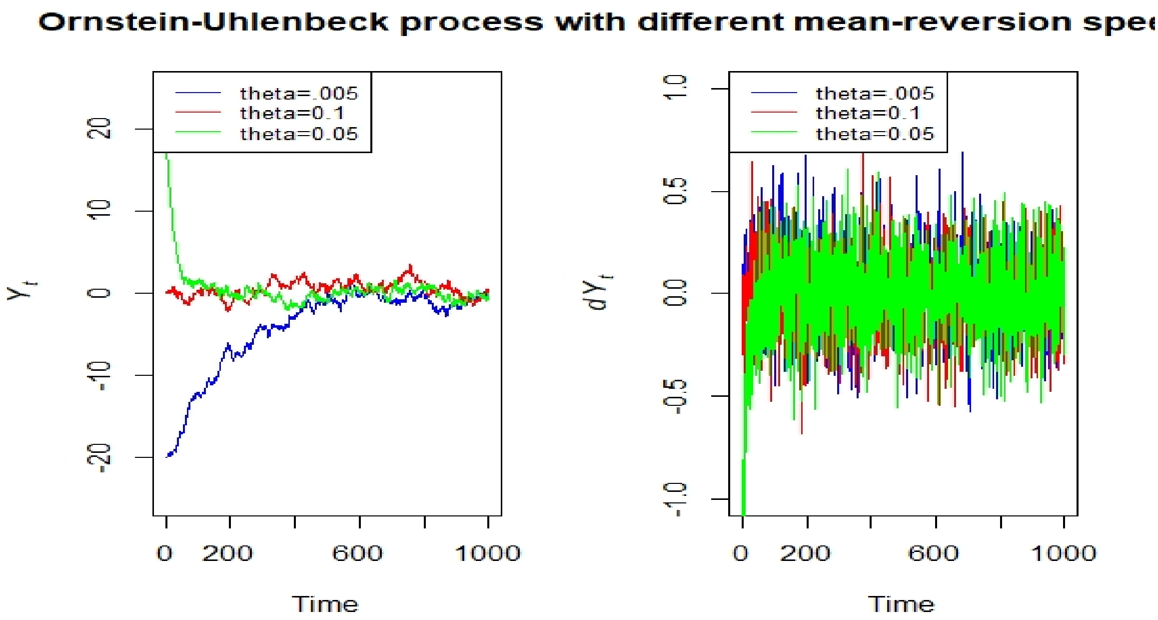
We see how smooth the transformed data is in comparison to the simulated random walk.

2.1 Mean-Reversion

Another important property of stationary time series that needs to be mentioned is mean-reversion - that is when time series tends to fall when its value is above its mean and rise when its value is below its mean. Ornstein-Uhlenbeck (OU) is a popular model (also known as Vasicek Model) in Stochastic Finance that exhibits this property:

$$dY_t = \theta(\mu - Y_t)dt + \sigma dW_t \tag{2.2}$$

where W_t denotes Wiener Process, μ and σ are the long term mean and variance resp. and θ is the speed of reversion. We have simulated and plotted below three OU processes with mean 0 but different speeds of mean reversion:



We see that all three processes tend to return to the long-term mean. On the plot above, in the right we have put also their first-differenced transformations. It is clear that those have reached the mean much faster than the original time series.

2.2 Autoregressive (AR) Time-series Model

Autoregressive models are models where current-period values are related to previous-period values. A simple AR(p) model can be expressed as:

$$x_t = b_0 + b_1 x_{t-1} + b_2 x_{t-2} + \ldots + b_n x_{t-p} + \varepsilon_t \tag{2.3}$$

whereas a simple AR(1) model is:

$$\triangle Y_t = (b-1)Y_{t-1} + \varepsilon_t = \phi Y_{t-1} + \varepsilon_t \tag{2.4}$$

When AR(1) model has already reached its long-term mean it is logical to expect to $x_t = x_{t-1}$. Substituting above yields:

$$x_t = \frac{b_0}{1 - b_1} \tag{2.5}$$

Thus, AR(1) model would likely increase if its current value is below $b_0/(1-b_1)$, and decrease if its current value is above $b_0/(1-b_1)$.

To briefly summarize so far we have seen that a random walk can be transformed to a stationary process and stationarity allows us to make statistical inference. Furthermore, if we know that time-series is stationary it could be modeled in mean-reverting process, where one can expect that after the time-series diverges from its mean it will come back to equilibrium just like in the tale of the dog following its owner. Before we jump to the section, where we attempt to exploit this property, we need to describe a more formal approach of detecting Unit Roots which is another way of saying non-stationarity.

3. Unit Roots and Statistical Tests to detect it

We can explain what is known as the unit root problem in the context of AR(1) model. If a time series comes from an AR(1) model then to be covariance stationary, the absolute value of the lag coefficient b_1, must be less than 1 (see equation 2.5). This is important because in a scenario where two time series have a unit root (and both are not cointegrated) linear regression of one variable on the other would be meaningless since the error term in the linear regression will not be covariance stationary. The impact of this could be inconsistent regression coefficients and standard errors, and violated regression assumptions. Often, this is the case when in the data there are some outliers/pulses/level shifts, trends or seasonal deterministic pulses.

The examination of the autocorrelations of a time series at various lags is a well-known prescription for inferring whether or not a time series is stationary. Subsequently, optimal lag is chosen that eliminates randomness. Typically, for a stationary time series, either autocorrelations at all lags are statistically indistinguishable from zero, or the autocorrelations drop off rapidly to zero as the number of lags becomes large. There are various tests that can be used to detect non-stationarity such as the Augmented Dickey-Fuller (ADF) test, Phillips-Perron (PP) test, Kalman Filters etc. We follow here the standard approach with applying ADF test.

The Dickey-Fuller test examines the null hypothesis of whether a unit root is present in the autoregressive model AR(p) of a time series. Using the notation from previous section (equation 2.4), if we put $g = b-1$ testing for unit root is equivalent to $g = 0$. The value of test statistic $\frac{\hat{g}}{std.err(\hat{g})}$ is compared to the respective critical values for the Dickey-Fuller (DF) test.

The Augmented Dickey-Fuller test (ADF) is an extension to the classical Dickey-Fuller test adding the terms $(b_k - 1)\Delta Y_{t-k}$:

$$\triangle Y_t = (b-1)Y_{t-1} + \sum_{k=1}^{p}(b_k - 1)\triangle Y_{t-k} + \varepsilon_t \tag{3.1}$$

Relative to the basic DF test, this enables us to analyze larger and more complicated set of time series models. The augmented Dickey-Fuller statistic used in the ADF test is a negative number. The more negative it is, the stronger the rejection of the hypothesis that there is a unit root.

One approach to confirm statistically non-stationarity is to compute information criteria such as the Akaike Information Criteria (AIC) or the Bayesian Information Criteria (BIC) to check whether there is any significant lag of higher order. Below we show the BIC values for the first 9 lags of the simulated data (codes available in the R script provided).

```
       [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7]    [,8]    [,9]
p    1.0000  2.0000  3.0000  4.0000  5.0000  6.0000  7.0000  8.0000  9.0000
BIC -1.9170 -1.9518 -1.9329 -1.9253 -1.9127 -1.8868 -1.8625 -1.8503 -1.8330
R^2  0.9397  0.9430  0.9431  0.9435  0.9437  0.9430  0.9424  0.9425  0.9421
```

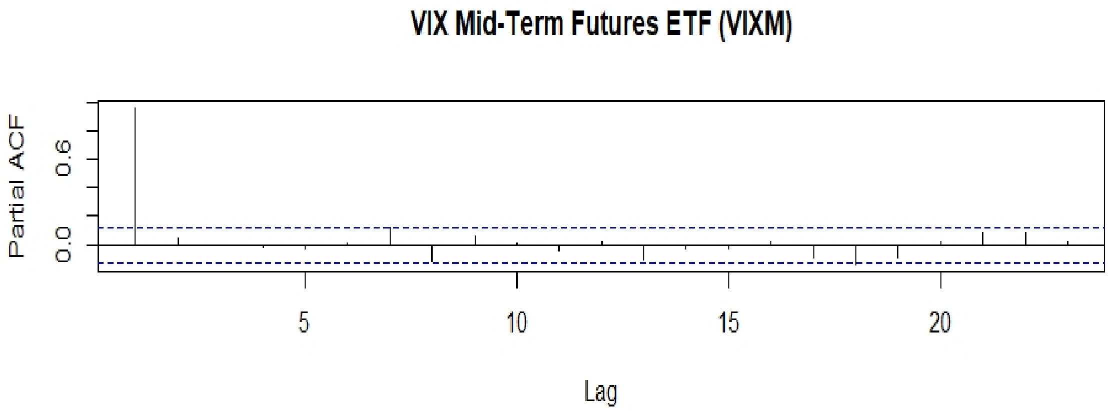
The slow decay of the autocorrelation function suggests the data follows a long-memory process. One should be wary that BIC or AIC do not necessarily pick the best model, it is rather they show the least worst model. This approach is less definite than a currently more popular approach of examining Partial Autocorrelation Function (PACF) plot, where significant lags appear above confidence limits. Next we are going to test for unit-root our real data in the one-year period 01/2019-01/2020 (the in-sample or train data).

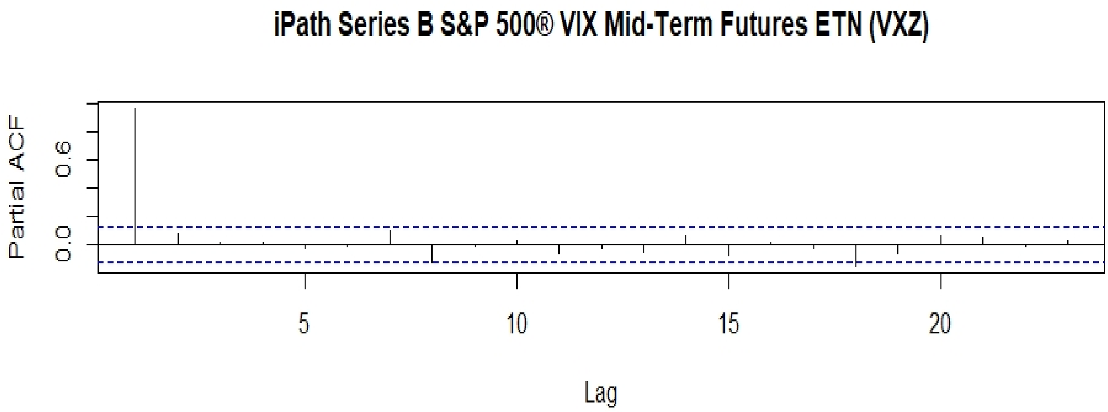
There are three different version of the ADF test, which are presented in the Appendix. We run here the pure Augmentedd Dickey-Fuller test without drift or trend included as we are aware that adding parameters might result in overfitted results due to temporary dependency. It can be noted from the R script that the results when using drift or trend tend to reject more often the null hypothesis of unit root. Here is the output when testing for unit root the in-sample data for VXZ.

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
z.lag.1    -0.001344   0.001042  -1.290   0.1983
z.diff.lag -0.114063   0.062532  -1.824   0.0693 .
---

Residual standard error: 0.3079 on 248 degrees of freedom
Multiple R-squared:  0.01893,        Adjusted R-squared:  0.01101
F-statistic: 2.392 on 2 and 248 DF,  p-value: 0.09355

Value of test-statistic is: -1.2899

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

The important coefficient we need to keep attention on is the **z.lag.1**. Its critical value (t value) is -1.29. This is less negative than the 5% critical value of -1.95, hence under 5% significance level we can not conclude that the time series is stationary. Next, we can make use of the partial autocorrelation function (PACF) to examine the spikes at each lag to determine their significance. The PACF is basically a measure of the correlations between x_t and x_{t-k} after adjustments for the presence of the lags of shorter lag within $x_{x-1}, x_{t-2}, ..., x_{t-k-1}$. On the PACF plot, there is a significant correlation at lag 1 followed by correlations that are not significant. This pattern indicates an autoregressive term of order 1.



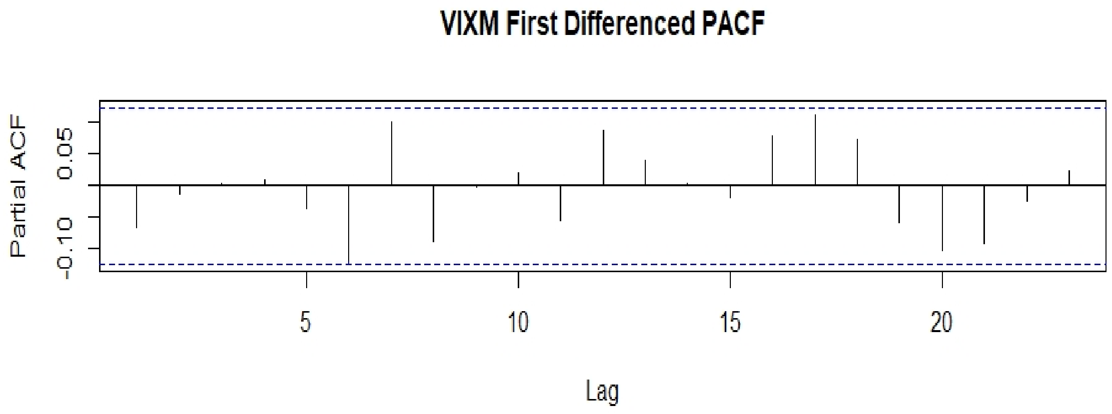


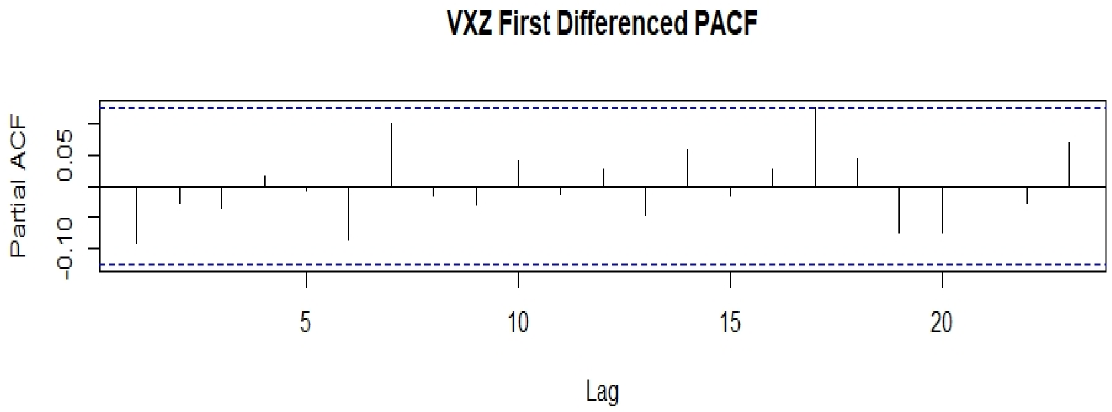
There is evidence of significant lag of order one in both VIXM and VXZ time series. In financial time-series there is rarely the need to add more than one lag to eliminate randomness. In fact, as we see in our case even lag one might be superfluous (eg. under 10% critical value). We can proceed, however, to see that when taken the first difference of the time-series there are no more significant lags. Firstly, the ADF test output after first differencing the VXZ (results similar also for VIXM):

```
Coefficients:
           Estimate Std. Error t value Pr(>|t|)
z.lag.1    -1.21809    0.09290 -13.112   <2e-16 ***
z.diff.lag  0.11375    0.06219   1.829   0.0686 .
---
```

```
Residual standard error: 0.3042 on 247 degrees of freedom
Multiple R-squared:  0.5576,        Adjusted R-squared:  0.554
F-statistic: 155.6 on 2 and 247 DF,  p-value: < 2.2e-16


Value of test-statistic is: -13.1119
```

We can conclude stationarity even under 1% significance level. This can be confirmed also from the new PACF plot:





The ACF and PACF show that all lags are now within the confidence bands. The lack of other persitent lags after taking first differences is a good sign that we can continue fitting AR(1). Last but not least, to confirm that we can fit an AR(p) model, it is necessary to check also the stability condition. This means that the eigenvalues of the estimated coefficients need to be smaller than one (i.e. to be inside the unit circle):

$$|\lambda I - \hat{b}| = 0 \tag{3.2}$$

Equivalently, we can check whether the roots of the characteristic polynomial of the AR(p) system are bigger than 1 (to be outside the unit circle). The original time-series of VIXM and VXZ have roots on the border of the unit circle, whereas the first-difference transformations have eigenvalues inside the unit circle which confirms stability. Summarized results of the analysis can be found in the next table:

Table 3.1: Testing for stationarity - VIXM and VXZ

Process	ADF t-stat	5% Critical Value	p-value	Stationary	Stable
VIXM Yt	-1.357	-1.95	0.0713	No	No
VXZ Yt	-1.29	-1.95	0.0935	No	No
VIXM dYt	-12.91	-1.95	$< 2.2e-16$	Yes	Yes
VXZ dYt	-13.11	-1.95	$< 2.2e-16$	Yes	Yes

The results prove again that most often successful is to model the first-differenced series as an autoregressive time series to eliminate randomness. Up to now, we have conducted time-series models test for one time series independently from another. In the next section we will establish whether the two time series VIXM and VXZ are cointegrated.

4. Identifying Cointegration

After the introduction of the fundamental concepts of stationarity and AR(p) models, we pass to the more technical part when more time series are involved together in the analysis, namely when cointegration is tested.

Let $Y_t = (y_{1t}, \dots y_{nt})^T$ denotes an (n x 1) vector of $I(1)$ time series. We say that Y_t is cointegrated if there exists an (n x 1) vector $\beta = (\beta_1, ..., \beta_n)^T$ such that

$$\beta^T Y_t = \beta_1 y_{1t} + \dots + \beta_n y_{nt} \sim I(0) \tag{4.1}$$

In other words, the nonstationary time series in Y_t are cointegrated if there is a linear combination of them that is stationary or $I(0)$. The linear combination $\beta^T Y_t$ is referred to as a long-run equilibrium relationship in economic theory. The intuition is that I(1) time series with a long-run equilibrium relationship cannot move away too far apart from the equilibrium because economic forces will act to restore the equilibrium relationship.[1]

So in case that both time series have unit root and are not cointegrated we are not allowed to use hypothesis tests on the regression coefficients because of inconsistent standard errors. However, in scenario where two time-series both have a unit root but are cointegrated the error term in the linear regression of one time series on the other will be covariance stationary. Accordingly, the regression coefficients and standard errors will be consistent, and we can use them for hypothesis tests. One should be still very cautious in interpreting the results of a regression with cointegrated variables since the cointegrated regression estimates the long-term relation between the two series but may not be the best model of the short-term relation between the two series. There are three main statistical tests to detect cointegration:

- Engle-Granger - the main method used in this report. There are two steps:
  1. Estimate the regression $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$. and the cointegrating error term $\hat{\varepsilon}$
  2. Test whether the error term from the regression in Step 1 has a unit root using a Dickey-Fuller test.
- Phillips-Ouilaris test - Because the residuals are based on the estimated coefficients of the regression, we cannot use the standard critical values for the Dickey–Fuller test. Instead, we must use critical values, which take into account the effect of uncertainty about the regression parameters on the distribution of the Dickey–Fuller test. These distributions are known as the Phillips-Ouliaris (PO). This test is used only as complement to the Engle-Granger method.
- Johansen test - it allows for multivariate time-series analysis, in contrast to Engle-Granger method which could be used only for single relationship. Even though not used in the described later trading strategy, we have discussed briefly how its output can be used.

4.1 Engle–Granger Method

If the (Engle–Granger) Dickey–Fuller test rejects the null hypothesis that the error term has a unit root, then we may conclude that the error term in the regression is covariance stationary and that the two time series are cointegrated. The parameters and standard errors from linear regression will be consistent and will let us test hypotheses about the long-term relation between the two time series. Once this has been ensured the time-series analysis is relatively straightforward and logical. Given that VIXM is larger in Assets Under Management and more liquid than VXZ, we can assume it is the leading variable and better candidate for chosen as independent variable in regression involving VXZ. We will see later how this can be inferred statistically. Firstly, we present the output from linear regression $VXZ_t = const + \beta * VIXM_t + \varepsilon_t$

```
Call:
lm(formula = etfs_insample[, "VXZ"] ~ etfs_insample[, "VIXM"])


Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)             0.095040   0.052133   1.823   0.0695 .
etfs_insample[, "VIXM"] 0.829712   0.002323 357.139   <2e-16 ***
```

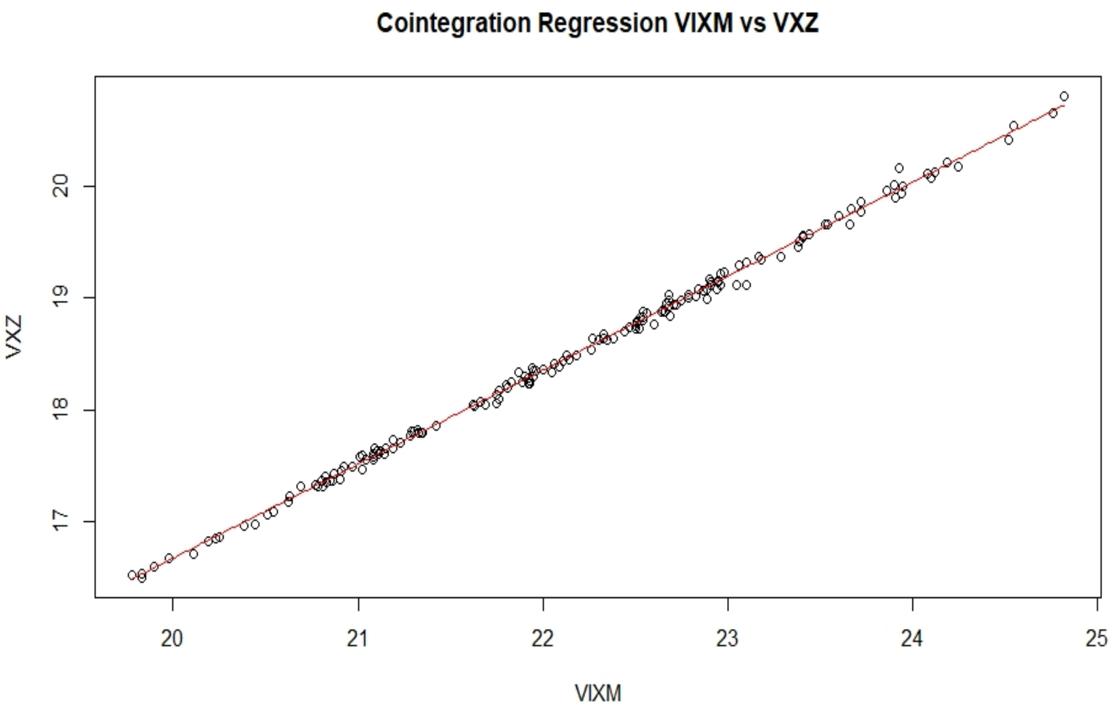
[1] Modeling Financial Time Series with S-Plus, Chapter Cointegration, E.Zhivot

```
Multiple R-squared:  0.998,	Adjusted R-squared:  0.998
```

The suggested equation (long-run equilibrium model) from the regression is:

$$V\hat{X}Z_t = 0.09 + 0.83 * VI\hat{X}M_t + \hat{e}_t \tag{4.2}$$

We can see that the OLS fit to the data is successful.

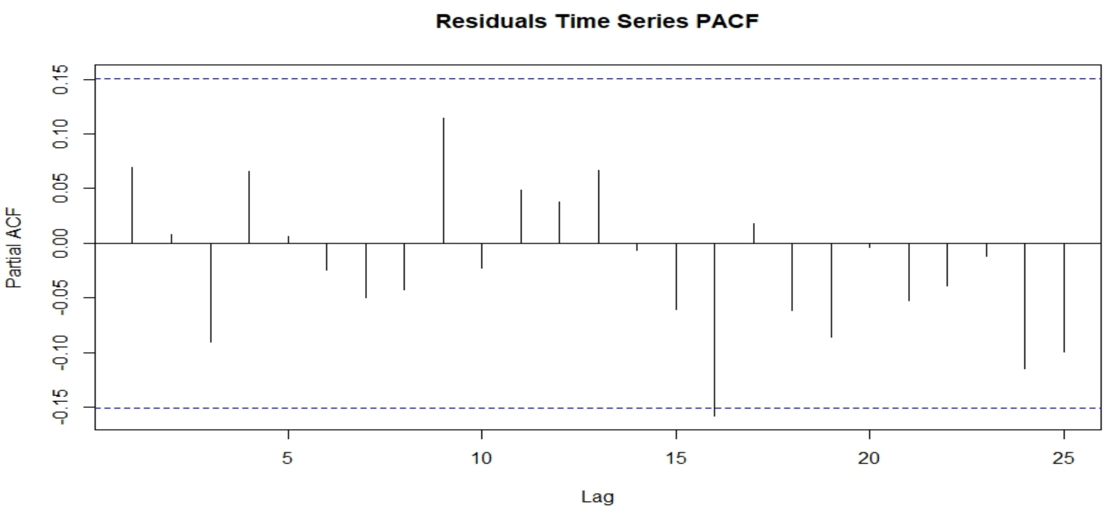


Another plot illustrates the real vs fitted values for VXZ:

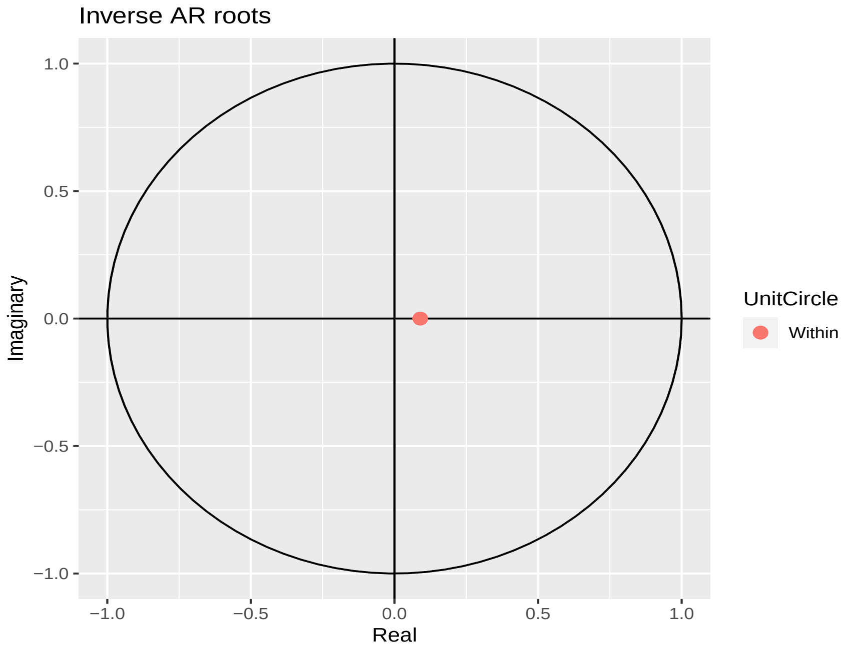
Again, it seems this is already very good match. Common pitfall encountered in practical applications is to conclude based on very high explanatory power the validity of the model. However, it is important to show that the relationship is not just a matter of optics, but it is indeed statistically reliable. For this purpose we do the second step of the Engle-Granger method, namely testing stationarity of the error-term $\hat{e}_t = V\hat{X}Z - 0.09 - 0.83 * VIXM_t$. We have applied Hansen's Covariate-Augmented Dickey Fuller (CADF) Test For Unit Roots to the residuals from our regression. The reason is that ADF test will only tell us whether, for a particular regression coefficient, the linear combination is stationary. In order to determine the optimal hedge ratio one can use Cointegrated Augmented Dickey-Fuller (CADF) test. It performs as well linear regression against the two time series and then tests for unit root under the linear combination:

```
################################################
# Augmented Dickey-Fuller Test Unit Root Test #
################################################

Test regression none

Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
      Min        1Q    Median        3Q       Max
-0.175191 -0.022255 -0.001798  0.026823  0.196651

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
z.lag.1    -0.78492    0.08311  -9.444   <2e-16 ***
z.diff.lag -0.10055    0.06323  -1.590    0.113
---

Residual standard error: 0.04356 on 248 degrees of freedom
Multiple R-squared:  0.4445,       Adjusted R-squared:   0.44
F-statistic: 99.22 on 2 and 248 DF,  p-value: < 2.2e-16

Value of test-statistic is: -9.4443

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

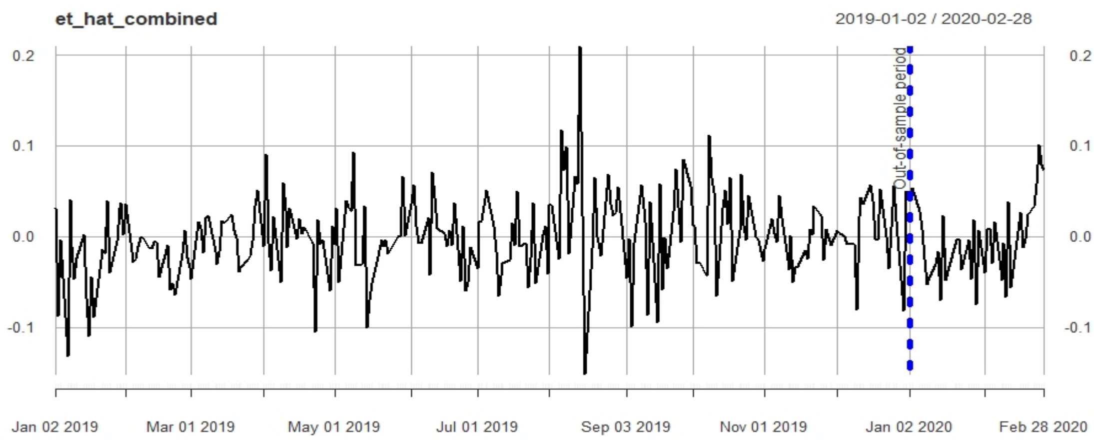
The CADF t-statistic (using one lag) confirms the stationarity of the error-term. The PACF plot also validates that there is no memory effect in the higher-order moments of the time series.



We see that the stability condition is also fulfilled:



Finally, the estimated cointegrating spread $\hat{e}_t$ series is plotted (train/test period splitted by the blue dotted line):



To make sure that we have the right choice of independent and dependent variable we can run Granger causality test. The null hypothesis in this test when running regression $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$ is x_t doesn't cause y_t

Table 4.1: Granger Causality Test

H_0	p-value
VIXM doesn't cause VXZ	0.0097
VXZ doesn't cause VIXM	0.0661

From the p-values we see that we can reject the null hypothesis that VIXM doesn't cause VXZ in 5% significance level test. This confirms that VIXM is the leading, whereas VXZ is the lagging variable.

4.2 Error Correction Model

Cointegration implies the existence of an Error Correction Model (ECM). This is another step to determine how variables are linked together by providing an adjustment to the long-run equilibrium from the short-run dynamics. This is particularly useful when modeling financial data, which can lead to spurious regression results due to non-stationarity. Dynamic regression is such that includes lagged values x_t and y_t:

$$y_t = \alpha y_{t-1} + \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \epsilon_t \tag{4.3}$$

Plugging the shifted error-term from the long-run equilibrium $y_t = b_0 + b_1 x_t + e_t$, it can be rewritten as:

$$\Delta y_t = \underbrace{\beta_1 \Delta x_t} - \underbrace{(1-\alpha)e_{t-1}} + \epsilon_t \tag{4.4}$$

where e_{t-1} is the lagged cointegrating spread from the equilbrium model. It is required to confirm the significance for $(1-\alpha)$.

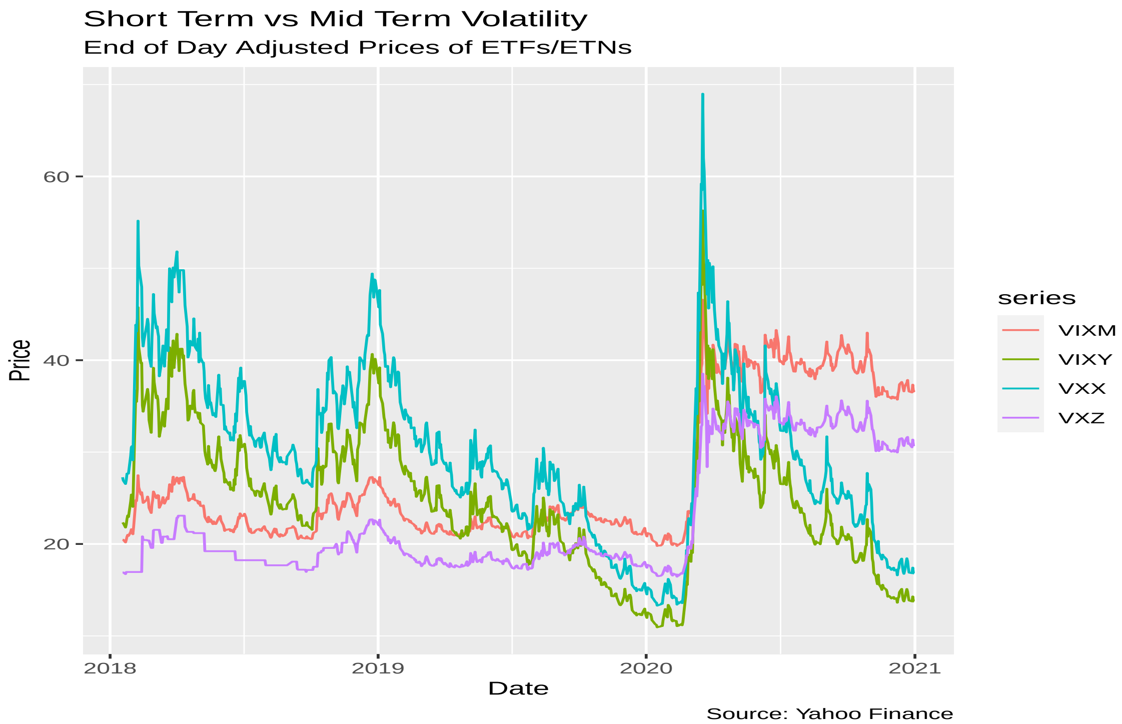
```
Call:
lm(formula = leg_VXZ ~ leg_VIXM + ec_term.lag + 0)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
leg_VIXM      0.74970    0.01408  53.227  < 2e-16 ***
ec_term.lag -0.10514    0.03784  -2.778  0.00561 **
---

Multiple R-squared:  0.8028,	Adjusted R-squared:  0.8022
F-statistic:  1417 on 2 and 696 DF,  p-value: < 2.2e-16
```

We see that the error correction term is significant. The idea of this check is to provide for correction in the equilibrium that could come in small moves $1 - \alpha \ll 1$ in the short run:

$$\Delta y_t = \beta_1^{shortrun} \Delta x_t + b_1^{longrun}(y_{t-1} - \beta_2 x_{t-2}) \tag{4.5}$$

4.3 Johansen Procedure

In this section we would like to briefly outline the results of the Johansen procedure before we present our piars trading strategy. Johansen procedure allows us to analyse whether multiple cointegrating relationship exists. Earlier we have seen the CADF test in essence tells us how much from two asset to long and short when carrying out a pair trade. The Johansen procedure checks for multiple linear combination and potentially one could build a portfolio of two or more securities traded in a mean reversion based strategy. We have included in our data the time series of iPath Series B S&P 500 VIX Short-Term Futures ETN (VXX) and ProShares VIX Short-Term Futures ETF (VIXY) which offer exposure to short-term volatility. Below the plot of the data:



As we can already guess from the plot the products are pair-wise cointegrated. Firstly, by applying the Johansen procedure only on VXZ and VIXM we can verify the results from the naive regression run earlier. The Cointegrating Equation (EC term) output is:

```
$beta$
                ect1
VXZ.l2    1.00000000
VIXM.l2  -0.83088734
constant -0.07029045
```

So, Johansen procedure suggests:

$$\hat{e}_t = r_{VXZ} - (0.07 + 0.83 r_{VIXM}) \tag{4.6}$$

This is very close to the coefficients from the naive regressions ($\hat{e}_t = r_{VXZ} - (0.09 + 0.83 r_{VIXM})$).

Now if we include all 4 assets in the analysis we obtain:

```
$beta$
               ect1
VXX.l2    1.0000000
VXZ.l2    0.2368250
VIXM.l2  -0.1904261
VIXY.l2  -1.2156304
constant -0.1754028
```

We see that Johansen test has found linear combination, which can be used to form portfolio following:

$$\hat{e}_t = r_{VXX} - (0.17 + 0.23 r_{VXZ} - 0.19 r_{VIXM} + 1.21 r_{VIXY}) \tag{4.7}$$

At first glance, it might seem surprising that VXZ and VIXM have different signs in the coefficients (since they both represent exposure to mid-term volatility). In fact it is logical as we have seen that building a stationary series of VXZ and VIXM involves shorting one and going long the other variable. Finally, we examine the eigenvalues and the critical tests to conclude the number of assets needed to build stationary portfolio:

```
######################
# Johansen-Procedure #
######################

Test type: maximal eigenvalue statistic (lambda max) , without linear trend and constant in cointegration

Eigenvalues (lambda):
[1]  3.598602e-01  2.871538e-01  8.624961e-02  2.394760e-02 -1.357884e-16

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 3 |  6.06  7.52  9.24 12.97
r <= 2 | 22.55 13.75 15.67 20.20
r <= 1 | 84.62 19.77 22.00 26.81
r = 0  |111.52 25.56 28.14 33.24
```

From the output we see that one can form stationary series by using 2 assets. We can already imagine that there is no enough evidence of short-term and mid-term cointegration for the observed period, however we see that the Johansen procedure can be useful tool in building portfolio with more complex relationships. We will continue in the section analyzing a strategy with our initial pair.

5. Trading Strategy

Having confirmed that the pair VIXM and VXZ are cointegrated, and having identified the leading/lag link we can finally construct our trading strategy. In the context of factor investing, it is easiest to think of the return of assets as a bundle of factor risks, where exposure to the different factors risks earns risk premiums. The final return can be presented as:

$$\triangle Y_t = \alpha + \sum_{i=1}^{n} \beta_i \Delta X_n + \varepsilon_t \tag{5.1}$$

If we assume that the factor risks are tradeable assets and we have determined correctly the coefficients β_i (hedging quantity) then by being long $\triangle Y_t$ and shorting the X-terms in the right quantity we can earn an excess return α:

$$\alpha = E|\Delta Y - \sum_{i=1}^{n} \beta_i \Delta X_n| \tag{5.2}$$

Admittedly, this is true when $E|\varepsilon_t| = 0$. Let us suppose that at some moment the error-term is not 0. If we have the proof that the residuals are not random and that they revert their long-term mean in fact we can exploit the drift from the mean by going short/long $\triangle Y_t$ / $\beta_i \triangle X_t$ depending on the sign of the error-term.

We have seen that the relationship in the in-sample period is:

$$V\hat{X}Z_t = 0.09 + 0.83 * VI\hat{X}M_t + \hat{e}_t \tag{5.3}$$

Assuming that this relationship hold, we can compute the error-term in the out-of-sample period (test period):

$$\hat{e}_t^{test} = \hat{Y}_{VXZ}^{test} - 0.09 - 0.83 * \hat{X}_{VIXM}^{test} \tag{5.4}$$

We have already demonstrated in the previous section when applying Engle-Grenger method that time series of the error-term is stationary. The next step is to use OU process to model it. Fitting to OU Process we have:

$$e_{t+\tau} = (1 - e^{-\theta\tau})\mu_e + e^{-\theta\tau} e_t + \epsilon_{t,\tau} \tag{5.5}$$

μ_e is the long-run equilibrium level of the OU process and θ the spead of mean-reversion.

To determine the reversion and autoregression terms from the above equation we run the regression:

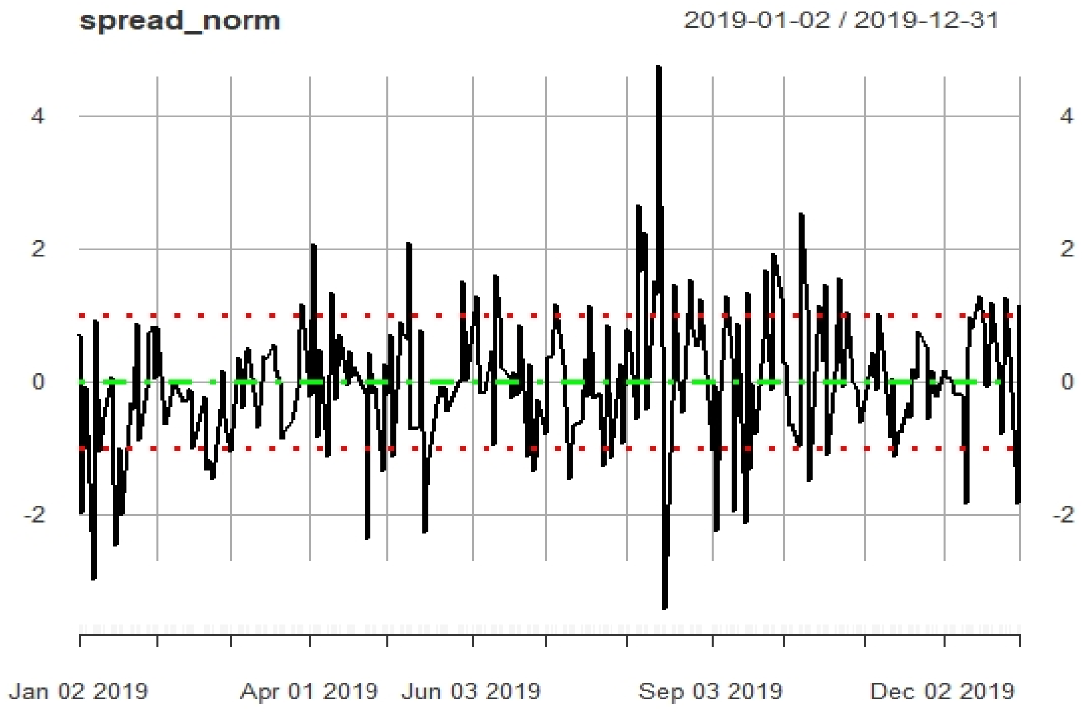
$e_{t+\tau} = C + Be_{t-1} + \epsilon_{t,\tau}$ where we substituted $e^{-\theta\tau} = B$ and $(1 - e^{-\theta\tau})\mu_e = C$. By solving for θ and μ_e we get:

$\theta = -\frac{lnB}{\tau}$ and $\mu_e = \frac{C}{1-B}$

The mean we computed is appr. 0. The standard deviation $\sigma_{eq} = \sqrt{\frac{Var[\epsilon_{t,\tau}]}{1-e^{-2\theta\tau}}}$ is 0.04. Having calculated this we can construct the bounds that will generate trading signals. For the purpose we normalize the spread as:

$$z = \frac{\hat{e}_t - \mu_e}{\sigma_{eq}} \tag{5.6}$$

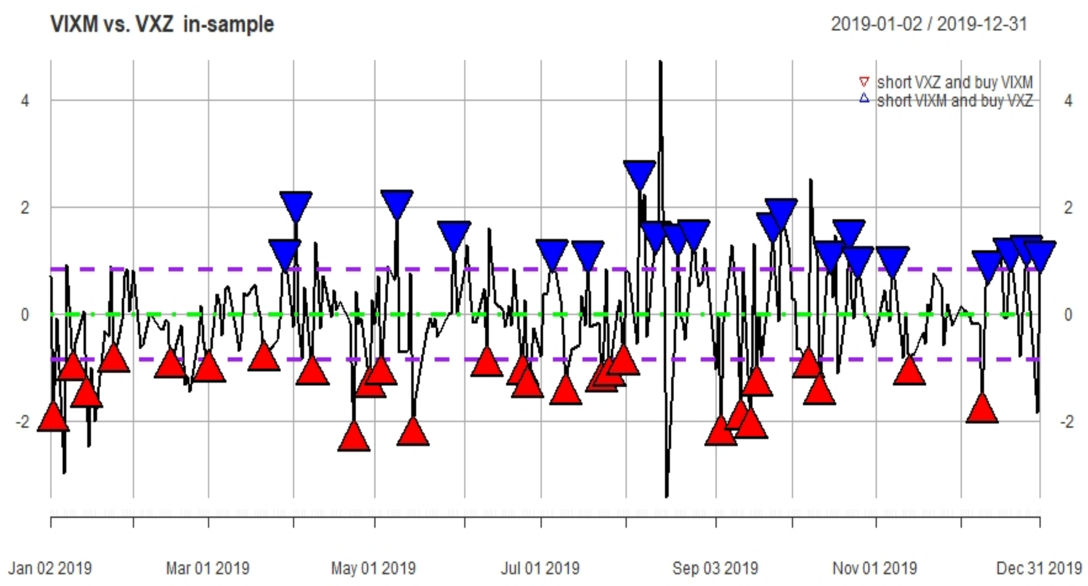
An indicator for the quality of the mean-reverting process it is worth looking at the half-life $h \propto \frac{ln2}{\theta}$. When looking for the half-life, by definition we are looking for the time t+h where the process is expected to halve its distance to the stationary mean, i.e. $E[Y_{t+h}] = \frac{1}{2}Y_t$. We computed half-life of 0.013 which given the τ of 252 days used (since our train period spans one year) should correspond to 3 days. Roughly speaking, in case the spread drifts we should not wait for more than a week for it to move back to its mean. In the next plot we can see the normalized spread.



Having constructed the bounds of expected mean-reversion we can finally define our trading rules:

- Go "Long" when the spread crosses (from above) the level $\mu_e - \sigma_{eq}$: this means buy Y and sell X
- Go "Short" when the spread crosses (from below) the level $\mu_e + \sigma_{eq}$: this means sell Y and buy X
- Exit positions when the spread reaches the level μ_e

So the trading rule we chose says we actually want to open a position as soon as the price ratio deviates with more than one standard deviation. This is in fact the more agressive strategy. Alternatively, the position could be opened not when the ratio breaks the for the first time, but instead when it crosses it to revert to the mean again. This presents more conservative approach at least from possible drawdown expected. However, this would expose us to buying at higher prices and cutting the profit margin. Important to mention is that a new trade is opened only when there is no already existing one (i.e. no increases of position size allowed). Below we plot the trading signals captured by the strategy during the train period:



Counting the number of exit signals (i.e. take profit) we have the following summary for the in-sample period.

```
do nothing short X and buy Y short Y and buy X        take profit
       159                    27                   20                  46
```

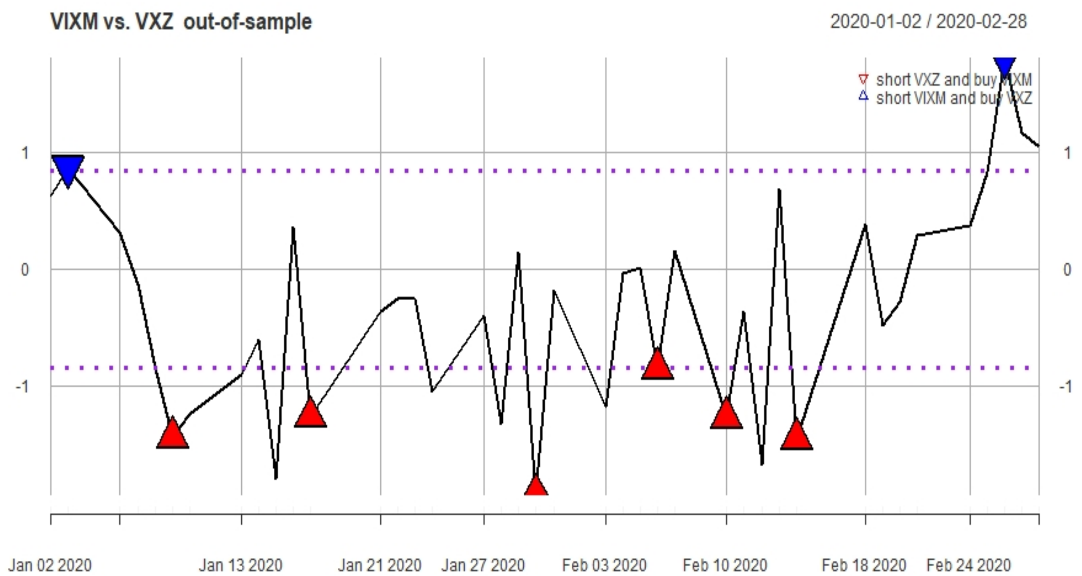
We see that the strategy is not excessively trading with an average of 1 trade per week. Another observation we can make is that some signals not immediately after exceeding the Z-spread bounds and the reason is that there can be jumps with higher magnitude than the spread itself. One of the key features of the strategy is that it is non-directional or market-neutral, i.e. it can be applied in bear or bull market regardless of the direction in which assets move. As long as there is any divergence this can be exploited no matter whether underlying prices rise or fall.

5.1 Optimization

Backtesting a strategy would not be complete if we do not analyze the strategy when varying the parameters of the strategy. For this reason we have tested the strategy for different thresholds that determine the trade signals. Some sort of optimization we have performed was looping through a sequence between 0 and 3 with steps of 0.05 where each number of the sequence was assigned to the threshold parameter setting the boundaries for position entries. At each step the profit and loss has been calculated. The output below shows the threshold for entering positions that maximizes the profit.

```
[1] "When threshold = 0.75"
[1] "The Return of the strategy for the period 2019-01-02 - 2019-12-31 is 16.816%"
[1] "When threshold = 0.80"
[1] "The Return of the strategy for the period 2019-01-02 - 2019-12-31 is 31.263%"
[1] "When threshold = 0.85"
[1] "The Return of the strategy for the period 2019-01-02 - 2019-12-31 is 39.887%"
[1] "When threshold = 0.9"
[1] "The Return of the strategy for the period 2019-01-02 - 2019-12-31 is 31.893%"
```

We have implemented here a very simplified optimization. However, one should not be confined only to finding most profitable strategy. Other parameters could be also included in the process, for instance the drawdowns which we discuss later. Intuitively, higher boundaries should imply lower drawdowns and one might think this is more conservative strategy. However, the more the spread diverges the higher is the risk of regime changes and cointegration ceases to exist. In fact, rather than increasing the threshold to make the strategy more conservative one should include stop loss parameter. Another solution would be to enter the position when the spread starts mean reverting back, i.e. instead of opening position when the spread exits the range determined by the threshold, one waits until the spread enters back to range. Using the optimized threshold parameter of 0.85 we plot the signals captured by the strategy for the out of sample period.



Using the optimized parameter we were able to capture more signals which led to better performance.

5.2 Performance Evaluation

To design a trade it is necessary to backtest its P&L and clarify magnitude the drawdowns. A solid strategy should be not be measured on profitability per se but also on some indicators of riskiness to determine its risk-reward ratio. Apart from the classic Sharpe Ratio we included also some downside risk measures. In this report, we made use of the R package 'Performance Analytics' to evaluate the performance. The downside risk measures[1] included in the package are recognized as a major improvement over traditional portfolio theory and a better measure of investment risk than standard deviation (which measures also upside risk). The idea is that we don't want to penalize a strategy for positive volatility, which is a benefit. As the industry standard for risk management, we have used the following metrics here to evaluate the performance of the strategies:

- Sharpe Ratio
- Sortino Ratio
- Return Over Maximum Drawdown (RoMaD)

The Sharpe and the Sortino Ratios can be expressed simply as $\frac{R_p - r_f}{\sigma_p}$ with the only difference that the latter uses the Downside Deviation in the denominator. The Downside Deviation eliminates positive returns when calculating risk. It uses a specified minimum acceptable return (idea proposed by Sharpe) instead of using the mean return of zero.

$$DownsideDeviation(R, MAR) = \delta_{MAR} = \sqrt{\frac{\sum_{t=1}^{n}(R_t - MAR)^2}{n}} \tag{5.7}$$

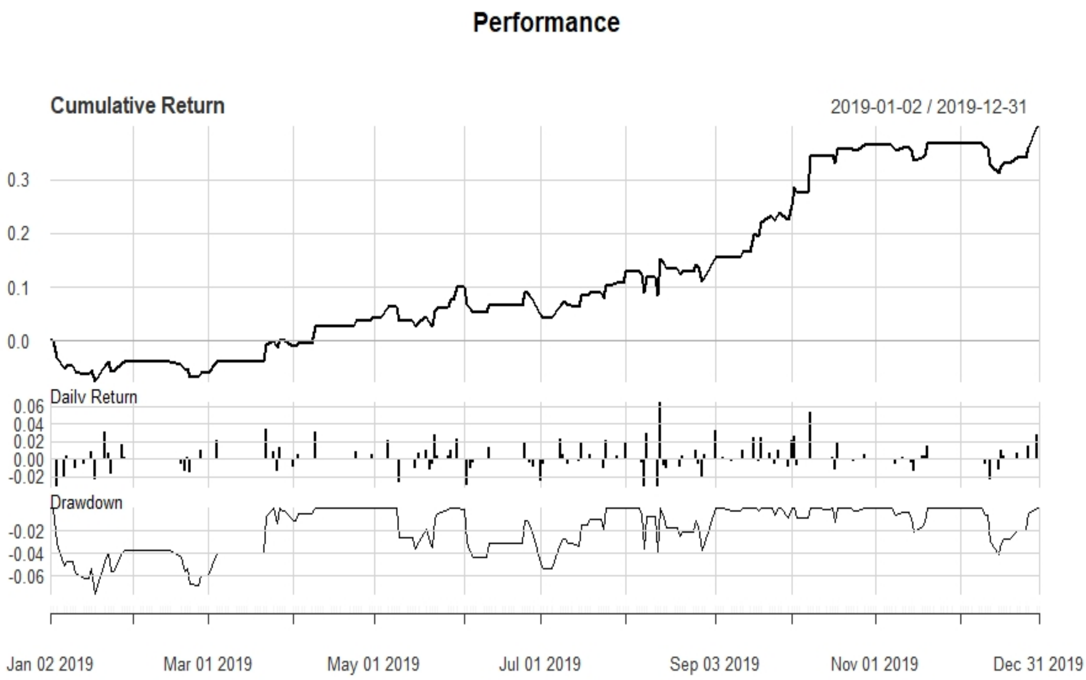
Return over maximum drawdown (RoMaD) is a risk-adjusted return metric used as an alternative to the Sharpe Ratio or Sortino Ratio. Return over maximum drawdown is used mainly when analyzing hedge funds. It can be expressed as:

RoMaD = portfolio return / maximum drawdown

Drawdown is the difference between a portfolio's point of maximum return (the "high-water" mark) and any subsequent low point of performance. Maximum drawdown, also called Max DD or MDD, is the largest difference between a high point and a low point.

5.2.1 In-Sample Performance

First, we present the performance for the train period (Jan.2019-Jan.2020). The plot below shows the evolution of the performance during that period:



[1] developed by Frank Sortino in the early eighties

We see that the strategy has started with a loss and after some sideways performance, it picks good signals around September 2019 to finish with performance of 39%. In fact, in September 2019 volatility rises with unexpected interest rate cuts by central banks. This could be actually the reason for triggering the spread to leave its bounds and create trading opportunities. After the market situation stabilizes, the strategy performance stay flat for the end of the training period. Next, we present the summary of the risk measures:

```
Semi Deviation                  0.0068
Gain Deviation                  0.0125
Loss Deviation                  0.0090
Downside Deviation (MAR=210%)   0.0114
Downside Deviation (Rf=63%)     0.0074
Downside Deviation (0%)         0.0062
Maximum Drawdown                0.0752
Historical VaR (95%)           -0.0133
Historical ES (95%)            -0.0238
```

With a return for the period of 39% and Maximum Drawdown of 7.5% the strategy shows RoMaD of around 5. This can be considered a very good result. As with any metric of evaluation, the performance expectations should be relatively to other products and stragtegies. In practice, investors want to see maximum drawdowns that are half the annual portfolio return or less. That means if the maximum drawdown is 10% over a given period, investors want a return of 20% (RoMaD = 2). So the larger a fund's drawdowns, the higher the expectation for returns. In the next table we have calculated also Sharpe and Sortino ratios

Table 5.1: In-sample Performance Results

Metric	Score
Return	39%
Sharpe Ratio	2.24
Sortino Ratio	4.04
RoMaD	5.3

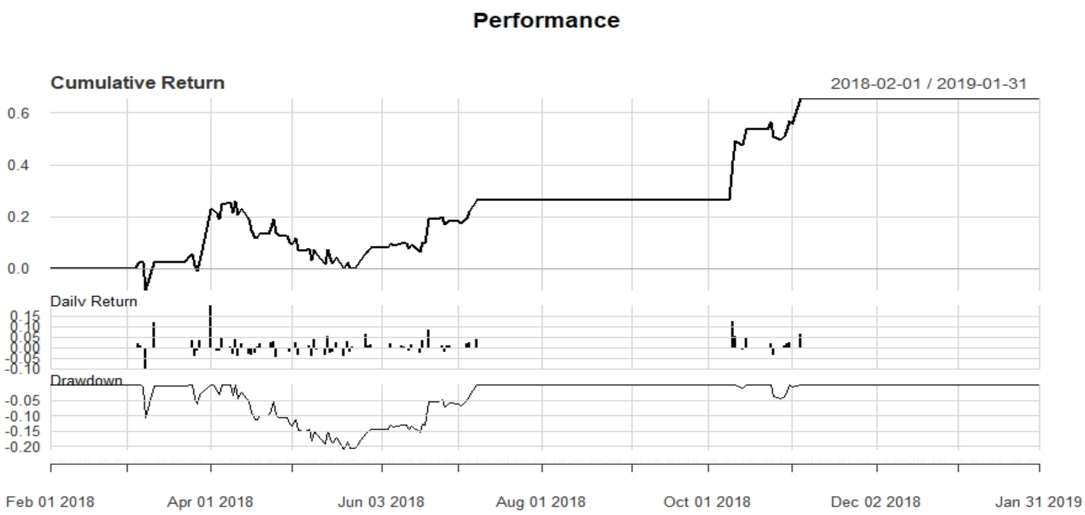
We see that Sortino score is higher than the Sharpe as we have explained that Sharpe includes also upside volatility. Interesting to mention is that the VaR is negative - this would imply the portfolio has a high probability of making a profit. Given it is daily periodicity, 5% VaR of negative 0.0357 implies the portfolio has a 95% chance of making more than 3.57% over the next day. In the section we can see how these statistics change.

5.2.2 Out-of-Sample Performance

During the test period, due to the shortened period, admittedly the strategy traded less with only 7 closed position.

```
do nothing short X and buy Y short Y and buy X       take profit
        25                 6                  2                7
```

The plot below shows the evolution during this period and the metrics are presented in table:



```
Semi Deviation                 0.0085
Gain Deviation                 0.0183
Loss Deviation                 0.0101
Downside Deviation (MAR=210%)  0.0121
Downside Deviation (Rf=63%)    0.0083
Downside Deviation (0%)        0.0071
Maximum Drawdown               0.0421
Historical VaR (95%)          -0.0184
Historical ES (95%)           -0.0270
```

The return for this two-months period of trading is 10.645%. We can see that the maximum drawdown of the strategy is 4.2% which might seem quite high, however relative to the cumulative return the strategy it can be still regarded as acceptable. The next table presents the performance metrics

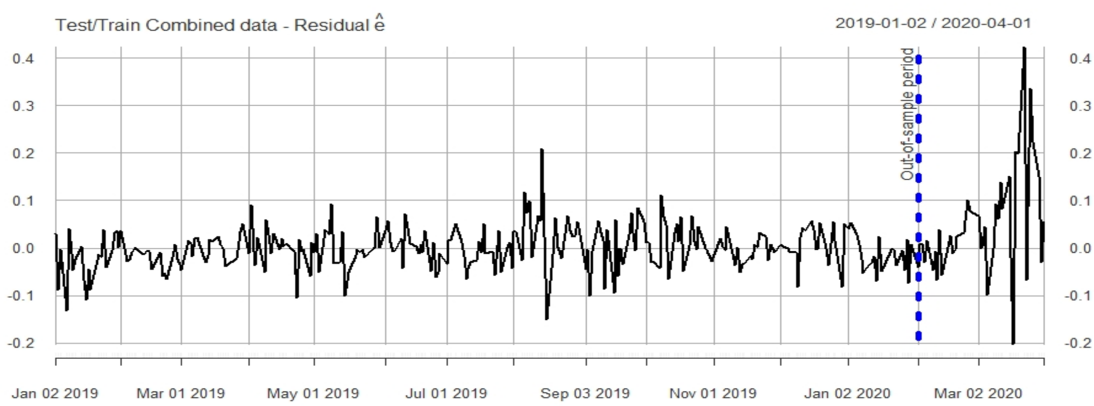
Table 5.2: Out-of-sample Performance Results

Metric	Score
Return	10.6%
Sharpe Ratio	1.20
Sortino Ratio	2.64
RoMaD	2.8

We have to admit that the out-of-sample performance is not so stellar relative to the in-sample. This can be due to the short period and lack of trading opportunities. Another reason for underperformance in the test relative to the train period can be changing market conditions causing in turn regime changes. Finally, if we consider a plain buy-and-hold strategy as a benchmark we can notice that in the in-sample period the cointegration would have overperfomed with return of 39% vs -19%, whereas in the out-of-sample it would have underperformed achieving 10.6% vs 13% of the plain buy-and-hold strategy. However, we should ask if buy-and-hold strategy is the right benchmark since it is simply proxy of volatility, when at the same time the cointegration strategy is market-neutral capturing signals in periods of both rising and declining volatility.

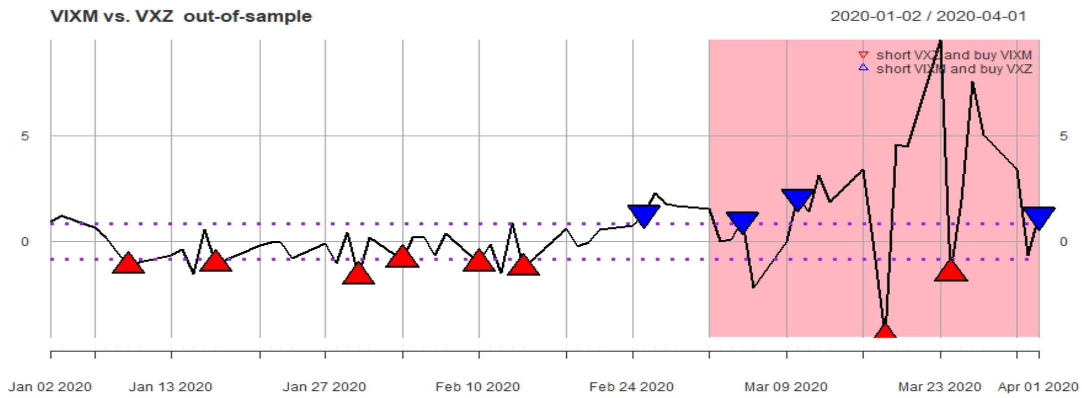
5.3 Regime Changes

Cointegration are not pure arbitrage strategies since the notion of uncertainty is still present - the spread produced by cointegrated time series is stationary and mean-reverting around some linear equilibrium level. The recent Covid pandemic is a clear example how previously firm relationships can be suddenly broken and divergence can last over prolonged period. Even if time series continue to exhibit mean-reversion, characteristics like half-life or bounds of expected divergence might have changed significantly. In March 2020 there was a sudden spike in volatility caused by the outbreak of Covid-19. To test the effect of



The residual in March jumped wildly in both extremes. A sudden increase in volatility caused by market crash and surprisingly rapid and strong comeback of equity prices.

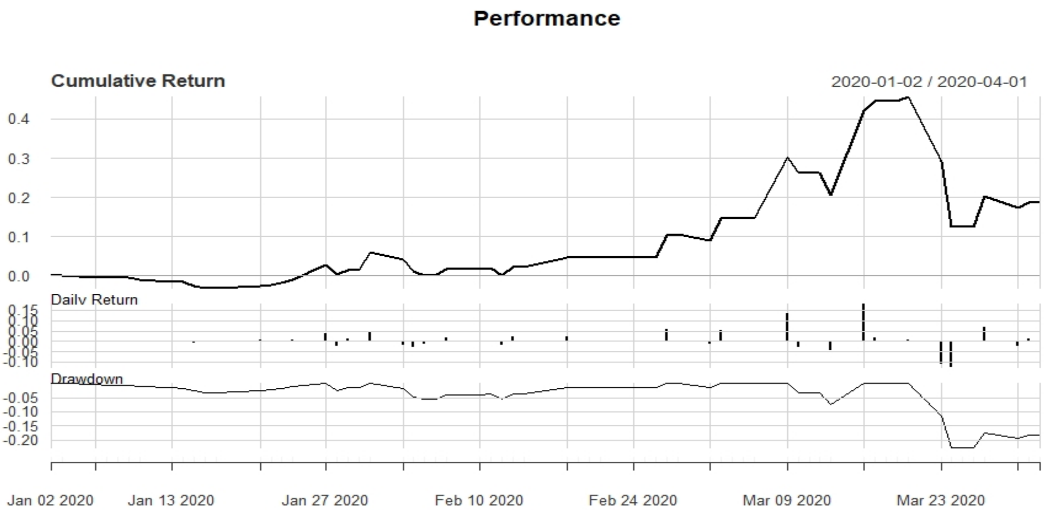
In the next plot we have highlighted the additional month data:



We see that the spread widens extremely in March 2020. At first it seems the strategy picks up the signal, however, when we look the the moves are so extreme afterwards that strategy can not enter a position since the rule is to trade only when the spread leaves the bounds from within the bounds (and not from beyond one bound into beyond another). In short, only by luck (or due to a fault in the strategy) the portfolio is not completely ruined in this case.

```
index       norm_spread     X      Y      diff         cash shares_Y shares_X mm_position           action
2020-03-18 -4.58731641 46.58 38.540 8.040001 1.239412e+01     8572    -3546   165204.60 short X and buy Y
2020-03-19  4.59577744 44.15 36.930 7.220002 1.600204e+05        0        0   160020.45       take profit
2020-03-20  4.47125594 44.47 37.190 7.280002 1.600204e+05        0        0   160020.45        do nothing
2020-03-23  9.54751137 39.74 33.490 6.250000 1.600204e+05        0        0   160020.45        do nothing
```

Performance is though not unaffected:



The maximum drawdown is now much higher even than that for the whole in-sample period. Clearly, the strategy needs to be recalibrated and thresholds optimzed again to align with very volatile period. One extreme move in March wipes out almost all the profit generated. Again we see that it is good for the strategy when the spread moves up and down - namely, this creates opportunities. However, when there are extreme moves, i.e. when instead of mean-reverting the spread even widens further, the strategy generates losses.

6. Conclusion

In this project, we have applied cointegration to time series data to see if a strategy based on the findings can generate value. Rather than studying correlation in returns, we have concentrated on co-movements in the prices of our pair of assets. By assessing mean-reversion and stationarity, cointegration reduces the risk of spurious regressions, which can be often the result when simple correlation analysis applied. The strategy based on cointegration is non-directional and has relatively low maximum drawdown when compared to the overall cumulative return.

The strategy has obtained return of 39% and 10.6% in resp. the in-sample and out-of-sample periods. For comparison a plain buy-and-hold strategy with VIXM would have returned loss of -19% during the in-sample and resp. profit of 13% in out-of-sample period. The overperformance of the buy-and-hold approach in the out-of-sample is due to the spiking volatility in mid February and on a risk-reward basis the cointegration strategy is still better. Nevertheless, this is to remind us that regardless of the strong results of in-sample performance, one should keep watch of pitfall in implementation. First of all, the old adage "correlation does not imply causation" is equally true also for cointegration. Regime changes might have prolonged impact on the spread of asset prices and negate previously valid conclusions with respect to the future. As the parameters are optimized with respect to in-sample data, it is logical that one needs to have at least a bit similar picture (as that of the test period) in the trading period for good results to be obtained, as any algorithm will trade to the detriment of out-of-sample optimization. The empirical results showed that, generally, with larger population sizes or a bigger time span (daily frequency) there is a bigger disparity between in-sample and out-of-sample performance. Thus, optimization and backtesting needs to be re-run on regular basis to check the consistency of the model given that the parameters might change over time or the relationship cease to exist.

Further factors that need to be taken into account are:

- Commissions and trading fees - our strategy has not produced high number of position, however, strategy that creates plenty of entry and exit signal can see returns easily evaporate when transactions costs included.
- Different tax treatments - practical issue that can have a huge impact when applied to different assets. Volatility ETFs and ETNs especially can be difficult to analyze. For instance, Volatility ETFs like VIXM are taxed as securities and unlike ETNs (such as VXZ), they make annual distributions of income and capital gains to shareholders. On the other hand, ETNs offer tax deferral options in contrast to ETFs.
- Choosing the right time-span. Choosing smaller periods can help confirming stationarity but it is possible that there are not enough trade signals
- Optimization - a more sophisticated optimization (for instance better than brute force) approach can be considered, including also more parameters (eg. stop-loss and testing with signals with different crossing techniques eg. spread crossing from above or below the threshold line). Instead of simply picking the bounds for which we have the maximum profit it could be attempted to maximize Sharpe ratio or minimize an objective function such as strategy drawdown/downside deviation.

Finding a methodology that never encounters loss is equivalent to finding the 'holy grail' in finance. We have shown that the cointegration may be unreliable metric in turbulent times. Far from believing the strategy is perfect, I have been nevertheless encouraged by the results and at the same time, made aware of the pitfalls that might appear and the investigations that need to be carried out. I believe that the idea examined herein, based on the empirical results obtained, is simple enough and has a significant probability of achieving similar, moderately successful results also in real-life trading. As far as I am concerned, the main contribution of this project is to draw the path how to apply the Engle and Granger procedure so that one can be reasonably sure that the model is correctly specified. I feel encouraged to look further for ways to screen rapidly candidates for cointegration and improve the process of parametrizing the optimal strategy.

References

[1] CQF, Module 6, Lecture 8: Co-Integration for Trading, **Diamond, R.** (2020)

[2] CQF, Final Project Workshop - Part I

[3] Learning and Trusting Cointegration in Statistical Arbitrage, **Diamond, R.** (2013)

[4] A drunk and her dog: An illustration of cointegration and error correction. **Murray, M.** (1994)

[5] Cointegration and asset allocation: A new active hedge fund strategy, **Alexander, C.**

[6] Quantitative Investment Analysis, **DeFusco, R.**, **McLeavey, D.**, **Pinto J, Runkle D. (2007)**

[7] Quantitative Trading with R, **Georgakopoulos, H.** (2015)

[8] Modelling Financial Time Series with S-PLUS, **Zhivot E.** (2002)

[9] Efficient Pair Selection for Pair-Trading Strategies, **McSharry, P.**(2015)

[10] Pairs Trading, Convergence Trading, Cointegration, **Herlemont. D**

[11] Ornstein-Uhlenbeck Process, Wikipedia

[12] Количественные методы в финансах, **Уотшем Т.Дж.** (1999)

[13] `https://etfdb.com/etfdb-category/volatility/`

[14] `https://www.quantstart.com/articles/Johansen-Test-for-Cointegrating-Time-Series-Analysis-in-R/`

A Appendix

A.1 Regression

Multiple linear regression model has the general form:

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + ... + b_k X_{ki} + \varepsilon_i, i = 1, 2, ...n \tag{6.1}$$

- Y_i is the dependent variable whereas X_{ji} are the independent variables (i-th observations)
- ε_i denotes the error term
- $b_0$is the intercept of the equation and $b_1 ... b_k$are the slope coefficients for each of the independent variables

A slope coefficient, b_j, measures how much the dependent variable, Y, changes when the independent variable, X_j changes by one unit, holding all other independent variables constant.

By minimizing the sum of squared residuals via the Maximum Likelihood Method (MLE), one can estimate the slope coefficients (providing that the matrix independent variables X has full rank). Presented in matrix notation this will be:

$$\hat{b} = (X^T X)^{-1} X^T Y \tag{6.2}$$

and the error term estimate will be

$$\hat{\varepsilon} = Y - X\hat{b} \tag{6.3}$$

This is the so-calles Ordinary Least Squares (OLS) method. Its Log Likelihood Function is:

$$log(L) = -\frac{n}{2} log(2\pi) - -\frac{n}{2} log(|\hat{\sum}|) - \frac{n}{2} \tag{6.4}$$

where $\hat{\sum} = \frac{\sum}{n} \varepsilon\varepsilon^T$or $\hat{\sum} = \frac{\sum}{n-kp} \varepsilon\varepsilon^T$using the MLE estimator or resp. the OLS estimator with k variables and p lags.

The multiple regression has been implemented in the r script for this project (function my_LM)

A.2 Autoregression models

An autoregressive model (AR), is simply time series regressed on its own past values. AR(p) autoregression for the variable x_t

$$x_t = b_0 + b_1 x_{t-1} + b_2 x_{t-2} + ... + b_p x_{t-p} + \varepsilon_t \tag{6.5}$$

In this equation, p past values of x_t are used to predict the current value of x_t. $b_0$is the drift term (a constant) and ε_t is the error term. This can be easily coded using the linear regression described previously (see function my_AR from the r script).

A.2.1Dickey-Fuller's test

Dickey-Fuller's test is a regression-based unit root test based on a transformed version of the AR(1) model $x_t = b_0 + b_1 x_{t-1} + \varepsilon_t$. Subtracting x_{t-1} from both sides gives:

$$x_t - x_{t-1} = b_0 + (b_1 - 1)x_{t-1} + \varepsilon_t \tag{6.6}$$

The null hypothesis of the Dickey-Fuller test is H_0: $b_1 - 1 = 0$, that is, that the time series has a unit root and is nonstationary—and the alternative hypothesis is H_a: $b_1 - 1 < 0$, that the time series does not have a unit root and is stationary.

There are three different version when testing for stationarity:

- Equation with no drift, nor trend: $\Delta Y_t = (b_1 - 1)Y_t + \varepsilon_t$
- Equation with drift: $\Delta Y_t = b_0 + (b_1 - 1)Y_t + \varepsilon_t$
- Equation with drift and determinstic trend: $\Delta Y_t = b_0 + b_1 t + (b_1 - 1)Y_t + \varepsilon_t$

Admittedly, the critical values are different for each version and in general testing with drift or drift and trend tends to reject the non-stationarity hypothesis more often (due to creating temporary dependence in the time series)

The Augmented Dickey-Fuller test (ADF) is an extension to the classical Dickey-Fuller test by adding the term $(b_p - 1)\Delta Y_{t-p}$. The idea of the ADF test is to remove possible autocorrelation effects.Using Information criteria, the optimal lag to be applied can be chosen.

A.2.2 Information Criteria

To choose between models one uses fit criteria like Bayesian Information Criteria(BIC) and Akaike Information Criteria(AIC). Iterating over different lag orders, the best predictor is the model with the lowest Information Criteria value. Both BIC and AIC are penalized-likelihood criteria and both approximately correct with respect to different goal and given different set of asymptotic assumptions. The definition used here for AIC in our AR(p) implementation is

$$AIC = log|\hat{\sum}| + 2\frac{1+k}{n} \tag{6.7}$$

where is k is the number of estimated parameters. Alternatively one can use the BIC which is defined as:

$$BIC = k * ln(n) - 2ln(\hat{L}) \tag{6.8}$$

where n is the number of observations, k - number of parameters and is the maximised likelihood of the model M, i.e. $\hat{L} = p(x|\hat{\theta}, M)$ ($\hat{\theta}$are the parameter values that maximize the likelihood function)

A.2.3 Johansen Test

To understand the Johansen Procedure, we need firstly to introduce the Vector Autoregressive Models (VAR). This is basically a multidimensional extension of the AR model described earlier. VAR(p) without drift can be expressed as:

$$x_t = \mu + A_1 x_{t-1} + ... + A_p x_{t-p} + w_t \tag{6.9}$$

where μ is the vector-valued mean of the series, A_i are the coefficient matrices for each lag and w_t is the multivariate Gaussian noise term with mean zero.

By differencing the series we can form the Vector Error Correction Model (VECM):

$$x_t - x_{t-1} = \mu + A_1 x_{t-1} + \Gamma_{t-1}(x_{t-1} - x_{t-2}) + ... + \Gamma_{t-1}(x_{t-p} - x_{t-p-1}) + w_t \tag{6.10}$$

where by Γ_i we noted the matrices for each differenced lag.

In case that $A_1 = 0$ we can conclude that there is no cointegration. Johansen test carries out eigenvalue decomposition of A. If the rank of A is given by r, then the null hypothesis of Johansen test is $r = 0$.

Johansen test sequentially tests how many assets are cointegrated by checking whether the rank of A, r $= 0, \leq 1, \leq 2$ and so on.

A.3 Performance Analytics

We start with Semi-deviation. It is defined by:

$$SD(X) = \left(E\left[(X - E(X)\right]^2 1_{\{X \leq E[X]\}}\right)^{\frac{1}{2}} \tag{6.11}$$

where $1_{\{X \leq E[X]\}}$ is the indicator function (1 if $X \leq E[X]$ else 0)

The Gain Standard Deviation is calculated in the same way as the simple standard deviation, except this statistic calculates an average (mean) return for only the periods with a gain and then measures the variation of only the gain periods around this gain mean. This statistic measures the volatility of upside performance. The Loss Deviations is defined similarly with the difference that the deviation is measured of only the loss period around a loss mean. (respectively it measures the volatility of downside performance).

Another measure that is considered is the Downside Deviation eliminates positive returns when calculating risk. It uses a specified minimum acceptable return (idea proposed by Sharpe) instead of using the mean return of zero.

$$DownsideDeviation(R, MAR) = \delta_{MAR} = \sqrt{\frac{\sum_{t=1}^{n}(R_t - MAR)^2}{n}} \tag{6.12}$$

where n is the number of observations in the series.

The maximum drawdown (MDD) as its name suggests, measures how sustained one's losses can be and is basically the maximum cumulative a strategy incurs during trading period. Large drawdown is of course a negative statistics of a model and low MDD is what traders want to see (other terms equal). Formally MDD up to time T is:

$$MDD(T) = Max_{\tau\epsilon(0,T)}[Max_{t\epsilon(0,\tau)}X(t) - X(\tau)] \tag{6.13}$$

where $X(T)$ is a random process $[X(0) = 0, t \geq 0]$. The Maximum Drawdown is namely the drawdown over the history of this variable.

Another important downside risk statistic that is calculated is of course Value at Risk (VaR). VaR provides an answer (within some probability bound) to the fundamental question what is the most one can lose on an investment. In its most general form VaR captures the potential loss of a risky asset in a traded portfolio from adverse market movements over a defined period for a given confidence interval.

$$MDD(T) = Max_{\tau\epsilon(0,T)}[Max_{t\epsilon(0,\tau)}X(t) - X(\tau)] \tag{6.14}$$

where $X(T)$ is a random process $[X(0) = 0, t \geq 0]$

Given a confindence value $\alpha\epsilon(0,1)$, the VaR at this confidence level is simply the smallest number l such that the probability that the loss exceed L is not bigger than $(1-\alpha)$. Mathematically , if L is the loss of a portfolio, then $VaR_\alpha(L)$ is the level $\alpha - quantile$, i.e.

$$VaR_\alpha(L) = inf\,\{l\epsilon R : P(L > l) \leq 1 - \alpha\} = inf\,\{l\epsilon R : F_L(l) \geq \alpha\} \tag{6.15}$$

The right equality is true if an underlying probability distribution exists. Although it virtually always represents a loss, VaR is conventionally reported as a positive number. Historical VaR is useful for calculating potential loss only when one knows that the returns of the risky asset have a distribution at least close to normal. However, this is not usually the case, as history shows that extreme events occur with much higher probability than that assigned by normal distributions. That's why an alternative method is developed, namely the modified value at risk (MVar). It accounts also for the skewness and kurtosis of PDFs. The formula is:

$$MVaR_p = \mu + \left[z_p + \frac{1}{6}(z_p^2 - 1)S + \frac{1}{24}(z_p^3 - 3z_p)K - \frac{1}{36}(2z_p^3 - 5z_p)S^2\right]\sigma \tag{6.16}$$

where S and K are skewness and kurtosis, μ is expected asset return and z is the distance the portfolio returns and its mean in terms of standard deviations (σ - st.deviation)

Historical Expected Shortfall (Hist. ES), known also as conditional VaR (or CVaR) tells us more about the magnitude of potential losses. We have:

$$ES_\alpha = \frac{1}{\alpha}\int_0^\alpha VaR_\gamma(X)d\gamma \tag{6.17}$$

where VaR_γ is the Value at Risk.

B Appendix

Results from March 2018 to June 2019 are interesting because we can contrast very volatile 2018 (especially the beginning and ending quarters of the year) with relatively quite first half of 2019. Taking as training period the period between March 2018/March 2019 we follow the described in the report approach. Both time series are non-stationary, but this time even after first-differencing there are left some memory effects:

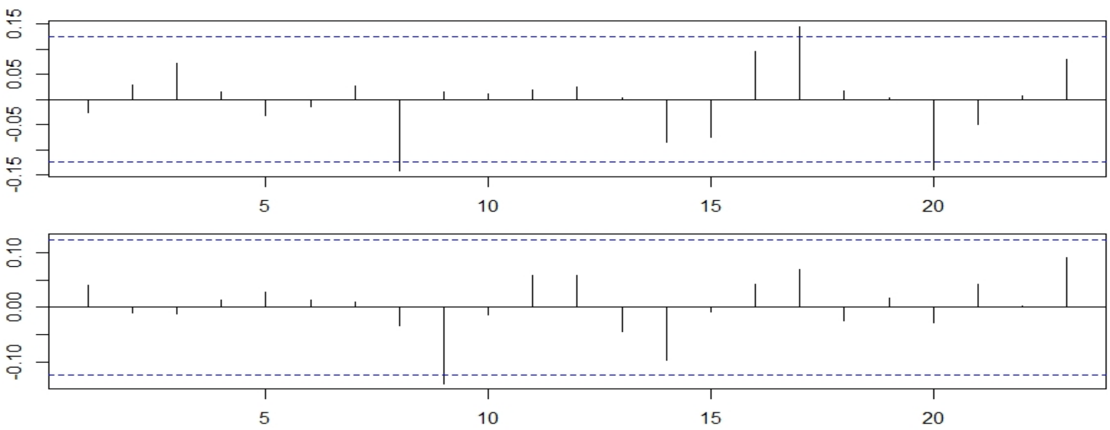


Figure 6.1: VIXM & VXZ first differences 03/2018-03/2019

It seems that there are still some significant lags, which could be caused by spiking volatility. 2018 started with a market correction and the worst point decline. The PACF shows that the duration of shocks is relatively persistent and influence the data several observations ahead. This should be a signal to be careful with such data due to big outliers. Splitting the data into smaller periods might be a good idea here. Nevertheless, we calculate the residuals and check for autocorellation:

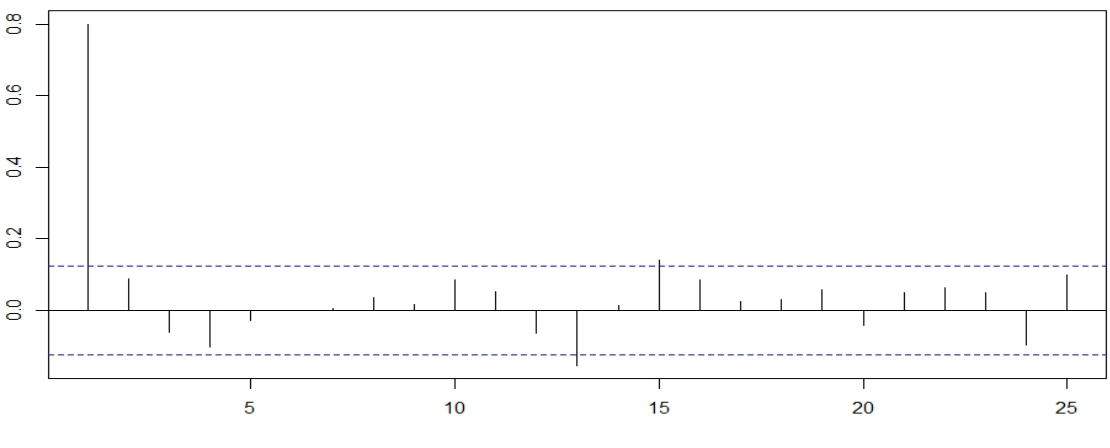
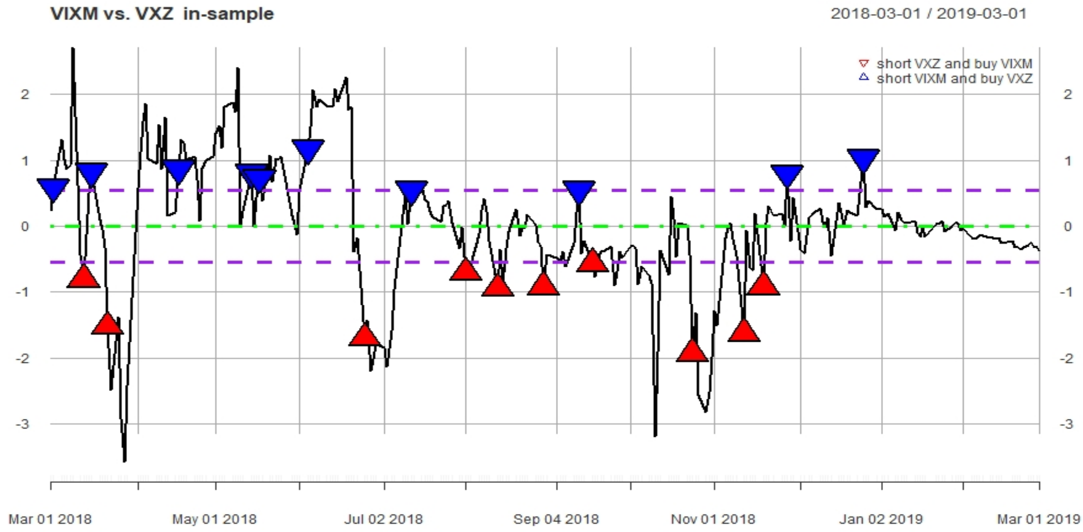
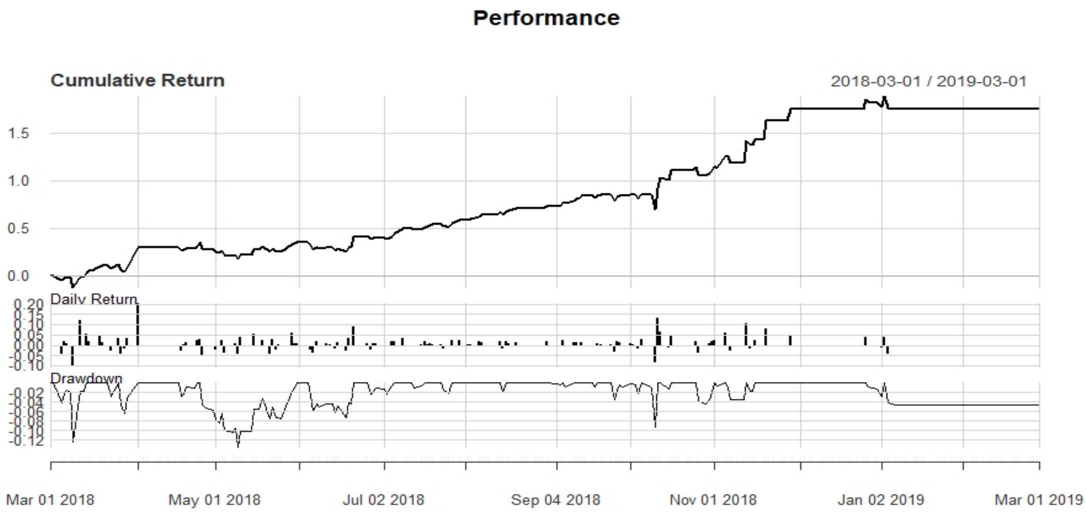


Figure 6.2: Residuals VIXM & VXZ first differences 03/2018-03/2019

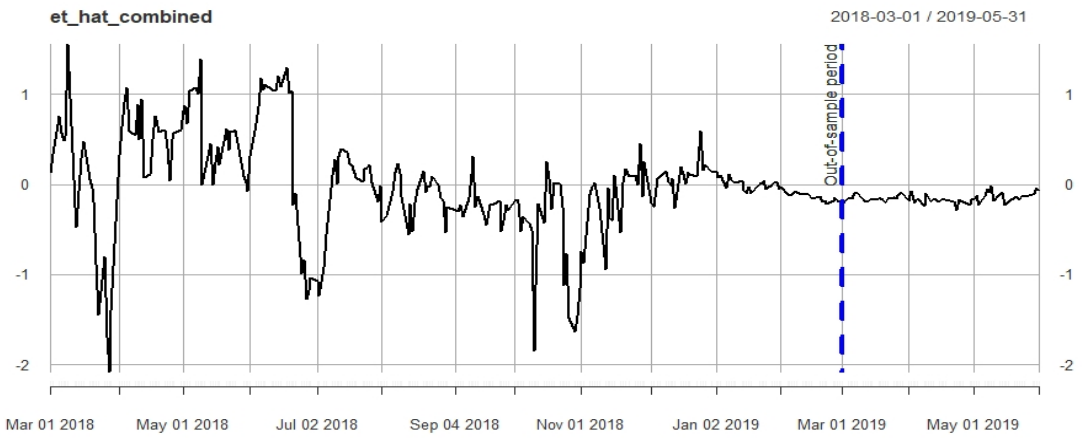
This looks still like AR(1) process but including lags of larger order should be analyzed as well. Fitting the process, we compute the spread and test our cointegration based strategy.



We can see that spread gyrates rapidly in both extremes in the beginning of 2018. However, in-sample performance after opitmization is exceptional:



The satisfaction of this is, however, short-lived. The spread in the in sample period oscilates around the 0 with no trading signals at all.



The lesson of this example is that calibrating the strategy on in-sample period that is profoundly different than the out-of-sample data is doomed to fail. In fact we can be happy that there are no losses.