Is there any relationship between bonds yields spreads and S&P 500?

February 24, 2021

Glukhov Nikita

gevnik08@gmail.com

2020-2021

Contents

1 Introduction

In the following work we aim to understand the possible relationship between bond yields spread and the S&P 500 index. Are they cointegrated? This is our primary research interest. While it seems like a relatively easy questions, we will see that due to the nature of data, obtaining a reliable result is challenging, so following work will require a comprehensive data analysis. Overall, with given data and our methodology, we obtain (SPOILER AHEAD) the result indicating the presence of cointegrating relationship. So, our secondary research interest will be in modelling the joint relationship using VECM and compare the predictions on series with the null models - random walks.

2 Data description and analysis

2.1 Data description

We consider three time series: SP 500, 3 month and 10 years bonds yields spread. The first was downloaded from Yahoo.Finance, the spread was based upon bond yields from FRED. Series at daily frequency. There was data preprocessing before the initial analysis. Some unimportant part of this preprocessing was done in Python (main work is still done in R) and included aligning data with each other based on dates and dropping observations like '.' contained in FRED data. After preprocessing, a dataset with initial series and series such as yieilds spread, yields spread first difference and S&P 500 first difference. This dataset will be used in future analysis which is done in R.

2.2 Initial data analysis

Figure 1 shows the historical dynamics of the two main series - S&P 500 and bond yield spreads, starting from 1993-01-29 and until 2020-12-10. We break the dataset into two parts: the train and test parts. The test part contains the last 562 observations out of 6963. Test parts are indicated by yellow and black colors on Figure 1.

2.2.1 Spread series

A first glance at data reveals that the spread contains high level of heteroscedasticity or, possibly, different variance structures across all time period. Generally, we could somewhat fight heteroscedasciticy by using log transformations on yields and take the differences, however, there is some numerical instability issue happens, because there are periods of time when interest rates were very low and even 0 (Figures 2, 3).

Figure 1: Bond yields spread (blue and black are train/test samples), SP 500 index (red and yellow are train/test samples)

Our first major analysis should concern the spread series. Is the series stationary? To answer that questions, we should conduct several tests to identify the presence of co-integrating relationship. This requires to carefully identify and interpret appropriate tests depending on the null hypothesis. I suggest the following tests

- The ADF test H_0: unit root is present against H_a: no unit root present
- The KPSS test H_0: there is (trend) stationarity against H_a: unit root

For the ADF test, we use an implementation from *aTSA* R package. It considers three regression settings : 1. linear regression with trend and no drift, 2. LR with drift but no trend, 3. drift and trend. The same specifications can be considered in an alternative ADF test from *urca* package. Both tests do not reject the null hypothesis of the presence of unit root.

For the KPSS test, we use implementation *kpss.test* from the *aTSA* package and *ur.kpss* from *urca*package. The test considers three settings. In each, it decomposes a series to a sum of a trend, random walk and stationary part. Each setting differs by including the deterministic trend component with/without a drift or not. While the *aTSA* test does not reject the null

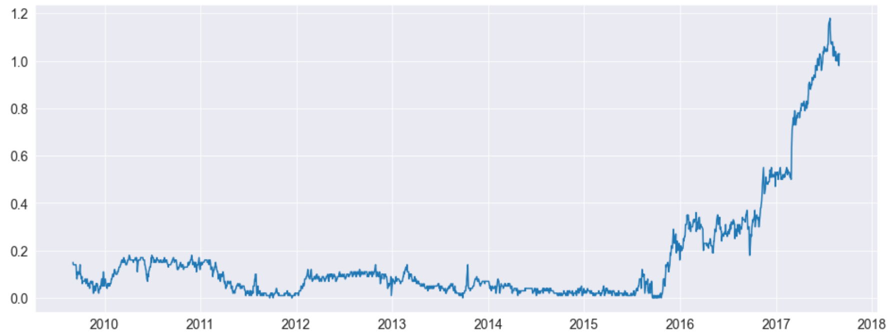

Figure 2: Low 3 month bond yields

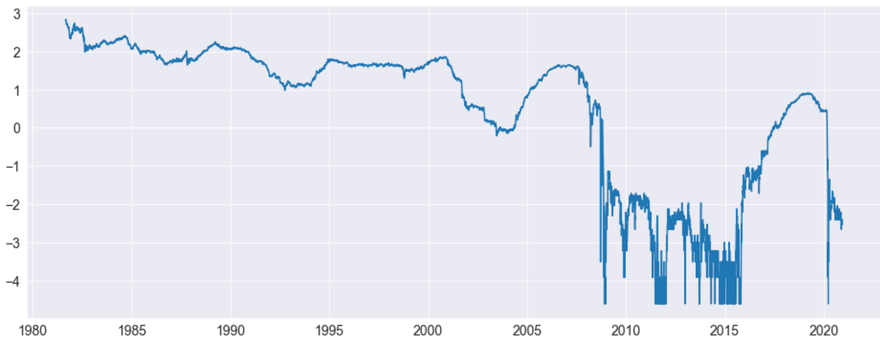

Figure 3: ACF plot for differenced spread series

hypothesis of stationarity, the test from *urca* package strongly rejects the hypothesis. Overall, we will assume that the series is not stationary. Therefore, for further analysis, our data will add the first difference of the spread series, call it spread returns. All the tests indicate stationarity of this series so the spread is I(1). Figures 4 and 5 show ACF and PACF plots correspondingly. While we would have expected about 2 lags to be significant, given 95% confidence, there are numerous significant lags up to the 29-th lag. So, there is presence of AR and MA structures. However, further order identification must be conducted carefully as there are higher order significant lags such as 19 and 21.

2.2.2 S&P 500 series

The next step is the analysis of S&P 500 price series. Figure 1 shows that S&P 500 is not stationary, at least, in mean. All test (that are mentioned previously) agree on the fact that the index series is non-stationary. The first difference series, call it returns,is stationary. So, S&P 500 is I(1). Additionally, Figures 6 and 7 show (partial) autocorrelation plots of returns of S&P 500. Small but significant (partial) autocorrelations are indicating a potential presence of ARMA(2,2) structure with p being potentially higher. We should note that including as higher lag as possible may lead to overfitting and greater computational complexity.

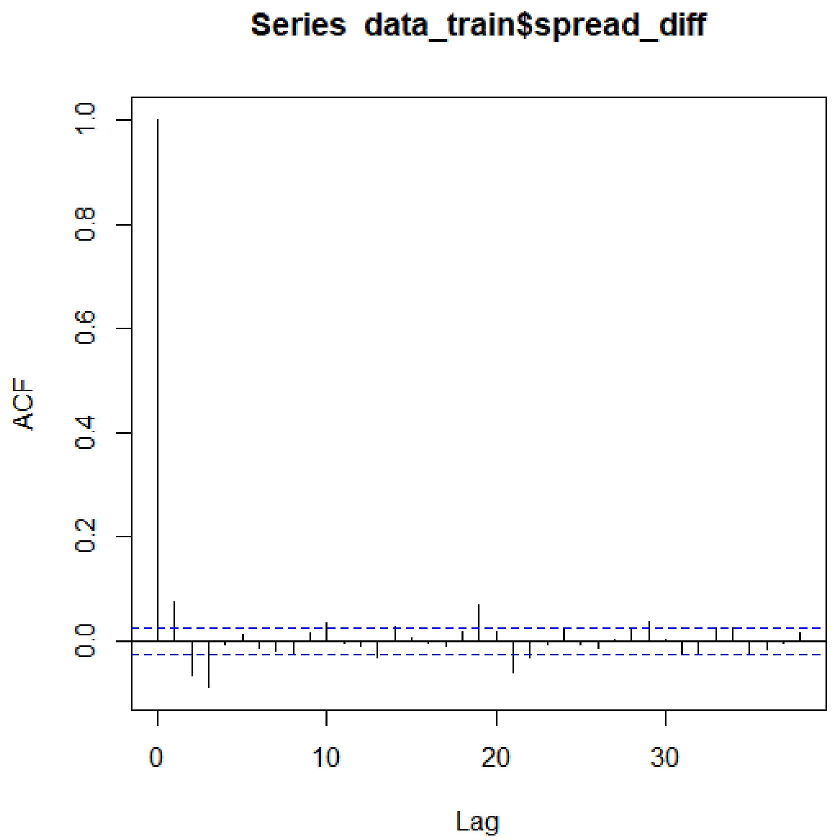

Figure 4: ACF plot for spread returns

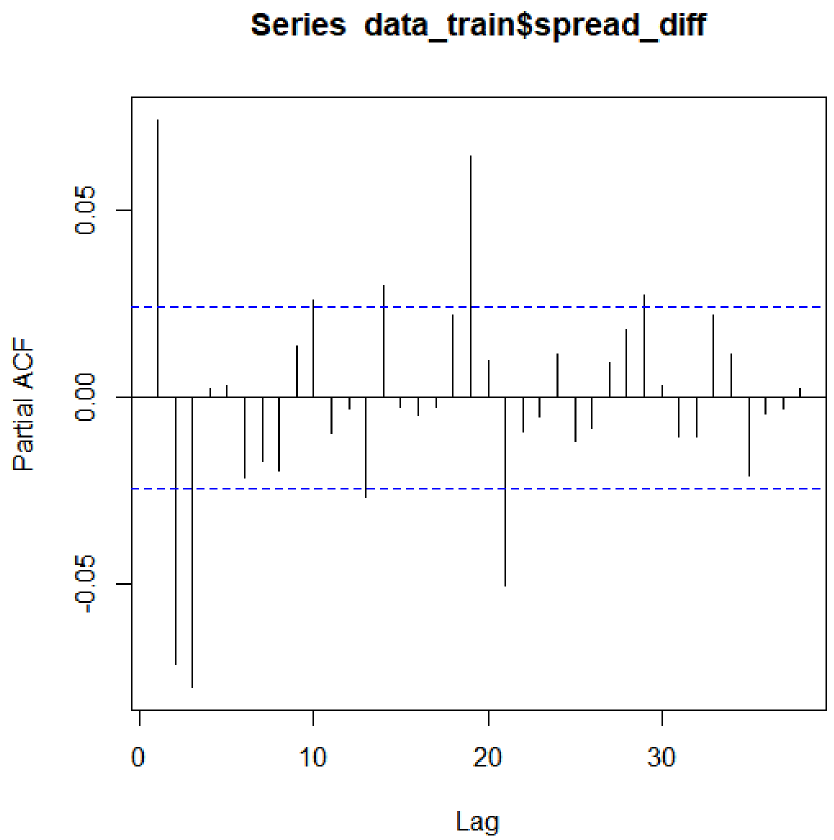

Figure 5: PACF plot for spread returns

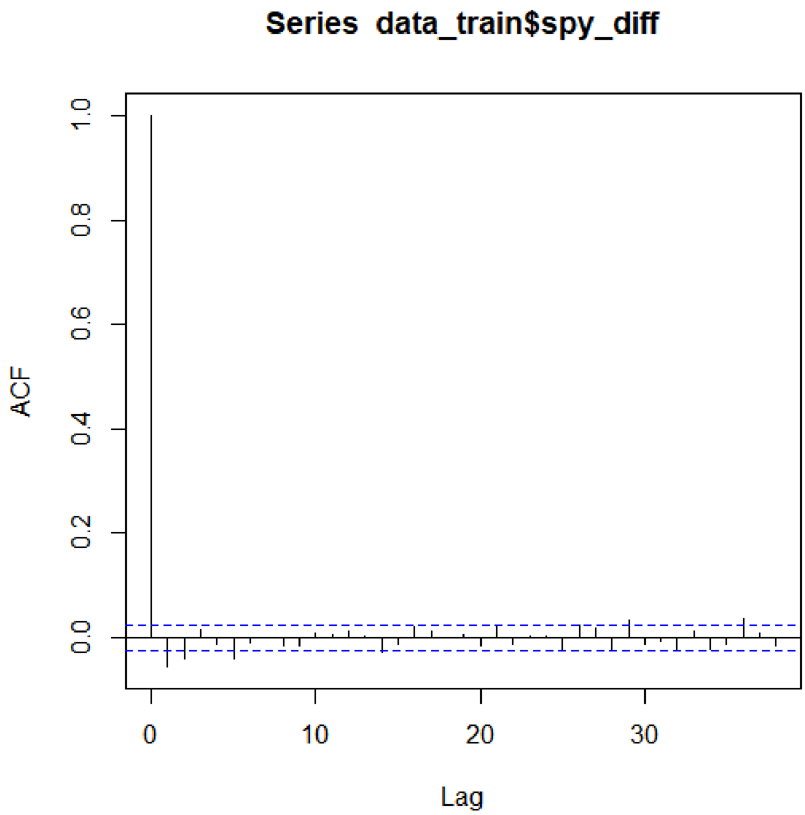

Figure 6: S&P 500 returns ACF plot

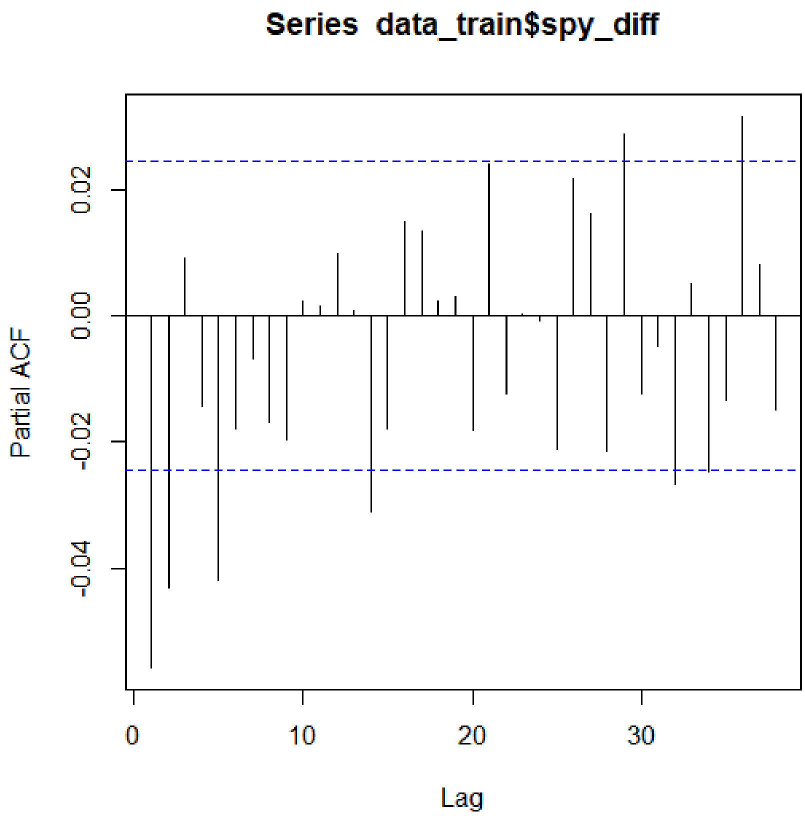

Figure 7: S&P 500 returns PACF plot

3 Modelling

3.1 Methodology

Recall our initial research interest: "Was there any relation between S&P 500 and bond yields spread?". If the answer is "yes" then we can potentially be better of jointly modelling them rather than each variable model on its own. So, now we have to carefully identify how can we analyze the joint relationship and how we measure the out-of-sample performance of joint and univariate models.

3.2 Relation between spread and S&P 500

Will the spread help us forecast the next S&P 500 values better? To answer this question, we need to understand the relationships between the series. Firstly, we know that their are both I(1). Intuitively, we may test the possible co-integrating relationship using the Johansen test. Further, we can conduct the Granger Causality test. Without a doubt, if there is a cointergrating relationship the there is some Granger causality, however, we conduct the two tests separately to double check our results. We mainly follow the algorithm described in Dave Giles' Blog (was mentioned in lecture)[1].

The first step was identifying the orders of integration of S&P 500 and yields spread which we have already did so far. The second step we set up a VAR model based on our two variables (not differences, actual series). We fit our VAR model using the *vars* R package. We include the intercept parameter and select the optimal lag using AIC. However, we also restrict the number of the maximum lags. The choice of this parameter is done by testing the residuals of both series on joint autocorrelation using Ljung–Box with the null hypothesis of independence of residuals. The resulting p-values are shown in Table 1. Figures 8 and 9 show ACF plots of residuals of spread series after applying VAR model at some lag p, this is just for the sake of illustration of changes of autoroccrelation values so ACF plots of the index series are not shown. Table 1 shows that after lag 3 we have high p-values indicating the the null hypothesis of joint independence at p lags (p also is the choice of a parameter in VAR model). Notice that Figures 8 and 9 show that while there is indeed quite low autocorrelations with lags up-to p, there are still relatively high autocorrelations after p (some of them might be even significant). At this point, I assume that lags after 30 might be well-enough.

[1] https://davegiles.blogspot.com/2011/04/testing-for-granger-causality.html

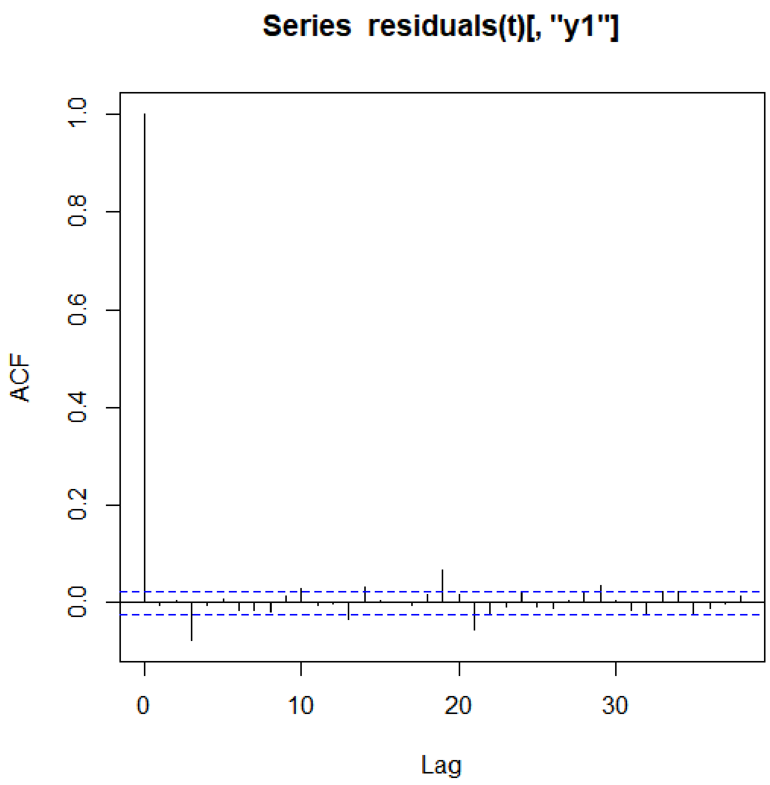

Figure 8: ACF plot of resiaduals of spread series with max lag = 3 in VAR model

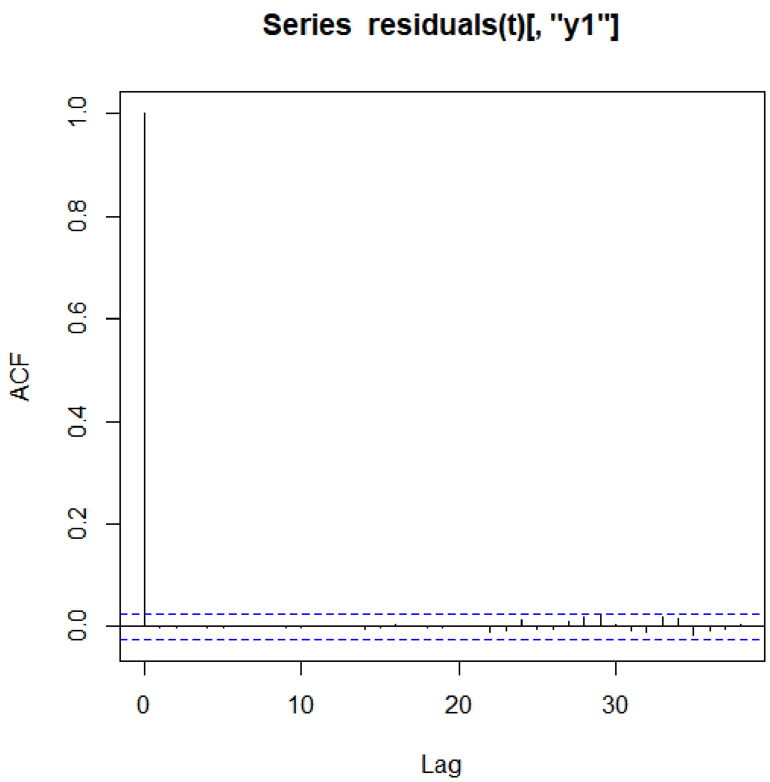

Figure 9: ACF plot of resiaduals of spread series with max lag = 30 in VAR model

Our next step is to test our original series on cointegration. We use the Johansen test, with function *ca.jo* from *urca* R package; we include the intercept and use the trace test. The null hypothesis is the number of cointegrating relationships r is r^*. We conduct two tests with $r^* = 0$ and $r^* \leq 1$. In the first case, the value of test-statistics is 20.07 which is higher than the 5pct threshold of 19.96. So, we can reject the hypothesis of $r = 0$ at 5pct significance. The second test results in the value of stat. of 7.98 which indicates significance at 10pct (stat. 7.52). Hence, we can only reject the null hypothesis at the 10pct significance level, while at 5pct (stat. 9.24) we do not reject the null hypothesis of $r \leq 1$ relationship. Overall, the test shows reasonable signs of the presence of a cointegrating relationship.
Finally, the Dave Giles blog suggests, we must then refit our VAR with $p + 1$ lag parameters and test the first p variables. There is no method that allows to do that automatically so I had to do it by myself. Here is the procedure [2]:

- 1. Get a shifted train sample data by $p + 1$ variables
- 2. Restrict the train sample by $p+1$ variables (drop $p+1$ as we get that number of NAs from the first step)
- 3. refit VAR with p lags and train sample with values of $p + 1$ lag as exogenous. All other params are the same as before.
- 4. Conduct the Wald test only only first p lags.[3]

The resulting p-value of the null-hyp that (SPREAD) DOES **NOT** GC (INDEX) is 0.2851 so we don't reject the null hypothesis, However, p-value of the null-hyp that (INDEX) DOES GC (SPREAD) is 4.148e-05 which results in that S&P 500 Granger cause the spread. Overall, I conclude that there is a cointegrating relationship between the series so we can use the VECM model.

3.3 Model comparison

We will test 4 + 1 models for each of the series, where 2 models are the Random walk model (null model), 2 ARIMA-GARCH models (on each model separately) and VECM. We

[2] see "Dave Gile procedure" in code attached

[3] As far as I understand, *causality* package tests without including the $p+1$-st exogenous lag. No other option I have not found. Please, don't punish hard if I messed up as I really tried doing my best here to verify all the details

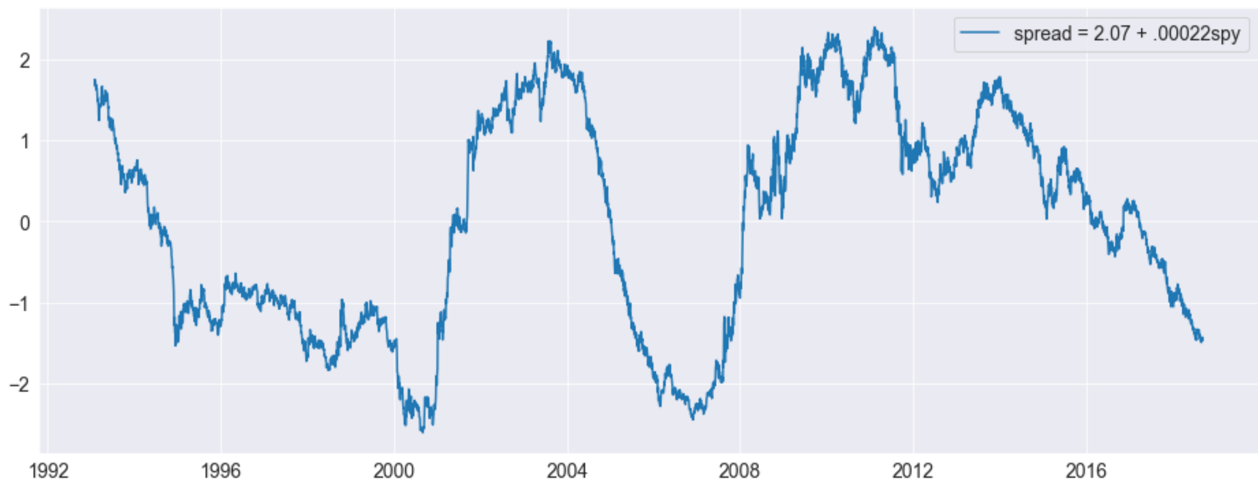

Figure 10: Spread between SP 500 and yields spread. Coeffitients were obtained using OLS

MaxLag	Spread	Index
1	1.744e-09	6.327e-06
3	0.6615	0.9601
5	0.993	0.996
30	0.9649	0.9859

Table 1: p-values of with Ljung–Box test for max lag p for associated series (null hypothesis: of independence of residuals)

will compare the resulting models by some measure of forecasting performance on the out-of-sample data. We define such measures to be MSE and the so called mean directional accuracy to account not only the value of our forecast but also the direction:

$$MSE = \frac{\sum_i (y_i - \hat{y}_i)^2}{N} \tag{3.1}$$

$$MDA = \frac{\sum_i \mathbf{I}_{sgn(y_t - y_{t-1}) = sgn(\hat{y}_t - y_{t-1})}}{N} \tag{3.2}$$

3.3.1 Model features

Recall, the RW model assumes the following specification

$$y_t = y_{t-1} + \epsilon_t$$

So, the optimal conditional forecast of such model will be y_{t-1}. In this model the previous values of prices or spreads will be used to forecast the next.

The second model is the ARIMA-GARCH model. Here, we have to define 4 parameters. We define GARCH(1,1) as standard. Now turn to ARIMA. Clearly, d should be 1 as we have already discussed, our series are $I(1)$. So, we train the model on returns. For spread returns, we mentioned earlier that the spread returns series shows significant higher order lags (see, Figures 4, 5). At first, p and q params were chosen by automatic procedure (with restricted $p \leq 20$ so that it converged in reasonable time) resulting in $p = 19$ and $q = 0$. However, such choice was not the best in terms of out of sample performance as there were better choices. Generally, there is a very large parameter space given significant lags for automatic procedure. So, I chose to $p = 3$ and $q = 3$ that resulted in relatively good performance, does not seem to overfit (as adding, for example, 20 lags may include other non-significant lags) and fast computation.

UPDATE: PLEASE, NOTE (!). Unfortunately, I could get reasonable rolling forecasts using ARIMA GARCH due to some technical issues. The code I googled does not work the way it is supposed to work. I had some very good results before however there were unreliable due to look-ahead bias introduced by mistake in parameters of the code. Now, the results are completly unreliable and I just have no idea how to fix it. I found it just before the deadline when validate the code and results. I could using not rolling forecasts but it won't be very consistent with other models (as their performance is measured with 521 data points). I chose not to restrict the data sample. So, I dismissed the ARIMA GARCH model. You still can find it in my code just to show that I tried...

Model	SP 500	spread
RW	19.19411	0.0025303
ARIMA	-	-
VECM	18.96260	0.002373

Table 2: MSE table)

Model	SP 500	spread
RW	0.507142	0.426785
ARIMA	-	-
VECM	0.503571	0.441071

Table 3: MDA table

We fit VECM using tsDyn package. lag parameter is 4, drift is included. The cointergrating coeffitient is obtained automatically by 2OLS procedure. We conduct rolling forecasts with 1 values ahead and refit each 5 step. Such params resulted in best error metrics. The results are reported in tables 2 and 3. It is observed that VECM did better than random for the spread series in terms of both metrics. S&P 500 obtained relatively better performance in forecasting values while loosing not much in terms of mean accuracy. So, unfortunately, I was not able to conduct the comparison with the ARIMA model but overall, we see that VECM did a little better than random. However, the philosophical questions is whether such complexity is worth.

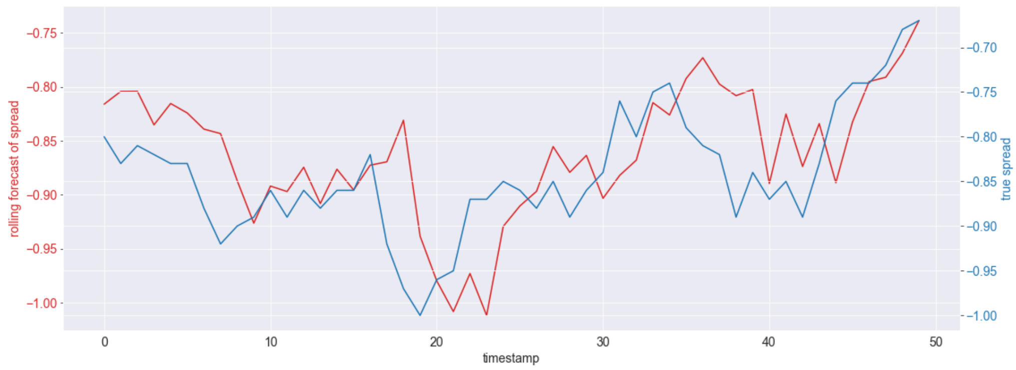

Figure 11: VECM, spread series, first 50 test data points

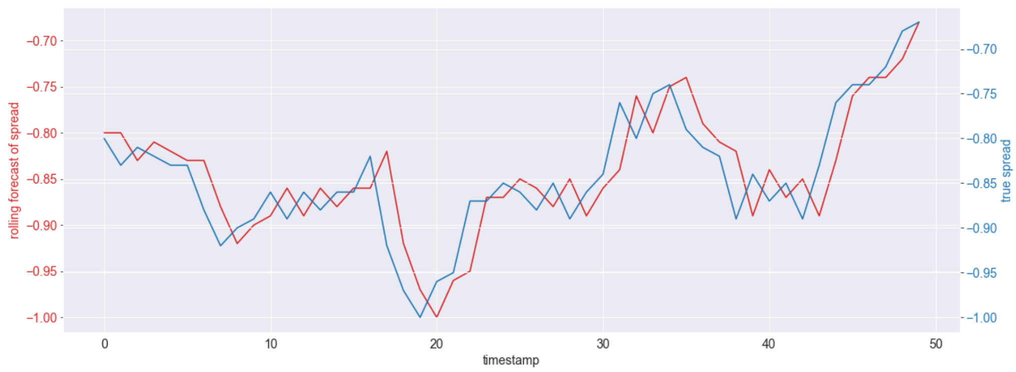

Figure 12: Random walk, spread series, first 50 test data points

4 Conclusion

Let us sum up what we have done. First of all, recall our initial research interest: "Was there any relation between S&P 500 and bond yields spread?". Now, recall our methedology. Firstly, we divided data into train and test sets. The choice was not dependent strongly dependent on some factors, this is a pitfall we should notice. Secondly, we analyzed the spread series on stationarity and found out that while there were some contradicting results, we concluded that it was non-stationary and integrated with order 1. We should be aware of the fact that we observed high heteroscedasticity in different periods of time. The choice of dataset and appropriate tests has a huge impact on the result of stationarity test. Our next major step was actually to study the relationships between our series. As both series were integrated of order 1, we had to analyze the cointegrating relationship extremely carefully, specifying all possible parameteres and initial hypotheses. Also, we carefully tested Granger Causality. Finally, we answered the question. The full answer should have been: "Using the described methodology, given this particular dataset and this particular split of data, we conclude that there is relationship between the series. In fact, there are chances that they are cointegrated.". While the answer is yes we must understand main limitations. I want to again emphasize that the choice of time period is really important. Moreover, I spent much time in reading descriptions of the tests procedures of R implementations as this is also very important when data is very "volatile".

At the last step we tried forecasting models based on our results. VECM showed, overall, better perfomance than the null model (Random walk) but not with great absolute improvement. So, there are still questions like "is the model specification right?", "is, in fact, the relationship that strong?", "is such complexity worth time?" or "are there other models that can perfom better?". These are still open questions.