Prediction of credit spreads

Anastasia Yaschenko

November 18, 2020

1 Problem statement

Credit spread is the difference between the yield on two debt instruments with the same maturity but different risks. Usually it is calculated between risk-free government and corporate bonds, but instead of government bond investors can choose some other benchmark bond (such as AAA-rated bond). Essentially, the spread captures the difference in credit quality since investors demand a premium for bearing additional credit risk. The higher is the credit spread of a corporate bond, the riskier it is.

Credit spread changes over time as economic conditions change. It widens in times of economic downturn and decreases in times of growth. The explanation is as follows: when market conditions deteriorate, investors prefer safer assets – government bonds. Increased demand increases prices of Treasuries and so reduces their yield. At the same time, the yield on corporate bonds increases because their price falls as investors sell them. Therefore, yields on government and corporate bonds move in different directions, and the spread widens. Reduced spread in times of economic prosperity can be explained similarly.

Credit spread is an important indicator used in a variety of situations. First, it is an indicator of the state of the economy. Secondly, since it accounts for credit risks (rating downgrades, default, etc.), it is widely used in risk management. For instance, it is used to calculate credit Value at Risk (VaR) for determining capital charges for credit risk. Finally, credit spreads are used in pricing of credit derivatives.

Given the applications of credit spread, it is important to know its future dynamics. It would allow to take actions to mitigate increased credit risk in advance, reducing possible losses. Therefore, this research paper aims to answer the question if credit spread can be reliably forecasted with basic time series models.

2 Data

The data for ICE BofA US High Yield Index Option-Adjusted Spread is taken from the site of Federal Reserve Bank of St. Louis (FRED). It is a spread index between spot Treasury yield and capitalization-weighted yield of US corporate bonds, which are rated BB or below (below investment grade). The data range is 6 years, from November 12, 2014 to November 12, 2020. The purpose of taking such range is to make data representative and to include dramatic increase in credit spread in 2016, so that the model could account for such situation. Two additional variables were taken for analysis: S&P500 and 13 week Treasury bill rate. To avoid having a situation when there are non-missing values in credit spread series and missing ones in the any of the other two series or vice versa, credit spread was matched with each of the two variables by date so as to remove missing values. Data on S&P500 index is also taken from FRED for the same period. Finally, data on 13 week Treasury bill has been obtained from Yahoo Finance for the same period as the other two time series. The frequency of all time series is daily. 11 last days (October 30, 2020 – November 12, 2020) are taken as test set, resulting in 10 daily forecasts. In order to reduce possible heteroscedasticity, logarithmic transformation is applied to time series.

3 Literature review

There is rather limited literature focused on prediction of credit spread. Most of the authors seem to agree about its determinants.

To start with, Avino and Nneji 2012 investigate whether credit spread is predictable at all. They use iTraxx Europe index data and find that it can be forecasted as a function of risk-free interest rate, slope of yield curve, firm‘s stock returns and volatility of assets. Also, it is found that linear models (more precisely, AR(1)) tend to produce forecasts superior to those of non-linear Markov models regime switching. Audzeyeva and Fuertes 2018 study predictability of credit spread in emerging markets. They use 2 models: one exploits information content in the current credit spread curve only, the other one includes several exogenous variables. These variables are level, slope, and curvature of US yield curve, US Treasury interest rate and its short-term volatility. Other predictors include domestic economy indicators, such as trade balance and its volatility and financial risk rating of a country. The authors conclude that the first model, in contrast to the second one, cannot produce results superior to random walk model. The second model performs better.

Bhar and Handzic 2011 use state-space models and Kalman Filter to find factors affecting credit spread.They find that VIX index, 10-year risk-free bond yield and S&P500 returns help to explain variation in credit spread. VIX is found to have a positive effect on credit spread, and this effect increases with lower bond ratings. Overall, they find that low-grade bonds are more sensitive to stock market variables. Increases in Treasury yields cause a fall in credit spread, which is consistent with definition the spread. This sensitivity falls monotonically as investment grade of a bond increases. Interestingly, they conclude that the correlation between S&P500 index and credit spread of investment grade bonds is positive, whereas it is negative for below investment grade bonds.

Clark and Baccar 2018 use asymmetric generalized autoregressive conditional heteroskedasticity (GARCH) models (AGARCH and GJR-GARCH) to find that factors that determine credit spread are: 10-year Treasury interest rate, interest rate volatility, slope of the yield curve, stock market returns and volatility, liquidity in the corporate bond market (proxied by long-term US dollar swap spreads) and foreign exchange rate. Treasury yield, slope of yield curve and liquidity have negative impact.

Krishnan, Ritchken, and Thomson 2010 use level, slope, and curvature factors of credit spread curve, Real Activity Index, Inflation Index, stock market momentum, stock market volatility and several firm-specific variables, such as book-to-market ratio. They find that given variables describing the shape of credit spread curve are present in the model, macroeconomic and firm-specific factors do not enhance predictions. Contrary to them, Tang and Yan 2008 examine the impact of economic conditions on credit spread. They find evidence of negative influence of GDP growth rate and investor sentiment and positive – of GDP growth volatility, systematic jump risk and firm cash flow volatility.

Brahimi 2017 makes an extensive study of the predictability of credit spreads, considering high yield and investment grade bonds separately. He finds that in both cases the spread is cointegrated with 10-year and 2-year government bond yield, VIX and ISM manufacturing index (monthly proxy for GDP). For high yield bonds, 10-year and 2-year government bond yields, earnings per share (EPS), lending standards, financial conditions and price-to-earnings (PE) ratio are found to be significant.

Bierens, Huang, and Kong 2005 find credit spreads to be mean-reverting in the long run. Mean and volatility are higher for lower grade bonds and bonds of longer maturity. The authors estimate ARX(1)-ARCH(1)-Jump model to find slope of yield curve, Russell 2000 index and VIX to be significant predictors of the spread.

Tauchen and Zhou 2011 claim that jump volatility has more explanatory power than systematic risk, volatility and interest rate factors. Jump volatility can be seen as a proxy for long-term macroeconomic and financial risk.

Güntay and Hackbarth 2010 assess the role of dispersion, or degree of variability, in analysts‘ forecasts of earnings. This dispersion acts as a proxy for uncertainty in cash flows, therefore, higher dispersions leads to higher credit spread. This result holds after controlling for bond-specific, firm-specific and macroeconomic variables.

4 Stationarity

First, time series are checked for stationarity. Two tests are used: Augmented Dickey–Fuller (ADF) test (H_0: non-stationary) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test (H_0: stationary). ADF test statistic has its own distribution. If this statistic is less than critical value, H_0 is rejected. KPSS also has non-standard distribution and its null hypothesis is rejected if its statistic is larger than critical value. The number of lags in ADF test was chosen by automatically BIC, since sample size is large. Specification without trend was used in all cases.

Critical values and results can be found in tables 1, 2 and 3. Based on these tests, all time series are found to be non-stationary. However, differencing them once makes them stationary (I multiply differenced data by 100 to avoid computational errors). Thus, all time series are integrated of order 1.

5 ARIMA model

First, I try to fit ARMA(p,q) model to differenced logarithmic credit spread data. By examining autocorrelation (ACF) and partial autocorrelation (PACF) plots in fig.4 and fig.5, I identify that lags for AR part can be p = 2, 7 or 10; for MA part q = 1, 7, 10. Overall, I fit 9 models with all possible combinations of p and q (results are in table 4). The model with the smallest BIC is ARMA(2,1), with smallest AIC – ARMA(7, 10).

Additionally, I fit ARMA model with automatic lag selection on the basis of BIC, which gives ARMA(2, 0). This model results in BIC = 6432.132 and AIC = 6416.081 – lower than any of the estimated models.

Next, I do residuals diagnostics. For the model to be good, residuals should be white noise. In order to check this, I use Ljung-Box test for serial autocorrelation. It has null hypothesis of absence of autocorrelation up to a specified lag. The number of lags tested should be large enough at first, and as long as null hypotheses are not rejected, it should be reduced. I start testing with 20 lags. When this test is used for residuals, there should be adjustment to degrees of freedom, equal to number of parameters (p + q + constant). For ARMA(2, 0), p-value of this test is 0.00465, thus H_0 is rejected at any reasonable significance level. Performing the same test for ARMA(2,1), chosen by BIC, results in a similar conclusion, with p-value = 0.003038. Checking residuals of ARMA(7,10) gives p-value = 0.121. We can test further only at 19 lags, since there are 18 parameters in this model and testing at 18 or more lags will result in zero or negative values of degrees of freedom. P-value = 0.04991 at 19 lags, allowing not to reject H_0 at 1% significance level. Therefore, ARMA(7, 10) for differenced logarithmic credit spread, or ARIMA(7, 1, 10) for logarithmic spread (since we difference once), is the best model.

To assess fit of the model, I use Mean Squared Error (MSE) and Mean Absolute Error (MAE). The formula for MSE is $\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, where y_i is true value, $\hat{y}_i$ is prediction. Formula for MAE is $\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$.

The in-sample MSE for ARIMA(7, 1, 10) is 3.48, MAE is 1.31. Fitted values are plotted in fig. 6. It can be seen they account for the variation in spread but do not account for its magnitude. Next, I make out-of sample predictions for the last 10 days (October 31, 2020 to November 12, 2020). The model obtained after fit is used to predict all 10 observations at once. The reasoning for such choice of forecasting procedure is that often long-term credit spread forecasts are more useful than short-term, since they allow enough time for the actions related to credit risk to be taken. Here, I make prediction only for 1 such period, but a more robust procedure would be to use a rolling window with prediction horizon for the next 10 days. The out-of-sample fit is MSE=21.214198 and MAE=3.649822. MSE has increased significantly since it amplifies errors by squaring them, whereas MAE preserves the original magnitude of errors. The visualisation can be seen in fig.7. It can be seen that predictions (red) are centred around 0 and do not account for the volatility in spread. However, I have also plotted predictions multiplied by 10 (blue) in order to better observe predicted movements of spread. Although the forecast does not follow the shape of real spread in all cases, it correctly represents the dynamics for the last 4 days.

6 Structural breaks

Next, I investigate with Quandt Likelihood Ratio (QLR) test whether there are structural breaks in level with unknown date present in the data. First and last 15% of observations are dropped when considering potential break dates. For every possible break date, Chow test statistic is computed. The maximum value of these statistics is QLR-statistic. In our case, QLR statistic = 4.9277, with p-value = 0.2548. Null hypothesis of no structural break in levels is not rejected.

Next, I test whether there is structural break in variance. I extract residuals from ARIMA(7,1,10) model, square them and regress on a constant. Essentially, this is a test for a break in levels of squared residuals, and the procedure is outlined above. QLR statistic is 74.268, p-value is 4.441e-16, making it possible to reject null hypothesis of no break in variance. The date for the likely break is December 2, 2019 (marked by red line in fig.1), linked to the start of coronavirus pandemic.

7 ARIMA model: accounting for structural breaks

In order to account for this break in variance, I fit two ARIMA(7,1,10) models. The first one is fit on the period November 12, 2014 – December 2, 2019, the second one – on the period December 2, 2019 – November 12, 2020. New predictions are made with the second model. Residuals of both models are white noise. For the first model, Ljung-Box test for no autocorrelation at 20 lags gives p-value=0.0518, not rejecting null hypothesis at 5% significance level. For the second model p-value = 0.1785, not rejecting null hypothesis of no autocorrelation at 20 lags for any reasonable significance level.

The forecasts can be seen in fig. 8. It can be seen that the mean of predictions is higher and they are more variable. Also, there is a small improvement in the forecast of the direction of credit spread percentage change. However, the in-sample and out-of-sample performance is worse than for ARIMA(7,1,10) model fit on whole dataset. In-sample MSE for the first model is 2.501726, for second – 7.823194. Although the model seems to improve fit for the period before coronavirus, the fit for coronavirus period is significantly worse. Out-of-sample MSE =24.410859 and MAE = 3.730919 – worse than for the previous ARIMA(7,1,10). Probably it could be explained by the fact that dataset for the second model accounting for structural breaks does not include similar periods of economic distress from the past. When ARIMA(7,1,10) was fit on the whole dataset, such period was included, resulting in better out-of-sample predictions. Thus, it can be concluded that advantages of accounting for breaks in variance do not outweigh advantages of a representative dataset. Therefore, next models are trained on the whole period of 6 years.

8 VAR model, Granger causality and cointegration

The literature has documented that performance of S&P500 index influences credit spread. Thus, it is possible that constructing vector autoregression (VAR) model with credit spreads and S&P500 can improve forecasts. Based on AIC criterion, best model contains 10 lags, based on BIC – 2 lags. For estimated VAR(2), in-sample MSE = 3.683811 , MAE = 1.329887 for credit spread series. For VAR(10), in-sample MSE = 3.546667, MAE = 1.327036. Out-of-sample MSE for VAR(2) is 21.196214, MAE is 3.457809. For VAR(10), MSE = 23.569692, MAE = 3.698996. Thus, VAR(10) has better in-sample fit, but it seems to overfit since its predictions are inferior to forecasts of VAR(2).

Using VAR(2), we can test for Granger causality. Null hypothesis is that there is no Granger causality for 2 variables. Testing for H_0: S&P500 does not Granger cause credit spread results in F-statistic = 9.813 with p-value = 5.652e-05, so null hypothesis can be rejected at 1% significance level. Testing for H_0: credit spread does not Granger cause S&P500 results in F-statistic = 3.5183 with p-value = 0.02977. Null hypothesis can be rejected at 5% significance level. Given these results, we might be interested in testing if these time series have cointegrating relationship. If so, there would be no need to difference time series and lose some fraction of information, since this relationship can be used to construct stationary linear combination of these variables.

To run tests on cointegration, I use non-differenced series of logarithmic values. The first test is Johansen test, which consists in iterative testing for rank of coefficient matrix related to first lags in Vector Error Correction Model (VECM). If rank = 0, there is no cointegration; if rank = 1, there is cointegration, if rank = 2, the time series are initially stationary. Thus, Johansen tests starts with testing H_0: rank = 0 against H_1: rank = 1. If H_0 is rejected, new pair of hypotheses is tested: H_0: rank = 1, H_1: rank = 2. From table 7 it can be seen that statistic = 4.09 – less than any critical value, so H_0: rank = 0 is not rejected and time series are not cointegrated. Using Engle-Granger test with H_0: no cointegration and specification without trend, p-value ≥ 0.10, also resulting in evidence of no cointegration.

Next, I test for cointegration between credit spread and 13 week Treasury bill rate. In Johansen test with H_0: rank = 0, statistic is 16.11, which is more than critical value of 14.90 at 5% significance level, so null hypothesis is rejected. Testing H_0: rank = 1 results in statistic = 3.10, which is less than any critical value. In this case, null hypothesis is not rejected, so matrix rank is 1 and there is cointegrating relationship. However, using Engle-Granger test results in p-value ≥ 0.10, meaning there is no cointegration. The results are contradictory, but let us assume than Johansen test result is correct and the time series are indeed cointegrated, so that it is possible to construct VECM model. From fig.11 it can be noted that in period when credit spread is low, Treasury rate is high, and when credit spread widens, Treasury rate falls. This is consistent with definition of credit spread as difference in yield on risk-free and risky debt instruments.

9 VECM model

I choose 2 lags for VECM model, since this is the number of lags which produced the best results for VAR model. This model is estimated using undifferenced logarithmic credit spread and 13 week Treasury bill rate. As can be seen in fig.12, VECM predictions systematically underestimate logarithm of credit spread (by more than 4%). In-sample MSE = 2.373239, MAE = 1.521309. Out-of-sample MSE = 15.775360, MAE = 3.970785.

10 Conclusion

Comparison of in-sample and out-of-sample performance of the models is represented in tables 5 and 6. Out of all models, VECM(2) seems to have the best in-sample fit based on MSE. Judging by MAE, the best in-sample fit is produced by ARIMA(7,1,10) trained on the whole dataset. As MSE amplifies large errors, it can be the case that VECM(2) makes fewer large errors, so it has lower MSE. However, in total it makes more errors, and so has higher MAE, which preserves magnitude of the data. What regards out-of-sample performance, based on MSE VECM is again the best model, but based on MAE, the best model is VAR(2). Therefore, the optimal model depends on the purpose of research and cost of errors. If the purpose is to understand which variables influence credit spread, then in-sample performance should be paid attention to. If the goal is prediction, then out-of-sample performance should be maximized. Similarly, if the goal is to make as few errors overall as possible, the model should be chosen on the basis of MAE, if the goal is to make as few large errors as possible – on the basis of MSE. It is worth noting that predictions of the best models considered are rather poor, so more complex models should be investigated to produce reliable forecasts.

I have performed tests to find cointegrating relationship between credit spread and S&P500 index and 13 week Treasury bill. Whereas there is no cointegration of credit spread with S&P500, there is long-run relationship with 13 week Treasury bill. This finding complements previous research, where authors find cointegration with 2 year Treasury notes and 10 year Treasury bonds.

Also, by comparing results of ARIMA(7,1,10) models trained on the whole dataset and on two parts of dataset separately, I find that inclusion of representative period of data may be more important than accounting for structural break in variance. This arises from the fact that ARIMA model trained on data covering coronavirus pandemic captured dynamics only of current crisis, but it could not recognise how it evolves in time (especially, how it finishes, since the data is available for approximately start and middle of crisis). Therefore, inclusion of full crisis period in data helps the model learn relevant dynamics and dependencies.

11 Appendix

Figure 1: ICE BofA US High Yield Index Option-Adjusted Spread. Source: FRED

Figure 2: S&P 500. Source: FRED

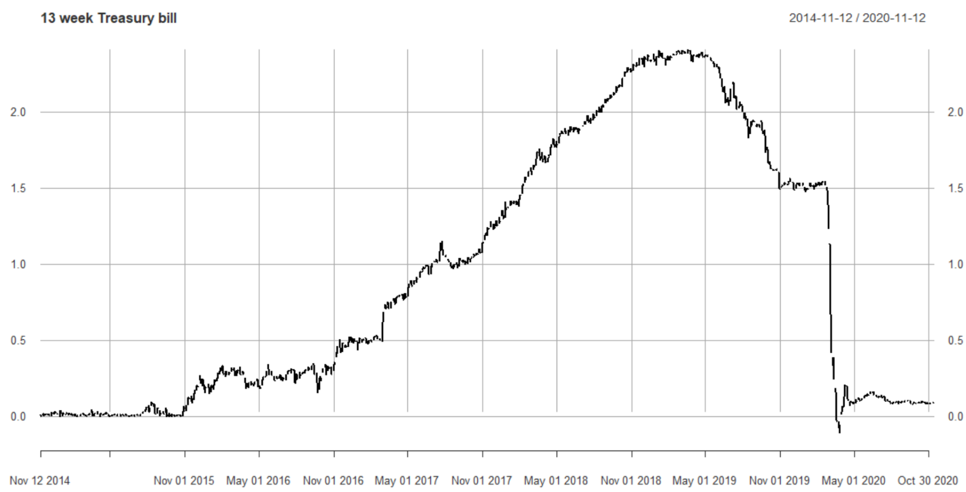

Figure 3: 13 week Treasury bill. Source: Yahoo Finance

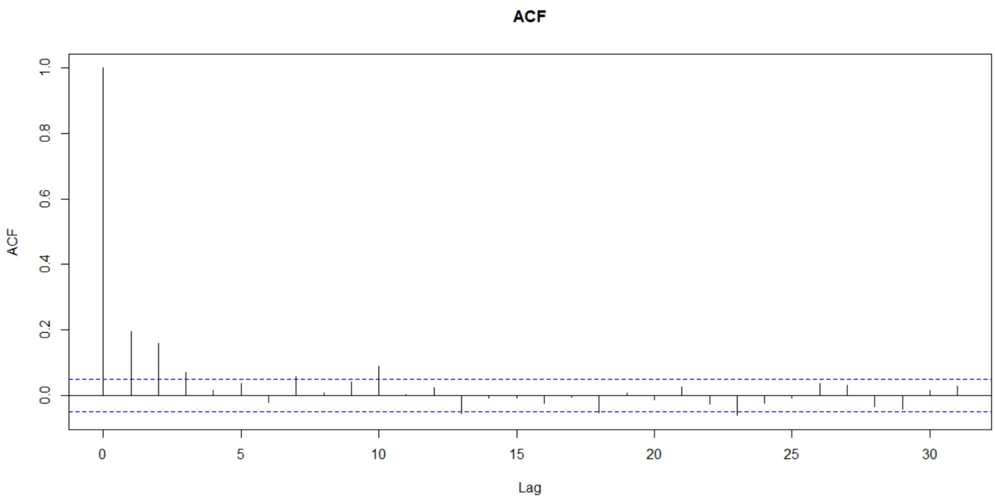

Figure 4: Autocorrelation plot for differenced log(spread)

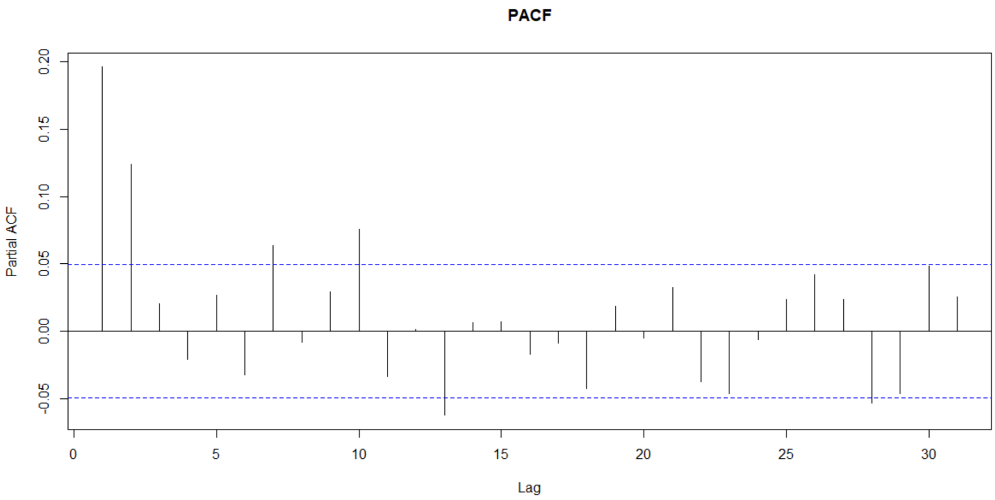

Figure 5: Partial autocorrelation plot for differenced log(spread)

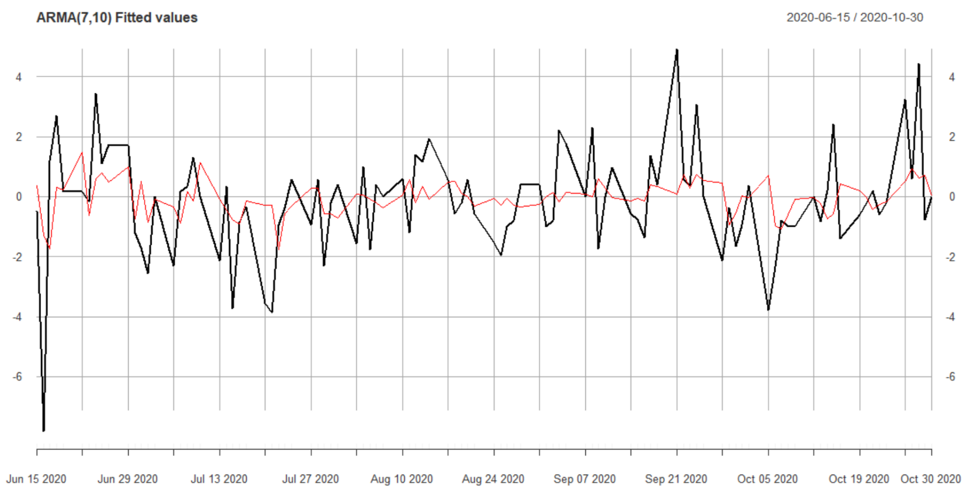

Figure 6: Black: last 100 observations of differenced log(spread). Red: fitted ARIMA(7,1,10) values.

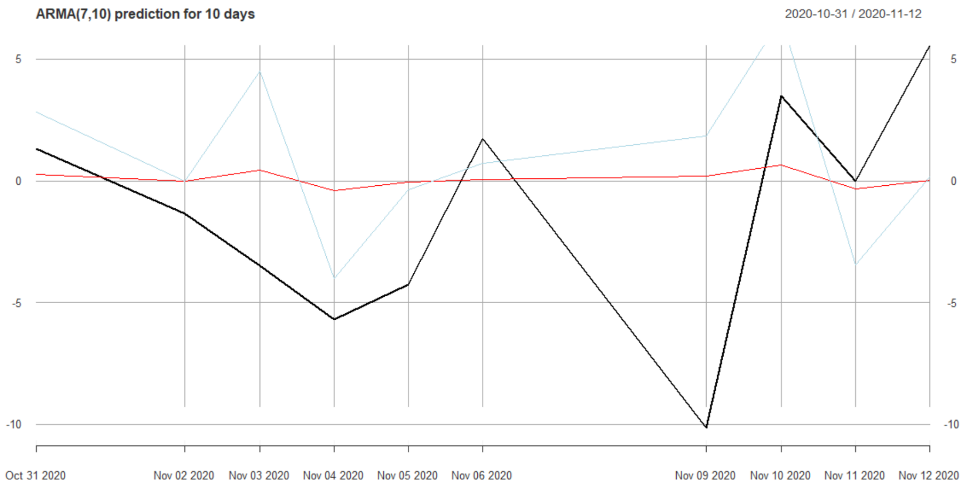

Figure 7: Black: differenced log(spread) for the last 10 days. Red: ARIMA(7,1,10) predictions. Blue: ARIMA(7,1,10) predictions multiplied by 10.

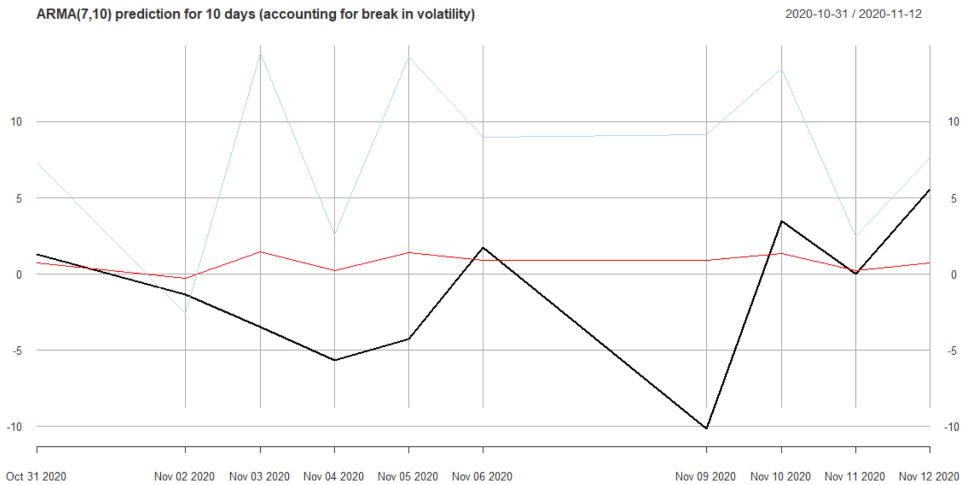

Figure 8: ARIMA(7,1,10) model, trained on the period December 2, 2019 – November 12, 2020. Black: differenced log(spread) for the last 10 days. Red: ARIMA(7,1,10) predictions. Blue: ARIMA(7,1,10) predictions multiplied by 10.

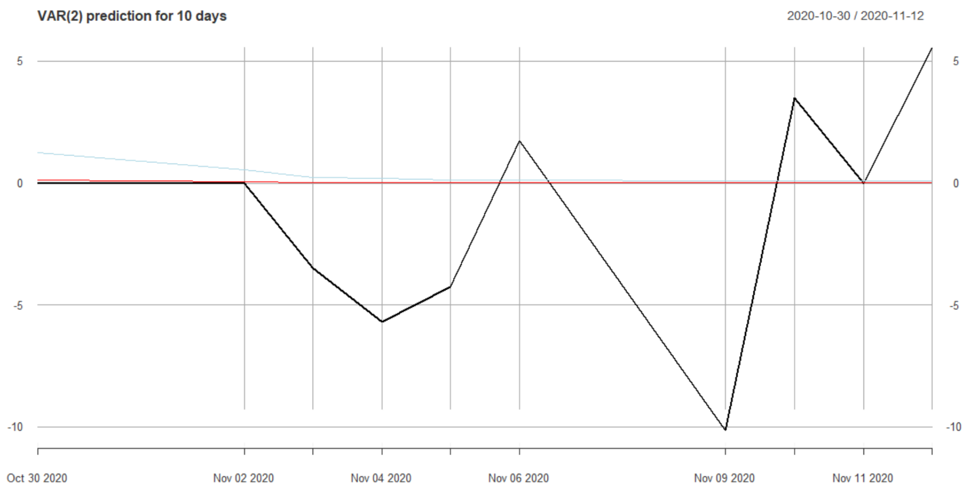

Figure 9: Black: differenced log(spread) for the last 10 days. Red: VAR(2) predictions. Blue: VAR(2) predictions multiplied by 10.

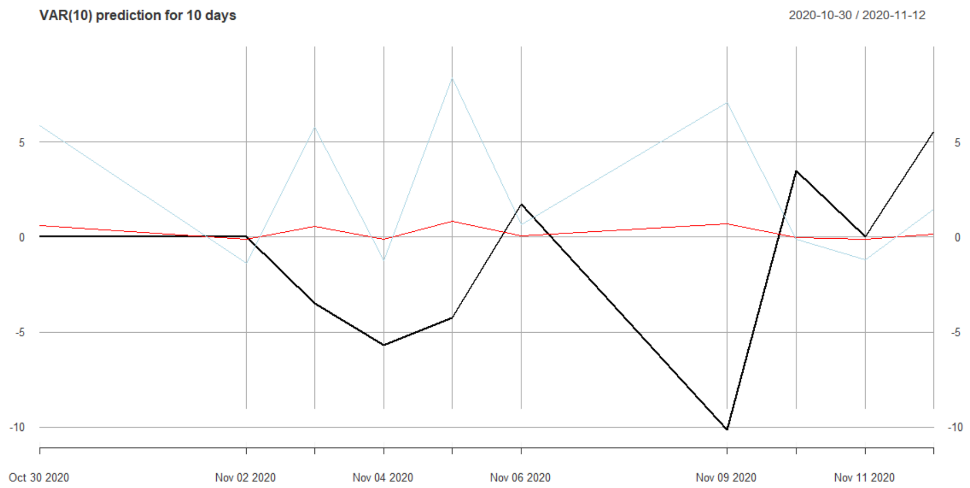

Figure 10: Black: differenced log(spread) for the last 10 days. Red: VAR(10) predictions. Blue: VAR(10) predictions multiplied by 10.

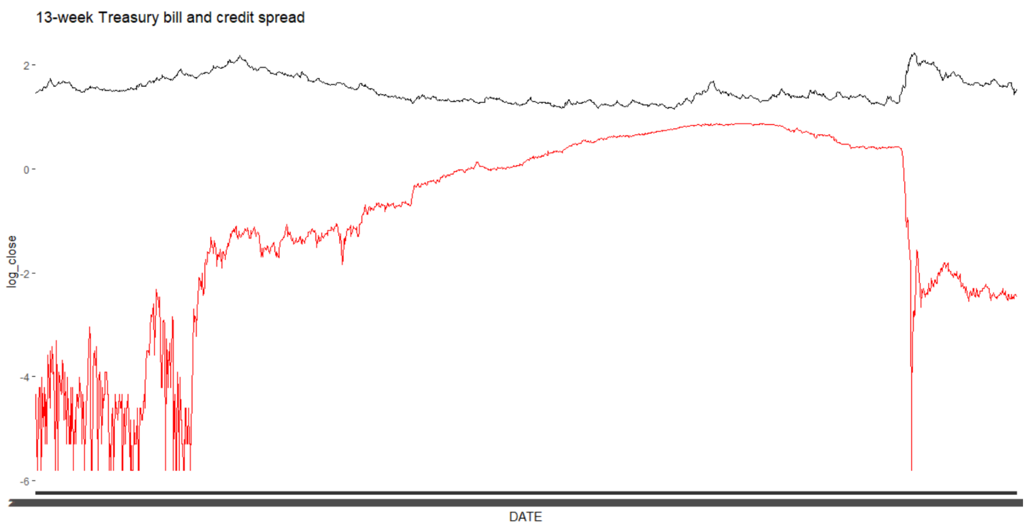

Figure 11: Black: logarithm of spread. Red: logarithm of 13 week Treasury bill rate.

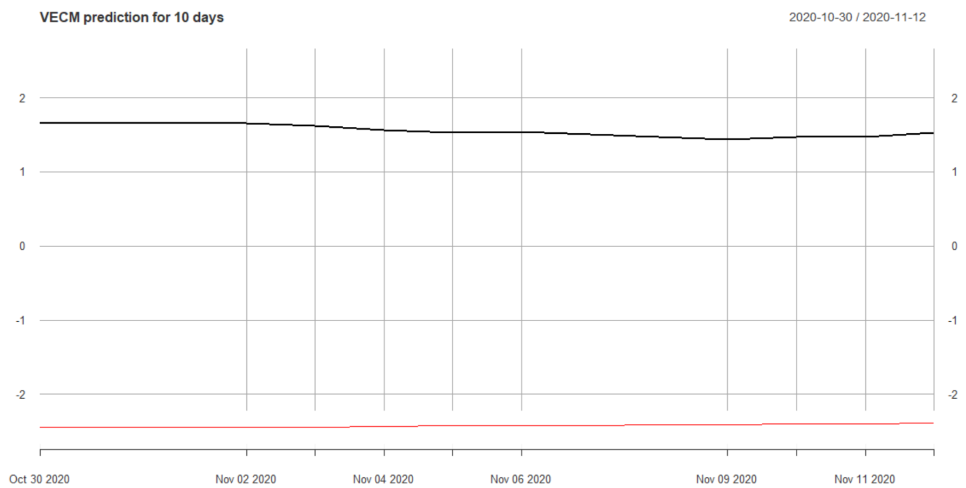

Figure 12: Black: logarithm of spread for the last 10 days. Red: VECM(2) prediction.

Table 1: Critical values for ADF and KPSS tests

	10%	5%	1%
ADF critical values	-3.43	-2.86	-2.57
KPSS critical values	0.347	0.463	0.739

Table 2: Statistics for undifferenced time series

	Spread	S&P500	Treasury bill
ADF statistics	-2.0745	-0.6501	-1.9481
KPSS statistics	3.7609	17.5938	8.9419

Table 3: Statistics for differenced time series

	Spread	S&P500	Treasury bill
ADF statistics	-21.9097	-11.9222	-20.4743
KPSS statistics	0.0707	0.0512	0.1743

Table 4: ARMA model specifications

	AR	MA	BIC	AIC
1	2.00	1.00	6446.29	6419.54
2	7.00	1.00	6472.98	6419.48
3	10.00	1.00	6481.92	6412.36
4	2.00	7.00	6481.69	6422.83
5	7.00	7.00	6498.02	6412.42
6	10.00	7.00	6506.09	6404.43
7	2.00	10.00	6490.43	6415.52
8	7.00	10.00	6504.09	6402.43
9	10.00	10.00	6524.48	6406.77

Table 5: In-sample fit of models

	ARIMA(7,1,10)	ARIMA(7,1,10) (total fit for 2 models)	VAR(2)	VAR(10)	VECM(2)
MSE	3.48	10.32492	3.683811	3.546667	2.373239
MAE	1.31	3.17043	1.329887	1.327036	1.521309

Table 6: Out-of-sample fit of models

	ARIMA(7,1,10)	ARIMA(7,1,10) (total fit for 2 models)	VAR(2)	VAR(10)	VECM(2)
MSE	21.214198	24.410859	21.196214	23.569692	15.775360
MAE	3.649822	3.730919	3.457809	3.698996	3.970785

Table 7: Johansen cointegration test results

	critical value 10%	critical value 5%	critical value 1%	S&P500	Treasury rate
rank = 1	6.50	8.18	11.65	0.00	3.10
rank = 0	12.91	14.90	19.19	4.09	16.11

List of Figures

List of Tables

References

Audzeyeva, Alena and Ana Maria Fuertes (2018). "On the predictability of emerging market sovereign credit spreads". In: *Journal of International Money and Finance* 88, pp. 140–157.

Avino, Davide and Ogonna Nneji (2012). "Are CDS Spreads Predictable? An Analysis of Linear and Non-Linear Forecasting Models". In: *SSRN Electronic Journal* 42848.

Bhar, Ramaprasad and Nedim Handzic (2011). "A Multifactor Model of Credit Spreads". In: *Asia-Pacific Financial Markets* 18.1, pp. 105–127.

Bierens, Herman, Jing-zhi Huang, and Weipeng Kong (2005). "Time-Series Estimation of Aggregate Corporate Bond Credit Spreads". In: *Working Paper, Penn State University*, pp. 1–44.

Brahimi, Marouane (2017). "The impact of macro- economic indicators on credit spreads". In:

Clark, Ephraim and Selima Baccar (2018). "Modelling credit spreads with time volatility, skewness, and kurtosis". In: *Annals of Operations Research* 262.2, pp. 431–461.

Güntay, Levent and Dirk Hackbarth (2010). "Corporate bond credit spreads and forecast dispersion". In: *Journal of Banking and Finance* 34.10, pp. 2328–2345.

Krishnan, C. N.V., Peter H. Ritchken, and James B. Thomson (2010). "Predicting credit spreads". In: *Journal of Financial Intermediation* 19.4, pp. 529–563.

Tang, Dragon Yongjun and Hong Yan (2008). "Market conditions , default risk and credit spreads Dragon Yongjun Tang Discussion Paper Series 2 : Banking and Financial Studies". In: *Journal of Banking & Finance* 34.08, pp. 743–753.

Tauchen, George and Hao Zhou (2011). "Realized jumps on financial markets and predicting credit spreads". In: *Journal of Econometrics* 160.1, pp. 102–118.