Reflection Model: A New Predictor for Credit Spread

Part 1 Introduction

As one of the key factors to measure market risk, credit spread has a significant influence on predicting economic trends and further constructing investment strategies. For the purpose of this research, we obtain 21 features after feature engineering. In this paper, we implement three ways for predicting credit spread: "Basic Prediction", "Long-run Trend Prediction" and "Accurate Daily Prediction".

In Basic Prediction, we constructed the prediction model by applying Long-Short Term Memory ("LSTM") model. Although the outcome seemed acceptable, it still had two major shortcomings: 1. Prediction only for the next day; 2. Low precision. To solve the first problem, we derived Long-Run Trend Prediction model to get prediction without future information.

To work out the second problem, we established Accurate Daily Prediction model (we also name it as "Reflection Model"), analyzing the difference between the predicted value from the Basic Prediction and true value based on Kalman Filter ("KF").

We used the credit spread of AAA grade bonds over 10-Year Treasury Bond as the label when constructing the models. Further, we applied the model to predict the credit spread of all US investment grade bonds. The three prediction models will be introduced in Part 3. Results of models will be presented in Part 4. Statistical tests on the predictor will be applied in Part 5. Finally, we simulate an investment stategy and try to predict economic state using the models. This application verifies that our research and these models are meaningful and innovative.

Part 2 Data and Feature Engineering

From WRDS, OECD, CRSP, and FRED databases, we collected 23 features of daily frequency from Jan. 1st, 1997 to Jan. 15th, 2019, including 3-Month Financial Commercial Paper Rate, TED spread, Dow Jones Industrial Average, CBOE Volatility Index, and Fama-French 3 factors. All missing values of data are filled with previous ones.

We calculated the slope and curvature as a representation of the term structure of interest rate using 10-Year, 2-Year, and 3-Month Treasury Constant Maturity Rate. Inspiring from traditional time series analysis models, we added credit spread from the last three days as our new features. The credit spread is defined as the difference between the effective yield of Corporate Bond and 10-Year Treasury Constant Maturity Rate.

To select valid features for credit spread, we conducted Information Coefficient ("IC") Analysis on original features. For each feature and credit spread, we calculated the average of 30-Day moving coefficients. We also considered correlation coefficient decay by calculating 1-Day to 30-Day lag coefficient between features and credit spread.

According to the result, we dropped 5 features with the least correlations (less than 0.1), including Fama-French 3 factors, risk-free return rate (1-Month treasury bill rate) and G-10 Index of the exchange value of the dollar. Meanwhile, to avoid linear dependence trap, we selected the feature with the largest coefficient among those with similar economic meanings. We eventually obtained 21 features.

Part 3 Models

3.1 Basic Prediction: Long-Short Term Memory

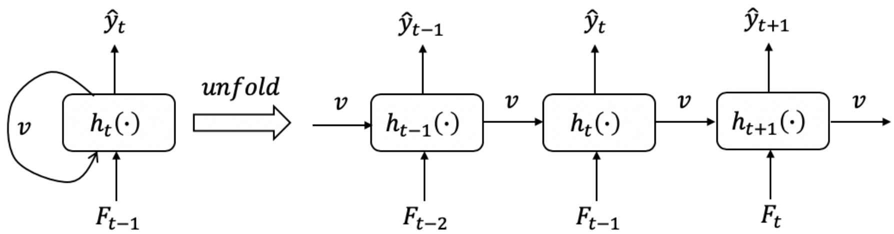
To predict credit spread based on collected features and historical data, one proper method is Long-Short Term Memory model. The LSTM model is a special type of Recurrent Neural Network ("RNN"). RNN is a typical model dealing with time series problem. Different from common RNN, it further considers the relationship between historical information and current state using memory cells in the neural network. Here, we skip basic descriptions of Neural Network and use the following notations to construct the network.

Let y_t denote the response variable, $i.e.$ credit spread; F_t denote the vector of all features, y_{t-k}: y_t denote terms from $t-k$ to t in a specific time series. The model aims to find a function $f(\cdot)$ such that

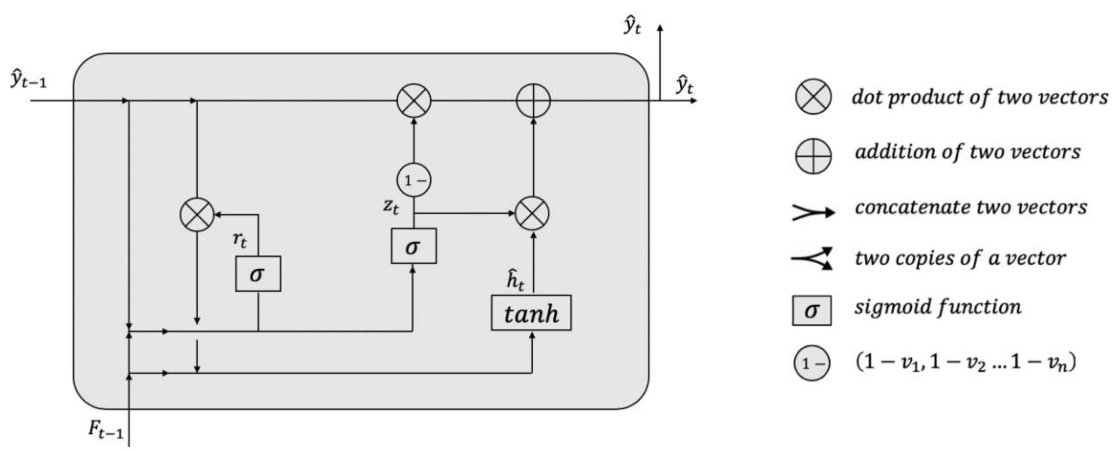
$$y_t = f(F_{t-1}, y_{t-m}: y_{t-1}) + e_t, \quad E[e_t | \mathcal{F}_{t-1}] = 0$$

where $\mathcal{F}_t$ is a filtration that represents the Information Set up to time t.

The Neural Network model assumes the function has a form of transition network like human's neural network. The Recurrent Neural Network follows such a form:



Common RNN always suffers from the exponential increase (decrease) of transition variable v. By introducing Constant Error Carousel ("CEC") units, LSTM model deals with the exploding and vanishing gradient problems. We also used LSTM model with Gated Recurrent Unit ("GRU"), a simpler version of LSTM model, to predict credit spread.



In the above graph, all layer functions can be calculated as follows:

$$z_t = \sigma(W_z \cdot [\hat{y}_{t-1}, F_{t-1}]), \qquad r_t = \sigma(W_r \cdot [\hat{y}_{t-1}, F_{t-1}])$$

$$\hat{h}_t = \tanh(W \cdot [r_t * \hat{y}_{t-1}, F_{t-1}]), \qquad \hat{y}_t = (1 - z_t) * \hat{y}_{t-1} + z_t * \hat{h}_t$$

where W_z, W_r, W are the parameters learned from the training process.

There are three reasons to choose LSTM as the predictor: 1. In the economic world, fluctuations are caused by events. The effect of such events is sometimes retentive. It's impossible for conventional time series model to convey this effect. 2. In LSTM, there won't be items like w_i^n, and thus exploding and vanishing gradient problems, which promises the predictor can perform better on a big data set. 3. The memory that is controlled by a sigmoid layer may be long or short according to whether there is a new significant event happened. This characteristic is very similar to economic time series.

When applying this model to predict credit spread, we used m-day past information of credit spread and current information of features, $e.g.$ assume at time n, we use the credit spread at time $n, n-1, \ldots n-m+1$ and features at time n to predict credit spread at time $n+1$.

The Basic Model showed a good prediction in the long run, but if we focused on the short run, it would show obvious hysteresis. In this study, we chose $m = 3$ concerning the workloads, algorithm complexity, and statistic test. The hysteresis is about one or two days. The result is to be shown in the next part. To conquer the problem, we need to improve the model.

3.2 Long-Run Trend Prediction with EMD

Another big problem of the Basic Model is the lack of practicability. If we predict the value of credit spread with this model, it will not be profitable in practice. Even though it performs well in the long run, the predictor still restricts us from predicting on the next day, providing no further insights.

A more practical method is to learn parameters purely from the training set without features from the test set. A simple method is to treat y_t as future ten days' credit spread $i.e.$ a series prediction. However, this could be challenging due to the amplification of noise in the iterations (the model is difficult to converge). In practice, any high-frequency signals will ruin the convergence significantly. Thus, we introduced the method of empirical mode decomposition ("EMD").

EMD is a method of breaking down a signal while not leaving the time domain. It filters out functions that form a complete and nearly orthogonal basis for the original signal. The functions, known as Intrinsic Mode Functions (IMF: a function that has only one extreme between zero crossings, and has a mean value of zero), are therefore sufficient to describe the signal, even though they are not necessarily orthogonal. The fact that the functions into which a signal is decomposed are all in the time-domain and of the same length as the original signal allows for varying frequency in time to be preserved. Signal processes often have multiple causes, and each of these causes may happen at specific time intervals.

For a signal $X(t)$, let m_1 be the mean of its upper and lower envelopes as determined from a cubic-spline interpolation of local maxima and minima. The locality is determined by an arbitrary parameter; the effectiveness of the EMD depends greatly on such a parameter.

More specifically, let local maxima points be $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$, the start and end point of the signal be $(x_s, y_s), (x_e, y_e)$. We apply cubic-spline interpolation towards these points to calculate upper envelopes. Between every two local maxima points, we use $f_i(x) = a_i x^3 + b_i x^2 + c_i x + d_i$ to fit the function in the interval $[x_i, x_{i+1}]$ and need to satisfy conditions below:

1. $f_i(x_i) = f_{i+1}(x_i)$ for $i \in (2, \dots n-1)$ and $f_j(x) = a_j x^3 + b_j x^2 + c_j x + d_j$ for $j \in (1, n)$
2. $f(x_s) = y_s$ and $f(x_e) = y_e$
3. $f_i{}'(x_i) = f_{i+1}{}'(x_i)$ for $i \in (2, \dots n-1)$
4. $f_i{}''(x_i) = f_{i+1}{}''(x_i)$ for $i \in (2, \dots n-1)$
5. $f''(x_s) = 0$ and $f''(x_e) = 0$

These five equations form a matrix function that can solve all the (a_i, b_i, c_i, d_i) and thus the function of upper envelopes in interval $[x_s, x_e]$. Similarly, we obtain the lower envelopes and calculate the means of upper and lower envelopes to generate a new sequence: m_1.

The first component h_1 is computed as: $h_1 = X(t) - m_1$. In the second sifting process, h_1 is treated as the signal, and m_2 is the mean of h_1's upper and lower envelopes and we compute $h_2 = h_1 - m_2$. This sifting procedure is repeated k times until negative local maxima and positive local minima don't appear in new generated sequence h_k .

Applying EMD on all features and historical credit spread, we try to predict on the test set using the Long-Run Trend Prediction model (to predict on the test set without updating input of model). The performance (showed and discussed in the next part) is not satisfactory, thus requiring improvement. A possible explanation is that the rule trained from the training set is effective only in a few days, instead of building a strong connection between spread and features. Nonetheless, the model is still active in predicting the economic state.

3.3 Accurate Daily Prediction with Kalman Filter

Another improvement for the Basic Prediction model is to enhance the accuracy of $\hat{y}_t$. A proper measure accuracy is the variance of deviation, $i.e.$ difference between the true value and predicted value. This model is called Accurate Daily Prediction, we also call it Reflection Model.

Define $\hat{d}_t = y_t - \hat{y}_t$, where $\hat{d}_t$ acts like white noise. If $\hat{d}_t$ was indeed white noise, then the Basic Prediction model would be the best estimator for credit spread. In fact, $\hat{d}_t$ follows some unknown process. Thus, we need to find an adjustment process d_t to improve the predictor.

$$y_t^* = \hat{y}_t(F_t, y_{t-1}\colon y_{t-m}) + d_t$$

where y_t^* denotes a better estimator. Kalman filter is a typical way to deal with noise reduction problems. The filter can be described as a discrete dynamic system.

$$d_t = A d_{t-1} + B u_t + W_t, \qquad \hat{d}_t = H d_t + V_t$$

This means the true value of d_t is determined by d_{t-1} and control vector u_t. W_t and V_t are normal white noises. $\hat{d}_t$ is observed value, determined by true value. A, B, H are parameters matrices. According to our model, we show the five important iteration equations used in programming.

$$\begin{aligned} d_{t|t-1} &= A d_{t-1|t-1} + B u_t, & P_{t|t-1} &= A P_{t-1|t-1} A^T + Q \\ d_{t|t} &= d_{t|t-1} + Kal_t \cdot \left(\hat{d}_t - H \cdot d_{t|t-1}\right), & P_{t|t} &= (1 - Kal_t \cdot H) \cdot P_{t|t-1} \\ Kal_t &= P_{t|t-1} \cdot \frac{H^t}{\left(H \cdot P_{t|t-1} \cdot H^T + R\right)} \end{aligned}$$

where $d_{t|t-1}$ denotes prediction at time t given information at time $t-1$; $d_{t-1|t-1}$ denotes the best result at time $t-1$; $P_{t|t-1}$ and $P_{t|t}$ denotes covariance matrix of estimation error corresponding to $d_{t|t-1}$ and $d_{t|t}$; Q denotes covariance of systematic noise; Kal_t denotes Kalman gain at time t.

In the next part, we will showKalman filter on the deviation and adjust prediction based on the filtered process. Theoretically, the method will reduce the variance of deviation. Through this method, we get two distributions, one from prediction, and one from observation. We assume that both of them follow a normal distribution.

Assume that $\mathcal{N}(\mu, \sigma^2)$ is the probability density function of normal distribution.

$$\mathcal{N}(\mu_0, \sigma_0^2) \cdot \mathcal{N}(\mu_1, \sigma_1^2) = \mathcal{N}\left(\mu', {\sigma'}^2\right)$$

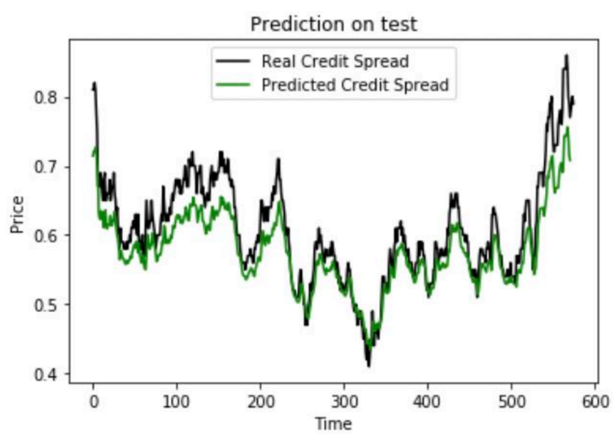
$$\Rightarrow \mu' = \mu_0 + Kal(\mu_1 - \mu_0), \qquad {\sigma'}^2 = \sigma_0^2 - Kal \cdot \sigma_0^2, \qquad Kal = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_1^2}$$

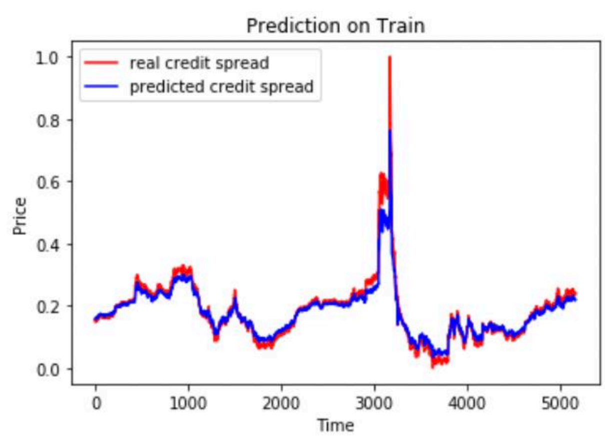
Because Kalman Gain must belong to $(0,1)$, when we coordinate two normal distributions, the variance must decrease. This conclusion will be confirmed in the next part.

Part 4 Results

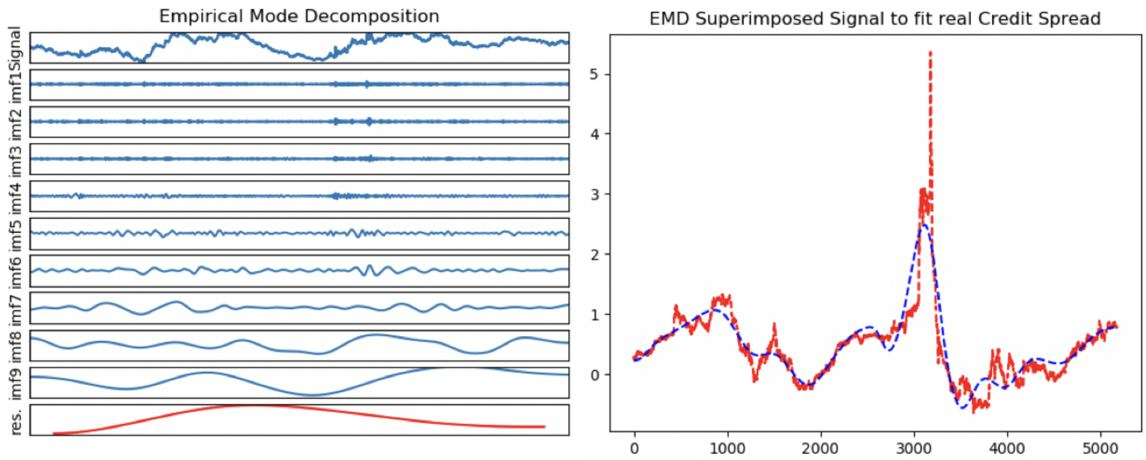
4.1 Results of three models

The Basic Prediction model uses credit spread of last three days and features of previous day to predict the next day's credit spread via LSTM model. We split all 5749-day data into the training set of 5174 days (use the last 10% as validation sets) and the test set of 575 days. We set five LSTM network (each contains 100 layers) and drop off 30% layers after training. The result is presented as follows:

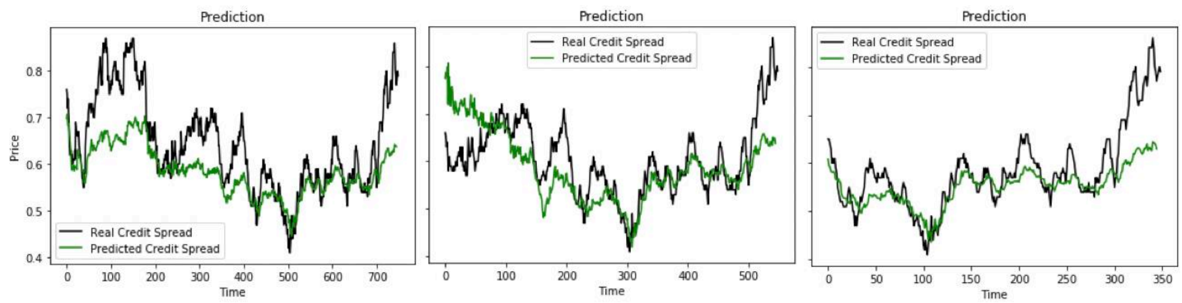




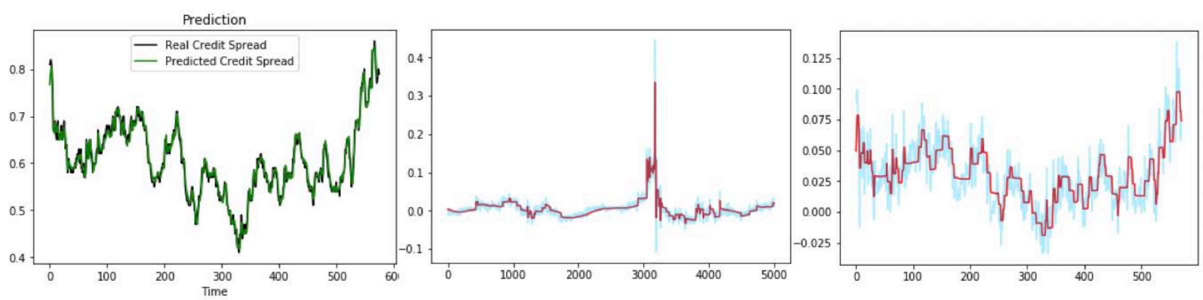
The Long-Run Trend Prediction model predicts credit spread without any information from the test set. We use EMD to decompose all features (here we only present EMD on credit spread: the left graph is IMF decomposed from credit spread, the right graph is the addition of IMF9-IMF6)



Using Long-Run Trend Prediction model, we obtain 10 results, here we present only three of them, correspondingly with the training set of 5000 days, 5200 days, and 5400 days.



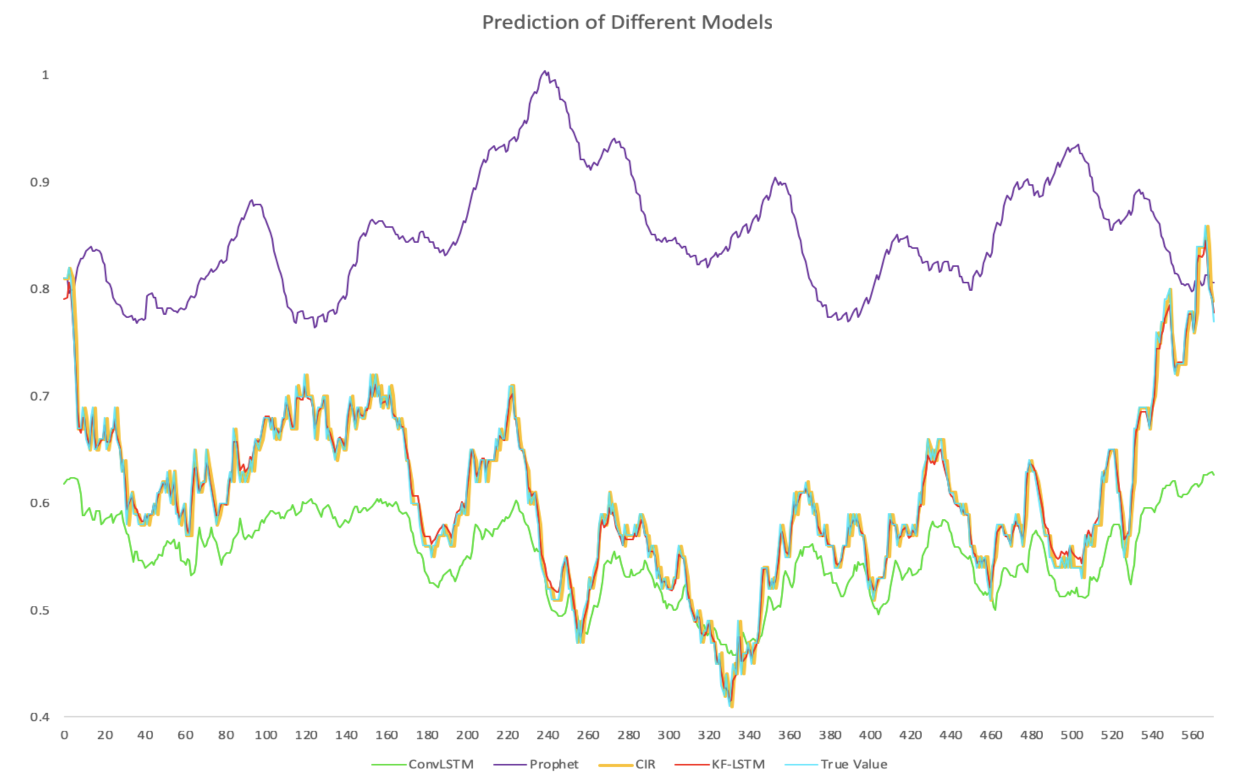
The model only describes a rough trend in the long run, with greater deviation. The Reflection Model is by Kalman Filter. The right two graphs below are filters on the deviation of prediction on the training set and the test set. The left graph below is the adjusted prediction on the test set.



4.2 Comparison with Benchmarks

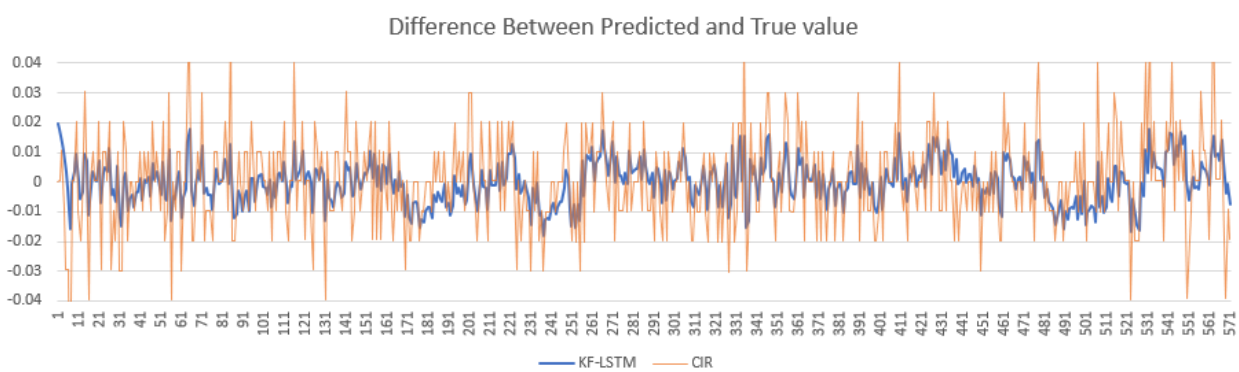
We use three benchmarks to compare with the Reflection Model (based on LSTM and Kalman Filter, noted KF-LSTM). The three benchmarks are 1. PROPHET model (a prediction model developed by Facebook); 2. Conventional LSTM model (not using deep layers and only using historical credit spread as inputs); 3. CIR model, a classical stochastic process for interest rate (the equation is as follows, where a, b, σ are parameters to be trained)

$$dR_t = (a - bR_t)dt + \sigma\sqrt{R_t}dW_t$$



Obviously, Conventional LSTM and Prophet have poor performances. Prophet may be more powerful for business analysis than financial analysis. Conventional LSTM may not learn high frequent data well. Both are not fed with external features.

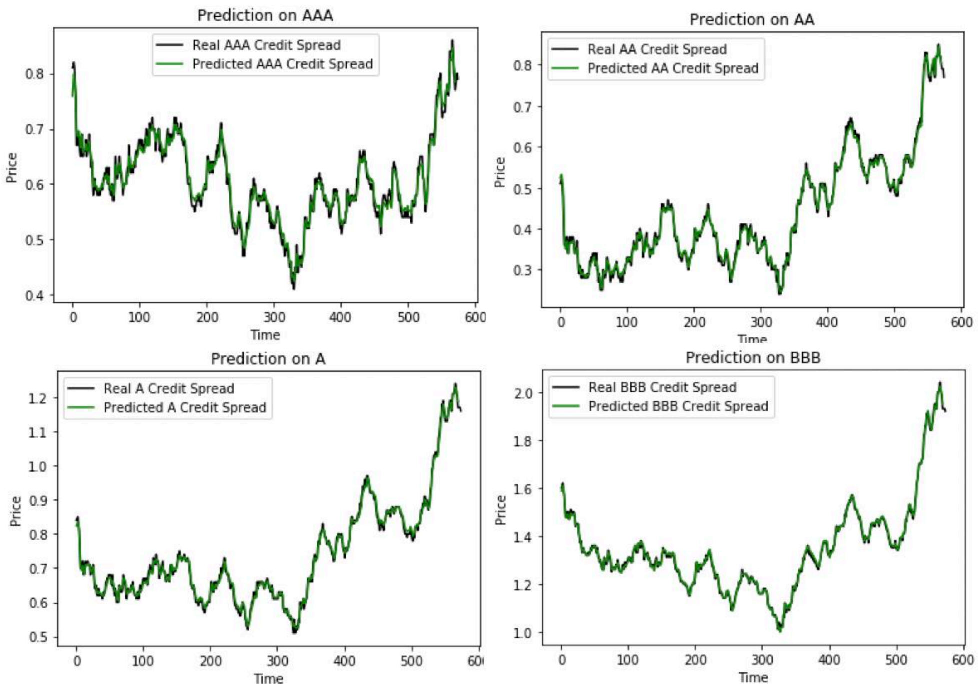
CIR performs almost as well as Reflection Model . To compare these two models, we observe the difference between the true value and predicted value of credit spread, $i.e.$ $y - \hat{y}$. We can see from the following graph, CIR has almost $-2 \sim +2$ basis point error, while Reflection Model has only $-1 \sim +1$ basis point error. It is a great breakthrough!



4.3 Results of US investment grade corporate bonds

When applying the predictor to all different types of US investment grade corporate bonds, $i.e.$ AAA grade, AA grade, A grade and BBB grade, we use the Reflection Model .

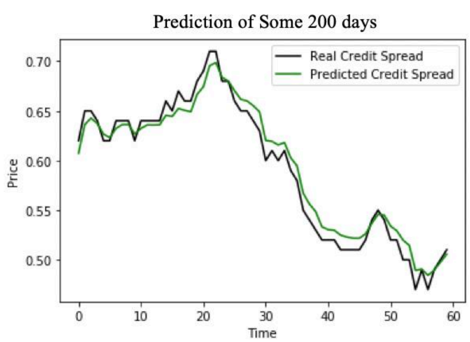
The results are as follows:



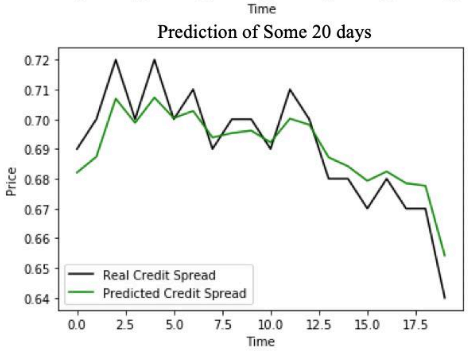
Part 5 Performance

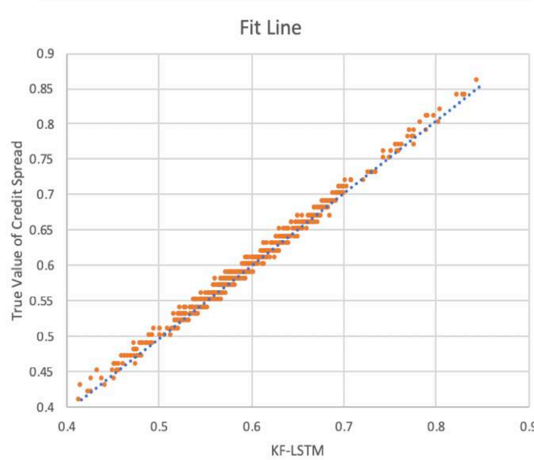
5.1 Performance Evaluation

To evaluate performance, we compare Reflection Model with CIR model. The following evaluation indicators are calculated from predictions on the test set. We amplify the results to show the details of performance, $i.e.$ arbitrary 200 days and 20 days of the test set. The figure on the lower right side is a fitting line comparing predicted values with true values.



	KF-LSTM	CIR
Maximal Deviation	0.0198	0.0794
Mean Square Error	0.0053%	0.0248%
Mean Absolute Error	0.0058	0.0115
R Square	99.108%	95.819%
Variance of Predictor	0.0019%	0.0116%
Mean Absolute Percentage Error	0.978	1.914



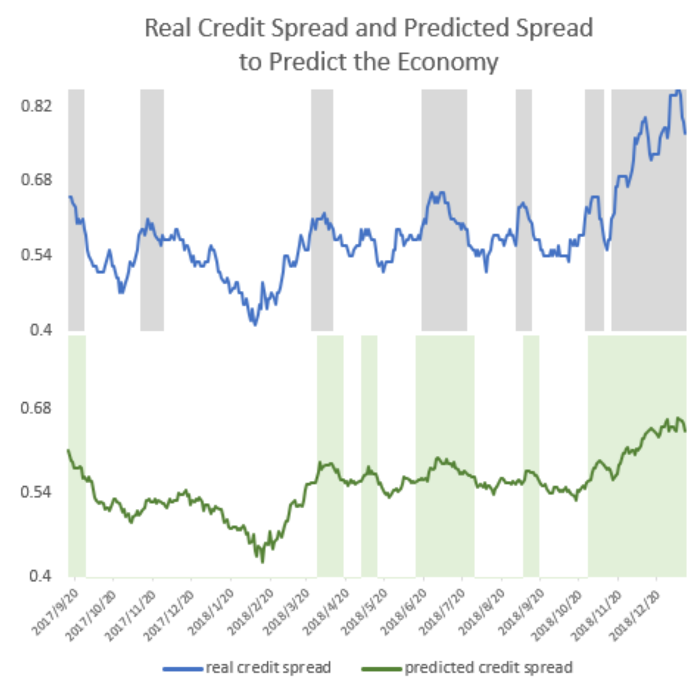


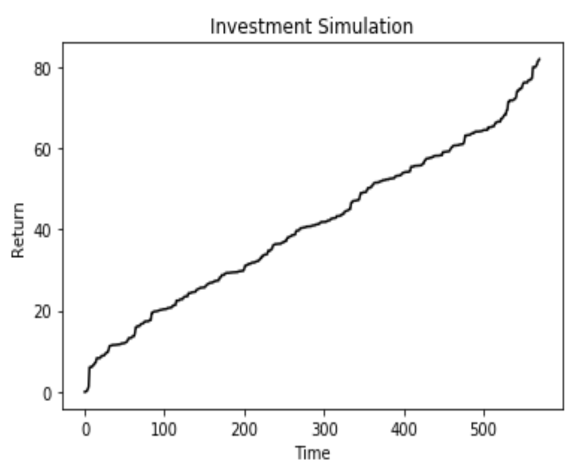
Based on the results, Reflection Model is obviously outperforming: **1. Accuracy:** the error is limited to a range of $(-0.0198\%, 0.0198\%)$. **2. Robustness:** the variance of the predictor is very small. **3. Practicability:** we can apply this model to investment and economic prediction. **4. No hysteresis:** the prediction shows no hysteresis

However, some shortcomings still exist 1. Lack of Explanation: LSTM is a type of neural network model and we do not have theoretical economic support for trained function. 2. Black Box: the generating process is inexplicable; the predictor will fluctuate at different training processes. 3. Complexity: the model is very complicated, it requires high computing capacity and large memory, and takes much training time.

5.2 Applications

One application is to simulate an investment. Assume that there is a financial derivative whose underlying asset is the credit spread of AAA grade bond index over 10 years treasury bond. The quoted price of this derivative is $(100 - credit\ spread)$. Then based on the next day prediction, we execute the following strategy: buy $(predicted\ value - today'value)/0.01$ of such derivatives if predicted value is higher than the current value and short sell the same amount otherwise. The daily return will depend on the true value of credit spread. The cumulative return is figured as the graph on the right side.





The return reaches as high as 8,000% in around two years without considering liquidity, trading fees, and other trading frictions.

Another application is to predict economic state. Here we use the Long-run Trend Prediction model, $i.e.$ without future information. The graph is shown on the upper left. Grey areas mean higher or increasing credit spread and deteriorating economy.

For example, on July 6th, 2017, American government publicized 46.5 billion dollars trading deficit. On Sep 21st, 2017, S&P downgraded China's sovereignty. On Dec 2nd, 2017, President announced Tax Reduction Bill. In April 2018, a trade war between China and American starts. On July 11th, 2018, President announced to increase tariffs on China's $200 billion in goods. On December 19th, 2018, the Fed announced that it will raise the federal funds rate target range by 25 basis points to 2.25% to 2.5%. On December 22nd, 2018, due to the failure of the US Congress to reach an agreement on the provisional appropriations bill, federal government agencies "stopped" from 0:00 pm EST.

Green areas are the predicted economic states. The prediction is 85% accurate.

Part 6 Conclusion

LSTM is a powerful model to analyze time series problem. We set it as the Basic Model and try to improve it by using EMD to smoothe the inputs and KF to amend the deviation. The two methods make LSTM more practical and accurate. This is the origin of our Reflection Model. Through our tests, Reflection Model outperforms the classical CIR model and this remarks a great breakthrough. Reflection Model can be illusrtrated as the formula:

$$y_t^* = f\left(\dot{F}_{t-1}, \dot{y}_{t-m}\colon \dot{y}_{t-1}\right) + d_t(y_{t-1}, \hat{y}_t) + e_t, \quad e_t \sim \mathcal{N}(0, \sigma^2)$$

where $f(\cdot)$ is trained by LSTM, $\dot{F}_t$ and $\dot{y}_t$ are features and credit spread with EMD, d_t is the adjustment from KF, $\hat{y} = f(F_{t-1}, y_{t-m}\colon y_{t-1})$ is an initial and rough estimator. The error term e_t must follow a normal distribution, which is required by KF model. y_t^* is an optimized predictor.

The procedure of EMD has no significant influence on the accuracy, but it strengthens the robustness of the prediction. Therefore, we still keep it and admit its significance.

The procedure of KF generates a satisfactory modification. A proper explanation is that we coordinate the information from what we observe and what we predict based on historical data and external features. If all information is orthongonal and accurate, it could give a better prediction. It is like coordinating the information from both fundamental and technical analysis to precisely estimate stock price. That is the basic idea for the Reflection Model.

In addition, Artificial Neural Network means to simulate human's brain to learn knowledge. But in reality, we rarely learn the goal very well. We make faults and even repeat them. A practical solution is to memorize lessons from past failure and try to perform as well as possible, $i.e.$ reflecting on ourselves. This is the origin of the name, "Reflection Model".

Reference

Cavallo E A, Valenzuela P. The Determinants of Corporate Risk in Emerging Markets: An Option-Adjusted Spread Analysis[J]. Social Science Electronic Publishing.

Clark E, Baccar S. Modelling credit spreads with time volatility, skewness, and kurtosis[J]. Annals of Operations Research, 2015.

Cinar Y G, Mirisaee H, Goswami P, et al. Position-based Content Attention for Time Series Forecasting with Sequence-to-sequence RNNs[J]. 2017.

Gilchrist S, Zakrajsek E. Credit Spreads and Business Cycle Fluctuations[J]. American Economic Review, 2012, 102(4):1692-1720.

Krause B, Kahembwe E, Murray I, et al. Dynamic Evaluation of Neural Sequence Models[J]. 2017.

Krishnan C N V, Ritchken P H, Thomson J B. Predicting credit spreads[J]. Journal of Financial Intermediation, 2010, 19(4):529-563.

Kuchaiev O, Ginsburg B. Factorization tricks for LSTM networks[J]. 2017.

Levent Güntay, Hackbarth D. Corporate bond credit spreads and forecast dispersion[J]. Journal of Banking & Finance, 2010, 34(10):0-2345.

Max Lambert, Andrew Engroff, Matt Dyer, Ben Byer. Empirical Mode Decomposition[OL]. https://www.clear.rice.edu/elec301/Projects02/empiricalMode/process.html, 2019.

NOZAWA, YOSHIO. What Drives the Cross-Section of Credit Spreads?: A Variance Decomposition Approach[J]. The Journal of Finance, 2017.

Son Y, Byun H, Lee J. Nonparametric machine learning models for predicting the credit default swaps: An empirical study[M]. Pergamon Press, Inc. 2016.

Thakur B P S, Kannadhasan M, Goyal V. Determinants of corporate credit spread: evidence from India[J]. DECISION, 2018.

Xu Y, Haan J D. The time-varying relationship between credit spreads and employment growth[J]. Applied Economics, 2018(3):1-15.