Machine learning-aided modeling of fixed income instruments

Daniel Martin, Barnabás Póczos
Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA 15213
{dlmartin, bapoczos}@cs.cmu.edu

Burton Hollifield
Tepper School of Business
Carnegie Mellon University
Pittsburgh, PA 15213
burtonh@andrew.cmu.edu

Abstract

The fixed income market is very important to the economy. Sovereign issues are influenced by central bank policy, and corporate issues are viewed relative to these sovereign issues. Compared to equities, bonds have lower liquidity and transparency; hence, there is less public data available. We demonstrate the applicability of machine learning models in forecasting interest rates for U.S. Treasuries of varying maturities, as well as clean prices of corporate issues.

1 The Fixed Income Market

Fixed income markets have strategic importance to the global economy. The global bond market is valued over $100 trillion, as compared to S&P's $64 trillion valuation of the global equity markets. [14] Although most bonds are available to individual investors, the majority are purchased by large institutional funds and insurance companies for the purpose of hedging risk.

A bond is issued with a defined term (lifespan) and coupon (interest) rate. A credit rating agency rates bonds based on the issuer's creditworthiness. As there is a greater variety of bonds than stocks, secondary trades of a bond do not take place in a market, but through direct trader-to-trader negotiations. This system reduces liquidity and transparency for trades, which increases risk for both the buyer and the seller. Having an improved ability to forecast prices would mitigate some pricing risk, and, given the size of institutional trades, avoid large losses.

Sovereign Bonds. Sovereign bonds are fixed income debt issued by a sovereign government. These bonds are usually considered the lowest risk available in their respective countries, and thus offer the lowest rates in the country. The U.S. government currently holds a top AAA rating from major credit agencies. The interest rates are influenced by the monetary policies of the country's central bank, which aims to avoid sudden changes.

Corporate Bonds. Corporate bonds are issued by companies, and are usually considered higher risk than sovereign bonds. The perceived additional risk of investing in a particular company leads to higher interest rates for corporate bonds as compared to sovereign bonds. Prior to 2002, U.S. corporate bond trade prices were not publicly available except through compulsory reporting of large trades by insurance companies. [6] As buyers had limited information, this caused bond prices to be relatively volatile. To combat this information void, the Securities and Exchanges Commission (SEC), a government body issued new reporting requirements. To comply with these new requrirements, the Financial Industry Regulatory Authority (FINRA), a trade group, set up a public database (TRACE) where licensed security dealers would be required to report trade information. It officially launched in 2006.

An expanded version of a paper presented at the NIPS 2018 Workshop on Challenges and Opportunities for AI in Financial Services: the Impact of Fairness, Explainability, Accuracy, and Privacy, Montréal, Canada.

Matrix Pricing. Many bonds trade infrequently. This complicates the task of determining a suitable purchase price. The most popular technique used in this scenario is called "matrix pricing" [45]. Matrix pricing involves adding an arbitrary spread to a roughly-comparable bond: a more-actively traded bond from the same company, a similar bond from a different company, or a U.S. Treasury bond [44]. Dealers will usually hold this quote barring large shifts in the price of the comparable [23]. The arbitrary nature of this technique can lead to discrepancies in valuation. Using machine learning on existing data, our goal is to forecast the prices of such bonds without finding comparables.

Summary. The rest of our paper will be organized as follows. We will begin by examining prior forecasting work in Section 2. In Section 3 we introduce the machine learning models used in our experiments. We will continue with our bond experiments in Section 4. Finally, we discuss these results in Section 5, and delve into further avenues of exploration.

2 Related Work

Predicting the next trade price of a corporate bond has not been widely studied. Therefore, in addition to the papers that directly address this problem, we will also examine papers that cover bond credit rating as well as papers that cover stock prediction.

2.1 Bond-based Studies

Price Prediction using Machine Learning. An early paper [10] to use machine learning for bond price prediction used an artificial neural network (ANN) to predict the price of a 50-year U.S. Treasury bond. The author used 4 input variables: transaction settlement date, coupon rate, yield, and maturity date. The output of the model would be the bond's quoted price. Five hundred prices were calculated with the input variables as training, then the model would output five hundred prices using the testing data. The author used 1 hidden layer, and varied its size to how the error decreased with additional nodes.

Another study [8] attempted to predict the monthly yield of 10-year Treasury bonds using macroeconomic indicators such as the Consumer Price Index (CPI). Predicting the yield would allow calculation of the price. As the indicators used for input are only issued on a monthly basis, the input dataset was too sparse, and the tested models did not outperform ordinary regression.

A later study [16] did a more comprehensive analysis. This study used a dataset of 762,678 U.S. corporate bonds that contained 61 attributes for each bond. Some of these attributes described the bond itself, such as its coupon rate or years to maturity. Other attributes were extracted from the bond's previous 10 transactions, including the time for the trade to be reported, number of units sold, the real price, as well as a price from a regression curve. The authors describe various classification methods, ranging from linear regression models and ARIMA to more complex models such as random forests and feed-forward ANNs. The best models were ANNs. While this study covers a variety of models, the dataset gave a very limited history of the bond's price. Given a longer time period, some of these models may perform differently.

An additional study [22] took a different approach to predict prices of Japanese corporate bonds. Instead of previous prices, the authors used fundamental and macroeconomic factors as inputs: the current net earnings, the management's forecasted earnings for the next period, an index of business conditions, and the bond's credit rating. These inputs were fed to a support vector machine (SVM) to predict whether a bond would have a normal or abnormal negative return. Six years of data were used for training, and two years were used for testing Using a Gaussian SVM, the authors were able to identify 65 percent of the negative abnormal returns and 62 percent of the normal returns.

Besides these studies, another study [12] used Monte Carlo methods to price a bond with more exotic features that are difficult to model analytically. These features include early-redemption clauses for the issuer and conversion clauses that allow bondholders to exchange for stock. There has also been work [25] in predicting prices for futures contracts on Treasury bonds.

Credit Rating and Default Prediction. Credit rating prediction for bonds has a longer history. An early study [26] used stepwise regression to extract relevant financial variables from a given dataset, then compared results from neural networks and other techniques. The best results came from a partitioned neural network.

A later study [3] used various ensembles of ANNs and SVMs to predict credit rating from company fundamentals.

Finally, a third study [19] used the Random Forest (RF) model to predict defaults of junk bonds. The inputs included yield, credit rating, time to maturity, returns, and volatilities. The Random Forest was shown to be more accurate and consistent than the classical metric used to predict defaults.

2.2 Stock-based Studies

Many different techniques have been used to predict stock pricing. We will attempt to cover some major model types here.

Generalized Linear Regression and Filtering. The most basic models in this category are linear regression variants. These models include nonparametric regression or the generalized autoregressive conditional heteroskedasticity (GARCH) model [33]. They can also be combined with filters, like the Kalman filter [39]. By reducing the effects of noise, these filters allow the models to converge more quickly. However, these models are limited in forecast length and predictive power.

A further refinement of this model is quantile regression. Several studies [34][38][5][9] have attempted to see whether predicting based on return quantiles is more effective. While they are better than ordinary regression, they still require pre-selection into appropriate quantiles to ensure accurate prediction.

Classic Models. Nonlinear models have been used with greater predictive success. Earlier work [48] showed the benefits of time-varying margin for Support Vector Regression (SVR) as compared to a fixed margin. More recent work [1] has confirmed that SVR and feed-forward ANNs provide good, consistent predictive performance. Gaussian Process (GP) regression has compared favorably to ARIMA [42].

Tree-based classifiers have also shown good performance. One study [46] shows boosted trees having higher accuracy than ANNs and SVMs for hybrid news and price inputs. Random Forests have also been used [24] to predict stock price direction from a set of common stock price indicators.

Two studies [36][37] have used a hierarchical semi-supervised learning (SSL) model for price prediction. Their models have two layers: the first layer contains prices of important commodities and major stock indices, while the second layer contains prices of individual stocks.

Some studies have used the Kohonen self-organizing map (SOM) [13]. The SOM is used to learn patterns in the data, which are then used in forecasting. While its learning rate is slower than an ANN, it is not a black box and also provides decent accuracy [2].

Some studies [21] [32] have also used the Relevance Vector Machine (RVM), finding it to have similar performance to an SVM, but with a simpler model and shorter training time.

Hidden Markov Models have been used [18] for stock prediction. Different changes in stock price were represented as state transitions. The authors applied it to three stocks, and compared their model to ARIMA and ANN models from a previous paper. Their model outperformed on two out of the three stocks.

Deep Learning. Deep learning techniques have also been used for stock prediction. However, as they all utilize very large datasets, these techniques may not translate well to the problem of bond pricing.

Some of these models operate only on price history. Entangled-RNN [49] uses both forward and backward training of a recurrent ANN for price prediction. Another study [20] used an LSTM network for daily price prediction. Further studies use the recurrent Boltzmann machine (RBM) [29], hybrid neural nets [30], and undecimated fully-convolutional neural nets [15].

Other models incorporate external data for prediction. One study[51] used 8-K report filings to report uptrends and downtrends. Recurrent Neural Networks-Restricted Boltzmann Machines (RNNRBM) were best at predicting downward trends, while Stacked Denoising Autoencoders (SDA) could predict uptrends. Another study [4] attempts to use forecasts of company fundamentals to train RNNs for predicting future prices.

Clustering and Other Preprocessing. For better predictive accuracy, independent component analysis (ICA) has been used to extract unique features. One study [35] used ICA on the principal components that explained 95 percent of the return, calling the returned components "news" predictors. Another study [17] used ICA to choose financial variables for training an SVM to predict a stock's price.

Many clustering methods have been used on stock data. PCA [50] and Isomap [31] have been used for clustering on prices. Tucker decomposition [28] has been used on financial news data to use in predicting prices. Hierarchical clustering [27] has been used in a hybrid approach that took both textual (earnings releases) and quantitative factors (company fundamentals) for prediction.

Finally, some studies have tried to address the issue of sparsity in input data. Conditional sparse coding [47] was used to reconstruct quotes for missing stocks at varying time periods. Another approach [43] used dictionary learning to form a representation of sparse data series that could be easily processed and analyzed.

3 Forecasting Models

Our problem is to be able to forecast a future series of interest rates (for Treasuries) and a future series of trade prices (for corporate bonds). As more work has been done with stock forecasting, we will choose two models that have been used successfully, and adapt them for our experiments.

Random Forests (RF). The random forest [7][46] model is considered an ensemble-based learning method. The training process entails constructing multiple decision trees (hence, the "forest"). The output is the mean output of the decision trees, so that the overall variance is lower than that of any individual estimator. This is a concern, since deep decision trees tend to have high variance.

When constructing the forest, we consider each tree as a *bootstrap* estimator: we select a random subsample with replacement from the data set and then train the tree. Furthermore, when creating a decision node in the tree, we find the best split given a random subset of the input features. The final averaging is as follows, where B is the number of trees:

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B} f_b(x')$$

In our tests, we set the number of decision trees to 100.

Support Vector Regression (SVR). By contrast, the support vector regression [11][48][1] model seeks to fit a curve to the data while minimizing the distance from the points to be within a certain margin. The SVR model will find a higher-dimensional space where the data is linear, and return the line of best fit in that space. For our SVR model, we set our slack penalty $C = 0.8$ and our margin size $\varepsilon = 0.2$. We used the radial basis function (RBF) as our kernel with $\sigma = 1$:

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{||\mathbf{x} - \mathbf{x}'||^2}{2\sigma^2}\right)$$

The SVR model is an optimization problem. The primal form is as follows:

$$\begin{aligned} &\min_{w,b,\zeta,\zeta^*} \frac{1}{2}w^T w + C\sum_{i=1}^{n}(\zeta_i + \zeta_i^*) \\ \text{subject to } & y_i - w^T\phi(x_i) - b \leq \varepsilon + \zeta_i, \\ & w^T\phi(x_i) + b - y_i \leq \varepsilon + \zeta_i^*, \\ & \zeta_i, \zeta_i^* \geq 0, i = 1, ..., n \end{aligned}$$

Its dual form:

$$\begin{aligned} &\min_{\alpha,\alpha^*} \frac{1}{2}(\alpha - \alpha^*)^T K(\alpha - \alpha^*) + \varepsilon\mathbf{1}^T(\alpha + \alpha^*) - y^T(\alpha - \alpha^*) \\ \text{subject to } & \mathbf{1}^T(\alpha - \alpha^*) = 0 \\ & 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, ..., n \end{aligned}$$

(a) Window 1

(b) Window 6

Figure 1: Equal-length data for RF on 3-year Treasury

Solving this problem gives us the following regression line:

$$\sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x) + \rho$$

Gaussian Process (GP). The gaussian process [40][42] regression method involves conditioning on the data. A gaussian process is a set of random variables such that any subset is a multivariate normal variable centered at the origin with covariance defined according to a given function. This gaussian process is used as the prior for finding the data's posterior. This posterior is then used for prediction.

4 Experiments

We use a lag method of forecasting: given the previous *n* prices, can we predict the next price? For our experiments, we used a lag of 5 points.

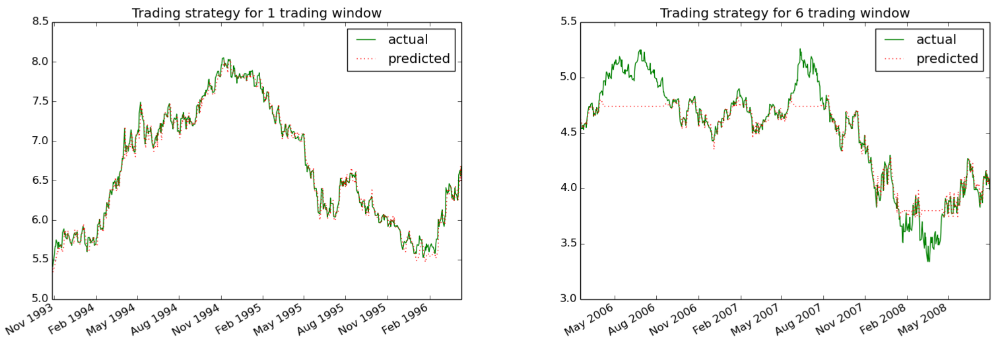
4.1 Treasuries

Data. We considered U.S. Treasuries of 3 year, 10 year, and 30 year terms. The daily nominal interest rates were obtained from CRSP for the period of October 27, 1993 to June 5, 2018.

Training Data Ratio. We first conducted preliminary tests using smaller proportions of training data (70:10, 40:10, and so on). Here, we include a portion of the results from training our models on Treasury bonds using a 10:10 train-test data ratio. In figures (1) and (2), we see that the RF model can fit well, with some missing patches. However, figure (3b) shows a case where the training data was sufficiently different to be misleading: instead of following the line, the model only senses level changes. Our SVR results in figures (4), (5), and (6) are similar, with the distortions being more pronounced. The GP model seems most vulnerable to changes in price, to the point of producing sharp changes of its own as shown in figures (7), (8), and (9). While some of these trials had good results, the trained models were much more vulnerable to regime changes in price. If the the testing or training intervals contained extreme rises and falls, the model would give distorted predictions. To minimize these disruptions, we decided to use 90:10 by default.

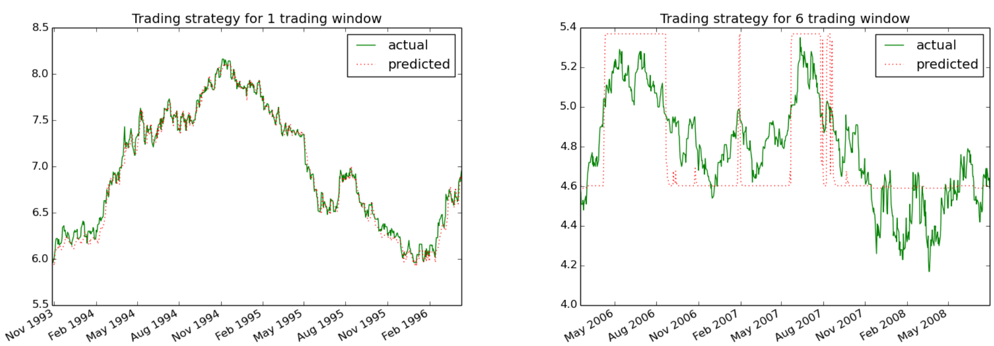
Results. In Figures (10), (11), and (12), we see the results of the RF and SVR models. We use the R^2 coefficient and the mean squared error (MSE) to measure our accuracy. Tables 1 and 2 show these results on our test data.

Figure (13b) shows the results of the GP model, and table 3 lists its metrics. While these results are better than those of the 10:10 data model, the outliers of the prediction result are much higher than the data range.



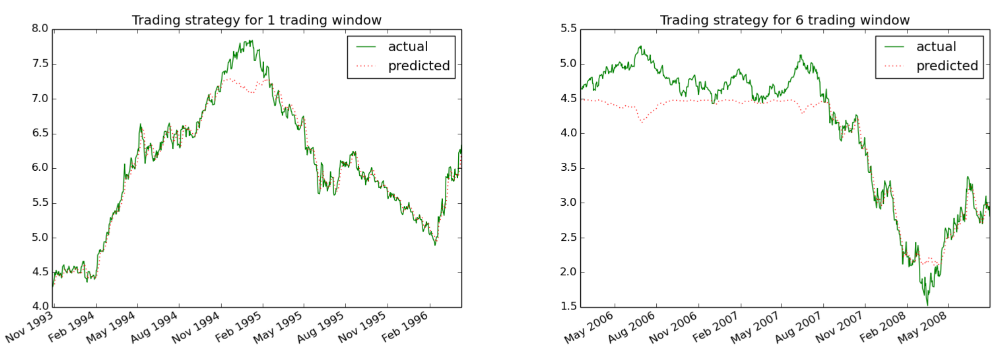
(a) Window 1 (b) Window 6

Figure 2: Equal-length data for RF on 10-year Treasury



(a) Window 1 (b) Window 6

Figure 3: Equal-length data for RF on 30-year Treasury



(a) Window 1 (b) Window 6

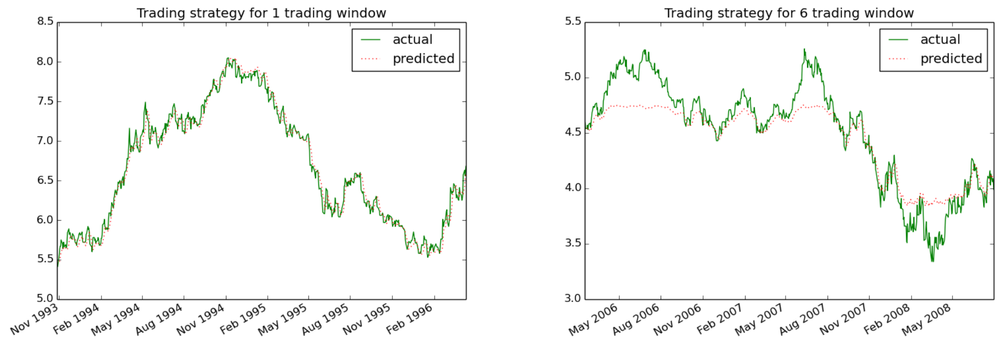
Figure 4: Equal-length data for SVR on 3-year Treasury

Table 1: RF prediction metrics

Term	Test R^2	Test MSE
3	0.9943	0.0016
10	0.9876	0.0021
30	0.9618	0.0026

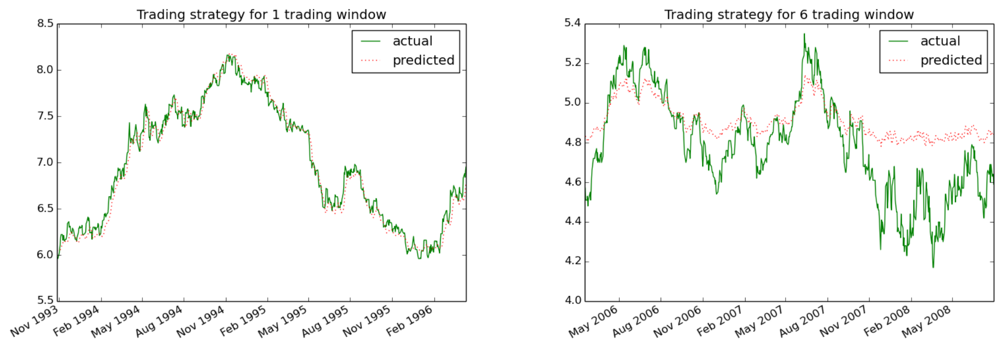
Table 2: SVR prediction metrics

Term	Test R^2	Test MSE
3	0.9907	0.0026
10	0.9754	0.0042
30	0.796	0.0139



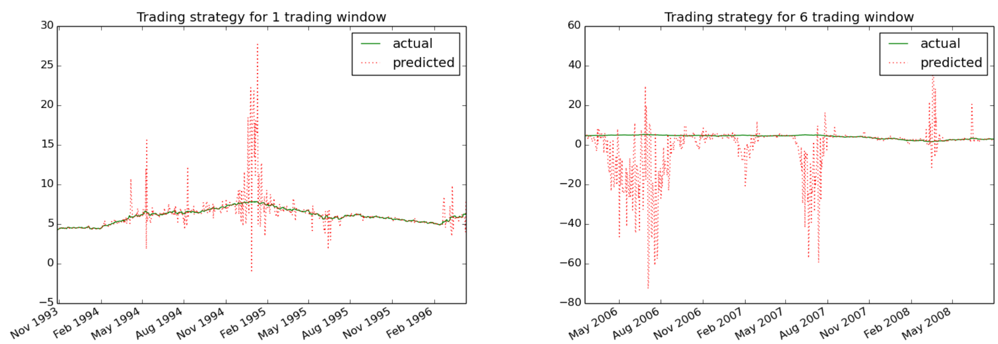
(a) Window 1 (b) Window 6

Figure 5: Equal-length data for SVR on 10-year Treasury



(a) Window 1 (b) Window 6

Figure 6: Equal-length data for SVR on 30-year Treasury

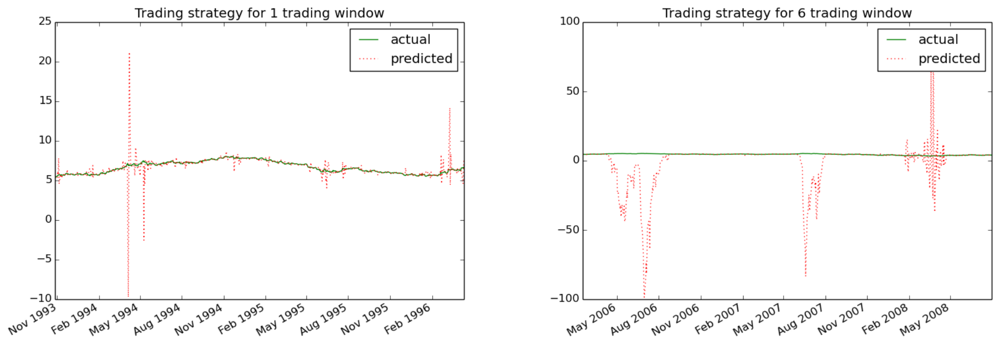


(a) Window 1 (b) Window 6

Figure 7: Equal-length data for GP on 3-year Treasury

Table 3: GP prediction metrics

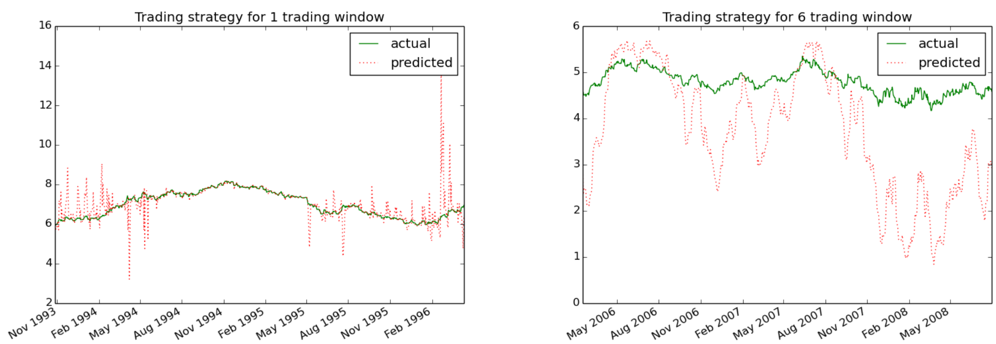
Term	Test R^2	Test MSE
3	0.9939	0.0017
10	0.8656	0.0226
30	-16.8108	1.2146



(a) Window 1

(b) Window 6

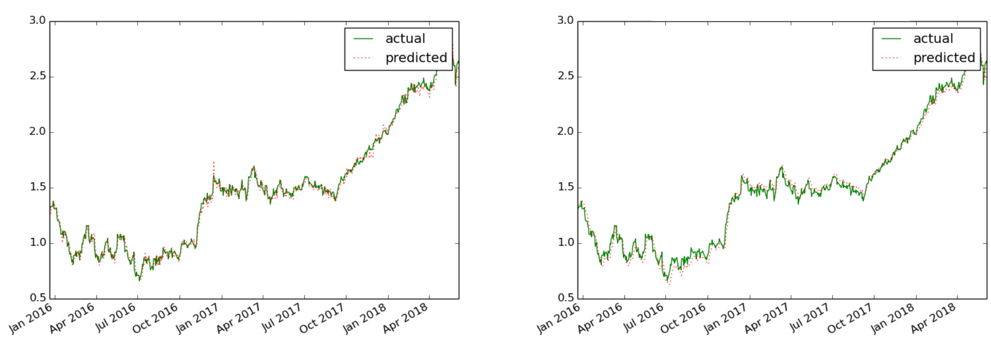
Figure 8: Equal-length data for GP on 10-year Treasury



(a) Window 1

(b) Window 6

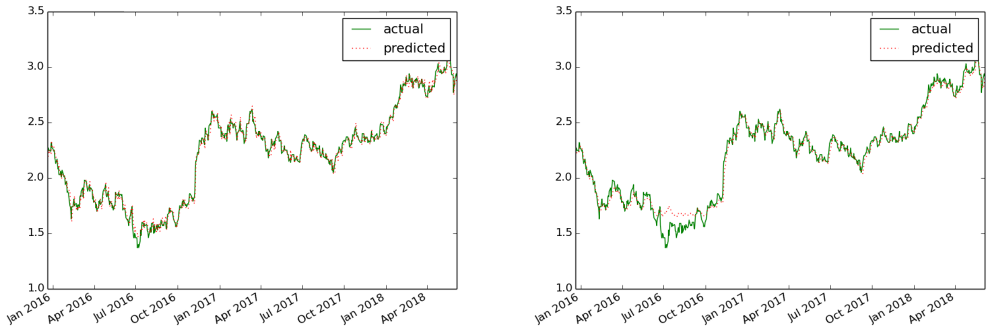
Figure 9: Equal-length data for GP on 30-year Treasury



(a) RF

(b) SVR

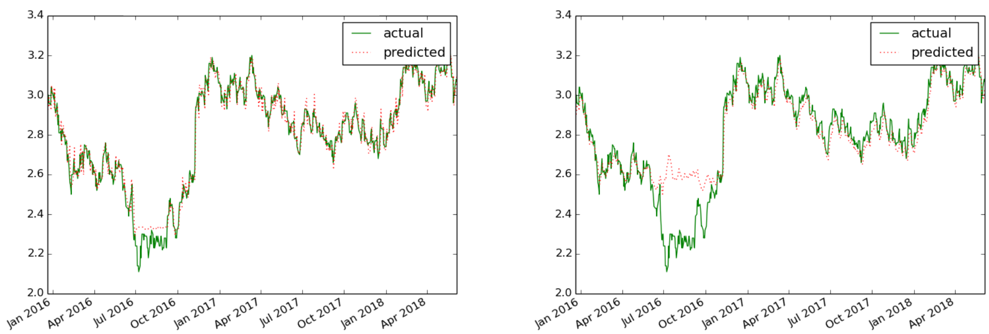
Figure 10: Predictive performance for a 3-year Treasury bond



(a) RF

(b) SVR

Figure 11: Predictive performance for a 10-year Treasury bond



(a) RF

(b) SVR

Figure 12: Predictive performance for a 30-year Treasury bond

4.2 Corporate Bonds

We considered three bonds: a 30-year Exxon bond issued in 1991, a 30-year Johnson&Johnson bond issued in 1993, and a 15-year Alcoa bond issued in 2007. The first two bonds are rated Aaa by Moody's. The Alcoa bond is rated Ba2 (below investment grade) by Moody's. Figures (14), (15), and (16) show the prediction results.

4.3 Winsorizing

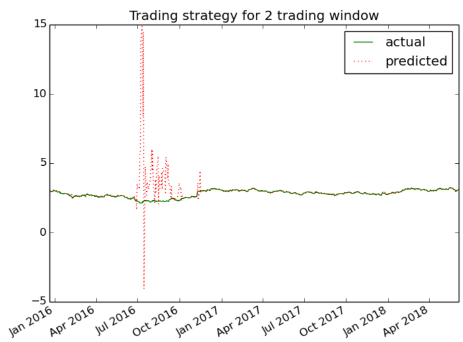
As the corporate bond data consists of manual reports from dealers, there is the possibility of data input errors. There also may be individual trades that diverge from the overall trend. In this situation, winsorizing the data should minimize disruption caused by the degree of divergence, simplifying the task of fitting the trend. This process entails finding outliers, and replacing them with a chosen high or low margin point.

For our margins, we calculated the moving median standard deviation. First, we obtained the moving median using a median filter with a kernel size of 5 points. We then found the absolute deviation of each of the points from the median, and standardized it to the 75th percentile. We winsorized the data to lie within $\pm$ 2 deviations. Subsequently, we trained and tested our models as before.

In tables 4 and 5, we see how both our metrics are greatly improved when using the winsorized data. We confirm this using Figures (17), (18), and (19). The winsorized data preserves most of the peaks and troughs of the original line, while scaling down the more extreme points.

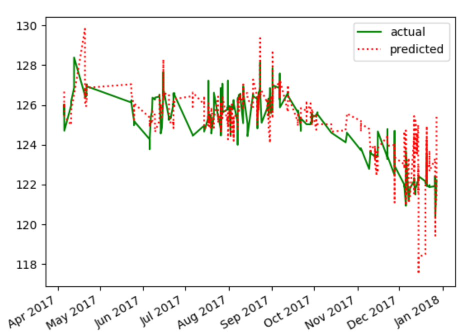
(a) 3-year Treasury

(b) 10-year Treasury

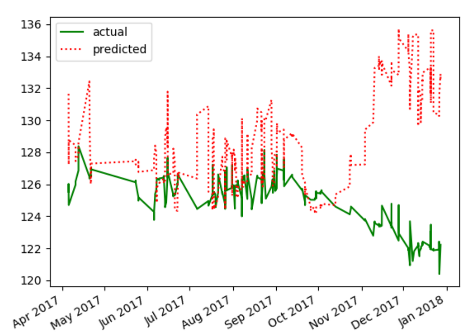


(c) 30-year Treasury

Figure 13: GP predictive performance



(a) RF

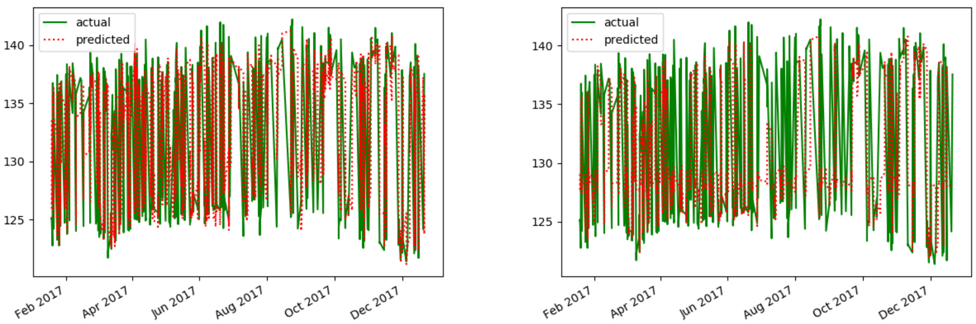


(b) SVR

Figure 14: Predictive performance for a 30-year Exxon bond issued in 1991

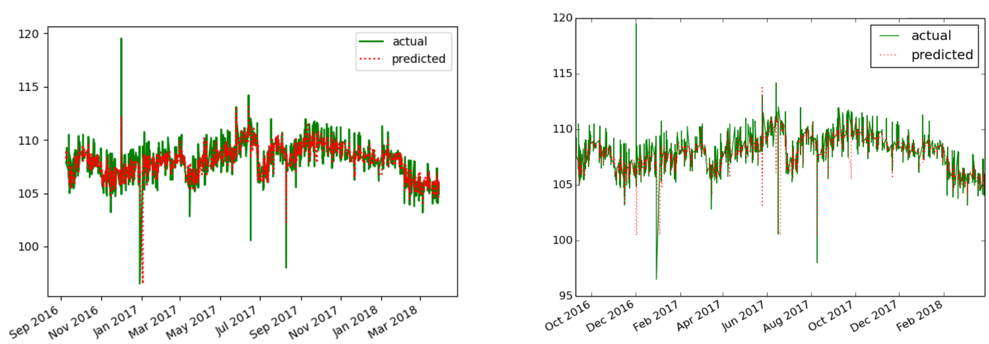
Table 4: RF corporate metrics

Bond	Exxon		J&J		Alcoa	
Winsorized?	No	Yes	No	Yes	No	Yes
Test R^2	0.446	0.779	0.32	0.594	0.649	0.942
Test MSE	1.815	0.711	28.902	15.895	0.96	0.128
Test MAPE	0.794	0.504	2.707	1.567	0.568	0.194



(a) RF (b) SVR

Figure 15: Predictive performance for a 30-year J&J bond issued in 1993

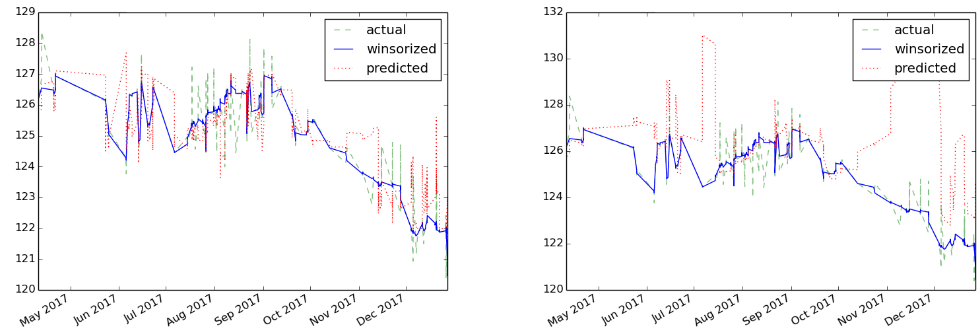


(a) RF (b) SVR

Figure 16: Predictive performance for a 15-year Alcoa bond issued in 2007

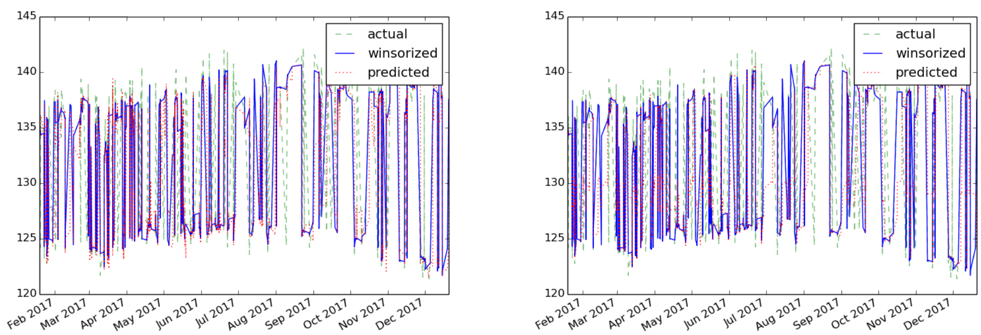
Table 5: SVR corporate metrics

Bond	Exxon		J&J		Alcoa	
Winsorized?	No	Yes	No	Yes	No	Yes
Test R^2	-10.682	-1.289	0.083	0.347	0.379	0.727
Test MSE	38.883	7.373	38.953	25.587	1.7	0.602
Test MAPE	3.687	1.373	3.373	2.44	0.601	0.215



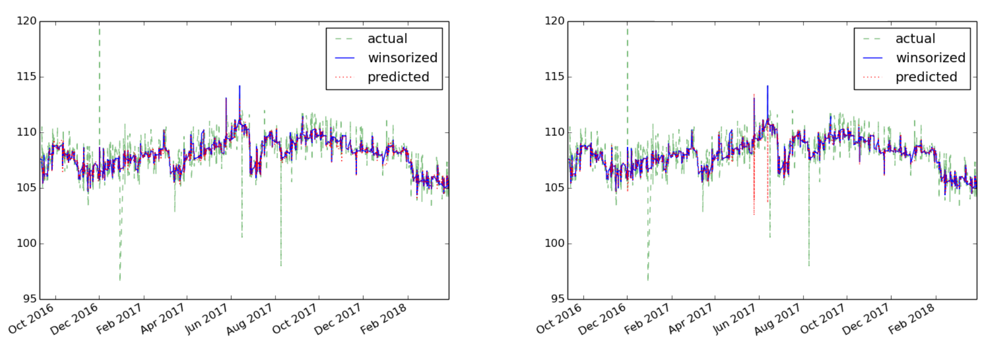
(a) RF (b) SVR

Figure 17: Winsorized performance for a 30-year Exxon bond issued in 1991



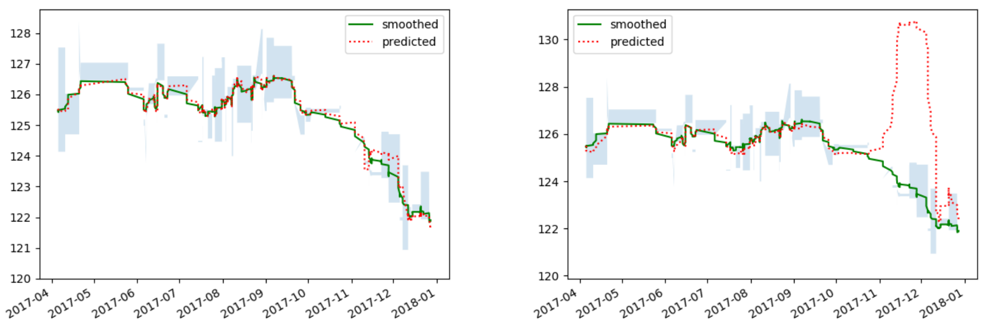
(a) RF (b) SVR

Figure 18: Winsorized performance for a 30-year J&J bond issued in 1993



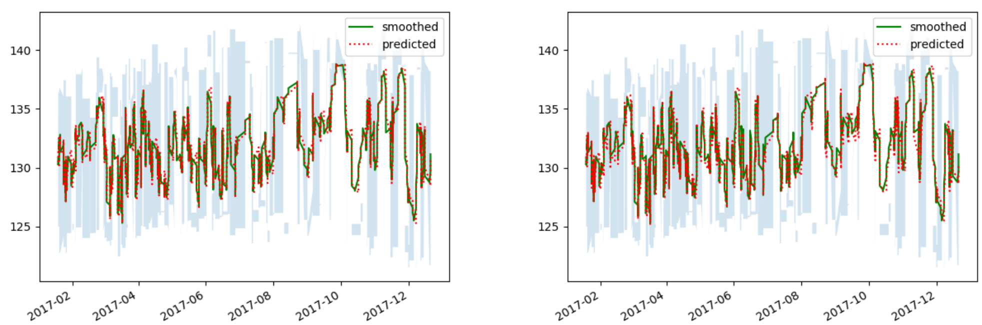
(a) RF (b) SVR

Figure 19: Winsorized performance for a 15-year Alcoa bond issued in 2007



(a) RF (b) SVR

Figure 20: Smoothed performance for a 30-year Exxon bond issued in 1991



(a) RF (b) SVR

Figure 21: Smoothed performance for a 30-year J&J bond issued in 1993

4.4 Smoothing

Another method of handling noisy data is exponential smoothing. This method is used to bring out the common trends in data; for instance, the Holt-Winters model for determining seasonality uses three rounds of smoothing. As our data is relatively noisy, we can apply a single smoothing round before training to demonstrate that we can predict the overall trend.

The procedure is essentially a stepwise weighted average. Each point is replaced with a weighted average of itself and the previous point. The weight of the present point is called the *smoothing factor*, which we set to 0.1. Due to this averaging, the resulting line is smoother than that of the winsorized model. Hence, it will not capture all the inflections of the data. This can be seen in the metrics: while the MSE and MAPE of the smoothed models is lower, the R^2 value is also lower.

Table 6: Smoothed RF metrics

Bond	Exxon	J&J	Alcoa
Test R^2	0.991	0.956	0.994
Test MSE	0.025	0.34	0.011
Test MAPE	0.088	0.353	0.062

Table 7: Smoothed SVR metrics

Bond	Exxon	J&J	Alcoa
Test R^2	-0.42	0.952	0.981
Test MSE	3.923	0.36	0.035
Test MAPE	0.716	0.34	0.087

(a) RF (b) SVR

Figure 22: Smoothed performance for a 15-year Alcoa bond issued in 2007

5 Discussion

Model Efficacy. In our tests, the RF model performs distinctly better than the SVR. The RF achieved a better fit to the data when it was close and had fewer divergent areas. Furthermore, we note that training models with a 5 point lag achieves good results in many cases, providing a credible data-agnostic alternative to matrix pricing.

The GP model would probably fare better in shorter forecasting intervals than we used.

Treasuries vs. Corporates. Treasury bond rates were not very divergent, and hence easier to fit for both models. This can be attributed to central bank policies to avoid sudden large shifts in rates, resulting in a sort of data filter.

On the other hand, the individual corporate transactions had a higher degree of divergence, as there are fewer filtering influences. Taking this variation into account, we note that the quantity of datapoints is key for maintaining accuracy. We had about 5x as many data points for Alcoa as for Exxon, so the models for Alcoa were more accurate. Alcoa is a junk-rated bond and has a shorter lifespan than the other two bonds, so holders have less of an incentive to hold.

6 Further Directions

There are several further directions in which this work can be expanded upon.

Machine Learning. While we covered the RF, SVR, and GP models, there are other forecasting models than can be used for this problem of forecasting. Long short-term memory (LSTM) [20], a type of deep learning model, can be used when larger datasets are available. Variations of these models may also be helpful. While GP regression was sensitive to outliers, Student's-t regression [41] has been shown to be more robust. While random forests use averaging over their regressor ensemble, it is possible to use boosted forests instead. Boosting adds regressors one-by-one, looking to decrease the overall bias of the model with each new regressor.

Besides different models, we can do further experiments with hyperparameter optimization, so one can arrive at a model better adjusted to the particular dataset. For RF, this would include adjusting the number of trees, the splitting thresholds, and the tree lenghts. For SVR, this would include the margin size, the slack penalty, and the kernel function. For GP, this would be the choice of kernel function.

It is also possible to use multiple models in a cascading fashion to reduce the overall forecasting error.

Data Preprocessing. Further preprocessing of the corporate bond data can be useful in prediction. As in other instruments, there may be seasonality and other recurring patterns. These patterns can be extracted using smoothing or other procedures. We can also alter the lag length of input data in case longer-range patterns are needed. Furthermore, additional provided trade information can be used, such as the size of the trade.

Effects of the Market. Our model can be further refined by incorporating the effects of factors external to the issuer. Many stock models take into account the effects of company announcements and investor sentiment when making price predictions, as well as company fundamentals [28]. Bond prices may also be affected by these and other factors. For instance, we can examine the effects of stock price changes on bond price, as well as how changes within the overall market sector and credit class of the issuer are reflected.

Asset Variations. The models in this paper could be extended to measure the predictive quality of bond covenant provisions. For example, what effects do issuer call options have on the market trade price? What about conversion clauses, or insurance? How much the availability of options and derivatives affect the price of the base asset?

Sovereign Issuer Relationships. Besides the U.S., many other countries release nominal rates for their outstanding sovereign debt. This model could be applied to forecast their rates as well; one can even take into account currency and trade relationships for these models. Such a model can be useful when forecasting in foreign corporate bond markets. Many countries have secondary corporate bond markets, but their secondary sovereign market is much more active.

7 Conclusion

In this paper, we have demonstrated two types of fixed-income forecasting problems that can be handled using machine learning models. Treasury interest rates can be easily forecasted from past data given a long enough history. Corporate bond transaction prices, although noiser, can also be predicted to a degree, depending on the amount and quality of data available. From this work, we can test more complex models to better explain corporate bond prices. Overally, having an effective forecasting tool can help increase market liquidity at accurate prices, aiding investors and issuers alike.

References

[1] Torkil Aamodt. Predicting stock markets with neural networks: A comparative study. Master's thesis, University of Oslo, 2015.

[2] Mark O Afolabi and Olatoyosi Olude. Predicting stock prices using a hybrid Kohonen self organizing map (SOM). In *40th Annual Hawaii International Conference on System Sciences*, pages 48–48. IEEE, 2007.

[3] Hyunchal Ahn and Kyoung-Jae Kim. Corporate credit rating using multiclass classification models with order information. *International Journal of Economics and Management Engineering*, 5(12):1783–1788, 2011.

[4] John Alberg and Zachary C Lipton. Improving factor-based quantitative investing by forecasting company fundamentals. *arXiv preprint arXiv:1711.04837*, 2017.

[5] Michelle L. Barnes and Anthony W. Hughes. A quantile regression analysis of the cross section of stock market returns. Working Papers 02-2, Federal Reserve Bank of Boston, 2002.

[6] Hendrik Bessembinder and William Maxwell. Transparency and the corporate bond market. *Journal of Economic Perspectives*, 22(2):217–234, 2008.

[7] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[8] Marco Castellani and Emanuel Augusto dos Santos. Forecasting long-term government bond yields: an application of statistical and ai models. Technical report, ISEG–Departamento de Economia, 2006.

[9] Chia-Chang Chuang, Chung-Ming Kuan, and Hsin-Yi Lin. Causality in quantiles and dynamic stock return–volume relations. *Journal of Banking & Finance*, 33:1351–1360, 2009.

[10] Andrew Colin. Fixed income attribution with minimum raw material. *Journal of Performance Measurement*, 11(2):50, 2006.

[11] Harris Drucker, Christopher J. C. Burges, Linda Kaufman, Alex J. Smola, and Vladimir Vapnik. Support vector regression machines. In M. C. Mozer, M. I. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems 9*, pages 155–161. MIT Press, 1997.

[12] Bella Dubrov. Monte carlo simulation with machine learning for pricing american options and convertible bonds. *SSRN*, 2015.

[13] Lars Edler. *Analysing Economic Data with Self-Organizing Maps: A Geometric Neural Network Approach*. PhD thesis, Christian-Albrechts-University of Kiel, 2007.

[14] Federated Investors, Inc. Infographic: The bond market. Online, June 2017.

[15] John Cristian Borges Gamboa. Deep learning for time-series analysis. *arXiv preprint arXiv:1701.01887*, 2017.

[16] Swetava Ganguli and Jared Dunnmon. Machine learning for better models for predicting bond prices. *arXiv preprint arXiv:1705.01142*, 2017.

[17] Hakob Grigoryan. A Stock Market Prediction Method Based on Support Vector Machines (SVM) and Independent Component Analysis (ICA). *Database Systems Journal*, 7(1):12–21, 2016.

[18] Aditya Gupta and Bhuwan Dhingra. Stock market prediction using hidden markov models. In *2012 Students Conference on Engineering and Systems (SCES)*, pages 1–4. IEEE, 2012.

[19] Alan Hanna. Corporate bond default risk: Seeing the wood through the trees? Technical report, Queen's University Belfast, 2016.

[20] Magnus Hansson. On stock return prediction with LSTM networks. Master's thesis, Lund University, 2017.

[21] Altaf Hossain and Mohammed Nasser. Recurrent support and relevance vector machines based model with application to forecasting volatility of financial returns. *Journal of Intelligent Learning Systems and Applications*, 3:230–241, 2011.

[22] Hiroaki Jotaki, Yasuo Yamashita, and Hiroshi Takahashi. Predicting corporate bond prices in japan using a support vector machine. In *eKNOW 2017 : The Ninth International Conference on Information, Process, and Knowledge Management*, pages 13–19, 2017.

[23] Yusho Kagraoka. Corporate bond liquidity and matrix pricing. *Physica A*, 355:158–164, 2005.

[24] Luckyson Khaidem, Snehanshu Saha, and Sudeepa Roy Dey. Predicting the direction of stock market prices using random forest. *arXiv preprint arXiv:1605.00003*, 2016.

[25] Ajitabh Kishore. Machine learning and algorithmic trading: In fixed income markets. *SSRN*, 2013.

[26] Young S. Kwon and Sungsic Lim. Performance comparison of machine learning approaches to bond rating. In *IEEE International Conference on Systems, Man, and Cybernetics*, 1998.

[27] Anthony J.T. Lee, Ming-Chih Lin, Rung-Tai Kao, and Kuo-Tay Chen. An effective clustering approach to stock market prediction. In *PACIS 2010 Proceedings*, 2010.

[28] Qing Li, LiLing Jiang, Ping Li, and Hsinchun Chen. Tensor-based learning for predicting stock movements. In *AAAI*, pages 1784–1790, 2015.

[29] Qiubin Liang, Wenge Rong, Jiayi Zhang, Jingshuang Liu, and Zhang Xiong. Restricted boltzmann machine based stock market trend prediction. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 1380–1387. IEEE, 2017.

[30] Tao Lin, Tian Guo, and Karl Aberer. Hybrid neural networks for learning the trend in time series. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, 2017.

[31] Ruiling Liu, Hengjin Cai, and Cheng Luo. Clustering Analysis of Stocks of CSI 300 Index Based on Manifold Learning. *Journal of Intelligent Learning Systems and Applications*, 4:120–126, 2012.

[32] Ivana P. Markovic, Milos B. Stojanovic, and Milos M. Bozic. Stock market trend prediction using a sparse bayesian framework. In *12th Symposium on Neural Network Applications in Electrical Engineering (NEUREL)*, 2014.

[33] Thomas Mazzoni. Are short term stock asset returns predictable? an extended empirical analysis. Technical report, University of Hagen, 2010.

[34] Loukia Meligkotsidou, Ekaterini Panopoulou, Ioannis D. Vrontos, and Spyridon D. Vrontos. A quantile regression approach to equity premium prediction. *Journal of Forecasting*, 33:558–576, 2014.

[35] PY Mok, KP Lam, and HS Ng. An ICA design of intraday stock prediction models with automatic variable selection. In *Proceedings of 2004 IEEE International Joint Conference on Neural Networks*, volume 3, pages 2135–2140. IEEE, 2004.

[36] Kanghee Park and Hyunjung Shin. Forecasting stock price movement with semi-supervised learning. In *Proceedings of the International Conference on Data Mining (DMIN)*, 2012.

[37] Kanghee Park and Hyunjung Shin. Stock price prediction based on hierarchical structure of financial networks. In *International Conference on Neural Information Processing*, pages 456–464. Springer, 2013.

[38] Thomas Q. Pedersen. Predictable return distributions. *Journal of Forecasting*, 34:114–132, 2015.

[39] James Martin Rankin. *Kalman filtering approach to market price forecasting*. PhD thesis, Iowa State University, 1986.

[40] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2005.

[41] Arno Solin and Simo Särkkä. State space methods for efficient inference in Student-t process regression. In *Artificial Intelligence and Statistics*, pages 885–893, 2015.

[42] Bagas Abisena Swastanto. Gaussian process regression for long-term time series forecasting. Master's thesis, Delft University of Technology, 2016.

[43] George Tzagkarakis, Juliana Caicedo-Llano, and Thomas Dionysopoulos. Sparse modeling of volatile financial time series via low-dimensional patterns over learned dictionaries. *Algorithmic Finance*, 4(3-4):139–158, 2015.

[44] Arthur Warga and Ivo Welch. Bondholder losses in leveraged buyouts. *Review of Financial Studies*, 6(4):959–82, 1993.

[45] Arthur D. Warga. Corporate bond price discrepancies in the dealer and exchange markets. *The Journal of Fixed Income*, 1(3):7–16, 1991.

[46] Bin Weng. *Application of machine learning techniques for stock market prediction*. PhD thesis, Auburn University, 2017.

[47] Min Xu and John Lafferty. Conditional sparse coding and multiple regression for grouped data. Technical report, Carnegie Mellon University, 2012.

[48] Haiqin Yang. Margin variations in support vector regression for stock market prediction. Master's thesis, Chinese University of Hong Kong, 2003.

[49] Jinsung Yoon and Mihaela van der Schaar. E-RNN: Entangled Recurrent Neural Networks for Causal Prediction. In *Proceedings of the International Conference on Machine Learning*, 2017.

[50] Huanhuan Yu, Rongda Chen, and Guoping Zhang. A SVM Stock Selection Model within PCA. *Procedia Computer Science*, 31:406 – 412, 2014.

[51] Yao Zhang. Using financial reports to predict stock market trends with machine learning techniques. Master's thesis, University of Oxford, 2015.