Computer Science
&
Data Science

Capstone Report - Spring 2022

# Application of Attention-Based Architectures in Dynamic Aircraft Valuation

*Adib El Ounani,*
*Paule Kairyte*

supervised by
Guo Li & David Yu

## Preface

This research paper has been written as a capstone project for undergraduate studies at New York University Shanghai, by Adib El Ounani (Computer Science major) and Paule Kairyte (Data Science and Business and Finance double-major). Supervising the research were Assistant Professor of Practice in Data Science Li Guo and CFA, ISTAT Certified Aviation Appraiser, Adjunct Professor of Finance David Yu. The project was inspired by the large turbulence, financial risk of the aviation sector and the inaccuracy of conventional aircraft valuation methods. The methods described in this paper are aimed to prove to be useful for aircraft appraisers as well as of interest to anyone in the aircraft financing field.

## Acknowledgements

Special thanks to fellow student Robert Melikyan who consistently provided guidance and support throughout the length of the research. We are also very grateful to professor David Yu who allowed us to join on this ongoing research project and provided with all the necessary background in aircraft financing. Thank you to assistant professor Li Guo, who has helped us with maintaining the progress of the research and assisted with technical aspects of the project. Lastly, we would like to thank Sihan Liu, who was part of our research team in the first half of our research and has contributed to the project.

**Abstract**

*Aircraft appraisers, amongst other parties in the aviation industry, rely the most on the accurate and efficient aircraft valuation metrics and predictions. However, lack of data accessibility and standardization, inaccuracies of deep rooted traditional theoretical valuation approaches, has become a barrier for quantitative aircraft valuation techniques. This project's aim is to get over those boundaries, collect and consolidate data from different sources, such as, reports from Appraisers, Airlines, FAA, and OEMs, in order to create a workable data set for quantitative solutions. The approaches used are building on results from previous machine learning models implemented, and takes on deep-learning models, such as Convolution Neural Networks, Long-Short Term Memory, and attention based Long-Short Term Memory models to address the hyper-correlation of aircraft status data.*

**Keywords**

**Capstone; Computer science; New York University Shanghai; Data Science; Aviation; Aircraft Valuation; Convolution Neural Networks; Long-Short Term Memory Network; Machine Learning; Deep Learning; Attention-based architecture; Attention-based LSTM; Senior Project**

## Contents

# 1 Introduction

The air travel industry has been experiencing considerable financial risk and disruption in the past few years. While it is mainly due to COVID-19 restrictions and its consequences in reducing air travel, even without any external influences aviation sector is prone to turbulence and financial risk. While there are many factor contributing to it, aircraft valuation still remains one of the most important factors that impact the industry. Aircraft are large and complex assets, each with its own specific use and models, sometimes even specific components[1]. In his book, David Yu sheds light on valuation differences along with the factors that influence such values[2], and while there have been many valuation methods used before, none have successfully addressed the diverse nature of each aircraft based on their build and specifications. As the basis for the air transportation industry, a universal model to navigate the diversity of the aircraft types and their special features has been lacking. Thus, the main goal is to create a model to help us tackle this conundrum.

Traditional methods for valuation, such as equity and cash-glow focused techniques, are not airtight due to the assumptions and inconsistencies involved in the valuation process. In the past two years, some New York University Shanghai students have already made an effort to solve this problem [3][4]. They built machine learning models such as random forest, decision trees, k-means, ensemble model and neural network model in an attempt to predict the price of a new aircraft accurately. Although, the work proved to be fruitful and they were able to provide analytical suggestions and efficient value predictions, there is a room for improvement of accuracy by implementing more advance deep learning techniques.

Our research team, has set a goal to continue building on the previous work, improve the quality of the data used, and try different deep learning architectures, such as Convolutional Neural Networks and Long-Short Term Memory Network, for this task. The main issue in the previous studies was that certain features of the data set appeared to be highly correlated[3]. Since attention layer showed good performance in feature selection, we decided to test attention-based Long-Short Term Memory model and Convolutional Neural Network model to solve this issue[5].

The process to achieving our goal consisted of the following steps:

1. Expand the existing data by merging more diversified sources.
2. Clean and transform new data using naive imputation techniques.
3. Compare mean and median imputation of missing values.
4. Build and test LSTM, CNN, and attention based LSTM models and compare results.

## 2 Related Work

Aircraft valuation is important to all main parties of the aviation industry like OEMs, MROs, Appraisers, Leasing Companies, and Airlines. However, most popular methodologies for aircraft valuation come from methods and views used by aircraft operators and aviation investors. Three most significant approaches of modeling methodologies are net present value (NPV), return on invested capital (ROIC), and real option analysis. [6]

The net present value (NPV) methodology considers operating cash flows positions and provides a present value by discounting a series of projected future cash flows generated by the specific aircraft.[7] The issue with this method is that due to geographic and economic features, it is hard to accurately predict the operating cash, furthermore, even using relatively flexible WACC-based net present value approach with Monte-Carlo Simulation, it fails to capture the risks of owning and operating and aircraft and the flexibility of operating leases.

From the perspective of the long-term economic return, the return on invested capital (ROIC) methodology values the aircraft based on external market demand (i.e., number of aircraft in service, firm backlog, customer base, and time in the aviation cycle), offering a more robust remarketing perspective.[8] The flaws of this approach are that it is difficult to quantify such factors like new market expansion, replacement and retirement decisions, making them hard to embed into models and use to correctly estimate the cash flows.[9]

Despite the shortcomings of these approaches, leasing and airline companies still focus on cash flow-based valuation techniques, such as NPV and ROIC, as it provides a link between real-time value and profit, while taking into consideration the changing condition of aircraft operation. Due to this factor these two methods perform better than accounting-based approaches, thus making them more widely used.

The real option analysis method mainly used by OEM, who design and manufacture aircraft and performance improvement packages (PIP), allows them to decide more accurately on the production choices and pricing of aircraft and PIP. [10] This approach captures the opportunities aircraft operators have to expand into new markets, develop new routes, form new fleet combinations, and increase competitive advantage with the purchase of new aircraft. The real option analysis method is suggested since traditional cash-based valuation approaches fail to account for the flexibility offered to airline management to steer programs into profitable directions. [11]However, it is trivial to quantify each opportunity adopted in real option analysis and the process involves many assumptions.

As technology progressed, so did the valuation approaches, and a new trend is visible in the publications in the field. There has been a shift away from above-mentioned cash flow-based approaches to regressive methods, raising a question of which value is being calculated and how it can be defined.[12]

The appraising industry rarely adopts the quantitative valuation approach and instead focuses on the valuation framework, measurement metrics, and the historical influence of single factors on aircraft valuation, allowing them to better interpret the valuation. According to ISTAT (International Society of Transport Aircraft Trading), appraisers face different value definitions based on their economic inclusion of factors. The four main kinds of "Value" referred by them are the "Base Value", the "Market Value", the "Residual Value", and the "Distressed Value". For the purposes of our study we will focus on the prediction of the "Base Value" and the "Market Value", since according to the bluebooks of appraisers, they provide the most direct standard for a value of an aircraft. The definitions of these two values are as follows[11]:

**“Base Value"** is the appraiser’s opinion of the underlying economic value of an aircraft in an open, unrestricted, stable market environment with a reasonable balance of supply and demand, and assumes full consideration of its "highest and best use". An aircraft’s Base Value is founded in the historical trend of values and in the projection of value trends and presumes an arm’s-length cash transaction between willing, able and knowledgeable parties, acting prudently, with an absence of duress and with a reasonable period of time available for marketing. in most cases, the Base Value of an aircraft assumes its physical condition is average for an aircraft of its type and age, and its maintenance time status is at mid-life, mid-time.

**“Market Value"** is the appraiser’s opinion of the most likely trading price that may be generated for an aircraft under the market circumstances that are perceived to exist at the time in question. Market Value assumes that the aircraft is valued at its highest, best use, that the parties to the hypothetical sale transaction are willing, able, prudent and knowledgeable, and under no unusual pressure for a prompt sale, and that the transaction would be negotiated in an open and unrestricted market on an arm’s-length basis, for cash or equivalent consideration, and given an adequate amount of time for effective exposure to prospective buyers.

Although the above mentioned methods provide comprehensive interpretation of how various factors affect the price, they are still limited by assumptions. Recent advancements in data science and computer science allow the implementation of more advanced statistical methods, machine and deep learning models to quantify performance and depreciation data and estimate the value of an aircraft. Previous works by David Yu, Yuxin Zhang, Robert Melikyan, and Liu Houze demonstrate the advantages of conditioned regression, machine and deep learning methods.[2] [4] [3] Although, previous research shows improvement in aircraft price valuation, most recent results show necessity for further work with the size and diversity of data, as well as fine tuning of the models.

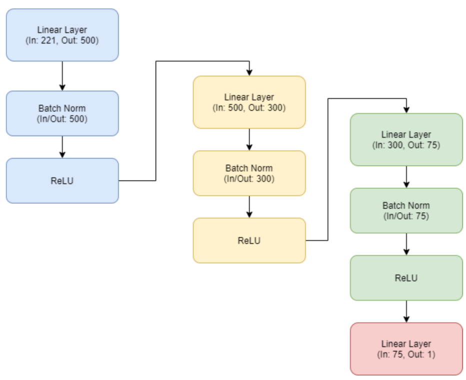


Figure 1: The ANN architecture used by Robert Melikyan and Liu Houze

The paper written by Vaheb attempted to forecast stock prices of Goldman Sachs and General Electric companies using a special type of deep learning architecture, the Long-Short Term Memory architecture.[13] Although the final result of his model was not very satisfactory, this paper has its strength in explaining why we should continue working with deep learning models, especially LSTM model for the goal of asset pricing. To add on work of our peers regarding this model, based off of the results of asset pricing achieved by Luyang Chen, Markus Pelger, and Jason Zhuwe, we implemented a similar recurrent LSTM network to better reflect the market and capture macroeconomic information affecting the seasonal cycles of the aviation industry.[8]

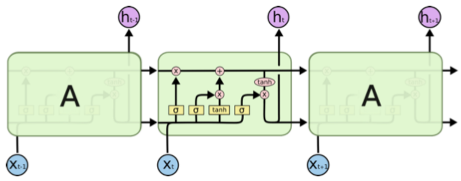


Figure 2: The LSTM architecture used by Vaheb

Research paper *Convolutional neural networks applied to high-frequency market microstructure forecasting* investigates the advantages of Convolutional Neural Networks (CNN) for big data in financial forecasting.[14] Their work indicates that CNN architectures deal well with forecasting tasks and extracting meaningful features. Research presented in the paper *Attention Is All You Need*, further explains the advantages of attention-based architectures, encouraging us to use Attention-Based Convolution Neural Networks architecture in our work.

# 3 Solution

## 3.1 Data Engineering and Cleaning

The lack of more diversified data, noted in previous work done, required scrapping from wider range of data sources, processing and merging with the old data set. The original data set shape was 59786 rows and 118 columns, the newly collected data set shape was 55605 rows and 72 columns. After merging the two data sets and cleaning the attributes, we ended up with a final merged data set of 115392 rows and 114 columns. After combining the new data with the previous data, the new merged data set had an increase of 86.47 percent compared to the old data set.

For initial collection of data, we had to scrape it form PDF documents shared by different airlines and appraisers. We used OCR (Optical Character Recognition) to read the values from the documents, however, due to low quality of some of the documents, there was high level of inaccuracy in the initially collected data. Part of the data was cleaned utilizing python and it's libraries like pandas, however, there was a relatively big part of the new data that had to be cleaned manually. One of the main issues with the newly collected data, was a lack of uniformity between the data sources, which resulted in duplicate columns with slightly different labels (sometimes due to encoding standards). After adjusting the data manually and using Pandas library, we merged the two data sets, and conducted an additional check if the is any need more need for data scrubbing.

## 3.2 Data processing

The nature of the data itself, due to the diversity in the formats of the data sources, made it necessary for extensive data processing. The main challenge was due to lack of data uniformity, there were many missing values, however, we could not just opt to delete entire rows, because that would result in decreasing the data by approximately 90 percent. Filling in the missing values with NaN value was also not an appropriate solution, since it would create inaccuracy in the predictions of the models.

To solve the issue of missing values we tested different variations of data imputation. We maintained the original data with the missing values for baseline, and compared it with data that used mean and updating mean for imputation of continuous variables and most frequent value for categorical variables. Although, the previous year's work used the K-Nearest Neighbors for data imputation, we decided to go with the mean and updating mean this time, since the data is already significantly more diversified and elaborate than it was last year, therefore, we decided that for our purposes of testing the efficiency of attention-based architectures in aircraft valuation it would be enough. However, utilizing K-Nearest Neighbors together with attention-based architectures could yield improvement in accuracy and is something we are considering for future work.

## 3.3 Model Training and Analysis

Since the new data set was substantially larger than the original data set used in the previous year, in order to save time and computational power, we decided to first test all models on the old data set, and then re-run the best performing models on the combined data for the final results. During the initial tests we found that CNN had significantly worse performance than Attention-based LSTM and our baseline LSTM model, therefore, we did not run the CNN model on the new data set.

## a) Long-Short Term Memory Networks

In addition to previously built and tested models, this year's work also expanded on Recurrent Neural Network architectures, more specifically Long-Short Term Memory Networks. After preprocessing the data the dimensions of our data frame were 59786x119, which for this architecture was used with train-validation-test split. For testing we used 10 percent of the data, train and validation split was determined through iterations of batches from data with a split of 80-20. In order to avoid over-fitting of the model, we also implemented an early stop callback to monitor the loss of accuracy with an increase of epochs, which resulted in the final model having 100 epochs. The final architecture is illustrated in Figure 3.

```
Model: "sequential_7"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm_7 (LSTM)               (None, 300, 10)           4800

 time_distributed_7 (TimeDis  (None, 300, 1)           11
 tributed)

=================================================================
Total params: 4,811
Trainable params: 4,811
Non-trainable params: 0
_________________________________________________________________
```

Figure 3: The LSTM Architecture in Text

## b) Convolutional Neural Network

Another architecture, that we tested, based on it's success in related fields and big data is Convolutional Neural Network. For the task intended, a 2D CNN architecture. To reduce the computational costs a Max pooling layer was added. To avoid over-fitting the drop rate of the dropout layer is 0.1. For the activation function a Sigmoid function was used for the data. However, the CNN model we build, didn't provide significant results and needs more fine-tuning to be done in the future. The final architecture is illustrated in Figure 4.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 300, 1, 8)         880

 conv2d_1 (Conv2D)           (None, 298, 1, 8)         200

 max_pooling2d (MaxPooling2D  (None, 149, 1, 8)        0
 )

 conv2d_2 (Conv2D)           (None, 147, 1, 8)         200

 max_pooling2d_1 (MaxPooling  (None, 73, 1, 8)         0
 2D)

 flatten (Flatten)           (None, 584)               0

 dropout (Dropout)           (None, 584)               0

 dense (Dense)               (None, 1)                 585

=================================================================
Total params: 1,865
Trainable params: 1,865
Non-trainable params: 0
```

Figure 4: The CNN Architecture in Text

## c) Attention based Long-Short Term Memory Networks

Since, the LSTM model performed better than CNN, we decided to add an attention mechanism for processing sequential data to our LSTM architecture. Attention layer, acts as an interface connecting the encoder and decoder, providing it with information from all hidden states of the encoder, as such, allowing the model to learn the association between parts of the input sequence, thus solving the issue of hyper-correlation in certain features of our data.

$$h_{t,t'} = tanh(x_t^T W_t + x_{t'}^T W_x + b_t)$$

$$e_{t,t'} = \sigma(W_a h_{t,t'} + b_a)$$

$$a_t = softmax(e_t)$$

$$l_t = \sum_{t'} a_{t,t'} x_{t'}$$

The LSTM models focus more on short-term memory, as the name suggests, however combining it with an attention layer gives more context to the algorithm. For that we used the SeqSelfAttention library from Keras. The final architecture is illustrated in Figure 5.

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm_1 (LSTM)               (None, 300, 10)           4800

 seq_self_attention_1 (SeqSe  (None, 300, 10)          705
 lfAttention)

 time_distributed_1 (TimeDis  (None, 300, 1)           11
 tributed)

=================================================================
Total params: 5,516
Trainable params: 5,516
Non-trainable params: 0
```

Figure 5: The Attention-based LSTM Architecture in Text

# 4 Results

## 4.1 Experimentation Protocol

For our purposes, mean square error is the most suitable way to measure performance, because the predictions must be compared to market and base values. For deep learning models we built the data split is 90-10, with 90 percent training allocation, 10 percent testing. The 90 percent of the data set that was allocated to training, was then split 80-20 between training and validation. The above mentioned split of training and testing data was done to avoid any significant overfitting problems.

## 4.2 Testing results

As we have mentioned before the CNN model did not discern the underlying patterns in the old data, as can be seen in the figure below. Due to relatively low accuracy of the model we decided to move forward without running it on the new consolidated data, and focus more on Long-Short Term Memory architecture for the application of attention layers. The results of our Convolutional Neural Networks model for train validation loss can be seen in Figure 6.

For the LSTM data without any type of imputation for the missing values, the results appears to be decent and reasonable (See Figure 7). However, it appears that our dynamical imputation of the mean does to not yield better success rates, as expected initially (See Figure 8).

Since LSTM showed better results than CNN, we decided to apply attention layers to this architecture, expecting to see even higher level of accuracy. However, after running the attention based LSTM it seems that it doesn't perform any better than the previously tested LSTM

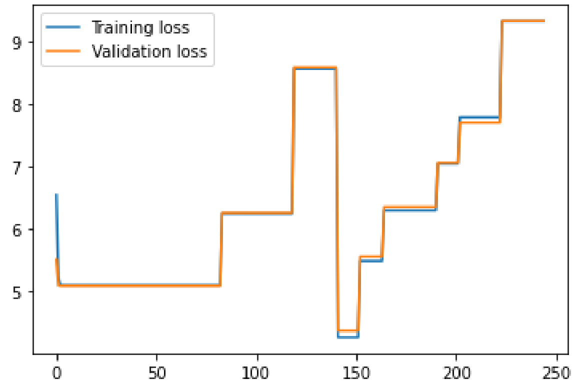


Figure 6: Train Validation Loss for CNN

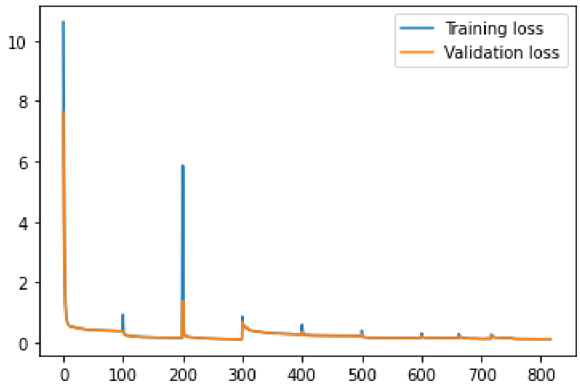


Figure 7: Train Validation Loss for LSTM

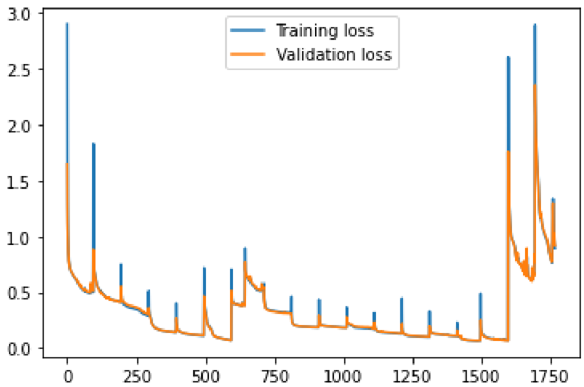


Figure 8: Train Validation Loss for LSTM with Dynamic Imputation

architecture (See Figure 9).

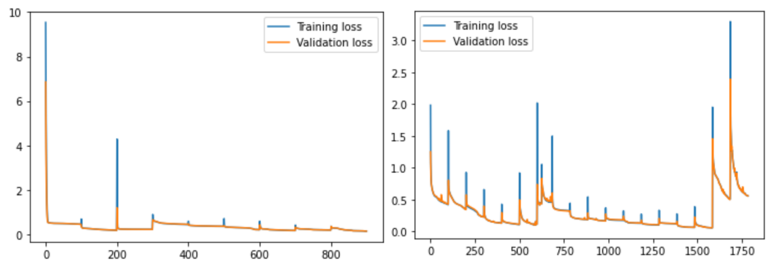


Figure 9: Train Validation Loss for Attention based LSTM

| Metrics\Algos | LSTM | Attention | LSTM (Imputed) | Attention (Imputed) |
|---|---|---|---|---|
| Training MSE | 0.1177 | 0.1447 | 0.8931 | 0.5560 |
| Validation MSE | 0.1240 | 0.1449 | 0.9194 | 0.5553 |
| Testing MSE | 0.8354 | 0.9652 | 3.6590 | 4.0438 |

# 5 Discussion

The project faced several key issues. First of all, the data collection and processing steps took significant portion of the whole project, since there were many discrepancies in the value entries, fixing which could not be done fully by automation and had partially be done manually. Furthermore, around 90 percent of the entries had some missing values present, which had to be addressed in a proper way. To see the full impact of data imputation on the prediction results, we decided to keep the initial data and compare it to data with mean/updating mean values for missing continuous variables and most frequent for categorical values. This provided a better understanding on how the different methods of imputation affected the accuracy of the models, and can be implemented in future work.

The second issue the project faced, is the size of the data and the computational power required to process it, train and test the models. To solve this, we resorted to using old data set (which is almost half the size of the new one), for all the new models, and only run the most successful ones on the new data set. However, due to the increased diversification of the new data set, the assumption that the models would perform the same cannot be fully backed, therefore, given all the necessary resources it would be better to train/test all of the models on the new data.

Lastly, comparing the results with the performance of the last year's models, only Long-Short Term Memory model performed better, although it still did not surpass gradient boosting performance, further backing the hypothesis drawn by the previous students, that deep learning methods do not fully utilize the ability to dig out the hyper-correlation of the attributes without specific structure [3]. However, to verify that, there should be more testing of different complexity of models done.

## 6 Conclusion

In this project, we first expanded the data, since it was one of the key suspected factors affecting the accuracy of the models in the previous work done. We then used three deep learning architectures for comparison of results: Convolutional Neural Networks, Long-Short Term Memory, Attention-based Long-Short Term Memory. First we ran these models on the old data set, since it was significantly smaller and, thus more efficient, to check which models yield significant results. We abandoned CNN model after the first run, because it didn't show promising results and other two models had better performance. The model that displayed the highest level of accuracy in predictions, was LSTM without imputation (missing values were dropped), the training mean squared error was 0.1177, validation mean squared error 0.1240, and testing mean squared error 0.8354. Although, we had the hypothesis, that attention based architectures would show increased accuracy, results of this project showed, that it was even less successful than simple LSTM architecture. Therefore, aggregating the results of this year and previous years, gradient boosting showed the most accurate prediction results. For our future work, we should focus on more elaborate data imputation techniques, since although, the dynamic mean did not show good results, we are limiting the diversity of the data by dropping the empty values. Furthermore, it would be beneficial, to work on a more advanced LSTM model, as well as attempt an attention-based CNN architecture.

## References


[1] R. T. B. Vasigh and D. Jenkins, *Aircraft finance: strategies for managing capital cost in turbulent industry*, 2012.

[2] D. Yu, *Aircraft Valuation: Airplane Investments as an Asset Class.* Palgrave Macmillan, 2020.

[3] R. Melikyan and H. Liu, "Deep learning applied in dynamic aircraft valuation," *Tech. Rep*, 2021.

[4] Y. Zhang, "Dynamic pricing for commercial aircraft," *Tech. Rep*, 2020.

[5] A. Vaswani, et al, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. vol.30, pp. 5998–6008, 2017. [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

[6] B. Vasigh and Z. Rowe, *Foundations of airline finance: methodology and practice.* Routledge, Taylor Francis Group, 2020.

[7] S. Mason and J. Morrison, *The value of new-technology single-aisles.* CIT Group Inc., 2016.

[8] M. P. L. Chen and J. Zhu, "Deep learning in asset pricing," 2021.

[9] J. Gorjidooz, "Aircraft valuation in dynamic air transport industry," *Journal of Business Economics Research*, vol. vol.8, 12 2010.

[10] C. Justin and D. Mavris, "Valuation of real options with flexible early exercise in a competitive environment: The case of performance improvement packages," 7 2014.

[11] W. Gibson and P. Morrell, "Airline finance and aircraft financial evaluation- evidence from the field," *ATRS World Conference*, pp. 6–9, 6 2005.

[12] N. Hallerstrom, *Modeling Aircraft Loan Lease Portfolios.* Online, https://hallerstrom.com/wp-content/uploads/2020/09/Modelling-Aircraft-Loan-Lease-Portfolios-January-2020.pdf.

[13] E. Salavati and H. Vaheb, "Asset price forecasting using recurrent neural networks," 2020.

[14] M. F. J. Doering and S. Markose, "Convolutional neural networks applied to high-frequency market microstructure forecasting," *Computer Science and Electronic Engineering (CEEC)*, vol. vol.9, pp. 31–36, 2017. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8101595