Investigating the predictive ability of LSTM networks on industrial metal prices

OLIVER BOOM, EXTERNAL SUPERVISOR: T. FLETCHER, and INTERNAL SUPERVISOR: S. NEETHLING

A proposal for the investigation of price prediction of Industrial Metals using long-short term memory networks (LSTMs) is presented. Two time-series methods; ARIMA and additive modelling, and two variants of a stochastic method; uni-variate and multi-variate Gaussian Process prediction, have been selected as benchmark tests. The considerations required to evaluate the prognostic ability of uni-variate and multi-variate LSTM network are assessed. And the opportunity to exploit the shared information within the Industrial Metals universe by using Multi-Task Learning is explored, in line with the econometric literature's theory of excessive co-movement theory and the existence of a common factor between commodities.

1 INTRODUCTION

In this paper a brief introduction to Industrial Metals (IMs) and why their price prediction is of interest will be discussed. Then a short survey outlining the major challenges of price prediction and available techniques is given. The data available is explored with a discussion on why IMs should benefit from a Multi-Task Learning framework.

Once this background has been provided the project objectives and proposed methodology are presented. Here, some of the key considerations in how the project will be approached, and the manner in which progress will be evaluated is discussed.

1.1 Industrial Metals

Industrial Metals are resources of paramount importance across economies of all stages of development, and are the raw materials that unpins vast sectors. IMs form a subset of commodity sector, and their price fluctuations have significant repercussions on macroeconomic performance[1][2].

There is a diverse range of parties that have a strong interest in forecasting IM prices. Mining profit has high sensitivity to IM spot price, therefore price forecasting is important to determine whether a mining activities can be economically exploited[3]. Further down the supply chain in manufacturing, price forecasting can advise time-to-buy strategy for material procurement for smaller businesses, or influence whether engaging in hedging strategies is advisable for larger businesses. Larger still, both policymakers and corporate managers follow prices closely as underlying price has a large impact on trade[4][5]. Due to the considerable interest in price forecasting, significant effort is made by researchers to increase their predictive ability.

1.2 Price Forecasting

Price forecasting is challenging due to the high volatility and non-linearity of the underlying instruments and the extensive quantity of parameters both observed and hidden[6]. It is unknown whether the challenges in modelling arise from the technical challenge of modelling non-linearity or whether the driving factors are exogenously based and therefore not captured by the models.

An extensive range of methodologies are available to researchers wishing to forecast prices. The most prevalent techniques can be categorized into; qualitative, econometric, stochastic, time-series and machine learning based[3][7]. The technical skill, accuracy and background knowledge varies widely across these methodologies, with each technique having advantages and disadvantages. Econometric and qualitative methods were historically more important, but there has been a practical move towards time-series methods more recently. However, the ability of machine learning techniques to model dynamic, non-linear time-series means that there is significant potential for this methodology to outperform other techniques[7][8].

One subset of machine learning, deep learning, has the ability to capture complex non-linear relationships within data. However, the traditional neural networks architectures which perform well on regression and classification problems fail to adapt well to deal with temporal problems due to the loss of the temporal structure of the data. A neural architecture the Recurrent Neural Network (RNN) was developed to better model dynamic, temporal behaviour. However, there are often issues with the vanishing gradient problem due to the recurrent structure of the networks. This results in either the network vanishing to zero, or exponentially blowing up.

This was addressed by re-engineering the neuron to cleverly avoid the vanishing gradient problem[9], with this new structure of network being called a Long Short Term Memory network (LSTM). Their primary application has been within speech recognition, and while their application to econometric time-series is certainly a prominent area of research, it is still a relatively new application. And 2009 V. Ahti stated that feed forward neural networks were still the "neural architecture of choice in time-series econometrics"[4].

Price forecasting can be framed as either an auto-regressive, uni-variate or multi-variate problem. If all of the information to predict the time-series is contained within an instruments own history, then auto-regressive techniques are appropriate. Often with financial prediction there may be other financial time-series which are highly correlated but lagged with the instrument you wish to predict. An ideal situation for a researcher is to find one or more financial time-series that have a highly correlated relationship/s and predictable lag/s with the IM of interest. There are numerous complex econometric factors that impact the price of commodities so uni-variate prediction would be too simplistic, while multi-variate prediction might have more success[1]. The more driving factors that are included in the dataset the greater the predictive power.

The hypothesis for this investigation is that high capacity models are required to to tackle the complex problem of econometric time series prediction. The state-of-the-art Long Short Term Memory networks are a sophisticated deep learning multi-variate regression technique. And it is the hypothesis of this investigation that these LSTM networks should outperform other conventional computational methods in predicting IM prices. To test this hypothesis the predictions made by LSTM regression will be compared against several benchmark methods. These benchmark methods will be; the classical time-series model ARIMA, a more sophisticated additive regression model (FbProphet) designed to capture seasonality (both of which are auto-regressive) and also stochastic methods which could take the form of an auto-regressive (AR) and/or multi-variate Gaussian Processes (GP) regression. As the multi-variate GP could potentially be much more complex than the ARIMAm, FbProphet and AR GP regression models it is highlighted as a desirable but not essential benchmark test.

1.3 Datasets

Through a partnership with ChAI extensive econometric data has been acquired, some of which is believed to have prognostic ability on IM metals prices. These are financial time-series dating back to 2006. Through a literature review and discussions with researchers at ChAI, the instruments which are believed to have the best predictive ability on IM instruments of choice will be selected to be include in the training dataset.

The financial data available is of high quality, with most instruments being sampled daily. Examples of the econometric data available is foreign exchange prices, commodity spot prices, bond prices, implied volatility rates and interest rates etc.

One of the major challenges in this project is determining whether there is enough data to make high quality predictions using deep learning. There are certainly numerous exogenous factors that will be driving prices of IMs both as a collective and idiosyncratically, that cannot be included in the dataset due to unavailability of data. However, the question is whether enough of key driving factors for IM prices is contained within the dataset to capture a large enough percentage of the variance within IM prices, in order to allow for accurate forecasting. Even if many of the key driving factors are contained within the dataset it is not clear whether there is enough data to train LSTMs and not over-fit to the data. It is certainly probably that there might not be enough data within the dataset to allow for a well generalised LSTM predictor. There is no hard and fast rule as to the quantity of data required for LSTMs to perform well. And this is one of the core questions of the project; is there enough data contained within the dataset for a LSTM to outperform simpler models.

Alongside the general overall volume of the dataset being potentially too small, there could potentially be issues regarding the length, in particular, it being insufficiently long. The data dates back to 2006, but through discussions with Stephen, it was noted that some of the time lags for the driving factors of IM prices movement can be of the orders of 10 years. As these factors would only appear once within the dataset it is unlikely they would be captured by the models. However, this is not a serious concern because the long lag driving factors are generally more applicable to long terms forecasts i.e within the span of years rather than the 1 day to 1 year time window. Therefore initial analysis would lead to the belief that the shorter window is not an issue.

1.4 Multi-Task Learning

There is considerable evidence that Mineral Commodities exhibit significant co-movement[10]. This is a significant area of economic research with the observation that a "common factor" exists within the Mineral Commodities universe[1]. There is a shared common factor both within the overall commodities universe and within the closer related sub-universes i.e precious metals, industrial metals etc. And also research into which financial variables are correlated with the common factor to a statistically significant degree[1]. These investigation will help aid the selection of financial time-series as features for the dataset.

In predicting a singular metal e.g copper, there will be components of the prediction which are shared with the other IMs such as market risk-on/risk-off attitudes, US dollar effective exchange rate, global demand etc. But there will also be an idiosyncratic component of the price prediction which is unique to copper. There is potential to exploit this shared information between these IMs using Multi-Task Learning.

Multi-Task Learning (MTL) is a machine learning framework which aims to leverage shared information between a collection of related tasks to improve the predictive performance in all tasks[11][12]. MTL is more a paradigm than a specific technique. A series of tasks are trained together simultaneously with the tasks being optimized for as a collective. The objective here would to have a shared dataset of features used for training, and the prices of several metals as target variables. With each metal price being an individual task, these tasks can be optimised simultaneously, sharing common information to give better predictive performance. Section 1.3 noted the challenges of the relatively small dataset. MTL is a technique that increases generalisation of models, and it is hoped that its use will decrease over-fitting.

2 PROJECT OBJECTIVES

The primary objective of this project is to investigate the predictive ability of LSTM networks on industrial metal prices. This is split into two primary areas; firstly to ascertain whether LSTM networks outperform conventional time-series and stochastic techniques and then secondly whether Multi-Task Learning can be used with LSTMs to improve predictive performance. To achieve this objective there are several sub-objectives which have been identified and discussed in this Section. A top level view of the objectives looks as follows:

1. Feature Selection
2. Pre-processing
   - Data Cleaning
   - Feature Engineering
   - Dimensionality Reduction
3. Benchmark Creation
   - ARIMA
   - Additive Modelling (FbProphet)
   - Auto-Regressive Gaussian Process Regression
   - Multi-variate Gaussian Process Regression
4. LSTM Investigation
   - Auto-regression
   - Multi-Input Single-Output (MISO)
   - Multi-Input Multi Output (MIMO)

3 PROPOSED METHODOLOGY

In this Section, considerations for the scheme of work are presented alongside some initial technical decisions.

The software development language will be Python3 and best practices in sustainable software development will be followed. The individual sub-modules will be developed in a Google Colabs environment, while being integrated with Github

3.1 Project Management

To track the timeline of this project, assess progress and manage risks on delivery, a Gaant chart is being used for project management. This is contained within Figure 1. As is seen in Section 2 the project is relatively modular in it's nature. The different regression techniques can be considered neat, independent pieces of development. This is useful in detailing the scheme of work, and the project scheduling has been based off of this list.

Each individual benchmark technique has an output prediction which can be considered a deliverable. The objective is to be able to perform a prediction at the end of each development period scheduled within Figure 1. After discussion with Stephen, it was decided that this does not need to be a physically submitted deliverable, but can take place in the form of a review of the code and discussion regarding the results.

In addition to these prediction reviews there are also several draft deadlines for the literature review and the report. In Figure 1 the week is generally scheduled with specific tasks, while the weekends have been left as free days to catch-up on any aspect of the project that requires more attention. In the Figure the blue periods are periods of work while the red diamonds are physical deliverables.

3.2 Technical Decisions

3.2.1 Target Variable Format. The common target variable that is predicted in price forecasting is the log return of the instrument of interest. Choosing the log return instead of the return has several useful statistical properties in time-series forecasting[13] and all reference to predictions should assume that the quantity being predicted is the log return of the IM of choice, relative to the date of prediction.

3.2.2 Forecast Length. The forecast length is how far into the future is being predicted. The time-frame of the prediction is important and some financial time-series might have prognostic ability over some time-horizons but not others. It is generally considered that as the forecast length increases so does the inaccuracy of the forecast. Short term prediction i.e 1 day to 1 week is likely to be dominated by noise, whereas long term predictions is likely to be unreliable. Time horizons of a week, a month and a quarter have been selected as periods on interest for the forecast length. And were selected partially due to them being periods of interest for ChAIs clients.

3.2.3 Rolling Windows. The way the LSTM networks will be evaluated is by training them on a fraction of the overall dataset, yet to be decided, then using a rolling window approach on held out data. A subset of the test dataset of a given window size is used to predict one forecast length into the future. This price prediction is then compared to the observed price for the decided error metrics. Then the window is moved forward by one time step and the process repeated. This is carried out for the entire test set and the model is not retrained while carrying out the predictions.

3.2.4 Error metric. Several metrics will be used to evaluate the performance of the prediction methods. Two of the most popular metrics within regression are Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE)[14]. MSE was selected due to its simplicity and intuitive nature. Also its sensitivity to outliers is a desirable quality in price prediction[15]. MAPE was selected as it achieves the same outcome as Mean Absolute Error evaluation but has the additional benefit of showing the relative size of the error. In addition to MSE and MAPE the directionality of the prediction will also be evaluated using the Mean Directional Accuracy (MDA) metric.

3.2.5 Feature Selection. There are many different individual instruments within the IM universe. Aluminium and Copper have been highlighted as good target variables for prediction. This is as they are extensively used/traded and are arguably the two most important non-precious metals in the world by any metric (extraction, price, consumption etc[16]) and accurate forecasts for these would be well received. Additionally there is a rich source of predictive literature with these as their subject, on which to build[17][18].

When the problem expands to a larger set of target variables for Multi-Input Multi-Output MTL, more metals must be chosen as individual tasks. This is an early further objective; to select what other metals would be best suited for inclusion in the LSTM MIMO target prediction set.

In addition to further target variables selection, more analysis must be conducted into what financial time-series should be included as features in the dataset. The features vary depending on which financial exchange is being predicted for. As an example, for Aluminium (Al) on the LME the input time series are; Al SHFE stocks, Al LME prices, Al COMEX stocks, Al LME stocks, Al COMEX stocks, TED spread, Baltic Dry Index, CBOE Volatility Index, SKEW Index, Goldman Sachs Commodity Index, Yuan Spot Price.

3.2.6 Pre-processing. Prior to using the regression models, it will be necessary to clean and engineer the data to remove bad data, and improve model performance. The data cleaning will be fairly simple, looking to remove data points which are significantly far from the mean (3 standard deviations).

Through discussions with members of the ChAI team, potential transformations of the input time-series were explored. Several techniques were advised from empirical industry experience within hedge funds. The aim of these techniques is, for each input financial time-series, to spawn transformations that will augment the data and enhance the information gained from that singular series.

A more extensive literature review of these techniques is required to show peer-reviewed justification for their use. The methods that were advised were for each input were; percentage changes, quantiles of percentage changes, deciles, exponentially weighted moving averages (with a series of different half lives) and a rolling volatility curve. The exponentially weighted moving averages should act to de-noise the data. While the rolling volatility transform can serve as an indicator of risk appetites. It might be unnecessary to incorporate all of these transforms and further investigation regarding which to use will be carried out.

Once these transforms have been performed the number of potential inputs to the model will have risen dramatically. If 8 transforms were applied (4 exponentially weighted average half lives and one of each of the others), then it has increased the dataset by a factor of 8. There will be added value within this data but also significant increase in dimensionality. Therefore it will be necessary to perform dimensionality reduction to capture the majority of the variance while decreasing the number of inputs. To do this principal component analysis (PCA) will be conducted to reduce the features down to the principal components which will contain 95% of the variance within the data. This was an empirically chosen number, which was deemed to contain a large amount of variance, while likely being able to significantly reduce the dimensionality of the data.

4 PRELIMINARY LITERATURE REVIEW

4.1 Economic Theory & Price Prediction

Industrial metals are essential economic drivers not just for producing countries, where they can be the most significant exported product for some countries; e.g Zambia with 98% of it's exports were copper in the 90s[4] but also for consuming countries due to the widespread use as an input into many industries[19]. There is a deep and extensive pool of literature within economics detailing the driving factors for IM price movements. And significant interest from researchers goes into price prediction [3], with a wide range of methodologies being explored. One of the major challenges in price prediction is that metal prices are non-linear time-series that form as some combination of market dynamics and stochastic economic forces, with the relative contributing factor of these being an unknown[20][6].

"An Assessment of time-series methods in metal price forecasting" provides a good discussion of why price prediction if important, an overview of the key metals exchanges and their role in price discovery and a discussion of some of the simpler and most prevalent computational prediction methods; static time-series forecasting[3]. The ARIMA model appears in this and many other papers, as the de facto model used in price prediction[17].

Historically qualitative economic analysis was the most widespread form of price prediction, until the relatively sophisticated time-series methods became more common. This was due to the relative ease of making time-series predictions compared to the extensive economic education required to make qualitative predictions. In "Alternative techniques for forecasting mineral commodity prices" a discussion of the relative merits of the available methodologies is presented[7]. Alongside time-series and qualitative methods stochastic Gaussian models are presented, as well as a general treatment of Machine Learning. Cortez et al conclude that machine learning appears the most promising computational technique available for price prediction. The justification for this being that econometric, stochastic techniques and time-series methods do not capture the dynamic qualities earlier discussed[6]. Cortez et al say this is due to these methods assuming linearity[7]. While this is certainly the case for the time-series methods, it is slightly surprising in regards to the stochastic techniques, as the key stochastic technique I am investigating; Gaussian Processes appear as a tool to model non-linear time-series frequently in literature[21][17].

An observation made in the economic literature is that commodities prices exhibit excess co-movement[1][19]. Pindyck and Rotemberg argued that commodities which are intuitively unrelated, display excessive co-movement and a tendency to move together[22]. In "How important are common factors in driving commodity prices?" Vansteenkiste analyses which econometric factors are statistically significant with non-fuel commodity price movements and determines that commodity prices are well correlated with other commodities within the same category e.g industrial metals with other industrial metals. But they also display co-movement with other commodity sub-universes. She determines that there is a statistically significant common factor to these commodities. It is this which prompted the idea of using Multi-Task Learning as as a potential approach to exploit the existence of this common factor.

4.2 Time-Series

In his book "Stationary and non-stationary time-series" G. Nason details the technical considerations required for time-series prediction. The concept of stationary is explained and why it is important for time-series forecasting. The majority of common time-series models require stationarity as a pre-requisite. Therefore, if a time-series is not stationary if it necessary for the analyst to transform the time-series into a stationary signal[23]. In the book "An Introductory Study on Time Series Modeling and Forecasting", ARIMA is noted as being one of the most popular models due to its flexibility and methodical approach[24] but also is described as extremely limited due to the assumption that the underlying series is linear, clearly an incorrect assumption in IM prediction. These books detail the steps required to make an ARIMA prediction, and they will serve as the guide to create the benchmark prediction. Due to space limitations in this plan, the details of the workings of these models and will not be discussed in depth.

4.3 Gaussian Processes

"Gaussian processes for time-series modelling" provides a good introduction to GPs and show GPs to have several desirable qualities for modelling[25]. They are described as being good models for dynamic and noisy time series, and are good at dealing with uncertainty with modelling. GPs allow for a more methodical approach to uncertainty, and using Bayesian inference allows for "robust modelling even in highly uncertain situations"[25]. One of the very powerful aspects of GPs is that they are deal with probability directly and give both the predictions and variance of the predictions, allowing you to place confidence intervals on your predictions[26]. Another desirable quality of GPs is that they are interpretable and easily explained. In financial forecasting this is very desirable and a premium quality that could be essential to some researchers[27]. In "Multi-task Learning with Gaussian Processes", Adam Chai also details how to use GPs in a MTL framework[27].

4.4 Long Short Term Memory Networks

"Deep Learning with Long Short-Term Memory for Time Series Prediction" details how LSTMs were developed as an answer to the challenge of wishing to use deep learning on time series data[28]. They are Recurrent Neural Network (RNN) that addresses the tendency of RNNs to either decay to nothing, or exponentially blow up by engineering the composition of the individual neurons that the networks are built from. It is shown that LSTMs significantly outperform ARIMA models in the case of traffic data[28]. Several other use-case papers were explored [29][8] with the later paper being as recent as March 2019. These papers detail positive performance of LSTMs, with the second paper having sophisticated deep learning methods as it's benchmark tests.

4.5 Multi-Task Learning

"A brief review on multi-task learning" and "A survey on Multi-Task Learning" provide an description of the state-of-the-art within MTL. As a paradigm it has only existed since the late 90s, and is similar to transfer learning, but in MTL all the task are optimized for simultaneously and there is no distinction between tasks[12][11].

Thung and Wee's survey detail the distinctions between the different data fidelity terms of MTL; MISO (Multi-Input Single-Output), SIMO (Single-Input Multi-Output), MIMO (Multi-Input Multi-Output) and their mathematical formulation[11]. It also details the mathematical formulation of regularisation approaches and other specific cases, as well as its application using incomplete data and deep learning, both of which might be applicable in this project.

There are some relevant papers inspecting the application of MTL within deep learning. In "Multi-Task Learning for Stock Selection" J. Ghosn et al. [30] found that using MTL increased generalisation performance and outperformed benchmarks significantly, although it was only compared to two buy-and-hold benchmarks, which is limited competition. However, it is a useful paper as it is addressing the same non-linear financial time series problem, using neural networks and MTL, and yielded positive results. In this paper the neural network architecture was a basic feed forward Artificial Neural Network (ANN), which should theoretically perform worse on time-series data than the proposed LSTM model.

"An Overview of Multi-Task Learning in Deep Neural Networks" is a survey of MTL specifically within deep learning[31]. It is most useful in discussing the nuances of parameter sharing and detailing state-of-the-art MTL deep learning architectures. This paper notes that MTL is effective within deep learning due to it's effect of "implicit data augmentation".

4.6 Libraries

As this is a data science/application focused project it is important to leverage the best suited libraries and adapt them for this use case. Aside from the ubiquitous numpy, pandas and matplotlib the following libraries have been highlighted as potentially useful:

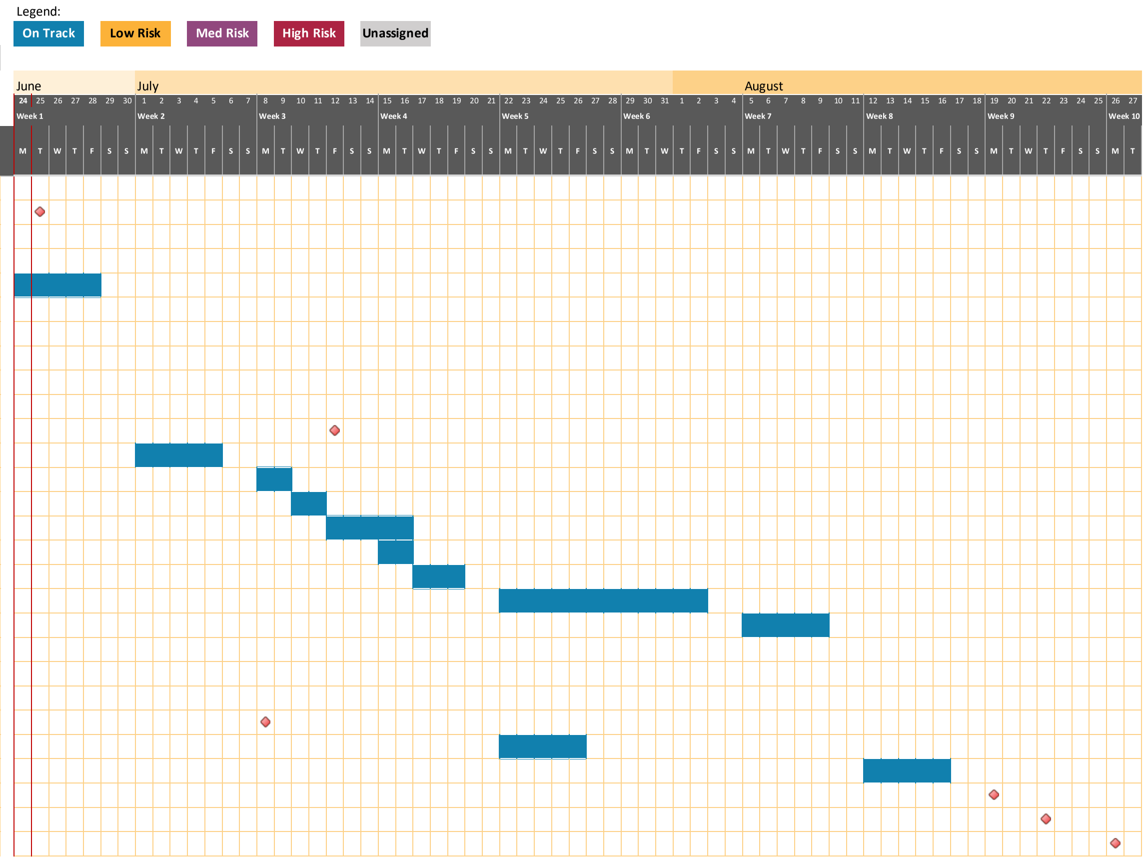
(1) Statsmodel: This will be primarily used in ARIMA modelling. Has useful functions for determining stationarity and an ARIMA module.
(2) FbProphet: Facebooks seasonality regression package that will provide the additive modelling regression benchmark. A time-series module that could outperform ARIMA if the signal has underlying seasonal components.
(3) GPy: A well documented Gaussian Process regression library. Has extensive documentation, a wide variety of models and good visualization tools[32].
(4) GPFlow: A Gaussian Process library built on Tensorflow. If there are challenges regarding computational power then this could be a more scalable option than GPy[33].
(5) Keras: Has significant support for recurrent neural networks with LSTM layer implementations[34].
(6) Pytorch: Also has significant support for recurrent neural networks with LSTM layer implementations, but at a lower level [34]. Might be required for MTL due to the additional control. Further analysis is required to choose between the Pytorch and Keras packages. It would be unnecessary to develop the LSTM networks on a lower level package than is required if there is no clear benefit in performance.

Fig. 1. Proposed Project Schedule

23/06/2019

1

Description	Category	Progress	Start	No. Days
Project Plan				
Project Hand In	Goal	100%	25/6/19	1
Project Outline				
Lit Review	On Track	100%	1/6/19	28
High Level Objectives	On Track	100%	3/6/19	3
Requirements Analysis	On Track	100%	12/6/19	5
High Level Design	On Track	100%	17/6/19	7
Software				
Code Inspection	Goal	0%	12/7/19	1
Preprocessing	On Track	10%	1/7/19	5
Additive Model	On Track	0%	8/7/19	2
ARIMA	On Track	10%	10/7/19	2
Univariate GP	On Track	0%	12/7/19	5
Mutli Variate GP	On Track	0%	15/7/19	2
LSTM AR	On Track	0%	17/7/19	3
LSTM MISO	On Track	0%	22/7/19	12
LSTM MIMO	On Track	0%	5/8/19	5
Report				
Lit Review Draft	Goal	0%	8/7/19	1
Report + Lit Review	On Track	0%	22/7/19	5
Report Write	On Track	0%	12/8/19	5
First Draft Report	Goal	0%	19/8/19	1
Second Draft Report	Goal	0%	22/8/19	1
Final Draft Report	Goal	0%	26/8/19	1



REFERENCES

[1] Isabel Vansteenkiste. How important are common factors in driving commodity prices? *IMF Working Paper Series*, 2009.

[2] Christian Pierdzioch, Jan Christoph Rülke, and Georg Stadtmann. Forecasting metal prices: Do forecasters herd? *Journal of Banking and Finance*, 37(1):150–158, 2013.

[3] Gillian Dooley and Helena Lenihan. An assessment of time series methods in metal price forecasting. *Resources Policy*, 2005.

[4] Valtteri Ahti. Forecasting commodity prices with nonlinear models. *Helsinki Centre of Economic Research*, 17(268), 2009.

[5] Pablo M. Pincheira and Nicolas Hardy. Forecasting Base Metal Prices with Commodity Currencies. *Ssrn*, 2018.

[6] W. Labys. *Modeling and Forecasting Primary Commodity Prices*. Ashgate, 2006.

[7] C. A. Tapia Cortez, S. Saydam, J. Coulton, and C. Sammut. Alternative techniques for forecasting mineral commodity prices. *International Journal of Mining Science and Technology*, 2018.

[8] Sangyeon Kim and Myungjoo Kang. Financial series prediction using Attention LSTM. 2019.

[9] Hochreiter Sepp and Schmidhuber J?urgen. Long Short Term Memory. *Neural Computation*, 9(8):1–32, 1997.

[10] Joseph P Byrne, Giorgio Fazio, and Norbert Fiess. Comovements , Common Factors and Fundamentals. *World Bank Policy Research Working Paper*, (February):1–33, 2011.

[11] Kim Han Thung and Chong Yaw Wee. A brief review on multi-task learning. *Multimedia Tools and Applications*, 77(22):29705–29725, 2018.

[12] Yu Zhang and Qiang Yang. A Survey on Multi-Task Learning. pages 1–20, 2017.

[13] Robert S. Hudson and Andros Gregoriou. Calculating and comparing security returns is harder than you think: A comparison between logarithmic and simple returns. *International Review of Financial Analysis*, 38(0):151–162, 2014.

[14] Alexei Botchkarev. Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology. pages 1–37, 2018.

[15] Georgios Drakos. Selecting Evaluation Metrics for Machine Learning Models, 2018.

[16] Hélyette Geman and William O. Smith. Theory of storage, inventory and volatility in the LME base metals. *Resources Policy*, 38(1):18–28, 2013.

[17] Shuhao Zhang. Copper & Aluminium Price Forecasting. 2019.

[18] Siddharth Chatterjee. Predicting Commodity Futures Prices Using Online Learning. 2018.

[19] A and Rossen. What are metal prices like? Co-movement, price cycles and long-run trends. 2014.

[20] Thomas Kriechbaumer, Andrew Angus, David Parsons, and Monica Rivas Casado. An improved wavelet–ARIMA approach for forecasting metal prices. *Resources Policy*, 39:32–41, 3 2014.

[21] Andrew McHutchon. Nonlinear Modelling and Control using Gaussian Processes. *Thesis*, (August), 2014.

[22] Robert S Pindyck and Julio J Rotemberg. The Excess Co-Movement of Commodity Prices. *The Economic Journal,*, 100(403):1173–1189, 2008.

[23] G P Nason. *Stationary and non-stationary time series*. Number 1994. 2003.

[24] Ratnadip Adhikari. *An Introductory Study on Time Series Modeling and Forecasting*, volume 1302.6613. LAP Lambert Academic Publishing, 2013.

[25] S. Roberts, M. Osborne, M. Ebden, S. Reece, N. Gibson, and S. Aigrain. Gaussian processes for time-series modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984):1–27, 2013.

[26] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*, volume 11. 2006.

[27] Kian Ming Adam Chai. Multi-task learning with Gaussian processes. *PQDT - UK & Ireland*, page 1, 2010.

[28] Yuxiu Hua, Zhifeng Zhao, Rongpeng Li, Xianfu Chen, Zhiming Liu, and Honggang Zhang. Deep Learning with Long Short-Term Memory for Time Series Prediction. *IEEE Communications Magazine*, pages 1–9, 2019.

[29] Yu Zhang, Pengyuan Zhang, and Yonghong Yan. Attention-based LSTM with multi-task learning for distant speech recognition. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2017-Augus:3857–3861, 2017.

[30] J Ghosn and Yoshua Bengio. Multi-Task Learning for Stock Selection. *Conference on Neural Information Processing Systems, NIPS 1996*, pages 946–952, 1996.

[31] Sebastian Ruder. An Overview of Multi-Task Learning in Deep Neural Networks. (May), 2017.

[32] Sheffield ML Department. GPy Homepage, 2015.

[33] Alexander G. de G. Matthews, Mark van der Wilk, Tom Nickson, Keisuke Fujii, Alexis Boukouvalas, Pablo León-Villagrá, Zoubin Ghahramani, and James Hensman. GPflow: A Gaussian process library using TensorFlow. *Journal of Machine Learning Research*, pages 1–6, 2017.

[34] Franois Chollet et al. Keras Documentation: Recurrent Layers. 2015.