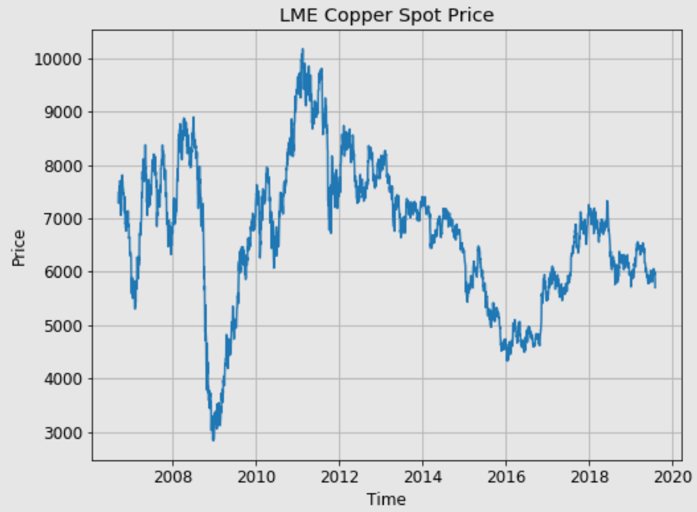
Investigating the predictive ability of LSTMs on industrial metal prices

Oliver Boom

Imperial College London

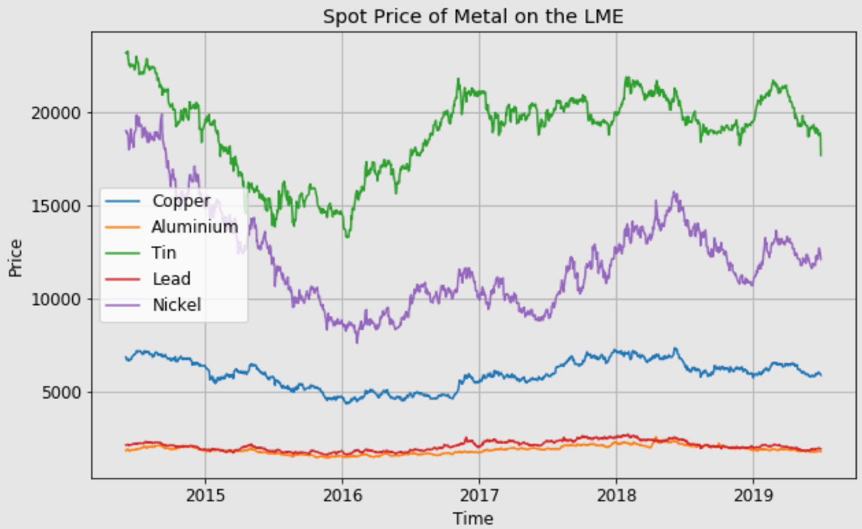
Introduction

- Highly desired information
- Price forecasting is challenging → Non-linear, volatile, dynamic
- Traditional techniques limited → ML can address challenges
- Long Short-Term Memory networks (LSTMs) achieve best-in-class results on other sequence predictions problems.
- Limited data, exogenous driving factors. Are LSTMs relevant for this context?



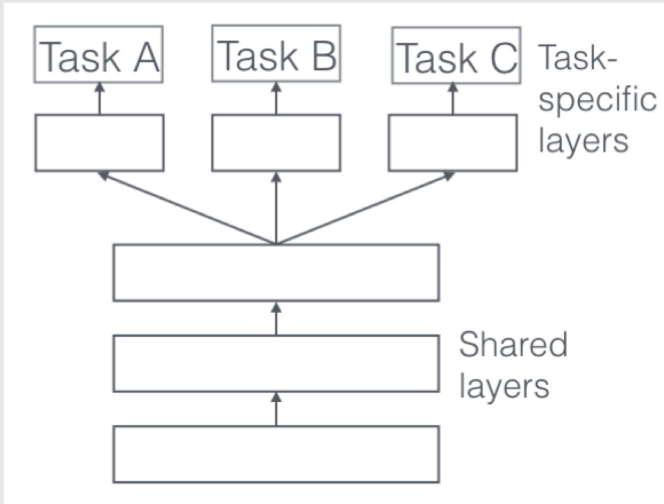
Co-movement of commodities

- Common Factor + Idiosyncratic factor
- Excessive co-movement of prices
- Commodity sub-universe → Industrial Metals



Multi-Task Learning

- Multi-Task Learning: A machine learning framework which aims to leverage shared information between a collection of related tasks to improve the predictive performance in all tasks
- Individual Tasks → Price of individual metals
- Shared layers → Neural network architecture

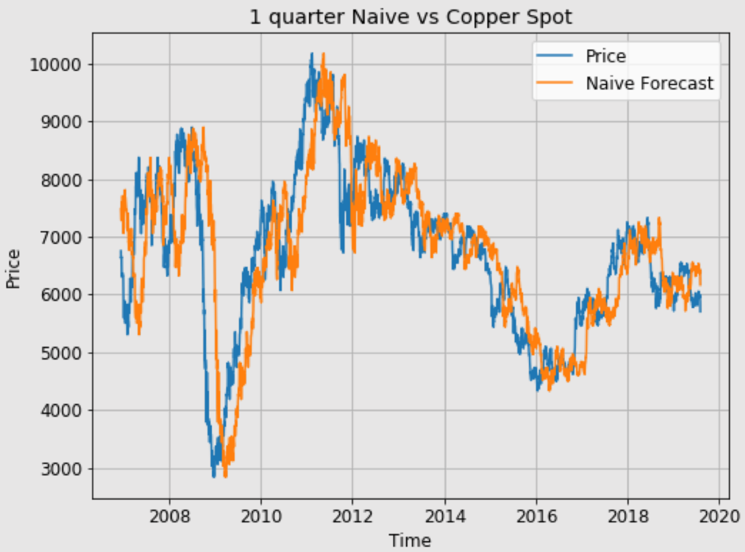


Benchmarks

- ARIMA forecast - Decomposes signal into different temporal components

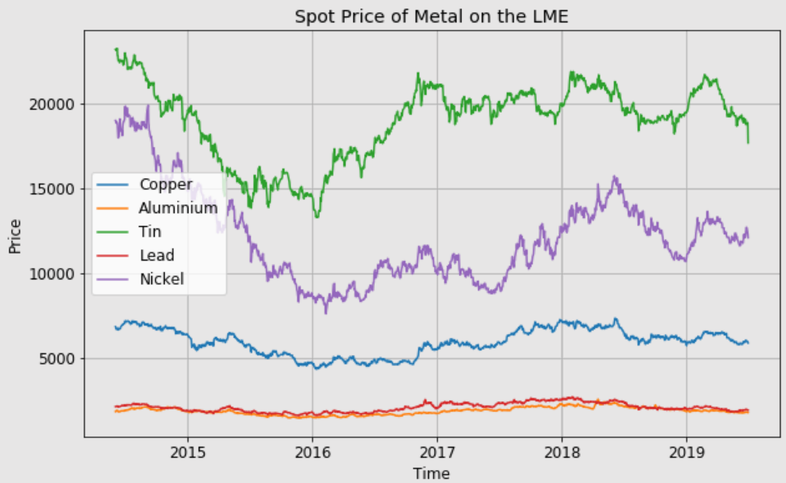
- ARIMA
  - Autoregressive Model → Target depends linearly on it's own previous values
  - Moving Average Model → Uses the past errors as explanatory variables in the model

- Naïve forecast - Take the price today as the price for the next forecast length



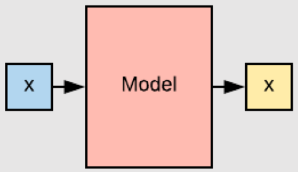
Objectives

- Do LSTMs outperform ARIMA and naïve forecasts?
- Does Multi-Task Leaning outperform single task?
- Does a multi-variate outperform autoregressive?
- Foresight is a collection of tools built to investigate the predictive ability of LSTMs on industrial metal prices

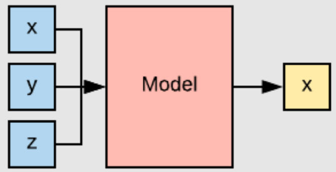


Regression Frameworks

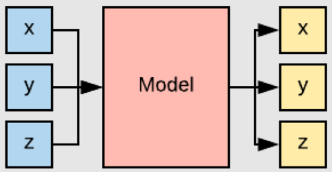
Single Input Single Output
Autoregressive



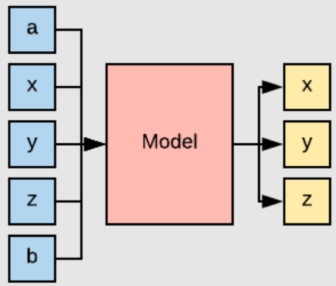
Multi Input Single Output
Multivariate



Multi Input Multi Output
Autoregressive



Multi Input Multi Output
Multivariate



Evaluation Metrics

- Forecast lengths: 1 Week, 1 Month, 1 Quarter, ½ year

- Commodity Selected: Copper
- Multi Task Learning Complex: Copper, Aluminium, Tin, Lead, Nickel

- Mean Squared Error (MSE) → Higher impact of outliers
- Mean Absolute Error (MAE) → More robust to outliers
- Mean Directional Accuracy (MDA)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 \qquad MAE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)$$

$$MDA = \frac{1}{n}\sum_{i=1}^{n}(sign(Y_i - Y_{i-1}) == sign(\widehat{Y}_i - \widehat{Y}_{i-1}))$$

Preprocessing

- Normalization → Log returns vs Scaling

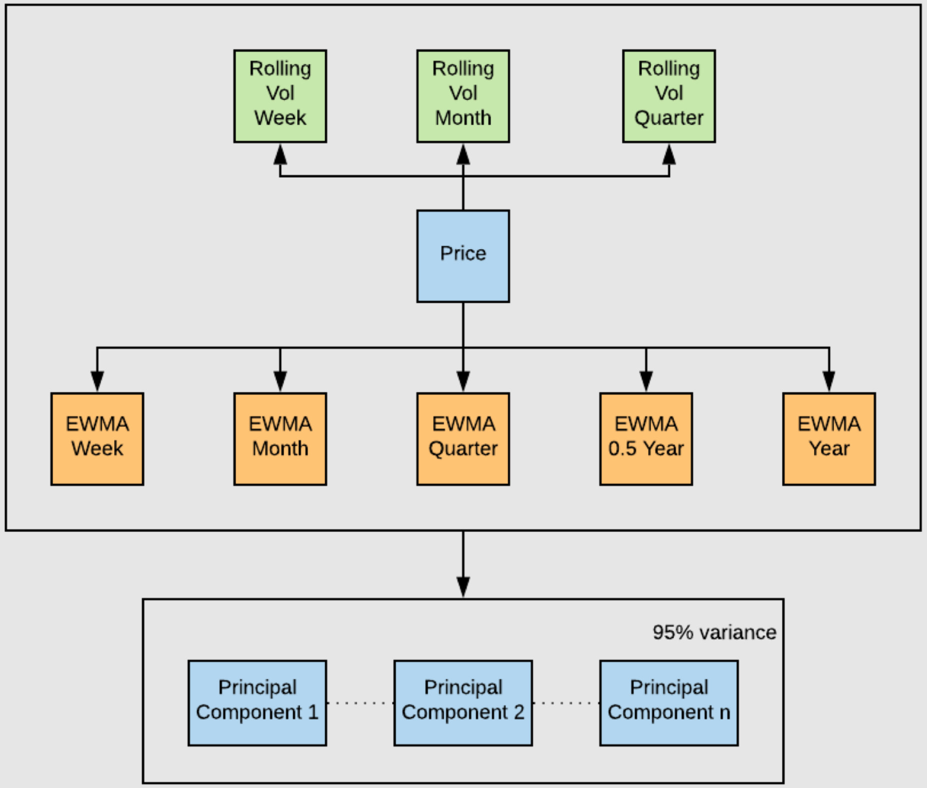
$$\rightarrow LR = Log\left(\frac{Y_i}{Y_{i-1}}\right)$$

- Feature Engineering
  - Feature Spawning

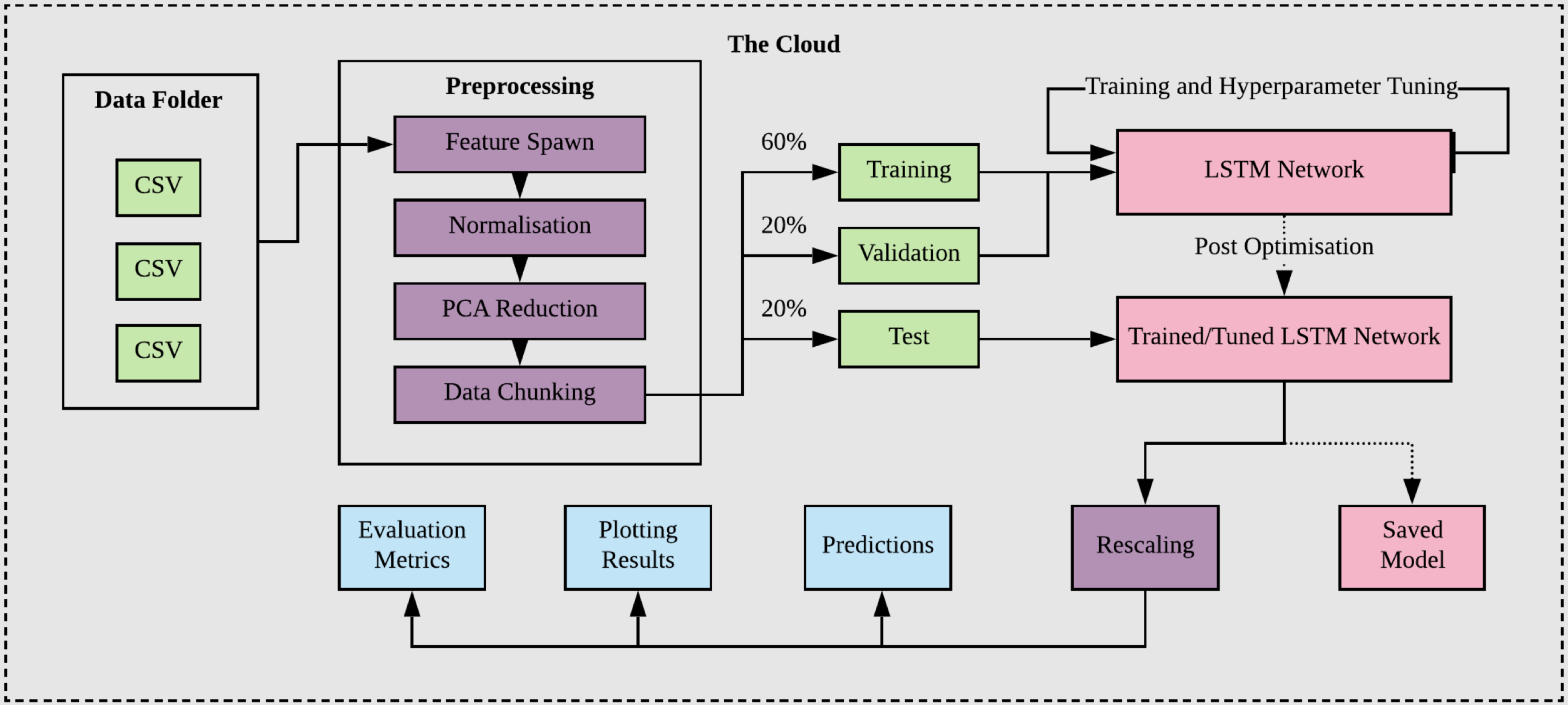
  → Exponentially weighted moving average

  → Rolling volatility

- PCA → Do LSTMs improve of suffer from the extra variance.



Architecture



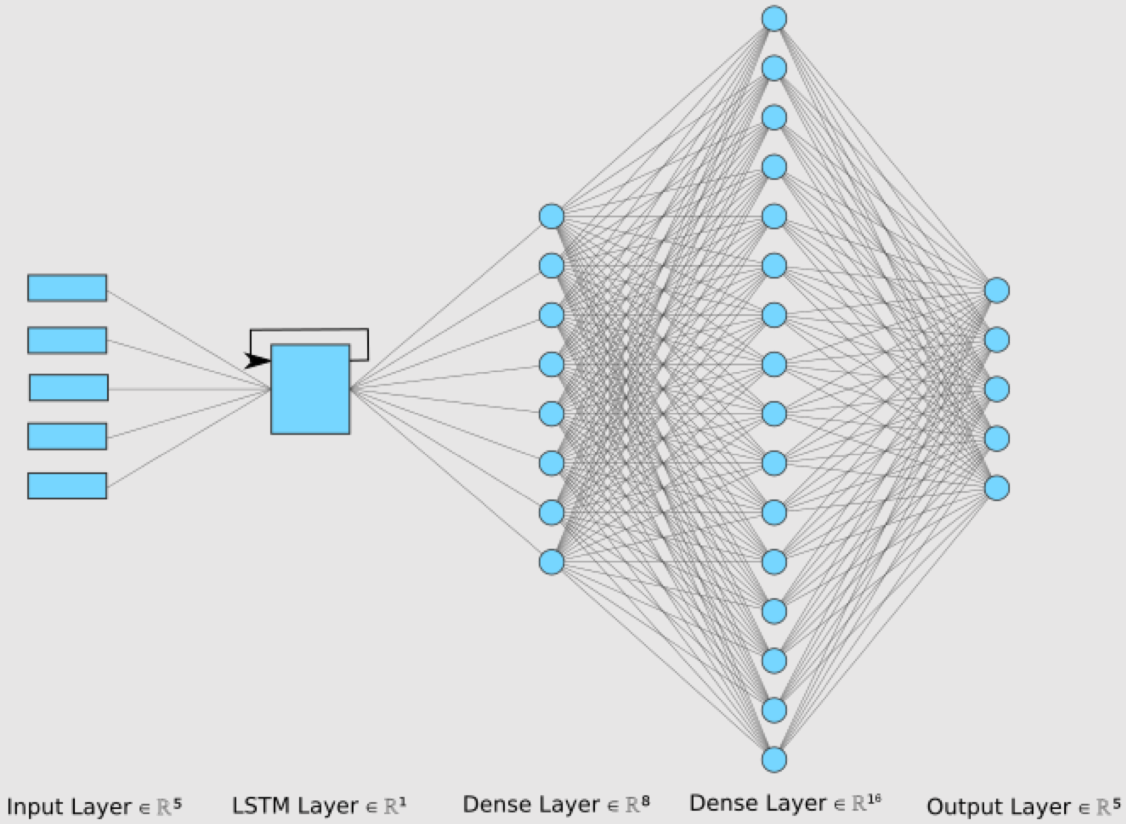
Neural Architecture

- Hyperparameter Tuning

→ Begin with simplest model and add complexity

- Window size of inputs

→ How far back is relevant for memory



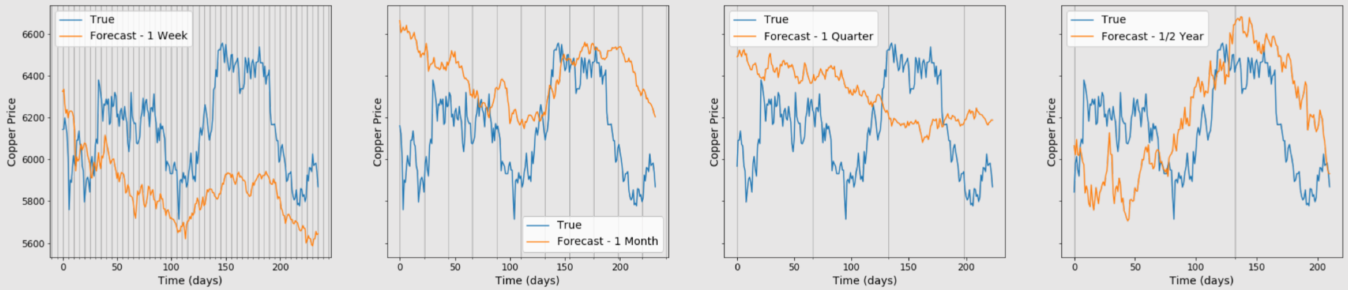
Results

Ranked for each metric on each forecast length and summed the ranking number

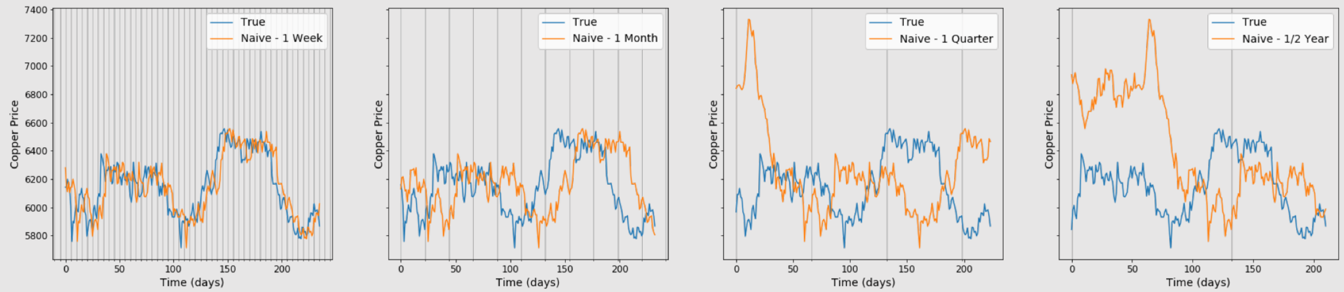
→ Low score = Highly ranked

Framework	MAE/MSE Only	All
AR MTL	27	36
MV MTL	24	37
Naïve	17	39
ARIMA	26	40
AR	34	46
MV	40	53

MV MTL LSTM Predictions



Naïve Predictions



What has been learnt?

- Across all metrics LSTMs outperformed benchmarks, but can be framed in different ways (regression only, cumulative sum etc.)
- Multi task is effective over longer time frames
- Multivariate outperforms autoregressive but only in the multi task case.
- Feature spawn not effective for LSTMs but PCA is
- Scaling shown to be more effective than the naturally transforming log returns
- Optimal window size:
  - Short term (1 week-1 month) → ½ a year
  - Long term (1 month to half a year) → 1 year

Questions