Deliverable 2

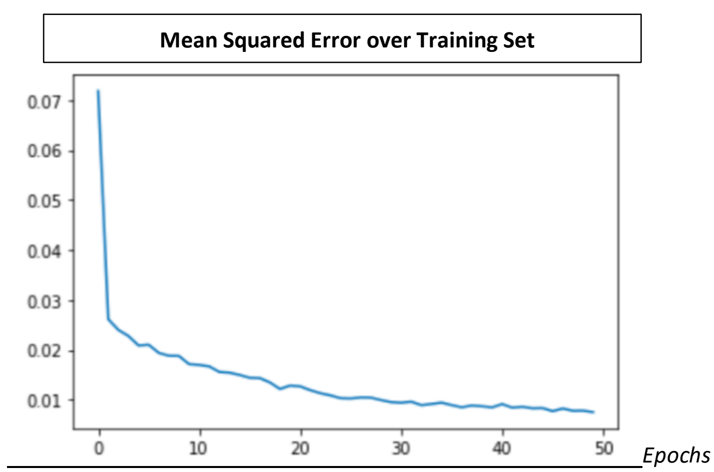
Our project consists in using a LSTM model to forecast CPI inflation growth rates 12 periods ahead. We would like to build a model that can capitalize on past inflation dynamics to predict future trends.

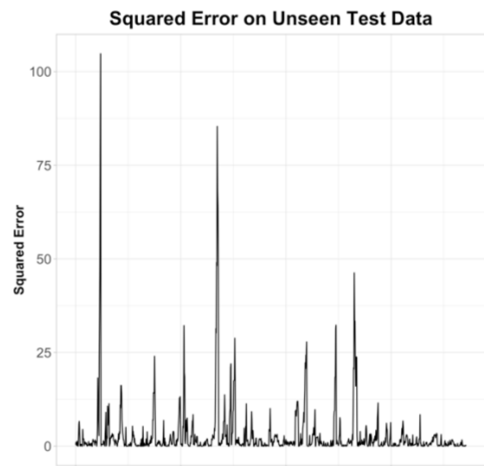
We will be working with CPI (all items) from Bloomberg, at YoY % Growth Rate units to obtain stationarity in our data. To augment the accuracy of our forecasts, we will also consider exogenous explanatory variables such as WTI oil prices, unemployment YoY % growth rates, all of which are available through Bloomberg (Deliverable 3). We use monthly data starting from 1940 up to August 2021 (around 900 observations).

Regarding data preprocessing, we scaled all of our data using the MinMaxScaler() method in Sklearn. We omitted the last 20 observations of CPI for our out-of-sample forecast and used 10 % of the remaining data for testing. Also, inputs and outputs (each of length L = 20 and H = 12 respectively) were shuffled randomly in time for both the test and training data. Doing so applies a restriction on our model: short term inflation dynamics are invariant through time and could avoid overfitting issues. We plan to test if the model performs better if we drop this assumption (Deliverable 3).

To obtain the best set of hyperparameter values, we applied k = 5 cross-fold validation over our training data and use the average set of MSE's as our metric of performance. While the ML algorithm is non-deterministic and results may vary slightly each time, we obtain our best model with 50 epochs, a dropout rate of 0.2 and 75 hidden units.

Preliminary Results:





<u>Code Access:</u>

You can access the code via the following link:
https://colab.research.google.com/drive/1MXZjgz4xaRx-46owGS4zVSqbLY8ZKHSR#scrollTo=KSWOCSgjRn2V

<u>Looking Forward</u>:

Here is a list of the ideas we would like to explore in more depth:

1. Test whether our assumption of shuffling data at random (with respect to time) avoids overfitting and increases model accuracy.
2. Vary the Lag length L (test L = 12 and L = 24) to see if the model achieves higher accuracy.
3. Include exogenous variables such as WTI, unemployment, (and potentially US GDP and the Federal Funds Rate).
4. Data Augmentation to increase our sample size. We do so by creating noise in the data that simulates inflation trends.
5. Finally, we could also consider other features in the LSTM architecture (have not given much thought to that yet however).
6. Compare the efficiency of our models to alternative time-series methods (such as (V)AR(X)) widely used in econometrics.