BAS 475: Final Project

Jack Hutchins, Jack Kidder, Johnathan Williams, Cecil Walsh

The Problem

The supply chain of the Empire is under attack by evil rebels. During the battle, key members of our analytics team were taken captive and can no longer perform their galactic duties.

The fate of the Empire hinges upon our ability to accurately predict our monthly Imperial Credit.

With each month our Empire grows stronger and our credits have seen a steady increase monthly. Enough with the analytics talk. Get working so we can catch those Rebels!

As a result our team was selected to choose an compare different forecasting models in order to choose the best one for predicting the Empire’s future Imperial Credit balance.

The models selected for analysis were the ARIMA model, ETS model and Long-Short term memory model.

Data Preparation

The data received required preparation before it could be analyzed, most notably that the data was in reverse time order.

We know this because we were explicitly told that credit usage has been growing steadily.

As such, our first step in forecasting our credit budget for the coming months was to assign each data point to a month in descending order, this allowed us to properly analyze our data.

Data Preparation cont.

	ï..credit_in_millions
1	1.9419
2	1.8006
3	1.8962
4	1.8983
5	1.9097
6	1.8099
7	1.8734
8	1.9049
9	1.9117
10	1.8806

	credits	month
1	1.9419	492
2	1.8006	491
3	1.8962	490
4	1.8983	489
5	1.9097	488
6	1.8099	487
7	1.8734	486
8	1.9049	485
9	1.9117	484
10	1.8806	483

How the Models Were Compared

- To compare the models, we used the last 12 time steps of the data as holdout data.
- This data is not used to train the model, but is used to test its accuracy. We then used the most accurate model to forecast the next 12 time steps.
- The models were compared on root-mean squared error (RMSE) and mean-absolute percentage error (MAPE). RMSE is our primary metric, since that is how our results will be compared to others in the class. MAPE is used for explainability, since it is much easier to explain than RMSE.

The ARIMA Model

ARIMA is a model used in forecasting that uses its own lagged values in order to predict future values. It is most often used on non-seasonal data, but can be very easily modified to work on seasonal data as well. A variety of factors are taken into account when creating an ARIMA model most of which revolve around making the data as stationary as possible and limiting the model's error through use of lag and regression.

AR-I-MA

(Autoregression) - The number of lags to use in predicting credits

(Integrated) - The number of differences required to make the series stationary

(Moving Average) - regression built from lagged deviations from average

ARIMA Results

Credits ~ pdq(0,1,1)

This is the model we chose to predict future credits, in simpler terms we found no significant use for auto-regression in our analysis, but we did find significant reason to perform differencing and take lag into account when creating our model.

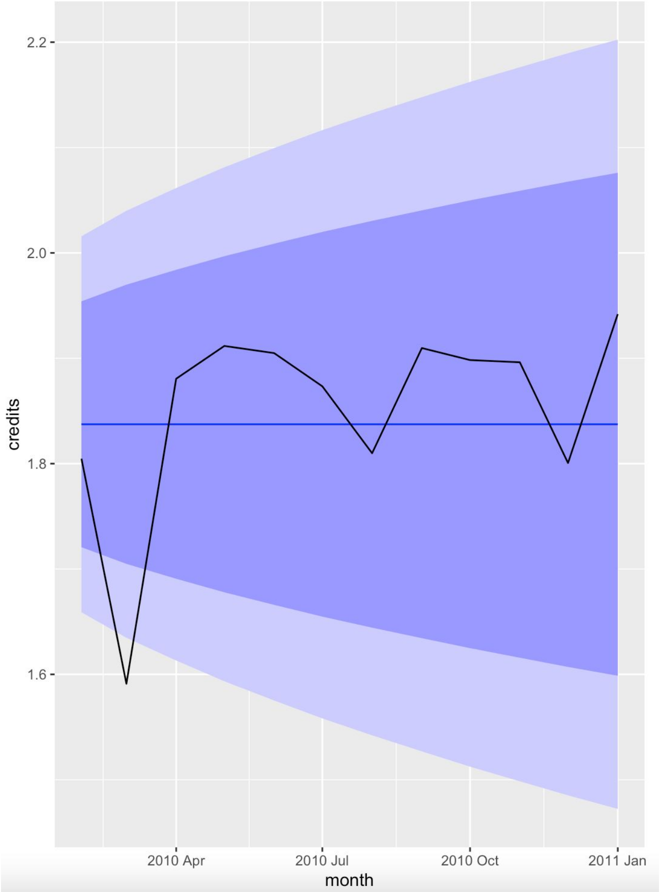
Our model produced an RMSE value of .0915 when compared against the holdout, and produced the smallest mean absolute percentage error of 3.9%

The ETS Model

The ETS Model is an exponential smoothing model that combines error (E), trend (T), and seasonal (S) components. The model decides whether to account for each of the components using additive, multiplicative, or not at all.

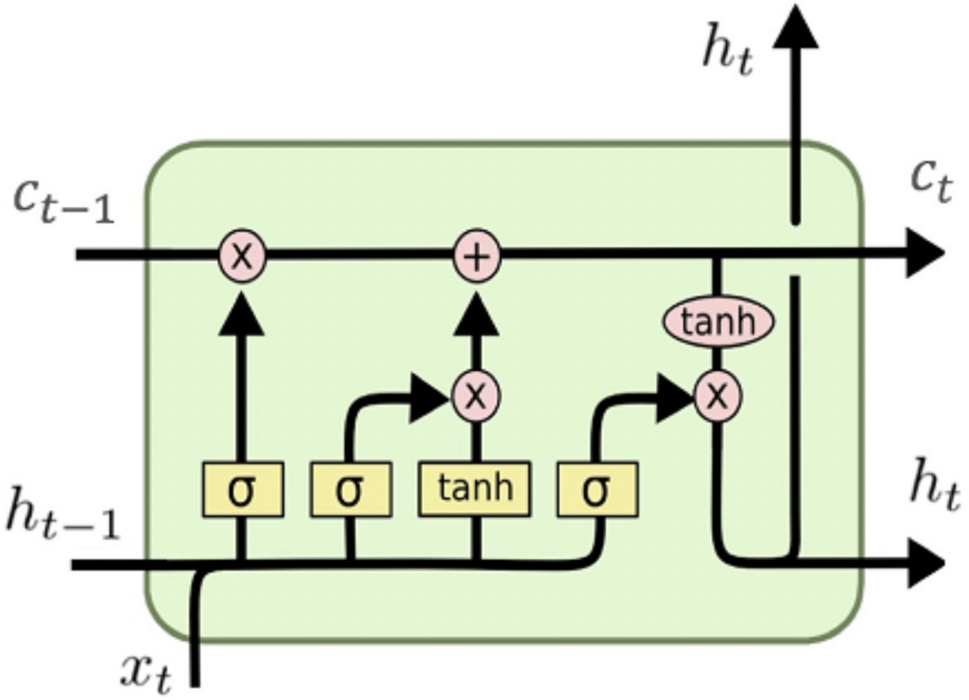
ETS Results

Our model uses additive errors and did not account for trend or seasonality (ANN). It had an alpha parameter of 0.53, meaning it gives much more weight to recent data points. Our results were very similar to those of the ARIMA model.



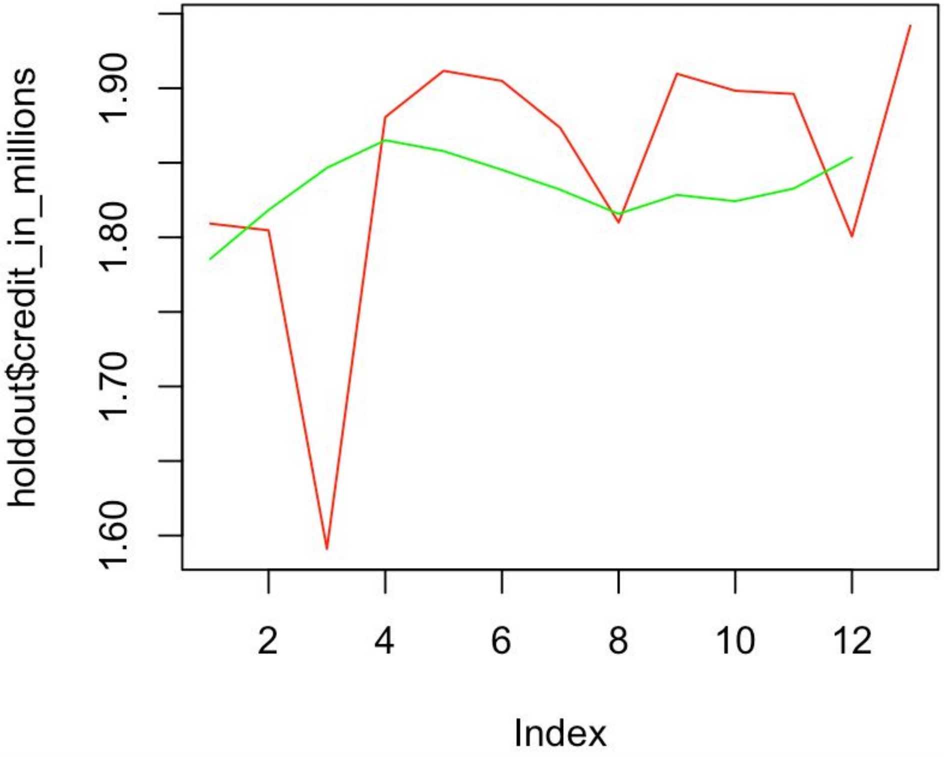
The LSTM Model

The Long Short Term Memory (LSTM) model is a recurrent neural network (RNN) with improvements over a vanilla RNN that improves its performance with long-term dependencies and seasonality. LSTM and RNN models are specifically designed for sequential data, which makes them perform much better than standard neural networks for time series forecasting.



LSTM Results

To the right is the LSTM model's forecast vs the holdout data. The model performs relatively well, only being off by around 4.14% on average. The final iteration of the LSTM model takes the last 24 values as input to predict the next value with a slight drift added to the output. The Model also takes the month as input to hopefully capture seasonal trends in the data



Results Comparison

Model	RMSE	MA-Percentage Error
ARIMA	0.0915	**3.90%**
ETS	0.0915	7.17%
LSTM	**0.0849**	4.14%

Conclusion

We have decided to go forward with the LSTM model as it had the lowest RMSE value when compared against the holdout, this means that on average our LSTM model is making smaller errors when predicting future forecast than those of ARIMA and ETS.

It should be noted, however, that ARIMA had the lowest reported value of mean absolute percentage error, but we felt the difference in RMSE values was significant enough to warrant choosing the LSTM model over its competitors.

Thus, we have gone forward and used the LSTM model to produce our final prediction of the coming months.

Final Prediction