Deliverable 3
Matthieu Fisher & Steve Wen

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.

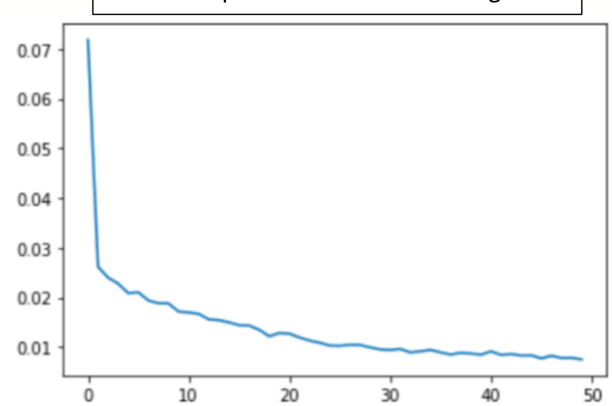
We will be working with CPI (all items) from Bloomberg, at YoY % Growth Rate units to obtain stationarity in our data. To augment the accuracy of our forecasts, we will also consider exogenous explanatory variables such as WTI oil prices and unemployment rates, all of which are available through Bloomberg. We use monthly data starting from 1982 up to August 2021 for our multivariate model, and data starting at 1940 for our univariate model.

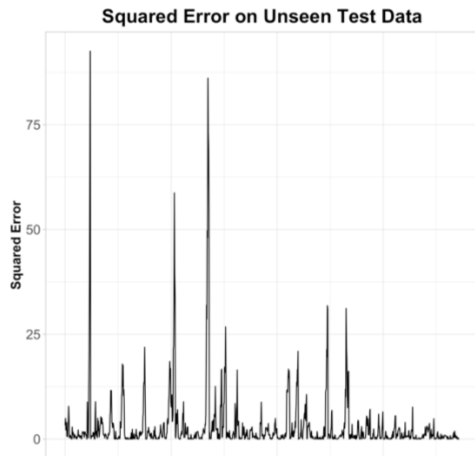
Regarding data preprocessing, we scaled all of our data using the MinMaxScaler() method in Sklearn. We omitted the last 20 observations of CPI for our out-of-sample forecast and used 10 % of the remaining data for testing.

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.

The first counterfactual experiment we test is to assess whether inflation dynamics in the short run are time invariant. At first, we chose to shuffle our input data (which was ordered sequentially through time) before feeding it into the model. We tuned our hyperparameters to obtain the best performing model (on the validation set), which sets the number of epochs at 50, the number of hidden units at 75, and a dropout rate of 0.2. The results shown in Deliverable 2 are shown below. We obtain a MSE of the test data of 3.56 and squared errors bounded reaching close to 100 units. These metrics indicate that there is definitely room for improvement.

Mean Squared Error over Training Set

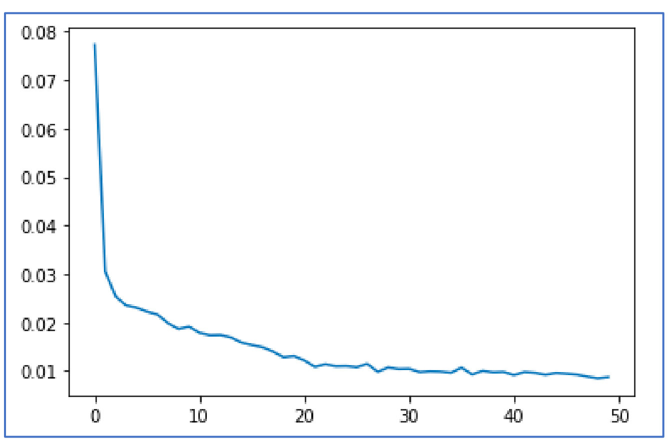


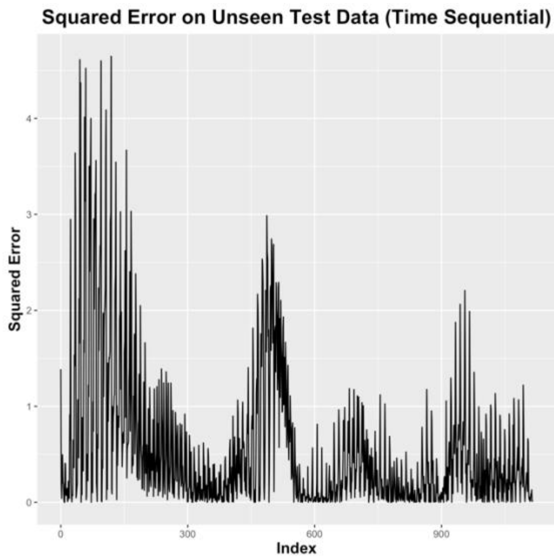


Mean Squared Error = 3.536

However, once we drop this assumption by feeding the inputs in a time-sequential order, we find that the accuracy on our test data is much improved. In a model containing 75 hidden units, 50 epochs and a dropout rate of 0.2 (chosen after applying k=5 cross-fold validation), we find that the squared errors on the test data are bounded above by 6 units (as opposed to 100, when we fed the data in a non-time-sequential manner). The mean squared error equal to 0.486, which is 7 times smaller than in our model above.

Mean Squared Error over Training Set
(Time Sequential Inputs)



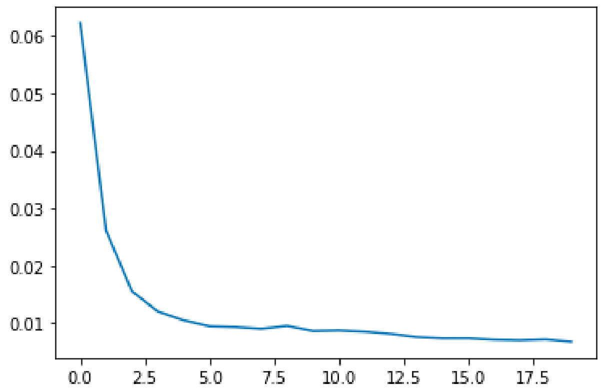


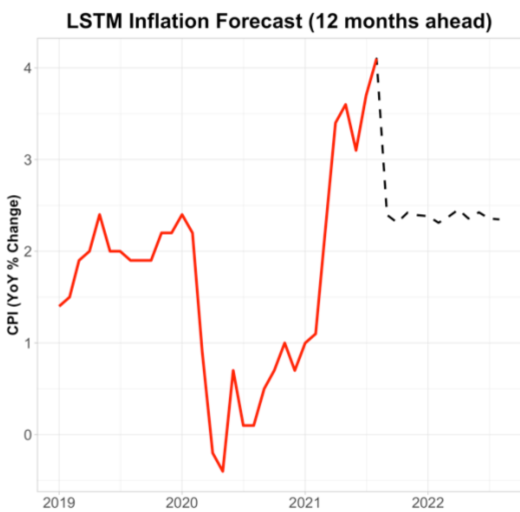
Mean Squared Error = 0.486

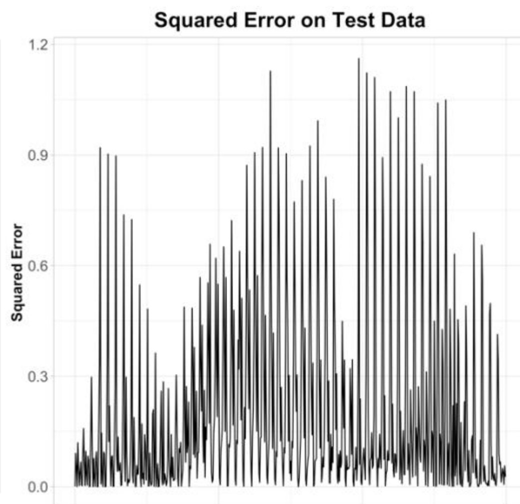
Based on those results, it makes sense to feed the inputs into our model sequentially through time. We also recognize that variations in the price level are sensitive to shocks in other macroeconomic variables. The traditional Phillips Curve empirically shows a strong negative relationship between the unemployment rate and inflation. On the other hand, changes in oil prices, a major component of Canadian exports, is highly (positively) correlated to price levels. In that line of thought, we now chose to include historical data on the WTI (oil prices) and Canadian unemployment rate into our forecast on inflation and assess the potential increase in the model's performance.

The model performs very well on the test data. Setting the hyperparameters to the following: {*number of epochs: 20; dropout rate: 0.2; hidden units : 75},* we obtain an mean squared error on the test data of 0.195 and squared errors being bounded above by 1.2 units. The mean results and our out of sample forecast are depicted below:

Mean Squared Error over Training Set
(Time Sequential Inputs)







Mean Squared Error = 0.195

While the multivariate model yields promising results on the unseen test data compared to our 2 previous assessments, the out-of-sample forecast is somewhat disappointing. Inflation is expected to die off to around 2 % range immediately, and stay within the 1-3% target range over the next few months. Those results seem highly unrealistic and contradict the robustness of the model on the test set.

Code Access:

You can access the code via the following link:

https://colab.research.google.com/drive/1MXZjgz4xaRx-46owGS4zVSqbLY8ZKHSR#scrollTo=KSWOCSgjRn2V

Looking Forward:

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

1. Understand the forecast yielded by our multivariate model, and potentially correct any flaws in the model
2. Compare the efficiency of our models to alternative time-series methods (such as (V)AR(X)) widely used in econometrics.