Slide 2 - Problem identification

For this project we have decided to use a relatively well-researched problem so that we can focus more on the methods we apply. Hence, the goal of our project is to build a workflow for forecasting industrial production in the US, the UK, and Canada, apply it using the benchmark big macroeconomic datasets, and possibly find some spillover effects between the economies.

We decided to focus on industrial production since its forecasting is highly relevant for macroeconomic policy and for portfolio management in investments. Besides, given that we have large datasets with possible some reliable leading indicators, the task of extracting meaningful signals from such indicators seems to be a reasonable problem for machine learning application.

Slide 3 - Literature Review

The working paper by Buckman and Andreas from Bank of England inspired us for setting up our machine learning workflow. In that paper they applied similar approach to what we have done in our project but with the goal of forecasting unemployment.

Also, the sources of the benchmark macroeconomic datasets like FRED-MD for the countries we analyze are mentioned on this slide.

Slide 4 - Framework

Our baseline dataset is FRED-MD and we aim to forecast year-over-year change in industrial production 12-month ahead. We use mean absolute error as our target metric since it is more robust to outliers than root mean squared error and is better suited for target variable in percentages than mean absolute percentage error.

We applied the transformations suggested by the authors of the original paper to induce stationarity and also applied standardization based on the data available up to each datapoint.

We test six ML models and as a benchmark use random walk process.

Our training sample includes 38 years and we subdivided it so that we have an additional validation subset for hyperparameter tuning. Consequently, we have almost 20 years of data for testing.

Slide 5 - Baseline results

Given our baseline goal of 12-month ahead forecast, we have the following results. As you can see our hyperparameter tuning subset is much less volatile than testing subset with errors for all the models and random walk being substantially lower. In the test subset Gradient Boosting Regressor achieves the lowest mean absolute error. Interestingly enough, in the period that excludes the global financial crisis and the pandemic, Lasso model performs best.

Slide 6 - Performance at different horizons

When we vary our forecast horizon, we observe that random walk performs best at shorter horizons, up to 6 months. At the same time, at longer horizons, the models struggle to extract meaningful signals that would improve forecasts.

Slide 7 - Performance with richer lag structure

Testing models with different number of lagged predictors reveal that almost all models yield the best performance when given more lags. Interest observation here is that neural network struggles with too many lags, probably due to small sample.

Slide 8 - Forecast quality

Visual inspection of some of the results indicate that the models tend to have better results when forecasting economic recovery after a crisis has already taken place, like in 2010 and 2021. At the same time, the models do poorly in predicting economic downturns and the forecasts are quite rigid around the unconditional mean.

Slide 9 - Performance with PCA

We have also tested the models using principal components extracted from the data. For these purposes we had to adjust our framework slightly and ensure that the transformations both ensure continuity of the features and don’t use future data at each datapoint.

As we can see, the models do perform better when given principal components instead of all features and the optimal number of principal components is around 8.

Slide 10 - Reconnecting the PCs and raw variables: Part 1

Here we tried to find the relationship between the principal components and raw variables. Reported in the tables are the incremental changes in R squared of the linear regressions of each variable against principal components when the principal component at the top of the table is added to the regression. Based on the list of top 10 variables we have assigned to each principal component an economic concept it seems represent. Here we have manufacturing and employment, interest rate spreads, short term interest rates and housing, and corporate bond yields.

Slide 11 - Reconnecting the PCs and raw variables: Part 2

And on this slide, we have credit, inflation, stock market and labor market.

Slide 12 - Feature importance analysis

Based on such analysis and given one of the models, here it is Gradient Boosting Regressor, we estimated Shapley values to gain insight into feature importance. As you can see, interest rate spreads show the highest impact on the model output. Interestingly, we see that high values of the principal component that corresponds to interest rate spreads lead to higher forecast, which is a bit counter intuitive, since interest rate spreads rise when economy is expected to slow down. A possible explanation is that the model incorporates the idea that if an economy is close to slowdown now, 12-month ahead it will probably be in recovery.

Slide 13 - Large macroeconomic dataset for Canada

Here we applied our framework to the large macroeconomic dataset for Canada with appropriate changes to the training and testing subsets. At first, we estimated the models using principal components extracted from the data for Canada and observe that the best performance is achieved with 21 principal components. After that we augmented the feature space with the first eight principal components for the US. As you can see, for some of the models there is an improvement in forecast quality given the US data too.

Slide 14 - Large macroeconomic dataset for the UK

We applied the same approach to the large macroeconomic dataset for the UK, though here the feature space after adjusting for missing values included only 26 variables. Here we don’t find an improvement in forecast quality when adding the US data to the feature space.

Slide 15 - Conclusions

To conclude, we find that Gradient Boosting Regressor achieves the best result for 12-month ahead industrial production forecasting for the US, which is consistent with prior findings. We observe improvements of forecast quality when using principal components instead of all features and when using more lags of predictors.

We observe worse forecast quality for the UK and Canada, partly due to a more limited training and testing subsets. We have identified meaningful spillover effect from the US economy to Canadian economy but not to the UK economy.