EC4304 Group Project - Group 4

Name	Matriculation Number
Kam Keng Han	A0218546Y
Ruben Wu Fengying	A0233019N
Kwan Sin Min	A0225678R
Liew Jun Yang	A0236555Y

Introduction

How effective are we in forecasting short-term gross domestic product (GDP)? Which model will do best in forecasting the GDP of a country in the next 12 months? This paper aims to answer these questions to shed light on which models are suitable for forecasting GDP and contribute to the body of literature evaluating different models that are used to forecast GDP.

Effectively forecasting the GDP will help central banks, fiscal authorities and private sectors make better decisions (Chauvet & Potter, 2013). Thus, it is important to be able to accurately forecast GDP so that decisions made meet their intended objectives. However, forecasting problems are widespread and ever-present in all areas of economics and finance where an agent's decision is dependent on the uncertain future value of one or more variables of interest (Elliott & Timmermann, 2008). GDP forecasting is no exception to this forecasting problem and therefore, this paper contributes by helping to find out the model that performs the best when used for forecasting GDP in Canada. Specifically, we are interested in finding out if autoregressive distributed lag models will be able to outperform the benchmark autoregressive model.

In this paper, we will apply various forecasting techniques to predict the future GDP in Canada. We find that over the period of January 2010 to December 2019, across the 1-step, 3-step, 6-step and 12-step horizons, the forecast of the ADL model does not outperform the forecast of the AR model in forecasting Canada's GDP. The forecasts of the simple mean model and simple average forecast combination can outperform the forecast of the AR model for the 1-step and 3-step forecasts. The Granger-Ramanathan forecast combination was also able to outperform the forecast of the AR model for the 3-step forecast. The forecast of the AR model is not outperformed by any other forecasts used in this paper for 6-step and 12-step forecasts.

The paper is organized in this manner. The next section will present the methods followed by the data. The last section will present the results.

Methods

The present study considers the forecasting performance of the benchmark autoregressive model (AR) against autoregressive distributed lag (ADL) models. This AR model served as our initial benchmark, providing a basis for comparing with alternative models. To enhance our forecasting accuracy, we adopted Autoregressive Distributed Lag (ADL) models, which incorporate the influence of four leading economic indicators: housing starts, building permits, term spread (10-year - 3-months), and durable goods orders. These economic indicators are commonly used to predict GDP in studies (Rudebusch & Williams, 2009;

Christiansen et al., 2014; Paul, 1969). Combinations of these models are also considered: simple average and the Granger-Ramanathan approach, as combining forecasts often produces better forecast performance.

In this project, we will conduct forecasts at various intervals, including 1-step, 3-step, 6-step, and 12-step ahead predictions, and we will assess the performance of each model at these different forecast horizons. The forecasts are made using a rolling window estimation of fixed length. The exact window size varies with the forecasting horizon and the chosen number of lags. We use the rolling window scheme as Giacomini and White (2006) showed that in the case of nested models, the Diebold–Mariano (DM) test maintained its normal distribution when a rolling window framework was used. To avoid problems when comparing the predictive performance of the AR and ADL models (i.e., nested models), we use the rolling window scheme.

To assess the performance of each model, we will conduct a Mincer-Zarnowitz (MZ) test. This test is a valuable tool for evaluating the unbiasedness of our forecasts. We also check if the forecast errors are white noise (for 1-step ahead forecasts) or at most MA(h-1) (for h-step ahead forecasts). Lastly, we will employ the Diebold-Mariano (DM) test to determine whether the differences between various forecasts are statistically significant at 5%. This information will guide us in making well-informed decisions and reaching our conclusions about whether the forecasts of the comparison models can beat out the forecast of the benchmark AR(p) model.

Models

Benchmark Model

Autoregressive (AR) Model. The forecast equation for the AR model of order *p* is:

$$\hat{\pi}_{t+h|t} = \hat{\phi}_{0,h} + \hat{\phi}_{1,h}\pi_t + \ldots + \hat{\phi}_{p,h}\pi_{t-p+1}$$

The order, p, is chosen using the Akaike Information Criterion (AIC) and the parameters, $\hat{\phi}_{p,h}$, are estimated by the Ordinary Least Squares (OLS) method. As we are utilising monthly data, we consider up to 12 autoregressive lags. The model is fixed for each forecast horizon.

Comparison Models

Autoregressive Distributed Lag (ADL) Model. The forecast equation for the ADL model of order (*p*, *q*) is:

$$\hat{\pi}_{t+h|t} = \hat{\phi}_{0,h} + \hat{\phi}_{1,h}\pi_t + \ldots + \hat{\phi}_{p,h}\pi_{t-p+1} + \hat{\beta}_{1,h}x_t + \ldots + \hat{\beta}_{q,h}x_{t-q+1}$$

The order p follows the order selected for the benchmark AR(p) model. The order q, is chosen using the Akaike Information Criterion (AIC) and the parameters, $\hat{\phi}_{p,h}$ and $\hat{\beta}_{q,h}$, are estimated by the Ordinary Least Squares (OLS) method. As we are utilising monthly data, we consider up to 12 lags of the leading indicators: housing starts, building permits, term spread, and durable goods orders. We also consider the combination of the better-performing leading indicators. The model is fixed for each forecast horizon.

Simple Mean. We also consider the simple mean model by regressing GDP on a constant.

Simple Average Forecast Combination. We take the simple average of all the model predictions. In this procedure, all forecasts are given equal weight.

$$Y_t = (f_{1t} + f_{2t} + \ldots + \beta_N f_{Nt})/n$$

Granger-Ramanathan Forecast Combination. The Granger-Ramanathan procedure allows for the selection of optimal weights to combine forecasts. We employ a constrained regression: omit intercept, force non-negative betas and make the coefficients sum to 1. We then regress the actual values of the validation set on the forecasted values:

$$Y_t = \beta_1 f_{1t} + \beta_2 f_{2t} + \ldots + \beta_N f_{Nt} + e_t$$

Data

The dataset under consideration is constructed by Fortin-Gagnon et al. (2022) and includes data on Canada's monthly GDP, housing starts, building permits, term spread (10 years minus 3 months) and durable goods order. We sample the data from January 1981 (1981m1) to December 2019 (2019m12), yielding a total of 468 observations. We exclude the periods affected by the COVID-19 pandemic to prevent potential 'contamination' of COVID-19 on the evaluation of the models' relative performance. It is important to note that this dataset is already seasonally adjusted and corrected for stationarity by Fortin-Gagnon et al. (2022). The exception is for term spread, which we corrected for stationarity using the first difference. Figures 1 and 2 show the plots of GDP and leading indicators over time. As seen, the variables do not exhibit trend or seasonality.

To facilitate our analysis and evaluation, we divided the dataset into three distinct subsets: 1) training set, comprising observations from January 1981 to December 2000, encompassing 240 observations, 2) validation set, containing data from January 2001 to December 2009, which includes 108 observations, and 3) test set, covering the period from January 2010 to December 2019, consisting of 120 observations. This division of the data allows us to perform training, validation, and testing of models, which is a common and robust approach in time series analysis and forecasting.

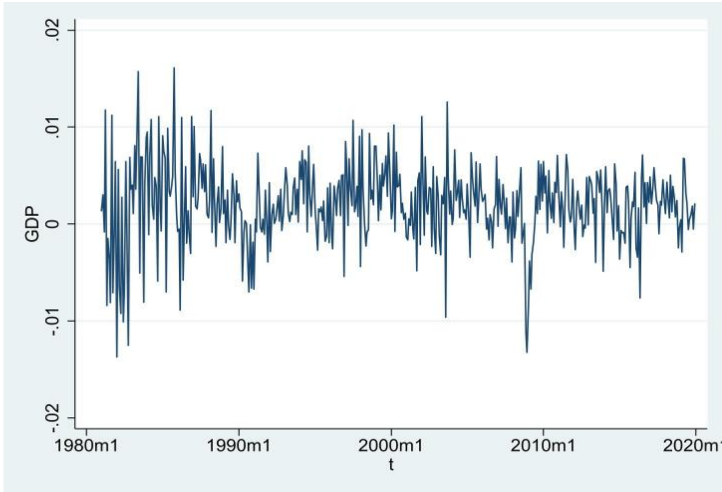

Figure 1: Plot of GDP over time

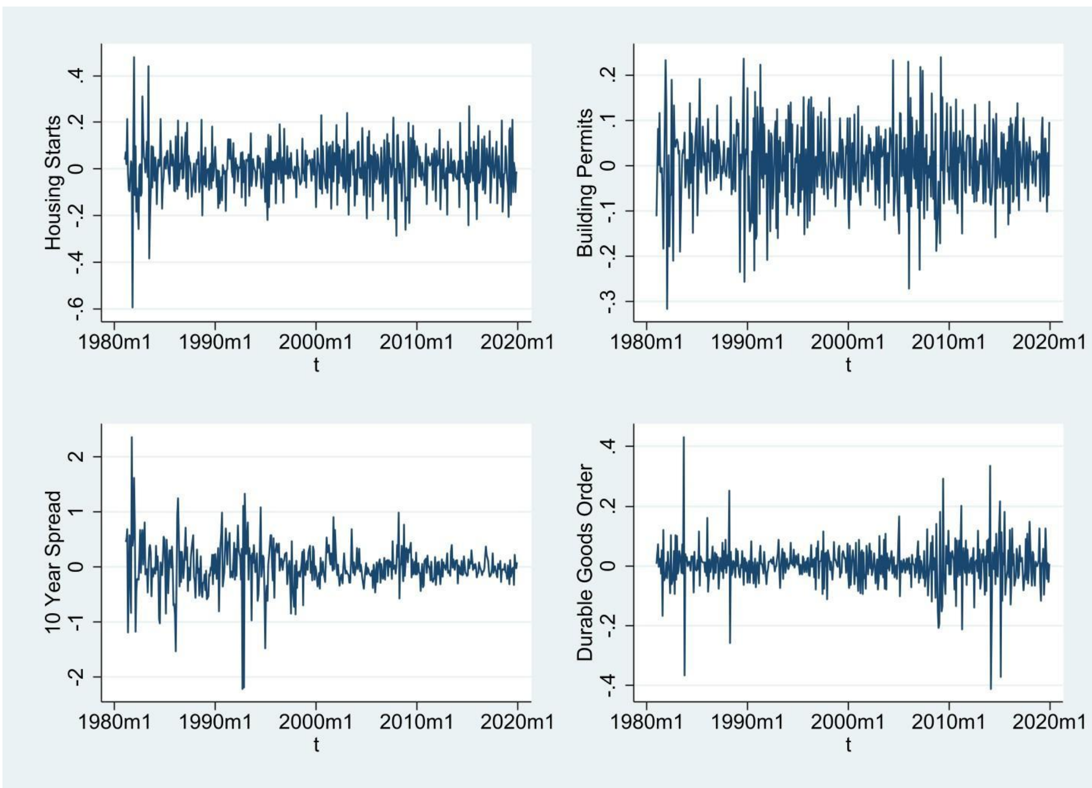

Figure 2: Plots of leading indicators over time

Results

Model Selection

To select the benchmark and comparison models, we considered up to 12 autoregressive lags and 12 lags of four leading indicators: housing starts, building permits, term spread and durable good orders. Models are chosen based on AIC.

Benchmark AR(p) Model

The AIC value for AR(3) is the lowest (-1830.07). Thus, AIC selects AR(3) as our benchmark model.

Model	**AIC**
AR(1)	-1803.069
AR(2)	-1804.979
AR(3)	**-1830.07**
AR(4)	-1828.942
AR(5)	-1827.686
AR(6)	-1826.271
AR(7)	-1824.847
AR(8)	-1822.865
AR(9)	-1821.132
AR(10)	-1819.278
AR(11)	-1817.324
AR(12)	-1823.168

Table 1: AIC values for AR(p)

Comparison ADL Models

Single Leading Indicator. AIC selects 1 ADL model: AR(3) augmented with 6 lags of building permits. For housing starts, term spread and durable goods orders, AIC selects the benchmark model augmented with 0 lags of each leading indicator. Hence, we stop considering housing starts, term spread and durable goods orders as leading indicators for our comparison models. As only 1 leading indicator is chosen, we will not consider ADL models augmented with multiple leading indicators.

Model/AIC	Housing Starts	Building Permits	Term Spread	Durable Goods Orders
AR(3)	**-1830.07**	-1830.07	**-1829.635**	**-1830.07**
ADL(3,1)	-1828.122	-1828.171	-1828.525	-1829.865
ADL(3,2)	-1828.44	-1828.908	-1827.944	-1827.869
ADL(3,3)	-1828.895	-1828.212	-1825.961	-1825.881
ADL(3,4)	-1827.049	-1826.594	-1824.208	-1824.464
ADL(3,5)	-1827.74	-1824.666	-1822.339	-1825.856
ADL(3,6)	-1827.542	**-1844.255**	-1820.355	-1823.851
ADL(3,7)	-1825.756	-1842.276	-1818.375	-1827.233
ADL(3,8)	-1826.085	-1841.029	-1816.848	-1825.892
ADL(3,9)	-1824.406	-1839.882	-1814.905	-1824.49
ADL(3,10)	-1822.969	-1838.23	-1812.949	-1823.978
ADL(3,11)	-1821	-1836.986	-1811.055	-1826.805
ADL(3,12)	-1819.024	-1835.094	-1810.432	-1825.643

Table 2: AIC values for AR(3) augmented with 0-12 lags of all 4 predictors. Model selected per predictor is highlighted in yellow

Granger-Ramanathan Forecast Combination

To achieve the Granger-Ramanathan forecast combination, we iteratively apply the aforementioned constrained regression on the forecasts from the 2 selected models and the simple mean model by eliminating the forecast with negative weight(s) after each iteration, until all weights are positive. The resulting weighted linear combinations are:

Forecasts Models	**1-Step Forecast**	**3-Step Forecast**	**6-Step Forecast**	**12-Step Forecast**
Simple mean	-	-	0.7386556	0.1581143
AR(3)	0.5005113	0.7525372	0.2613444	0.4283444
AR(3) + 6 lags of building permits	0.4994887	0.2474628	-	0.4135413

Table 3: Weights of forecasts to form Granger-Ramanathan Forecast Combination

Forecast Evaluation

We report the out-of-sample root-mean-squared errors (OOS RMSE) for the different models across the 1-, 3-, 6- and 12-step forecast horizons. We also report the p-values of the Mincer-Zarnowitz (MZ) test of unbiasedness and whether the forecast errors are white noise (for 1-step ahead forecasts) or at most MA(h-1) (for 3-, 6- and 12-step ahead forecasts). To examine if the differences in OOS RMSE are significant enough to declare one forecast as performing better than the other, we also report the p-values of the Diebold-Mariano (DM) test. In total, we consider 5 forecasts: AR(3) as the benchmark, AR(3) with 6 lags of building permits, simple mean, simple average forecast combination and Granger-Ramanathan forecast combination.

1-Step Forecast

Overall, most forecasts performed equally well as the benchmark, AR(3) model at the 5% significance level according to the DM test. More specifically, the DM test against the benchmark AR(3) model suggests that the forecasts from the simple mean and simple average model have better predictive ability. Looking at OOS RMSE alone, all other models in consideration have lower RMSE than the AR(3) model, with both the simple mean and simple average model yielding the smallest 2 RMSE, reiterating the predictive ability of the forecasts from those models. However, forecasts from all models, apart from the simple mean model, are biased at the 5% significance level according to the MZ test. Noticeably, the forecast errors cannot be regarded as white noise when looking at the ACF plot and Q-statistics as serial correlation can be observed. This thus suggests that the 1-step forecasts are not optimal under the squared loss.

Model	**OOS RMSE**	**MZ Test p-value**	**DM Test against AR(3)**
AR(3)	0.0029954	**<u>0.0000</u>**	**-**
AR(3) + 6 lags of building permits	0.0029718	**<u>0.0010</u>**	0.5945
Simple mean	0.0028172	0.3365	**<u>0.0281</u>**
Simple average	0.0028852	**<u>0.0213</u>**	**<u>0.0013</u>**
Granger-Ramanathan	0.0029671	**<u>0.0006</u>**	0.2128

Table 4: 1-step ahead results for all models: OOS RMSE, MZ test results, DM test results against AR(3) benchmark (values significant at the 5% level are bolded and underlined)

Figure 3: Plot of 1-step ahead forecasts (AR(3), AR(3) + 6 lags of building permits, Granger-Ramanathan combination, simple average combination) vs. actual values with 68% forecast intervals

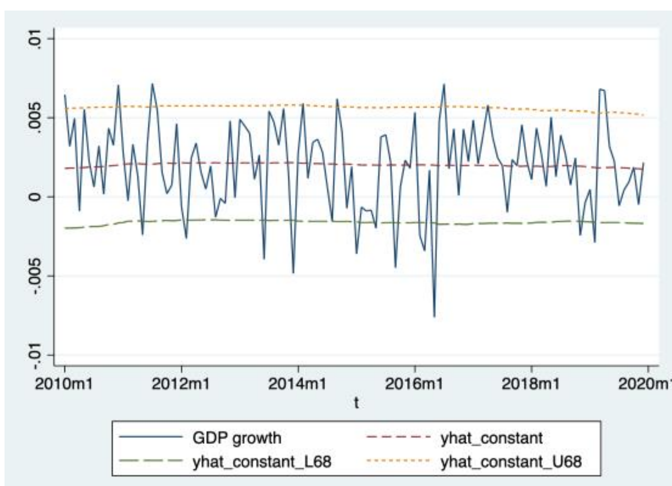

Figure 4: Plot of simple mean forecasts vs. actual values with 68% forecast intervals

3-Step Forecast

The forecasts from AR(3) augmented with 6 lags of building permits performed equally well as the benchmark, AR(3) model at the 5% significance level according to the DM test. More specifically, the DM test against the benchmark AR(3) model suggests that the forecasts from the simple mean, simple average and Granger-Ramanathan model have better predictive ability. Looking at OOS RMSE alone, all models in consideration have lower RMSE than the AR(3) model, with both the simple mean and simple average model yielding the smallest 2 RMSE, reiterating the predictive ability of the forecasts from those models. However, forecasts from all models, apart from the simple mean model, are biased at the 5% significance level according to the MZ test. It should also be noted that the forecast errors are not at most MA(2) when looking at the ACF plot and Q-statistics as serial correlation can be observed. This thus suggests that the 3-step forecasts are not optimal under the squared loss.

Model	OOS RMSE	MZ Test p-value	DM Test against AR(3)
AR(3)	0.0029617	**<u>0.0000</u>**	-
AR(3) + 6 lags of building permits	0.0029003	**<u>0.0013</u>**	0.1850
Simple mean	0.0028172	0.3942	**<u>0.0076</u>**
Simple average	0.0028549	**<u>0.0217</u>**	**<u>0.0001</u>**
Granger-Ramanathan	0.002934	**<u>0.0000</u>**	**<u>0.0195</u>**

Table 5: 3-step ahead results for all models: OOS RMSE, MZ test results, DM test results against AR(3) benchmark (values significant at the 5% level are bolded and underlined)

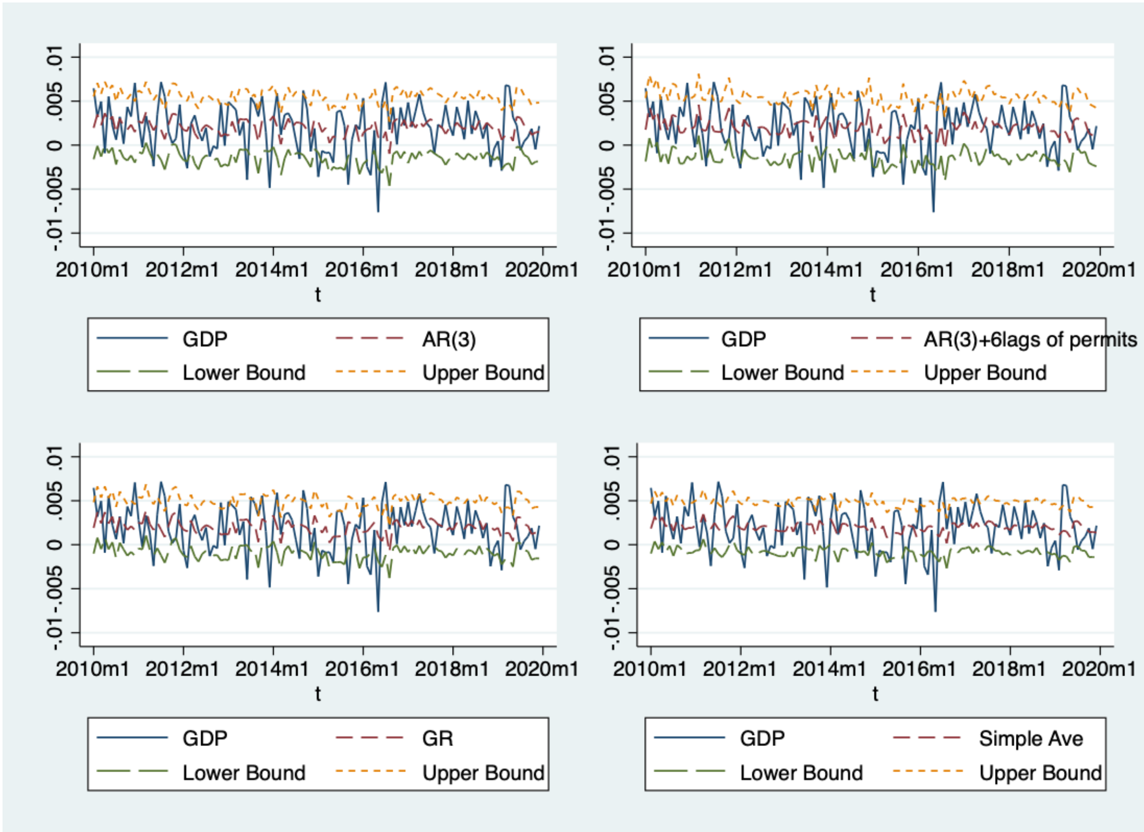

Figure 5: Plot of 3-step ahead forecasts (AR(3), AR(3) + 6 lags of building permits, Granger-Ramanathan combination, simple average combination) vs. actual values with 68% forecast intervals

6-Step Forecast

Overall, all forecasts performed equally well as the benchmark, AR(3) model at the 5% significance level according to the DM test. Looking at OOS RMSE alone, the AR(3) augmented with 6 lags of building permits and simple mean had larger values than AR(3). However, these differences are not judged to be significantly different at the 5% level by the DM test. Forecasts are unbiased at the 5% significance level according to the MZ test. Forecast errors are also at most MA(5) by looking at the ACF plot and Q-statistics. This thus suggests that the 6-step forecasts are optimal under the squared loss.

Model	OOS RMSE	MZ Test p-value	DM Test against AR(3)
AR(3)	0.0028126	0.4716	-
AR(3) + 6 lags of building permits	0.002826	0.2527	0.8082
Simple mean	0.0028172	0.2940	0.9171
Simple average	0.0027969	0.8513	0.5005
Granger-Ramanathan	0.0028007	0.9647	0.7392

Table 6: 6-step ahead results for all models: OOS RMSE, MZ test results, DM test results against AR(3) benchmark

Figure 6: Plot of 6-step ahead forecasts (AR(3), AR(3) + 6 lags of building permits, Granger-Ramanathan combination, simple average combination) vs. actual values with 68% forecast intervals

12-Step Forecast

Overall, all forecasts performed equally well as the benchmark, AR(3) model at the 5% significance level according to the DM test. Looking at OOS RMSE alone, only the AR(3) augmented with 6 lags of building permits and the Granger-Ramanathan forecast combination had larger values than AR(3). However, these differences are not judged to be significantly different at the 5% level by the DM test. Forecasts are unbiased at the 5% significance level according to the MZ test. The only exception is the forecast errors for the ADL(3,6) model with building permits which has a p-value = .0081 < 0.05. Forecast errors are also at most MA(11) by looking at the ACF plot and Q-statistics. This thus suggests that the 12-step forecasts, except for the forecast yielded by the AR(3) with 6 lags of building permits, are optimal under the squared loss.

Model	**OOS RMSE**	**MZ Test p-value**	**DM Test against AR(3)**
AR(3)	0.002851	0.0533	-
AR(3) + 6 lags of building permits	0.0029093	<u>**0.0081**</u>	0.1901
Simple mean	0.0028172	0.2715	0.3292
Simple average	0.0028426	0.1284	0.6073
Granger-Ramanathan	0.0028544	0.0732	0.8438

Table 7: 12-step ahead results for all models: OOS RMSE, MZ test results, DM test results against AR(3) benchmark (values significant at the 5% level are bolded and underlined)

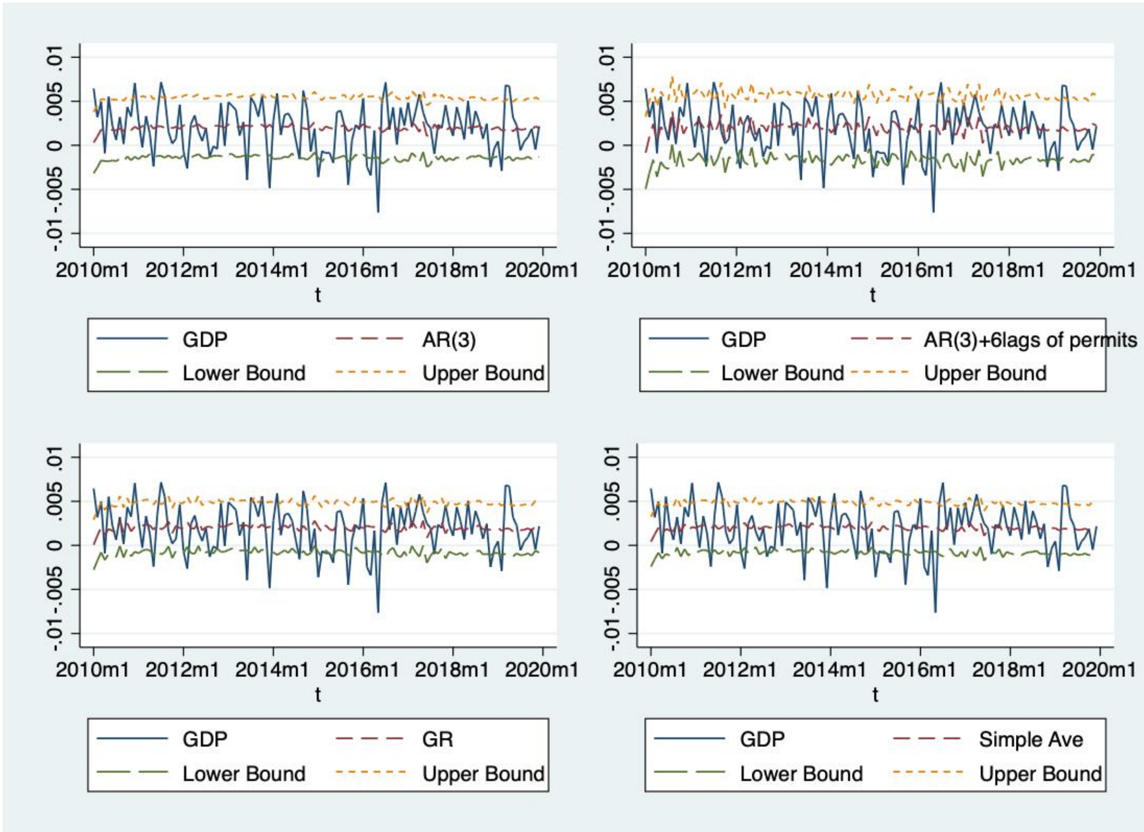

Figure 7: Plot of 12-step ahead forecasts (AR(3), AR(3) + 6 lags of building permits, Granger-Ramanathan combination, simple average combination) vs. actual values with 68% forecast intervals

Conclusion

Forecasts of the autoregressive distributed lag model (ADL) augmented with building permits do not outperform the benchmark autoregressive model (AR) in forecasting Canadian real GDP for January 2010 to December 2019. Across the 1-, 3-, 6- and 12-step forecast horizons, the forecast of the AR model performs equally well as the ADL model forecasts. In some cases, forecasts of a simple mean model and forecast combinations can outperform the benchmark forecast. For the 1-step forecast, forecasts of the simple mean and simple average forecast combination are found to outperform the AR(3) forecast. For the 3-step forecast, forecasts of the simple mean, simple average forecast combination and Granger-Ramanathan forecast combination can outperform the AR(3) forecast. At further forecast horizons, however, neither forecasts of the simple mean nor any forecast combination outperformed the benchmark forecast, performing equally well. As such, we can infer that the simple mean and forecast combination methods may perform especially well for shorter forecast horizons. The finding that an easy-to-compute simple mean model is not outperformed by the benchmark AR model is surprising. It is also surprising that employing a more complicated method of combining forecasts (i.e., Granger-Ramanathan) does not prove to be superior to taking a simple average of all forecasts. Overall, the performance of the simple mean model, benchmark autoregressive model and simple average forecast combination method proves that simple methods are useful and should be considered. However, it is important to note that this study is done specific to the data from Canada. It is still possible that the AR(p) model is outperformed by ADL models when data from other countries is used.

A limitation of the paper is that we only considered 4 leading indicators, 3 of which were dropped as AIC selected 0 lags of these indicators. Considering a larger number of leading indicators and a combination of these potentially better-performing leading indicators could result in an ADL model which performs better than our benchmark. For instance, future studies might consider including high-yield spreads and other term spreads not included in our current study. This limitation may also hinder the performance of the forecast combination models, given that the averaging out process is only done with, at best, 3 models. A greater number of models would potentially allow for better performance, as the combination can smooth out the weaknesses of each model and take advantage of their strengths. Another limitation is that we did not allow the forecasting models to differ for each forecast horizon due to time constraints. Future studies might consider conducting model selection for each of the forecast horizons for potentially more optimal h-step forecasts. Lastly, due to data constraints, we were unable to incorporate forecast errors into the forecasts to correct for the non-white noise property of our 1-step ahead forecast errors and the more than MA(2) property of our 3-step ahead forecast errors. To adequately correct the errors, one would need to hold out another set of observations to construct the forecast errors that will then be included in the forecasting. We attempted to add one to two lags of GDP for our 1-step ahead forecasts to resolve the issue but it was insufficient to correct the errors. As such, a key limitation is that our 1- and 3-step ahead forecasts are not optimal and we are unclear of what the source of it is.

In this paper, we omitted data after December 2019 due to worries over the potential impact of COVID-19 on forecasting performance. Future studies may wish to forecast Canada's GDP using a fixed window ending in December 2019 and see if our ranking of the models' performance changes. Future studies may also consider applying machine learning methods to forecasting Canada's GDP (e.g., ridge, LASSO, elastic net, random forest). As mentioned, there are multiple possible leading indicators of GDP. Machine learning methods can perform well in such high-dimensional settings and might yield better forecasts of GDP than the standard econometric models considered in this paper. Recent papers on Canadian GDP growth forecasting have found that boosting methods can outperform autoregressive models (Qureshi et al., 2021). A recent study by Goulet Coulombe et al. (2021) also found that non-linear machine learning methods had an edge in extrapolating key UK indicators during the COVID-19 period. This suggests that there is value in employing machine learning methods, even for potentially hard-to-predict periods (e.g., the COVID-19 period).

References

Chauvet, M., & Potter, S. (2013). Forecasting output. *Handbook of Economic Forecasting*, 141–194. https://doi.org/10.1016/b978-0-444-53683-9.00003-7

Christiansen, C., Eriksen, J. N., & Møller, S. V. (2014). Forecasting US recessions: The role of sentiment. *Journal of Banking & Finance*, *49*, 459–468. https://doi.org/10.1016/j.jbankfin.2014.06.017

Elliott, G., & Timmermann, A. (2008). Economic Forecasting. *Journal of Economic Literature*, *46*(1), 3–56. http://www.jstor.org/stable/27646946

Fortin-Gagnon, O., Leroux, M., Stevanovic, D., & Surprenant, S. (2022). A large Canadian database for macroeconomic analysis. *Canadian Journal of Economics/Revue Canadienne d'économique*, *55*(4), 1799–1833. https://doi.org/10.1111/caje.12618

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*(6), 1545-1578. https://doi.org/10.1111/j.1468-0262.2006.00718.x

Goulet Coulombe, P., Marcellino, M., & Stevanovic, D. (2021). Can machine learning catch the covid-19 recession? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3796421

Paul, S. (1969). Short-Term Economic Forecasting: Some Leading Indicators. *Economic and Political Weekly*, *4*(8), M7–M12. http://www.jstor.org/stable/40737342

Qureshi, S., Chu, B., & Demers, F.S. (2021). *Forecasting Canadian GDP Growth with Machine Learning* (Carleton Economics Working Papers No. 21-05). https://carleton.ca/economics/wp-content/uploads/cewp21-05.pdf

Rudebusch, G. D., & Williams, J. C. (2009). Forecasting Recessions: The Puzzle of the Enduring Power of the Yield Curve. *Journal of Business & Economic Statistics*, *27*(4), 492–503. http://www.jstor.org/stable/27799101