Forecasting Inflation in Canada

EC4304

2023/2024 Semester 1

1. Introduction

Forecasting inflation is crucial for policy making and financial planning. Accurate inflation forecasts guide monetary policy decisions and stabilizing economies. Inaccuracies in inflation forecasting by policy makers such as central banks could undermine their credibility and hence, their ability to anchor inflation expectations. However, inflation has been considered hard to forecast in the literature, and it’s difficult to beat simple univariate models (Stock & Watson, 2009). According to Stock & Watson (2010), “it is exceedingly difficult to improve systematically upon simple univariate forecasting models, such as the Atkeson-Ohanian (2001) random walk model”.

Considering this, this report focuses on the effectiveness of Autoregressive (AR), Autoregressive Distributed Lag (ADL) models, and forecast combinations in the context of inflation forecasting, particularly on the Canadian economy. These models are chosen due to their relative simplicity yet potential performance in forecasting. Simpler models may be preferred, particularly in contexts where more complex models do not necessarily yield significantly better results, due to concerns like overfitting and ease of interpretation. Canada was chosen due to its unique inflation dynamics. According to Matheson (2019), inflation in Canada, has not only the trends of other advanced economies, but also displayed distinct characteristics such as lower volatility and less persistence compared to its OECD counterparts. Therefore, Canada provides an interesting avenue to examine these models in a stable, yet distinct inflationary environment. This is especially true for testing the effectiveness of AR type models, in an economy where inflation is less persistent.

In this report, our primary objective is to evaluate the pseudo out-of-sample effectiveness of AR, ADL, and various forecast combinations in accurately forecasting inflation against a random walk benchmark, particularly within the Canadian economic context. We aim to investigate not only the overall predictive power of these models but also to discern for different forecasting horizons (h= 1, 3, 6, 12 months), which specific models or techniques may exhibit superior performance based on pseudo out-of-sample root mean squared error. Pseudo out-of-sample forecasts were also statistically compared using the Diebold-Mariano (1995) test of equal predictive ability.

We find that simpler forecasting models such as Autoregressive (AR) and Autoregressive Distributed Lag (ADL) demonstrate great effectiveness. These models surpass the benchmark Atkeson-Ohanian (2001) Random Walk model in terms of predictive accuracy at h = 1,3,6. Furthermore, we also found that the use of forecast combinations led to worse predictive accuracy compared to “pure” forecasts.

The remainder of this report is structured as follows: Section 2 describes the models and methodology employed in this report, Section 3 outlines the data source, Section 4 provides an analysis of the findings and Section 5 and 6 discusses the limitations and provides a conclusion to this report.

2. Methodology

Explanation of models chosen.

2.1.1 Benchmark Forecasts: Random Walk (RW) AO model

We choose the random walk model to be the benchmark in our analysis. A random walk forecast assumes that the future values of a time series will be equal to the most recent observed value, with no regard for trends, seasonality, or any underlying patterns. In other words, it assumes that the best estimate for the future is the current value, and future changes are purely random. The simplicity of this model makes it a good benchmark to compare against our more complicated models at various forecast intervals.

For our forecasting purposes, we adopted the Atkeson-Ohanian (2001) model, where we forecasted a h-step ahead forecast by averaging the past h values of our RW model.

$$Y^{h}_{t+h|t} = \frac{1}{h}(Y_t + Y_{t-1} + \cdots + Y_{t-h+1})$$

- h is the forecast horizon
- t is the current time period
- Y_t is the value at time period t

We compared all our models to this benchmark by Out-of-sample Root-mean-square-error (RMSE) to determine if our more complicated models beat this benchmark.

2.1.2 AR

An Autoregressive (AR) model uses a linear combination of the past values of time series data for forecasting predictions. The idea behind an AR model is that the current value of a time series is linearly dependent on its previous values, with the degree of dependence determined by the order of the model. The order of the model, denoted as ‘p’, determines how many past values are included in the model.

The general form of an AR(p) model is as follows:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$

- Y_t is the value at time period t
- ϕ_1, ϕ_2, ϕ_p are autoregressive coefficients

- $Y_{t-1}, Y_{t-2}, Y_{t-p}$ are lagged values of the time series
- ε_t is the white noise error term at time t

2.1.3 ADL

The Autoregressive Distributed Lag (ADL) model incorporates both AR components as well as lagged values of other relevant independent variables in the form of distributed lags. ADL models are useful in studying dynamic relationships over time as they allow for the analysis of how changes in independent variables may affect the dependent variables. These independent variables are chosen as potential leading indicators for our dependent variable (inflation) based on established economic theory.

The general form of an ADL(p, k) model is as follows:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \gamma_1 X_{t-1} + \gamma_2 X_{t-2} + \cdots + \gamma_k X_{t-k} + \varepsilon_t$$

- Y_t is the value at time period t
- $\phi_1,\ \phi_2,\ \phi_p$ are autoregressive coefficients
- $Y_{t-1}, Y_{t-2}, Y_{t-p}$ are lagged values of the time series
- X_t is the value of the independent variable at time period t
- $\gamma_1, \gamma_2, \gamma_k$ are coefficients of the independent variable
- $X_{t-1}, X_{t-2}, X_{t-k}$ are lagged values of the independent variable
- ε_t is the white noise error term at time t

Our chosen independent variables will be explained further in Section 3.1.

2.1.4 Forecast Combination: Simple Averaging

We utilize forecast combination for better results as forecast combinations have shown to have improved performances over individual forecasting models (Atiya, 2020). Combining forecasts from multiple methods can help mitigate individual errors and produce a more accurate and robust prediction by hedging the inaccuracies derived from a single forecast model.

The simplest form of forecast combination is the simple averaging where we take the average forecast of several models to create new forecast results. In this case, we take the average forecast of our Random Walk forecast, AR(1), AR(AIC), and top 3 ADL models. All our models have the same weights in the simple average.

$$Y_{SA} = (Y_{RW} + Y_{AR1} + Y_{ARAIC} + Y_{ADL1} + Y_{ADL2} + Y_{ADL3})/6$$

- Y_{SA} is our simple averaging forecast

- Y_{RW} is our random walk forecast
- Y_{AR1} is our AR(1) forecast
- Y_{ARAIC} is our AR(AIC) forecast
- $Y_{ADL1}, Y_{ADL2}, Y_{ADL3}$ are our top 3 ADL forecasts respectively

2.1.5 Forecast Combinations: Bates-Granger

Bates and Granger (1969), suggested using empirical weights based on out-of-sample forecast variances. We did this by firstly using a pseudo out-of-sample forecast to get the forecast error for each model. Assuming our models were unbiased, the mean-squared-error (MSE) is equivalent to the variance of the model, thus we computed the MSE for each model and inverted them. Finally, we normalized these inverted values to get the weights of each of the models.

The benefit of this over simple averaging is that we were able to account for the variances in each of the model and give a higher weight to more accurate (lower variance) models, which would hopefully improve forecast accuracy.

$$w_j^* = \frac{\sigma_j^{-2}}{\sigma_1^{-2} + \sigma_2^{-2} + \cdots + \sigma_N^{-2}}$$

- σ_j is the Standard Deviation of Model j
- w_j^* is the weight of model j in the averaging forecast

2.1.6 Forecast Combinations: Granger-Ramanathan

Granger and Ramanathan (1984) introduced a regression method to combine forecasts. This is done by regressing our out-of-sample forecasted values of each model on the true values, while also adding a constraint that all the coefficients of the regressions must add up to 1. For easier implementation, we then remove the forecasts with negative coefficients and run the regression again, until only positive coefficients remain. It is assumed that forecasts with negative coefficients do not add much predictive power to other forecasts. The remaining positive coefficients will be the weights of averaging.

The advantage of this over Bates-Granger is that it allows for correlations among our forecasts, and takes that into account when producing the weights, producing a more robust forecast combination.

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

Constraint: $\beta_1 + \beta_2 + \cdots + \beta_N = 1$

- Y_t^{true} is the true values at time t
- β_i are coefficients of regression of forecast model i
- f_{it} are forecast values of model i at time t

2.1.7 Forecast Combinations: Bayesian Model Averaging (BMA)

Bayesian Model Averaging (BMA) is an averaging model where the weights are calculated using BIC. BMA puts most weights on the model with the lowest BIC, and very little to none on those with BIC very different from the best model. The formula below shows how the weights are calculated, these weights are then normalized before being used in averaging.

$$w_m^* = e^{\left(-\frac{BIC_m - BIC_{Lowest}}{2}\right)}$$

$$w_m = \frac{w_m^*}{\sum_{m=1}^{M} w_m^*}$$

- BIC_m, BIC_{Lowest} are the BIC values of model m and lowest BIC of all the models respectively
- w_m^* are the pre-normalized weights of model m
- w_m are the weights of model m after normalization

2.1.8 Forecast Combinations: Weighted Akaike Information Criterion (WAIC)

Weighted Akaike Information Criterion (WAIC) uses the same formula for calculating weights as BMA, but instead of comparing models on BIC, we compare them on AIC. Using both BMA and WAIC allows us a more comprehensive assessment of our forecast combinations as AIC and BIC may favour different models.

$$w_m^* = e^{\left(-\frac{AIC_m - AIC_{Lowest}}{2}\right)}$$

$$w_m = \frac{w_m^*}{\sum_{m=1}^{M} w_m^*}$$

- AIC_m, AIC_{Lowest} are the AIC values of model m and lowest BIC of all the models respectively
- w_m^* are the pre-normalized weights of model m
- w_m are the weights of model m after normalization

Forecasting Scheme:

2.2 Forecasting with different horizons.

To account for different time horizons in our analysis, we generate direct forecasts for our AR and ADL models. We do this by changing the lags of the autoregressive and distributed lag components of the forecast model to correspond to the respective h-period ahead forecast.

Examples of the models used are shown below:

3-step ahead forecast using an AR(1) model: $y_t = y_{t-3} + \varepsilon_t$

3-step ahead forecast using an ADL(1,1) model: $y_t = y_{t-3} + x_{t-3} + \varepsilon_t$

Where y_t refers to the logarithmic difference in consumer price index (CPI) between the h-period ahead and current period in Canada and x_t is a vector of leading indicators with predictive capabilities on CPI data. The specific leading indicators will be introduced in Section 3.1.

2.3 Model Selection

2.3.1 AR Model

For the AR model, we started off by plotting the partial autocorrelation plots of the logarithmic difference between the h-step ahead and current period CPI data (LDh_CPI_ALL_CAN) to determine the maximum lags of logarithmic first difference CPI data (LD1_CPI_ALL_CAN) that we should potentially include in the AR(p) model. Next, for each time horizon, we ran a regression of LDh_CPI_ALL_CAN on the first lag until the max lag of LD1_CPI_ALL_CAN. By comparing their AIC, we selected the model with the lowest value as the best AR(p) model as a baseline moving forward.

For our forecasting, we used AR(1) and AR(AIC) where AR(AIC) is the best AR model chosen by AIC. We chose to use these models as AR(1) is the simplest form of AR models which allows us to use it as a pseudo-benchmark against more complex models, while AR(AIC) is the best of all possible AR models as denoted by AIC testing.

2.3.2 ADL Model

To create our ADL models, we first conducted the Granger Causality Test on the independent variables, that will be introduced in Section 3.1, to ensure that they are truly helpful in predicting inflation rates. Initial Granger Causality tests tested up to 12 lags (1 year of lagged data) of each independent variable. If up to 12 lags were statistically significant at the 5% level, we would then expand the Granger Causality test to include up to 18 lags (1.5 years of lagged data). By testing against the null of no causality at a 5%

significance level, we managed to eliminate a few leading indicators and conclude that they do not help in predicting inflation.

We then included distributed lags of the independent variables that Granger Causes inflation into the best AR(p) model found for each time horizon. The independent variables were added one at a time, such that every ADL model only consisted of an AR component and the distributed lag of one leading indicator. The maximum number of lags for the leading indicator in the ADL model corresponded to the number of lags tested in the Granger Causality test. After appending the first lag until the max lag of the leading indicator into the best AR model found earlier, we compare their AIC values and choose the smallest AIC as the best "pure" ADL model for that specific leading indicator. Some ADL models failed to beat the purely AR models and they were removed from the pool of variables for the next step.

In the next step, we took the best ADL models for every leading indicator and combined them to create ADL models of 2 or more leading indicators. After going through every possible combination, the AIC values of these "combined" models were then compared amongst themselves and the "pure" ADL models and the 3 best models with the lowest AIC values were chosen. We chose 3 ADL models as the single best model according to AIC does not always translate to the best model for Out-of-sample RSME, so the top 3 models would give us a more representative look at how ADL models compare to our benchmark.

2.3.3 Forecast Combinations

The main forecast combinations used were Simple Averaging, Bates-Granger, Granger Ramanathan, BMA and WAIC. The implementation follows as described in Sections 2.1.4 – Section 2.1.8.

2.4 Rolling Window

With reference to the models stated in sections 2.1.3 – 2.1.8, these models are built upon the AR(AIC) model that was mentioned in section 2.1.2. Consequently, these models have a nested AR(AIC) component within them. As a Diebold-Mariano Test (DM test) will be employed to evaluate whether the forecast models beat the benchmark AO and AR(1) models, an expanding window scheme should not be used to ensure that the normality of the DM test statistic holds. Furthermore, the expanding data sample size used to make later forecasts will reduce parameter uncertainties associated with these later forecasts which suggests that forecast error variances may be reduced with time, violating the DM assumption that loss differentials are covariance stationary.

To avoid the above scenario, we made use of a rolling window scheme where the forecast models are constantly re-estimated and predicted using the most recent N number of observations. To obtain the next period point forecast, we add in the N+1 observation while dropping the first observation, ensuring that the sample size used for estimating point forecasts remains fixed at N for every subsequent window.

By adopting a rolling window scheme, we ensure that the DM test assumptions are not violated. There is also the added benefit of creating more accurate forecasts as the forecast models are estimated using recent data which considers possible changes in the data pattern.

Explanation of Model Comparison: RMSE and DM Test

2.5.1 Out-of-sample (OOS) Forecast Evaluation: Mean Squared Error (MSE)

In the context of forecast modeling, the optimal conditional point forecast is one that minimizes the expected loss or risk. The OOS-MSE serves as an indicator of this risk when working under the squared loss function. It can be broken down into variance and squared-bias components, along with some irreducible error. Therefore, our initial evaluation of forecasting models for Canadian inflation involves comparing their out of sample root mean squared error (OOS-RMSE) performance against the AO random walk benchmark. This comparison aims to determine whether more complex ADL and combined weighted models can significantly outperform the Random Walk benchmark. The equation of MSE under a squared loss function is defined below:

$$MSE = \frac{1}{n}(Y - \hat{Y})^2$$

2.5.2 Out-of-sample (OOS) Forecast Evaluation: Diebold-Mariano Test (DM test)

To evaluate the statistical significance of the difference in losses between the benchmark and selected models, we employ the DM test, as suggested by Diebold and Mariano (1995). The DM test functions under the null hypothesis that the predictive abilities of the chosen and benchmark models are equal. A rejection of the null hypothesis from the DM test provides an indication of the superiority of a chosen forecast over the benchmark if the chosen forecast has a lower loss criterion (in our case lower MSE). Assuming that the differences in the loss functions are covariance stationary, the test statistic is calculated as follows:

$$d = L\left(e_{t+h|t}^{(1)}\right) - L\left(e_{t+h|t}^{(2)}\right)$$

$$DM = \frac{\bar{d}}{\hat{\sigma}_{\bar{d}}} \sim N(0,1)$$

Where $\bar{d}$ is the sample mean of loss differential and $\widehat{\sigma_{\bar{d}}}$ is a consistent estimate of its standard deviation. We use Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors, due to the serial correlation in loss differentials. The truncation parameter for the lag length is observed by plotting the autocorrelation plot of the loss function differentials and taking the greatest lag with significant autocorrelations.

By setting $L\left(e_{t+h|t}^{(1)}\right)$ and $L\left(e_{t+h|t}^{(2)}\right)$ as the loss function for the benchmark model and chosen model respectively, positive loss differentials imply that the benchmark model has a greater loss than the chosen model. Hence, positive loss differentials reflect outperformance of the benchmark and a small p-value corresponding to these positive loss differentials highlights outperformance at statistically significant levels.

3. Data Source

3.1 CAN-MD dataset

The data that we use in this analysis is sourced from the CAN-MD dataset, a large monthly dataset for macroeconomic analysis of the Canadian economy prepared by Fortin-Gagnon et al. (2022). From this dataset, our dependent variable is CPI_ALL_CAN which is the consumer price index across Canada.

For the distributed lags, we consider the following six variables from money supply, interest rate, T-bill bank rate spread, unemployment and industrial production price index as possible leading indicators of inflation: i) M_BASE1 (monetary base) ii) M2p (M2+ gross) iii) M3 (M3 gross) iv) BANK_RATE_L (bank rate) v) TBILL_6M-Bank_rate (6 Month Treasury bond – Bank Rate) vi) UNEMP_CAN (Unemployment Rate) vii) Industrial Production Price Index (IPPI). The variables used for the distributed lags, taken directly from the balanced and stationary dataset, and starts from 1981 January.

3.2 Transformation of CPI_ALL_CAN

To accommodate forecasting across various time horizons, we transform CPI_ALL_CAN from the available raw data provided by Fortin-Gagnon et al. (2022) into distinct stationary series corresponding to each forecasting horizon, all starting from 1981 January. This stationary transformation is achieved by taking the h-step logarithmic first difference of CPI_ALL_CAN. For instance, for the 3-month ahead horizon, LD3_CPI_ALL_CAN represents the log difference in the raw Consumer Price Index over three months calculated as i.e. $\log(P_t) - \log(P_{t-3})$. Logarithmic first differences was chosen following the methodology in Fortin-Gagnon et al. (2022).

This approach ensures that our dependent variable directly corresponds to the h-period ahead log inflation difference, rather than a one-step ahead log difference projected h-periods forward.

Importantly, while these serve as the dependent variables in our forecast, the autoregressive predictors (for AR models) remain as the 1-month log difference of CPI_ALL_CAN consistently across all forecasting horizons. For example, in a 3-month ahead autoregressive forecast, we would regress LD3_CPI_ALL_CAN against the lagged values of the 1-month log differenced index, LD1_CPI_ALL_CAN.

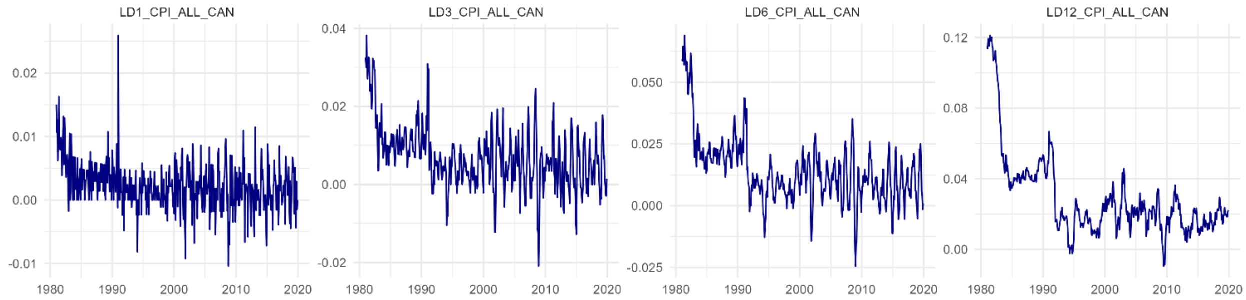

Figure 1: CPI_ALL_CAN log-differenced at h= 1, 3, 6 and 12.

3.2 Time Period, Horizons and Training-Validation-Test Split:

For our analysis, we restrict the samples to the pre-pandemic period ending 2019 December, yielding a total of 468 observations. We consider the forecast horizons of 1-month, 3-months, 6-months, and 12-months to assess the short-term to long-term performance of our forecasting methods compared against the benchmark. Figure 1 above shows the plots of the CPI_ALL_CAN log differenced for different horizons.

The dataset is then split into three subsets: training (comprising 268 observations), validation (100 observations), and testing (100 observations). The validation sample will be used for the computation of weights for the Bates-Granger and Granger-Ramanathan forecast combinations following section 2.1.5 to 2.1.6. After calculating the weights, we use the data up till the first 368 observations to generate our point forecasts as well as the forecast combinations, which will be used to comparatively test the out-of-sample performance of each of the different models against the test set.

All models employ a fixed rolling window approach, as outlined in section 2.4. The window's size is fixed for each horizon accounting for horizon shift as well as the autoregressive orders.

4. Results and Evaluation

4.1 Best AR(p) model for each horizon

Table 1: Best AR(p) models chosen by AIC

Horizon	AIC Best
1	AR12
3	AR12
6	AR12
12	AR15

Table 1 shows the best p orders for the AR models chosen by AIC for the different horizons. Notably, AR(12) was chosen for h = 1, 3, 6 while AR(15) was chosen for h = 12. This high AR order suggest that despite Canadian inflation not being very persistent compared to other OECD countries, as explained in Matheson (2019), past lags of inflation still help predict future inflation.

On another note, for horizons, h=1,3 and 6, the best AR order chosen by BIC was AR(1). Motivated by this, we also included evaluating the AR(1) forecast into our analysis, to see if the simplest case of AR(1) will perform as well as the models chosen by AIC.

4.2 Best 3 ADL models

The top three ADL models were selected following the methodology in section 2.3.2.

Table 2: **1 Step ahead ADL models**

Model	Model Notation	AIC	BIC
Best AR(AIC) model	AR12	-2179.02	-2133.04
M1/M2/Bank Rate/UE rate/IPPI	ADL1: m1m2brunip	-2201.99	-2106.48
M1/M2/UE rate/IPPI	ADL2: m1m2unip	-2201.76	-2109.79
M1/Bank Rate/UE rate/IPPI	ADL3: m1brunip	-2199.66	-2111.23

Table 3: **3 Step ahead ADL models**

Model	Model Notation	AIC	BIC
Best AR(AIC) model	AR12	-1839.02	-1793.45
M2/BankRate/UE rate	ADL1: m2brun	-1866.24	-1775.11
M2/UE rate	ADL2: m2un	-1864.48	-1776.84
M2/BankRate	ADL3: m2br	-1864.31	-1797.71

Table 4: **6 Step ahead ADL models**

Model	Model Notation	AIC	BIC
Best AR(AIC) model	AR12	-1599.43	-1553.97
M2	ADL1: m2	-1654.37	-1577.43
M3/M2	ADL2: m3m2	-1647.64	-1549.72
M3	ADL3: m3	-1635.19	-1568.74

Table 5: **12 Step ahead ADL models**

Model	Model Notation	AIC	BIC
Best AR(AIC) model	AR15	-1380.32	-1324.76
M2/BankRate	ADL1: m2br	-1538.32	-1420.27
M2/UE rate	ADL2: m2un	-1522.81	-1383.92
M2/BankRate/UE rate	ADL3: m2brun	-1520.62	-1343.53

As seen in tables 2 to 5, at lower horizons of h=1 and 3, more of the leading indicators were selected to be in the best ADL models, while h= 6 and 12 selected fewer leading indicators. For instance, the best ADL model in h=1 contains five leading indicators, while for h=12, the best model contained two leading indicators.

As for forecast combinations, the weights of the different combinations are detailed in the appendix.

4.3 Main Results and Discussion: RMSE and DM test results

Table 6 shows the main results of our analysis and the accompanying p-values from the Diebold-Mariano test of the forecasts compared against the Random Walk Benchmark.

For h=1 and 3, the lowest pseudo out-of-sample RMSE models were the ADL models. They had a lower RMSE compared to the random walk benchmark by 30% and 45% respectively. For h = 6, the model with the lowest RMSE was the AR(AIC) model, beating the benchmark by 37.9%. For all three horizons h=1,3,6 they were also found to be statistically significantly superior based on the results of the Diebold-Mariano test. These findings are consistent with Stock and Watson (2009), who also observed that AR(AIC) as well as ADL models having lower RMSE compared to the Atkeson-Ohanian (2001) Random walk model.

For h=12, none of the models evaluated in this report were statistically significantly different from the random walk benchmark, although some models had lower RMSE. Therefore, no models had superior forecasting ability compared to the random walk benchmark.

As for AR(1) forecasts (chosen by BIC for h=1,3,6), despite beating the Random Walk benchmark, they consistently underperform AR(AIC) forecasts by RMSE across all horizons. This suggests that AIC may be a more effective criterion for model selection in forecasting scenarios compared to BIC.

In terms of forecast combinations, they generally performed worse than the “pure” forecasts by RMSE across forecast horizons. This may be largely driven by the poor performance of the RW and AR(1) models which may contribute to worse performance when added into a forecast combination. Furthermore, this implies that the more accurate "pure" forecasts were perhaps already sufficiently comprehensive, diminishing the value added by combining them with lesser-performing models. This outcome prompts a reevaluation of the criteria for selecting models for combination, suggesting a potential refinement in the aggregation process or the exclusion of models below a certain efficacy threshold.

Table 6: **Random Walk Benchmark**

	h=1	**h=3**	**h=6**	**h=12**
RW	0.004356	0.009767	0.010954	0.008933
AR1	0.003456	0.006832	0.009017	0.009285
	(0.0026)	(0.0000)	(0.0068)	(0.7381)
AR(AIC)	0.003080	0.005421	**0.006801**	0.008377
	(0.0001)	(0.0000)	(0.0002)	(0.3752)
ADL1	0.003048	**0.005326**	0.006919	0.008703
	(0.0001)	(0.0000)	(0.0010)	(0.7290)
ADL2	0.003050	0.005336	0.007009	0.008890
	(0.0001)	(0.0000)	(0.0024)	(0.9454)
ADL3	**0.003042**	0.005378	0.006965	0.008900
	(0.0001)	(0.0000)	(0.0015)	(0.9592)
FA	0.003096	0.005980	0.007433	0.008363
	(0.0000)	(0.0000)	(0.0004)	(0.3305)
BG	0.003073	0.005769	0.007371	**0.008300**
	(0.0000)	(0.0000)	(0.0005)	(0.2902)
GR	0.003152	0.005768	0.007123	0.008474
	(0.0001)	(0.0000)	(0.0004)	(0.5562)
BMA	0.003455	0.006594	0.006802	0.008703
	(0.0026)	(0.0000)	(0.0002)	(0.7287)
WAIC	0.003043	0.005359	0.006907	0.008837
	(0.0001)	(0.0000)	(0.0011)	(0.8797)

Note: The displayed values represent the root mean squared error (RMSE) of the forecast horizon (h=1, h=3, h=6, h=12) for each forecast. Values in parentheses are the p-values from the Diebold-Mariano (DM) test against the Random Walk benchmark for the specific forecast. Values in red show forecasts that were not statistically significantly different from the Random Walk benchmark Bolded values represent the lowest RMSE forecast.

Upon comparing the RMSE values across the different horizons in Table 6, we observed that most of the RMSE of the top performing models, had similar RMSE values compared to the AR(AIC) model. Therefore, this prompted further statistical investigation. We use the DM test to test if these other models were significantly different than the AR(AIC) model.

Table 7 below shows the RMSE of the forecasts as well as the results of the DM tests against the AR(AIC) forecasts across different horizons. We observe that for h = 1, none of the better RMSE models were statistically different compared to the AR(AIC) model. Notably, at h=1 the BMA forecast combination is found to be statistically significantly worse than the AR(AIC) model. For h = 3, this trend continued, with the lower RMSE models not differing significantly from the AR(AIC). Additionally, for h = 3, most of the forecast combinations—including Forecast Averaging, Bates-Granger, Granger-Ramanathan, and Bayesian Model Averaging—were statistically worse than the AR(AIC) forecast based on the DM test. The findings for the 6-month horizon mirrored those for 3 months. These findings are consistent with Stock and Watson's (2009) literature review which found the lack of success found by attempts in the literature to obtain gains over individual indicator forecasts using forecast combinations for U.S. inflation. For h=12, the performance of other models was not statistically distinguishable from the AR(AIC), though as previously mentioned, none of these models, including AR(AIC), outperformed the Random Walk benchmark.

Table 7: **DM Tests against AR(AIC)**

	h=1	**h=3**	**h=6**	**h=12**
AR(AIC)	0.003080	0.005421	**0.006801**	0.008377
AR1	0.003456 (0.0001)	0.006832 (0.0000)	0.009017 (0.0000)	0.009285 (0.0902)
ADL1	0.003048 (0.7412)	**0.005326** (0.6735)	0.006919 (0.6227)	0.008703 (0.3991)
ADL2	0.003050 (0.7547)	0.005336 (0.5105)	0.007009 (0.4789)	0.008890 (0.2699)
ADL3	**0.003042** (0.6813)	0.005378 (0.6735)	0.006965 (0.5186)	0.008900 (0.2748)
FA	0.003096 (0.8467)	0.005980 (0.0001)	0.007433 (0.0009)	0.008363 (0.9586)
BG	0.003073 (0.9163)	0.005769 (0.0006)	0.007371 (0.0003)	**0.008300** (0.7060)
GR	0.003152 (0.3180)	0.005768 (0.0000)	0.007123 (0.0000)	0.008474 (0.6399)
BMA	0.003455 (0.0001)	0.006594 (0.0000)	0.006802 (0.8536)	0.008703 (0.3994)
WAIC	0.003043 (0.6976)	0.005359 (0.5515)	0.006907 (0.6603)	0.008837 (0.3049)

Note: The displayed values represent the root mean squared error (RMSE) of the forecast horizon (h=1, h=3, h=6, h=12). Values in parentheses are the p-values from the Diebold-Mariano (DM) test against the AR(AIC) forecast for the specific forecast. Values in red show forecasts that were not statistically significantly different from the Random Walk benchmark. Bolded values represent the lowest RMSE forecast.

Also, following from the discussion earlier, Table 7 demonstrates that AR(1) is statistically significantly worse compared to the AR(AIC) model, reaffirming that AIC is the more effective criterion for forecasting model selection.

These results from table 6 and 7 suggest that AR(AIC) and its better performance over ADL and forecast combination is robust across horizons 1, 3 and 6. The use of forecast combinations did not generally enhance forecasting accuracy in this context. Furthermore, ADL models are not statistically significantly better than AR(AIC) forecasts. This indicates that for inflation forecasting in Canada, a well specified AR(AIC) model is sufficient in forecasting inflation across short to medium term inflation (h=1 to 6) compared to ADL forecasts and their combinations. Other methods may be needed to beat the AO random walk benchmark for longer forecasting horizons of 12 months and beyond.

5. Limitations

1) Exhaustive Grid Search Consideration:

One limitation of our study lies in the decision to refrain from an exhaustive grid search when creating our ADL models. We worked with the assumption that the best ADL models for each individual independent variable by AIC testing would form some of the best possible ADL models when these variables are combined. However, this assumption may not be correct, and there could be an ADL combination model that does not use the exact variables or lags of the best individual ADL models that performs better than our Top 3 ADL models selected. As a result, there might exist undiscovered optimal ADL models that could provide a higher forecasting accuracy.

2) Model Selection Scope:

Another limitation arises from the narrow scope of models considered in this study. While we have evaluated traditional models such as Random Walk, AR, ADL. The recent studies in the field of inflation forecasting have suggested that Machine Learning (ML) models such as tree-based methods can often result in better forecasts than traditional models, especially over longer forecast horizons (Araujo, 2023). The exclusion of ML models thus restricts the comprehensiveness of our analysis.

3) Pandemic data constraints:

We limited our dataset to only pre-COVID-19 pandemic period to prevent our analysis from being potentially hindered by the economic disruptions caused by the pandemic. This was to ensure that our findings were not biased and could be generalized to display findings for non-pandemic economic periods. However, the transferability of our forecasting models to the post-pandemic period remains uncertain as the dynamics of the economy may have drastically changed from the pre-pandemic economy. Since we are only using data from the pre-pandemic period, we should be careful when applying these findings to the post-pandemic period.

6. Conclusion

In conclusion, our research has evaluated several models for forecasting inflation in Canada and compared it to our simple benchmark model at various time horizons. Our analysis revealed that simpler Autoregressive (AR) and Autoregressive Distributed Lag (ADL) models are quite effective in forecasting inflation within the Canadian context, often outperforming the Random Walk benchmark, especially for lower forecasting horizons of h = 1, 3 and 6. Particularly, the AR model selected by the Akaike Information Criterion (AIC) demonstrated consistent accuracy across these forecasting horizons. Forecast combinations did not yield any improvement in prediction accuracy. Based on DM tests, we conclude that ADL and forecast combinations are statistically indistinguishable or perform worse compared to AR(AIC). These findings emphasize the effectiveness of simpler models in capturing the unique dynamics of Canadian inflation and suggest that complexity in model selection does not necessarily guarantee enhanced forecasting precision.

To acknowledge our limitations, some possible extensions for our research could be using an exhaustive grid search on both AIC and BIC to determine the best ADL models for evaluation, to potentially discover better ADL models than the ones we selected. We could also apply various Machine Learning models such as random forest and XGBoost to provide a more comprehensive analysis. Finally, we could update our models with relevant post-pandemic data for a more robust analysis of inflation forecasting in this post-pandemic period.

References:

Atiya, A. F. (2020). Why does forecast combination work so well? International Journal of Forecasting, 36(1), 197-200. https://doi.org/10.1016/j.ijforecast.2019.03.010.

Araujo, G. S., & Gaglianone, W. P. (2023). Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models. Latin American Journal of Central Banking, 4(2), 100087. https://doi.org/10.1016/j.latcb.2023.100087.

Atkeson, A., & Ohanian, L. E. (2001). "Are Phillips curves useful for forecasting inflation?" Quarterly Review, Federal Reserve Bank of Minneapolis, 25(Win), 2-11. https://ideas.repec.org/a/fip/fedmqr/y2001iwinp2-11nv.25no.1.html

Bates, J. M., & Granger, C. W. J. (1969). The Combination of Forecasts. OR, 20(4), 451–468. https://doi.org/10.2307/3008764

Diebold, F., & Mariano, R. (1995). Comparing Predictive Accuracy. Journal of Business & Economic Statistics, 13(3), 253–263.

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

Granger, C.W.J. and Ramanathan, R. (1984), Improved methods of combining forecasts. J. Forecast., 3: 197-204. https://doi.org/10.1002/for.3980030207

Matheson, T. D. (2019). *How to Improve Inflation Forecasting in Canada*. IMF. Retrieved November 12, 2023, from https://www.imf.org/en/Publications/WP/Issues/2019/09/13/How-to-Improve-Inflation-Forecasting-in-Canada-48541

Stock, J. H., & Watson, M. W. (2003). Forecasting Output and Inflation: The Role of Asset Prices. Journal of Economic Literature, 41(3), 788–829. http://www.jstor.org/stable/3217532

Stock, J. H., & Watson, M. W. (2009). Phillips Curve Inflation Forecasts. In J. Fuhrer, Y. K. Kodrzycki, J. S. Little, & G. P. Olivei (Eds.), Understanding Inflation and the Implications for Monetary Policy: A Phillips Curve Retrospective (p. 0). The MIT Press. https://doi.org/10.7551/mitpress/9780262013635.003.0003

Stock, J., & Watson, M. (2010). Modeling Inflation After the Crisis (w16488; p. w16488). National Bureau of Economic Research. https://doi.org/10.3386/w16488

Appendix

Weights of Forecast Combinations

Table 8: **1 Step ahead models**

Models	RW_h1	ar1_h1	ar12_h1	m1m2brunip	m1m2unip	m1brunip
BG Weights	0.1054	0.1784	0.1769	0.1804	0.1828	0.1761
GR Weights	0.0273	0.4502	-	-	0.5225	-

Table 9: **3 Step ahead models**

Models	RW_h3	ar1_h3	ar12_h3	m2brun	m2un	m2br
BG Weights	0.0897	0.1983	0.2093	0.1613	0.1623	0.1790
GR Weights	-	0.3301	0.6184	-	-	0.0515

Table 10: **6 Step ahead models**

Models	RW_h6	ar1_h6	ar12_h6	m2	m3m2	m3
BG Weights	0.1181	0.1984	0.2127	0.1462	0.1492	0.1755
GR Weights	-	0.1968	0.7153	-	-	0.0879

Table 11: **12 Step ahead models**

Models	RW_h12	ar1_h12	ar15_h12	m2br	m2un	m2brun
BG Weights	0.1901	0.2381	0.2473	0.1158	0.1018	0.1069
GR Weights	-	0.3907	0.6093	-	-	-

Table 12: **1 Step ahead models**

Model	**AR1**	**AR12**	**m1m2brunip_1**	**m1m2unip_h1**	**m1brunip_h1**
BMA Weights	0.99866	0.00134	-	-	-
WAIC Weights	-	-	0.38522	0.44622	0.16856

Table 13: **3 Step ahead models**

Model	**AR1**	**AR12**	**m2brun_h3**	**m2un_h3**	**m2br_h3**
BMA Weights	0.86808	0.13147	-	-	0.00045
WAIC Weights	-	0.00225	0.08300	0.14824	0.76651

Table 14: **6 Step ahead models**

Model	**AR1**	**AR12**	**m2_h6**	**m3m2_h6**	**m3_h6**
BMA Weights	0.00269	0.97935	-	-	0.01794
WAIC Weights	-	-	0.83893	0.15380	0.00727

Table 15: **12 Step ahead models**

Model	**AR1**	**AR15**	**m2br_h12**	**m2un_h12**	**m2brun_h12**
BMA Weights	-	0.00014	0.99980	0.00007	-
WAIC Weights	-	-	0.14490	0.83887	0.01623