PAYMENTS CANADA

Nowcasting macroeconomic variables using payments data

payments.ca

TABLE OF CONTENTS

Nowcasting macroeconomic variables using payment data

Anita Smirnova[a], Youssef Hassine[a], Oluwasegun Bewaji[a],
Ramakrishna Chittari[b], Pranav Arvind Joshi[b], Carolyn Knight[b], and Princess Uchegbulam[b]

[a] *Payments Canada*
[b] *Georgian College, Big Data Analytics*
June 2022

Abstract

The recent COVID-19 crisis underscored the importance of having more frequent access to estimates reflecting the current state of the economy, as these estimates aid policymakers in directing monetary and other types of economic policy. This paper aims to determine whether payment data can be used to forecast ('nowcast') macroeconomic indicators in the near future, in order to avoid the lag that official gross domestic product (GDP) and consumer price index (CPI) figures have. Payments Canada has near-real-time access to the payment system variables, which cover a wide range of spending activities and allow for timely macroeconomic forecasting. We use machine learning (ML) techniques to build the nowcasting model. The paper presents preliminary findings on the use of payment data to improve GDP and CPI forecasts. We find that selected macroeconomic indexes and payments system data are co-integrated, and ML modeling demonstrates significant predictive ability for real GDP and CPI. The paper was carried out in collaboration with Georgian college students as part of their internship program at Payments Canada. The project is part of the early stages of Payments Canada's research into nowcasting using payments data.

Keywords: nowcasting, payment data, machine learning, macroeconomic indicators, GDP, CPI.

1. Introduction

The COVID-19 crisis caused severe disruptions in Canadian as well as global economic activity. Frequent economic fluctuations require a rapid response from policy-makers and other decision-makers to support the financial system. The first step in developing any kind of policy or making short-term decisions is to make observations about the current state of the economy.

Macroeconomic variables are important drivers of asset price fluctuations and returns, affecting business cycle analysis and assisting policymakers in directing monetary policy and other types of economic policy. According to Galbraith and Tkacz (2018), the most important indicator of economic performance is the Gross Domestic Product (GDP) growth rate. This indicator provides information on the health and stability of a country and is often a major determinant of whether a country is growing or experiencing a recession. The Consumer Price Index (CPI), which measures the average change in the prices consumers pay for a basket of goods and services over time, is another useful indicator of economic performance. However, policymakers also use other macroeconomic indicators to effectively monitor economic activity, such as the Canadian Financial Stress Index (CFSI), VIX (Adjusted Close), Canadian American exchange rate, S&P/TSX (Adjusted Closed), unemployment rate (percentage), and employment.

In normal circumstances, policy institutions use lagged macro variables in linear models to forecast the actual state of the economy. While CPI is released with a lag of two weeks, GDP is released with a lag of two months and is subject to extensive revision (Bańbura et al. 2013). Due to its slow response and limited capacity to reflect sudden and large effects, this lagged low frequency data and linear regression model became unreliable during economically stressful periods such as the world experienced as a result of the COVID-19 pandemic (Carvalho et al. 2020, Coulombe et al. 2021).

We will employ nowcasting, an approach introduced into economics by Evans (2005) and Giannone et al. (2008), to avoid the delays and uncertainty associated with lagged macroeconomic indicators. This approach entails predicting the recent past, present, and near future of key lagging variables in order to obtain current results about global economic activities. Furthermore, nowcasting assists investors in forecasting real-time economic activity so that they can maximize their investment returns.

The current state of the economy was explored using payments system data in machine learning (ML) models. The Automated Clearing Settlement System (ACSS) was considered for this project because the data captures a variety of transactions, including consumer income and expenditures, business-to-business payments, and government transfer payments in Canada. It has previously been demonstrated that such data contains timely economic information and is thus useful for forecasting (Barnett et al. 2016, Duarte et al. 2017, Aprigliano et al. 2019, Chapman & Desai, 2020). Special interest was focused on retail data as consumption is a component of GDP and CPI, thus payments system data provides an incomplete but direct source of information on changes in these macroeconomic indicators.

Nowcasting research has advanced rapidly in recent years, despite having a long history that dates back to the Mitchell and Burns (1938) classification of hundreds of variables as leading, coincident, and lagging indicators. The availability of new data sources has accelerated the development and application of various econometric techniques. Dimension reduction of a wide cross-section of time series, for instance, is one of the frequently used techniques. This concept served as the foundation for the nowcasting dynamic factor model (DFM) created by Giannone et al. in 2008. Another strategy involves combining data from different frequencies to create a more accurate forecast, such as the mixed-data sampling (MIDAS) models created by Ghysels et al. (2007). Chernis and Sekkel (2017) demonstrate that both models are competitive in their ability to forecast Canada's GDP, even though DFM performs marginally better than the MIDAS model.

Economists now have access to a wider range of data, including financial market data, Google search data, and satellite data. Despite its non-traditional, high-frequency, and large-scale nature, such data has been demonstrated to be useful for economic forecasting and nowcasting (Choi and Varian 2012; Andreou et al. 2013; Li 2016; Donaldson and Storeygard 2016; Buono et al. 2017). Since it has been demonstrated that ML-based prediction approaches fit non-traditional datasets better than traditional econometric tools (Chakraborty & Joseph, 2017; Richardson et al. 2018, Maehashi & Shintani 2020, Chapman & Desai 2022), researchers have begun to employ them.

The research presented in this paper contributes to the existing literature on nowcasting by serving as an investigation into how machine learning techniques can be used in nowcasting as well as a comparison of the performance of ML techniques vis-a-vis the traditional statistical approach. We describe how payment systems work in Section 2 as well as the relationship between payments and economic activity. Section 3 covers data and methodology for machine learning, as well as data exploration. In Section 4, we compare and contrast different models, and in Section 5, we discuss the research's limitations. Many references provide details about payments and explain the different models that we used for nowcasting in this paper.

2. Macroeconomic indexes and the COVID-19

The spread of COVID-19 and the efforts to minimize transmission of the virus have had far-reaching impacts on the economy and the financial system. The growing number of cases necessitated immediate measures such as lockdowns, travel bans, social isolation, and small business closures. During economically unstable times, every measure or decision leads to a fluctuation in the economy (Zamfir and Iordache, 2022). Moreover, these measures were not just an isolated incident but a set of recurrent events dependent on multiple waves of increased COVID-19 cases.

All these fluctuations are well reflected in the macroeconomic index trends (see appendix). In a time of crisis, macroeconomic indexes play an essential role in helping conduct monetary, fiscal, and government policy and making projections for supply-side economics and business cycle analysis. There are two important macroeconomic variables: Gross Domestic Product (GDP) and the Consumer Price Index (CPI). GDP provides information on the health and stability of a country and determines whether a country is growing or experiencing a recession. The CPI measures the average change over time in the prices consumers pay for a basket of goods and services. However, a huge limitation here is that both of these parameters are released with a delay: eight weeks in the case of GDP and two weeks in the case of CPI (see Figure 1). This delay might be acceptable for a lot of people and institutions. However, the latest crisis showed us how frequently disruptions take place and how crucial it is to react quickly to everything that is going on, especially for policy and decision makers.

Macroeconomic Index	Published by	Frequency
Consumer Price Index (CPI)	Statistics Canada	Monthly
Gross Domestic Product (GDP)	Statistics Canada	Quarterly
Volatility Index (VIX)	Chicago Board Options Exchange (CBOE)	Real-time
Unemployment Rate (UNEMPLOY)	Statistics Canada	Monthly
Canadian Financial Stress Index (CFSI)	Bank of Canada	Monthly
Treasury Bills (TBILL)	Federal and provincial governments	Twice a week

Figure 1 – The Overview of Macroeconomic Indexes

It is important to note that we are not underestimating the traditional approach. Lagged GDP and CPI work great in normal circumstances. However, during economically unstable times, using lagged data results in a slow response to unexpected fluctuations. Additionally, macroeconomic indexes are subject to multiple revisions, which lowers their reliability during financially unstable times. Lastly, the traditional approach uses linear regressions that have limited capacity to reflect large effects.

In order to address the limitations of the traditional approach, we use nowcasting, which is considered to be a more non-traditional approach. Nowcasting is defined as the prediction of the present, the very recent past, and the very near future state of an economic indicator. In other words, it is a way to monitor financial and economic conditions in near-real time. It is important to make a very clear distinction between nowcasting and forecasting. Often, confusion occurs between the two definitions: forecasting and nowcasting. Forecasting is more forward-looking and covers longer periods of time, whereas nowcasting focuses on the current moment or very near future. Nowcasting helps to avoid the limitations of the traditional approach by providing near-real time forecasts that help to avoid delays. It uses payment data available on a timely basis, utilizes machine learning models, and captures non-linear correlations.

3. Payment system data and its influence on nowcasting

Payment system data captures a broad range of spending activities. As the data is available on a timely basis, it makes it easier to use for nowcasting purposes. Some research suggests that using payment data can even improve the accuracy of nowcasting. For example, John W. Galbraith and Greg Tkacz (2016) found evidence that debit card transactions can significantly lower errors related to nowcasting.

As we consider consumption expenditure part of GDP, payment system data contains very important features for tracking GDP growth. For instance, the correlation matrix shown below represents relationships between different payment stream variables and macroeconomic indexes. Red in this graph represents a positive correlation, whereas blue indicates a negative one. As we can see, both GDP and CPI are highly correlated to the majority of payment variables. For example, AFT debit and credit, ICP items, online and POS payments, government direct deposits, EDIs, and electronic remittances all have a positive correlation of above 0.7 with GDP and CPI.

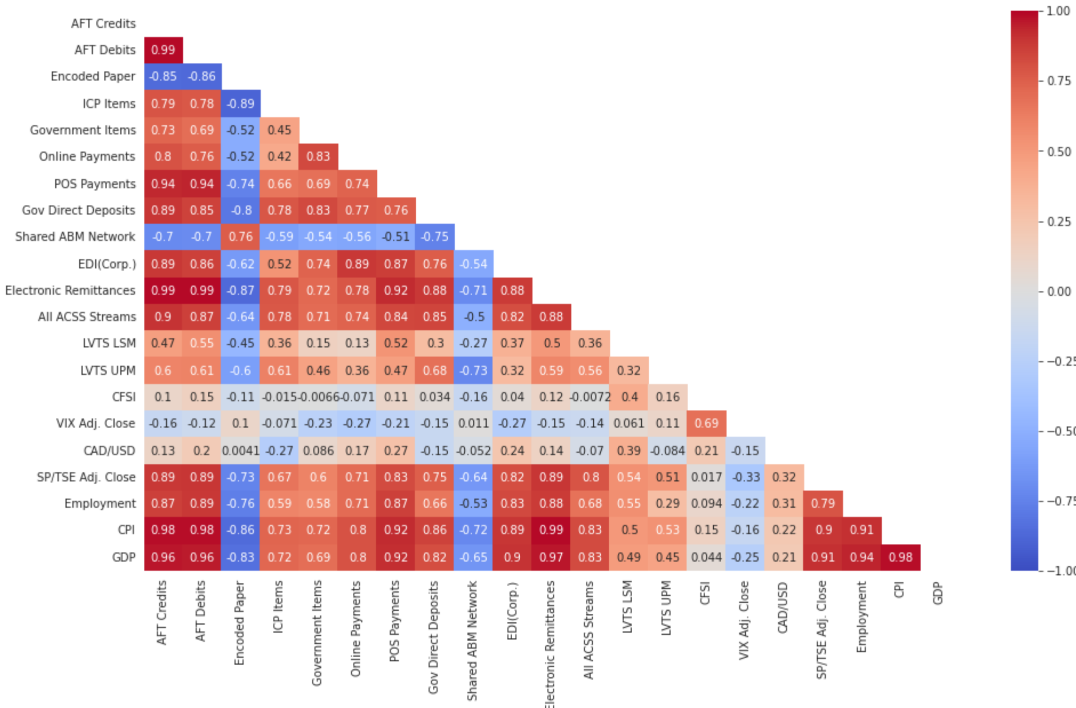


Figure 2 – Correlation Heatmap of payment streams, settlement mechanisms, economic indicators, and macroeconomic variables.

Besides the correlation aspect, Chapman and Desai (2020) in their paper on nowcasting, concluded that there are some other advantages to payment data. Payment data, in particular, has a higher frequency and can be tracked on a daily basis, allowing for near-real-time forecasting. Additionally, retail data carries timely information about the economy (e.g. income and expenditure data, B2B payments).

However, it is important to highlight that consumers switch between different payment methods from time to time for several reasons, which can bias our results if we use few payment system data variables. Due to this assumption, we use several technologies in the model to endogenize consumers' choices (John W. Galbraith and Greg Tkacz, 2016).

3.1 Retail batch payment system

There are two main payment systems: the retail batch payment system and the high-value payment system. The retail system consists of the Automated Clearing Settlement System

(ACSS) and the United States Bulk Exchange application (UBSE). The retail batch payment system clears paper-based and electronic payments[1] (Figure 2).

Paper-based	Electronic
● Cheques ● Paper remittances ● Government items ● Government cheques ● Redeemed bonds ● Treasury bills ● Coupons	● Direct deposits ● Electronic data interchange ● Electronic remittances ● Image paper items ● Online debits and credits ● Point of service debit and credit ● Pre-authorized debits ● Shared ABM Networks

Figure 2 - ACSS paper-based and electronic payment streams

3.2 High-value payment system

Payments Canada introduced Lynx, a new high-value payment system that replaced the Large Value Transfer System (LVTS) in 2021. It is an electronic wire system responsible for the transfer of payments between participating financial institutions, in Canadian dollars. Payments within Lynx are both final and irrevocable. As an economy, Canada depends on the exchange of billions of dollars each day. Payments Canada ensures these financial transactions are carried out safely and securely. Part of this work includes looking into ways to improve the payment clearing and settlement process. To do so, Payments Canada monitors payments data from Lynx to keep track of the country's economic activity. This is critical because the value and volume that flow through our systems can influence the performance of its member financial institutions, policies, payment system development, and overall decision-making.

4. Data and methodology

In this paper, we will use the nowcasting approach, which entails predicting key lagging variables in the recent past, present, and near future in order to obtain current results about global economic activities. The current state of the economy was explored using payment system data in machine learning models. ACSS was considered for this project because the data captures a variety of transactions, including consumer income and expenditures, business-to-business payments, and government transfer payments in Canada. It has previously been demonstrated

[1] For more, please see: https://www.payments.ca/retail-payments

that such data contains timely economic information and is thus useful for forecasting (Chapman & Desai, 2020). Special interest was focused on retail data as consumption is a component of GDP and CPI, thus payment system data provides an incomplete but direct source of information on changes in GDP.

4.1. Machine learning

Machine learning (ML) utilizes computer algorithms for learning to perform tasks. Humans, for example, have a proclivity for learning how to complete a task, making accurate predictions, and behaving intelligently. This type of learning is almost always based on observations, experiences, data, or information. As a result, machine learning is concerned with teaching machines to perform better in the future based on previous experiences (Schapire, 2008).

The essence of machine learning is automation. In other words, its objective is to devise learning algorithms that do the learning automatically without human intervention or assistance. "Programming by example" can be applied to any machine learning process. Often, we are required to accomplish certain tasks, like spam filtering. But instead of programming the computer to solve the task directly, in machine learning, we employ techniques by which the computer develops its own program based on examples that we provide. Machine learning is a sub-area of artificial intelligence (AI) that involves other domains, especially statistics, mathematics, physics, computer science, and more (Schapire, 2008).

4.2. Payment data

The Automated Clearing Settlement System (ACSS) and the Large Value Transfer System (LVTS) are two of Payments Canada's payment systems that were used in this project. Payments Canada's modernized high-value payment system, Lynx, became operational during the development of this project (Payments Canada, 2017a). We only used LVTS for high-value payment data because Lynx was launched in September 2021 and there were not enough data points to work with during this project. The data was collected between January 1, 2000 and August 31, 2021 for both payment systems. Saturdays, Sundays, and certain statutory holidays (e.g., New Year's Day, Good Friday, etc.) are non-operational days for Payments Canada systems (Payments Canada, 2017c). As a result, these dates are not included in the payment data for both payment systems.

4.2.1. Automated Clearing Settlement System data

The data obtained from the Automated Clearing Settlement System (ACSS) was aggregated at the daily level by the stream. Each row item includes:

- Date (formatted as Year-Month-Day): it's the day when the transaction is made
- Stream: stream-ID
- Stream description: stream label
- Stream type: Debit Payments, Electronic Fund Transfer Payments (EFT), or Paper Payments.
- Sent, received, and total volume are the daily number of transactions completed within the ACSS for a particular stream on a given day
- Sent, received, and total value are the dollar amounts of the transactions completed within the ACSS for a particular stream on a given day

The sent volume and received volume are equal, as are the sent value and received value, because if a direct clearer sends x volume/value, the receiver will receive the same amount. Total volume was the sum of sent volume and received volume, while total value was the sum of sent value and received value.

4.2.2. Large Value Transfer System data

The Large Value Transfer System (LVTS) data was aggregated at the daily level by different settlement mechanisms:

- Date (formatted as Year-Month-Day): it's the day when the transaction is made
- Settlement mechanism: UPM, LSM, and RTM
- Stream: SWIFT message (MT103/MT205)
- Sent, received, and total volume are the daily number of transactions completed within the ACSS for a particular stream on a given day.
- Sent, received, and total value are the dollar amounts of the transactions completed within the ACSS for a particular stream on a given day.

The settlement mechanisms within the LVTS are the Urgent Payment Mechanism (UPM)[2], Real-Time Mechanism (RTM)[3], and Liquidity Saving Mechanism (LSM)[4], which are used to establish credit for the exchange of payments between different participants. Therefore, they are new with the implementation of Lynx. The stream variable contained many missing values, so it was not included in any analysis.

Unlike the ACSS, sent volumes, received volumes, and total volumes, as well as sent values, received values, and total values, are all equal within the LVTS.

4.3. Initial exploratory data analysis

The Python libraries NumPy, Pandas, matplotlib, datetime, and seaborn were used for exploratory data analysis. The integrated development environment for exploratory data analysis, feature engineering, and machine learning model deployment and evaluation was Jupyter Notebooks.

Histograms and box plots were used for ACSS data to explore the distribution of values for the variables sent volume, received volume, total volume, sent value, received value, and total value. Visualization of the data confirmed that values for sent volume and received volume were equivalent, as were the values for sent value and received value (Figure 3). All variables were skewed to the right.

[2] The Urgent Payment Mechanism (UPM) is designed to settle payments on an individual gross basis. By definition, payments routed to this mechanism are urgent and must be settled without any delay to provide certainty of settling the payment before a specific time or as soon as possible.

[3] Real-Time Mechanism (RTM) is designed to settle payments during the finalization window each day when participants are attempting to flatten their positions using interbank payment messages.

[4] Liquidity Saving Mechanism (LSM) is designed to reduce the amount of intraday liquidity required to settle a payment where the participant does not require the payment to settle immediately.

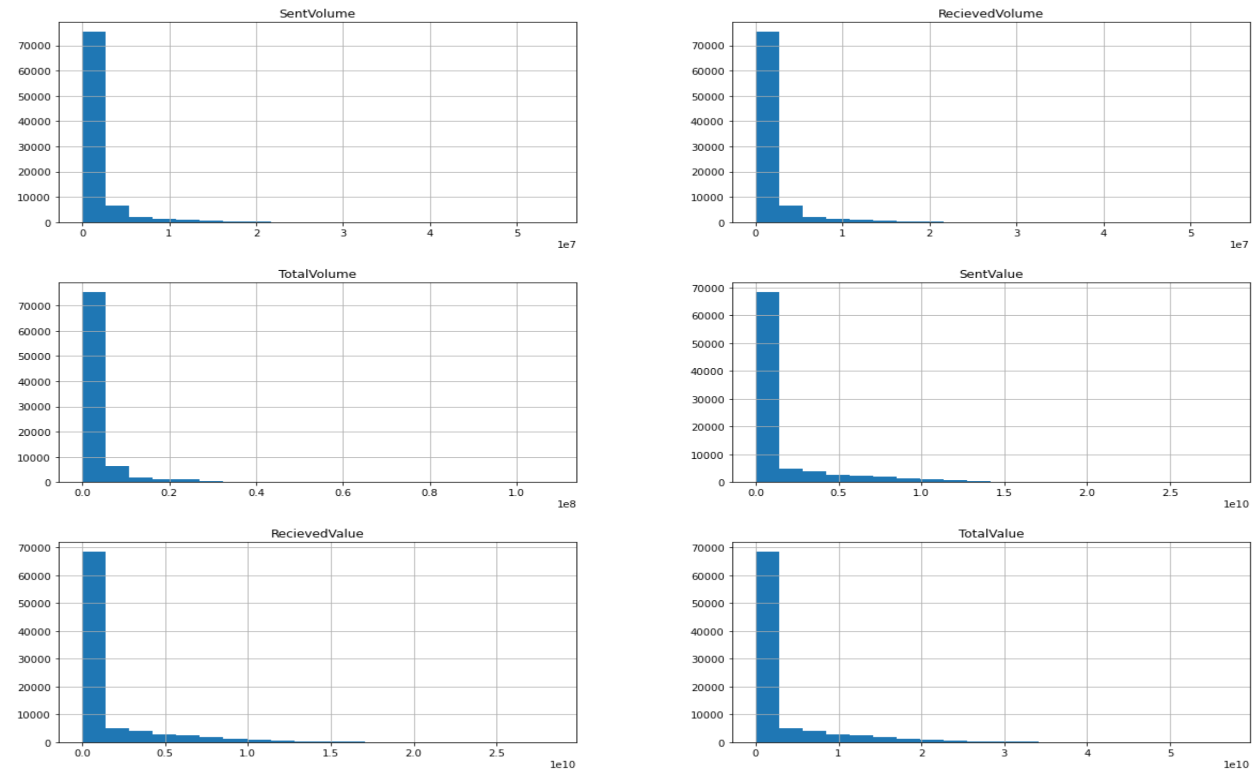
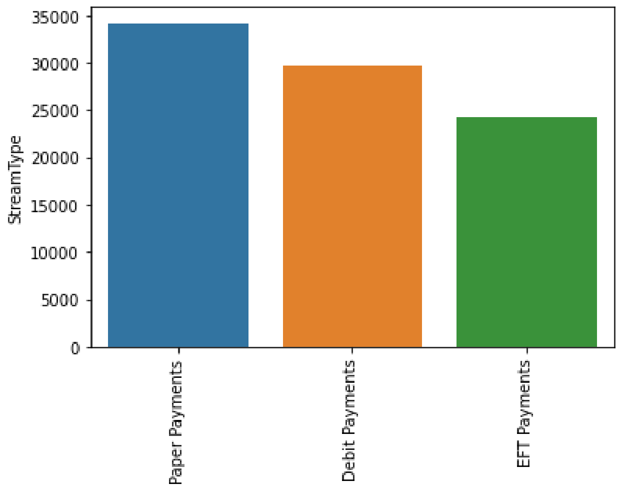


Figure 3 – Distribution of Sent, Received and Total Volumes and Sent, Received and Total Values within ACSS data.

Paper Payments were the most prevalent stream group, followed by Debit Payments, and finally Electronic Fund Transfer Payments, according to the box plot for counts of records by stream groups joined together based on stream type similarities (Figure 4).

Bar plots for counts of records by ACSS stream (Figure 5) were plotted. Payment streams that appeared most frequently were AFT Debits, Shared ABM Networks, Point of Service – Credit, Point of Service – Debit, Unqualified <$50,000, EDI (Corporate to Corporate), Computer Rejects, Electronic Remittances, Paper Based Remittances, Encoded <$50,000 and AFT Credits. Paper Image Returns (National), SETs, and Government Image Exchange were the payment streams that were used the least within the ACSS data. It should be noted that streams within ACSS have not been static within the 20 years that the data encompasses. Some streams have been discontinued (i.e. Large, >$50,000 in 2014) while others have been implemented more recently (i.e. SETs in 2020).



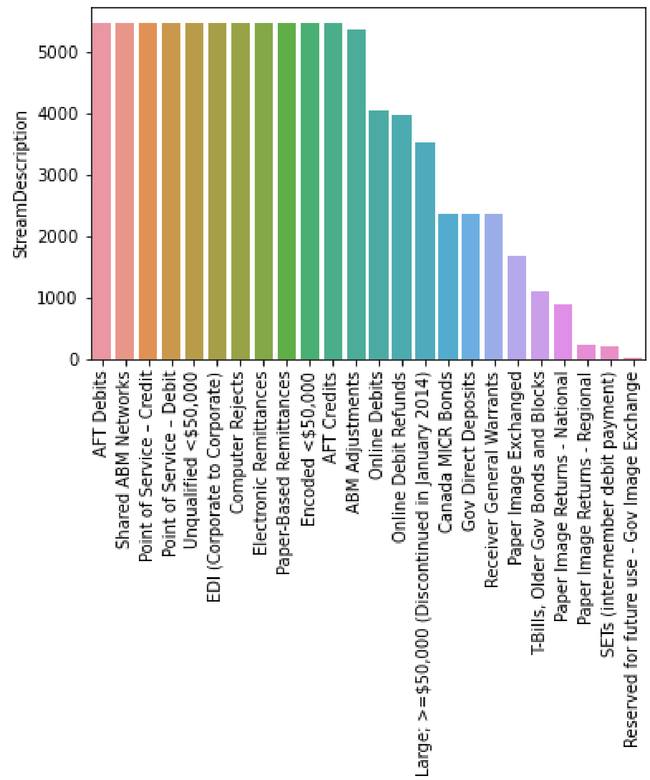


Figure 4 - Counts of records by stream groups within ACSS data.

Figure 5 - Counts of records of streams within ACSS data.

For the remaining analysis, only received values and received volumes were used. Additional analysis revealed the payment stream Point of Service (Debit) had the amount of received volume at 13 million payments (Figure 6). However, for received value, the payment stream Large, >=$50,000, settled the greatest value of payments at nearly 8.7 billion dollars (Figure 7).

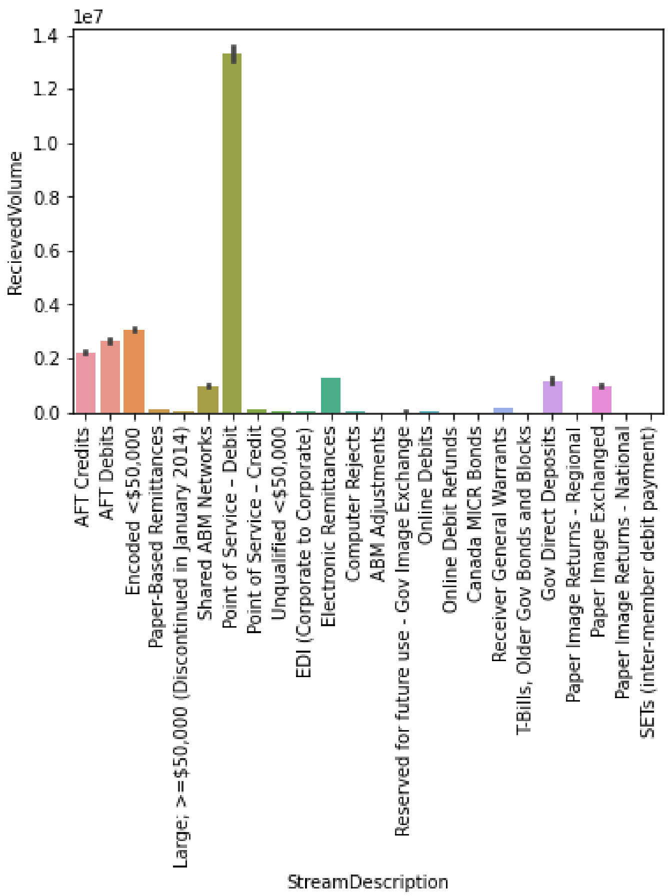


Figure 6 - Received Volumes by Payment Stream within ACSS

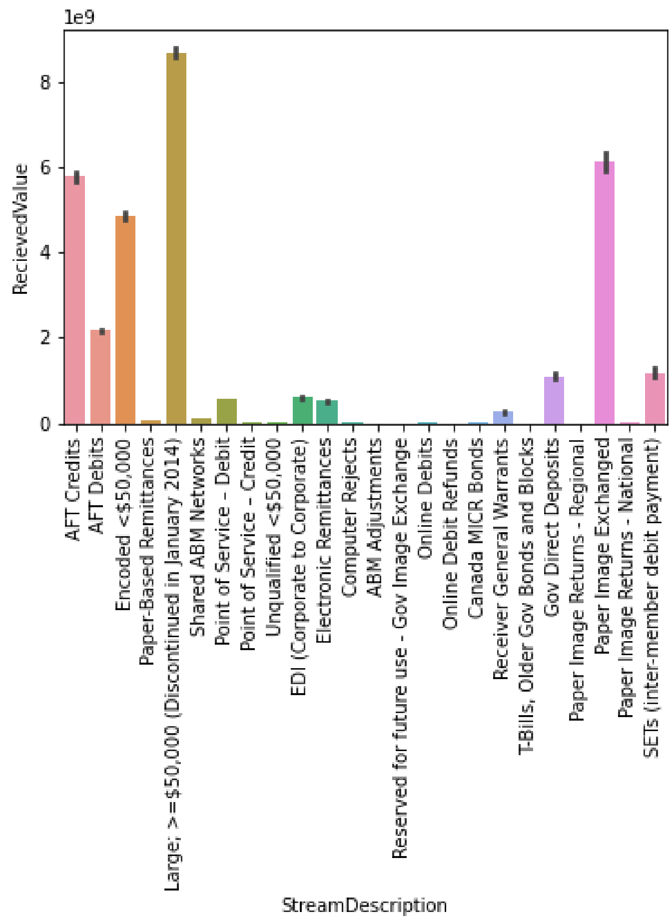


Figure 7 - Received Value by Payment Stream within ACSS

Unlike the manner in which payment streams were explored within the ACSS data, the settlement mechanisms from the LVTS data were additionally evaluated by looking at average values and average volumes to get more insights on how these mechanisms perform and whether they can contribute to the accuracy of nowcasting. The Real-Time Mechanism (RTM) had the highest average value of the three settlement mechanisms in terms of received value. The Liquidity Saving Mechanism (LSM) received the most volume on average (Figure 8). The data used in the machine learning models was not limited to just one of the high-value payment systems because settlement mechanisms were used.

Settlement Mechanism	Average received volume	Average received value
Liquidity Saving Mechanism	28401.9	8.799e+10
Real-Time Mechanism	15.5	2.664e+11
Urgent Payment Mechanism	328.9	3.415e+10

Figure 8 - Average received volumes and average received values for settlement mechanisms.

4.4. Feature Engineering

A new DataFrame was created using the pandas library, containing the received values of each stream description aggregated at the monthly level. This DataFrame consisted of the 24 ACSS stream descriptions found within the original data set plus an additional column which was the sum of all ACSS payment streams (called "All Streams"). All other variables (i.e. total values, received volumes, etc.) were excluded from here onwards.

GDP data was obtained from Statistics Canada (Statistics Canada, n.d.) and ranged from January 2000 to June 2021. The GDP data was merged into the ACSS DataFrame. The GDP values obtained were in millions and using chained dollars[5]. Additional economic indicators and macroeconomic variables were obtained from Yahoo! Finance and merged into the DataFrame.

From the LVTS data, the received values were aggregated at a monthly level for each settlement mechanism. This aggregated data was then merged into the main DataFrame. The ACSS payment streams, LVTS settlement mechanisms, GDP, and additional economic indicators and macroeconomic variables were all measured at a monthly level from January 2000 to June 2021. Since there was no payment data included past June 2021, the RTM settlement mechanism was excluded from the dataset.

As was discussed earlier, the different payment streams within the ACSS payment system were not static within the dataset. Certain ACSS payment streams were grouped together to account for the additions and removals of these streams. These groupings were based on similar

[5] Chained dollars is a method used to adjust real dollar amounts for inflation over time, to allow the comparison between different years using 2012 as the base year.

groupings used by Chapman and Desai (2021) in a similar study, as well as the rules governing ACSS (Payments Canada, n.d.). The ACSS payment streams were grouped as follows:

- AFT Debit: sum of ‘AFT Debits’ and ‘Paper Based Remittances’
- Encoded Paper: sum of ‘Encoded <\$50,000’, ‘Large; >\$50,000’, ‘SETs’, ‘Unqualified <\$50,000’
- ICP Items: this takes the sum of ‘Paper Image Exchange’ (Regional & National) and subtracts ‘Paper Image Returns’ (Regional & National)
- Government Items: sum of ‘Canada MICR Bonds’, ‘Receiver General Warrants’, ‘T-Bills, Older Government Bonds and Blocks’ and ‘Reserved for future use-Government Image Exchange’
- Online Payments: the difference between ‘Online Debits’ and ‘Online Debit Refunds’
- POS Payments: the difference between ‘POS Debit’ and ‘POS Credit’
- Shared ABM Networks: the difference between ‘Shared ABM Networks’ and ‘ABM Networks’

After the ACSS payment streams were grouped, certain ungrouped streams were removed from the DataFrame. The final variables used are listed in Figure 9.

ACSS Variables	LVTS Variables	Other Variables
AFT Credits AFT Debits Encoded Paper Image Captured Payment (ICP) Items Government Items Online Payments Point of Sale (POS) Payments Government Direct Deposits Shared ABM Network EDI Corporate to Corporate Electronic Remittances All ACSS Streams	Settlement mechanisms	CFSI VIX (Adj. Close) CAD/USD SP/TSE (Adj. Close) GDP CPI Employment

Figure 9 – Final variables after completion of feature engineering.

4.5. Additional Exploratory Data Analysis

Once feature engineering was completed, further exploratory data analysis was conducted on the entire dataset. Histograms of the values (received values from ACSS and LVTS data) were plotted, as were boxplots (see Appendix). The distributions amongst the variables were all quite different.

Distributions appeared to be skewed left, skewed right, and bimodal (Figure 32). No variables had a normal distribution, but some appeared to be approaching normal shape (i.e. POS Payments) (Figure 32).

Historically, Sweden and South Korea, two of the most cashless societies in the world, have shifted to digital payments, which can lead to an increase in GDP over time. Some research shows that the switch to digital payments can increase the annual GDP by three percentage points (Mass et al., 2019), which explains some of the correlations that we found in this project (Figure 10).

Correlations were conducted between each of the finalized variables (Figure 10). All ACSS stream groupings had a moderate to strong correlation with both GDP and CPI. Encoded Paper and Shared ABM Networks were the only ACSS streams that had negative correlations. Both AFT Credits and AFT Debits had very strong positive correlations with GDP, each at 0.96. Historically, the switch to digital payments. These two groupings also had a strong positive correlation with CPI, each at 0.99. Settlement mechanisms UPM and LSM only had a moderate positive correlation with GDP, at 0.45 and 0.49, respectively. Similarly, Settlement mechanisms UPM and LSM had moderate positive correlation with CPI at 0.53 and 0.5, respectively. All payment variables had poor correlations with the CFSI, VIX, and CAD/USD (see Figure 2).

Finally, we plotted all our economic variables as a time series (Figure 11). However, with the start of the COVID-19 pandemic in early 2020, there was a huge increase in value within the settlement mechanisms related to the series of significant actions that the Bank of Canada has undertaken to support the financial system. Many of which involved corresponding payments in the high-value payment system. As a result, the Bank of Canada sent an extraordinary value of payments in 2021, disrupting historical trends and making it difficult to discern between those payments and "business as usual" payments processed via the high-value payment system. This increase has yet to resume to pre-pandemic levels.

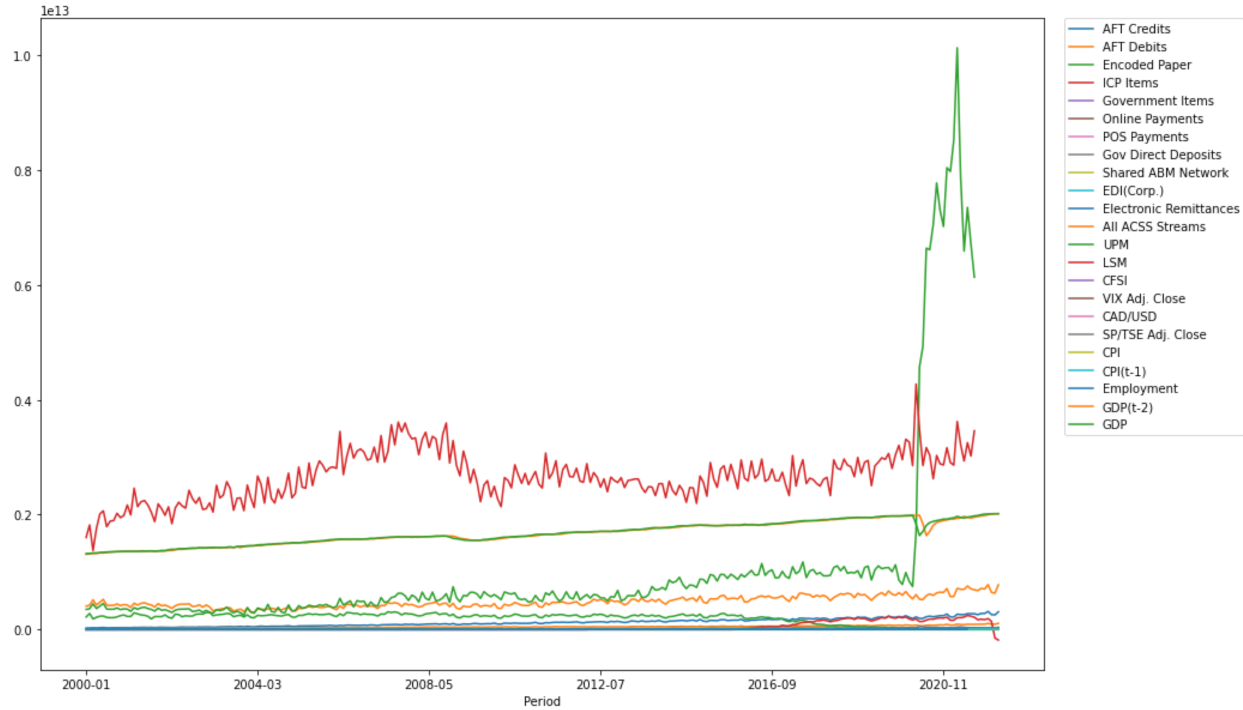


Figure 11-Time series plot of values of payment streams, settlement mechanisms, GDP, and other variables.

5. Machine learning model evaluation

After the collected data was cleaned, formatted, feature-engineered, analyzed, and visualized, it was ready to be fed into an optimal machine learning algorithm that would produce the most accurate nowcasting predictions for GDP and CPI.

It is important to note that since both GDP and CPI are continuous variables, only regression ML models were used for this major research project. For both GDP and CPI nowcasting, eight regression models were considered for analysis. These include Multi-Linear Regression (MLR), Ridge (L2), Lasso (L1), Elastic Net Regression, Decision Trees, Random Forests, LightGBM, LightGBM, and XGBoost.

Initially, we start by importing several packages such as Python's scikit-learn, NumPy, Pandas, and matplotlib and conducting the seasonal adjustment of macroeconomic as well as payment variables. Then, the data schema was split along the row axis into training and test data using sequential timestamp data. There are 13 features in the independent or predictor variables, 12 of which are representative of the grouped payment streams in ACSS, as well as an aggregate of

ACSS payment streams (Figure 9). Furthermore, we mentioned in Section 2 that GDP is quarterly lagged while the CPI is monthly lagged, so their historical values would be useful for nowcasting (Figure 1). Therefore, we used GDP(t-2) and CPI(t-1) to predict the target variables GDP(t) and CPI(t), respectively:

$$\widehat{GDP}_t = f(payments_t\,,\ GDP_{t-2}),$$

$$\widehat{CPI}_t = f(payments_t\,,\ CPI_{t-1}).$$

All the candidate models were evaluated using five-fold cross-validation and the coefficient of determination (R^2) scores were observed as measures of baseline performance for both GDP and CPI. All used models were then fine-tuned (i.e., hyperparameters were optimized) using the grid search technique. The baseline and fine-tuned performance of the various regression models used to calculate GDP and CPI are shown in the appendix.

R^2 is a statistical measure that represents the proportion of variance for a dependent variable that can be explained by an independent variable(s). The R^2test score for various models is an effective measure of their capabilities to nowcast GDP or CPI. To generate the most accurate model, all models were fine-tuned using grid search, and the best model hyperparameters were determined. After fine-tuning, the R^2 scores of the models were evaluated. The Ridge Regression, Lasso Regression, and Elastic Net Regression models were the most accurate fine-tuned models for both GDP and CPI (Figure 13). Therefore, we selected these three models for deployment for sequential timestamp data.

6. Model deployment

6.1 Sequential data

For predicting the GDP and CPI, we used sequential data to investigate its impact on the accuracy of our predictions. As a result, we came up with three models: Ridge Regression, Lasso Regression, and Elastic Net Regression, which we fine-tuned using the grid search method. After splitting the data into 55% for training (from 2000 until 2011) and 45% for testing (from 2012 until 2022), non-shuffled this time, we came up with a test dataset that consisted of 120 records out of 267 records. After the model was implemented, its results were displayed in the form of line or time-series plots for actual and predicted figures.

Working with sequential data is extremely important while dealing with data problems where the ordering matters, which is the case in time series. We consider different ML models as each method gives slightly different results so we need to be mindful while selecting the right model for nowcasting purposes. However, it is important to note that ML models have their own drawbacks and limitations in forecasting even if we work with sequential data (Section 6).

6.1.1 Ridge Regression

The fine-tuned Ridge Regression model consisted of the following hyper-parameter settings in Python 3 code:

1. alpha : 0.03
2. fit_intercept : True
3. normalize : True
4. solver : sag

The importance of each feature when it comes to predicting the GDP figures using test data between 2017 and 2022, with the fine-tuned Ridge Regression model was evaluated (Figure 18). It can be observed that GDP(t-2) has the highest weightage, while POS Payments and Electronic remittances also have considerable importance (Appendix).

Figure 19 shows the trends for GDP actual numbers and the predicted GDP values using the Ridge Regression model. We notice that the predicted curve closely follows the actual curve which results in higher accuracy of our predictions.

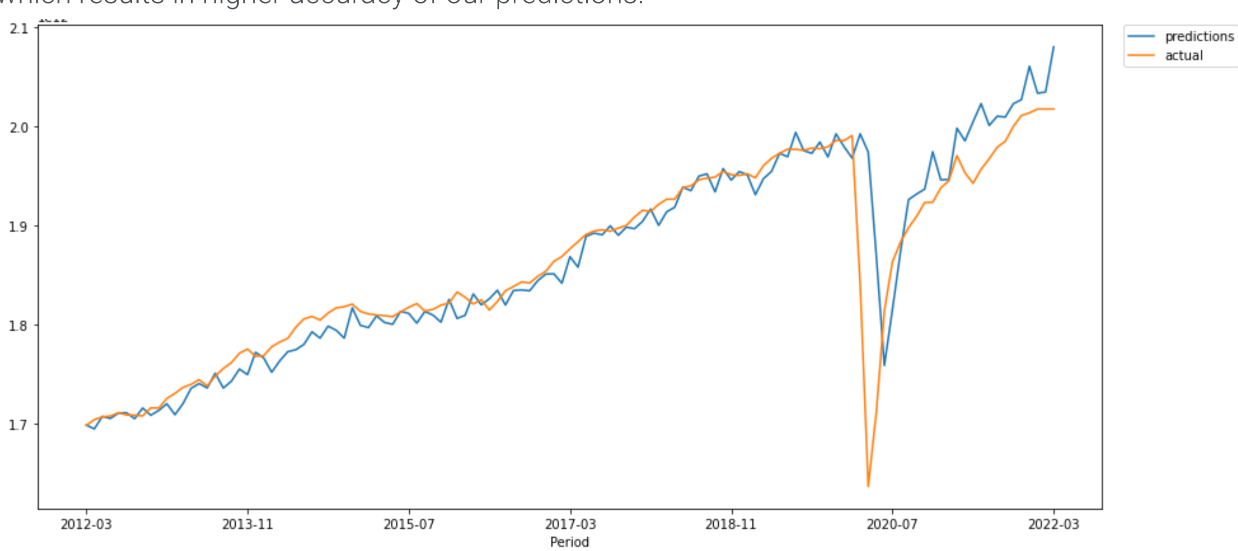


Figure 19 - Line plot for GDP Actuals vs. Predictions

The importance of each feature when it comes to predicting the CPI figures using test data between 2017 and 2022, with the fine-tuned Ridge Regression model was evaluated (Figure 20). It can be observed that CPI(t-1) has the highest weightage by far compared to all other ACSS payments streams (Appendix).

Figure 21 shows the trends for CPI actual numbers and the predicted CPI values using the Ridge Regression model. The predicted curve does behave similarly with the actual curve but we notice from 2016, our predictions started increasing significantly compared to the actual values which is an indication that it is a good model even though it's overperforming.

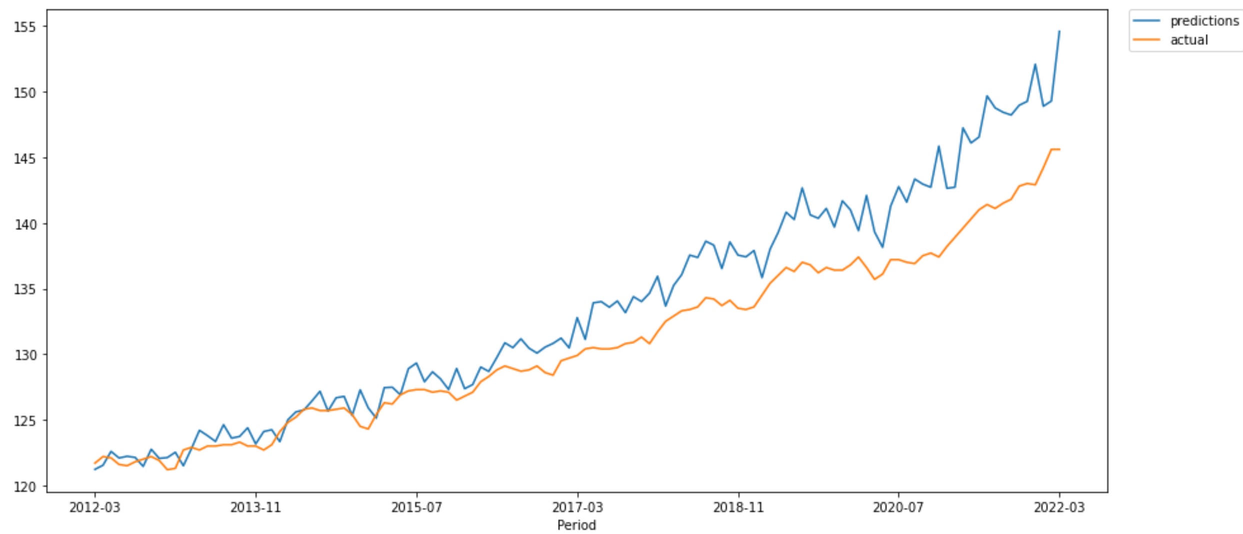


Figure 21-Lines plot for CPI Actuals vs. Predictions

6.1.2 Lasso Regression

The fine-tuned Lasso Regression model consisted of the following hyper-parameter settings in Python 3 code:

1. alpha : 0.01
2. fit_intercept : True
3. normalize : True
4. selection : cyclic

The significance of each feature in predicting GDP figures using test data from 2017 to 2022 with the fine-tuned Lasso Regression model was assessed (Figure 22). It can be observed that

GDP(t-2) has the highest weightage compared to all other ACSS payments streams, even though AFT Debits and ICP items have considerable importance (Appendix).

Figure 23 shows the trends for GDP actual numbers and the predicted GDP values using the Lasso Regression model. The predicted curve behaves similarly to the actual curve, indicating that we have very good predictions because both curves follow nearly the same pattern; even during the COVID-19 period, the model predicted the abrupt decrease in GDP fairly well.

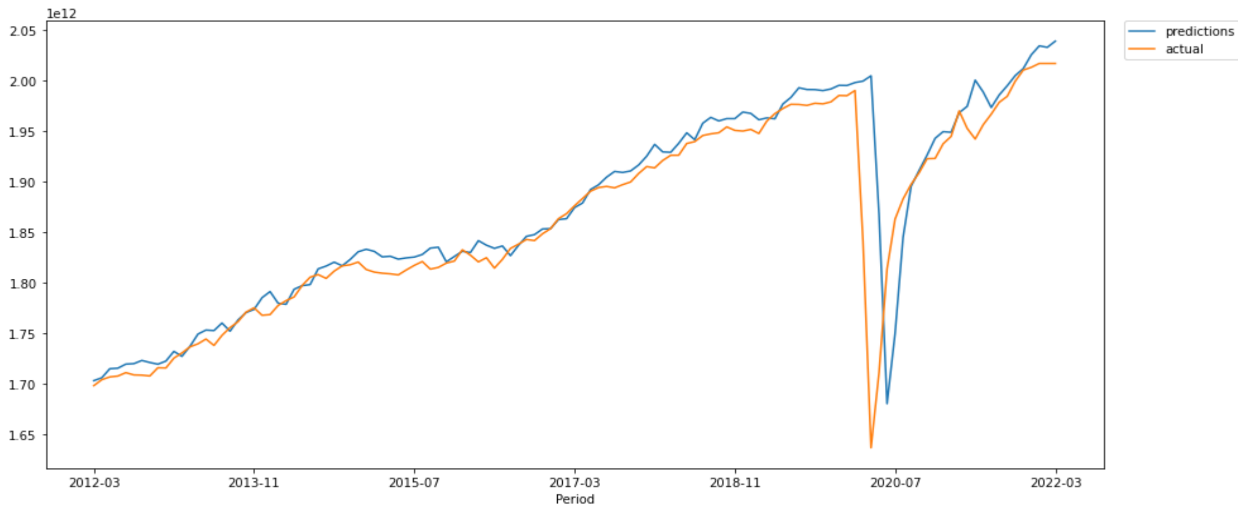


Figure 23 - Line Plot for GDP Actuals vs. Predictions

The significance of each feature in predicting CPI figures using test data from 2017 to 2022 using the fine-tuned Lasso Regression model was assessed (Figure 24). It can be observed that CPI(t-1) has the highest weightage by far compared to all other ACSS payments streams (Appendix).

Figure 25 shows the trends for CPI actual numbers and the predicted CPI values using the Lasso Regression model. The predicted curve behaves similarly to the actual curve, indicating that we have made very good predictions because both curves follow nearly the same pattern.

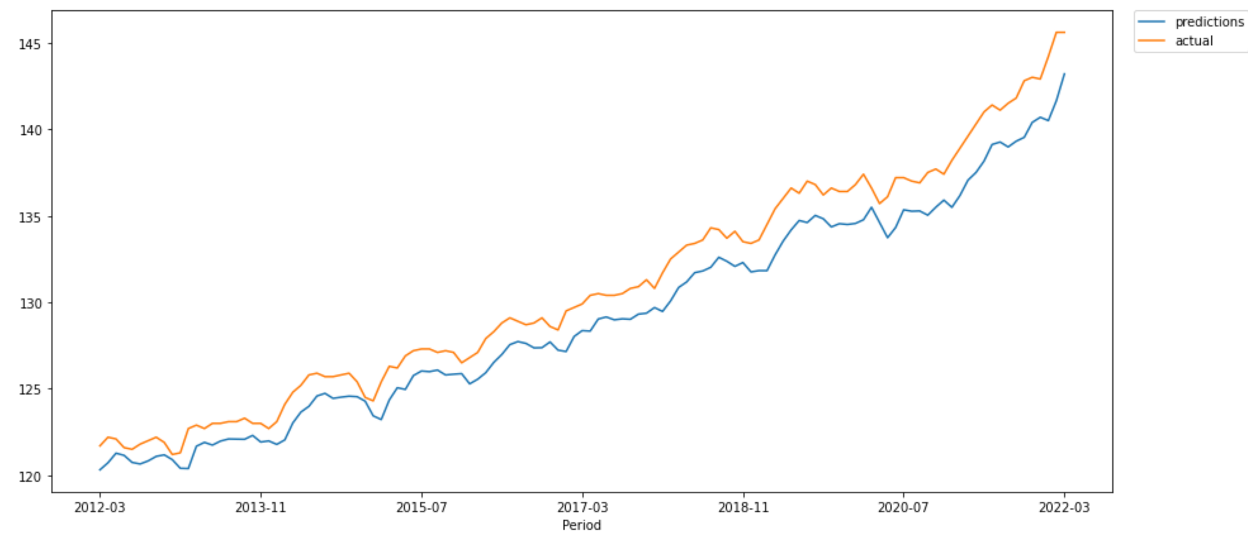


Figure 25-Lines plot for CPI Actuals vs. Predictions

6.1.3 Elastic Net Regression

The fine-tuned Elastic Net Regression model consisted of the following hyper-parameter settings in Python 3 code:

1. alpha : 0.0001
2. fit_intercept : True
3. l1_ratio : 0.7
4. normalize : True
5. selection : cyclic

The importance of each feature when it comes to predicting the GDP figures using test data between 2017 and 2022, with the fine-tuned Elastic Net Regression model was evaluated (Figure 26). It can be observed that GDP(t-2) has the highest weightage, while POS Payments and Electronic remittances also have considerable importance (Appendix).

Figure 27 shows the trends for GDP actual numbers and the predicted GDP values using the Elastic Net Regression model. We notice that the predicted curve behaves similarly to the actual curve, which means that our predictions are fairly accurate.

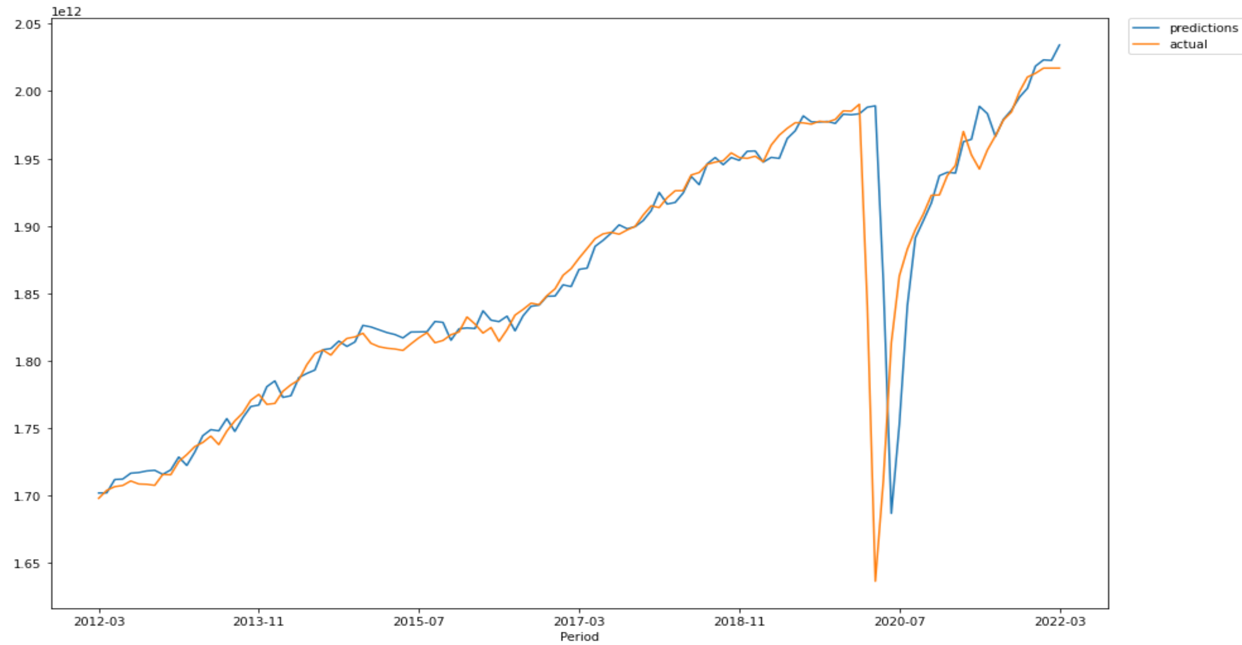


Figure 27 - Line Plot for GDP Actuals vs. Predictions

The significance of each feature in predicting CPI figures using test data between 2017 and 2022 with the fine-tuned Elastic Net Regression model was assessed (Figure 28). It can be observed that CPI(t-1) has the highest weightage by far compared to all other ACSS payments streams (Appendix).

Figure 29 shows the trends for CPI actual numbers and the predicted CPI values using the Elastic Net Regression model. The predicted curve follows almost the same pattern as the actual curve, but starting in 2016, our predictions began increasing significantly in comparison to the actual values, indicating that we may have a good model that behaves very similarly to the Ridge Regression model that we used previously.

Figure 29-Lines plot for CPI Actuals vs. Predictions

7. Drawbacks and limitations

This research paper is subject to several limitations. First of all, there is a possible drawback that can be related to the loss of interpretability and overfitting problems. When it comes to using machine learning models, it is important to choose the right hyperparameters while fine-tuning, and accurately split the data using the right subsets while using cross-validation techniques. Otherwise, there is a chance of the model results being too closely aligned to a limited dataset, which results in overfitting problems that are technically unavoidable due to the dependency between GDP parameters and payment data.

Predicting macroeconomic variables during a crisis, such as COVID-19, is a serious problem because the economy is experiencing a large and unprecedented shock, making our historical data less informative in terms of forecasting and causing disruptions in nowcasting accuracy. Furthermore, the COVID-19 has fundamentally altered consumer behavior, making forecasting more difficult. CPI, for example, is a measure of expenditures reported by Canadians based on a predetermined basket of goods and services. Quantifying the cost of a fixed basket over time allows for a consistent measurement of pure price change. The fixed-basket approach has worked relatively well under normal economic conditions. However, consumers have changed their spending patterns because of the most recent pandemic crisis, which results in a misalignment between consumer actual spending and its macroeconomic estimation.

Despite the variety of ACSS payments system data that involve a broad range of spending activities, they still do not capture all retail payment instruments such as credit card and e-transfer methods which saw an important growth during COVID-19 period, pointing to some limitations of these payment methods in nowcasting GDP growth and CPI growth (Chapman and Desai, 2020)[6]. We should also mention that ACSS payment streams are sometimes facing policy changes and technological advancements, which makes them unstable, so we need to be cautious and aware of any changes that could mislead our predictions.

8. Conclusion

Based on the case of the most recent COVID-19 crisis, this research project demonstrates the importance of timely macroeconomic forecasts for policy and decision makers. During the course of this paper's analysis, we discovered a strong correlation between payment system data and macroeconomic indicators, specifically GDP and CPI.

We used a sequential model followed by fine-tuning of certain hyperparameters. By comparing eight different machine learning models, we discovered the best-tested models, which included Ridge Regression, Lasso Regression, and Elastic Net Regression. Although we obtained statistically accurate predictions for both GDP and CPI growth, the Elastic Net performed the best in terms of predicting fluctuations. The top three dominant features for GDP predictions were GDP(t-2), Online Payments, and AFT Credits. The top three dominant features for CPI predictions were CPI(t-1), Online Payments, and AFT Credits. When compared to ACSS payment data, LVTS settlement mechanism data proved to be far less important.

However, this research project is subject to multiple limitations. It is suggested that future research seeks a better balance between maximizing model accuracy and overfitting the model. Furthermore, adding stress testing to this research topic could benefit it by observing how the developed model responds to various economic disruptions. Finally, looking into correlations between payment data and other macroeconomic indicators like unemployment, the Canadian Financial Stress Index (CFSI), and the Product Price Index (PPI) could provide more insight into

[6] For more, you can see: https://www.bankofcanada.ca/wp-content/uploads/2021/01/swp2021-2.pdf

the methodology and potentially broaden the model's application to forecasting other macroeconomic indicators.

9. References

Andreou, E., E. Ghysels, and A. Kourtellos (2013): “Should Macroeconomic Forecasters Use Daily Financial Data and How?,” Journal of Business & Economics Statistics, 31(2), 240–251.

Aprigliano, V., G. Ardizzi, L. Monteforte, et al. (2019). Using the payment system data to forecast the economic activity. International Journal of Central Banking 15 (4), 55–80. https://www.ijcb. org/journal/ijcb19q4a2.pdf.

Bańbura, M., Giannone, D., Modugno, M, and Reichlin, L. (2013). Now-casting and the Real-Time Data Flow. European Central Bank Working Paper No. 1564, July 2013, https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1564.pdf

Banbura, ´ M., D. Giannone, and L. Reichlin (2011): “Nowcasting,” in Oxford Handbook on Economic Forecasting, ed. by M. P. Clements, and D. F. Hendry, pp. 63–90. Oxford University Press.

Barnett, W., M. Chauvet, D. Leiva-Leon, L. Su, et al. (2016). Nowcasting nominal GDP with the credit-card augmented divisia monetary. Technical report, The Johns Hopkins Institute for Applied Economics. https://ideas.repec.org/p/pra/mprapa/73246.html.

Carvalho, V. M., S. Hansen, A. Ortiz, J. R. Garcia, T. Rodrigo, S. Rodriguez Mora, and P. Ruiz de Aguirre (2020). Tracking the COVID-19 crisis with high-resolution transaction data. https: //www.repository.cam.ac.uk/bitstream/handle/1810/310898/cwpe2030.pdf?sequence=5.

Chakraborty, C. and A. Joseph (2017). Machine learning at central banks. Technical report, Bank of England Working Paper No. 674. https://ssrn.com/abstract=3031796.

Chapman, J. T. and A. Desai (2020). Using payments data to nowcast macroeconomic variables during the onset of COVID-19. Journal of Financial Market Infrastructures 9 (1). www.bankofcanada. ca/2021/01/staff-working-paper-2021-2.

Chapman, James, and Ajit Desai. Macroeconomic Predictions Using Payments Data and Machine Learning. No. 2022-10. Bank of Canada, 2022

Chernis, T. and R. Sekkel (2017). A dynamic factor model for nowcasting Canadian GDP growth. Empirical Economics 53 (1), 217–234.

Choi, H. and H. Varian (2012). Predicting the present with Google Trends. Economic Record 88, 2–9.

Coulombe, P. G., M. Marcellino, and D. Stevanovic (2021). Can machine learning catch the COVID19 recession? National Institute Economic Review 256, 71–109. doi: https://doi.org/10.1017/nie.2021.10.

Donaldson, D. and A. Storeygard (2016). The view from above: Applications of satellite data in economics. Journal of Economic Perspectives 30 (4), 171–198.

Duarte, C., P. M. Rodrigues, and A. Rua (2017). A mixed frequency approach to the forecasting of private consumption with ATM/POS data. International Journal of Forecasting 33 (1), 61–75. doi: 10.1016/j.ijforecast.2016.08.003.

Evans, M. D. D. (2005): "Where Are We Now? Real-Time Estimates of the Macroeconomy," International Journal of Central Banking, 1(2).

Galbraith, J. and G. Tkacz (2007). *Electronic transactions as high-frequency indicators of economic activity. Technical report, Bank of Canada.* https://www.bankofcanada.ca/wp-content/uploads/2010/02/wp07-58.pdf

Galbraith, J. W., & Tkacz, G. (2016). *Nowcasting with payments system data.*

http://johngalbraith-economics.mcgill.ca/adebit_2016.pdf

Galbraith, J. W., & Tkacz, G. (2018). *Nowcasting with payments system data. International Journal of Forecasting*, *34*(2), 366–376. https://doi.org/10.1016/j.ijforecast.2016.10.002

Ghysels, E., A. Sinko, and R. Valkanov (2007). Midas regressions: Further results and new directions. Econometric Reviews 26 (1), 53–90. https://doi.org/10.1080/07474930600972467.

Giannone, D., L. Reichlin, and D. Small (2008): "Nowcasting: The Real-Time Informational Content Of Macroeconomic Data," Journal of Monetary Economics, 55(4), 665–676.

Higgins, P. (2014). GDPNow: A Model for GDP "Nowcasting", Federal Reserve Bank Of Atlanta Working Paper 2014-7, July 2014 https://www.atlantafed.org/-/media/documents/research/publications/wp/2014/wp1407.pdf

Li, X. (2016). Nowcasting with big data: Is Google useful in the presence of other information. Policy Research Working Paper 7398.

Maehashi, K. and M. Shintani (2020). Macroeconomic forecasting using factor models and machine learning: an application to Japan. Journal of the Japanese and International Economies 58, 101104. doi: 10.1016/j.jjie.2020.101104

Markus Massi, Godfrey Sullivan, Michael Strauß, and Mohammad Khan (2019). *How Cashless Payments Help Economies Grow* https://www.bcg.com/publications/2019/cashless-payments-help-economies-grow

Mitchell, W. C. and Burns, A. F., (1938) Statistical Indicators of Cyclical Revivals. National Bureau of Economic Research (NEBR) https://www.nber.org/system/files/chapters/c4251/c4251.pdf

Payments Canada. (2017a). *High-value payment system—LVTS*. https://www.payments.ca/high-value-payments

Payments Canada. (2017b). *Retail batch payment system*. https://www.payments.ca/retail-payments

Payments Canada. (2017c). *Systems Closure Schedule*. Payments Canada. https://www.payments.ca/about-us/updates/system-closure-schedule

Payments Canada. (2017d). *What We Do*. https://www.payments.ca/about-us/what-we-do

Payments Canada. (2017e). *Who We Are*. https://www.payments.ca/about-us/who-we-are

Payments Canada, D., A. (n.d.). *List of ACSS Stream Identifiers*.

Richardson, A., T. Mulder, and T. Vehbi (2018). Nowcasting New Zealand GDP using machine learning algorithms. Technical report, CAMA Working Paper No. 47/2018. https://ssrn.com/abstract=3256578.

Smirnova, A., & Bewaji, S. (2021). *Payments Economics—Q3 2021* (Payment Economics). Payments Canada.

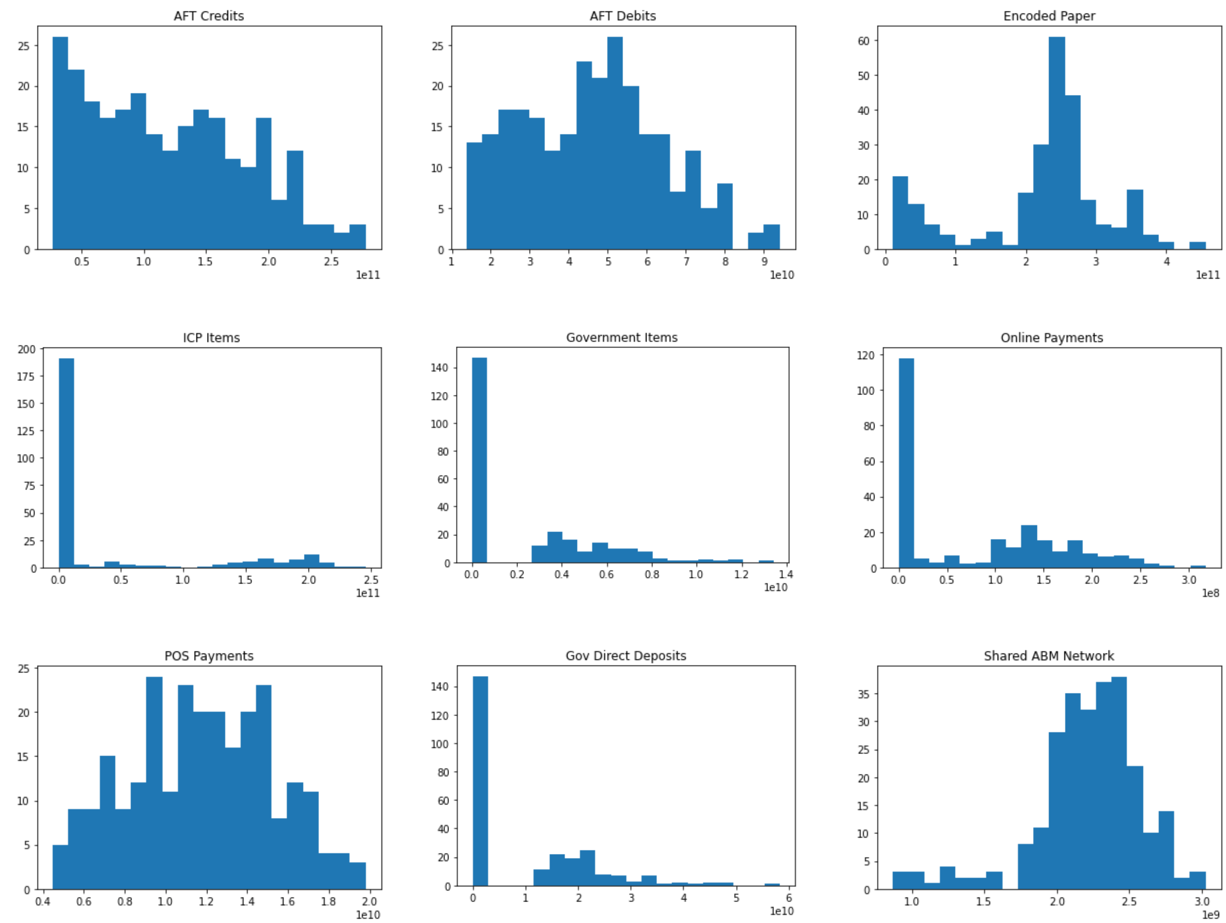
Statistics Canada. (n.d.). *Gross domestic product (GDP) at basic prices, by industry, monthly growth rates (x 1,000,000)*. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610043402

Yahoo! Finance. (n.d.-a). *CBOE Volatility Index (^VIX)*. CBOE Volatility Index (^VIX). https://finance.yahoo.com/quote/%5EVIX/history

Yahoo! Finance. (n.d.-b). *S&P/TXT Composite index (^GSPTSE)*. S&P/TXT Composite Index (^GSPTSE). https://ca.finance.yahoo.com/quote/%5EGSPTSE/history/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuYmluZy5jb20v&guce_referrer_sig=AQAAAJ0-dNplsfnyDQJ9j0Qq85zQ6S7efGcbcsRKYNwR-RhQM0OCBVg9Alhl1b79tP63L42JFtgf0bWuCfogbUEdHZbRoaziU4pVV2p_DJOCINvOivn0k_gcWZ0HDeX

Zamfir Ionela-Cătălina & Iordache Ana-Maria Mihaela (2022). *The influences of covid-19 pandemic on macroeconomic indexes for European countries, Applied Economics*, DOI: 10.1080/00036846.2022.2031858

Appendix



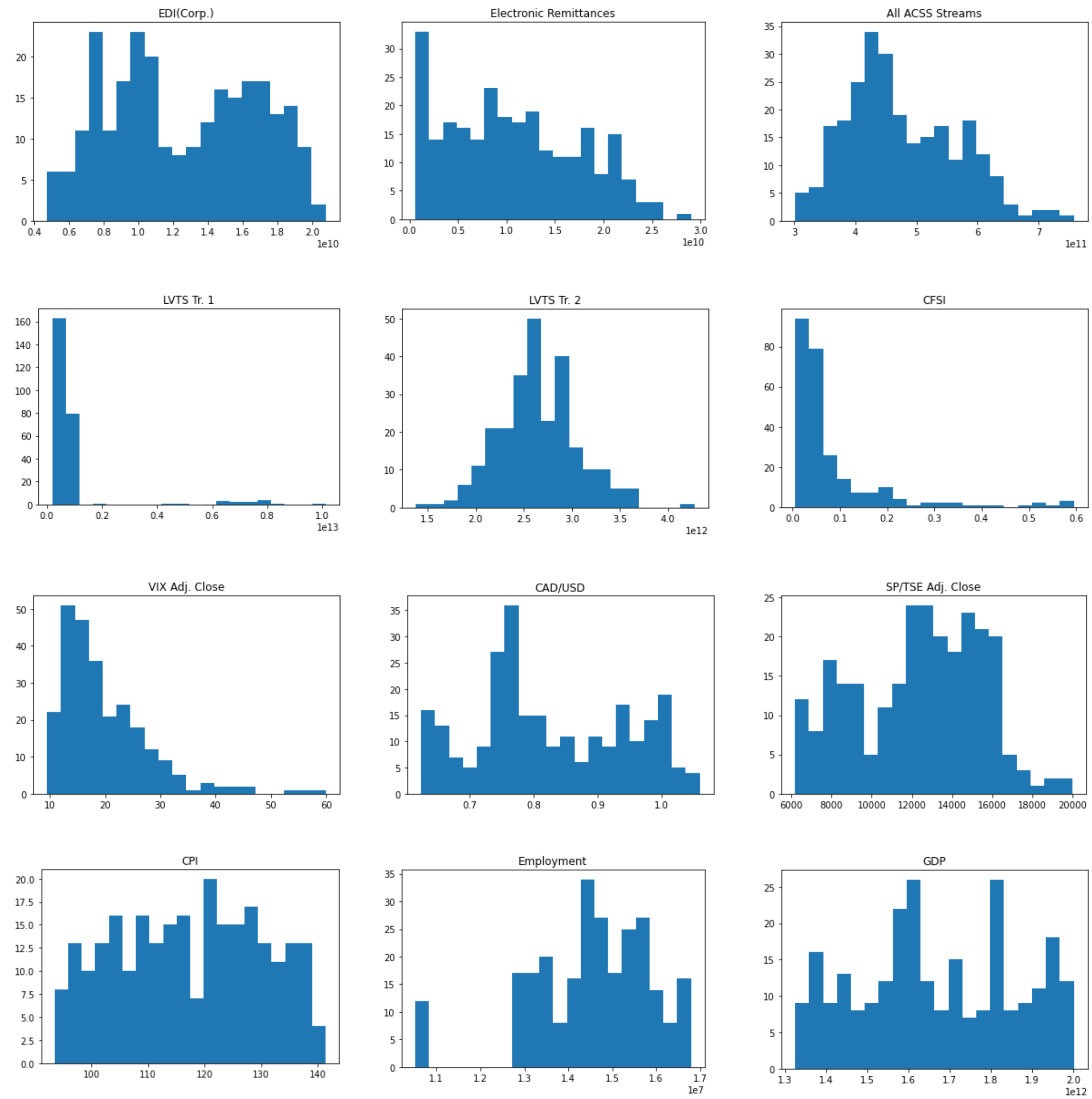
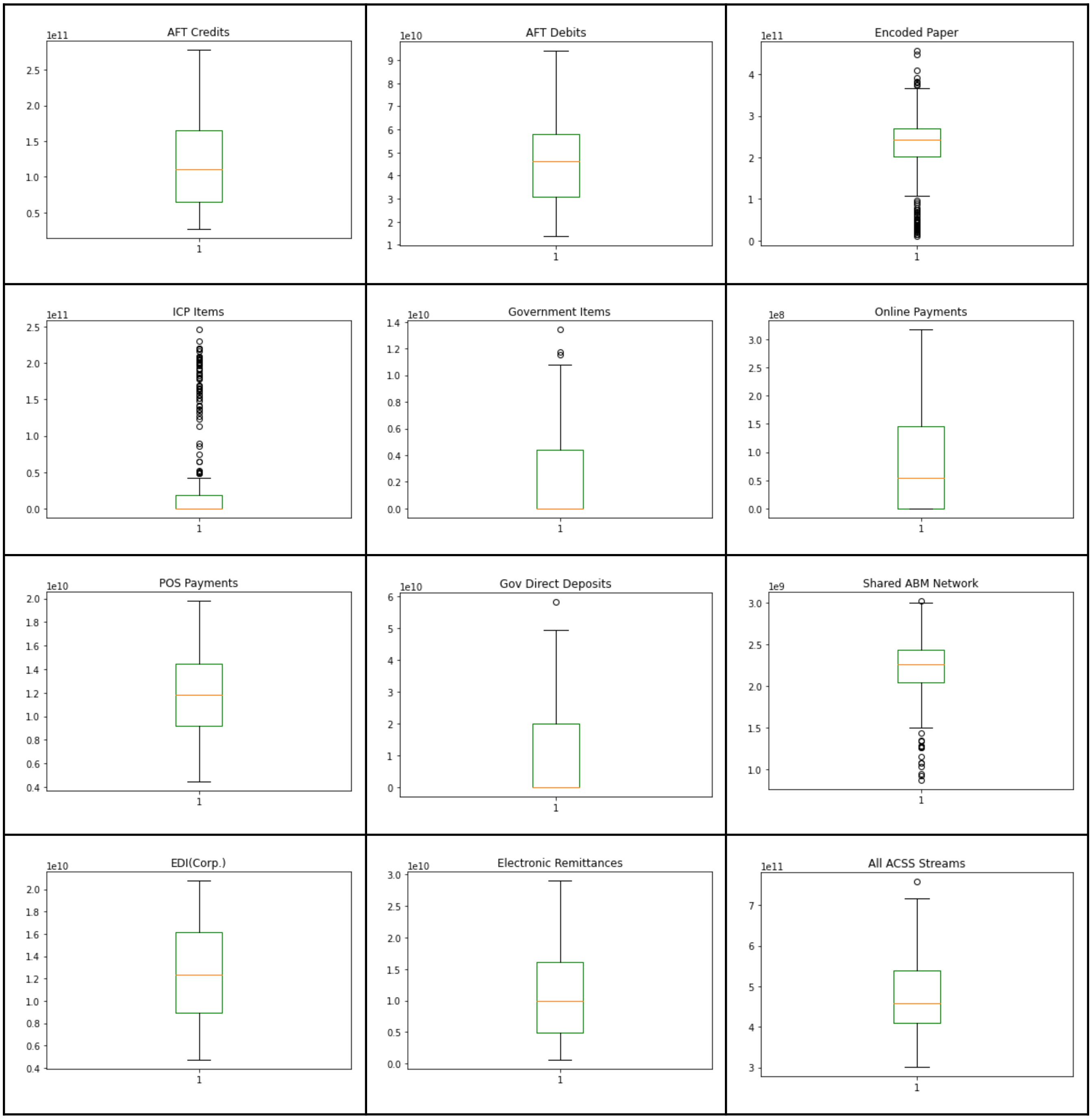


Figure 30 – Histogram plots for values of ACSS, LVTS, GDP and other economic indicators.


AFT Credits
1e11
AFT Debits
1e10
Encoded Paper
1e11
ICP Items
1e11
Government Items
1e10
Online Payments
1e8
POS Payments
1e10
Gov Direct Deposits
1e10
Shared ABM Network
1e9
EDI(Corp.)
1e10
Electronic Remittances
1e10
All ACSS Streams
1e11

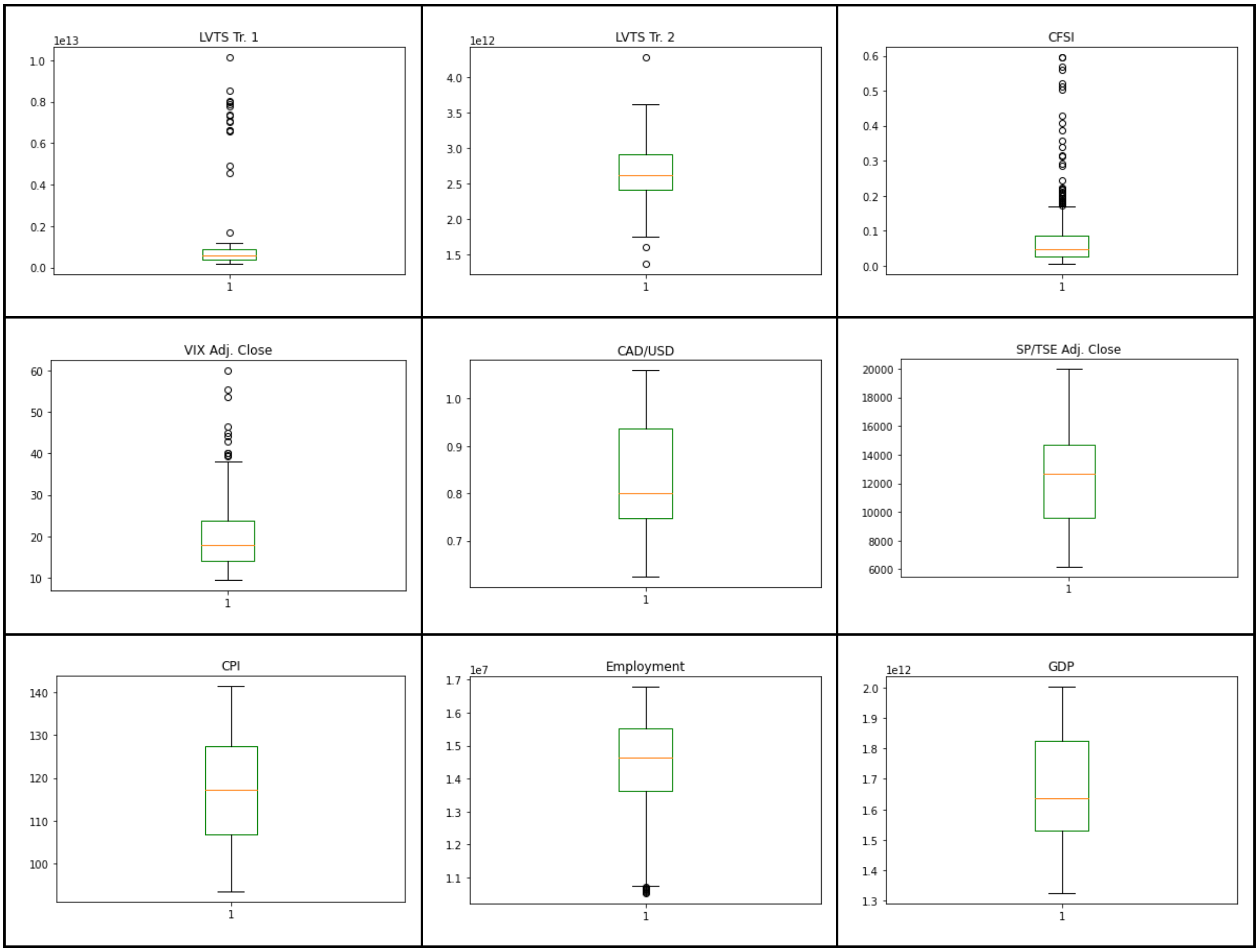


Figure 31– Boxplots for values of ACSS, LVTS, GDP and other economic indicators

S.N.	Regression Model	Baseline GDP Model R^2	Fine-Tuned GDP Model R^2	Baseline CPI Model R^2	Fine-Tuned CPI Model R^2
1	Multi-Linear Regression	86.65%	99.04%	85.92%	99.92%
2	Ridge Regression	86.65%	98.69%	82.66%	99.66%
3	Lasso Regression	99.76%	99.04%	97.65%	99.71%
4	Elastic Net regression	99.56%	99.05%	99.02%	99.77%
5	Decision Trees	94.74%	96.63%	97%	99.69%
6	Random Forests	99.16%	99.21%	99%	99.82%
7	XGBoost	98.43%	97.50%	98.75%	98.07%
8	LightGBM	99.44%	97.87%	99.64%	98.87%

Figure 12-Regression Models for GDP and CPI using randomized timestamp data – Baseline vs. Fine-Tuned Model Performance

S.N.	Regression Model	Baseline GDP Model R^2	Fine-Tuned GDP Model R^2	Baseline CPI Model R^2	Fine-Tuned CPI Model R^2
1	Multi-Linear Regression	<0	<0	<0	<0
2	Ridge Regression	<0	80.20%	<0	70.47%
3	Lasso Regression	77.56%	77.56%	<0	92.08%
4	Elastic Net regression	<0	80.29%	<0	90.35%
5	Decision Trees	<0	<0	<0	<0
6	Random Forests	<0	<0	<0	<0
7	XGBoost	<0	<0	<0	<0
8	LightGBM	<0	<0	<0	<0

Figure 13-Regression Models for GDP and CPI using sequential timestamp data – Baseline vs. Fine-Tuned Model Performance

ACSS payments streams details

• A: ABM Adjustments - POS payment items used to correct errors from shared ABM network transactions (Stream N).

• B: Canada Savings Bond - Part of Government items. It includes bonds (Series 32 and up and Premium Bonds) issued by the Government of Canada.

• C: AFT Credit - Direct deposit such as payroll, account transfers, government social payments, business to consumer non-payroll payments, etc.

• D: AFT Debit - Pre-authorized debit (PAD) payments such as bills, mortgages, utility payments, membership dues, charitable donations, RRSP investments, etc.

• E: Encoded Paper - Paper bills of exchange which includes cheques, inter-member debits, money orders, bank drafts, settlement vouchers, paper PAD, etc.

• F: Paper-Based Remittances - These are used for bill payments and are identical to electronic bill payments (Stream Y).

• G: Receiver General Warrants - Part of Government Items. Paper payment items payable by the Receiver General for Canada.

• H: Treasury Bills and Old-style Bonds - Part of Government paper items. Certain Government of Canada paper payment items such as Treasury bills, old-style Canada Savings Bonds, coupons, etc.

• I: ICP Regional Image Captured Payment - Items entered into the ACSS/USBE on a regional basis.

• J: On-line Payments - Electronic payments initiated using a debit card through an open network, most commonly the internet, to purchase goods and services.

• K: On-line Payment Refunds - Credit payments used to credit a Cardholder's Account in the case of refunds or returns of an Online Payment (Stream J).

• L: Large-value Paper - This is similar to Stream E; starting in Jan 2014, this stream merged into E 34.

• M: Government Direct Deposit - Recurring social payments such as payroll, pension, child tax benefits, social security, and tax refunds.

• N: Shared ABM Network - POS debit payments used to withdraw cash from a card activated device.

• O: ICP National - Image Captured Payments are electronically imaged paper items that can be used to replace the physical paper item: cheques, bank drafts, etc.

• P: POS Payments - Point-of-service payment items resulting from the point-of-sale purchase of goods or services using a debit card.

• Q: POS Return - Credit payments used to credit a cardholder's account in the case of refunds or returns of a POS payment (Stream P).

• R: ICP Returns - Image captured payment returned items entered into the ACSS/USBE on a national basis.

• S: ICP Returns National - National image captured payment returned items entered into the ACSS/USBE on a national basis.

• U: Unqualified Paper Payment - Paper items that are all other bills of exchange which do not meet Canada Payments Association requirements for Encoded Paper classification.

• X: EDI Payment - Electronic data interchanges are an exchange of corporate-to-corporate payments such as purchase orders, invoices, and shipping notices.

• Y: EDI Remittances - Electronic data interchange remittances are used for Electronic Bill Payments such as online bill payments and telephone bill payments.

• Z: Computer Rejects - Encoded paper items whose identification and tracking information could not be verified through automated processes.

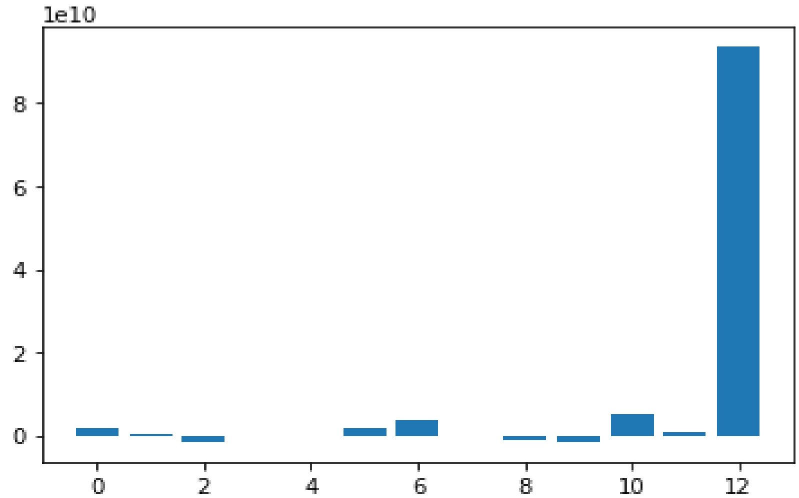


Figure 18 - Feature Importance – Fine-Tuned Ridge Regression for GDP

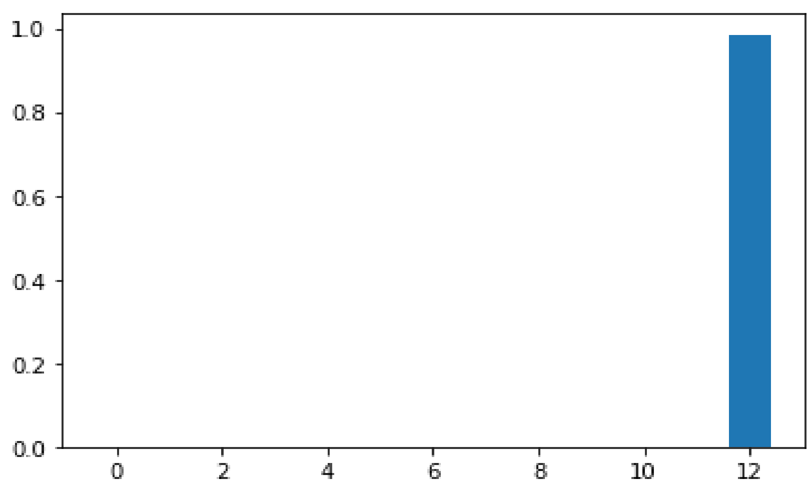


Figure 20 - Feature Importance – Fine-Tuned Ridge Regression for CPI

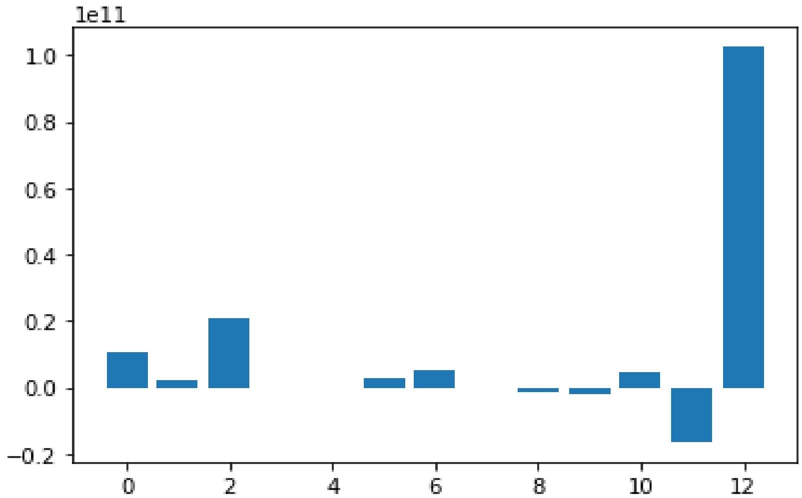


Figure 22 - Feature Importance – Fine-Tuned Lasso Regression for GDP

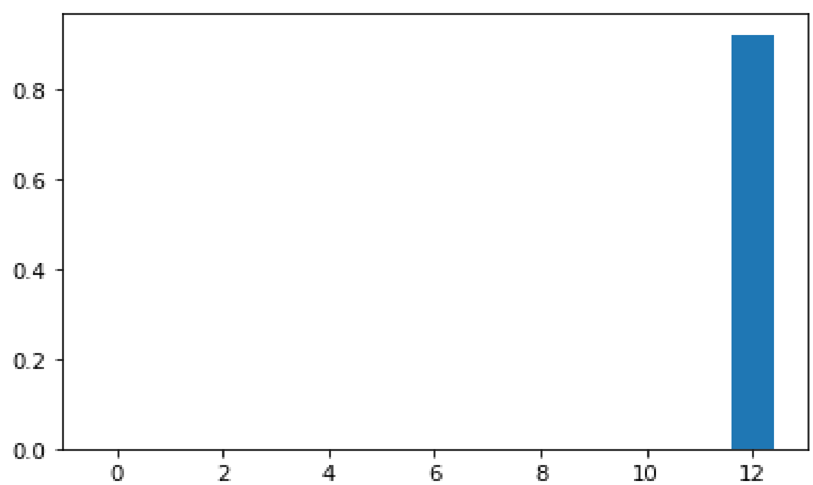


Figure 24 - Feature Importance – Fine-Tuned Lasso Regression for CPI

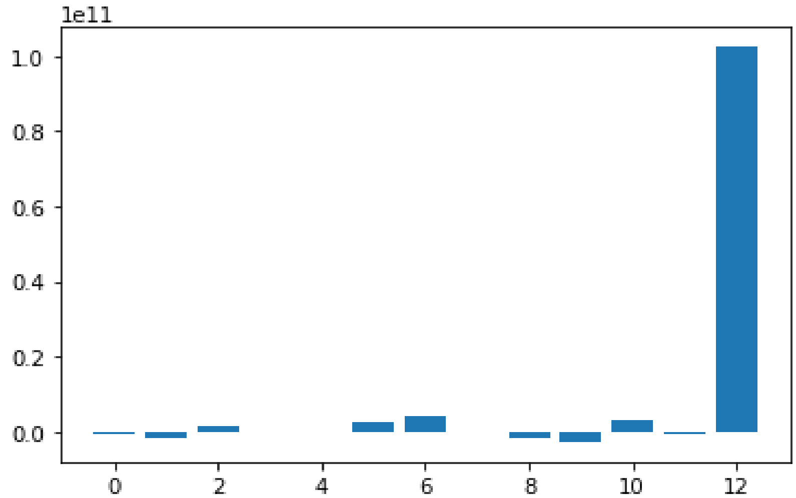


Figure 26 - Feature Importance – Fine-Tuned Ridge Regression for GDP

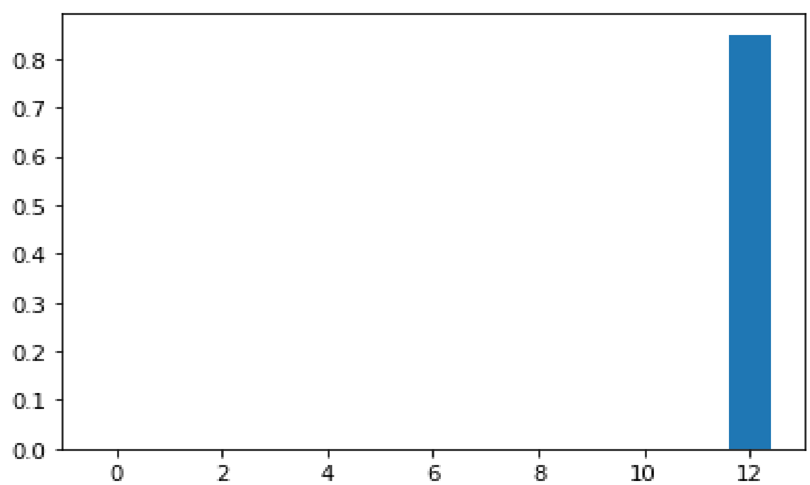


Figure 28 - Feature Importance – Fine-Tuned Ridge Regression for CPI

Independent variables	Dependent variable
0: AFT Credits 1: AFT Debits 2: Encoded paper 3: ICP Items 4: Government items 5: Online Payments 6: POS Payments 7: Government Direct Deposits 8: Shared ABM Network 9: EDI payment 10: Electronic Remittances 11: All ACSS Streams 12: GDP(t-2)	GDP(t)
0: AFT Credits 1: AFT Debits 2: Encoded paper 3: ICP Items 4: Government items 5: Online Payments 6: POS Payments 7: Government Direct Deposits 8: Shared ABM Network 9: EDI payment 10: Electronic Remittances 11: All ACSS Streams 12: CPI(t-1)	CPI(t)