Now-Casting Job Vacancies of Industries Across Canada

Mohammad Alipourlangouri[1] & SeyedehZahra (Zahra) Mousavi[3]

Abstract

This report serves as the culmination of the final project undertaken in the class DATA5001, focusing on the task of nowcasting job vacancies across 21 industries within Canada. Our collaborative effort, as a team, has been dedicated to employing the (S)ARIMA(X)[6] modeling approach to forecast job vacancies with precision and accuracy. In this comprehensive report, we delve into the intricacies of our modeling framework, beginning with an analysis of baseline results and subsequently exploring the impact of external real-time variables on our forecasts. Furthermore, we investigate the reliability of these variables from an economic standpoint and elucidate how governments can harness this information to inform strategic decision-making processes.

1 Introduction

Recognizing the demand for more timely labor market insights, particularly in light of the significant lag in the publication of Labour Market Indicators by Statistics Canada[4]—often spanning 40-50 days post-reference period to accommodate data collection, processing, validation, and analysis—this project addresses the need for expedited information access. With data users expressing a keen interest in more immediate labor market updates, our focus shifts towards forecasting job vacancies across 21 industries within Canada. By delving into the intricacies of data pre-processing, including innovative approaches to impute missing values during the six-month period amidst the COVID-19 pandemic, we aim to lay a robust foundation for subsequent analyses. Our journey progresses to baseline modeling, where we employ predictive methodologies to anticipate job vacancy values across industries. Through a meticulous examination of results and the incorporation of external variables, we aim not only to provide comparable insights but also to illuminate the profound impact of these variables on forecasting outcomes, ultimately contributing to the quest for timelier and more actionable labor market information.

2 Pre processing

In the preprocessing phase, our primary objective was to ensure the dataset's integrity and readiness for analysis. A critical step involved creating the "`Job_vacancies_Value`" column, where we meticulously combined the "`value`" and "`statistics`" columns for each month and industry. This consolidation was essential for structuring the data appropriately, setting a solid foundation for subsequent analyses. However, a notable challenge emerged with the data for the months spanning April to September 2020. During this period, a gap in the dataset was identified, prompting us to address this issue using a thoughtful approach. Rather than omitting the 2020 data entirely, which could lead to information loss, we opted to predict the missing job vacancy values for these months using a baseline ARIMA model.

Our decision to predict the job vacancy values for the April to September 2020 period was motivated by the recognition of bias in the data post-2020. It became apparent that the patterns observed in the data after 2020 deviated significantly from the historical trends. By predicting these values, we effectively filled the gap in our dataset, enabling us to create a more comprehensive and representative dataset for analysis. This approach not only enhanced the robustness of our model but also improved its ability to accurately capture the underlying patterns in the data. Overall, this preprocessing step was crucial in ensuring that our analyses and predictions remained reliable and actionable, setting a strong foundation for the subsequent stages of our research.

3 External Data

In our quest for external data sources, we uncovered the invaluable role of the Toronto stock market[5] in predicting job vacancy values. The stock market serves as a leading economic indicator, mirroring investor sentiment and overall business performance. Fluctuations in stock prices can serve as early signals of changes in economic conditions, influencing businesses' hiring decisions and, consequently, impacting job vacancy rates. A thriving stock market often fosters investor confidence, driving increased business investments and expansions, thereby creating more job vacancies. Conversely, a downturn in the stock market may indicate economic uncertainty, prompting businesses to exercise caution in their hiring practices. Stock market volume, which reflects investor activity and market sentiment, is another crucial factor influencing job vacancy rates. A robust stock market with high trading volume typically signifies economic growth and heightened investor confidence, leading to increased business investments and expansions, and consequently, more job vacancies. Conversely, a decrease in stock market volume may suggest economic uncertainty, prompting businesses to be more prudent in their hiring decisions. The availability of stock market data on a daily basis makes it a real-time and highly valuable resource, enriching our analysis with up-to-date insights. To integrate this impactful external factor into our analysis, we converted the daily volume of the stock market to the mean value of each month, ensuring alignment with our original dataset and enhancing the relevance and reliability of our analysis. Overall, the Canadian stock market data has proven to be a valuable asset, providing crucial insights into the dynamics of job vacancy rates and contributing significantly to the depth and accuracy of our analysis.

In our analysis, we utilized the money market data from the Bank of Canada [2] as another external factor. The impact of the money market, particularly the policies implemented by the Bank of Canada, on job vacancies in Canada is substantial and multifaceted. The Bank of Canada plays a crucial role in shaping the economic environment for businesses, primarily through mechanisms such as influencing interest rates and borrowing costs. Lower interest rates, for example, can stimulate borrowing for expansion and investment, leading to increased job creation. Additionally, the Bank of Canada's strategies, such as purchasing government securities like Treasury bills, inject liquidity into the economy, stimulating economic activity and bolstering business confidence. The stability of the money market, ensured by the Bank of Canada's policies, is essential for businesses planning expansions and hiring, as it fosters confidence and encourages long-term investments that can translate into more job opportunities. Overall, the Bank of Canada's influence extends beyond interest rates and borrowing costs to include economic stimulus, market stability, business confidence, and overall economic growth, all of which play crucial roles in shaping job vacancies in Canada. In this dataset, we included Canadian Treasury bills with maturities ranging from one month to one year. This data is published in real-time on business days, and we converted it to monthly values to align with our original data for job vacancies value.

In Figure1, the heat map illustrates significant correlations between job vacancies value and specific industries, particularly evident in the "`mean_TB.CDN.1Y.MID`" column, representing Canadian Treasury bills over one year[2]. Additionally, a notable correlation is observed with the "`Mean_Volume`" column from the TSX stock market data[5]. These findings led us to select these parameters as crucial external variables for predicting job vacancies value. Due to the high correlation observed in the "utilities" industry between the job vacancy Value and these two external values, we have chosen to showcase our results specifically for this industry.

4 Method and Results

4.1 Baseline Model

The Autoregressive Integrated Moving Average (ARIMA) model is widely recognized as a powerful and adaptable tool for analyzing and forecasting time series data. It is particularly well-suited for our work, offering three main components—autoregression (AR), differencing (I), and moving average

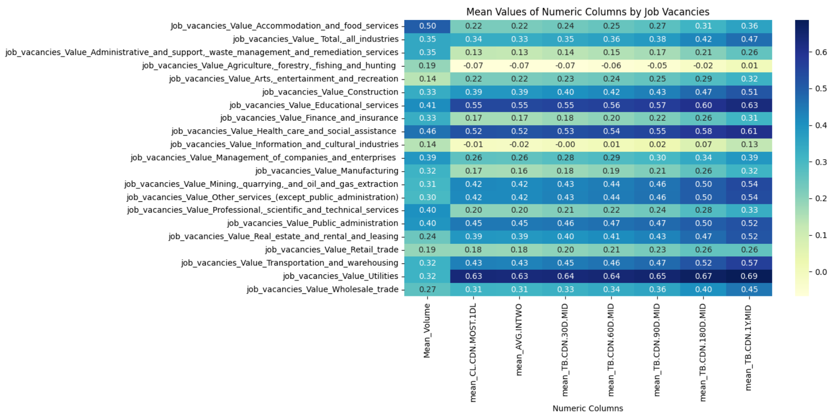

Figure 1: Correlation Matrix between Job vacancy by Industry and External data

(MA)—that can effectively capture the intricate patterns often found in economic and business data. ARIMA's flexibility allows it to handle diverse time series data characteristics, whether they exhibit trends, seasonality, or irregular patterns, making it adaptable to the nuanced trends often observed in job market dynamics. ARIMA's high interpretability is another key advantage, as it can explain the underlying dynamics of job vacancy values clearly, aiding stakeholders and decision-makers in strategic workforce planning and policy formulation. Overall, ARIMA's capacity to model complex patterns, flexibility in handling diverse data characteristics, and interpretability make it a valuable model for our research.

To focus on the utilities industry, we subsetted the data based on the NAICS code. We then created moving averages over 30, 7, and 3 months (for seasonal analysis) to compare the cleaned data with the original values, aiming to uncover underlying trends or patterns(Figure2).

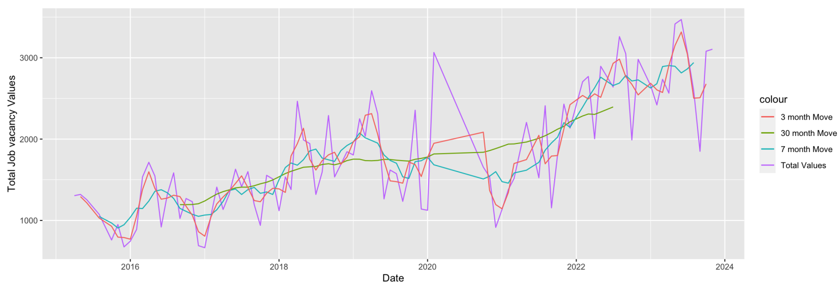

Figure 2: Plot of Moving Averages

Visualizing both the uncleaned and cleaned data, along with the moving averages, provided insights into the data's behavior over time. These visualizations were instrumental in understanding the data's characteristics and identifying any anomalies or trends that could influence our analysis. Subsequently, we decomposed the time series into three distinct components: trend, seasonality, and remainder (or residual) Figure3. The trend component illustrates the overarching long-term behavior of the time series, while the seasonal component identifies the periodic fluctuations. The remainder represents the noise or irregular variations that cannot be attributed to the trend or seasonality. This decomposition clearly reveals the seasonality present within the data. We extracted the deseasonalized component, effectively removing seasonal fluctuations from the time series data. This step offers several advantages: it allows for a clearer focus on the underlying trend or for making forecasts, as seasonal fluctuations are no longer present to obscure the trend component; it renders the data more amenable to modeling techniques that assume stationary data, by eliminating the auto-correlation introduced by seasonality; and it facilitates comparisons between different time periods, especially when the goal is to understand underlying patterns or trends unaffected by seasonal effects. Overall, by removing seasonality, we were able to better discern the underlying

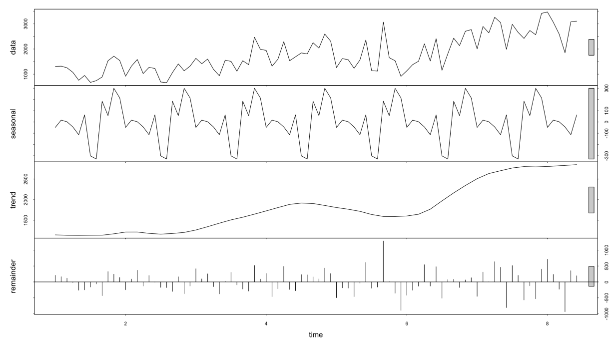

Figure 3: Decomposed Time Series Behaviour

patterns and trends present in the time series data, enhancing our analysis and forecasting capabilities. To ensure the stationarity of the time series, we conducted the Augmented Dickey-Fuller test. If the data was not stationary, we took the first differences to achieve stationarity, aligning with our model assumptions. Using the autocorrelation function (ACF) and partial autocorrelation function (PACF) achieved by tsdisplay() function for auto.arima model, we identified lag points to fit our custom ARIMA model's "q" parameter as the result in Figure4 is shown in Figure4, in the left hand side, auto arima function tried to get the lags but because of the limitations of the function, there are still some lags bouncing out of the line, alerting us that "q" parameter should be set as 24, where the bounce is in the ACF of the auto arima model. We employed an auto ARIMA model on the desea-

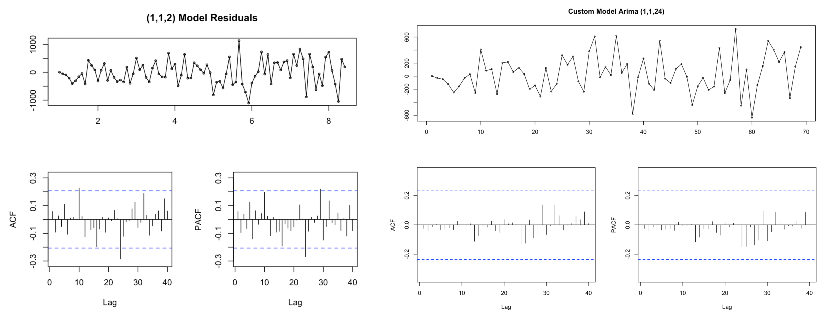

Figure 4: Auto ARIMA Residual for finding tunning parameters

sonalized data and examined its residuals. To validate our models, we tested them on a holdout set without seasonality, assessing their performance and suitability for forecasting job vacancy values, but no important change was observed. Additionally, we explored various ARIMA models with specific lag points to experiment with different modeling approaches which is shown in Figure5. By comparing several models of ARIMA using accuracy measures like mean absolute error, root mean squared error, and mean absolute percentage error(Table 1), we determined that the ARIMA model with a (1,1,24) parameter was the most suitable for predicting future job vacancy values in the utilities industry and we picked this model as the baseline model for the next step which is using exogenous variables.

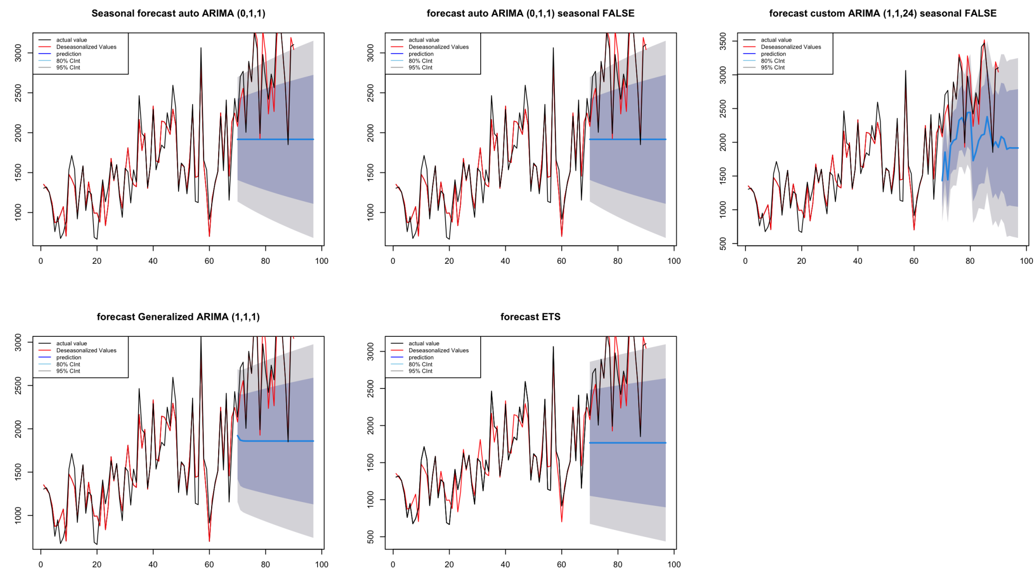

Figure 5: Auto ARIMA Residual for finding tunning parameters

Table 1: Numeric Accuracy Measures

Model	**ME**	**RMSE**	**MAE**	**MPE**	**MAPE**	**MASE**	**ACF1**
With Seasonality (0,1,2)	38.61208	391.3881	302.7623	-3.504134	21.0434	0.8269861	0.07753553
Without Seasonality (0,1,2)	38.61208	391.3881	302.7623	-3.504134	21.0434	0.8269861	0.07753553
Custom ARIMA (1,1,24)	34.83098	284.1032	222.1887	-1.478795	15.62463	0.6069018	-0.02444862
Default ARIMA (1,1,1)	44.15393	388.1766	300.1568	-3.016814	20.65705	0.8198692	-0.01534106

4.2 Final Model

In enhancing the baseline model for improved accuracy, we incorporated external variables. Firstly, we included the monthly mean of the Toronto Stock Market [5] volume. Despite its weak linear correlation with the target variable, the influence of the stock market on sales rates, and consequently, job vacancy rates (our target variable), is undeniable. Additionally, we integrated Money Market data (specifically, Canadian Treasury Bills over five time periods) [2] to observe their impact on the model. Referring to Figure 1, we selected Canadian Treasury Bills over six months and one year, as no significant changes were observed with the inclusion of other Treasury Bills variables. The results presented in Figure 6 demonstrate that the use of exogenous variables as auxiliary predictors has improved the model's ability to capture patterns and fluctuations, as well as to follow trends more accurately.

Table 2: Model Accuracy Comparison

Model	**ME**	**RMSE**	**MAE**	**MPE**	**MAPE**	**MASE**	**ACF1**
Baseline	34.83098	284.1032	222.1887	-1.478795	15.62463	0.6069018	-0.02444862
Stock Market	34.90279	274.3583	209.1265	-1.125757	14.4246	0.5712228	-0.03845909
Stock Market + 1Y TB	39.6225	259.4482	204.3765	-0.9051123	14.08047	0.5582483	-0.03448572
Stock Market + 1Y + 6M TB	28.77308	257.0768	190.8331	-0.5477825	13.1938	0.5212549	-0.01965309

Table 2 showcases the incremental enhancement in model accuracy with the incorporation of external variables. Despite this improvement, Figure 6 highlights some concerns regarding the ARIMA model's reliability when solely considering the Stock Market variable, as it tends to exhibit a horizontal trend. On the other hand, when including Stock Market, one-year Treasury Bills, and six-month Treasury Bills, the model displays a spike in the prediction, which is inconsistent with the data-set's behavior, according to the 2, model with Stock Market + 6 month TB + 1 year TB has

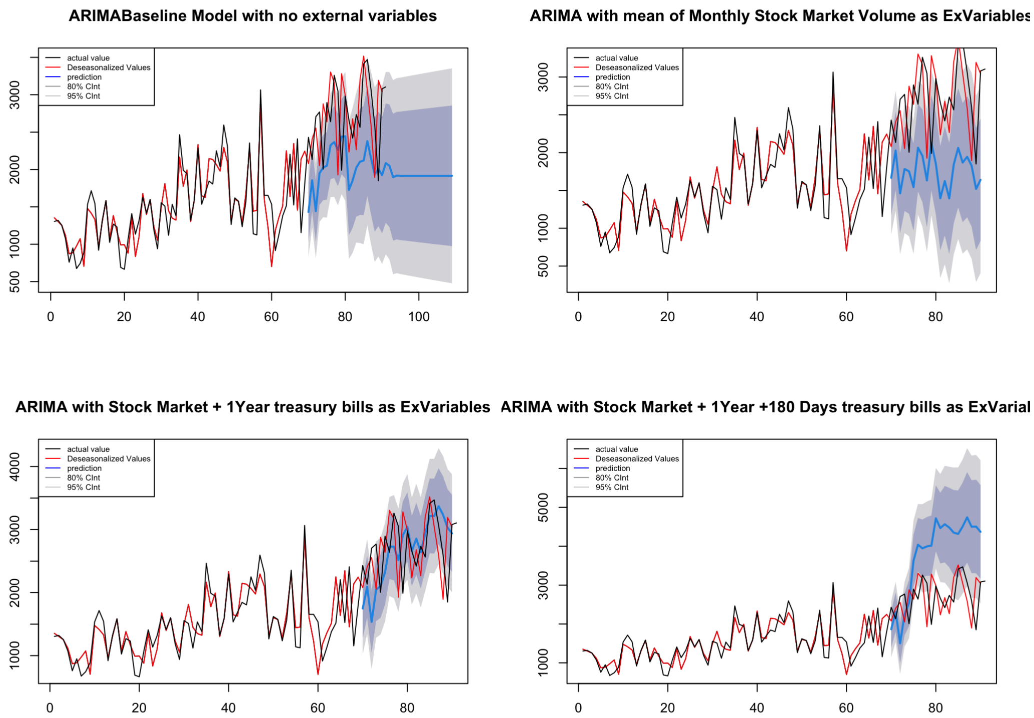

Figure 6: comparison between 4 ARIMA models with and without external variables

the best fit metrics, but in the prediction phase it did not have a good performance due to potential over-fitting. In contrast, the model utilizing Stock Market and one-year Treasury Bills as external variables demonstrates a good fit to the observed pattern and better prediction on unseen data in Figure 6. This model mimics the trend more closely, suggesting it may be a more reliable choice for predicting job vacancy values in the ”Utilities” industry.

5 Limitations and Future Works

In Section 4, we detailed the implementation of our model and demonstrated the significant impact of external variables on its performance. This analysis was conducted across several industries, consistently yielding similar results. Our experiments underscore the potential for improved model accuracy through the integration of reliable external datasets. However, a major challenge lies in accessing such data, as it requires expertise in economic analysis to identify relevant exogenous variables for each specific industry. In our approach, we selected external datasets somewhat blindly, hoping to uncover relationships with our target variable. Nonetheless, there exist high-quality general datasets with a maximum delay of one week that could enhance our model's performance. Unfortunately, accessing these external variables often requires substantial funding and budget allocation. Some potentially beneficial external datasets include Canadian consumer confidence, Canada's unemployment rate, and Canada's business confidence. It is evident that these values can greatly impact the model's accuracy.In our future endeavors, our team is turning its attention to potential additional target variables sourced from the StatsCanada website. Specifically, we aim to now-cast Total Employment, Payrolls, and Hours Worked. We are confident that by using our model ”with more external variables”, we can achieve reliable and impactful outcomes.

6 Additional results

In this section we briefly show how our model works on other industries in Figure 7 for Educational Services, and Figure 8 for Health care and social assistance.

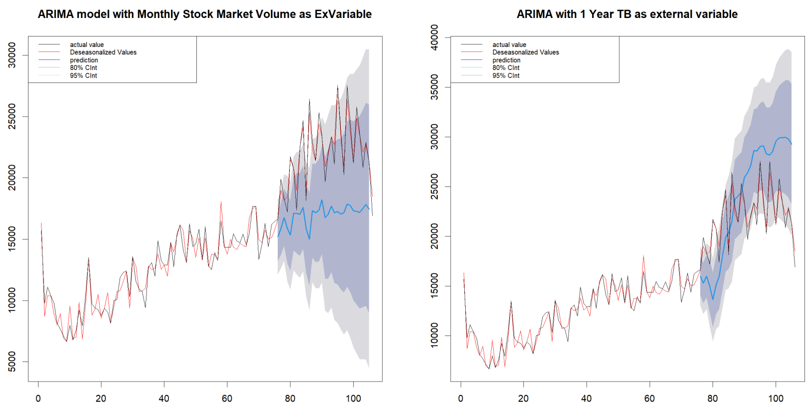

Figure 7: now-casting Job vacancy values for Educational Services

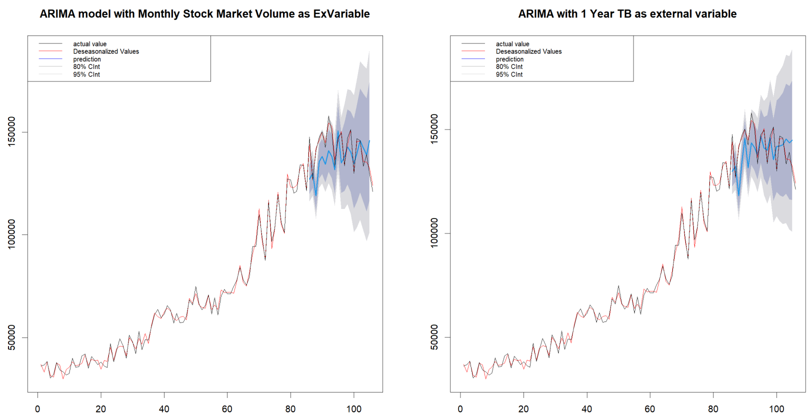

Figure 8: now-casting Job vacancy values for Health care and social assistance

References

[1] M. Alipour. Linkedin profile. www.linkedin.com/in/mohammad-alipourlangouri-320069215. Accessed: April 12, 2024.

[2] Bank of Canada. Canadian treasury bills dataset. Bank of Canada. Accessed: April 12, 2024.

[3] Z. Mousavi. Linkedin profile. https://www.linkedin.com/in/iamzahramousavi/. Accessed: April 12, 2024.

[4] Statistics Canada. Labour market indicator. Statistics Canada. Accessed: April 12, 2024.

[5] Toronto Stock Exchange. Daily closing prices of tsx composite index. Toronto Stock Exchange. Accessed: April 12, 2024.

[6] M. Zekri and L. Souici-Meslati. Immunological approach for intrusion detection. *ARIMA J.*, 17, 2014.