[Page de garde]

HEC MONTRÉAL

Now-Casting Canadian Job Vacancies using Scraped Web Job Postings

Lucas Loboguerrero

Benoit Dostie
HEC Montréal
Directeur de recherche

Anne-Lore Fraikin
Conseil de L'Information sur le Marché du Travail
Codirectrice de recherche

Sciences de la gestion
(Spécialisation Économie Appliquée)

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences en gestion
(M. Sc.)

Summary

This research project examines the relationship between online job postings and official job vacancies in Canada. We use Seasonal and Trend decomposition via Loess (STL) to dissect the time series data, and carry out statistical tests for assessing stationarity and cointegration. The study reveals that job postings tend to remain available for longer than official vacancies, and that there is a lag relationship between the two data sets. The study also highlights a key distinction: job postings are characterized as inflows, while job vacancies are seen as stocks.

However, the representativeness of this data is limited, posing challenges to its direct use in estimating job vacancies. Despite this limitation, the results suggest that job postings can be used to accurately predict official vacancies in the long run. This study offers valuable insights into labor market dynamics and recommends developing an Error Correction Model (ECM) for forecasting purposes.

Résumé

Ce projet de recherche examine la relation entre les offres d'emploi en ligne et les offres d'emploi officielles au Canada. Nous utilisons la décomposition saisonnière et de tendance via Loess (STL) pour disséquer les données de séries chronologiques, et réalisons des tests statistiques pour évaluer la stationnarité et la cointégration. L'étude révèle que les offres d'emploi ont tendance à rester disponibles plus longtemps que les offres officielles, et qu'il existe un décalage entre les deux ensembles de données. L'étude souligne également une distinction clé : les offres d'emploi sont caractérisées comme des flux, tandis que les offres d'emploi officielles sont considérées comme des stocks. Cependant, la représentativité de ces données est limitée, ce qui pose des défis à leur utilisation directe pour estimer les offres d'emploi. Malgré cette limitation, les résultats suggèrent que les offres d'emploi peuvent être utilisées pour prédire avec précision les offres d'emploi officielles à long terme. Cette étude offre des aperçus précieux sur la dynamique du marché du travail et recommande le développement d'un Modèle de Correction d'Erreur (ECM) à des fins de prévision.

Keywords: online job postings, official job vacancies, Cointegration, STL decomposition, NLP, data integration, data visualization.

Remerciements

A mon professeur Benoit Dostie, son travail de relecture et ses idées de modélisation ont été d'une grande valeur pour mon projet.

A ma gestionnaire Anne Lore Fraikin, Je n'aurais pas pu le faire sans sa connaissance du sujet et des bonnes pratiques. Sa disponibilité et bonne humeur ont été d'un soutien sans faille.

A Lassad Damak a qui je suis reconnaissant de m'avoir présenté aux bonnes personnes et de m'avoir sensibilisé aux enjeux des pénuries de main d'œuvre. Il a su aborder avec tact cette thématique humaine.

To the people at LMIC, I would like to take this opportunity to express my heartfelt appreciation to Lorena, Michael, and Sukriti, whose invaluable contributions have played a pivotal role in the success of this project.

Chapter 1: Introduction

1.1 Historical context of labour shortages in Canada

As of September 2022, if you walk through the streets of Montréal, you'll notice many "Nous embauchons" signs. These "We are hiring" notices are displayed on café doors and in front of factories.

"We are hiring" in French.

(source photo) https://lactualite.com/politique/penurie-de-main-doeuvre-attirer-former-retenir/?
Photo : Rodolphe Beaulieu

Although there seems to be newfound panic in political and media circles, the concern about labor shortages is not new in the Canadian economic literature. In 1998, a report by Human Resources Development Canada highlighted the growing labor shortages in skilled professions. However, it didn't suggest that this was a widespread issue. (Gingras and Roy 1998). As early as 2005, the Institut de la statistique du Québec indicated that the growth of the active population would stagnate in the future mostly due to the low fertility rates and an ageing population (Asselin 2005).

On October 9th, 2022, Bank of Canada governor Tiff Macklem alerted to an "exceptionally high number" of job vacancies in the labor market.[1] In the second quarter of 2022 they hit a record number of 1 million (997,000)[2], a trend that has been steadily increasing in the last two decades with the initial months of the pandemic serving as a momentary break from this tendency.

While other recessions were caused by shocks in the economy or the stock market, the covid-19 recession was a public health crisis. Therefore, it affected the labor market in unusual ways. When looking at US data from the Job Openings and Labor Turnover Survey, we notice that the drop in vacancies has been significantly less pronounced compared to the recession in 2008. While they dropped by half in 2008, from their highest to their lowest point, they only dropped by one third during during covid.[3]

Because the Job Vacancy and Wage Survey was interrupted by Statistics Canada for the second and third quarter of 2020, we must turn to job postings data to infer the state of the demand side of the labor market during that period. One study using postings from Burning Glass found that it took until May of 2020 for them to recover to 80% of their pre-pandemic levels which is also an unusually rapid pace of recovery (Jones, Lange et al. 2020).

---

[1] https://www.reuters.com/markets/europe/canadas-economy-has-scope-slow-with-exceptionally-high-vacant-jobs-central-bank-2022-10-09/
[2] https://www150.statcan.gc.ca/n1/daily-quotidien/220920/dq220920b-eng.htm
[3] There is no available official survey data for vacancies in Canada predating march, therefore we would need to rely on Job postings from the Job Bank to make the same comparison with Canada.

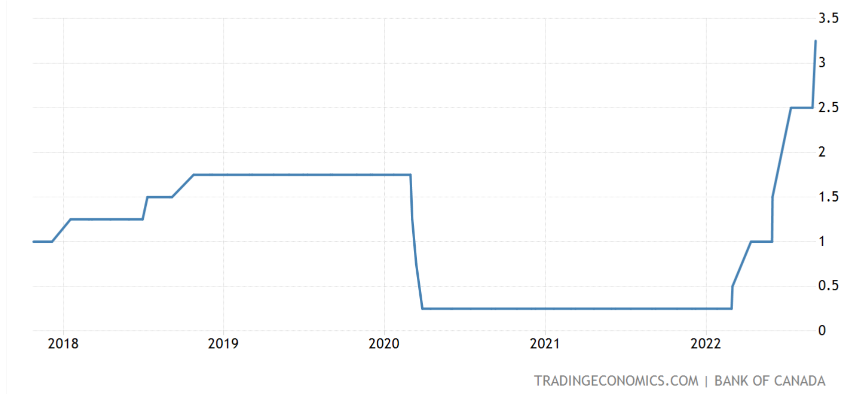

Figure 1- Bank of Canada interest Rates 2018-2022

While inflation currently remains the primary reason for increasing rates, central bankers are also closely monitoring tight labor markets. They put businesses under pressure by forcing them to raise compensation to recruit difficult-to-find candidates. Bank of Canada senior deputy governor Carolyn Rogers pointed out that this phenomenon combined with inflation could lead to a wage-price spiral. This view can be easily understood if we look at the standard Philips curve model, where the ratio of Vacancy/Unemployment seems to outperforms other indicators of slack for measuring both PCE core inflation and wage inflation. (Barnichon and Shapiro 2022).

Estimating vacancy rate is a serious matter. In 2014, the conservative government had to lower their job vacancy rate estimates from 4% to 1.5% due to faulty forecasts. The inaccurate projections were used to justify the creation of the Canada Job Grant to combat job shortages which led to the mismanagement of 3 billion dollars of public funds. Kijiji job postings which contained an extensive number of duplicates were to blame for dire job shortage warnings from Ottawa.[4]

[4] https://www.theglobeandmail.com/news/politics/how-kijijis-wonky-data-threw-off-ottawas-math-on-skills-shortages/article17675622/

1.2 The contemporary challenge: the need for accurate Labor Market Data

As of today, labor shortages are again portrayed in the media as a widespread phenomenon, which raises questions about their true nature and legitimacy. The problem Canada has seems to be the lack of frequent data. Statistics Canada produces the Job Vacancy and Wage Survey (JVWS) which gives us a monthly overview of the amount of job openings per province and a more refined quarterly picture including occupation, industry, and economic region.[5] This is a problem when economic shocks such as a pandemic arise because we do not have the possibility to assess potential short-run labor market squeezes.

However, there are plenty of other sources of data that can be exploited as a potential real time proxies for official labor market statistics. On the supply side of the labor market, job searches have been shown to be an accurate predictor of unemployment rate (Askitas and Zimmermann 2009, Choi and Varian 2009). Conversely, in this thesis we will focus on analyzing the demand side of the labor market by looking at job postings that employers publish on various job boards as a substitute for the official vacancy statistic from the JVWS.

In this study, we will delve into the time series characteristics of vacancies derived from online job postings, specifically those aggregated by Vicinity Jobs, and compare them with official data gathered by Statistics Canada's Job Vacancy and Wage Survey from January 2018 to December 2022. Our motivation stems from a report by the Labour Market Information Council (LMIC), which posits that online job posting data might exhibit biases based on occupation and geography, making such data less reliable for vacancy estimations.

1.3 The design and objective of our study

The LMIC report claims that while online job postings can offer distinct insights, there's a pronounced bias towards certain sectors, occupations, and regions. For

[5] An economic region (ER) is a grouping of complete census divisions (CDs) (with one exception in Ontario) created as a standard geographic unit for analysis of regional economic activity.

instance, postings often disproportionately represent professions in business, finance, administration, and management, while there is an underrepresentation in trades and sales fields. Given this backdrop, the core objective of our paper is to critically assess this claim. We aim to determine the extent to which job vacancy time series are indeed representative at both the 10 NOC level and at the provincial level. In the ensuing chapters, we will present the contents of the LMIC report and interrogate its primary assertion.

Our methodology aims to determine whether the Official and Job postings time series are influenced by the same underlying factor. The first step is to decompose the time series into three main components trend, seasonality, and remainder.

Our goal with trend decomposition is to analyze the general movement of the time series pairs, including whether they increase, decrease, remain stable, or move in different directions over time. By looking at the seasonality we hope to detect similar pattern variation between the pairs over time. This implies that job postings of capturing fluctuations in labor demand that are typical in industries or regions where seasonality is a factor. When removing trend and seasonality, we expect the pairs of remainders with variations of similar magnitude, no discernible pattern and little persistence.

As a second step, we apply cointegration to the corresponding time series to ascertain whether they are influenced by the same underlying factor. Unlike the Seasonal and Trend decomposition using Loess, which relies heavily on visual analysis of time series characteristics, cointegration is a more rigorous and systematic technique for determining whether the series share a common long-term trend and co-movement.

We follow the methodological basis that Lovaglio et al. (2019) used to assess the Italian labor market by comparing quarterly time series of job vacancies published by Eurostat and the real-time obtained from a private aggregator - a real-time monitor that collects Italian online job postings equivalent to Vicinity Jobs.[6] We aim to evaluate if job postings can provide a reliable measure of official vacancies with the benefit of being available in real time with granularity (province and type of occupation).

Our analysis finds a strong and significant long-run relationship between official job vacancies and online job postings at the national and provincial levels. However, the results vary across different aggregation levels and occupations, with the 60-day aggregation level demonstrating the strongest cointegration across the greatest number of occupations. Despite some limitations, our study provides valuable insights into the applicability of using online job posting data to understand labor market dynamics at a granular level. As a next step, we recommend developing an Error Correction Model (ECM) for forecasting purposes. The ECM would combine the short-run dynamics and the long-run equilibrium relationship of the variables, allowing for more accurate and reliable forecasts.

[6] https://www.theglobeandmail.com/report-on-business/economy/jobs/ottawa-adjust-labor-data-raising-questions-about-national-skills-shortage/article18457198/

In the next chapter we explain the concept of nowcasting and do a general literature review of the studies using internet data for economic forecasting. We put an emphasis on the three relevant studies, (including the one on which we base our methodology on) that compare the characteristics time series of online job posting with the official data produced by their respective countries' statistical offices. The third chapter elaborates on the data source and sample construction used in the thesis. We also summarize the LMIC report 36 which details the biases of online job postings over JVWS data.

The fourth chapter delves deeper into the methodology used in our analysis. The fifth chapter is devoted to presenting the results obtained in our analysis including various data visualizations we produced. Finally, the last chapter presents the conclusions we draw from this thesis.

Chapter 2: Literature Review

2.1 Real time internet data in economic literature

There is often a long delay in the publication of key economic measures, making it more difficult for policymakers to get a correct picture of present conditions. Therefore, a new strand of economic literature seeks to leverage the internet as a new source of data to produce real-time indicators. It is often referred to as "now-casting" [7] and is a contraction of forecasting and now. Researchers have frequently used Google search keywords as a data source.

They have been used to construct consumer confidence indexes (Della Penna and Huang 2009), housing distress indexes (Chauvet et al. 2013) or predict "early warning signs" of stock market moves (Preis et al. 2013). However, The Economist (2020) underscores the importance of caution when using real-time data sources, as they may not fully represent the broader economy due to potential biases, data quality issues, and their limited scope. The article notes that while some indicators, such as mobility data or restaurant visits, can provide insights into specific sectors, they may not offer a comprehensive understanding of the overall economic activity. Seasonal variations, substitution effects, and the narrow focus of certain indicators can introduce biases and distortions in the data. The article advises that real-time data should be used carefully and in conjunction with established official measures to obtain a more accurate representation of economic activity.

In a similar vein, Koenig, Dolmas, and Piger from MIT present their paper, "The Use and Abuse of Real-time Data in Economic Forecasting," which argues that relying solely on the most recent data for forecasting can lead to inaccurate predictions. The authors recommend using real-time vintage data, incorporating

[7] The term was initially used in meteorology to refer "to the description of the current state of the weather in detail and the prediction of changes that can be expected on a timescale of a few hours". It was coined in 1981 by meteorologist Keith Browning https://public.wmo.int/en/resources/bulletinnow-casting-guidelines-–-summary

only the information available at a given point in time, rather than updated or revised data. Furthermore, they advocate for the use of first release left-side data, as it often serves as an efficient estimate of subsequent releases and provides a more accurate representation of early relationships between variables.

2.2 Real time internet data in labor economics

A 2002 survey revealed that 50% of Canadian adults online used the internet to view job postings (Ipsos-Reid 2002). The latest Canadian Internet Use Survey, conducted in 2020 by Statistics Canada, shows that 93% of Canadians use the internet, marking a significant 44% increase since 2002 (Statistics Canada 2001, Statistics Canada 2020) [9]. A similar trend is to be expected in most developed nations.

One of the first order effects of the internet's expansion is its increasing use as a job matchmaker. Computerization provides the opportunity for the seeker and employer to lower the cost of search and increase the efficiency of job searching (Kuhn 2014) and generates a substantial amount of data on the recruitment process.
The earliest example available in the literature of the use of internet data to predict a labor statistic comes from Ettredge (2005) which attempted to regress the US unemployment rate by using the number of web searches with job related keywords and found that job searches are highly associated with the unemployment rate of males over 20 but not associated with their female counterpart's unemployment rate. The article concluded that males were more likely than females to conduct internet job searches.

A similar study was conducted in Canada finding a long-term relationship between Job listing keywords from Google Search and the unemployment levels of young people aged from 15-24 years old (Mitchell 2015).

2.3 Comparing the trends of Job posting and official Vacancies

The approach of comparing job posting aggregates with official vacancy numbers has been explored in studies from three countries: the Netherlands, Italy, and Slovakia (de Pedraza, Visintin et al. 2019, Lovaglio, Mezzanzanica et al. 2020, Štefánik, Lyócsa et al. 2022). The Italian and Dutch study rely on a private service that scrapes vacancy postings from the internet and classifies and removes duplicate postings using natural language processing data in a very similar way to what Vicinity Jobs does for us. The Slovakian study on the other hand, uses data from a private job board that holds 78% of the market share for online job searches. While the sources are slightly different, the initial concern remains the same. Are online job postings representative of vacancies?

To answer that question, the three studies used STL trend cycle decomposition. However, they differ in subsequent statistical analysis. Both the Italian and Dutch study use the Augmented Dickey Fuller to test the null hypothesis of the presence of a unit root, while the Slovakian study uses Kwiatkowski–Phillips–Schmidt–Shin (KPSS) to test for the null hypothesis of stationarity around a trend against the presence of a unit root.

To assess whether Job posting, and Official vacancy time series are influenced by a common underlying phenomenon, the Dutch study uses a cross-spectral analysis by examining patterns of autocorrelation, cross-correlation, cross-spectral density, and squared coherency. The Slovakian study examines Autocorrelation and partial autocorrelation and use autoregressive models, they subsequently create a year of job posting forecasts via autoregressive models and compare its predictive ability of vacancies, employment, and unemployment rate. The Italian study that we will attempt to replicate detects the presence of cointegration to determine whether job postings represent real vacancies in a reliable manner. While the Dutch and Slovakian papers rely on a single country wide pair of time series, the Italian study also analyses time series pairs by industry sectors.

To our knowledge, we are the first study to examine representativity of job postings at occupation and regional levels.

The Dutch study highlights a pronounced similarity between official data and job vacancy data. Although the official data is consistently higher, it follows a parallel trend. The difference in seasonal patterns suggests a lag-lead in the relationship between official and job posting data, while a decrease in volatility over time for both series is observed. Furthermore, the studies cross spectral analysis shows a positive, significant, and nearly simultaneous co-movement between the series at time lags close to 0.

The Slovakian study identifies similarities in the trends and seasonality between official vacancies and job postings. They find that both series seem to present strong autocorrelation which is indicative of a unit root. When detrended the series show a decline in autocorrelation which would lead us to believe that the unit root is caused by the linear trend. They argue that any model that attempts to predict the official number of vacancies will not be able to do so because the strong persistence means they would need the latest value and that the data is released long after the reporting period. Using an autoregressive model, they discovered that online job vacancies effectively predict future values of official vacancies, and their predictive accuracy increases as the forecasting horizon expands.

The extensive Italian research into labor sectors uncovers notable disparities across different economic sectors. Industry and professional, scientific, and technical activities are overrepresented, while the construction sector is underrepresented. Interestingly, there are noticeable differences between official and web data counts across all sectors.

Despite these discrepancies, both official and web data exhibit similar seasonality and trends at the country level. Linear regression analysis confirms these trends, as both data sets display similar coefficients. However, the official data shows a greater magnitude in its seasonality. When examining the components of the STL, both series appear to be non-stationary, further highlighting their similarities and differences.

All series cointegrate, with R-squares higher than 0.9 for all sectors, showcasing a strong relationship between the data sets. However, exceptions can be observed in both the Accommodation and Food Service Activities sector and the Financial and Insurance Activities sector, where the official vacancies series are I(0) integrated of order zero, indicating that these series are not integrated. Consequently, the pairs cannot be cointegrated in these economic sectors. In contrast, at the country level, the series do cointegrate, and a long-run relationship coefficient of 0.616 signifies that official data accounts for 62% of the web vacancies' movements during a quarter.

Chapter 3 : The Data Sources and Sample Construction

In Canada, two primary sources of data are used to analyze the job market: the official vacancy data from the Vacancy and Wage Survey and the online aggregator, Vicinity Jobs. In this chapter, we aim to provide a comprehensive analysis of the Canadian job market by examining these two data sources and linking them together using the National Occupation Classification methodology. This standardized framework provides a consistent way to categorize occupations across the country.

3.1 The Official Statistics

In the eyes of Statistics Canada[10], a job is considered vacant if it meets with the 3 following criteria:

- a specific position exists
- work could start within 30 days
- the employer is actively seeking workers from outside the organization to fill the position.

3.2 **About NOCs**

Occupations are defined using the National Occupation Classification system, which structures them in a 5-digit hierarchical format. The first digit represents the 10 broad occupational categories going from 0 through 9. The second digit represents the 6 levels of training, education, experience, and responsibilities required for the role. A digit of 5 means that the role requires no formal training while a 0 represents a management role. This second digit is an innovation added in the latest version of the NOC in September 2021. The last three digits are used to categorize each occupation into more refined subgroups and unit groups.

[10] https://www150.statcan.gc.ca/n1/pub/72-210-g/2013001/part-partie3-eng.htm

For instance, the code for the title "Economists and economic policy researchers and analysts" is 41401. The first digit is 4 for "Occupations in education, law and social, community and government services", the second digit is 1 because the degree necessitates the completion of a university degree.

Major group

41 – Professional occupations in law, education, social, community and government services

Sub-major group

414 – Professional occupations in government services

Minor group

4140 – Policy and program researchers, consultants and officers

In this paper we will only focus on the 10 major NOCs at the 1-digit level of aggregation for at the country level contained in the table below:

Skill Level (DIGIT)	Occupation
0	Management occupations
1	Business, finance and administration occupations
2	Natural and applied sciences and related occupations
3	Health occupations
4	Occupations in education, law and social, community and government services
5	Occupations in art, culture, recreation and sport
6	Sales and service occupations
7	Trades, transport and equipment operators and related occupations
8	Natural resources, agriculture and related production occupations
9	Occupations in manufacturing and utilities

Table 1-List of National Occupation Classification- 1 digit - Broad Occupational Categories

3.3 Vicinity Jobs Postings

In this paper, we utilize data from a Canadian company, Vicinity Jobs, which has been aggregating job postings in both French and English from job boards across Canada since January 2018. They process the data by linking roughly 200 000 individual posts to their corresponding National Occupational Number and Economic Region and removing duplicate posts with their own proprietary Natural Language Processing (NLP) technology.

A report detailing the sources of job postings collected by Vicinity Jobs in Nova Scotia for October 2022 indicates that 45.8% (3,769 postings) originated from employer corporate websites. [12] Other significant sources of online job postings included Indeed, with 1,779 postings accounting for 21.6%, the Service Canada Job Bank with 970 postings making up 11.8%, and Monster with 563 postings contributing to 6.8% of the total postings. We show in the following chapter that job postings produced by Vicinity Jobs seem to present systematic biases. The database only considers online job postings, which means that smaller companies that do not advertise their vacancies online may not be adequately represented.

In addition to Vicinity Jobs, there are other providers of online job posting data available in the Canadian market, such as LightCast and Burning Glass Technologies. One such provider is BioTalent Canada, which specializes in providing labor market information (LMI) data for the biotechnology sector.

Indeed provides publicly available data for Canada for sectors across industry sectors and provinces monthly since February of 2020.[13] In a subsequent investigation, we could use this data to perform cointegration on job postings across various industry sectors throughout Canada. This is because Vicinity Jobs ability to

---

[12] 20221115 Vicinity Online Job Postings Monthly Report -- October 2022.pdf (novascotia.ca)

[13] https://github.com/hiring-lab/job_postings_tracker/tree/master/CA

match jobs with the correct industry is low, despite providing data on job postings by industry.

3.4 PostingsLack of Representativeness in Job Postings

LMI's Outlook Report No. 36 evaluates the representativeness of online job postings from Vicinity Jobs compared to JVWS job vacancies. The report identifies a significant bias favoring management and service sector occupations that require a university education, while underrepresenting trades and other roles that demand secondary education or on-the-job training. It points out one of the major drawbacks of using online job posting data is the lack of representativity in geographic and sectorial categories.

Regarding geography, job postings seem to be less prevalent in the main metropolitan areas of Toronto, Vancouver, and Montreal than in other areas. According to the JVWS they represent 45% of the vacancies while only accounting for 38% of job postings.

Across occupations at the NOC level 1, we find that job postings tend to favor professional and service sector professions while underrepresenting trades and manual labor professions. Figure 2

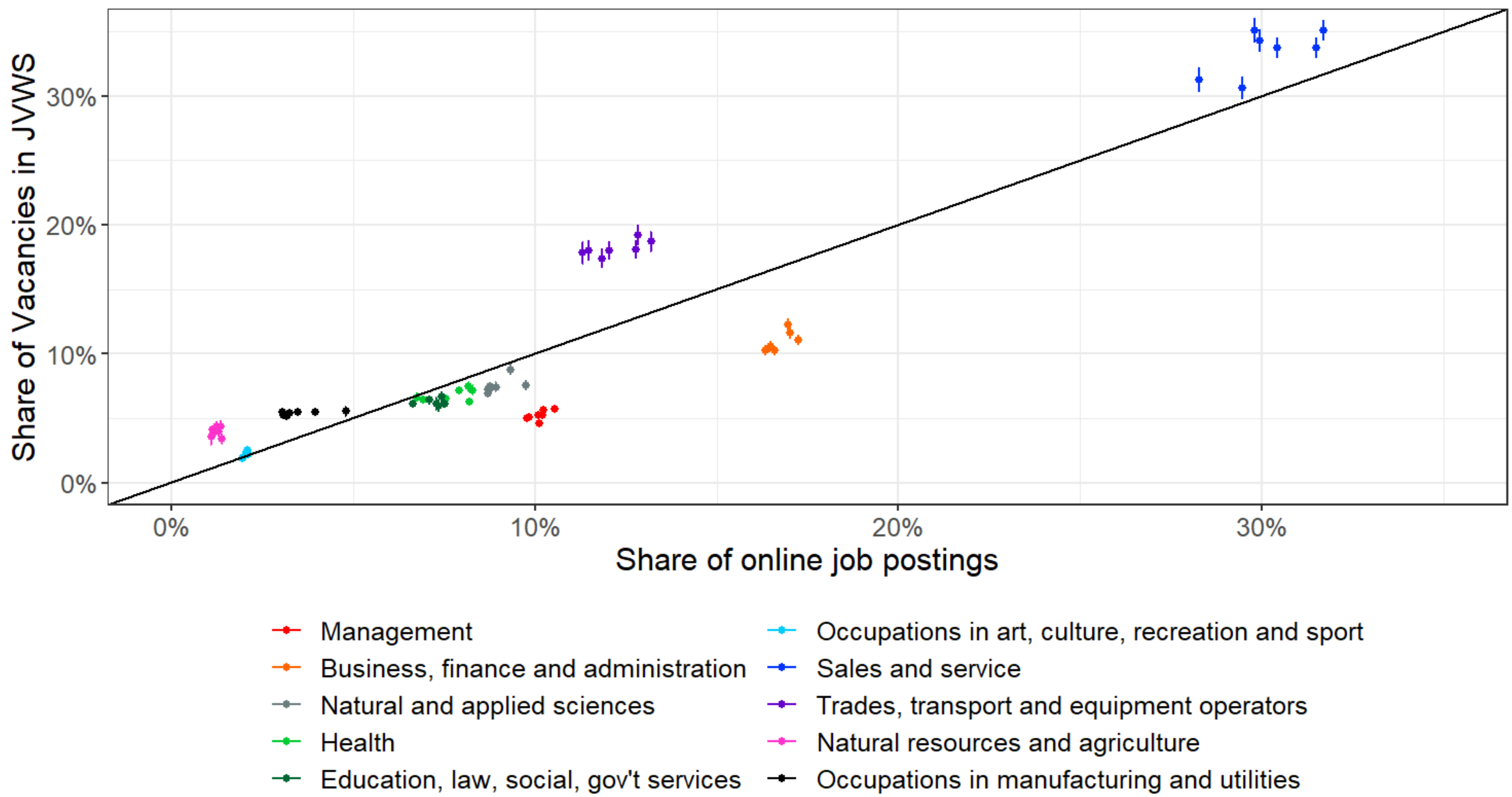

Figure 2 -Distribution of online job postings and JVWS vacancies, Canada-wide, quarterly (source LMIC)

This report presents claims that job postings are consistently higher than estimated job postings across the 7 quarters studies. Two reasons are provided to explain this inconsistency: the omission of public administration positions that are not included in the JVWS, accounting for 2% of the workforce, and the possibility of duplicate job postings from various job boards that have been counted several times.

At the provincial level, this claim appears valid only for Nova Scotia, Saskatchewan, and Prince Edward Island. Figure 4 Furthermore, when visualising Vacancies vs Job postings monthly, we can see that the relationship changes entirely, where job postings are now consistently lower than job vacancies. Figure 5

This discrepancy arises because job postings represent inflows, while job vacancies are stocks. Thus, comparing their levels is akin to comparing apples and oranges. We will go more into details in the methodology where we will explain how we will convert a flow of job postings into a stock. The report states that the lack of representativeness prevents the use of this data to estimate vacancies. We will attempt to question this assertion at the provincial and NOC-1 level for Canada

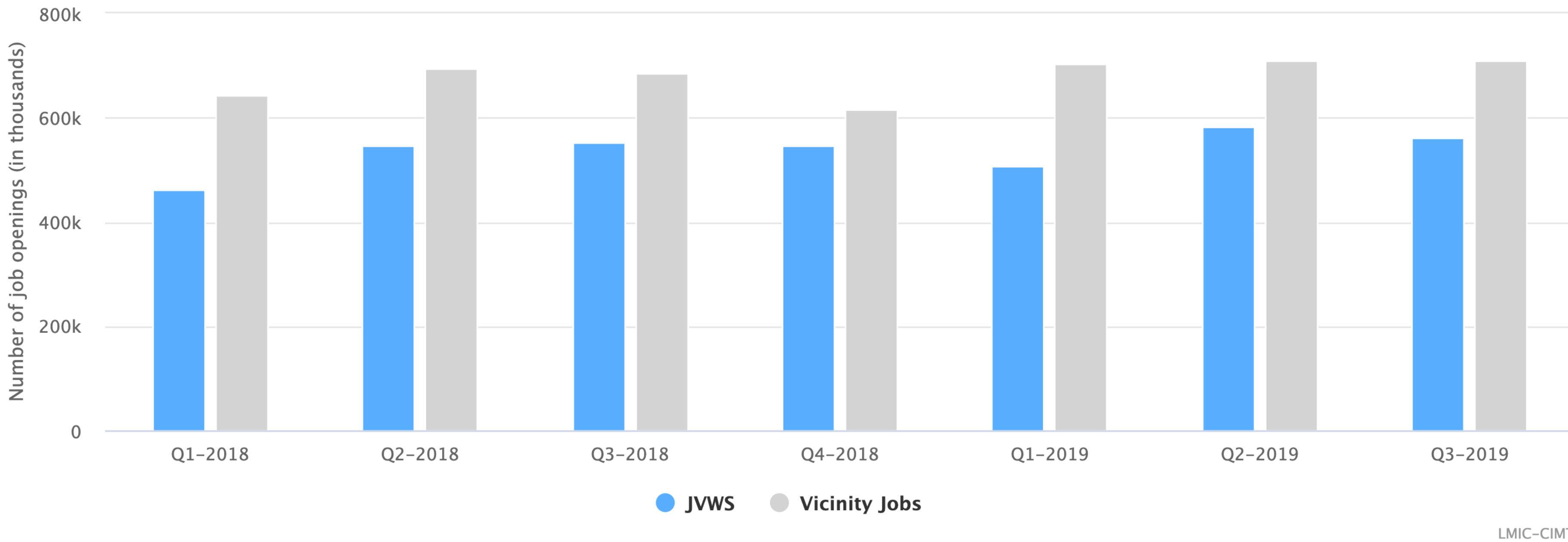

Figure 3-Total online job postings are consistently higher than estimated numbers of vacancies 2018 Q1 to 2019 Q3 (source LMIC)

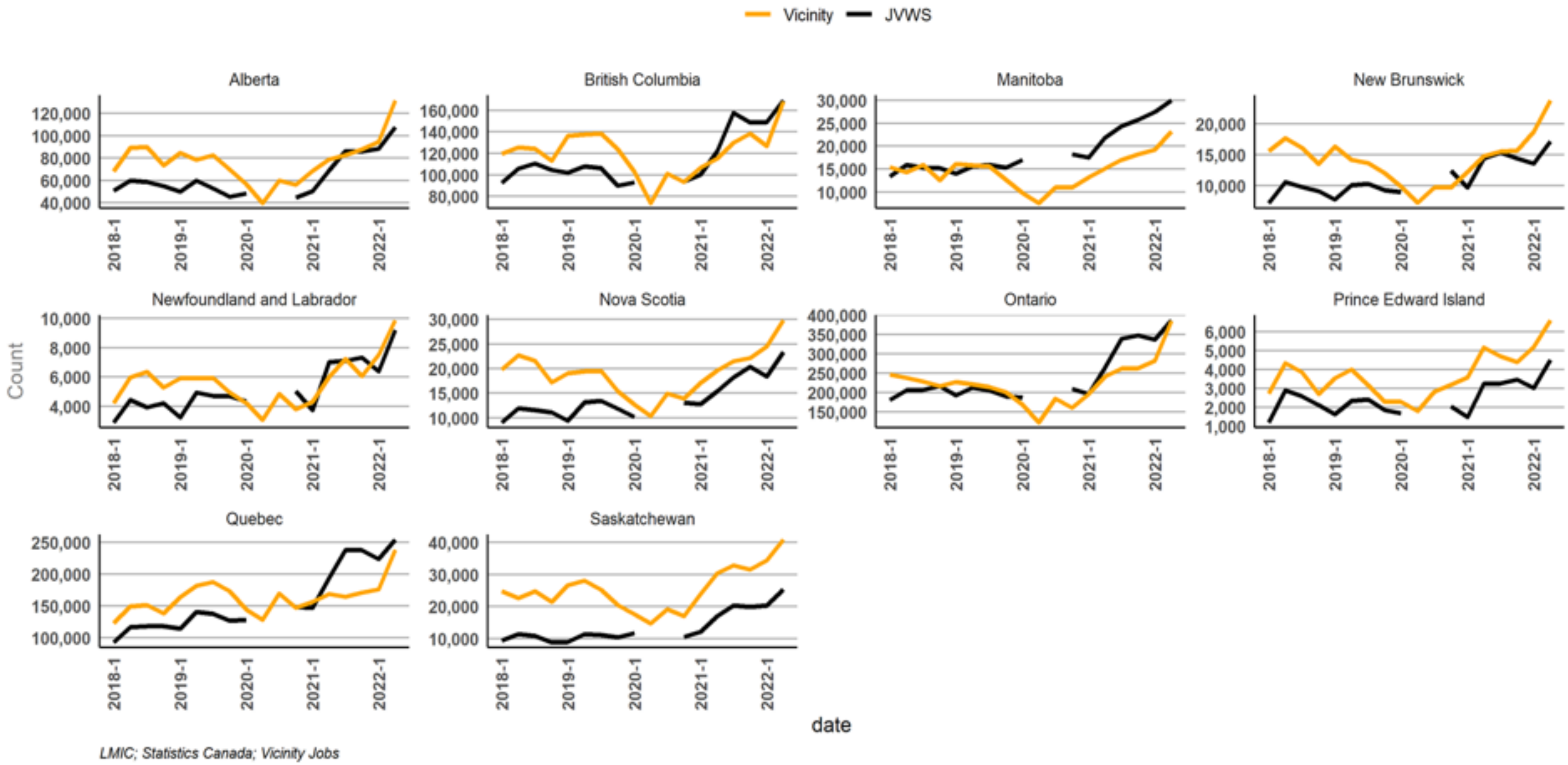

Figure 4-Total online job postings vs numbers of vacancies 2018 Q1 to 2022 Q1 (source LMIC)

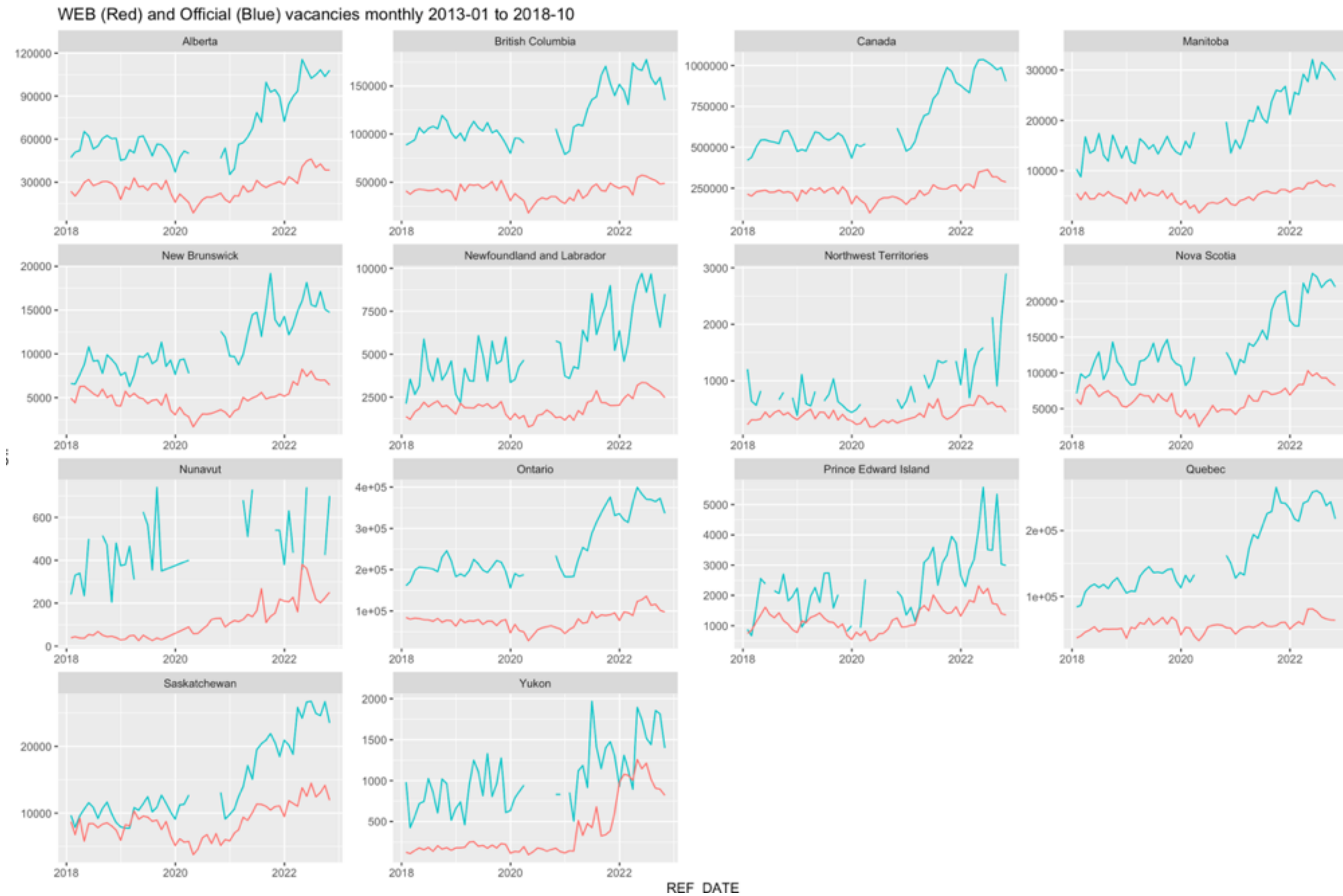

Figure 5-Total online job postings vs numbers of vacancies at province and country level January 2018 to December 2022

2

Chapter 4: Methodology

The primary goal of this study is to examine the extent to which job vacancy and job posting time series in Canada are influenced by the same fundamental process, as well as to uncover potential cointegration between the two series. We employ the "Seasonal and Trend decomposition using Loess" decomposition technique for graphical analysis of the data, while utilizing statistical tests to assess stationarity and cointegration, ultimately shedding light on the relationship between job vacancies and job postings in the Canadian context.

Our methodology will involve a three-step process: initially, we conduct an STL decomposition to analyze the time series data; subsequently, administer integration tests to evaluate stationarity; and finally, we perform cointegration tests to determine the presence of any long-term relationships between the variables under consideration.

4.1 STL

First, we employ the "Seasonal and Trend decomposition using Loess" method (Cleveland, Cleveland et al. 1990). When applying STL decomposition to our time series data for job postings WEB_t and official vacancies $JVWS_t$, we decomposed it in three separate components: seasonality, trend, and remainder.

Loess smoothing is a nonparametric method of smoothing a time series. We do not assume that our data follows any specific distribution, such as a normal distribution. By using this technique, we enhance the robustness of our model, minimizing the influence of outliers when analyzing trends. Figure 6

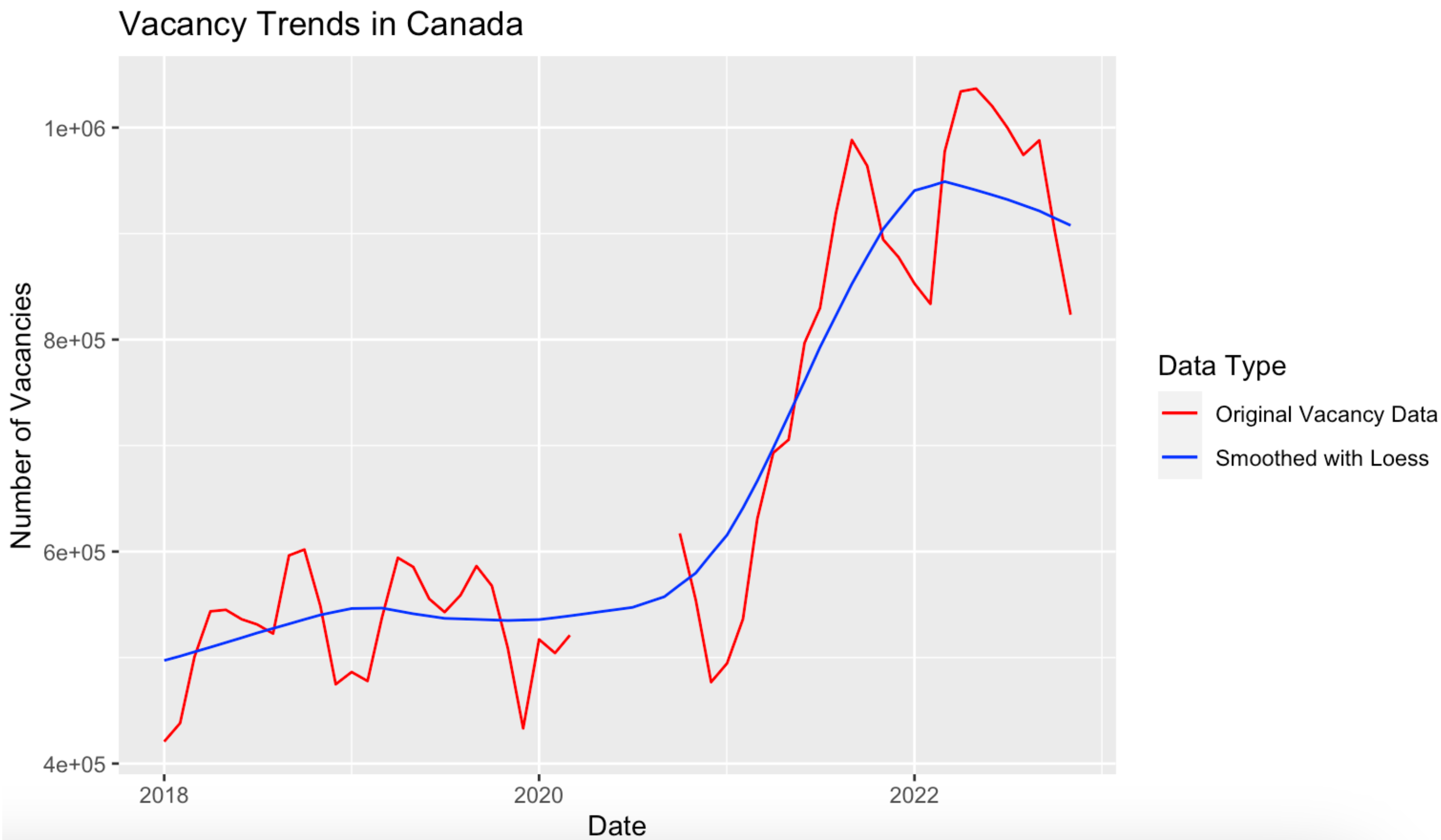

Figure 6- Example of Loess Curve fitting of Canadian JVWS data- 2018-01 to 2022-12

The general formula for STL decomposition is as follows:

$$Y_t = T_t + S_t + R_t \tag{1}$$

Y_t is the value of the time series at time t, T_t the trend component, S_t is the seasonal component, and R_t is the remainder.

4.2 Order of integration and Cointegration

In our analysis, we start with the assumption that the time series for job postings WEB_t and official job vacancies $JVWS_t$ are two random walks with stochastic processes integrated I(1):

$$WEB_t = \mu_t + \epsilon_t \sim I(1) \tag{2}$$

$$JVWS_t = \delta_t + N_t \sim I(1) \tag{3}$$

Where:

WEB_t is the stock of job postings at time t.
$JVWS_t$ is the stock of job vacancies at time t.
μ_t, δ_t are constants or intercept terms
ϵ_t, N_t are error terms which are evenly distributed over time, $N(0, \sigma^2)$

We first need to determine if these time series data are stationary. To do this, we use the Augmented Dickey-Fuller (ADF) test to examine each time series. If both series are non-stationary, we take their first differences to obtain stationary series.

Nelson and Plosser (1982) highlighted the issue of high persistence in macroeconomic time series, such as job postings and job vacancies. They demonstrated that many macroeconomic time series exhibit strong autocorrelation, meaning that their current values are highly dependent on their previous values. This high persistence complicates the analysis of these time series, as traditional regression methods are dependent on assumptions of stationarity.

To tackle this challenge, our initial step is to ascertain the stationarity of these time series data. We use the Augmented Dickey-Fuller (ADF) test to examine each time series. If both series are non-stationary, we take their first differences to obtain stationary series.

To verify the order of integration and stationarity, we use the auto.arima algorithm, which selects the best ARIMA model with the lowest Akaike Information Criterion (AIC) based on various combinations of autoregressive and moving average lags. The AIC helps us choose the most appropriate model by balancing the goodness of fit and model complexity.

Once we have determined the appropriate order of integration, we run a regression on the first differences of the two series and obtain the residuals. By testing the residuals for stationarity with the ADF test, we can determine if the two series are cointegrated. Cointegration implies that the two series are driven by the same underlying process.

If the two series are cointegrated, there must be a linear combination that results in a stationary I(0) process:

$$q_t \equiv \alpha WEB_t + \pi JVWS_t \sim I(0) \quad (4)$$

To establish that, we use the Engle-Granger 1987 method by regressing $WEB_t = \beta JVWS_t + \varepsilon_t$ and we use the ADF test on the least square residuals we estimated.

If the series are cointegrated, we expect them to be stationary $\widehat{\varepsilon_t} = \hat{\beta} JVWS_t - WEB_t \sim I(0)$.

4.3 Minimum Observations Requirement

A foundational aspect of time series forecasting is determining the minimum number of observations required for the chosen models. This not only ensures the reliability and accuracy of the forecasts but also guards against overfitting, where the model might capture noise rather than the underlying pattern.

For the STL Decomposition, while there isn't a strict minimum observation requirement, the clarity of the decomposition—particularly in distinguishing between seasonality, trend, and remainder components—benefits from a longer series. For monthly data, a dataset spanning at least two years (24 observations) is often recommended to adequately capture recurring seasonal patterns.

When considering the ARIMA model, especially the airline model variant, a theoretical minimum of 16 observations is required (Hyndman & Kostenko, 2007). However, this is a baseline. Given the intricacies of economic indicators like job postings and vacancies, practical application often demands multiple seasonal cycles (i.e., several years of data) to yield robust and reliable forecasts.

For our study, a particular challenge arises with the NOC data, which spans only the quarters between 2018 to 2022. This limited timeframe poses potential limitations in capturing the full seasonal patterns and long-term trends, which could impact the robustness of our models and analyses.

Lastly, for Cointegration Analysis, while there isn't a strict minimum observation requirement, the power of tests like the Engle-Granger method is amplified with a more extended dataset. This ensures a higher likelihood of detecting genuine long-term relationships between non-stationary series.

In summary, while theoretical minimums provide a starting point, the dynamic nature of job postings and vacancies in Canada, especially during unprecedented times like the pandemic, underscores the value of a more extended dataset. It offers richer context, captures more variability, and leads to more insightful and reliable forecasts.

4.4 Converting job postings from flow to stock

The job vacancy data from Vicinity Jobs provides us with a snapshot of the daily influx of job postings. This data is collected by scraping various job boards throughout Canada on any given day. In contrast, the job vacancy data from JVWS doesn't just provide daily figures. Instead, it offers a comprehensive picture or "stock" of job vacancies in Canada. This data is compiled by Statistics Canada, which determines the total number of job vacancies across different regions, occupations, and industries. They achieve this by surveying a carefully chosen group of employers from every sector. (Rassart and Patak)

In simpler terms, envision a stock as the water already in a bathtub, while the inflow represents the water pouring in from the faucet each minute.. It's important to understand that these two are fundamentally distinct, so it wouldn't be appropriate to directly compare them.

To obtain the stock of Job postings we need to assume on the average duration of a vacancy.

Prior studies have used a fixed duration corresponding to the length of time the offer was visible on a single website. (Turrell, Speigner et al. 2019, Garasto, Djumalieva et al. 2021) However, they did not have information on whether the vacancy was filled or not, as the website would automatically remove the ad after a predetermined period.

Currently, Vicinity Jobs does not offer details on the outflow of postings, meaning there's no data on how many positions have been filled or removed from job boards. Therefore, we tested 30, 60 and 90 days as durations for their hypothetical expiration. We computed net daily flows of posting by adding the postings published on the current day and subtracting expired ones.

$$Stock_m = Stock_{m-1} + \sum_{j \epsilon m} \left(Inflow_j - Outflow_{j-30} \right) \tag{5}$$

This is the formula when the duration of aggregation is set at 30 days where:
$Stock_m$ is the stock of job postings for the current month.
$Inflow_j$ is the inflow of job postings for the current day.
$Outflow_{j-30}$ is the inflow of job postings from 30 days before the current day that we considered to be flushed out.

Instead of assuming an initial $Stock_0$ of job postings we initialize the first value at 0. This should not affect the result of our analysis because the same information on variation of the time series will be comparable over time.

In what follows, we refer to the stock job posting time series computed from job posting inflows at 30, 60 and 90 days as Web30, Web60 and Web90 respectively.

4.5 Interpolation of missing data

The pandemic led to JVWS having missing data from March 1st, 2020, to October 1st, 2020. This absence poses a challenge due to the lack of a benchmark during that span. Observing the job posting time series during this timeframe reveals a general decline, which aligns with the period of mass layoffs and soaring unemployment that was widely reported. Considering the widespread uncertainty of that period, it is expected that job vacancies would have decreased significantly.

While a polynomial interpolation method might have more accurately reflected the concave nature of the gap, we opted for a linear method due to its simplicity, conservative assumptions, and the presumption of a constant rate of change.

Essentially, we're drawing a straight line from March to October 1st, 2020. By doing this we are aware that the variation of job postings during this period will be mostly unexplained and therefore lead to a higher degree of error.

We also need to account for the missing data in small provinces such as Nunavut, Prince Edward Island, Yukon, and Northern Territories which have shown to have unreliable reports of official vacancies. Figure 5

Chapter 5 : Results

We start our analysis with the trend, seasonality, and remainder components of the STL decomposition. Subsequently, we probe for unit roots in both official and web vacancies, and assess cointegration for each time series pair.

Our comprehensive analysis encompasses the national, provincial, and occupational tiers. For clarity's sake, the analysis is reiterated across each level. We evaluate which aggregation duration of job posting flow into stock most accurately represents each job posting series.

5.1 Job postings and job vacancies at the country level

Between January 1st, 2018, and December 31st, 2022, Canada saw a collection of 14 million job postings, precisely 13,900,872. There is a noticeable pattern in the data, with official vacancies peaking in May 2022 and online job postings peaking a month later in June 2022. This could suggest that online job postings tend to remain available for longer than official vacancies. Additionally, during April and May of 2020, online job postings had the lowest values, which corresponds to the period when official vacancies were not available. It's noteworthy that although official vacancies seemed to rise slightly in the months preceding the pandemic, there was a marked decline in job postings beginning in November 2019.

Moreover, while the official data shows a significant peak in September 2021, this peak isn't mirrored in the online job posting data. Figure 7

5.1.1 STL

The trend-cycle component, as depicted in Figure 8 , filters out effects from outliers and seasonality, revealing a consistent trend over time, though with certain noticeable deviations. From 2018 until the middle of 2019, both job postings and official statistics show a constant level. However, job postings seem to have declined from the middle of 2019 to the middle of 2020. This decline is not observed in the trend of the official data, mainly due to the missing data caused by the Covid pandemic, which has been replaced in our dataset by linear interpolation. Additionally, there was no significant decline in the raw data for official vacancies in 2019, unlike the decline observed in job postings.

Both the job postings and the official vacancies exhibit an upward trajectory from October 2020 through January 2022. However, while official vacancies appear to level off quickly and even slightly decrease in the remaining period of observation, job postings take more time to slow down.

In terms of seasonality, distinct differences are evident. The magnitude of the seasonality in our job postings is lower in the job posting data than the official data, this would be expected if each reported vacancy doesn’t always lead to a job posting. As per the LMIC report 36 we cited earlier, some professions are overrepresented or underrepresented in job postings for each vacancy. It's anticipated that jobs in underrepresented occupations might be filled without formal job postings, perhaps through word of mouth or references. Conversely, overrepresented occupations, especially in specialized fields, might produce more job postings than actual vacancies, as companies proactively search for top-tier candidates even before a position becomes available. Therefore, we can infer that at the aggregated country level, a lot of seasonal professions might be underrepresented in job postings.

We can also observe a lag relationship between the official and the job posting data. While the seasonal maximum for official data seems to arrive in September. The seasonal maximums for job postings aggregated at 30 and 60 days is in June and July for those aggregated at 90 days. Seasonal minimums between official and job posting data seem to be in synch.

The remainder components in our series reveal patterns that are not accounted for by trend or seasonality. We observe some similarities between the official and job posting data remainders with notable peaks at the end of 2019 and through the year 2022 which will be expected in series representing the same underlying phenomenon. Additionally, there doesn’t seem to be a high degree of persistence in either series.

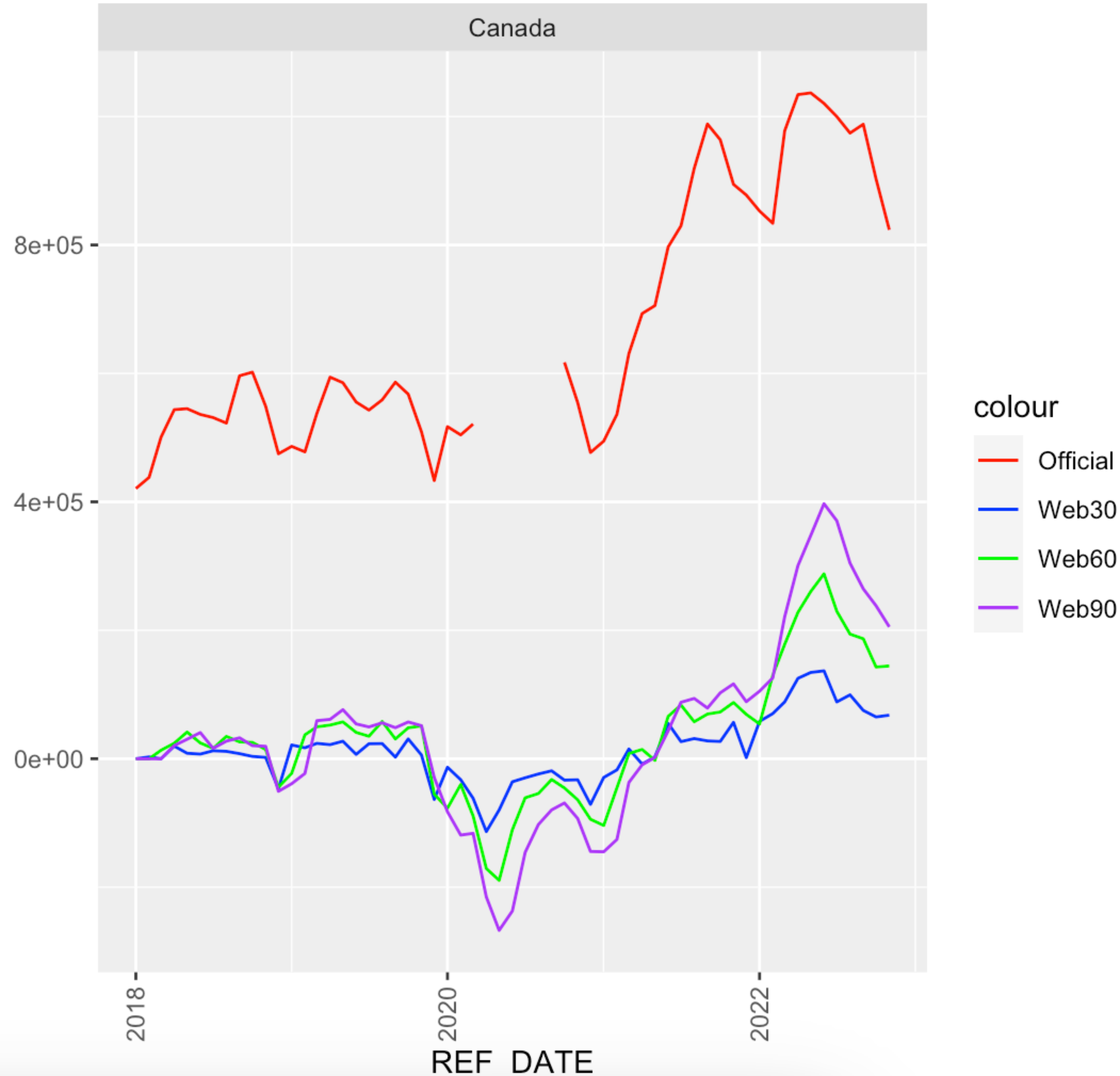

Figure 7 Official and Online data on vacancies in Canada, January 2018 to December 2022.

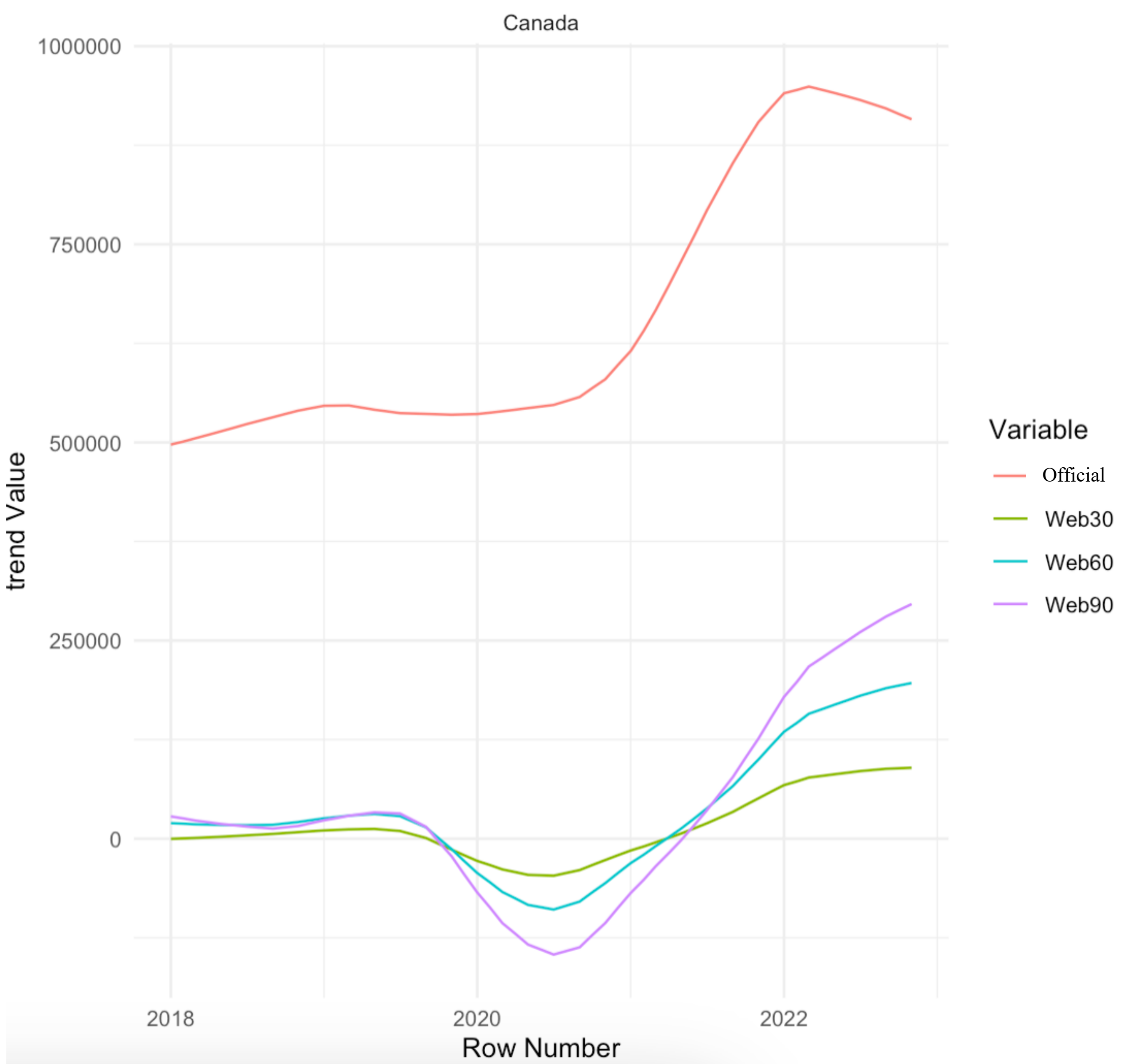

Figure 8 Trend cycle component of Official and Online data on vacancies in Canada, January 2018 to December 2022.

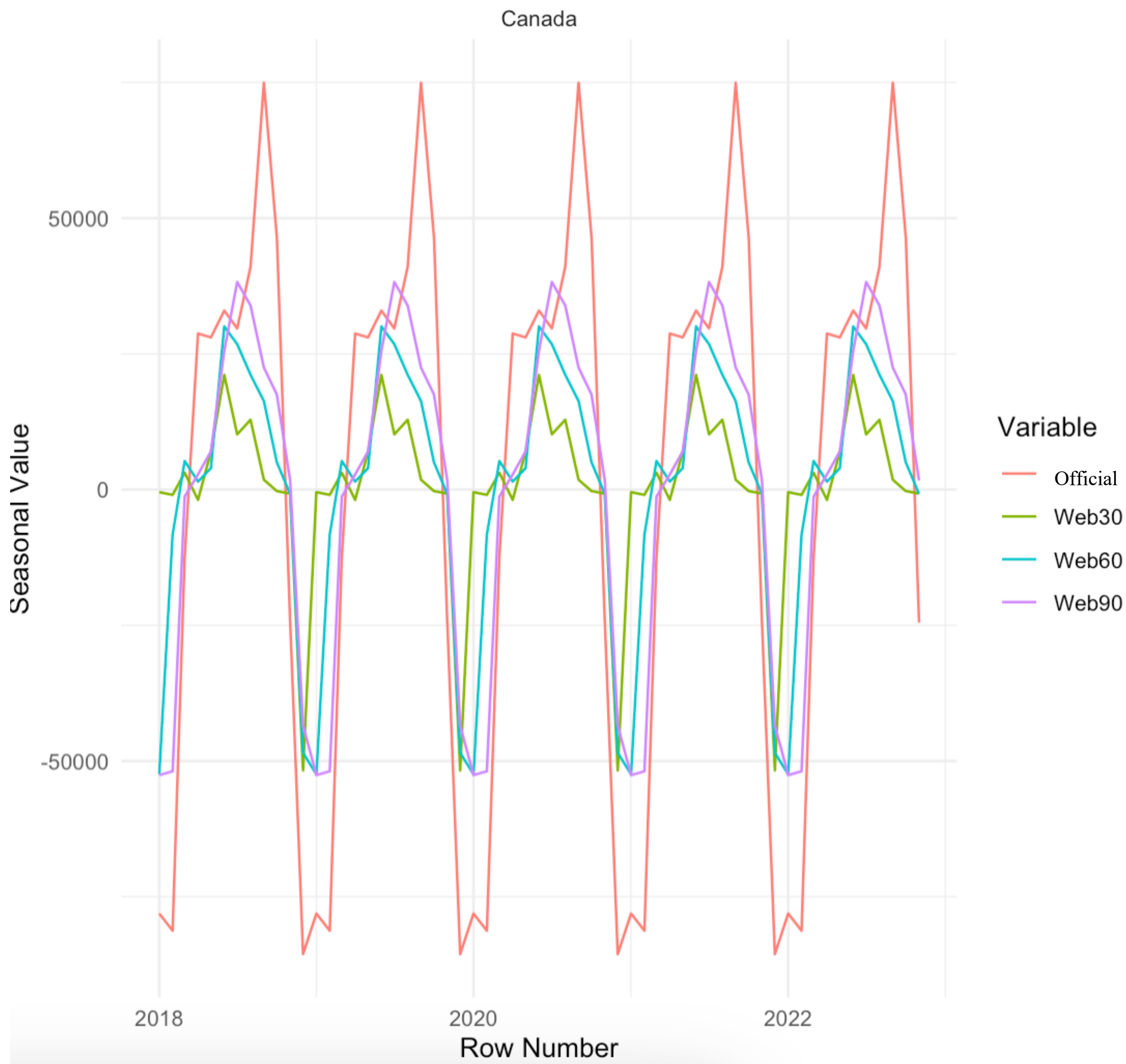

Figure 9 Seasonal component of Official and Online data on vacancies in Canada, January 2018 to December 2022.

Figure 10 Irregular component of Official and Online data on vacancies in Canada, January 2018 to December 2022.

5.1.2 Integration and Cointegration

Table 2 presents the cointegration results for Canada at the three levels of aggregation for job postings. The table shows that job postings aggregated at 30 days and official vacancies are cointegrated, indicating that they have a long-run relationship. The long-run coefficient estimate for this pair is 2.8874 (with a standard error of 0.3222) and is statistically significant at the 1% level. This suggests that a 1% increase in job postings aggregated at 30 days is associated with a 2.8874% increase in official vacancies in the long run, holding other factors constant.

Similarly, job postings aggregated at 60 days and official vacancies are also cointegrated, with a long-run coefficient estimate of 1.5158 (with a standard error of 0.1599) and statistically significant at the 1% level. This implies that a 1% increase in job postings aggregated at 60 days is associated with a 1.5158% increase in official vacancies in the long run, holding other factors constant.

In addition, official vacancies and job postings aggregated at 90 days are also cointegrated, with a long-run coefficient estimate of 1.0338 (with a standard error of 0.1102) and statistically significant at the 1% level. This suggests that a 1% increase in job postings aggregated at 90 days is associated with a 1.0338% increase in official vacancies in the long run, holding other factors constant.

The Table 2-Cointegration results, monthly at country level, shows that job postings aggregated at different intervals have varying levels of accuracy in predicting official vacancies. Job postings aggregated at 30 days have the highest long-run coefficient estimate (2.8874), indicating a strong relationship with official vacancies compared to other levels of aggregation. Meanwhile, job postings aggregated at 60 days have a lower coefficient estimate (1.5158) but has the best fit to the data with an R-Squared value of 0.6118 which is higher than the one at 30 days. This indicates that this model provides a more accurate representation of the relationship between job postings and official vacancies. Job postings aggregated at 90 days have a coefficient estimate (1.0338) that is lower than both 30 and 60 days, with an R-Squared value of 0.6071, which is slightly lower than the 60-day aggregation but still indicating a good fit to the data.

Regarding the unit root tests, all the series are integrated of order 1 (i.e., I(1)), which means that all series present unit roots and which implies characteristics of non-stationarity. The unit root t-test for job postings aggregated at 30 days resulted in a value of -2.4735, which is statistically significant at the 5% level of significance, suggesting that this series is stationary in first differences. For job postings aggregated at 60 days and 90 days, the unit root t-test resulted in values of -2.3203 and -2.2599, respectively, which are also statistically significant at the 5% level of significance, indicating that these series are also stationary in first differences. Therefore, we can conclude that all three pairs of time series are cointegrated. Considering a critical value of −1.95 at the 5% level of significance, these results suggest that the null hypothesis of no cointegration can be rejected for all three pairs, and we can infer that the two-time series in each pair have a long-run relationship.

In conclusion, the cointegration findings indicate a sustained relationship between job postings and official vacancies in Canada across the 30-day, 60-day, and 90-day aggregation periods. An increase in job postings is associated with an increase in official vacancies in the long run, but it is important to note that the causality may not necessarily be in this direction.

Province	Series	I(d)	Cointegration	Long Run Coeff.	R^2	Unit root t test
Canada	JVWS	I(1)	Y	2.8874 (0.3222)***	0.5849	-2.4735
	WEB30	I(1)				
Canada	JVWS	I(1)	Y	1.5158 (0.1599)***	0.6118	-2.3203
	WEB60	I(1)				
Canada	JVWS	I(1)	Y	1.0338 (0.1102)***	0.6071	-2.2599
	WEB90	I(1)				

*p value $<$ 0.010 , **p value $<$ 0.005, ***p value $<$ 0.001

Table 2-Cointegration results, monthly at country level

5.2 Job postings and job vacancies at the province level

At the provincial level, there was a relatively consistent period of growth from January 2018 to December 2019. However, this was interrupted by the profound effects of the COVID-19 pandemic from January 2020 to December 2022, a factor not reflected in the official data. Additionally, we have noticed significant gaps in the official data for series in smaller provinces such as Yukon, Prince Edward Island, Northwest Territories, and Nunavut.

This discrepancy can be attributed to the relatively smaller sample sizes that some provinces contend with. While a larger province like Ontario may report job vacancies ranging from 200,000 to 400,000 during the reported period, smaller provinces may only report in the thousands or even the hundreds, with the smallest provinces like Nunavut reporting in the low hundreds. Figure 11

Overall if we aggregate the data at monthly levels, we can see that the stock of job posting seem to be systematically lower than the official statistics.

After analyzing the three methods of aggregation, we have observed that constructing the stocks with a longer frequency result in the series capturing a greater degree of volatility in their fluctuations. This might be because aggregating the series that way smooth out smaller fluctuations giving us a clearer picture of the underlying trend. Figure 4 &Figure 5.

5.2.1 STL

Generally, trends mirror those at the national level (as seen in Figure 12), but there are notable exceptions warranting closer examination. While it would be time-consuming to discuss each trend individually, we can focus on certain outliers. One particular example is the Yukon, where there has been a substantial increase in job postings despite official vacancies remaining stagnant during the same period.

The seasonality of job postings varies across different regions in Canada. Figure 13. Some provinces, such as Prince Edward Island, have a clear pattern of seasonality that matches up well with official data, while others, like Newfoundland Labrador, Manitoba, and Ontario, show lower levels of seasonality in job postings than in official data. In Nunavut, employers tend to plan ahead with a six-month lead time for job postings, likely due to seasonal fluctuations in workforce needs. Meanwhile, the situation is more variable in the Northwestern Territories, where seasonality is apparent but lacks a consistent pattern. Aggregating job postings over different time periods can help to adjust for seasonality and bring the data closer in magnitude to official figures. However, there are trade-offs to consider. Longer aggregation periods may reveal more volatility in job postings, while shorter periods may smooth out the data and obscure underlying trends.

The remainder, Figure 14,component can be useful in understanding the discrepancies between official data and job postings in smaller provinces such as the Northwestern Territories, Nunavut, Prince Edward Island, and Newfoundland Labrador. In these regions, official data shows more volatility than job postings, which could be due to the small sample size used in the job vacancy and wage survey. While this can lead to a lot of volatility, job postings may provide a more accurate picture of the labor market in smaller provinces.

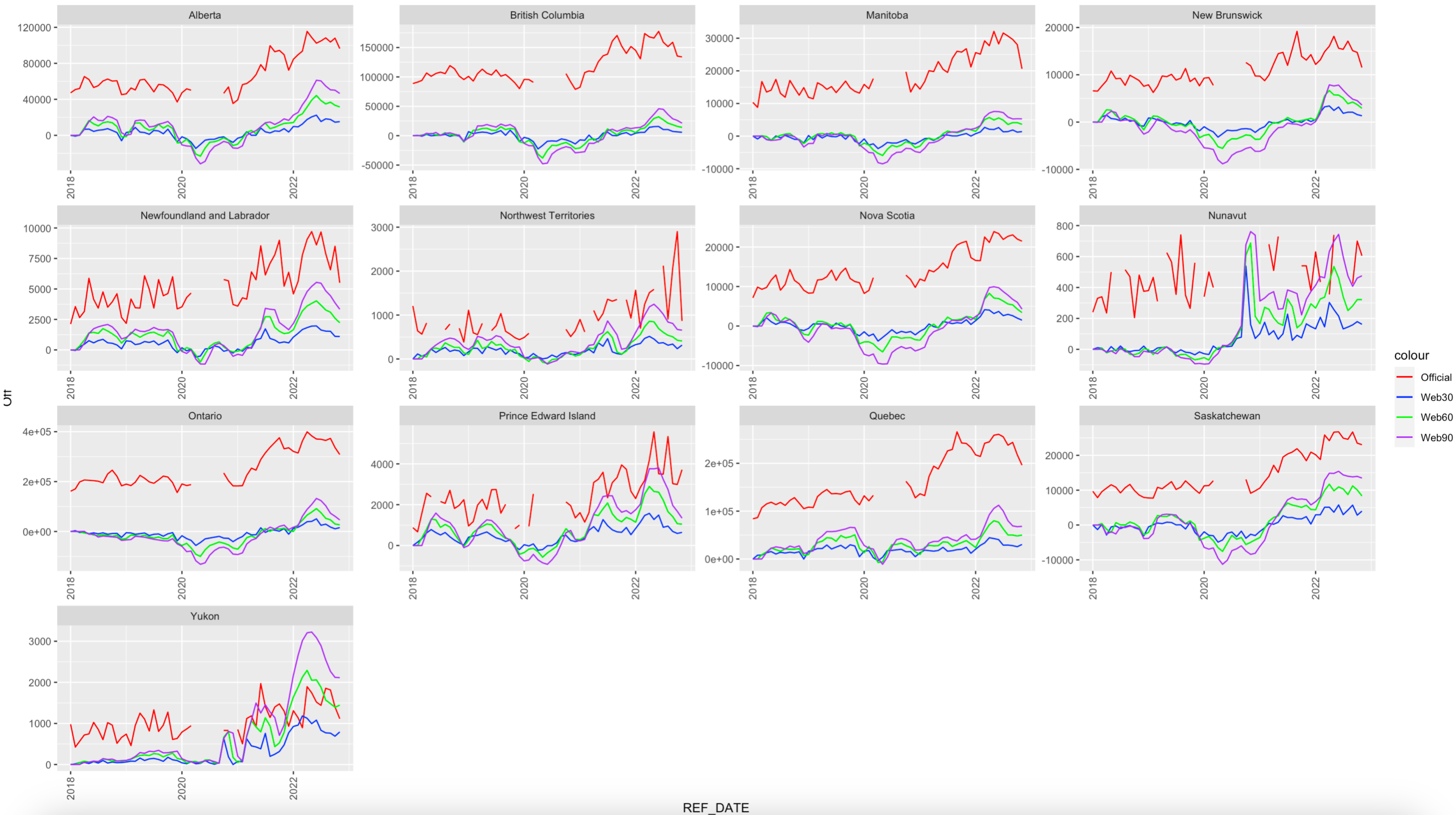

Figure 11Official and Online data on vacancies in Canadian Provinces, January 2018 to December 2022.

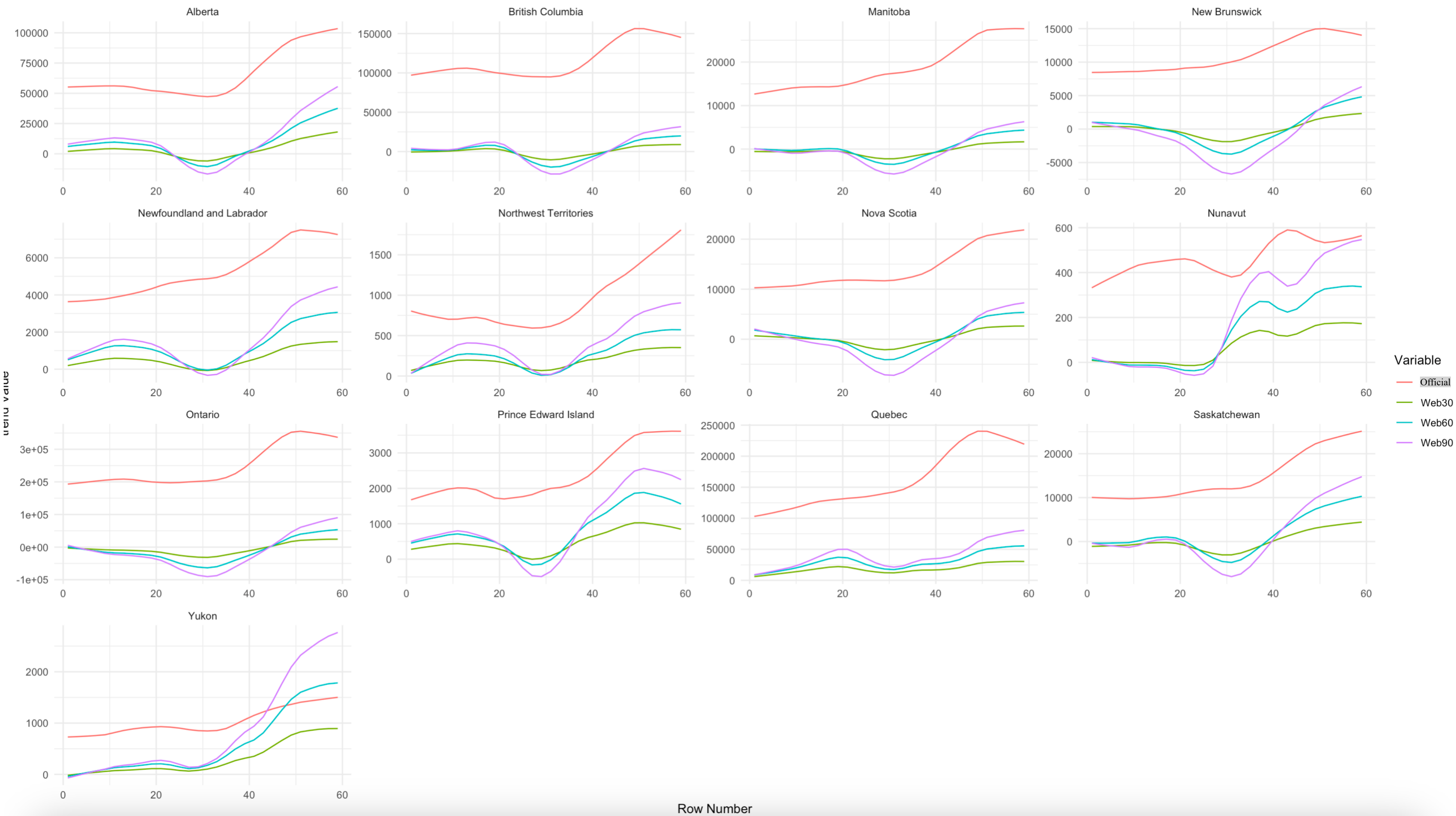

Figure 12 Trend cycle component of Official and Online data on vacancies in Canadian Provinces, January 2018 to December 2022.

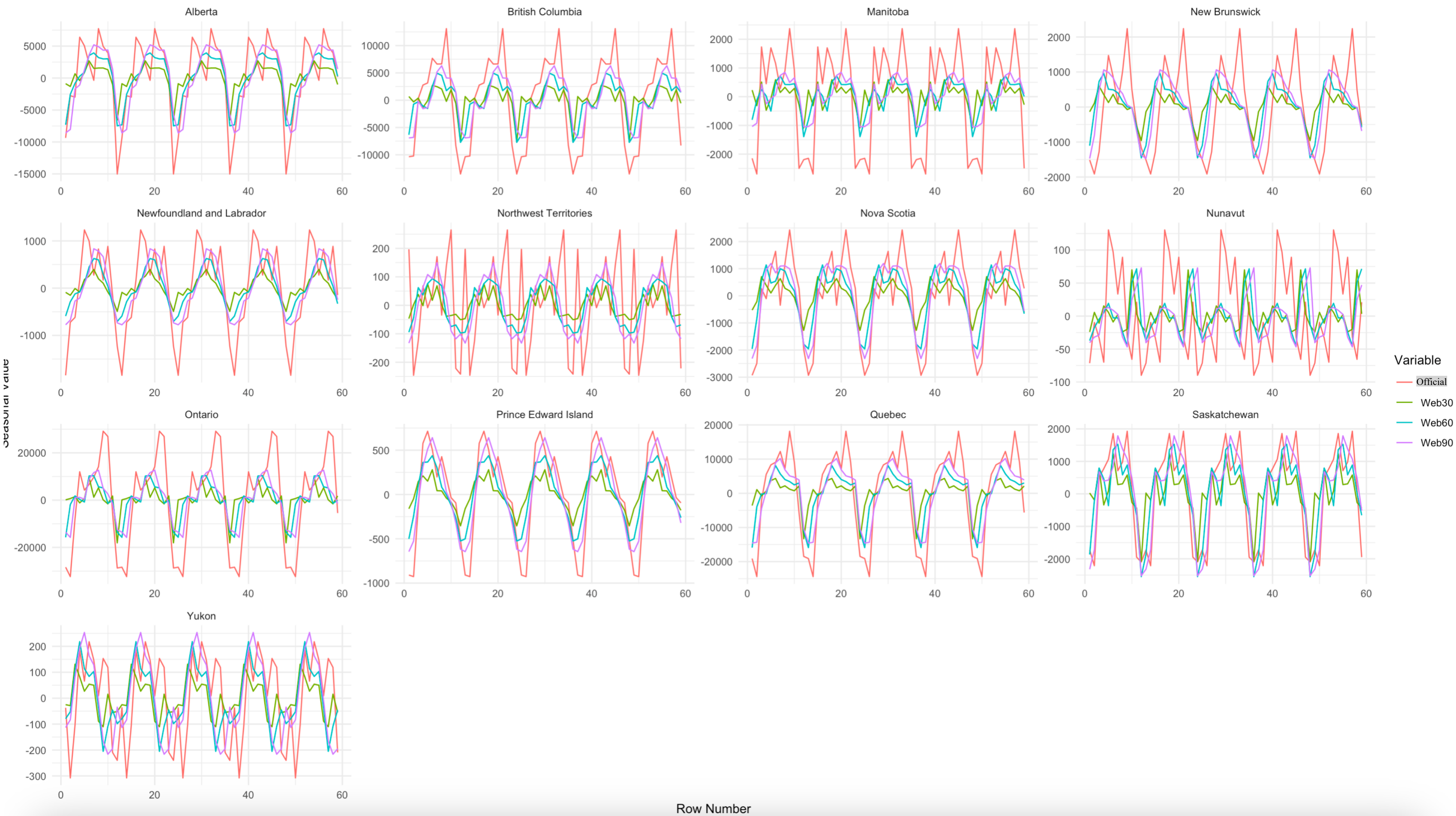

Figure 13 Seasonal component of Official and Online data on vacancies in Canadian Provinces, January 2018 to December 2022.

Figure 14 Remainder component of Official and Online data on vacancies in Canadian Provinces, January 2018 to December 2022.

5.2.2 Integration and Cointegration

Cointegration tests indicate a persistent relationship between online job postings and official vacancies across all Canadian provinces. The long-run coefficient estimates are positive and statistically significant at all three aggregation intervals (30, 60, and 90 days), indicating that an increase in official vacancies leads to an increase in job postings in the long run.

The R-squared values vary across the provinces, ranging from 24% to 69%. The highest values are observed in provinces with larger populations, such as Ontario and Quebec, suggesting that the number of job postings on online platforms is more closely related to the total number of job vacancies in larger provinces.

When examining the R-Squared values across the tables, we observe the following ranges: in the 30-day table, R-Squared values range from 0.05418 for Nunavut to 0.6475 for Saskatchewan; in the 60-day table, R-Squared values range from 0.08598 for Nunavut to 0.6848 for Saskatchewan; and in the 90-day table, R-Squared values range from 0.09228 for Nunavut to 0.6903 for Saskatchewan.

The R-squared values in the 60-day and 90-day tables are generally lower compared to the R-squared values in the 30-day table, indicating a weaker fit of the model for some provinces. However, the R-squared values are still significant for most provinces, suggesting a stable and predictable relationship between the two-time series in the long run.

If we examine long run coefficient at 30 days, we can say that a 1% increase in job postings in Alberta for example, will result in a 2.29% increase in official vacancies (with a standard deviation of 0.2144).

The unit root t tests indicate that the residuals of the regression of the official vacancies on the job postings are non-stationary. It is worth noting that the critical values for the unit root t tests are -1.95 at a 5% significance level and -1.61 at a 10%

significance level. All provinces in the analysis show t statistics with absolute values larger than these critical values, indicating that the null hypothesis of no cointegration can be rejected with high confidence.

Province	Series	I(d)	Cointegration	Long Run Coeff.	R^2	Unit root t test
Alberta	JVWS	I(1)	Y	2.2869(0.2144) ***	0.6662	-2.579
	WEB	I(1)				
British Columbia	JVWS	I(1)	Y	2.184(0.354)***	0.4004	-2.3113
	WEB	I(1)				
Manitoba	JVWS	I(1)	Y	2.5512(0.4298)***	0.382	-2.6468
	WEB	I(1)				
New Brunswick	JVWS	I(1)	Y	1.0505(0.2464)***	0.2417	-2.6954
	WEB	I(1)				
Newfoundland and Labrador	JVWS	I(1)	Y	1.9527(0.3229)***	0.3908	-3.1547
	WEB	I(1)				
Northwest Territories	JVWS	I(1)	Y	1.9482(0.4442)***	0.2523	-3.9058
	WEB	I(1)				
Nova Scotia	JVWS	I(1)	Y	1.7976(0.2591)***	0.4579	-2.1969
	WEB	I(1)				
Nunavut	JVWS	I(1)	Y	0.31212(0.17272)	0.05418	-3.7838
	WEB	I(1)				
Ontario	JVWS	I(1)	Y	2.5972(0.2839)***	0.5949	-3.0364
	WEB	I(1)				
Prince Edward Island	JVWS	I(1)	Y	1.6341(0.2365)***	0.4559	-4.3907
	WEB	I(1)				
Quebec	JVWS	I(1)	Y	3.3237(0.5438)***	0.3959	-2.0934
	WEB	I(1)				
Saskatchewan	JVWS	I(1)	Y	1.7895(0.1749)***	0.6475	-2.1241
	WEB	I(1)				
Yukon	JVWS	I(1)	Y	0.66423(0.10496)***	0.4024	-5.023
	WEB	I(1)				

*p value < 0.010 , **p value < 0.005, ***p value < 0.001

Table 3-Cointegration results, monthly at Province level- 30 day aggregation

Province	Series	I(d)	Cointegration	Long Run Coeff.	R^2	Unit root t test
Alberta	JVWS	I(1)	Y	1.2081(0.1086)***	0.6846	-2.478
	WEB	I(1)				
British Columbia	JVWS	I(1)	Y	1.1926(0.1848)***	0.4221	-2.2758
	WEB	I(1)				
Manitoba	JVWS	I(1)	Y	1.3826(0.2260)***	0.3963	-2.6922
	WEB	I(1)				
New Brunswick	JVWS	I(1)	Y	0.5716(0.1245)***	0.27	-2.6091
	WEB	I(1)				
Newfoundland and Labrador	JVWS	I(1)	Y	1.1231(0.1551)***	0.4791	-3.0046
	WEB	I(1)				
Northwest Territories	JVWS	I(1)	Y	1.17874(0.22792)***	0.3194	-4.2203
	WEB	I(1)				
Nova Scotia	JVWS	I(1)	Y	0.9676(0.1282)***	0.4997	-1.9238
	WEB	I(1)				
Nunavut	JVWS	I(1)	Y	0.22129(0.09557)	0.08598	-3.789
	WEB	I(1)				
Ontario	JVWS	I(1)	Y	1.3961(0.1419)***	0.6294	-2.9732
	WEB	I(1)				
Prince Edward Island	JVWS	I(1)	Y	0.8931(0.1155)***	0.5121	-4.3557
	WEB	I(1)				
Quebec	JVWS	I(1)	Y	1.890(0.275)***	0.453	-1.9688
	WEB	I(1)				
Saskatchewan	JVWS	I(1)	Y	0.95937(0.08621)***	0.6848	2.1143
	WEB	I(1)				
Yukon	JVWS	I(1)	Y	0.36104(0.05313)***	0.4378	-5.5155
	WEB	I(1)				
	WEB	I(1)				

*p value < 0.010 , **p value < 0.005, ***p value < 0.001

Table 4 Cointegration results, monthly at Province level- 60 day aggregation

Province	Series	I(d)	Cointegration	Long Run Coeff.	R^2	Unit root t test
Alberta	JVWS	I(1)	Y	0.83278(0.07645)***	0.6755	-2.3687
	WEB	I(1)				
British Columbia	JVWS	I(1)	Y	0.8297(0.1285)***	0.4224	-2.0996
	WEB	I(1)				
Manitoba	JVWS	I(1)	Y	0.9464(0.1558)***	0.3929	-2.6625
	WEB	I(1)				
New Brunswick	JVWS	I(1)	Y	0.38645(0.08604)***	0.2614	-2.5483
	WEB	I(1)				
Newfoundland and Labrador	JVWS	I(1)	Y	0.7664(0.1081)***	0.4687	-2.8494
	WEB	I(1)				
Northwest Territories	JVWS	I(1)	Y	0.89683(0.14803)***	0.3917	-4.2744
	WEB	I(1)				
Nova Scotia	JVWS	I(1)	Y	0.67329(0.087520)***	0.5094	-1.8317
	WEB	I(1)				
Nunavut	JVWS	I(1)	Y	0.16127(0.06700)*	0.09228	-3.8777
	WEB	I(1)				
Ontario	JVWS	I(1)	Y	0.95364(0.09722)***	0.628	-2.7803
	WEB	I(1)				
Prince Edward Island	JVWS	I(1)	Y	0.6373(0.0754)***	0.5563	-4.5463
	WEB	I(1)				
Quebec	JVWS	I(1)	Y	1.3616(0.1881)***	0.479	-1.8713
	WEB	I(1)				
Saskatchewan	JVWS	I(1)	Y	0.65881(0.65881)***	0.6903	-2.0819
	WEB	I(1)				
Yukon	JVWS	I(1)	Y	0.25515(0.03494)***	0.4833	-5.557
	WEB	I(1)				

*p value $<$ 0.010 , **p value $<$ 0.005, ***p value $<$ 0.001

Table 5 Cointegration results, monthly at level- 90 day aggregation

5.3 Job postings and job vacancies at Occupation levels

In this section, we will analyze the relationship between online job postings and official job vacancy data at the occupational level. It is important to note that the data available for this level of analysis is less extensive than at the country, or provincial level as we have access to quarterly data rather than monthly data, for the Job vacancy and wage survey which may affect the granularity of our findings.

5.3.1 STL

When examining trend discrepancies between official data and job postings, some notable differences emerge across various occupational fields. For natural and applied sciences and related occupations (4), official and job posting trends appear to diverge over time, suggesting potential discrepancies in the representation of these occupations.

In the case of natural resources and agriculture (5) and occupations in manufacturing and utilities (7), job postings remain constant over time, while trends in official data show variation. This indicates that job posting data may not fully capture the dynamics of these sectors.

On the other hand, sales and service occupations (8), as well as education, law, social, community, and government services occupations (2) seem to be accurately portrayed, with job posting trends aligning closely with official data.

A stronger magnitude of seasonality in job postings than official data is observed for business, finance, and administration occupations (0), management occupations (3), and natural and applied sciences and related occupations (4). Conversely, official vacancy seasonality exhibits a greater magnitude for natural resources and agriculture (5), occupations in art, culture, recreation, and sport (6), and health occupations (2).

The differences in the magnitude of seasonality can provide valuable insights into which occupations are more or less represented in the job posting data.

When observing the peak of seasonality job postings seem to lead for finance and administration occupations (0), management occupations (3), natural and applied sciences and related occupations (4), natural resources and agriculture (5), and occupations in manufacturing and utilities (7). Notably, for health occupations (2) and sales and service occupations (8) , this lead is only observed at 30-day aggregation.

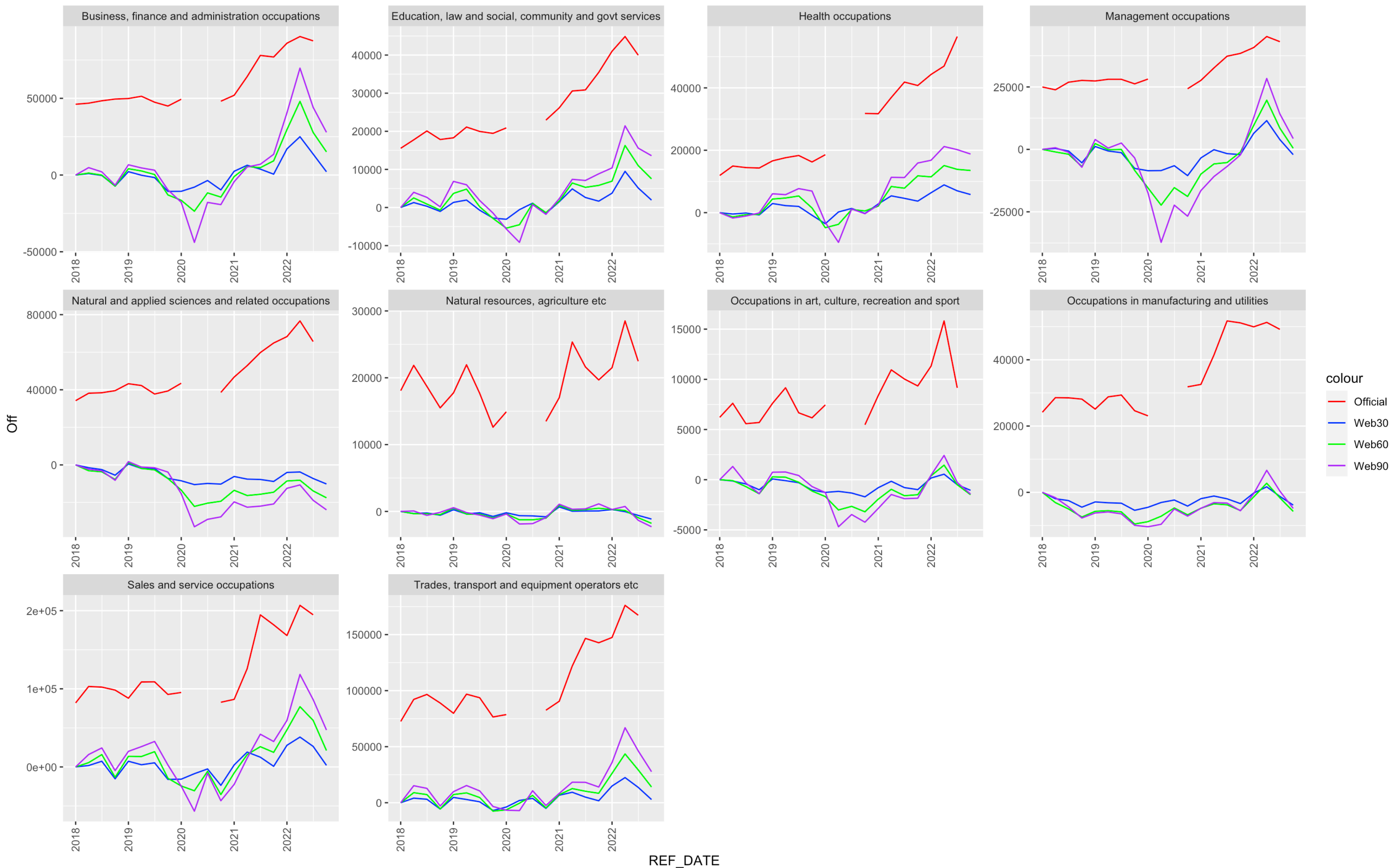

Figure 15- Quarterly Official and Online Job Vacancies in the 10 1-digit NOC from Q1 2018 to Q4 2022

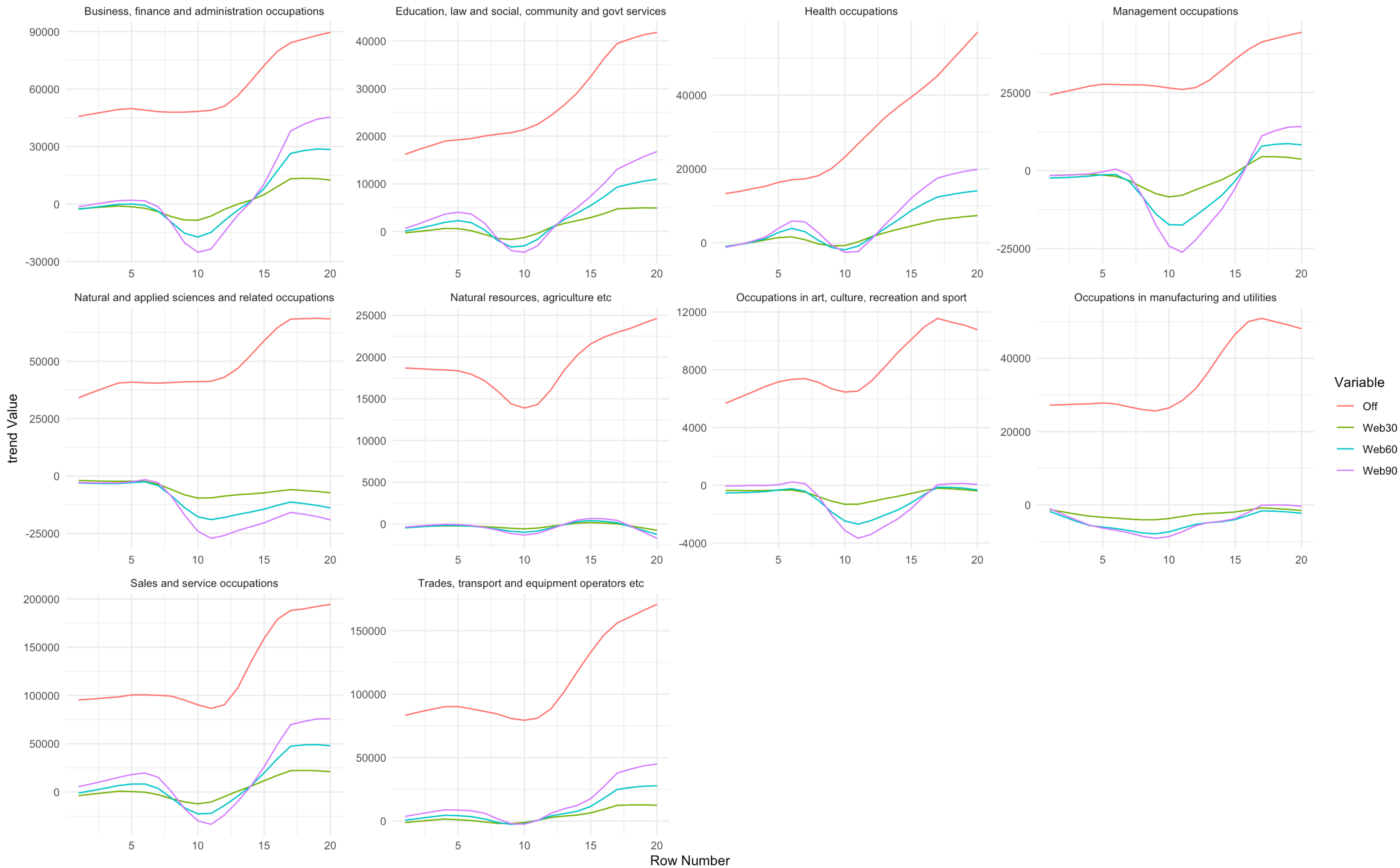

Figure 16-Trend Cycle of Quarterly Official and Online Job Vacancies in the 10 1-digit NOC from Q1 2018 to Q4 2022

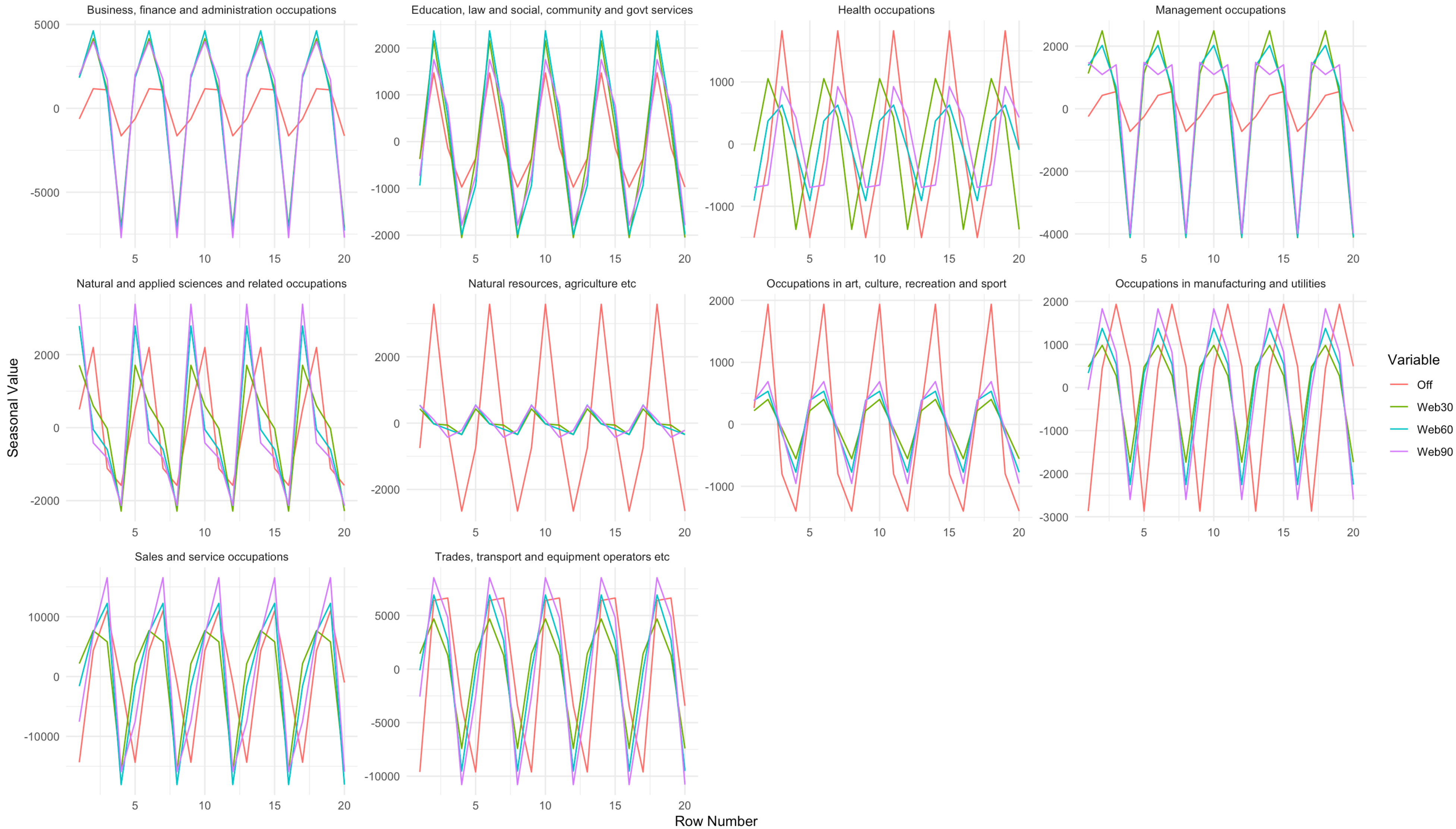

Figure 17- Seasonality of Quarterly Official and Online Job Vacancies in the 10 1-digit NOC from Q1 2018 to Q4 2022

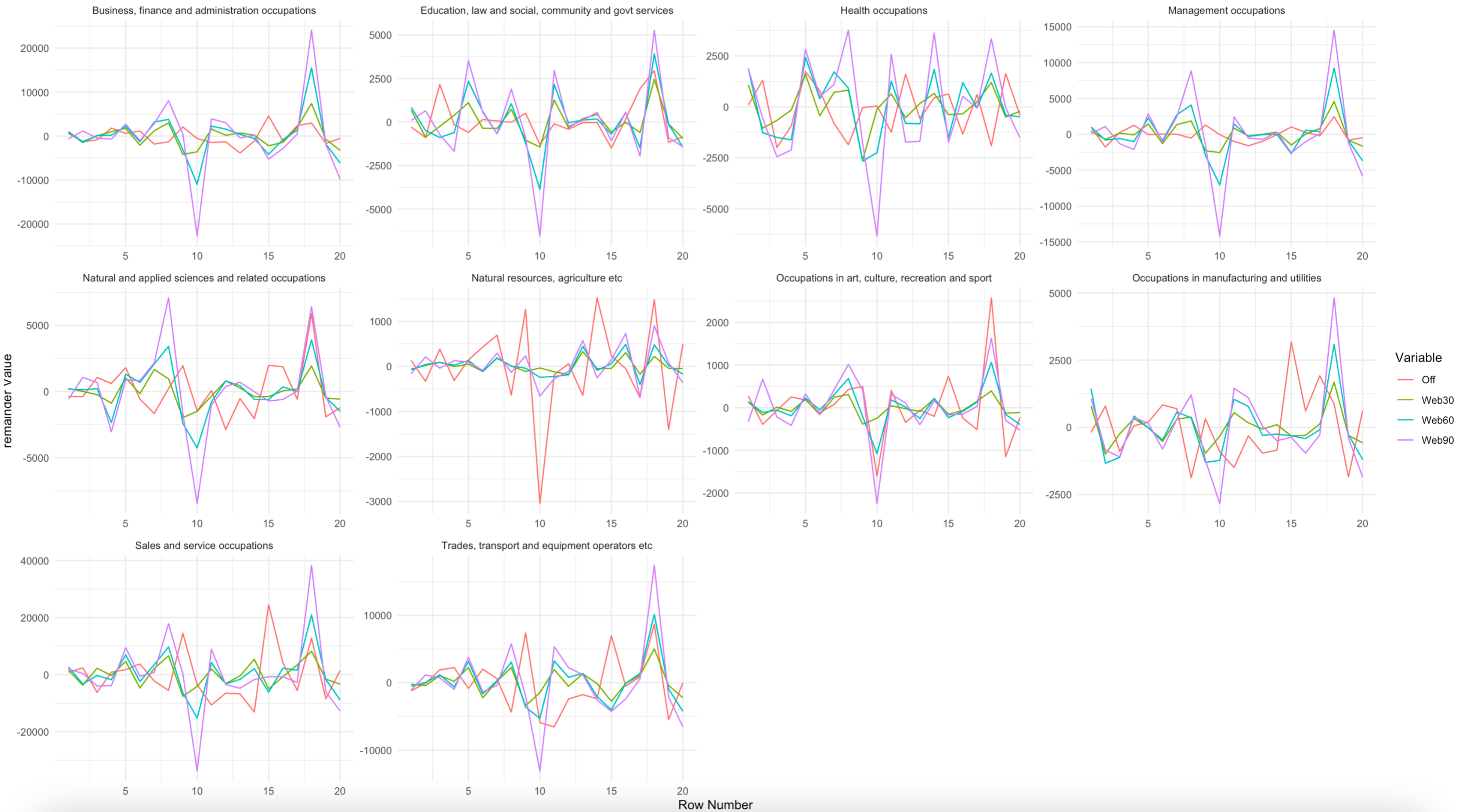

Figure 18 - Remainder of Quarterly Official and Online Job Vacancies in the 10 1-digit NOC from Q1 2018 to Q4 2022

5.3.2 Integration and Cointegration at occupation level

The cointegration results for job postings and official series across various occupations reveal interesting insights, depending on the aggregation level of job postings. At a 30-day aggregation, we find that business, finance and administration occupations, health occupations, occupations in education, law, social, community and government services, trades, transport and equipment operators and related occupations, and natural resources, agriculture and related production occupations exhibit cointegration with their respective official series.

When job postings are aggregated at 60 days, cointegration is observed for business, finance and administration occupations, occupations in education, law and social, community and government services, sales and service occupations, trades, transport and equipment operators and related occupations, and natural resources, agriculture and related production occupations.

At a 90-day aggregation level, cointegration is found for occupations in education, law and social, community and government services, and trades, transport and equipment operators and related occupations with their respective official series.

Upon examining the results, it appears that the 60-day aggregation level demonstrates the strongest cointegration across the greatest number of occupations. Among the various occupations, trades, transport and equipment operators and related occupations, and occupations in education, law, and social, community, and government services show the most consistent cointegration across different aggregation levels. This suggests that job posting trends in these occupational fields may provide the best predictions for official vacancy data when compared to other occupations.

At this level the R-squared values for cointegrated occupations display considerable variation, reflecting differing levels of correlation between the official vacancies and job postings time series. The strongest relationships are observed in business, finance, and administration occupations; education, law, and social, community, and government services; sales and service occupations; and trades, transport, and equipment operators and related occupations. These occupations exhibit R-squared values of 0.6953, 0.6686, 0.6428, and 0.6895, respectively, indicating a relatively strong fit for the models.

For natural resources, agriculture, and related production occupations, the R-squared value is lower at 0.09994, suggesting a weaker relationship between the official vacancies and job postings time series. Nonetheless, the cointegration in this occupation category indicates a long-term equilibrium between the two variables.

The long-run coefficients of the cointegrated occupations provide further insight into the relationship between job postings and official vacancies at the 60-day aggregation. These coefficients can be interpreted as the response of official vacancies to a change in job postings, indicating the extent to which official vacancies and job postings move together in the long run.

For instance, in business, finance, and administration occupations, the long-run coefficient is 0.8274, suggesting that a 1% increase in job postings is associated with an approximately 0.83% increase in official vacancies, implying a strong positive relationship between the two variables. However, this relationship is weaker compared to the occupations in education, law, and social, community, and government services, where the long-run coefficient is 1.4329. This higher coefficient indicates that a 1% increase in job postings corresponds to a 1.43% increase in official vacancies, highlighting a stronger relationship between the variables in the education, law, and social, community, and government services category as compared to the business, finance, and administration occupations.

In sales and service occupations, the long-run coefficient is 1.2739, signifying that a 1% increase in job postings is linked to a 1.27% increase in official vacancies.

This also demonstrates a strong positive relationship between job postings and official vacancies.

NOC	Series	I(d)	Cointegration	Long Run Coeff.	R^2	Unit root t test
Management occupations	JVWS	I(1)	N			
	WEB	I(1)				
Business, finance and administration occupations	JVWS	I(1)	Y	1.4382(0.2741)***	0.6047	-2.2383
	WEB	I(1)				
Natural and applied sciences and related occupations	JVWS	I(1)	N			
	WEB	I(1)				
Health occupations	JVWS	I(1)	Y	3.7815(0.6019)***	0.6868	-2.1139
	WEB	I(1)				
Occupations in education, law and social, community and government services	JVWS	I(1)	Y	2.416(0.476)***	0.5887	-2.27
	WEB	I(1)				
Occupations in art, culture, recreation and sport	JVWS	I(1)	N			
	WEB	I(1)				
Sales and service occupations	JVWS	I(0)	N			
	WEB	I(1)				
Trades, transport and equipment operators and related occupations	JVWS	I(1)	Y	3.4723(0.8117)***	0.5041	-1.8686
	WEB	I(1)				
Natural resources, agriculture and related production occupations	JVWS	I(1)	Y	2.589(2.181)	0.07263	-1.6657
	WEB	I(1)				

*p value $<$ 0.010 , **p value $<$ 0.005, ***p value $<$ 0.001

Table 6 Cointegration results, Quarterly at occupation level- 30 day aggregation

NOC	Series	I(d)	Cointegration	Long Run Coeff.	R^2	Unit root t test
Management	JVWS	I(1)	N			
	WEB	I(1)				
Business, finance and administration	JVWS	I(1)	Y	0.8274(0.1291)***	0.6953	-2.2473
	WEB	I(1)				
Natural and applied sciences and related	JVWS	I(1)	N			
	WEB	I(1)				
Health	JVWS	I(1)	N			
	WEB	I(1)				
Occupations in education, law and social, community and government services	JVWS	I(1)	Y	1.4329(0.2378)***	0.6686	-2.166
	WEB	I(1)				
Occupations in art, culture, recreation and sport	JVWS	I(1)	N			
	WEB	I(1)				
Sales and service	JVWS	I(0)	N			
	WEB	I(1)				
Trades, transport and equipment operators and related	JVWS	I(1)	Y	2.308(0.365)***	0.6895	-2.8661
	WEB	I(1)				
Natural resources, agriculture and related production	JVWS	I(1)	Y	1.912(1.352)	0.09994	-1.7831
	WEB	I(1)				
Occupations in manufacturing and utilities	JVWS	I(1)	N			
	WEB	I(1)				

*p value < 0.010 , **p value < 0.005, ***p value < 0.001

Table 7 Cointegration results, monthly at Province level- 60 day aggregation

Province	Series	I(d)	Cointegration	Long Run Coeff.	R^2	Unit root t test
Management occupations	JVWS	I(1)	N			
	WEB	I(1)				
Business, finance and administration occupations	JVWS	I(1)	N			
	WEB	I(0)				
Natural and applied sciences and related occupations	JVWS	I(1)	N			
	WEB	I(1)				
Health occupations	JVWS	I(1)	N			
	WEB	I(1)				
Occupations in education, law and social, community and government services	JVWS	I(1)	Y	1.0278(0.1758)***	0.6552	-1.8547
	WEB	I(1)				
Occupations in art, culture, recreation and sport	JVWS	I(1)	N			
	WEB	I(1)				
Sales and service occupations	JVWS	I(1)	N			
	WEB	I(0)				
Trades, transport and equipment operators and related occupations	JVWS	I(1)	Y	1.6393(0.2251)***	0.7465	-2.9441
	WEB	I(1)				
Natural resources, agriculture and related production occupations	JVWS	I(1)	N			
	WEB	I(0)				
Occupations in manufacturing and utilities	JVWS	I(1)	N			
	WEB	I(0)				

*p value < 0.010 , **p value < 0.005, ***p value < 0.001

Table 8 Cointegration results, monthly at Province level- 90 day aggregation

Chapter 5 : Conclusion

In conclusion, our study underscores the value of web sources for understanding Canada's labor market dynamics. Specifically, we utilized Vicinity Jobs, a job posting aggregator, to gather data on job vacancies at the national, provincial, and occupational levels. Our study employed cointegration analysis, trend and seasonality comparisons, and various aggregation levels to assess the accuracy and representation of online job postings compared to official statistics. Despite our research spanning a relatively brief four-year period, our statistical findings indicate that job vacancy data sourced online reliably mirrors actual job vacancy trends in Canada.

On a national scale, we identified a robust and significant long-term correlation between official job vacancies and online job postings. We found the 60-day aggregation to be the most accurate.

At the provincial level, our findings suggest a sustained relationship between online and official job postings across all Canadian provinces. The 30-day aggregation, boasting the highest R-Squared values, emerged as the most precise.

Occupationally, our research revealed both congruencies and variances in trends and seasonality across diverse occupational sectors when comparing job postings to official data. The cointegration results varied across different aggregation levels and occupations, with the 60-day aggregation level demonstrating the strongest cointegration across the greatest number of occupations.

However, our study faced challenges, notably the absence of official data during the COVID-19 pandemic months. This data void might have compromised our model's robustness. Future endeavors could address this by integrating a structural break in the analysis for those particular months, accommodating the unique disruptions induced by the pandemic. Another constraint was the quarterly frequency of occupational data. This limited data frequency might have curtailed the depth of our occupational insights, somewhat diminishing the breadth of our analysis.

Building upon the foundation of our cointegration analysis, a logical next step would involve forecasting future job vacancy trends using our model and subsequently comparing these predictions with real-time data as they become available. This will serve a dual purpose: validating the reliability and robustness of our model and providing stakeholders with actionable intelligence on future labor market dynamics.

To do this, we could employ Vector Error Correction Models (VECM). Given that our variables are cointegrated, a VECM would be apt, as it leverages both the long-run equilibrium relationship and short-run dynamics of the series. By forecasting job vacancies using this model, we can obtain out-of-sample predictions for both online and official job vacancies. Subsequently, as new official job vacancy data gets published, we can compare our forecasts against this real-time data. This would not only test the accuracy of our predictions but also affirm the validity of our model in capturing the underlying dynamics of the labor market.

Furthermore, complementing this with rolling-forecast error variance decomposition could give us insights into which variables contribute most to forecast errors. This would aid in refining our model and improving its forecasting capabilities. In essence, this iterative process of forecasting and validation will help in ensuring that our model remains relevant and accurate in ever-changing labor market scenarios.

In our exploration, we operated at a relatively broad level of detail, centering on provincial and major occupational tiers. However, there is potential for further exploration by examining sub-major and minor groups of occupations, as well as by investigating more geographically granular areas such as economic regions. This approach could help to identify any major discrepancies between metropolitan and rural economic regions, offering a more comprehensive understanding of labor market dynamics across different areas and occupational categories. While this deeper dive would undoubtedly demand more computational resources, the resultant enriched insights could greatly enhance our grasp of the labor market's intricacies.

Bibliography

Askitas, N. and K. F. Zimmermann (2009). "Google econometrics and unemployment forecasting." Applied Economics Quarterly **55**(2): 107-120.

Asselin, S. (2005). "Qui Travaille selon des horaires quotidiens non standard." Conditions de Vie **9**(3).

Barnichon, R. and A. H. Shapiro (2022). "What's the Best Measure of Economic Slack?" FRBSF Economic Letter **4**.

Choi, H. and H. Varian (2009). "Predicting initial claims for unemployment benefits." Google Inc **1**(2009): 1-5.

Cleveland, R. B., et al. (1990). "STL: A seasonal-trend decomposition." J. Off. Stat **6**(1): 3-73.

de Pedraza, P., et al. (2019). "Survey vs scraped data: comparing time series properties of web and survey vacancy data." IZA Journal of Labor Economics **8**(1).

Garasto, S., et al. (2021). "Developing experimental estimates of regional skill demand."

Gingras, Y. and R. Roy (1998). Y at-il une pénurie de main-d'oeuvre qualifiée au Canada?, Développement des ressources humaines Canada, Direction générale de la ….

Hyndman, R. J., & Kostenko, A. V. (2007). "MINIMUM SAMPLE SIZE REQUIREMENTS FOR SEASONAL FORECASTING MODELS." FORESIGHT, Spring Issue 6.

Ipsos-Reid (2002). ONLINE JOB HUNTING NOW AMONG MOST POPULAR ONLINE ACTIVITIES, Ipsos.

Jones, S. R., et al. (2020). "Waiting for recovery: The Canadian labour market in June 2020." Canadian Public Policy **46**(S2): S102-S118.

Koenig, E. F., Dolmas, S., & Piger, J. (2003). The Use and Abuse of Real-Time Data in Economic Forecasting. The Review of Economics and Statistics, 85(3), 618-628. Retrieved from http://www.jstor.org/stable/3211701

Kuhn, P. J. (2014). "The internet as a labor market matchmaker." IZA World of Labor.

Lovaglio, P. G., et al. (2020). "Comparing time series characteristics of official and web job vacancy data." Quality & Quantity **54**(1): 85-98.

Labour Market Information Council. (2020). "How Representative Are Online Job Postings?" LMI Insight Report no. 36, October. Retrieved from https://lmic-cimt.ca/fr/des-publications/rapport-de-perspectives-de-limt-36/.

Mitchell, D. (2015). "Forecasting the Canadian Unemployment Rate Using Internet Searches."

Rassart, E. and Z. Patak "Producing labour statistics by detailed geography and occupation: Experiences from the new Canadian Job Vacancy and Wage Survey."

Statistics Canada (2001). "Canadian Internet Survey Use."

Statistics Canada (2020). "Canadian Internet Use Survey." from https://www150.statcan.gc.ca/n1/daily-quotidien/210622/dq210622b-eng.htm.

Štefánik, M., et al. (2022). "Using online job postings to predict key labour market indicators." Social Science Computer Review: 08944393221085705.

Turrell, A., et al. (2019). Transforming naturally occurring text data into economic statistics: The case of online job vacancy postings, National Bureau of Economic Research.

The Economist. (2020, July 23). Why real-time economic data need to be treated with caution. Retrieved from https://www.economist.com/finance-and-economics/2020/07/23/why-real-time-economic-data-need-to-be-treated-with-caution.