BIG DATA APPLICATIONS IN LABOR ECONOMICS, PART B

Edited by Benjamin Elsner
and Solomon W. Polachek

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

RESEARCH IN LABOR ECONOMICS

VOLUME 52B

BIG DATA APPLICATIONS IN LABOR ECONOMICS, PART B

RESEARCH IN LABOR ECONOMICS

Series Editor: Solomon W. Polachek and Konstantinos Tatsiramos

Recent Volumes:

Volume 37: Labor Market Issues in China
Edited by Corrado Giulietti, Konstantinos Tatsiramos and Klaus F. Zimmermann

Volume 38: New Analyses in Worker Well-Being
Edited by Solomon W. Polachek and Konstantinos Tatsiramos

Volume 39: Safety Nets and Benefit Dependence
Edited by Stéphane Carcillo, Herwig Immervoll, Stephen P. Jenkins, Sebastian Königs and Konstantinos Tatsiramos

Volume 40: Factors Affecting Worker Well-Being: The Impact of Change in the Labor Market
Edited by Solomon W. Polachek and Konstantinos Tatsiramos

Volume 41: Gender Convergence in the Labor Market
Edited by Solomon W. Polachek, Konstantinos Tatsiramos and Klaus F. Zimmermann

Volume 42: Gender and the Labor Market
Edited by Solomon W. Polachek, Konstantinos Tatsiramos and Klaus F. Zimmermann

Volume 43: Inequality: Causes and Consequences
Edited by Lorenzo Cappellari, Solomon W. Polachek, and Konstantinos Tatsiramos

Volume 44: Income Inequality Around the World
Edited by Lorenzo Cappellari, Solomon W. Polachek, and Konstantinos Tatsiramos

Volume 45: Skill Mismatch in Labor Markets
Edited by Solomon W. Polachek, Konstantinos Pouliakas, Giovanni Russo, and Konstantinos Tatsiramos

Volume 46: Transitions Through the Labor Market: Work, Occupation, Earnings and Retirement
Edited by Solomon W. Polachek and Konstantinos Tatsiramos

Volume 47: Health and Labor Markets
Edited by Solomon W. Polachek and Konstantinos Tatsiramos

Volume 48: Change at Home, in the Labor Market, and On the Job
Edited by Solomon W. Polachek and Konstantinos Tatsiramos

Volume 49: Workplace Productivity and Management Practices
Edited by Solomon W. Polachek and Konstantinos Tatsiramos

Volume 50: Fiftieth Celebratory Volume
Edited by Solomon W. Polachek and Konstantinos Tatsiramos

Volume 51: Time Use in Economics
Edited by Daniel S. Hamermesh and Solomon W. Polachek

Volume 52A: Big Data Applications in Labor Economics, Part A
Edited by Benjamin Elsner and Solomon W. Polachek

This page intentionally left blank

EDITORIAL ADVISORY BOARD

This page intentionally left blank

RESEARCH IN LABOR ECONOMICS VOLUME 52B

BIG DATA APPLICATIONS IN LABOR ECONOMICS, PART B

EDITED BY

BENJAMIN ELSNER
University College Dublin, Ireland and IZA, Germany

AND

SOLOMON W. POLACHEK
State University of New York at Binghamton, USA and IZA, Germany

United Kingdom – North America – Japan
India – Malaysia – China

Emerald Publishing Limited
Emerald Publishing, Floor 5, Northspring, 21-23 Wellington Street, Leeds LS1 4DL

First edition 2025

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-83608-713-7 (Print)
ISBN: 978-1-83608-712-0 (Online)
ISBN: 978-1-83608-714-4 (Epub)

ISSN: 0147-9121 (Series)

Certificate number 1985

Awarded in recognition of Emerald's production department's adherence to quality systems and processes when preparing scholarly journals for print

INVESTOR IN PEOPLE

CONTENTS

This page intentionally left blank

LIST OF CONTRIBUTORS

Alejandra Bellatin	PricewaterhouseCoopers, Canada
Brian Callaci	Open Markets Institute, USA
Sugat Chaturvedi	Ahmedabad University, India
Vivian Chu	Bank of Canada, Canada
Verónica Escudero	International Labour Organization, Switzerland; Center for Affective Global Action, USA
Gabriela Galassi	Bank of Canada, Canada
Andreas Kuhn	Swiss Federal University for Vocational Education and Training, Switzerland
Hannah Liepmann	International Labour Organization, Switzerland; IZA, Germany
Kanika Mahajan	Ashoka University, India
Peter Norlander	Loyola University Chicago, USA
Sérgio Pinto	University of Maryland at College Park, USA; Instituto Universitário de Lisboa (ISCTE-IUL), DINAMIA'CET, Portugal
Ana Podjanin	International Labour Organization, Switzerland
Zahra Siddique	University of Bristol, UK
Marshall Steinbaum	University of Utah, USA
Matt Walsh	Lightcast, USA
Lennart Ziegler	University of Vienna, Austria

This page intentionally left blank

PREFACE

Over the past century, economics has undergone a remarkable transformation in the way economists utilize data. Early 20th-century research was largely theoretical, often limited by the paucity of data and formidable computational constraints. Agricultural economists, then at the forefront of the field, pioneered econometrics by utilizing limited crop yield data to estimate production functions, and more aggregated consumer data to estimate consumer demand. The advent of computers beginning in the 1960s and 70s revolutionized data processing capabilities, enabling economists to utilize, and governments to provide larger, often more micro, datasets. More detailed data and better computers stimulated the development of econometrics as a discipline and facilitated the integration of statistical methods into economics research, thereby improving the precision of empirical analyses. More recently, the proliferation of big data and technological advances in data collection methods further revolutionized the landscape. The widespread availability of big data allowed economists to access diverse and detailed information, ranging from social media interactions to high-frequency job data to detailed job ads. Advanced data collection methods, such as web scraping, artificial intelligence, and sensor technologies, further expanded the scope of empirical research. This volume (Parts A and B) contains 13 new leading-edge research articles that utilize state-of-the-art methods of accessing and analyzing big data. Four articles use the data to address economy-wide issues. These encompass labor demand, unemployment, wage growth, and job vacancies. Five articles deal with measuring skills. These encompass current and changing work requirements as well as how job tasks mesh with personality traits. Two articles measure gender differences, and, finally, two articles use big data to uncover the characteristics of franchise agreements.

Perhaps no question is more important to an economy than accurately tallying unemployment rate statistics in a timely manner. Unemployment rate data are used by the federal government to determine optimal monetary policy, by financiers to determine their investments, and by the press to accurately report the state of the economy. In the United States, the Bureau of Labor Statistics (BLS) relies on monthly the Current Population Survey (CPS) to compute unemployment rates. Data are collected mid-month and reported about a month later. Most other countries follow a similar approach. However, the one-month lag between collecting and reporting the data can lead to imprecision, as much can change in a month. Thus, one important question is whether or not machine-readable data of some kind can be utilized to obtain accurate unemployment rates in real time. In the first article, Kailing Shen and Yanran Zhu compare the effectiveness of big-data-based *Job Posting Counts (JPCs)* obtained

from Lightcast (formerly Burning Glass) to the now used traditional Current Population Survey *unemployment rate* data in capturing labor market transitions in the United States. They seek to answer: whether big-data-based JPCs effectively capture labor market transitions, whether JPCs better correlate with broad labor market shifts than the official unemployment rate, whether JPCs outperform local labor market dynamics than the unemployment rate, whether policymakers could benefit from utilizing both JPCs and the unemployment rate when responding to labor market dynamics.

Unemployment rates are one metric to assess the state of a labor market, but this metric can be misleading if it does not consider part-time employment. Hyeri Choi and Ioana Marinescu use data from Burning Glass and the Current Population Survey spanning 18 years to investigate what factors contribute to involuntary part-time employment in the United States. According to their findings, local labor supply and demand factors play a significant role in the prevalence of involuntary part-time employment. An increase in the local unemployment rate leads to more involuntary part-time work, whereas an increase in job vacancies reduces it. Additionally, areas where firms have a higher degree of labor market power tend to have higher rates of involuntary part-time employment. Overall, these findings emphasize the importance of involuntary part-time employment as a form of labor market slack. During bad economic times, many workers are unfortunately forced to work less than they want, which adds to the already existing economic burden of high unemployment.

Another labor market metric is wage growth. Higher wage growth could be a sign of future inflation, and lower wage growth could signal an impending recession. Further, asymmetric wage growth, in which wages grow more quickly in one sector compared to another, could have implications for the earnings distribution. But wage growth or the reverse, wage stagnation, could be a result of an unprecedented shock to an economy overall. In the next article, Pawel Adrjan and Reamonn Lydon utilize data from millions of online Indeed Wage Tracker job postings to analyze monthly estimates of annual growth in advertised wages across several European countries and the United States from 2019 to 2022, a period spanning the pandemic. The study assesses how online job posting data compare to official sources on job vacancies, new hires, and wage levels. Further, it examines how wage growth has varied during this period, particularly exploring the impact of the COVID-19 pandemic. Among the questions addressed are: What factors contribute to the observed wage growth trends? Is there a correlation between wage growth and other economic indicators? What are the trends in wage growth for different income levels or job categories? And how does the study account for the heterogeneity in post-pandemic wage growth?

All labor markets involve a search process where firms look for workers and workers look for firms. An efficient labor market is one where this search process has minimal friction, and workers can easily match with firms. A common measure of labor market efficiency is the Beveridge curve, which depicts the relationship between job vacancies and unemployment. It shows that as job vacancies increase, unemployment rates decrease. The emergence of new technologies such as online job search can improve the efficiency of matching in the

labor market, making it easier for firms and workers to find one another. This greater efficiency shifts the Beveridge curve inward, resulting in a higher vacancy rate for a given unemployment rate. In the next article, Leonardo Fabio Morales, Carlos Ospino, and Nicole Amaral use data from Public Employment Services in Colombia to study how a change in the requirement to post jobs online affects the Beveridge curve. They perform a difference-in-difference analysis, comparing the efficiency in sectors that introduced this requirement to sectors that did not. Their results show that online job postings significantly increased labor market efficiency. In affected sectors, the introduction of the requirement led to a decrease in the vacancy rate and increases in the hiring and employment growth rates.

Job ad data can also be used on a micro level. The next five articles analyze the type of skills demanded, how the demand for these skills changed over time – particularly during the COVID-19 pandemic – how the demand for particular skills varies across countries, and how the demand for particular skills relates to employee personality. The next article by Vera Brenčič and Andrew McGee concentrates on the latter, whether personality traits are associated with specific task requirements. Based on over 100,000 US job advertisements posted on Monster.com from June 26, 2006, to July 8, 2006, the article finds that employers often emphasize personality traits in job advertisements. As such, it explores whether particular tasks require certain personality attributes. It finds employers seek specific traits (extroversion for communication tasks) and workers match based on both personality and tasks, even within occupations. Though tasks, not just occupations, drive this "sorting," financial rewards for such matches are unclear as is how personality traits explain overall career paths. Further, the article explores how a combination of personality traits and a set of within-occupation task variations might be crucial for fully understanding their respective roles in the job market.

In the next article, Luiza Antonie, Laura Gatto, Sarah Oloumi, and Miana Plesca similarly explore the relationship between personality traits and skill requirements in job postings. This interdisciplinary research involved computer scientists and economists, and they used a keyword extraction technique on 39,487 descriptions posted on a Canadian University Co-op and Career online job board from May 1, 2013, to May 1, 2020. The main focus of the study is on the "Big Five" personality traits, namely openness, conscientiousness, extroversion, agreeableness, and neuroticism, as these traits are believed to be vital in the modern work environment. The findings show that two-thirds of all jobs refer to at least one of the Big Five traits, although different occupations require different traits. Technical or business-oriented jobs typically require openness, accounting jobs require conscientiousness, and healthcare and childcare jobs require agreeableness and emotional stability. Although these results may not appear surprising, it is remarkable that these results come out so clearly even though the job ads rarely mention the traits explicitly.

The demand for specific skills can change over time due to a variety of factors that reflect shifts in the economy, technology, and societal needs. This is especially true for digitalization skills during unexpected events, such as the COVID-19 pandemic, which could have accelerated as remote work and online

services became necessary. In such circumstances, employers may seek individuals with skills related to remote collaboration tools, digital communication, and cybersecurity. In the next article, Gabriela Galassi, Alejandra Bellatin and Vivian Chu critique traditional labor market research for its slow data and analysis, highlighting the limitations it posed during the fast-paced COVID-19 pandemic. As such, they use text analytics to construct a new dataset from Canadian Indeed data. They study trends in digitalization throughout the pandemic to analyze how different lockdown measures impact the demand for digital production jobs. They then compare how the severity of lockdown measures compares to pre-pandemic levels in terms of digital job openings.

The process of measuring skill demand is relatively straightforward in North America and Europe. These regions have access to frequent surveys that cover different types of skills. Moreover, databases such as O*Net allow researchers to map skills to tasks and occupations. However, obtaining similar data in low and middle-income countries can be challenging, making it difficult to measure and document shifts in skill demand in the labor market. A common approach is to apply skill definitions from North America and Europe to occupations in other parts of the world. However, this approach can be misleading because the same job may require different skills in different countries. The next article by Verónica Escudero, Hannah Liepmann, and Ana Podjanin proposes an alternative approach. They use data from a job board in Uruguay and use natural language processing methods to develop a country-specific taxonomy of different skills. This method can be easily applied to measure skills in many countries around the world, given that comparable job board data are available in most countries. By doing so, researchers can obtain more accurate insights into the skill requirements of different jobs in different countries.

Although there are many estimates of skill demand in Europe and North America, it is often unclear how it is related to wages and the time it takes to fill a vacancy. If companies require a complex set of skills, does this mean they have to offer higher wages and experience more difficulties in finding a suitable candidate? The next article by Lennart Ziegler seeks to answer these questions with data from 1.5 million job postings by the Austrian public employment service. He uses machine learning to extract the most common skill requirements mentioned in the postings. Even among jobs with similar characteristics, the number of skill requirements is positively associated with higher wages. This effect is mainly driven by managerial and analytical skills, whereas soft skills play a minor role. Vacancies with many skill requirements also take considerably longer to fill. These findings suggest that firms face a trade-off in their search for workers. More skill requirements may attract workers that better match the skill profile of a job, but this comes at the cost of a higher wage and greater difficulty in finding such a worker.

Some changes in skill demand over time are related to gender. It is well known that fertility rates have been declining, the age of first marriage has increased, and the husband–wife age gap has narrowed over the past 60 years. Concomitant with these demographic changes has been an increase in female education and female lifetime labor force participation, both coupled with a slight decline in male labor

force participation. In short, societal norms have been changing. But not yet studied is how firms have responded, especially concerning job postings. In the next article, Andreas Kuhn demonstrates how a combination of diverse data sources sheds light on the transformation of gender norms over time. Examining Swiss job advertisements from 1950 to 2020, his study reveals a significant shift in employers' preferences, with the proportion of gender-neutral job posts rising from 5% to nearly 95%. To reinforce and contextualize this finding, the analysis incorporates time series data from Google's German book corpus, focusing on specific queries related to gender equality. The additional data align with the evolution of gender-neutral job posts, indicating two distinct narratives – one centered on personal relationships and identity and the other on the political and public realm. Notably, the narrative on personal relations precedes the change in gender-neutral job ads. The article quantifies the substantive changes in gender norms over recent decades and characterizes its complexity.

In the next article, Sugat Chaturvedi, Kanika Mahajan, and Zahra Siddique investigate gender differences in skill demand by analyzing job descriptions from a large volume of ads posted on a job portal in India. They utilize domain-specific unlabeled data to create word vector representations (word embeddings). They examine how various required skill categories correlate with log posted wages and explore how skill demand varies with firm size. Importantly, they relate these skill demands to gender. Among the questions they address are how firms' skill demand varies by gender and how the associated skill categories correlate with posted wages. Their conclusions have strong implications regarding the gender pay gap within the economy, and at large compared to small firms.

The final two articles make use of big data to investigate how anticompetitive practices in the franchise sector affect workers. In this sector, it is common for firms to enter into no-poach agreements with other firms and noncompete agreements with their workers. Both of these agreements weaken the position of workers. If firms agree to not "poach" workers from competitors, this gives firms a higher monopsony power, as workers have fewer outside options. The same is true for noncompete agreements. If workers are unable to move to a direct competitor after ending their employment, this severely limits their outside options. A challenge for researchers and regulators is that such agreements are not public. The article by Peter Norlander presents a machine-learning-based method to identify no-poach and noncompete agreements in the franchise industry. He analyzes text data from Franchise Disclosure Documents in the United States and applies supervised machine learning to uncover evidence of such agreements in an unstructured text corpus. The study reveals that no-poach and noncompete agreements are very common in the United States. However, as regulation became stricter around 2017, no-poach agreements became less common, whereas noncompete agreements remained prevalent.

The article by Brian Callaci, Sérgio Pinto, Marshall Steinbaum, and Matt Walsh provides additional evidence on agreements in the franchise industry and their effects on workers. The researchers combine text data from Franchise Disclosure Documents with job advertisements from Burning Glass Technologies to examine a wide range of agreements known as *vertical restraints*. These

agreements restrict the decisions of workers and firms regarding whom to hire, who to do business with, what products to sell, or in what area to operate. The analysis reveals that vertical restraints are highly prevalent in the franchise sector, with franchisors exerting significant control over how franchisees operate their businesses. The authors use wage regressions to provide suggestive evidence of the negative impact of vertical restraints on wages. Restrictions on how firms operate their businesses, who they hire, and the use of no-poach and noncompete agreements are associated with lower wages.

The contributions to this volume resulted from an open Call for Abstracts that elicited 28 responses, out of which we solicited complete articles from authors of 13 submissions. All were then refereed, each by two scholars whose identity was held anonymous from the contributors. For insightful editorial advice, we thank Silke Anger, Nicolas Apfel, Patrick Arni, Philippe Askenazy, Xue Bai, Maria Balgova, Audinga Baltrunaite. Martin Biewen, Ian Burn, Patrick Button, Fabrizio Colella, Bart Leo Wim Cockx. Jonathan Cribb, Ron Diris, Miriam Gensowski, Giuseppe Grasso, Brad Hershbein, Bassier, Ihsaan, Hannah Illing, Melanie Jones, Gaurav Kankahalli, Michael Koch, Thomas Krumel, Etienne Lalé, Caecilia Lipowski, Michael Lipsitz, Jonas Maibom, Arianna Marchetti, Jaime Montana, Paul Muller, Lucas Navarro, Francisco Parro, Tho Pham, Kevin Rinz, Rodimiro Rodrigo, Rafael Sanchez, Hannes Schwandt, Ocampo, Sergio, Wei Si, Daphne Skandalis, Yang Song, Todd Sorensen, Brad Speigner, Anthony Stittmatter, Elisa Taveras-Pena, Simon Trenkle, C. Y. Kelvin Yuen, Brandon Vick, Catherine Weinberger, Marc Witte, Liqiu Zhao, and Christian Zimpelmann. This volume would not have been possible without their thorough and expeditious help.

Benjamin Elsner
Solomon W. Polachek
Volume Editors

LETTING JOB POSTINGS TALK: RECENT TRENDS IN DIGITALIZATION

Gabriela Galassi[a], Alejandra Bellatin[b] and Vivian Chu[a]

[a]*Bank of Canada, Canada*
[b]*PricewaterhouseCoopers, Canada*

ABSTRACT

We construct a novel dataset of Canadian online job postings, classified by occupation. The data, provided by Indeed, an online job board, represents vacancies advertised by employers across Canada. We have classified these job postings into standard occupations using text analytics. This dataset has been used to study changes in the demand for jobs linked to digitalization over the COVID-19 pandemic. To this end, we leverage time-series and cross-sectional variations in COVID-19 containment policies, examining their impact on jobs broadly related to digitalization. Our findings reveal that vacancies in digital production jobs increased more substantially than in traditional jobs during the reopening phases. However, no substantial differences were observed when considering different types of vacancies according to the use of digital technologies (i.e., occupations at low risk of automation or those that allow remote work). Overall, our results do not support the popular idea that the COVID-19 pandemic marked a significant turning point in digitalization trends, but rather document a modest shift in this direction.

Keywords: Job postings; labor demand; digitalization; text analytics; COVID-19

1. INTRODUCTION

Historically, the empirical labor market research has focused on data derived from labor force surveys, population censuses, and various consumer and firm surveys. The data are typically low-frequency, subject to publication lags, and

Big Data Applications in Labor Economics, Part B
Research in Labor Economics, Volume 52B, 1–33

ISSN: 0147-9121/doi:10.1108/S0147-91212024000052B022

often relatively aggregated. Such characteristics hinder the analysis of labor market developments in real time and the identification of trends within disaggregated market segments. As was particularly evident during the COVID-19 pandemic, these limitations represent substantial challenges for both policymakers and academics.

This article leverages comprehensive data on job postings from a leading job board, indeed, to offer insights into the Canadian labor market during the COVID-19 pandemic with a particular focus on digitalization. Job posting data have recently shown to be useful in a variety of contexts. For example, Hensvik et al. (2021), Marinescu et al. (2021), and Bernstein et al. (2023) analyze the evolution of job vacancies, search behavior, and the resulting labor market tightness at different stages of the pandemic using several job boards. Forsythe et al. (2020) documented a substantial drop in job postings in the US at the onset of the pandemic using data from Burning Glass Technologies. Jones et al. (2023) use Burning Glass Technologies data for Canada to complement traditional data sources and to understand the magnitude of the flows in the labor market at the onset of the pandemic.

Our main contribution lies in constructing a new dataset of online job postings, classified by standard occupations using text analytics. We enhanced an algorithm originally developed by Turrell et al. (2022), adapting it to the data provided by Indeed and to the bilingual setting of the Canadian labor market. This adaptation has allowed us to achieve an acceptable level of accuracy at the 4-digit level of the Canadian National Occupation Classification (NOC), the most granular level of our analysis.

We then demonstrate the economic usefulness of our new dataset by employing it to examine trends in digitalization during the pandemic. To achieve this, we group job titles based on their role in digital production (such as software development, hardware production, information technology support), and in digital adoption (namely, occupations that offer the possibility for remote work, or have low risk of automation).

In our application to the COVID-19 context, we leverage the variation in the containment measures across Canadian provinces to estimate the effects of the pandemic on digitalization. To mitigate biases arising from the correlation between the disease spread and other local shocks, we adopt a differences-in-differences approach and an event-study approach. These methods allow us to exploit variations in the timing of lockdowns and reopenings across provinces.

Our findings indicate that less stringent containment measures lead to stronger recovery in openings for jobs involved in the production of digital technology than for other jobs. Postings for jobs that can be done remotely and that are not at high risk of automation, also increase slightly more during reopenings; however, the difference compared to other types of job postings is not significant. Similarly, no significant differences are observed during lockdowns. These findings complement and contrast with those by Soh et al. (2024) with the same data for the US; whereas they document changes favoring digital workers relative to non-digital ones, these changes are driven by a relatively small decline in the demand for digital workers compared to others. In contrast, we find that demand

for digital production occupations increases more than for other occupations during reopenings. We also offer suggestive evidence that, during the pandemic, firms posted fewer low-wage and female-oriented vacancies in occupations not linked to the digital economy, as opposed to those in the digital economy where this phenomenon was not observed.

Our article contributes to a growing body of literature using online job postings data to analyze labor market outcomes. While the use of these data has become particularly prolific during the pandemic, there are notable precedents. Turrell et al. (2022) used the online vacancies from the job site Reed to examine labor mismatch in relation to the productivity puzzle in the UK. They found that regional mismatches played a more prominent role than occupational mismatches in explaining productivity statistics. Using data from CareerBuilder, Marinescu and Wolthoff (2020) document that job titles explain over 90% of the wage variance. The Indeed data featured in our study has also been previously used in labor market research. For instance, Gimbel and Sinclair (2020) analyzed mismatches between job seekers and employers in the US, and Adrjan and Lydon (2019) showed that labor market tightness, as measured by job postings and clicks, correlates with posted wages. Our article combines text analytics with occupation descriptors to systematically structure job posting data into occupations within the Canadian context.

The application to digitalization during the COVID-19 pandemic aligns with the literature of acceleration of technological change with recessions. Hershbein and Kahn (2018) showed that firms in areas severely impacted by the 2008 crisis persistently increase both their skill requirements and capital investments.[1] Consistent with these findings, Jaimovich and Siu (2020) report that job losses in routine occupations are concentrated in recessions, without corresponding employment gains in these occupations during recoveries. Foote and Ryan (2014) also observe that middle-skill workers, predominantly in routine occupations, are concentrated in cycle-sensitive industries (like manufacturing and construction), leading to cyclical fluctuations in their employment levels. The COVID-19 crisis not only reduced the opportunity costs of technological change typically associated with economic crises but also made disease-control measures more conducive to the adoption of digital technologies. Our article focuses on job postings related to the production of digital technologies, extending beyond the concept of automation risk that has been predominant in this literature.[2]

Although there has been much discussion about the acceleration of technology adoption during the pandemic, the evidence is still limited. Alexopoulos and Lyons (2021) analyzed various unstructured data sources to assess trends in the adoption of digital technologies in Canada, both before and after the pandemic. These technologies include artificial intelligence, data science, and robotics. While some indicators suggest technological sectors have been outpacing others during the pandemic, others point to a slowdown in technology adoption during the recessionary periods. Barrero et al. (2021) document a substantial increase in new US patent applications related to remote work technologies since the onset of the pandemic. Our results complement these studies. While we do not find conclusive evidence of accelerated technology adoption during the pandemic, our analysis reveals that firms notably increased their demand for jobs in digital production as restrictions were eased.

The remainder of the article is organized as follows. Section 2 presents the data on job postings and discusses its usefulness for labor market research. Section 3 explains the algorithm we built to classify the data into occupations. We then turn to the application to digitalization during the pandemic. Section 4.1 presents how we group the data using the classifications relevant to analyze digitalization, and Section 4.2 shows the recent trends in these groups. We present the event study that leverages province-level variation in lockdowns and reopenings in Section 4.3. Section 5 concludes.

2. ONLINE JOB POSTINGS DATA

We use job postings collected by Indeed, the largest job site in Canada.[3] Indeed advertises job postings by employers directly on its website, as well as postings collected from employers' websites, which are treated to avoid duplication. The data include the job title, the first and last day visible, and the city and province, and they are available from 2018 on.

Fig. 1 shows the annual growth rate of the smoothed number of new job postings (7 day moving average) published each day in Canada. Growth rates for 2021 and 2022 refer to 2019 to avoid base-year effects caused by large drops at the beginning of the pandemic. The volume of online job postings in Indeed closely follows the trend in online job board vacancies from the Job Vacancy and Wage Survey (JVWS, dots in Fig. 1), and with total employment from the Labour Force Survey (LFS) (Fig. 2). These are both collected by Statistics Canada, which is the official statistics office. Reassuringly, the data on job postings show a distribution across provinces that is similar to that of employment in the LFS in Canada (Fig. 3).

Overall, the data on online job postings provided by Indeed appear representative and hence useful for labor market research on online vacancies. The data are unstructured, containing over four million unique job titles. Fig. 4 shows a word cloud of the text in the job postings. We next explain how we classify the data into occupations in the Canadian context.

3. CLASSIFYING JOB POSTINGS INTO OCCUPATIONS

We construct a text analytics algorithm to classify job titles into relevant occupation classifications. The Indeed data lack occupational variables that can be directly mapped into the standard occupation categories. Instead, we utilize the job title and company name in the job postings to classify them into Canada's four-digit NOC, version 2016.3. This version was the most recent and offered the highest level of disaggregation of the NOCs at the time of our analysis.

We build upon an algorithm developed by Turrell et al. (2022). Unlike the approach of Turrell et al. (2022), our job ads data do not include job descriptions; instead, we only have access to the job title and company name for each posting. Our adaptation of the algorithm performs adequately, achieving accuracy at the

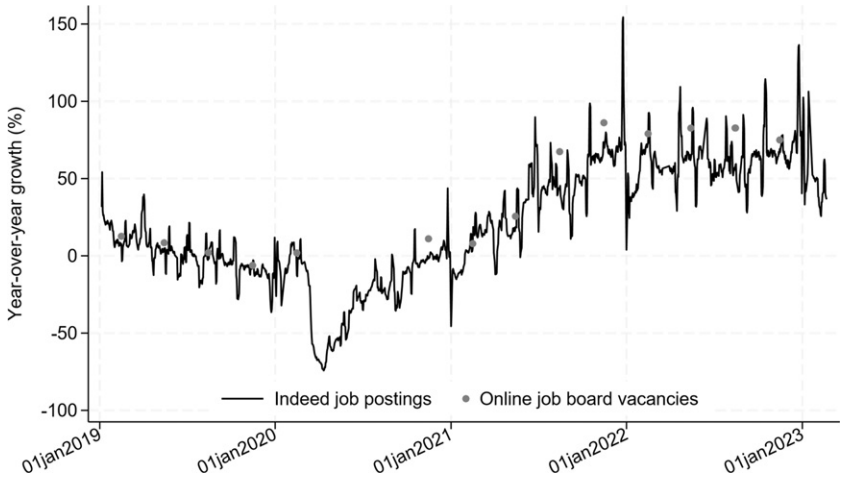

Fig. 1. Year-Over-Year Growth of Online Job Postings on Indeed and Online Job Board Vacancies Collected by Statistics Canada.
Note: Job postings are daily data, smoothed with a 7-day moving average. Vacancy data are collected on a quarterly basis. To avoid base-year effects associated with the COVID-19 pandemic, the year-over-year growth rate calculations use 2019 values for 2021 and 2022.
Source: Indeed job postings, Job Vacancy and Wage Survey (Statistics Canada), and own calculations.

high end of comparable classifications. The main modifications we made to the algorithm by Turrell et al. (2022) include (i) mapping jobs to the Canadian NOC, for which we create dictionaries in English and in French, (ii) expanding the list of abbreviations, and (iii) using company names, which we believe offer some insight about the sector of each vacancy.

We first clean the text in job postings following standard text analytic techniques.[4] Abbreviations in the job postings are expanded using an adaptation of Turrell et al. (2022)'s dictionary that adds abbreviations from human resources websites.

To perform the matches, we compile two types of dictionaries by scraping text data from the Government of Canada website: the job title dictionary and the broader text dictionary.[5] The job title dictionary contains sample job titles for each of the 500 NOC titles. The broader text dictionary has information on descriptions and main tasks for each job title. Dictionaries are constructed in English, French, and bilingual (appending the previous two). Due to efficiency considerations, job postings in English are classified using the English-only dictionary. Job postings either from Quebec (Canada's main francophone province) or those using francophone special characters, such as accents or cedilla, are classified using the bilingual dictionary. This is because some job postings in Quebec are in English.

There are two stages in the matching algorithm. First, it looks for an exact match between the title in the job posting and the title from the NOC. If found,

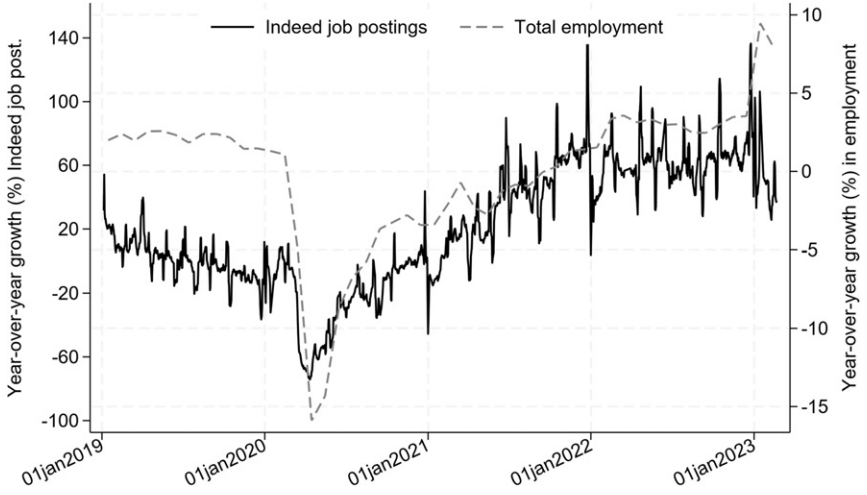

Fig. 2. Year-Over-Year Growth in Online Job Postings on Indeed and Total Employment.

Note: Job postings are daily data, smoothed with a 7-day moving average. Employment data are collected monthly. To avoid base-year effects associated with the COVID-19 pandemic, the year-over-year growth rate calculations use 2019 values for 2021 and 2022.

Source: Indeed job postings, Labour Force Survey (Statistics Canada), and own calculations.

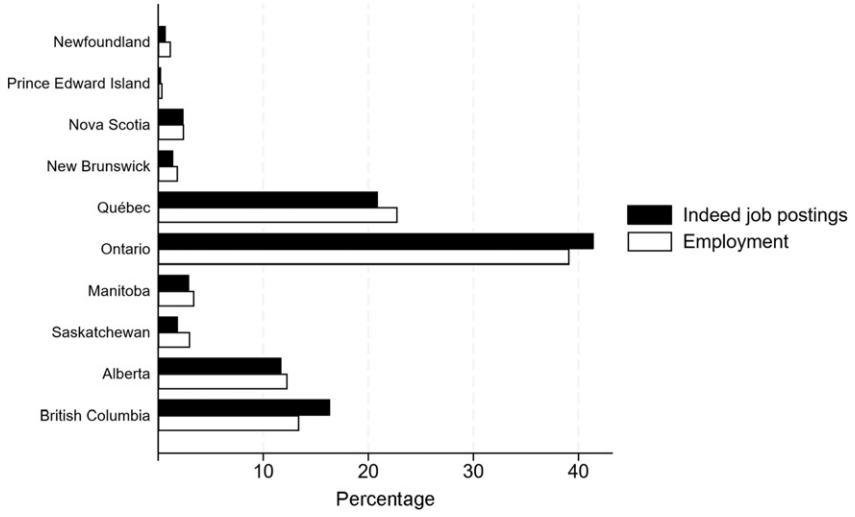

Fig. 3. Percentage of Online Job Postings and Employment by Canadian Provinces in 2019.

Note: Job postings data are collected daily. Employment data are collected monthly.

Source: Indeed job postings, Labour Force Survey (Statistics Canada), and own calculations.

Fig. 4. Word Cloud of Online Job Postings Since March 2020.
Source: Indeed job postings and own calculations.

the relevant four-digit NOC is returned. If no exact match is found, the algorithm proceeds to a second stage, the so-called fuzzy match, comparing both the job title and the company name to the broader NOC dictionaries. Our algorithm uses the term frequency-inverse document frequency (*tf-idf*) technique. It calculates the cosine similarity between the job-posting text and the NOC category. It then returns the job posting that has the smallest cosine distance, ensuring that no job postings remain unmatched. More details are in Appendix A.

Accuracy of the classification algorithm: We evaluated the performance of the algorithm by manually verifying the classification produced by the algorithm in 100 random job titles. Accuracy according to this procedure (percentage of correct matches) is 70% for job postings in English. By incorporating the company name into the algorithm, we gain a three-percentage-point increase in accuracy compared to using only the job title. For our French sample with the

bilingual dictionary, the accuracy is 66%. And when we manually delete the English job postings in our French sample and use only the French dictionary, we obtain an accuracy of 74.5%.

The accuracy values we obtain are adequate for a 4-digit automatic classification according to Turrell et al. (2022), particularly when job ad text is missing, as in our case.[6] Also, using the NOC for classifying job postings in Canada seems to be a good choice. In a classification exercise using Turrell et al. (2022)'s algorithm to classify our job postings into the 3-digit UK Standard Occupation Classification, as they do in their article, we achieved a decreased accuracy level (64%) in our English sample.

Fig. 5 provides reassuring insights into the distribution of Indeed job postings across broad occupational groups (NOC 1 digit) in comparison to employment levels in these groups, as indicated by the LFS data. The distributions look fairly similar. Certain groups, like managers, are overrepresented in the Indeed job postings, while others, such as operators, appear to be underrepresented. This seems in line with the presence of a larger proportion of high-skilled vacancies in online job postings than in vacancies in general.

4. APPLICATION: DID DIGITALIZATION ACCELERATE DURING THE COVID-19 PANDEMIC?

The first confirmed case of COVID-19 in Canada was reported on January 27, 2020. On March 11, 2020, the World Health Organization declared COVID-19 a global pandemic, leading to the implementation of travel bans and local

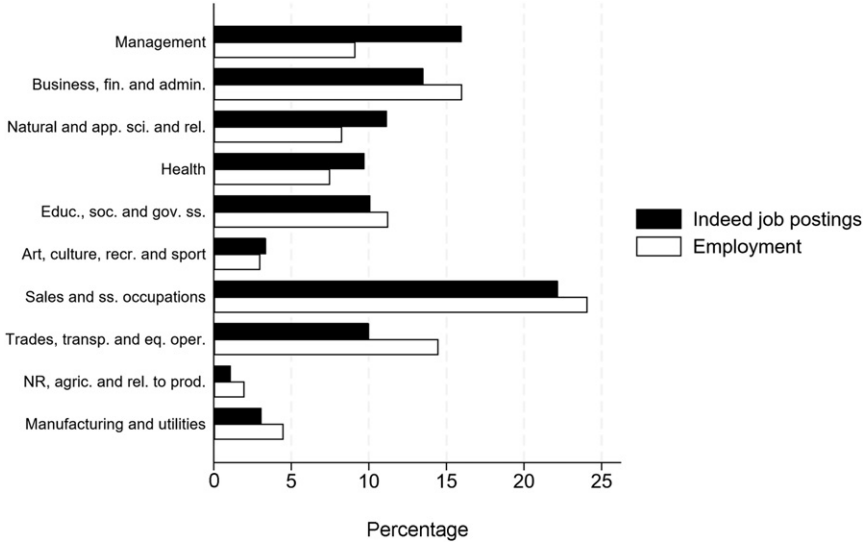

Fig. 5. Percentage of Job Postings and Employment by Occupation Groups in Canada in 2019.
Source: Indeed job postings, Labour Force Survey (Statistics Canada) and own calculations.