BIG DATA APPLICATIONS IN LABOR ECONOMICS, PART A

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 52B: Big Data Applications in Labor Economics, Part B
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 52A

BIG DATA APPLICATIONS IN LABOR ECONOMICS, PART A

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

emerald PUBLISHING

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: 978-1-83549-975-7 (Print)
ISBN: 978-1-83549-974-0 (Online)
ISBN: 978-1-83549-976-4 (Epub)

ISSN: 0147-9121 (Series)

ISOQAR QUALITY ASSURED ISO 9001
UKAS QUALITY MANAGEMENT 026
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

Pawel Adrjan	Indeed Hiring Lab, Spain and University of Oxford, UK
Nicole Amaral	Coursera, USA
Luiza Antonie	University of Guelph, Canada
Vera Brenčič	University of Alberta, Canada
Hyeri Choi	University of Illinois Urbana-Champaign, USA
Laura Gatto	University of Guelph, Canada
Reamonn Lydon	Central Bank of Ireland, Ireland
Ioana Marinescu	University of Pennsylvania, USA
Andrew McGee	University of Alberta, Canada
Leonardo Fabio Morales	Banco de la República, Colombia
Sarah Oloumi	University of Guelph, Canada
Carlos Ospino	World Bank, USA
Miana Plesca	University of Guelph, Canada
Kailing Shen	The Australian National University, Australia
Yanran Zhu	The Australian National University, Australia

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 1970s 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 (AI), 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 CPS 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 CPS 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 inwards 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 (ML) 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 aligns 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 anti-competitive 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 ML-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 ML 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