

# BIG DATA APPLICATIONS IN LABOR ECONOMICS, PART A

**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 52A**

# 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

Orley C. Ashenfelter  
*Princeton University*

Francine D. Blau  
*Cornell University*

Alison L. Booth  
*Australian National University*

Richard Blundell  
*University College London*

David Card  
*University of California*

Ronald G. Ehrenberg  
*Cornell University*

Richard Freeman  
*Harvard University*

Reuben Gronau  
*Bank of Israel*

Daniel S. Hamermesh  
*University of Texas at Austin*

James J. Heckman  
*University of Chicago*

Christopher A. Pissarides  
*London School of Economics*

Yoram Weiss  
*Tel-Aviv University*

Klaus Zimmermann  
*UNU-MERIT, Maastricht  
University*

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*



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

Editorial matter and selection © 2025 Benjamin Elsner and Solomon W. Polachek.  
Published under exclusive licence by Emerald Publishing Limited.  
Individual chapters © 2025 by Emerald Publishing Limited.

**Reprints and permissions service**

Contact: [www.copyright.com](http://www.copyright.com)

No part of this book may be reproduced, stored in a retrieval system, transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without either the prior written permission of the publisher or a licence permitting restricted copying issued in the UK by The Copyright Licensing Agency and in the USA by The Copyright Clearance Center. Any opinions expressed in the chapters are those of the authors. Whilst Emerald makes every effort to ensure the quality and accuracy of its content, Emerald makes no representation implied or otherwise, as to the chapters' suitability and application and disclaims any warranties, express or implied, to their use.

**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)



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

<i>List of Contributors</i>	<i>xi</i>
<i>Preface</i>	<i>xiii</i>
<b>Labor Force Transition Dynamics: Unemployment Rate or Job Posting Counts?</b>	<b>1</b>
<i>Kailing Shen and Yanran Zhu</i>	
<b>The Labor Demand Side of Involuntary Part-Time Employment</b>	<b>35</b>
<i>Hyeri Choi and Ioana Marinescu</i>	
<b>What Do Wages in Online Job Postings Tell Us About Wage Growth?</b>	<b>71</b>
<i>Pawel Adrjan and Reamonn Lydon</i>	
<b>Online Vacancies and Their Role in Labor Market Performance</b>	<b>119</b>
<i>Leonardo Fabio Morales, Carlos Ospino and Nicole Amaral</i>	
<b>Demand for Personality Traits, Tasks, and Sorting</b>	<b>161</b>
<i>Vera Brenčič and Andrew McGee</i>	
<b>Personality Characteristics Demanded by Employers: Analysis of Job Descriptions From University Job Boards</b>	<b>213</b>
<i>Luiza Antonie, Laura Gatto, Sarah Oloumi and Miana Plesca</i>	

This page intentionally left blank

# LIST OF CONTRIBUTORS

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

# LABOR FORCE TRANSITION DYNAMICS: UNEMPLOYMENT RATE OR JOB POSTING COUNTS?

Kailing Shen and Yanran Zhu

*The Australian National University, Australia*

## ABSTRACT

*Job posting counts (JPCs) are emerging as indicators of employment dynamics, yet their validity requires assessment. This study evaluates the effectiveness of big-data-based JPCs compared to the traditional survey-based unemployment rate in capturing labor market transitions in the United States. Using the Current Population Survey, our comparison focuses on their ability to predict individuals' transitions between employment and unemployment. We explore not only monthly national JPCs but also four additional versions that measure labor demand at various granularities. Our findings suggest comparable predictive power for JPCs and the unemployment rate, with each capturing different aspects of the variation in these transitions. Coefficients for both metrics remain statistically significant when considered together. Notably, the unemployment rate's correlation with transitions changes signs when adding year fixed effects, a phenomenon not observed for JPCs. Among more granular levels of JPCs, the most refined measure – those by state, occupation, and industry – exhibits the strongest predictive capabilities. Furthermore, our main results remain robust when applying alternative JPC measures (total unique and total JPCs).*

**Keywords:** Job posting counts; unemployment rate; labor market dynamics; granularity; predictive power

## 1. INTRODUCTION

The unemployment rate has long been the cornerstone of labor market analysis. However, its limitations are becoming increasingly apparent. Its reliance on surveys can lead to delays and limited granularity. This calls for exploring alternative metrics. Job posting counts (JPCs) have emerged as a promising candidate, providing real-time data on labor demand across various sectors and regions. This study investigates the effectiveness of JPCs in capturing labor market transitions, comparing them directly to the traditional unemployment rate.

Building upon a rich body of research that has explored the time-series characteristics of the unemployment rate and its effect on labor market transitions during business cycles and economic shocks (Bentolila & Bertola, 1990; Caballero & Hammour, 1994; Davis & Haltiwanger, 2001; Howe, 1990), this study aims to fill a critical gap in similar studies examining JPCs. Instead of drawing definitive causal inferences, this study focuses on exploring how JPCs can complement the unemployment rate in predicting labor market dynamics. This aligns closely with the original purposes of these metrics and contributes to the growing interest in integrating diverse data sources (Lohr & Raghunathan, 2017) and big data-driven approaches like nowcasting (e.g., Moriwaki, 2020). By leveraging JPCs alongside the unemployment rate, we move beyond considering it as the sole “correct” measure, instead treating both metrics as equally valuable. In doing so, this research has the potential to significantly improve our ability to predict and respond to changes in the labor market.

Our analysis uses the United States labor market data from 2010 to 2019, drawing upon the Current Population Survey (CPS) as the source of ground truth employment-unemployment transitions and Lightcast (formerly Burning Glass) data for job postings. Given the complexities described by Rivera Drew et al. (2014) preceding our study period, our data management process is straightforward. We employ the reweighting-estimation-transformation (RWET) approach (Shen & Taska, 2020) to adjust JPCs. This allows us to construct four adjusted JPCs, capturing labor demand at the following local levels: state, state-occupation, state-industry, and state-occupation-industry. For comparison, we utilize the seasonally adjusted monthly unemployment rate for individuals aged 25 to 54.

Our empirical evidence reveals consistent trends over time between the unemployment rate, JPCs, and the probabilities of transitioning between employment and unemployment. Labor market transitions exhibit significant variations across demographic groups, with males, younger individuals, and those with lower education experiencing greater fluctuations in employment status. During our ten-year study period, which also witnessed a general labor market improvement, the number of individuals transitioning from unemployment to employment marginally exceeded the number transitioning from employment to unemployment. This is primarily driven by the high number of individuals who became employed after being unemployed due to voluntary quits or (re-)entry.

Our main empirical analysis uses new unique JPCs, which are constructed from un-deduplicated Lightcast job posting data, using a multi-step data management process. As discussed in detail later, the undeduplicated data allows the

same job posting to be listed multiple times. New unique JPCs better capture the flow of unique job opportunities.

Specifically, our results reveal distinct relationships between new unique JPCs and the unemployment rate with labor market transitions. While the unemployment rate has short-term negative and long-term positive correlations with the chances of transitioning from employment to unemployment, new unique JPCs demonstrate stable negative correlations regardless of time frame or controls. Conversely, while the unemployment rate has short-term null and long-term negative correlations with the chances of transitioning from unemployment to employment, new unique JPCs demonstrate stable positive correlations regardless of time frame or controls. These contrasting patterns highlight that the two types of metrics, although comparable in their predictive power, capture different aspects of the variations in labor market transitions.

Furthermore, among more granular levels of JPCs, the most refined measure – JPCs by state, occupation, and industry – exhibits the strongest predictive capabilities, outperforming other JPC measures and the survey-based unemployment rate. Such stronger correlations of granular JPCs support the RWET approach and underscore the local nature of labor market dynamics. Notably, regression results using both metrics, alternative unemployment rate versions (seasonal and unadjusted), and subsamples divided by the types of unemployment or education, gender and age groups, other JPC measures (total unique and total) further confirm the robustness of these findings.

In summary, this study validates the big-data-based JPCs in their capacity to capture labor market transitions. Our study also finds that JPCs can be a more robust metric than the unemployment rate in their correlation with overall labor market transitions. Our analysis further reveals that JPCs, particularly those adjusted for local labor market dynamics, outperform the unemployment rate in predicting transitions, suggesting their potential for informing effective policy decisions. Specifically, policymakers can leverage granular JPCs to gain valuable insights into specific labor market trends at the local level. This information can be used to tailor interventions and policies to address specific needs, such as skills development programs or targeted job placement initiatives. Ultimately, utilizing both JPCs and the unemployment rate can lead to a more efficient and responsive labor market that better serves the needs of diverse workers and businesses.

However, it is important to acknowledge that job postings are user-generated and thus contain a certain level of noise and potential errors. The validity of JPCs therefore relies heavily on the aggregator's processing methods, which are not always fully transparent. Despite these limitations, our study demonstrates the potential of JPCs to complement traditional measures and improve our understanding of labor market dynamics.

The remainder of this article is organized as follows: Section 2 discusses the three sets of data used in the study; Section 3 presents the descriptive evidence; Section 4 presents the main regression findings; Section 5 examines the robustness of the results using subsamples and alternative measures of JPCs. Section 6 concludes the study.

## 2. DATA

Three data sources from the United States are used in this study: the monthly microdata from the CPS, the monthly unemployment rates from the Bureau of Labor Statistics, and the job posting data supplied by Lightcast.

### 2.1 CPS Individual Transition Sample

Specifically, we construct labor market transitions using the monthly CPS microdata. We identify individuals across consecutive months using combinations of household identification and personal identification. Given that our job posting data are sequentially available starting from 2010 and given that the COVID-19 pandemic dramatically reshaped the labor market in 2020, our study period spans the 10 years in between, that is, from 2010 to 2019. The raw CPS data in this period have 17,911,401 observations covering 3,498,678 individuals in 1,400,452 households.

We keep individuals in interviewed households, excluding those associated with the armed forces or living in group quarters. Our sample is further limited to the reference person, partner of the reference person, or child of the reference person of the primary families or individuals within households. Our transition sample is constructed from the monthly dataset, where each observation represents an individual's labor market status transition record across two consecutive calendar months, referred to as month  $t$  and  $t + 1$ . In the design of CPS, respondents are surveyed for four months, rotated out for eight months, and then surveyed for a second four-month period. Therefore, one individual could correspond to multiple observations.<sup>1</sup>

To maintain data consistency, we exclude the very small fraction of transition samples where individuals experienced an age drop or age change greater than 1, underwent a gender or race change, or displayed an implausible change in their education level. We kept transition records only if the individual was part of the labor force, that is, either employed ( $E$ ) or unemployed ( $U$ ) in both month  $t$  and  $t + 1$ . All 623,042 individuals in the final sample were aged 25 to 54, had educational attainment lower than the doctoral level, and were neither part of the armed forces nor self-employed.

In total, we recorded 1,610,285 transitions between employment and unemployment, of which 1,540,887 (or 95.7%) began from an employment status ( $E_t E_{t+1}$  or  $E_t U_{t+1}$ ), and 69,398 (4.3%) began from an unemployment status ( $U_t E_{t+1}$ , or  $U_t U_{t+1}$ ).<sup>2</sup> In general, most individuals in the labor force are employed.

### 2.2 Unemployment Rate

We examined both seasonally adjusted and unadjusted unemployment rates for individuals aged 25–54 years. Although the results slightly vary, the seasonally adjusted unemployment rate consistently displayed a stronger correlation than the unadjusted rates with the labor market transitions we investigated. Therefore, to subject the JPCs to a more rigorous test, we present and discuss results based solely on the seasonally adjusted unemployment rates.

### 2.3 Lightcast Job Posting Counts

Our JPCs are constructed using the unduplicated version of Lightcast’s weekly job posting data for the United States. Lightcast is a major supplier of job posting data sets in the United States. Their data has been extensively used in previous labor market studies (e.g., [Acemoglu et al., 2022](#); [Braxton & Taska, 2023](#); [Deming & Noray, 2020](#); [Hershbein & Kahn, 2018](#)). In the unduplicated Lightcast job posting data, the same job posting could be listed multiple times if it appeared through multiple channels (e.g., company website, job search portals). This duplication could occur across different calendar dates. This “multi-homing” phenomenon is discussed in detail by [Krumel et al. \(2023\)](#) as well as [Evans et al. \(2023\)](#). Specifically, [Krumel et al. \(2023\)](#) show the “collapse” of labor demand found in [Forsythe et al. \(2020\)](#) might be an artifact of using the new job posting counts rather than total job posting counts.

In light of this debate, we constructed three versions of job posting counts for each month based on the duplicate flag provided: total JPC, unique JPC, and new unique JPC.<sup>3</sup> Recognizing that most researchers lack access to the unduplicated version and typically use the deduplicated version, we *therefore selected new unique JPCs as our preferred measure*, which offers a more accurate representation of the true number of job postings. This choice aligns with both our intuitive preference and the empirical evidence presented in the appendix, which suggests the validity of our findings using either the total or unique JPC.

Using the new unique JPCs, we further ensure alignment between the job posting datasets and those of the CPS to facilitate comparison exercises. Specifically, we match four variables across these two sets of data: calendar month, state, industry, and occupation. To align with the CPS transition sample, we project all job postings dated in calendar dates into their corresponding calendar months. We also limit the dataset to the job postings of the 50 US states and the District of Columbia, excluding territories, as well as postings related to the armed forces industry (CsdInferredNAICS = 928,110), postings related to armed forces occupations (CsdONET = 55), or postings with a missing state.

The first two digits of both the industry and occupation codes are used to match job postings with individuals in the CPS transition sample. The industry categories used in the CPS and the job posting data are largely similar, with several exceptions. Specifically, durable and nondurable goods manufacturing in CPS are merged to align with the manufacturing industry in the job posting data. Similarly, services for private households and other services, excluding public administration in CPS, are consolidated to match the corresponding category in the job posting data. Consequently, our data cover 19 industries. The occupation categories in CPS and the job posting data are identical, resulting in 22 occupations. The concordance [Tables A1 and A2](#) are included in the Appendix.

In total, we have a very large number of postings: 253,570,684. Noticing these job postings corresponds to a much smaller number of unique combinations even when we consider all four variables of interest, we instead use the combination level data with frequency weights. In particular, the raw posting data corresponds to 2,277,880

combinations of month  $\times$  state  $\times$  occupation  $\times$  industry. This much smaller data set ( $2,277,880/253,570,684 = 0.90\%$ ) significantly reduces our computing costs.

While new unique JPCs offer a significant improvement over total and unique JPCs, they still provide a national-level perspective, which may not adequately capture the nuances of labor market dynamics across diverse regions, industries, and occupations. To address this limitation and gain deeper insights into these variations, we introduce four additional JPC measures using the RWET method. This method is particularly valuable in addressing the issue of “thin markets,” where traditional JPCs may be unreliable due to small sample sizes. By leveraging all job postings across the country simultaneously, the RWET method enables us to estimate changes in labor market demand even for relatively rare combinations of state, occupation, and industry.

The implementation of the RWET approach in our study is slightly different from that described in [Shen and Taska \(2020\)](#) because we are using a collapsed dataset of job postings. Specifically, let there be  $N_0$  job posting combinations for the 3,652 days for the period 2010–2019; call this set A. Let the frequent weight for set A be  $w_{mn, s, occ, ind}$ , which is the number of job postings for calendar month  $mn$ , state  $s$ , occupation  $occ$ , and industry  $ind$ .

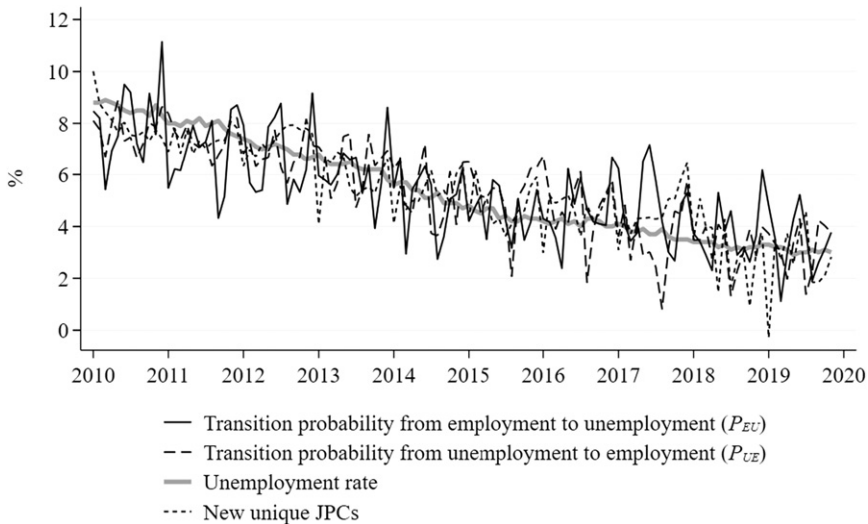
Without loss of generality, let us construct JPCs adjusted by state, occupation, and industry for 2010 January. Let there be  $N_1$  job posting combinations for the 31 days for 2010 January; call this set B. Obviously, set B is a subset of set A. First, we can pool the two sets together and set dummy  $Y$  to be one if the observation comes from set B. For every observation from A, we set its weight,  $w'$  to be  $w_{mn, s, occ, ind}/3652$ ; for every observation from set B,  $w'$  is  $w_{2010.1, s, occ, ind}/31$ .

Next, we conduct a probit estimation control for dummies of state, occupation, and industry, with  $w'$  as the weight. Post-estimation, we can calculate,  $\hat{p}_{2010.1, s, ind, occ}$ , the predicted probability of any combination of state, occupation, and industry to come from 2010 January instead of a random date of the 10-year period. Finally, if we take the JPCs of set A as unit, then the relative JPCs adjusted by state, occupation, and industry are then  $\hat{p}_{2010.1, s, ind, occ} / (1 - \hat{p}_{2010.1, s, ind, occ})$ .

We can derive such adjusted JPCs for every month, state, occupation, and industry combinations. Furthermore, we can use the same approach to construct JPCs adjusted by the following: (1) state; (2) state  $\times$  occupation; and (3) state  $\times$  industry, in addition to the JPCs adjusted by (4) state  $\times$  occupation  $\times$  industry, as shown above. These are the four adjusted JPC measures used in addition to the raw national JPCs in this study.

### 3. DESCRIPTIVE EXPLORATIONS

We begin our empirical analysis by presenting two complementary sets of figures. In both [Figs. 1](#) and [2](#), the transitional probabilities between employment and unemployment,  $P_{EU}$  and  $P_{UE}$ , are used as benchmarks. Here  $P_{EU}$  refers to the probability of an employed individual in month  $t$  being unemployed in month  $t + 1$ ;  $P_{UE}$  refers to the probability of an unemployed individual in month  $t$  being employed in month  $t + 1$ . [Fig. 1](#) plots the longitudinal trajectories of  $P_{EU}$  and



*Fig. 1.* Longitudinal Dynamics of Transition Probabilities, Unemployment Rate, and New Unique JPCs.

*Note:* Three data sources are used in this figure.  $P_{EU}$  and  $P_{UE}$  are derived from monthly CPS microdata; monthly unemployment rate comes from the Bureau of Labor Statistics (BLS); the new unique JPCs are constructed from the unduplicated version of Lightcast’s weekly job posting datasets.  $P_{EU}$  refers to the probability of an employed individual in month  $t$  being unemployed in month  $t + 1$ ;  $P_{UE}$  refers to the probability of an unemployed individual in month  $t$  being employed in month  $t + 1$ . The mean and standard deviation (percent) of raw  $P_{EU}$  and  $P_{UE}$  are 0.954(0.240) and 27.501(6.782), respectively. Those of the unemployment rate (percent) and the new unique JPCs (millions) are 5.398(1.890) and 2.015(0.711), respectively. To ensure the comparability, we use the negative values of  $P_{UE}$  and the new unique JPCs.

Additionally, we adjust  $P_{EU}$ ,  $(-P_{UE})$ , and  $(-JPC)$  to have the same mean and standard deviation as the unemployment rate.

$P_{UE}$  against the unemployment rate and JPCs; **Fig. 2** focuses on the correlations between pairs of these measures using scatter plots.

**Fig. 1** adapts  $P_{EU}$ ,  $P_{UE}$ , and JPCs to account for their divergent time-series properties relative to the unemployment rate, thereby enabling insightful comparisons. The negative values of  $P_{UE}$  and JPCs are used, and  $P_{EU}$ ,  $(-P_{UE})$ , and  $(-JPC)$  are also adjusted to have the same mean and standard deviation as the unemployment rate. **Fig. 1** shows a consistent decrease in the unemployment rate throughout the study period, 2010–2019, reflecting a continuous improvement in the labor market. This trend coincides with a concurrent increase in  $P_{UE}$  and JPCs, and a decrease in  $P_{EU}$ .

In addition to the overall consistency of the trends in these four aggregate measures, **Fig. 1** reveals that the unemployment rate exhibits smoother variability compared with the other three measures. This difference in variability stems from the

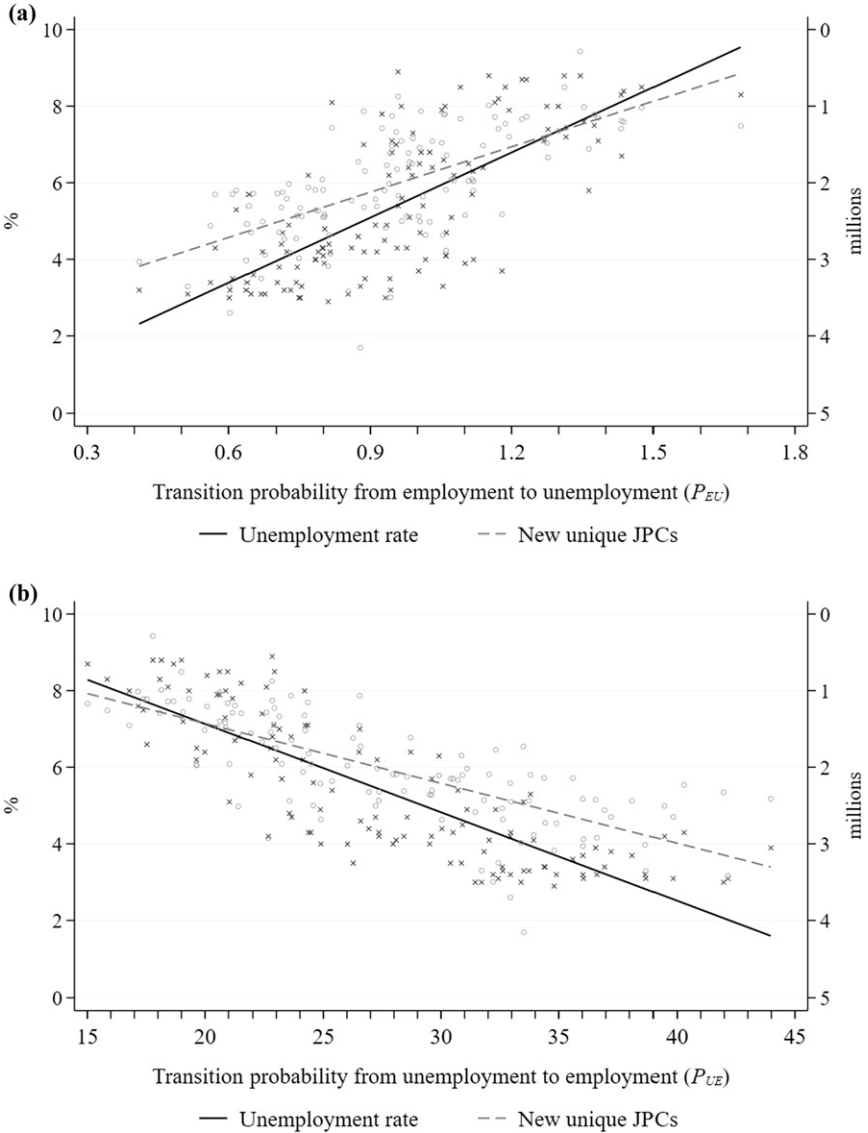


Fig. 2. Correlation Between Transition Probabilities and Unemployment Rate and New Unique JPCs. (a). Transitions From Employment to Unemployment. (b). Transitions From Unemployment to Employment.

Note: Data sources and variables definitions are the same as explained in the note of Fig. 1. The unemployment rate is presented on the left y-axis (an increasing sequence), while the new unique JPC is presented on the right y-axis (a decreasing sequence).