This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: Big Data for Twenty-First-Century Economic Statistics

Volume Authors/Editors: Katharine G. Abraham, Ron S. Jarmin, Brian Moyer, and Matthew D. Shapiro, editors

Volume Publisher: University of Chicago Press

Volume ISBNs: 978-0-226-80125-4 (cloth), 978-0-226-80139-1 (electronic)

Volume URL: https://www.nber.org/books-and-chapters/big-data-twenty-first-century-economic-statistics

Conference Date: March 15-16, 2019

Publication Date: Februrary 2022

Chapter Title: Front matter, Prefatory note

Chapter Author(s):

Chapter URL: https://www.nber.org/books-and-chapters/big-data-twenty-first-century-economic-statistics/front-matter-prefatory-note

Chapter pages in book: p. –11 – 0

Big Data for Twenty-First-Century Economic Statistics

NBER
NATIONAL BUREAU of ECONOMIC RESEARCH

Studies in Income and Wealth
Volume 79

National Bureau of Economic Research
Conference on Research in Income and Wealth

NBER
NATIONAL BUREAU of ECONOMIC RESEARCH

Big Data for Twenty-First-Century Economic Statistics

Edited by **Katharine G. Abraham, Ron S. Jarmin, Brian C. Moyer, and Matthew D. Shapiro**

The University of Chicago Press

Chicago and London

The University of Chicago Press, Chicago 60637
The University of Chicago Press, Ltd., London
© 2022 by the National Bureau of Economic Research
Chapter 6, "Transforming Naturally Occurring Text Data into Economic Statistics: The Case of Online Job Vacancy Postings," by Arthur Turrell, Bradley Speigner, Jyldyz Djumalieva, David Copple, and James Thurgood © Bank of England
All rights reserved. No part of this book may be used or reproduced in any manner whatsoever without written permission, except in the case of brief quotations in critical articles and reviews. For more information, contact the University of Chicago Press, 1427 E. 60th St., Chicago, IL 60637.
Published 2022
Printed in the United States of America

31 30 29 28 27 26 25 24 23 22 1 2 3 4 5

ISBN-13: 978-0-226-80125-4 (cloth)
ISBN-13: 978-0-226-80139-1 (e-book)
DOI: https://doi.org/10.7208/chicago/9780226801391.001.0001

Library of Congress Cataloging-in-Publication Data

Names: Abraham, Katharine G., editor. | Jarmin, Ronald S., 1964–, editor. | Moyer, Brian, editor. | Shapiro, Matthew D. (Matthew David), editor.
Title: Big data for twenty-first-century economic statistics / edited by Katharine G. Abraham, Ron S. Jarmin, Brian C. Moyer, Matthew D. Shapiro.
Other titles: Big data for 21st century economic statistics | Studies in income and wealth ; v. 79.
Description: Chicago : University of Chicago Press, 2022. | Series: Studies in income and wealth ; volume 79
Identifiers: LCCN 2021030585 | ISBN 9780226801254 (cloth) | ISBN 9780226801391 (ebook)
Subjects: LCSH: Economics—Statistical methods—Data processing. | Big data.
Classification: LCC HB143.5 .B54 2022 | DDC 330.072/7—dc23
LC record available at https://lccn.loc.gov/2021030585

♾ This paper meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).

National Bureau of Economic Research

Officers

John Lipsky, *Chair*
Peter Blair Henry, *Vice-Chair*
James M. Poterba, *President and Chief Executive Officer*
Robert Mednick, *Treasurer*
Kelly Horak, *Controller and Assistant Corporate Secretary*
Alterra Milone, *Corporate Secretary*
Denis Healy, *Assistant Corporate Secretary*

Directors At Large

Susan M. Collins
Kathleen B. Cooper
Charles H. Dallara
George C. Eads
Jessica P. Einhorn
Mohamed El-Erian
Diana Farrell
Helena Foulkes
Jacob A. Frenkel
Robert S. Hamada
Peter Blair Henry
Karen N. Horn
Lisa Jordan
John Lipsky
Laurence H. Meyer
Karen Mills
Michael H. Moskow
Alicia H. Munnell
Robert T. Parry
Douglas Peterson
James M. Poterba
John S. Reed
Hal Varian
Mark Weinberger
Martin B. Zimmerman

Directors by University Appointment

Timothy Bresnahan, *Stanford*
Pierre-André Chiappori, *Columbia*
Maureen Cropper, *Maryland*
Alan V. Deardorff, *Michigan*
Graham Elliott, *California, San Diego*
Edward Foster, *Minnesota*
Bruce Hansen, *Wisconsin-Madison*
Benjamin Hermalin, *California, Berkeley*
Samuel Kortum, *Yale*
George Mailath, *Pennsylvania*
Joel Mokyr, *Northwestern*
Richard L. Schmalensee, *Massachusetts Institute of Technology*
Lars Stole, *Chicago*
Ingo Walter, *New York*
David B. Yoffie, *Harvard*

Directors by Appointment of Other Organizations

Timothy Beatty, *Agricultural and Applied Economics Association*
Martin J. Gruber, *American Finance Association*
Philip Hoffman, *Economic History Association*
Arthur Kennickell, *American Statistical Association*
Robert Mednick, *American Institute of Certified Public Accountants*
Dana Peterson, *The Conference Board*
Lynn Reaser, *National Association for Business Economics*
Peter L. Rousseau, *American Economic Association*
Gregor W. Smith, *Canadian Economics Association*
William Spriggs, *American Federation of Labor and Congress of Industrial Organizations*

Directors Emeriti

George Akerlof
Peter C. Aldrich
Elizabeth E. Bailey
Jagdish Bhagwati
John H. Biggs
Don R. Conlan
Ray C. Fair
Saul H. Hymans
Marjorie B. McElroy
Rudolph A. Oswald
Andrew Postlewaite
John J. Siegfried
Craig Swan
Marina v. N. Whitman

Relation of the Directors to the Work and Publications of the NBER

1. The object of the NBER is to ascertain and present to the economics profession, and to the public more generally, important economic facts and their interpretation in a scientific manner without policy recommendations. The Board of Directors is charged with the responsibility of ensuring that the work of the NBER is carried on in strict conformity with this object.

2. The President shall establish an internal review process to ensure that book manuscripts proposed for publication DO NOT contain policy recommendations. This shall apply both to the proceedings of conferences and to manuscripts by a single author or by one or more co-authors but shall not apply to authors of comments at NBER conferences who are not NBER affiliates.

3. No book manuscript reporting research shall be published by the NBER until the President has sent to each member of the Board a notice that a manuscript is recommended for publication and that in the President's opinion it is suitable for publication in accordance with the above principles of the NBER. Such notification will include a table of contents and an abstract or summary of the manuscript's content, a list of contributors if applicable, and a response form for use by Directors who desire a copy of the manuscript for review. Each manuscript shall contain a summary drawing attention to the nature and treatment of the problem studied and the main conclusions reached.

4. No volume shall be published until forty-five days have elapsed from the above notification of intention to publish it. During this period a copy shall be sent to any Director requesting it, and if any Director objects to publication on the grounds that the manuscript contains policy recommendations, the objection will be presented to the author(s) or editor(s). In case of dispute, all members of the Board shall be notified, and the President shall appoint an ad hoc committee of the Board to decide the matter; thirty days additional shall be granted for this purpose.

5. The President shall present annually to the Board a report describing the internal manuscript review process, any objections made by Directors before publication or by anyone after publication, any disputes about such matters, and how they were handled.

6. Publications of the NBER issued for informational purposes concerning the work of the Bureau, or issued to inform the public of the activities at the Bureau, including but not limited to the NBER Digest and Reporter, shall be consistent with the object stated in paragraph 1. They shall contain a specific disclaimer noting that they have not passed through the review procedures required in this resolution. The Executive Committee of the Board is charged with the review of all such publications from time to time.

7. NBER working papers and manuscripts distributed on the Bureau's web site are not deemed to be publications for the purpose of this resolution, but they shall be consistent with the object stated in paragraph 1. Working papers shall contain a specific disclaimer noting that they have not passed through the review procedures required in this resolution. The NBER's web site shall contain a similar disclaimer. The President shall establish an internal review process to ensure that the working papers and the web site do not contain policy recommendations, and shall report annually to the Board on this process and any concerns raised in connection with it.

8. Unless otherwise determined by the Board or exempted by the terms of paragraphs 6 and 7, a copy of this resolution shall be printed in each NBER publication as described in paragraph 2 above.

Contents

Prefatory Note xi

Introduction: Big Data for Twenty-First-Century Economic Statistics: The Future Is Now 1
Katharine G. Abraham, Ron S. Jarmin, Brian C. Moyer, and Matthew D. Shapiro

I. Toward Comprehensive Use of Big Data in Economic Statistics

1. **Reengineering Key National Economic Indicators** 25
Gabriel Ehrlich, John C. Haltiwanger, Ron S. Jarmin, David Johnson, and Matthew D. Shapiro

2. **Big Data in the US Consumer Price Index: Experiences and Plans** 69
Crystal G. Konny, Brendan K. Williams, and David M. Friedman

3. **Improving Retail Trade Data Products Using Alternative Data Sources** 99
Rebecca J. Hutchinson

4. **From Transaction Data to Economic Statistics: Constructing Real-Time, High-Frequency, Geographic Measures of Consumer Spending** 115
Aditya Aladangady, Shifrah Aron-Dine, Wendy Dunn, Laura Feiveson, Paul Lengermann, and Claudia Sahm

5. **Improving the Accuracy of Economic Measurement with Multiple Data Sources: The Case of Payroll Employment Data** 147
Tomaz Cajner, Leland D. Crane, Ryan A. Decker, Adrian Hamins-Puertolas, and Christopher Kurz

II. Uses of Big Data for Classification

6. **Transforming Naturally Occurring Text Data into Economic Statistics: The Case of Online Job Vacancy Postings** 173
Arthur Turrell, Bradley Speigner, Jyldyz Djumalieva, David Copple, and James Thurgood

7. **Automating Response Evaluation for Franchising Questions on the 2017 Economic Census** 209
Joseph Staudt, Yifang Wei, Lisa Singh, Shawn Klimek, J. Bradford Jensen, and Andrew Baer

8. **Using Public Data to Generate Industrial Classification Codes** 229
John Cuffe, Sudip Bhattacharjee, Ugochukwu Etudo, Justin C. Smith, Nevada Basdeo, Nathaniel Burbank, and Shawn R. Roberts

III. Uses of Big Data for Sectoral Measurement

9. **Nowcasting the Local Economy: Using Yelp Data to Measure Economic Activity** 249
Edward L. Glaeser, Hyunjin Kim, and Michael Luca

10. **Unit Values for Import and Export Price Indexes: A Proof of Concept** 275
Don A. Fast and Susan E. Fleck

11. **Quantifying Productivity Growth in the Delivery of Important Episodes of Care within the Medicare Program Using Insurance Claims and Administrative Data** 297
John A. Romley, Abe Dunn, Dana Goldman, and Neeraj Sood

12. Valuing Housing Services in the Era of Big Data: A User Cost Approach Leveraging Zillow Microdata 339
Marina Gindelsky, Jeremy G. Moulton, and Scott A. Wentland

IV. Methodological Challenges and Advances

13. Off to the Races: A Comparison of Machine Learning and Alternative Data for Predicting Economic Indicators 373
Jeffrey C. Chen, Abe Dunn, Kyle Hood, Alexander Driessen, and Andrea Batch

14. A Machine Learning Analysis of Seasonal and Cyclical Sales in Weekly Scanner Data 403
Rishab Guha and Serena Ng

15. Estimating the Benefits of New Products 437
W. Erwin Diewert and Robert C. Feenstra

Contributors 475
Author Index 479
Subject Index 485

Prefatory Note

This volume contains revised versions of the papers presented at the Conference on Research in Income and Wealth entitled "Big Data for 21st Century Economic Statistics," held in Washington, DC, on March 15–16, 2019.

We gratefully acknowledge the financial support for this conference provided by the Alfred P. Sloan Foundation under grant G-2018-11019. Support for the general activities of the Conference on Research in Income and Wealth is provided by the following agencies: Bureau of Economic Analysis, Bureau of Labor Statistics, the Census Bureau, the Board of Governors of the Federal Reserve System, the Statistics of Income/Internal Revenue Service, and Statistics Canada.

We thank Katharine G. Abraham, Ron S. Jarmin, Brian C. Moyer, and Matthew D. Shapiro, who served as conference organizers and as editors of the volume.

Executive Committee, November 2020

John M. Abowd
Katharine G. Abraham (chair)
Susanto Basu
Ernst R. Berndt
Carol A. Corrado
Alberto Cavallo
Lucy Eldridge
John C. Haltiwanger
Ron S. Jarmin
J. Bradford Jensen
Barry Johnson
Greg Peterson
Valerie A. Ramey
Peter Schott
Daniel Sichel
Erich Strassner
William Wascher