

# Optimal Probabilistic Record Linkage: Best Practice for Linking Employers in Survey and Administrative Data\*

John M. Abowd<sup>† ‡</sup>    Joelle Abramowitz<sup>§</sup>    Margaret C. Levenstein<sup>§</sup>    Kristin McCue<sup>†</sup>  
Dhiren Patki<sup>¶</sup>    Trivellore Raghunathan<sup>§||</sup>    Ann M. Rodgers<sup>§</sup>    Matthew D. Shapiro<sup>¶§\*\*</sup>  
Nada Wasi<sup>††</sup>

January 2019

## Abstract

This paper illustrates an application of record linkage between a household-level survey and an establishment-level frame in the absence of unique identifiers. Linkage between frames in this setting is challenging because the distribution of employment across firms is highly asymmetric. To address these difficulties, this paper uses a supervised machine learning model to probabilistically link survey respondents in the Health and Retirement Study (HRS) with employers and establishments in the Census Business Register (BR) to create a new data source which we call the CenHRS. Multiple imputation is used to propagate uncertainty from the linkage step into subsequent analyses of the linked data. The linked data reveal new evidence that survey respondents' misreporting and selective nonresponse about employer characteristics are systematically correlated with wages.

**Keywords:** Probabilistic record linkage; survey data; administrative data; multiple imputation; measurement error; nonresponse

---

\* Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed (DRB release number 6961 and 7357). This research is supported by the Alfred P. Sloan Foundation through the CenHRS project at the University of Michigan with additional support from the Michigan Node of the NSF-Census Research Network (NCRN) under NSF SES 1131500.

<sup>†</sup> U.S. Census Bureau

<sup>‡</sup> Labor Dynamics Institute, Cornell University

<sup>§</sup> Institute for Social Research, University of Michigan

<sup>¶</sup> Department of Economics, University of Michigan

<sup>||</sup> Department of Biostatistics, University of Michigan

<sup>\*\*</sup> NBER

<sup>††</sup> Puey Ungphakorn Institute for Economic Research, Bank of Thailand