Bird's eye view of literature

Record linkage surveys

- Books on record linkage: Christen (2012); Herzog et al. (2007); Harron et al. (2015)
- Review of free software tools: Köpcke and Rahm (2010)
- Survey of RL techniques in economics : Bailey et al. (2017), Abramitzky et al. (2019a)
- Surveys by the"father" and "mother" of modern record linkage (Winkler, 1999)
- Comparison of deterministic vs. probabilistic matching methods (Doidge and Harron, 2018)

Matching methods

- Probabilistic Record Linkage (Fellegi and Sunter, 1969; Abramitzky et al., 2018)
- Prior-informed imputation (Goldstein et al., 2012)
- Bayesian "Beta" Record Linkage (Sadinle, 2017)
- Deterministic methods for economics (Ferrie, 1996; Abramitzky et al., 2012, 2019b)
- Deterministic methods for bounding parameter (Nix and Qian, 2015)
- Allow multiple matches by imputing equal weight values (Bleakley and Ferrie, 2016)

Estimation methods

- Integrated matching and estimation procedures
  - Iterative matching, regression analysis, imputation (Scheuren and Winkler, 1997)
  - Bayesian matching and population size (Tancredi and Liseo, 2015)
- Bias corrections for regression with matched data
  - Bias correction for regression (Scheuren and Winkler, 1993; Lahiri and Larsen, 2005)
  - Weighted least squares-based bias correction for linear and logistic regression (Hof and Zwinderman, 2012)
  - Bias correction for NN-matching (Hirukawa and Prokhorov, 2018)
- GMM with multiple matches (Anderson et al., 2019)

References

P. Christen, *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Springer Publishing Company, Incorporated, 2012.

T. N. Herzog, F. J. Scheuren, and W. E. Winkler, *Data Quality and Record Linkage Techniques*, 1st ed. Springer Publishing Company, Incorporated, 2007.

K. Harron, H. Goldstein, and C. Dibben, *Methodological Developments in Data Linkage*. United States: John Wiley Sons Inc., 2015.

H. Köpcke and E. Rahm, "Frameworks for entity matching: A comparison," *Data Knowledge Engineering*, vol. 69, pp. 197–210, 02 2010.

M. Bailey, C. Cole, M. Henderson, and C. Massey, "How well do automated linking methods perform? lessons from u.s. historical data," National Bureau of Economic Research, Working Paper 24019, November 2017. [Online]. Available: http://www.nber.org/papers/w24019

R. Abramitzky, L. Boustan, K. Eriksson, J. Feigenbaum, and S. Perez, "Automated linking of historical data," *NBER Working Paper*, 2019.

W. Winkler, "The state of record linkage and current research problems," *Statist. Med.*, vol. 14, 10 1999.

J. Doidge and K. Harron, "Demystifying probabilistic linkage," *International Journal for Population Data Science*, vol. 3, 01 2018.

I. P. Fellegi and A. B. Sunter, "A theory for record linkage," *Journal of the American Statistical Association*, vol. 64, pp. 1183–1210, 1969.

R. Abramitzky, R. Mill, and S. Perez, "Linking individuals across historical sources: a fully automated approach," National Bureau of Economic Research, Working Paper 24324, February 2018. [Online]. Available: http://www.nber.org/papers/w24324

H. Goldstein, K. L. Harron, and A. M. Wade, "The analysis of record-linked data using multiple imputation with data value priors." *Statistics in medicine*, vol. 31 28, pp. 3481–93, 2012.

M. Sadinle, "Bayesian estimation of bipartite matchings for record linkage," *Journal of the American Statistical Association*, vol. 112, no. 518, p. 600?612, 2017.

J. P. Ferrie, "A new sample of males linked from the public use microdata sample of the 1850 u.s. federal census of population to the 1860 u.s. federal census manuscript schedules," *Historical Methods*, vol. 29, no. 4, pp. 141–156, 1 1996.

R. Abramitzky, L. P. Boustan, and K. Eriksson, "Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration," *American Economic Review*, vol. 102, no. 5, pp. 1832–56, May 2012. [Online]. Available: http://www.aeaweb.org/articles?id=10.1257/aer.102.5.1832

——, "To the new world and back again: Return migrants in the age of mass migration," *ILR Review*, vol. 72, no. 2, pp. 300–322, 2019. [Online]. Available: https://doi.org/10.1177/0019793917726981

E. Nix and N. Qian, "The fluidity of race: "Passing" in the united states, 1880-1940," National Bureau of Economic Research, Working Paper 20828, January 2015. [Online]. Available: http://www.nber.org/papers/w20828

H. Bleakley and J. Ferrie, "Shocking behavior: Random wealth in antebellum georgia and human capital across generations," *The Quarterly Journal of Economics*, vol. 131, no. 3, pp. 1455–1495, 2016. [Online]. Available: https://doi.org/10.1093/qje/qjw014

F. Scheuren and W. Winkler, "Regression analysis of data files that are computer matched - part ii," *Survey Methodology*, vol. 23, 01 1997.

A. Tancredi and B. Liseo, "Regression analysis with linked data: problems and possible solutions," *Statistica*, vol. 75, no. 1, pp. 19–35, 2015. [Online]. Available: https://rivista-statistica.unibo.it/article/view/5821

F. Scheuren and W. Winkler, "Regression analysis of data files that are computer matched," *Survey Methodology*, vol. 19, 01 1993.

P. Lahiri and M. D. Larsen, "Regression analysis with linked data," *Journal of the American Statistical Association*, vol. 100, no. 469, pp. 222–230, 2005. [Online]. Available: http://www.jstor.org/stable/27590532

M. H. P. Hof and A. H. Zwinderman, "Methods for analyzing data from probabilistic linkage strategies based on partially identifying variables," *Statistics in Medicine*, vol. 31, no. 30, pp. 4231–4242, 2012. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.5498

M. Hirukawa and A. Prokhorov, "Consistent estimation of linear regression models using matched data," *Journal of Econometrics*, vol. 203, no. 2, pp. 344 – 358, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304407617302464

R. Anderson, B. Honore, and A. Lleras-Muney, "Estimation and inference using imperfectly matched data," *Working paper*, August 2019. [Online]. Available: http://www.github.com/rachelsanderson/ImperfectMatching