Birth, Death, and Record Linkage: Conducting Survival Analysis in the Presence of Linkage Error

Elan Segarra[*]

April 2020

Abstract

Longitudinal panels have a long history of use across the social sciences; however, they can be imperfect representations of reality when fuzzy linkage methods are employed during their creation. In this paper I study survival estimation (e.g. firm death, mortality, or emigration) when missed linkages induce error in the observed lifetime durations, and thus inconsistency in standard survival estimators. Importantly, the error introduced does not take the form of a standard competing risks model, and the methods developed here illustrate that consistency of the parameters of interest can be restored without correcting the linkages. This work makes three distinct theoretical contributions: First, under a known independent linkage error process I show that the marginal distribution of time to death is non-parametrically identified from linkage error induced durations, and I provide consistent estimators. Second, I provide sharp informative bounds on the marginal distribution of death when independence is relaxed. Third, when start and end periods are also observed, I show the marginal distribution of death can be point identified without imposing any dependence structure on the linkage error. The methods are then applied to longitudinal business data (where linkage error occurs due to firm relocation) to show that traditional estimates of survival rates of new firms are significantly overestimated. Finally, I discuss additional applications to the estimation of household migration and mortality where linkage error is induced by family name changes at marriage.

---

[*]University of Wisconsin–Madison, Department of Economics, United States

This is just a placeholder. Paper coming soon.