Simply omitting or better using proxies? – Some guidance for multiple regression analysis with administrative labour market data.

Shihan Du

*Copenhagen Business School*

Administrative or register data are often characterised by a large number of observations but only a small set of operations related variables. In order to mitigate any possible omitted variable bias, empirical labour market research regularly uses proxy variables from individual employment biographies. We provide an in-depth analysis about the role of omitted variables and the usefulness of such proxy variables using a sample of German administrative data that is linked with rich household level survey data. We find that results of a wage regressions possess a sizable omitted variable bias. Proxy variables tend to reduce this bias but only by a small amount. In contrast, there is no evidence that results for a labour market transition model are sizably biased. We observe only in small number of cases a statistical partial relationship between omitted variables and the work history variables.