Ghanaian worker data tutorial

In this tutorial we exploit a unique matched worker and firm dataset, often called linked employer-employee data (LEED), to investigate how the influence of individual characteristics, like education, differs once we control for firm characteristics.

These types of data generally come from two sources. The first is interviews of workers as part of a broader firm survey. This dataset comes from a Ghanaian survey of this nature. Other examples are the work I’m currently doing in Zimbabwe and the Enterprise Surveys of the World Bank (<http://www.enterprisesurveys.org/>). The second is administrative data (often tax records) which links firms and workers. There are a number of datasets of this type for European countries, for example: Germany (<http://fdz.iab.de/en/Integrated_Establishment_and_Individual_Data/LIAB.aspx>); Denmark (<http://www.ike.aau.dk/Research/Data-Sources/>); and France. Other countries are also making these types of data available (see <https://www.academia.edu/536917/A_Guide_to_Linked_Employer-Employee_Data_Sources_in_the_EU_and_Beyond_1st_Edition_Feb_2008_> for a listing of these countries, and South Africa is also in the process of constructing a dataset of this nature.

We will undertake the tutorial in both Stata and R. Information on R can be found here (<https://www.r-project.org/>), it may be useful to load R Studio (<https://www.rstudio.com/)>., and here is a bit on one of the people heavily involved in developing packages for R (<http://priceonomics.com/hadley-wickham-the-man-who-revolutionized-r/>).

Questions to investigate:

* Do large firms and small firms hire different types of individuals?
* Do they pay different amounts?
* If yes, is this because they hire different types of people or is it something else?
* Do returns to education differ between different sized firms?
* How would you go about disentangling the firm from the individual effect?