

**Read me File for Replication of “Subsidizing vocational Training for
Disadvantaged Youth in Colombia: Evidence from a Randomized Trial”**

by Orazio Attanasio, Adriana Kugler, and Costas Meghir

The Stata dataset AEJApp-20090168_data.dta, created with Stata 10, includes all variables needed to replicate the analysis reported in Tables 1-6 in the paper. The do file AEJApp-20090168_program.txt includes Stata commands to generate any necessary variables not in the original data set which are needed for the analysis as well as commands to replicate the results in Tables 1-6. The names of the variables in the data which include “lb” or “04” in their names are baseline or pre treatment variables, while those which include “s” or “06” are the follow up or post treatment variables. The names of the variables (which are described in Appendix A of the paper) are:

dwomen - An indicator for whether the person is a woman.

age_lb and age_s - Age.

dmarried_lb and dmarried_s - Married status dummy.

educ_lb and educ_s - Years of formal education.

tenure_04 and tenure_06 - Tenure on the job.

empl_04 and empl_06 - Employment indicator.

pempl_04 and pempl_06 - Paid employment indicator.

salary_04 and salary_06 - Real salary in Colombian Pesos.

profit_04 and profit_06 - Real net profits in Colombian Pesos.

dformal_04 and dformal_06 - Formal employment indicator.

contract_04 and contract_06 - Written contract indicator.

days_04 and days_06 - Days of work per month in levels.

hours_04 and hours_06 - Hours worked per week.

dcontinue - A dummy for whether the person continues in the sample in the follow up survey.

City - Name of the city where the person lives.

codigo_ecap - Training institution identifier.

codigo_curs - Identifier for the course in which the person registered.

p_selec - administrative data on probability of randomly being offered training at a given training institution.