Labor Economics Project. Returns to additional training

Petrakov Sergey, Semeneev Danat, Kuleshov Gleb

Problem

- The price of additional education has grown substantially over last few decades, people may spend enormous sums of money on training without any evidence of significant payoff to their future earnings.

Questions

- Does additional training affect future returns?
- How does additional training affect individual wage?

Data

- Cross-sectional data - RLMS 2015 year (10881 observations in General Model).
- Restrictions on the dataset:

1. Respondents with no information about training omitted
2. ‘Working’ respondents with no information on salary omitted
3. Respondents refusing to assess their skills with 0-9 scale omitted
4. Respondents refusing to call their education level omitted

Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
fact_wage	6,947	21722.36	12774.06	50	68000
wage	10,903	13840.71	14596.66	0	68000
fact_avg_hs	417	7.63789	11.42012	1	96
avg_train_~y	10,903	.2921214	2.66882	0	96
fact_tr_days	405	24.80494	31.06611	1	210
real_train~s	10,903	.9213978	7.600821	0	210
skills	10,033	6.067378	2.045545	1	9
age	10,903	40.68642	13.58811	16	65
trained	10,902	.0398092	.1955197	0	1
fem	10,903	.5544346	.4970508	0	1
EDUC1	10,890	.5073462	.499969	0	1
EDUC2	10,890	.2568411	.4369111	0	1
EDUC3	10,890	.0141414	.1180793	0	1

General Model

Two-Stage Heckman procedure

Heckman selection model -- two-step estimates (regression model with sample selection)

Number of obs	= 10,881
Selected	= 6,707
Nonselected	= 4,174
Wald chi2(9)	= 841.01
Prob > chi2	= 0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnwage						
wrkgov	-.1104459	.0145465	-7.59	0.000	-.1389566	-.0819353
trained	.1687497	.0312764	5.40	0.000	.1074491	.2300503
disab	-.4476685	.1231784	-3.63	0.000	-.6890936	-.2062433
age	.0702515	.0181955	3.86	0.000	.0345889	.1059141
age2	-.0009158	.0002209	-4.15	0.000	-.0013488	-.0004828
EDUC1	.1170522	.0459323	2.55	0.011	.0270265	.2070779
EDUC2	.4890591	.0594601	8.22	0.000	.3725193	.6055988
EDUC3	.5723178	.0890156	6.43	0.000	.3978504	.7467853
fem	-.3859805	.0305605	-12.63	0.000	-.4458779	-.3260831
_cons	8.504126	.4385414	19.39	0.000	7.644601	9.363651

Selection equation (analysis of the fact of work via the first stage of Heckman procedure)

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
arbeit						
age	.2296983	.0069883	32.87	0.000	.2160015	.2433951
age2	-.0027858	.0000828	-33.64	0.000	-.0029481	-.0026235
EDUC1	.5140918	.0345394	14.88	0.000	.4463958	.5817877
EDUC2	.7264003	.0396133	18.34	0.000	.6487596	.804041
EDUC3	.9330593	.1221935	7.64	0.000	.6935645	1.172554
disab	-1.239042	.0663036	-18.69	0.000	-1.368995	-1.10909
fem	-.3914831	.0273198	-14.33	0.000	-.445029	-.3379372
non_sal_inc	-9.77e-07	2.87e-07	-3.40	0.001	-1.54e-06	-4.14e-07
_cons	-4.068176	.1330967	-30.57	0.000	-4.329041	-3.807311
/mills						
lambda	.2391808	.1420618	1.68	0.092	-.0392552	.5176168
rho	0.40540					
sigma	.58998437					

```
--------------------------------------------
Tobit 2 model (sample selection model)
2-step Heckman / heckit estimation
10877 observations (4169 censored and 6708 observed)
23 free parameters (df = 10855)
Probit selection equation:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.060e+00  1.331e-01 -30.503  < 2e-16 ***
age          2.293e-01  6.988e-03  32.814  < 2e-16 ***
age2        -2.781e-03  8.281e-05 -33.580  < 2e-16 ***
EDUC1        5.136e-01  3.455e-02  14.867  < 2e-16 ***
EDUC2        7.256e-01  3.962e-02  18.315  < 2e-16 ***
EDUC3        9.428e-01  1.219e-01   7.733 1.15e-14 ***
fem         -3.921e-01  2.733e-02 -14.347  < 2e-16 ***
non_sal_inc -9.774e-07  2.871e-07  -3.405 0.000664 ***
disab       -1.240e+00  6.630e-02 -18.701  < 2e-16 ***
Outcome equation:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     8.501e+00  4.381e-01  19.405  < 2e-16 ***
age             7.037e-02  1.817e-02   3.873 0.000108 ***
age2           -9.175e-04  2.206e-04  -4.159 3.23e-05 ***
EDUC1           1.181e-01  4.592e-02   2.571 0.010145 *
EDUC2           4.903e-01  5.944e-02   8.248  < 2e-16 ***
EDUC3           5.746e-01  8.938e-02   6.429 1.34e-10 ***
trained         1.612e-01  3.829e-02   4.210 2.58e-05 ***
disab          -4.495e-01  1.233e-01  -3.645 0.000269 ***
fem            -3.864e-01  3.061e-02 -12.623  < 2e-16 ***
govwork        -1.104e-01  1.456e-02  -7.586 3.58e-14 ***
real_train_hours 6.171e-05  1.532e-04   0.403 0.687162
Multiple R-Squared:0.1821,      Adjusted R-Squared:0.1808
   Error terms:
              Estimate Std. Error t value Pr(>|t|)
invMillsRatio   0.2408     0.1422   1.694   0.0903 .
sigma           0.5903         NA      NA       NA
rho             0.4080         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
```

Extension 1. *We refined our general model adding a variable for real training hours in regression*

Restrictions imposed on dataset:

- Age between 16 and 65
- Wage $\leq$ 70 000 rubles
- Evaluation of ln(wage)

Collolaries: the *real_train_hours* variable is statistically insignificant, while *trained* variable is significant at the 0,1% level

Extension 2. We refined our general model adding a variable for real training days in regression

```
10877 observations (4169 censored and 6708 observed)
23 free parameters (df = 10855)
Probit selection equation:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.060e+00  1.331e-01 -30.503  < 2e-16 ***
age          2.293e-01  6.988e-03  32.814  < 2e-16 ***
age2        -2.781e-03  8.281e-05 -33.580  < 2e-16 ***
EDUC1        5.136e-01  3.455e-02  14.867  < 2e-16 ***
EDUC2        7.256e-01  3.962e-02  18.315  < 2e-16 ***
EDUC3        9.428e-01  1.219e-01   7.733 1.15e-14 ***
fem         -3.921e-01  2.733e-02 -14.347  < 2e-16 ***
non_sal_inc -9.774e-07  2.871e-07  -3.405 0.000664 ***
disab       -1.240e+00  6.630e-02 -18.701  < 2e-16 ***
Outcome equation:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.4912955  0.4382734  19.374  < 2e-16 ***
age          0.0707621  0.0181793   3.892 9.98e-05 ***
age2        -0.0009221  0.0002207  -4.178 2.96e-05 ***
EDUC1        0.1191758  0.0459444   2.594 0.009502 **
EDUC2        0.4917977  0.0594703   8.270  < 2e-16 ***
EDUC3        0.5763973  0.0894078   6.447 1.19e-10 ***
trained      0.1439434  0.0398158   3.615 0.000301 ***
disab       -0.4514763  0.1233417  -3.660 0.000253 ***
fem         -0.3868647  0.0306205 -12.634  < 2e-16 ***
govwork     -0.1103672  0.0145462  -7.587 3.53e-14 ***
real_train_days 0.0012008  0.0010932   1.098 0.272014
Multiple R-Squared:0.1822,      Adjusted R-Squared:0.1809
   Error terms:
              Estimate Std. Error t value Pr(>|t|)
invMillsRatio   0.2436     0.1422   1.713   0.0867 .
sigma           0.5908         NA      NA       NA
rho             0.4124         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```

Restrictions imposed on dataset:

- Age between 16 and 65
- Wage $\leq$ 70 000 rubles
- Evaluation of ln(wage)

Collolaries: the *real_train_days* variable is statistically insignificant, while *trained* variable is significant at the 0,1% level

Extension 3. We added variable for real training days in regression

```
10016 observations (3661 censored and 6355 observed)
23 free parameters (df = 9994)
Probit selection equation:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.013e+00  1.418e-01 -28.305  < 2e-16 ***
age          2.294e-01  7.383e-03  31.068  < 2e-16 ***
age2        -2.788e-03  8.718e-05 -31.975  < 2e-16 ***
EDUC1        5.120e-01  3.679e-02  13.916  < 2e-16 ***
EDUC2        7.013e-01  4.167e-02  16.829  < 2e-16 ***
EDUC3        9.531e-01  1.269e-01   7.510 6.39e-14 ***
fem         -3.926e-01  2.861e-02 -13.722  < 2e-16 ***
non_sal_inc -9.786e-07  2.889e-07  -3.388 0.000708 ***
disab       -1.228e+00  6.907e-02 -17.776  < 2e-16 ***
Outcome equation:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.4300839  0.4310638  19.556  < 2e-16 ***
age          0.0578601  0.0179885   3.217 0.001302 **
age2        -0.0008058  0.0002188  -3.682 0.000233 ***
EDUC1        0.0647477  0.0455886   1.420 0.155563
EDUC2        0.3989857  0.0576844   6.917 4.91e-12 ***
EDUC3        0.4900731  0.0885742   5.533 3.23e-08 ***
trained      0.1323658  0.0318195   4.160 3.21e-05 ***
disab       -0.4086934  0.1216273  -3.360 0.000782 ***
fem         -0.3642027  0.0303823 -11.987  < 2e-16 ***
govwork     -0.1185852  0.0145901  -8.128 4.89e-16 ***
skills       0.0695077  0.0041092  16.915  < 2e-16 ***
Multiple R-Squared:0.2137,      Adjusted R-Squared:0.2123
   Error terms:
              Estimate Std. Error t value Pr(>|t|)
invMillsRatio   0.2351     0.1431   1.643      0.1
sigma           0.5758         NA      NA       NA
rho             0.4084         NA      NA       NA
```

Restrictions imposed on dataset:

- Age between 16 and 65
- Wage $\leq$ 70 000 rubles
- Evaluation of ln(wage)

Collolaries: the *skills* variable is statistically significant, and *trained* variable is significant at the 0,1% level. However, here is a high possibility of the endogeneity of variable *skills.*

Result

Training matters.

However, the question of the causal relationship remains open, in the future it is interesting to check it between training and wage, as well as between self skills and wage using instrumental variable.

Theoretical basement

- Behrman (1977) – skills
- Ben-Porath Model – education & age
- Heckman (1979) two stage procedure

Articles:

Brunello, Giorgio, On the Complementarity between Education and Training in Europe (June 2001).

Gerfin, Work-Related Training and Wages: An Empirical Analysis for Male Workers in Switzerland. (March 2004)