STATA ‘tricks’: Creating professional graphs to enhance communication of statistical results

Dr. Paul Sirma – Economist, AIR

Prof. Ashu Handa, AIR Institute Fellow and Kenan Eminent Professor of Public Policy, University of North Carolina

Session II: April 2, 2025

AIR®
Advancing Evidence.
Improving Lives.

UAPS UEPA
Union for African Population Studies
Union pour l'Étude de la Population Africaine

AERC CREA

Outline for today

- Results
  - Table of regression output (required)
  - Summary graphs of coefficients
    - Plot key coefficients from one regression
    - Plot key coefficients from the same regression (Y) for different sub-groups
    - Plot the same coefficient for different outcomes (Y)

Introduction

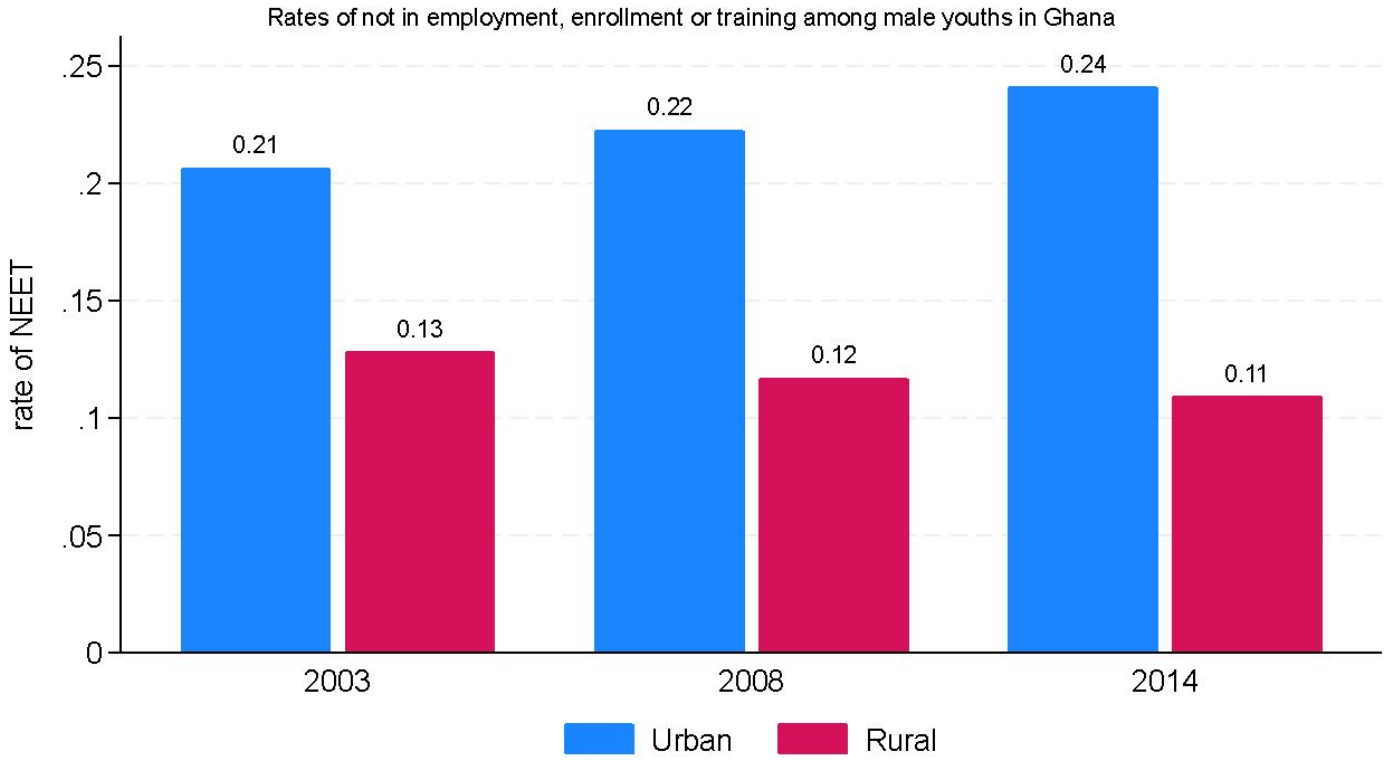
- Package needed: coefplot
  - Displaying regression coefficients and more
  - Author: Ben Jann
  - Resources: https://repec.sowi.unibe.ch/stata/coefplot/

- Installing coefplot package in your computer
  - *ssc install coefplot, replace*

- To understand the syntax of coefplot
  - Call for help in stata: *help coefplot*

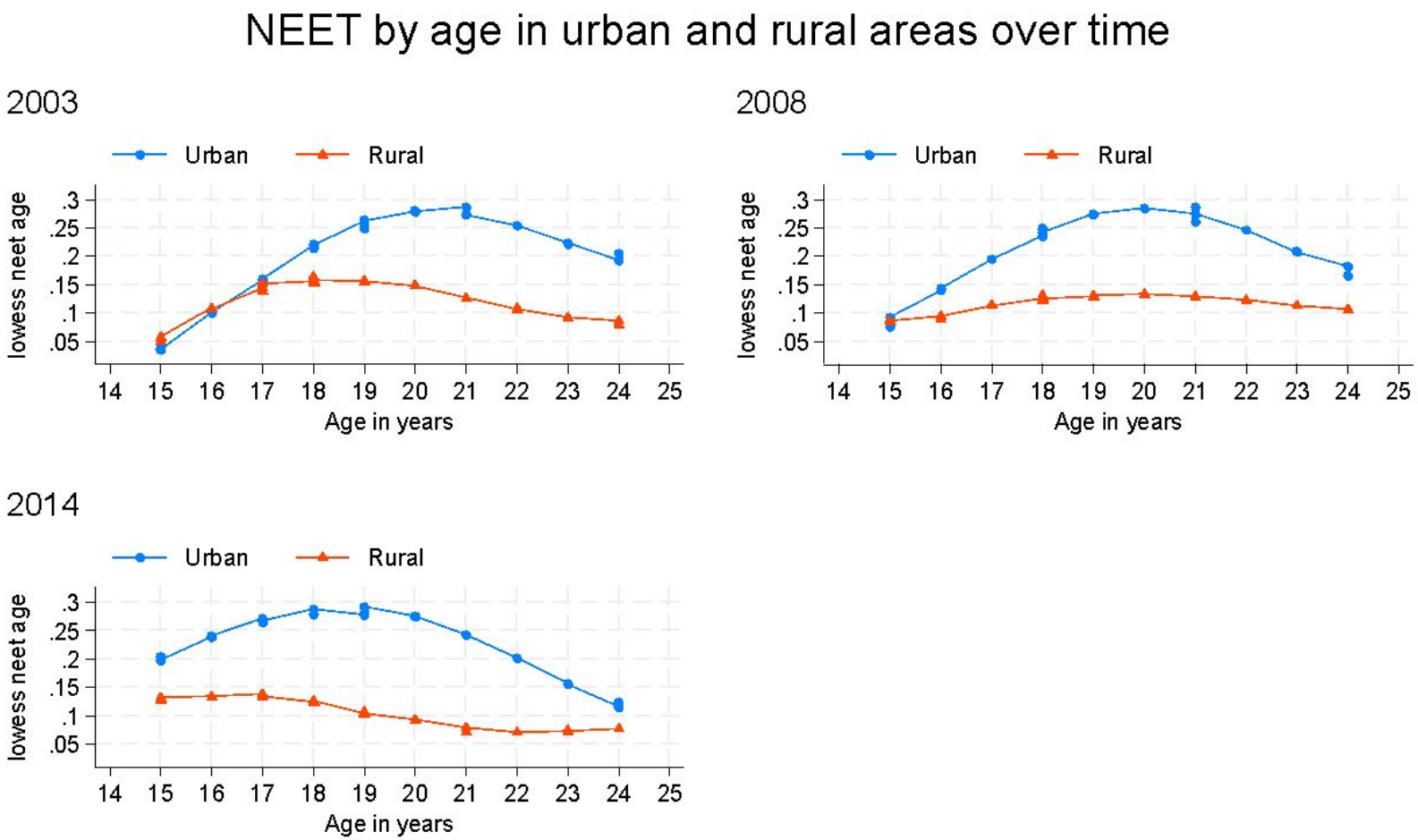
Recap from last session

Rate of not in employment, enrollment or training (NEET) is increasing in urban areas over time among young men in Ghana



Source: Author's calculations using Ghana's Demographic and Household Survey data.

The gap is widening among younger men



Goal: graphical representation of the regression output

Urban:

reg neet year2008 year2014 age_2 age_3 age_4 age_5 age_6 age_7 age_8 age_9 age_10 if rural ==0, r

neet	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
year2008	**.0205034**	**.0215487**	**0.95**	**0.341**	**-.0217556**	**.0627625**
year2014	**.0376215**	**.0222254**	**1.69**	**0.091**	**-.0059646**	**.0812077**
age_2	.0685399	.0317051	2.16	0.031	.0063631	.1307168
age_3	.1212137	.0353165	3.43	0.001	.0519545	.1904728
age_4	.1198266	.0333794	3.59	0.000	.0543662	.1852869
age_5	.2416338	.0382977	6.31	0.000	.1665282	.3167395
age_6	.1992317	.038176	5.22	0.000	.1243649	.2740986
age_7	.2334269	.0411552	5.67	0.000	.1527174	.3141364
age_8	.1142619	.0368932	3.10	0.002	.0419106	.1866132
age_9	.0528976	.0332896	1.59	0.112	-.0123866	.1181817
age_10	.0897838	.0357166	2.51	0.012	.0197399	.1598276
_cons	.0809203	.0231158	3.50	0.000	.0355879	.1262528

Rural:

reg neet year2008 year2014 age_2 age_3 age_4 age_5 age_6 age_7 age_8 age_9 age_10 if rural ==1, r

neet	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
year2008	**-.0132159**	**.01463**	**-0.90**	**0.366**	**-.0419025**	**.0154707**
year2014	**-.020363**	**.0149014**	**-1.37**	**0.172**	**-.0495817**	**.0088558**
age_2	.0510082	.021771	2.34	0.019	.0083196	.0936967
age_3	.0277575	.0223311	1.24	0.214	-.0160294	.0715444
age_4	.0975914	.0234785	4.16	0.000	.0515548	.1436281
age_5	.068804	.0251296	2.74	0.006	.0195299	.1180781
age_6	.0513173	.0246995	2.08	0.038	.0028864	.0997482
age_7	.0190586	.0245162	0.78	0.437	-.0290128	.0671299
age_8	.0046135	.0240916	0.19	0.848	-.0426253	.0518523
age_9	.0289109	.0266241	1.09	0.278	-.0232937	.0811156
age_10	.0100647	.0247859	0.41	0.685	-.0385356	.0586649
_cons	.0891896	.0151379	5.89	0.000	.0595072	.118872

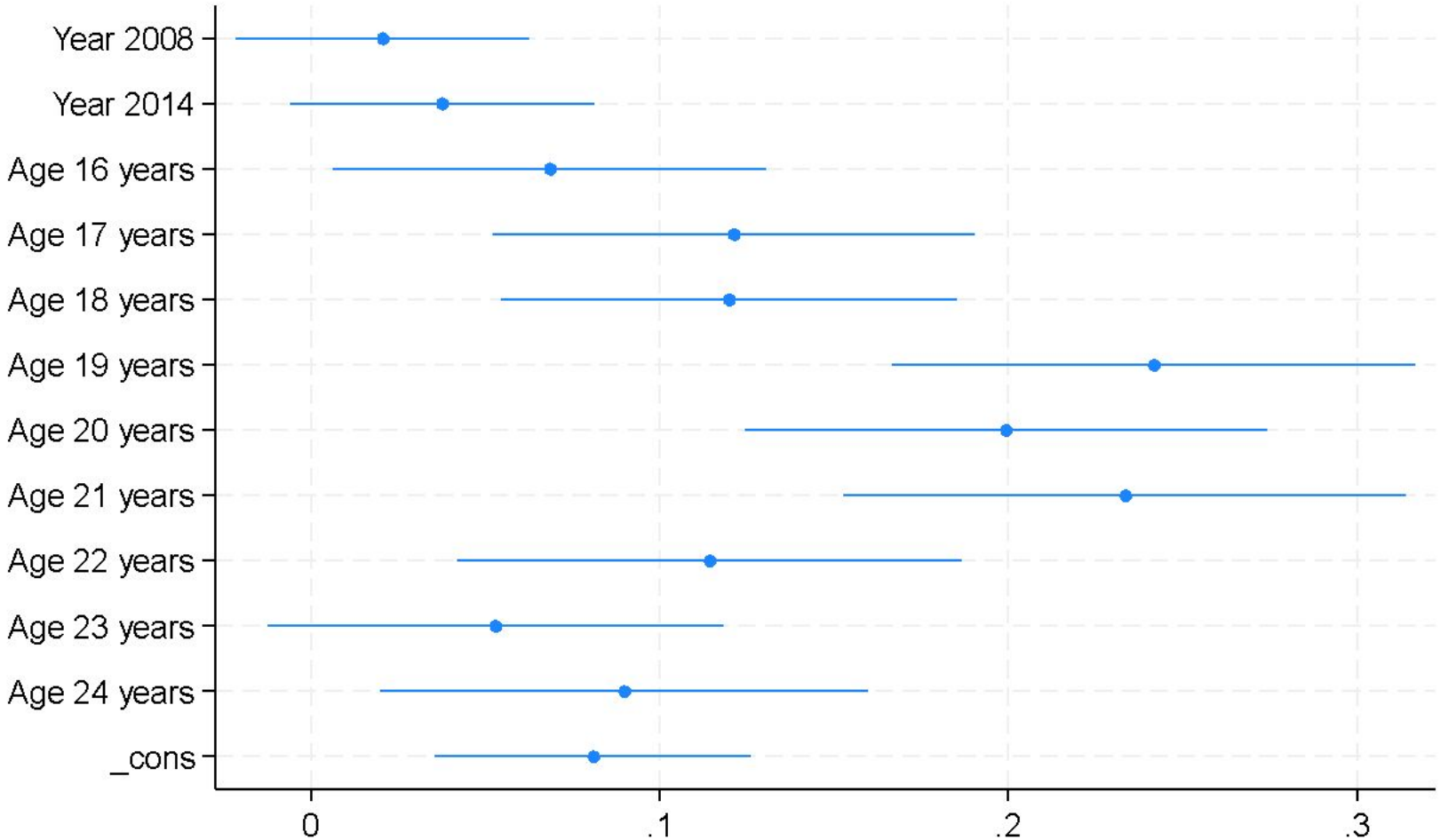
Plotting coefficients from single regression

Urban:

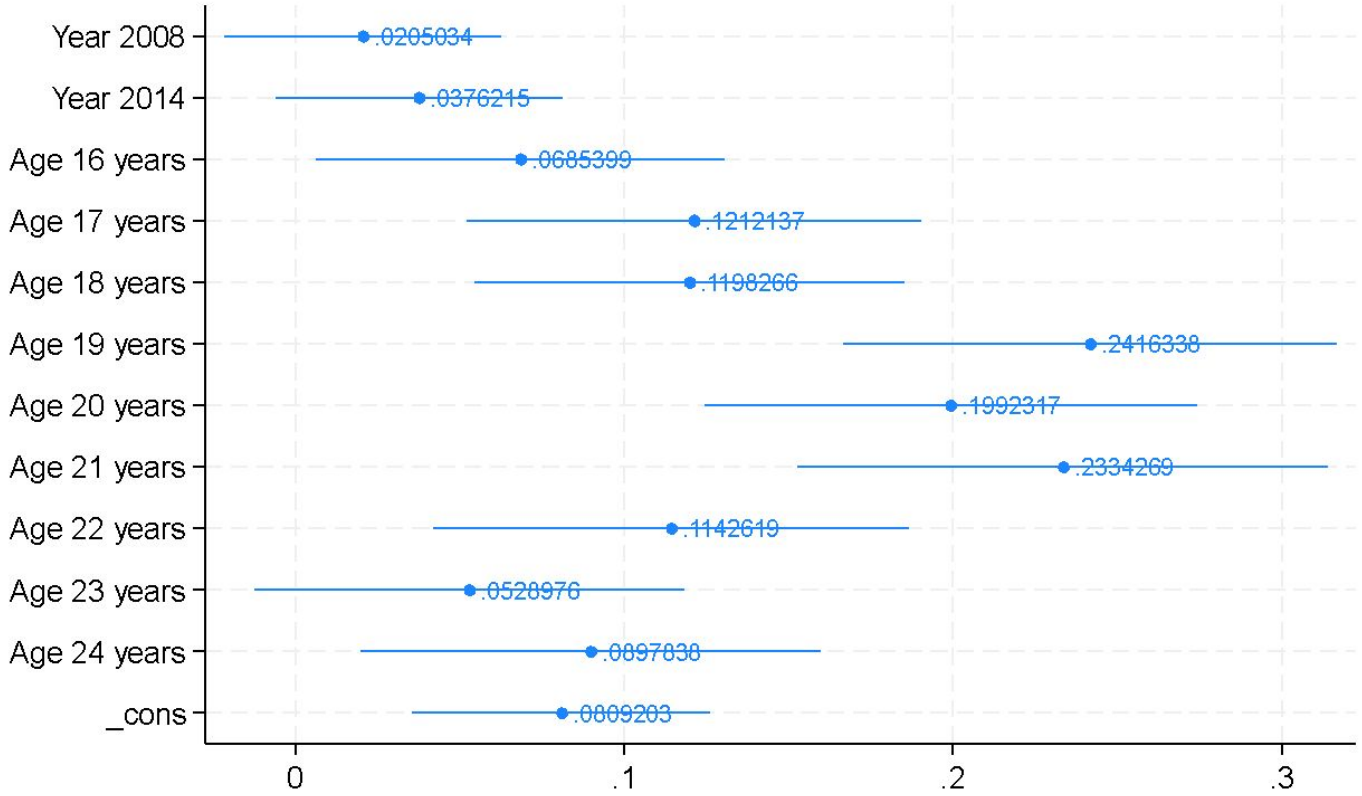
reg neet year2008 year2014 age_2 age_3 age_4 age_5 age_6 age_7 age_8 age_9 age_10 ***if rural ==0****, r*

neet	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
year2008	**.0205034**	**.0215487**	**0.95**	**0.341**	**-.0217556**	**.0627625**
year2014	**.0376215**	**.0222254**	**1.69**	**0.091**	**-.0059646**	**.0812077**
age_2	.0685399	.0317051	2.16	0.031	.0063631	.1307168
age_3	.1212137	.0353165	3.43	0.001	.0519545	.1904728
age_4	.1198266	.0333794	3.59	0.000	.0543662	.1852869
age_5	.2416338	.0382977	6.31	0.000	.1665282	.3167395
age_6	.1992317	.038176	5.22	0.000	.1243649	.2740986
age_7	.2334269	.0411552	5.67	0.000	.1527174	.3141364
age_8	.1142619	.0368932	3.10	0.002	.0419106	.1866132
age_9	.0528976	.0332896	1.59	0.112	-.0123866	.1181817
age_10	.0897838	.0357166	2.51	0.012	.0197399	.1598276
_cons	.0809203	.0231158	3.50	0.000	.0355879	.1262528

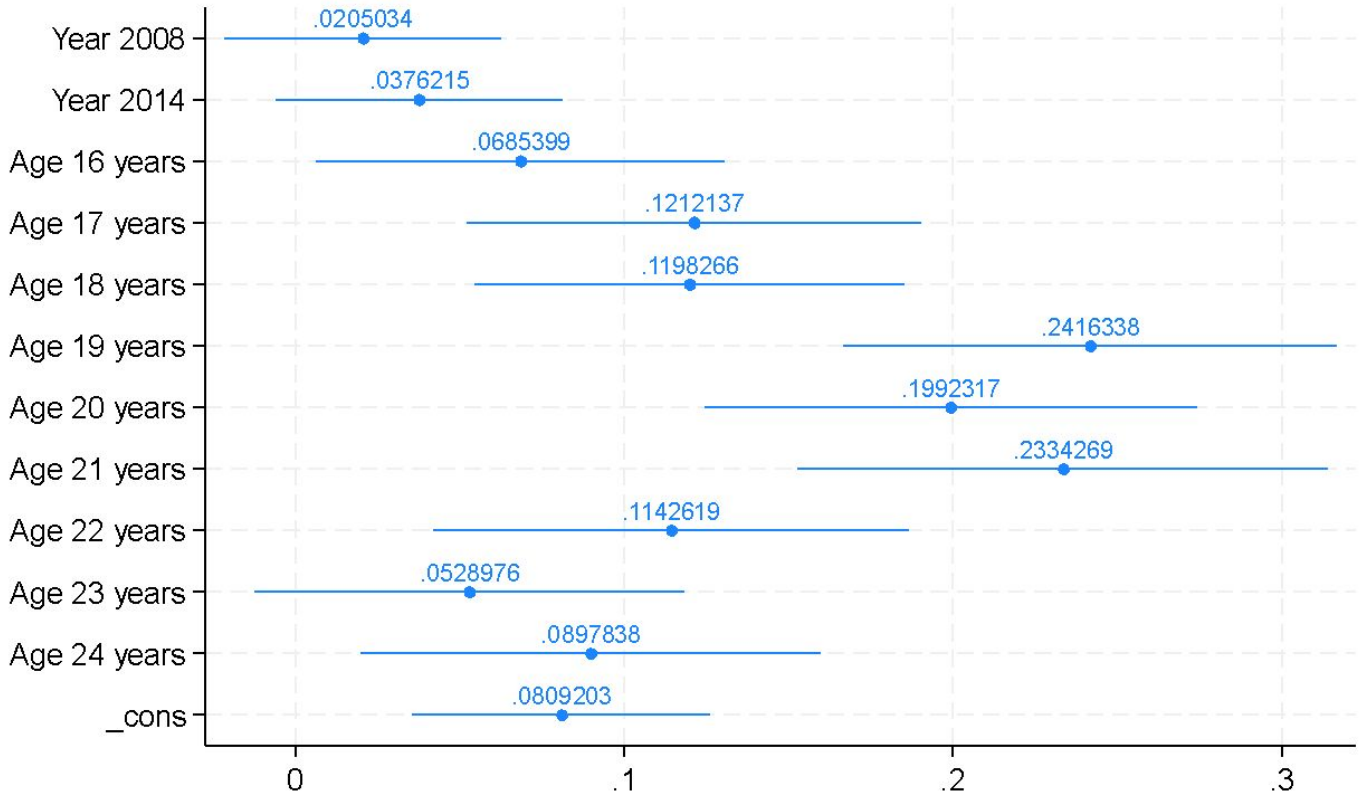
coefplot


Year 2008
Year 2014
Age 16 years
Age 17 years
Age 18 years
Age 19 years
Age 20 years
Age 21 years
Age 22 years
Age 23 years
Age 24 years
_cons
0
.1
.2
.3

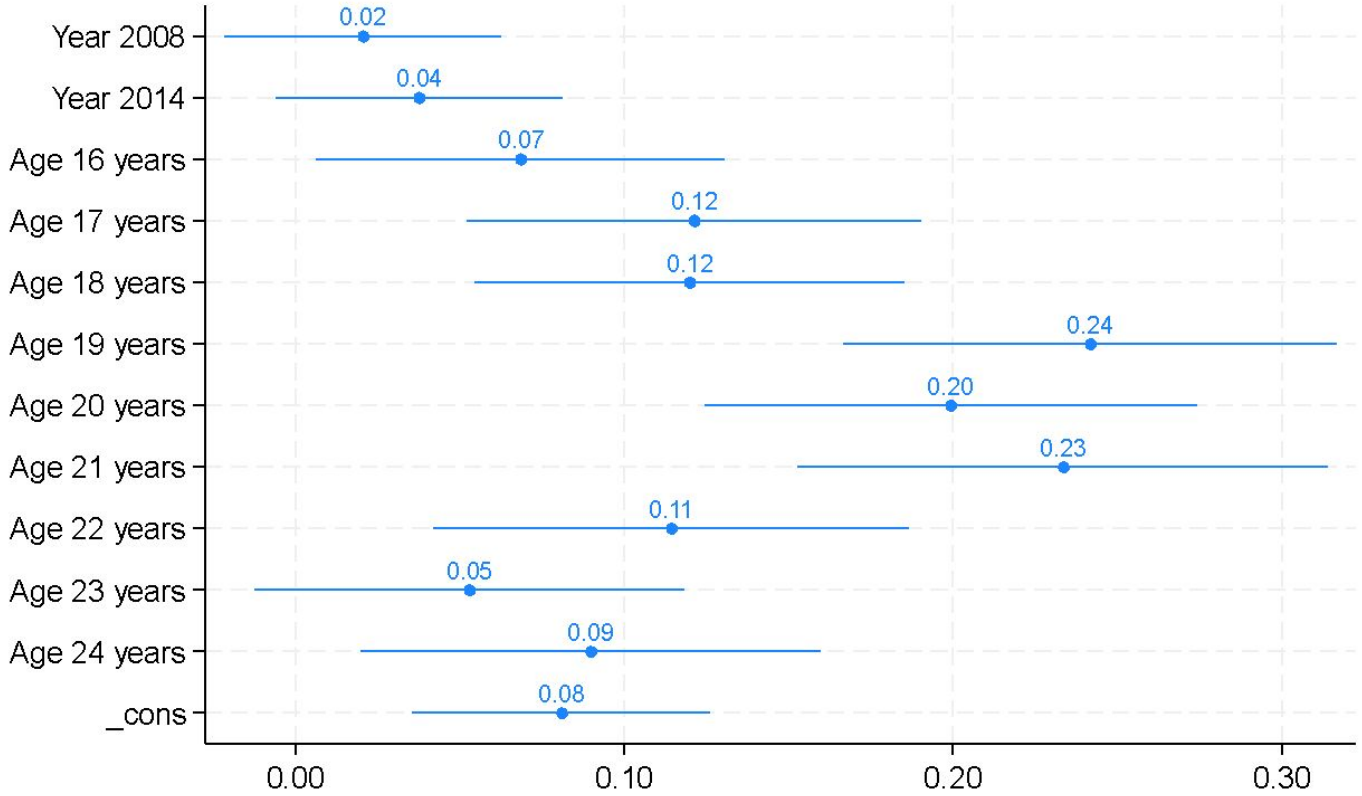
coefplot,mlabel



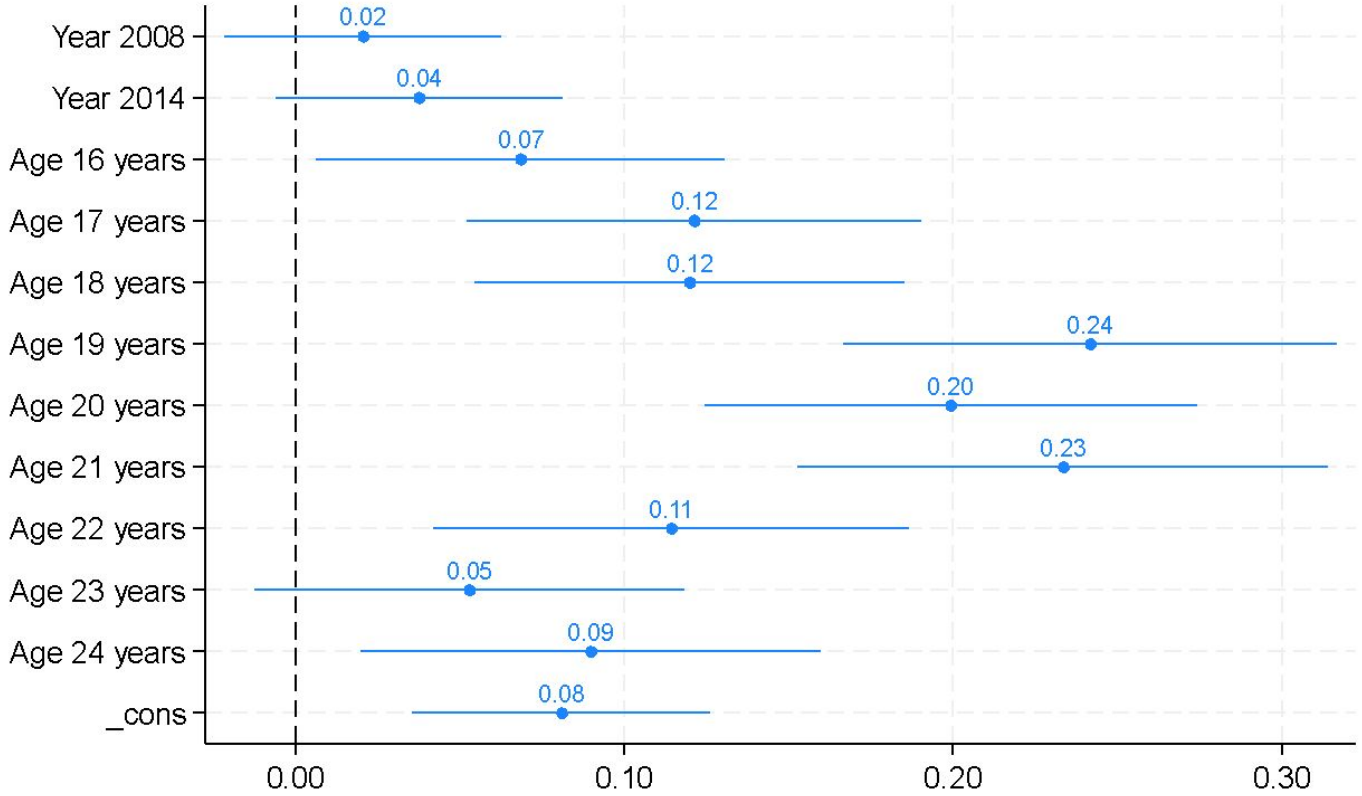
coefplot , mlabel mlabposition(12)



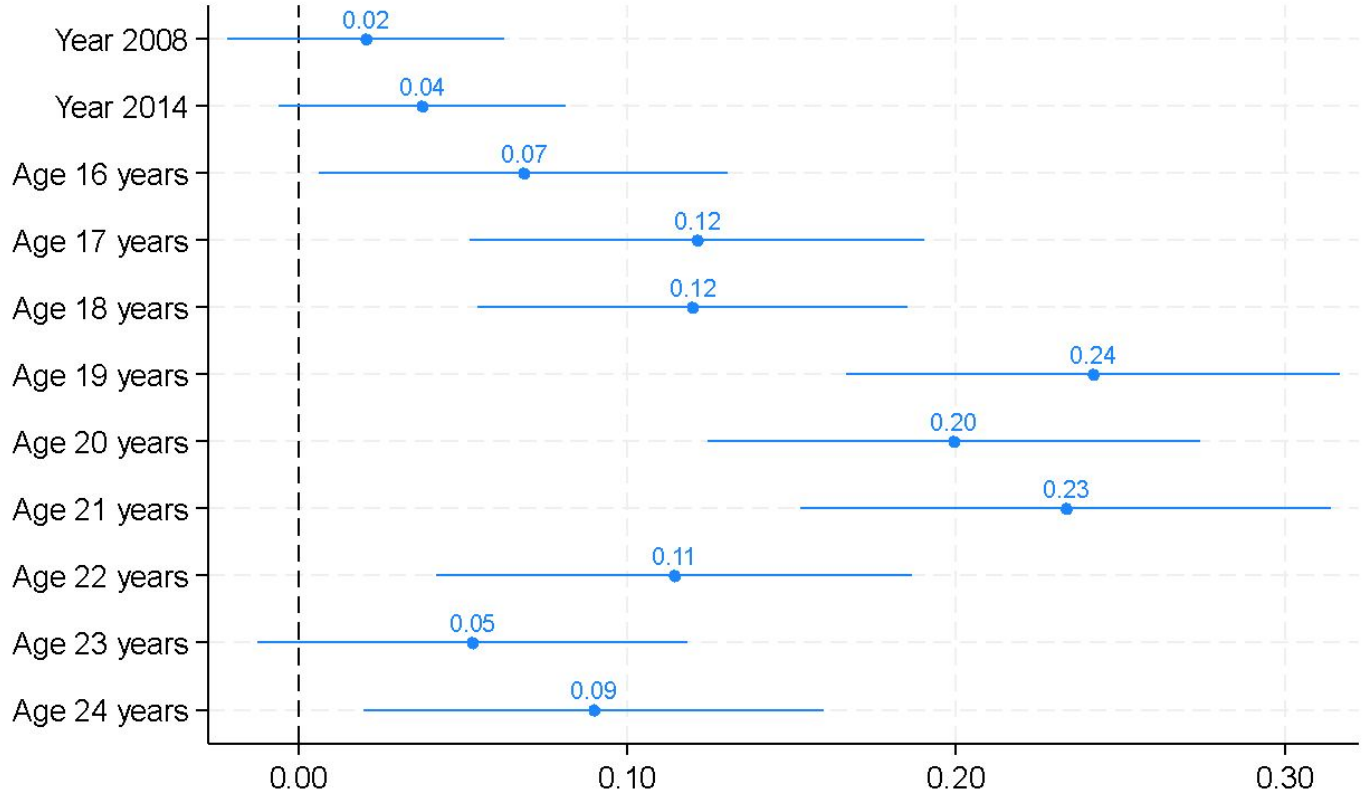
coefplot , mlabel mlabposition(12) format(%9.2f)



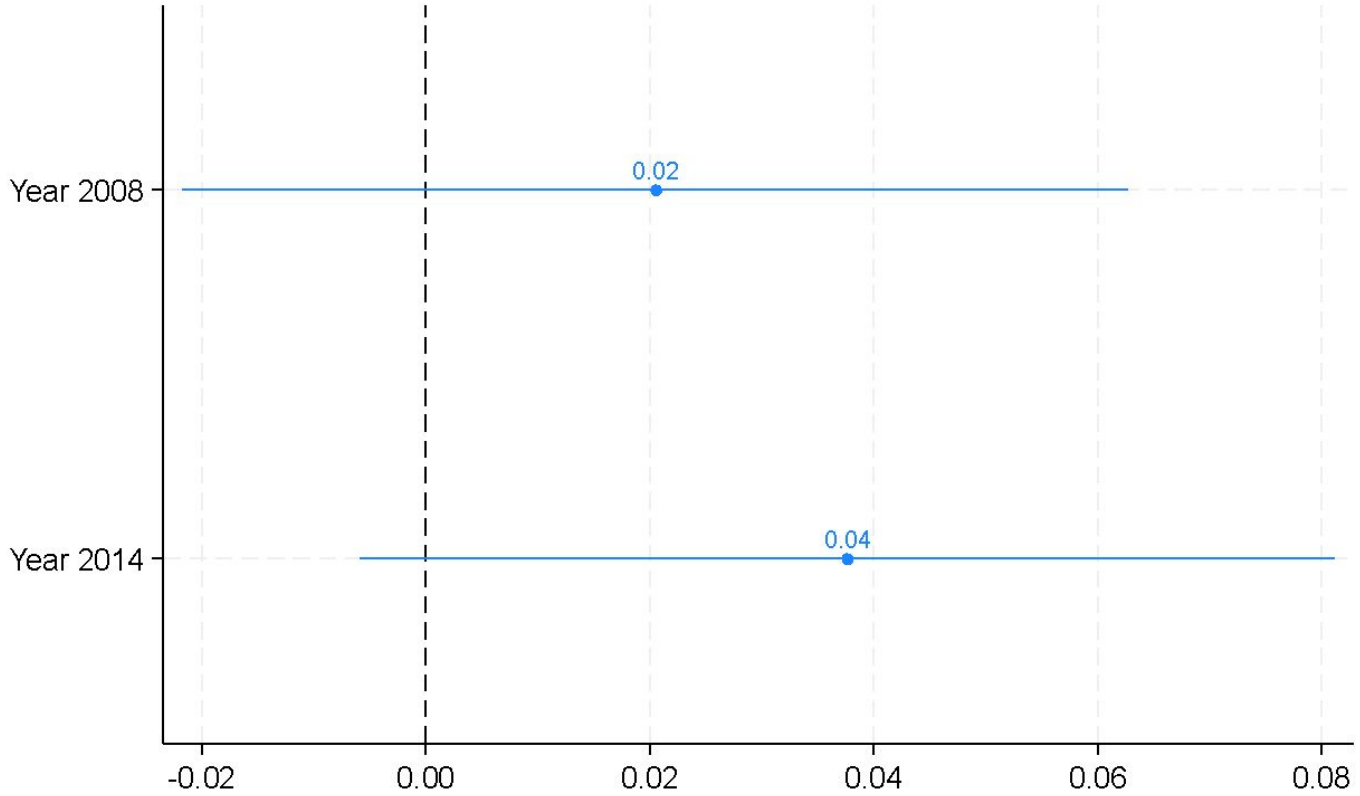
coefplot , mlabel mlabposition(12) format(%9.2f) xline(0)



coefplot , mlabel mlabposition(12) format(%9.2f) xline(0) drop(_cons)



coefplot , mlabel mlabposition(12) format(%9.2f) xline(0) keep(year2008 year2014)



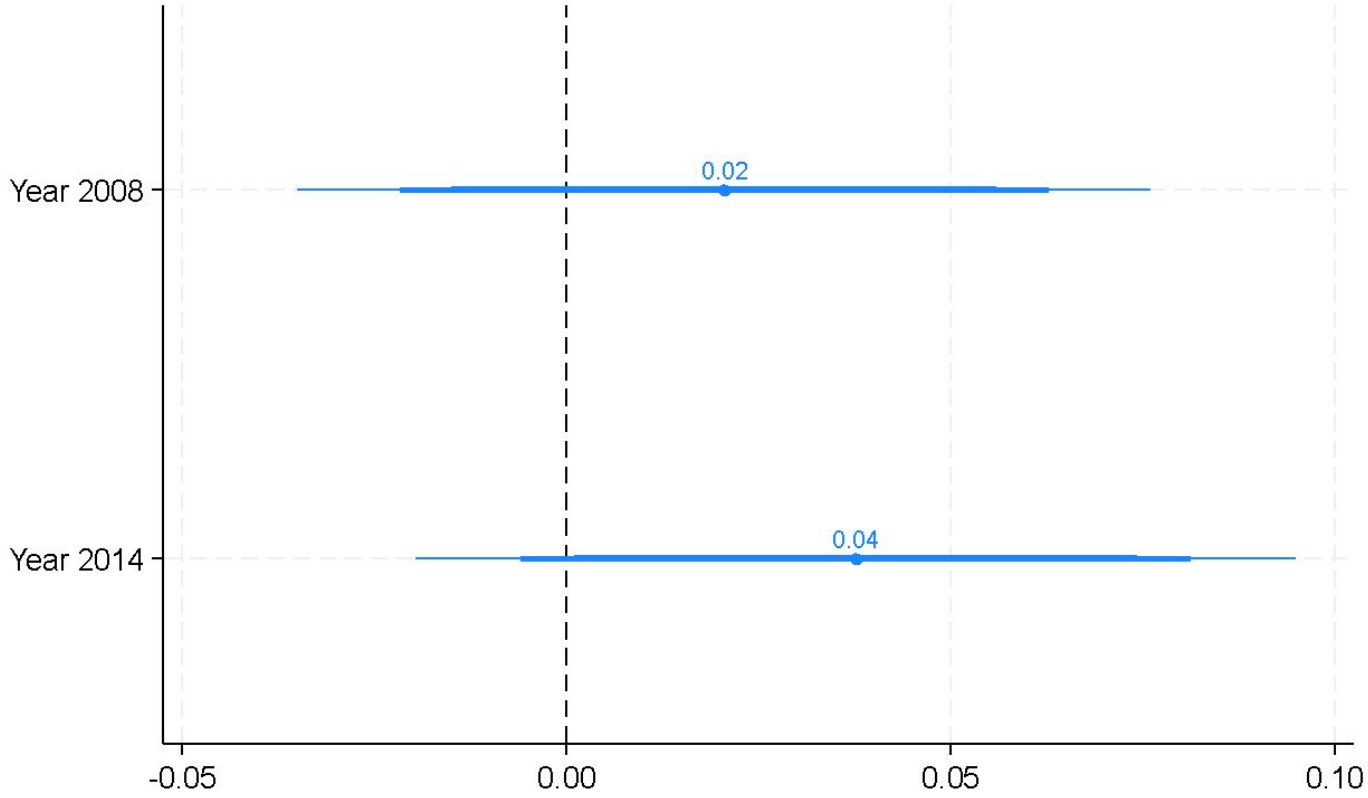
What’s with the horizontal lines on the coefficients?

help coefplot

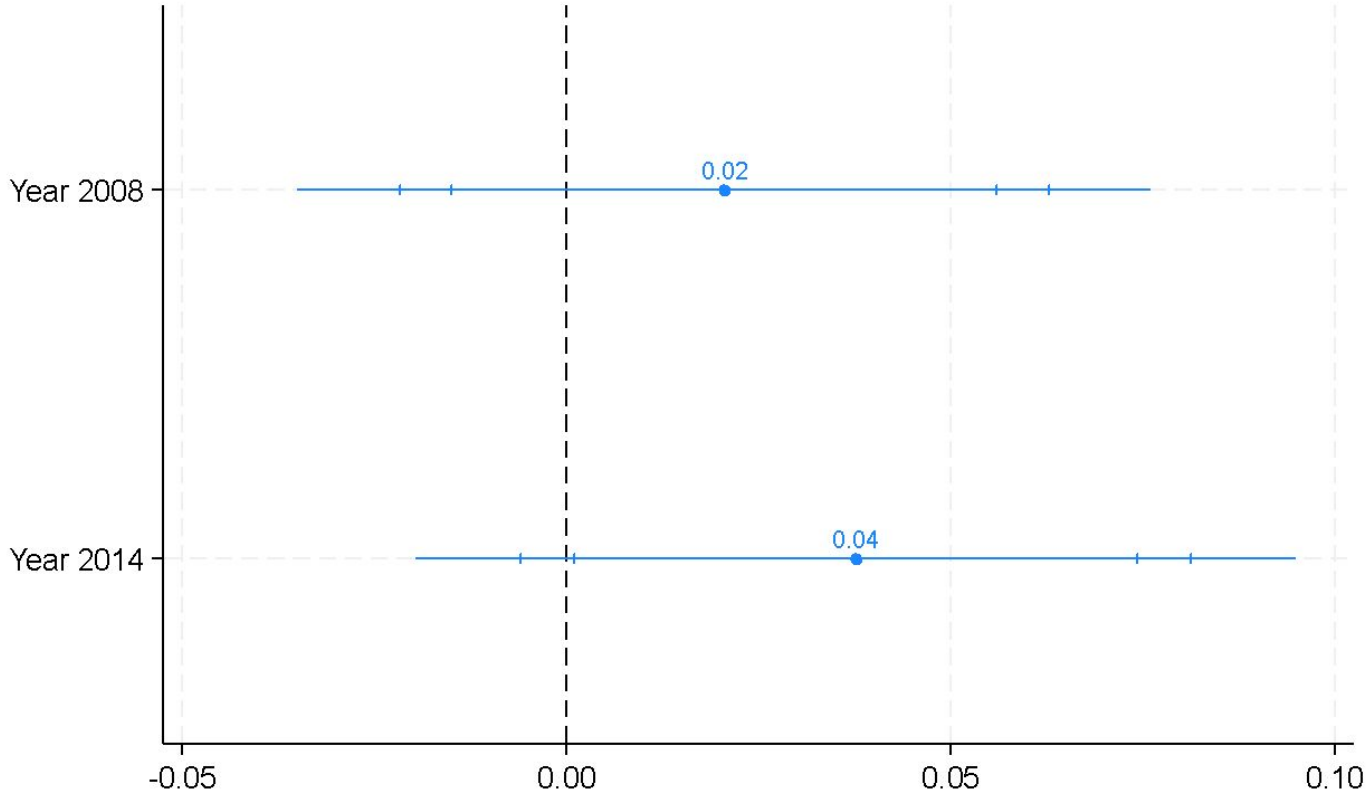
“levels(numlist) sets the level(s), as percentages, for confidence intervals. Specified values may be between 10.00 and 99.99 and can have at most two digits after the decimal point. The default is **levels(95)** or as set by set level. If multiple values are specified, multiple confidence intervals are plotted. For example, type levels(99.9 99 95) to plot the 99.9%, 99%, and 95% confidence intervals...

ci(spec) specifies the source from which to collect confidence intervals. Default is to compute confidence intervals for the levels specified in levels() using variances/standard errors (and, possibly, degrees of freedom). The ci() option is useful to plot confidence intervals that have been provided by the estimation command”

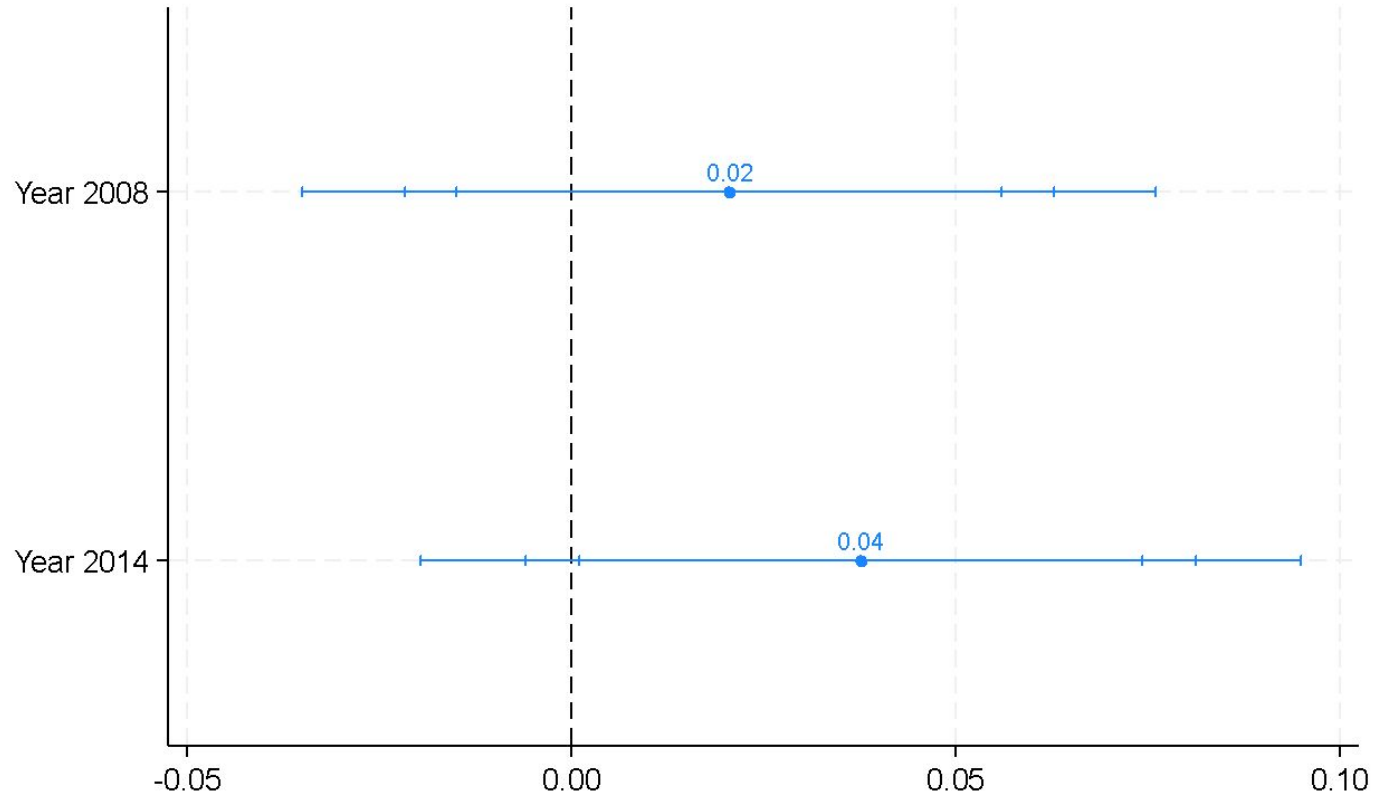
coefplot , mlabel mlabposition(12) format(%9.2f) xline(0) keep(year2008 year2014) ci(99 95 90)



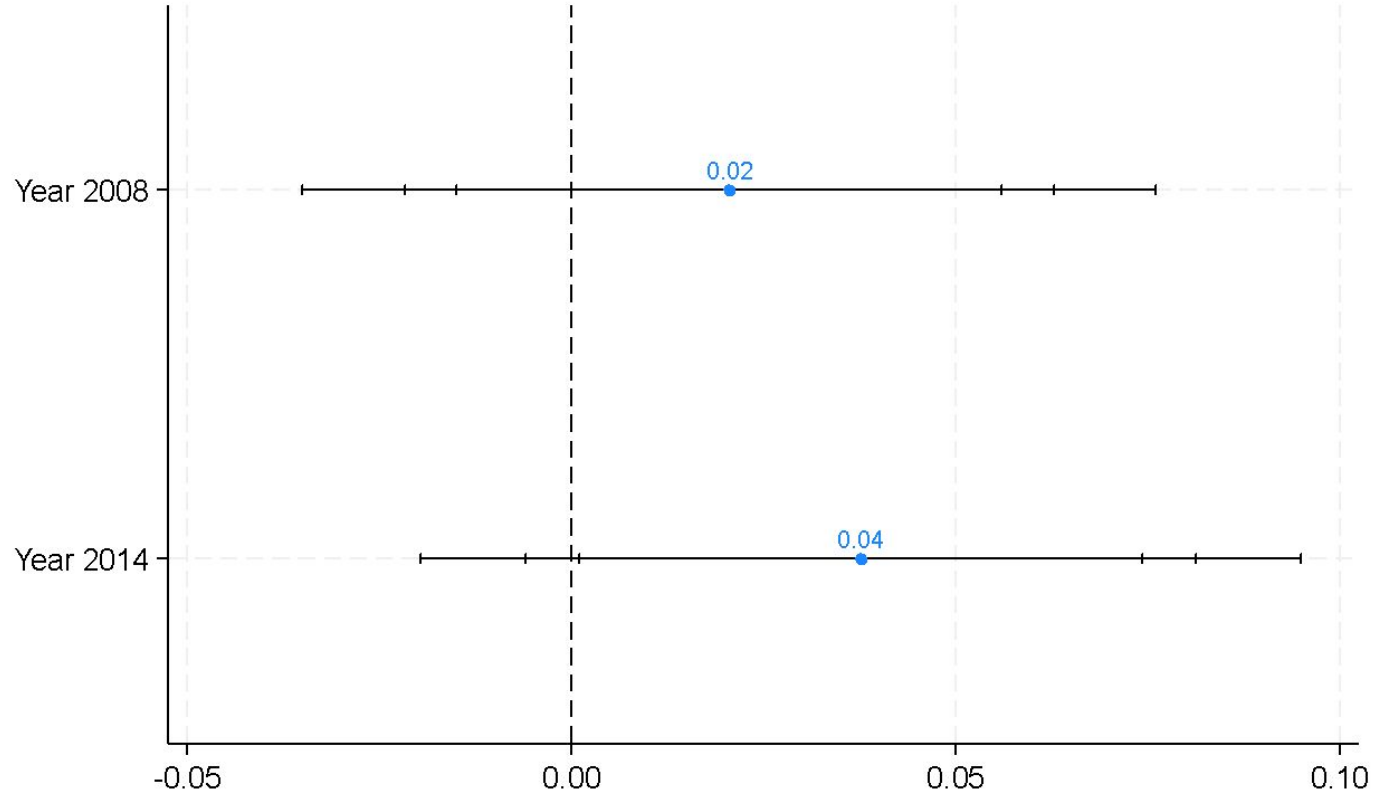
coefplot , mlabel mlabposition(12) format(%9.2f) xline(0) keep(year2008 year2014) ci(99 95 90) ciopts(recast(rspike rcap rcap))



coefplot , mlabel mlabposition(12) format(%9.2f) xline(0) keep(year2008 year2014) ci(99 95 90) ciopts(recast(rcap ..))

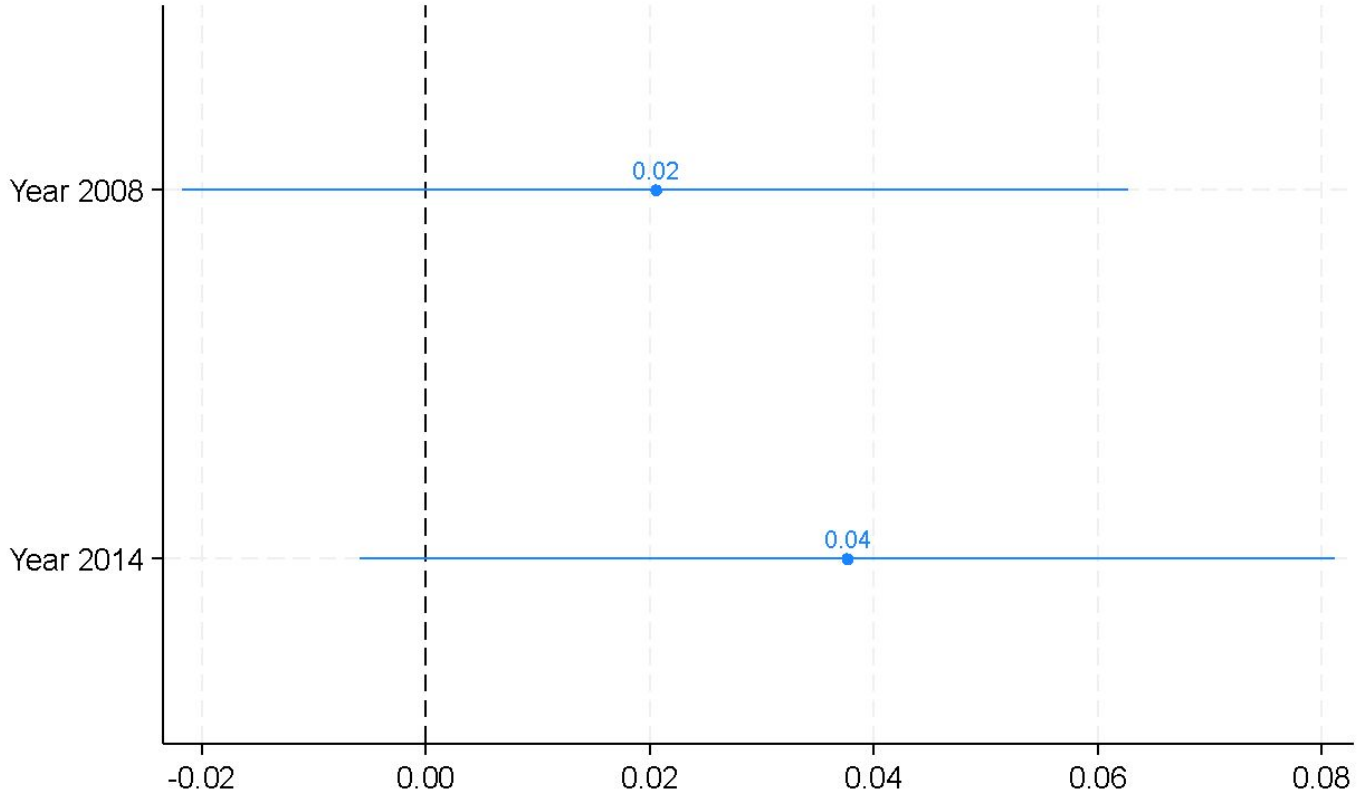


coefplot , mlabel mlabposition(12) format(%9.2f) xline(0) keep(year2008 year2014) ci(99 95 90) ciopts(recast(rcap ..) lc(black ..) lw(medium..))



Back to our default 95% CIs

Urban: coefplot , mlabel mlabposition(12) format(%9.2f) xline(0)



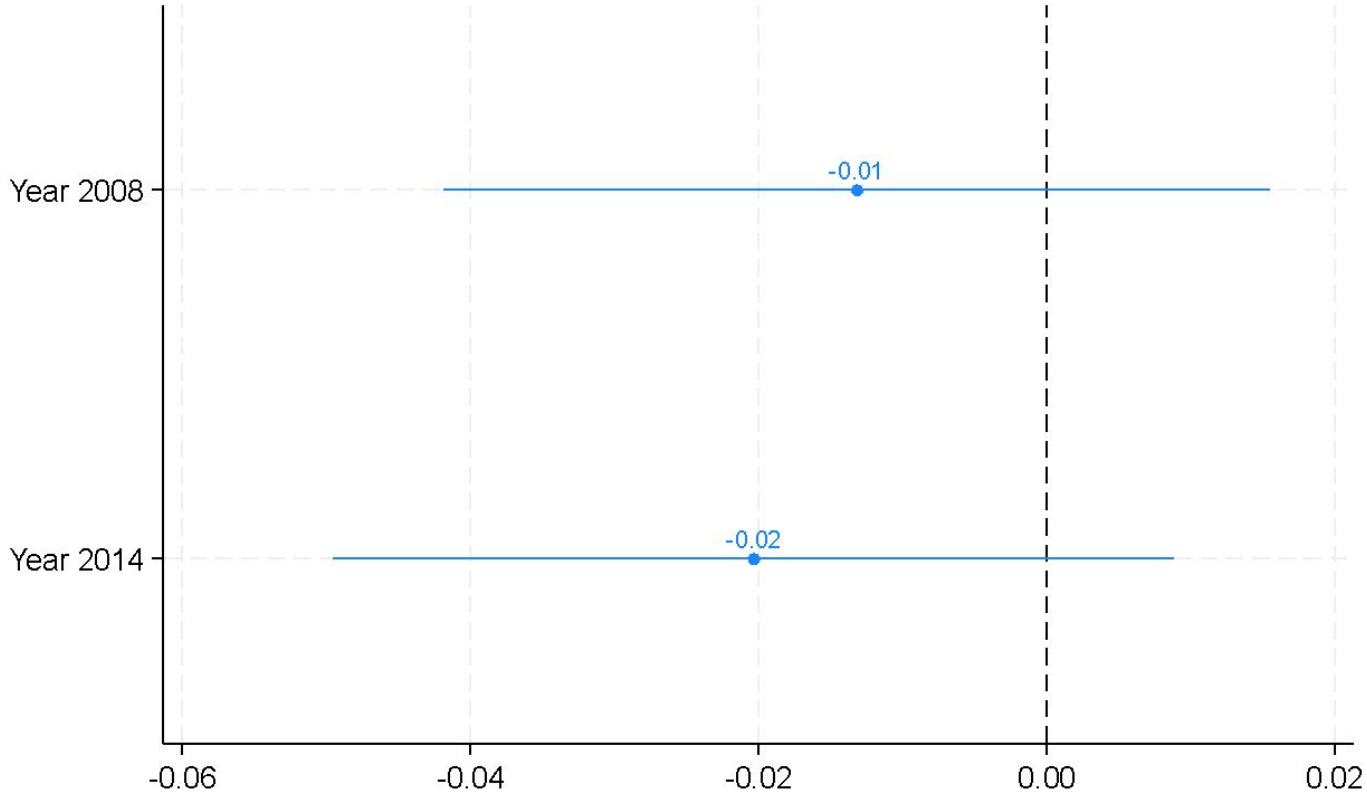
We can run the same regression analysis for rural areas

Rural:

*reg neet year2008 year2014 age_2 age_3 age_4 age_5 age_6 age_7 age_8 age_9 age_10 if **rural ==1**, r*

neet	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
year2008	**-.0132159**	**.01463**	**-0.90**	**0.366**	**-.0419025**	**.0154707**
year2014	**-.020363**	**.0149014**	**-1.37**	**0.172**	**-.0495817**	**.0088558**
age_2	.0510082	.021771	2.34	0.019	.0083196	.0936967
age_3	.0277575	.0223311	1.24	0.214	-.0160294	.0715444
age_4	.0975914	.0234785	4.16	0.000	.0515548	.1436281
age_5	.068804	.0251296	2.74	0.006	.0195299	.1180781
age_6	.0513173	.0246995	2.08	0.038	.0028864	.0997482
age_7	.0190586	.0245162	0.78	0.437	-.0290128	.0671299
age_8	.0046135	.0240916	0.19	0.848	-.0426253	.0518523
age_9	.0289109	.0266241	1.09	0.278	-.0232937	.0811156
age_10	.0100647	.0247859	0.41	0.685	-.0385356	.0586649
_cons	.0891896	.0151379	5.89	0.000	.0595072	.118872

coefplot , mlabel mlabposition(12) format(%9.2f) xline(0) keep(year2008 year2014)



Plot rural and urban models on the same graph

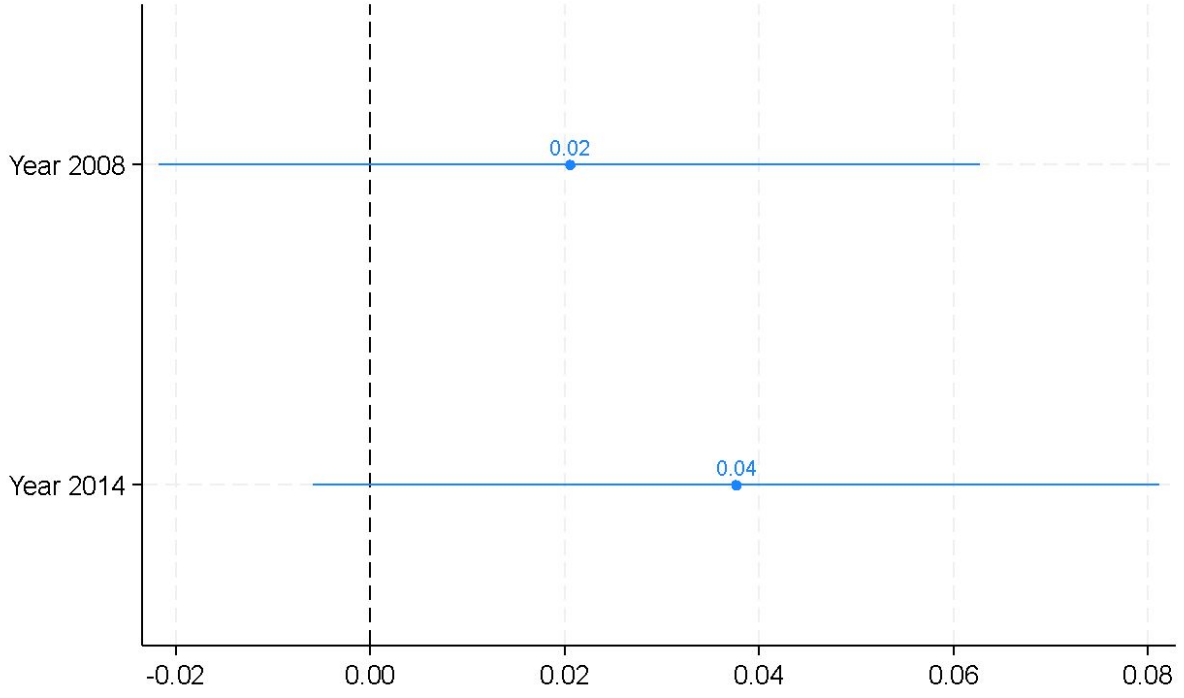
One outcome variable (neet)

Two subgroups (rural and urban)

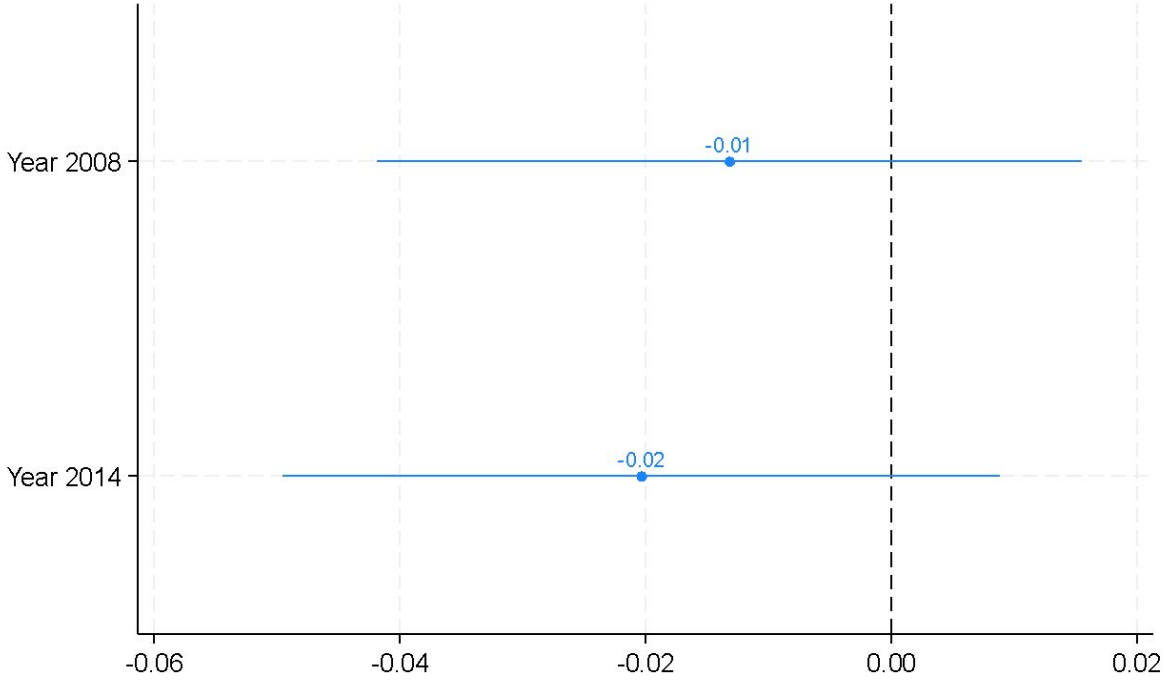
reg neet year2008 year2014 age_2 age_3 age_4 age_5 age_6 age_7 age_8 age_9 age_10 if rural==0, r

reg neet year2008 year2014 age_2 age_3 age_4 age_5 age_6 age_7 age_8 age_9 age_10 if rural==1, r

Urban



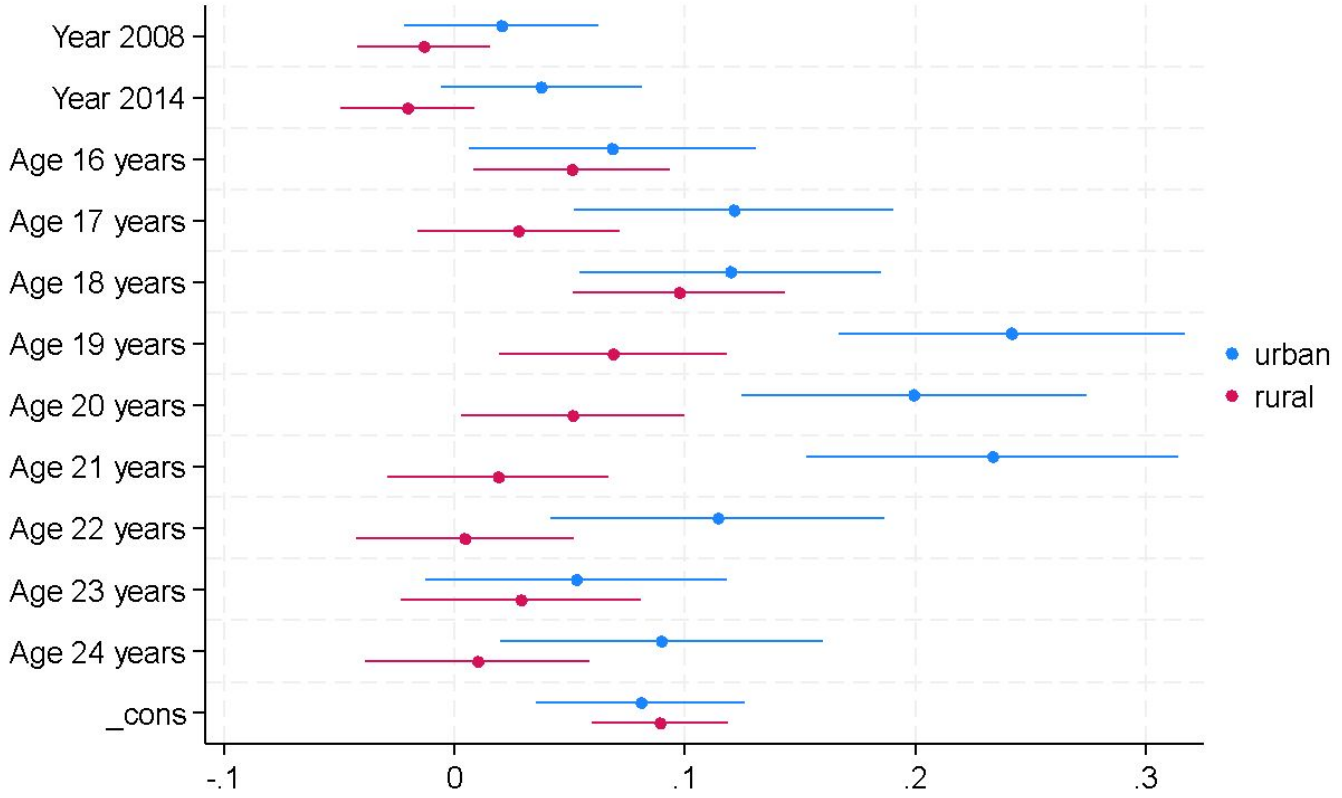
Rural



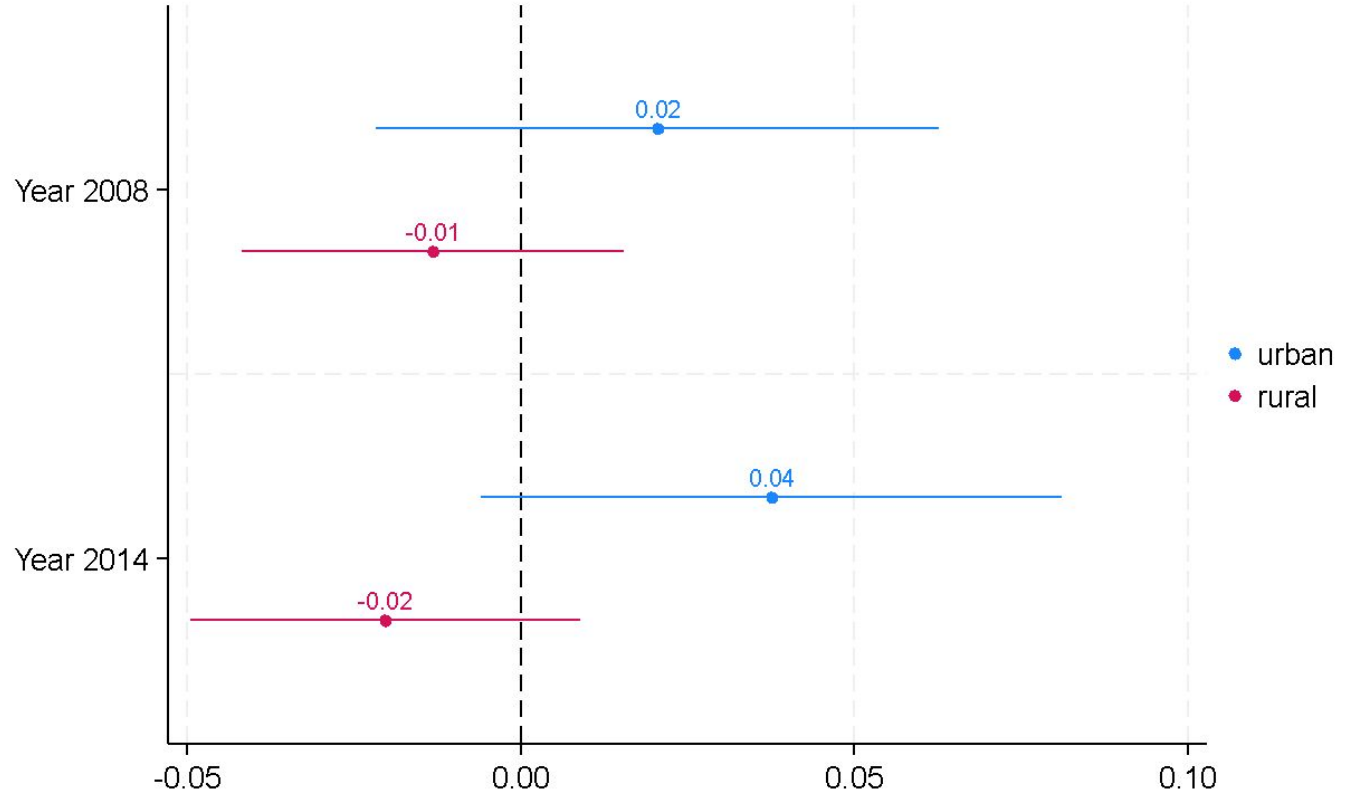
Storing estimation results

- After estimating a model (regression/summary statistics), you can store the estimated results and access them later
  - Stata command: *estimates store* ***NAME***
    - For example:
      - *reg neet year2008 year2014 age_2 age_3 age_4 age_5 age_6 age_7 age_8 age_9 age_10 if rural==0, r*
      - *estimates store urban*
      - *reg neet year2008 year2014 age_2 age_3 age_4 age_5 age_6 age_7 age_8 age_9 age_10 if rural==1, r*
      - *estimates store rural*
- Storing estimates uses up your Stata memory space. Delete your stored estimates when you are done using them to free up space using
  - ***estimates clear***

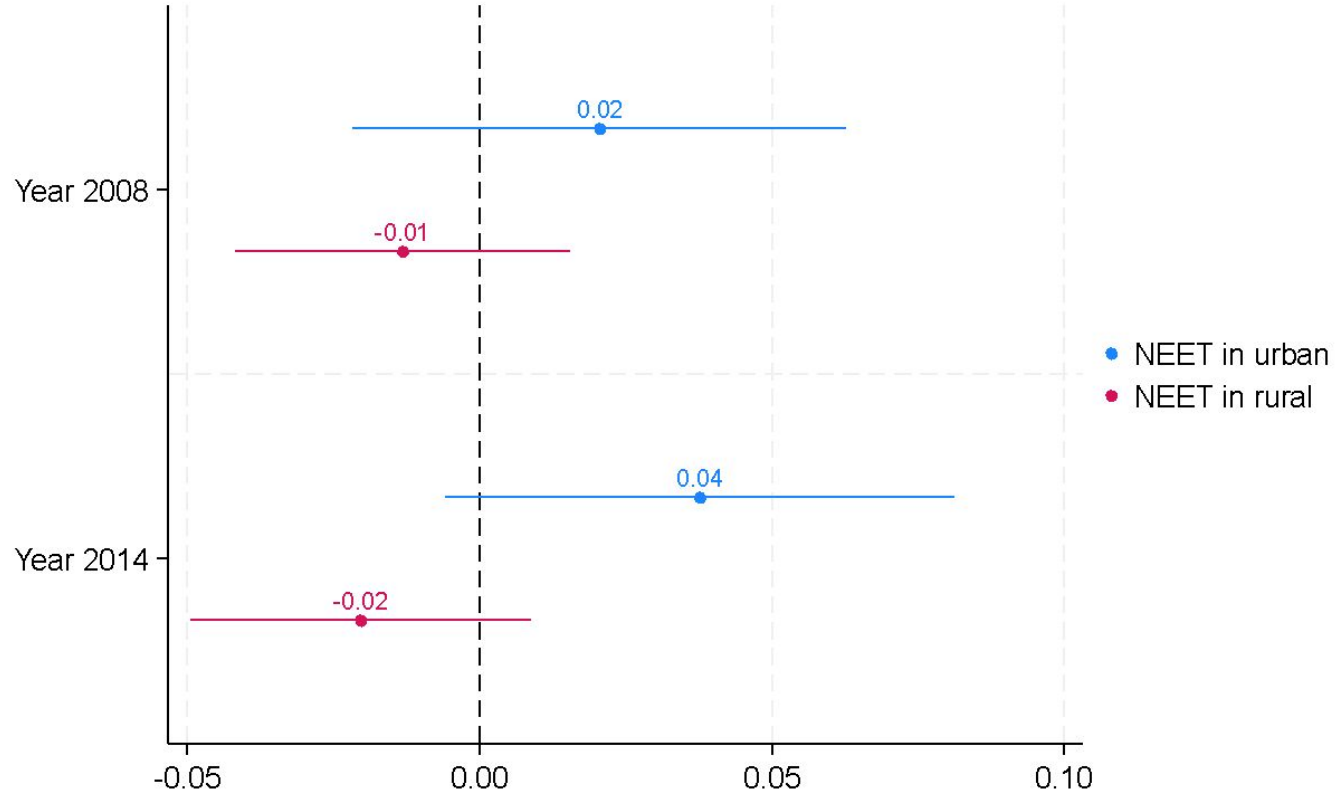
coefplot (urban) (rural)



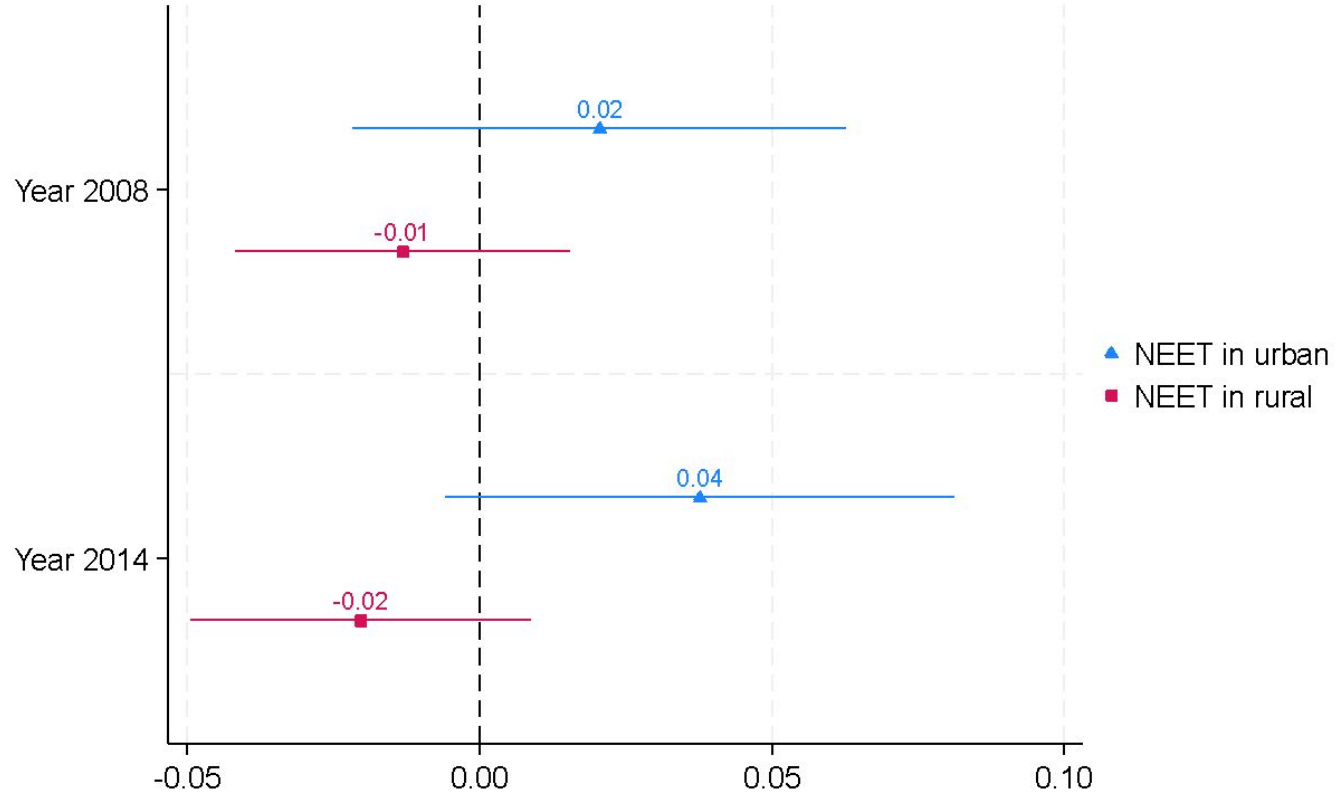
coefplot (urban) (rural) , mlabel mlabposition(12) format(%9.2f) xline(0) keep(year2008 year2014)



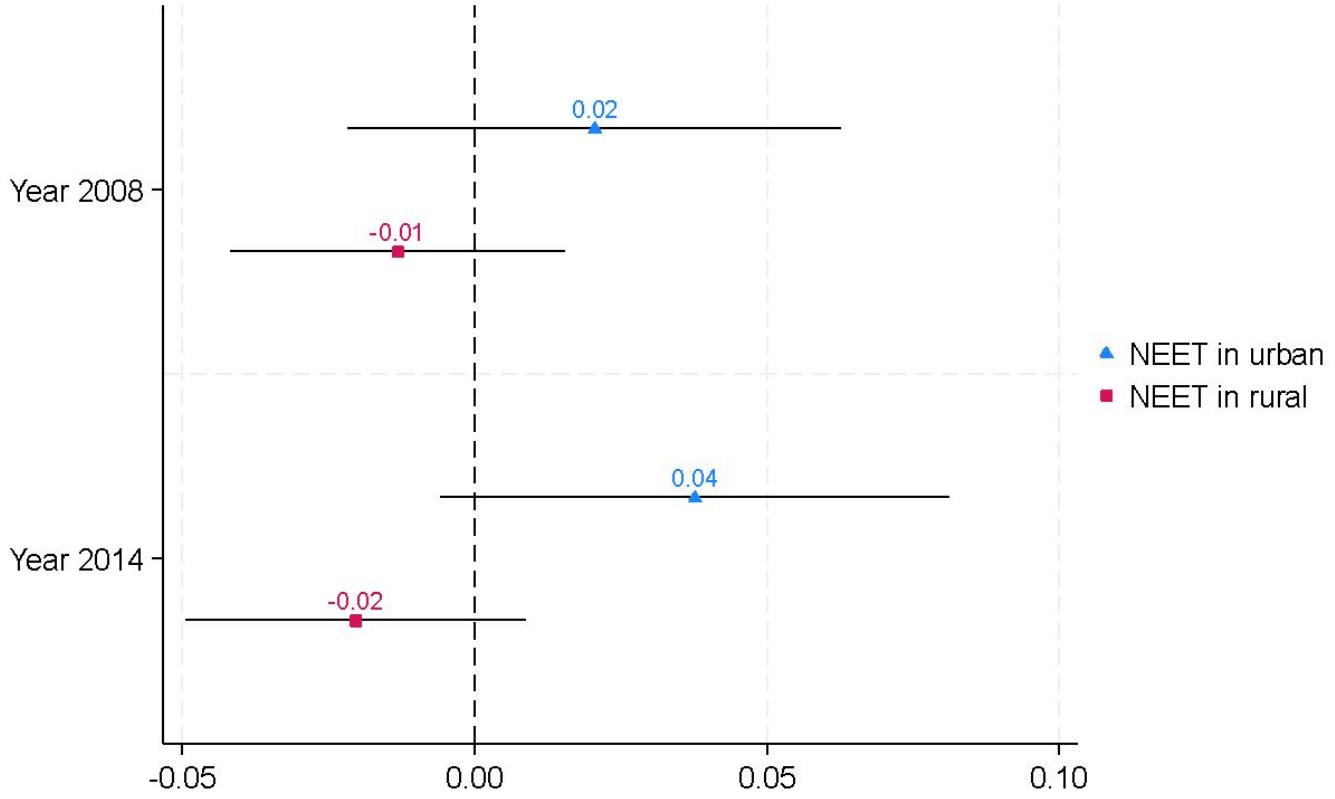
coefplot (urban, label("NEET in urban")) (rural, label("NEET in rural")), mlabel mlabposition(12) format(%9.2f) xline(0) keep(year2008 year2014)



coefplot (urban, label("NEET in urban") msymbol(triangle)) (rural, label("NEET in rural") msymbol(square)), mlabel mlabposition(12) format(%9.2f) xline(0) keep(year2008 year2014)



coefplot (urban, label("NEET in urban") msymbol(triangle)) (rural, label("NEET in rural") msymbol(square)), mlabel mlabposition(12) format(%9.2f) xline(0) keep(year2008 year2014) ciopts(lc(black))



Multiple outcomes with the same coefficients

Two outcomes (NEET and employment)

- NEET (0/1) indicator
  - = 1if NOT in employment, education or training
  - =0 if in employment, education or training

- offfarm (0/1) indicator that is
  - = 1 if employed outside the farm and
  - =0 if employed on the farm

One coefficient

- NEET Rural:
  - *reg **neet** year2008 year2014 age_2 age_3 age_4 age_5 age_6 age_7 age_8 age_9 age_10 if rural==1, r*
- Off-farm Rural:
  - *reg **offfarm** year2008 year2014 age_2 age_3 age_4 age_5 age_6 age_7 age_8 age_9 age_10 if rural==1, r*

NEET Rural: *reg neet year2008 year2014 age_2 age_3 age_4 age_5 age_6 age_7 age_8 age_9 age_10 if rural==1, r*

neet	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
year2008	-.0132159	.01463	-0.90	0.366	-.0419025	.0154707
year2014	-.020363	.0149014	-1.37	0.172	-.0495817	.0088558
age_2	.0510082	.021771	2.34	0.019	.0083196	.0936967
age_3	.0277575	.0223311	1.24	0.214	-.0160294	.0715444
age_4	.0975914	.0234785	4.16	0.000	.0515548	.1436281
age_5	.068804	.0251296	2.74	0.006	.0195299	.1180781
age_6	.0513173	.0246995	2.08	0.038	.0028864	.0997482
age_7	.0190586	.0245162	0.78	0.437	-.0290128	.0671299
age_8	.0046135	.0240916	0.19	0.848	-.0426253	.0518523
age_9	.0289109	.0266241	1.09	0.278	-.0232937	.0811156
age_10	.0100647	.0247859	0.41	0.685	-.0385356	.0586649
_cons	.0891896	.0151379	5.89	0.000	.0595072	.118872

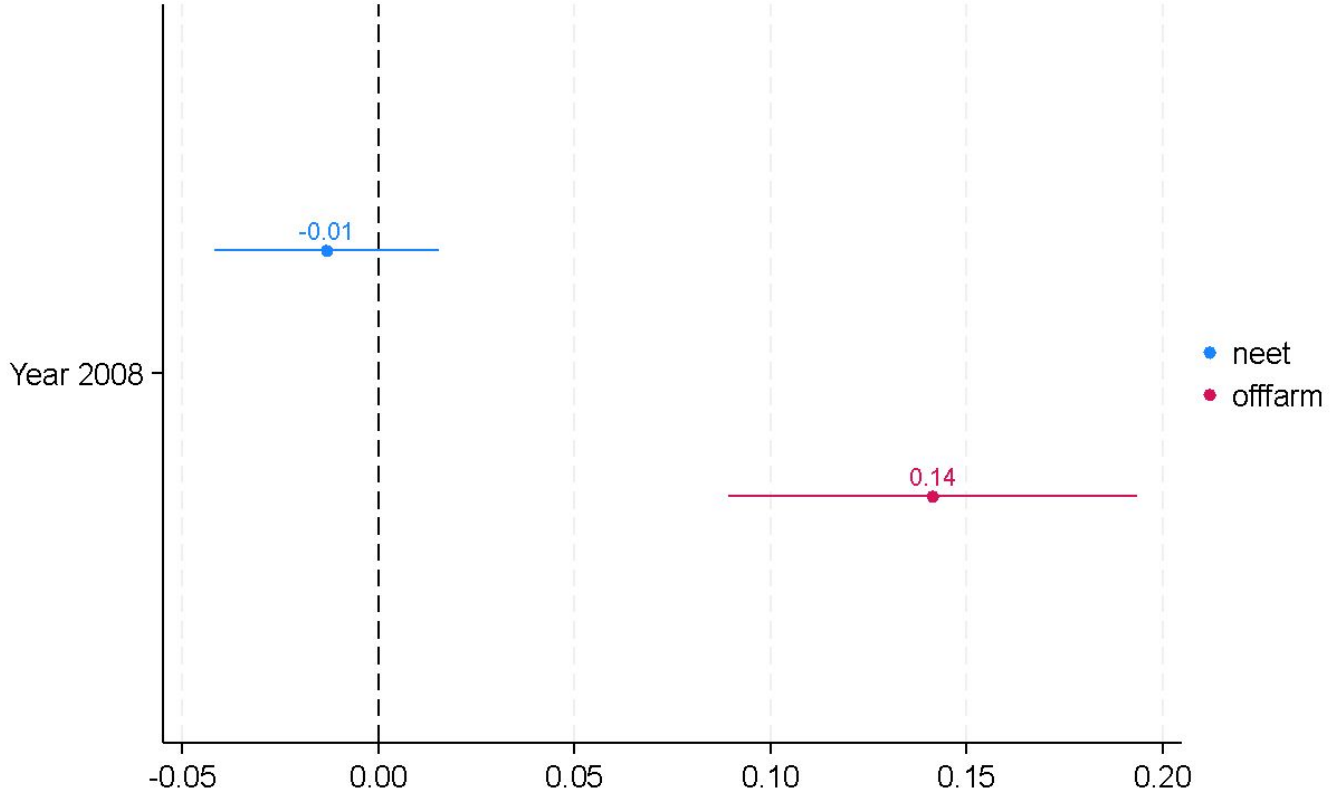
estimates store neet

Off-farm Rural: *reg offfarm year2008 year2014 age_2 age_3 age_4 age_5 age_6 age_7 age_8 age_9 age_10 if rural==1, r*

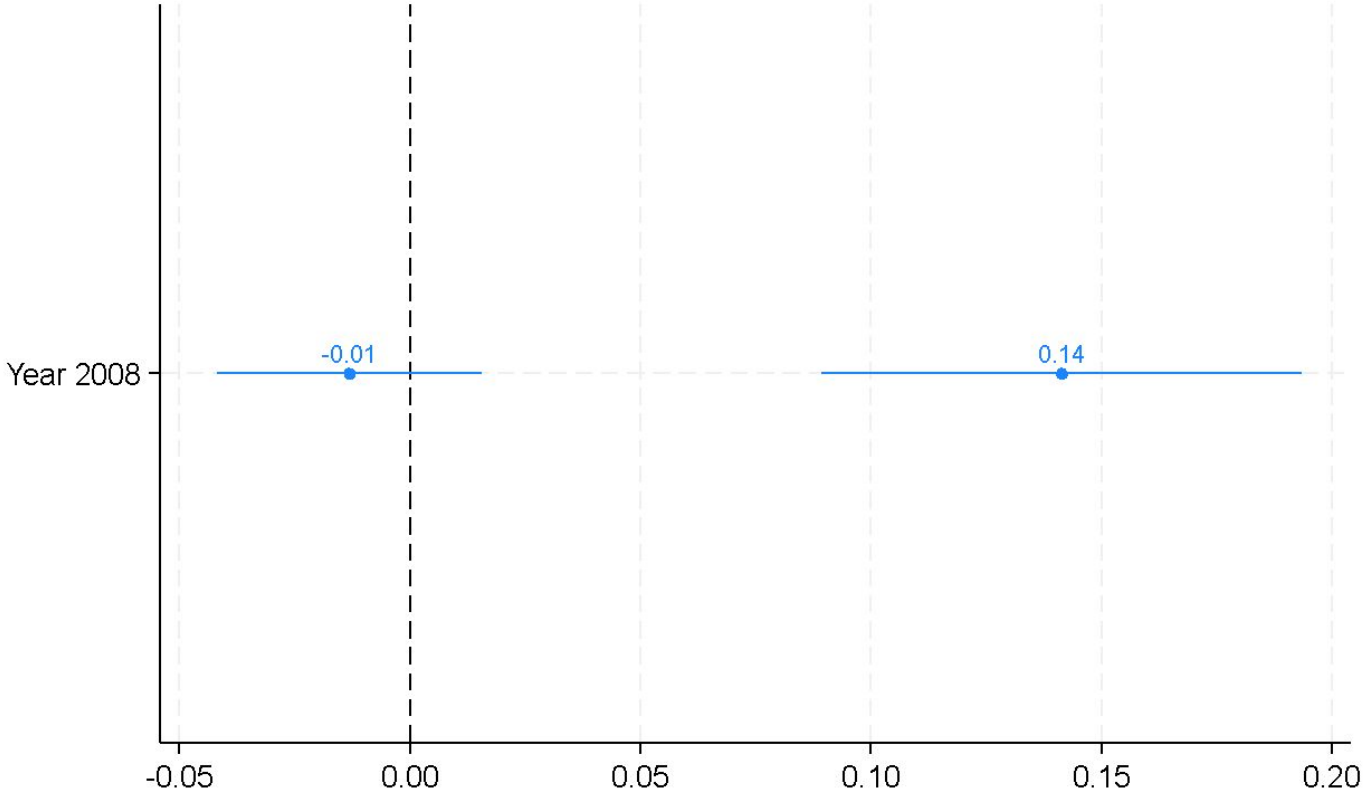
offfarm	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
year2008	.1412723	.0265845	5.31	0.000	.089129	.1934157
year2014	.1353617	.0264465	5.12	0.000	.083489	.1872343
age_2	.0492852	.0482447	1.02	0.307	-.0453431	.1439135
age_3	.0040602	.0483476	0.08	0.933	-.0907698	.0988902
age_4	.0801057	.0453404	1.77	0.077	-.008826	.1690373
age_5	.1215867	.0494473	2.46	0.014	.0245997	.2185737
age_6	.0211714	.0448207	0.47	0.637	-.066741	.1090837
age_7	.1751065	.0510227	3.43	0.001	.0750295	.2751835
age_8	.1366944	.0496191	2.75	0.006	.0393704	.2340185
age_9	.220034	.051742	4.25	0.000	.1185461	.3215218
age_10	.1901072	.0506619	3.75	0.000	.0907379	.2894766
_cons	.106935	.0356311	3.00	0.003	.0370474	.1768226

estimates store offfarm

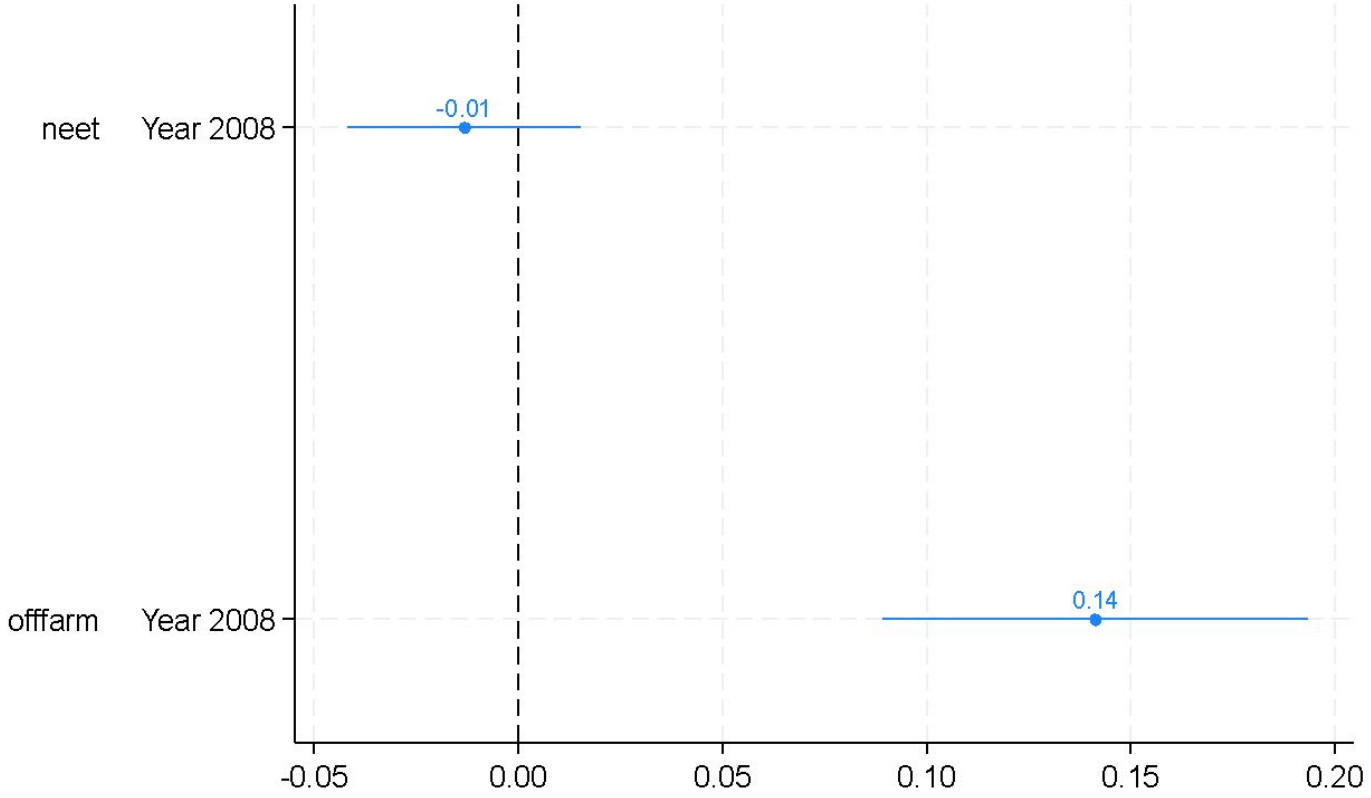
coefplot (neet) (offfarm), mlabel mlabposition(12) format(%9.2f) xline(0) keep(year2008)



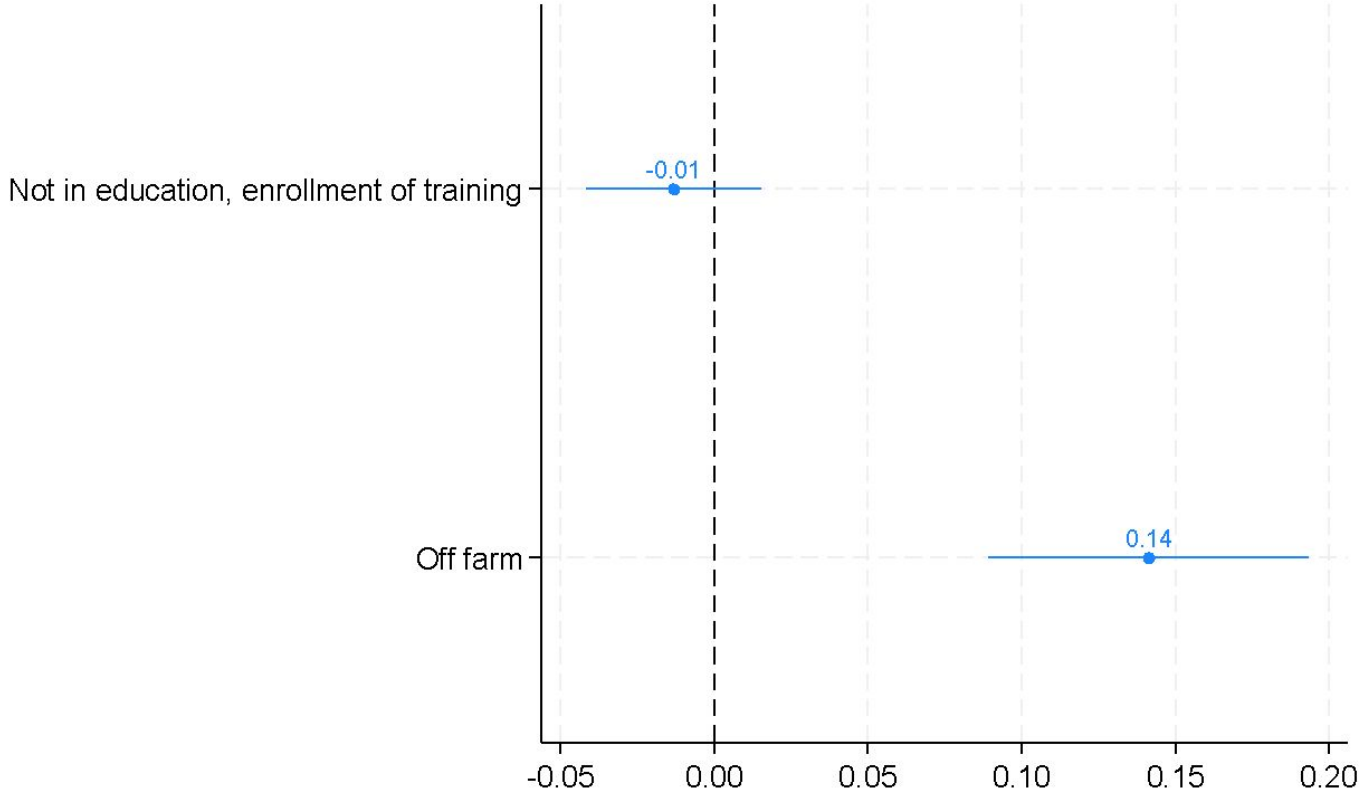
coefplot (neet offfarm) , mlabel mlabposition(12) format(%9.2f) xline(0) keep(year2008)



coefplot (neet offfarm), mlabel mlabposition(12) format(%9.2f) xline(0) keep(year2008) asequation



coefplot (neet offfarm), mlabel mlabposition(12) format(%9.2f) xline(0) keep(year2008) asequation swapname



Adding a third outcome

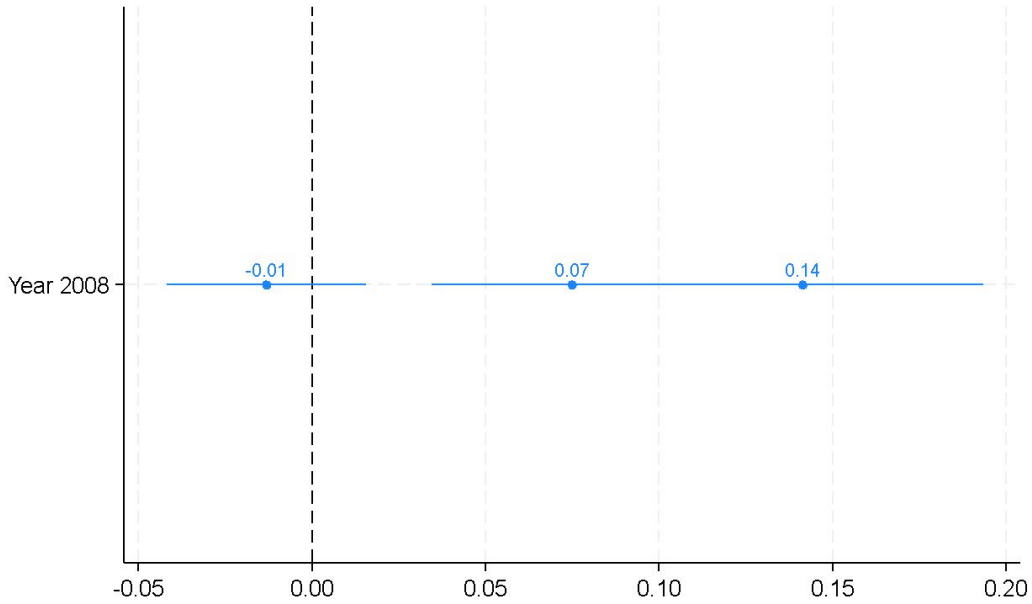
reg employed year2008 year2014 age_2 age_3 age_4 age_5 age_6 age_7 age_8 age_9 age_10 if rural==1, r

employed	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
year2008	.0748356	.0205892	3.63	0.000	.0344643	.1152068
year2014	.1307284	.0210776	6.20	0.000	.0893993	.1720574
age_2	.0301977	.0338482	0.89	0.372	-.036172	.0965674
age_3	.0840416	.0368515	2.28	0.023	.0117831	.1563
age_4	.1625156	.0345181	4.71	0.000	.0948324	.2301988
age_5	.2378392	.0373011	6.38	0.000	.164699	.3109793
age_6	.3596835	.0365082	9.85	0.000	.2880981	.4312689
age_7	.3820857	.0386875	9.88	0.000	.3062272	.4579442
age_8	.4787038	.0357186	13.40	0.000	.4086667	.5487409
age_9	.4872589	.0362585	13.44	0.000	.4161632	.5583546
age_10	.5062429	.034975	14.47	0.000	.4376639	.5748219
_cons	.2901619	.0257977	11.25	0.000	.2395776	.3407462

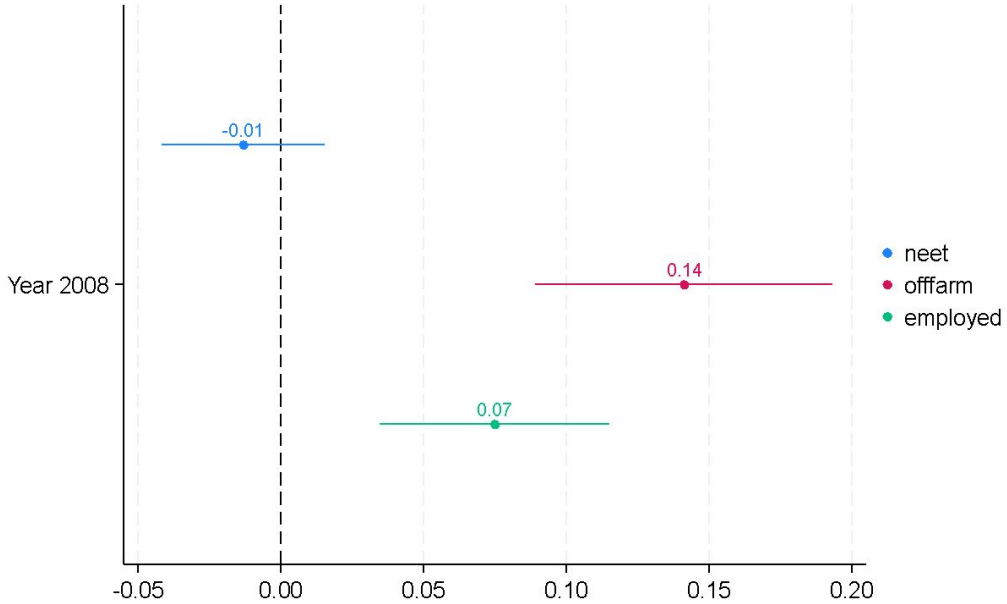
estimate store employed

Using options '*asequation*' and '*swapnames*' produce nicer graphs when you have multiple outcomes

coefplot (neet offfarm employed) , mlabel mlabposition(12) format(%9.2f) xline(0) keep(year2008)

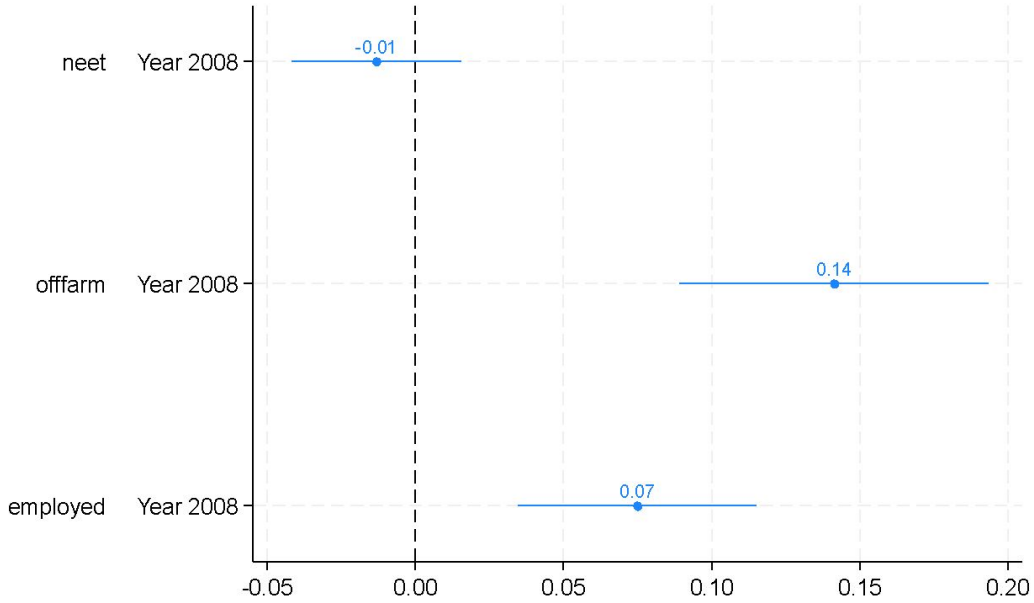


coefplot neet offfarm employed , mlabel mlabposition(12) format(%9.2f) xline(0) keep(year2008)



Nicer graph by swapping equation names with coefficient names

coefplot (neet offfarm employed) , mlabel mlabposition(12) format(%9.2f) xline(0) keep(year2008) asequation



coefplot (neet offfarm employed) , mlabel mlabposition(12) format(%9.2f) xline(0) keep(year2008) asequation swapnames