Decomposition methods in the social sciences

Bamberg Graduate School of Social Sciences, June 7–8, 2018

Ben Jann

University of Bern, Institut of Sociology

Postestimation tools

Contents

Post-estimation commands

- Similar to other estimation commands in Stata, `oaxaca` leaves results behind in `e(b)` and `e(V)` so that they can be processed by post-estimation commands.
- Examples are:
  - Command `test` and `testnl` to perform hypothesis tests.
  - Commands `lincom` and `nlcom` to compute linear and non-linear combinations (and the corresponding standard errors).
  - Commands such as `esttab` (Jann 2007) and `coefplot` (Jann 2014) to make tables and graphs from results.
- For many of these commands it is important to know how the elements in `e(b)` are named. Type

  ```
  . ereturn display, coeflegend
  ```

  after running `oaxaca` to display the names.

Hypothesis tests

- In its standard output, `oaxaca` displays tests of the individual components against zero.
- Depending on context, tests against other values might be required and you might also want to perform joint tests of multiple hypotheses.
- A general command to perform so-called Wald tests of simple and composite linear hypotheses, is `test`. A command for nonlinear hypotheses is `testnl`.

Hypothesis tests

```
. oaxaca lnwage schooling ft_experience ft_experience2, by(bcsex) weight(1)
  (output omitted)
. ereturn display, coeflegend
```

lnwage	Coef.	Legend
overall		
group_1	2.749054	_b[overall:group_1]
group_2	2.498484	_b[overall:group_2]
difference	.2505696	_b[overall:difference]
explained	.1492473	_b[overall:explained]
unexplained	.1013223	_b[overall:unexplained]
explained		
schooling	-.008201	_b[explained:schooling]
ft_experience	.3672357	_b[explained:ft_experience]
ft_experience2	-.2097875	_b[explained:ft_experience2]
unexplained		
schooling	.0852652	_b[unexplained:schooling]
ft_experience	.1784167	_b[unexplained:ft_experience]
ft_experience2	-.050762	_b[unexplained:ft_experience2]
_cons	-.1115975	_b[unexplained:_cons]

Examples

- Test that the explained part is different from the unexplained part:

```
. test _b[overall:explained] = _b[overall:unexplained]
 ( 1)  [overall]explained - [overall]unexplained = 0
           chi2(  1) =    6.55
         Prob > chi2 =    0.0105
```

- Joint test of the contributions of `ft_experience` and `ft_experience2` to the explained part against zero:

```
. test _b[explained:ft_experience] = 0
  (output omitted)
. test _b[explained:ft_experience2] = 0, accum
 ( 1)  [explained]ft_experience = 0
 ( 2)  [explained]ft_experience2 = 0
           chi2(  2) =  393.78
         Prob > chi2 =    0.0000
```

- This is a different test than testing their joint contribution:

```
. test _b[explained:ft_experience] + _b[explained:ft_experience2] = 0
 ( 1)  [explained]ft_experience + [explained]ft_experience2 = 0
           chi2(  1) =  383.93
         Prob > chi2 =    0.0000
```

Linear and nonlinear combinations

- Close cousins of `test` and `testnl` are commands `lincom` and `nlcom`.
- I find `nlcom` extremely useful because it can generate arbitrary combinations and transformations of results. Standard errors (and covariances between multiple results) are computed by the so-called "delta methods" (linearization; first order Taylor series approximation; see, e.g., Feiveson 1999, Oehlert 1992).
- `lincom` is similar, but can only be used for linear combinations (and only computes one result at the time).

- Express the explained part and the unexplained part as percentage of the overall gap.

```
. nlcom (Percent_explained:   _b[overall:explained] /_b[overall:difference]*100) ///
>       (Percent_unexplained: _b[overall:unexplained]/_b[overall:difference]*100)
Percent_ex~d:  _b[overall:explained] /_b[overall:difference]*100
Percent_un~d:  _b[overall:unexplained]/_b[overall:difference]*100
```

lnwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Percent_explained	59.56321	3.93531	15.14	0.000	51.85014	67.27627
Percent_unexplained	40.43679	3.93531	10.28	0.000	32.72373	48.14986

- Compute the percentage of the overall gap that is explained by schooling, and the percentage that is explained by work experience.

```
. nlcom (schooling:  _b[explained:schooling] / _b[overall:difference]*100) ///
>       (experience: (_b[explained:ft_experience] + _b[explained:ft_experience2]) ///
>                     / _b[overall:difference]*100)
   schooling:  _b[explained:schooling] / _b[overall:difference]*100
  experience:  (_b[explained:ft_experience] + _b[explained:ft_experience2])
>       / _b[overall:difference]*100
```

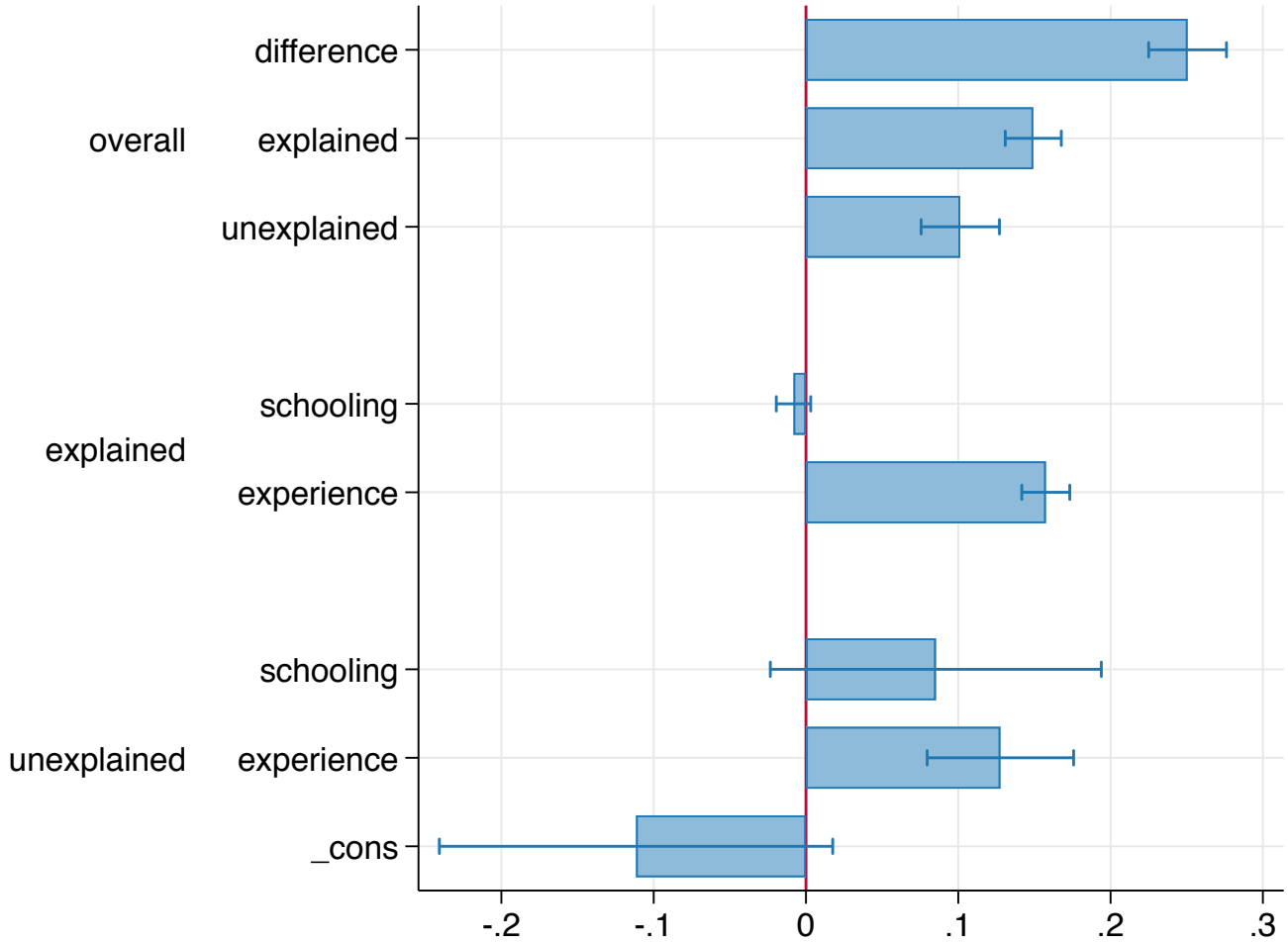
lnwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
schooling	-3.272932	2.369246	-1.38	0.167	-7.916569	1.370706
experience	62.83614	4.24111	14.82	0.000	54.52371	71.14856

Tables and graphs

- To tabulate results from `oaxaca` (and export the table to LaTeX or Word etc.) you can use, for example, command `esttab` (Jann 2007). There are also various other user commands that could be employed.
- For graphs, try `coefplot` (Jann 2014).
- Both commands support including results from multiple calls to `oaxaca` or `nlcom` that have been stored using `estimates store`.
- For `nlcom`, you need to specify the `post` option before tabulation and graphing is possible.

Example: graph

```
. oaxaca lnwage schooling (experience: ft_experience ft_experience2), by(bcsex) weight(1)
  (output omitted)
. coefplot, drop(overall:group*) xline(0) ///
>     recast(bar) barwidth(.7) base(0) citop ciopts(recast(rcap))
```



Example: display means and coefficients

- Note that `oaxaca` returns the coefficients and means that are used for the decomposition in `e(b0)` and `e(V0)`. Use option `xb` to display these auxiliary statistics.

```
. oaxaca, xb
Blinder-Oaxaca decomposition                    Number of obs     =      7,860
                                                  Model           =     linear
Group 1: bcsex = 1                                N of obs 1      =       3877
Group 2: bcsex = 2                                N of obs 2      =       3983
```

lnwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
overall						
group_1	2.749054	.009236	297.64	0.000	2.730951	2.767156
group_2	2.498484	.0092013	271.54	0.000	2.48045	2.516518
difference	.2505696	.0130372	19.22	0.000	.2250172	.276122
explained	.1492473	.009391	15.89	0.000	.1308412	.1676533
unexplained	.1013223	.0131188	7.72	0.000	.07561	.1270346
explained						
schooling	-.008201	.0057638	-1.42	0.155	-.0194978	.0030958
experience	.1574483	.0080355	19.59	0.000	.1416989	.1731976
unexplained						

Example: display means and coefficients

schooling	.0852652	.0554512	1.54	0.124	-.0234172	.1939476
experience	.1276546	.0245238	5.21	0.000	.0795889	.1757204
_cons	-.1115975	.0658889	-1.69	0.090	-.2407374	.0175423

experience: ft_experience ft_experience2

Coefficients (b) and means (x)

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
b1						
schooling	.0933227	.0029897	31.21	0.000	.0874629	.0991825
ft_experience	.0516494	.0031	16.66	0.000	.0455736	.0577253
ft_experience2	-.0009358	.0000859	-10.89	0.000	-.0011042	-.0007674
_cons	1.000596	.0487866	20.51	0.000	.904976	1.096216
b2						
schooling	.086751	.003054	28.41	0.000	.0807653	.0927367
ft_experience	.0358245	.0029841	12.01	0.000	.0299759	.0416732
ft_experience2	-.0006908	.0000953	-7.25	0.000	-.0008777	-.000504
_cons	1.112193	.0442856	25.11	0.000	1.025395	1.198992
b_ref						
schooling	.0933227	.0029897	31.21	0.000	.0874629	.0991825
ft_experience	.0516494	.0031	16.66	0.000	.0455736	.0577253
ft_experience2	-.0009358	.0000859	-10.89	0.000	-.0011042	-.0007674
_cons	1.000596	.0487866	20.51	0.000	.904976	1.096216

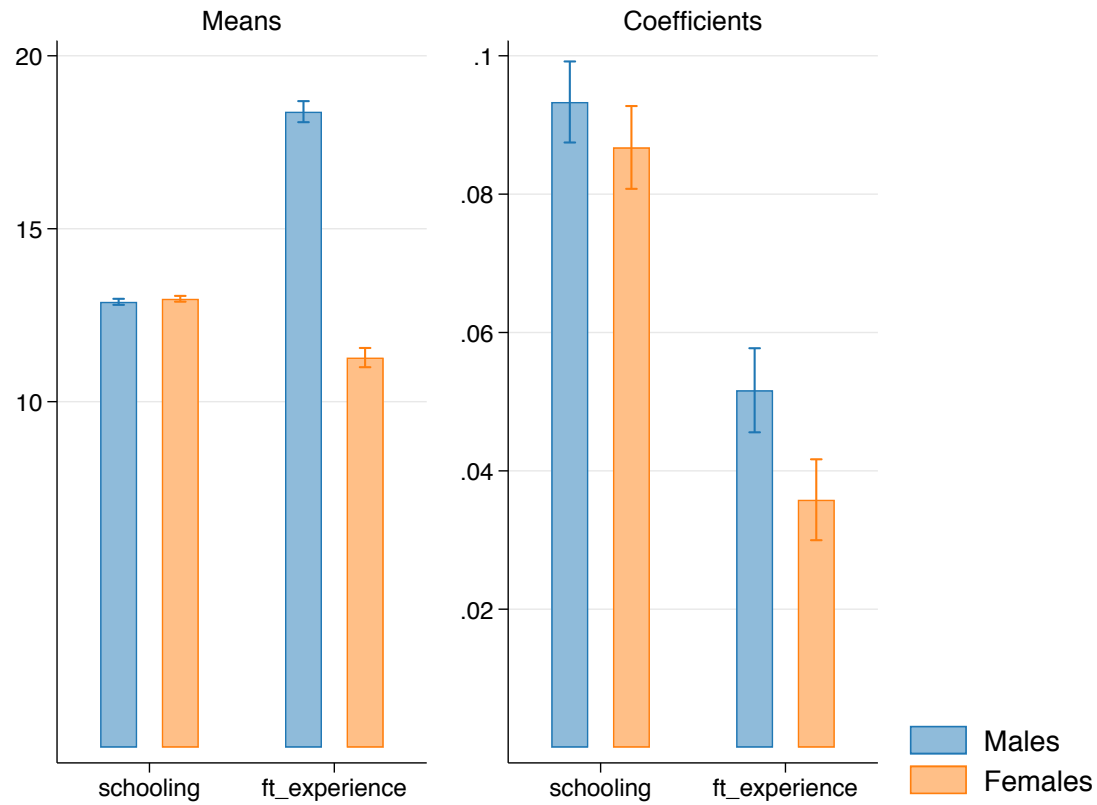
Example: display means and coefficients

x1						
schooling	12.88664	.0445749	289.10	0.000	12.79927	12.974
ft_experience	18.38458	.1552555	118.41	0.000	18.08028	18.68887
ft_experience2	431.4208	5.604688	76.97	0.000	420.4358	442.4058
_cons	1	.	.	.	.	.
x2						
schooling	12.97452	.0426577	304.15	0.000	12.89091	13.05812
ft_experience	11.27442	.1418485	79.48	0.000	10.9964	11.55243
ft_experience2	207.2343	4.439645	46.68	0.000	198.5328	215.9358
_cons	1	.	.	.	.	.

Example: display means and coefficients

- You can use `coefplot` to draw a graph:

```
. coefplot (. , keep(x1:) drop(_cons ft_experience2)) ///
>          (. , keep(x2:) drop(_cons ft_experience2)), bylabel(Means) ///
>     ||   (. , keep(b1:) drop(_cons ft_experience2)) ///
>          (. , keep(b2:) drop(_cons ft_experience2)), bylabel(Coefficients) ///
> || , b(b0) v(V0) byopts(yrescale) plotlabels(Males Females) ///
> recast(bar) barwidth(.2) base(0) citop ciopts(recast(rcap)) vertical
```



Example: graphing results from `nlcom`

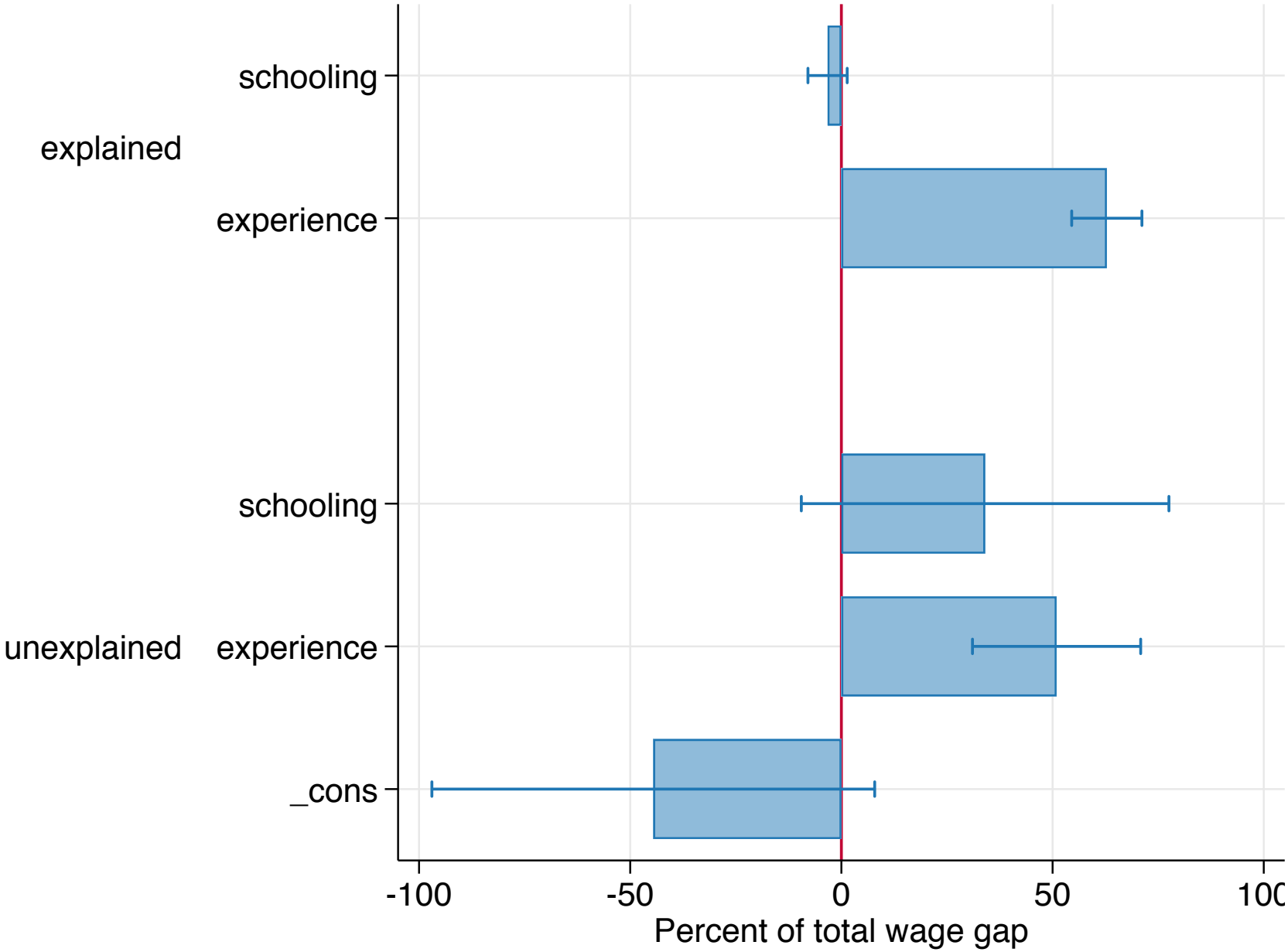
- Use the `post` option in `nlcom` to move the results to `e()` so that they can be tabulated (but be aware that this will delete original results unless they have been saved using `estimates store`).
- In the following example the detailed decomposition results are displayed as percentages of the overall gap.

```
. nlcom (e_schooling:  _b[explained:schooling]/_b[overall:difference]*100) ///
>       (e_experience: _b[explained:experience]/_b[overall:difference]*100) ///
>       (u_schooling:  _b[unexplained:schooling]/_b[overall:difference]*100) ///
>       (u_experience: _b[unexplained:experience]/_b[overall:difference]*100) ///
>       (u__cons:      _b[unexplained:_cons]/_b[overall:difference]*100) ///
>       , post
```

(*output omitted*)

```
. coefplot (., keep(e_*) asequation(explained) rename(e_* = "") ///
>         \ ., keep(u_*) asequation(unexplained) rename(u_* = "")) ///
>     , xline(0) recast(bar) barwidth(.7) base(0) citop ciopts(recast(rcap)) ///
>     xtitle("Percent of total wage gap")
```

Example: graphing results from `nlcom`



Example: graphing results from multiple decompositions

- Use `estimates store` to hold on to results from a decomposition for later processing.
- Example: wage gap in private sector vs. in public sector.

```
. oaxaca lnwage schooling (experience: ft_experience ft_experience2) if oeffd12!=1 ///
>     , by(bcsex) weight(1)
```

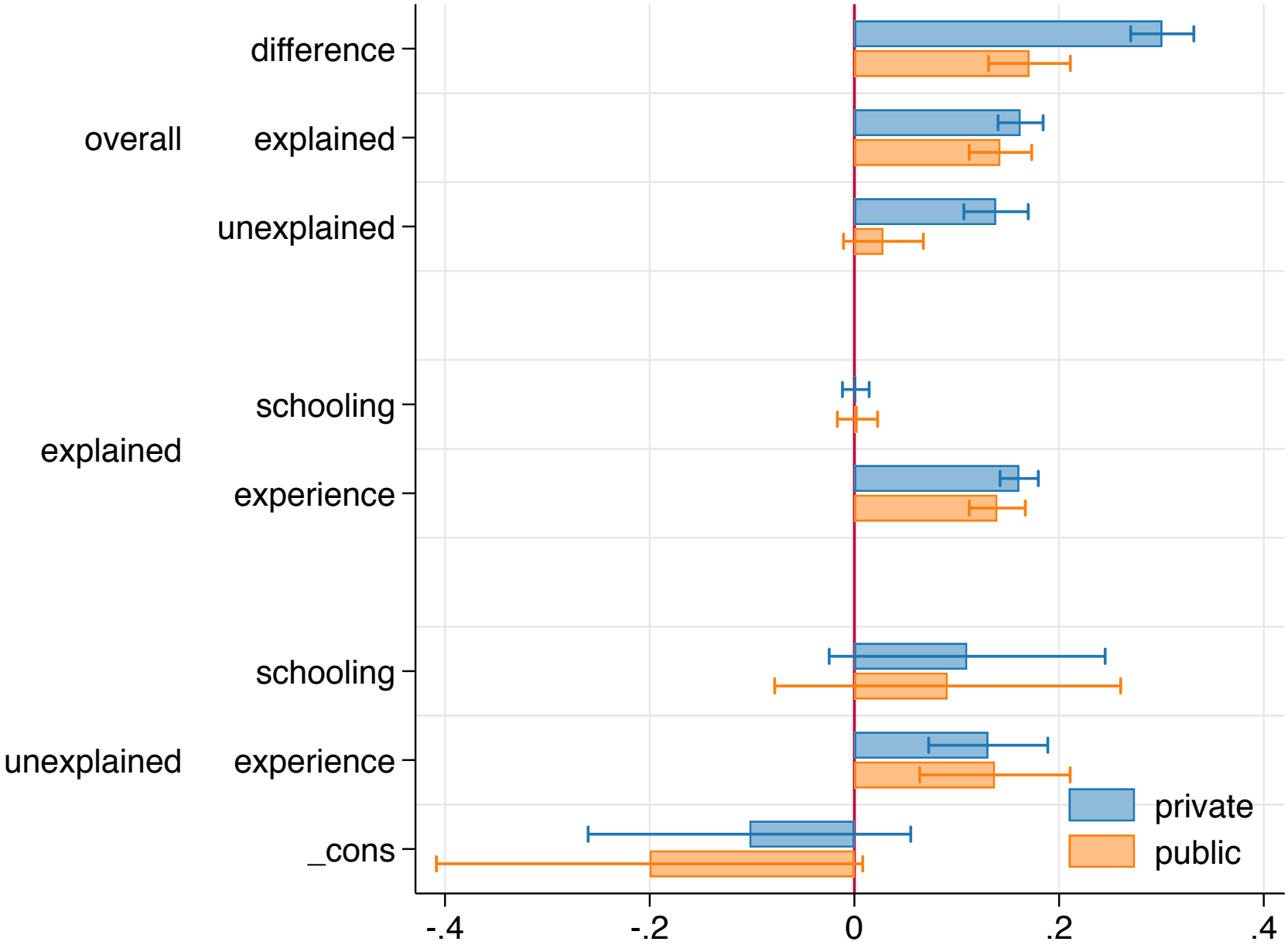
(*output omitted*)

```
. estimate store private
. oaxaca lnwage schooling (experience: ft_experience ft_experience2) if oeffd12==1 ///
>     , by(bcsex) weight(1)
```

(*output omitted*)

```
. estimate store public
. coefplot private public, drop(overall:group*) xline(0) ///
>     recast(bar) barwidth(.3) base(0) citop ciopts(recast(rcap))
```

Example: graphing results from multiple decompositions



Example: table

```
. oaxaca lnwage schooling ft_experience ft_experience2, by(bcsex) weight(1) nodetail
  (output omitted)
. estimates store raw
. nlcom (explained:   _b[overall:explained]  /_b[overall:difference]*100) ///
>       (unexplained: _b[overall:unexplained]/_b[overall:difference]*100), post
  (output omitted)
. estimates store pct
. esttab raw pct using mytable.tex, replace ///
>     keep(difference explained unexplained) nostar ci wide ///
>     noobs nonumber mtitle("Decomposition" "In percent") eqlab(none)
(output written to mytable.tex)
```

- The table looks like this:

	Decomposition		In percent	
difference	0.251	[0.225,0.276]		
explained	0.149	[0.131,0.168]	59.56	[51.85,67.28]
unexplained	0.101	[0.0756,0.127]	40.44	[32.72,48.15]

95% confidence intervals in brackets

Exercise 2

- Using the extended decomposition from Exercise 1, compute the percentage contribution of each variable to the explained part of the decomposition.
- To illustrate the effect of the survey design, draw a graph that displays these results with and without taking the survey design into account.
- Create a Word table (`rtf` file type in `esttab`) of the underlying decomposition results (with and without survey design) that could be included in the appendix of your paper.

References

- Feiveson, Alan H. (1999). FAQ: What is the delta method and how is it used to estimate the standard error of a transformed parameter? http://www.stata.com/support/faqs/stat/deltam.html
- Jann, Ben (2007). Making regression tables simplified. The Stata Journal 7(2):227–244.
- Jann, Ben (2014). Plotting regression coefficients and other estimates. The Stata Journal 14(4):708–737.
- Oehlert, Gary W. (1992). A Note on the Delta Method. The American Statistician 46(1):27–29.