

RIF decomposition

1. Repeat the example analysis of the private–public gap in wage inequality. This time, use the Gini coefficient as well as the D9/D1 ratio, the D5/D1 ratio, and the D9/D5 ratio as inequality measures. If possible, use `rifreg`, `rifvar`, and `oaxaca_rif` to calculate these results.

Be aware that the Gini coefficient is defined at the level of wages, not log wages. That is, you need to use original variable `wage` in the Gini decomposition, not variable `lnwage`. In contrast, for a decile-ratio decomposition, an option is to use the influence function of the inter-quantile range of `lnwage` (rather than the influence function of the inter-quantile ratio of `wage`).

2. Combine your RIF decomposition for the Gini coefficient with reweighting (analogous to the reweighted OB decomposition) and calculate the specification error. Use `oaxaca_rif` for this exercise.
3. Optional: Try to replicate the results for the reweighted Gini decomposition manually by first computing the RIF and then applying `oaxaca` to the RIF taking reweighting into account. You will need two calls to `oaxaca` to compute all results.

The whole procedure is somewhat involved. You could proceed about as follows:

Step 1 Generate the sector-specific RIF of the Gini using command `dstat` with option `rif()`. See slide 15 from the session on RIF regressions.

Step 2 Generate balancing weights that adjust the distribution of covariates among people in the private sector to the distribution observed among people in the public sector (this is the type of decomposition computed by `oaxaca_rif` with option `wgt(1)`, i.e. the people in the private sector are reweighted). You can use command `kmatch` to generate the weights or compute the weights manually by working with predictions from logistic regression.

Step 3 Compute the reweighted RIF of the Gini for the private sector. Similar to Step 1, but with the balancing weights applied to `dstat`.

Step 4 Apply `oaxaca` to the raw RIF of the private sector and the reweighted RIF of the private sector to obtain the “pure explained part” and the “specification error” (the “explained part” of this decomposition quantifies the “pure explained part”; the “unexplained part” quantifies the “specification error”).

This is not straightforward to do because it is a decomposition between private sector and reweighted private sector (i.e. twice the same subsample) and we need to duplicate the observations to be able to apply `oaxaca`. The `expand` command can be used for this purpose. It is also a good idea to do the expansion only temporarily using command `preserve` so that the data will not be changed permanently. Here is a brief sketch of what you need to do:

```
. preserve // preserve the data so that they can be restored later
. generate ID = _n // add a unique ID for each observation
. expand 2 if public==0 // duplicate each private sector observation
. bysort ID: generate byte G = (_n==2) if public==0 // generate a 0/1 variable
// that tags the duplicates; we can then use this as the group variable in
// oaxaca
// [here: code to perform the analysis
// - generate a RIF variable containing the raw RIF for G==0 and the reweighted
// RIF for G==1
```

```

// - generate weights containing 1 for G==0 and the balancing weights for G==1
// - apply oaxaca to the RIF variable while applying the weights; G is the group
//   variable]
. restore    // restore original data

```

Step 5 Apply `oaxaca` to the reweighted RIF of the private sector and the raw RIF of public sector to obtain the “pure unexplained part” and the “reweighting error” (the “explained part” of this decomposition quantifies the “reweighting error”; the “unexplained part” quantifies the “pure unexplained part”). This is straightforward; no duplication of observations is needed.

4. Optional: Repeat the analysis using entropy balancing for the reweighting. How do the results change?

Hint: You only need to modify Step 2 where the balancing weights are computed.