Decomposition methods in economics

Nicole Fortin, UBC

Thomas Lemieux, UBC

Sergio Firpo, EESP-FGV

This chapter

- Uses the classic work of Oaxaca (1973) and Blinder (1973) for the **mean** as its point of departure
- Focuses on recent developments (last 15 years) on how to go beyond the mean
- Provides empirical illustrations and discusses applications throughout
- Suggests a “user guide” of best practices

What is new since Oaxaca?

- Methods for going beyond the mean motivated by:
  - Inequality literature (JMP, DFL, etc.)
  - Interest for "what happens where", e.g. gender gap and glass ceiling
- Connection with treatment effect literature
  - Helps formalize and clarify some aspects of decompositions
  - Structural vs. unstructural modelling
- Ongoing issues
  - Base group problem (Oaxaca and Ransom, 1999)
  - Selection (Neal and Johnson, Petrongolo and Olivetti, etc.)

Plan of the presentation

- Status of the paper…
- Quick refresher on the Oaxaca decomposition
- Cover the main contribution of the chapter as a set of six main “take away” points
- Explain where we are in terms of writing and what is the “to do” list

Refresher on Oaxaca decomposition

- Want to decompose the difference in the mean of an outcome variable *Y* between two groups *A* and *B*
- Groups could also be periods, regions, etc.
- Postulate linear model for *Y*, with conditionally independent errors:

$$Y_{iG} = \beta_{G0} + \sum_{k=1}^{K} X_{iGk}\beta_{Gk} + \varepsilon_{iG}, G = A, B,$$

- The difference $\Delta = E(Y_B) - E(Y_A)$ can be decomposed as

$$\Delta = \underbrace{(\beta_{B0} - \beta_{A0}) + \sum_{k=1}^{K} \mathbb{E}(X_{Bk})(\beta_{Bk} - \beta_{Ak})}_{\Delta_S \text{ (Unexplained)}} + \underbrace{\sum_{k=1}^{K} (\mathbb{E}(X_{Bk}) - \mathbb{E}(X_{Ak}))\beta_{Ak}}_{\Delta_X \text{ (Explained)}}$$

A few remarks

- We focus on this particular decomposition, but we could also change the order, show the interaction term, etc. Does not affect the substance of the argument in most cases.
- In the "aggregate" decomposition, we only divide Δ into its two components Δ_S (wage structure effect) and Δ_X (composition effect).
- In the "detailed" decomposition we also look at the contribution of each individual covariate (or corresponding β)
- The "intercept" component of Δ_S, $\beta_{B0} - \beta_{A0}$, is the wage structure effect for the base group. Unless the other β's are the same in group *A* and *B*, $\beta_{B0} - \beta_{A0}$ will arbitrarily depend on the base group chosen.

The six take-away points

1. The wage structure effect (Δ_S) can be interpreted as a treatment effect
2. Going beyond the mean is a "solved" problem for the aggregate decomposition
3. Going beyond the mean is more difficult for the detailed decomposition
4. The analogy between quantile and standard (mean) regressions is not helpful
5. Decomposing proportions is easier than decomposing quantiles
6. There is no econometric solution to the base group problem

1. The wage structure effect (Δ_S) can be interpreted as a treatment effect

- The conditional independence assumption ($E(\varepsilon|X)=0$) usually invoked in Oaxaca decompositions can be replaced by the weaker ignorability assumption to compute the aggregate decomposition
- For example, ability (ε) can be correlated with education (X) as long as the correlation is the same in groups A and B.
- Also helps provide a slightly more structural foundation to the decomposition.
- If we have $Y_G=m_G(X, \varepsilon)$ and ignorability, then:
  - Δ_S solely reflects changes in the m(.) functions (ATET)
  - Δ_X solely reflects changes in the distribution of X and ε (ignorability key for this last result).

1. The wage structure effect (Δ_S) can be interpreted as a treatment effect

- A number of estimators for ATET= Δ_S have been proposed in the treatment effect literature
  - Inverse probability weighting (IPW), matching, etc.
- Formal results exist, e.g. IPW is efficient for
  - ATET (Hirano, Imbens, and Ridder, 2003)
  - Quantile treatment effects (Firpo, 2007)
- We like this since it provides a theoretical justification for DFL's approach…

1. The wage structure effect (Δ_S) can be interpreted as a treatment effect

- When the treatment effect $Y_{iB}-Y_{iA}$ is heterogenous, the ATET depends on the characteristics of group B.
- The difference in intercepts $\beta_{B0}- \beta_{A0}$ can be interpreted as the ATE in the base group
- Each component $E(X_{Bk})(\beta_{Bk}- \beta_{Ak})$ indicates by how much the ATE changes when we switch from $X_{Bk}=0$ (base group) to $X_{Bk}=E(X_{Bk})$.
- Not clear this is, in general, a sensible way of describing heterogeneity in the treatment effect.
- Needs some economics to help here, for instance A=blacks, B=whites, and X_k is union status dummy (Ashenfelter)

Going beyond the mean is a "solved" problem for the aggregate decomposition

- Can directly apply non-parametric methods (IPW, matching, etc.) from the treatment effect literature.
- Ignorability is crucial, but $m_G(X, \varepsilon)$ does not need to be linear
- Inference by bootstrap or analytical standard errors in the case of IPW ("generated regressor" correction required)
- IPW/DFL very easy to use with large and well behaved (no support problem) data sets.

Going beyond the mean is more difficult for the detailed decomposition

- Until recently, there were only a few partial (and not always satisfactory) ways of performing a detailed decomposition for general distributional measures (quantiles in particular):
  - DFL conditional reweighting for the components of Δ_X linked to dummy covariates (e.g. unions)
  - Machado-Mata quantile regressions for components of Δ_S.
  - Sequential DFL-type reweighting, adding one covariate at a time. Sensitive to order used as in a simple regression.
- A more promising approach is to estimate for proportions, and invert back to quantiles. RIF regression of Firpo, Fortin and Lemieux (2009) is one possible way of doing so (more on this soon)

Going beyond the mean is more difficult for the detailed decomposition

- We also propose a more general conditional reweighting approach
- Intuition for components of Δ_X:
  - When sequentially adding regressors, the effect for the last one is consistent since all other covariates have been controlled for.
  - Similarly, comparing the effect obtained by reweighting on all X's vs. all X's except X_k gives the correct effect of X_k.
  - Repeating the procedure for each X_k gives the right marginal contribution of each X_k, though the k effects do not sum up to the total (interaction effects).
- A reweighting approach can also be used to compute the components of Δ_S (as in DiNardo and Lemieux, 1997):
  - Restrict sample to X_k=0 (or other base group value)
  - Reweight on the X_{-k} other covariates to have the same distribution as in the full sample.
  - Gives the distribution when the wage structure effect of X_k has been set to zero.

Quantile regressions do not help

- Tempting to run quantile regressions (say for the median) and perform a decomposition as in the case of the mean (Oaxaca)
- Does not work because there are two interpretations to β for the mean
  - Conditional mean: $E(Y|X) = X\beta$
  - Uncond. mean (LIE): $E(Y) = E_X(E(Y|X)) = E_X(X)\beta$
- But the LIE does not work for quantiles
  - Conditional quantile: $Q\tau(Y|X) = X\beta\tau$
  - Uncond. quantile: $Q\tau \neq E_X(Q\tau(Y|X)) = E_X(X)\beta\tau$
- Only the first interpretation works for $\beta\tau$, which is not useful for decomposing unconditional quantiles

Decomposing proportions is easier than decomposing quantiles

- Example: 10 percent of men earn more than 80K a year, but only 5 percent of women do so.
- Easy to do a decomposition by running LP models for the probability of earning less (or more) than 80K, and perform a Oaxaca decomposition on the proportions.
- By contrast, much less obvious how to decompose the difference between the 90th quantile for men (80K) and women (say 65K)
- But function linking proportions and quantiles is the cumulative distribution.
- Counterfactual proportions → Counterfactual cumulative → Counterfactual quantiles
- Can be illustrated graphically

Figure 1: Relationship Between Proportions and Quantiles

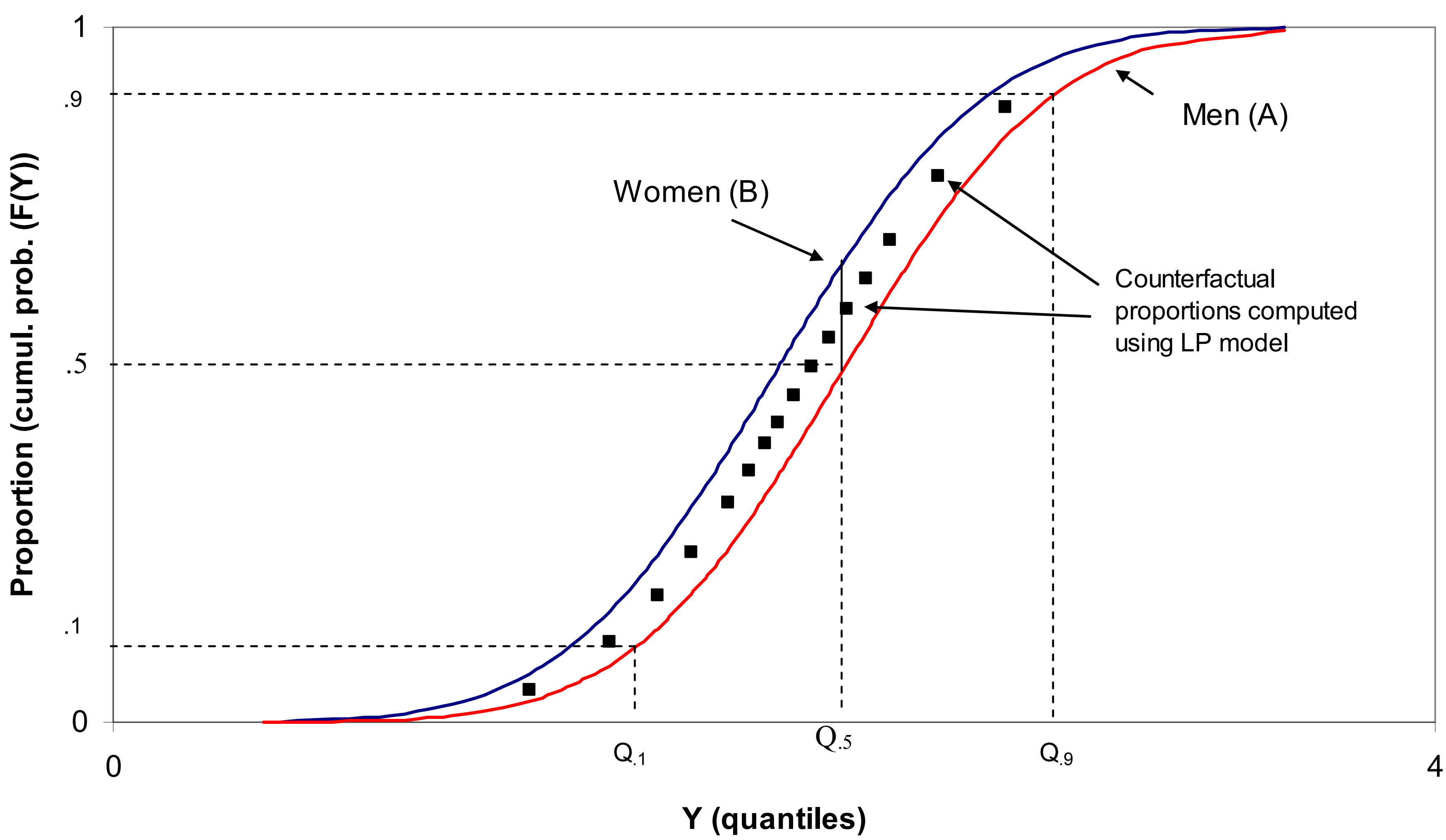


Figure 2: RIF Regressions: Inverting Locally

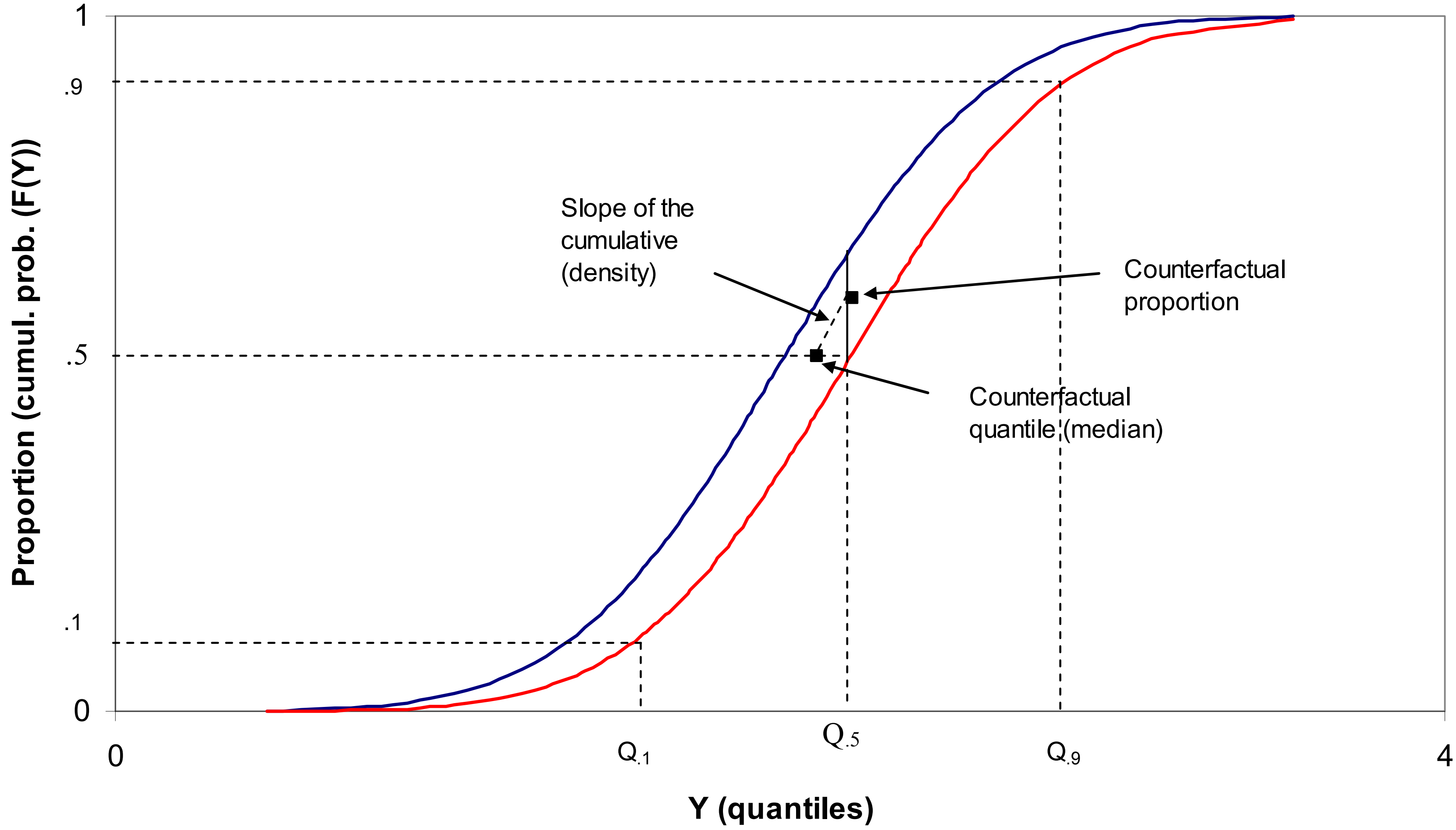
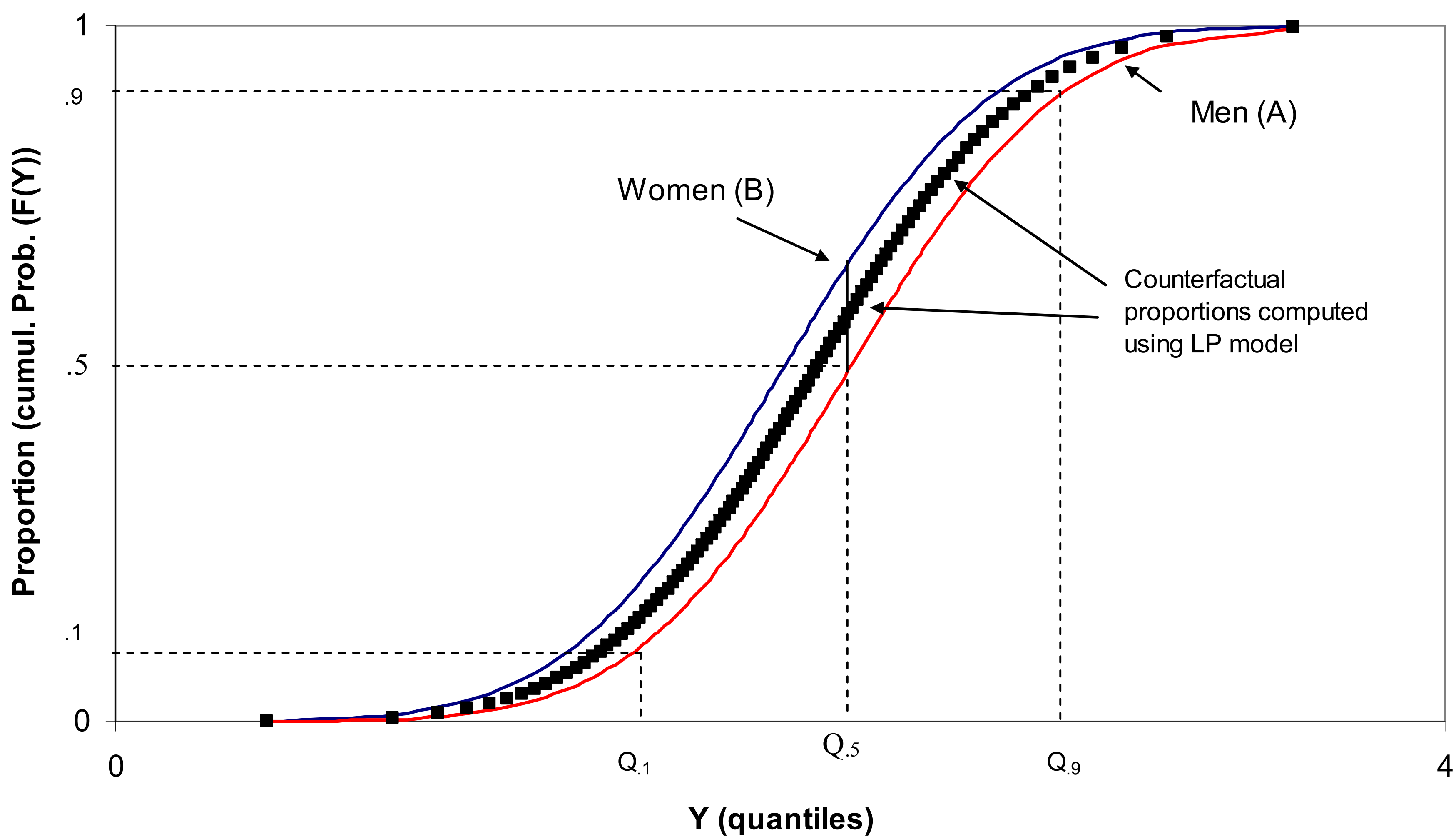


Figure 3: Inverting Globally



Decomposing proportions is easier than decomposing quantiles

- FFL recentered influence function (RIF) regressions
  - Run LP models (or logit/probit) for being below a given quantiles, and divide by density (slope of cumulative) to locally invert.
  - Dependent variable is dummy $1(Y<Q_\tau)$ divided by density → influence function for the quantile.
  - RIF approach works for other distributional measures (Gini, variance, etc.)
- Chernozhukov et al. (2009)
  - Estimate "distributional regressions" (LP, logit or probit) for each value of Y (say at each percentile)
  - Invert back globally to recover counterfactual quantiles

There is no econometric solution to he base group problem

- Elements of the detailed decomposition are well defined for Δ_X.
- Effect of changing the distribution of X_k (group A to group B) holding the distribution of the other covariates constant
- No base group problem for elements Δ_X.
- For Δ_S, however, there are as many detailed decompositions as there are base groups
- OK when the base group is of particular economic interest. For example, if base group = unskilled (0 experience, primary education)
- Otherwise the whole exercise is not very useful
- Better to find interesting ways of characterizing the heterogeneity in the treatment effect to give some guidance on what are the interesting economic factors at work.
- (Theoretical) example: gender gap small in most occupations, but large in a few “top-end” occupations. Can then compute counterfactual gender gap if nobody was in these few top-end occupations.

To-do list

- Finish the write-up
- Add empirical application(s)
- Discuss empirical applications