Summary:
Chapter 1 Decomposition Methods in Economics

Nicole Fortin et al.

https://github.com/s-saisw/readingSummary

February 12, 2021

Review

Keywords:

1 Introduction

- Decomposition methods are often used to examine differences between two groups e.g. wage gap between union VS non-union members, races, men VS women etc.
- Limits of decomposition methods:
  - Decomposition methods follow a partial equilibrium approach, ignoring the general equilibrium effect.
  - Decomposition methods do not seek to recover "deep" structural parameters.
- Setting: suppose there are 2 groups (g), A and B. Assume that
  1. Outcome variable Y is linearly correlated to covariates X.
  2. Error term ν is conditionally dependent of X, i.e. $\mathbb{E}(\nu_{gi}|X_i) = 0$.

$$Y_{gi} = \beta_{g0} + \sum_{k=1}^{K} X_{ik}\beta_{gk} + \nu_{gi}, \;\; g = A, B \tag{1}$$

Then the average difference between 2 groups, $\hat{\Delta}^{\mu}_{O}$, can be written as

$$\begin{aligned}
\hat{\Delta}^{\mu}_{O} =& \hat{\beta}_{B0} + \sum_{k=1}^{K} \bar{X}_{Bk}\hat{\beta}_{Bk} - (\hat{\beta}_{A0} + \sum_{k=1}^{K} \bar{X}_{Ak}\hat{\beta}_{Ak}) \\
=& \hat{\beta}_{B0} - \hat{\beta}_{A0} + \sum_{k=1}^{K} \bar{X}_{Bk}\hat{\beta}_{Bk} - \sum_{k=1}^{K} \bar{X}_{Ak}\hat{\beta}_{Ak} - \sum_{k=1}^{K} \bar{X}_{Bk}\hat{\beta}_{Ak} + \sum_{k=1}^{K} \bar{X}_{Bk}\hat{\beta}_{Ak} \\
=& \left[(\hat{\beta}_{B0} - \hat{\beta}_{A0}) + \sum_{k=1}^{K} \bar{X}_{Bk}(\hat{\beta}_{Bk} - \hat{\beta}_{Ak}) \right] + \sum_{k=1}^{K} (\bar{X}_{Bk} - \bar{X}_{Ak})\hat{\beta}_{Ak} \\
=& \hat{\Delta}^{\mu}_{S} + \hat{\Delta}^{\mu}_{X},
\end{aligned}$$

where $\hat{\Delta}^{\mu}_{S}$ is the *unexplained (structural) part*, and $\hat{\Delta}^{\mu}_{X}$ the *explained* (by covariates X) part.

- *Aggregate decomposition* refers to decomposing $\hat{\Delta}^{\mu}_{O} = \hat{\Delta}^{\mu}_{S} + \hat{\Delta}^{\mu}_{X}$, while *detailed decomposition* refers to subdividing the effect for each $k = 1, ..., K$

- Take away messages from this chapter

  1. The wage structure effect can be interpreted as a treatment effect.
  2. Going beyond the mean is a "solved" problem for 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 general solution to the "omitted" group problem.

2 Identification

Assumption 1 ***(Mutually Exclusive Groups).*** *The population of agents can be divided into two mutually exclusive groups: A and B. Thus, for an agent i, $D_{Ai} + D_{Bi} = 1$, where $D_{gi} = \mathbb{1}\{i \text{ is in } g\}$.*

- This rules out decomposing wage differentials between overlapping groups e.g. Blacks, Whites, and Hispanics, who can be Black or White. (In this case, dummy variable method is a more natural approach.)

- Overall, this assumption is not very restrictive.

2.1 Aggregate decomposition

We would like to decompose the observable "overall difference" (e.g. wage gap) into "four decomposition terms" under "overlapping support assumption" and "ignorability assumption".

2.1.1 Overall difference

- Let the distributional statistic of interest be $v(F_{Y_g|D_s})$, where $v : F \to \mathbb{R}$. $F_{Y_g|D_s}$ can be both observed (when $g = s$) and counterfactual (when $g \neq s$) distributions.

- The overall v-difference in wages between the two groups in terms of the distributional statistic v is

$$\Delta^{v}_{O} = v(F_{Y_B|D_B}) - v(F_{Y_A|D_A}) \tag{2}$$

  – Overall difference is generally not what we want to uncover. We are more interested in the counterfactuals, e.g. what if group A workers' pay follow group B workers' wage profile.

Assumption 2 ***(Structural Form).*** *A worker i from group g is paid according to the wage structure m_g, which are functions of worker i's observable (X) and unobservable (ϵ) characteristics.*

$$Y_{Ai} = M_A(X_i, \epsilon) \text{ and } Y_{Bi} = M_B(X_i, \epsilon) \tag{3}$$

- This assumption implies there are only 3 sources of wage differentials

  1. wage setting functions m
  2. distribution of X

3. distribution of ϵ

Assumption 3 *(Simple Counterfactual Treatment).* *A counterfactual wage structure for workers in group B, $m^c(\cdot,\cdot) \equiv m_A(\cdot,\cdot)$, and vice versa.*

- For example, counterfactual wage of a worker in group B is $Y^C_{A|D_B,i} = m^C(X_i, \epsilon_i) = m_A(X_i, \epsilon_i)$.
- This assumption rules out the existence of another counterfactual wage distribution e.g. what would happen to the pay structure of female in a world in which there is no gender.
- This assumption makes decomposition framework a partial equilibrium framework.

2.1.2 Four decomposition terms

- Overall difference between two groups can be attributed to
  1. Differences in returns to X
  2. Differences in returns to ϵ
  3. Differences in the distribution of X
  4. Differences in the distribution of ϵ
- 1 and 2 can be summed up as structural difference, Δ^v_s. This is also called "v-wage structure effect." Then we have

$$\Delta^v_O = \Delta^v_s + \Delta^v_X + \Delta^v_\epsilon. \tag{4}$$

2.1.3 Overlapping support

Assumption 4 *(Overlapping Support).* *Let the support of wage setting factors $[X', \epsilon']'$ be $\mathcal{X} \times \mathcal{E}$. For any $[x', e']'$ in $\mathcal{X} \times \mathcal{E}$, $0 < \Pr[D_B = 1 | X = x, \epsilon = e] < 1$.*

- This assumption implies there is no $X = x$ or $\epsilon = e$ that divides the population into groups. A counter example is when only high-ability people are union workers and only low-ability people are non-union workers.
- In the decomposition of gender wage differential, this assumption is often violated.

2.1.4 Ignorability

- Because there are some unobserved components involved, even if we fix the distribution of X to be the same across 2 groups, we do not observe what would happen to ϵ. Thus, we need to impose a further assumption to make sure that changes in X are not confounded by changes in ϵ.
- For workers in group B, observed wage follows the following distribution

$$F_{Y_B|D_B}(y) = \int F_{Y_B|X,D_B}(y|X=x) \cdot dF_{X|D_B}(x) \tag{5}$$

  The counterfactual wage follows

$$F_{Y^C_A|X=X:D_B}(y) = \int F_{Y_A|X,D_A}(y|X=x) \cdot dF_{X|D_B}(x). \tag{6}$$

That is, we integrate the counterfactual distribution over the distribution of observable characteristics of workers in group B[1].

- Consider the RHS. For $g = A, B$, conditional distribution is defined as follows.

$$F_{Y_g|X,D_g}(y|X = x) = \Pr(m_g(X, \epsilon) \leq y|X = x, D_g = 1) \tag{7}$$

- According to (7), the distribution of wages depends only on the distribution of ϵ and the wage structure $m_g(\cdot)$. Therefore, when we replace the conditional distribution on the RHS of (5) with the counterfactual one as in (6), we are replacing both functional form and distribution of ϵ. $\Rightarrow$ We need to separate the two from each other, i.e. ignorability assumption

Assumption 5 ***(Ignorability/Conditional independence).*** *For $g = A, B$, let (D_g, X, ϵ) have a joint distribution. For all x in $\mathcal{X}$, $D_g \perp \epsilon|X$.*

- It implies that $\Delta_S^v = v(F_{Y_B|D_B}) - v(F_{Y_A^C:X=X|D_B})$, i.e. difference between actual and counterfactual depends solely on wage structure (i.e. structural functions $m_B(\cdot, \cdot)$ and $m_A(\cdot, \cdot)$).
- This assumption is often called "unconfoundedness" or "selection on observables" in the program evaluation literature.
- It is somewhat strong assumption. There are 3 cases under which it may not hold.
    1. Differential selection into labor marker
       The decision to participate may be different for men and women.
    2. Self-selection into groups A and B based on unobservables.
    3. Choice of X and ϵ.

Proposition 1 ***(Identification of the Aggregate Decomposition).*** *Under simple counterfactual, overlapping support, and ignorability, the overall v-gap can be written as*

$$\Delta_O^v = \Delta_s^v + \Delta_X^v,$$

where

1. *$\Delta_S^v = v(F_{Y_B|D_B}) - v(F_{Y_A^C:X=X|D_B})$, i.e. wage structure term solely reflects the difference between the structural functions $m_B(\cdot, \cdot)$ and $m_A(\cdot, \cdot)$.*
2. *$\Delta_X^v = v(F_{Y_A^C:X=X|D_B}) - v(F_{Y_A|D_A})$, i.e. composition effect term solely reflects the difference between the X and ϵ between 2 groups.*

- Explanation:
    - $v(F_{Y_B|D_B}) - v(F_{Y_A^C:X=X|D_B})$ is the difference of two distributions that have different shapes but the same Xs.
    - $v(F_{Y_A^C:X=X|D_B}) - v(F_{Y_A|D_A})$ is the difference of two distributions that have the same shape but different Xs.

[1] Note that we can easily replace $F_{Y_B|X,D_B}(y|X = x)$ with $F_{Y_A|X,D_A}(y|X = x)$ under simple counterfactual assumption.

- Note that

$$\begin{aligned}
\Delta_O^v =& v(F_{Y_B|D_B}) - v(F_{Y_A|D_A}) \\
=& v(F_{Y_B|D_B}) - v(F_{Y_A^C:X=X|D_B}) + v(F_{Y_A^C:X=X|D_B}) - v(F_{Y_A|D_A}) \\
=& \Delta_S^v + \Delta_X^v
\end{aligned}$$

Assumption 6 ***(Invariance of Conditional Distributions).*** *When replace the conditional distribution of* (5) *as in* (6), $F_{Y_A|X,D_A}(y|X=x)$ *remains valid for any* $x \in \mathcal{X}$

- This assumption is not empirically testable. Thus, it must be justified using the economic context.

References