Inequality in Health

Lecture III: Measuring and Decomposing Health Inequality I

Dr Martin Karlsson

University of Duisburg-Essen

Winter semester 2022-23

Outline

1. Recap of Last Lecture
2. Introduction
3. Oaxaca-Blinder Decomposition
4. Extensions to Oaxaca-Blinder Decomposition
5. Mediation Analysis
6. Summary and Conclusions
7. Appendix

Recap of Last Lecture

Recap of Last Lecture

- The **Lorenz curve** captures all quantile share information of a distribution and can be used to compare distributions.
- **Gini & Theil indices** summarize this information and give a complete ordering of the income distribution.
- The **concentration curve** is useful to analyse socioeconomic disparities in health.
- When we want to test **concentration curve dominance** empirically, we need a **decision rule**.
- A **concentration index** is a summary statistic for the inequality in income-related health differences.
- The choice of health indicator is important as subjective measures might overstate actual SES-related health inequality.

Introduction

Introduction

- Health sector inequalities measured through bivariate relationship between health variable and SES.
- Once we have **measured** inequalities, natural next step is to seek to **account** for them.
- To go beyond measurement of inequalities, we need **multivariate analysis**, like:
  - Finer description of inequality through **standardization** for age, gender, etc.
  - Explanation of **differences between groups** through **decomposition** analysis.
  - Identification of **causal relationship** between health variable and SES.

Decomposition Analysis

- We examine methods of decomposing inequality into its **contributing factors**.
- Idea: to explain or describe the outcome variable by a set of factors that vary systematically with SES.
- Example: **Child malnutrition** as outcome variable. Differences in malnutrition may be explained by differences in **income** (poorer children more malnourished than richer children), but also by differences in:
  - **distance** to the closest **hospital**,
  - parental **education**,
  - **insurance** coverage,
  - access to clean **water** etc.
- We want to know to what extent inequalities in health are due to inequalities **in each factor**.

Interpretation of Results

- We now focus on decomposition methods based on **regression analysis**.

1. If regressions are purely **descriptive**, they reveal the **associations** that characterise health inequality.
   - We may not gain further knowledge about the **mechanisms** underlying the relationship between factors and outcomes.
   - Inequality explained in a **statistical** sense but limited implications for **policies** aimed at reducing inequality.
2. If data allow identification of **causal effects**, we identify the factors that **generate** inequality.
   - Conclusions about how policies would **impact** on inequality.

Oaxaca-Blinder Decomposition

Leading Example: Gender Pay Gap

Support The Guardian | Subscribe | Search jobs | Sign in | Search | International edition

The Guardian

News | Opinion | Sport | Culture | Lifestyle | More

World ▸ Europe US Americas Asia Australia Middle East Africa Inequality Cities Global development

Gender pay gap

Gender pay gap narrows but TUC calls for greater pressure

Companies should be required to explain how they will close the gap, says Frances O'Grady

Alexandra Topping *and* Patrick Collinson

Fri 26 Oct 2018 06.00 BST

60

▲ Gender pay gap: women now earn 17.9% less than men, down from 18.4% last year. Photograph: Joe Giddens/PA

The gender pay gap has dropped slightly this year to a low of 17.9%, according to figures from the Office for National Statistics.

Women now earn 17.9% less than men, down from 18.4% last year. The rate for full-time workers has dropped from 9.1% to a low of 8.6%.

most viewed

New discovery throws light on mystery of pyramids' construction

Defiant Donald Trump warns of 'caravans and crime' in last minute push for midterm votes

'Freakish miracle': New Zealand fisherman pulls drowning toddler from ocean

Britain has never looked so foolish in the world's eyes *Jan Fleischhauer*

Canada debates assisted death laws after woman is forced to end life early

Figure 1. Gender Pay Gap.

Oaxaca-Blinder Decomposition

- Idea: to explain the distribution of Y taking into account some variables X that vary **systematically** with **gender**.
- **Example**: female earnings may be less responsive to education than male earnings.
- Oaxaca decomposes the gap (in mean) of the outcome variable between two groups, M and F:

$$Y_l = X'\beta_l + \epsilon_l$$

where

l group indicator, $l \in \{M, F\}$
Y outcome variable
X set of determinants

and $\mathbb{E}(\epsilon_l) = 0$.

Gap in Mean Outcomes

- Oaxaca-Blinder (OB) decomposition assumes that inequalities are caused by differences in the **magnitudes** of determinants, but also by differences in their **effects**.
- Regression model for **males** (M) and **females** (F):

$$Y_l = \begin{cases} X'\beta_M & +\epsilon_M \\ X'\beta_F & +\epsilon_F \end{cases} \quad (1)$$

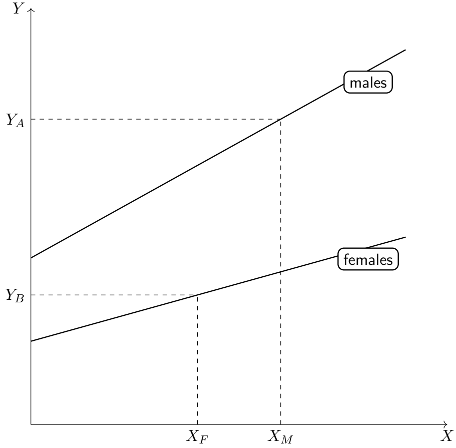

Figure 2. Differences in slopes.

Gap between Mean Outcomes

- Gap between the mean outcomes:

$$\mathbb{E}\left(Y_M - Y_F\right) = \mathbb{E}\left(X_M\right)' \beta_M - \mathbb{E}\left(X_F\right)' \beta_F \tag{2}$$

- The outcome gap is decomposed into two parts:
  - One partition due to the **magnitude** of the determinants, e.g. difference in **covariates**: $\Delta X = X^M - X^F$
  - Another part due to the **effect** of the determinants, e.g. difference in **coefficients**: $\Delta\beta = \beta^M - \beta^F$.
- Eq. (2) can be expressed in two ways:

$$\begin{aligned} \mathbb{E}\left(Y_M - Y_F\right) &= \mathbb{E}\left(\Delta X\right)' \beta_F + \mathbb{E}\left(X_M\right)' \Delta\beta && (3) \\ &= \mathbb{E}\left(\Delta X\right)' \beta_M + \mathbb{E}\left(X_F\right)' \Delta\beta && (4) \end{aligned}$$

Parts of the Oaxaca Decomposition

- The two decompositions in Eq. (3) and (4) are special cases of the **general** decomposition;

$$\mathbb{E}(Y_M - Y_F) = \underbrace{\mathbb{E}(\Delta X)' \beta_F}_{E} + \underbrace{\mathbb{E}(X_F)' \Delta\beta}_{C} + \underbrace{\mathbb{E}(\Delta X)' \Delta\beta}_{CE} \qquad (5)$$

E Gap in "**endowments**" (explained)
C Gap in **coefficients** (unexplained)
CE **Interaction** of these two gaps.

Oaxaca Decomposition: Components

- E: part of inequality that can be traced back to observed factors.
- C: part of inequality left unexplained by the observed factors.
- The interaction CE represents differences in levels **and** in effects between the groups that **exist at the same time**: they cannot be pinpointed.
- Example: Gender pay gap
  - E: differences in **educational attainment** (levels)
  - C: differences in the **effects** of educational attainment
  - CE: **interaction** between the levels and the effect of educational attainment.
- CE can be allocated to either C or E, or distributed between them.

Empirical Application: Averkamp et al. (2020)

- In a recent study, Averkamp et al. (2020) decompose the gender wage gap in the US.
- They evaluate to what extent the **gap in dual-earner households** can be explained by the partner's characteristics.
- The authors argue that when conflicts between careers occur, families choose to promote the career of the spouse with better **labor market opportunities**.
- The results show that the **partner's characteristics** explain a substantial share of the wage gap.

Averkamp et al. (2020): Results

Table 1. Decomposition Results by Year

	1980	1989	1998	2010
Wage gap	0.430	0.327	0.265	0.250
Standard decomposition				
total explained	0.225 (52%)	0.206 (63%)	0.188 (71%)	0.114 (46%)
unexplained	0.205 (48%)	0.121 (37%)	0.077 (29%)	0.136 (54%)
Extended decomposition				
total explained	0.357 (83%)	0.331 (101%)	0.241 (91%)	0.172 (69%)
own characteristics	0.229 (53%)	0.22 (67%)	0.201 (76%)	0.145 (58%)
partner characteristics	0.128 (30%)	0.111 (34%)	0.040 (15%)	0.027 (11%)
unexplained	0.072 (17%)	-0.004 (-1%)	0.024 (9%)	0.078 (31%)

Notes: First line shows log differences, second line (in parentheses) gives percentage of total wage gap.

Extensions to Oaxaca-Blinder Decomposition

Machado & Mata Decomposition

- OB: decomposition of effects on **mean** outcome (Oaxaca, 1973).
- It can be more interesting to look at effects on the **entire distribution**.
- Example: gender wage gap.
  - Consider the education distribution: how large is the gap when we look at individuals with low education?
  - Is the gap bigger/smaller when we consider highly educated individuals instead?
- I.e. the **effect of covariates** is allowed to **differ** over conditional outcome distribution.
- Machado and Mata (2005, MM) analyze this issue using **quantile regression**.

Marginal and Conditional Distributions

Compare the wage distribution ω to the wage distribution conditional just on **one covariate** z, a university degree dummy ($= 1$ if individual has degree; $= 0$ if they have no degree).

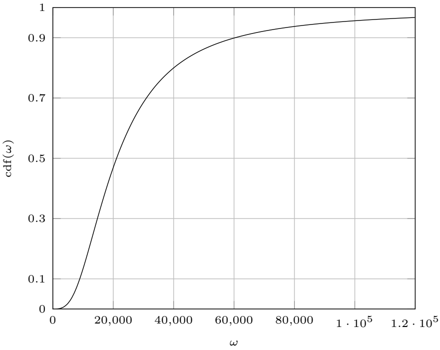

Figure 3. Marginal wage distribution.

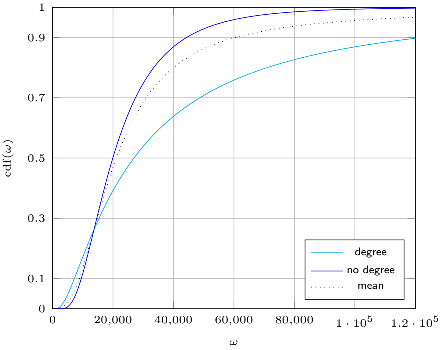

Figure 4. Conditional wage distributions.

Conditional and Unconditional Quantiles

- We consider one specific statistic, the quantile function $Q_\theta(\cdot)$.

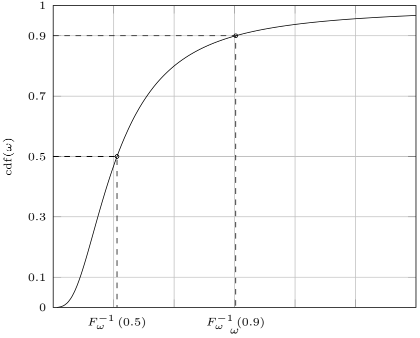

Figure 5. Marginal wage distribution.

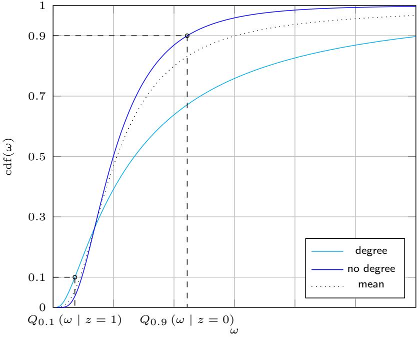

Figure 6. Conditional wage distributions.

Time Trends

- We consider the wage distribution $\omega(t)$ in **two** time **periods**: $t = 1986, 1995$.

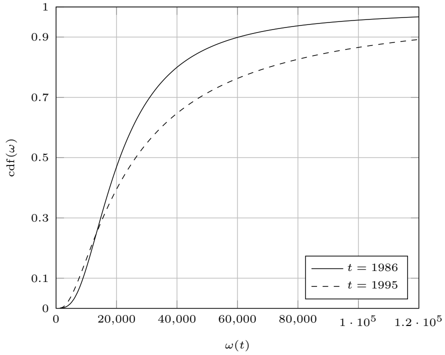

Figure 7. Marginal wage distributions.

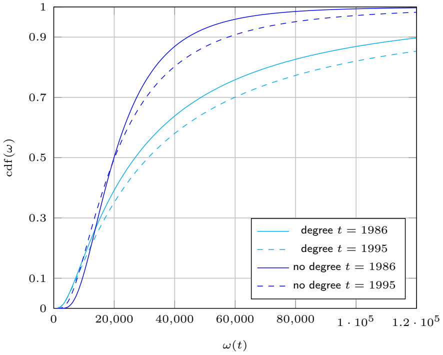

Figure 8. Conditional wage distributions.

Counterfactual Distribution

- Consider the fictitious wage distribution that would have prevailed at time $t = 1995$ if the covariate z was distributed as in $t = 1986$: $f(\omega_{1995}; z_{1986})$.
- This is a **counterfactual** distribution.

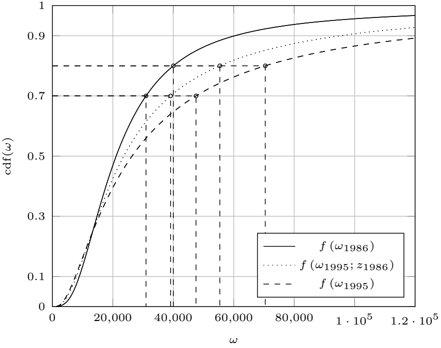

Figure 9. Marginal wage distributions.

Counterfactual Quantiles

- Once we have the cdf:s of this counterfactual distribution:

$$F(\omega_{1995} \mid Z_{1986}) = F_0(\omega_{1995}) \Pr(z_{1986} = 0) + F_1(\omega_{1995}) \Pr(z_{1986} = 1)$$

- ...can derive a **counterfactual quantile function** for any quantile θ

$$Q_\theta(\omega_{1995} \mid Z_{1986}) = F^{-1}(\omega_{1995} \mid Z_{1986})$$

- A comparison of $Q_\theta(\omega_{1995} \mid Z_{1995})$, $Q_\theta(\omega_{1995} \mid Z_{1986})$ and $Q_\theta(\omega_{1986} \mid Z_{1986})$ would then tell us
  - How much of the change is due to **changes in** z? $(Q_\theta(\omega_{1995} \mid Z_{1995}) - Q_\theta(\omega_{1995} \mid Z_{1986}))$
  - How much is due to changes in **returns to education**? $(Q_\theta(\omega_{1995} \mid Z_{1986}) - Q_\theta(\omega_{1986} \mid Z_{1986}))$
  - And this for **any quantile** of the distribution $\theta \in (0, 1)$.

The Curse of Dimensionality

- There is only one problem.
- We normally have a **set of variables** in Z.
- And we won't be able to estimate the full cdf $F\left(\omega_{1995}\right)$ for any $z \in Z$.
- Machado & Mata propose a **simulation algorithm** that allows to estimate counterfactual distributions for many combinations of the observable characteristics Z.

Counterfactual Distributions II

- Consider outcome ω, groups A, B and the respective characteristic vectors Z_A and Z_B.
- Regression quantiles for group A and B are respectively $\beta^A(\theta)$ and $\beta^B(\theta)$ for $\theta \in (0,1)$.
- Consider **quantile regressions**:

$$Q_\theta\left(\omega_A \mid Z_A = z_A\right) = z_A \beta^A(\theta) \tag{6}$$

$$Q_\theta\left(\omega_B \mid Z_B = z_B\right) = z_B \beta^B(\theta) \tag{7}$$

- If Eq. (6) is correctly specified, $Q_\theta\left(\omega_A \mid Z_A = z_A\right)$ gives a full characterization of the distribution of ω_A given z_A (same for group B – Eq. 7).

Counterfactual Distributions III

- Consider a **counterfactual** random variable ω_{AB} with conditional quantiles

$$Q_\theta\left(\omega_{AB} \mid Z_A = z_A\right) = z_A \beta^B(\theta)$$

- MM generate a sample from the unconditional distribution of ω_{AB} as follows:
  1. Sample θ from a standard uniform distribution.
  2. Estimate $\hat{\beta}^B(\theta)$
  3. Sample x_A from its empirical distribution.
  4. Compute $\hat{\omega}_{AB} = x_A \hat{\beta}^B(\theta)$.
  5. Repeat previous steps n times.
- Then use **bootstrapping** to estimate standard errors for the quantiles of the counterfactual distribution.

Machado & Mata: Empirical Application

- Decomposition of effects on the entire wage distribution in Portugal considering two points in time: $t = 1986, 1995$.
- Observable characteristics Z: sex, education, age, tenure.
- Counterfactual distributions useful to disentangle the effect of **two types of changes**:
  1. In the **stock** of human capital:
     - By increasing the number of educated workers their wages decrease.
  2. In the **returns** to human capital's components:
     - If more educated individuals experience greater wage differentials, increased education contributes to increasing wage inequality.
- Aim: to decompose the difference

$$Q_\theta\left(\omega_{1995} \mid Z_{1995}\right) - Q_\theta\left(\omega_{1986} \mid Z_{1986}\right).$$

Decomposing the Changes

- General decomposition of the changes in the quantile function Q_θ:

$$
\begin{aligned}
Q_\theta(\omega_{1995} \mid Z_{1995}) - Q_\theta(\omega_{1986} \mid Z_{1986}) = & \\
& \underbrace{Q_\theta(\omega_{1995} \mid Z_{1986}) - Q_\theta(\omega_{1995} \mid Z_{1995})}_{\text{coefficient component}} \\
& + \underbrace{Q_\theta(\omega_{1986} \mid Z_{1986}) - Q_\theta(\omega_{1995} \mid Z_{1986})}_{\text{covariate component}} \\
& + \text{residual component}
\end{aligned}
$$

where $Q_\theta(\omega_{1995} \mid Z_{1986})$ is the counterfactual density of ω at time 1995 with covariates distributed as they were at time 1986.

The Components

- The **coefficient component**

$$Q_\theta\left(\omega_{1995} \mid Z_{1986}\right) - Q_\theta\left(\omega_{1995} \mid Z_{1995}\right)$$

  captures the contribution of the quantile regression coefficients to the **overall change** (remuneration/returns to skills).
- The **covariate component**

$$Q_\theta\left(\omega_{1986} \mid Z_{1986}\right) - Q_\theta\left(\omega_{1995} \mid Z_{1986}\right)$$

  captures the contribution of the covariates to the **changes in the wage density** (individual characteristics).
- The **residual** component is unexplained.
- The contribution of an **individual covariate** z is given by:

$$Q_\theta\left(\omega_{1995} \mid Z_{1995}\right) - Q_\theta\left(\omega_{1995} \mid Z_{1995}; z_{1986}\right)$$

Results

Table 2. Decomposition of Changes in the Wage Distribution in Portugal, 86-95.

	Marginals			Aggregate contributions			Individual covariates			
	1986	1995	Change	Covariates	Coefficients	Residual	Sex	Education	Age	Tenure
10th quant.	5.535	5.721	0.186	0.049	0.187	-0.050	-0.021	0.027	0.015	-0.027
				[0.263]	[1.005]	[-0.269]	[-0.115]	[0.147]	[0.078]	[-0.146]
25th quant.	5.705	5.894	0.189	0.049	0.148	-0.008	-0.014	0.035	0.013	-0.015
				[0.259]	[0.783]	[-0.042]	[-0.073]	[0.185]	[0.070]	[-0.078]
Median	5.957	6.201	0.244	0.064	0.163	0.017	-0.015	0.058	0.015	-0.017
				[0.262]	[0.668]	[0.070]	[-0.063]	[0.236]	[0.060]	[-0.071]
75th quant.	6.350	6.621	0.271	0.092	0.197	-0.018	-0.023	0.092	0.024	-0.015
				[0.340]	[0.727]	[-0.066]	[-0.086]	[0.341]	[0.087]	[-0.054]
90th quant.	6.786	7.168	0.382	0.093	0.238	0.051	-0.022	0.092	0.017	-0.053
				[0.243]	[0.623]	[0.134]	[-0.057]	[0.240]	[0.045]	[-0.140]

Estimates for selected quantiles of the distribution; [proportion of the total change explained by the factor].

Machado & Mata: Limitations & Further Extensions

- It is **path dependent**: the decomposition results depend on the **order** in which the decomposition is performed.
- The decomposition of the outcome variable's structure still depends on the choice of the **omitted group**.
- **Sample selection** is potentially a relevant issue in studies on wage inequality. Albrecht et al. (2009) extend the MM method to tackle this issue.
- **Computationally** demanding, even though the procedure can be simplified (Albrecht et al., 2003).
- It assumes **linearity**, but a linear specification may be too restrictive.

Decomposition of Non-Linear Models

- In case of limited dependent variables, linear decomposition techniques might be inappropriate.
- Thus, several approaches were developed which take non-linearity into account.
- Yun (2004) introduces an approach which uses linearisation of the model; however, the non-linear model structure is ignored.
- The method proposed by Fairlie (2005) is based on matching between groups but suffers from path dependence (i.e. decomposition is not unique).
- A recent approach by Schwiebert (2015) is applicable to a wide range of non-linear models and takes the weaknesses of the other two methods into account.

Mediation Analysis

Heckman & Pinto Mediation Analysis

- Heckman and Pinto (2015) developed a technique for decomposing treatment effects in RCT settings.
- The approach allows to decompose the causal effect into **direct effects** mediated by known factors and **indirect effects** due to unmeasured variables.
- They define an outcome Y_d as

$$Y_d = f_d\left(\theta_d^p, \theta_d^u, X\right) \tag{8}$$

with

f_d production function for treatment $d \in 0, 1$,

θ_d^p proxied ($=$ observed) inputs,

θ_d^u unmeasured inputs, and

X baseline variables unaffected by treatment.

Average Treatment Effect

- The average treatment effect (ATE) of treatment D is given by

$$ATE = \mathbb{E}\left(Y_1 - Y_0\right) \tag{9}$$
$$= \mathbb{E}\left(f_1\left(\theta_1^p, \theta_1^u, X\right) - f_0\left(\theta_0^p, \theta_0^u, X\right)\right), \tag{10}$$

i.e., the expected difference between the outcome in the treated and untreated state.

- This ATE can be decomposed into an indirect effect (IE),

$$IE\left(d\right) = \mathbb{E}\left(f_d\left(\theta_1^p, \theta_d^u, X\right) - f_d\left(\theta_0^p, \theta_d^u, X\right)\right) \tag{11}$$

and a direct effect (DE),

$$DE\left(d\right) = \mathbb{E}\left(f_1\left(\theta_d^p, \theta_1^u, X\right) - f_0\left(\theta_d^p, \theta_0^u, X\right)\right) \tag{12}$$

- Thus, we get

$$ATE = DE(1) + IE(0) \tag{13}$$
$$= DE(0) + IE(1). \tag{14}$$

Application: Heckman et al. (2013)

- Heckman et al. (2013) decompose the average treatment effect of an early life intervention.
- Application: the **Perry Preschool programme** – an early childhood intervention targeted at disadvantaged African Americans.
- RQ: through which channels did the intervention affect adult outcomes?
- The intervention was conducted as an RCT and caused significant improvements in education, earnings, and healthy behaviour (Heckman et al., 2010).
- Heckman et al. (2013) find that experimentally induced changes in **personality traits** explain a sizable portion of adult treatment effects.

Empirical Strategy

- For the outcomes Y_d, the following linear specification is used:

$$Y_d = \kappa_d + \underbrace{\sum_{j \in J_p} \alpha_d^j \theta_d^j}_{\text{Measured skills}} + \underbrace{\sum_{j \in J \setminus J_p} \alpha_d^j \theta_d^j}_{\text{Unobserved skills}} + \beta_d X + \tilde{\epsilon}_d \tag{15}$$

$$= \tau_d + \sum_{j \in J_p} \alpha_d^j \theta_d^j + \beta_d X + \epsilon_d \tag{16}$$

where

$$\tau_d = \kappa_d + \sum_{j \in J \setminus J_p} \alpha_d^j E\left(\theta_d^j\right), \text{ and} \tag{17}$$

$$\epsilon_d = \tilde{\epsilon}_d + \sum_{j \in J \setminus J_p} \alpha_d^j \left(\theta_d^j - E\left(\theta_d^j\right)\right) \tag{18}$$

Empirical Strategy

- Under the assumption that treatment affects skills but not the **impact** of skills and background variables X on outcomes (i.e., $\beta_1 = \beta_0$, $\alpha_1^j = \alpha_0^j \; \forall j \in J_p$), we get:

$$Y_d = \tau_d + \sum_{j \in J_p} \alpha^j \theta_d^j + \beta X + \epsilon_d \tag{19}$$

- Thus, we can write the observed outcome Y as

$$\begin{aligned} Y &= DY_1 + (1-D)\,Y_0 && (20) \\ &= \tau_0 + \tau D + \sum_{j \in J_p} \alpha^j \theta^j + \beta X + \epsilon && (21) \end{aligned}$$

where $\tau = \tau_1 - \tau_0$, $\epsilon = D\epsilon_1 + (1-D)\,\epsilon_0$, and $\theta^j = D\theta_1^j + (1-D)\,\theta_0^j$.

Empirical Strategy

- If $\theta^j \; \forall j \in J_p$ are independent of the error term ϵ, the OLS estimates of (21) for $\alpha^j \; \forall j \in J_p$ are unbiased.
- Now the ATE can be decomposed as

$$ATE = \underbrace{(\tau_1 - \tau_0)}_{\text{treatment effect due to unmeasured skills}} + \underbrace{\sum_{j \in J \setminus J_p} \alpha^j E\left(\theta_1^j - \theta_0^j\right)}_{\text{treatment effect due to measured skills}} \qquad (22)$$

- The second term is constructed by using the estimates of the α^j and the effects of the treatment on skills.
- If the unmeasured skills are correlated with measured skills and outcomes, least squares estimators of α^j are **biased**.

Results

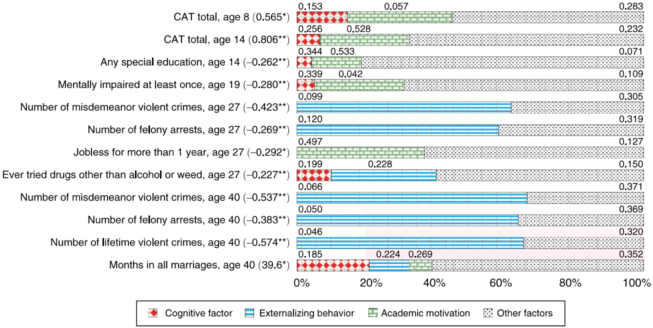

Figure 10. Decomposition of Effects (Females), Heckman et al. (2013).

Summary and Conclusions

Summary and Conclusions

- The **Oaxaca-Blinder decomposition** allows to attribute differences in the outcome variable (ex. health) between two groups to several explanatory factors.
- The decomposition partitions differences into:
  - Gap in **endowments** (explained)
  - Gap in **coefficients** (unexplained)
  - Interaction between gaps.
- Machado & Mata propose a decomposition method based on quantile regression.
  - It allows to "go beyond the mean" by performing a **detailed** decomposition by quantiles.
  - It **simulates counterfactual** distributions of the covariates.

Literature I

ALBRECHT, J., A. BJÖRKLUND, AND S. VROMAN (2003): "Is there a glass ceiling in Sweden?" *Journal of Labor economics*, 21, 145–177.

ALBRECHT, J., A. VAN VUUREN, AND S. VROMAN (2009): "Counterfactual distributions with sample selection adjustments: Econometric theory and an application to the Netherlands," *Labour Economics*, 16, 383–396.

AVERKAMP, D., C. BREDEMEIER, AND F. JUESSEN (2020): "Decomposing Gender Wage Gaps: A Family Economics Perspective," .

FAIRLIE, R. W. (2005): "An extension of the Blinder-Oaxaca decomposition technique to logit and probit models," *Journal of economic and social measurement*, 30, 305–316.

HECKMAN, J., S. H. MOON, R. PINTO, P. SAVELYEV, AND A. YAVITZ (2010): "Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program," *Quantitative economics*, 1, 1–46.

HECKMAN, J., R. PINTO, AND P. SAVELYEV (2013): "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes," *American Economic Review*, 103, 2052–86.

HECKMAN, J. J. AND R. PINTO (2015): "Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs," *Econometric reviews*, 34, 6–31.

MACHADO, J. A. AND J. MATA (2005): "Counterfactual decomposition of changes in wage distributions using quantile regression," *Journal of applied Econometrics*, 20, 445–465.

OAXACA, R. (1973): "Male-female wage differentials in urban labor markets," *International economic review*, 693–709.

SCHWIEBERT, J. (2015): "A detailed decomposition for nonlinear econometric models," *The Journal of Economic Inequality*, 13, 53–67.

YUN, M.-S. (2004): "Decomposing differences in the first moment," *Economics letters*, 82, 275–280.

Appendix

Oaxaca Decomposition 1

$$\mathbb{E}\left(Y_M - Y_F\right) = \mathbb{E}\left(\Delta X\right)' \beta_M + \mathbb{E}\left(X_M\right)' \Delta\beta - \mathbb{E}\left(\Delta X\right)' \Delta\beta$$

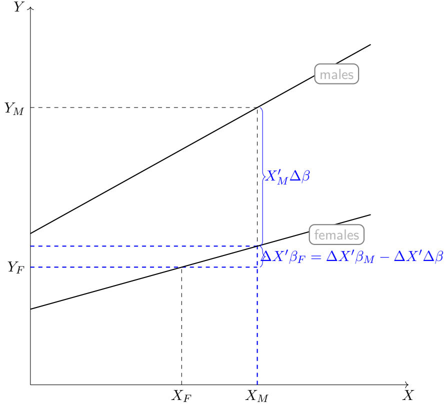

Figure 11. First Decomposition.

Oaxaca Decomposition 2

$$\mathbb{E}\left(Y_M - Y_F\right) = \mathbb{E}\left(\Delta X\right)' \beta_F + \mathbb{E}\left(X_F\right)' \Delta\beta + \mathbb{E}\left(\Delta X\right)' \Delta\beta$$

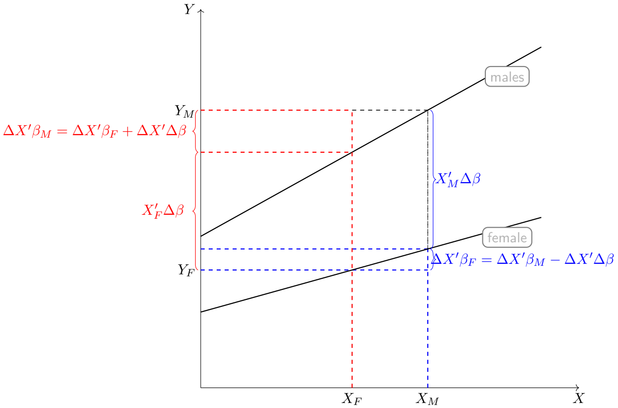

Figure 12. Second Decomposition.

Combinations

- A generalised version, that combines both decomposition approaches, would be:

$$\begin{aligned}\mathbb{E}\left(Y_A - Y_B\right) &= \overbrace{\Delta X'\beta_B}^{E} + \overbrace{D \times \Delta X'\Delta\beta}^{CE} + \overbrace{X_B'\Delta\beta}^{C} + \overbrace{(I-D) \times \Delta X'\Delta\beta}^{CE} \\ &= \Delta X'\left[D\beta_A + (I-D)\,\beta_B\right] + \left[X_A\,(I-D) + X_B D\right]\Delta\beta\end{aligned}$$

  where I is the identity matrix and D a matrix of weights.
- $D = 1 \Rightarrow$ Oaxaca decomposition 1.
- $D = 0 \Rightarrow$ Oaxaca decomposition 2.

Cotton's Method

- Choice of the weighting matrix D somewhat **arbitrary**.
- Among others Cotton (1988) proposes to weight the coefficients by mean of the **proportion of observations** in each group:

$$\begin{aligned}\beta^{\star} &= D\beta_A + (I - D)\,\beta_B \\ &= \frac{n_A}{n_A + n_B}\beta_A + \frac{n_B}{n_A + n_B}\beta_B\end{aligned}$$

  where n_l is the number of observations in group $l \in (A, B)$.
- $D = sI$, with s relative sample size of the majority group.

Cotton's Method

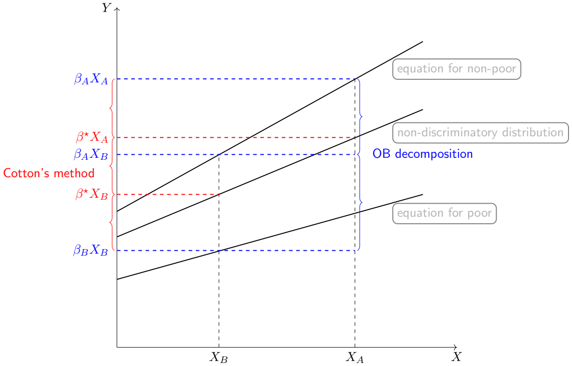

Figure 13. Decomposition of the non-poor - poor differential.

- OB decomposition: $(\beta_A X_A - \beta_A X_B) + (\beta_A X_B - \beta_B X_B)$.
- Cotton's: $(\beta_A X_A - \beta^\star X_A) + (\beta^\star X_A - \beta^\star X_B) + (\beta^\star X_B - \beta_B X_B)$.