Inequality in Health

Lecture IV: Measuring and Decomposing Health Inequality II

Dr Martin Karlsson

University of Duisburg-Essen

Winter semester 2022-23

Outline

Recap of Last Lecture

Recap of Last Lecture

- The **Oaxaca-Blinder decomposition** allows to attribute differences in the outcome variable (ex. health) between two groups to several explanatory factors.

Recap of Last Lecture

- 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:

Recap of Last Lecture

- 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)

Recap of Last Lecture

- 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)

Recap of Last Lecture

- 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.

Recap of Last Lecture

- 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.

Recap of Last Lecture

- 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.

Recap of Last Lecture

- 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.

Introduction

Introduction

- Decomposition methods like OB and MM allow for a **limited** understanding of socioeconomic-related inequality in health: explain differences between **two groups** only (e.g. rich-poor).

Introduction

- Decomposition methods like OB and MM allow for a **limited** understanding of socioeconomic-related inequality in health: explain differences between **two groups** only (e.g. rich-poor).
- Alternative decomposition methods explain socioeconomic-related inequality across the **entire distribution** of the SES variable (ex. income distribution).

Introduction

- Decomposition methods like OB and MM allow for a **limited** understanding of socioeconomic-related inequality in health: explain differences between **two groups** only (e.g. rich-poor).
- Alternative decomposition methods explain socioeconomic-related inequality across the **entire distribution** of the SES variable (ex. income distribution).
- Popular measures of inequality: the Gini coefficient, the concentration curve and the concentration index.

Introduction

- Decomposition methods like OB and MM allow for a **limited** understanding of socioeconomic-related inequality in health: explain differences between **two groups** only (e.g. rich-poor).
- Alternative decomposition methods explain socioeconomic-related inequality across the **entire distribution** of the SES variable (ex. income distribution).
- Popular measures of inequality: the Gini coefficient, the concentration curve and the concentration index.
- Focus on the **concentration index**: it measures the socioeconomic inequality of health taking into account:

Introduction

- Decomposition methods like OB and MM allow for a **limited** understanding of socioeconomic-related inequality in health: explain differences between **two groups** only (e.g. rich-poor).
- Alternative decomposition methods explain socioeconomic-related inequality across the **entire distribution** of the SES variable (ex. income distribution).
- Popular measures of inequality: the Gini coefficient, the concentration curve and the concentration index.
- Focus on the **concentration index**: it measures the socioeconomic inequality of health taking into account:
    - the **level of health** of each individual

Introduction

- Decomposition methods like OB and MM allow for a **limited** understanding of socioeconomic-related inequality in health: explain differences between **two groups** only (e.g. rich-poor).
- Alternative decomposition methods explain socioeconomic-related inequality across the **entire distribution** of the SES variable (ex. income distribution).
- Popular measures of inequality: the Gini coefficient, the concentration curve and the concentration index.
- Focus on the **concentration index**: it measures the socioeconomic inequality of health taking into account:
  - the **level of health** of each individual
  - their socioeconomic **rank**.

Correcting the Concentration Index

Introduction

- The **choice** of the inequality indicator to use may influence the results generated by the analysis.

Introduction

- The **choice** of the inequality indicator to use may influence the results generated by the analysis.
- Over the years researchers proposed alternative ways to compute the **concentration index**, providing **improved** alternatives to tackle some shortcomings of the original CI.

Introduction

- The **choice** of the inequality indicator to use may influence the results generated by the analysis.
- Over the years researchers proposed alternative ways to compute the **concentration index**, providing **improved** alternatives to tackle some shortcomings of the original CI.
- We consider:

Introduction

- The **choice** of the inequality indicator to use may influence the results generated by the analysis.
- Over the years researchers proposed alternative ways to compute the **concentration index**, providing **improved** alternatives to tackle some shortcomings of the original CI.
- We consider:
  1. A modified concentration index

Introduction

- The **choice** of the inequality indicator to use may influence the results generated by the analysis.
- Over the years researchers proposed alternative ways to compute the **concentration index**, providing **improved** alternatives to tackle some shortcomings of the original CI.
- We consider:
  1. A modified concentration index
  2. The Wagstaff normalization

Introduction

- The **choice** of the inequality indicator to use may influence the results generated by the analysis.
- Over the years researchers proposed alternative ways to compute the **concentration index**, providing **improved** alternatives to tackle some shortcomings of the original CI.
- We consider:
    1. A modified concentration index
    2. The Wagstaff normalization
    3. The Erreygers index.

Health Variable: Level of Measurement

- According to the type and the amount of **information** available we can express health status through different type of variables:

Health Variable: Level of Measurement

- According to the type and the amount of **information** available we can express health status through different type of variables:
  1. **Ordinal**: we are just able to **rank** individuals.

Health Variable: Level of Measurement

- According to the type and the amount of **information** available we can express health status through different type of variables:
  1. **Ordinal**: we are just able to **rank** individuals.
     - Example: Self-Assessed Health – SAH (1 – excellent; 2 – fair; 3 – poor health).

Health Variable: Level of Measurement

- According to the type and the amount of **information** available we can express health status through different type of variables:
  1. **Ordinal**: we are just able to **rank** individuals.
     - Example: Self-Assessed Health – SAH (1 – excellent; 2 – fair; 3 – poor health).
  2. **Cardinal**: possible to **compare differences** between health states.

Health Variable: Level of Measurement

- According to the type and the amount of **information** available we can express health status through different type of variables:
  1. **Ordinal**: we are just able to **rank** individuals.
     - Example: Self-Assessed Health – SAH (1 – excellent; 2 – fair; 3 – poor health).
  2. **Cardinal**: possible to **compare differences** between health states.
     - Example: body temperature.

Health Variable: Level of Measurement

- According to the type and the amount of **information** available we can express health status through different type of variables:
  1. **Ordinal**: we are just able to **rank** individuals.
     - Example: Self-Assessed Health – SAH (1 – excellent; 2 – fair; 3 – poor health).
  2. **Cardinal**: possible to **compare differences** between health states.
     - Example: body temperature.
  3. **Ratio-scale**: "zero health" fixed in an unambiguous way (reference).

Health Variable: Level of Measurement

- According to the type and the amount of **information** available we can express health status through different type of variables:
  1. **Ordinal**: we are just able to **rank** individuals.
     - Example: Self-Assessed Health – SAH (1 – excellent; 2 – fair; 3 – poor health).
  2. **Cardinal**: possible to **compare differences** between health states.
     - Example: body temperature.
  3. **Ratio-scale**: "zero health" fixed in an unambiguous way (reference).
     - Example: life expectancy; pulse.

Health Variable: Level of Measurement

- According to the type and the amount of **information** available we can express health status through different type of variables:
  1. **Ordinal**: we are just able to **rank** individuals.
     - Example: Self-Assessed Health – SAH (1 – excellent; 2 – fair; 3 – poor health).
  2. **Cardinal**: possible to **compare differences** between health states.
     - Example: body temperature.
  3. **Ratio-scale**: “zero health” fixed in an unambiguous way (reference).
     - Example: life expectancy; pulse.
- **Range** of the health variable:

Health Variable: Level of Measurement

- According to the type and the amount of **information** available we can express health status through different type of variables:
  1. **Ordinal**: we are just able to **rank** individuals.
     - Example: Self-Assessed Health – SAH (1 – excellent; 2 – fair; 3 – poor health).
  2. **Cardinal**: possible to **compare differences** between health states.
     - Example: body temperature.
  3. **Ratio-scale**: "zero health" fixed in an unambiguous way (reference).
     - Example: life expectancy; pulse.
- **Range** of the health variable:
  - **Bounded**: have both a finite upper and lower bound.

Health Variable: Level of Measurement

- According to the type and the amount of **information** available we can express health status through different type of variables:
  1. **Ordinal**: we are just able to **rank** individuals.
     - Example: Self-Assessed Health – SAH (1 – excellent; 2 – fair; 3 – poor health).
  2. **Cardinal**: possible to **compare differences** between health states.
     - Example: body temperature.
  3. **Ratio-scale**: "zero health" fixed in an unambiguous way (reference).
     - Example: life expectancy; pulse.
- **Range** of the health variable:
  - **Bounded**: have both a finite upper and lower bound.
  - **Unbounded**: we assume here that unbounded variables have a finite lower bound.

Limitations of $C(h)$ by Health Variable

1. **Ordinal** variable:

Limitations of $C(h)$ by Health Variable

1. **Ordinal** variable:
   - Any **positive monotone** transformation captures the same information.

Limitations of $C(h)$ by Health Variable

1. **Ordinal** variable:
   - Any **positive monotone** transformation captures the same information.
   - $C(h)$ is not invariant to positive monotone transformations.

Limitations of $C(h)$ by Health Variable

1. **Ordinal** variable:
   - Any **positive monotone** transformation captures the same information.
   - $C(h)$ is not invariant to positive monotone transformations.
   - With **qualitative** health variables $C(h)$ is to a large extent **arbitrary**, because differences between individuals cannot be compared.

Limitations of $C(h)$ by Health Variable

1. **Ordinal** variable:
   - Any **positive monotone** transformation captures the same information.
   - $C(h)$ is not invariant to positive monotone transformations.
   - With **qualitative** health variables $C(h)$ is to a large extent **arbitrary**, because differences between individuals cannot be compared.
2. **Cardinal** variable:

Limitations of $C(h)$ by Health Variable

1. **Ordinal** variable:
    - Any **positive monotone** transformation captures the same information.
    - $C(h)$ is not invariant to positive monotone transformations.
    - With **qualitative** health variables $C(h)$ is to a large extent **arbitrary**, because differences between individuals cannot be compared.
2. **Cardinal** variable:
    - Any **positive linear** transformation captures the same information.

Limitations of $C(h)$ by Health Variable

1. **Ordinal** variable:
   - Any **positive monotone** transformation captures the same information.
   - $C(h)$ is not invariant to positive monotone transformations.
   - With **qualitative** health variables $C(h)$ is to a large extent **arbitrary**, because differences between individuals cannot be compared.
2. **Cardinal** variable:
   - Any **positive linear** transformation captures the same information.
   - $C(h)$ is not invariant to positive linear transformations.

Limitations of $C(h)$ by Health Variable

1. **Ordinal** variable:
   - Any **positive monotone** transformation captures the same information.
   - $C(h)$ is not invariant to positive monotone transformations.
   - With **qualitative** health variables $C(h)$ is to a large extent **arbitrary**, because differences between individuals cannot be compared.
2. **Cardinal** variable:
   - Any **positive linear** transformation captures the same information.
   - $C(h)$ is not invariant to positive linear transformations.
3. **Ratio-scale** variable:

Limitations of $C(h)$ by Health Variable

1. **Ordinal** variable:
   - Any **positive monotone** transformation captures the same information.
   - $C(h)$ is not invariant to positive monotone transformations.
   - With **qualitative** health variables $C(h)$ is to a large extent **arbitrary**, because differences between individuals cannot be compared.
2. **Cardinal** variable:
   - Any **positive linear** transformation captures the same information.
   - $C(h)$ is not invariant to positive linear transformations.
3. **Ratio-scale** variable:
   - Any **positive proportional** transformation captures the same information.

Limitations of $C(h)$ by Health Variable

1. **Ordinal** variable:
   - Any **positive monotone** transformation captures the same information.
   - $C(h)$ is not invariant to positive monotone transformations.
   - With **qualitative** health variables $C(h)$ is to a large extent **arbitrary**, because differences between individuals cannot be compared.
2. **Cardinal** variable:
   - Any **positive linear** transformation captures the same information.
   - $C(h)$ is not invariant to positive linear transformations.
3. **Ratio-scale** variable:
   - Any **positive proportional** transformation captures the same information.
   - $C(h)$ is only invariant to positive proportional transformations, implying that health should be measured on a **ratio scale**.

Limitations of $C(h)$: Bounds

- Bounds of $C(h)$: $C\left(h\right) \in \left(-\frac{(n-1)}{n}; \frac{(n-1)}{n}\right)$.

Limitations of $C(h)$: Bounds

- Bounds of $C(h)$: $C\left(h\right) \in \left(-\frac{(n-1)}{n}; \frac{(n-1)}{n}\right)$.
- For any **bounded** health variable (with a finite upper value or a positive lower value), $C(h)$ has **varying bounds**:

$$-\frac{(b_h-\mu_h)(\mu_h-a_h)}{\mu_h(b_h-a_h)} \leq C(h) \leq \frac{(b_h-\mu_h)(\mu_h-a_h)}{\mu_h(b_h-a_h)}$$

where

Limitations of $C(h)$: Bounds

- Bounds of $C(h)$: $C(h) \in \left(-\frac{(n-1)}{n}; \frac{(n-1)}{n}\right)$.
- For any **bounded** health variable (with a finite upper value or a positive lower value), $C(h)$ has **varying bounds**:

$$-\frac{(b_h - \mu_h)(\mu_h - a_h)}{\mu_h(b_h - a_h)} \leq C(h) \leq \frac{(b_h - \mu_h)(\mu_h - a_h)}{\mu_h(b_h - a_h)}$$

where

a_h lower bound of the health variable h

Limitations of $C(h)$: Bounds

- Bounds of $C(h)$: $C\left(h\right) \in \left(-\frac{(n-1)}{n}; \frac{(n-1)}{n}\right)$.
- For any **bounded** health variable (with a finite upper value or a positive lower value), $C(h)$ has **varying bounds**:

$$-\frac{(b_h - \mu_h)(\mu_h - a_h)}{\mu_h(b_h - a_h)} \leq C(h) \leq \frac{(b_h - \mu_h)(\mu_h - a_h)}{\mu_h(b_h - a_h)}$$

where

a_h lower bound of the health variable h
b_h upper bound of the health variable h

Limitations of $C(h)$: Bounds

- Bounds of $C(h)$: $C\left(h\right) \in \left(-\frac{(n-1)}{n}; \frac{(n-1)}{n}\right)$.
- For any **bounded** health variable (with a finite upper value or a positive lower value), $C(h)$ has **varying bounds**:

$$-\frac{(b_h-\mu_h)(\mu_h-a_h)}{\mu_h(b_h-a_h)} \leq C(h) \leq \frac{(b_h-\mu_h)(\mu_h-a_h)}{\mu_h(b_h-a_h)}$$

where

a_h lower bound of the health variable h
b_h upper bound of the health variable h

and with

$$0 \leq a_h < b_h < +\infty.$$

Limitations of $C(h)$: Bounds

- Bounds of $C(h)$: $C(h) \in \left(-\frac{(n-1)}{n}; \frac{(n-1)}{n}\right)$.
- For any **bounded** health variable (with a finite upper value or a positive lower value), $C(h)$ has **varying bounds**:

$$-\frac{(b_h - \mu_h)(\mu_h - a_h)}{\mu_h(b_h - a_h)} \leq C(h) \leq \frac{(b_h - \mu_h)(\mu_h - a_h)}{\mu_h(b_h - a_h)}$$

where

a_h lower bound of the health variable h

b_h upper bound of the health variable h

and with

$$0 \leq a_h < b_h < +\infty.$$

- **Comparing** populations with **different mean** health levels is problematic, even if health is measured on a ratio-scale level.

Limitations of $C(h)$: Health and Ill-Health

- Example: health levels measure **health**, while malnutrition measures **ill health**.

Limitations of $C(h)$: Health and Ill-Health

- Example: health levels measure **health**, while malnutrition measures **ill health**.
- If the health variable h is **bounded**, define a corresponding ill health variable s:

$$
\begin{aligned}
s_i &\equiv b_h - h_i \\
\mu_s &= b_h - \mu_h
\end{aligned}
$$

Limitations of $C(h)$: Health and Ill-Health

- Example: health levels measure **health**, while malnutrition measures **ill health**.
- If the health variable h is **bounded**, define a corresponding ill health variable s:

$$\begin{aligned} s_i &\equiv b_h - h_i \\ \mu_s &= b_h - \mu_h \end{aligned}$$

where

Limitations of $C(h)$: Health and Ill-Health

- Example: health levels measure **health**, while malnutrition measures **ill health**.
- If the health variable h is **bounded**, define a corresponding ill health variable s:

$$\begin{aligned} s_i &\equiv b_h - h_i \\ \mu_s &= b_h - \mu_h \end{aligned}$$

where

s_i ill-health status of individual i;

Limitations of $C(h)$: Health and Ill-Health

- Example: health levels measure **health**, while malnutrition measures **ill health**.
- If the health variable h is **bounded**, define a corresponding ill health variable s:

$$\begin{aligned} s_i &\equiv b_h - h_i \\ \mu_s &= b_h - \mu_h \end{aligned}$$

where

s_i ill-health status of individual i;
μ_s average ill health of the population.

Limitations of $C(h)$: Health and Ill-Health

- Example: health levels measure **health**, while malnutrition measures **ill health**.
- If the health variable h is **bounded**, define a corresponding ill health variable s:

$$\begin{aligned} s_i &\equiv b_h - h_i \\ \mu_s &= b_h - \mu_h \end{aligned}$$

where

s_i ill-health status of individual i;
μ_s average ill health of the population.

- The ill-health Concentration Index $C(s)$ is defined by analogy with the health Concentration Index $C(h)$:

$$C(s) \equiv 1 - \frac{\sum_{i=1}^{n}(2\lambda_i - 1)s_i}{n^2\mu_s}$$

Health and Ill-Health

- Health and ill-health are *mirrors* of one another $\Rightarrow C(h)$ and $C(s)$ should give "mirror images" of inequality.

Health and Ill-Health

- Health and ill-health are *mirrors* of one another $\Rightarrow C(h)$ and $C(s)$ should give "mirror images" of inequality.
- Comparing the two:

$$C(s) = 1 - \frac{(\mu_h + \mu_s)}{\mu_s} + \frac{\mu_h}{\mu_s} \frac{\sum_{i=1}^{n} (2\lambda_i - 1) h_i}{n^2 \mu_h} = -\frac{\mu_h}{\mu_s} C(h)$$

Health and Ill-Health

- Health and ill-health are *mirrors* of one another $\Rightarrow C(h)$ and $C(s)$ should give "mirror images" of inequality.
- Comparing the two:

$$C(s) = 1 - \frac{(\mu_h + \mu_s)}{\mu_s} + \frac{\mu_h}{\mu_s}\frac{\sum_{i=1}^{n}(2\lambda_i - 1)h_i}{n^2\mu_h} = -\frac{\mu_h}{\mu_s}C(h)$$

- Consider two health distributions: h_1 and h_2.

Health and Ill-Health

- Health and ill-health are *mirrors* of one another $\Rightarrow C(h)$ and $C(s)$ should give "mirror images" of inequality.
- Comparing the two:

$$C(s) = 1 - \frac{(\mu_h + \mu_s)}{\mu_s} + \frac{\mu_h}{\mu_s} \frac{\sum_{i=1}^{n}(2\lambda_i - 1)h_i}{n^2 \mu_h} = -\frac{\mu_h}{\mu_s} C(h)$$

- Consider two health distributions: h_1 and h_2.
- If $\mu_{h_1} = \mu_{h_2}$, then $C(h_1) > C(h_2) \Leftrightarrow C(s_1) < C(s_2)$.

Health and Ill-Health

- Health and ill-health are *mirrors* of one another $\Rightarrow C(h)$ and $C(s)$ should give “mirror images” of inequality.
- Comparing the two:

$$C(s) = 1 - \frac{(\mu_h + \mu_s)}{\mu_s} + \frac{\mu_h}{\mu_s} \frac{\sum_{i=1}^{n} (2\lambda_i - 1) h_i}{n^2 \mu_h} = -\frac{\mu_h}{\mu_s} C(h)$$

- Consider two health distributions: h_1 and h_2.
- If $\mu_{h_1} = \mu_{h_2}$, then $C(h_1) > C(h_2) \Leftrightarrow C(s_1) < C(s_2)$.
- **Relative differences** correctly measured: $\frac{C(h_1)}{C(h_2)} = \frac{C(s_1)}{C(s_2)}$.

Health and Ill-Health

- Health and ill-health are *mirrors* of one another $\Rightarrow C(h)$ and $C(s)$ should give "mirror images" of inequality.
- Comparing the two:

$$C(s) = 1 - \frac{(\mu_h + \mu_s)}{\mu_s} + \frac{\mu_h}{\mu_s} \frac{\sum_{i=1}^{n}(2\lambda_i - 1)h_i}{n^2 \mu_h} = -\frac{\mu_h}{\mu_s} C(h)$$

- Consider two health distributions: h_1 and h_2.
- If $\mu_{h_1} = \mu_{h_2}$, then $C(h_1) > C(h_2) \Leftrightarrow C(s_1) < C(s_2)$.
- **Relative differences** correctly measured: $\frac{C(h_1)}{C(h_2)} = \frac{C(s_1)}{C(s_2)}$.
- If $\mu_{h_1} \neq \mu_{h_2}$, these properties no longer hold: inequalities in **ill health** may give different rankings than inequalities in **health**.

Introduction

- Define a generic **family of indices** I for some distribution h:

$$I(h) = f(a_h, b_h, \mu_h, n) \sum_{i=1}^{n} r_i h_i$$

for a continuous function $f(\cdot)$.

Introduction

- Define a generic **family of indices** I for some distribution h:

$$I(h) = f(a_h, b_h, \mu_h, n) \sum_{i=1}^{n} r_i h_i$$

  for a continuous function $f(\cdot)$.
- For unbounded variables, $b_h = +\infty$ and $f(\cdot) = f(a_h, \mu_h, n)$.

Introduction

- Define a generic **family of indices** I for some distribution h:

$$I(h) = f(a_h, b_h, \mu_h, n) \sum_{i=1}^{n} r_i h_i$$

  for a continuous function $f(\cdot)$.
- For unbounded variables, $b_h = +\infty$ and $f(\cdot) = f(a_h, \mu_h, n)$.
- We may impose restrictions on the form of $f(\cdot)$ depending on the properties that we want $I(h)$ to satisfy.

Sign Condition

- By convention positive (negative; 0) values of $I(h)$ should signal a pro-rich (pro-poor; no) **bias** in the distribution.

Sign Condition

- By convention positive (negative; 0) values of $I(h)$ should signal a pro-rich (pro-poor; no) **bias** in the distribution.
- **Sign condition**: the sign of $I(h)$ coincides with the sign of $\sum_{i=1}^{n} r_i h_i$.

Sign Condition

- By convention positive (negative; 0) values of $I(h)$ should signal a pro-rich (pro-poor; no) **bias** in the distribution.
- **Sign condition**: the sign of $I(h)$ coincides with the sign of $\sum_{i=1}^{n} r_i h_i$.

Proposition 1

The sign condition is satisfied if and only if:

- *For h **unbounded**: $f(a_h, \mu_h, n) > 0$ for $n > 0$ and $a_h < \mu_h < +\infty$;*
- *For h **bounded**: $f(a_h, b_h, \mu_h, n) > 0$ for $n > 0$ and $a_h < \mu_h < b_h$.*

Scale Invariance

- $I(h)$ should be independent of the unit of measurement of h.

Scale Invariance

- $I(h)$ should be independent of the unit of measurement of h.
- **Scale invariance**: for a change in the unit of measurement of h that transform the distribution h into $\tilde{h}$, $I(\tilde{h}) = I(h)$.

Scale Invariance

- $I(h)$ should be independent of the unit of measurement of h.
- **Scale invariance**: for a change in the unit of measurement of h that transform the distribution h into $\tilde{h}$, $I(\tilde{h}) = I(h)$.
    - If h **cardinal**: a positive linear transformation,

$$\tilde{h}_i = \alpha + \beta h_i; \qquad \tilde{a}_h = \alpha + \beta a_h; \qquad \tilde{b}_h = \alpha + \beta b_h.$$

Scale Invariance

- $I(h)$ should be independent of the unit of measurement of h.
- **Scale invariance**: for a change in the unit of measurement of h that transform the distribution h into $\tilde{h}$, $I(\tilde{h}) = I(h)$.
  - If h **cardinal**: a positive linear transformation,

$$\tilde{h}_i = \alpha + \beta h_i; \qquad \tilde{a}_h = \alpha + \beta a_h; \qquad \tilde{b}_h = \alpha + \beta b_h.$$

  - If h **ratio-scale**: a positive proportional transformation,

$$\tilde{h}_i = \beta h_i; \qquad \tilde{a}_h = \beta a_h; \qquad \tilde{b}_h = \beta b_h.$$

Proposition 2

$I(h)$ has the scale invariance property if and only if:

- *For h **unbounded**: $f(a_h, \mu_h, n) = \frac{1}{\mu_h - a_h} k(n)$;*
- *For h **bounded**: $f(a_h, b_h, \mu_h, n) = \frac{1}{b_h - a_h} g\left(\frac{\mu_h - a_h}{b_h - a_h}, n\right)$.*

Modified Concentration Index

- For h **unbounded**, if we fix $I(h)$'s bounds to -1 and 1, we can write a **modified** version $\hat{C}(h)$ of the standard **concentration index** $C(h)$:

$$\hat{C}(h) = \frac{2}{n^2\left(\mu_h - a_h\right)} \sum_{i=1}^{n} r_i h_i$$

that satisfies the sign condition and the scale invariance property.

Modified Concentration Index

- For h **unbounded**, if we fix $I(h)$'s bounds to -1 and 1, we can write a **modified** version $\hat{C}(h)$ of the standard **concentration index** $C(h)$:

$$\hat{C}(h) = \frac{2}{n^2\left(\mu_h - a_h\right)} \sum_{i=1}^{n} r_i h_i$$

  that satisfies the sign condition and the scale invariance property.
- For h **bounded**, it is more convenient to write a standardized definition h_i^* of h_i:

$$h_i^* \equiv \frac{h_i - a_h}{b_h - a_h}$$

  where $h_i^* \in [a_{h^*} = 0; b_{h^*} = 1]$ and $\mu_{h^*} = \frac{\mu_h - a_h}{b_h - a_h}$.

Modified Concentration Index

- For h **unbounded**, if we fix $I(h)$'s bounds to -1 and 1, we can write a **modified** version $\hat{C}(h)$ of the standard **concentration index** $C(h)$:

$$\hat{C}(h) = \frac{2}{n^2\left(\mu_h - a_h\right)} \sum_{i=1}^{n} r_i h_i$$

that satisfies the sign condition and the scale invariance property.

- For h **bounded**, it is more convenient to write a standardized definition h_i^* of h_i:

$$h_i^* \equiv \frac{h_i - a_h}{b_h - a_h}$$

where $h_i^* \in [a_{h^*} = 0; b_{h^*} = 1]$ and $\mu_{h^*} = \frac{\mu_h - a_h}{b_h - a_h}$.

- We can write a generic scale-invariant, rank dependent index $I(h^*)$ for bounded **cardinal** and **ratio-scale** variables:

$$I(h^*) = g(\mu_{h^*}, n) \sum_{i=1}^{n} r_i h_i^*$$

Mirror Property (Bounded Vars)

- The **ranking** of the distribution of the health index should be the opposite of the distribution of the corresponding ill-health index.

Mirror Property (Bounded Vars)

- The **ranking** of the distribution of the health index should be the opposite of the distribution of the corresponding ill-health index.
- **Mirror property**: for a health distribution h and the associated ill health distribution s, $I(h) = -I(s)$.

Wagstaff Index (Bounded Vars)

- Wagstaff (2005) suggested a **normalization** formula aimed at remedying the **bounds** issue.

Wagstaff Index (Bounded Vars)

- Wagstaff (2005) suggested a **normalization** formula aimed at remedying the **bounds** issue.
- We can define the **Wagstaff-normalized CI** $W(h)$ as:

$$
\begin{aligned}
W(h^*) &\equiv \frac{\mu_{h^*}(b_{h^*} - a_{h^*})}{(b_{h^*} - \mu_{h^*})(\mu_{h^*} - a_{h^*})} C(h^*) \\
&= \frac{2(b_{h^*} - a_{h^*})}{n^2(b_{h^*} - \mu_{h^*})(\mu_{h^*} - a_{h^*})} \sum_{i=1}^{n} r_i h_i^*
\end{aligned}
$$

Wagstaff Index (Bounded Vars)

- Wagstaff (2005) suggested a **normalization** formula aimed at remedying the **bounds** issue.
- We can define the **Wagstaff-normalized CI** $W(h)$ as:

$$\begin{aligned} W(h^*) &\equiv \frac{\mu_{h^*}(b_{h^*} - a_{h^*})}{(b_{h^*} - \mu_{h^*})(\mu_{h^*} - a_{h^*})} C(h^*) \\ &= \frac{2(b_{h^*} - a_{h^*})}{n^2(b_{h^*} - \mu_{h^*})(\mu_{h^*} - a_{h^*})} \sum_{i=1}^{n} r_i h_i^* \end{aligned}$$

- $W(h^*) \in [-1, 1]$.

Wagstaff Index (Bounded Vars)

- Wagstaff (2005) suggested a **normalization** formula aimed at remedying the **bounds** issue.
- We can define the **Wagstaff-normalized CI** $W(h)$ as:

$$
\begin{aligned}
W(h^*) &\equiv \frac{\mu_{h^*}(b_{h^*} - a_{h^*})}{(b_{h^*} - \mu_{h^*})(\mu_{h^*} - a_{h^*})} C(h^*) \\
&= \frac{2(b_{h^*} - a_{h^*})}{n^2(b_{h^*} - \mu_{h^*})(\mu_{h^*} - a_{h^*})} \sum_{i=1}^{n} r_i h_i^*
\end{aligned}
$$

- $W(h^*) \in [-1, 1]$.
- $W(h^*)$ satisfies the **mirror condition**.

Wagstaff Index (Bounded Vars)

- Wagstaff (2005) suggested a **normalization** formula aimed at remedying the **bounds** issue.
- We can define the **Wagstaff-normalized CI** $W(h)$ as:

$$
\begin{aligned}
W(h^*) &\equiv \frac{\mu_{h^*}(b_{h^*} - a_{h^*})}{(b_{h^*} - \mu_{h^*})(\mu_{h^*} - a_{h^*})} C(h^*) \\
&= \frac{2(b_{h^*} - a_{h^*})}{n^2(b_{h^*} - \mu_{h^*})(\mu_{h^*} - a_{h^*})} \sum_{i=1}^{n} r_i h_i^*
\end{aligned}
$$

- $W(h^*) \in [-1, 1]$.
- $W(h^*)$ satisfies the **mirror condition**.
- Advantage over $C(h^*)$: $W(h^*)$ is invariant to a positive linear transformation of h^*, so h^* can be also measured on a **cardinal scale**.

Erreygers Index (Bounded Vars)

- Erreygers (2009) proposes the following index:

$$E(h^*) = \frac{8}{n^2(b_{h^*} - a_{h^*})} \sum_{i=1}^{n} r_i h_i^*$$

Erreygers Index (Bounded Vars)

- Erreygers (2009) proposes the following index:

$$E(h^*) = \frac{8}{n^2(b_{h^*} - a_{h^*})} \sum_{i=1}^{n} r_i h_i^*$$

- Formally $E(h^*)$ and $W(h^*)$ differ only with respect to the **normalization** applied to the weighted sum of h.

Erreygers Index (Bounded Vars)

- Erreygers (2009) proposes the following index:

$$E(h^*) = \frac{8}{n^2(b_{h^*} - a_{h^*})} \sum_{i=1}^{n} r_i h_i^*$$

- Formally $E(h^*)$ and $W(h^*)$ differ only with respect to the **normalization** applied to the weighted sum of h.
- $E(h^*)$ is the only index that is insensitive to any feasible **equal addition** to the standardized health variable h^*:

Erreygers Index (Bounded Vars)

- Erreygers (2009) proposes the following index:

$$E(h^*) = \frac{8}{n^2(b_{h^*} - a_{h^*})} \sum_{i=1}^{n} r_i h_i^*$$

- Formally $E(h^*)$ and $W(h^*)$ differ only with respect to the **normalization** applied to the weighted sum of h.
- $E(h^*)$ is the only index that is insensitive to any feasible **equal addition** to the standardized health variable h^*:
  - an equal increment of h^* for all individuals – **keeping its bounds constant** – does not affect the value of the index, all other things equal.

Erreygers Index (Bounded Vars)

- Erreygers (2009) proposes the following index:

$$E(h^*) = \frac{8}{n^2(b_{h^*} - a_{h^*})} \sum_{i=1}^{n} r_i h_i^*$$

- Formally $E(h^*)$ and $W(h^*)$ differ only with respect to the **normalization** applied to the weighted sum of h.
- $E(h^*)$ is the only index that is insensitive to any feasible **equal addition** to the standardized health variable h^*:
  - an equal increment of h^* for all individuals – **keeping its bounds constant** – does not affect the value of the index, all other things equal.
- However, there is a debate in the literature on whether one index between $E(h^*)$ and $W(h^*)$ is superior to the other.

Indices and Their Properties: A Recap

Table 1. Concentration indices and their properties by level of measurement of the health variable.

	Variable Range			
	Unbounded		*Bounded*	
Variable Level	Index	Property	Index	Property
Ordinal	Concentration index and its variants in principle meaningless			
Cardinal	Modified CI	Sign condition Scale invariance	Erreygers and Wagstaff indices	Sign condition Scale invariance
Ratio-scale	Concentration index			Mirror property

Source: Erreygers and Van Ourti (2011).

Comparing Indices: An Example

- **Stunting** is a (ill-) health measure reflecting chronic **malnutrition**.

Comparing Indices: An Example

- **Stunting** is a (ill-) health measure reflecting chronic **malnutrition**.
- Binary **ill-health** variable s: $s_i = 1$ if child i is stunted, 0 otherwise.

Comparing Indices: An Example

- **Stunting** is a (ill-) health measure reflecting chronic **malnutrition**.
- Binary **ill-health** variable s: $s_i = 1$ if child i is stunted, 0 otherwise.
- **Health** variable h: $h_i = 1$ if child i is not stunted, 0 otherwise.

Comparing Indices: An Example

- **Stunting** is a (ill-) health measure reflecting chronic **malnutrition**.
- Binary **ill-health** variable s: $s_i = 1$ if child i is stunted, 0 otherwise.
- **Health** variable h: $h_i = 1$ if child i is not stunted, 0 otherwise.

Table 2. Stunting in some developing countries. Source: Erreygers (2009).

Country	μ_s	μ_h	$-C(s)$	$C(h)$	$W(h)$	$E(h)$
Nigeria (2003)	0.3845	0.6155	0.1612	**0.1007**	**0.2619**	0.2479
Cameroon (2004)	0.3165	0.6835	0.1698	**0.0786**	**0.2484**	0.2150
Kenya (2003)	0.3056	0.6944	0.1265	0.0557	0.1822	0.1546
Ghana (2003)	0.2943	0.7057	0.1743	**0.0727**	**0.2470**	0.2052
Cambodia (2000)	0.2943	0.7057	0.0887	0.0370	0.1257	0.1044
Bolivia (2003)	0.0680	0.9320	0.2739	**0.0200**	**0.2939**	**0.0745**
Peru (2000)	0.0501	0.9499	0.3676	**0.0194**	**0.3870**	**0.0737**
Nicaragua (2001)	0.0422	0.9578	0.3304	**0.0146**	**0.3450**	**0.0558**
Colombia (2005)	0.0215	0.9785	0.2699	**0.0059**	**0.2758**	**0.0232**

Decomposition of the Concentration Index

Introduction

- Decomposition of the CI: provide a specification for the health outcome:

$$h_{it} = \alpha_t + \sum_{j=1}^{J} \beta_j X_{jit} + \varepsilon_{it}. \tag{1}$$

Introduction

- Decomposition of the CI: provide a specification for the health outcome:

$$h_{it} = \alpha_t + \sum_{j=1}^{J} \beta_j X_{jit} + \varepsilon_{it}. \tag{1}$$

- Consider the Erreygers index and adjust it to include multiple observations per individual (Kjellsson, 2018):

$$E(h) = \frac{8}{(nt)^2} \sum_{t=1}^{T} \sum_{i=1}^{n} r_i h_i \tag{2}$$

Decomposition of the Concentration Index

- Substitute (1) in (2) to obtain

$$E(h) = 4 \sum_{j=1}^{J} \beta_j V(X_j) + 4V^{\varepsilon} \tag{3}$$

where

$$V(X_j) = \frac{2}{(nt)^2} \sum_{t=1}^{T} \sum_{i=1}^{n} r_i X_{jit} \qquad V^{\varepsilon} = \frac{2}{(nt)^2} \sum_{t=1}^{T} \sum_{i=1}^{n} r_i \varepsilon_{it} \tag{4}$$

denote the generalized concentration indices for X_j and the error term, respectively.

Decomposition of the Concentration Index

- Substitute (1) in (2) to obtain

$$E(h) = 4\sum_{j=1}^{J}\beta_j V(X_j) + 4V^{\varepsilon} \tag{3}$$

where

$$V(X_j) = \frac{2}{(nt)^2}\sum_{t=1}^{T}\sum_{i=1}^{n} r_i X_{jit} \qquad V^{\varepsilon} = \frac{2}{(nt)^2}\sum_{t=1}^{T}\sum_{i=1}^{n} r_i \varepsilon_{it} \tag{4}$$

denote the generalized concentration indices for X_j and the error term, respectively.

- So $E(h)$ is the weighted sum of the **CI for the J regressors** plus a **residual component**.

Decomposition of the Concentration Index

- Substitute (1) in (2) to obtain

$$E(h) = 4\sum_{j=1}^{J}\beta_j V(X_j) + 4V^{\varepsilon} \tag{3}$$

where

$$V(X_j) = \frac{2}{(nt)^2}\sum_{t=1}^{T}\sum_{i=1}^{n} r_i X_{jit} \qquad V^{\varepsilon} = \frac{2}{(nt)^2}\sum_{t=1}^{T}\sum_{i=1}^{n} r_i \varepsilon_{it} \tag{4}$$

denote the generalized concentration indices for X_j and the error term, respectively.

- So $E(h)$ is the weighted sum of the **CI for the J regressors** plus a **residual component**.
- A similar decomposition applies to the other indices.

Empirical Application: Kjellsson (2018)

- In a recent study, Kjellsson (2018) investigates **income-related smoking inequality** among Swedish women.

Empirical Application: Kjellsson (2018)

- In a recent study, Kjellsson (2018) investigates **income-related smoking inequality** among Swedish women.
- The author uses a random effects probit model to determine the β_j coefficients from equation (1).

Empirical Application: Kjellsson (2018)

- In a recent study, Kjellsson (2018) investigates **income-related smoking inequality** among Swedish women.
- The author uses a random effects probit model to determine the β_j coefficients from equation (1).
- To account for **persistence** in smoking behavior, **lagged outcomes** h_{it-1} are included in the regressions.

Empirical Application: Kjellsson (2018)

- In a recent study, Kjellsson (2018) investigates **income-related smoking inequality** among Swedish women.
- The author uses a random effects probit model to determine the β_j coefficients from equation (1).
- To account for **persistence** in smoking behavior, **lagged outcomes** h_{it-1} are included in the regressions.
- Within-individual means of time-variant variables allow to control for unobserved time-invariant heterogeneity (**Mundlak**-type specification).

Empirical Application: Kjellsson (2018)

- In a recent study, Kjellsson (2018) investigates **income-related smoking inequality** among Swedish women.
- The author uses a random effects probit model to determine the β_j coefficients from equation (1).
- To account for **persistence** in smoking behavior, **lagged outcomes** h_{it-1} are included in the regressions.
- Within-individual means of time-variant variables allow to control for unobserved time-invariant heterogeneity (**Mundlak**-type specification).
- The results show an income related smoking inequality in favor of the rich ($E = -0.084$) which is persistent over time.

Empirical Application: Kjellsson (2018)

- In a recent study, Kjellsson (2018) investigates **income-related smoking inequality** among Swedish women.
- The author uses a random effects probit model to determine the β_j coefficients from equation (1).
- To account for **persistence** in smoking behavior, **lagged outcomes** h_{it-1} are included in the regressions.
- Within-individual means of time-variant variables allow to control for unobserved time-invariant heterogeneity (**Mundlak**-type specification).
- The results show an income related smoking inequality in favor of the rich ($E = -0.084$) which is persistent over time.
- Main drivers are **education** and living in a **single-adult household**.

Empirical Application: Kjellsson (2018)

	V_k	Static RE probit (Mundlak) PE	Contribution	%	Dynamic RE probit (Mundlak) PE	Contribution	%
fath_white_high	0.034	−0.051	−0.007	8.4	−0.015	−0.002	2.5
	(.000)	(.024)	(.045)		(.403)	(.417)	
fath_white_low	0.020	−0.029	−0.002	2.8	−0.002	−0.000	0.2
	(.001)	(.270)	(.312)		(.932)	(.936)	
fath_farm	−0.023	−0.057	0.005	−6.2	−0.005	0.000	−0.6
	(.000)	(.038)	(.082)		(.812)	(.817)	
im2	−0.005	0.057	−0.001	1.5	0.011	−0.000	0.3
	(.113)	(.198)	(.356)		(.699)	(.753)	
cohort40	0.013	0.074	0.004	−4.7	0.037	0.002	−2.4
	(.036)	(.030)	(.113)		(.113)	(.202)	
cohort50	0.025	0.135	0.013	−15.9	0.108	0.011	−12.7
	(.000)	(.025)	(.050)		(.018)	(.038)	
cohort60	−0.015	0.100	−0.006	7.3	0.071	−0.004	5.2
	(.037)	(.253)	(.326)		(.290)	(.391)	
age	−0.104	−0.007	0.003	−3.4	−0.002	0.001	−1.2
	(.611)	(.048)	(.653)		(.368)	(.745)	
yrsschool	0.540	−0.026	−0.056	66.6	−0.015	−0.033	39.0
	(.000)	(.000)	(.000)		(.000)	(.000)	
child1	0.004	−0.019	−0.000	0.4	−0.015	−0.000	0.3
	(.224)	(.116)	(.392)		(.353)	(.533)	
child2plus	−0.002	−0.041	0.000	−0.4	−0.035	0.000	−0.3
	(.749)	(.006)	(.753)		(.040)	(.761)	
single	−0.068	0.032	−0.009	10.3	0.050	−0.014	16.3
	(.000)	(.032)	(.034)		(.018)	(.020)	
lninc	0.156	0.007	0.004	−4.9	0.009	0.006	−6.6
	(.000)	(.572)	(.571)		(.547)	(.546)	
ln_LIFEinc	0.172	−0.004	−0.003	3.4	−0.019	−0.013	15.7
	(.000)	(.905)	(.904)		(.470)	(.467)	
m_child2plus	0.004	−0.076	−0.001	1.5	−0.036	−0.001	0.7
	(.224)	(.141)	(.411)		(.347)	(.524)	
m_child1	−0.002	−0.195	0.001	−1.7	−0.074	0.001	−0.7
	(.749)	(.000)	(.766)		(.064)	(.791)	
m_single	−0.068	0.052	−0.014	16.9	−0.012	0.003	−3.8
	(.000)	(.079)	(.082)		(.674)	(.676)	
Y_{i0}	−0.018				0.450	−0.032	37.8
	(.006)				(.000)	(.008)	
Y_{it-1}	−0.021				0.092	−0.008	9.1
	(.000)				(.001)	(.014)	
Residual			−0.012	14.7		0.000	0.2
Erreygers' index			−0.084	100		−0.084	100

Level-Dependent Indices

Introduction

- Literature on decomposition of health inequality mainly focuses on **rank-dependent indices**.

Introduction

- Literature on decomposition of health inequality mainly focuses on **rank-dependent indices**.
- However, if one is interested in decomposing an inequality index I into

Introduction

- Literature on decomposition of health inequality mainly focuses on **rank-dependent indices**.
- However, if one is interested in decomposing an inequality index I into
    - within-subgroup inequality I_W

Introduction

- Literature on decomposition of health inequality mainly focuses on **rank-dependent indices**.
- However, if one is interested in decomposing an inequality index I into
  - within-subgroup inequality I_W
  - and the between-subgroup inequality I_B

  rank-dependent indices usually produce a non-zero residual term $I_X = I - I_W - I_B$.

Introduction

- Literature on decomposition of health inequality mainly focuses on **rank-dependent indices**.
- However, if one is interested in decomposing an inequality index I into
  - within-subgroup inequality I_W
  - and the between-subgroup inequality I_B

  rank-dependent indices usually produce a non-zero residual term $I_X = I - I_W - I_B$.
- Erreygers et al. (2018) developed a level-dependent index with the property of **subgroup decomposability**.

Introduction

Within-subgroup inequality I_W:

$$I_j = \frac{1}{n_j} \sum_{i \in G_j} w_i\left(\mathbf{y}_j\right) h_i \tag{5}$$

$$I_W = \sum_{j=1}^{k} s_j I_j \tag{6}$$

where

I_j	inequality index for subgroup j
n_j	number of individuals in subgroup j
G_j	individuals in subgroup j
$w_i\left(\mathbf{y}_j\right)$	weight of individual i for position within group j
h_i	health status of individual i
s_j	weight of subgroup j.

Introduction

Between-subgroup inequality I_B:

$$I_B = \frac{1}{n} \sum_{j=1}^{k} n_j w_j \left(\boldsymbol{\mu}_y \right) \mu_{h_j} \tag{7}$$

where

n total number of individuals

$w_j \left(\boldsymbol{\mu}_y \right)$ weight reflecting average situation of individuals in subgroup j

μ_{h_j} mean health status in subgroup j.

Rank-Dependent Measures

The standard concentration index I^R can be written as

$$I^R = \frac{1}{n}\sum_{i=1}^{n} w_i\left(\mathbf{y}\right) h_i \tag{8}$$

with

$$w_i\left(\mathbf{y}\right) = \frac{2r_i\left(\mathbf{y}\right) - n - 1}{n} \tag{9}$$

where

$w_i\left(\mathbf{y}\right)$ weight reflecting situation of individual i in total population.

Rank-Dependent Measures

Defining

$$\begin{aligned} w_i\left(\mathbf{y}_j\right) &= \frac{2r_i\left(\mathbf{y}_j\right) - n_j - 1}{n_j} \\ w_j\left(\boldsymbol{\mu}_y\right) &= \frac{2r_j\left(\boldsymbol{\mu}_y\right) - n - 1}{n} \end{aligned}$$

with

$$r_j\left(\boldsymbol{\mu}_y\right) = \frac{n_j + 1}{2} + \sum_{l=0}^{j-1} n_l, \tag{10}$$

$\mu_{y_1} < \ldots < \mu_{y_k}$, $s_j = \frac{n_j}{n}^2$ and $n_0 = 0$, the residual term

$$I_X^R = \frac{1}{n} \sum_{j=1}^{k} \sum_{i \in G_j} \left[w_i\left(\mathbf{y}\right) h_i - \frac{n_j}{n} w_i\left(\mathbf{y}_j\right) h_i - w_j\left(\boldsymbol{\mu}_y\right) \mu_{h_j} \right] \tag{11}$$

only equals zero if the subgroup income ranges do not overlap.

Level-Dependent Measures

If we instead consider **level-dependent** weights

$$\begin{aligned}
w_i\left(\mathbf{y}\right) &= \frac{y_i - \mu_y}{\mu_y} \\
w_i\left(\mathbf{y}_j\right) &= \frac{y_i - \mu_{y_j}}{\mu_{y_j}} \\
w_j\left(\boldsymbol{\mu}_y\right) &= \frac{\mu_{y_j} - \mu_y}{\mu_y},
\end{aligned}$$

we obtain the total population inequality I^L as the sum of the within- (I_W^L) and the between-subgroup inequality (I_B^L):

$$\begin{aligned}
I^L &= I_W^L + I_B^L \\
&= \frac{1}{n\mu_y}\sum_{i=1}^{n} y_i h_i - \mu_h
\end{aligned}$$

Comparing Decompositions: An Example

- Erreygers et al. (2018) estimate inequality in health – measured as SF-6D health score – due to equivalised income for Australian population aged 15+.

Source: Erreygers et al. (2018).

Comparing Decompositions: An Example

- Erreygers et al. (2018) estimate inequality in health – measured as SF-6D health score – due to equivalised income for Australian population aged 15+.
- Decompose inequality according to sex and compare rank- and level-dependent indices.

Table 3. Decomposition of health inequality by sex.

	I^R		I^L	
	Values	%	Values	%
Within	0.0325	49.47	0.0141	98.30
Between	0.0210	31.92	0.0002	1.70
Residual	0.0122	18.61	-	-
Total	0.0657	100.00	0.0144	100.00

Source: Erreygers et al. (2018).

Two-Dimensional Decomposition

Introduction

- The previously introduced CI decomposition relies on specifying the determinants of health h_i.

Introduction

- The previously introduced CI decomposition relies on specifying the determinants of health h_i.
- However, **determinants of socioeconomic status** are ignored by this approach.

Introduction

- The previously introduced CI decomposition relies on specifying the determinants of health h_i.
- However, **determinants of socioeconomic status** are ignored by this approach.
- Kessels and Erreygers (2019) propose an approach which takes determinants of both health and SES into account.

Introduction

- The previously introduced CI decomposition relies on specifying the determinants of health h_i.
- However, **determinants of socioeconomic status** are ignored by this approach.
- Kessels and Erreygers (2019) propose an approach which takes determinants of both health and SES into account.
- This direct regression approach is easy to implement and applicable for rank- and level-dependent indices.

Implementation

- We can rewrite the specific index I as

$$I = \frac{1}{n}\sum_{i=1}^{n} w_i h_i = \frac{1}{n}\sum_{i=1}^{n} u_i = \mu_u \qquad (12)$$

where w_i are the specific index's weights.

Implementation

- We can rewrite the specific index I as

$$I = \frac{1}{n}\sum_{i=1}^{n} w_i h_i = \frac{1}{n}\sum_{i=1}^{n} u_i = \mu_u \tag{12}$$

  where w_i are the specific index's weights.
- The u_i values combines both health and socioeconomic performance and can be written as

$$u_i = \beta_0 + \sum_{j=1}^{J} \beta_j x_{ij} + \eta_i \tag{13}$$

  where x_{ij} are determinants of health and/or SES.

Implementation

- We can rewrite the specific index I as

$$I = \frac{1}{n}\sum_{i=1}^{n} w_i h_i = \frac{1}{n}\sum_{i=1}^{n} u_i = \mu_u \tag{12}$$

  where w_i are the specific index's weights.
- The u_i values combines both health and socioeconomic performance and can be written as

$$u_i = \beta_0 + \sum_{j=1}^{J} \beta_j x_{ij} + \eta_i \tag{13}$$

  where x_{ij} are determinants of health and/or SES.
- Estimating 13 via OLS and inserting into 12, we obtain

$$I = \widehat{\beta}_0 + \sum_{j=1}^{J} \widehat{\beta}_j \mu_{x_j} \tag{14}$$

Interpretation

- For marginal changes $\Delta\mu_{x_j}$, the effect on I is equal to $\widehat{\beta}_j\Delta\mu_{x_j}$.

Interpretation

- For marginal changes $\Delta\mu_{x_j}$, the effect on I is equal to $\widehat{\beta}_j \Delta\mu_{x_j}$.
- $\widehat{\beta}_j \mu_{x_j}$ is **not** the contribution of x_j to the level of inequality.

Interpretation

- For marginal changes $\Delta\mu_{x_j}$, the effect on I is equal to $\widehat{\beta}_j\Delta\mu_{x_j}$.
- $\widehat{\beta}_j\mu_{x_j}$ is **not** the contribution of x_j to the level of inequality.
- The importance of specific variables can be evaluated via the logworth statistic defined as

$$\text{Logworth} = -\log_{10}(p) \tag{15}$$

where p is the p value of the F test conducted on a certain set of variables.

Interpretation

- For marginal changes $\Delta\mu_{x_j}$, the effect on I is equal to $\widehat{\beta}_j \Delta\mu_{x_j}$.
- $\widehat{\beta}_j \mu_{x_j}$ is **not** the contribution of x_j to the level of inequality.
- The importance of specific variables can be evaluated via the logworth statistic defined as

$$\text{Logworth} = -\log_{10}(p) \tag{15}$$

  where p is the p value of the F test conducted on a certain set of variables.
- Kessels and Erreygers (2019) apply this approach to Australian data with SF-6D as health indicator and equivalent income to denote SES.

Empirical Application - Kessels and Erreygers (2019)

Variable	$\hat{\chi}_j$	Prob $> \lvert t \rvert$	Prob $> F$	Logworth
Male	0.0260	0.0033		2.488
Indigenous	−0.0967	0.0004		3.418
Age	−0.0015	< 0.0001		7.027
Not married	−0.1145	< 0.0001		30.566
Children 0-4	−0.0653	< 0.0001		12.526
Children 5-14	−0.0450	< 0.0001		12.480
Semi-detached house	0.0065	0.6999		
Flat	−0.0007	0.9611		
Non-private dwelling	−0.2691	0.0016		
Other dwelling	−0.1340	0.0249	0.0044	2.361
Managers & professionals	0.2266	< 0.0001		
Manual workers	−0.0893	< 0.0001		
Unemployed	−0.1702	< 0.0001		
Not in labour force	−0.2250	< 0.0001	< 0.0001	279.915
Living poorly	−0.2272	< 0.0001		
Just getting along	−0.2163	< 0.0001	< 0.0001	103.965
Smoking	−0.0481	< 0.0001		4.521
Very good sleep quality	0.0232	0.0398		
Fairly bad sleep quality	−0.0283	0.0113		
Very bad sleep quality	−0.0547	0.0155		
Not reported	−0.0505	0.1951	0.0006	3.213
Almost always stressed	−0.0197	0.2252		
Often stressed	−0.0049	0.6489		
Rarely stressed	−0.0302	0.0090		
Never stressed	−0.0276	0.2919	0.0895	1.048
Life satisfaction	0.0301	< 0.0001		18.983
Very satisfied with weight	0.0252	0.1268		
Satisfied with weight	−0.0072	0.5408		
Dissatisfied with weight	0.0342	0.0028		
Very dissatisfied with weight	−0.0022	0.8994	0.0012	2.910
No physical activity	−0.0886	< 0.0001		
Some physical activity	−0.0470	< 0.0001	< 0.0001	10.845
Constant	0.0368	0.2773		
R^2	0.2111			

Figure 1. Results for Level-Dependent Index

Summary and Conclusions

Summary and Conclusions

- The properties of the concentration index depend on the measurement **characteristics** of the health variable of interest.

Summary and Conclusions

- The properties of the concentration index depend on the measurement **characteristics** of the health variable of interest.
- When the health variable is cardinal and has finite upper and lower bounds, the **Erreygers index** $E(h)$ and the **Wagstaff index** $W(h)$ satisfy the desired properties (sign, scale invariance, mirror property) and are superior to the CI.

Summary and Conclusions

- The properties of the concentration index depend on the measurement **characteristics** of the health variable of interest.
- When the health variable is cardinal and has finite upper and lower bounds, the **Erreygers index** $E(h)$ and the **Wagstaff index** $W(h)$ satisfy the desired properties (sign, scale invariance, mirror property) and are superior to the CI.
- Level-dependent indices allow for decomposition into within- and between-subgroup inequality.

Summary and Conclusions

- The properties of the concentration index depend on the measurement **characteristics** of the health variable of interest.
- When the health variable is cardinal and has finite upper and lower bounds, the **Erreygers index** $E(h)$ and the **Wagstaff index** $W(h)$ satisfy the desired properties (sign, scale invariance, mirror property) and are superior to the CI.
- Level-dependent indices allow for decomposition into within- and between-subgroup inequality.
- Applying decomposition methods to inequality indicators (like the CI) allows to analyse income-related inequalities in health across the **entire** income **distribution** (income proxying for SES).

Summary and Conclusions

- The properties of the concentration index depend on the measurement **characteristics** of the health variable of interest.
- When the health variable is cardinal and has finite upper and lower bounds, the **Erreygers index** $E(h)$ and the **Wagstaff index** $W(h)$ satisfy the desired properties (sign, scale invariance, mirror property) and are superior to the CI.
- Level-dependent indices allow for decomposition into within- and between-subgroup inequality.
- Applying decomposition methods to inequality indicators (like the CI) allows to analyse income-related inequalities in health across the **entire** income **distribution** (income proxying for SES).
- In this case, each source of inequality is quantified – and not just the difference between the two groups.

Literature I

Erreygers, G. (2009): "Correcting the concentration index," *Journal of health economics*, 28, 504–515.

Erreygers, G., R. Kessels, L. Chen, and P. Clarke (2018): "Subgroup Decomposability of Income-Related Inequality of Health, with an Application to Australia," *Economic Record*, 94, 39–50.

Erreygers, G. and T. Van Ourti (2011): "Measuring socioeconomic inequality in health, health care and health financing by means of rank-dependent indices: a recipe for good practice," *Journal of health economics*, 30, 685–694.

Kessels, R. and G. Erreygers (2019): "A direct regression approach to decomposing socioeconomic inequality of health," *Health economics*, 28, 884–905.

Kjellsson, G. (2018): "Extending decomposition analysis to account for unobserved heterogeneity and persistence in health behavior: Income-related smoking inequality among Swedish women," *Health Economics*, 27, 440–447.

Wagstaff, A. (2005): "The bounds of the concentration index when the variable of interest is binary, with an application to immunization inequality," *Health economics*, 14, 429–432.

Change in the CI: Oaxaca-Style Approach

- Suppose we want to compare and decompose the **differences** between concentration indices of two populations, for example to analyse how inequality changes in one country over time t.

Change in the CI: Oaxaca-Style Approach

- Suppose we want to compare and decompose the **differences** between concentration indices of two populations, for example to analyse how inequality changes in one country over time t.
- Consider the decomposition of the CI:

$$C = \sum_k \eta_k C_k + \frac{GC_\epsilon}{\mu} \tag{16}$$

Change in the CI: Oaxaca-Style Approach

- Suppose we want to compare and decompose the **differences** between concentration indices of two populations, for example to analyse how inequality changes in one country over time t.
- Consider the decomposition of the CI:

$$C = \sum_k \eta_k C_k + \frac{GC_\epsilon}{\mu} \tag{16}$$

- We may apply a Oaxaca-style decomposition:

$$\Delta C = \sum_k \eta_{kt} \left(C_{kt} - C_{k,t-1}\right) + \sum_k C_{k,t-1} \left(\eta_{kt} - \eta_{k,t-1}\right) + \Delta \left(\frac{GC_{\eta t}}{\mu_t}\right) \tag{17}$$

Change in the CI: Oaxaca-Style Approach

- Suppose we want to compare and decompose the **differences** between concentration indices of two populations, for example to analyse how inequality changes in one country over time t.
- Consider the decomposition of the CI:

$$C = \sum_k \eta_k C_k + \frac{GC_\epsilon}{\mu} \tag{16}$$

- We may apply a Oaxaca-style decomposition:

$$\Delta C = \sum_k \eta_{kt} \left(C_{kt} - C_{k,t-1} \right) + \sum_k C_{k,t-1} \left(\eta_{kt} - \eta_{k,t-1} \right) + \Delta \left(\frac{GC_{\eta t}}{\mu_t} \right) \tag{17}$$

- It gives us differences in inequality at different t as a sum of:

Change in the CI: Oaxaca-Style Approach

- Suppose we want to compare and decompose the **differences** between concentration indices of two populations, for example to analyse how inequality changes in one country over time t.
- Consider the decomposition of the CI:

$$C = \sum_k \eta_k C_k + \frac{GC_\epsilon}{\mu} \tag{16}$$

- We may apply a Oaxaca-style decomposition:

$$\Delta C = \sum_k \eta_{kt} \left(C_{kt} - C_{k,t-1} \right) + \sum_k C_{k,t-1} \left(\eta_{kt} - \eta_{k,t-1} \right) + \Delta \left(\frac{GC_{\eta t}}{\mu_t} \right) \tag{17}$$

- It gives us differences in inequality at different t as a sum of:
    1. CIs for determinants k weighted by their elasticities.

Change in the CI: Oaxaca-Style Approach

- Suppose we want to compare and decompose the **differences** between concentration indices of two populations, for example to analyse how inequality changes in one country over time t.
- Consider the decomposition of the CI:

$$C = \sum_k \eta_k C_k + \frac{GC_\epsilon}{\mu} \tag{16}$$

- We may apply a Oaxaca-style decomposition:

$$\Delta C = \sum_k \eta_{kt} \left(C_{kt} - C_{k,t-1}\right) + \sum_k C_{k,t-1} \left(\eta_{kt} - \eta_{k,t-1}\right) + \Delta \left(\frac{GC_{\eta t}}{\mu_t}\right) \tag{17}$$

- It gives us differences in inequality at different t as a sum of:
  1. CIs for determinants k weighted by their elasticities.
  2. Elasticities weighted by the respective CIs.

Change in the CI: Oaxaca-Style Approach

- Suppose we want to compare and decompose the **differences** between concentration indices of two populations, for example to analyse how inequality changes in one country over time t.
- Consider the decomposition of the CI:

$$C = \sum_k \eta_k C_k + \frac{GC_\epsilon}{\mu} \tag{16}$$

- We may apply a Oaxaca-style decomposition:

$$\Delta C = \sum_k \eta_{kt} \left(C_{kt} - C_{k,t-1} \right) + \sum_k C_{k,t-1} \left(\eta_{kt} - \eta_{k,t-1} \right) + \Delta \left(\frac{GC_{\eta t}}{\mu_t} \right) \tag{17}$$

- It gives us differences in inequality at different t as a sum of:
  1. CIs for determinants k weighted by their elasticities.
  2. Elasticities weighted by the respective CIs.
  3. Generalised CIs of the residuals.

Example: Change in the CI, Different Components

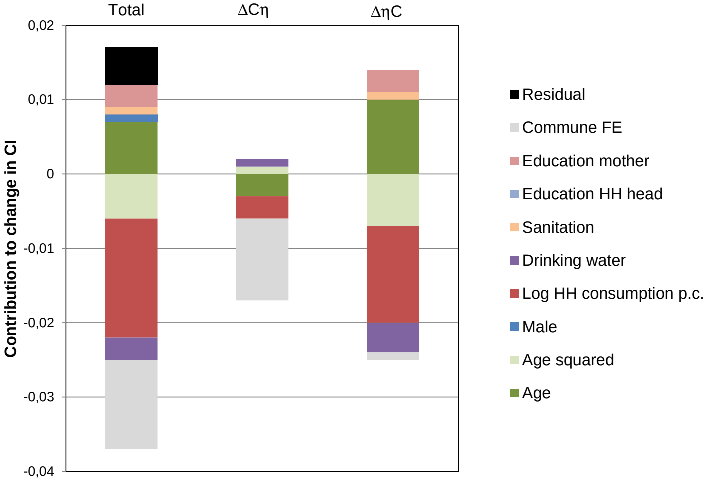

Figure 2. Decomposition of change in CI for HAZ-scores of children < 10 y.o. in Vietnam, 1993-98.

Change in the CI: Total Differential Approach

- However decomposition (17) does not allow to distinguish how changes within the elasticities η_k affect changes in socioeconomic-related inequality.

Change in the CI: Total Differential Approach

- However decomposition (17) does not allow to distinguish how changes within the elasticities η_k affect changes in socioeconomic-related inequality.
- Consider the derivatives w.r.t. changes in β_k and $\bar{x}_k$:

$$\frac{\mathrm{d}C}{\mathrm{d}\beta_k} = \frac{\partial C}{\partial \beta_k} + \frac{\partial C}{\partial \mu}\frac{\mathrm{d}\mu}{\mathrm{d}\beta_k} = \frac{\bar{x}_k C_k}{\mu} - \frac{\bar{x}_k}{\mu} C$$

$$\frac{\mathrm{d}C}{\mathrm{d}\bar{x}_k} = \frac{\beta_k}{\mu}\left(C_k - C\right)$$

Change in the CI: Total Differential Approach

- Get the **total differential** of Eq. (16):

$$\mathrm{d}C = -\frac{C}{\mu}\mathrm{d}\alpha + \sum_k \frac{\bar{x}_k}{\mu}\left(C_k - C\right)\mathrm{d}\beta_k + \sum_k \frac{\beta_k}{\mu}\left(C_k - C\right)\mathrm{d}\bar{x}_k + \sum_k \frac{\beta_k \bar{x}_k}{\mu}\mathrm{d}C_k + \mathrm{d}\frac{GC_\epsilon}{\mu} \tag{18}$$

Change in the CI: Total Differential Approach

- Get the **total differential** of Eq. (16):

$$\mathrm{d}C = -\frac{C}{\mu}\mathrm{d}\alpha + \sum_k \frac{\bar{x}_k}{\mu}\left(C_k - C\right)\mathrm{d}\beta_k + \sum_k \frac{\beta_k}{\mu}\left(C_k - C\right)\mathrm{d}\bar{x}_k + \sum_k \frac{\beta_k \bar{x}_k}{\mu}\mathrm{d}C_k + \mathrm{d}\frac{GC_\epsilon}{\mu} \tag{18}$$

- Changes in β_k and $\bar{x}_k$ have a **direct** effect on changes in C.

Change in the CI: Total Differential Approach

- Get the **total differential** of Eq. (16):

$$\mathrm{d}C = -\frac{C}{\mu}\mathrm{d}\alpha + \sum_k \frac{\bar{x}_k}{\mu}\left(C_k - C\right)\mathrm{d}\beta_k + \sum_k \frac{\beta_k}{\mu}\left(C_k - C\right)\mathrm{d}\bar{x}_k + \sum_k \frac{\beta_k \bar{x}_k}{\mu}\mathrm{d}C_k + \mathrm{d}\frac{GC_\epsilon}{\mu} \tag{18}$$

- Changes in β_k and $\bar{x}_k$ have a **direct** effect on changes in C.
- They also have an **indirect** effect through μ: an increase in inequality in $\bar{x}_k$ increases the degree of inequality in h.

Change in the CI: Total Differential Approach

- Get the **total differential** of Eq. (16):

$$\mathrm{d}C = -\frac{C}{\mu}\mathrm{d}\alpha + \sum_k \frac{\bar{x}_k}{\mu}\left(C_k - C\right)\mathrm{d}\beta_k + \sum_k \frac{\beta_k}{\mu}\left(C_k - C\right)\mathrm{d}\bar{x}_k + \sum_k \frac{\beta_k \bar{x}_k}{\mu}\mathrm{d}C_k + \mathrm{d}\frac{GC_\epsilon}{\mu} \tag{18}$$

- Changes in β_k and $\bar{x}_k$ have a **direct** effect on changes in C.
- They also have an **indirect** effect through μ: an increase in inequality in $\bar{x}_k$ increases the degree of inequality in h.
- C increases for increases in β_k and $\bar{x}_k$; C decreases for increases in μ.

Example: Change in HAZ-Scores

Table 4. Decomposition of changes in the CI for HAZ-scores: Comparison between total differential and Oaxaca-style approach.

	Total differential approach (Eq. ??)					Oaxaca-style approach (Eq. 17)	
Variable	β	$\bar{x}$	CI	Total	Percent	Total	Percent
Child's age (in months)	0.003	0.011	-0.002	0.012	-57	0.007	-30
Child's age squared	0.003	-0.010	0.001	-0.006	29	-0.006	26
Male	0.001	0.000	0.000	0.001	-5	0.001	-3
Household consumption	-0.005	-0.005	-0.002	-0.011	52	-0.016	74
Safe drinking water	-0.002	0.000	0.000	-0.003	14	-0.003	16
Satisfactory sanitation	0.003	-0.002	0.000	0.001	-5	0.001	-5
Years schooling household head	0.001	0.000	-0.001	0.000	0	0.000	1
Years schooling mother	0.005	0.000	-0.001	0.004	-19	0.003	-11
Commune (fixed effects)	0.000	-0.014	-0.010	-0.025	119	-0.012	55
Residual				0.005	-24	0.005	-24
Total	0.010	-0.021	-0.016	-0.021	100	-0.022	100