xtoaxaca - Extending the Kitagawa-Oaxaca-Blinder Decomposition Approach to longitudinal data analyses

Hannes Kröger, Jörg Hartmann

German Institute for Economic Research (DIW), Berlin
University of Göttingen

24.05.2019

Who would benefit from using xtoaxaca

- You have at least two time points
- You want a flexible way to decompose the level over time
- You want a (counterfactual) decomposition of change over time

Prior approaches

- Based on different research questions very many different decompositions can be chosen
- At the moment we cannot make a systematic comparison
- We present our preferred solution

Traditional approach Kitagawa-Oaxaca-Blinder decomposition

- $R = E(Y_A) - E(Y_B)$

Traditional approach Kitagawa-Oaxaca-Blinder decomposition

- $R = E(Y_A) - E(Y_B)$
- $Y_l = X_l \beta_l + \epsilon_l, E(\epsilon_l) = 0, l \in \{A, B\}$

Traditional approach Kitagawa-Oaxaca-Blinder decomposition

- $R = E(Y_A) - E(Y_B)$
- $Y_l = X_l \beta_l + \epsilon_l, E(\epsilon_l) = 0, l \in \{A, B\}$
- $R = E + C + I$
- $R = [E(X_B) - E(X_A)]\beta_A + E(X_A)(\beta_B - \beta_A) + [E(X_B) - E(X_A)](\beta_B - \beta_A)$

Traditional approach Kitagawa-Oaxaca-Blinder decomposition

- $R = E(Y_A) - E(Y_B)$
- $Y_l = X_l \beta_l + \epsilon_l, E(\epsilon_l) = 0, l \in \{A, B\}$
- $R = E + C + I$
- $R = [E(X_B) - E(X_A)]\beta_A + E(X_A)(\beta_B - \beta_A) + [E(X_B) - E(X_A)](\beta_B - \beta_A)$

Traditional approach Kitagawa-Oaxaca-Blinder decomposition

- $R = E(Y_A) - E(Y_B)$
- $Y_l = X_l\beta_l + \epsilon_l, E(\epsilon_l) = 0, l \in \{A, B\}$
- $R = E + C + I$
- $R = [E(X_B) - E(X_A)]\beta_A + E(X_A)(\beta_B - \beta_A) + [E(X_B) - E(X_A)](\beta_B - \beta_A)$

Traditional approach Kitagawa-Oaxaca-Blinder decomposition

- $R = E(Y_A) - E(Y_B)$
- $Y_l = X_l\beta_l + \epsilon_l, E(\epsilon_l) = 0, l \in \{A, B\}$
- $R = E + C + I$
- $R = [E(X_B) - E(X_A)]\beta_A + E(X_A)(\beta_B - \beta_A) + [E(X_B) - E(X_A)](\beta_B - \beta_A)$

Research questions for the parts of the decomposition

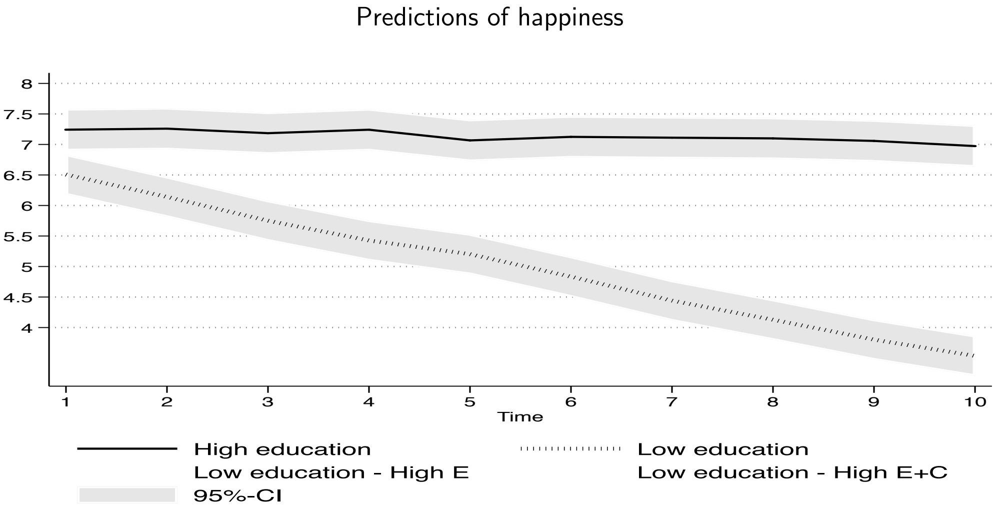
- **E**: How much smaller/bigger would the gap be, if the endowments of group A were the same as for group B?

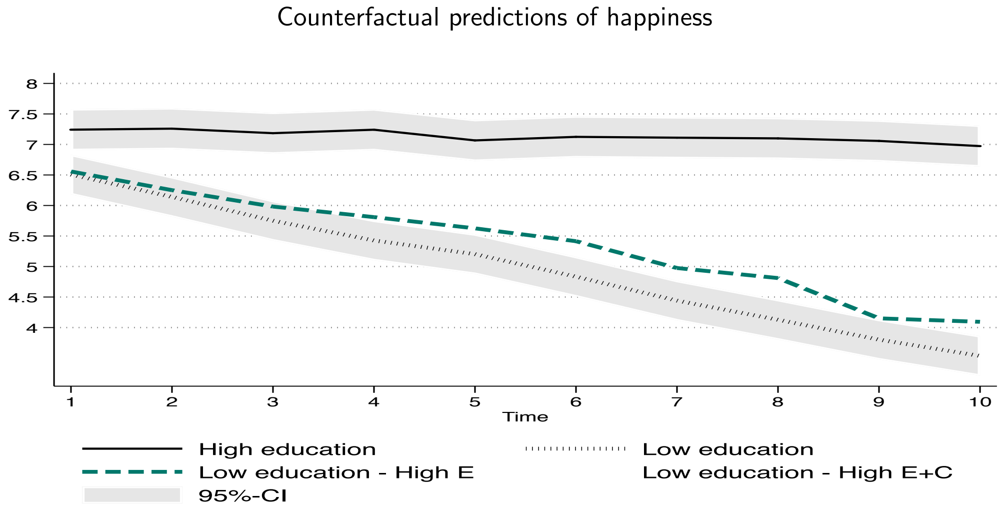
Research questions for the parts of the decomposition

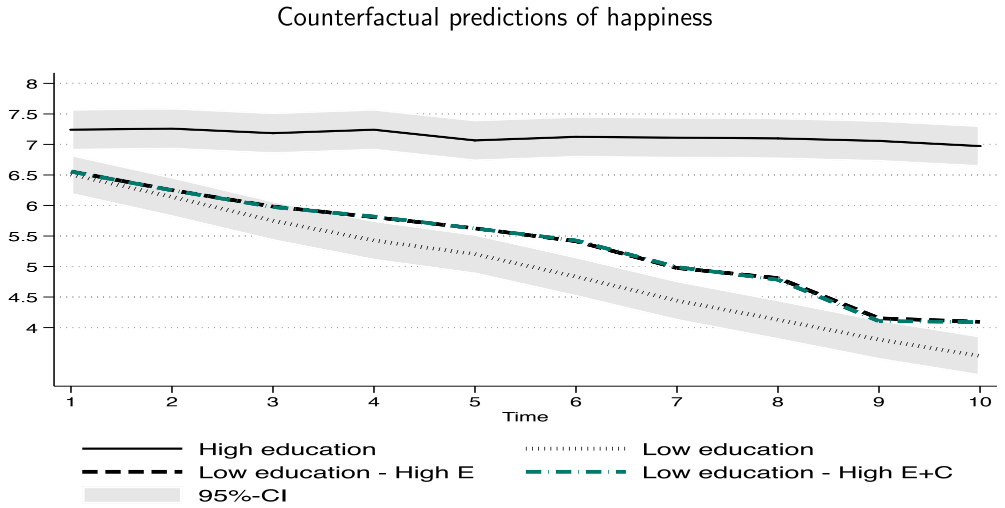
- **E**: How much smaller/bigger would the gap be, if the endowments of group A were the same as for group B?
- **C**: How much smaller/bigger would the gap be, if the effect of the explanatory variables of group A were the same as the effects for group B?

GC prediction approach

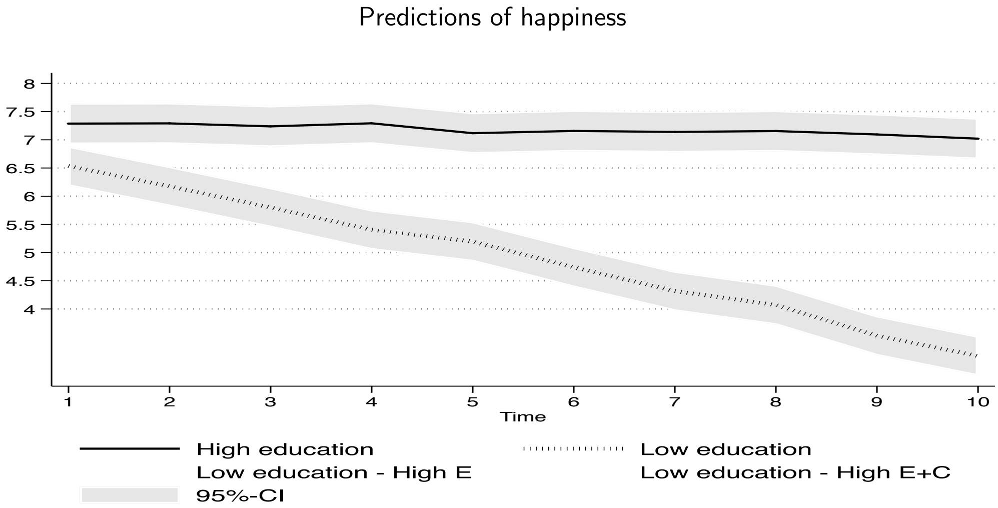
- $R(t) = E(Y_B(t) - Y_A(t))$
- Latent growth curve model for change in happiness
- parametric, semi-parametric, non-parametric

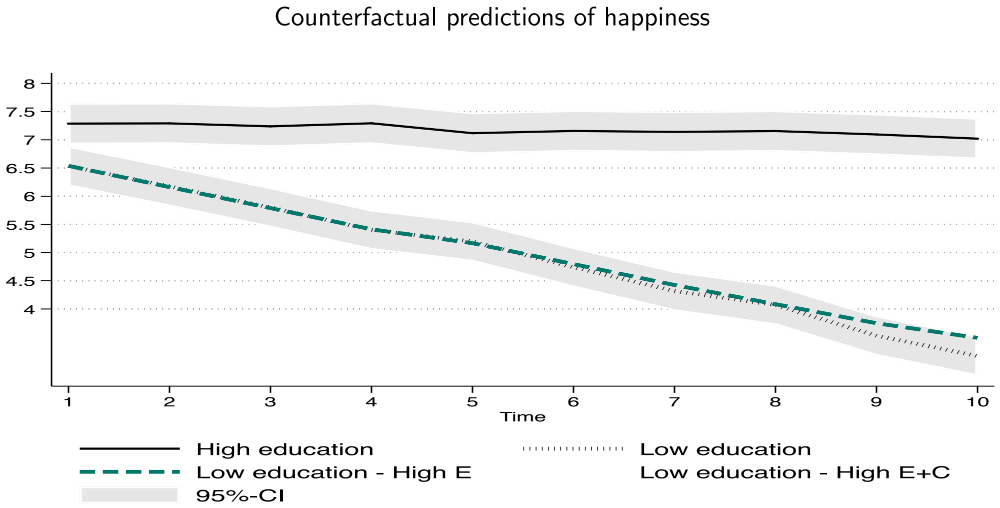

Predictions of happiness
8
7.5
7
6.5
6
5.5
5
4.5
4
1
2
3
4
5
6
7
8
9
10
Time
High education
Low education
Low education - High E
Low education - High E+C
95%-CI


Counterfactual predictions of happiness
8
7.5
7
6.5
6
5.5
5
4.5
4
1
2
3
4
5
6
7
8
9
10
Time
High education
Low education
Low education - High E
Low education - High E+C
95%-CI

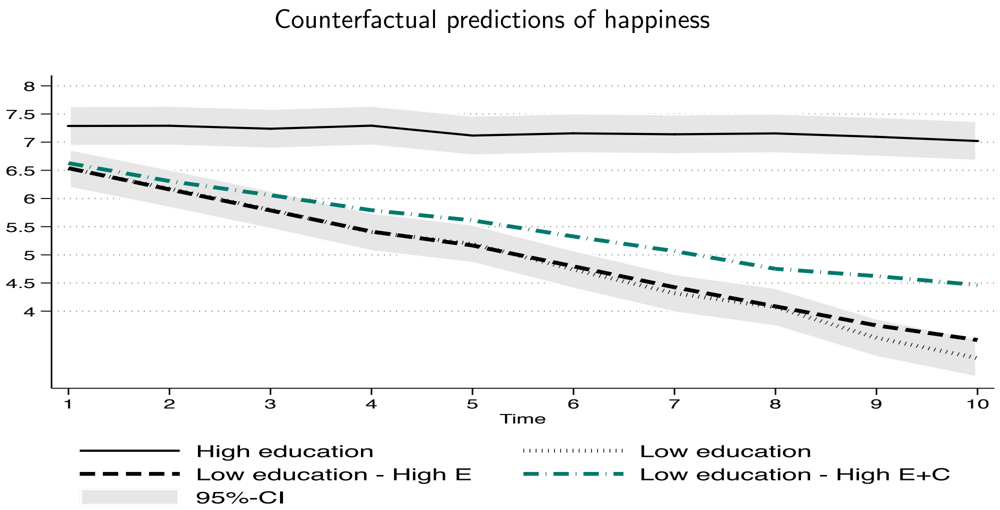


Counterfactual predictions of happiness
8
7.5
7
6.5
6
5.5
5
4.5
4
1
2
3
4
5
6
7
8
9
10
Time
High education
Low education
Low education - High E
Low education - High E+C
95%-CI


Predictions of happiness
8
7.5
7
6.5
6
5.5
5
4.5
4
1
2
3
4
5
6
7
8
9
10
Time
High education
Low education
Low education - High E
Low education - High E+C
95%-CI



Counterfactual predictions of happiness
8
7.5
7
6.5
6
5.5
5
4.5
4
1
2
3
4
5
6
7
8
9
10
Time
High education
Low education
Low education - High E
Low education - High E+C
95%-CI



Counterfactual predictions of happiness
8
7.5
7
6.5
6
5.5
5
4.5
4
1
2
3
4
5
6
7
8
9
10
Time
High education
Low education
Low education - High E
Low education - High E+C
95%-CI

GC decomposition approach

- $R(t) = E(t) + C(t) + I(t)$
- Decomposition is time dependent

GC decomposition approach

- $R(t) = E(t) + C(t) + I(t)$
- Decomposition is time dependent
- $R = [E(X_{ht}) - E(X_{lt})]\beta_{lt} + E(X_{lt})(\beta_{ht} - \beta_{lt}) + [E(X_{ht}) - E(X_{lt})](\beta_{ht} - \beta_{lt})$
- Conditional on functional form of time

GC decomposition approach

- $R(t) = E(t) + C(t) + I(t)$
- Decomposition is time dependent
- $R = [E(X_{ht}) - E(X_{lt})]\beta_{lt} + E(X_{lt})(\beta_{ht} - \beta_{lt}) + [E(X_{ht}) - E(X_{lt})](\beta_{ht} - \beta_{lt})$
- Conditional on functional form of time

GC decomposition approach

- $R(t) = E(t) + C(t) + I(t)$
- Decomposition is time dependent
- $R = [E(X_{ht}) - E(X_{lt})]\beta_{lt} + E(X_{lt})(\beta_{ht} - \beta_{lt}) + [E(X_{ht}) - E(X_{lt})](\beta_{ht} - \beta_{lt})$
- Conditional on functional form of time

GC decomposition approach

- $R(t) = E(t) + C(t) + I(t)$
- Decomposition is time dependent
- $R = [E(X_{ht}) - E(X_{lt})]\beta_{lt} + E(X_{lt})(\beta_{ht} - \beta_{lt}) + [E(X_{ht}) - E(X_{lt})](\beta_{ht} - \beta_{lt})$
- Conditional on functional form of time

Decomposition of change

- Decomposing the change in happiness
- it always needs two time points to compare

Decomposition of change

$$\begin{aligned}\Delta Y^I &= Y_t^I - Y_s^I \\ &= \bar{X}_t^I \beta_t^I - \bar{X}_s^I \beta_s^I\end{aligned}$$

Decomposition of change

$$\Delta Y^l = Y^l_t - Y^l_s$$
$$= \bar{X}^l_t \beta^l_t - \bar{X}^l_s \beta^l_s$$

and the change of the group difference over time then is

$$\Delta Y = \Delta Y^A - \Delta Y^B$$
$$= (\bar{X}^A_t \beta^A_t - \bar{X}^A_s \beta^A_s) - (\bar{X}^B_t \beta^B_t - \bar{X}^B_s \beta^B_s)$$
$$= \bar{X}^A_t \beta^A_t - \bar{X}^A_s \beta^A_s - \bar{X}^B_t \beta^B_t + \bar{X}^B_s \beta^B_s$$

Change due to endowment

$$\Delta Y_E = (\bar{X}_t^A - \bar{X}_s^A)\beta_s^A - (\bar{X}_t^B - \bar{X}_s^B)\beta_s^B$$

Change due to endowment

$$\Delta Y_E = (\bar{X}_t^A - \bar{X}_s^A)\beta_{s}^A - (\bar{X}_t^B - \bar{X}_s^B)\beta_{s}^B$$

Change due to coefficients

$$\Delta Y_C = (\beta_t^A - \beta_s^A)\bar{X}_s^A - (\beta_t^B - \beta_s^B)\bar{X}_s^B$$

Change due to coefficients

$$\Delta Y_C = (\beta_t^A - \beta_s^A)\bar{X}_{\color{red}{s}}^A - (\beta_t^B - \beta_s^B)\bar{X}_{\color{red}{s}}^B$$

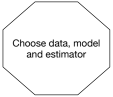
Research questions for the parts of the decomposition

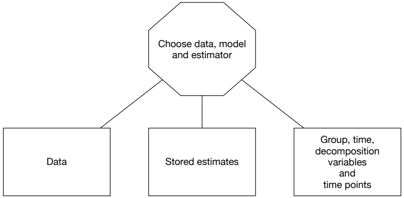
- **dE**: How much smaller/bigger would the **change** in the gap be, if the endowments of group A had **changed** in the same way as for group B (and the difference in coefficients had stayed the same)?

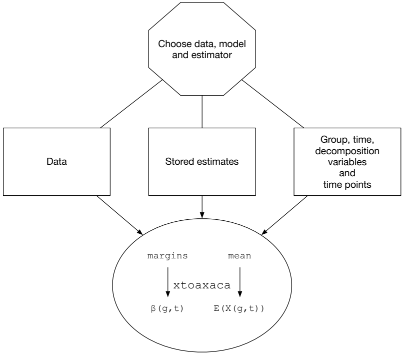
Research questions for the parts of the decomposition

- **dE**: How much smaller/bigger would the **change** in the gap be, if the endowments of group A had **changed** in the same way as for group B (and the difference in coefficients had stayed the same)?

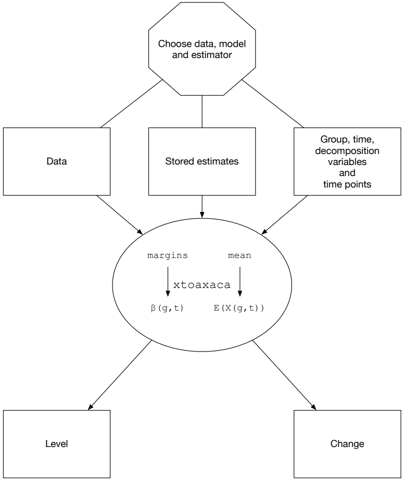
- **dC**: How much smaller/bigger would the **change** in the gap be, if the coefficients of group A had **changed** in the same way as for group B (and the difference in endowments had stayed the same)?


Choose data, model and estimator


Choose data, model and estimator
Data
Stored estimates
Group, time, decomposition variables and time points



Choose data, model and estimator
Data
Stored estimates
Group, time, decomposition variables and time points
margins
mean
xtoaxaca
β(g,t)
E(X(g,t))


Choose data, model and estimator
Data
Stored estimates
Group, time, decomposition variables and time points
margins
mean
xtoaxaca
β(g,t)
E(X(g,t))
Level
Change

Quick note on the use of `margins`

- all decompositions are done at means or other specified values
- calculation of average over the population not necessary
- margins conducted only on one observation
- speed of the calculation of `margins` (not `mean`) independent of sample size

xtoaxaca in action

```
xtoaxaca exp, groupvar(group2) groupcat(1 2) timevar(time)
times(1 3 5 ) /// model1(base) model2(control) ///
timebandwidth(1) basecontrols(edu) change timeref(1)
```

xtoaxaca in action

```
xtoaxaca exp, groupvar(group2) groupcat(1 2) timevar(time)
times(1 3 5 ) /// model1(base) model2(control) ///
timebandwidth(1) basecontrols(edu) change timeref(1)
```

xtoaxaca in action

```
xtoaxaca exp, groupvar(group2) groupcat(1 2) timevar(time)
times(1 3 5 ) /// model1(base) model2(control) ///
timebandwidth(1) basecontrols(edu) change timeref(1)
```

xtoaxaca in action

```
xtoaxaca exp, groupvar(group2) groupcat(1 2) timevar(time)
times(1 3 5 ) /// model1(base) model2(control) ///
timebandwidth(1) basecontrols(edu) change timeref(1)
```

xtoaxaca in action

```
xtoaxaca exp, groupvar(group2) groupcat(1 2) timevar(time)
times(1 3 5 ) /// model1(base) model2(control) ///
timebandwidth(1) basecontrols(edu) change timeref(1)
```

xtoaxaca in action

```
xtoaxaca exp, groupvar(group2) groupcat(1 2) timevar(time)
times(1 3 5 ) /// model1(base) model2(control) ///
timebandwidth(1) basecontrols(edu) change timeref(1)
```

xtoaxaca in action

```
xtoaxaca exp, groupvar(group2) groupcat(1 2) timevar(time)
times(1 3 5 ) /// model1(base) model2(control) ///
timebandwidth(1) basecontrols(edu) change timeref(1)
```

xtoaxaca in action

```
xtoaxaca exp, groupvar(group2) groupcat(1 2) timevar(time)
times(1 3 5 ) /// model1(base) model2(control) ///
timebandwidth(1) basecontrols(edu) change timeref(1)
```

xtoaxaca in action

```
xtoaxaca exp, groupvar(group2) groupcat(1 2) timevar(time)
times(1 3 5 ) /// model1(base) model2(control) ///
timebandwidth(1) basecontrols(edu) change timeref(1)
```

xtoaxaca in action

```
xtoaxaca exp, groupvar(group2) groupcat(1 2) timevar(time)
times(1 3 5 ) /// model1(base) model2(control) ///
timebandwidth(1) basecontrols(edu) change timeref(1)
```

xtoaxaca in action

```
Stata example
```

General problems with decomposition

- Reference group
- Reference group of decomposition variables

Extensions?

- Different forms of change decomposition (e.g. Kim, Makepiece)
- Different counterfactual scenarios
- Reduce to one model estimation
- Extension for SEM


DANKE!