

Decomposition Methods in the Social Sciences

GESIS Training Course

January 29 – February 1, 2024, Cologne

Johannes Giesecke (Humboldt University Berlin)

Ben Jann (University of Bern)

5. Difference-in-difference decompositions

Contents

- 1 Smith-Welch decomposition
- 2 Example analysis
- 3 Juhn-Murphy-Pierce 1991 decomposition
- 4 Example analysis
- 5 Interventionist approach (Kröger/Hartmann)

Difference-in-difference decompositions

- Up to now we were concerned with a single outcome differential (e.g. a gender wage gap) at a specific point in time and in a specific region or population.
- Often, however, comparisons over time or between countries or regions are of interest.
 - ▶ How did the gender wage gap change over time and how much of this change is due to changes with respect to covariates?
 - ▶ How would the gender wage gap in country A look like if it had the wage structure of country B?
- One way of analyzing changes over time or between populations is to compare separate decomposition results. Some questions, however, require a “double” or “difference-in-difference” decomposition.
- A very famous application of such methodology, for example, is the “Swimming upstream” paper by Blau and Kahn (1997).

- 1 Smith-Welch decomposition
- 2 Example analysis
- 3 Juhn-Murphy-Pierce 1991 decomposition
- 4 Example analysis
- 5 Interventionist approach (Kröger/Hartmann)

Smith-Welch decomposition

(Smith and Welch 1987, also see e.g. Heckman et al. 2000, Kröger and Hartmann 2021)

- Given is a linear model

$$Y^{gt} = X^{gt}\beta^{gt} + \epsilon^{gt}, \quad E(\epsilon^{gt}|X^{gt}) = 0$$

for two groups, $g = m, f$ (males and females), at two time points, $t = 0, 1$.

- Using the male coefficients, β^{mt} , as reference, the decomposition of the group difference in average Y at time t can be written as

$$\Delta^{\mu t} = (\bar{X}^{mt} - \bar{X}^{ft})\beta^{mt} + \bar{X}^{ft}(\beta^{mt} - \beta^{ft}) = \Delta_X^{\mu t} + \Delta_S^{\mu t}$$

- We are now interested in decomposing the change in the wage gap over time.

Smith-Welch decomposition

- Let $\Delta\bar{X}^t = \bar{X}^{mt} - \bar{X}^{ft}$ and $\Delta\beta^t = \beta^{mt} - \beta^{ft}$. Using the male coefficients from the first time point, β^{m0} , as reference, this double decomposition can be written as

$$\begin{aligned}d\Delta^\mu &= \Delta^{\mu 1} - \Delta^{\mu 0} = \{(\Delta\bar{X}^1 - \Delta\bar{X}^0)\beta^{m0} + \Delta\bar{X}^1(\beta^{m1} - \beta^{m0})\} \\&\quad + \{\bar{X}^{f1}(\Delta\beta^1 - \Delta\beta^0) + (\bar{X}^{f1} - \bar{X}^{f0})\Delta\beta^0\} \\&= d\Delta_X^\mu + d\Delta_S^\mu\end{aligned}$$

- Interpretation:

$(\Delta\bar{X}^1 - \Delta\bar{X}^0)\beta^{m0}$ main endowments effect: shows how the wage gap changed because men and women became more similar or dissimilar in X (negative, if they became more similar; positive, if they became more dissimilar)

$\Delta\bar{X}^1(\beta^{m1} - \beta^{m0})$ secondary endowments effect due to change in reference wage structure over time

Smith-Welch decomposition

$\bar{X}^{f1}(\Delta\beta^1 - \Delta\beta^0)$ primary coefficients effect: effect of change in wage structure difference between men and women (negative, if coefficients became more similar; positive, if coefficients became more dissimilar)

$(\bar{X}^{f1} - \bar{X}^{f0})\Delta\beta^0$ secondary coefficients effect due to change in reference endowments over time

- Of course, various other types of decompositions are possible depending on the choice of the reference group and the reference year. The index problem of the standard OB decomposition is now a double index problem, which can make it hard to keep an overview.
- See `help smithwelch` for a systematic discussion. It starts with the threefold decomposition and then shows how the formulas change if a reference group and/or a reference year is introduced. Of course, reference groups/years can also be results from pooled or averaged models. See Kröger and Hartmann (2021) for further variants of difference-in-differences decompositions.

- 1 Smith-Welch decomposition
- 2 Example analysis
- 3 Juhn-Murphy-Pierce 1991 decomposition
- 4 Example analysis
- 5 Interventionist approach (Kröger/Hartmann)

Example analysis (using smithwelmch by Jann 2005b)

```
. use gsoep-extract, clear
(Example data based on the German Socio-Economic Panel)

. keep if inlist(wave,1995,2015)
(23,792 observations deleted)

. keep if inrange(age, 25, 55)
(8,147 observations deleted)

. generate lnwage = ln(wage)
(2,734 missing values generated)

. generate expft2 = expft^2
(56 missing values generated)

. generate byte t = wave==2015 // 0 = 1995, 1 = 2015

. generate byte female = sex==2 // 0 = male, 1 = female

. summarize lnwage yeduc expft expft2 t female
```

| Variable | Obs | Mean | Std. dev. | Min | Max |
|----------|--------|----------|-----------|----------|---------|
| lnwage | 8,277 | 2.71872 | .483595 | 1.108563 | 4.86638 |
| yeduc | 10,735 | 12.04332 | 2.700811 | 7 | 18 |
| expft | 10,955 | 12.21353 | 9.640926 | 0 | 40.5 |
| expft2 | 10,955 | 242.1094 | 305.9972 | 0 | 1640.25 |
| t | 11,011 | .6637908 | .4724329 | 0 | 1 |
| female | 11,011 | .5466352 | .497843 | 0 | 1 |

```
. drop if missing(lnwage,yeduc,expft) // remove unused observation
(2,940 observations deleted)
```

```
. svyset psu [pw=weight], strata(strata)
```

Sampling weights: weight

VCE: linearized

Single unit: missing

Strata 1: strata

Sampling unit 1: psu

FPC 1: <zero>

Example analysis

- Outcome by time point and gender

```
. svy: mean lnwage if !missing(yeduc, expft), over(t female)
(running mean on estimation sample)
```

Survey: Mean estimation

Number of strata = 15

Number of PSUs = 2,459

Number of obs = 8,071

Population size = 23,859,587

Design df = 2,444

| | Linearized | | | |
|-------------------|------------|-----------|----------------------|----------|
| | Mean | std. err. | [95% conf. interval] | |
| c.lnwage@t#female | | | | |
| 0 0 | 2.84812 | .0132636 | 2.822111 | 2.874129 |
| 0 1 | 2.627384 | .0186491 | 2.590815 | 2.663954 |
| 1 0 | 2.862735 | .0164148 | 2.830547 | 2.894923 |
| 1 1 | 2.657428 | .0151483 | 2.627723 | 2.687132 |

Example analysis

- Characteristics by time point and gender

```
. svy: mean yeduc expft if !missing(lnwage), over(t female)
(running mean on estimation sample)
```

Survey: Mean estimation

| | | | |
|--------------------|-------|-------------------|------------|
| Number of strata = | 15 | Number of obs = | 8,071 |
| Number of PSUs = | 2,459 | Population size = | 23,859,587 |
| | | Design df = | 2,444 |

| | Linearized | | | |
|------------------|------------|-----------|----------------------|----------|
| | Mean | std. err. | [95% conf. interval] | |
| c.yeduc@t#female | | | | |
| 0 0 | 12.04901 | .0924429 | 11.86774 | 12.23029 |
| 0 1 | 11.79327 | .091392 | 11.61406 | 11.97249 |
| 1 0 | 12.7098 | .0980728 | 12.51749 | 12.90212 |
| 1 1 | 12.94121 | .0904373 | 12.76387 | 13.11855 |
| c.expft@t#female | | | | |
| 0 0 | 17.49546 | .3094411 | 16.88867 | 18.10226 |
| 0 1 | 12.39256 | .3700076 | 11.667 | 13.11812 |
| 1 0 | 17.17692 | .3427645 | 16.50478 | 17.84906 |
| 1 1 | 10.79882 | .2760263 | 10.25755 | 11.34009 |

Example analysis

- OB decomposition in 1995

```
. oaxaca lnwage yeduc (experience: expft expft2) ///  
> if t==0, by(female) weight(1) nodetail svy
```

Blinder-Oaxaca decomposition

Number of strata = 4
Number of PSUs = 794

Number of obs = 2,609
Population size = 11,707,370
Design df = 790
Model = linear
N of obs 1 = 1,486
N of obs 2 = 1,123

Group 1: female = 0

Group 2: female = 1

explained: $(X1 - X2) * b1$

unexplained: $X2 * (b1 - b2)$

| lnwage | Linearized | | t | P> t | [95% conf. interval] | |
|-------------|-------------|-----------|--------|-------|----------------------|----------|
| | Coefficient | std. err. | | | | |
| overall | | | | | | |
| group_1 | 2.84812 | .013042 | 218.38 | 0.000 | 2.822519 | 2.873721 |
| group_2 | 2.627384 | .0178997 | 146.78 | 0.000 | 2.592248 | 2.662521 |
| difference | .2207358 | .0206432 | 10.69 | 0.000 | .1802138 | .2612579 |
| explained | .0758792 | .0115696 | 6.56 | 0.000 | .0531684 | .0985899 |
| unexplained | .1448567 | .0216315 | 6.70 | 0.000 | .1023946 | .1873188 |

Example analysis

- OB decomposition in 2015

```
. oaxaca lnwage yeduc (experience: expft expft2) ///  
> if t==1, by(female) weight(1) nodetail svy
```

Blinder-Oaxaca decomposition

Number of strata = 15

Number of PSUs = 2,037

Number of obs = 5,462

Population size = 12,152,217

Design df = 2,022

Model = linear

Group 1: female = 0

N of obs 1 = 2,642

Group 2: female = 1

N of obs 2 = 2,820

explained: $(X1 - X2) * b1$

unexplained: $X2 * (b1 - b2)$

| lnwage | Linearized | | t | P> t | [95% conf. interval] | |
|-------------|-------------|-----------|--------|-------|----------------------|----------|
| | Coefficient | std. err. | | | | |
| overall | | | | | | |
| group_1 | 2.862735 | .0162749 | 175.90 | 0.000 | 2.830818 | 2.894652 |
| group_2 | 2.657428 | .0149503 | 177.75 | 0.000 | 2.628108 | 2.686747 |
| difference | .2053074 | .0204701 | 10.03 | 0.000 | .1651628 | .245452 |
| explained | .0904872 | .0151554 | 5.97 | 0.000 | .0607654 | .120209 |
| unexplained | .1148202 | .0211416 | 5.43 | 0.000 | .0733585 | .1562819 |

Example analysis

- Estimate outcome models by time point and gender; these model will be used as input to the Smith-Welch decomposition.

```
. svy: regress lnwage yeduc expft expft2 if female==0 & t==0  
  (output omitted)  
. estimates store male_t0  
. svy: regress lnwage yeduc expft expft2 if female==1 & t==0  
  (output omitted)  
. estimates store female_t0  
. svy: regress lnwage yeduc expft expft2 if female==0 & t==1  
  (output omitted)  
. estimates store male_t1  
. svy: regress lnwage yeduc expft expft2 if female==1 & t==1  
  (output omitted)  
. estimates store female_t1
```

Example analysis

- Overview of models.

```
. esttab male_t0 female_t0 male_t1 female_t1, nogap mti nonum r2
```

| | male_t0 | female_t0 | male_t1 | female_t1 |
|--------|-------------------------|----------------------|-------------------------|-------------------------|
| yeduc | 0.0567*** (9.51) | 0.0586*** (9.63) | 0.0829*** (14.88) | 0.0789*** (14.91) |
| expft | 0.0318*** (5.62) | 0.0206** (3.11) | 0.0357*** (6.04) | 0.0313*** (6.36) |
| expft2 | -0.000614*** (-4.09) | -0.000373 (-1.93) | -0.000593*** (-3.79) | -0.000541*** (-3.67) |
| _cons | 1.855*** (20.65) | 1.769*** (20.61) | 1.430*** (14.96) | 1.404*** (18.93) |
| N | 1486 | 1123 | 2642 | 2820 |
| R-sq | 0.145 | 0.123 | 0.280 | 0.277 |

t statistics in parentheses

* p<0.05, ** p<0.01, *** p<0.001

Example analysis

- Smith-Welch decomposition

```
. smithwelch male_t0 female_t0 male_t1 female_t1, reference(1) benchmark(1)
```

Decompositions of individual differentials:

| | D | E | C |
|----------|----------|----------|----------|
| Sample 1 | .2207358 | .0758792 | .1448567 |
| Sample 2 | .2053074 | .0904872 | .1148202 |

Difference in (components of) differentials:

| | dD | dE | dC |
|--|-----------|---------|-----------|
| | -.0154284 | .014608 | -.0300365 |

Decomposition of difference in differentials:

| | D | E | C |
|----|-----------|-----------|-----------|
| dE | .014608 | -.0087573 | .0233653 |
| dC | -.0300365 | -.0105462 | -.0194903 |

D = differential / difference in component of differential

E = part of D due to differences in endowments

C = part of D due to differences in coefficients

● Detailed Smith-Welch decomposition

```
. smithwelch male_t0 female_t0 male_t1 female_t1, reference(1) benchmark(1) ///
> detail(schooling=yeduc, experience=expft*)
```

Decompositions of individual differentials:

| Sample 1 | D | E | C |
|------------|-----------|-----------|-----------|
| schooling | -.0084441 | .0144885 | -.0229326 |
| experience | .1426075 | .0613907 | .0812168 |
| _cons | .0865725 | 0 | .0865725 |
| Total | .2207358 | .0758792 | .1448567 |
| Sample 2 | D | E | C |
| schooling | .0328791 | -.0191954 | .0520745 |
| experience | .1466832 | .1096826 | .0370006 |
| _cons | .0257451 | 0 | .0257451 |
| Total | .2053074 | .0904872 | .1148202 |

Difference in (components of) differentials:

| | dD | dE | dC |
|------------|-----------|-----------|-----------|
| schooling | .0413232 | -.0336838 | .0750071 |
| experience | .0040757 | .0482919 | -.0442162 |
| _cons | -.0608274 | 0 | -.0608274 |

| | | | |
|-------|-----------|---------|-----------|
| Total | -.0154284 | .014608 | -.0300365 |
|-------|-----------|---------|-----------|

Decomposition of difference in differentials:

| dE | D | E | C |
|------------|-----------|-----------|-----------|
| schooling | -.0336838 | -.0275985 | -.0060853 |
| experience | .0482919 | .0188412 | .0294507 |
| _cons | 0 | 0 | 0 |
| Total | .014608 | -.0087573 | .0233653 |
| dC | D | E | C |
| schooling | .0750071 | -.0022322 | .0772393 |
| experience | -.0442162 | -.0083139 | -.0359022 |
| _cons | -.0608274 | 0 | -.0608274 |
| Total | -.0300365 | -.0105462 | -.0194903 |

D = differential / difference in component of differential

E = part of D due to differences in endowments

C = part of D due to differences in coefficients

- 1 Smith-Welch decomposition
- 2 Example analysis
- 3 Juhn-Murphy-Pierce 1991 decomposition
- 4 Example analysis
- 5 Interventionist approach (Kröger/Hartmann)

Juhn-Murphy-Pierce 1991 decomposition

(Juhn et al. 1991)

- Juhn et al. (1991) use an alternative setup by considering changes in the residual variance.
- The argument is that the gender wage gap will be large if the residual variance, that is, the variance of wages once controlling for observables such as education or work experience, is large.
- Conceptually, the residual variance can be viewed as the price of unobservables. The idea is, that there may be differences between men and women in such unobservables. If the prices increase, the gender wage gap will increase as well.
- For example, Blau and Kahn (1997) use the method in an analysis in which they argue that the gender wage gap would have declined more than it actually did, hadn't there been a strong increase in general wage inequality that had nothing to do with gender.

Juhn-Murphy-Pierce 1991 decomposition

- Consider a model

$$Y^{gt} = X^{gt}\beta^t + \epsilon^{gt}$$

for two groups, $g = m, f$ (males and females), at two time points, $t = 0, 1$, where β^t are some reference regression parameters (non-discriminatory prices of observables).

- The model can also be expressed as

$$Y^{gt} = X^{gt}\beta^t + r^{gt}\sigma^t$$

where σ is a reference residual standard deviation (non-discriminatory prices of unobservables) and $r = (Y - X\beta^t)/\sigma^t$ is a standardized residual.

- Thus, the equation now has a two-component error term. The residuals are expressed as a function of the general residual inequality at time t and the positions of the residuals in the residual distribution.

Juhn-Murphy-Pierce 1991 decomposition

- The mean outcome differential between men and women at time t can then be decomposed as follows:

$$\begin{aligned}\Delta^{\mu,t} &= \bar{Y}^{mt} - \bar{Y}^{ft} = (\bar{X}^{mt} - \bar{X}^{ft})\beta^t + (\bar{r}^{mt} - \bar{r}^{ft})\sigma^t \\ &= \Delta_X^\mu + \Delta_S^\mu\end{aligned}$$

- The first term is the “predicted gap” and the second term is the “residual gap”. They are equal to the “explained” part and the “unexplained” part in a standard OLS decomposition using β^t as reference coefficients.
- Let $\Delta\bar{X}^t = (\bar{X}^{mt} - \bar{X}^{ft})$ and $\Delta\bar{r}^t = (\bar{r}^{mt} - \bar{r}^{ft})$. Given two time points $t = 0$ and $t = 1$ the change in the outcome differential can then be written as

$$\begin{aligned}\Delta^{\mu,1} - \Delta^{\mu,0} &= (\Delta\bar{X}^1\beta^1 - \Delta\bar{X}^0\beta^0) + (\Delta\bar{r}^1\sigma^1 - \Delta\bar{r}^0\sigma^0) \\ &= d\Delta_X^\mu + d\Delta_S^\mu\end{aligned}$$

Juhn-Murphy-Pierce 1991 decomposition

- The change in the “predicted gap” can be further decomposed as

$$d\Delta_X^\mu = (\Delta\bar{X}^1 - \Delta\bar{X}^0)\beta^0 + \Delta\bar{X}^0(\beta^1 - \beta^0) + (\Delta\bar{X}^1 - \Delta\bar{X}^0)(\beta^1 - \beta^0)$$

where the first term is the main “observed quantities” effect due to a change in gender differences in X , the second term is a secondary effect due a change in “prices” for observed quantities, and the third term is an interaction term.

- Likewise, the change in the “residual gap” can be decomposed as

$$d\Delta_S^\mu = (\Delta\bar{r}^1 - \Delta\bar{r}^0)\sigma^0 + \Delta\bar{r}^0(\sigma^1 - \sigma^0) + (\Delta\bar{r}^1 - \Delta\bar{r}^0)(\sigma^1 - \sigma^0)$$

where the first term is the so-called “gap effect” due to changes in the group differences in residual positions (i.e. changes in the group differences in “unobserved quantities” and changes in discrimination), the second term is the part due to changes in residual inequality (i.e. changes in “prices” for unobserved quantities), and the third term is again an interaction term.

Juhn-Murphy-Pierce 1991 decomposition

- Similar to other decompositions, it is common practice to use the “prices” of one of the years as the reference prices (or use some average or pooled results). For example, if we use $t = 0$ as the reference, the decomposition simplifies to:

$$d\Delta_X^\mu = (\Delta\bar{X}^1 - \Delta\bar{X}^0)\beta^0 + \Delta\bar{X}^1(\beta^1 - \beta^0)$$

$$d\Delta_S^\mu = (\Delta\bar{r}^1 - \Delta\bar{r}^0)\sigma^0 + \Delta\bar{r}^1(\sigma^1 - \sigma^0)$$

- Furthermore, a detailed decomposition can be obtained for the components of $d\Delta_X^\mu$ in the usual way (but obviously not for $d\Delta_S^\mu$).
- Note that results for $d\Delta_X^\mu$ are the same as for the Smith-Welch decomposition (if using the same setup). For $d\Delta_S^\mu$ only the total is the same; that is, Smith-Welch and Juhn-Murphy-Pierce lead to a different breakup of $d\Delta_S^\mu$.

Juhn-Murphy-Pierce 1991 decomposition: estimation

- Estimation of the components of $d\Delta_X^\mu$ is straightforward.
- Estimation of the components of $d\Delta_S^\mu$ is more involved and requires some discussion. Two approaches are used in the literature, a **parametric** approach and a **nonparametric** approach.
- Parametric
 - ▶ Since, by definition, $\epsilon^t = r^t \sigma^t$, expression $\Delta \bar{r}^t \sigma^t$ can simply be estimated as the mean difference in residuals ϵ between men and women at time t .
 - ▶ But what about expressions such as $\Delta \bar{r}^1 \sigma^0$?
 - ▶ An obvious solution is to estimate the residual standard deviation at time $t = 0$ and then multiply it by the mean difference in standardized residuals of $t = 1$.
- Nonparametric
 - ▶ The parametric approach is simple, but neglects changes in the distributional shape (apart from the variance).
 - ▶ The following nonparametric procedure has therefore been proposed.

Juhn-Murphy-Pierce 1991 decomposition: estimation

- ▶ Let $F^t()$ be the distribution function of the residuals at time t . Furthermore, let p^t represent the relative positions of the residuals in the residual distribution at time t , that is

$$p^{gt} = F^t(\epsilon^{gt}) \quad \text{and thus} \quad \epsilon^{gt} = Q^t(p^{gt})$$

where $Q() = F^{-1}()$ is the quantile function (inverse of $F()$).

- ▶ The solution now is to apply the quantile function of one time point to the residual ranks of the other time point.
- ▶ For example, $\Delta \bar{r}^1 \sigma^0$ is estimated by assigning each individual at $t = 1$ a percentile number corresponding to its position in the residual distribution of $t = 1$ (i.e., compute p^1), then using these relative ranks to derive hypothetical residuals given the $t = 0$ residual distribution (i.e. compute $Q^0(p^1)$), and finally taking the mean difference in these hypothetical residuals between men and women.
- The JMP 1991 procedure relies on some strong assumptions and is not free of critique (e.g. Yun 2009).

- 1 Smith-Welch decomposition
- 2 Example analysis
- 3 Juhn-Murphy-Pierce 1991 decomposition
- 4 Example analysis**
- 5 Interventionist approach (Kröger/Hartmann)

Example analysis (using jmpierce2 by Jann 2005b)

- Estimate the outcome models to be used in the JMP 1991 decomposition.
 - ▶ jmpierce2 does not fully support models estimated with the svy prefix; this is why we just apply the weights rather than svy.

```
. regress lnwage yeduc expft expft2 [pw=weight] if female==0 & t==0  
  (output omitted)  
. estimates store male_t0  
. regress lnwage yeduc expft expft2 [pw=weight] if female==1 & t==0  
  (output omitted)  
. estimates store female_t0  
. regress lnwage yeduc expft expft2 [pw=weight] if female==0 & t==1  
  (output omitted)  
. estimates store male_t1  
. regress lnwage yeduc expft expft2 [pw=weight] if female==1 & t==1  
  (output omitted)  
. estimates store female_t1
```

• Nonparametric JMP 1991 decomposition

```
. jmpierce2 male_t0 female_t0 male_t1 female_t1, reference(1) benchmark(1)
```

Decomposition of individual differentials:

| | raw dif- ferential | quantity effect | residual gap |
|----------|-----------------------|--------------------|-----------------|
| Sample 1 | .2207358 | .0758792 | .1448567 |
| Sample 2 | .2053074 | .0904872 | .1148202 |

Difference in (components of) differentials:

| | D | E | U |
|-------|-----------|---------|-----------|
| Total | -.0154284 | .014608 | -.0300365 |

Decomposition of difference in predicted gap:

| | E | Q | P |
|-------|---------|-----------|----------|
| Total | .014608 | -.0087573 | .0233653 |

Decomposition of difference in residual gap:

| | U | Q | P |
|-------|-----------|-----------|----------|
| Total | -.0300365 | -.0394444 | .0094079 |

D = difference in differential
E = difference in predicted gap
U = difference in residual gap
Q = quantity effect
P = price effect

● Parametric JMP 1991 decomposition

```
. jmpierce2 male_t0 female_t0 male_t1 female_t1, reference(1) benchmark(1) parametric
```

Decomposition of individual differentials:

| | raw dif- ferential | quantity effect | residual gap |
|----------|-----------------------|--------------------|-----------------|
| Sample 1 | .2207358 | .0758792 | .1448567 |
| Sample 2 | .2053074 | .0904872 | .1148202 |

Difference in (components of) differentials:

| | D | E | U |
|-------|-----------|---------|-----------|
| Total | -.0154284 | .014608 | -.0300365 |

Decomposition of difference in predicted gap:

| | E | Q | P |
|-------|---------|-----------|----------|
| Total | .014608 | -.0087573 | .0233653 |

Decomposition of difference in residual gap:

| | U | Q | P |
|-------|-----------|-----------|----------|
| Total | -.0300365 | -.0385607 | .0085242 |

D = difference in differential
 E = difference in predicted gap
 U = difference in residual gap
 Q = quantity effect
 P = price effect

• Detailed JMP 1991 decomposition

```
. jmpierce2 male_t0 female_t0 male_t1 female_t1, reference(1) benchmark(1) ///
> detail(schooling=yeduc, experience=expft*)
```

Decomposition of individual differentials:

| | raw dif- ferential | quantity effect | residual gap |
|----------|-----------------------|--------------------|-----------------|
| Sample 1 | .2207358 | .0758792 | .1448567 |
| Sample 2 | .2053074 | .0904872 | .1148202 |

Difference in (components of) differentials:

| | D | E | U |
|-------|-----------|---------|-----------|
| Total | -.0154284 | .014608 | -.0300365 |

Decomposition of difference in predicted gap:

| | E | Q | P |
|------------|-----------|-----------|-----------|
| Total | .014608 | -.0087573 | .0233653 |
| schooling | -.0336838 | -.0275985 | -.0060853 |
| experience | .0482919 | .0188412 | .0294507 |

Decomposition of difference in residual gap:

| | U | Q | P |
|-------|-----------|-----------|----------|
| Total | -.0300365 | -.0394444 | .0094079 |

D = difference in differential
 E = difference in predicted gap
 U = difference in residual gap
 Q = quantity effect
 P = price effect

- 1 Smith-Welch decomposition
- 2 Example analysis
- 3 Juhn-Murphy-Pierce 1991 decomposition
- 4 Example analysis
- 5 Interventionist approach (Kröger/Hartmann)

Interventionist approach (Kröger/Hartmann)

- Kröger and Hartmann (2021) give an overview of several difference-in-differences decompositions and propagate an easy to interpret interventionist approach.
- The approach is somewhat different from Smith-Welch and Juhn-Murphy-Pierce in that it does not try to provide separate breakups of the change/difference in the explained part and the change/difference in the unexplained part.
- Again, let there be two groups two groups, $g \in \{m, f\}$ (males and females), and two time points, $t \in \{0, 1\}$.
- The decomposition then asks:
 1. How does the group difference change if endowments are adjusted to the level of time 1, but coefficients are kept at their values of time 0 (endowments effect).
 2. How does the group difference change if coefficients are adjusted to the values of time 1, but endowments are kept at their level of time 0 (coefficients effect)

Interventionist approach (Kröger/Hartmann)

- That is, the decomposition is as follows:

$$\Delta E = (\bar{X}^{m1} - \bar{X}^{m0})\beta^{m0} - (\bar{X}^{f1} - \bar{X}^{f0})\beta^{f0}$$

$$\Delta C = \bar{X}^{m0}(\beta^{m1} - \beta^{m0}) - \bar{X}^{f0}(\beta^{f1} - \beta^{f0})$$

$$\Delta I = (\bar{X}^{m1} - \bar{X}^{m0})(\beta^{m1} - \beta^{m0}) - (\bar{X}^{f1} - \bar{X}^{f0})(\beta^{f1} - \beta^{f0})$$

- The sum of ΔE , ΔC , and ΔI is equal to the overall change in the group difference between time 0 and time 1.
- Primary interest lies in ΔE (endowments effect) and ΔC (coefficients effect). As usual, the interaction term ΔI is less straightforward to interpret.

Exercise 5

References

- Blau, F.D., L.M. Kahn (1997). Swimming Upstream: Trends in the Gender Wage Differential in the 1980s. *Journal of Labor Economics* 15(1):1–42.
- Heckman, J.J., T.M. Lyons, Petra E. Todd (2000). Understanding Black-White Wage Differentials. *American Economic Review* 90(2): 344–349.
- Jann, B. (2005a). jmpierce2: Stata module to compute trend decomposition of outcome differentials. Available from <https://ideas.repec.org/c/boc/bocode/s448804.html>.
- Jann, B. (2005b). smithwelch: Stata module to compute trend decomposition of outcome differentials. Available from <http://ideas.repec.org/c/boc/bocode/s448805.html>.
- Juhn, C., K.M. Murphy, Brooks Pierce (1991). Accounting for the Slowdown in Black-White Wage Convergence. In Marvin Kosters (Ed.), *Workers and Their Wages* (pp. 107–143). Washington, DC: AEI Press.
- Kröger, H., J. Hartmann (2021). Extending the Kitagawa-Oaxaca-Blinder decomposition approach to panel data. *The Stata Journal* 21(2):360–410.
- Smith, J.P., F.R. Welch (1989). Black Economic Progress After Myrdal. *Journal of Economic Literature* 27(2):519–564.
- Yun, M. (2009). Wage Differentials, Discrimination and Inequality: A Cautionary Note on the Juhn, Murphy and Pierce Decomposition Method. *Scottish Journal of Political Economy* 56(1):123–137.