Assignment3 msb104

Frida Alendal og Katinca Valvatne

Table of contents


```
# Laster nødvendige pakker for del A
library(dplyr)
library(ggplot2)
```

```
# Last inn data_B3 fra Assignment 2
data_B3 <- readRDS("data/data_B3.rds")
```

```
list.files("data")
```

```
[1] "base_year.rds"       "coverage.rds"        "cs_2017.rds"
[4] "data_B3.rds"         "dev_2000_2017.rds"   "gdpc_clean.rds"
[7] "gini_nuts2_year.rds" "y0.rds"              "y17.rds"
```

```r
# Laster inn datasett som ble lagret i Assignment 1/2
gini_nuts2_year <- readRDS("data/gini_nuts2_year.rds")
dev_2000_2017   <- readRDS("data/dev_2000_2017.rds")
```

1 Del A: Testing av utviklingseffekter på tvers av undergrupper

```r
# Forutsetter at data_B3 er laget som i Assignment 2
# Vi lager tre grupper basert på tertiler av høyere utdanning

data_A <- data_B3 %>%
  mutate(
    edu_group3 = cut(
      edu_tertiary,
      breaks = quantile(edu_tertiary, probs = c(0, 1/3, 2/3, 1)),
      include.lowest = TRUE,
      labels = c("Low edu", "Mid edu", "High edu")
    )
  )

# Sjekk at gruppene ble rimelige
data_A %>%
  count(edu_group3)
```

```
# A tibble: 3 x 2
  edu_group3     n
  <fct>      <int>
1 Low edu        7
2 Mid edu        7
3 High edu       7
```

```r
# Regresjon gini ~ utvikling (dev_log) i hver utdanningsgruppe

mod_low  <- lm(gini_w ~ dev_log, data = filter(data_A, edu_group3 == "Low edu"))
mod_mid  <- lm(gini_w ~ dev_log, data = filter(data_A, edu_group3 == "Mid edu"))
mod_high <- lm(gini_w ~ dev_log, data = filter(data_A, edu_group3 == "High edu"))

summary(mod_low)
```

```
Call:
lm(formula = gini_w ~ dev_log, data = filter(data_A, edu_group3 ==
    "Low edu"))
```

```
Residuals:
        1         2         3         4         5         6         7 
-0.024778  0.053360  0.053795 -0.049411 -0.005867 -0.032891  0.005792 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.10722    0.06767   1.584    0.174
dev_log     -0.08148    0.12607  -0.646    0.547

Residual standard error: 0.0446 on 5 degrees of freedom
Multiple R-squared:  0.0771,	Adjusted R-squared:  -0.1075 
F-statistic: 0.4177 on 1 and 5 DF,  p-value: 0.5466
```

```
summary(mod_mid)
```

```
Call:
lm(formula = gini_w ~ dev_log, data = filter(data_A, edu_group3 == 
    "Mid edu"))

Residuals:
        1         2         3         4         5         6         7 
 0.054665  0.018725 -0.023187 -0.005104 -0.005780 -0.043830  0.004511 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.006553   0.043882  -0.149    0.887
dev_log      0.188705   0.112207   1.682    0.153

Residual standard error: 0.03429 on 5 degrees of freedom
Multiple R-squared:  0.3613,	Adjusted R-squared:  0.2336 
F-statistic: 2.828 on 1 and 5 DF,  p-value: 0.1534
```

```
summary(mod_high)
```

```
Call:
lm(formula = gini_w ~ dev_log, data = filter(data_A, edu_group3 == 
    "High edu"))

Residuals:
        1         2         3         4         5         6         7 
 0.005983 -0.040335 -0.008859 -0.029213  0.038335 -0.027167  0.061256 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) -0.01246    0.03280  -0.380   0.7196  
dev_log      0.12366    0.04672   2.646   0.0456 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04136 on 5 degrees of freedom
Multiple R-squared:  0.5835,	Adjusted R-squared:  0.5002 
F-statistic: 7.004 on 1 and 5 DF,  p-value: 0.04562
```

```r
# Modell uten interaksjon: samme dev_log-effekt i alle grupper
mod_no_int <- lm(gini_w ~ dev_log + edu_group3, data = data_A)

# Modell MED interaksjon: dev_log-effekt kan være ulik i gruppene
mod_int <- lm(gini_w ~ dev_log * edu_group3, data = data_A)

summary(mod_int)
```

```
Call:
lm(formula = gini_w ~ dev_log * edu_group3, data = data_A)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.04941 -0.02717 -0.00578  0.01872  0.06126 

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                 0.10722    0.06117   1.753    0.100
dev_log                    -0.08148    0.11395  -0.715    0.486
edu_group3Mid edu          -0.11378    0.08002  -1.422    0.176
edu_group3High edu         -0.11969    0.06902  -1.734    0.103
dev_log:edu_group3Mid edu   0.27018    0.17433   1.550    0.142
dev_log:edu_group3High edu  0.20514    0.12271   1.672    0.115

Residual standard error: 0.04031 on 15 degrees of freedom
Multiple R-squared:  0.3984,	Adjusted R-squared:  0.1978 
F-statistic: 1.987 on 5 and 15 DF,  p-value: 0.1392
```

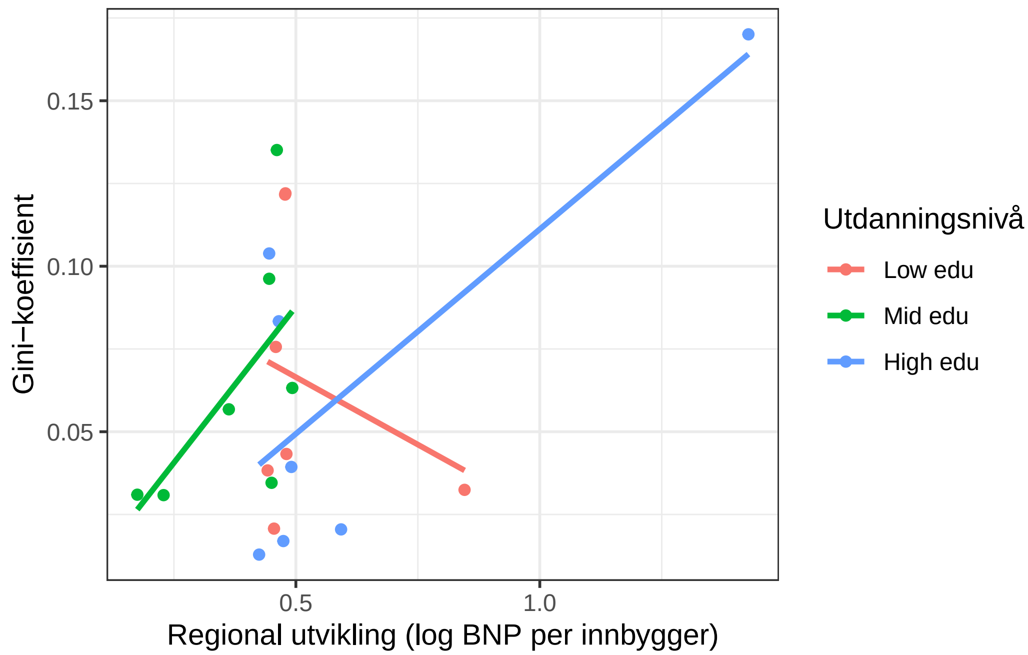
```r
# Likhetstest for helninger:
# H0: samme dev_log-effekt i alle grupper
anova(mod_no_int, mod_int)
```

```
Analysis of Variance Table

Model 1: gini_w ~ dev_log + edu_group3
Model 2: gini_w ~ dev_log * edu_group3
```

```
  Res.Df      RSS Df Sum of Sq     F Pr(>F)
1     17 0.029657
2     15 0.024378  2 0.0052788 1.624 0.2299
```

```r
ggplot(data_A, aes(x = dev_log, y = gini_w, colour = edu_group3)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    x = "Regional utvikling (log BNP per innbygger)",
    y = "Gini-koeffisient",
    colour = "Utdanningsnivå"
  ) +
  theme_bw()
```



1.1 Drøfting av delutvalgsanalyse

2 Del B: Utforskning av alternative funksjonelle former

2.1 Funksjonell form-utforskning

TEKST AV KATINCA: Kort tekst om hvorfor vi prøver kvadratisk/kubisk i stedet for bare lineær.

```r
# Lager et rent datasett for del B med bare variablene vi trenger
data_B <- data_B3 |>
  dplyr::select(NUTS2, gini_w, dev_log)
```

```r
# Lager kvadratisk og kubisk form av dev_log
data_B <- data_B |>
  dplyr::mutate(
    dev_log2 = dev_log^2,       # dev_log i andre
    dev_log3 = dev_log^3        # dev_log i tredje
  )

# Sjekk at variablene ser fornuftige ut
dplyr::glimpse(data_B)
```

```
Rows: 21
Columns: 5
$ NUTS2    <chr> "AT11", "AT12", "AT21", "AT22", "AT31", "AT32", "AT33", "CZ05~
$ gini_w   <dbl> 0.04327867, 0.13510424, 0.09620401, 0.12164526, 0.12201980, 0~
$ dev_log  <dbl> 0.4806890, 0.4609958, 0.4453069, 0.4778743, 0.4786247, 0.4926~
$ dev_log2 <dbl> 0.23106192, 0.21251712, 0.19829823, 0.22836380, 0.22908160, 0~
$ dev_log3 <dbl> 0.111068923, 0.097969498, 0.088303566, 0.109129182, 0.1096441~
```

```r
# Referansemodell: lineær sammenheng mellom utvikling og ulikhet
mod_lin <- lm(gini_w ~ dev_log, data = data_B)

# Alternativ 1: kvadratisk funksjon (tillater bøyd kurve)
mod_quad <- lm(gini_w ~ dev_log + dev_log2, data = data_B)

# Alternativ 2: kubisk funksjon (enda mer fleksibel kurve)
mod_cubic <- lm(gini_w ~ dev_log + dev_log2 + dev_log3, data = data_B)

# Viser resultatene (Katinca bruker dette til å beskrive koeffisienter og signifikans)
summary(mod_lin)
```

```
Call:
lm(formula = gini_w ~ dev_log, data = data_B)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.061231 -0.025062 -0.004973  0.026312  0.074543 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)  0.02089    0.02064   1.012   0.3243  
dev_log      0.08606    0.03700   2.326   0.0313 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04074 on 19 degrees of freedom
```

```
Multiple R-squared:  0.2216,    Adjusted R-squared:  0.1806 
F-statistic: 5.409 on 1 and 19 DF,  p-value: 0.03125
```

```r
summary(mod_quad)
```

```
Call:
lm(formula = gini_w ~ dev_log + dev_log2, data = data_B)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.05059 -0.02364 -0.01610  0.02461  0.07649 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.05567    0.05160   1.079    0.295
dev_log     -0.02473    0.15489  -0.160    0.875
dev_log2     0.06749    0.09156   0.737    0.471

Residual standard error: 0.04124 on 18 degrees of freedom
Multiple R-squared:  0.2444,    Adjusted R-squared:  0.1605 
F-statistic: 2.911 on 2 and 18 DF,  p-value: 0.08027
```

```r
summary(mod_cubic)
```

```
Call:
lm(formula = gini_w ~ dev_log + dev_log2 + dev_log3, data = data_B)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.052177 -0.026784 -0.000492  0.018642  0.070277 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) -0.07024    0.08412  -0.835   0.4153  
dev_log      0.73033    0.43706   1.671   0.1130  
dev_log2    -1.21343    0.70435  -1.723   0.1031  
dev_log3     0.57427    0.31341   1.832   0.0845 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03878 on 17 degrees of freedom
Multiple R-squared:  0.369, Adjusted R-squared:  0.2577 
F-statistic: 3.314 on 3 and 17 DF,  p-value: 0.0451
```

```
# Samler nøkkeltall fra modellene i en liten tabell
model_comp <- tibble::tibble(
  modell   = c("Lineær", "Kvadratisk", "Kubisk"),
  R2       = c(summary(mod_lin)$r.squared,
               summary(mod_quad)$r.squared,
               summary(mod_cubic)$r.squared),
  adj_R2   = c(summary(mod_lin)$adj.r.squared,
               summary(mod_quad)$adj.r.squared,
               summary(mod_cubic)$adj.r.squared),
  AIC      = c(AIC(mod_lin), AIC(mod_quad), AIC(mod_cubic))
)

model_comp
```

```
# A tibble: 3 x 4
  modell        R2 adj_R2   AIC
  <chr>      <dbl>  <dbl> <dbl>
1 Lineær     0.222  0.181 -70.9
2 Kvadratisk 0.244  0.160 -69.5
3 Kubisk     0.369  0.258 -71.3
```

2.2 Estemering og visualisering

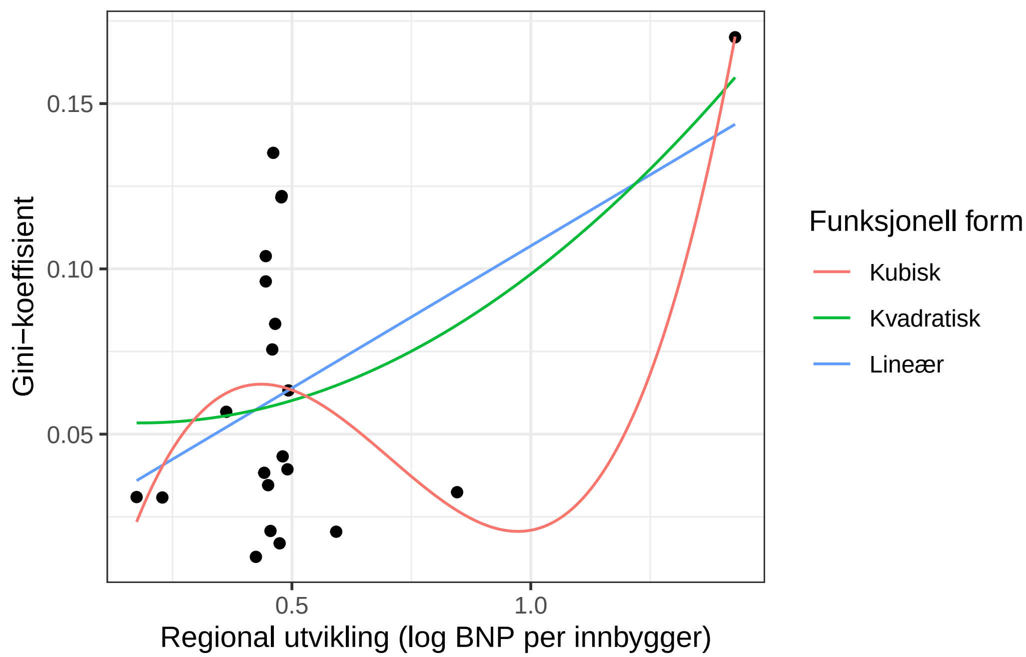
TEKST AV KATINCA: Skriv her om figur + hvordan kurvene ser ut, og hva de betyr

```
# Lager et "grid" av dev_log-verdier for å plotte glatte prediksjonskurver
pred_grid <- data.frame(
  dev_log = seq(
    from = min(data_B$dev_log, na.rm = TRUE),
    to   = max(data_B$dev_log, na.rm = TRUE),
    length.out = 200
  )
)

# Lag kvadratisk og kubisk variant også i pred_grid
pred_grid <- pred_grid |>
  dplyr::mutate(
    dev_log2 = dev_log^2,
    dev_log3 = dev_log^3
  )

# Predikert Gini for hver modell på samme grid
pred_grid$gini_lin   <- predict(mod_lin,   newdata = pred_grid)
pred_grid$gini_quad  <- predict(mod_quad,  newdata = pred_grid)
pred_grid$gini_cubic <- predict(mod_cubic, newdata = pred_grid)
```

```
# Plotter datapunkter + tre kurver for de ulike funksjonelle formene
ggplot(data_B, aes(x = dev_log, y = gini_w)) +
  geom_point() +
  geom_line(data = pred_grid,
            aes(y = gini_lin, colour = "Lineær")) +
  geom_line(data = pred_grid,
            aes(y = gini_quad, colour = "Kvadratisk")) +
  geom_line(data = pred_grid,
            aes(y = gini_cubic, colour = "Kubisk")) +
  labs(
    x      = "Regional utvikling (log BNP per innbygger)",
    y      = "Gini-koeffisient",
    colour = "Funksjonell form"
  ) +
  theme_bw()
```



2.3 Tolkning av resultater

TEKST AV KATINCA: Her skrives 200–400 ord om koeffisienter, implikasjoner, hvilket funksjonsvalg som passer best, og sammenligning med den opprinnelige lineære modellen. Hvilken funksjonell form som passer best (bruk tabellen model_comp + figuren), sammenligning med den lineære modellen.

3 Del C: Test for heteroskedastisitet og drøfting av kausalitet

3.1 Heteroskedastisitetstesting

TEKST AV KATINCA: Kort forklaring på hva heteroskedastisitet er og hvorfor vi tester for det.

```
# Laster inn pakker for heteroskedastisitetstest
library(lmtest)
library(sandwich)

# Breusch-Pagan-test for lineær, kvadratisk og kubisk modell
bp_lin   <- bptest(mod_lin)
bp_quad  <- bptest(mod_quad)
bp_cubic <- bptest(mod_cubic)

bp_lin
```

```
	studentized Breusch-Pagan test

data:  mod_lin
BP = 0.44556, df = 1, p-value = 0.5045
```

```
bp_quad
```

```
	studentized Breusch-Pagan test

data:  mod_quad
BP = 2.5602, df = 2, p-value = 0.278
```

```
bp_cubic
```

```
	studentized Breusch-Pagan test

data:  mod_cubic
BP = 3.9673, df = 3, p-value = 0.265
```

```
# Samler Breusch-Pagan-resultater i en liten tabell
het_tests <- tibble::tibble(
  modell  = c("Lineær", "Kvadratisk", "Kubisk"),
  BP_stat = c(bp_lin$statistic,  bp_quad$statistic,  bp_cubic$statistic),
  df      = c(bp_lin$parameter,  bp_quad$parameter,  bp_cubic$parameter),
  p_value = c(bp_lin$p.value,    bp_quad$p.value,    bp_cubic$p.value)
```

```
)

het_tests  # Tabell Katinca kan bruke i teksten
```

```
# A tibble: 3 x 4
  modell     BP_stat    df p_value
  <chr>        <dbl> <dbl>   <dbl>
1 Lineær       0.446     1   0.504
2 Kvadratisk   2.56      2   0.278
3 Kubisk       3.97      3   0.265
```

3.2 Kausalitetsdrøfting

TEKST AV KATINCA: 200–400 ord om kausal tolkning, mulige problemer (omvendt kausalitet, utelatte variabler, målefeil), og hva man kunne gjort i videre forskning.

4 Del D: Panelestimat

4.1 Panelestimeringsoppgave

TEKST AV KATINCA: Kort intro om at vi bruker hele panelet (alle år og regioner) for å estimere effekten av regional utvikling (dev_log) på ulikhet (gini_w) med faste effekter for land, år og NUTS2-regioner.

```
#Laster inn pakke for paneldata
library(plm)
```

```
# Lager paneldatasett med variablene vi trenger
panel_data <- gini_nuts2_year %>%
  # slår inn utviklingsvariabelen dev_log fra dev_2000_2017
  dplyr::left_join(
    dev_2000_2017 %>%
      dplyr::select(NUTS2, dev_log),   # bare NUTS2 og dev_log her
    by = "NUTS2"                       # join KUN på NUTS2
  ) %>%
  # tar bare med de variablene vi trenger videre
  dplyr::select(country, NUTS2, year, gini_w, dev_log)

dplyr::glimpse(panel_data)
```

```
Rows: 360
Columns: 5
$ country <chr> "AT", "AT", "AT", "AT", "AT", "AT", "AT", "AT", "AT", "AT", "A~
$ NUTS2   <chr> "AT11", "AT11", "AT11", "AT11", "AT11", "AT11", "AT11", "AT11"~
```

```
$ year    <int> 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 20~
$ gini_w  <dbl> 0.02463680, 0.03210136, 0.03346170, 0.05061639, 0.04616263, 0.~
$ dev_log <dbl> 0.4806890, 0.4806890, 0.4806890, 0.4806890, 0.4806890, 0.48068~
```

```r
# Definerer panel-struktur

# Panel 1: NUTS2 × år (region-år-panel)
panel_p_nuts2 <- plm::pdata.frame(
  panel_data,
  index = c("NUTS2", "year")
)

# Panel 2: land × år (land-år-panel)
panel_p_country <- plm::pdata.frame(
  panel_data,
  index = c("country", "year")
)

# Sjekk at alt ser greit ut
dplyr::glimpse(panel_p_nuts2)
```

```
Rows: 360
Columns: 5
$ country <pseries> "AT", "AT", "AT", "AT", "AT", "AT", "AT", "AT", "AT", "AT"~
$ NUTS2   <fct> AT11, AT11, AT11, AT11, AT11, AT11, AT11, AT11, AT11, AT11, AT~
$ year    <fct> 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 20~
$ gini_w  <pseries> 0.02463680, 0.03210136, 0.03346170, 0.05061639, 0.04616263~
$ dev_log <pseries> 0.4806890, 0.4806890, 0.4806890, 0.4806890, 0.4806890, 0.4~
```

```r
# Estimerer panelmodeller med faste effekter via dummier (lm)

# Modell 1: Land-faste effekter (country-FE)
mod_fe_country <- lm(
  gini_w ~ dev_log + factor(country),
  data = panel_data
)

# Modell 2: Års-faste effekter (year-FE)
mod_fe_year <- lm(
  gini_w ~ dev_log + factor(year),
  data = panel_data
)

# Modell 3: NUTS2-faste effekter (region-FE)
mod_fe_nuts2 <- lm(
  gini_w ~ dev_log + factor(NUTS2),
  data = panel_data
```

```r
)

# Modell 4: NUTS2- og år-faste effekter samtidig
mod_fe_nuts2_year <- lm(
  gini_w ~ dev_log + factor(NUTS2) + factor(year),
  data = panel_data
)

# Viser resultatene (Katinca bruker dette i teksten)
summary(mod_fe_country)
```

```

Call:
lm(formula = gini_w ~ dev_log + factor(country), data = panel_data)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.077916 -0.019129 -0.001334  0.016662  0.075219 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)        0.103756   0.009144  11.347  < 2e-16 ***
dev_log           -0.011242   0.018625  -0.604    0.546    
factor(country)CZ -0.074855   0.010211  -7.331 1.57e-12 ***
factor(country)EE  0.090028   0.019253   4.676 4.16e-06 ***
factor(country)ES -0.066504   0.003555 -18.708  < 2e-16 ***
factor(country)SI -0.006238   0.005871  -1.063    0.289    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02869 on 354 degrees of freedom
Multiple R-squared:  0.6749,	Adjusted R-squared:  0.6703 
F-statistic:   147 on 5 and 354 DF,  p-value: < 2.2e-16
```

```r
summary(mod_fe_year)
```

```

Call:
lm(formula = gini_w ~ dev_log + factor(year), data = panel_data)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.094620 -0.032092 -0.005021  0.029292  0.109388 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
```

```
(Intercept)      -0.001664   0.016326  -0.102    0.919    
dev_log           0.090199   0.009962   9.054   <2e-16 ***
factor(year)2001  0.022699   0.018931   1.199    0.231    
factor(year)2002  0.022678   0.018591   1.220    0.223    
factor(year)2003  0.024808   0.018307   1.355    0.176    
factor(year)2004  0.026974   0.018307   1.473    0.142    
factor(year)2005  0.026054   0.018307   1.423    0.156    
factor(year)2006  0.024830   0.018307   1.356    0.176    
factor(year)2007  0.023945   0.018307   1.308    0.192    
factor(year)2008  0.022391   0.018307   1.223    0.222    
factor(year)2009  0.021219   0.018307   1.159    0.247    
factor(year)2010  0.021756   0.018307   1.188    0.235    
factor(year)2011  0.021901   0.018307   1.196    0.232    
factor(year)2012  0.022189   0.018307   1.212    0.226    
factor(year)2013  0.020337   0.018307   1.111    0.267    
factor(year)2014  0.019330   0.018307   1.056    0.292    
factor(year)2015  0.020470   0.018307   1.118    0.264    
factor(year)2016  0.019426   0.018307   1.061    0.289    
factor(year)2017  0.020468   0.018307   1.118    0.264    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04592 on 341 degrees of freedom
Multiple R-squared:  0.1978,    Adjusted R-squared:  0.1554 
F-statistic:  4.67 on 18 and 341 DF,  p-value: 3.014e-09
```

```r
summary(mod_fe_nuts2)
```

```
Call:
lm(formula = gini_w ~ dev_log + factor(NUTS2), data = panel_data)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.029720 -0.004402  0.000137  0.005004  0.021855 

Coefficients: (1 not defined because of singularities)
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)        1.000137   0.039494  25.324  < 2e-16 ***
dev_log           -1.986759   0.085036 -23.364  < 2e-16 ***
factor(NUTS2)AT12  0.046630   0.002563  18.191  < 2e-16 ***
factor(NUTS2)AT21 -0.019985   0.003009  -6.642 1.24e-10 ***
factor(NUTS2)AT22  0.089775   0.002806  31.995  < 2e-16 ***
factor(NUTS2)AT31  0.102514   0.002833  36.184  < 2e-16 ***
factor(NUTS2)AT32  0.054881   0.003521  15.587  < 2e-16 ***
factor(NUTS2)AT33 -0.046387   0.002660 -17.440  < 2e-16 ***
factor(NUTS2)CZ05  0.699607   0.032556  21.489  < 2e-16 ***
```

```
factor(NUTS2)EE00  2.014044   0.081977  24.568  < 2e-16 ***
factor(NUTS2)ES11  0.202236   0.011216  18.031  < 2e-16 ***
factor(NUTS2)ES21  0.011255   0.003376   3.334  0.00095 ***
factor(NUTS2)ES24 -0.049270   0.002649 -18.596  < 2e-16 ***
factor(NUTS2)ES41 -0.002794   0.002504  -1.116  0.26530
factor(NUTS2)ES42 -0.093765   0.003126 -29.998  < 2e-16 ***
factor(NUTS2)ES51 -0.149600   0.004185 -35.744  < 2e-16 ***
factor(NUTS2)ES52 -0.234658   0.008987 -26.111  < 2e-16 ***
factor(NUTS2)ES53 -0.513469   0.020177 -25.449  < 2e-16 ***
factor(NUTS2)ES61 -0.068702   0.002769 -24.814  < 2e-16 ***
factor(NUTS2)ES70 -0.619073   0.024725 -25.039  < 2e-16 ***
factor(NUTS2)SI03 -0.019015   0.002695  -7.055 9.75e-12 ***
factor(NUTS2)SI04        NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.008496 on 339 degrees of freedom
Multiple R-squared:  0.9727,    Adjusted R-squared:  0.9711
F-statistic:   604 on 20 and 339 DF,  p-value: < 2.2e-16
```

```r
summary(mod_fe_nuts2_year)
```

```
Call:
lm(formula = gini_w ~ dev_log + factor(NUTS2) + factor(year),
    data = panel_data)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0273865 -0.0046302  0.0005091  0.0050317  0.0197112

Coefficients: (1 not defined because of singularities)
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        1.007e+00  3.896e-02  25.841  < 2e-16 ***
dev_log           -1.998e+00  8.364e-02 -23.887  < 2e-16 ***
factor(NUTS2)AT12  4.641e-02  2.519e-03  18.426  < 2e-16 ***
factor(NUTS2)AT21 -2.038e-02  2.960e-03  -6.886 3.02e-11 ***
factor(NUTS2)AT22  8.974e-02  2.755e-03  32.574  < 2e-16 ***
factor(NUTS2)AT31  1.025e-01  2.782e-03  36.844  < 2e-16 ***
factor(NUTS2)AT32  5.501e-02  3.458e-03  15.910  < 2e-16 ***
factor(NUTS2)AT33 -4.667e-02  2.615e-03 -17.850  < 2e-16 ***
factor(NUTS2)CZ05  7.037e-01  3.202e-02  21.978  < 2e-16 ***
factor(NUTS2)EE00  2.025e+00  8.063e-02  25.110  < 2e-16 ***
factor(NUTS2)ES11  2.034e-01  1.103e-02  18.446  < 2e-16 ***
factor(NUTS2)ES21  1.130e-02  3.319e-03   3.406 0.000742 ***
factor(NUTS2)ES24 -4.940e-02  2.606e-03 -18.954  < 2e-16 ***
factor(NUTS2)ES41 -3.032e-03  2.465e-03  -1.230 0.219595
```

```
factor(NUTS2)ES42 -9.426e-02  3.079e-03 -30.613  < 2e-16 ***
factor(NUTS2)ES51 -1.503e-01  4.122e-03 -36.462  < 2e-16 ***
factor(NUTS2)ES52 -2.360e-01  8.845e-03 -26.687  < 2e-16 ***
factor(NUTS2)ES53 -5.161e-01  1.984e-02 -26.007  < 2e-16 ***
factor(NUTS2)ES61 -6.911e-02  2.727e-03 -25.339  < 2e-16 ***
factor(NUTS2)ES70 -6.223e-01  2.432e-02 -25.590  < 2e-16 ***
factor(NUTS2)SI03 -1.886e-02  2.647e-03  -7.126 6.81e-12 ***
factor(NUTS2)SI04         NA         NA      NA       NA
factor(year)2001   1.644e-03  3.470e-03   0.474 0.636039
factor(year)2002   1.993e-03  3.411e-03   0.584 0.559472
factor(year)2003   8.686e-04  3.363e-03   0.258 0.796354
factor(year)2004   3.034e-03  3.363e-03   0.902 0.367602
factor(year)2005   2.114e-03  3.363e-03   0.629 0.530016
factor(year)2006   8.901e-04  3.363e-03   0.265 0.791429
factor(year)2007   5.234e-06  3.363e-03   0.002 0.998759
factor(year)2008  -1.549e-03  3.363e-03  -0.461 0.645425
factor(year)2009  -2.721e-03  3.363e-03  -0.809 0.419027
factor(year)2010  -2.183e-03  3.363e-03  -0.649 0.516642
factor(year)2011  -2.039e-03  3.363e-03  -0.606 0.544752
factor(year)2012  -1.751e-03  3.363e-03  -0.521 0.602926
factor(year)2013  -3.603e-03  3.363e-03  -1.071 0.284824
factor(year)2014  -4.609e-03  3.363e-03  -1.371 0.171431
factor(year)2015  -3.470e-03  3.363e-03  -1.032 0.302923
factor(year)2016  -4.513e-03  3.363e-03  -1.342 0.180500
factor(year)2017  -3.472e-03  3.363e-03  -1.033 0.302608
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.008342 on 322 degrees of freedom
Multiple R-squared:  0.975, Adjusted R-squared:  0.9721
F-statistic: 339.4 on 37 and 322 DF,  p-value: < 2.2e-16
```

4.2 Sammenlikningstabell (presentasjon av resultatene)

```r
# Samler nøkkeltall fra panelmodellene i en tabell
panel_comp <- tibble::tibble(
  modell    = c("Land-FE", "År-FE", "NUTS2-FE", "NUTS2+år-FE"),
  beta_dev  = c(
    coef(mod_fe_country)["dev_log"],
    coef(mod_fe_year)["dev_log"],
    coef(mod_fe_nuts2)["dev_log"],
    coef(mod_fe_nuts2_year)["dev_log"]
  ),
  se_dev    = c(
    summary(mod_fe_country)$coef["dev_log", "Std. Error"],
    summary(mod_fe_year)$coef["dev_log", "Std. Error"],
```

```r
    summary(mod_fe_nuts2)$coef["dev_log", "Std. Error"],
    summary(mod_fe_nuts2_year)$coef["dev_log", "Std. Error"]
  ),
  p_dev     = c(
    summary(mod_fe_country)$coef["dev_log", "Pr(>|t|)"],
    summary(mod_fe_year)$coef["dev_log", "Pr(>|t|)"],
    summary(mod_fe_nuts2)$coef["dev_log", "Pr(>|t|)"],
    summary(mod_fe_nuts2_year)$coef["dev_log", "Pr(>|t|)"]
  ),
  R2        = c(
    summary(mod_fe_country)$r.squared,
    summary(mod_fe_year)$r.squared,
    summary(mod_fe_nuts2)$r.squared,
    summary(mod_fe_nuts2_year)$r.squared
  ),
  adj_R2    = c(
    summary(mod_fe_country)$adj.r.squared,
    summary(mod_fe_year)$adj.r.squared,
    summary(mod_fe_nuts2)$adj.r.squared,
    summary(mod_fe_nuts2_year)$adj.r.squared
  )
)

panel_comp
```

```
# A tibble: 4 x 6
  modell      beta_dev  se_dev    p_dev    R2 adj_R2
  <chr>          <dbl>   <dbl>    <dbl> <dbl>  <dbl>
1 Land-FE      -0.0112 0.0186  5.46e- 1 0.675  0.670
2 År-FE         0.0902 0.00996 1.08e-17 0.198  0.155
3 NUTS2-FE     -1.99   0.0850  1.30e-72 0.973  0.971
4 NUTS2+år-FE  -2.00   0.0836  2.85e-73 0.975  0.972
```

TEKST AV KATINCA: 200-400 ord. Hvordan dev_log endrer seg på tvers av modellene hva tegnet og størrelsen på koeffisienten betyr hvilken spesifikasjon (f.eks. NUTS2+år-FE) som virker mest troverdig og hvorfor.

5 Del E: Kildebruk

5.1 Kilder

5.2 Bruk av KI-verktøy (I tråd med HVL sin bruk av KI)