Human capital panel data analysis

Code ▾

Hide

```
library(readxl)
pwt100 <- read_excel("C:/Users/Stoycho/Downloads/pwt100.xlsx",
    sheet = "Data")
pwt100<-pwt100[,-c(1,3)]
head(pwt100)
```

country	year	rgdpe	rgdpo	pop	emp	avh	hc	ccon	cda
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Aruba	1950	NA	NA	NA	NA	NA	NA	NA	NA
Aruba	1951	NA	NA	NA	NA	NA	NA	NA	NA
Aruba	1952	NA	NA	NA	NA	NA	NA	NA	NA
Aruba	1953	NA	NA	NA	NA	NA	NA	NA	NA
Aruba	1954	NA	NA	NA	NA	NA	NA	NA	NA
Aruba	1955	NA	NA	NA	NA	NA	NA	NA	NA

6 rows | 1-10 of 50 columns

Hide

```
library(plm)
df<-pwt100
df<-plm.data(df, index=c("country", "year"))
```

```
Warning: use of 'plm.data' is discouraged, better use 'pdata.frame' instead
```

Hide

```
model1<-lm(rgdpo~pop+hc+ctfp+rnna, data=df)
stargazer::stargazer(model1, type="text")
```

```

================================================
                        Dependent variable:
                    ----------------------------
                               rgdpo
------------------------------------------------
pop                         1,150.465***
                              (38.670)

hc                         -51,617.210***
                            (6,996.340)

ctfp                       106,608.200***
                            (17,040.690)

rnna                          0.227***
                              (0.001)

Constant                   -42,718.060**
                            (19,037.980)

------------------------------------------------
Observations                   6,412
R2                             0.941
Adjusted R2                    0.941
Residual Std. Error   369,447.400 (df = 6407)
F Statistic         25,710.380*** (df = 4; 6407)
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

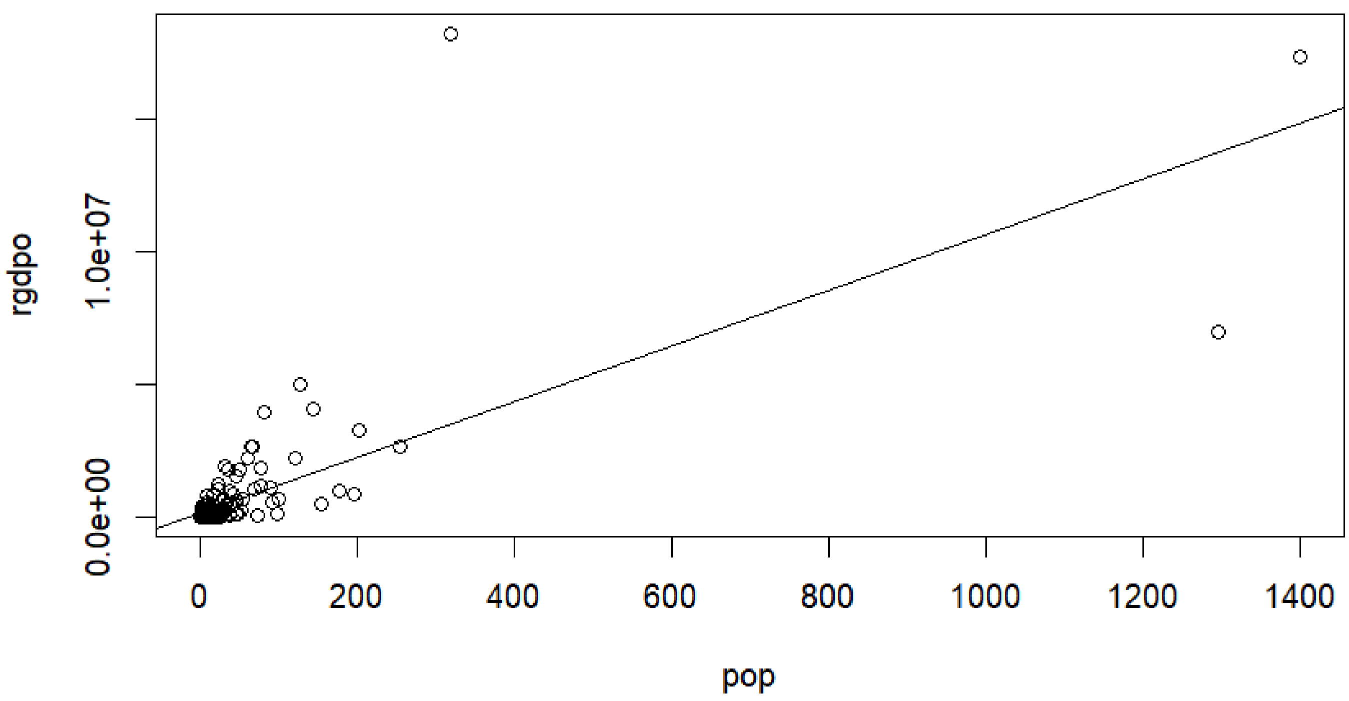
Hide

```
df2 <- df[df$year == 2014, ]
model2<-lm(rgdpo~hc+pop+rnna, data=df2)
stargazer::stargazer(model2, type="text")
```

```

===============================================
                        Dependent variable:    
                    ---------------------------
                               rgdpo           
-----------------------------------------------
hc                         -95,710.690*        
                           (53,409.810)        

pop                         839.608***         
                             (319.076)         

rnna                         0.246***          
                              (0.006)          

Constant                    173,349.400        
                           (140,566.100)       

-----------------------------------------------
Observations                    144            
R2                             0.969           
Adjusted R2                    0.968           
Residual Std. Error   404,680.300 (df = 140)   
F Statistic         1,445.395*** (df = 3; 140) 
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

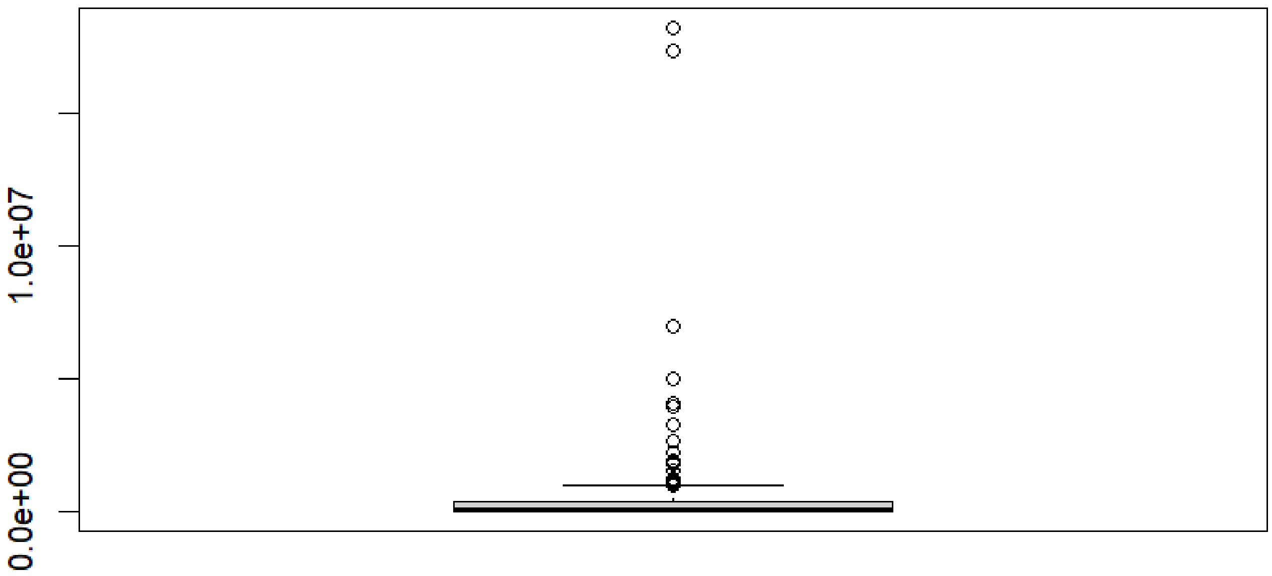
Hide

```
library(ggplot2 )
model3<-lm(rgdpo ~ pop, data = df2)
plot(rgdpo ~ pop, data = df2)
abline(model3)
```



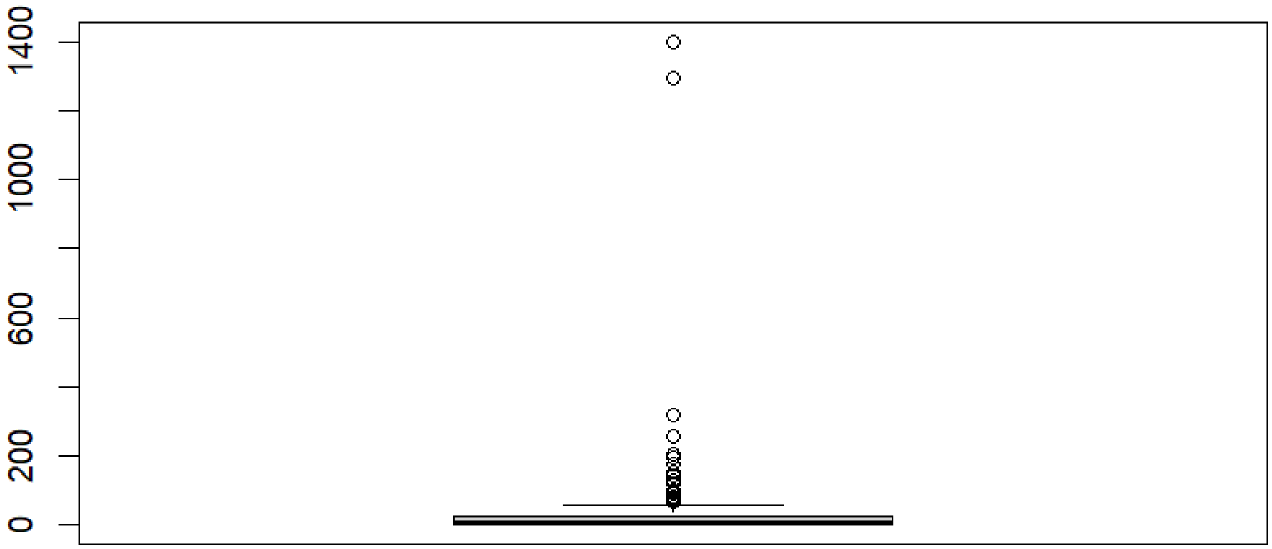
Hide

```
boxplot(df2$rgdpo)
```



Hide

```
boxplot(df2$pop)
```



Hide

```
df3<-data.frame(
  GDP=df2$rgdpo,
  hc=df2$hc,
  pop=df2$pop,
  rnna=df2$rnna
)
```

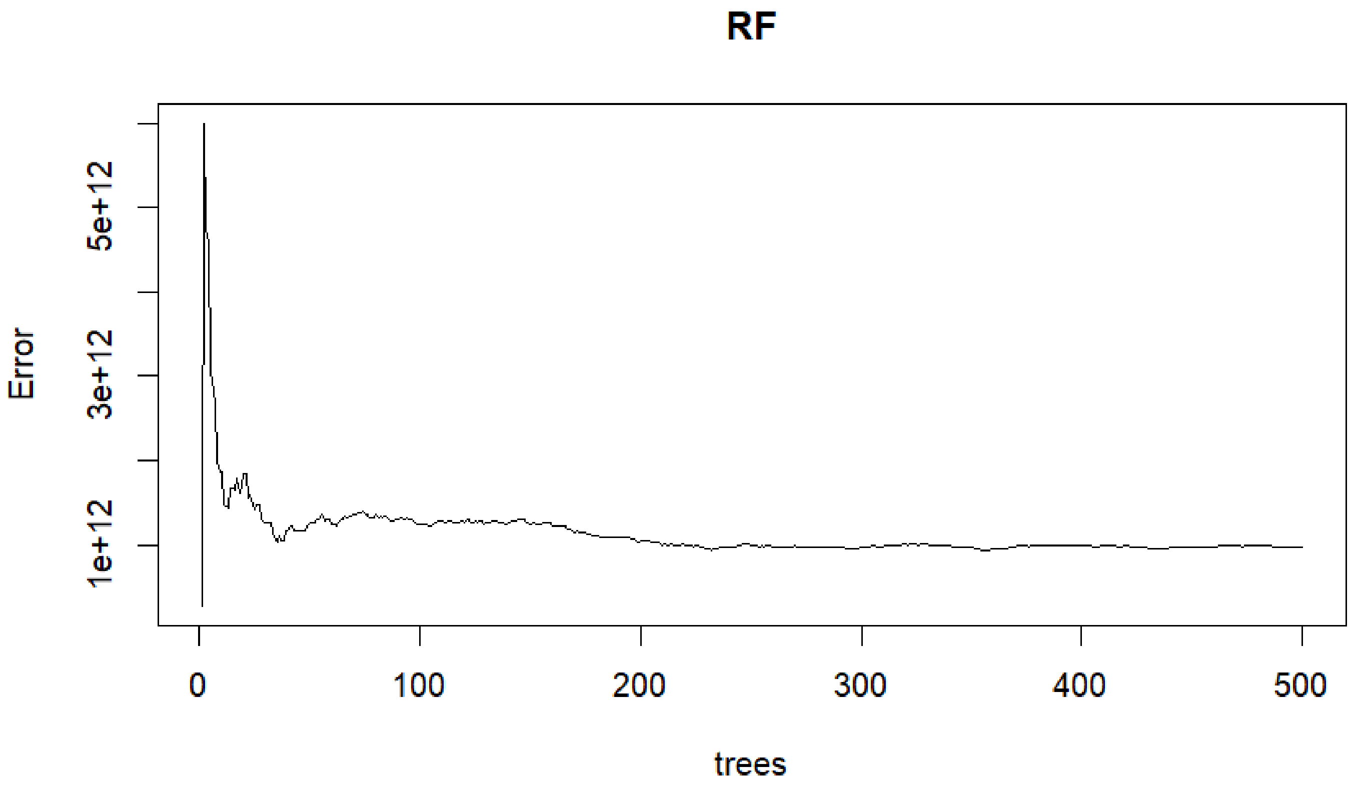
Hide

```
library(randomForest)
df3<-na.omit(df3)
RF<-randomForest(GDP~pop+rnna, data=df3)
print(RF)
```

```


Call:
 randomForest(formula = GDP ~ pop + rnna, data = df3) 
               Type of random forest: regression
                     Number of trees: 500
No. of variables tried at each split: 1

          Mean of squared residuals: 982202421661
                    % Var explained: 80.71
```

Hide

```
plot(RF)
```



Hide

```
importance(RF)
```

```
     IncNodePurity
pop   3.167396e+14
rnna  4.124764e+14
```

Hide

```

model_qr <- quantreg::rq(GDP ~ pop + rnna, data = df3, tau = 0.25)
summary(model_qr)
```

```

Call: quantreg::rq(formula = GDP ~ pop + rnna, tau = 0.25, data = df3)

tau: [1] 0.25

Coefficients:
            coefficients lower bd     upper bd
(Intercept) -11777.21594 -16154.42690  -7612.50206
pop           1695.91514    572.49549   3267.29625
rnna             0.18295      0.13149      0.19505
```

Hide

```
rho <- sum(abs(residuals(model_qr)))
model_qr_null <- quantreg::rq(GDP ~ 1, data = df3, tau = 0.5)
```

```
Warning in rq.fit.br(x, y, tau = tau, ...) : Solution may be nonunique
```

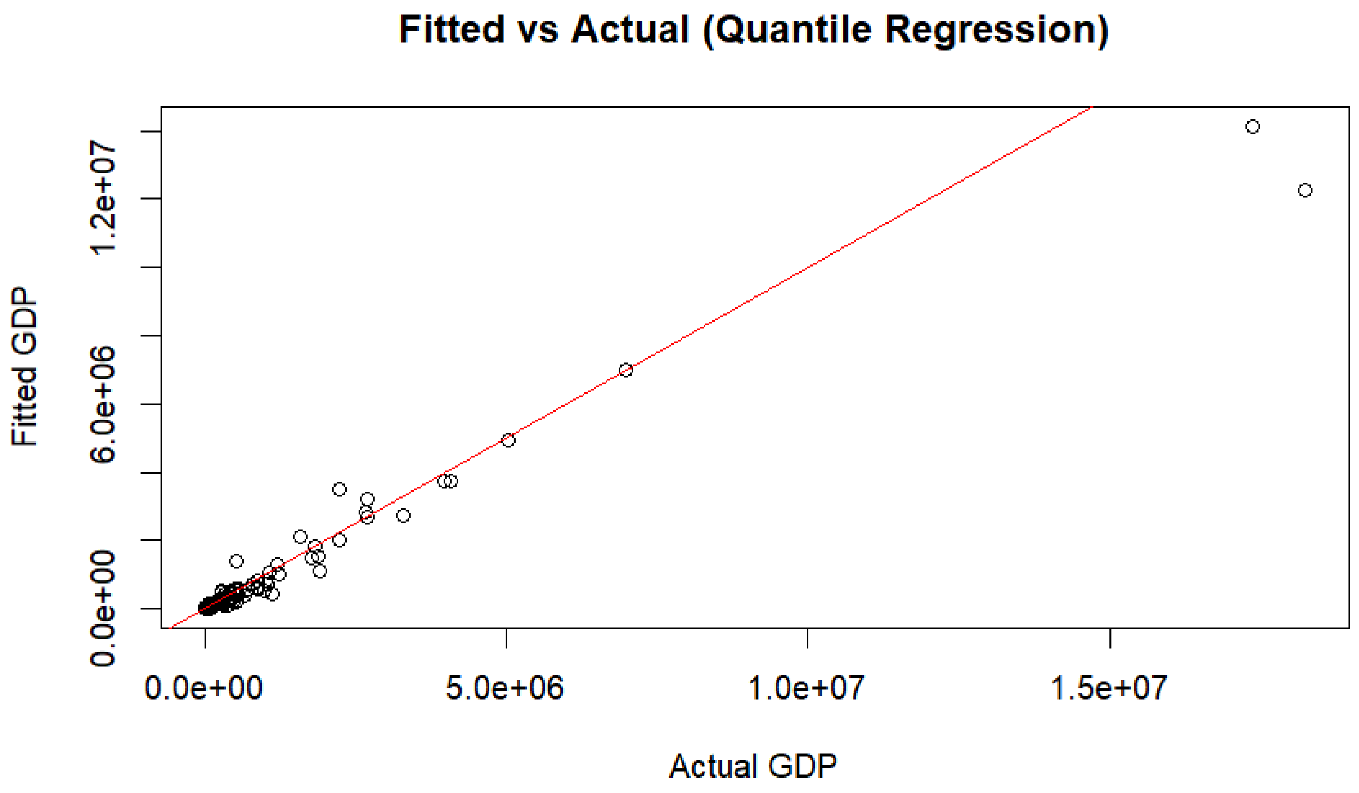
Hide

```
rho_null <- sum(abs(residuals(model_qr_null)))
pseudo_r2 <- 1 - (rho / rho_null)
pseudo_r2
```

```
[1] 0.7711102
```

Hide

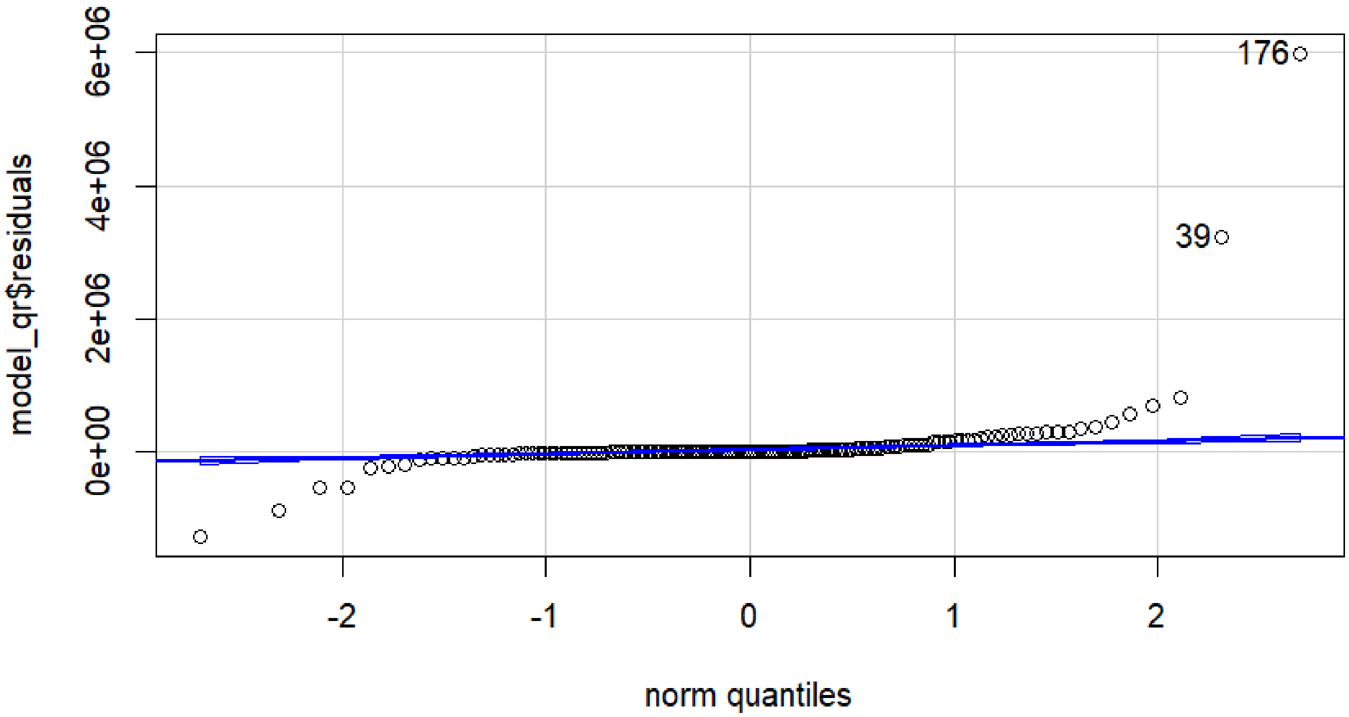
```
plot(df3$GDP, fitted(model_qr), main = "Fitted vs Actual (Quantile Regression)", xlab = "Actual GDP", ylab = "Fitt
ed GDP")
abline(0, 1, col = "red")
```



Hide

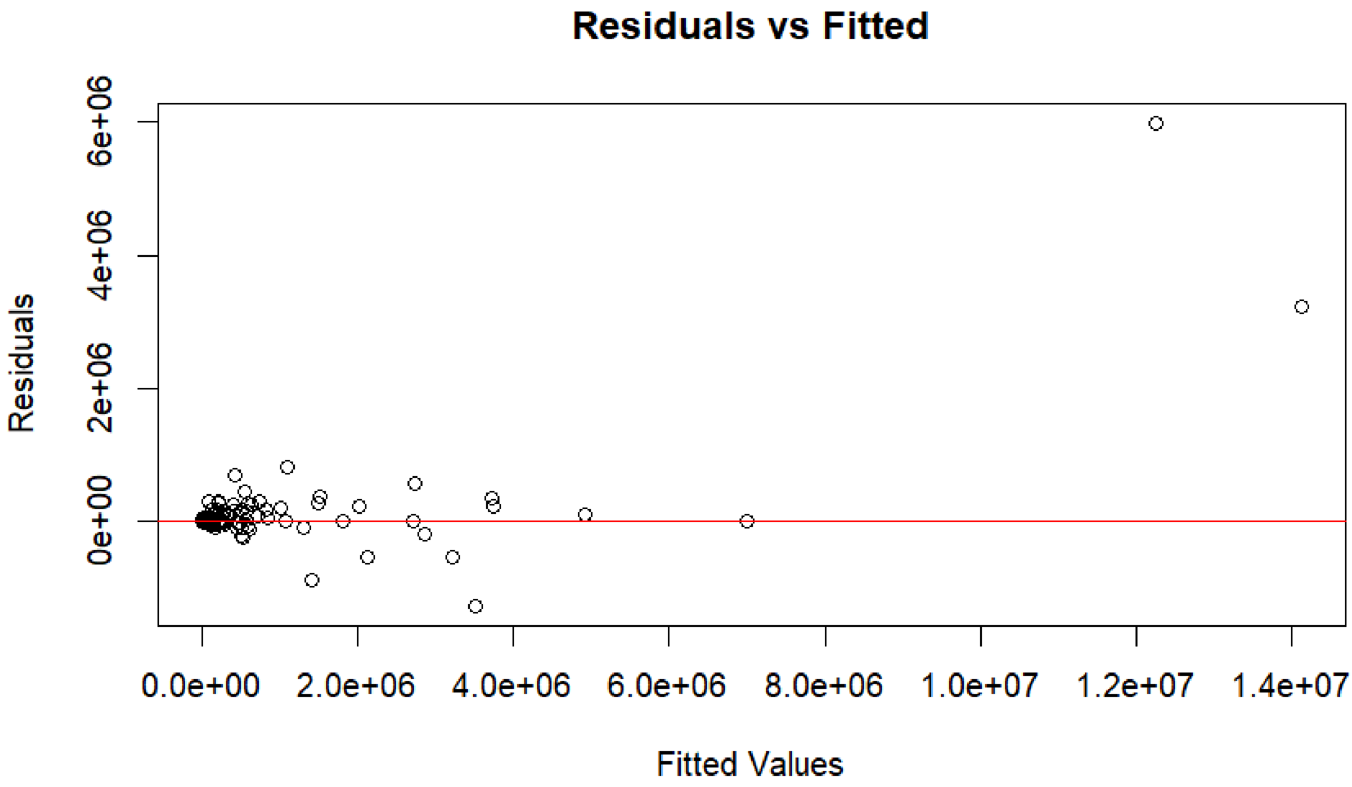
```
car::qqPlot(model_qr$residuals)
```

```
176  39
138  26
```



Hide

```
plot(fitted(model_qr), residuals(model_qr), main="Residuals vs Fitted",
     xlab="Fitted Values", ylab="Residuals")
abline(h = 0, col = "red")
```



Hide

```
# Bootstrapping to get more robust standard errors
boot_qr <- summary(model_qr, se="boot", R=1000)  # 1000 bootstrap replicates
boot_qr
```

```


Call: quantreg::rq(formula = GDP ~ pop + rnna, tau = 0.25, data = df3)

tau: [1] 0.25

Coefficients:
            Value        Std. Error   t value      Pr(>|t|)
(Intercept) -11777.21594  11070.07811     -1.06388      0.28920
pop           1695.91514    790.36519      2.14574      0.03361
rnna             0.18295      0.01689     10.83110      0.00000
```

Hide

```
boot_qr2 <- summary(model2, se="boot", R=1000)  # 1000 bootstrap replicates
boot_qr2
```

```

Call:
lm(formula = rgdpo ~ hc + pop + rnna, data = df2)

Residuals:
     Min       1Q   Median       3Q      Max 
-2287441   -25473    50921   121985  2383972 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.733e+05  1.406e+05   1.233  0.21956    
hc          -9.571e+04  5.341e+04  -1.792  0.07529 .  
pop          8.396e+02  3.191e+02   2.631  0.00946 ** 
rnna         2.460e-01  6.119e-03  40.208  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 404700 on 140 degrees of freedom
  (39 observations deleted due to missingness)
Multiple R-squared:  0.9687,    Adjusted R-squared:  0.9681 
F-statistic:  1445 on 3 and 140 DF,  p-value: < 2.2e-16
```

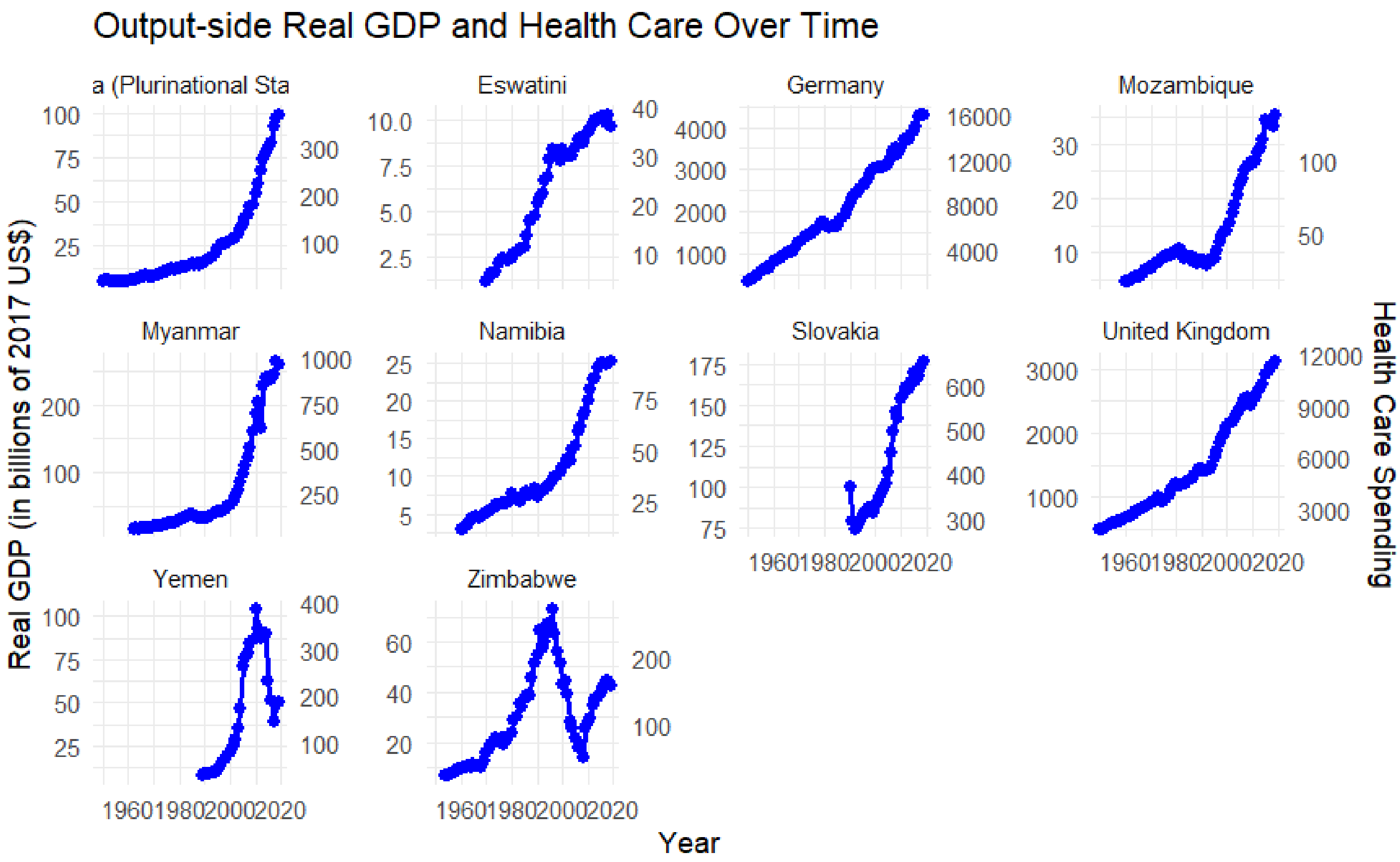
Hide

```
stargazer::stargazer(model2, model_qr,type="text")
```

```

% Error: Unrecognized object type.
```

Hide

```
library(plm)
library(ggplot2)
library(dplyr)

panel_data <- plm.data(pwt100, index = c("country"))

panel_data <- data.frame(
  year = panel_data$year,
  GDP = panel_data$rgdpe,
  HC = panel_data$hc,
  country = panel_data$country
)

panel_data <- na.omit(panel_data)

set.seed(123)
selected_countries <- sample(unique(panel_data$country), 10)

panel_data <- panel_data %>%
  filter(country %in% selected_countries)

panel_data$year <- as.numeric(panel_data$year)
panel_data$GDP <- as.numeric(panel_data$GDP)
panel_data$HC <- as.numeric(panel_data$HC)

gdp_max <- max(panel_data$GDP, na.rm = TRUE)
hc_max <- max(panel_data$HC, na.rm = TRUE)
transformation_factor <- hc_min / hc_max

plot <- ggplot(data = panel_data) +
  geom_point(aes(x = year, y = GDP / 1000), color = "blue", size = 2) +
  geom_line(aes(x = year, y = GDP / 1000), color = "blue", size = 1)  +
  facet_wrap(~country, scales = "free_y") +
  scale_y_continuous(
    name = "Real GDP (in billions of 2017 US$)",
    sec.axis = sec_axis(~ . / transformation_factor, name = "Health Care Spending")
  ) +
  labs(title = "Output-side Real GDP and Health Care Over Time",
       x = "Year") +
  theme_minimal()

print(plot)
```



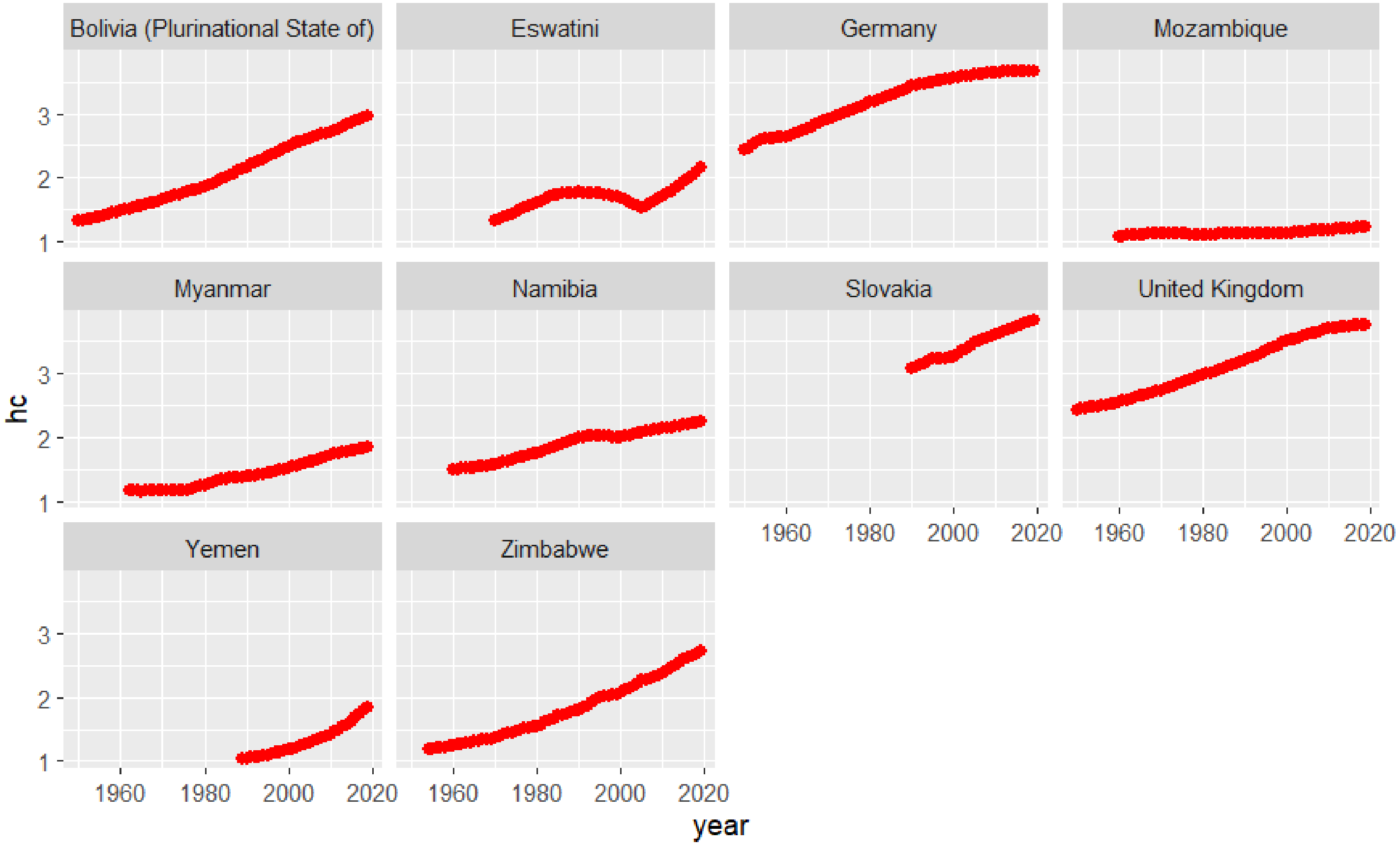
Hide

```
library(plm)
library(ggplot2)
library(dplyr)




set.seed(123)
selected_countries <- sample(unique(panel_data$country), 10)

panel_data <- panel_data %>%
  filter(country %in% selected_countries)

panel_data <- plm.data(panel_data, indexes = c("country"))

panel_data$year <- as.numeric(panel_data$year)
panel_data$hc <- as.numeric(panel_data$hc)

plot <- ggplot(data = panel_data, aes(x = year, y = hc)) +
  geom_point(color = "red", size = 2) +
  geom_line(color = "red", size = 1) +
  facet_wrap(~country)
  labs(title = "Health Care Spending Over Time",
       x = "Year",
       y = "Health Care Spending (in relevant units)") +
  theme_minimal()
```

```
NULL
```

Hide

```
print(plot)
```


Bolivia (Plurinational State of)
Eswatini
Germany
Mozambique
Myanmar
Namibia
Slovakia
United Kingdom
Yemen
Zimbabwe
hc
1
2
3
year
1960
1980
2000
2020