Liam Henson ECO1400

2025-12-01

```
rm(list=ls())
set.seed(198)
library(mlr3tuning)
```

```
## Loading required package: mlr3
```

```
## Loading required package: paradox
```

```
library(haven)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.2
## v ggplot2   4.0.0     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```
library(ivreg)
library(DoubleML)
library(mlr3)
library(mlr3learners)
library(paradox)
library(xgboost)
```

```
##
## Attaching package: 'xgboost'
##
## The following object is masked from 'package:dplyr':
##
##     slice
```

```
library(huxtable)
```

```
##
## Attaching package: 'huxtable'
```

```
##
## The following object is masked from 'package:dplyr':
##
##     add_rownames
##
## The following object is masked from 'package:ggplot2':
##
##     theme_grey
```

```
library(stargazer)
```

```
##
## Please cite as:
##
##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
library(knitr)
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'
##
## The following object is masked from 'package:huxtable':
##
##     add_footnote
##
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
library(tinytex)
library(knitr)
```

```
#reads, selects relevant variables and filterers data
data = read_dta("combined-general-household-survey.dta")
filtered_data = data %>%
  select("earn",
         "agelfted",
         "yobirth",
         "age",
         "datyear",
         "sex",
         "nireland") %>%
  filter(!is.na(agelfted) & !is.na(earn) & age <= 64 & agelfted >= 10 & agelfted <= 18) %>%
  mutate(
    yearat14 = yobirth + 14,
    yearat15 = yobirth + 15,
    learn = log(earn)
  )
```

```r
# Defines treatment and instrument variables
filtered_data = filtered_data %>%
  mutate(
    Law14 = ifelse((nireland ==0 & yearat14 >= 47) | (nireland ==1 & yearat14 >= 57),1,0),
    Law15 = ifelse(yearat15>= 73,1,0),
    School14 = as.integer(agelfted >= 15),
    School15 = as.integer(agelfted >= 16)
  )

write.csv(filtered_data,"filtered_data.csv")
filtered_data = read.csv("filtered_data.csv")

# Specifies window used to construct samples
# (band = 4 produces main results in the paper, band = 3, band = 5 produces results given in appendix)
band = 4

# Constructs samples for each reform/region
GB_47_sample = filtered_data %>%
  filter(yearat14 >= 47 - band & yearat14 <= 47 + band & nireland ==0)

NI_57_sample = filtered_data %>%
  filter(yearat14 >= 57 - band & yearat14 <= 57 + band  & nireland ==1)

GB_72_sample = filtered_data %>%
  filter(yearat15 >= 72 - band, yearat15 <= 72 + band & nireland ==0)

NI_72_sample = filtered_data %>%
  filter(yearat15 >= 72 - band, yearat15 <= 72 + band & nireland == 1)
```

```r
# Create a function to create tables to summary statistics for each region/reform sample
summary_stats = function(data, vars, var_names,Law,caption) {

  #Creates two separate data frames for pre and post law expose
  pre_treat_data = data[data[[Law]]==0,vars]
  post_treat_data = data[data[[Law]]==1,vars]

  # Computes mean and sd for each data frame rounding to 2 decimal places
  pre_teat_mean = round(apply(pre_treat_data, 2, mean),2)
  pre_teat_sd = round(apply(pre_treat_data, 2, sd),2)
  post_teat_mean = round(apply(post_treat_data, 2, mean),2)
  post_treat_sd = round(apply(post_treat_data, 2, sd),2)

  # Creates a table of estimated means and sds for each variable in each a sample
  stats_data = data.frame(
    Variable = var_names,
    Non_exposed = paste0(pre_teat_mean, " (",pre_teat_sd,")"),
    Exposed = paste0(post_teat_mean," (",post_treat_sd,")")
  )

  # Output as Latex table
  stats_table = kable(
    stats_data,
    col.names = c(
```

```
      "Variable",
      "Pre Reform Non Exposed Cohorts",
      "Post Reform Exposed Cohorts"
    ), format = "latex",
    booktabs = TRUE,
    align = "lcc",
    caption = caption
  )
  stats_table = footnote(stats_table,
                         "Values are in presented in mean (sd). Real annual earnings reflect 1998 U.K. p
  return(cat(stats_table))
}

#Defines variables and table titles
vars15 = c("learn","agelfted","age","School14")
labels15 = c("Log Real Annual Earnings","Age Left Education",
             "Age at Time of Survey","Proportion Stayed Past 14")

vars16 = c("learn","agelfted","age","School15")
labels16 = c("Log Real Annual Earnings","Age Left Education",
             "Age at Time of Survey","Proportion Stayed Past 15")


# Creates tables 1..4 in the report
captionGB1947 = paste0("Descriptive Statistics: Great Britain 1947 Sample 14 to 15 reform",
                       " (N=", nrow(GB_47_sample),")")
summary_stats(GB_47_sample, vars15, labels15,"Law14",captionGB1947)
```

```
## \begin{table}
##
## \caption{\label{tab:unnamed-chunk-4}Descriptive Statistics: Great Britain 1947 Sample 14 to 15 refor
## \centering
## \begin{tabular}[t]{lcc}
## \toprule
## Variable & Pre Reform Non Exposed Cohorts & Post Reform Exposed Cohorts\\
## \midrule
## Log Real Annual Earnings & 8.66 (1.02) & 8.79 (0.97)\\
## Age Left Education & 14.64 (1.11) & 15.33 (0.9)\\
## Age at Time of Survey & 57.37 (3.05) & 54.54 (4.33)\\
## Proportion Stayed Past 14 & 0.34 (0.47) & 0.9 (0.3)\\
## \bottomrule
## \multicolumn{3}{l}{\rule{0pt}{1em}\textit{Note: }}\\
## \multicolumn{3}{l}{\rule{0pt}{1em}Values are in presented in mean (sd). Real annual earnings reflect
## \end{tabular}
## \end{table}
```

```
captionNI1957 = paste0("Descriptive Statistics: Northern Ireland 1957 Sample 14 to 15 reform",
                       " (N=", nrow(NI_57_sample),")")
summary_stats(NI_57_sample, vars15, labels15,"Law14",captionNI1957)
```

```
## \begin{table}
##
```

```
## \caption{\label{tab:unnamed-chunk-4}Descriptive Statistics: Northern Ireland 1957 Sample 14 to 15 re
## \centering
## \begin{tabular}[t]{lcc}
## \toprule
## Variable & Pre Reform Non Exposed Cohorts & Post Reform Exposed Cohorts\\
## \midrule
## Log Real Annual Earnings & 8.7 (0.89) & 8.76 (0.9)\\
## Age Left Education & 15.09 (1.31) & 15.62 (1.09)\\
## Age at Time of Survey & 50.11 (4.16) & 45.8 (4.47)\\
## Proportion Stayed Past 14 & 0.54 (0.5) & 0.89 (0.31)\\
## \bottomrule
## \multicolumn{3}{l}{\rule{0pt}{1em}\textit{Note: }}\\
## \multicolumn{3}{l}{\rule{0pt}{1em}Values are in presented in mean (sd). Real annual earnings reflect
## \end{tabular}
## \end{table}
```

```
captionGB1973 = paste0("Descriptive Statistics: Great Britain 1972 Sample 15 to 16 reform",
                       " (N=", nrow(GB_72_sample),")")
summary_stats(GB_72_sample, vars16, labels16,"Law15",captionGB1973)
```

```
## \begin{table}
##
## \caption{\label{tab:unnamed-chunk-4}Descriptive Statistics: Great Britain 1972 Sample 15 to 16 refor
## \centering
## \begin{tabular}[t]{lcc}
## \toprule
## Variable & Pre Reform Non Exposed Cohorts & Post Reform Exposed Cohorts\\
## \midrule
## Log Real Annual Earnings & 9.05 (1.11) & 9.06 (1.1)\\
## Age Left Education & 15.94 (1.02) & 16.33 (0.84)\\
## Age at Time of Survey & 39.78 (7.27) & 35.45 (7.23)\\
## Proportion Stayed Past 15 & 0.61 (0.49) & 0.92 (0.27)\\
## \bottomrule
## \multicolumn{3}{l}{\rule{0pt}{1em}\textit{Note: }}\\
## \multicolumn{3}{l}{\rule{0pt}{1em}Values are in presented in mean (sd). Real annual earnings reflect
## \end{tabular}
## \end{table}
```

```
captionNI173 = paste0("Descriptive Statistics: Northern Ireland 1972 Sample 15 to 16 reform",
                      " (N=", nrow(NI_72_sample),")")
summary_stats(NI_72_sample, vars16, labels16,"Law15",captionNI173)
```

```
## \begin{table}
##
## \caption{\label{tab:unnamed-chunk-4}Descriptive Statistics: Northern Ireland 1972 Sample 15 to 16 re
## \centering
## \begin{tabular}[t]{lcc}
## \toprule
## Variable & Pre Reform Non Exposed Cohorts & Post Reform Exposed Cohorts\\
## \midrule
## Log Real Annual Earnings & 8.9 (0.79) & 8.88 (0.73)\\
## Age Left Education & 16.16 (1.05) & 16.5 (0.84)\\
## Age at Time of Survey & 36.56 (4.7) & 32.05 (4.66)\\
```

```
## Proportion Stayed Past 15 & 0.7 (0.46) & 0.95 (0.23)\\
## \bottomrule
## \multicolumn{3}{l}{\rule{0pt}{1em}\textit{Note: }}\\
## \multicolumn{3}{l}{\rule{0pt}{1em}Values are in presented in mean (sd). Real annual earnings reflect
## \end{tabular}
## \end{table}
```

```
# A function to plot proportion of students which stay in school past given legal threshold over time
year_plot = function(data,year_var,school_var,min_year,max_year,ylab,reform_year) {

  # Filters which years we want to include in the plot based on min_year and max_year
  data = data[data[[year_var]]>=min_year & data[[year_var]]<=max_year,]

  # Defines years plotted as a factor for easier indexing as defining as a factor automatically orders
  data[[year_var]] = factor(data[[year_var]])
  years = levels(data[[year_var]])

  # Creates empty vector of proportions for each year
  avg_left_school = rep(0, length(years))

  # Creates vector of proportions for each year
  for (i in 1:length(years)) {
    avg_left_school[i] = mean(data[data[[year_var]]==years[i],school_var])
  }


  plot(as.numeric(years) + 1900, avg_left_school,xlab = "Year",ylab = ylab,type = "o",ylim = c(0.2,1))

  #Adds vertical ling to the plot to specify the year the reform was implemented
  abline(v=reform_year,lty = 5)

  #Adds year od reform next to vertical line on the plot
  text(reform_year-3, 0.92, labels = reform_year, pos = 4, offset = 0.3, cex = 0.9)
}
```

```
#Filters data to have just Great Britain observations
filtered_data_GB = filtered_data %>% filter(nireland ==0)

# Creates Figure 1 in the report
png("GB_Figure1.png", width = 10, height = 4,units = "in", res = 500)
par(mfrow = c(1,2), mar = c(5, 5, 2, 2))
ylab = expression("Proportion staying in school past age 14")
year_plot(filtered_data_GB,"yearat14","School14",37,57,ylab,1947)


ylab = expression("Proportion staying in school past age 15")
year_plot(filtered_data_GB,"yearat15","School15",62,82,ylab,1972)
dev.off()
```

```
## pdf
##   2
```

```r
#Filters data to have just Great Britain observations
filtered_data_GB = filtered_data %>% filter(nireland ==1)

# Creates Figure 2 in report
png("NI_Figure2.png", width = 10, height = 4,units = "in", res = 500)
par(mfrow = c(1,2), mar = c(5, 5, 2, 2))
ylab = expression("Proportion staying in school past age 14")
year_plot(filtered_data_GB,"yearat14","School14",44,67,ylab,1957)


ylab = expression("Proportion staying in school past age 15")
year_plot(filtered_data_GB,"yearat15","School15",60,80,ylab,1972)
dev.off()
```

```
## pdf
##   2
```

```r
#Defines the learners to be used in the analysis

#Defines random forest
ml_gRf = lrn("regr.ranger")
ml_mRf = lrn("classif.ranger", predict_type = "prob")
ml_rRf = lrn("classif.ranger", predict_type = "prob")

#Defines XGboost
ml_gXG = lrn("regr.xgboost")
ml_mXG = lrn("classif.xgboost", predict_type = "prob")
ml_rXG = lrn("classif.xgboost", predict_type = "prob")

#Defines elastic net
ml_gGlm = lrn("regr.cv_glmnet")
ml_mGlm = lrn("classif.cv_glmnet", predict_type = "prob")
ml_rGlm = lrn("classif.cv_glmnet", predict_type = "prob")
```

```r
# Define tuning parameters for Random Forest, XGBoost and Elastic Net
Rf_tuning_params = ps(
  num.trees = p_int(lower = 150, upper = 650),
  max.depth = p_int(lower = 3, upper = 9),
  min.node.size = p_int(lower = 55, upper = 220)
)

Xgb_tuning_params = ps(
  nrounds = p_int(lower = 60, upper = 250),
  max_depth = p_int(lower = 3, upper = 9),
  min_child_weight = p_dbl(lower = 1, upper = 8),
  gamma = p_dbl(lower = 0, upper = 4.),
  eta = p_dbl(lower = 0.01, upper = 0.4)
)

Net_tuning_param = ps(alpha = p_dbl(lower = 0.01, upper = 0.99))
```

```r
# Fits IIVM models for Great Britain 1947 reform 5-fold cross-fitting
x_15 = c("yearat14", "age", 'datyear', 'sex')
IIVM_14_GB_data = DoubleMLData$new(
  data = GB_47_sample,
  y_col = "learn",
  d_cols = "School14",
  z_cols = "Law14",
  x_cols = x_15
)

#Fits Random Forest model
IIVM_14_GB_RF_5_fold = DoubleMLIIVM$new(
  IIVM_14_GB_data,
  ml_gRf,
  ml_mRf,
  ml_rRf,
  n_folds = 5,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 10,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_14_GB_RF_5_fold$tune(param_set = list("ml_g" = Rf_tuning_params,
                                           "ml_m" = Rf_tuning_params,
                                           "ml_r" = Rf_tuning_params))
IIVM_14_GB_RF_5_fold$fit()
IIVM_14_GB_RF_5_fold$coef


#Fits XGBoost model
IIVM_14_GB_XG_5_fold = DoubleMLIIVM$new(
  IIVM_14_GB_data,
  ml_gXG,
  ml_mXG,
  ml_rXG,
  n_folds = 5,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 10,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_14_GB_XG_5_fold$tune(param_set = list("ml_g" = Xgb_tuning_params,
                                           "ml_m" = Xgb_tuning_params,
                                           "ml_r" = Xgb_tuning_params))
IIVM_14_GB_XG_5_fold$fit()
IIVM_14_GB_XG_5_fold$coef

#Fits Elastic Net model
IIVM_14_GB_Glm_5_fold = DoubleMLIIVM$new(
```

```
  IIVM_14_GB_data,
  ml_gGlm,
  ml_mGlm,
  ml_rGlm,
  n_folds = 10,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 5,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_14_GB_Glm_5_fold$tune(param_set = list("ml_g" = Net_tuning_param,
                                            "ml_m" = Net_tuning_param,
                                            "ml_r" = Net_tuning_param))
IIVM_14_GB_Glm_5_fold$fit()
IIVM_14_GB_Glm_5_fold$coef



# Fits IIVM model for Great Britain 1947 reform 10-fold cross-fitting

#Fits Random Forest model
IIVM_14_GB_RF_10_fold = DoubleMLIIVM$new(
  IIVM_14_GB_data,
  ml_gRf,
  ml_mRf,
  ml_rRf,
  n_folds = 10,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 5,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_14_GB_RF_10_fold$tune(param_set = list("ml_g" = Rf_tuning_params,
                                            "ml_m" = Rf_tuning_params,
                                            "ml_r" = Rf_tuning_params))
IIVM_14_GB_RF_10_fold$fit()
IIVM_14_GB_RF_10_fold$coef


#Fits XGBoost  model
IIVM_14_GB_XG_10_fold = DoubleMLIIVM$new(
  IIVM_14_GB_data,
  ml_gXG,
  ml_mXG,
  ml_rXG,
  n_folds = 10,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 5,
```

```
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_14_GB_XG_10_fold$tune(param_set = list("ml_g" = Xgb_tuning_params,
                                            "ml_m" = Xgb_tuning_params,
                                            "ml_r" = Xgb_tuning_params))
IIVM_14_GB_XG_10_fold$fit()
IIVM_14_GB_XG_10_fold$coef

#Fits Elastic Net model
IIVM_14_GB_Glm_10_fold = DoubleMLIIVM$new(
  IIVM_14_GB_data,
  ml_gGlm,
  ml_mGlm,
  ml_rGlm,
  n_folds = 10,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 5,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_14_GB_Glm_10_fold$tune(param_set = list("ml_g" = Net_tuning_param,
                                            "ml_m" = Net_tuning_param,
                                            "ml_r" = Net_tuning_param))
IIVM_14_GB_Glm_10_fold$fit()
IIVM_14_GB_Glm_10_fold$coef
```

```
# Fits IIVM model for Northern Ireland 1957 reform 5-fold cross-fitting
x_15 = c("yearat14", "age", 'datyear', 'sex')
IIVM_14_NI_data = DoubleMLData$new(
  data  = NI_57_sample,
  y_col = "learn",
  d_cols = "School14",
  z_cols = "Law14",
  x_cols = x_15
)

#Fits Random Forest model
IIVM_14_NI_RF_5_fold = DoubleMLIIVM$new(
  IIVM_14_NI_data,
  ml_gRf,
  ml_mRf,
  ml_rRf,
  n_folds = 5,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 10,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)
```

```
IIVM_14_NI_RF_5_fold$tune(param_set = list("ml_g" = Rf_tuning_params,
                                           "ml_m" = Rf_tuning_params,
                                           "ml_r" = Rf_tuning_params))
IIVM_14_NI_RF_5_fold$fit()
IIVM_14_NI_RF_5_fold$coef


#Fits XGBoost  model
IIVM_14_NI_XG_5_fold = DoubleMLIIVM$new(
  IIVM_14_NI_data,
  ml_gXG,
  ml_mXG,
  ml_rXG,
  n_folds = 5,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 10,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_14_NI_XG_5_fold$tune(param_set = list("ml_g" = Xgb_tuning_params,
                                           "ml_m" = Xgb_tuning_params,
                                           "ml_r" = Xgb_tuning_params))
IIVM_14_NI_XG_5_fold$fit()
IIVM_14_NI_XG_5_fold$coef

#Fits Elastic Net model
IIVM_14_NI_Glm_5_fold = DoubleMLIIVM$new(
  IIVM_14_NI_data,
  ml_gGlm,
  ml_mGlm,
  ml_rGlm,
  n_folds = 10,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 5,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_14_NI_Glm_5_fold$tune(param_set = list("ml_g" = Net_tuning_param,
                                           "ml_m" = Net_tuning_param,
                                           "ml_r" = Net_tuning_param))
IIVM_14_NI_Glm_5_fold$fit()
IIVM_14_NI_Glm_5_fold$coef




# Fits IIVM model for Northern Ireland 1957 reform 10-fold cross-fitting

#Fits Random Forest model
```

```
IIVM_14_NI_RF_10_fold = DoubleMLIIVM$new(
  IIVM_14_NI_data,
  ml_gRf,
  ml_mRf,
  ml_rRf,
  n_folds = 10,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 5,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_14_NI_RF_10_fold$tune(param_set = list("ml_g" = Rf_tuning_params,
                                            "ml_m" = Rf_tuning_params,
                                            "ml_r" = Rf_tuning_params))
IIVM_14_NI_RF_10_fold$fit()
IIVM_14_NI_RF_10_fold$coef


#Fits XGBoost  model
IIVM_14_NI_XG_10_fold = DoubleMLIIVM$new(
  IIVM_14_NI_data,
  ml_gXG,
  ml_mXG,
  ml_rXG,
  n_folds = 10,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 5,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_14_NI_XG_10_fold$tune(param_set = list("ml_g" = Xgb_tuning_params,
                                            "ml_m" = Xgb_tuning_params,
                                            "ml_r" = Xgb_tuning_params))
IIVM_14_NI_XG_10_fold$fit()
IIVM_14_NI_XG_10_fold$coef

#Fits Elastic Net model
IIVM_14_NI_Glm_10_fold = DoubleMLIIVM$new(
  IIVM_14_NI_data,
  ml_gGlm,
  ml_mGlm,
  ml_rGlm,
  n_folds = 10,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 5,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)
```

```
IIVM_14_NI_Glm_10_fold$tune(param_set = list("ml_g" = Net_tuning_param,
                                             "ml_m" = Net_tuning_param,
                                             "ml_r" = Net_tuning_param))
IIVM_14_NI_Glm_10_fold$fit()
IIVM_14_NI_Glm_10_fold$coef
```

```
# Fits IIVM model for Great Britain 1973 reform 10-fold cross-fitting
x_16 = c("yearat15", "age", 'datyear', 'sex')
IIVM_15_GB_data = DoubleMLData$new(
  data  = GB_72_sample,
  y_col = "learn",
  d_cols = "School15",
  z_cols = "Law15",
  x_cols = x_16
)

#Fits Random Forest model
IIVM_15_GB_RF_5_fold = DoubleMLIIVM$new(
  IIVM_15_GB_data,
  ml_gRf,
  ml_mRf,
  ml_rRf,
  n_folds = 5,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 10,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_15_GB_RF_5_fold$tune(param_set = list("ml_g" = Rf_tuning_params,
                                           "ml_m" = Rf_tuning_params,
                                           "ml_r" = Rf_tuning_params))
IIVM_15_GB_RF_5_fold$fit()
IIVM_15_GB_RF_5_fold$coef


#Fits XGBoost  model
IIVM_15_GB_XG_5_fold = DoubleMLIIVM$new(
  IIVM_15_GB_data,
  ml_gXG,
  ml_mXG,
  ml_rXG,
  n_folds = 5,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 10,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_15_GB_XG_5_fold$tune(param_set = list("ml_g" = Xgb_tuning_params,
                                           "ml_m" = Xgb_tuning_params,
```

```
                                      "ml_r" = Xgb_tuning_params))
IIVM_15_GB_XG_5_fold$fit()
IIVM_15_GB_XG_5_fold$coef

#Fits Elastic Net model
IIVM_15_GB_Glm_5_fold = DoubleMLIIVM$new(
  IIVM_15_GB_data,
  ml_gGlm,
  ml_mGlm,
  ml_rGlm,
  n_folds = 10,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 5,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_15_GB_Glm_5_fold$tune(param_set = list("ml_g" = Net_tuning_param,
                                      "ml_m" = Net_tuning_param,
                                      "ml_r" = Net_tuning_param))
IIVM_15_GB_Glm_5_fold$fit()
IIVM_15_GB_Glm_5_fold$coef




# Fits IIVM model for Great Britain 1973 reform 10-fold cross-fitting
#Fits Random Forest model
IIVM_15_GB_RF_10_fold = DoubleMLIIVM$new(
  IIVM_15_GB_data,
  ml_gRf,
  ml_mRf,
  ml_rRf,
  n_folds = 10,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 5,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_15_GB_RF_10_fold$tune(param_set = list("ml_g" = Rf_tuning_params,
                                      "ml_m" = Rf_tuning_params,
                                      "ml_r" = Rf_tuning_params))
IIVM_15_GB_RF_10_fold$fit()
IIVM_15_GB_RF_10_fold$coef


#Fits XGBoost  model
IIVM_15_GB_XG_10_fold = DoubleMLIIVM$new(
  IIVM_15_GB_data,
  ml_gXG,
```

```
  ml_mXG,
  ml_rXG,
  n_folds = 10,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 5,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_15_GB_XG_10_fold$tune(param_set = list("ml_g" = Xgb_tuning_params,
                                            "ml_m" = Xgb_tuning_params,
                                            "ml_r" = Xgb_tuning_params))
IIVM_15_GB_XG_10_fold$fit()
IIVM_15_GB_XG_10_fold$coef

#Fits Elastic Net model
IIVM_15_GB_Glm_10_fold = DoubleMLIIVM$new(
  IIVM_15_GB_data,
  ml_gGlm,
  ml_mGlm,
  ml_rGlm,
  n_folds = 10,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 5,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_15_GB_Glm_10_fold$tune(param_set = list("ml_g" = Net_tuning_param,
                                            "ml_m" = Net_tuning_param,
                                            "ml_r" = Net_tuning_param))
IIVM_15_GB_Glm_10_fold$fit()
IIVM_15_GB_Glm_10_fold$coef
```

```
# Fits IIVM model for Northern Ireland 1973 reform 5-fold cross-fitting
x_16 = c("yearat15", "age", 'datyear', 'sex')
IIVM_15_NI_data = DoubleMLData$new(
  data  = NI_72_sample,
  y_col = "learn",
  d_cols = "School15",
  z_cols = "Law15",
  x_cols = x_16
)

#Fits Random Forest model
IIVM_15_NI_RF_5_fold = DoubleMLIIVM$new(
  IIVM_15_NI_data,
  ml_gRf,
  ml_mRf,
  ml_rRf,
  n_folds = 5,
```

```
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 10,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_15_NI_RF_5_fold$tune(param_set = list("ml_g" = Rf_tuning_params,
                                           "ml_m" = Rf_tuning_params,
                                           "ml_r" = Rf_tuning_params))
IIVM_15_NI_RF_5_fold$fit()
IIVM_15_NI_RF_5_fold$coef


#Fits XGBoost  model
IIVM_15_NI_XG_5_fold = DoubleMLIIVM$new(
  IIVM_15_NI_data,
  ml_gXG,
  ml_mXG,
  ml_rXG,
  n_folds = 5,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 10,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_15_NI_XG_5_fold$tune(param_set = list("ml_g" = Xgb_tuning_params,
                                           "ml_m" = Xgb_tuning_params,
                                           "ml_r" = Xgb_tuning_params))
IIVM_15_NI_XG_5_fold$fit()
IIVM_15_NI_XG_5_fold$coef

#Fits Elastic Net model
IIVM_15_NI_Glm_5_fold = DoubleMLIIVM$new(
  IIVM_15_NI_data,
  ml_gGlm,
  ml_mGlm,
  ml_rGlm,
  n_folds = 10,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 5,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_15_NI_Glm_5_fold$tune(param_set = list("ml_g" = Net_tuning_param,
                                           "ml_m" = Net_tuning_param,
                                           "ml_r" = Net_tuning_param))
IIVM_15_NI_Glm_5_fold$fit()
IIVM_15_NI_Glm_5_fold$coef
```

```r
# Fits IIVM model for Northern Ireland 1973 reform 10-fold cross-fitting

#Fits Random Forest model
IIVM_15_NI_RF_10_fold = DoubleMLIIVM$new(
  IIVM_15_NI_data,
  ml_gRf,
  ml_mRf,
  ml_rRf,
  n_folds = 10,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 5,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_15_NI_RF_10_fold$tune(param_set = list("ml_g" = Rf_tuning_params,
                                            "ml_m" = Rf_tuning_params,
                                            "ml_r" = Rf_tuning_params))
IIVM_15_NI_RF_10_fold$fit()
IIVM_15_NI_RF_10_fold$coef


#Fits XGBoost  model
IIVM_15_NI_XG_10_fold = DoubleMLIIVM$new(
  IIVM_15_NI_data,
  ml_gXG,
  ml_mXG,
  ml_rXG,
  n_folds = 10,
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 5,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_15_NI_XG_10_fold$tune(param_set = list("ml_g" = Xgb_tuning_params,
                                            "ml_m" = Xgb_tuning_params,
                                            "ml_r" = Xgb_tuning_params))
IIVM_15_NI_XG_10_fold$fit()
IIVM_15_NI_XG_10_fold$coef

#Fits Elastic Net model
IIVM_15_NI_Glm_10_fold = DoubleMLIIVM$new(
  IIVM_15_NI_data,
  ml_gGlm,
  ml_mGlm,
  ml_rGlm,
  n_folds = 10,
```

```
  apply_cross_fitting = TRUE,
  score = "LATE",
  n_rep = 5,
  trimming_threshold = 0.05,
  subgroups = list(always_takers = TRUE, never_takers = TRUE)
)

IIVM_15_NI_Glm_10_fold$tune(param_set = list("ml_g" = Net_tuning_param,
                                             "ml_m" = Net_tuning_param,
                                             "ml_r" = Net_tuning_param))
IIVM_15_NI_Glm_10_fold$fit()
IIVM_15_NI_Glm_10_fold$coef
```

```
# Defines IV models for each region/reform sample
IvGB47 = ivreg(learn ~ School14 + age + sex + yearat14 + datyear
               | Law14 + age + sex + yearat14 + datyear,data = GB_47_sample)

IvNI57 = ivreg(learn ~ School14 + age + sex + yearat14 + datyear
               | Law14 + age + sex + yearat14 + datyear,data = NI_57_sample)

IvGB72 = ivreg(learn ~ School15 + age + sex + yearat15 + datyear
               | Law15 + age + sex + yearat15 + datyear,data = GB_72_sample)

IvNI72 = ivreg(learn ~ School15 + age + sex + yearat15 + datyear
               | Law15 + age + sex + yearat15 + datyear,data = NI_72_sample)
```

```
# A function to create table of results (coef, CI) for each IV model
get_params = function(model) {
  #Creates empty list of model parameters
  params = list()

  #gets p-value for variable of interest
  pval = summary(model)$coefficients[2,4]
  pval_F = as.numeric(summary(model, diagnostics = TRUE)$diagnostics["Weak instruments", "p-value"])
  #Adds parameters to the list and rounds to 3 decimal places
  params[[1]] = round(as.numeric(model$coefficients[2]), 3)
  params[[2]] = round(as.numeric(confint(model)[2, 1]), 3)
  params[[3]] = round(as.numeric(confint(model)[2, 2]), 3)
  params[[4]] = case_when(pval<0.001~"***",
                          pval<0.01~"**",
                          pval<0.05~"*",
                          .default = "")
  params[[5]] = round(as.numeric(summary(model,
                                 diagnostics = TRUE)$diagnostics["Weak instruments", "statistic"]), 3
  params[[6]] = case_when(pval_F<0.001~"***",
                          pval_F<0.01~"**",
                          pval_F<0.05~"*",
                          .default = "")
  return(params)
}

# Creates a list of each IV model
models = list(IvGB47, IvNI57, IvGB72, IvNI72)
```

```r
# Creates a matrix of model parameters
params = sapply(models, get_params)

#Creates data frame of, learners, estimated coefs, estimated CIs, and weak instrument F-stats
results = data.frame(
  Sample = c("GB 1947 Sample", "NI 1957 Sample", "GB 1972 Sample", "NI 1972 Sample"),
  Fstat = paste0(as.numeric(params[5, ]),params[6,]),
  Coef = paste0(as.numeric(params[1, ]),params[4,]),
  CI = paste0("[", params[2, ], ",", params[3, ], "]")
)

#Outputs results in a latex table (Creates table 5 in the report)
Iv_table = kable(
  results,
  col.names = c(
    "Sample",
    "Weak Instrument Test F-Stat",
    "Coefficient",
    "95 Confidence Interval"

  ),
  format = 'latex',
  booktabs = TRUE,
  align = "lccc",
  caption = "2SLS estimates for the effect of staying in school
  past the minimum school leaving age on log real earnings"
)

Iv_table = footnote(
  Iv_table,
  "Covariates included are age, year at time of reform, sex, year of survey"
)

cat(Iv_table)
```

```
## \begin{table}
##
## \caption{\label{tab:unnamed-chunk-15}2SLS estimates for the effect of staying in school
##   past the minimum school leaving age on log real earnings}
## \centering
## \begin{tabular}[t]{lccc}
## \toprule
## Sample & Weak Instrument Test F-Stat & Coefficient & 95 Confidence Interval\\
## \midrule
## GB 1947 Sample & 624.065*** & 0.227* & {}[0.051,0.403]\\
## NI 1957 Sample & 31.32*** & 0.105 & {}[-0.52,0.729]\\
## GB 1972 Sample & 634.86*** & -0.124 & {}[-0.309,0.062]\\
## NI 1972 Sample & 41.952*** & -0.332 & {}[-0.899,0.235]\\
## \bottomrule
## \multicolumn{4}{l}{\rule{0pt}{1em}\textit{Note: }}\\
## \multicolumn{4}{l}{\rule{0pt}{1em}Covariates included are age, year at time of reform, sex, year of
## \end{tabular}
## \end{table}
```

```
# A function to create tables of results coef, CI for IIVM model
dml_results = function(models, caption) {

  # A function that will be used to extract coefs and CIs from each given model
  get_params = function(model) {
    #Creates empty list of model parameters
    params = list()

    #Adds parameters to the list and rounds to 3 decimal places
    params[[1]] = round(as.numeric(model$coef), 3)
    params[[2]] = round(as.numeric(model$confint()[1]), 3)
    params[[3]] = round(as.numeric(model$confint()[2]), 3)
    params[[4]] = case_when(as.numeric(model$pval)<0.001~"***",
                            as.numeric(model$pval)<0.01~"**",
                          as.numeric(model$pval)<0.05~"*",,
                          .default = "")
    return(params)
  }

  # Creates a matrix of model parameters
  params = sapply(models, get_params)

  #Creates data frame of folds, learners, estimated coefs, estimated CIs
  results = data.frame(
    K = c("K = 5", "", "", "K = 10", "", ""),
    Learner = rep(c("Random Forest", "XGBoost", "Elastic Net"), 2),
    Coef = paste0(as.numeric(params[1, ]),params[4, ]),
    CI = paste0("[", params[2, ], ",", params[3, ], "]")
  )

  #Outputs results in a latex table
  dml_results = kable(
    results,
    col.names = c("Cross-fitting Folds", "Learner", "Coefficient", "95 Confidence Interval"),
    format = 'latex',
    booktabs = TRUE,
    align = "lcc",
    caption = caption
  )
  dml_results = footnote(dml_results,"Covariates included are age,
                         year at time of reform, sex, year of survey")
  return(cat(dml_results))
}
```

```
#Outputs IIVM model results in latex tables for each region/reform combination
#Creating tables 6..9 in the report
dml_results(
  list(
    IIVM_14_GB_RF_5_fold,
    IIVM_14_GB_XG_5_fold,
    IIVM_14_GB_Glm_5_fold,
    IIVM_14_GB_RF_10_fold,
    IIVM_14_GB_XG_10_fold,
```

```
    IIVM_14_GB_Glm_10_fold
  ),
  "Interactive IV LATE estimates for the effect of staying
  in school past age 15 on log real earnings: Great Britain, 1947 Reform"
)
```

```
## \begin{table}
##
## \caption{\label{tab:unnamed-chunk-17}Interactive IV LATE estimates for the effect of staying
##   in school past age 15 on log real earnings: Great Britain, 1947 Reform}
## \centering
## \begin{tabular}[t]{lccl}
## \toprule
## Cross-fitting Folds & Learner & Coefficient & 95 Confidence Interval\\
## \midrule
## K = 5 & Random Forest & 0.344*** & {}[0.31,0.378]\\
##  & XGBoost & 0.329*** & {}[0.295,0.363]\\
##  & Elastic Net & 0.317*** & {}[0.283,0.352]\\
## K = 10 & Random Forest & 0.347*** & {}[0.314,0.381]\\
##  & XGBoost & 0.299*** & {}[0.264,0.334]\\
## \addlinespace
##  & Elastic Net & 0.331*** & {}[0.296,0.366]\\
## \bottomrule
## \multicolumn{4}{l}{\rule{0pt}{1em}\textit{Note: }}\\
## \multicolumn{4}{l}{\rule{0pt}{1em}makecell[l]{Covariates included are age, \
## \end{tabular}
## \end{table}
```

```
dml_results(
  list(
    IIVM_14_NI_RF_5_fold,
    IIVM_14_NI_XG_5_fold,
    IIVM_14_NI_Glm_5_fold,
    IIVM_14_NI_RF_10_fold,
    IIVM_14_NI_XG_10_fold,
    IIVM_14_NI_Glm_10_fold
  ),
  "Interactive IV LATE estimates for the effect of staying
  in school past age 15 on log real earnings: Northern Ireland, 1957 Reform"
)
```

```
## \begin{table}
##
## \caption{\label{tab:unnamed-chunk-17}Interactive IV LATE estimates for the effect of staying
##   in school past age 15 on log real earnings: Northern Ireland, 1957 Reform}
## \centering
## \begin{tabular}[t]{lccl}
## \toprule
## Cross-fitting Folds & Learner & Coefficient & 95 Confidence Interval\\
## \midrule
## K = 5 & Random Forest & 0.279*** & {}[0.181,0.377]\\
##  & XGBoost & 0.241*** & {}[0.15,0.331]\\
##  & Elastic Net & 0.263*** & {}[0.177,0.35]\\
```

```
## K = 10 & Random Forest & 0.293*** & {}[0.19,0.396]\\
##  & XGBoost & 0.202*** & {}[0.11,0.295]\\
## \addlinespace
##  & Elastic Net & 0.216*** & {}[0.131,0.301]\\
## \bottomrule
## \multicolumn{4}{l}{\rule{0pt}{1em}\textit{Note: }}\\
## \multicolumn{4}{l}{\rule{0pt}{1em}makecell[l]{Covariates included are age, \
## \end{tabular}
## \end{table}
```

```
dml_results(
  list(
    IIVM_15_GB_RF_5_fold,
    IIVM_15_GB_XG_5_fold,
    IIVM_15_GB_Glm_5_fold,
    IIVM_15_GB_RF_10_fold,
    IIVM_15_GB_XG_10_fold,
    IIVM_15_GB_Glm_10_fold
  ),
  "Interactive IV LATE estimates for the effect of staying
  in school past age 15 on log real earnings: Great Britain, 1972 Reform"
)
```

```
## \begin{table}
##
## \caption{\label{tab:unnamed-chunk-17}Interactive IV LATE estimates for the effect of staying
##   in school past age 15 on log real earnings: Great Britain, 1972 Reform}
## \centering
## \begin{tabular}[t]{lccl}
## \toprule
## Cross-fitting Folds & Learner & Coefficient & 95 Confidence Interval\\
## \midrule
## K = 5 & Random Forest & 0.176*** & {}[0.134,0.219]\\
##  & XGBoost & 0.111*** & {}[0.054,0.168]\\
##  & Elastic Net & 0.174*** & {}[0.136,0.212]\\
## K = 10 & Random Forest & 0.179*** & {}[0.136,0.222]\\
##  & XGBoost & 0.105*** & {}[0.063,0.148]\\
## \addlinespace
##  & Elastic Net & 0.23*** & {}[0.192,0.268]\\
## \bottomrule
## \multicolumn{4}{l}{\rule{0pt}{1em}\textit{Note: }}\\
## \multicolumn{4}{l}{\rule{0pt}{1em}makecell[l]{Covariates included are age, \
## \end{tabular}
## \end{table}
```

```
dml_results(
  list(
    IIVM_15_NI_RF_5_fold,
    IIVM_15_NI_XG_5_fold,
    IIVM_15_NI_Glm_5_fold,
    IIVM_15_NI_RF_10_fold,
    IIVM_15_NI_XG_10_fold,
    IIVM_15_NI_Glm_10_fold
```

```
  ),
  "Interactive IV LATE estimates for the effect of staying
  in school past age 15 on log real earnings: Northern Ireland, 1972 Reform"
)
```

```
## \begin{table}
##
## \caption{\label{tab:unnamed-chunk-17}Interactive IV LATE estimates for the effect of staying
##   in school past age 15 on log real earnings: Northern Ireland, 1972 Reform}
## \centering
## \begin{tabular}[t]{lccl}
## \toprule
## Cross-fitting Folds & Learner & Coefficient & 95 Confidence Interval\\
## \midrule
## K = 5 & Random Forest & 0.252*** & {}[0.153,0.352]\\
##  & XGBoost & 0.274*** & {}[0.178,0.37]\\
##  & Elastic Net & 0.087 & {}[-0.001,0.175]\\
## K = 10 & Random Forest & 0.257*** & {}[0.154,0.36]\\
##  & XGBoost & 0.225*** & {}[0.147,0.303]\\
## \addlinespace
##  & Elastic Net & 0.113* & {}[0.025,0.2]\\
## \bottomrule
## \multicolumn{4}{l}{\rule{0pt}{1em}\textit{Note: }}\\
## \multicolumn{4}{l}{\rule{0pt}{1em}makecell[l]{Covariates included are age, \
## \end{tabular}
## \end{table}
```