Sample_Project

Tuffy Licciardi Issa

2025-10-25

This document is a minimal, reproducible sample of my empirical workflow
Full code and additional projects: **github.com/tuffyli**

Objectives

The main objective of this sample work is to explore the expansion of Brazil's alimony rights to women in stable unions and its effects on education, particularly among young women who became newly entitled to this right.
This analysis is inspired by Rangel's (2006) seminal work on the policy's expansion. The main difference in this study is the application of an additional Propensity Score Matching (PSM) method, with a particular focus on young mothers and their educational decisions.

Data

In the **repository**, I include supplementary code detailing the selection of key variables.
Due to the large size of the PNAD dataset, it is not possible to make the full data publicly available. Therefore, I provide a filtered version of the dataset containing only the variables relevant to the analysis.

```
# ------------------- #
# 1. Data Extraction ----

data <- readRDS("C:/Users/tuffy/Documents/Trabalhos/Ava_Pol/Bases/final_filtered_9295.rds")

#1.1 Summary ----
#' In the data manipulation code I created a summary table. Here I present it

summary <- readRDS("C:/Users/tuffy/Documents/Trabalhos/Ava_Pol/Bases/summary.rds")
print(summary)
```

```
##                              Mean        SD Min     Max
## Treatment                    0.45      0.50   0       1
## Female = 1                   1.00      0.00   1       1
## Race (White or Asian = 1)    0.55      0.50   0       1
## Highest Education            4.22      0.56   0       9
## Age                         21.04      2.41  15      24
## Years of education           6.61      3.37   1      17
## Last grade concluded         4.47      2.00   1       9
## Course enrollment            2.48      1.65   0       9
```

```
## Household per capita wage 173221.66 269756.48   0 9549387
## Pension (yes = 1)                0.00      0.05   0       1
## CBO Group                        4.34      2.58   0       8
## Male child (house)               0.57      0.72   0      12
## Female child (house)             0.55      0.71   0       6
## Total childs                     1.17      1.00   0      13
```

```
rm(summary)
```

Treated individuals are women aged 15–24 in non-civil unions (consensual or religious-only) who are household heads or spouses; controls are otherwise similar women in legally recognized unions (civil or civil + religious).

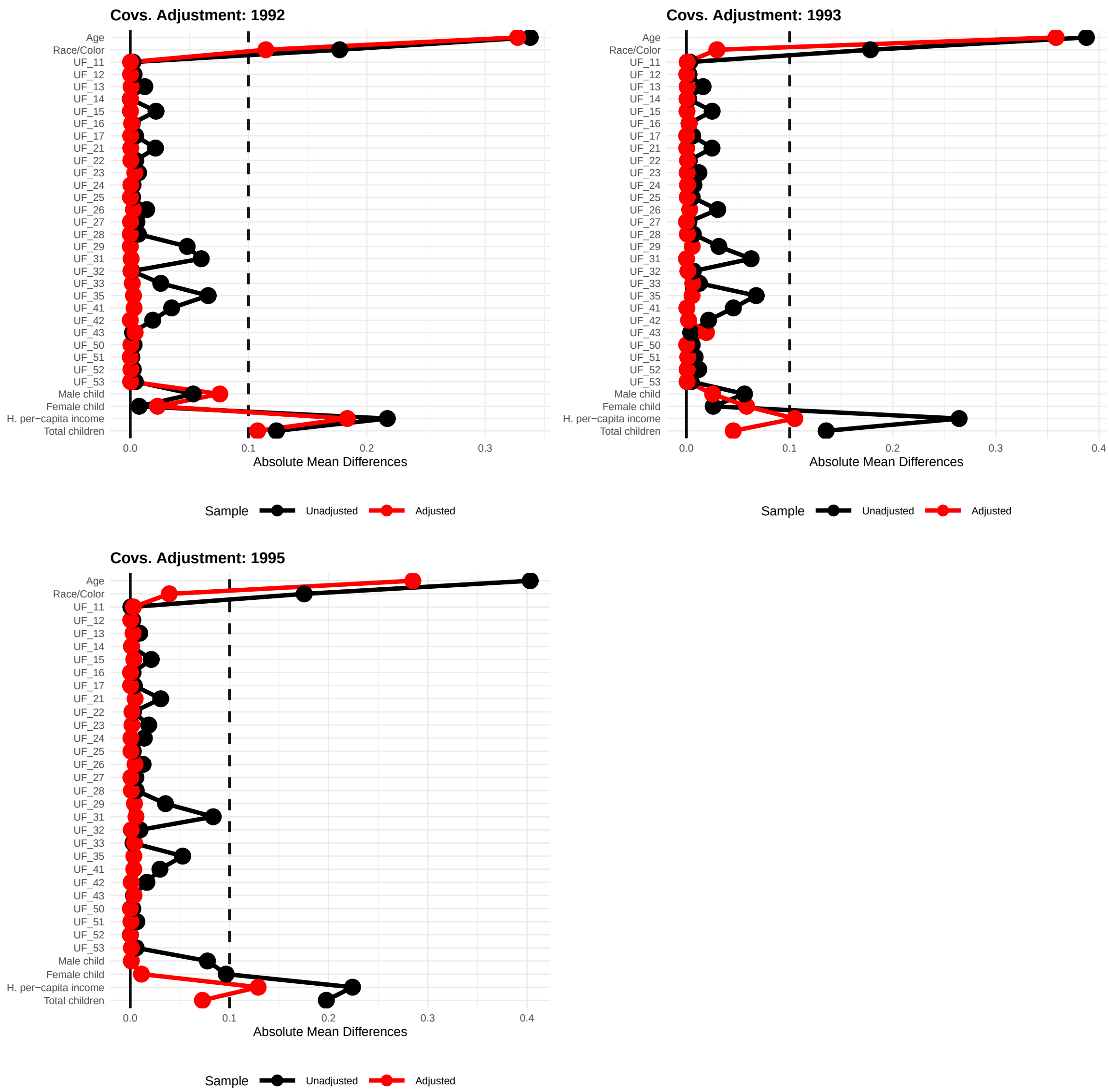
Propensity Score Matching

In this next step I will execute the propensity score matching between treated and control females.

```
# ------------------------------------------------------------------
# 2. PSM balance
# -----------------------------------------------------------------
plots <- list()

for (year in c(1992, 1993, 1995)) {
  df_psm <- data %>%
    filter(ano == year) %>%
    select(
      treatment, uf, age, cor, anos_estudo, peso_pessoa,
      ultima_serie_concluida, tipo_curso_frequenta, grupo_cbo,
      trabalhou_ultimo_ano, renda_dom_per_capita,
      dummy_filhos_homens_dom, dummy_filhos_mulheres_dom,
      pensao_dummy, total_filhos
    ) %>%
    mutate(uf = as.factor(uf)) %>%
    filter(
      !is.na(treatment), !is.na(uf), !is.na(cor), !is.na(peso_pessoa),
      !is.na(dummy_filhos_homens_dom), !is.na(dummy_filhos_mulheres_dom),
      !is.na(renda_dom_per_capita), !is.na(total_filhos)
    )

  match_model <- matchit(
    treatment ~ age + cor + uf + dummy_filhos_homens_dom +
      dummy_filhos_mulheres_dom + renda_dom_per_capita + total_filhos,
    data = df_psm,
    s.weights = df_psm$peso_pessoa,
    caliper = 0.05,
    method = "nearest",
    replace = TRUE
  )

  # Show the match summary counts in the document
  sum <- summary(match_model)
  latex_df <- as.data.frame(sum$nn)
```

```r
  latex_table <- kable(latex_df,
                       format = "latex",
                       booktabs = TRUE,
                       caption = "Estatísticas descritivas ponderadas",
                       align = "lccc")

  #The files are availabre in the repository
  writeLines(latex_table,
             paste0("C:/Users/tuffy/Documents/Trabalhos/", year,"_match_summary.tex"))


  # Love plot (INLINE)
  custom_labels <- c(
    distance_weighted         = "Distance (weighted)",
    uf                        = "UF",
    cor                       = "Race/Color",
    dummy_filhos_homens_dom   = "Male child",
    dummy_filhos_mulheres_dom = "Female child",
    age                       = "Age",
    renda_dom_per_capita      = "H. per-capita income",
    total_filhos              = "Total children"
  )

  love <- love.plot(
    match_model,
    stats         = "mean.diffs",
    abs           = TRUE,
    threshold     = 0.10,
    colors        = c("black", "red"),
    line          = TRUE,
    var.names     = custom_labels,
    drop.distance = TRUE
  ) +
    labs(
      title = paste0("Covs. Adjustment: ",year)
    ) +
   theme_minimal(base_size = 5) +
    theme(plot.title = element_text(face = "bold"),
                 legend.position = "bottom")
  #Grouping the plots
  plots[[as.character(year)]] <- love


  matched_data <- match.data(match_model) %>%
    mutate(ano = year)

  assign(paste0("df_matched_", year), matched_data)
  rm(year, custom_labels, sum, ess_table, latex_df, latex_table, love)

}

top    <- cowplot::plot_grid(plots[["1992"]], plots[["1993"]],
                              ncol = 2, align = "hv", rel_widths = c(1, 1))
```

```r
bottom <- cowplot::plot_grid(plots[["1995"]],
                             ncol = 2, align = "hv", rel_widths = c(1, 1))  # legend stays here

combined <- cowplot::plot_grid(top, bottom, ncol = 1, rel_heights = c(1, 1.05))
combined
```



Regression

To assess treatment effects, we fit progressively richer models (no FE, UF×year FE, and FE + controls) using the matched sample.

```r
# -------------------------------------------------------------------------- #
# 3. Regression -----
# -------------------------------------------------------------------------- #
```

```r
# Bind matched samples (1992, 1993, 1995) and prepare identifiers
df_psm <- bind_rows(df_matched_1992, df_matched_1993, df_matched_1995)

df_psm$ano    <- factor(df_psm$ano)          # ensure year is a factor
df_psm$ano    <- relevel(df_psm$ano, ref = "1992")    # reference year = 1992
df_psm$uf_ano <- interaction(df_psm$uf, df_psm$ano, sep = "_")  # UF × year FE key


# --- Model 1: no fixed effects (year-to-year comparison via i())
model_nc <- feols(
  anos_estudo ~ treatment + i(ano, treatment, ref = "1992"),
  data = df_psm, weights = df_psm$weights, vcov = "hetero"
)

# --- Model 2: add UF×year fixed effects (absorbs geography-by-year shocks)
model_fe <- feols(
  anos_estudo ~ treatment + i(ano, treatment, ref = "1992") | uf_ano,
  data = df_psm, weights = df_psm$weights, vcov = "hetero"
)

# --- Model 3: UF×year FE + basic controls
model_cc <- feols(
  anos_estudo ~ treatment + i(ano, treatment, ref = "1992") +
    total_filhos + cor + age | uf_ano,
  data = df_psm, weights = df_psm$weights, vcov = "hetero"
)
```

```r
fixest::etable(
  tex = T,
  list("No FE" = model_nc, "UF×Year FE" = model_fe, "FE + Controls" = model_cc),
  drop     = "Constant",
  dict = c(
    "anos_estudo"                    = "Years of education",
    "treatment$"                     = "Treatment (1992)",
    "i\\(ano, treatment.*\\)::1993"  = "Treatment × 1993",
    "i\\(ano, treatment.*\\)::1995"  = "Treatment × 1995",
    "age"                            = "Age",
    "cor"                            = "Race/Color",
    "total_filhos"                   = "Total children",
    "treatment x ano = 1993"         = "Treatment × 1993",
    "treatment x ano = 1995"         = "Treatment × 1995"
  ),
  se.below    = TRUE,
  signif.code = "letters",
  digits      = 3,
  fitstat     = c("n","r2","rmse"),
  title       = "ATT by Year (ref = 1992)",
  notes       = "SEs are heteroskedastic-robust. Weights from matching sample."
)
```

There is no statistically significant effect of alimony eligibility on the years of education among young women. However, the low R^2 values indicate limited explanatory power, these specifications explain only a small share of the variation in schooling outcomes, so the results should be interpreted with caution.

Tabela 1: ATT by Year (ref = 1992)

Dependent Variable:	Years of education		
Model:	(1)	(2)	(3)
Variables			
treatment	-1.77[a]	-1.66[a]	-1.25[a]
	(0.077)	(0.100)	(0.095)
treatment × ano = 1993	0.307[a]	0.230	0.112
	(0.080)	(0.146)	(0.136)
treatment × ano = 1995	0.452[a]	0.176	0.055
	(0.077)	(0.147)	(0.137)
Total children			-0.915[a]
			(0.028)
Race/Color			0.803[a]
			(0.057)
Age			0.260[a]
			(0.011)
Fixed-effects			
uf_ano		Yes	Yes
Fit statistics			
Observations	17,779	17,779	17,779
R^2	0.05238	0.11600	0.21156
RMSE	66.252	63.990	60.432

Heteroskedasticity-robust standard-errors in parentheses
Signif. Codes: a: 0.01, b: 0.05, c: 0.1

SEs are heteroskedastic-robust. Weights from matching sample.