MY457: Reappraisal of Tellez, Juan, 2021, Land, Opportunism, and Displacement in Civil Wars: Evidence from Colombia

Candidate 42433

mié/28/may.

Summary of the original paper

The paper authored by Téllez (2021) investigates a critical yet underexplored dimension of civil war dynamics: the role of opportunism in driving forced displacement. Téllez argues that, in a context of weak property rights and following an economic shock in this case, the rise in the value of African palm oil driven by international market prices an ideal scenario emerged in which rural elites colluded with armed groups to forcibly displace civilians and seize their land. He tests this theory using data from Colombia between 1993 and 2005, a period marked by both the rapid expansion of the palm oil industry and intense land-related conflict.

Research Design

Given the scope of the study, the research question can be formulated as: Did the expansion of the African palm oil industry in Colombia between 1993 and 2005 have an effect on rural displacement?

The main empirical approach is a difference in differences design leveraging variation in the timing of palm oil adoption across municipalities. The design includes municipality and year fixed effects to account for time invariant local characteristics and national shocks, and controls for time varying confounders such as guerrilla violence and coca cultivation.

$$\text{dispRate}_{it} = \alpha_i + \omega_t + \beta \cdot \text{palmOil}_{it} + \boldsymbol{\phi} \cdot \mathbf{X}_{it} + \varepsilon_{it} \tag{1}$$

Where:

- dispRate_{it} is the displacement rate in municipality i at time t.
- α_i are municipality fixed effects controlling for time-invariant characteristics.
- ω_t are year fixed effects capturing common shocks across all municipalities.
- palmOil_{it} is a treatment variable indicating whether municipality i produces palm oil at time t.
- $\mathbf{X}_{it}$ is a vector of time-varying covariates, specifically the number of guerrilla (FARC) attacks and the presence of coca cultivation.
- $\boldsymbol{\phi}$ are the corresponding coefficients for the control variables.
- ε_{it} is the error term.

The identification assumption is that, conditional on these controls, the timing of palm expansion is not systematically correlated with unobserved shocks to displacement i.e., the parallel trends assumption. This is test by the author using Sun and Abraham (2020) package, that can be found in the appendix of the paper.

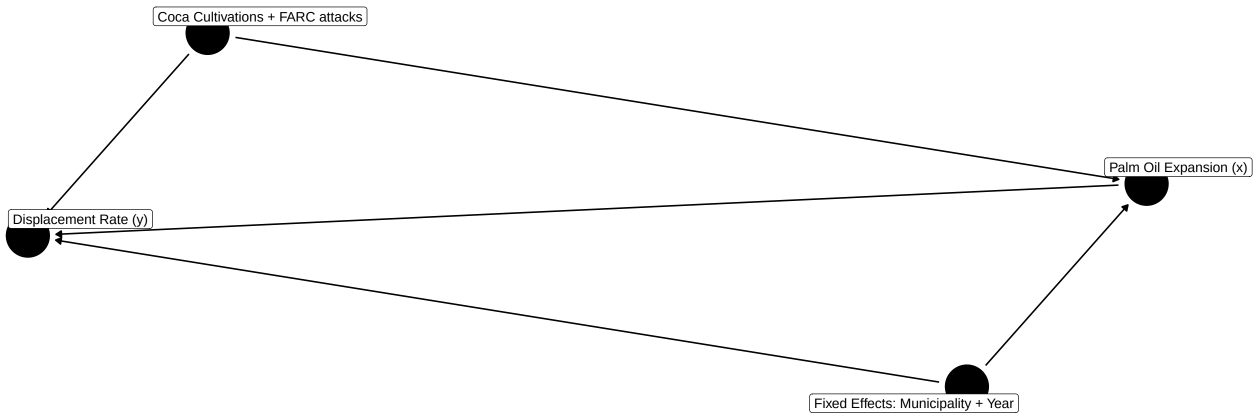


Figure 1: Causal Dag Design for Tellez 2021

The study draws on a combination of administrative panel data and an original household survey. The municipal level data include annual displacement rates from the Colombian Victim's Unit and palm oil cultivation data from FEDEPALMA, covering nearly all municipalities in the country. At the household level, survey responses provide information on displacement experiences and motivations, enabling a more granular analysis of mechanisms.

The results of the main approach show that the adoption of palm oil cultivation is significantly associated with an increase in displacement, particularly in municipalities with a paramilitary presence. This pattern is not observed in guerrilla controlled areas, reinforcing the idea that the displacement in palm growing regions was opportunistic rather than ideologically motivated. More specifically, the mechanism appears to be collusion between rural elites and armed groups namely, the paramilitaries who used violence strategically to appropriate land.

The complementary analysis at the municipal level data, where Téllez incorporates original household survey data to explore the motivations and consequences of displacement at the individual level. He finds that respondents from palm producing municipalities were more likely to report being forcibly displaced by elites or paramilitaries, often citing land seizure as the reason. These households also tended to own or work on larger plots of land, supporting the argument that economic accumulation rather than military strategy was the driving force behind the violence. Furthermore, municipalities with palm oil production today show significantly higher levels of land restitution claims, suggesting long term consequences of wartime land grabs. Overall, the study highlights the need to go beyond strategic accounts of displacement and consider how economic incentives and elite armed group alliances shape the trajectory of civil war violence and post conflict justice.

Replication

I chose to replicate the following table and figures because I considered them to be essential components of the paper. The replication of the main model presented in Equation (1) employs a two-way fixed effects specification within a difference-in-differences (DiD) framework.The table include interactions between the plantations and national production of palm-oil and the presence of armed groups.

I also replicated the key figures from the paper. Figure 1 (right) shows the displacement rate per 1,000 residents in Colombian municipalities, illustrating the divergence in displacement trends between palm and non-palm areas over time. The Event Study (Figure 1, left), included in the appendix, is replicated as well to assess the parallel trends assumption.

Additionally, I replicated Figure 2, which demonstrates both the intensity effect of palm cultivation and the conditional effect depending on the presence of armed groups. Finally, I reproduced Figure 3, which highlights how displacement dynamics vary across different scenarios, with significantly higher probabilities of displacement in municipalities engaged in palm production.

	Model 1	Model 2	Model 3	Model 4
Palm-oil plantation	0.382**	−2.056**	−1.213***	0.692
	(0.121)	(0.642)	(0.023)	(0.548)
FARC Attack	0.148***	0.148***	0.148***	
	(0.013)	(0.013)	(0.013)	
Coca plantation	0.541***	0.539***	0.541***	0.580***
	(0.056)	(0.056)	(0.056)	(0.058)
Palm x National Prod		0.248***		
		(0.063)		
Palm x AUC Presence			1.614***	
			(0.120)	
Palm x FARC Presence				−0.327
				(0.559)
Num. obs.	14192	14192	14192	14208
Num. groups: cod_dane	1118	1118	1118	1118
Num. groups: year	13	13	13	13

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 1: Effect of Palm-Oil Growth on Displacement

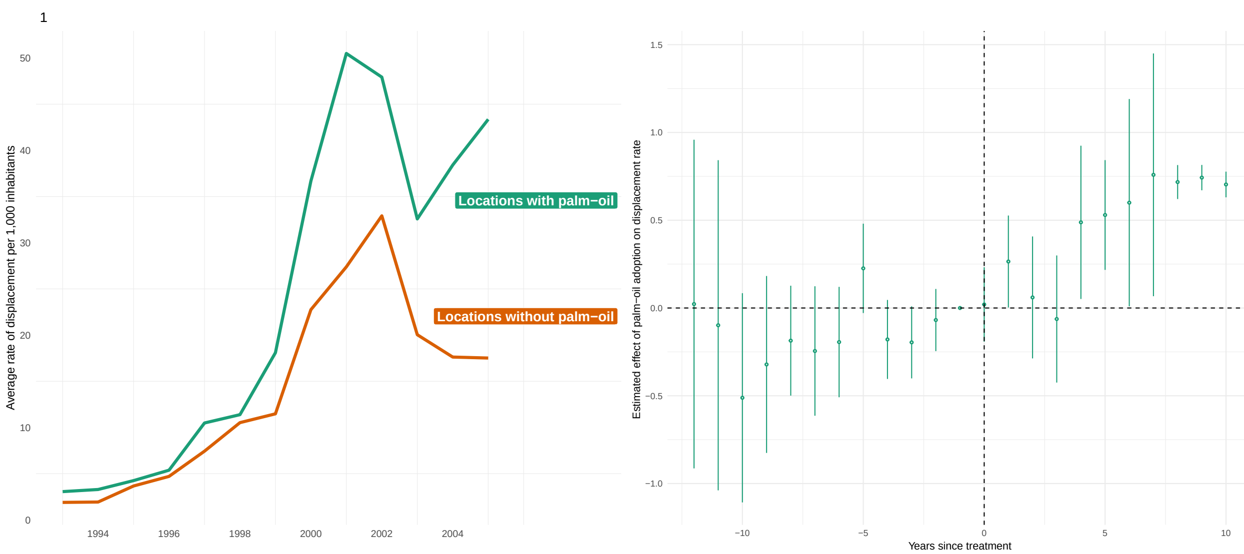


Figure 2: Rate of displacement per 1000 residents of Colombia Municipalities with and without Palm-Oil Cultivations and Event Study

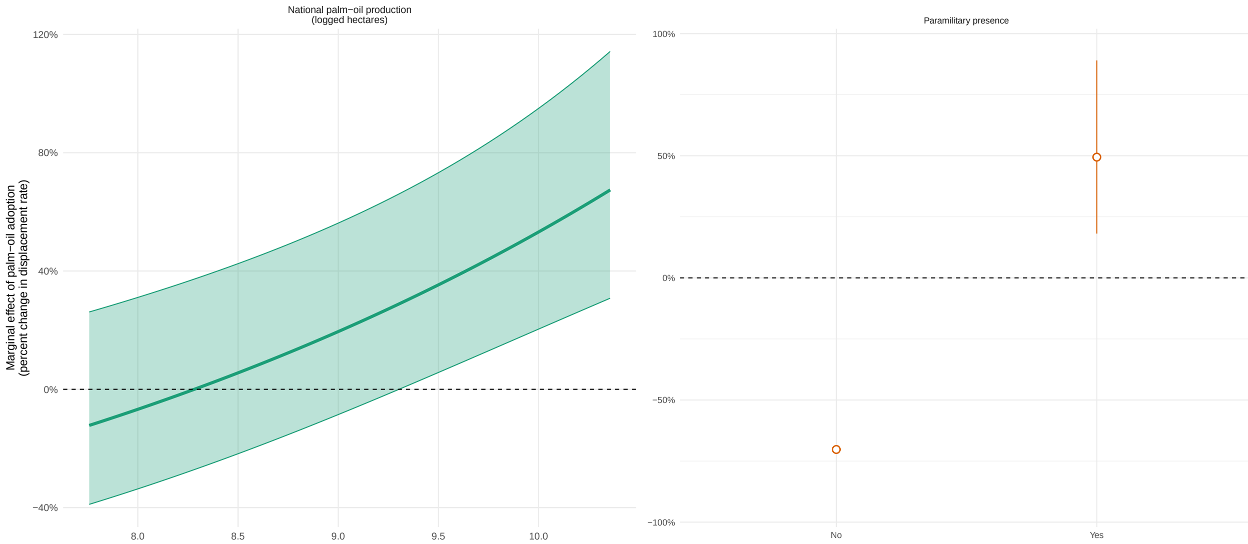


Figure 3: Effect of Palm Oil presence on displacement rate across National Oil Poduction and Presence / Absence of Paramilary Groups

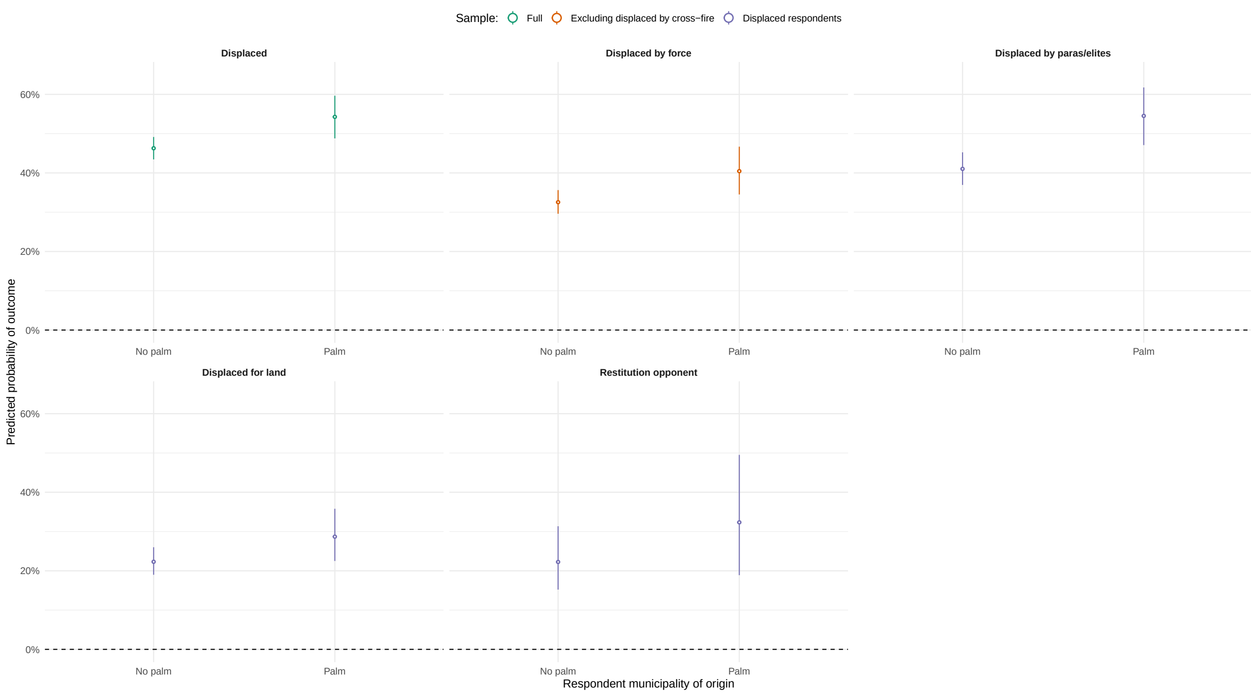


Figure 4: predicted probabilities from Logit Models using household Survey Data

Reappraisal

Critical Comments and Modifications

Use and adaptation of code to modern package

When replicating the results from the main model Table1, the original code using the `felm` function was replaced with `feols` because `feols` is a more recent and computationally efficient estimator for fixed effects models. It offers faster estimation, robust standard error options, and better support for high dimensional fixed effects and flexible inference procedures, making it well suited for modern empirical work .However, marginal effects visualizations retained the original `felm` framework since `predict_partial` does not support `feols` objects, ensuring exact replication of the figures.

Similarly, the graphs were adapted to the required format, excluding the Roboto font, which failed to compile in LaTeX.

Main model covariates

When replicating the main results of the core model, the author includes covariates such as palm oil production and left-wing guerrilla activity (FARC attacks). While the results section highlights the importance of AUC (paramilitary) presence, I argue that a direct measure, such as the number of AUC attacks, should be incorporated into the main specification, given the central role of paramilitary actors in forced displacement.

A potential proxy could be the number of deaths caused by paramilitary groups across regions, as reported by the Truth Commission of Colombia, though this might have its own limitations discuss further on the limitation section.

Colombia conjuncture explanation

Moreover, while the paper mentions the inclusion of guerrilla attacks and coca cultivation as potential confounders in the main model,the inclusion of these variables requires a deeper understanding of Colombia's armed conflict dynamics and territorial control.

It is important to acknowledge that some studies highlight the limitations of using illicit crop cultivation as a proxy for armed group presence. Arias and Ibáñez (2019) argue that such proxies may introduce bias if the location and extent of cultivation respond endogenously to conflict dynamics rather than purely reflecting armed group control . This potential endogeneity could affect the interpretation of the estimated effects if coca cultivation changes as a reaction to displacement or violence, rather than being a fixed indicator of armed presence.

SUTVA and No Anticipation discussion

Although the paper discusses and tests the parallel trends assumption, and later adds robustness checks, there is no discussion of SUTVA (Stable Unit Treatment Value Assumption) or no anticipation. These assumptions may plausibly be violated in this context.

For instance, it is plausible that municipalities neighboring those affected by forced displacement may have also experienced displacement due to fear of similar events. This hypothetical yet realistic scenario implies the presence of spillovers, which would violate SUTVA, and may also indicate a violation of the no anticipation assumption.

Reverse Causality

A deeper understanding of the armed conflict also opens the door to questioning potential reverse causality. That is, the causal relationship might run in the opposite direction: displacement could have preceded and enabled palm oil cultivation, rather than palm oil expansion being the driver of displacement.

The Colombian conflict caused widespread displacement for reasons unrelated to palm oil such as territorial control by armed groups, violence between guerrillas and paramilitaries, or general insecurity. This displacement left large tracts of land abandoned or under utilized.

The reverse causality theory could be supported by several observations from the paper itself:

- The timing of palm oil adoption and displacement could be interpreted differently , perhaps the first wave of displacements occurred for conflict related reasons, and palm oil expanded into these areas after initial displacement.

- The correlation between palm oil municipalities and land restitution claims could reflect that palm oil companies targeted areas with the weakest property rights and highest prior displacement rates, not that they caused the initial displacement.
- The association between paramilitary presence and palm oil displacement could be that paramilitaries first displaced people for strategic military reasons, then facilitated land sales to palm oil interests as a secondary benefit.

Alternative Estimators

To generate robustness checks, I re estimated the main model with modern estimators being this **Callaway & Sant'Anna Estimator (2021)** , **Sun & Abraham Estimator (2020) and FECT, IFE MC from the Liu, Wang, and Xu (2022)**

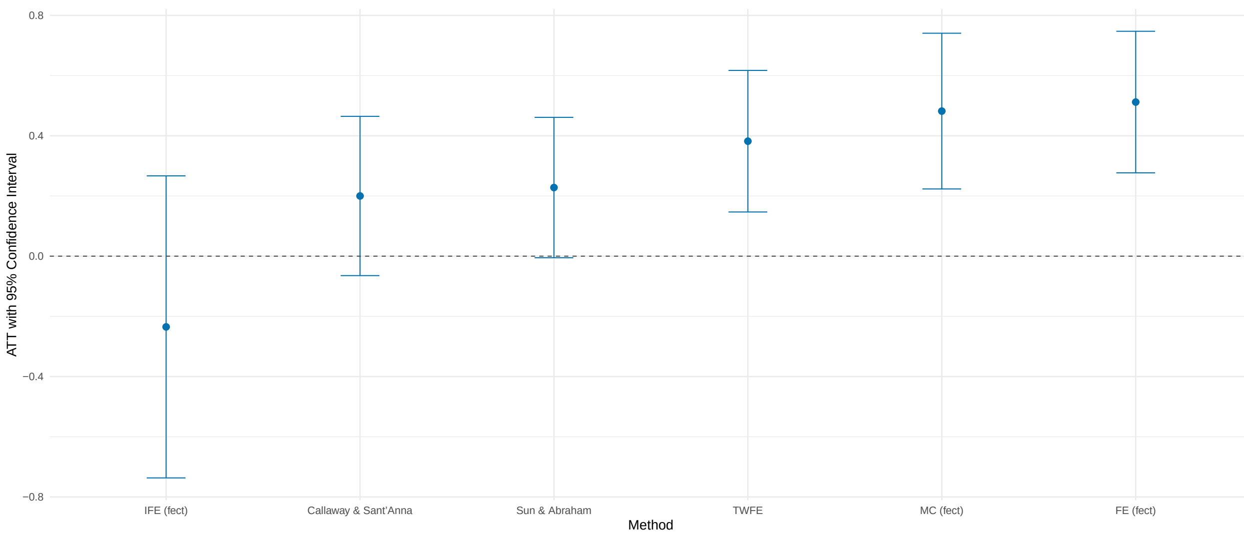


Figure 5: Comparison of ATT Estimates by Method

	IFE	Callaway & Sant'Anna	Sun & Abraham	TWFE	MC	FE
ATT	-0.235	0.200	0.228	0.382***	0.482***	0.512***
	(0.256)	(0.135)	(0.119)	(0.120)	(0.132)	(0.120)

Among the six estimators, only TWFE, FE, and Matching produce statistically significant and positive ATT estimates, while Callaway & Sant'Anna, Sun & Abraham, and IFE do not yield significant results. This discrepancy likely reflects the fact that conventional models like TWFE and FE, which assume homogeneous treatment effects and parallel trends, may overestimate the impact due to violations of their identifying assumptions in the presence of staggered adoption. In contrast, the more robust estimators Callaway & Sant'Anna and Sun & Abraham explicitly account for treatment effect heterogeneity and avoid biases from improper weighting, often resulting in more conservative and statistically non significant estimates.

Additional Parallel Trend Test

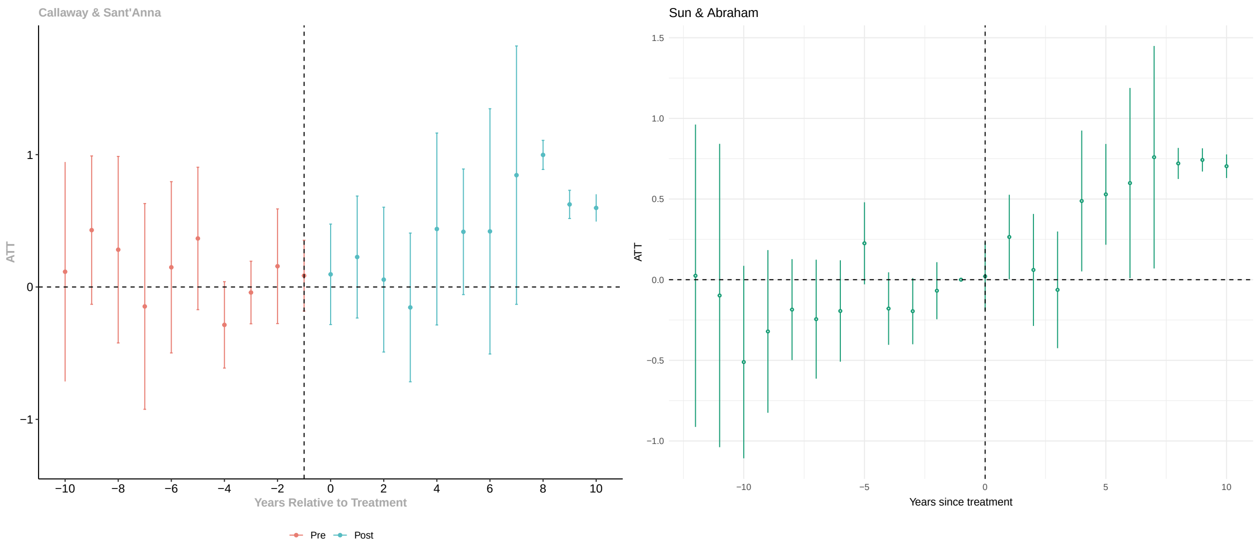


Figure 6: Event Study Callawa Sat'Anna - Sun and Abraham

The Callaway and Sant'Anna estimator shows relatively flat pre-treatment coefficients, with confidence intervals overlapping zero, suggesting the parallel trends assumption is not violated. After treatment, the estimated effect increase moderately and become statistically distinguishable from zero in later periods, indicating a delayed but positive effect of the treatment on displacement.

In the Sun and Abraham estimator, the pre-treatment coefficients also hover around zero, further supporting no strong evidence of pre-trends. Post-treatment, the estimated effects exhibit a clear upward trajectory, with increasing ATT values and tighter confidence intervals, suggesting a consistent and growing treatment effect over time.

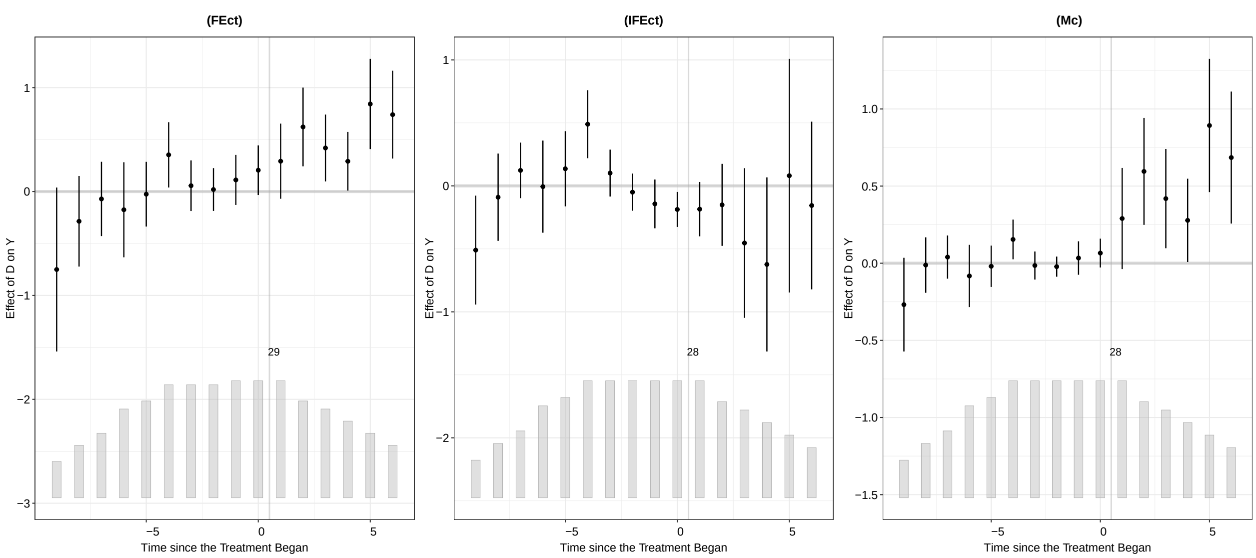


Figure 7: Event Study FEct, IFEct, MC

The FEct (Fixed Effects counterfactual trends) estimator shows relatively volatile effects before the treatment, but most estimates are statistically indistinguishable from zero. Post-treatment, the ATT estimates rise steadily, with clear statistical significance, pointing to a robust and persistent effect of the treatment.

In the IFEt (Interactive Fixed Effects) model, although some pre-treatment variation is observed, the coefficients mostly remain within confidence bounds around zero. Following treatment, the estimates show a sharp increase in the ATT, indicating a strong and significant response to treatment, especially in the immediate years after the intervention.

Lastly, the Matrix Completion (MC) method yields relatively stable pre-treatment estimates, suggesting well balanced counterfactuals. After treatment, the effect on the outcome becomes increasingly positive and statistically significant, consistent with other estimators in confirming the positive impact of the treatment on the outcome of interest.

Placebo Test

	Placebo Model -3	Placebo Model -10
Palm-oil placebo (-3)	0.249	
	(0.141)	
Palm-oil placebo (-10)		0.442
		(0.427)
FARC Attack	0.148^{***}	0.148^{***}
	(0.013)	(0.013)
Coca plantation	0.542^{***}	0.546^{***}
	(0.056)	(0.056)
Num. obs.	14192	14192
Num. groups: cod_dane	1118	1118
Num. groups: year	13	13

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3: Placebo Test

The placebo test, conducted by shifting the treatment to leads at -2 and -10 periods, did not yield statistically significant results. This finding supports the validity of the main estimate, indicating that the observed effect is unlikely to be driven by pre-existing trends or anticipatory behavior.

New Data Incorporation

This dataset was provided by the Center for Economic Development Studies (CEDE) at the Faculty of Economics, Universidad de los Andes, Colombia, under strict confidentiality conditions.

To deepen the analysis of forced displacement and palm oil cultivation, the study period is extended to cover 2007–2015. This time frame was chosen because it is relatively isolated from other major confounding factors in Colombia, such as the start of the peace negotiations in 2015.

Additionally, this extended period allows the inclusion of additional variables that may help explain the phenomenon, such as institutional effectiveness as a proxy of strong municipal institutions.The Specific definitions of these variables, provided by CEDE, are detailed in the appendix.

$$\text{dispRate}_{it} = \alpha_i + \omega_t + \beta \cdot \text{palmOil}_{it} + \boldsymbol{\phi} \cdot \mathbf{X}_{it} + \gamma \cdot \text{Institutionality}_{it} + \varepsilon_{it} \tag{2}$$

The guiding hypothesis of this analysis is that forced displacement occurred primarily in municipalities with weaker institutional frameworks, drawing on the work of Acemoglu, Johnson, and Robinson (2005)in contexts where institutions are weak or extractive, governance mechanisms fail to provide security, enforce contracts, or regulate conflicts effectively, thereby creating fertile ground for violence, displacement, and economic exclusion.

The regression results suggest that the presence of palm cultivation alone doesn´t have a statistically significant effect on the displacement rate in the mentioned years (2007-2015). However,the negative and significant

	Model 113
Palm Oil plantation	0.443
	(0.285)
Plantation x Institutional Quality Index	-0.011^{***}
	(0.003)
Num. obs.	3912
Num. groups: codmpio	1065
Num. groups: ano	4
R^2 (full model)	0.91
Adj. R^2 (full model)	0.87

Note: Model includes municipal and year fixed effects and controls for presence of coca and FARC attacks. Standard errors clustered at the municipality level.
$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.1$

Table 4: Regression Results for Model with Institutions Index

interaction between palm cultivation and the Institutional Quality Index suggests that in municipalities with palm cultivation, stronger administrative capacity mitigates the displacement effect, reflecting that institutional quality can reduce the negative impact of palm cultivation related displacement.

Limitations

- One important limitation while mentioned in the paper could be further explored: the absence of monthly level data. Given Colombia's climate, African palm can be cultivated year round. Monthly data could help assess more precisely whether displacement occurred before or after the planting of palm oil, strengthening or challenging the case for reverse causality.
- Displacement data quality is also a limitation, as it is a highly sensitive issue. According to the Truth Commission, in 67% of displacement cases, the perpetrator is either not reported or not identified. This significantly limits the ability to attribute displacement events to specific armed actors, thus weakening the interpretation of the conflict's causal mechanisms.
- As mentioned earlier, the number of paramilitary related deaths per municipality could serve as a proxy for paramilitary attacks. However, the lack of direct attack data for paramilitary groups unlike the detailed data available for FARC remains a limitation of the model.

References

-Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." American Economic Review 91 (5): 1369–1401. https://doi.org/10.1257/aer.91.5.1369.

-Arias, M.A., Ibáñez, A.M. and Zambrano, A., 2019. Agricultural production amid conflict: Separating the effects of conflict into shocks and uncertainty. World Development, 119, pp.165–184.

-Tellez, Juan Fernando. 2021. "Land, Opportunism, and Displacement in Civil Wars: Evidence from Colombia." American Political Science Review, 1–16. https://doi.org/10.1017/S0003055421001003.

-De Hauteclocque, Guillaume. 2023. fixest: Fast Fixed-Effects Estimations. R package version 0.12.0. https://cran.r-project.org/package=fixest.

-Ibáñez, Ana María, and Carlos Eduardo Vélez. 2008. "Civil Conflict and Forced Migration: The Micro Determinants and Welfare Losses of Displacement in Colombia." World Development 36 (4): 659–676. https://doi.org/10.1016/j.worlddev.2007.04.013.

-Dell, Melissa. 2015. "Trafficking Networks and the Mexican Drug War." American Economic Review 105 (6): 1738–1779.

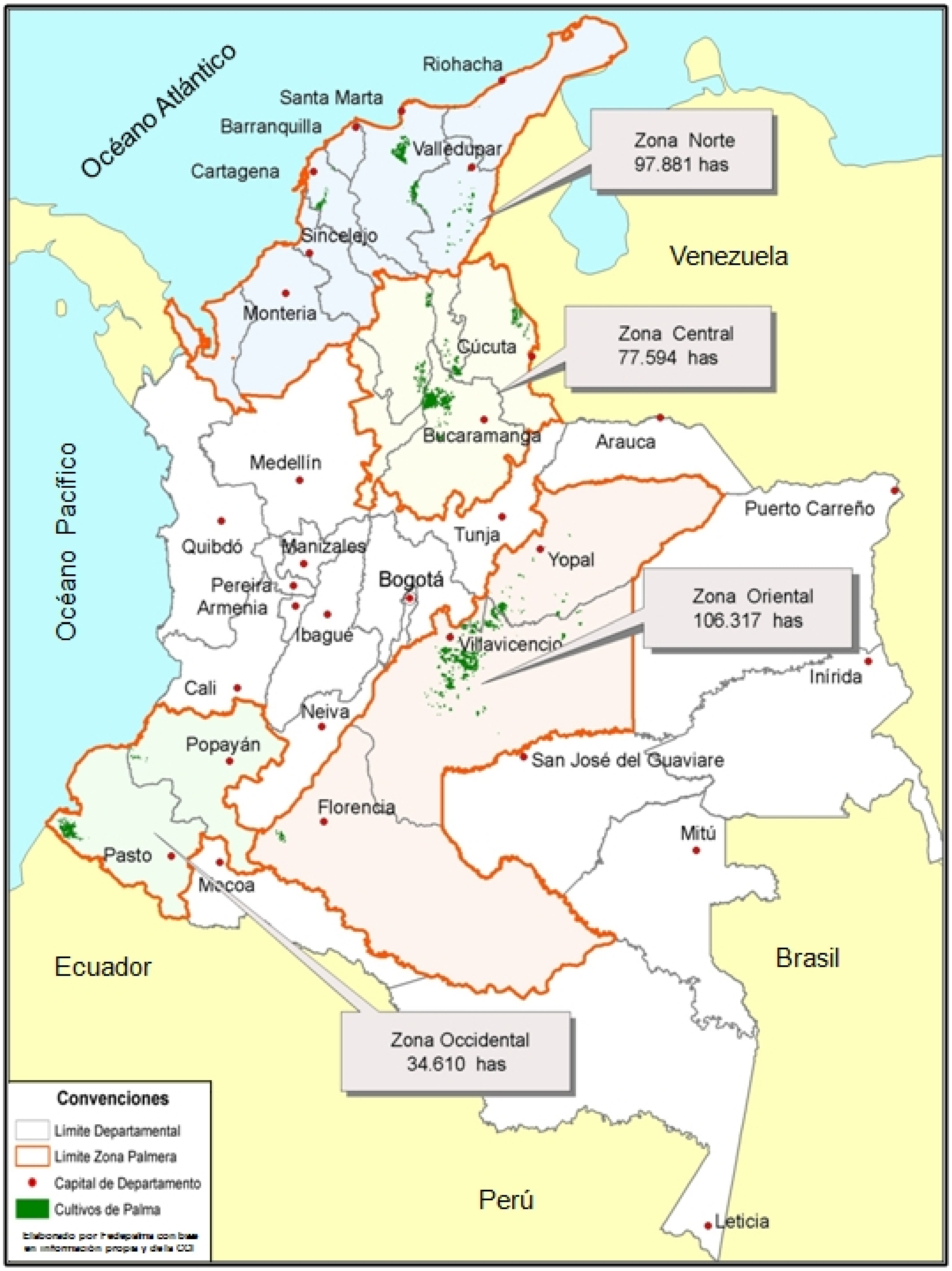
-González, Fernán. 2014. Poder y violencia en Colombia. Bogotá: CINEP.

Appendix

Institutional Quality Index Measurement

The index evaluates the management quality of territorial entities based on their effectiveness, efficiency, legal compliance, and administrative capacity. Effectiveness measures how well municipalities meet planned goals, efficiency assesses optimal use of resources in key sectors, legal compliance ensures adherence to relevant laws governing resource management, and administrative capacity reflects the availability and quality of human and technological resources supporting organizational functions. This index ranges from 0 (poor performance) to 100 (excellent performance), indicating the overall institutional quality of the municipality.

Palm Oil Cultivation Zones in Colombia



Code appendix

Please make sure that the vast majority of your code appears here and not in-line. However, if you are using functions that you have defined in a separate script, you can simply include the source call in the appendix, and not the function definition itself.

```
# this chunk contains code that sets global options for the entire .Rmd.
# we use include=FALSE to suppress it from the top of the document, but it will still appear in the app

knitr::opts_chunk$set(echo=FALSE, warning=FALSE, message=FALSE, linewidth=60)

# you can include your libraries here:
library(tidyverse)
library(dagitty)
library(ggdag)
library(dagitty)
library(ggplot2)
library(dplyr)
library(ggrepel)
library(hrbrthemes)
library(lfe)
library(interplot)
library(stargazer)
library(patchwork)
library(broom)
library(here)
library(paletteer)
library(texreg)
library(hrbrthemes)
library(patchwork)
library(scales)
library(extrafont)
library(ggeffects)
library(did)
library(staggered)
library(haven)
library(fixest)
library(gsynth)
library(fect)
library(cowplot)
library(grid)
library(gridGraphics)

# and any other options in R:
options(scipen=999)

setwd("C:/Users/Acer/Desktop/Maestria/LSE CLASES/MY457-Causal Inference/Summative/summative-reappraisal-


dag_minimal <- dagitty("dag {
  x -> y
  . -> x
  . -> y
  .. -> x
  .. -> y
```

```r
}")

# Add readable labels
tidy_dag <- tidy_dagitty(dag_minimal) %>%
  mutate(label = case_when(
    name == "x" ~ "Palm Oil Expansion (x)",
    name == "y" ~ "Displacement Rate (y)",
    name == "." ~ "Coca Cultivations + FARC attacks",
    name == ".." ~ "Fixed Effects: Municipality + Year"
  ))

# Plot the DAG in minimalist style (left-to-right)
ggdag(tidy_dag, text = FALSE, use_labels = "label", layout = "lr") +
  geom_dag_text(size = 3.8, color = "black") +
  theme_void() +
  labs(title = " ") +
  theme(
    plot.title = element_text(hjust = 0.5, size = 14, face = "bold")
  )
muni <- readRDS("C:/Users/Acer/Desktop/Maestria/LSE CLASES/MY457-Causal Inference/Summative/summative-r
view(muni)


#Change the package felm for the modern feols pacakage.

# Model 1
m1 <- feols(sdisp_rate ~ palm_presence + farc_attack + coca | cod_dane + year,
            data = muni,
            vcov = ~cod_dane)


# Model 2: interaction with logged national palm production
m2 <- feols(sdisp_rate ~ palm_presence * lnatprod + farc_attack + coca | cod_dane + year,
            data = muni,
            vcov = ~cod_dane)

# Model 3: interaction with AUC paramilitary presence
m3 <- feols(sdisp_rate ~ palm_presence * auc_dummy + farc_attack + coca | cod_dane + year,
            data = muni,
            vcov = ~cod_dane)
#Model 4 : interaction with Plantation and Farc Presence
m4 <- feols(sdisp_rate ~ palm_presence * farc_dummy + coca
  | cod_dane + year,         # municipality and year fixed effects
  data    = muni,
  cluster = ~cod_dane        # cluster-robust SE by municipality
)

texreg(
  l = list(m1, m2, m3, m4),
  custom.model.names = c("Model 1", "Model 2", "Model 3", "Model 4"),
  custom.coef.names = c(
    "Palm-oil plantation",       # palm_presence
    "FARC Attack",               # farc_attack
```

```r
    "Coca plantation",           # coca
    "Palm x National Prod",      # palm_presence:lnatprod
    "Palm x AUC Presence",       # palm_presence:auc_dummy
    "Palm x FARC Presence"       # palm_presence:farc_dummy
  ),
  digits = 3,
  include.nobs = TRUE,
  include.rsquared = FALSE,
  include.adjrs = FALSE,
  include.aic = FALSE,
  include.bic = FALSE
)
# read data
df<- readRDS("C:/Users/Acer/Desktop/Maestria/LSE CLASES/MY457-Causal Inference/Summative/summative-reap
df <- df %>%
  mutate(year = unclass(year))


# inv_sine function
inv_sine = function(x)
{
  log(x + sqrt((x^2 + 1)))
}

# broken down by palm presence
pDat = df %>%
  dplyr::select(year, palm_presence, disp_rate) %>%
  group_by(year, palm_presence) %>%
  summarise_all(~mean(.,na.rm = T)) %>%
  drop_na() %>%
  mutate(palm_presence = ifelse(palm_presence == 1,
                                "Locations with palm-oil",
                                "Locations without palm-oil"))

# for adding labels
labels =
  pDat %>%
  filter(year == 2003)


# marginal effects
## code from Patrick Baylis to deal with felm:
#https://www.patrickbaylis.com/blog/2021-01-22-predict-partial/
predict_partial <- function(object, newdata, se.fit = FALSE,
                            interval = "none",
                            level = 0.95){
  if(missing(newdata)) {
    stop("predict_partial requires newdata and predicts for all group effects = 0.")
  }

  newdata <- newdata; se.fit = T; interval = "confidence"; level = 0.95

  # Extract terms object, removing response variable
```

```r
  tt <- terms(object)
  Terms <- delete.response(tt)

  # Remove intercept
  attr(Terms, "intercept") <- 0

  X <- model.matrix(Terms, data = newdata)

  if (class(object) == "fixest") {
    B <- as.numeric(coef(object))
    df <- attributes(vcov(fit_feols, attr = T))$dof.K
  } else if (class(object) %in% c("lm", "felm")) {
    B <- as.numeric(object$coef)
    df <- object$df.residual
  } else {
    stop("class(object) should be lm, fixest, or felm.")
  }

  fit <- data.frame(fit = as.vector(X %*% B))

  if(se.fit | interval != "none") {
    sig <- vcov(object)
    se <- apply(X, MARGIN = 1, FUN = get_se, sig = sig)
  }

  if(interval == "confidence"){
    t_val <- qt((1 - level) / 2 + level, df = df)
    fit$lwr <- fit$fit - t_val * se
    fit$upr <- fit$fit + t_val * se
  } else if (interval == "prediction"){
    stop("interval = \"prediction\" not yet implemented")
  }
  if(se.fit){
    return(list(fit=fit, se.fit = se))
  } else {
    return(fit)
  }
}

get_se <- function(r, sig) {
  # linear combination, helper function for predict_partial
  # Given numeric vector r (the constants) and vcov sig (the ), compute SE
  r <- matrix(r, nrow = 1)
  sqrt(r %*% sig %*% t(r))
}


# new data for lnatprod
newdata = tibble(palm_presence = 1,
                 farc_attack = 0,
                 coca = 0,
                 lnatprod = seq(min(df$lnatprod, na.rm = TRUE),
                                max(df$lnatprod, na.rm = TRUE),
```

```r
                                by = .05))

## marginal effects

m2 = felm(sdisp_rate ~  palm_presence*lnatprod + farc_attack + coca - lnatprod
          | cod_dane + year | 0 | cod_dane, exactDOF = 'rM',
          data = df)


m5 = felm(sdisp_rate ~ palm_presence*auc_dummy + farc_attack + coca - auc_dummy
          | cod_dane + year | 0 | cod_dane, exactDOF = 'rM',
          data = df)

# get marginal effects
preds = predict_partial(m2, newdata = newdata,
                        se.fit = TRUE)

preds = bind_cols(newdata, preds) %>%
  # exponentiate to approximate change
  mutate(effect = (exp(fit) - 1),
         low = (exp(fit - 1.96*se.fit) - 1),
         hi = (exp(fit + 1.96*se.fit) - 1)) %>%
  mutate(group = "National palm-oil production\n(logged hectares)")


pt1 <- ggplot(preds, aes(x = lnatprod, y = effect, ymin = low, ymax = hi)) +
  geom_ribbon(alpha = .3, color = "#1b9e77", fill = "#1b9e77") +
  geom_line(size = 1.5, color = "#1b9e77") +
  labs(x = NULL,
       y = "Marginal effect of palm-oil adoption\n(percent change in displacement rate)") +
  theme_minimal(base_size = 12) +
  scale_y_continuous(labels = scales::percent_format()) +
  facet_wrap(vars(group)) +
  theme(panel.grid.minor = element_blank(),
        text = element_text()) +
  geom_hline(yintercept = 0, lty = 2)

# new data for auc
newdata = tibble(palm_presence = 1,
                 farc_attack = 0,
                 coca = 0,
                 auc_dummy = c(0, 1))

# get marginal effects
preds = predict_partial(m5, newdata = newdata,
                        se.fit = TRUE)

preds = bind_cols(newdata, preds) %>%
  # exponentiate to approximate change
  mutate(effect = (exp(fit) - 1),
         low = (exp(fit - 1.96*se.fit) - 1),
         hi = (exp(fit + 1.96*se.fit) - 1)) %>%
  mutate(group = "Paramilitary presence")
```

```r
pt2 = ggplot(preds, aes(x = factor(auc_dummy), y = effect,
                        ymin = low,
                        ymax = hi)) +
  geom_pointrange(fatten = 1, size = 3, shape = 21, fill = "white",
                  position = position_dodge(width = .2),
                  color = "#d95f02") +
  labs(x = NULL,
       y = NULL) +
  theme_minimal() +
  scale_y_percent(limits = c(-1, 1)) +
  scale_x_discrete(breaks = c(0, 1),
                   labels = c("No", "Yes")) +
  facet_wrap(vars(group)) +
  geom_hline(yintercept = 0, lty = 2)

ptt2<-(pt1+pt2)

figure1 <- ggplot(filter(pDat, year < 2006),
                  aes(x = year, y = disp_rate, color = palm_presence)) +
  geom_line(size = 1.5) +
  theme_minimal(base_size = 12) +  # base_size size
  labs(x = '',
       y = 'Average rate of displacement per 1,000 inhabitants',
       color = 'Palm oil cultivations:') +
  theme(legend.position = 'none',
        panel.grid.major = element_blank(),
        text = element_text()) +  # no family
  geom_label_repel(data = labels,
                   aes(x = year, y = disp_rate,
                       label = palm_presence,
                       fill = palm_presence),
                   fontface = "bold",
                   color = "white",
                   nudge_y = 2,
                   size = 5,
                   min.segment.length = Inf,
                   nudge_x = 5) +
  scale_color_brewer(palette = "Dark2") +
  scale_fill_brewer(palette = "Dark2") +
  scale_x_continuous(breaks = seq(1992, 2004, 2)) +
labs(title = " 1")


onset = muni %>%
 dplyr::select(cod_dane, year, palm_presence) %>%
  filter(palm_presence == 1) %>%
  group_by(cod_dane) %>%
  summarise(first_treat = min(year))

suns = muni %>%
  dplyr::select(cod_dane, year, sdisp_rate, palm_presence,
         farc_attack, coca) %>%
```

```r
  mutate(year = unclass(year)) %>%
  # year treated
  left_join(onset) %>%
  # set never-treated to number above maximum year
  # (any year not in year == never-treated)
  mutate(first_treat = replace_na(first_treat, 10000)) %>%
  mutate(first_treat = unclass(first_treat)) %>%
  # change year format to match suns
  mutate(year_rel = year - 1992,
         first_treat_rel = first_treat - 1992)

res_sa20 = fixest::feols(sdisp_rate ~ farc_attack + coca +
                           sunab(first_treat_rel, year_rel) |
                           cod_dane + year, suns)



pDat = fixest::iplot(res_sa20)$prms %>% as_tibble()
# Graphs Event Study
pt3 <- ggplot(pDat, aes(x = x, y = estimate, ymin = ci_low, ymax = ci_high)) +
  geom_pointrange(fatten = 0.9, size = 0.7,
                  color = "#1b9e77", shape = 21, fill = "white") +
  geom_hline(yintercept = 0, linetype = "dashed") +
  geom_vline(xintercept = 0, linetype = "dashed") +
  theme_minimal() +
  labs(
    x = "Years since treatment",
    y = "Estimated effect of palm-oil adoption on displacement rate ",
    title = " "
  )

## models

df <- readRDS("C:/Users/Acer/Desktop/Maestria/LSE CLASES/MY457-Causal Inference/Summative/summative-reap


m1 = glm(displace ~ palmDummy + age + ed + child + sex + hhsize,
         family = binomial(link = 'logit'), data = df)

m2 = glm(displace_force2 ~ palmDummy + age + ed + child + sex + hhsize,
         family = binomial(link = 'logit'), data = df)

m3 = glm(disp_why_land ~ palmDummy + age + ed + child + sex + hhsize,
         family = binomial(link = 'logit'),
         data = df[df$displace == 1,])

m4 = glm(disp_paras ~ palmDummy + age + ed + child + sex + hhsize,
         family = binomial(link = 'logit'),
         data = df[df$displace == 1,])

m5 = lm(land_owned ~ palmDummy + age + ed + child + sex + hhsize,
        data = df[df$displace == 1,])
```

```r
m6 = glm(claimant ~ palmDummy + age + ed + child + sex + hhsize,
         family = binomial(link = 'logit'),
         data = df[df$displace == 1,])




# predicted probabilities
preds <- tibble(
  outcomes = c("Displaced", "Displaced by force",
               "Displaced for land", "Displaced by paras/elites",
               "Restitution opponent"),
  models = list(m1, m2, m3, m4, m6)
) %>%
  mutate(preds = map(models, ~ ggeffect(.x, terms = "palmDummy") %>% as_tibble())) %>%
  unnest(preds) %>%
  mutate(
    sample = case_when(
      outcomes == "Displaced" ~ "Full",
      outcomes == "Displaced by force" ~ "Excluding displaced by cross-fire",
      TRUE ~ "Displaced respondents"
    ),
    sample = factor(sample, levels = c("Full", "Excluding displaced by cross-fire",
                                       "Displaced respondents")),
    palm = ifelse(x == 1, "Palm", "No palm")
  )


# Plot
pt4<- ggplot(preds, aes(x = palm, y = predicted, ymin = conf.low, ymax = conf.high,
                        color = sample)) +
 geom_pointrange(fatten = 0.9, size = 0.9, shape = 21, fill = "white") +
 geom_hline(yintercept = 0, linetype = "dashed") +
 theme_minimal(base_size = 12) +
 facet_wrap(vars(outcomes), scales = "free_x") +
 labs(
 x = "Respondent municipality of origin",
 y = "Predicted probability of outcome",
 color = "Sample:"
 ) +
 scale_color_brewer(palette = "Dark2") +
 scale_y_continuous(labels = percent_format(), limits = c(0, 0.65)) +
 theme(
 legend.position = "top",
 strip.text = element_text(face = "bold")
 )




(figure1+pt3)
```

```r
ptt2

pt4

muni <- muni %>%
  mutate(cod_dane_num = as.numeric(as.character(cod_dane))) %>%

  group_by(cod_dane_num) %>%
  mutate(first_treated = ifelse(sum(palm_presence) > 0,
                                min(year[palm_presence == 1]),
                                Inf)) %>%
  ungroup()



att_result <- att_gt(yname = "sdisp_rate",
                     gname = "first_treated",
                     idname = "cod_dane_num",
                     tname = "year",
                     data = muni,
                     allow_unbalanced_panel = TRUE)


CSE <- aggte (att_result, type = "simple", na.rm = T)



muni$palm_presence <- as.numeric(muni$palm_presence)

# I created the first_treated as a numeric variable
muni <- muni[order(muni$cod_dane, muni$year), ]

muni$first_treated <- NA_real_
for(i in unique(muni$cod_dane)) {
  subset_data <- muni[muni$cod_dane == i, ]
  if(any(subset_data$palm_presence == 1, na.rm = TRUE)) {
    first_treat <- min(subset_data$year[subset_data$palm_presence == 1], na.rm = TRUE)
    muni$first_treated[muni$cod_dane == i] <- first_treat
  } else {
    muni$first_treated[muni$cod_dane == i] <- Inf  #never treated
  }
}

#I treat also year and the other variables as a numeric
muni$year <- as.numeric(muni$year)
muni$farc_attack <- as.numeric(muni$farc_attack)
muni$coca <- as.numeric(muni$coca)
muni$sdisp_rate <- as.numeric(muni$sdisp_rate)
muni$cod_dane <- as.numeric(as.character(muni$cod_dane))
#Sunab Model
```

```
sunab_est <- feols(sdisp_rate ~ sunab(first_treated, year) + farc_attack + coca |
                  cod_dane + year,
                  data = muni,
                  vcov = ~cod_dane)
ggregate_sunab <- aggregate(sunab_est, agg = "att")



fect<-fect(sdisp_rate ~ palm_presence, data = muni, index = c("cod_dane","year"),
                X = farc_attack+coca,
          force = "two-way", method = "fe", CV = TRUE, r = c(0, 5),
          se = TRUE, nboots = 200, parallel=TRUE)


ifect<-fect(sdisp_rate ~ palm_presence, data = muni, index = c("cod_dane","year"),
                X = farc_attack+coca,
          force = "two-way", method = "ife", CV = TRUE, r = c(0, 5),
          se = TRUE, nboots = 200, parallel=TRUE)

mc<-fect(sdisp_rate ~ palm_presence, data = muni, index = c("cod_dane","year"),
                X = farc_attack+coca,
          force = "two-way", method = "mc", CV = TRUE, r = c(0, 5),
          se = TRUE, nboots = 200, parallel=TRUE)




# Data
df <- data.frame(
  Method = c("TWFE", "Callaway & Sant'Anna", "Sun & Abraham", "FE (fect)",
             "IFE (fect)", "MC (fect)"),
  ATT = c(0.382, 0.200, 0.228, 0.512, -0.235, 0.482),
  SE = c(0.120, 0.135, 0.119, 0.120, 0.256, 0.132)
)

df$CI_lower <- df$ATT - 1.96 * df$SE
df$CI_upper <- df$ATT + 1.96 * df$SE

# Plot
ggplot(df, aes(x = reorder(Method, ATT), y = ATT)) +
  geom_point(size = 3, color = "#0072B2") +
  geom_errorbar(aes(ymin = CI_lower, ymax = CI_upper), width = 0.2, color = "#0072B2") +
  geom_hline(yintercept = 0, linetype = "dashed", color = "gray30") +
  labs(
    title = "",
    x = "Method",
    y = "ATT with 95% Confidence Interval"
  ) +
  theme_minimal(base_size = 14)
att_result <- att_gt(yname = "sdisp_rate",
                     gname = "first_treated",
                     idname = "cod_dane_num",
                     tname = "year",
```

```r
                     data = muni,
                     allow_unbalanced_panel = TRUE)

CSE1 <- aggte (att_result, type = "dynamic", na.rm = T)

plot_dynamic <- ggdid(CSE1, ncol = 5) +
  ggtitle("Callaway & Sant'Anna") +
  xlab("Years Relative to Treatment") +
  ylab("ATT") +
  scale_x_continuous(limits = c(-10, 10), breaks = seq(-10, 10, 2)) +
  geom_vline(xintercept = -1, linetype = "dashed")

print(plot_dynamic)
sunab_est <- feols(sdisp_rate ~ sunab(first_treated, year) + farc_attack + coca |
                  cod_dane + year,
                  data = muni,
                  vcov = ~cod_dane)


plot2 = fixest::iplot(sunab_est)$prms %>% as_tibble()

graph2<- ggplot(plot2, aes(x = x, y = estimate, ymin = ci_low, ymax = ci_high)) +
  geom_pointrange(fatten = 0.9, size = 0.7,
                  color = "#1b9e77", shape = 21, fill = "white") +
  geom_hline(yintercept = 0, linetype = "dashed") +
  geom_vline(xintercept = 0, linetype = "dashed") +
  theme_minimal() +
  labs(
    x = "Years since treatment",
    y = "ATT",
    title = "Sun & Abraham"
  )


plot(fect, main = "Estimated ATT (FEct)", ylab = "Effect of D on Y",
  cex.main = 0.8, cex.lab = 0.8, cex.axis = 0.8)

plot(ifect, main = "Estimated ATT (IFEct)", ylab = "Effect of D on Y",
  cex.main = 0.8, cex.lab = 0.8, cex.axis = 0.8)

plot(mc, main = "Estimated ATT (Mc)", ylab = "Effect of D on Y",
  cex.main = 0.8, cex.lab = 0.8, cex.axis = 0.8)


p1 <- plot_dynamic

p2 <- graph2
(p1+p2)


p3 <-plot(fect, main = "(FEct)", ylab = "Effect of D on Y",
  cex.main = 0.8, cex.lab = 0.8, cex.axis = 0.8)
p4 <-plot(ifect, main = "(IFEct)", ylab = "Effect of D on Y",
```

```r
  cex.main = 0.8, cex.lab = 0.8, cex.axis = 0.8)
p5 <- plot(mc, main = " (Mc)", ylab = "Effect of D on Y",
  cex.main = 0.8, cex.lab = 0.8, cex.axis = 0.8)


# join graphs
final_plot <- plot_grid(
 p3,
  p4, p5,
  ncol = 3
)
final_plot


#Treatment year
muni <- muni %>%
  group_by(cod_dane) %>%
  mutate(first_treatment = ifelse(any(palm_presence == 1), min(year[palm_presence == 1]),
                                  NA)) %>%
  ungroup()

# Define placebo treatment year: 3 and 10 years earlier.
muni <- muni %>%
  mutate(fake_treat_year_3 = first_treatment - 3)

muni <- muni %>%
  mutate(fake_treat_year_10 = first_treatment - 10)

#Placebo test dummy
muni <- muni %>%
  mutate(placebo_treat_3 = ifelse(!is.na(fake_treat_year_3) & year >= fake_treat_year_3,
                                  1, 0))

muni <- muni %>%
  mutate(placebo_treat_10 = ifelse(!is.na(fake_treat_year_10) & year >= fake_treat_year_10,
                                   1, 0))

#DID for placebo

placebo_model_3 <- feols(sdisp_rate ~ placebo_treat_3 + farc_attack + coca |
                           cod_dane + year, data = muni)

placebo_model_10 <- feols(sdisp_rate ~ placebo_treat_10 + farc_attack + coca |
                            cod_dane + year, data = muni)

#table

coef_map <- list(
  "placebo_treat_3"    = "Palm-oil placebo (-3)",
  "placebo_treat_10"   = "Palm-oil placebo (-10)",
  "farc_attack"        = "FARC Attack",
  "coca"               = "Coca plantation"
)
```

```r
# table
texreg(
  l = list(placebo_model_3, placebo_model_10),
  custom.model.names = c("Placebo Model -3", "Placebo Model -10"),
  custom.coef.map    = coef_map,
  digits             = 3,
  include.nobs       = TRUE,
  include.rsquared   = FALSE,
  include.adjrs      = FALSE,
  include.aic        = FALSE,
  include.bic        = FALSE
)

#I import the 4 main data based with data 2007-2020

Agri <- read_dta("C:/Users/Acer/Desktop/Maestria/LSE CLASES/MY457-Causal Inference/Summative/summative-r

C_A <- read_dta("C:/Users/Acer/Desktop/Maestria/LSE CLASES/MY457-Causal Inference/Summative/summative-re

BG <- read_dta("C:/Users/Acer/Desktop/Maestria/LSE CLASES/MY457-Causal Inference/Summative/summative-rea

POB <- read_dta("C:/Users/Acer/Desktop/Maestria/LSE CLASES/MY457-Causal Inference/Summative/summative-re
#I subsets the years before 2015
base_C_A <- C_A %>%
  filter(ano >= 2007 & ano <= 2015) %>%
  dplyr::select(codmpio, coca, tpobc_FARC, desplazados_expulsion, ano)

base_BG <- BG %>%
  filter(ano >= 2007 & ano <= 2015) %>%
  dplyr::select(codmpio, DI_eficacia, DI_eficiencia, DI_capadmin, DI_desemp_int, ano)

base_Agri <- Agri %>%
  filter(ano >= 2007 & ano <= 2015) %>%
  dplyr::select(codmpio, r_palmaa, ano)

base_POB <- POB %>%
  filter(ano >= 2007 & ano <= 2015) %>%
  dplyr::select(codmpio, retro_pobl_tot, ano)

#I joined the database with the municipal code and the year
cbase_final <- base_C_A %>%
  left_join(base_BG, by = c("codmpio", "ano")) %>%
  left_join(base_Agri, by = c("codmpio", "ano")) %>%
  left_join(base_POB, by = c("codmpio", "ano"))

#I create the dummy variable for palm presence and coca plantation as the author did

cbase_final <- cbase_final %>%
  mutate(dummy_palmaa = ifelse(!is.na(r_palmaa) & r_palmaa > 0, 1, 0))

cbase_final <- cbase_final %>%
  mutate(coca = ifelse(!is.na(coca) & coca > 0, 1, 0))
```

```r
#I create the index of displacement dividing with the municipal population and
#then transform it into the inverse sine
cbase_final <- cbase_final %>%
  mutate(
    desplazamiento_rate = (desplazados_expulsion / retro_pobl_tot) * 1000,
    desplazamiento_rate_asinh = asinh(desplazamiento_rate)
  )


#model with interaction of quality institutions

m113<- feols(desplazamiento_rate_asinh ~ -1 + dummy_palmaa + coca + tpobc_FARC + dummy_palmaa*DI_desemp_

m113
texreg(m113,
       custom.model.names = "Model 113",
       custom.coef.names = c(
                              "dummy_palmaa" = "Palm Oil plantation",
                              "coca" = "Coca Cultivation Presence",
                              "tpobc_FARC" = "FARC Presence",
                              "DI_desemp_int" = "Institutional Quality Index",
                              "dummy_palmaa:DI_desemp_int" = "Plantation x Institutional Quality Index"

                            ),
       caption = "Regression Results for Model with Institutions",
       omit.coef="(ano|codmpio|DI_desemp_int|coca|tpobc_FARC)",
       stars = c(0.001, 0.01, 0.05, 0.1),
       gof = c("nobs", "r.squared")
)

# this chunk generates the complete code appendix.
# eval=FALSE tells R not to re-run (``evaluate'') the code here.
```