MY457: Reappraisal of Tellez (2022): Land, Opportunism, and Displacement in Civil Wars: Evidence from Colombia

Candidate Number: 43260

Fri/23/May

1 Introduction

Palm oil expansion has often been linked to environmental harm, elite land capture, and forced displacement. In a key contribution, Tellez (2022) argues that palm cultivation in Colombia drove displacement, particularly where paramilitaries aligned with palm-producing elites. Using a difference-in-differences (DiD) framework, the study estimates the effect of palm presence and national production trends on displacement rates.

While the theoretical logic is sound, the paper's causal claims rest on strong assumptions and modeling choices that merit closer scrutiny—especially the use of Two-Way Fixed Effects (TWFE) in a staggered treatment setting, and the reliance on national-level production as a proxy for treatment intensity. This reappraisal revisits the original analysis using updated estimators, new subnational data, and extended robustness checks.

I begin by re-estimating the main model with Interactive Fixed Effects (fect), showing consistent results with minor timing differences. I then assess the credibility of the identification strategy through placebo period and equivalence tests, finding no evidence of spurious pre-treatment effects. The central contribution is a new interaction model using original municipality-level soil quality data from IGAC and CEDE, which offers a more direct proxy for land value. The results reveal that displacement is not most intense in highly fertile areas—challenging a simple economic opportunism story and pointing instead to violence concentrated in marginal, institutionally weaker zones.

Finally, I extend the robustness analysis using new data on electoral violence and fraud from the Misión de Observación Electoral (MOE). With no effects detected on these placebo outcomes, the findings support the specificity of palm-related displacement rather than broader political instability.

2 Estimator Choices

This reappraisal first argues that the main findings in Tellez (2022) rely on a suboptimal estimator — Two-Way Fixed Effects (TWFE) via felm — which is known to produce biased estimates in settings with staggered treatment timing and heterogeneous treatment effects. Although the author acknowledges and implements the Sun & Abraham (2020) estimator, it is presented only as a robustness check, not the primary estimator. Given the data structure, the sunab() estimator is better aligned with the treatment design and should form the basis for the paper's main conclusions.

Tellez (2022) estimates the effect of palm oil expansion on forced displacement using a Difference-in-Differences framework implemented via the felm() estimator from the lfe package. While this approach is computationally efficient and historically popular in panel data settings, it is now well understood that TWFE models can yield biased estimates in the presence of staggered treatment adoption and heterogeneous treatment effects — both of which characterize the empirical context of palm expansion in Colombia. In particular, the treatment variable (municipal-level palm adoption) enters at different times across municipalities and is absorbing: once a unit adopts palm, it is considered treated in all subsequent years. In such settings, TWFE estimators may use already-treated units as controls for later-treated units, leading to contaminated comparisons and incorrect estimates of the Average Treatment Effect on the Treated (ATT) (Goodman-Bacon, 2021).

To further validate and extend these results, I re-estimate the main specifications using the Callaway & Sant'Anna (2021) estimator and the Interactive Fixed Effects (IFE) approach from the fect package. These estimators are designed to handle heterogeneous treatment effects and staggered timing while avoiding the bias of TWFE models. Unlike TWFE or GSCM (Generalized Synthetic Control), which do not natively align units by treatment timing, these methods offer a more flexible and robust framework for causal inference.

Moreover, while Sun & Abraham's approach treats treatment as binary and persistent, it does not account for variation in treatment intensity across post-treatment periods. This may oversimplify the dynamics of palm expansion, which likely intensifies incrementally after initial adoption. Methods like fect help address this by estimating unit-specific trends and capturing gradual treatment dynamics more flexibly.

The author's main results are replicated with no issues to flag in Table 1 and its dynamic treatment effects using the Sun & Abraham method are shown in Figure 1. Extending these results to more modern estimators using the fect package, I check whether Tellez's results are robust to this more modern estimator, plotting the dynamic treatment effect in Figure 2.

2.1 Main results replicated

Table 1: Effect of palm-oil growth on displacement. Models include municipal and year fixed effects and controls for time-varying presence of coca and FARC attacks.

	Dependent variable:			
	Displacement rate (inv-sine transformation)			
	(1)	(2)	(3)	(4)
Palm-oil plantation	0.382***	−2.056***	−1.213***	0.692
	(0.120)	(0.642)	(0.023)	(0.548)
Plantation X Natl Prod (log)		0.248***		
		(0.063)		
Plantation X AUC presence (dummy)			1.614***	
			(0.120)	
Plantation X FARC presence (dummy)				−0.327
				(0.559)
Observations	14,192	14,192	14,192	14,208
Note:	*p<0.1; **p<0.05; ***p<0.01			

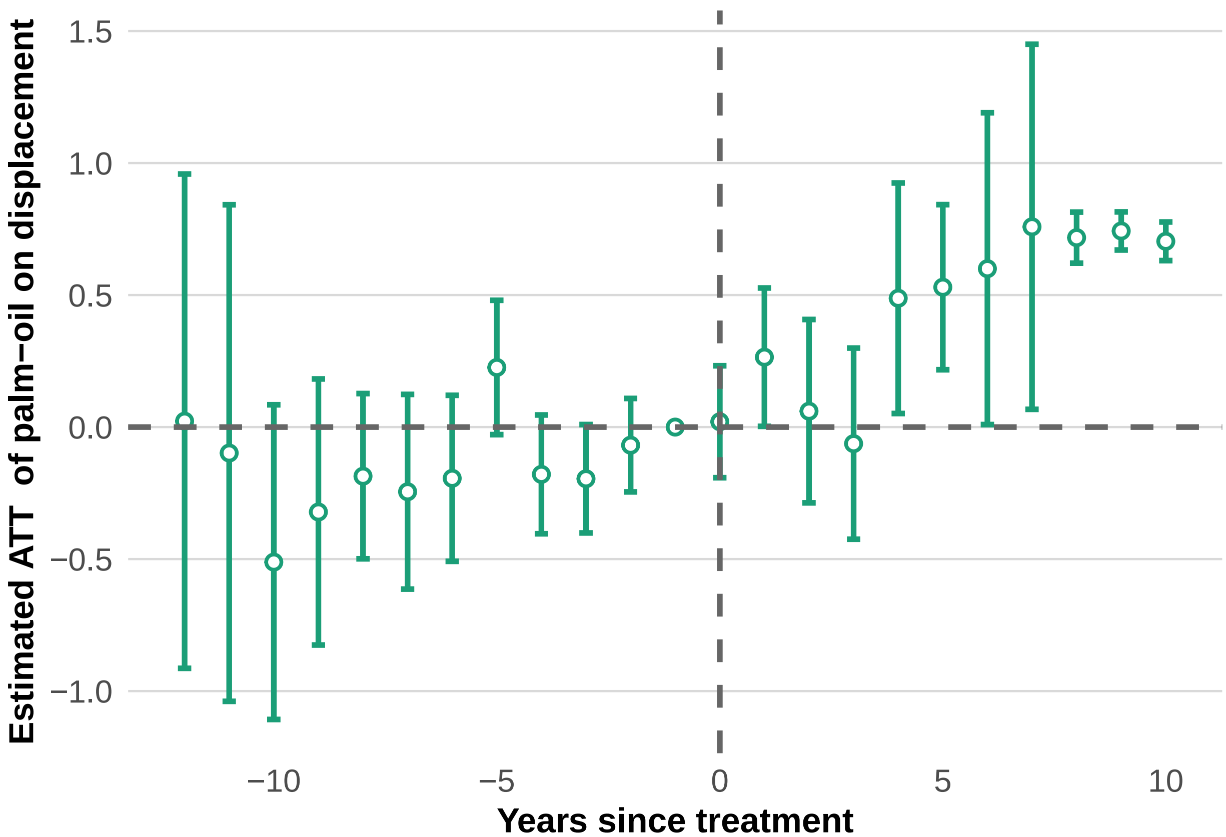

Figure 1: Dynamic treatment effects with the Sun & Abraham Method

2.2 Main results Extended with modern estimator

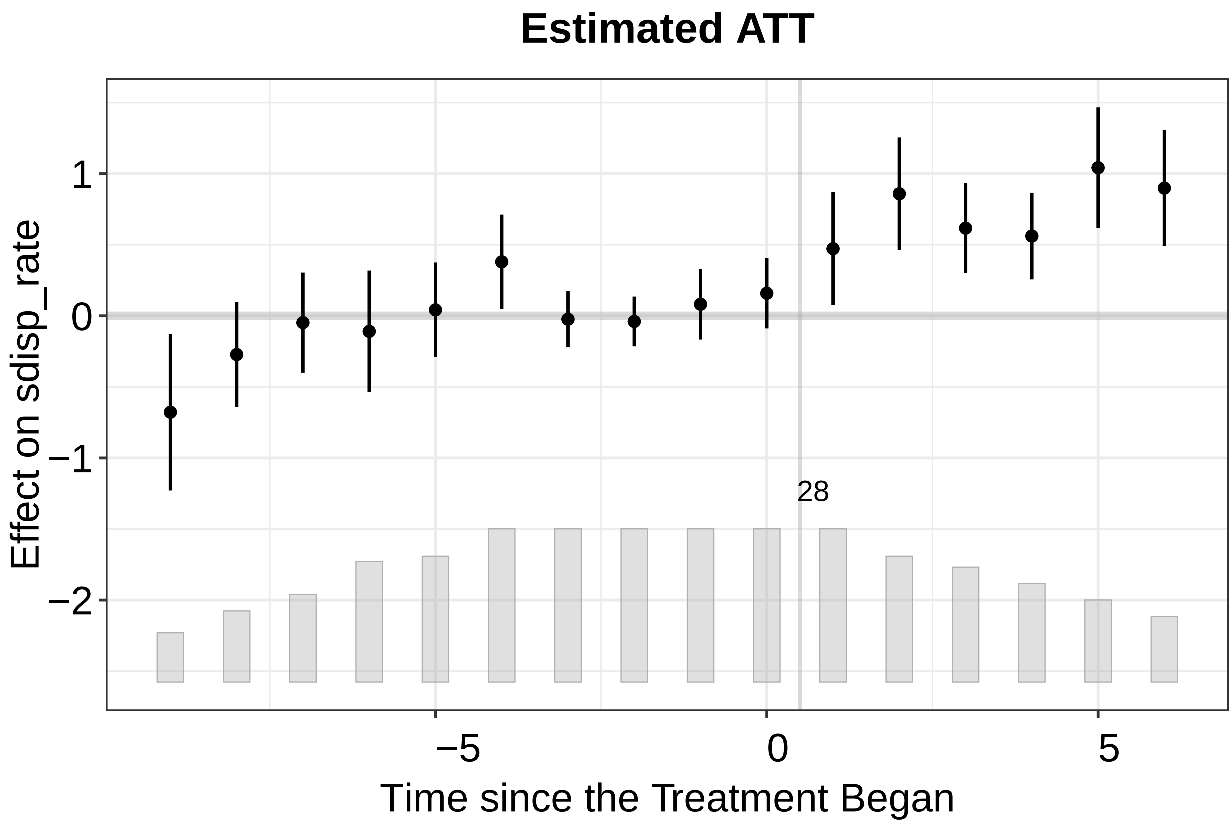

Figure 2: Dynamic treatment effects with modern fect estimator

Figure 1 presents the dynamic treatment effects estimated using the Sun & Abraham (2020) method implemented via sunab(), while Figure 2 shows the results from the Interactive Fixed Effects model (fect). Both approaches account for staggered treatment timing and heterogeneous effects, and their findings are broadly consistent: palm oil adoption is associated with a delayed but sustained increase in displacement rates. Both estimators were fitted using the model 2 in Table 1 for consistency.

The Sun & Abraham estimates show effects that become positive and statistically significant approximately five years after treatment, with pre-treatment coefficients centered near zero and is consistent with no parallel pre-trends violation. The fect estimator reveals a similar trajectory, with negative or null effects in the pre-treatment period and growing positive effects post-treatment. Although the magnitude of the post-treatment estimates under fect is slightly higher, both models detect a clear upward shift in displacement following palm expansion. This alignment across two modern estimators strengthens confidence in the robustness of the original findings and supports the interpretation that palm adoption contributes to forced displacement.

3 Addressing DiD Assumptions: Generalized Synthetic Control Method (GSCM)

As part of his robustness checks, Tellez (2022) estimates treatment effects using the Generalized Synthetic Control Method (GSCM) developed by Xu (2017), implemented via the gsynth package. This estimator extends the synthetic control approach to settings with multiple treated units and time periods, allowing for time-varying unobserved confounders through a latent factor model. GSCM represents a valuable addition to the identification strategy insofar as it addresses concerns about omitted variable bias in traditional Two-Way Fixed Effects (TWFE) models, particularly when unobserved heterogeneity varies over time and across units.

However, there are important limitations to how GSCM is applied in this case. While the plot presented in the appendix (reproduced here as Figure 3) is labeled as showing dynamic treatment effects "relative to treatment onset," this is somewhat misleading. The gsynth() function does not re-align municipalities by their individual treatment start dates. Instead, the method estimates average treatment effects in calendar time, pooling units treated in different years without adjusting for staggered adoption. As a result, the x-axis cannot be interpreted as event time (i.e., time relative to treatment), and the pattern of effects may reflect a mix of early- and late-treated units at different stages of treatment exposure. This undermines the dynamic interpretation of the plot and risks conflating variation across treatment cohorts.

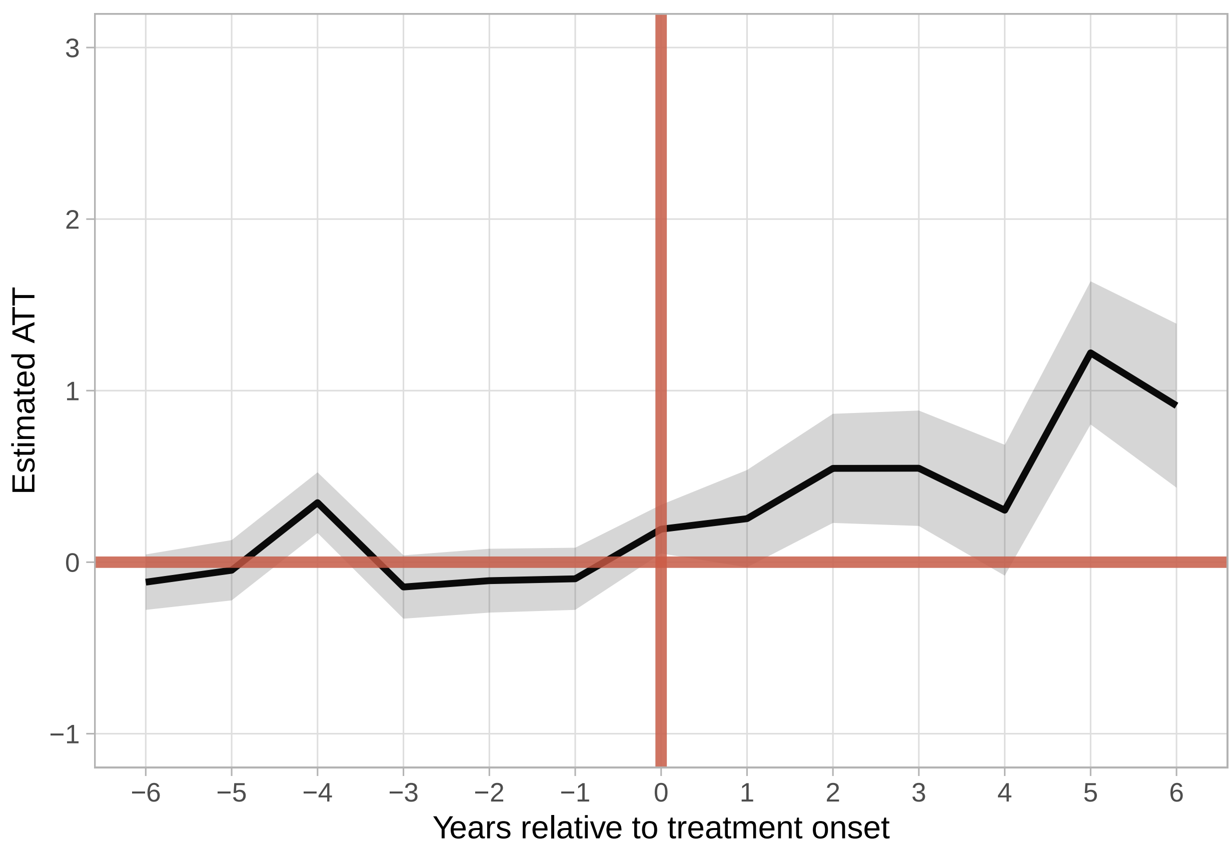

Figure 3: ATT relative to year of treatment onset

The plot in Figure 3 visualizes average treatment effects over time using the Generalized Synthetic Control Method (GSCM). The black line represents the estimated ATT (Average Treatment Effect on the Treated) in each period, and the grey band is the associated confidence interval.

The ATT estimates before treatment hover close to zero, with overlapping confidence intervals. This provides suggestive support for the parallel pre-trends assumption. In the post-treatment period, the ATT becomes positive and grows over time, peaking around 5 years after treatment, though the confidence intervals are wider, reflecting less precision.

4 Rethinking Interaction Effect Model: New Evidence from Colombian Soil Quality Data

In seeking to understand heterogeneity in the displacement effects of palm oil expansion, Tellez (2022) includes an interaction between municipal palm presence and the log of national palm oil production. This is a theoretically reasonable attempt to proxy for variation in production intensity, under the assumption that national-level expansion raises the profitability of cultivating palm at the municipal level. However, this strategy ultimately combines a binary, time-varying treatment indicator (palm presence) with a continuous but time-invariant contextual variable, and does not directly capture local variation in land suitability or palm productivity.

While Tellez's interaction term serves as an indirect proxy for intensity, it could conflate municipal treatment status with broader market trends, and cannot account for heterogeneity in land quality across treated municipalities. Crucially, this approach assumes that variation in displacement is driven by changes in national supply conditions, rather than underlying differences in land value or agricultural potential at the municipal level. Yet it is precisely this local variation—how profitable or attractive a municipality is for cultivation—that is most likely to shape violent displacement dynamics.

To address this limitation, I introduce a new interaction model that leverages municipality-specific data on soil quality, which directly measures land value and agricultural suitability. The soil quality indicator ranges from 1 to 8 and is constructed by CEDE based on data collected by IGAC (Instituto Geográfico Agustín Codazzi). It captures the fertility and suitability of land for agriculture at the municipal level. For municipalities with missing values in some years, I impute soil quality using the available observation for that municipality, which is valid given the static nature of the variable.

Figure 4 reproduces the author's marginal effects plot, showing the estimated effect of palm presence on displacement across values of national production (lnatprod). Figure 5 presents my alternative specification, in which the marginal effect of palm presence is plotted across values of soil quality. This soil-based interaction model provides a more direct test of the hypothesized mechanism. It complements the original analysis by grounding the heterogeneity in observable, spatially disaggregated data, and reinforces the conclusion that land dispossession was targeted and economically motivated.

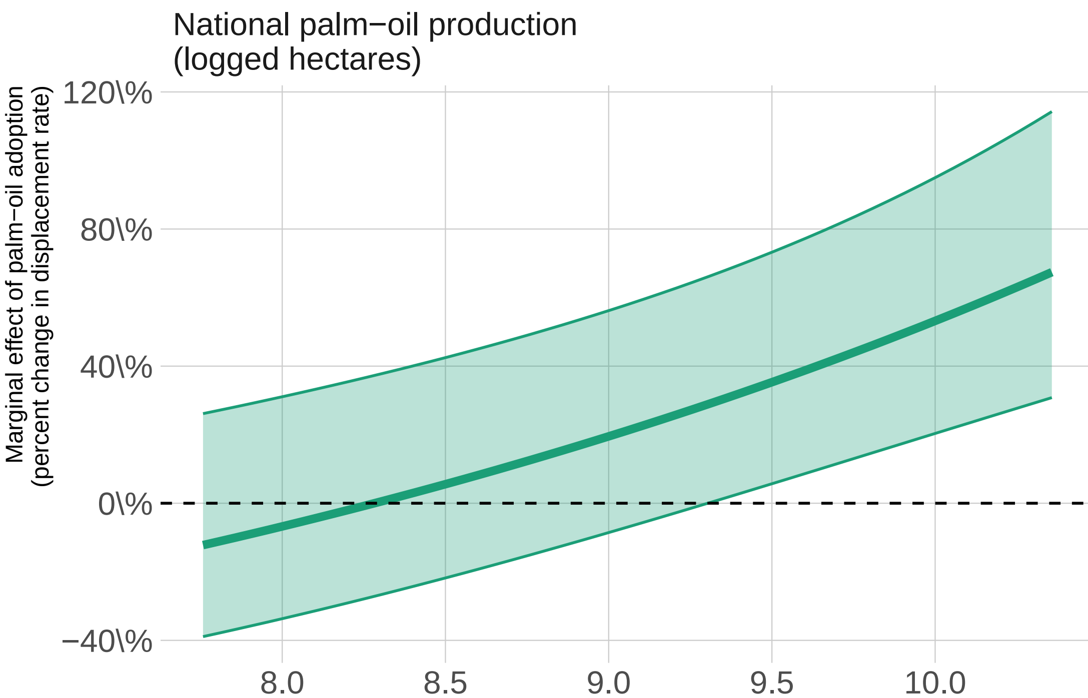

Figure 4: Effect of Palm-Oil Presence on Displacement Rate across National Palm-Oil Production Levels

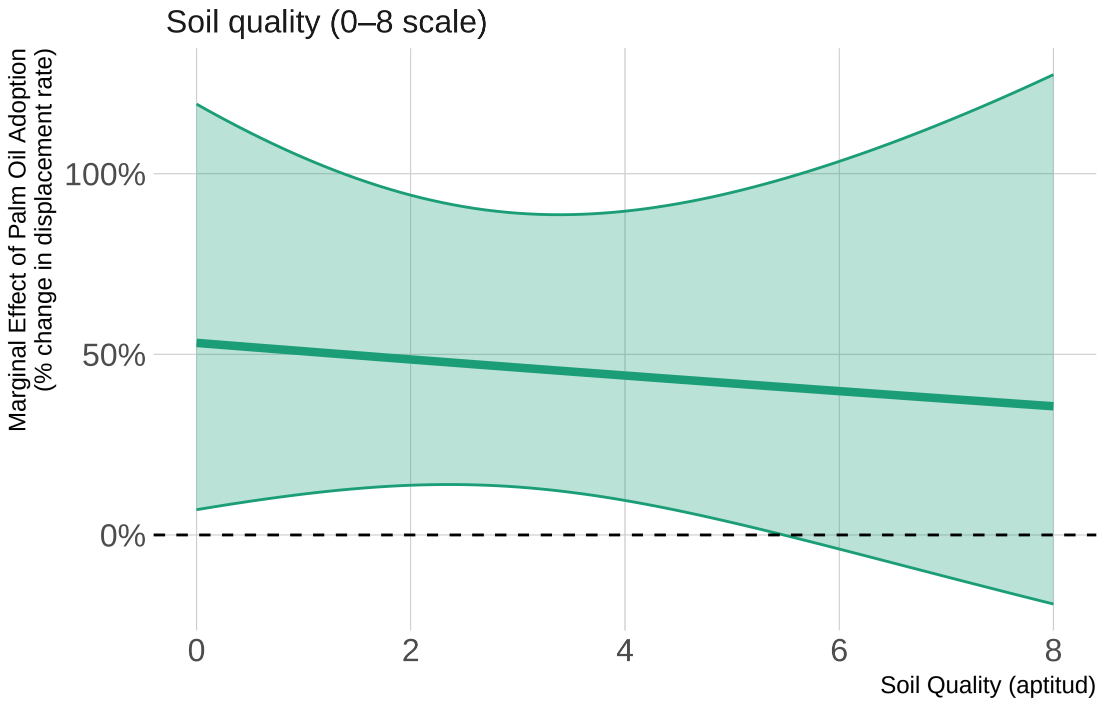

Figure 5: Effect of Palm-Oil Presence on Displacement Rate across Agricultural Soil Quality Levels

Figure 4 reproduces Tellez's marginal effects plot, showing that the displacement effect of palm oil adoption intensifies with rising national production. This suggests that broader national-level palm cultivation may have amplified the profitability of palm and, in turn, incentivised land grabs. In contrast, Figure 5 presents the results from my interaction model using municipality-specific soil quality. Here, I find that while palm oil adoption is consistently associated with increased displacement, the effect slightly declines as soil quality improves. The marginal effect remains positive across most of the range, but is noticeably weaker in municipalities with highly fertile land. Moreover, the confidence intervals widen and cross zero at higher soil quality levels, indicating increased uncertainty and weaker effects in these areas.

This finding challenges a simple opportunism hypothesis, which would predict stronger displacement effects in more valuable, agriculturally productive areas. Instead, it raises evidence that palm-related displacement has been concentrated in marginal or less-developed lands—areas, where tenure protections were possibly weaker, resistance was lower, or state capacity was minimal. Alternatively, if elite landholdings were already dominant in high-quality areas, there may have been less need for displacement to acquire land. In either case, these results suggest that palm-driven displacement is not purely a function of national agricultural demand, but may reflect strategic targeting in potentially more vulnerable municipalities.

5 Placebo Tests: New Evidence from Colombian Electoral Fraud and Violence Data

5.1 Author's Approach: Guerrilla Attacks as a Placebo Outcome

To address concerns that palm oil expansion may simply be correlated with general conflict or instability, Tellez (2022) includes a placebo test using FARC attacks against civilians. The logic is straightforward: if palm presence were simply a proxy for conflict intensity, we would expect it to predict not only displacement but also broader forms of violence. Finding no such relationship strengthens the claim that palm-driven displacement reflects a targeted process, not general violence escalation.

Tellez reports no significant relationship between palm expansion and guerrilla violence suggesting that the displacement effect is not confounded by simultaneous increases in conflict intensity. This is a valuable and well-motivated test. However, it addresses only a narrow form of violence (guerrilla attacks by a specific group), leaving open the possibility that palm adoption correlates with other forms of political or institutional instability.

5.2 My Extensions: Placebo Period and Equivalence Tests

To strengthen the credibility of the causal claims, I implement two additional placebo strategies using the fect estimator: a placebo period test and an equivalence test.

The placebo period test re-estimates the model using a fake treatment period in the pre-treatment window, specifically two years prior to actual treatmen, while the equivalence test formally assesses whether pre-treatment ATT estimates are statistically indistinguishable from zero. These are both shown in Figure 6 below.

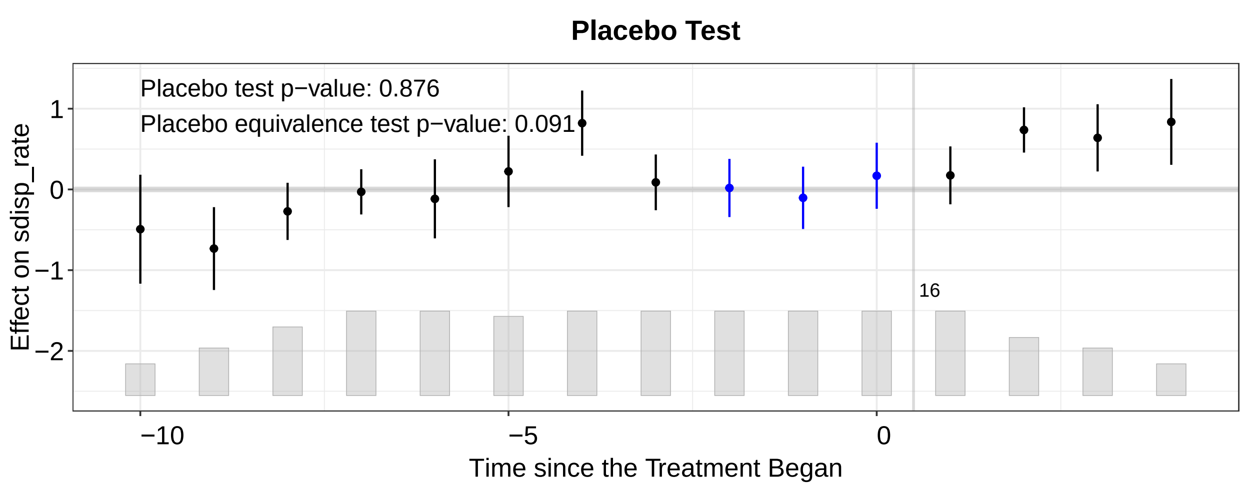

Figure 6: Placebo and Equivalence Test results

The placebo test p-value (0.876) indicates no significant treatment effect in the pre-treatment window, and the equivalence test p-value (0.091) suggests that these placebo estimates are statistically indistinguishable from zero within a reasonable tolerance band. Together, these findings support the assumption of parallel pre-treatment trends and provide reassurance that the model is not detecting spurious effects when treatment has not yet occurred.

I next turn to real-world placebo outcomes. Using newly collected data, I test whether palm oil adoption predicts changes in electoral violence and electoral fraud—outcomes that should not, in theory, be affected by land-driven palm expansion. These tests serve as strong falsification exercises: any significant effect would cast doubt on the interpretation that displacement is uniquely linked to palm-driven land dispossession.

5.3 Electoral Violence & Fraud

The data for these outcomes are drawn from the Misión de Observación Electoral (MOE), a Colombian electoral observation body. The electoral violence variable (elecviol) captures the number of violent incidents linked to electoral events in each municipality-year, while the electoral fraud variable (fraude) measures instances of documented election-related fraud. Both variables are only available for the years 2002 and 2003 at the municipality level which restricts the sample size and both datasets were accessed through the replication materials of Nieto-Matiz (2019).

To estimate the effects of palm expansion on these placebo outcomes, I employ the Callaway & Sant'Anna (2021) estimator, which is well-suited for this context. It explicitly accounts for staggered treatment adoption, allows for treatment effect heterogeneity, and estimates dynamic treatment effects across time relative to treatment. I use a doubly robust estimation strategy and set the base period to "universal".

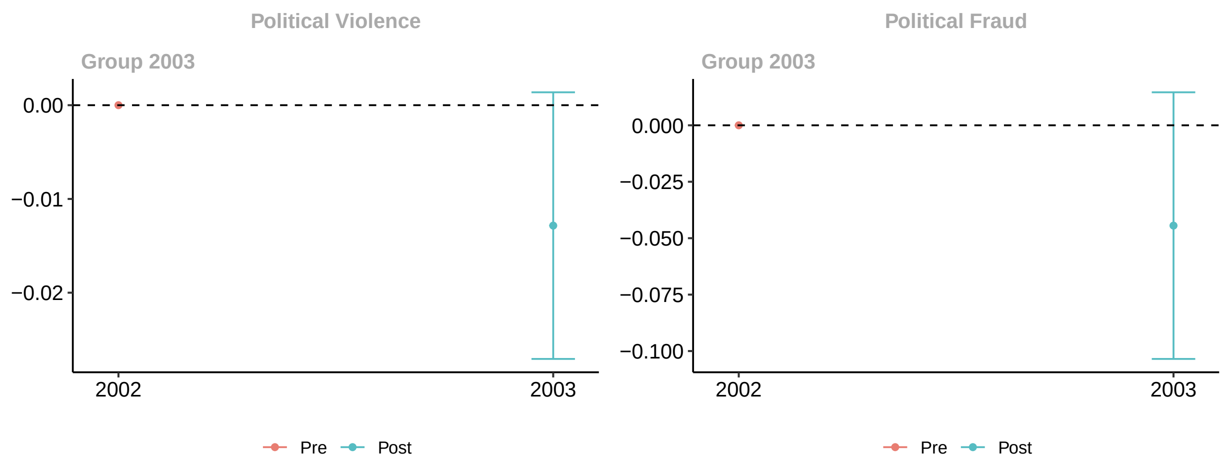

Figure 7: Placebo Effects of Palm Oil Adoption on Electoral Violence and Fraud (2002–2003)

Figure 7 presents placebo tests using electoral violence and fraud as outcomes, estimated with the Callaway & Sant'Anna (2021) method. In both cases, the point estimates are close to zero across pre- and post-treatment periods, and the confidence intervals are wide and consistently include zero. These results suggest that palm oil adoption did not lead to increases in electoral violence or fraud, nor were such effects already present before treatment. In short, palm expansion does not appear to generate broader institutional instability or political manipulation

The scope of these tests is constrained by limited data availability: electoral outcomes are only observed in 2002 and 2003, restricting the analysis to a single cohort. Nonetheless, the absence of any detectable effect, even under these conditions, offers supportive evidence for the specificity of palm-related displacement.

6 Conclusion

This reappraisal examined the empirical strategy used in Tellez (2022), focusing on estimator choice, assumption validation, and the interpretation of interaction effects. First, I highlighted that the paper's reliance on Two-Way Fixed Effects (TWFE) may be suboptimal given the staggered, absorbing nature of the treatment. By re-estimating the main results using a modern Interactive Fixed Effects model (fect), I show that the core findings are broadly robust to more appropriate estimators, though with some variation in the timing of effects.

In evaluating the assumptions underlying the causal design, I replicated the Generalized Synthetic Control Method (GSCM) used by the author. While this model provides visual support for the parallel pre-trends assumption, it lacks event-time alignment, limiting its interpretability in settings with heterogeneous treatment timing.

The most substantive contribution of this reappraisal lies in reassessing the interaction effect model. I argue that Tellez's use of national-level palm oil production as an interaction term provides only an indirect and locality-invariant proxy for land value. In contrast, I introduce a new interaction model using original data on municipality-level soil quality, which offers a direct measure of agricultural suitability and land value. The results reveal that while palm oil expansion increases displacement overall, the effect weakens in municipalities with more fertile land. This challenges a simple opportunism hypothesis and suggests that displacement may have been most strongly at play in marginal or institutionally weaker areas—where state capacity and land tenure protections may be looser.

Finally, I extend the robustness checks through placebo period and equivalence tests, and introduce real-world placebo outcomes—electoral violence and fraud—using new data and the Callaway & Sant'Anna (2021) estimator. The absence of any detectable effect on these outcomes reinforces the claim that palm-related displacement is a specific, targeted process rather than a reflection of broader political instability.

Overall, the reappraised results largely support the main findings of Tellez (2022), while extending it with updated estimators, new subnational data, and more rigorous placebo testing. At the same time, the analysis reveals important nuances in the heterogeneity of treatment effects. Specifically, the evidence suggests that displacement may have been more concentrated in agriculturally less valuable areas—pointing toward a pattern of armed groups targeting more vulnerable, institutionally weaker regions rather than the most fertile land. This interpretation does not contradict Tellez's conclusions, but it refines the understanding of where and why displacement occurred. These findings also highlight promising directions for future research. Improved access to fine-grained, time-varying data on land tenure and palm production intensity would enable a more precise identification of the underlying mechanisms.

7 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(ggeffects)
library(stargazer)
library(hrbrthemes)
library(here)
library("effects")
library(hrbrthemes)
library(ggrepel)
library(lfe)
library(interplot)
library(stargazer)
library(patchwork)
library(broom)
library(here)
library(paletteer)
library(gsynth)
library(tidyverse)
library(here)
library(broom)
library(stargazer)
library(hrbrthemes)

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



# palm presence models -----------------------------------------------------------------

# read data
df = read_rds("repFile/repl-data/muni.rds")


# fix haven_labelled variables
library(haven)
library(dplyr)

df <- df %>% mutate(across(where(is.labelled), as_factor = FALSE, as.numeric))

# stupid geom_label_repel problem
update_geom_defaults("label", list(family = "Roboto Condensed"))
update_geom_defaults("label_repel", list(family = "Roboto Condensed"))
```

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

# models
m1 = felm(sdisp_rate ~ palm_presence + farc_attack + coca
          | cod_dane + year | 0 | cod_dane, exactDOF = 'rM',
          data = df)

m1a = felm(ldisp_rate ~ palm_presence + farc_attack + coca
           | cod_dane + year | 0 | cod_dane, exactDOF = 'rM',
           data = df)


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


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

## farc attacks
m_farc = felm(sdisp_rate ~ palm_presence*farc_dummy + coca
              | cod_dane + year | 0 | cod_dane, exactDOF = 'rM',
              data = df)

# table output
stargazer(m1, m2, m5, m_farc,
          type = 'latex',
          keep = c('palm_presence',
                   ':lnatprod',
                   ':auc_dummy',
                   ':farc_dummy'),
          covariate.labels = c('Palm-oil plantation',
                               'Plantation X Natl Prod (log)',
                               'Plantation X AUC presence (dummy)',
                               'Plantation X FARC presence (dummy)'),
      keep.stat = 'n',
      title = 'Effect of palm-oil growth on displacement. Models include municipal and year fixed effec
      label = 'palm-pres-mods',
      out = 'repFile/paper/figures/palm-pres-mods.tex',
      dep.var.labels = c('Displacement rate (inv-sine transformation)'))
library(tidyverse)
library(hrbrthemes)

# load data
df = read_rds('repFile/repl-data/muni.rds')
```

```
## NEED: variable telling us time at which each unit is first treated
onset = df %>%
  dplyr::select(cod_dane, year, palm_presence) %>%
  filter(palm_presence == 1) %>%
  group_by(cod_dane) %>%
  summarise(first_treat = min(year))



# suns and abrams -------------------------------------------------------------------

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


x <- ggplot(pDat, aes(x = x, y = estimate)) +
  geom_errorbar(
    aes(ymin = ci_low, ymax = ci_high),
    width = 0.3,          # width of the bars (horizontal length)
    size = 1,           # thickness of the bars
    color = "#1b9e77"
  ) +
  geom_point(
    shape = 21,
    fill = "white",
    color = "#1b9e77",
    size = 2,             # size of the central point
    stroke = 1          # thickness of point border
  ) +
  geom_hline(yintercept = 0, linetype = 2, size = 1, color = "grey40") +
  geom_vline(xintercept = 0, linetype = 2, size = 1, color = "grey40") +
  theme_minimal(base_size = 13) +
  theme(
    axis.title = element_text(face = "bold", size = 13),
    axis.text = element_text(size = 12),
    panel.grid.minor = element_blank(),
    panel.grid.major.y = element_line(size = 0.4, color = "grey85"),
```

```r
    panel.grid.major.x = element_blank()
  ) +
  labs(
    x = "Years since treatment",
    y = "Estimated ATT  of palm-oil on displacement"
  )
x

#ggsave("repFile/paper/figures/suns.pdf", device = cairo_pdf)


df_merged <- readRDS("tellez_nieto_merged.rds")


df_merged$palm_soil_interaction <- df_merged$palm_presence * df_merged$aptitud_imputed

fect_out <- fect::fect(
  sdisp_rate ~ palm_presence + palm_soil_interaction + farc_attack + coca,
  data = df_merged,
  index = c("cod_dane", "year"),
  method = "ife",
  se = TRUE,
  nboots = 200
)

# Visualise the effect:
plot(fect_out)

# libraries
library(gsynth)
library(tidyverse)
library(here)

# load data
df = read_rds(here('repFile', 'repl-data', 'muni.rds')) %>%
  dplyr::select(sdisp_rate, palm_presence,
                coca, farc_attack,
                cod_dane, year) %>%
  drop_na()



unique(df$cod_dane)
##################################################My addition
#problematic variables handling

df <- df %>%
  mutate(year = haven::zap_labels(year))

df <- df %>%
  mutate(
    year = as.integer(year)           # or: factor(year, ordered = TRUE)
  )
```

```
df <- df %>%
  mutate(cod_dane = as.character(cod_dane))
length(unique(df$cod_dane))       # should be >1 if multiple municipalities
unique(df$cod_dane)

df <- df %>%
  mutate(cod_dane = factor(cod_dane))  # ensures multiple levels if data is good

nlevels(df$cod_dane)        # should return 1118
 # show first 10 levels
########################################################
# fit the model
out = gsynth(sdisp_rate ~ palm_presence + farc_attack + coca,
             data = df,
             index = c("cod_dane","year"),
             se = TRUE,
             inference = "parametric",
             r = c(0, 5),
             CV = TRUE,
             EM = TRUE,
             parallel = TRUE,
             cores = 12,
             force = "two-way",
             nboots = 1000,
             seed = 02139,
             min.T0 = 7)


# save output
save(out, file = here("repFile", "repl-data", "gsynth.rda"))


# load model output
load(here("repFile", "repl-data", "gsynth.rda"))


# plot
p <- plot(out)


print(p + labs(y = "Estimated ATT",
         title = NULL,
         x = "Years relative to treatment onset") +
  theme_light(base_size = 12) +
  geom_hline(yintercept = 0, color = "coral3", size = 2, alpha = .8) +
  geom_vline(xintercept = 0, color = "coral3", size = 2, alpha = .8) +
  theme(panel.grid.minor = element_blank()) +
  xlim(c(-6, 6)) +
  scale_x_continuous(breaks = -6:6) +
  ylim(c(-1, 3))
)
# marginal effects
```

```r
# read data
df = read_rds("repFile/repl-data/muni.rds")

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

## code from Patrick Baylis to deal with felm: https://www.patrickbaylis.com/blog/2021-01-22-predict-pa
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
  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))
```

```r
  } else {
    return(fit)
  }
}

get_se <- function(r, sig) {
  # Compute 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),
                                by = .05))



# 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)")

suppressWarnings({
  p1 = 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_ipsum(base_family = "sans") +
    scale_y_continuous(labels = function(x) paste0(x * 100, "\\%")) +
    facet_wrap(vars(group)) +
    theme(panel.grid.minor = element_blank()) +
    geom_hline(yintercept = 0, lty = 2)
  p1
})

#Merging Tellez (2022) and Nieto-Matiz (2019) data.

# Tellez (2022)

df = read_rds("repFile/repl-data/muni.rds")
```

```r
# Nieto Matiz (2019)
dat <- read.dta13("C:/Users/Lucca/causalinference/summative-reappraisal-luccarallovander/mydata/Nieto_Ma
unique(dat$codmpio)


# Load unique codes
tellez_codes <- unique(df$cod_dane)
nieto_codes <- unique(dat$codmpio)

# Ensure they're character or numeric for reliable comparison
tellez_codes <- as.character(tellez_codes)
nieto_codes <- as.character(nieto_codes)

# 1. Compare intersection
shared_codes <- intersect(tellez_codes, nieto_codes)

# 2. Codes only in Tellez
only_in_tellez <- setdiff(tellez_codes, nieto_codes)

# 3. Codes only in Nieto
only_in_nieto <- setdiff(nieto_codes, tellez_codes)

# 4. Summary
summary_df <- data.frame(
  Tellez_Total = length(tellez_codes),
  Nieto_Total = length(nieto_codes),
  Shared = length(shared_codes),
  Tellez_Only = length(only_in_tellez),
  Nieto_Only = length(only_in_nieto)
)

# Ensure codes and years have the same type
df$cod_dane <- as.character(df$cod_dane)
df$year <- as.integer(df$year)

dat$codmpio <- as.character(dat$codmpio)
dat$ano <- as.integer(dat$ano)

# Left join Nieto-Matiz onto Tellez
df_merged <- left_join(
  df,
  dat,
  by = c("cod_dane" = "codmpio", "year" = "ano")
)

names(df_merged)

# Choose one - let's say you want coca.x
df_merged <- df_merged %>%
  mutate(coca = coca)

saveRDS(df_merged, "tellez_nieto_merged.rds")
```

```r
#load merged data
df_merged <- readRDS("tellez_nieto_merged.rds")

# MODEL
library(lfe)

m_soil <- felm(
  sdisp_rate ~ palm_presence * aptitud_imputed + farc_attack + coca |
    cod_dane + year | 0 | cod_dane,
  data = df_merged
)

# HELPER FUNCTIONS (unchanged)
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
  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) {
  r <- matrix(r, nrow = 1)
  sqrt(r %*% sig %*% t(r))
}

# NEWDATA with coca
newdata <- tibble(
  palm_presence = 1,
  farc_attack = 0,
  coca = 0,
  aptitud_imputed = seq(min(df_merged$aptitud_imputed, na.rm = TRUE),
                        max(df_merged$aptitud_imputed, na.rm = TRUE),
                        by = 0.1)
)

# PREDICTIONS
preds <- predict_partial(m_soil, newdata = newdata, se.fit = TRUE)

preds <- bind_cols(newdata, preds) %>%
  mutate(
    effect = (exp(fit) - 1),
    low = (exp(fit - 1.96 * se.fit) - 1),
    hi = (exp(fit + 1.96 * se.fit) - 1),
    group = "Soil quality (0-8 scale)"
  )

# PLOT
library(ggplot2)
library(hrbrthemes)
library(scales)

p2 <- ggplot(preds, aes(x = aptitud_imputed, y = effect, ymin = low, ymax = hi)) +
  geom_ribbon(alpha = .3, color = "#1b9e77", fill = "#1b9e77") +
  geom_line(size = 1.5, color = "#1b9e77") +
  labs(
    x = "Soil Quality (aptitud)",
    y = "Marginal Effect of Palm Oil Adoption\n(% change in displacement rate)"
  ) +
  theme_ipsum(base_family = "sans") +
  scale_y_continuous(labels = percent_format(accuracy = 1)) +
  facet_wrap(vars(group)) +
  theme(panel.grid.minor = element_blank()) +
  geom_hline(yintercept = 0, lty = 2)
```

```
p2


# Test for pre-trends 1 -- joint F-test:
fect_out <- fect::fect(
  sdisp_rate ~ sdisp_rate ~  palm_presence +  palm_presence*palm_soil_interaction + farc_attack + coca,
  data = df_merged,
  index = c("cod_dane", "year"),
  method = "ife",
  se = TRUE,
  nboots = 200
)

# Placebo period tests:
fect_out_p <- fect::fect(sdisp_rate ~ palm_presence + palm_soil_interaction + farc_attack + coca, data =
saveRDS(fect_out_p, file = "fect_model_p.rds")

saveRDS(fect_out, file = "fect_model.rds")

fect_out <- readRDS("fect_model.rds")
fect_out_p <- readRDS("fect_model_p.rds")


plot(fect_out_p, stats = c("placebo.p","equiv.p"))

df_merged <- df_merged %>%
  group_by(cod_dane) %>%
  mutate(
    first_treat_year = if (any(palm_presence == 1)) {
      min(year[palm_presence == 1])
    } else {
      NA_real_
    }
  ) %>%
  ungroup() %>%
  mutate(
    G = if_else(is.na(first_treat_year), 0, first_treat_year)
  )


# This gives  the year each municipality was first treated, or 0 if never treated.
library(did)

# Filter only necessary variables
df_placebo <- df_merged %>%
  dplyr::select(cod_dane, year, G, elecviol, coca) %>%
  filter(!is.na(elecviol), !is.na(coca))

#required format by package
df_placebo <- df_placebo %>%
  mutate(cod_dane = as.numeric(cod_dane),
         coca = as.numeric(coca))
```

```r
did_proc_data <- did::pre_process_did(
  yname  = "elecviol",
  tname  = "year",
  idname = "cod_dane",
  gname  = "G",
  xformla = ~ coca,  # <- include coca here so it's retained
  allow_unbalanced_panel = TRUE,
  data = df_placebo
)


reg_placebo <- did::att_gt(
  yname = "elecviol",
  tname = "year",
  idname = "cod_dane",
  gname = "G",
  control_group = "nevertreated",
  est_method = "dr",
  base_period = "universal",
  xformla = ~ coca,
  allow_unbalanced_panel = TRUE,
  data = did_proc_data[['data']]
)

p1 <- ggdid(reg_placebo) + ggtitle("Political Violence")
fraude <- df_merged %>%
  filter(fraude == 1 | elecviol == 0)


# Filter only necessary variables
df_placebo <- df_merged %>%
  dplyr::select(cod_dane, year, G, fraude, coca) %>%
  filter(!is.na(coca), !is.na(fraude))

#required format by package
df_placebo <- df_placebo %>%
  mutate(cod_dane = as.numeric(cod_dane),
         coca = as.numeric(coca))

did_proc_data <- did::pre_process_did(
  yname  = "fraude",
  tname  = "year",
  idname = "cod_dane",
  gname  = "G",
  xformla = ~ coca,  # <- include coca here so it's retained
  allow_unbalanced_panel = TRUE,
  data = df_placebo
)


reg_placebo <- did::att_gt(
  yname = "fraude",
  tname = "year",
```

```r
  idname = "cod_dane",
  gname = "G",
  control_group = "nevertreated",
  est_method = "dr",
  base_period = "universal",
  xformla = ~ coca,
  allow_unbalanced_panel = TRUE,
  data = did_proc_data[['data']]
)

p2 <- ggdid(reg_placebo) + ggtitle("Political Fraud")

library(patchwork)

print((p1 + p2) & theme(plot.title = element_text(hjust = 0.5)))

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