Disrupting Compliance: The Impact of a Randomized Tax Holiday in Uruguay

2023-05-06

```
###############################################
message("Basic setup")
```

```
## Basic setup
```

```
rm(list=ls())
set.seed(1234)
options(scipen=999, digits=5)

###############################################
message("required libraries and setwd")
```

```
## required libraries and setwd
```

```
# Load/install packages --
if (!require("pacman")) install.packages("pacman")
```

```
## Loading required package: pacman
```

```
pacman::p_load(
  plyr,
  ggplot2,
  reshape2,
  zoo,
  sandwich,
  AER,
  xtable,
  stats,
  tidyr,
  dplyr,
  weights,
  estimatr,
  forcats,
  sf,
  ggpubr
)

#Set wd ---
home <-dirname(rstudioapi::getActiveDocumentContext()$path)
setwd(home)
```

```
##############################################
message("load data, functions, etc.")
```

load data, functions, etc.

```
load("panel_taxtime.Rda")
load("panel_goodtaxpayer.Rda")
load("naturalex_debt_gtp.Rda")
load("fieldex_data.Rda")
load("survey_data.Rda")

source("t_test.R")
source("test_diff.R")


#for maps

#Data for "Field Experiment: Geographic Distribution of Eligible and Ineligible Taxpayers"
datafx_map<-read_sf("v_mdg_parcelas_field_PROP_BP.shp")
shp<-read_sf("Marco2011_SEG_Montevideo_Total.shp")

#Data for "Natural Experiment: Property Plots of Winning Account Numbers"
naturalex_data_map<-read_sf("naturalex_map.shp")
shp2<-read_sf("v_mdg_manzanas.shp")

#Data for Reward Programs for Good Taxpayers in Latin America
latam_map_data<-read_sf("latinamerica.shp")


##############################################
message("Clean & add new variables")
```

Clean & add new variables

```
#cross section natural experiment data
naturalex<-buen_pagador_panel[buen_pagador_panel$centered==0,]

# New variables for the natural experiment data
taxes_panel$YEARMON_LOTT <- as.yearmon(taxes_panel$FECHA_SORTEO2) #year & month of the lottery
taxes_panel$YEAR_LOTT <- as.numeric(format(taxes_panel$YEARMON_LOTT, "%Y")) # year of the lottery
taxes_panel$missed_payment <- as.numeric(taxes_panel$en_fecha==0) #missed payment dummy
taxes_panel$compliance <- as.numeric(taxes_panel$nr_paymntsowed==0) #compliance dummy

# save data in the control group
control<-taxes_panel %>% filter(TREATMENT==0)

# add tax names in english
taxes_panel$tax <- taxes_panel$TRIBUTO
taxes_panel$tax <- as.factor(taxes_panel$tax)
levels(taxes_panel$tax) <-  c("Property", "Vehicle", "Sewage", "Head")

# dummy if it is a Vehicle tax o other tax
```

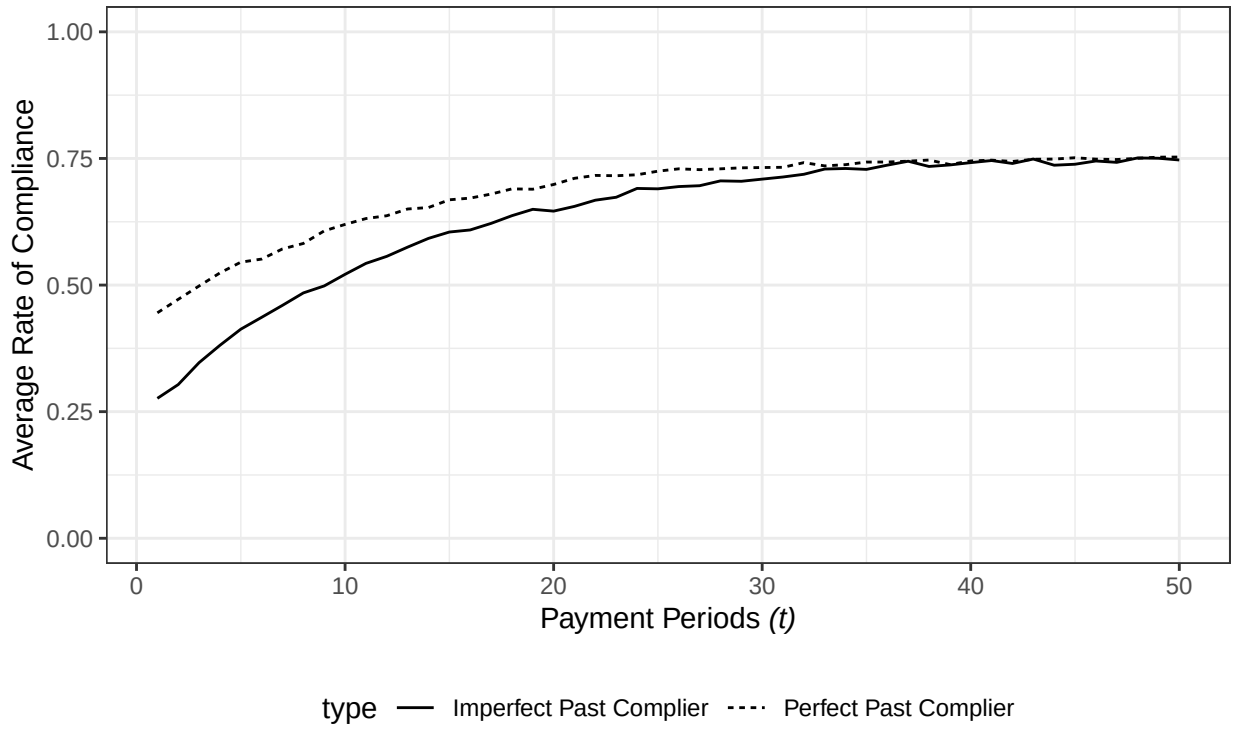
```
taxes_panel$holiday_type <- 1
taxes_panel$holiday_type[taxes_panel$TRIBUTO=="Patente de Rodados"] <- 0

# st & t_st variables (time standardized by payments of relevant lottery) ---
holiday <- taxes_panel[taxes_panel$cuota_exonerada==1,] #treated subset

#Contribucion Inmobiliaria
CI <- c(min(holiday$t[holiday$TRIBUTO=="Contribucion Inmobiliaria"],na.rm = T),
        max(holiday$t[holiday$TRIBUTO=="Contribucion Inmobiliaria"],na.rm = T))

taxes_panel <- taxes_panel[!(taxes_panel$TRIBUTO=="Contribucion Inmobiliaria" &
                               taxes_panel$ES_BP==1 &
                               taxes_panel$t>=CI[1] & taxes_panel$t<=CI[2]),]
taxes_panel$st[taxes_panel$TRIBUTO=="Contribucion Inmobiliaria"] <- CI[2] - 3
taxes_panel <- taxes_panel[!(taxes_panel$TRIBUTO=="Contribucion Inmobiliaria" &
                               taxes_panel$ES_BP==0 &
                               taxes_panel$t>=4 & taxes_panel$t<=CI[2]),]

#Patente de Rodados
PR <- c(min(holiday$t[holiday$TRIBUTO=="Patente de Rodados"],na.rm = T),
        max(holiday$t[holiday$TRIBUTO=="Patente de Rodados"],na.rm = T))
taxes_panel <- taxes_panel[!(taxes_panel$TRIBUTO=="Patente de Rodados" &
                               taxes_panel$ES_BP==1 &
                               taxes_panel$t>=PR[1] & taxes_panel$t<=PR[2]),]
taxes_panel$st[taxes_panel$TRIBUTO=="Patente de Rodados"] <- 0

#Saneamiento
TS <- c(min(holiday$t[holiday$TRIBUTO=="Saneamiento"],na.rm = T),
        max(holiday$t[holiday$TRIBUTO=="Saneamiento"],na.rm = T))
taxes_panel <- taxes_panel[!(taxes_panel$TRIBUTO=="Saneamiento" &
                               taxes_panel$ES_BP==1 &
                               taxes_panel$t>=TS[1] & taxes_panel$t<=TS[2]),]
taxes_panel$st[taxes_panel$TRIBUTO=="Saneamiento"] <- TS[2] - 3
taxes_panel <- taxes_panel[!(taxes_panel$TRIBUTO=="Saneamiento" &
                               taxes_panel$ES_BP==0 &
                               taxes_panel$t>=4 & taxes_panel$t<=TS[2]),]

#Tributos Domiciliarios
TD <- c(1,max(holiday$t[holiday$TRIBUTO=="Tributos Domiciliarios"],na.rm = T))
taxes_panel <- taxes_panel[!(taxes_panel$TRIBUTO=="Tributos Domiciliarios" &
                               taxes_panel$ES_BP==1 &
                               taxes_panel$t>=TD[1] & taxes_panel$t<=TD[2]),]
taxes_panel$st[taxes_panel$TRIBUTO=="Tributos Domiciliarios"] <- TD[2] - 3
taxes_panel <- taxes_panel[!(taxes_panel$TRIBUTO=="Tributos Domiciliarios" &
                               taxes_panel$ES_BP==0 &
                               taxes_panel$t>=4 & taxes_panel$t<=TD[2]),]

taxes_panel$t_st <- ifelse((taxes_panel$ES_BP==1 & taxes_panel$t > 0) |
                             (taxes_panel$ES_BP==0 & taxes_panel$t > 4),
                           taxes_panel$t - taxes_panel$st, taxes_panel$t)
rm(holiday, CI, TD, TS, PR)
```

```r
# add variables for the field experiment data

#treatment names variable
fieldex$tpooled <- NA
fieldex$tpooled[fieldex$treatment==6] <- "Control"
fieldex$tpooled[fieldex$treatment==0] <- "Reminder"
fieldex$tpooled[fieldex$treatment %in% c(1,2,4)] <- "Reminder+Info"


##############################################################################
# PAPER Tables & Figures
##############################################################################


###############################################
message("MAIN PAPER: Figure 1")
```

```
## MAIN PAPER: Figure 1
```

```r
set.seed(12345)

## SIMULATION
z <- 2 #tax payment
p <- 1 #probability of punishment
c <- 1 #cost of punishment
b <- .05 #intrinsic benefit of compliance

# function for the decision to comply in a single period
sim_function <- function(z, b, p, c, theta, lambda0){

  lambda <- NA
  for(i in 1:50){
    v <- rnorm(1) #random noise
    vector_theta <- (theta^((i+49):1))
    vector_lambda <- if(i==1){lambda0}else{c(lambda0, lambda)}
    lambda[i] <- ifelse((b + (p*c) - z + (1/5000)*z +
                           vector_theta %*% vector_lambda - v) > 0, 1, 0)}

  return(lambda)
}

# complete iteration including shock
sim_function1000 <- function(z, b, p,  c, theta, history, N, treat){
  sims <- matrix(NA, N, 50)
  for(i in 1:N){
    if (history=="perfect"){lambda0 <- rbinom(50,1,prob=1)}
    if (history=="marginal"){lambda0 <- rbinom(50,1,prob=.4)}
    if (treat==1) {lambda0[48:50] <- c(0, 0, 0)}
    sims[i, ] <- sim_function(z=z, b=b, p=p, c=c, theta=theta, lambda0=lambda0)}
  sims <- apply(sims, 2, mean)
  return(sims)
}
```

```r
paym <- rbind(
  cbind(sim_function1000(z=z, b=b, p=p, c=c, theta=0.7, history = "perfect",
                         treat=1, N=10000), "Perfect Past Complier"),
  cbind(sim_function1000(z=z, b=b, p=p, c=c, theta=0.7, history = "marginal",
                         treat=1, N=10000), "Imperfect Past Complier"))

paym <- as.data.frame(paym)
paym$t <- c(1:50, 1:50)
names(paym)[1:2] <- c("mean_payments","type")
paym$type <- as.factor(paym$type)
paym$mean_payments <- as.numeric(as.character(paym$mean_payments))

#Figure
ggplot(paym, aes(x=t, y=mean_payments, group=type)) +
  geom_line(aes(linetype = type)) +
  theme_bw() + ylim(0, 1) +
  ylab("Average Rate of Compliance") +
  xlab(expression(paste("Payment Periods ", italic("(t)")))) +
  labs(shape = "Taxpayer Type") +
  theme(legend.position = "bottom")
```



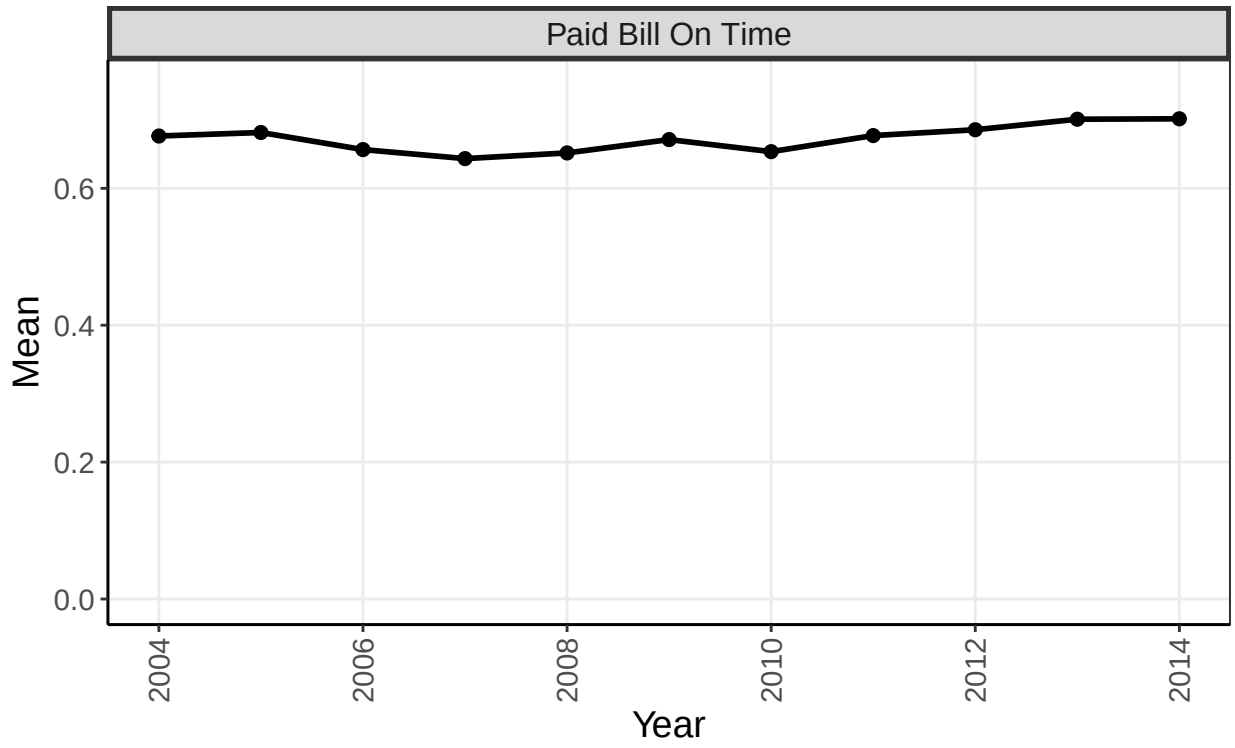

```r
##############################################
message("MAIN PAPER: Figure 2")
```

```
## MAIN PAPER: Figure 2
```

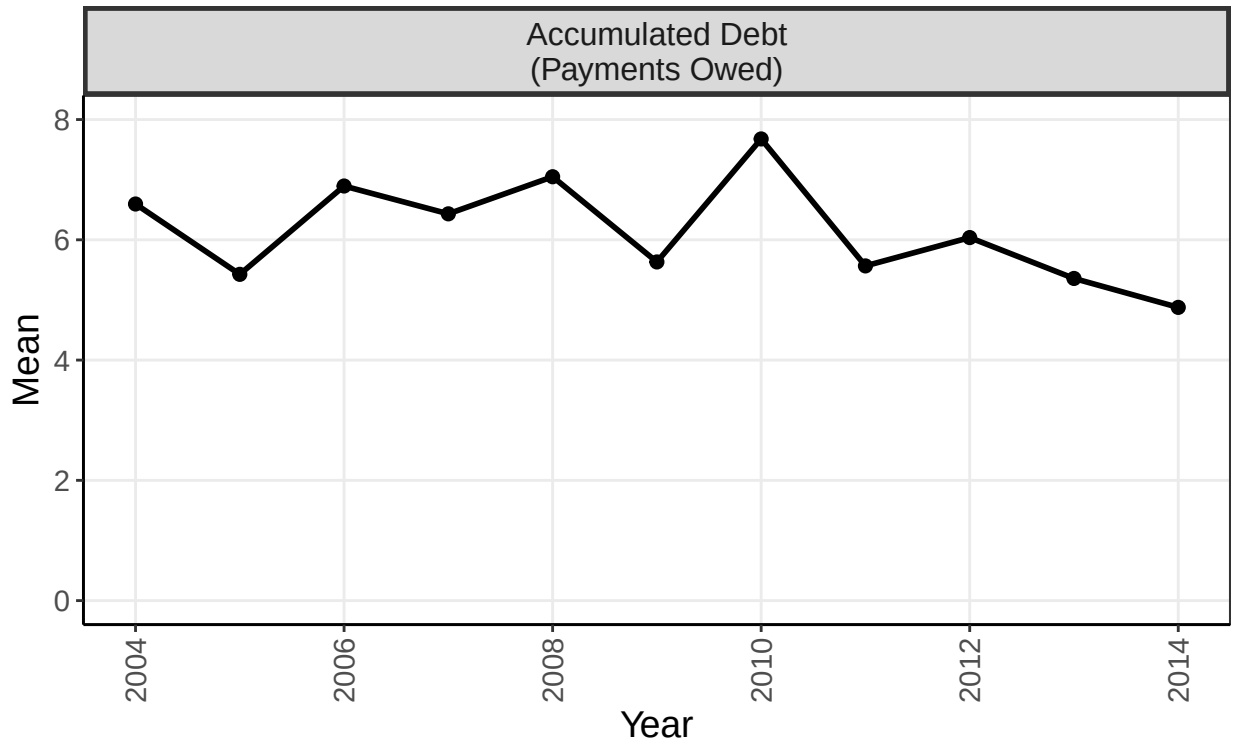
```r
# Paid on time figure
control %>% # We use only observations in the control group
  # group by year and take the mean of paid on time
  group_by(YEAR_LOTT) %>%
  dplyr::summarise(
    mean = mean(en_fecha, na.rm=T),
    outcome = "Paid Bill On Time"
  ) %>%
  ggplot(aes(YEAR_LOTT, mean)) +
  facet_wrap(~ outcome, scales="free_y") +
  geom_point(size=2) +
  geom_line(size=1) +
  xlab("Year") +
  ylab("Mean") +
  theme_bw() + ylim(c(0,.75)) +
  scale_colour_manual(values = c("black")) +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text.y = element_text(size = rel(1.25)),
        axis.title.y = element_text(size = rel(1.3)),
        axis.title.x = element_text(size = rel(1.3)),
        legend.text = element_text(size = rel(1.2)),
        strip.text.x = element_text(size = rel(1.4)),
        strip.text.y = element_text(size = rel(1.4)),
        strip.background = element_rect(size = 1.5),
        axis.text.x = element_text(size = rel(1.25), angle = 90, vjust = 0.5, hjust=1),
        legend.position = "bottom",
        legend.title=element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(colour = "black"))
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

```
## Warning: The ‘size‘ argument of ‘element_rect()‘ is deprecated as of ggplot2 3.4.0.
## i Please use the ‘linewidth‘ argument instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

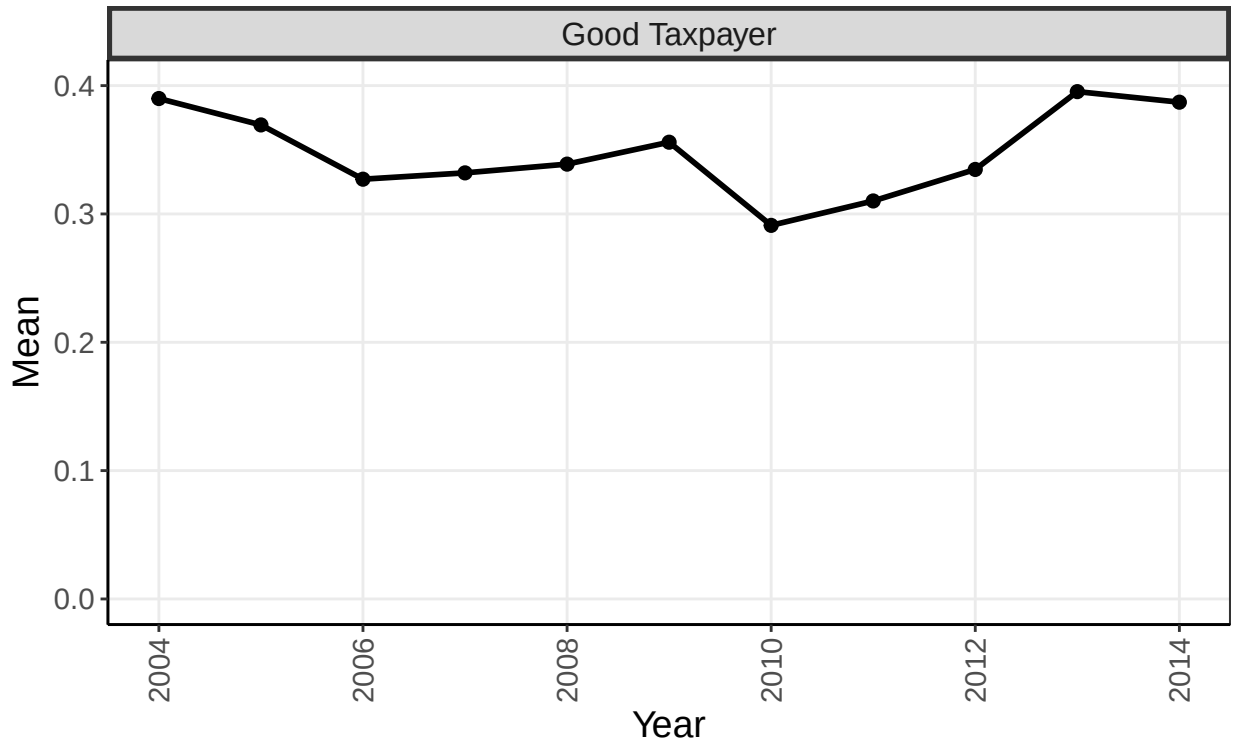



```r
# Accumulated Debt figure
control %>% # We use only observations in the control group
  # group by year and take the mean of accumulated debt
  group_by(YEAR_LOTT) %>%
  dplyr::summarise(
    mean = mean(nr_paymntsowed, na.rm=T),
    outcome = "Accumulated Debt\n(Payments Owed)"
  ) %>%
  ggplot(aes(YEAR_LOTT, mean)) +
  facet_wrap(~ outcome, scales="free_y") +
  geom_point(size=2) +
  geom_line(size=1) +
  xlab("Year") +
  ylab("Mean") +
  theme_bw() + ylim(c(0,8)) +
  scale_colour_manual(values = c("black")) +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text.y = element_text(size = rel(1.25)),
        axis.title.y = element_text(size = rel(1.3)),
        axis.title.x = element_text(size = rel(1.3)),
        legend.text = element_text(size = rel(1.2)),
        strip.text.x = element_text(size = rel(1.4)),
        strip.text.y = element_text(size = rel(1.4)),
        strip.background = element_rect(size = 1.5),
        axis.text.x = element_text(size = rel(1.25), angle = 90, vjust = 0.5, hjust=1),
        legend.position = "bottom",
        legend.title=element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(colour = "black"))
```




```
# Good taxpayer figure
control %>% # We use only observations in the control group
  filter(t==0) %>%  # We keep only observations for t=0
  # Share of good taxpayer by year
  group_by(YEAR_LOTT) %>%
  dplyr::summarise(
    mean = mean(ES_BP, na.rm=T),
    outcome = "Good Taxpayer"
  ) %>%
  ggplot(aes(YEAR_LOTT, mean)) +
  facet_wrap(~ outcome, scales="free_y") +
  geom_point(size=2) +
  geom_line(size=1) +
  xlab("Year") +
  ylab("Mean") +
  theme_bw() + ylim(c(0,.4)) +
  scale_colour_manual(values = c("black")) +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text.y = element_text(size = rel(1.25)),
        axis.title.y = element_text(size = rel(1.3)),
        axis.title.x = element_text(size = rel(1.3)),
        legend.text = element_text(size = rel(1.2)),
        strip.text.x = element_text(size = rel(1.4)),
        strip.text.y = element_text(size = rel(1.4)),
        strip.background = element_rect(size = 1.5),
        axis.text.x = element_text(size = rel(1.25), angle = 90, vjust = 0.5, hjust=1),
        legend.position = "bottom",
        legend.title=element_blank(),
        panel.grid.minor = element_blank(),
```

```
        axis.line = element_line(colour = "black"))
```




```
################################################
message("MAIN PAPER: Table 3.1")
```

```
## MAIN PAPER: Table 3.1
```

```
# Natural Experiment: Sample Sizes

# Sample sizes by Tax & Taxpayer Type
naturalex %>% # keep only the cross-section of observations at the time of treatment assignment
  dplyr::group_by(ES_BP, TRIBUTO) %>%
  dplyr::summarize(
    non_winning_accounts = sum(TREATMENT==0),
    winning_accounts = sum(TREATMENT==1),
    study_group = n()
  )
```

```
## `summarise()` has grouped output by 'ES_BP'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 8 x 5
## # Groups:   ES_BP [2]
##   ES_BP TRIBUTO                non_winning_accounts winning_accounts study_group
##   <int> <fct>                                 <int>            <int>       <int>
## 1     0 Contribucion Inmobili~                 1225             1211        2436
## 2     0 Patente de Rodados                     1924             1899        3823
```

```
## 3     0 Saneamiento                          939             915        1854
## 4     0 Tributos Domiciliarios              2062            2083        4145
## 5     1 Contribucion Inmobili~              1354            1339        2693
## 6     1 Patente de Rodados                   375             391         766
## 7     1 Saneamiento                          404             452         856
## 8     1 Tributos Domiciliarios              1041            1007        2048
```

```
# Sample sizes by Taxpayer Type
naturalex %>% # keep only the cross-section of observations at the time of treatment assignment
  dplyr::group_by(ES_BP) %>%
  dplyr::summarize(
    non_winning_accounts = sum(TREATMENT==0),
    winning_accounts = sum(TREATMENT==1),
    study_group = n()
  )
```

```
## # A tibble: 2 x 4
##   ES_BP non_winning_accounts winning_accounts study_group
##   <int>                <int>            <int>       <int>
## 1     0                 6150             6108       12258
## 2     1                 3174             3189        6363
```

```
##############################################
message("MAIN PAPER: Table 3.2")
```

```
## MAIN PAPER: Table 3.2
```

```
# table by treatment and taxpayer type
table(fieldex$tpooled, fieldex$type)
```

```
##
##                 good taxpayer bad taxpayer
##   Control                7243         3412
##   Reminder               1532         2080
##   Reminder+Info          3037         4150
```

```
# table by taxpayer type
table(fieldex$type)
```

```
##
## good taxpayer  bad taxpayer
##         14784         13862
```

```
##############################################
message("MAIN PAPER: Figure 3")
```

```
## MAIN PAPER: Figure 3
```

```r
# The Negative Impact of Holidays on Compliance

table(taxes_panel$TREATMENT, taxes_panel$TIENE_EXO)
```

```
##
##          0      1
##   0 550896      0
##   1 421618 128276
```

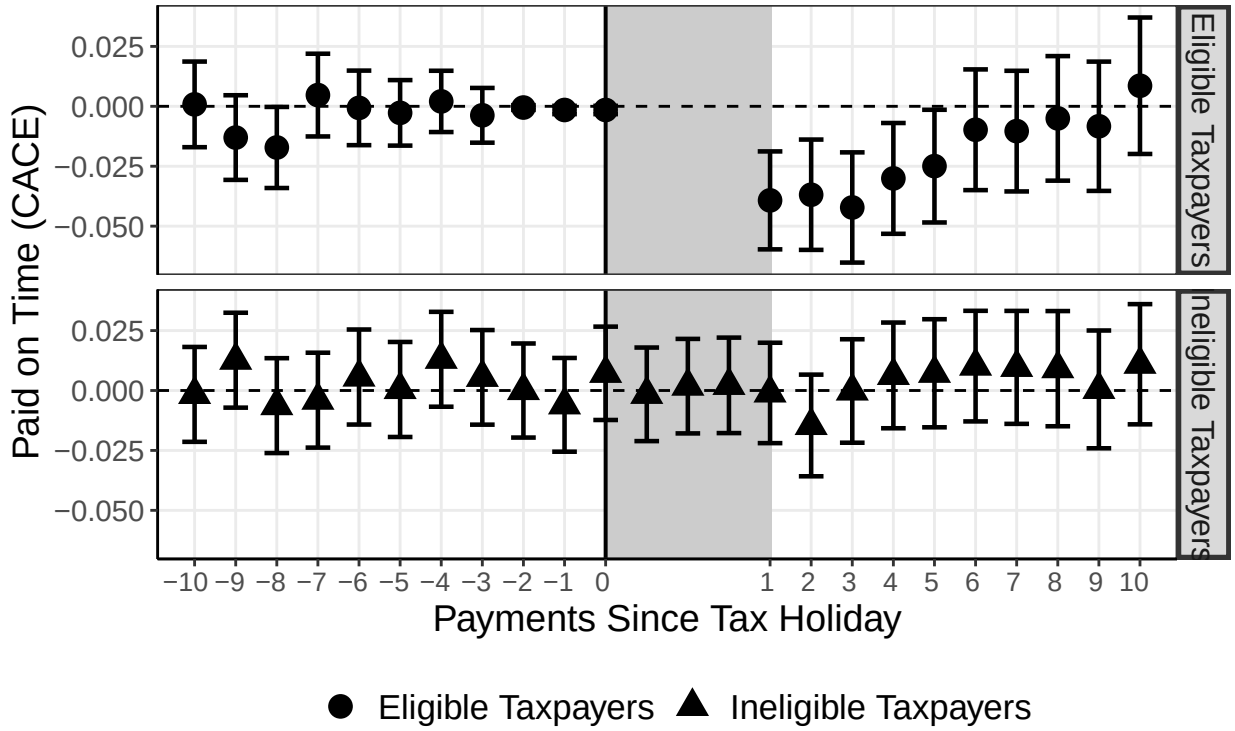
```r
t <- -10:13
btp <- NULL
gtp <- NULL

for (i in t){

  temp <- taxes_panel[taxes_panel$t_st == i, ]

  # For Ineligible Taxpayers
  btp_est <- tidy(lm_robust(en_fecha ~ TREATMENT,
                            data = filter(temp, ES_BP==0)))[2,]
  btp <- rbind.data.frame(btp, cbind.data.frame(i, "Ineligible Taxpayers", btp_est))

  # For Eligible Taxpayers
  # For good taxpayers, skip periods under the tax holiday
  if (nrow(temp[temp$TREATMENT==1 & temp$ES_BP==1,])==0) next

  gtp_est <- tidy(iv_robust(en_fecha ~ TIENE_EXO | TREATMENT,
                            data = filter(temp, ES_BP==1)))[2,]
  gtp <- rbind.data.frame(gtp, cbind.data.frame(i, "Eligible Taxpayers", gtp_est))

}
```

```
## Warning in sqrt(diag(vcov_fit$Vcov_hat)): NaNs produced
```

```r
names(gtp)[2] <- names(btp)[2] <- "type"
plot <- rbind.data.frame(gtp, btp); rm(gtp, btp)


#Figure
ggplot(plot, aes(x=i, y=estimate, shape=type)) +
  facet_grid(type~.) +
  geom_rect(data=NULL,aes(xmin=0, xmax=4, ymin=-Inf, ymax=Inf),
            fill="gray80", color="gray80") +
  geom_errorbar(aes(x=i,
                    ymin=conf.low,
                    ymax=conf.high),
                width=.6, size=.8, position = position_dodge(width = 0.6)) +
  geom_point(size=4, position = position_dodge(width = 0.6)) +
  xlab("Payments Since Tax Holiday") +
  ylab("Paid on Time (CACE)") +
  geom_vline(aes(xintercept=0), size=.7) +
  geom_hline(aes(yintercept=0), size=.5, linetype="dashed") +
```

```
  theme_bw() +
  scale_x_discrete(limit = t[!t %in% c(1,2,3)],
                   labels = as.character(c(t[t<1],
                                           t[!t%in%c(1,2,3) & t>0]-3))) +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text.y = element_text(size = rel(1.25)),
        axis.title.y = element_text(size = rel(1.3)),
        axis.title.x = element_text(size = rel(1.3)),
        legend.text = element_text(size = rel(1.2)),
        strip.text.x = element_text(size = rel(1.4)),
        strip.text.y = element_text(size = rel(1.4)),
        strip.background = element_rect(size = 1.5),
        axis.text.x = element_text(size = rel(1.1), hjust=.7),
        legend.position = "bottom",
        legend.title=element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(colour = "black"))
```

```
## Warning: Continuous limits supplied to discrete scale.
## i Did you mean 'limits = factor(...)' or 'scale_*_continuous()'?
```




```
##############################################
message("MAIN PAPER: Table 4.3")
```

```
## MAIN PAPER: Table 4.3
```

```
# Estimated Causal Effects of Tax Holidays

# estimate for 1st period after holiday (t_st==4)
t1_ace <- tidy(lm_robust(en_fecha ~ TREATMENT,
                         data = filter(taxes_panel, ES_BP==1 & t_st==4)))
t1_cace <- tidy(iv_robust(en_fecha ~ TIENE_EXO | TREATMENT,
                          data = filter(taxes_panel, ES_BP==1 & t_st==4)))
t1 <- c(t1_ace[1,2], t1_ace[2,2], t1_ace[2,3], t1_ace[2,5], t1_cace[2,2],t1_cace[2,3],t1_cace[2,5]);
rm(t1_ace, t1_cace)

# estimate for 5th period after holiday (t_st==8)
t5_ace <- tidy(lm_robust(en_fecha ~ TREATMENT,
                         data = filter(taxes_panel, ES_BP==1 & t_st==8)))
t5_cace <- tidy(iv_robust(en_fecha ~ TIENE_EXO | TREATMENT,
                          data = filter(taxes_panel, ES_BP==1 & t_st==8)))
t5 <- c(t5_ace[1,2], t5_ace[2,2], t5_ace[2,3], t5_ace[2,5], t5_cace[2,2],t5_cace[2,3],t5_cace[2,5]);
rm(t5_ace, t5_cace)

# estimate for 10th period after holiday (t_st==13)
t10_ace <- tidy(lm_robust(en_fecha ~ TREATMENT,
                          data = filter(taxes_panel, ES_BP==1 & t_st==13)))
t10_cace <- tidy(iv_robust(en_fecha ~ TIENE_EXO | TREATMENT,
                           data = filter(taxes_panel, ES_BP==1 & t_st==13)))
t10 <- c(t10_ace[1,2], t10_ace[2,2], t10_ace[2,3], t10_ace[2,5], t10_cace[2,2],t10_cace[2,3],t10_cace[2
rm(t10_ace, t10_cace)

# estimate for periods 1-10 after holiday (t_st==4-13)
t1_10_ace <- tidy(lm_robust(en_fecha ~ TREATMENT,
                            data = filter(taxes_panel, ES_BP==1 & t_st>3 & t_st<14)))
t1_10_cace <- tidy(iv_robust(en_fecha ~ TIENE_EXO | TREATMENT,
                             data = filter(taxes_panel, ES_BP==1 & t_st>3 & t_st<14)))
t1_10 <- c(t1_10_ace[1,2], t1_10_ace[2,2], t1_10_ace[2,3], t1_10_ace[2,5],
           t1_10_cace[2,2],t1_10_cace[2,3],t1_10_cace[2,5]);
rm(t1_10_ace, t1_10_cace)


# estimate for periods 1-10 after holiday (t_st==4-13) PROPERTY TAX
t1_10_ace <- tidy(lm_robust(en_fecha ~ TREATMENT,
                            data = filter(taxes_panel, ES_BP==1 & t_st>3 & t_st<14 &
                                            TRIBUTO == "Contribucion Inmobiliaria")))
t1_10_cace <- tidy(iv_robust(en_fecha ~ TIENE_EXO | TREATMENT,
                             data = filter(taxes_panel, ES_BP==1 & t_st>3 & t_st<14 &
                                             TRIBUTO == "Contribucion Inmobiliaria")))
t1_10_property <- c(t1_10_ace[1,2], t1_10_ace[2,2], t1_10_ace[2,3], t1_10_ace[2,5],t1_10_cace[2,2],t1_1
rm(t1_10_ace, t1_10_cace)


# estimate for periods 1-10 after holiday (t_st==4-13) HEAD
t1_10_ace <- tidy(lm_robust(en_fecha ~ TREATMENT,
                            data = filter(taxes_panel, ES_BP==1 & t_st>3 & t_st<14 &
                                            TRIBUTO == "Tributos Domiciliarios")))
t1_10_cace <- tidy(iv_robust(en_fecha ~ TIENE_EXO | TREATMENT,
                             data = filter(taxes_panel, ES_BP==1 & t_st>3 & t_st<14 &
```

```
                                              TRIBUTO == "Tributos Domiciliarios")))
t1_10_head <- c(t1_10_ace[1,2], t1_10_ace[2,2], t1_10_ace[2,3], t1_10_ace[2,5], t1_10_cace[2,2],t1_10_ca
rm(t1_10_ace, t1_10_cace)


# estimate for periods 1-10 after holiday (t_st==4-13) SEWAGE
t1_10_ace <- tidy(lm_robust(en_fecha ~ TREATMENT,
                            data = filter(taxes_panel, ES_BP==1 & t_st>3 & t_st<14 &
                                              TRIBUTO == "Saneamiento")))
t1_10_cace <- tidy(iv_robust(en_fecha ~ TIENE_EXO | TREATMENT,
                             data = filter(taxes_panel, ES_BP==1 & t_st>3 & t_st<14 &
                                              TRIBUTO == "Saneamiento")))
t1_10_sewage <- c(t1_10_ace[1,2], t1_10_ace[2,2], t1_10_ace[2,3], t1_10_ace[2,5],
                  t1_10_cace[2,2],t1_10_cace[2,3],t1_10_cace[2,5]);
rm(t1_10_ace, t1_10_cace)


# estimate for periods 1-10 after holiday (t_st==4-13) VEHICLE
t1_10_ace <- tidy(lm_robust(en_fecha ~ TREATMENT,
                            data = filter(taxes_panel, ES_BP==1 & t_st>3 & t_st<14 &
                                              TRIBUTO == "Patente de Rodados")))
t1_10_cace <- tidy(iv_robust(en_fecha ~ TIENE_EXO | TREATMENT,
                             data = filter(taxes_panel, ES_BP==1 & t_st>3 & t_st<14 &
                                              TRIBUTO == "Patente de Rodados")))
t1_10_vehicle <- c(t1_10_ace[1,2], t1_10_ace[2,2], t1_10_ace[2,3], t1_10_ace[2,5],
                   t1_10_cace[2,2],t1_10_cace[2,3],t1_10_cace[2,5]);
rm(t1_10_ace, t1_10_cace)



# Build table
table4.3 <- rbind.data.frame(t1,t5,t10,t1_10, t1_10_property, t1_10_head,
                             t1_10_sewage, t1_10_vehicle);
rm(t1,t5,t10,t1_10, t1_10_property, t1_10_head, t1_10_sewage, t1_10_vehicle)

rownames(table4.3) <- c("Post Tax Holiday Payment 1",
                        "Post Tax Holiday Payment 5",
                        "Post Tax Holiday Payment 10",
                        "Post Tax Holiday Payments 1-10",
                        "Post Tax Holiday Payments 1-10 (Property)",
                        "Post Tax Holiday Payments 1-10 (Head)",
                        "Post Tax Holiday Payments 1-10 (Sewage)",
                        "Post Tax Holiday Payments 1-10 (Vehicle)")

colnames(table4.3) <- c("Control Mean", "ACE", "SE_ACE", "p-value", "CACE","SE_CACE", "p-value")


# Table
round(table4.3,digits = 2)
```

```
##                                        Control Mean   ACE SE_ACE p-value
## Post Tax Holiday Payment 1                     0.93 -0.03   0.01    0.00
## Post Tax Holiday Payment 5                     0.91 -0.02   0.01    0.04
```

```
## Post Tax Holiday Payment 10                       0.87  0.01   0.01    0.55
## Post Tax Holiday Payments 1-10                    0.90 -0.02   0.00    0.00
## Post Tax Holiday Payments 1-10 (Property)         0.92 -0.02   0.00    0.00
## Post Tax Holiday Payments 1-10 (Head)             0.91 -0.01   0.00    0.01
## Post Tax Holiday Payments 1-10 (Sewage)           0.94 -0.02   0.01    0.00
## Post Tax Holiday Payments 1-10 (Vehicle)          0.75  0.00   0.01    0.80
##                                                  CACE SE_CACE p-value
## Post Tax Holiday Payment 1                      -0.04    0.01    0.00
## Post Tax Holiday Payment 5                      -0.02    0.01    0.04
## Post Tax Holiday Payment 10                      0.01    0.01    0.55
## Post Tax Holiday Payments 1-10                  -0.02    0.00    0.00
## Post Tax Holiday Payments 1-10 (Property) -0.03    0.01    0.00
## Post Tax Holiday Payments 1-10 (Head)     -0.02    0.01    0.01
## Post Tax Holiday Payments 1-10 (Sewage)   -0.03    0.01    0.00
## Post Tax Holiday Payments 1-10 (Vehicle)   0.00    0.01    0.80
```

```
##############################################
message("MAIN PAPER: Figure 4")
```

```
## MAIN PAPER: Figure 4
```

```
# Treatment Effects By Type of Tax: Holiday vs. No Holiday


gtp_taxes <- taxes_panel[taxes_panel$ES_BP==1, ]
t <- unique(gtp_taxes$t_st)
t <- t[order(t)]
t <- t[t>-11 & t<=21]

gtp_plot <- NULL

#No Holiday
for (i in 1:length(t)){

  temp <- gtp_taxes[gtp_taxes$t_st == t[i], ]

  if (nrow(temp[temp$TREATMENT==1 & temp$holiday_type==0,])==0) next
  if (nrow(temp[temp$TREATMENT==0 & temp$holiday_type==0,])==0) next

  on_time_noholiday <- robust.se(ivreg(en_fecha ~ TIENE_EXO, ~ TREATMENT,
                                       data=temp[temp$holiday_type==0,]))[2,1:2]

  gtp_plot <- rbind(gtp_plot,
                    as.vector(c("Paid on Time", t[i], on_time_noholiday, "No Holiday")))

  print(i)
}
```

```
## [1] "Robust Standard Errors"
## [1] 1
## [1] "Robust Standard Errors"
## [1] 2
## [1] "Robust Standard Errors"
```

```
## [1] 3
## [1] "Robust Standard Errors"
## [1] 4
## [1] "Robust Standard Errors"
## [1] 5
## [1] "Robust Standard Errors"
## [1] 6
## [1] "Robust Standard Errors"
## [1] 7
## [1] "Robust Standard Errors"
## [1] 8
## [1] "Robust Standard Errors"
## [1] 12
## [1] "Robust Standard Errors"
## [1] 13
## [1] "Robust Standard Errors"
## [1] 14
## [1] "Robust Standard Errors"
## [1] 15
## [1] "Robust Standard Errors"
## [1] 16
## [1] "Robust Standard Errors"
## [1] 17
## [1] "Robust Standard Errors"
## [1] 18
## [1] "Robust Standard Errors"
## [1] 19
## [1] "Robust Standard Errors"
## [1] 20
## [1] "Robust Standard Errors"
## [1] 21
## [1] "Robust Standard Errors"
## [1] 22
## [1] "Robust Standard Errors"
## [1] 23
## [1] "Robust Standard Errors"
## [1] 24
## [1] "Robust Standard Errors"
## [1] 25
## [1] "Robust Standard Errors"
## [1] 26
## [1] "Robust Standard Errors"
## [1] 27
## [1] "Robust Standard Errors"
## [1] 28
## [1] "Robust Standard Errors"
## [1] 29
```

```
#Holiday
for (i in 1:length(t)){

  temp <- gtp_taxes[gtp_taxes$t_st == t[i], ]

  if (nrow(temp[temp$TREATMENT==1 & temp$holiday_type==1,])==0) next
```

```
  if (nrow(temp[temp$TREATMENT==0 & temp$holiday_type==1,])==0) next

  on_time_holiday <- robust.se(ivreg(en_fecha ~ TIENE_EXO, ~ TREATMENT,
                                     data=temp[temp$holiday_type==1,]))[2,1:2]

  gtp_plot <- rbind(gtp_plot,
                    as.vector(c("Paid on Time", t[i], on_time_holiday, "Holiday")))
  print(i)
}
```

```
## [1] "Robust Standard Errors"
## [1] 1
## [1] "Robust Standard Errors"
## [1] 2
## [1] "Robust Standard Errors"
## [1] 3
## [1] "Robust Standard Errors"
## [1] 4
## [1] "Robust Standard Errors"
## [1] 5
## [1] "Robust Standard Errors"
## [1] 6
## [1] "Robust Standard Errors"
## [1] 7
## [1] "Robust Standard Errors"
## [1] 8
## [1] "Robust Standard Errors"
## [1] 9
## [1] "Robust Standard Errors"
## [1] 10
## [1] "Robust Standard Errors"
```

```
## Warning in sqrt(diag(se)): NaNs produced
```

```
## [1] 11
## [1] "Robust Standard Errors"
## [1] 12
## [1] "Robust Standard Errors"
## [1] 13
## [1] "Robust Standard Errors"
## [1] 14
## [1] "Robust Standard Errors"
## [1] 15
## [1] "Robust Standard Errors"
## [1] 16
## [1] "Robust Standard Errors"
## [1] 17
## [1] "Robust Standard Errors"
## [1] 18
## [1] "Robust Standard Errors"
## [1] 19
## [1] "Robust Standard Errors"
## [1] 20
```

```
## [1] "Robust Standard Errors"
## [1] 21
## [1] "Robust Standard Errors"
## [1] 22
## [1] "Robust Standard Errors"
## [1] 23
## [1] "Robust Standard Errors"
## [1] 24
## [1] "Robust Standard Errors"
## [1] 25
## [1] "Robust Standard Errors"
## [1] 26
## [1] "Robust Standard Errors"
## [1] 27
## [1] "Robust Standard Errors"
## [1] 28
## [1] "Robust Standard Errors"
## [1] 29
```

```r
gtp_plot <- as.data.frame(gtp_plot)
names(gtp_plot) <- c("outcome", "t", "CACE", "SE", "lottery_type")

gtp_plot$t <- as.numeric(as.character(gtp_plot$t))
gtp_plot$CACE <- as.numeric(as.character(gtp_plot$CACE))
gtp_plot$SE <- as.numeric(as.character(gtp_plot$SE))


#Difference
diff.hte <- NULL

for (i in 1:length(t)){

  temp <- gtp_plot[gtp_plot$t==t[i],]

  if (nrow(temp)!=2) next

  diff <- temp$CACE[temp$lottery_type=="Holiday"] - temp$CACE[temp$lottery_type=="No Holiday"]

  SE <- sqrt((temp$SE[temp$lottery_type=="Holiday"])^2 + (temp$CACE[temp$lottery_type=="No Holiday"])^2

  diff.hte <- rbind.data.frame(diff.hte, as.vector(c(t[i], diff, SE)))

  print(i)
}
```

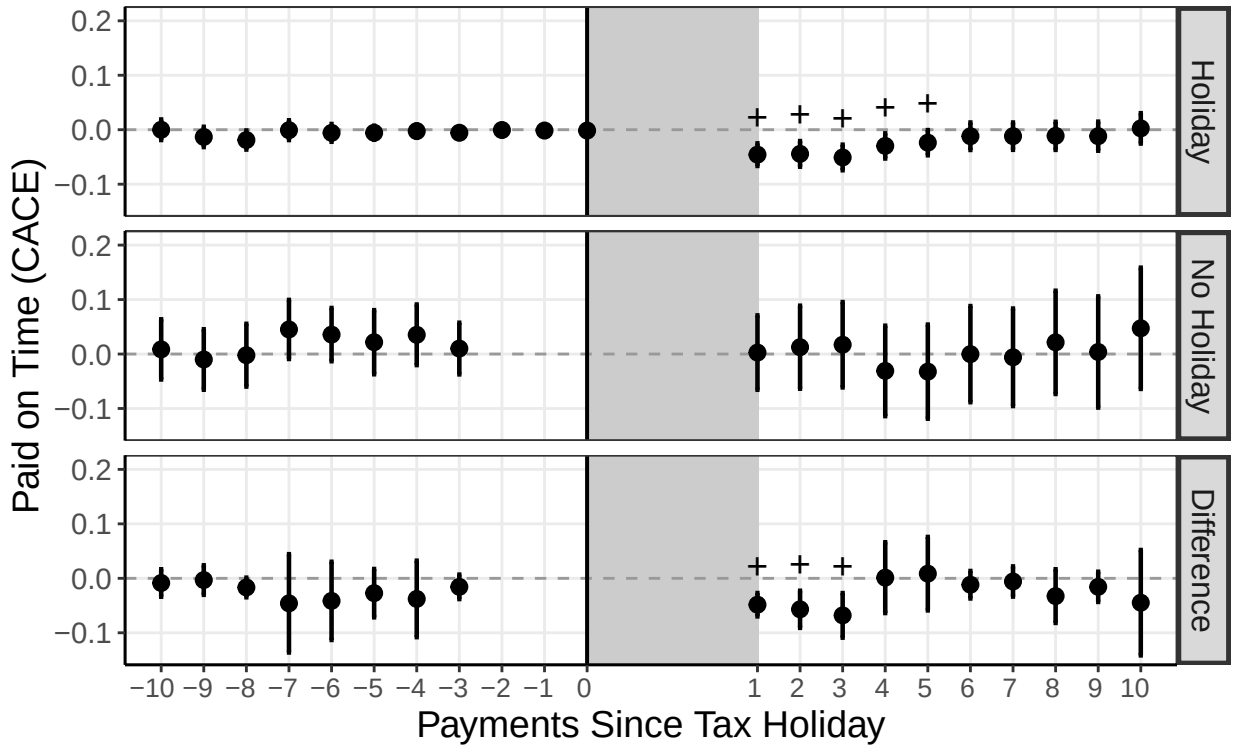
```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
## [1] 12
```

```
## [1] 13
## [1] 14
## [1] 15
## [1] 16
## [1] 17
## [1] 18
## [1] 19
## [1] 20
## [1] 21
## [1] 22
## [1] 23
## [1] 24
## [1] 25
## [1] 26
## [1] 27
## [1] 28
## [1] 29
```

```
names(diff.hte) <- c("t", "CACE", "SE")
diff.hte1 <- diff.hte
diff.hte$lottery_type <- "Difference"
diff.hte$outcome <- "Paid on Time"


gtp_plot <- rbind.data.frame(gtp_plot, diff.hte)

gtp_plot$upper <- gtp_plot$CACE + qnorm(.975) * gtp_plot$SE
gtp_plot$lower <- gtp_plot$CACE - qnorm(.975) * gtp_plot$SE

gtp_plot$lottery_type <- as.factor(gtp_plot$lottery_type)
gtp_plot$lottery_type <- relevel(gtp_plot$lottery_type, ref= "No Holiday")
gtp_plot$lottery_type <- relevel(gtp_plot$lottery_type, ref= "Holiday")

# Calculate the p-value for the difference (one and two-tailed tests)

gtp_plot$t.stat <- gtp_plot$CACE/gtp_plot$SE
gtp_plot$p_value.2 <- 2 * (1 - pnorm(abs(gtp_plot$t.stat)))  # Two-tailed test
gtp_plot$p_value.1 <- (1 - pnorm(abs(gtp_plot$t.stat)))  # One-tailed test
gtp_plot$stars <- ifelse(gtp_plot$p_value.1<=.05 &
                         gtp_plot$t>0, "+", " ") # add one-tailed for post-treatment periods

#Figure
p <- ggplot(gtp_plot[gtp_plot$t<14,], aes(x=t, y=CACE))
p + facet_grid(lottery_type~.) +
  geom_rect(data=NULL,aes(xmin=0, xmax=4, ymin=-Inf, ymax=Inf),
            fill="gray80", color="gray80") +
  geom_hline(aes(yintercept=0), size=.5, linetype="dashed", color="gray60") +
  geom_text(aes(x = t, y = (upper + .05), label = stars), size = 5) +
  geom_errorbar(aes(x=t,
                    ymin=lower,
                    ymax=upper),
                width=.1, size=.8, position = position_dodge(width = 0.5)) +
  geom_point(size=2.5, position = position_dodge(width = 0.5)) +
  xlab("Payments Since Tax Holiday") +
```

```
  ylab("Paid on Time (CACE)") +
  geom_vline(aes(xintercept=0), size=.7) +
  theme_bw() +
  scale_colour_manual(values = c("black","blue")) +
  scale_alpha_manual(values = c("FALSE"=0.8, "TRUE"=1), guide='none') +
  scale_x_discrete(limit = c(-10:0,4:13),
                   labels = as.character(-10:10)) +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text.x = element_text(size = rel(1.1), hjust=.7),
        axis.text.y = element_text(size = rel(1.25)),
        axis.title.y = element_text(size = rel(1.3)),
        axis.title.x = element_text(size = rel(1.3)),
        legend.text = element_text(size = rel(1.2)),
        strip.text.x = element_text(size = rel(1.4)),
        strip.text.y = element_text(size = rel(1.4)),
        strip.background = element_rect(size = 1.5),
        legend.position = "none",
        legend.title=element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(colour = "black"))
```

```
## Warning: Continuous limits supplied to discrete scale.
## i Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```




```
##############################################
message("MAIN PAPER: Figure 5")
```

```
## MAIN PAPER: Figure 5
```

```r
# Placebo test: Treatment Effects for Automatic vs. Manual Payers


gtp_taxes <- taxes_panel[taxes_panel$ES_BP==1, ]
gtp_taxes <- gtp_taxes[gtp_taxes$TRIBUTO!="Patente de Rodados",]
t <- unique(taxes_panel$t_st)
t <- t[order(t)]
t <- t[t>-11 & t<=28]
gtp_plot <- NULL

# Automatic Payment & Manual Payment
for (i in 1:length(t)){

  temp <- gtp_taxes[gtp_taxes$t_st == t[i], ]

  if (nrow(temp[temp$TREATMENT==1,])==0) next
  if (nrow(temp[temp$TREATMENT==0,])==0) next

  on_time_auto <- robust.se(ivreg(en_fecha ~ TIENE_EXO, ~ TREATMENT,
                                  data=temp[temp$autopay_win==1,]))[2,1:2]
  on_time_manual <- robust.se(ivreg(en_fecha ~ TIENE_EXO, ~ TREATMENT,
                                    data=temp[temp$autopay_win==0,]))[2,1:2]

  gtp_plot <- rbind(gtp_plot,
                    as.vector(c("Paid on Time", t[i], on_time_auto, "Automatic Payment")),
                    as.vector(c("Paid on Time", t[i], on_time_manual, "Manual Payment"))
  )
  print(i)
}
```

```
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 1
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 2
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 3
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 4
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 5
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 6
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 7
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
```

```
## [1] 8
## [1] "Robust Standard Errors"

## Warning in sqrt(diag(se)): NaNs produced

## [1] "Robust Standard Errors"
## [1] 9
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 10
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 11
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 15
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 16
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 17
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 18
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 19
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 20
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 21
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 22
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 23
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 24
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 25
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 26
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 27
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
```

```
## [1] 28
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 29
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 30
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 31
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 32
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 33
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 34
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 35
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 36
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 37
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 38
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 39
```

```r
gtp_plot <- as.data.frame(gtp_plot)
names(gtp_plot) <- c("outcome", "t", "CACE", "SE", "sample")

gtp_plot$t <- as.numeric(as.character(gtp_plot$t))
gtp_plot$CACE <- as.numeric(as.character(gtp_plot$CACE))
gtp_plot$SE <- as.numeric(as.character(gtp_plot$SE))

# Difference
diff.hte <- NULL

for (i in 1:length(t)){

  temp <- gtp_plot[gtp_plot$t==t[i],]

  if (nrow(temp)!=2) next

  diff <- temp$CACE[temp$sample=="Manual Payment"] - temp$CACE[temp$sample=="Automatic Payment"]

  SE <- sqrt((temp$SE[temp$sample=="Manual Payment"])^2 + (temp$CACE[temp$sample=="Automatic Payment"])
```

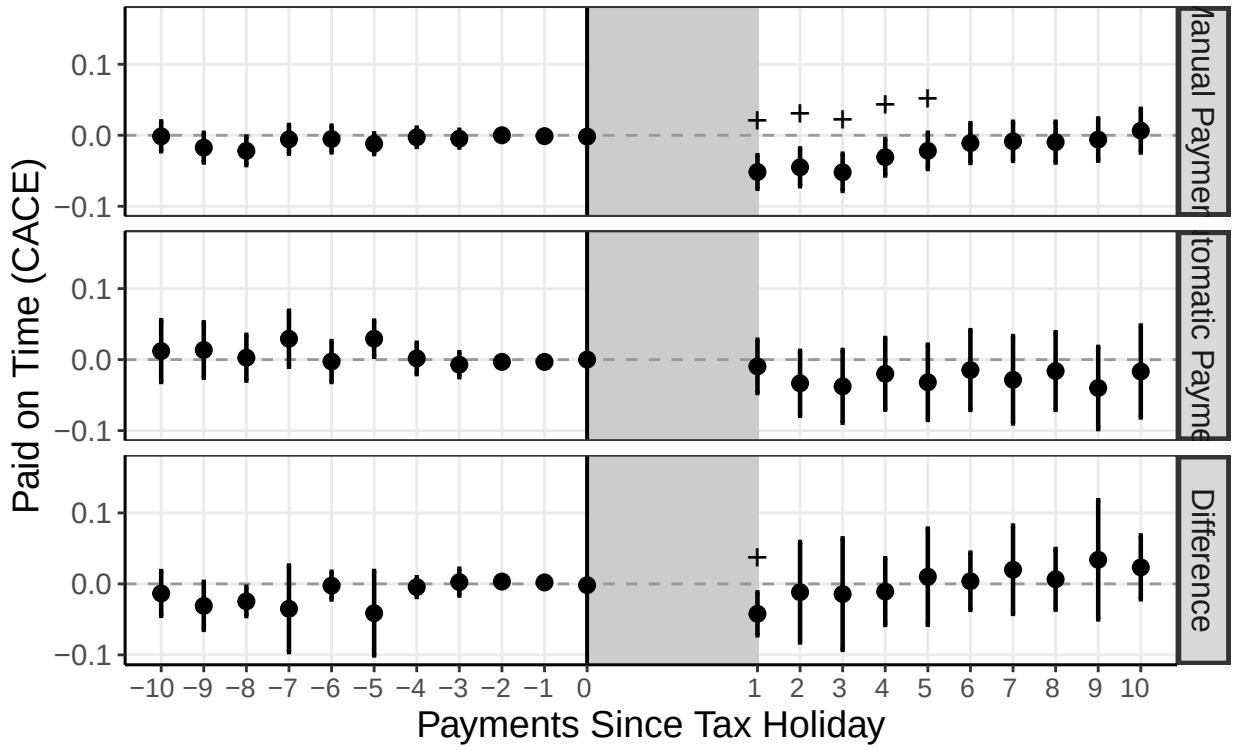
```
  diff.hte <- rbind.data.frame(diff.hte,
                               as.vector(c(t[i], diff, SE)))

  print(i)
}
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
## [1] 9
## [1] 10
## [1] 11
## [1] 15
## [1] 16
## [1] 17
## [1] 18
## [1] 19
## [1] 20
## [1] 21
## [1] 22
## [1] 23
## [1] 24
## [1] 25
## [1] 26
## [1] 27
## [1] 28
## [1] 29
## [1] 30
## [1] 31
## [1] 32
## [1] 33
## [1] 34
## [1] 35
## [1] 36
## [1] 37
## [1] 38
## [1] 39
```

```
names(diff.hte) <- c("t", "CACE", "SE")
diff.hte2 <- diff.hte
diff.hte$sample <- "Difference"
diff.hte$outcome <- "Paid on Time"


gtp_plot <- rbind.data.frame(gtp_plot, diff.hte)

gtp_plot$upper <- gtp_plot$CACE + qnorm(.975) * gtp_plot$SE
gtp_plot$lower <- gtp_plot$CACE - qnorm(.975) * gtp_plot$SE
```

```r
gtp_plot$sample <- as.factor(gtp_plot$sample)
gtp_plot$sample <- relevel(gtp_plot$sample, ref= "Automatic Payment")
gtp_plot$sample <- relevel(gtp_plot$sample, ref= "Manual Payment")

# Calculate the p-value for the difference (one and two-tailed tests)
gtp_plot$t.stat <- gtp_plot$CACE/gtp_plot$SE
gtp_plot$p_value.2 <- 2 * (1 - pnorm(abs(gtp_plot$t.stat)))  # Two-tailed test
gtp_plot$p_value.1 <- (1 - pnorm(abs(gtp_plot$t.stat)))  # One-tailed test
gtp_plot$stars <- ifelse(gtp_plot$p_value.1<=.05 &
                         gtp_plot$t>0, "+", " " ) # add one-tailed for post-treatment periods


# Figure
p <- ggplot(gtp_plot[gtp_plot$t<14,], aes(x=t, y=CACE))
p + facet_grid(sample~.) +
  geom_rect(data=NULL,aes(xmin=0, xmax=4, ymin=-Inf, ymax=Inf),
            fill="gray80", color="gray80") +
  geom_hline(aes(yintercept=0), size=.5, linetype="dashed", color="gray60") +
  geom_text(aes(x = t, y = (upper + .05), label = stars), size = 5) +
  geom_errorbar(aes(x=t,
                    ymin=lower,
                    ymax=upper),
                width=.1, size=.8, position = position_dodge(width = 0.5)) +
  geom_point(size=2.5, position = position_dodge(width = 0.5)) +
  xlab("Payments Since Tax Holiday") +
  ylab("Paid on Time (CACE)") +
  geom_vline(aes(xintercept=0), size=.7) +
  theme_bw() +
  scale_colour_manual(values = c("black","blue")) +
  scale_alpha_manual(values = c("FALSE"=0.8, "TRUE"=1), guide='none') +
  scale_x_discrete(limit = c(-10:0,4:13),
                   labels = as.character(-10:10)) +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text.x = element_text(size = rel(1.1), hjust=.7),
        axis.text.y = element_text(size = rel(1.25)),
        axis.title.y = element_text(size = rel(1.3)),
        axis.title.x = element_text(size = rel(1.3)),
        legend.text = element_text(size = rel(1.2)),
        strip.text.x = element_text(size = rel(1.4)),
        strip.text.y = element_text(size = rel(1.4)),
        strip.background = element_rect(size = 1.5),
        legend.position = "none",
        legend.title=element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(colour = "black"))
```

```
## Warning: Continuous limits supplied to discrete scale.
## i Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```




```
################################################
message("MAIN PAPER: Figure 6")
```

```
## MAIN PAPER: Figure 6
```

```
# The Stock of Habit: Perfect vs. Imperfect Past Compliers


taxpayer_type <- ddply(gtp_taxes[gtp_taxes$t_st<=0 & gtp_taxes$t_st>-15,],
                       "CUENTA", summarise,
                       type = mean(en_fecha, na.rm=T))

gtp_taxes <- merge(gtp_taxes, taxpayer_type, by="CUENTA", all.x=T)
t <- unique(gtp_taxes$t_st)
t <- t[order(t)]
t <- t[t>-11 & t<=21]
gtp_plot <- NULL

# Imperfect Past Complier
for (i in 1:length(t)){

  temp <- gtp_taxes[gtp_taxes$t_st == t[i], ]

  if (nrow(temp[temp$TREATMENT==1 & temp$type!=1,])==0) next
  if (nrow(temp[temp$TREATMENT==0 & temp$type!=1,])==0) next

  on_time_marginal <- robust.se(ivreg(en_fecha ~ TIENE_EXO, ~ TREATMENT,
                                      data=temp[temp$type!=1,]))[2,1:2]
```

```
  gtp_plot <- rbind(gtp_plot,
                    as.vector(c("Paid on Time", t[i],
                                on_time_marginal, "Imperfect Past Complier")))
  print(i)
}
```

```
## [1] "Robust Standard Errors"
## [1] 1
## [1] "Robust Standard Errors"
## [1] 2
## [1] "Robust Standard Errors"
## [1] 3
## [1] "Robust Standard Errors"
## [1] 4
## [1] "Robust Standard Errors"
## [1] 5
## [1] "Robust Standard Errors"
## [1] 6
## [1] "Robust Standard Errors"
## [1] 7
## [1] "Robust Standard Errors"
## [1] 8
## [1] "Robust Standard Errors"

## Warning in sqrt(diag(se)): NaNs produced

## [1] 9
## [1] "Robust Standard Errors"

## Warning in sqrt(diag(se)): NaNs produced

## [1] 10
## [1] "Robust Standard Errors"

## Warning in sqrt(diag(se)): NaNs produced

## [1] 11
## [1] "Robust Standard Errors"
## [1] 12
## [1] "Robust Standard Errors"
## [1] 13
## [1] "Robust Standard Errors"
## [1] 14
## [1] "Robust Standard Errors"
## [1] 15
## [1] "Robust Standard Errors"
## [1] 16
## [1] "Robust Standard Errors"
## [1] 17
## [1] "Robust Standard Errors"
## [1] 18
```

```
## [1] "Robust Standard Errors"
## [1] 19
## [1] "Robust Standard Errors"
## [1] 20
## [1] "Robust Standard Errors"
## [1] 21
## [1] "Robust Standard Errors"
## [1] 22
## [1] "Robust Standard Errors"
## [1] 23
## [1] "Robust Standard Errors"
## [1] 24
## [1] "Robust Standard Errors"
## [1] 25
## [1] "Robust Standard Errors"
## [1] 26
## [1] "Robust Standard Errors"
## [1] 27
## [1] "Robust Standard Errors"
## [1] 28
## [1] "Robust Standard Errors"
## [1] 29
```

```
# Perfect Past Complier
for (i in 1:length(t)){

  temp <- gtp_taxes[gtp_taxes$t_st == t[i], ]

  if (nrow(temp[temp$TREATMENT==1 & temp$type==1,])==0) next
  if (nrow(temp[temp$TREATMENT==0 & temp$type==1,])==0) next

  on_time_compliant <- robust.se(ivreg(en_fecha ~ TIENE_EXO, ~ TREATMENT,
                                       data=temp[temp$type==1,]))[2,1:2]

  gtp_plot <- rbind(gtp_plot,
                    as.vector(c("Paid on Time", t[i],
                                on_time_compliant, "Perfect Past Complier")))
  print(i)
}
```

```
## [1] "Robust Standard Errors"
## [1] 1
## [1] "Robust Standard Errors"
## [1] 2
## [1] "Robust Standard Errors"
## [1] 3
## [1] "Robust Standard Errors"
## [1] 4
## [1] "Robust Standard Errors"
## [1] 5
## [1] "Robust Standard Errors"
## [1] 6
## [1] "Robust Standard Errors"
## [1] 7
```

```
## [1] "Robust Standard Errors"
## [1] 8
## [1] "Robust Standard Errors"
## [1] 9
## [1] "Robust Standard Errors"
## [1] 10
## [1] "Robust Standard Errors"
## [1] 11
## [1] "Robust Standard Errors"
## [1] 12
## [1] "Robust Standard Errors"
## [1] 13
## [1] "Robust Standard Errors"
## [1] 14
## [1] "Robust Standard Errors"
## [1] 15
## [1] "Robust Standard Errors"
## [1] 16
## [1] "Robust Standard Errors"
## [1] 17
## [1] "Robust Standard Errors"
## [1] 18
## [1] "Robust Standard Errors"
## [1] 19
## [1] "Robust Standard Errors"
## [1] 20
## [1] "Robust Standard Errors"
## [1] 21
## [1] "Robust Standard Errors"
## [1] 22
## [1] "Robust Standard Errors"
## [1] 23
## [1] "Robust Standard Errors"
## [1] 24
## [1] "Robust Standard Errors"
## [1] 25
## [1] "Robust Standard Errors"
## [1] 26
## [1] "Robust Standard Errors"
## [1] 27
## [1] "Robust Standard Errors"
## [1] 28
## [1] "Robust Standard Errors"
## [1] 29
```

```r
gtp_plot <- as.data.frame(gtp_plot)
names(gtp_plot) <- c("outcome", "t", "CACE", "SE", "sample")

gtp_plot$t <- as.numeric(as.character(gtp_plot$t))
gtp_plot$CACE <- as.numeric(as.character(gtp_plot$CACE))
gtp_plot$SE <- as.numeric(as.character(gtp_plot$SE))

# Difference
diff.hte <- NULL
```

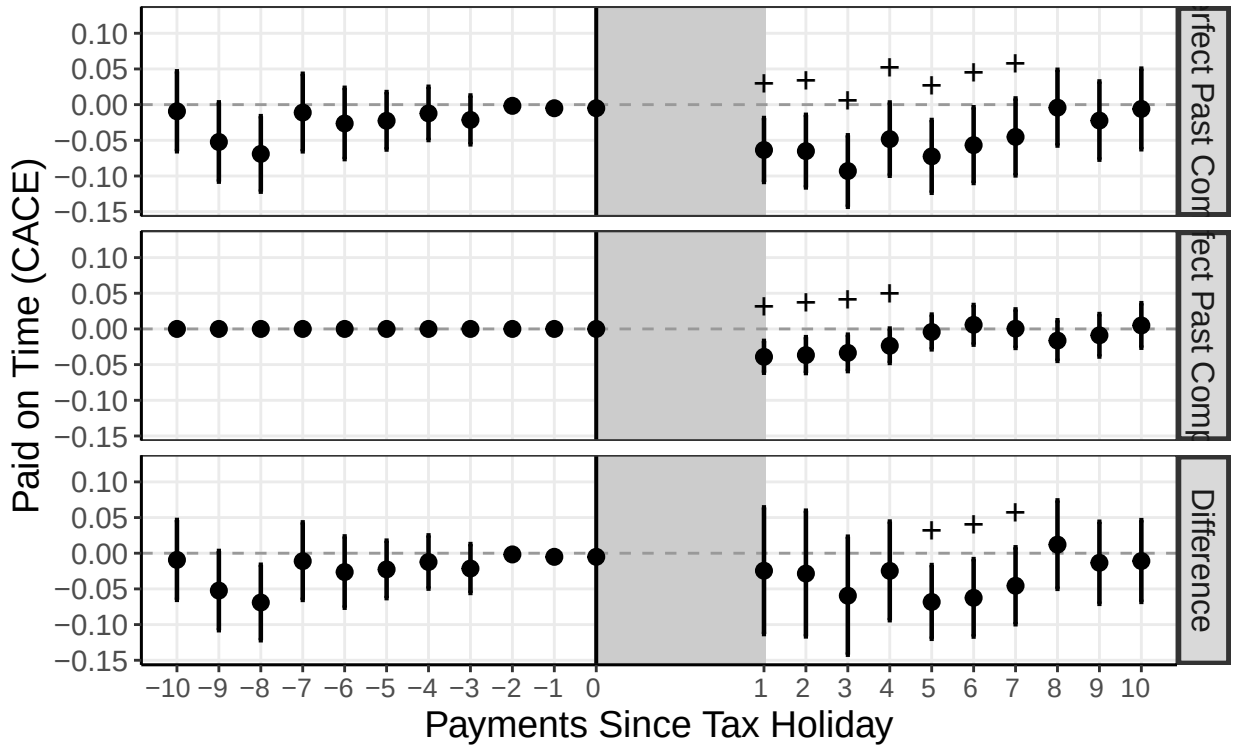
```r
for (i in 1:length(t)){

  temp <- gtp_plot[gtp_plot$t==t[i],]

  if (nrow(temp)!=2) next

  diff <- temp$CACE[temp$sample=="Imperfect Past Complier"] - temp$CACE[temp$sample=="Perfect Past Compl

  SE <- sqrt((temp$SE[temp$sample=="Imperfect Past Complier"])^2 + (temp$CACE[temp$sample=="Perfect Past

  diff.hte <- rbind.data.frame(diff.hte, as.vector(c(t[i], diff, SE)))

  print(i)
}
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
## [1] 9
## [1] 10
## [1] 11
## [1] 12
## [1] 13
## [1] 14
## [1] 15
## [1] 16
## [1] 17
## [1] 18
## [1] 19
## [1] 20
## [1] 21
## [1] 22
## [1] 23
## [1] 24
## [1] 25
## [1] 26
## [1] 27
## [1] 28
## [1] 29
```

```r
names(diff.hte) <- c("t", "CACE", "SE")
diff.hte3 <- diff.hte
diff.hte$sample <- "Difference"
diff.hte$outcome <- "Paid on Time"

gtp_plot <- rbind.data.frame(gtp_plot, diff.hte)

gtp_plot$upper <- gtp_plot$CACE + qnorm(.975) * gtp_plot$SE
```

```r
gtp_plot$lower <- gtp_plot$CACE - qnorm(.975) * gtp_plot$SE

gtp_plot$sample  <- as.factor(gtp_plot$sample)
gtp_plot$sample  <- relevel(gtp_plot$sample, ref= "Perfect Past Complier")
gtp_plot$sample  <- relevel(gtp_plot$sample, ref= "Imperfect Past Complier")

gtp_plot$upper <- gtp_plot$CACE + qnorm(.975) * gtp_plot$SE
gtp_plot$lower <- gtp_plot$CACE - qnorm(.975) * gtp_plot$SE

# Calculate the p-value for the difference (one and two-tailed tests)
gtp_plot$t.stat <- gtp_plot$CACE/gtp_plot$SE
gtp_plot$p_value.2 <- 2 * (1 - pnorm(abs(gtp_plot$t.stat)))  # Two-tailed test
gtp_plot$p_value.1 <- (1 - pnorm(abs(gtp_plot$t.stat)))  # One-tailed test
gtp_plot$stars <- ifelse(gtp_plot$p_value.1<=.05 &
                         gtp_plot$t>0, "+", " ") # add one-tailed for post-treatment periods


# Figure
p <- ggplot(gtp_plot[gtp_plot$t<14,], aes(x=t, y=CACE))
p + facet_grid(sample ~.) +
  geom_rect(data=NULL,aes(xmin=0, xmax=4, ymin=-Inf, ymax=Inf),
            fill="gray80", color="gray80") +
  geom_hline(aes(yintercept=0), size=.5, linetype="dashed", color="gray60") +
  geom_text(aes(x = t, y = (upper + .05), label = stars), size = 5) +
  geom_errorbar(aes(x=t,
                    ymin=lower,
                    ymax=upper),
                width=.1, size=.8, position = position_dodge(width = 0.5)) +
  geom_point(size=2.5, position = position_dodge(width = 0.5)) +
  xlab("Payments Since Tax Holiday") +
  ylab("Paid on Time (CACE)") +
  geom_vline(aes(xintercept=0), size=.7) +
  theme_bw() +
  scale_colour_manual(values = c("black","blue")) +
  scale_alpha_manual(values = c("FALSE"=0.8, "TRUE"=1), guide='none') +
  scale_x_discrete(limit = c(-10:0,4:13),
                   labels = as.character(-10:10)) +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text.x = element_text(size = rel(1.1), hjust=.7),
        axis.text.y = element_text(size = rel(1.25)),
        axis.title.y = element_text(size = rel(1.3)),
        axis.title.x = element_text(size = rel(1.3)),
        legend.text = element_text(size = rel(1.2)),
        strip.text.x = element_text(size = rel(1.4)),
        strip.text.y = element_text(size = rel(1.4)),
        strip.background = element_rect(size = 1.5),
        legend.position = "none",
        legend.title=element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(colour = "black"))
```

```
## Warning: Continuous limits supplied to discrete scale.
## i Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```




```
################################################
message("MAIN PAPER: Figure 7")
```

```
## MAIN PAPER: Figure 7
```

```
# Field Experiment: Effects of Information About the Tax Holiday on Compliance

# Weigths for treatment - placebo comparison
table(fieldex$type, fieldex$pooled_124_0)
```

```
##
##                     0    1
##   good taxpayer 1532 3037
##   bad taxpayer  2080 4150
```

```
table.weights <- 1/prop.table(table(fieldex$type, fieldex$pooled_124_0), 1)
table.weights <- melt(table.weights)
names(table.weights) <- c("type", "pooled_124_0", "pooled_124_0_wts")
fieldex <- fieldex %>% left_join(table.weights)
```

```
## Joining with `by = join_by(type, pooled_124_0)`
```

```
# Weights for treatment versus pure control
table(fieldex$type, fieldex$pooled_124_6)
```

```
##
```

```
##                     0    1
##   good taxpayer 7243 3037
##   bad taxpayer  3412 4150
```

```
table.weights <- 1/prop.table(table(fieldex$type, fieldex$pooled_124_6), 1)
table.weights <- melt(table.weights)
names(table.weights) <- c("type", "pooled_124_6", "pooled_124_6_wts")
fieldex <- fieldex %>% left_join(table.weights)
```

```
## Joining with `by = join_by(type, pooled_124_6)`
```

```
# Weights for placebo versus pure control
table(fieldex$type, fieldex$pooled_0_6)
```

```
##
##                     0    1
##   good taxpayer 7243 1532
##   bad taxpayer  3412 2080
```

```
table.weights <- 1/prop.table(table(fieldex$type, fieldex$pooled_0_6), 1)
table.weights <- melt(table.weights)
names(table.weights) <- c("type", "pooled_0_6", "pooled_0_6_wts")
fieldex <- fieldex %>% left_join(table.weights)
```

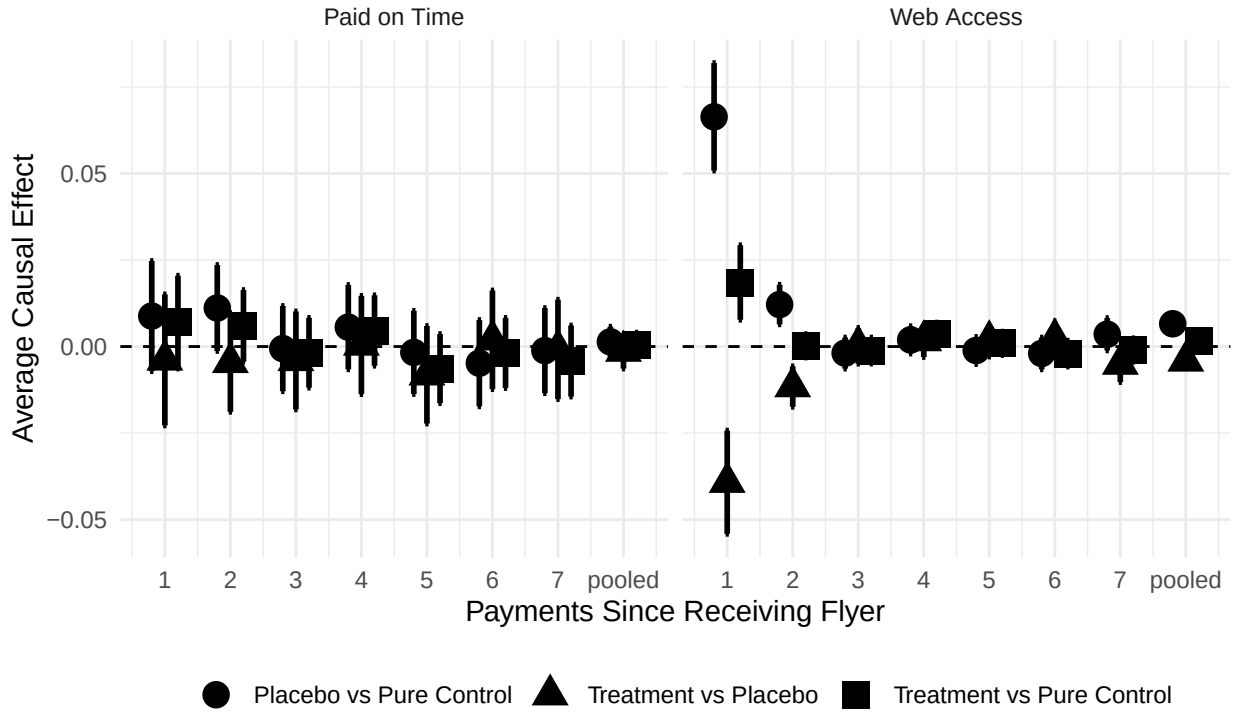
```
## Joining with `by = join_by(type, pooled_0_6)`
```

```
# Add variable to summarize compliance between 2014-2016
fieldex <-  fieldex %>% dplyr::mutate(
  compliance_1416 = (JUL_2014_ontime + NOV_2014_ontime +
                       MAR_2015_ontime + JUL_2015_ontime +
                       NOV_2015_ontime + MAR_2016_ontime +
                       JUL_2016_ontime)/12,
  intended_1416 = (JUL_2014_WEBACCESS + NOV_2014_WEBACCESS +
                     MAR_2015_WEBACCESS + JUL_2015_WEBACCESS +
                     NOV_2015_WEBACCESS + MAR_2016_WEBACCESS +
                     JUL_2016_WEBACCESS)/12
)


#### TREATMENT VERSUS PURE CONTROL
# a) compliance
comp_control <- rbind.data.frame(
  difference_in_means(JUL_2014_ontime ~ pooled_124_6, weights = pooled_124_6_wts,
                      data = fieldex),
  difference_in_means(NOV_2014_ontime ~ pooled_124_6, weights = pooled_124_6_wts,
                      data = fieldex),
  difference_in_means(MAR_2015_ontime ~ pooled_124_6, weights = pooled_124_6_wts,
                      data = fieldex),
  difference_in_means(JUL_2015_ontime ~ pooled_124_6, weights = pooled_124_6_wts,
                      data = fieldex),
  difference_in_means(NOV_2015_ontime ~ pooled_124_6, weights = pooled_124_6_wts,
```

```
                      data = fieldex),
  difference_in_means(MAR_2016_ontime ~ pooled_124_6, weights = pooled_124_6_wts,
                      data = fieldex),
  difference_in_means(JUL_2016_ontime ~ pooled_124_6, weights = pooled_124_6_wts,
                      data = fieldex),
  difference_in_means(compliance_1416 ~ pooled_124_6, weights = pooled_124_6_wts,
                      data = fieldex)
)
comp_control$outcome <- "Paid on Time"
comp_control$control <- "Treatment vs Pure Control"

# b) Intended compliance
intcomp_control <- rbind.data.frame(
  difference_in_means(JUL_2014_WEBACCESS ~ pooled_124_6, weights = pooled_124_6_wts,
                      data = fieldex),
  difference_in_means(NOV_2014_WEBACCESS ~ pooled_124_6, weights = pooled_124_6_wts,
                      data = fieldex),
  difference_in_means(MAR_2015_WEBACCESS ~ pooled_124_6, weights = pooled_124_6_wts,
                      data = fieldex),
  difference_in_means(JUL_2015_WEBACCESS ~ pooled_124_6, weights = pooled_124_6_wts,
                      data = fieldex),
  difference_in_means(NOV_2015_WEBACCESS ~ pooled_124_6, weights = pooled_124_6_wts,
                      data = fieldex),
  difference_in_means(MAR_2016_WEBACCESS ~ pooled_124_6, weights = pooled_124_6_wts,
                      data = fieldex),
  difference_in_means(JUL_2016_WEBACCESS ~ pooled_124_6, weights = pooled_124_6_wts,
                      data = fieldex),
  difference_in_means(intended_1416 ~ pooled_124_6, weights = pooled_124_6_wts,
                      data = fieldex)
)
intcomp_control$outcome <- "Web Access"
intcomp_control$control <- "Treatment vs Pure Control"


#### TREATMENT VERSUS PLACEBO
# a) Compliance
comp_placebo <- rbind.data.frame(
  difference_in_means(JUL_2014_ontime ~ pooled_124_0, weights = pooled_124_0_wts,
                      data = fieldex),
  difference_in_means(NOV_2014_ontime ~ pooled_124_0, weights = pooled_124_0_wts,
                      data = fieldex),
  difference_in_means(MAR_2015_ontime ~ pooled_124_0, weights = pooled_124_0_wts,
                      data = fieldex),
  difference_in_means(JUL_2015_ontime ~ pooled_124_0, weights = pooled_124_0_wts,
                      data = fieldex),
  difference_in_means(NOV_2015_ontime ~ pooled_124_0, weights = pooled_124_0_wts,
                      data = fieldex),
  difference_in_means(MAR_2016_ontime ~ pooled_124_0, weights = pooled_124_0_wts,
                      data = fieldex),
  difference_in_means(JUL_2016_ontime ~ pooled_124_0, weights = pooled_124_0_wts,
                      data = fieldex),
  difference_in_means(compliance_1416 ~ pooled_124_0, weights = pooled_124_0_wts,
                      data = fieldex)
```

```
)
comp_placebo$outcome <- "Paid on Time"
comp_placebo$control <- "Treatment vs Placebo"

# b) Intended compliance
intcomp_placebo <- rbind.data.frame(
  difference_in_means(JUL_2014_WEBACCESS ~ pooled_124_0,
                      weights = pooled_124_0_wts, data = fieldex),
  difference_in_means(NOV_2014_WEBACCESS ~ pooled_124_0,
                      weights = pooled_124_0_wts, data = fieldex),
  difference_in_means(MAR_2015_WEBACCESS ~ pooled_124_0,
                      weights = pooled_124_0_wts, data = fieldex),
  difference_in_means(JUL_2015_WEBACCESS ~ pooled_124_0,
                      weights = pooled_124_0_wts, data = fieldex),
  difference_in_means(NOV_2015_WEBACCESS ~ pooled_124_0,
                      weights = pooled_124_0_wts, data = fieldex),
  difference_in_means(MAR_2016_WEBACCESS ~ pooled_124_0,
                      weights = pooled_124_0_wts, data = fieldex),
  difference_in_means(JUL_2016_WEBACCESS ~ pooled_124_0,
                      weights = pooled_124_0_wts, data = fieldex),
  difference_in_means(intended_1416 ~ pooled_124_0,
                      weights = pooled_124_0_wts, data = fieldex)
)
intcomp_placebo$outcome <- "Web Access"
intcomp_placebo$control <- "Treatment vs Placebo"


#### PURE CONTROL VERSUS PLACEBO
# a) Compliance
comp_control_pla <- rbind.data.frame(
  difference_in_means(JUL_2014_ontime ~ pooled_0_6, weights = pooled_0_6_wts,
                      data = fieldex),
  difference_in_means(NOV_2014_ontime ~ pooled_0_6, weights = pooled_0_6_wts,
                      data = fieldex),
  difference_in_means(MAR_2015_ontime ~ pooled_0_6, weights = pooled_0_6_wts,
                      data = fieldex),
  difference_in_means(JUL_2015_ontime ~ pooled_0_6, weights = pooled_0_6_wts,
                      data = fieldex),
  difference_in_means(NOV_2015_ontime ~ pooled_0_6, weights = pooled_0_6_wts,
                      data = fieldex),
  difference_in_means(MAR_2016_ontime ~ pooled_0_6, weights = pooled_0_6_wts,
                      data = fieldex),
  difference_in_means(JUL_2016_ontime ~ pooled_0_6, weights = pooled_0_6_wts,
                      data = fieldex),
  difference_in_means(compliance_1416 ~ pooled_0_6, weights = pooled_0_6_wts,
                      data = fieldex)
)
comp_control_pla$outcome <- "Paid on Time"
comp_control_pla$control <- "Placebo vs Pure Control"

# b) Intended compliance
intcomp_control_pla <- rbind.data.frame(
  difference_in_means(JUL_2014_WEBACCESS ~ pooled_0_6,
```

```
                      weights = pooled_0_6_wts, data = fieldex),
  difference_in_means(NOV_2014_WEBACCESS ~ pooled_0_6,
                      weights = pooled_0_6_wts, data = fieldex),
  difference_in_means(MAR_2015_WEBACCESS ~ pooled_0_6,
                      weights = pooled_0_6_wts, data = fieldex),
  difference_in_means(JUL_2015_WEBACCESS ~ pooled_0_6,
                      weights = pooled_0_6_wts, data = fieldex),
  difference_in_means(NOV_2015_WEBACCESS ~ pooled_0_6,
                      weights = pooled_0_6_wts, data = fieldex),
  difference_in_means(MAR_2016_WEBACCESS ~ pooled_0_6,
                      weights = pooled_0_6_wts, data = fieldex),
  difference_in_means(JUL_2016_WEBACCESS ~ pooled_0_6,
                      weights = pooled_0_6_wts, data = fieldex),
  difference_in_means(intended_1416 ~ pooled_0_6,
                      weights = pooled_0_6_wts, data = fieldex)
)
intcomp_control_pla$outcome <- "Web Access"
intcomp_control_pla$control <- "Placebo vs Pure Control"

## Combine and plot
plotdata <- rbind.data.frame(comp_placebo, comp_control, comp_control_pla,
                             intcomp_placebo, intcomp_control,
                             intcomp_control_pla)
rm(comp_placebo, comp_control,
   intcomp_placebo, intcomp_control,
   comp_control_pla, intcomp_control_pla)

plotdata$time <- rep(1:8, 6)

plotdata$control <- as.factor(plotdata$control)

pd <- position_dodge(width = 0.6)
ggplot(plotdata, aes(x=time, y=coefficients, group = control, shape = control)) +
  facet_wrap( ~ outcome) +
  geom_point(size=4.5, position=pd) +
  geom_hline(aes(yintercept=0), size=.5, linetype="dashed") +
  geom_errorbar(aes(x=time,
                    ymin=conf.low,
                    ymax=conf.high),
                width=.15, size=1, position=pd) +
  xlab("Payments Since Receiving Flyer") + ylab("Average Causal Effect") +
  theme_minimal() +
  scale_colour_manual(values = c("black", "black")) +
  scale_x_continuous(breaks=1:8,
                     labels=c(as.character(1:7), "pooled")) +
  theme(legend.position = "bottom",
        legend.title=element_blank())
```



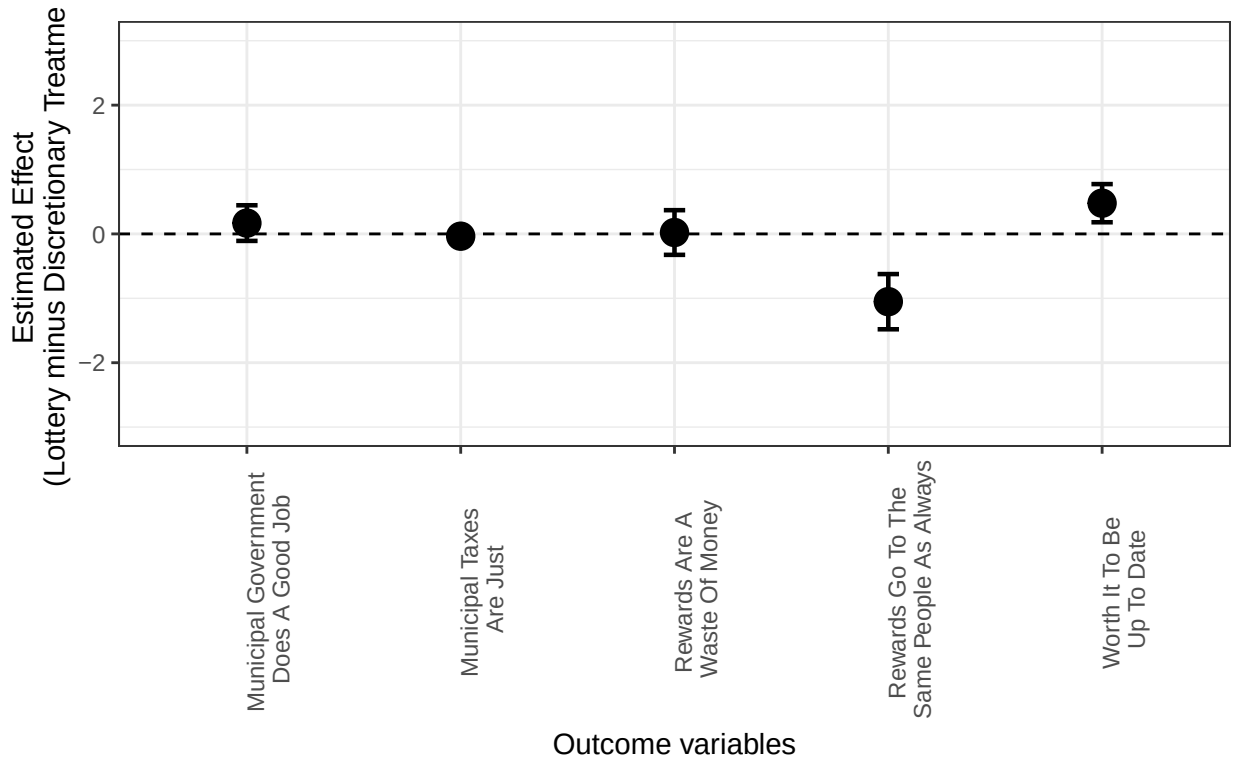

```
################################################
message("MAIN PAPER: Figure 8")
```

```
## MAIN PAPER: Figure 8
```

```
#  Survey Experiment: Effects of Information about the Lottery on Attitudes Towards Taxation

# Select all variables and take the mean, N & SE for
discretion1 <- ddply(survey_data, "treat_discretion",
                     summarise,
                     mean=mean(S1p4,na.rm=T),
                     N=length(na.omit(S1p4,na.rm=T)),
                     se=sd(S1p4,na.rm=T)/sqrt(N))

discretion2 <- ddply(survey_data, "treat_discretion",
                     summarise,
                     mean=mean(S1p1,na.rm=T),
                     N=length(na.omit(S1p1,na.rm=T)),
                     se=sd(S1p1,na.rm=T)/sqrt(N))

discretion3 <-  ddply(survey_data, "treat_discretion",
                      summarise,
                      mean=mean(S1p3,na.rm=T),
                      N=length(na.omit(S1p3,na.rm=T)),
                      se=sd(S1p3,na.rm=T)/sqrt(N))

discretion4 <- ddply(survey_data, "treat_discretion",
                     summarise,
```

```
                    mean=mean(S1p2,na.rm=T),
                    N=length(na.omit(S1p2,na.rm=T)),
                    se=sd(S1p2, na.rm=T)/sqrt(N))

discretion5 <- ddply(survey_data, "treat_discretion",
                    summarise,
                    mean=mean(S1p5,na.rm=T),
                    N=length(na.omit(S1p5,na.rm=T)),
                    se=sd(S1p5, na.rm=T)/sqrt(N))

# all together
discretion <- as.data.frame(rbind(discretion1, discretion2, discretion3,
                                  discretion4, discretion5))
rm(discretion1, discretion2, discretion3,discretion4, discretion5)

discretion$outcome <- rep(c("Rewards Go To The\n Same People As Always",
                            "Rewards Are A\n Waste Of Money",
                            "Worth It To Be\n Up To Date",
                            "Municipal Government\n Does A Good Job",
                            "Municipal Taxes\n Are Just"), each=2)


# Calculate the difference in means for all variables between the Lottery and Discretionary group
discretion_lot<-discretion[discretion$treat_discretion==0,] #Lottery
discretion_dis<-discretion[discretion$treat_discretion==1,] #Discretionary

diff<-data.frame(matrix(ncol=6,nrow=5))
colnames(diff)<-c('outcome','Estimate','Std. Error','p value','upper','lower')

diff$outcome <- c("Rewards Go To The\n Same People As Always",
                  "Rewards Are A\n Waste Of Money",
                  "Worth It To Be\n Up To Date",
                  "Municipal Government\n Does A Good Job",
                  "Municipal Taxes\n Are Just")
for (i in 1:5){

  aux_1<-discretion_lot[i,] #Lottery
  aux_0<-discretion_dis[i,] #Discretionary

  # Estimate
  mean<-aux_1$mean-aux_0$mean
  # Calculate the standard error of the difference between the means
  se_diff <- sqrt((aux_1$se^2) + (aux_0$se^2))
  # Calculate the t-statistic
  t_stat <- (aux_1$mean - aux_0$mean) / se_diff
  # Calculate the degrees of freedom
  df <- aux_1$N+ aux_0$N - 2
  # Calculate the p-value (two-tailed test)
  p_value <- 2 * pt(-abs(t_stat), df)
  p_value
  # calculate upper and lower CI
  upper <- mean + 1.96*se_diff
  lower <- mean - 1.96*se_diff
```

```
  diff[i,'Estimate']<-mean
  diff[i,'Std. Error']<-se_diff
  diff[i,'p value']<-p_value
  diff[i,'upper']<-upper
  diff[i,'lower']<-lower
}

# Figure
p <- ggplot(diff, aes(x=outcome, y=Estimate))
p + geom_point(size=4.5) +
  xlab("Outcome variables") +
  ylab("Estimated Effect \n (Lottery minus Discretionary Treatment)") +
  geom_errorbar(aes(x=outcome,
                    ymin=lower,
                    ymax=upper),
                width=.1, size=.8) +
  coord_cartesian(ylim = c(-3, 3))+
  geom_hline(yintercept = 0, linetype = "dashed", color = 'black')+
  theme_bw() +
  theme(legend.position = "none",
        axis.text.x = element_text(angle = 90, hjust = 0.5))
```



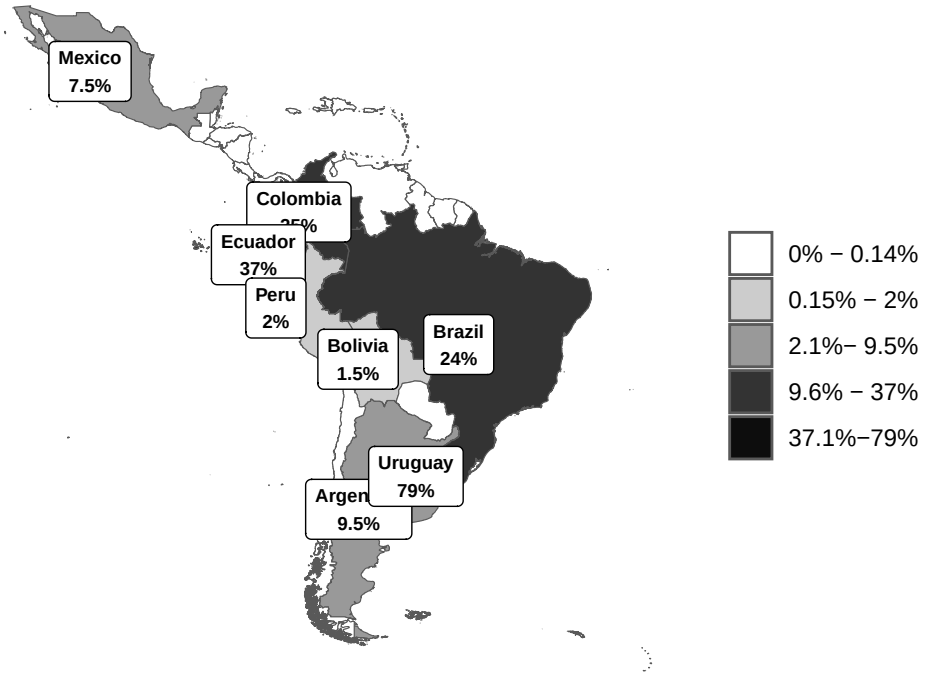

```
##############################################################################
# APPENDIX Tables & Figures
##############################################################################

##############################################
message("APPENDIX: Figure A1")
```

APPENDIX: Figure A1

```
# Reward Programs for Good Taxpayers in Latin America

# prep data ---
latam_map_data$per4<-as.numeric(gsub('%','',latam_map_data$per3))
latam_map_data$per4<-ifelse(is.na(latam_map_data$per4),0,latam_map_data$per4)
latam_map_data$group<-NA
latam_map_data$group[latam_map_data$per4==0]<-1
latam_map_data$group[latam_map_data$per4>=0.15 & latam_map_data$per4<=2]<-2
latam_map_data$group[latam_map_data$per4>=2.1 & latam_map_data$per4<=9.5]<-3
latam_map_data$group[latam_map_data$per4>=9.6 & latam_map_data$per4<=37]<-4
latam_map_data$group[latam_map_data$per4>=37.1 & latam_map_data$per4<=79]<-5

latam_map_data$label<-paste0(latam_map_data$NAME,'\n',latam_map_data$per3)
latam_map_data$label<-ifelse(is.na(latam_map_data$per3),NA,latam_map_data$label)


# Figure

labels<-c('0% - 0.14%','0.15% - 2%','2.1%- 9.5%','9.6% - 37%','37.1%-79%')

map<-ggplot() +
  geom_sf(data = latam_map_data,aes(fill=factor(group),  geometry = geometry))+
  scale_fill_manual(values = c("white", "grey80", "grey60", "grey20", "grey5"), name = "group", labels =
  geom_sf_label(data = latam_map_data, aes(label = label),
                fill='white',size = 2.5,nudge_x = -4,fontface = "bold") + #label.size = 0,
  theme_classic()+
  theme(plot.background = element_rect(fill = "white"),
        panel.background = element_rect(colour = "white",),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        axis.line = element_blank(),
        axis.ticks = element_blank(),
        axis.text = element_blank(),
        axis.title = element_blank(),
        legend.position = 'right',
        legend.title = element_blank())+
  coord_sf()

map
```

```
## Warning in st_point_on_surface.sfc(sf::st_zm(x)): st_point_on_surface may not
## give correct results for longitude/latitude data
```

```
## Warning: Removed 40 rows containing missing values (‘geom_label()‘).
```




```
################################################
message("APPENDIX: Table A1")
```

```
## APPENDIX: Table A1
```

```
#Balance test

gtp_panel <- buen_pagador_panel
gtp_panel <- gtp_panel[abs(gtp_panel$centered)<58,]

taxes_panel$t[taxes_panel$TRIBUTO=="TS" | taxes_panel$TRIBUTO=="TD"] <- taxes_panel$t/2


balance <- rbind(
  with(gtp_panel[gtp_panel$centered==0,], ttest(BP_time, TREATMENT)),
  with(gtp_panel[gtp_panel$centered==-1,], ttest(BP_time, TREATMENT)),
  with(gtp_panel[gtp_panel$centered==-2,], ttest(BP_time, TREATMENT)),
  with(gtp_panel[gtp_panel$centered==-3,], ttest(BP_time, TREATMENT)),
  with(gtp_panel[gtp_panel$centered==-4,], ttest(BP_time, TREATMENT)),
  with(gtp_panel[gtp_panel$centered==-5,], ttest(BP_time, TREATMENT)),
  with(gtp_panel[gtp_panel$centered==0,], ttest(autopay_win, TREATMENT)),
  with(gtp_panel[gtp_panel$centered==-1,], ttest(autopay_win, TREATMENT)),
  with(gtp_panel[gtp_panel$centered==-2,], ttest(autopay_win, TREATMENT)),
  with(gtp_panel[gtp_panel$centered==-3,], ttest(autopay_win, TREATMENT)),
  ttest(naturalex$VALOR_CAT2004, naturalex$TREATMENT),
  ttest(naturalex$VALOR_CATASTRALACTUAL, naturalex$TREATMENT),
  ttest(naturalex$rental, naturalex$TREATMENT),
  ttest(naturalex$JUBILADO, naturalex$TREATMENT),
```

```r
  ttest(naturalex$TODOPAGO, naturalex$TREATMENT))

balance <- round(balance, 3)
balance <- as.data.frame(balance)
rownames(balance) <-  c("Good taxpayer at t=0",
                        "Good taxpayer at t=-1",
                        "Good taxpayer at t=-2",
                        "Good taxpayer at t=-3",
                        "Good taxpayer at t=-4",
                        "Good taxpayer at t=-5",
                        "Automatic debit at t=0",
                        "Automatic debit at t=-1",
                        "Automatic debit at t=-2",
                        "Automatic debit at t=-3",
                        "2004 Property value",
                        "Current property value",
                        "Rented Property",
                        "Retiree",
                        "Paid year in full")

balance$sample <- c(rep("All Taxes", 6),
                    rep("Property, head & sewage", 7),
                    rep("Property", 2))

balance[,c(2:4,6,8,9)]
```

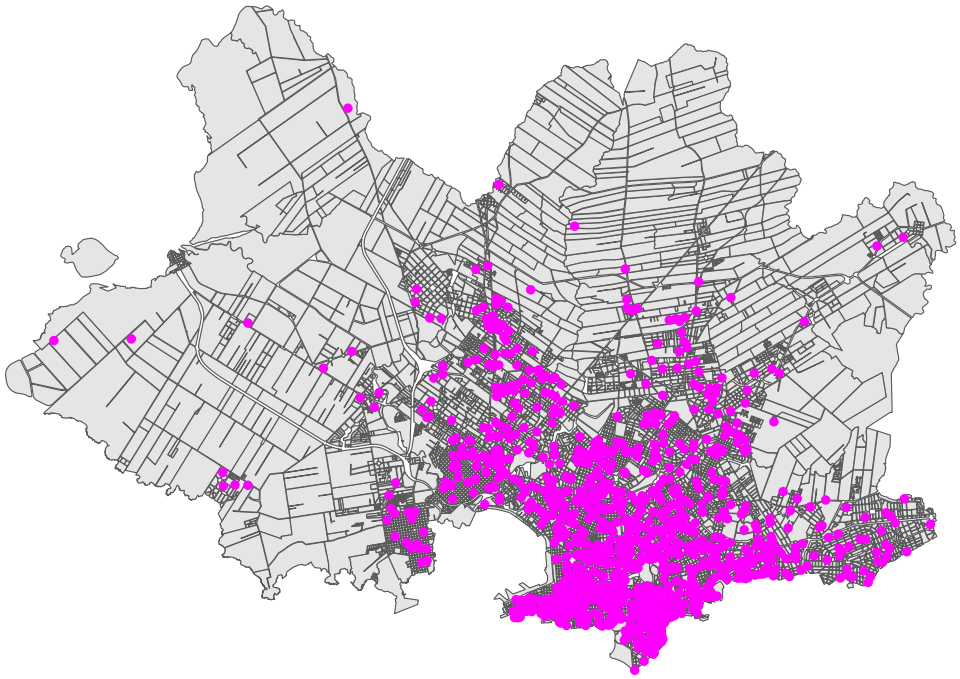
```
##                             Mean 0  Difference    SE_Diff     N p-value
## Good taxpayer at t=0         0.346       0.003      0.007 18621   0.651
## Good taxpayer at t=-1        0.347       0.001      0.007 18309   0.934
## Good taxpayer at t=-2        0.345       0.002      0.007 17991   0.823
## Good taxpayer at t=-3        0.346       0.002      0.007 17670   0.769
## Good taxpayer at t=-4        0.345       0.005      0.007 17350   0.514
## Good taxpayer at t=-5        0.345       0.003      0.007 17021   0.689
## Automatic debit at t=0       0.106      -0.003      0.005 14027   0.506
## Automatic debit at t=-1      0.106      -0.003      0.005 13807   0.503
## Automatic debit at t=-2      0.106      -0.004      0.005 13580   0.498
## Automatic debit at t=-3      0.106      -0.004      0.005 13350   0.447
## 2004 Property value     1259241.960 -199730.648 164141.207 13462   0.224
## Current property value  2747841.019 -305151.051 345883.758 13998   0.378
## Rented Property              0.243      -0.009      0.007 13998   0.229
## Retiree                      0.008      -0.001      0.002  5129   0.655
## Paid year in full            0.287      -0.010      0.013  5129   0.441
##                                          sample
## Good taxpayer at t=0                  All Taxes
## Good taxpayer at t=-1                 All Taxes
## Good taxpayer at t=-2                 All Taxes
## Good taxpayer at t=-3                 All Taxes
## Good taxpayer at t=-4                 All Taxes
## Good taxpayer at t=-5                 All Taxes
## Automatic debit at t=0  Property, head & sewage
## Automatic debit at t=-1 Property, head & sewage
## Automatic debit at t=-2 Property, head & sewage
## Automatic debit at t=-3 Property, head & sewage
```

```
## 2004 Property value     Property, head & sewage
## Current property value  Property, head & sewage
## Rented Property         Property, head & sewage
## Retiree                                Property
## Paid year in full                      Property
```

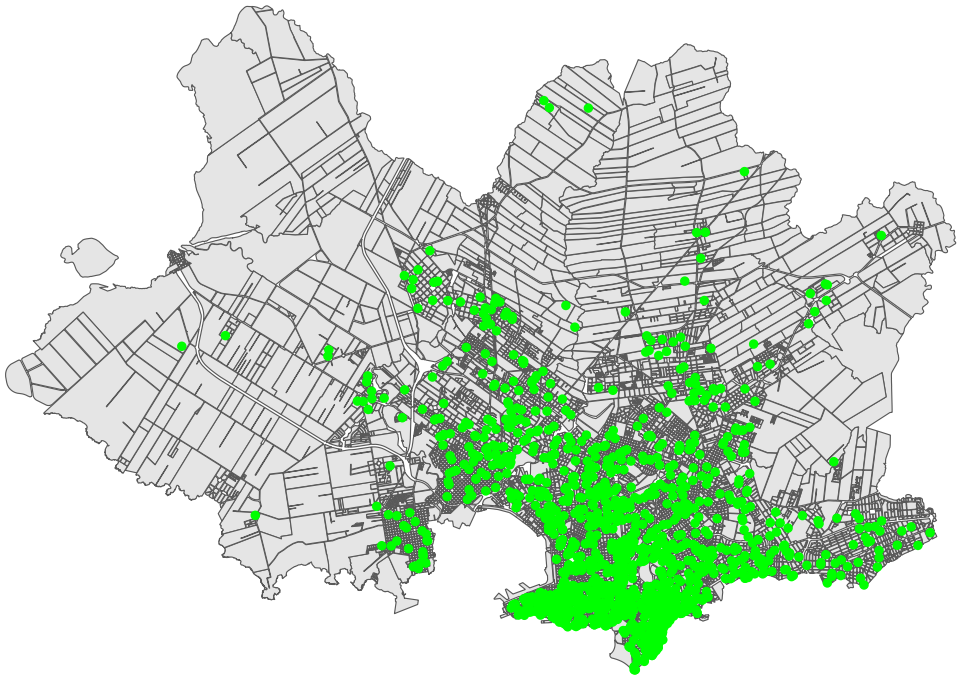
```
##############################################
message('APPENDIX: Figure A3')
```

```
## APPENDIX: Figure A3
```

```
# Natural Experiment: Property Plots of Winning Account Numbers.


# Winning Account number
#map
map<-ggplot() +
  geom_sf(data = shp2, aes(geometry = geometry))+
  geom_sf(data = naturalex_data_map[naturalex_data_map$TREATMENT==1,], aes(geometry = geometry),color="r
  #facet_wrap(~TRIBUTO)+
  theme_classic()+
  theme(plot.background = element_rect(fill = "white"),
        panel.background = element_rect(colour = "white",),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        axis.line = element_blank(),
        axis.ticks = element_blank(),
        axis.text = element_blank(),
        axis.title = element_blank(),
        legend.direction = 'horizontal',
        legend.position = 'bottom',
        legend.background = element_blank())+
  coord_sf()


map
```



```
# Non-Winning Account number
map<-ggplot() +
  geom_sf(data = shp2, aes(geometry = geometry))+
  geom_sf(data = naturalex_data_map[naturalex_data_map$TREATMENT==0,], aes(geometry = geometry),color="
  #facet_wrap(~TRIBUTO)+
  theme_classic()+
  theme(plot.background = element_rect(fill = "white"),
        panel.background = element_rect(colour = "white",),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        axis.line = element_blank(),
        axis.ticks = element_blank(),
        axis.text = element_blank(),
        axis.title = element_blank(),
        legend.direction = 'horizontal',
        legend.position = 'bottom',
        legend.background = element_blank())+
  coord_sf()


map
```



```
################################################
message("APPENDIX: Table A2")
```

```
## APPENDIX: Table A2
```

```
#Natural Experiment: Placebo Tests with Post-Treatment Variables - Ineligible Taxpayers.
#Winning vs. Non-Winning Account Numbers.

btp <- gtp_panel %>% filter(ES_BP==0)

placebo <- rbind.data.frame(
  with(btp %>% filter(centered==4), ttest(BP_time, TREATMENT)),
  with(btp %>% filter(centered==5), ttest(BP_time, TREATMENT)),
  with(btp %>% filter(centered==6), ttest(BP_time, TREATMENT)),
  with(btp %>% filter(centered==7), ttest(BP_time, TREATMENT)),
  with(btp %>% filter(centered==8), ttest(BP_time, TREATMENT)),
  with(btp %>% filter(centered %in% 4:8) %>% group_by(CUENTA, TREATMENT) %>%
         dplyr::summarize(BP_time = mean(BP_time)),
       ttest(BP_time, TREATMENT))
)
```

```
## `summarise()` has grouped output by 'CUENTA'. You can override using the
## `.groups` argument.
```

```
names(placebo) <- c("Mean 1", "Mean 0", "Difference", "SE_Diff",
                    "t-stat","N","df","p-value")
```

```
placebo$outcome <- c("Good taxpayer t=1",
                     "Good taxpayer t=2",
                     "Good taxpayer t=3",
                     "Good taxpayer t=4",
                     "Good taxpayer t=5",
                     "Good taxpayer t=1-5")

placebo
```

```
##     Mean 1   Mean 0 Difference   SE_Diff  t-stat     N    df p-value
## 1 0.090345 0.084775  0.0055700 0.0052404 1.06290 11639 11622 0.28785
## 2 0.102573 0.095222  0.0073515 0.0055708 1.31964 11489 11465 0.18698
## 3 0.116531 0.114190  0.0023402 0.0060008 0.38998 11339 11332 0.69656
## 4 0.118282 0.115829  0.0024528 0.0060785 0.40352 11192 11185 0.68658
## 5 0.119650 0.115447  0.0042035 0.0061281 0.68594 11052 11041 0.49277
## 6 0.109675 0.104818  0.0048571 0.0052155 0.93129 11639 11622 0.35172
##               outcome
## 1   Good taxpayer t=1
## 2   Good taxpayer t=2
## 3   Good taxpayer t=3
## 4   Good taxpayer t=4
## 5   Good taxpayer t=5
## 6 Good taxpayer t=1-5
```

```
################################################
message("APPENDIX: Figure A4(a)")
```

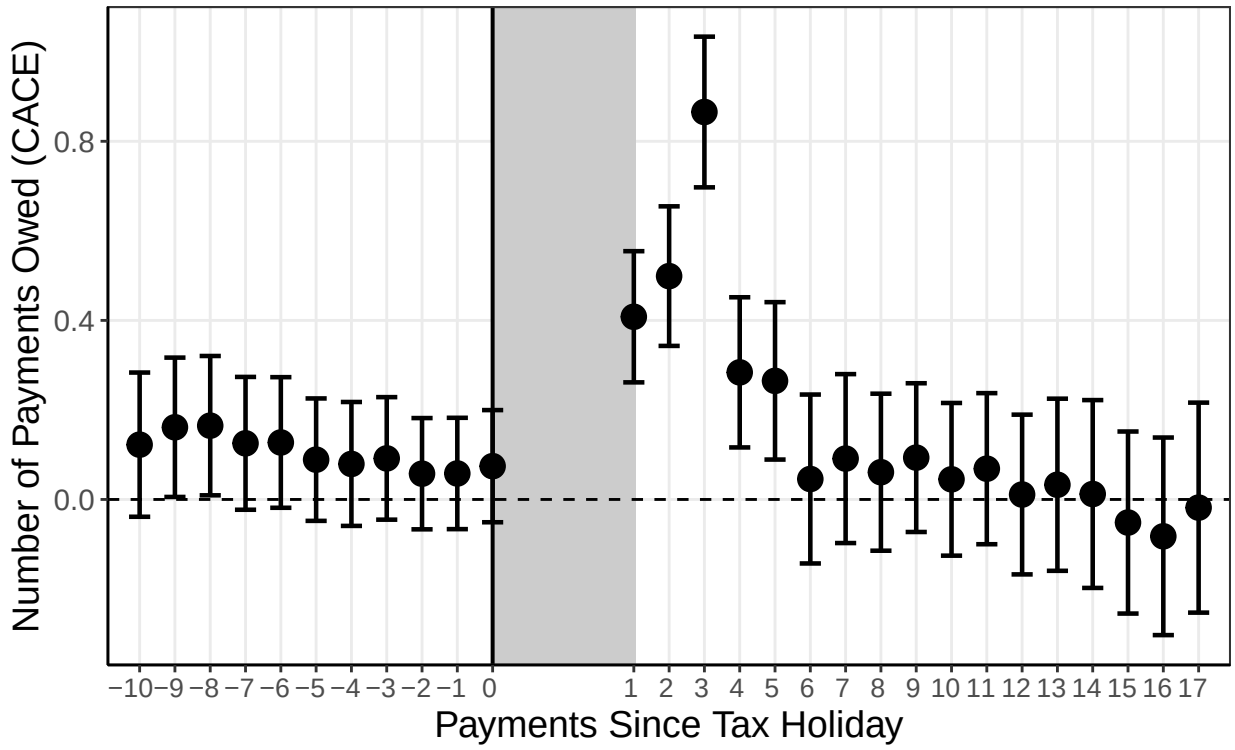
```
## APPENDIX: Figure A4(a)
```

```
#Natural Experiment: The Negative Impact of Holidays on Compliance.
#Effects on the Number of Payments Owed

t <- unique(taxes_panel$t_st)
t <- t[order(t)]
t <- t[t %in% -10:20]
gtp_plot <- NULL

#CACE estimation for all periods
for (i in 1:length(t)){

  temp <- taxes_panel[taxes_panel$t_st == t[i], ]
  if (nrow(temp[temp$TREATMENT==1 & temp$ES_BP==0,])==0) next
  if (nrow(temp[temp$TREATMENT==0 & temp$ES_BP==0,])==0) next

  on_time <- t.test(nr_paymntsowed ~ TREATMENT,
                    data=temp[temp$ES_BP==0,])
  on_time <- c(on_time$estimate[2]-on_time$estimate[1],
               -on_time$conf.int[1:2])

  gtp_plot <- rbind(gtp_plot,
                    as.vector(c("Paid on Time", t[i], 0, on_time)))
```

```
  if (nrow(temp[temp$TREATMENT==1 & temp$ES_BP==1,])==0) next
  if (nrow(temp[temp$TREATMENT==0 & temp$ES_BP==1,])==0) next

  ivest <- tidy(iv_robust(nr_paymntsowed ~ TIENE_EXO | TREATMENT,
                          data=temp[temp$ES_BP==1,]))
  on_time <- c(ivest$estimate[2], ivest$std.error[2])



  on_time <- c(on_time[1], on_time[1]-1.96*on_time[2], on_time[1]+1.96*on_time[2])

  gtp_plot <- rbind(gtp_plot,
                    as.vector(c("Paid on Time", t[i], 1, on_time)))

}

gtp_plot <- as.data.frame(gtp_plot)
names(gtp_plot) <- c("outcome", "t", "ES_BP", "CACE", "upper", "lower")

gtp_plot$t <- as.numeric(as.character(gtp_plot$t))
gtp_plot$CACE <- as.numeric(as.character(gtp_plot$CACE))
gtp_plot$upper <- as.numeric(as.character(gtp_plot$upper))
gtp_plot$lower <- as.numeric(as.character(gtp_plot$lower))

#Figure
ggplot(gtp_plot[gtp_plot$ES_BP==1,], aes(x=t, y=CACE)) +
  geom_rect(data=NULL,aes(xmin=0, xmax=4, ymin=-Inf, ymax=Inf),
            fill="gray80", color="gray80") +
  geom_errorbar(aes(x=t,
                    ymin=lower,
                    ymax=upper),
                width=.6, size=.8, position = position_dodge(width = 0.6)) +
  geom_point(size=4, position = position_dodge(width = 0.6)) +
  xlab("Payments Since Tax Holiday") +
  ylab("Number of Payments Owed (CACE)") +
  geom_vline(aes(xintercept=0), size=.7) +
  geom_hline(aes(yintercept=0), size=.5, linetype="dashed") +
  theme_bw() +
  scale_x_discrete(limit = c(-10:0,4:20),
                   labels = as.character(-10:17)) +
  scale_colour_manual(values = c("black","blue")) +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text.y = element_text(size = rel(1.25)),
        axis.title.y = element_text(size = rel(1.3)),
        axis.title.x = element_text(size = rel(1.3)),
        legend.text = element_text(size = rel(1.2)),
        strip.text.x = element_text(size = rel(1.4)),
        strip.text.y = element_text(size = rel(1.4)),
        strip.background = element_rect(size = 1.5),
        axis.text.x = element_text(size = rel(1.1), hjust=.7),
        legend.position = "bottom",
        legend.title=element_blank(),
        panel.grid.minor = element_blank(),
```

```
        axis.line = element_line(colour = "black"))
```

```
## Warning: Continuous limits supplied to discrete scale.
## i Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```




```
##############################################
message("APPENDIX: Figure A4(b)")
```

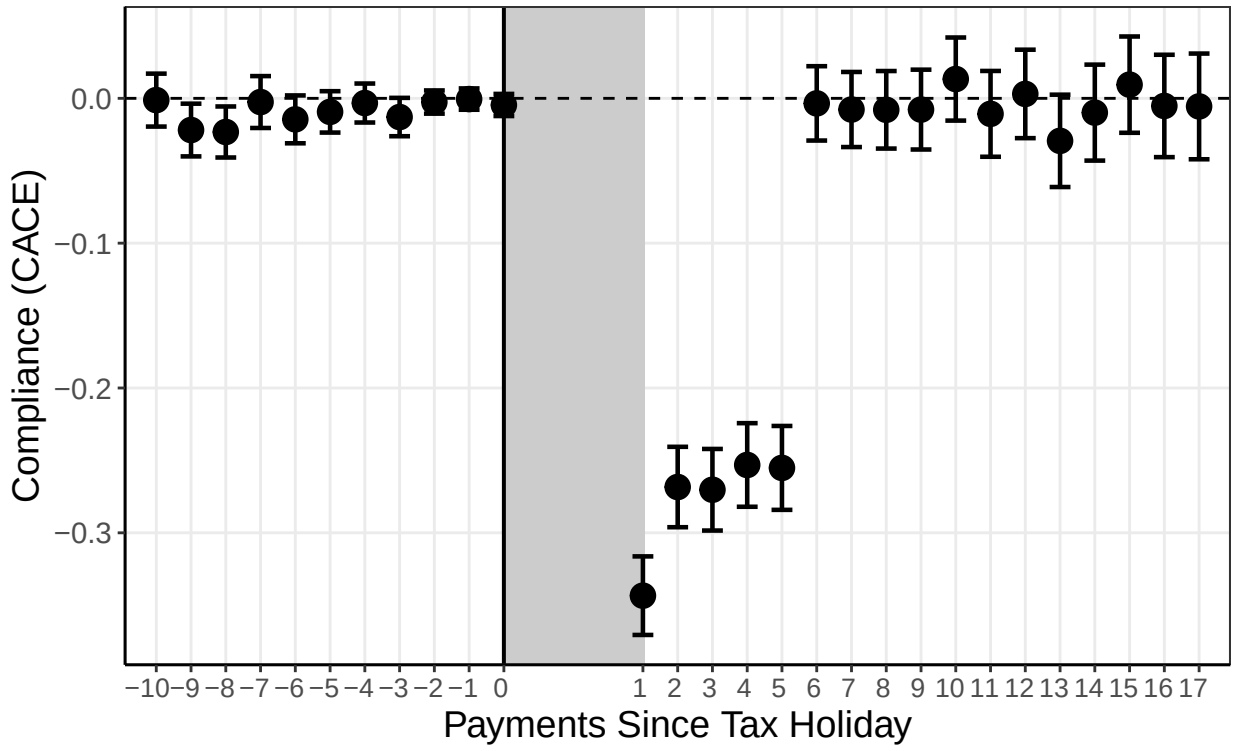
```
## APPENDIX: Figure A4(b)
```

```
#Natural Experiment: The Negative Impact of Holidays on Compliance.
#Effects on the Compliance

gtp_plot <- NULL

#CACE estimation for all periods
for (i in 1:length(t)){

  temp <- taxes_panel[taxes_panel$t_st == t[i], ]
  if (nrow(temp[temp$TREATMENT==1 & temp$ES_BP==0,])==0) next
  if (nrow(temp[temp$TREATMENT==0 & temp$ES_BP==0,])==0) next

  on_time <- t.test(compliance ~ TREATMENT,
                    data=temp[temp$ES_BP==0,])
  on_time <- c(on_time$estimate[2]-on_time$estimate[1],
               -on_time$conf.int[1:2])
```

```
  gtp_plot <- rbind(gtp_plot,
                    as.vector(c("Paid on Time", t[i], 0, on_time)))

  if (nrow(temp[temp$TREATMENT==1 & temp$ES_BP==1,])==0) next
  if (nrow(temp[temp$TREATMENT==0 & temp$ES_BP==1,])==0) next

  ivest <- tidy(iv_robust(compliance ~ TIENE_EXO | TREATMENT,
                          data=temp[temp$ES_BP==1,]))
  on_time <- c(ivest$estimate[2], ivest$std.error[2])

  on_time <- c(on_time[1], on_time[1]-1.96*on_time[2], on_time[1]+1.96*on_time[2])

  gtp_plot <- rbind(gtp_plot,
                    as.vector(c("Paid on Time", t[i], 1, on_time)))

}

gtp_plot <- as.data.frame(gtp_plot)
names(gtp_plot) <- c("outcome", "t", "ES_BP", "CACE", "upper", "lower")

gtp_plot$t <- as.numeric(as.character(gtp_plot$t))
gtp_plot$CACE <- as.numeric(as.character(gtp_plot$CACE))
gtp_plot$upper <- as.numeric(as.character(gtp_plot$upper))
gtp_plot$lower <- as.numeric(as.character(gtp_plot$lower))

#Figure
ggplot(gtp_plot[gtp_plot$ES_BP==1,], aes(x=t, y=CACE)) +
  geom_rect(data=NULL,aes(xmin=0, xmax=4, ymin=-Inf, ymax=Inf),
            fill="gray80", color="gray80") +
  geom_errorbar(aes(x=t,
                    ymin=lower,
                    ymax=upper),
                width=.6, size=.8, position = position_dodge(width = 0.6)) +
  geom_point(size=4, position = position_dodge(width = 0.6)) +
  xlab("Payments Since Tax Holiday") +
  ylab("Compliance (CACE)") +
  geom_vline(aes(xintercept=0), size=.7) +
  geom_hline(aes(yintercept=0), size=.5, linetype="dashed") +
  theme_bw() +
  scale_x_discrete(limit = c(-10:0,4:20),
                   labels = as.character(-10:17)) +
  scale_colour_manual(values = c("black","blue")) +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text.y = element_text(size = rel(1.25)),
        axis.title.y = element_text(size = rel(1.3)),
        axis.title.x = element_text(size = rel(1.3)),
        legend.text = element_text(size = rel(1.2)),
        strip.text.x = element_text(size = rel(1.4)),
        strip.text.y = element_text(size = rel(1.4)),
        strip.background = element_rect(size = 1.5),
        axis.text.x = element_text(size = rel(1.1), hjust=.7),
        legend.position = "bottom",
        legend.title=element_blank(),
```

```
        panel.grid.minor = element_blank(),
        axis.line = element_line(colour = "black"))
```

```
## Warning: Continuous limits supplied to discrete scale.
## i Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```




```
##############################################
message("APPENDIX: Figure A4(c)")
```

```
## APPENDIX: Figure A4(c)
```

```
#Natural Experiment: The Negative Impact of Holidays on Compliance.
#Effects on the Total Debt as of October 2014

difference_in_means(debt_amount ~ won_lottery,
                    data = naturalex_debt_gtp)
```

```
## Design:  Standard
##             Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper     DF
## won_lottery   183.64     296.68 0.61898  0.53596  -397.98   765.26 4947.9
```

```
#N
nrow(naturalex_debt_gtp)
```

```
## [1] 6355
```

```
##############################################
message("APPENDIX: Figure A5")
```

```
## APPENDIX: Figure A5
```

```
# Main Results - Full Post Treatment Period
# The Negative Impact of Holidays on Compliance

table(taxes_panel$TREATMENT, taxes_panel$TIENE_EXO)
```

```
##
##          0      1
##   0 550896      0
##   1 421618 128276
```

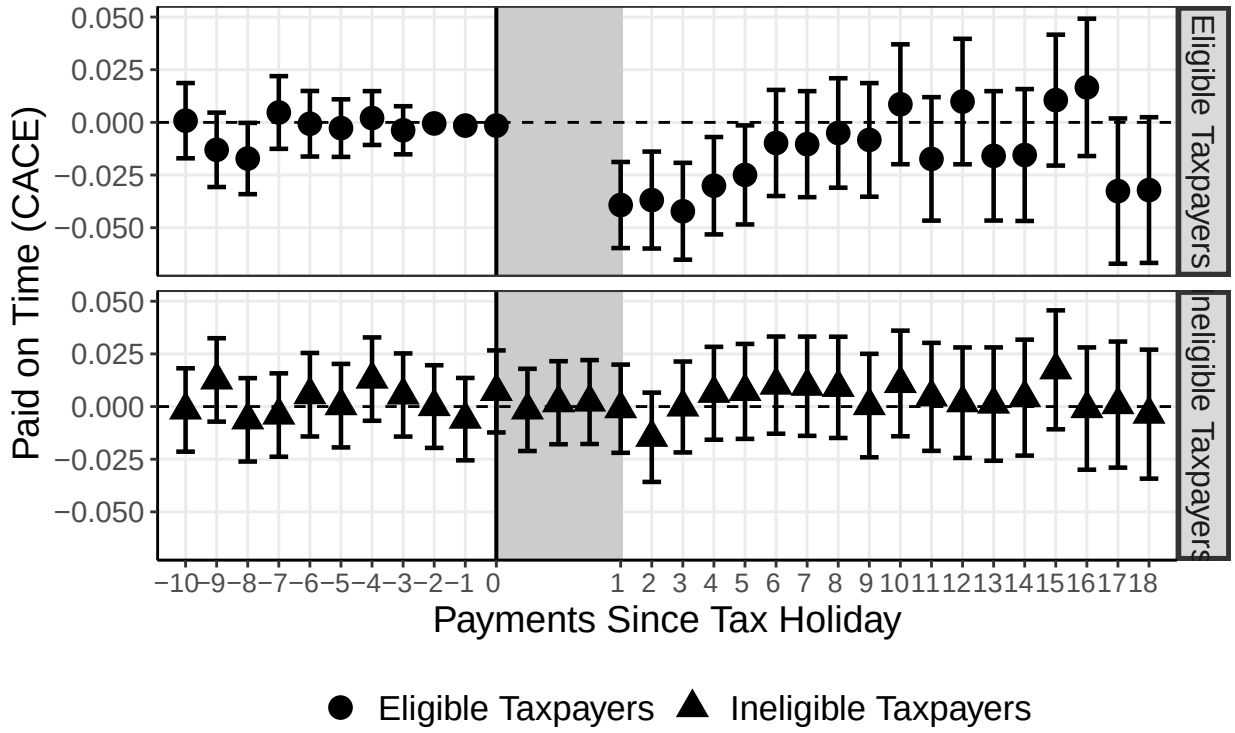
```
t <- -10:21
btp <- NULL
gtp <- NULL

for (i in t){

  temp <- taxes_panel[taxes_panel$t_st == i, ]

  # For Ineligible Taxpayers
  btp_est <- tidy(lm_robust(en_fecha ~ TREATMENT,
                            data = filter(temp, ES_BP==0)))[2,]
  btp <- rbind.data.frame(btp, cbind.data.frame(i, "Ineligible Taxpayers", btp_est))

  # For Eligible Taxpayers
  # For good taxpayers, skip periods under the tax holiday
  if (nrow(temp[temp$TREATMENT==1 & temp$ES_BP==1,])==0) next

  gtp_est <- tidy(iv_robust(en_fecha ~ TIENE_EXO | TREATMENT,
                            data = filter(temp, ES_BP==1)))[2,]
  gtp <- rbind.data.frame(gtp, cbind.data.frame(i, "Eligible Taxpayers", gtp_est))

}
```

```
## Warning in sqrt(diag(vcov_fit$Vcov_hat)): NaNs produced
```

```
names(gtp)[2] <- names(btp)[2] <- "type"
plot <- rbind.data.frame(gtp, btp); rm(gtp, btp)


#Figure
ggplot(plot, aes(x=i, y=estimate, shape=type)) +
  facet_grid(type~.) +
  geom_rect(data=NULL,aes(xmin=0, xmax=4, ymin=-Inf, ymax=Inf),
            fill="gray80", color="gray80") +
  geom_errorbar(aes(x=i,
                    ymin=conf.low,
```

```
                    ymax=conf.high),
                width=.6, size=.8, position = position_dodge(width = 0.6)) +
  geom_point(size=4, position = position_dodge(width = 0.6)) +
  xlab("Payments Since Tax Holiday") +
  ylab("Paid on Time (CACE)") +
  geom_vline(aes(xintercept=0), size=.7) +
  geom_hline(aes(yintercept=0), size=.5, linetype="dashed") +
  theme_bw() +
  scale_x_discrete(limit = t[!t %in% c(1,2,3)],
                   labels = as.character(c(t[t<1],
                                           t[!t%in%c(1,2,3) & t>0]-3))) +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text.y = element_text(size = rel(1.25)),
        axis.title.y = element_text(size = rel(1.3)),
        axis.title.x = element_text(size = rel(1.3)),
        legend.text = element_text(size = rel(1.2)),
        strip.text.x = element_text(size = rel(1.4)),
        strip.text.y = element_text(size = rel(1.4)),
        strip.background = element_rect(size = 1.5),
        axis.text.x = element_text(size = rel(1.1), hjust=.7),
        legend.position = "bottom",
        legend.title=element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(colour = "black"))
```

```
## Warning: Continuous limits supplied to discrete scale.
## i Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```



```
################################################
message("APPENDIX: Figure A6")
```

```
## APPENDIX: Figure A6
```

```
# Treatment Effects By Type of Tax: Holiday vs. No Holiday
# Full Post Treatment Period

gtp_taxes <- taxes_panel[taxes_panel$ES_BP==1, ]
t <- unique(gtp_taxes$t_st)
t <- t[order(t)]
t <- t[t>-11 & t<=21]
gtp_plot <- NULL

# No Holiday
for (i in 1:length(t)){

  temp <- gtp_taxes[gtp_taxes$t_st == t[i], ]

  if (nrow(temp[temp$TREATMENT==1 & temp$holiday_type==0,])==0) next
  if (nrow(temp[temp$TREATMENT==0 & temp$holiday_type==0,])==0) next

  on_time_noholiday <- robust.se(ivreg(en_fecha ~ TIENE_EXO, ~ TREATMENT,
                                       data=temp[temp$holiday_type==0,]))[2,1:2]


  gtp_plot <- rbind(gtp_plot,
                    as.vector(c("Paid on Time", t[i], on_time_noholiday, "No Holiday")))

  print(i)
}
```

```
## [1] "Robust Standard Errors"
## [1] 1
## [1] "Robust Standard Errors"
## [1] 2
## [1] "Robust Standard Errors"
## [1] 3
## [1] "Robust Standard Errors"
## [1] 4
## [1] "Robust Standard Errors"
## [1] 5
## [1] "Robust Standard Errors"
## [1] 6
## [1] "Robust Standard Errors"
## [1] 7
## [1] "Robust Standard Errors"
## [1] 8
## [1] "Robust Standard Errors"
## [1] 12
## [1] "Robust Standard Errors"
## [1] 13
## [1] "Robust Standard Errors"
```

```
## [1] 14
## [1] "Robust Standard Errors"
## [1] 15
## [1] "Robust Standard Errors"
## [1] 16
## [1] "Robust Standard Errors"
## [1] 17
## [1] "Robust Standard Errors"
## [1] 18
## [1] "Robust Standard Errors"
## [1] 19
## [1] "Robust Standard Errors"
## [1] 20
## [1] "Robust Standard Errors"
## [1] 21
## [1] "Robust Standard Errors"
## [1] 22
## [1] "Robust Standard Errors"
## [1] 23
## [1] "Robust Standard Errors"
## [1] 24
## [1] "Robust Standard Errors"
## [1] 25
## [1] "Robust Standard Errors"
## [1] 26
## [1] "Robust Standard Errors"
## [1] 27
## [1] "Robust Standard Errors"
## [1] 28
## [1] "Robust Standard Errors"
## [1] 29
```

```r
#Holiday
for (i in 1:length(t)){

  temp <- gtp_taxes[gtp_taxes$t_st == t[i], ]

  if (nrow(temp[temp$TREATMENT==1 & temp$holiday_type==1,])==0) next
  if (nrow(temp[temp$TREATMENT==0 & temp$holiday_type==1,])==0) next

  on_time_holiday <- robust.se(ivreg(en_fecha ~ TIENE_EXO, ~ TREATMENT,
                                     data=temp[temp$holiday_type==1,]))[2,1:2]

  gtp_plot <- rbind(gtp_plot,
                    as.vector(c("Paid on Time", t[i], on_time_holiday, "Holiday")))

  print(i)
}
```

```
## [1] "Robust Standard Errors"
## [1] 1
## [1] "Robust Standard Errors"
## [1] 2
## [1] "Robust Standard Errors"
```

```
## [1] 3
## [1] "Robust Standard Errors"
## [1] 4
## [1] "Robust Standard Errors"
## [1] 5
## [1] "Robust Standard Errors"
## [1] 6
## [1] "Robust Standard Errors"
## [1] 7
## [1] "Robust Standard Errors"
## [1] 8
## [1] "Robust Standard Errors"
## [1] 9
## [1] "Robust Standard Errors"
## [1] 10
## [1] "Robust Standard Errors"

## Warning in sqrt(diag(se)): NaNs produced

## [1] 11
## [1] "Robust Standard Errors"
## [1] 12
## [1] "Robust Standard Errors"
## [1] 13
## [1] "Robust Standard Errors"
## [1] 14
## [1] "Robust Standard Errors"
## [1] 15
## [1] "Robust Standard Errors"
## [1] 16
## [1] "Robust Standard Errors"
## [1] 17
## [1] "Robust Standard Errors"
## [1] 18
## [1] "Robust Standard Errors"
## [1] 19
## [1] "Robust Standard Errors"
## [1] 20
## [1] "Robust Standard Errors"
## [1] 21
## [1] "Robust Standard Errors"
## [1] 22
## [1] "Robust Standard Errors"
## [1] 23
## [1] "Robust Standard Errors"
## [1] 24
## [1] "Robust Standard Errors"
## [1] 25
## [1] "Robust Standard Errors"
## [1] 26
## [1] "Robust Standard Errors"
## [1] 27
## [1] "Robust Standard Errors"
## [1] 28
```

```
## [1] "Robust Standard Errors"
## [1] 29
```

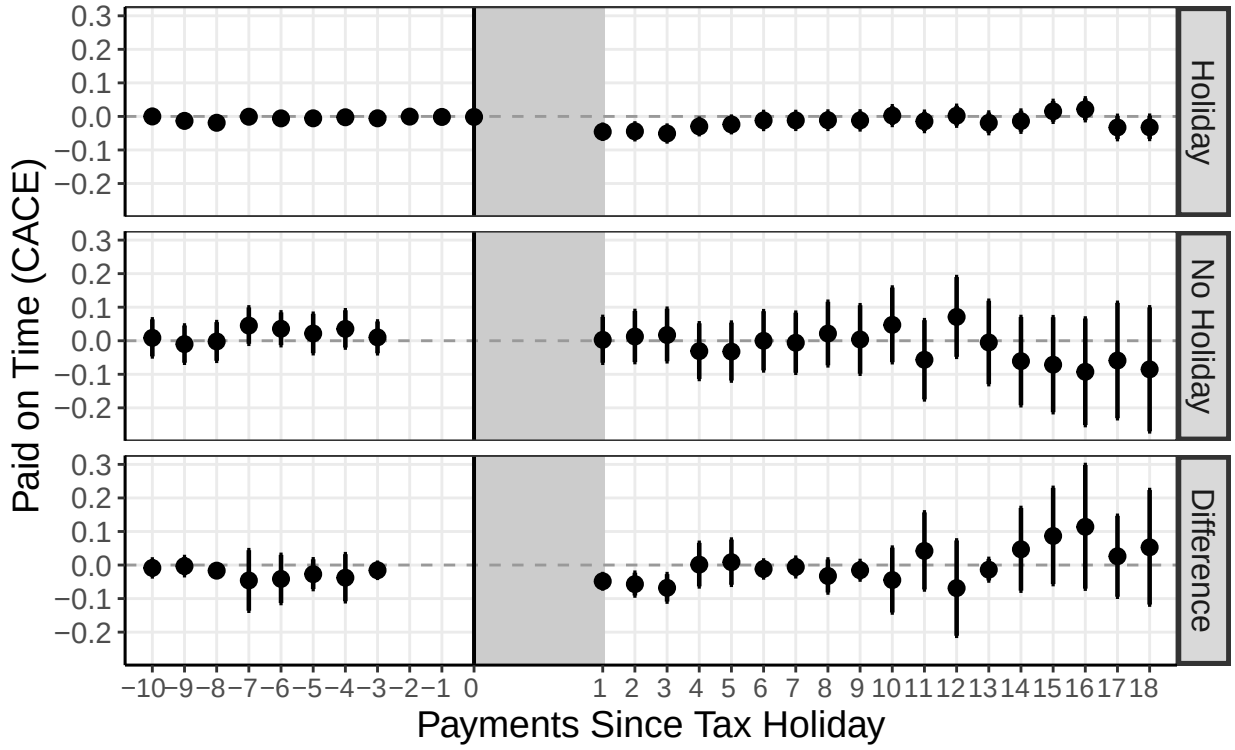
```
gtp_plot <- as.data.frame(gtp_plot)
names(gtp_plot) <- c("outcome", "t", "CACE", "SE", "lottery_type")

gtp_plot$t <- as.numeric(as.character(gtp_plot$t))
gtp_plot$CACE <- as.numeric(as.character(gtp_plot$CACE))
gtp_plot$SE <- as.numeric(as.character(gtp_plot$SE))

# Difference
diff.hte <- NULL

for (i in 1:length(t)){

  temp <- gtp_plot[gtp_plot$t==t[i],]

  if (nrow(temp)!=2) next

  diff <- temp$CACE[temp$lottery_type=="Holiday"] - temp$CACE[temp$lottery_type=="No Holiday"]

  SE <- sqrt((temp$SE[temp$lottery_type=="Holiday"])^2 + (temp$CACE[temp$lottery_type=="No Holiday"])^2

  diff.hte <- rbind.data.frame(diff.hte, as.vector(c(t[i], diff, SE)))

  print(i)
}
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
## [1] 12
## [1] 13
## [1] 14
## [1] 15
## [1] 16
## [1] 17
## [1] 18
## [1] 19
## [1] 20
## [1] 21
## [1] 22
## [1] 23
## [1] 24
## [1] 25
## [1] 26
## [1] 27
## [1] 28
## [1] 29
```

```
names(diff.hte) <- c("t", "CACE", "SE")
diff.hte1 <- diff.hte
diff.hte$lottery_type <- "Difference"
diff.hte$outcome <- "Paid on Time"


gtp_plot <- rbind.data.frame(gtp_plot, diff.hte)

gtp_plot$upper <- gtp_plot$CACE + qnorm(.975) * gtp_plot$SE
gtp_plot$lower <- gtp_plot$CACE - qnorm(.975) * gtp_plot$SE

gtp_plot$lottery_type <- as.factor(gtp_plot$lottery_type)
gtp_plot$lottery_type <- relevel(gtp_plot$lottery_type, ref= "No Holiday")
gtp_plot$lottery_type <- relevel(gtp_plot$lottery_type, ref= "Holiday")

# Calculate the p-value for the difference (one and two-tailed tests)
gtp_plot$t.stat <- gtp_plot$CACE/gtp_plot$SE
gtp_plot$p_value.2 <- 2 * (1 - pnorm(abs(gtp_plot$t.stat)))  # Two-tailed test
gtp_plot$p_value.1 <- (1 - pnorm(abs(gtp_plot$t.stat)))  # One-tailed test
gtp_plot$stars <- ifelse(gtp_plot$p_value.1<=.05 &
                         gtp_plot$t>0, "+", " ") # add one-tailed for post-treatment periods


# Figure
p <- ggplot(gtp_plot[gtp_plot$t<14+8,], aes(x=t, y=CACE))
p + facet_grid(lottery_type~.) +
  geom_rect(data=NULL,aes(xmin=0, xmax=4, ymin=-Inf, ymax=Inf),
            fill="gray80", color="gray80") +
  geom_hline(aes(yintercept=0), size=.5, linetype="dashed", color="gray60") +
  #geom_text(aes(x = t, y = (upper + .02), label = stars), size = 5) +
  geom_errorbar(aes(x=t,
                    ymin=lower,
                    ymax=upper),
                width=.1, size=.8, position = position_dodge(width = 0.5)) +
  geom_point(size=2.5, position = position_dodge(width = 0.5)) +
  xlab("Payments Since Tax Holiday") +
  ylab("Paid on Time (CACE)") +
  geom_vline(aes(xintercept=0), size=.7) +
  theme_bw() +
  scale_colour_manual(values = c("black","blue")) +
  scale_alpha_manual(values = c("FALSE"=0.8, "TRUE"=1), guide='none') +
  scale_x_discrete(limit = c(-10:0,4:21),
                   labels = as.character(-10:18)) +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text.x = element_text(size = rel(1.1), hjust=.7),
        axis.text.y = element_text(size = rel(1.25)),
        axis.title.y = element_text(size = rel(1.3)),
        axis.title.x = element_text(size = rel(1.3)),
        legend.text = element_text(size = rel(1.2)),
        strip.text.x = element_text(size = rel(1.4)),
        strip.text.y = element_text(size = rel(1.4)),
        strip.background = element_rect(size = 1.5),
        legend.position = "none",
```

```
        legend.title=element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(colour = "black"))
```

```
## Warning: Continuous limits supplied to discrete scale.
## i Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```




```
##############################################
message("APPENDIX: Figure A7")
```

```
## APPENDIX: Figure A7
```

```
# Placebo test: Treatment Effects for Automatic vs. Manual Payers
# Full Post Treatment Period

gtp_taxes <- taxes_panel[taxes_panel$ES_BP==1, ]
gtp_taxes <- gtp_taxes[gtp_taxes$TRIBUTO!="Patente de Rodados",]
t <- unique(taxes_panel$t_st)
t <- t[order(t)]
t <- t[t>-11 & t<=28]
gtp_plot <- NULL


# Automatic Payment & Manual Payment
for (i in 1:length(t)){

  temp <- gtp_taxes[gtp_taxes$t_st == t[i], ]
```

```
  if (nrow(temp[temp$TREATMENT==1,])==0) next
  if (nrow(temp[temp$TREATMENT==0,])==0) next

  on_time_auto <- robust.se(ivreg(en_fecha ~ TIENE_EXO, ~ TREATMENT,
                                  data=temp[temp$autopay_win==1,]))[2,1:2]
  on_time_manual <- robust.se(ivreg(en_fecha ~ TIENE_EXO, ~ TREATMENT,
                                    data=temp[temp$autopay_win==0,]))[2,1:2]

  gtp_plot <- rbind(gtp_plot,
                    as.vector(c("Paid on Time", t[i], on_time_auto, "Automatic Payment")),
                    as.vector(c("Paid on Time", t[i], on_time_manual, "Manual Payment")))
  print(i)
}
```

```
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 1
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 2
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 3
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 4
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 5
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 6
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 7
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 8
## [1] "Robust Standard Errors"

## Warning in sqrt(diag(se)): NaNs produced

## [1] "Robust Standard Errors"
## [1] 9
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 10
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 11
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
```

```
## [1] 15
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 16
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 17
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 18
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 19
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 20
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 21
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 22
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 23
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 24
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 25
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 26
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 27
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 28
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 29
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 30
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 31
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 32
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
```

```
## [1] 33
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 34
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 35
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 36
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 37
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 38
## [1] "Robust Standard Errors"
## [1] "Robust Standard Errors"
## [1] 39
```

```r
gtp_plot <- as.data.frame(gtp_plot)
names(gtp_plot) <- c("outcome", "t", "CACE", "SE", "sample")

gtp_plot$t <- as.numeric(as.character(gtp_plot$t))
gtp_plot$CACE <- as.numeric(as.character(gtp_plot$CACE))
gtp_plot$SE <- as.numeric(as.character(gtp_plot$SE))

# Difference
diff.hte <- NULL

for (i in 1:length(t)){

  temp <- gtp_plot[gtp_plot$t==t[i],]

  if (nrow(temp)!=2) next

  diff <- temp$CACE[temp$sample=="Manual Payment"] - temp$CACE[temp$sample=="Automatic Payment"]

  SE <- sqrt((temp$SE[temp$sample=="Manual Payment"])^2 + (temp$CACE[temp$sample=="Automatic Payment"])

  diff.hte <- rbind.data.frame(diff.hte, as.vector(c(t[i], diff, SE)))

  print(i)
}
```

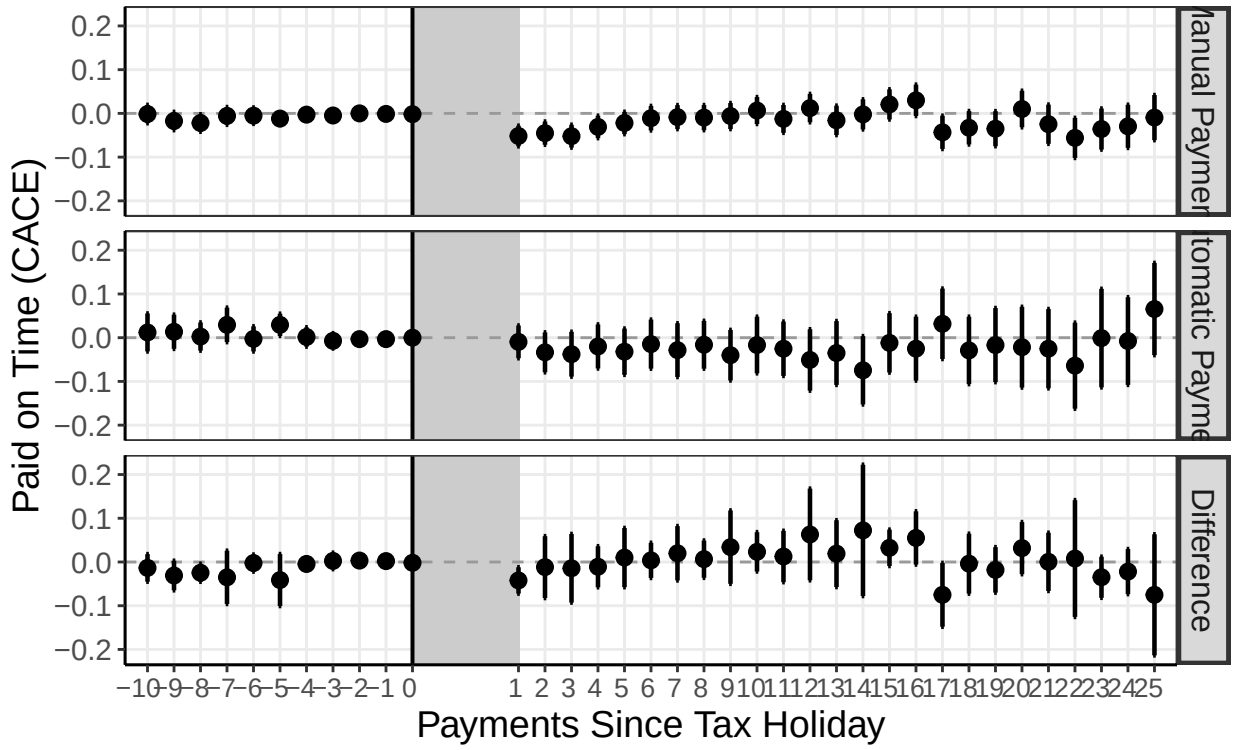
```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
## [1] 9
```

```
## [1] 10
## [1] 11
## [1] 15
## [1] 16
## [1] 17
## [1] 18
## [1] 19
## [1] 20
## [1] 21
## [1] 22
## [1] 23
## [1] 24
## [1] 25
## [1] 26
## [1] 27
## [1] 28
## [1] 29
## [1] 30
## [1] 31
## [1] 32
## [1] 33
## [1] 34
## [1] 35
## [1] 36
## [1] 37
## [1] 38
## [1] 39
```

```
names(diff.hte) <- c("t", "CACE", "SE")
diff.hte2 <- diff.hte
diff.hte$sample <- "Difference"
diff.hte$outcome <- "Paid on Time"


gtp_plot <- rbind.data.frame(gtp_plot, diff.hte)

gtp_plot$upper <- gtp_plot$CACE + qnorm(.975) * gtp_plot$SE
gtp_plot$lower <- gtp_plot$CACE - qnorm(.975) * gtp_plot$SE

gtp_plot$sample <- as.factor(gtp_plot$sample)
gtp_plot$sample <- relevel(gtp_plot$sample, ref= "Automatic Payment")
gtp_plot$sample <- relevel(gtp_plot$sample, ref= "Manual Payment")

# Calculate the p-value for the difference (one and two-tailed tests)
gtp_plot$t.stat <- gtp_plot$CACE/gtp_plot$SE
gtp_plot$p_value.2 <- 2 * (1 - pnorm(abs(gtp_plot$t.stat)))  # Two-tailed test
gtp_plot$p_value.1 <- (1 - pnorm(abs(gtp_plot$t.stat)))  # One-tailed test
gtp_plot$stars <- ifelse(gtp_plot$p_value.1<=.05 &
                           gtp_plot$t>0, "+", " " ) # add one-tailed for post-treatment periods


# Figure
p <- ggplot(gtp_plot[gtp_plot$t<14+15,], aes(x=t, y=CACE))
p + facet_grid(sample~.) +
```

```
  geom_rect(data=NULL,aes(xmin=0, xmax=4, ymin=-Inf, ymax=Inf),
            fill="gray80", color="gray80") +
  geom_hline(aes(yintercept=0), size=.5, linetype="dashed", color="gray60") +
  #geom_text(aes(x = t, y = (upper + .02), label = stars), size = 5) +
  geom_errorbar(aes(x=t,
                    ymin=lower,
                    ymax=upper),
                width=.1, size=.8, position = position_dodge(width = 0.5)) +
  geom_point(size=2.5, position = position_dodge(width = 0.5)) +
  xlab("Payments Since Tax Holiday") +
  ylab("Paid on Time (CACE)") +
  geom_vline(aes(xintercept=0), size=.7) +
  theme_bw() +
  scale_colour_manual(values = c("black","blue")) +
  scale_alpha_manual(values = c("FALSE"=0.8, "TRUE"=1), guide='none') +
  scale_x_discrete(limit = c(-10:0,4:28),
                   labels = as.character(-10:25)) +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text.x = element_text(size = rel(1.1), hjust=.7),
        axis.text.y = element_text(size = rel(1.25)),
        axis.title.y = element_text(size = rel(1.3)),
        axis.title.x = element_text(size = rel(1.3)),
        legend.text = element_text(size = rel(1.2)),
        strip.text.x = element_text(size = rel(1.4)),
        strip.text.y = element_text(size = rel(1.4)),
        strip.background = element_rect(size = 1.5),
        legend.position = "none",
        legend.title=element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(colour = "black"))
```

```
## Warning: Continuous limits supplied to discrete scale.
## i Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```




```
################################################
message("APPENDIX: Figure A8")
```

```
## APPENDIX: Figure A8
```

```
# The Stock of Habit: Perfect vs. Imperfect Past Compliers
# Full Post Treatment Period


taxpayer_type <- ddply(gtp_taxes[gtp_taxes$t_st<=0 & gtp_taxes$t_st>-15,],
                       "CUENTA", summarise,
                       type = mean(en_fecha, na.rm=T))

gtp_taxes <- merge(gtp_taxes, taxpayer_type, by="CUENTA", all.x=T)
t <- unique(gtp_taxes$t_st)
t <- t[order(t)]
t <- t[t>-11 & t<=21]
gtp_plot <- NULL

# Imperfect Past Complier
for (i in 1:length(t)){

  temp <- gtp_taxes[gtp_taxes$t_st == t[i], ]

  if (nrow(temp[temp$TREATMENT==1 & temp$type!=1,])==0) next
  if (nrow(temp[temp$TREATMENT==0 & temp$type!=1,])==0) next

  on_time_marginal <- robust.se(ivreg(en_fecha ~ TIENE_EXO, ~ TREATMENT,
```

```
                                    data=temp[temp$type!=1,]))[2,1:2]


  gtp_plot <- rbind(gtp_plot,
                    as.vector(c("Paid on Time", t[i],
                                on_time_marginal, "Imperfect Past Complier")))
  print(i)
}
```

```
## [1] "Robust Standard Errors"
## [1] 1
## [1] "Robust Standard Errors"
## [1] 2
## [1] "Robust Standard Errors"
## [1] 3
## [1] "Robust Standard Errors"
## [1] 4
## [1] "Robust Standard Errors"
## [1] 5
## [1] "Robust Standard Errors"
## [1] 6
## [1] "Robust Standard Errors"
## [1] 7
## [1] "Robust Standard Errors"
## [1] 8
## [1] "Robust Standard Errors"

## Warning in sqrt(diag(se)): NaNs produced

## [1] 9
## [1] "Robust Standard Errors"

## Warning in sqrt(diag(se)): NaNs produced

## [1] 10
## [1] "Robust Standard Errors"

## Warning in sqrt(diag(se)): NaNs produced

## [1] 11
## [1] "Robust Standard Errors"
## [1] 12
## [1] "Robust Standard Errors"
## [1] 13
## [1] "Robust Standard Errors"
## [1] 14
## [1] "Robust Standard Errors"
## [1] 15
## [1] "Robust Standard Errors"
## [1] 16
## [1] "Robust Standard Errors"
```

```
## [1] 17
## [1] "Robust Standard Errors"
## [1] 18
## [1] "Robust Standard Errors"
## [1] 19
## [1] "Robust Standard Errors"
## [1] 20
## [1] "Robust Standard Errors"
## [1] 21
## [1] "Robust Standard Errors"
## [1] 22
## [1] "Robust Standard Errors"
## [1] 23
## [1] "Robust Standard Errors"
## [1] 24
## [1] "Robust Standard Errors"
## [1] 25
## [1] "Robust Standard Errors"
## [1] 26
## [1] "Robust Standard Errors"
## [1] 27
## [1] "Robust Standard Errors"
## [1] 28
## [1] "Robust Standard Errors"
## [1] 29
```

```r
# Perfect Past Complier
for (i in 1:length(t)){

  temp <- gtp_taxes[gtp_taxes$t_st == t[i], ]

  if (nrow(temp[temp$TREATMENT==1 & temp$type==1,])==0) next
  if (nrow(temp[temp$TREATMENT==0 & temp$type==1,])==0) next

  on_time_compliant <- robust.se(ivreg(en_fecha ~ TIENE_EXO, ~ TREATMENT,
                                       data=temp[temp$type==1,]))[2,1:2]


  gtp_plot <- rbind(gtp_plot,
                    as.vector(c("Paid on Time", t[i],
                                on_time_compliant, "Perfect Past Complier")))
  print(i)
}
```

```
## [1] "Robust Standard Errors"
## [1] 1
## [1] "Robust Standard Errors"
## [1] 2
## [1] "Robust Standard Errors"
## [1] 3
## [1] "Robust Standard Errors"
## [1] 4
## [1] "Robust Standard Errors"
## [1] 5
```

```
## [1] "Robust Standard Errors"
## [1] 6
## [1] "Robust Standard Errors"
## [1] 7
## [1] "Robust Standard Errors"
## [1] 8
## [1] "Robust Standard Errors"
## [1] 9
## [1] "Robust Standard Errors"
## [1] 10
## [1] "Robust Standard Errors"
## [1] 11
## [1] "Robust Standard Errors"
## [1] 12
## [1] "Robust Standard Errors"
## [1] 13
## [1] "Robust Standard Errors"
## [1] 14
## [1] "Robust Standard Errors"
## [1] 15
## [1] "Robust Standard Errors"
## [1] 16
## [1] "Robust Standard Errors"
## [1] 17
## [1] "Robust Standard Errors"
## [1] 18
## [1] "Robust Standard Errors"
## [1] 19
## [1] "Robust Standard Errors"
## [1] 20
## [1] "Robust Standard Errors"
## [1] 21
## [1] "Robust Standard Errors"
## [1] 22
## [1] "Robust Standard Errors"
## [1] 23
## [1] "Robust Standard Errors"
## [1] 24
## [1] "Robust Standard Errors"
## [1] 25
## [1] "Robust Standard Errors"
## [1] 26
## [1] "Robust Standard Errors"
## [1] 27
## [1] "Robust Standard Errors"
## [1] 28
## [1] "Robust Standard Errors"
## [1] 29
```

```r
gtp_plot <- as.data.frame(gtp_plot)
names(gtp_plot) <- c("outcome", "t", "CACE", "SE", "sample")

gtp_plot$t <- as.numeric(as.character(gtp_plot$t))
gtp_plot$CACE <- as.numeric(as.character(gtp_plot$CACE))
```

```r
gtp_plot$SE <- as.numeric(as.character(gtp_plot$SE))

# Difference
diff.hte <- NULL

for (i in 1:length(t)){

  temp <- gtp_plot[gtp_plot$t==t[i],]

  if (nrow(temp)!=2) next

  diff <- temp$CACE[temp$sample=="Imperfect Past Complier"] - temp$CACE[temp$sample=="Perfect Past Compl

  SE <- sqrt((temp$SE[temp$sample=="Imperfect Past Complier"])^2 + (temp$CACE[temp$sample=="Perfect Past

  diff.hte <- rbind.data.frame(diff.hte, as.vector(c(t[i], diff, SE)))

  print(i)
}
```

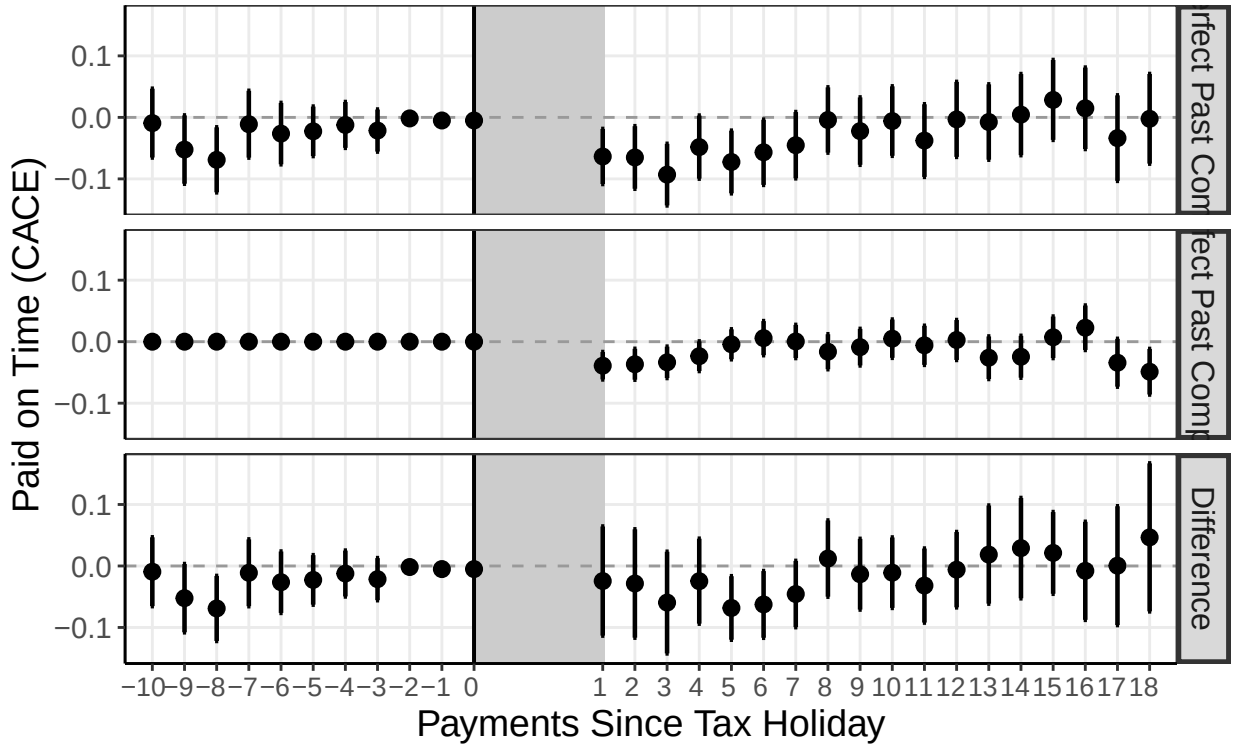
```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
## [1] 9
## [1] 10
## [1] 11
## [1] 12
## [1] 13
## [1] 14
## [1] 15
## [1] 16
## [1] 17
## [1] 18
## [1] 19
## [1] 20
## [1] 21
## [1] 22
## [1] 23
## [1] 24
## [1] 25
## [1] 26
## [1] 27
## [1] 28
## [1] 29
```

```r
names(diff.hte) <- c("t", "CACE", "SE")
diff.hte3 <- diff.hte
diff.hte$sample <- "Difference"
```

```r
diff.hte$outcome <- "Paid on Time"

gtp_plot <- rbind.data.frame(gtp_plot, diff.hte)

gtp_plot$upper <- gtp_plot$CACE + qnorm(.975) * gtp_plot$SE
gtp_plot$lower <- gtp_plot$CACE - qnorm(.975) * gtp_plot$SE

gtp_plot$sample  <- as.factor(gtp_plot$sample)
gtp_plot$sample  <- relevel(gtp_plot$sample, ref= "Perfect Past Complier")
gtp_plot$sample  <- relevel(gtp_plot$sample, ref= "Imperfect Past Complier")

gtp_plot$upper <- gtp_plot$CACE + qnorm(.975) * gtp_plot$SE
gtp_plot$lower <- gtp_plot$CACE - qnorm(.975) * gtp_plot$SE


# Calculate the p-value for the difference (one and two-tailed tests)

gtp_plot$t.stat <- gtp_plot$CACE/gtp_plot$SE
gtp_plot$p_value.2 <- 2 * (1 - pnorm(abs(gtp_plot$t.stat)))  # Two-tailed test
gtp_plot$p_value.1 <- (1 - pnorm(abs(gtp_plot$t.stat)))  # One-tailed test
gtp_plot$stars <- ifelse(gtp_plot$p_value.1<=.05 &
                         gtp_plot$t>0, "+", " ") # add one-tailed for post-treatment periods


# Figure

p <- ggplot(gtp_plot[gtp_plot$t<22,], aes(x=t, y=CACE))
p + facet_grid(sample ~.) +
  geom_rect(data=NULL,aes(xmin=0, xmax=4, ymin=-Inf, ymax=Inf),
            fill="gray80", color="gray80") +
  geom_hline(aes(yintercept=0), size=.5, linetype="dashed", color="gray60") +
  #geom_text(aes(x = t, y = (upper + .02), label = stars), size = 5) +
  geom_errorbar(aes(x=t,
                    ymin=lower,
                    ymax=upper),
                width=.1, size=.8, position = position_dodge(width = 0.5)) +
  geom_point(size=2.5, position = position_dodge(width = 0.5)) +
  xlab("Payments Since Tax Holiday") +
  ylab("Paid on Time (CACE)") +
  geom_vline(aes(xintercept=0), size=.7) +
  theme_bw() +
  scale_colour_manual(values = c("black","blue")) +
  scale_alpha_manual(values = c("FALSE"=0.8, "TRUE"=1), guide='none') +
  scale_x_discrete(limit = c(-10:0,4:21),
                   labels = as.character(-10:18)) +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text.x = element_text(size = rel(1.1), hjust=.7),
        axis.text.y = element_text(size = rel(1.25)),
        axis.title.y = element_text(size = rel(1.3)),
        axis.title.x = element_text(size = rel(1.3)),
        legend.text = element_text(size = rel(1.2)),
        strip.text.x = element_text(size = rel(1.4)),
        strip.text.y = element_text(size = rel(1.4)),
```

```
        strip.background = element_rect(size = 1.5),
        legend.position = "none",
        legend.title=element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(colour = "black"))
```

```
## Warning: Continuous limits supplied to discrete scale.
## i Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```




```
################################################
message("APPENDIX: Figure A9")
```

```
## APPENDIX: Figure A9
```

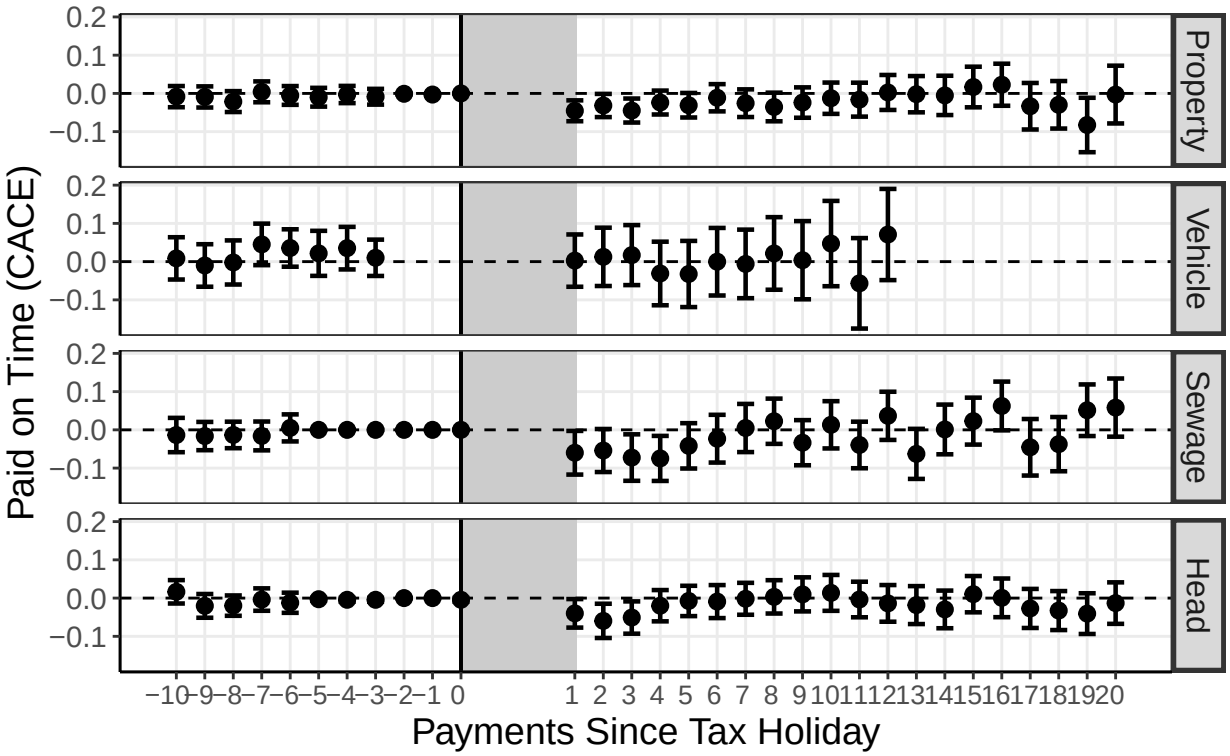
```
# Natural Experiment: Heterogeneous Treatment Effects by Type of Tax.

gtp_taxes <- taxes_panel[taxes_panel$ES_BP==1, ]

t <- unique(gtp_taxes$t_st[gtp_taxes$TREATMENT==1])
t <- t[order(t)]
t <- t[t>-11 & t<24]
taxes <- as.character(unique(taxes_panel$TRIBUTO))
gtp_plot <- NULL

#CACE estimation for all periods
for (i in 1:length(t)){
```

```
  for (j in 1:4){

    temp <- gtp_taxes[gtp_taxes$t_st == t[i] & gtp_taxes$TRIBUTO==taxes[j], ]
    temp$en_fecha[temp$TREATMENT==1]

    if (nrow(temp[temp$TREATMENT==1,])==0) next
    if (nrow(temp[temp$TREATMENT==0,])==0) next

    ivest <- tidy(iv_robust(en_fecha ~ TIENE_EXO | TREATMENT,
                            data=temp))
    on_time <- c(ivest$estimate[2], ivest$std.error[2])

    ivest <- tidy(iv_robust(nr_paymntsowed ~ TIENE_EXO | TREATMENT,
                            data=temp))
    bills_owed <- c(ivest$estimate[2], ivest$std.error[2])


    gtp_plot <- rbind(gtp_plot,
                      as.vector(c("Paid on Time", t[i], on_time, taxes[j])),
                      as.vector(c("Nr. of Bills Owed", t[i], bills_owed, taxes[j]))
    )

  }
}
```

```
## Warning in sqrt(diag(vcov_fit$Vcov_hat)): NaNs produced
```

```
gtp_plot <- as.data.frame(gtp_plot)
names(gtp_plot) <- c("outcome", "t", "CACE", "SE", "tax")

gtp_plot$t <- as.numeric(as.character(gtp_plot$t))
gtp_plot <- gtp_plot[!(gtp_plot$tax=="Patente de Rodados" &
                         gtp_plot$t>15),]
gtp_plot <- gtp_plot[!((gtp_plot$tax=="Contribucion Inmobiliaria" |
                          gtp_plot$tax=="Patente de Rodados") &
                         gtp_plot$t>25),]

gtp_plot$CACE <- as.numeric(as.character(gtp_plot$CACE))
gtp_plot$SE <- as.numeric(as.character(gtp_plot$SE))
gtp_plot$upper <- gtp_plot$CACE + qnorm(.975) * gtp_plot$SE
gtp_plot$lower <- gtp_plot$CACE - qnorm(.975) * gtp_plot$SE

gtp_plot$tax <- as.factor(gtp_plot$tax)
levels(gtp_plot$tax) <- c("Property", "Vehicle", "Sewage", "Head")
gtp_plot$t_label <- ifelse(gtp_plot$t <= 0, as.character(gtp_plot$t),
                           ifelse(gtp_plot$t %in% 1:3, NA, as.character(gtp_plot$t - 3)))


#Figure
ggplot(gtp_plot[gtp_plot$outcome=="Paid on Time",], aes(x=t, y=CACE)) +
  facet_grid(tax~.) +
  geom_rect(data=NULL,aes(xmin=0, xmax=4, ymin=-Inf, ymax=Inf),
            fill="gray80", color="gray80") +
```

```
  geom_errorbar(aes(x=t,
                    ymin=lower,
                    ymax=upper),
                width=.6, size=.8, position = position_dodge(width = 0.5)) +
  geom_point(size=2.5, position = position_dodge(width = 0.5)) +
  xlab("Payments Since Tax Holiday") +
  ylab("Paid on Time (CACE)") +
  geom_vline(aes(xintercept=0), size=.7) +
  geom_hline(aes(yintercept=0), size=.5, linetype="dashed") +
  theme_bw() +
  scale_x_continuous(breaks = unique(gtp_plot$t),
                     labels = unique(gtp_plot$t_label)) +
  scale_colour_manual(values = c("black","blue")) +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text.x = element_text(size = rel(1.1), hjust=.7),
        axis.text.y = element_text(size = rel(1.25)),
        axis.title.y = element_text(size = rel(1.3)),
        axis.title.x = element_text(size = rel(1.3)),
        legend.text = element_text(size = rel(1.2)),
        strip.text.x = element_text(size = rel(1.4)),
        strip.text.y = element_text(size = rel(1.4)),
        strip.background = element_rect(size = 1.5),
        legend.position = "none",
        legend.title=element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(colour = "black"))
```



```r
##############################################
message("APPENDIX: Figure A10")
```

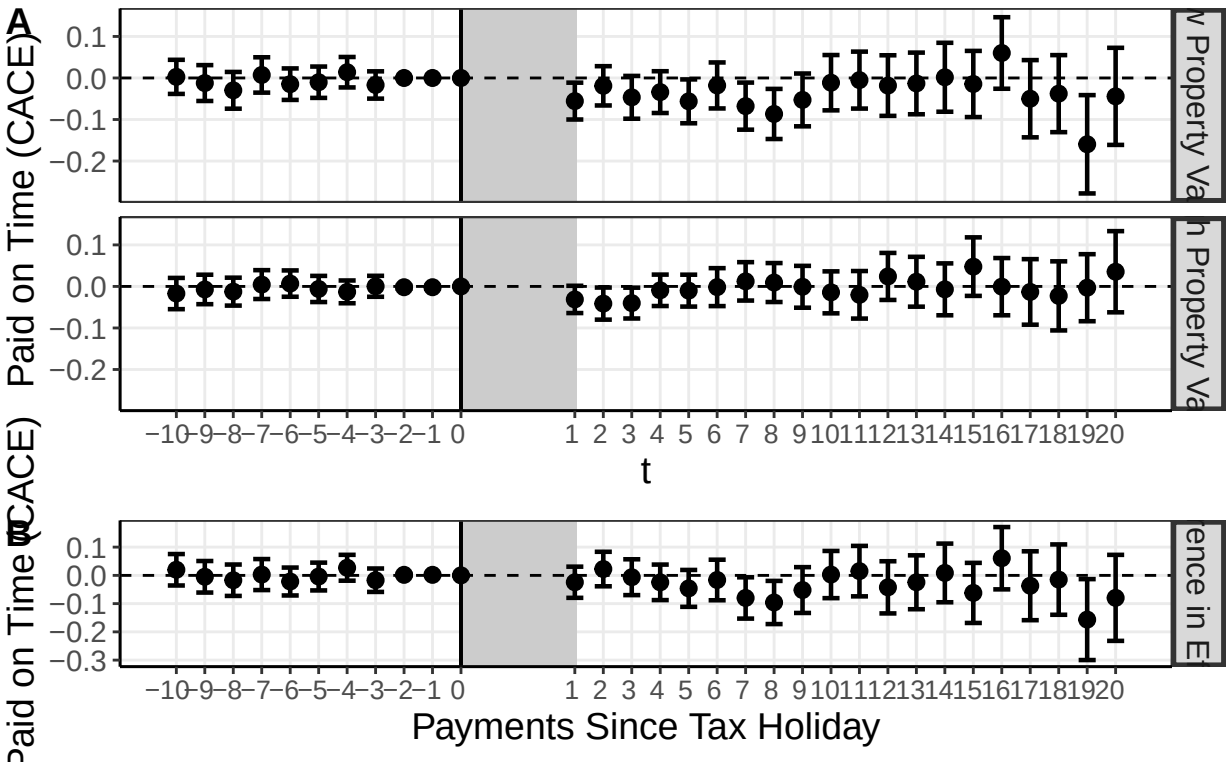
APPENDIX: Figure A10

```r
# Natural Experiment: Heterogeneous Treatment Effects by Property Value.

gtp_taxes <- taxes_panel[taxes_panel$ES_BP==1 &
                           taxes_panel$TRIBUTO=="Contribucion Inmobiliaria", ]

gtp_taxes$high_propvalue <- ifelse(gtp_taxes$VALOR_CAT2004 >
                                     median(gtp_taxes$VALOR_CAT2004, na.rm=T),
                                   1, 0)

t <- unique(gtp_taxes$t_st[gtp_taxes$TREATMENT==1])
t <- t[order(t)]
t <- t[t>-11 & t<24]
pay_month <- as.character(unique(gtp_taxes$bill_month))
gtp_plot <- NULL

gtp_taxes <- gtp_taxes[!is.na(gtp_taxes$high_propvalue),]

#CACE estimation for all periods
for (i in 1:length(t)){

  for (j in c(1,0)){

    temp <- gtp_taxes[gtp_taxes$t_st == t[i] &
                        gtp_taxes$high_propvalue==j, ]

    if (nrow(temp[temp$TREATMENT==1,])==0) next
    if (nrow(temp[temp$TREATMENT==0,])==0) next

    ivest <- tidy(iv_robust(en_fecha ~ TIENE_EXO | TREATMENT,
                            data=temp))
    on_time  <- c(ivest$estimate[2], ivest$std.error[2])

    gtp_plot <- rbind(gtp_plot,
                      as.vector(c("Paid on Time", t[i], on_time, j)))


  }

}

gtp_plot <- as.data.frame(gtp_plot)
names(gtp_plot) <- c("outcome", "t", "CACE", "SE", "prop_value")

gtp_plot$t <- as.numeric(as.character(gtp_plot$t))

gtp_plot$CACE <- as.numeric(as.character(gtp_plot$CACE))
gtp_plot$SE <- as.numeric(as.character(gtp_plot$SE))
gtp_plot$upper <- gtp_plot$CACE + qnorm(.975) * gtp_plot$SE
```

```
gtp_plot$lower <- gtp_plot$CACE - qnorm(.975) * gtp_plot$SE

# Difference
low<-gtp_plot[gtp_plot$prop_value==0,]
high<-gtp_plot[gtp_plot$prop_value==1,]

diff<-data.frame(matrix(ncol=7))
colnames(diff)<-c("outcome","t","CACE","SE","prop_value","upper","lower")

t=1
for (i in unique(gtp_plot$t)){

  aux_1<-low[low$t==i,]
  aux_0<-high[high$t==i,]

  # Estimate
  CACE<-aux_1$CACE-aux_0$CACE
  # Calculate the standard error of the difference between the means
  se_diff <- sqrt((aux_1$SE^2) + (aux_0$SE^2))
  # calculate upper and lower CI
  upper <- CACE + 1.96*se_diff
  lower <- CACE - 1.96*se_diff

  diff[t,'CACE']<-CACE
  diff[t,'SE']<-se_diff
  diff[t,'upper']<-upper
  diff[t,'lower']<-lower
  diff[t,'t']<-paste0(i)
  diff[t,'prop_value']<-'Difference in Effects'
  diff[t,'outcome']<-'Paid on Time'
  t<-t+1
}
rm(aux_1,aux_0,low,high)

# Add to data
gtp_plot<-rbind(gtp_plot,diff);rm(diff)

gtp_plot$t <- as.numeric(as.character(gtp_plot$t))
gtp_plot$t_label <- ifelse(gtp_plot$t <= 0, as.character(gtp_plot$t),
                           ifelse(gtp_plot$t %in% 1:3, NA, as.character(gtp_plot$t - 3)))
gtp_plot$prop_value <- ifelse(gtp_plot$prop_value=='0',"Low Property Value",gtp_plot$prop_value)
gtp_plot$prop_value <- ifelse(gtp_plot$prop_value=='1',"High Property Value",gtp_plot$prop_value)


gtp_plot$prop_value<-forcats::fct_relevel(gtp_plot$prop_value, "Low Property Value","High Property Valu
gtp_plot$t_label <- ifelse(gtp_plot$t <= 0, as.character(gtp_plot$t),
                           ifelse(gtp_plot$t %in% 1:3, NA, as.character(gtp_plot$t - 3)))

#Figure A
fig_a<-ggplot(gtp_plot[gtp_plot$prop_value %in% c("Low Property Value","High Property Value"),], aes(x=
  facet_grid(prop_value ~ .) +
  geom_rect(data=NULL,aes(xmin=0, xmax=4, ymin=-Inf, ymax=Inf),
            fill="gray80", color="gray80") +
```

```
  geom_errorbar(aes(x=t,
                    ymin=lower,
                    ymax=upper),
                width=.6, size=.8, position = position_dodge(width = 0.5)) +
  geom_point(size=2.5, position = position_dodge(width = 0.5)) +
  #xlab("Payments Since Tax Holiday") +
  ylab("Paid on Time (CACE)") +
  scale_x_continuous(breaks = unique(gtp_plot$t),
                     labels = unique(gtp_plot$t_label)) +
  geom_vline(aes(xintercept=0), size=.7) +
  geom_hline(aes(yintercept=0), size=.5, linetype="dashed") +
  theme_bw() +
  scale_colour_manual(values = c("black","blue")) +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text.x = element_text(size = rel(1.1), hjust=.7),
        axis.text.y = element_text(size = rel(1.25)),
        axis.title.y = element_text(size = rel(1.3)),
        axis.title.x = element_text(size = rel(1.3)),
        legend.text = element_text(size = rel(1.2)),
        strip.text.x = element_text(size = rel(1.4)),
        strip.text.y = element_text(size = rel(1.4)),
        strip.background = element_rect(size = 1.5),
        legend.position = "none",
        legend.title=element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(colour = "black"))

#Figure B
fig_b<-ggplot(gtp_plot[gtp_plot$prop_value %in% c('Difference in Effects'),], aes(x=t, y=CACE)) +
  facet_grid(prop_value ~ .) +
  geom_rect(data=NULL,aes(xmin=0, xmax=4, ymin=-Inf, ymax=Inf),
            fill="gray80", color="gray80") +
  geom_errorbar(aes(x=t,
                    ymin=lower,
                    ymax=upper),
                width=.6, size=.8, position = position_dodge(width = 0.5)) +
  geom_point(size=2.5, position = position_dodge(width = 0.5)) +
  xlab("Payments Since Tax Holiday") +
  ylab("Paid on Time (CACE)") +
  scale_x_continuous(breaks = unique(gtp_plot$t),
                     labels = unique(gtp_plot$t_label)) +
  geom_vline(aes(xintercept=0), size=.7) +
  geom_hline(aes(yintercept=0), size=.5, linetype="dashed") +
  theme_bw() +
  scale_colour_manual(values = c("black","blue")) +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text.x = element_text(size = rel(1.1), hjust=.7),
        axis.text.y = element_text(size = rel(1.25)),
        axis.title.y = element_text(size = rel(1.3)),
        axis.title.x = element_text(size = rel(1.3)),
        legend.text = element_text(size = rel(1.2)),
        strip.text.x = element_text(size = rel(1.4)),
        strip.text.y = element_text(size = rel(1.4)),
```

```
        strip.background = element_rect(size = 1.5),
        legend.position = "none",
        legend.title=element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(colour = "black"))

#Final figure
fig<-ggarrange(fig_a,fig_b,ncol = 1,nrow = 2,labels = c('A','B'),heights = c(2, 1))
fig
```




```
############################################
message("APPENDIX TABLE A3")
```

```
## APPENDIX TABLE A3
```

```
# Multiple comparison adjustments for the Natural Experiment

# Main test: Diff-in-diff (three outcomes): 1 year DiD for all taxes pooled

# rescaling the time variable to account for the taxes that have twice as
# many payments per year
taxes_panel$t_st_2 <- ifelse(taxes_panel$tax=="Sewage" | taxes_panel$tax=="Head",
                             taxes_panel$t_st/2, taxes_panel$t_st)


# 1 year diff in diff setup
dd_data <-  rbind.data.frame(taxes_panel %>% filter(ES_BP==1) %>%
```

```
                                group_by(CUENTA, TRIBUTO, TREATMENT) %>% dplyr::summarise(
                                  compliance_mean_DiD_1yr =
                                    mean(compliance[t_st>3 & t_st<=6], na.rm=T) -
                                    mean(compliance[t_st<0 & t_st>=(-3)], na.rm=T),
                                  missed_payment_mean_DiD_1yr =
                                    mean(missed_payment[t_st>3 & t_st<=6], na.rm=T)-
                                    mean(missed_payment[t_st<0 & t_st>=(-3)], na.rm=T),
                                  nr_missed_payments_mean_DiD_1yr =
                                    mean(nr_paymntsowed[t_st> 3 & t_st<=6], na.rm=T)-
                                    mean(nr_paymntsowed[t_st<0 & t_st>=(-3)], na.rm=T),
                                  compliance_mean_DiD_1yr.yr2 =
                                    mean(compliance[t_st>6 & t_st<=9], na.rm=T) -
                                    mean(compliance[t_st<0 & t_st>=(-3)], na.rm=T),
                                  missed_payment_mean_DiD_1yr.yr2 =
                                    mean(missed_payment[t_st>6 & t_st<=9], na.rm=T)-
                                    mean(missed_payment[t_st<0 & t_st>=(-3)], na.rm=T),
                                  nr_missed_payments_mean_DiD_1yr.yr2 =
                                    mean(nr_paymntsowed[t_st> 6 & t_st<=9], na.rm=T)-
                                    mean(nr_paymntsowed[t_st<0 & t_st>=(-3)], na.rm=T)
                                )
)
```

```
## `summarise()` has grouped output by 'CUENTA', 'TRIBUTO'. You can override using
## the `.groups` argument.
```

```
DiD_1yr_compliance <- difference_in_means(compliance_mean_DiD_1yr ~ TREATMENT,
                                          data = dd_data)
DiD_1yr_compliance
```

```
## Design:  Standard
##           Estimate Std. Error t value
## TREATMENT  -0.2061  0.0094951 -21.706
##
## TREATMENT 0.000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
##           CI Lower CI Upper     DF
## TREATMENT -0.22471 -0.18748 4772.5
```

```
DiD_1yr_missed_payment <- difference_in_means(missed_payment_mean_DiD_1yr ~ TREATMENT,
                                              data = dd_data)
DiD_1yr_missed_payment
```

```
## Design:  Standard
##           Estimate Std. Error t value    Pr(>|t|) CI Lower CI Upper     DF
## TREATMENT  0.02783  0.0067317  4.1341 0.000036166 0.014633 0.041026 5413.9
```

```
DiD_1yr_nrmissed_payments <- difference_in_means(nr_missed_payments_mean_DiD_1yr ~ TREATMENT,
                                                 data = dd_data)
DiD_1yr_nrmissed_payments
```

```
## Design:  Standard
##           Estimate Std. Error t value                  Pr(>|t|) CI Lower
```

```
## TREATMENT  0.35802   0.039006  9.1787 0.00000000000000000065324  0.28155
##           CI Upper     DF
## TREATMENT  0.43449 4445.8
```

```
main <- rbind.data.frame(c("Main 1yr DiD - Missed Payment", DiD_1yr_missed_payment$p.value),
                         c("Main 1yr DiD - Nr Missed Payments", DiD_1yr_nrmissed_payments$p.value),
                         c("Main 1yr DiD - Compliance", DiD_1yr_compliance$p.value))

# Persistence of effects, heterogeneous effects (three outcomes)
# Difference between years 1 and 2 for the three outcomes.

persistence_missed <- comp.eff(difference_in_means(missed_payment_mean_DiD_1yr ~ TREATMENT,
                                                   data = dd_data),
                               difference_in_means(missed_payment_mean_DiD_1yr.yr2 ~ TREATMENT,
                                                   data = dd_data))
```

```
## [1] "Difference in Means 1"
## Design:  Standard
##           Estimate Std. Error t value    Pr(>|t|) CI Lower CI Upper     DF
## TREATMENT  0.02783  0.0067317  4.1341 0.000036166 0.014633 0.041026 5413.9
## [1] "Difference in Means 2"
## Design:  Standard
##           Estimate Std. Error t value Pr(>|t|)  CI Lower CI Upper     DF
## TREATMENT 0.017565  0.0070116  2.5051 0.012272 0.0038191 0.031311 5014.2
## [1] "#####. Difference in Effects"
## Diff in effects              SE               t         p-value
##       0.0102649       0.0097199       1.0560633       0.2909636
```

```
persistence_nrowed <- comp.eff(difference_in_means(nr_missed_payments_mean_DiD_1yr ~ TREATMENT,
                                                   data = dd_data),
                               difference_in_means(nr_missed_payments_mean_DiD_1yr.yr2 ~ TREATMENT,
                                                   data = dd_data))
```

```
## [1] "Difference in Means 1"
## Design:  Standard
##           Estimate Std. Error t value                   Pr(>|t|) CI Lower
## TREATMENT  0.35802   0.039006  9.1787 0.00000000000000000065324  0.28155
##           CI Upper     DF
## TREATMENT  0.43449 4445.8
## [1] "Difference in Means 2"
## Design:  Standard
##           Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper   DF
## TREATMENT  0.11354   0.040471  2.8054 0.0050456 0.034195  0.19288 4946
## [1] "#####. Difference in Effects"
## Diff in effects              SE               t         p-value
##      0.24448466      0.05620823      4.34962394      0.00001376
```

```
persistence_compliance <- comp.eff(difference_in_means(compliance_mean_DiD_1yr ~ TREATMENT,
                                                       data = dd_data),
                                   difference_in_means(compliance_mean_DiD_1yr.yr2 ~ TREATMENT,
                                                       data = dd_data))
```

```
## [1] "Difference in Means 1"
## Design:  Standard
##           Estimate Std. Error t value
## TREATMENT  -0.2061  0.0094951 -21.706
##
## TREATMENT 0.0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
##           CI Lower CI Upper     DF
## TREATMENT -0.22471 -0.18748 4772.5
## [1] "Difference in Means 2"
## Design:  Standard
##           Estimate Std. Error t value
## TREATMENT -0.12163  0.0088733 -13.708
##                                                    Pr(>|t|) CI Lower CI Upper
## TREATMENT 0.000000000000000000000000000000000000000051821 -0.13903 -0.10424
##               DF
## TREATMENT 5019.8
## [1] "#####. Difference in Effects"
##    Diff in effects                 SE                  t            p-value
## -0.084466442099960  0.012995886315108 -6.499475299484973  0.000000000084091
```

```
persistence <- rbind.data.frame(c("Persistence yr 1 vs 2  - Missed Payment", persistence_missed[4]),
                                c("Persistence yr 1 vs 2  - Nr Payments Missed", persistence_nrowed[4]),
                                c("Persistence yr 1 vs 2 - Compliance", persistence_compliance[4])
)



# Heterogeneous effects, by cost of payment

# 1 year diff in diff setup
dd_data_vc <-  rbind.data.frame(taxes_panel %>% filter(ES_BP==1 & TRIBUTO=="Contribucion Inmobiliaria") %>%
                                  group_by(CUENTA, TREATMENT, VALOR_CAT2004) %>% dplyr::summarise(
                                    compliance_mean_DiD_1yr =
                                      mean(compliance[t_st>3 & t_st<=6], na.rm=T) -
                                      mean(compliance[t_st<0 & t_st>=(-3)], na.rm=T),
                                    missed_payment_mean_DiD_1yr =
                                      mean(missed_payment[t_st>3 & t_st<=6], na.rm=T)-
                                      mean(missed_payment[t_st<0 & t_st>=(-3)], na.rm=T),
                                    nr_missed_payments_mean_DiD_1yr =
                                      mean(nr_paymntsowed[t_st> 3 & t_st<=6], na.rm=T)-
                                      mean(nr_paymntsowed[t_st<0 & t_st>=(-3)], na.rm=T)
                                  )
)
```

```
## `summarise()` has grouped output by 'CUENTA', 'TREATMENT'. You can override
## using the `.groups` argument.
```

```
dd_data_vc$high_propvalue <- ifelse(dd_data_vc$VALOR_CAT2004 >
                                      median(dd_data_vc$VALOR_CAT2004, na.rm=T),
                                    1, 0)

income_missed <- comp.eff(difference_in_means(missed_payment_mean_DiD_1yr ~ TREATMENT,
                                              data = filter(dd_data_vc, high_propvalue == 1)),
```

```r
                          difference_in_means(missed_payment_mean_DiD_1yr ~ TREATMENT,
                                              data = filter(dd_data_vc, high_propvalue == 0)))
```

```
## [1] "Difference in Means 1"
## Design:  Standard
##           Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper     DF
## TREATMENT 0.034689   0.010999  3.1538 0.0016582 0.013106 0.056273 1031.9
## [1] "Difference in Means 2"
## Design:  Standard
##           Estimate Std. Error t value Pr(>|t|)   CI Lower CI Upper     DF
## TREATMENT 0.018827   0.013383  1.4067   0.1598 -0.0074338 0.045087 1069.1
## [1] "#####. Difference in Effects"
## Diff in effects              SE               t         p-value
##        0.015863        0.017323        0.915687        0.359930
```

```r
income_nrmissed <- comp.eff(difference_in_means(nr_missed_payments_mean_DiD_1yr ~ TREATMENT,
                                                data = filter(dd_data_vc, high_propvalue == 1)),
                            difference_in_means(nr_missed_payments_mean_DiD_1yr ~ TREATMENT,
                                                data = filter(dd_data_vc, high_propvalue == 0)))
```

```
## [1] "Difference in Means 1"
## Design:  Standard
##           Estimate Std. Error t value
## TREATMENT  0.94033   0.047994  19.592
##                                                                                Pr(>|t|)
## TREATMENT 0.00000000000000000000000000000000000000000000000000000000000000000000064371
##           CI Lower CI Upper     DF
## TREATMENT   0.8461   1.0346 724.79
## [1] "Difference in Means 2"
## Design:  Standard
##           Estimate Std. Error t value                              Pr(>|t|)
## TREATMENT  0.81161   0.066725  12.164 0.00000000000000000000000000000000044187
##           CI Lower CI Upper     DF
## TREATMENT  0.68069  0.94253 1132.5
## [1] "#####. Difference in Effects"
## Diff in effects              SE               t         p-value
##        0.128721        0.082193        1.566085        0.117464
```

```r
income_compliance <- comp.eff(difference_in_means(compliance_mean_DiD_1yr ~ TREATMENT,
                                                  data = filter(dd_data_vc, high_propvalue == 1)),
                              difference_in_means(compliance_mean_DiD_1yr ~ TREATMENT,
                                                  data = filter(dd_data_vc, high_propvalue == 0)))
```

```
## [1] "Difference in Means 1"
## Design:  Standard
##           Estimate Std. Error t value
## TREATMENT  -0.5051   0.021872 -23.093
##                                                                                                 P
## TREATMENT 0.000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
##           CI Lower CI Upper    DF
## TREATMENT -0.54803 -0.46217 820.7
```

```
## [1] "Difference in Means 2"
## Design:  Standard
##           Estimate Std. Error t value
## TREATMENT -0.41839   0.022015 -19.005
##                                                                             Pr(>|t|)
## TREATMENT 0.0000000000000000000000000000000000000000000000000000000000000000018416
##           CI Lower CI Upper     DF
## TREATMENT  -0.4616 -0.37518 855.95
## [1] "#####. Difference in Effects"
## Diff in effects             SE              t         p-value
##      -0.0867107       0.0310331      -2.7941387       0.0052463
```

```
income <- rbind.data.frame(c("Income HTE - Missed Payment", income_missed[4]),
                           c("Income HTE - Nr Payments Missed", income_nrmissed[4]),
                           c("Income HTE - Compliance", income_compliance[4]))


# Comparing natural and field experiments
natvsfield <- comp.eff(difference_in_means(missed_payment ~ TREATMENT,
                                           data = filter(taxes_panel, ES_BP==1 & t_st==4)),
                       difference_in_means(JUL_2014_ontime ~ pooled_124_0, weights = pooled_124_0_wts,
                                           data = filter(fieldex, type=="good taxpayer")))
```

```
## [1] "Difference in Means 1"
## Design:  Standard
##           Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper     DF
## TREATMENT 0.028635  0.0075862  3.7746 0.00016197 0.013763 0.043507 5362.7
## [1] "Difference in Means 2"
## Design:  Standard (weighted)
##                Estimate Std. Error t value Pr(>|t|)  CI Lower CI Upper   DF
## pooled_124_0 0.00079172  0.0075116  0.1054  0.91606 -0.013935 0.015518 4565
## [1] "#####. Difference in Effects"
## Diff in effects             SE              t         p-value
##        0.027844        0.010676        2.608069        0.009119
```

```
natvsfield <- rbind.data.frame(c("Natural vs Field Experiment", natvsfield[4]))


names(main) <- names(persistence) <- names(income) <- names(natvsfield) <- c("H","p")
all <- rbind(main, persistence, income, natvsfield)
all$p <- as.numeric(all$p)
all$p.bonferroni <- p.adjust(all$p, method = "bonferroni")
all$p.fdr <- p.adjust(all$p, method = "fdr")

# Round the numeric columns to 3 decimal places
all[, 2:4] <- round(all[, 2:4], 5)

#Table
all
```

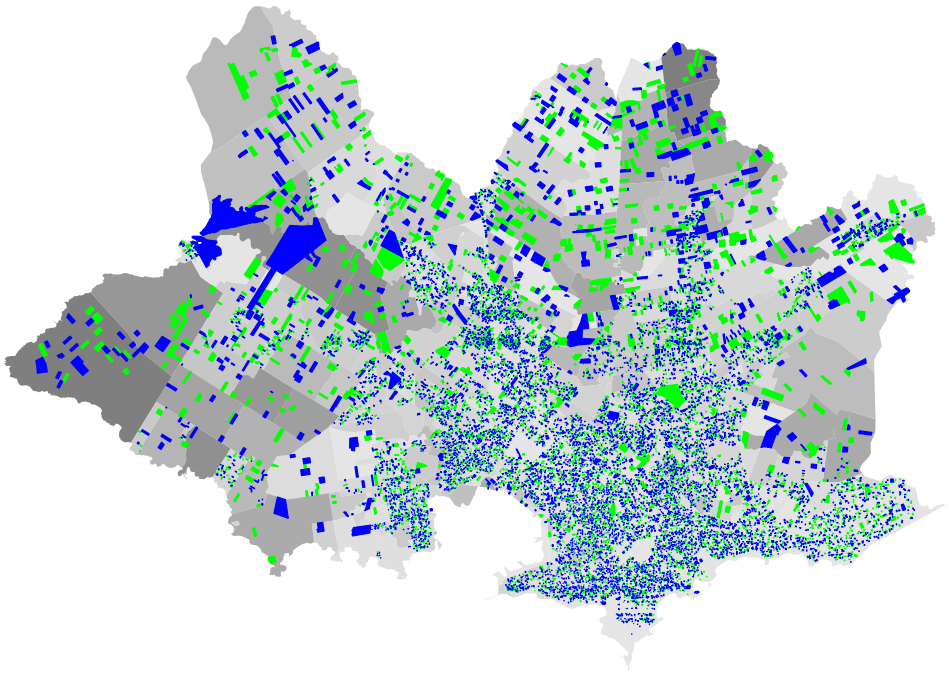
```
##                                              H       p p.bonferroni   p.fdr
## 1               Main 1yr DiD - Missed Payment 0.00004      0.00036 0.00007
## 2            Main 1yr DiD - Nr Missed Payments 0.00000      0.00000 0.00000
```

```
## 3                    Main 1yr DiD - Compliance 0.00000      0.00000 0.00000
## 4      Persistence yr 1 vs 2  - Missed Payment 0.29096      1.00000 0.32329
## 5  Persistence yr 1 vs 2  - Nr Payments Missed 0.00001      0.00014 0.00003
## 6           Persistence yr 1 vs 2 - Compliance 0.00000      0.00000 0.00000
## 7                  Income HTE - Missed Payment 0.35993      1.00000 0.35993
## 8              Income HTE - Nr Payments Missed 0.11746      1.00000 0.14683
## 9                     Income HTE - Compliance 0.00525      0.05246 0.00874
## 10                Natural vs Field Experiment 0.00912      0.09119 0.01303
```

```
###############################################
message('APPENDIX: Figure A18')
```

```
## APPENDIX: Figure A18
```

```
# Field Experiment: Geographic Distribution of
# Eligible and Ineligible Taxpayers


# Map
map<-ggplot() +
  geom_sf(data = shp,aes(fill=PORC_NBI,  geometry = geometry), colour = NA)+
  scale_fill_gradient(low = "grey90", high = "grey50",name = "PORC_NBI") +
  geom_sf(data = datafx_map[datafx_map$ES_BP==1,], aes(geometry = geometry),color='green',fill='green')+
  geom_sf(data = datafx_map[datafx_map$ES_BP==0,], aes(geometry = geometry),color='blue',fill='blue') +
  theme_classic()+
  theme(plot.background = element_rect(fill = "white"),
        panel.background = element_rect(colour = "white",),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        axis.line = element_blank(),
        axis.ticks = element_blank(),
        axis.text = element_blank(),
        axis.title = element_blank(),
        legend.position = 'none')+
  coord_sf()

map
```



```
################################################
message("APPENDIX TABLE A4")
```

```
## APPENDIX TABLE A4
```

```
# Field Experiment: Pre-Treatment Balance

out <- cbind.data.frame(
  c("MAR_2010_ontime", "JUL_2010_ontime", "NOV_2010_ontime", "MAR_2011_ontime",
    "JUL_2011_ontime", "NOV_2011_ontime", "MAR_2012_ontime", "JUL_2012_ontime",
    "NOV_2012_ontime", "MAR_2013_ontime", "JUL_2013_ontime", "NOV_2013_ontime",
    "MAR_2014_ontime"),
  c("Paid on Time MAR 2010", "Paid on Time JUL 2010", "Paid on Time NOV 2010",
    "Paid on Time MAR 2011", "Paid on Time JUL 2011", "Paid on Time NOV 2011",
    "Paid on Time MAR 2012", "Paid on Time JUL 2012", "Paid on Time NOV 2012",
    "Paid on Time MAR 2013", "Paid on Time JUL 2013", "Paid on Time NOV 2013",
    "Paid on Time MAR 2014")
)

treat <- rbind.data.frame(
  c("pooled_124_6", "pooled_124_6_wts", "Treatment versus Pure Control"),
  c("pooled_124_0", "pooled_124_0_wts", "Treatment versus Placebo"),
  c("pooled_0_6", "pooled_0_6_wts", "Placebo versus Pure Control")
)

bal <- merge(treat, out)
names(bal) <- c("treat_var", "wts", "treat_label", "out_var", "out_label")
```

```r
balance <- NULL
for (i in 1:nrow(bal)){

  out <- fieldex[,bal$out_var[i]]
  treat <- fieldex[,bal$treat_var[i]]
  wts <- fieldex[,bal$wts[i]]
  balance <- rbind.data.frame(balance, difference_in_means(out ~ treat,
                                                           weights = wts))
  rm(out, treat, wts)
}

balance <- cbind(bal[,c("treat_label", "out_label")],
                 balance[,c("coefficients", "std.error", "nobs", "p.value")])

#Table
balance
```

```
##                      treat_label             out_label coefficients std.error
## 1  Treatment versus Pure Control Paid on Time MAR 2010   0.00144937 0.0071966
## 2       Treatment versus Placebo Paid on Time MAR 2010   0.00204241 0.0100170
## 3    Placebo versus Pure Control Paid on Time MAR 2010  -0.00123572 0.0087109
## 4  Treatment versus Pure Control Paid on Time JUL 2010   0.00449265 0.0072252
## 5       Treatment versus Placebo Paid on Time JUL 2010   0.00193801 0.0100599
## 6    Placebo versus Pure Control Paid on Time JUL 2010   0.00499282 0.0086563
## 7  Treatment versus Pure Control Paid on Time NOV 2010   0.01951457 0.0072176
## 8       Treatment versus Placebo Paid on Time NOV 2010  -0.00180374 0.0100054
## 9    Placebo versus Pure Control Paid on Time NOV 2010   0.02171193 0.0086319
## 10 Treatment versus Pure Control Paid on Time MAR 2011   0.00844649 0.0071248
## 11      Treatment versus Placebo Paid on Time MAR 2011  -0.00240279 0.0099315
## 12   Placebo versus Pure Control Paid on Time MAR 2011   0.01056837 0.0085213
## 13 Treatment versus Pure Control Paid on Time JUL 2011   0.00759820 0.0070888
## 14      Treatment versus Placebo Paid on Time JUL 2011  -0.00219970 0.0098723
## 15   Placebo versus Pure Control Paid on Time JUL 2011   0.00820028 0.0085152
## 16 Treatment versus Pure Control Paid on Time NOV 2011   0.00091900 0.0070400
## 17      Treatment versus Placebo Paid on Time NOV 2011  -0.00184469 0.0098361
## 18   Placebo versus Pure Control Paid on Time NOV 2011   0.00117599 0.0084731
## 19 Treatment versus Pure Control Paid on Time MAR 2012   0.01151766 0.0069545
## 20      Treatment versus Placebo Paid on Time MAR 2012  -0.00583782 0.0097552
## 21   Placebo versus Pure Control Paid on Time MAR 2012   0.01193612 0.0083481
## 22 Treatment versus Pure Control Paid on Time JUL 2012  -0.00140435 0.0070199
## 23      Treatment versus Placebo Paid on Time JUL 2012   0.00081256 0.0098653
## 24   Placebo versus Pure Control Paid on Time JUL 2012  -0.00567193 0.0085353
## 25 Treatment versus Pure Control Paid on Time NOV 2012  -0.00237368 0.0069520
## 26      Treatment versus Placebo Paid on Time NOV 2012  -0.00711907 0.0097933
## 27   Placebo versus Pure Control Paid on Time NOV 2012   0.00425777 0.0082334
## 28 Treatment versus Pure Control Paid on Time MAR 2013   0.00310250 0.0069097
## 29      Treatment versus Placebo Paid on Time MAR 2013  -0.00737321 0.0097620
## 30   Placebo versus Pure Control Paid on Time MAR 2013   0.00966121 0.0081435
## 31 Treatment versus Pure Control Paid on Time JUL 2013  -0.00703562 0.0067432
## 32      Treatment versus Placebo Paid on Time JUL 2013  -0.00836457 0.0098008
## 33   Placebo versus Pure Control Paid on Time JUL 2013  -0.00083002 0.0077181
## 34 Treatment versus Pure Control Paid on Time NOV 2013   0.00425667 0.0068100
## 35      Treatment versus Placebo Paid on Time NOV 2013   0.00146086 0.0098633
## 36   Placebo versus Pure Control Paid on Time NOV 2013   0.00291111 0.0078577
```

```
## 37 Treatment versus Pure Control Paid on Time MAR 2014  -0.00167924 0.0068263
## 38      Treatment versus Placebo Paid on Time MAR 2014  -0.00314534 0.0098393
## 39   Placebo versus Pure Control Paid on Time MAR 2014   0.00031359 0.0078903
##     nobs   p.value
## 1  16531 0.8403899
## 2  10022 0.8384393
## 3  13241 0.8871936
## 4  16582 0.5340776
## 5  10060 0.8472387
## 6  13282 0.5640932
## 7  16616 0.0068632
## 8  10091 0.8569390
## 9  13297 0.0119041
## 10 16896 0.2358369
## 11 10245 0.8088339
## 12 13507 0.2149139
## 13 16937 0.2838001
## 14 10270 0.8236834
## 15 13541 0.3355571
## 16 16984 0.8961412
## 17 10295 0.8512382
## 18 13583 0.8896167
## 19 17234 0.0977111
## 20 10447 0.5495631
## 21 13777 0.1527959
## 22 17367 0.8414413
## 23 10494 0.9343578
## 24 13883 0.5063650
## 25 17378 0.7327759
## 26 10503 0.4672810
## 27 13895 0.6050703
## 28 17567 0.6534322
## 29 10598 0.4500852
## 30 14065 0.2354949
## 31 17660 0.2967938
## 32 10654 0.3934254
## 33 14134 0.9143599
## 34 17686 0.5319366
## 35 10669 0.8822584
## 36 14155 0.7110303
## 37 17808 0.8056894
## 38 10778 0.7492231
## 39 14244 0.9682983
```

```
################################################
message("APPENDIX TABLE A5")
```

```
## APPENDIX TABLE A5
```

```
#Field Experiment: Treatment Conditions and Sample
## Sizes (Full Experimental Design)
```

```
#Table by type
tab_1<-as.data.frame(table(fieldex$treatment, fieldex$type)) %>%
  pivot_wider(.,names_from = "Var2",values_from = "Freq") %>%
  mutate(Var1= as.character(Var1),
         Var1=case_when(Var1=='0'~'Reminder of Taxes Due (Placebo Control)',
                        Var1=='1'~'Reminder + Lottery/Individual Reward',
                        Var1=='2'~ 'Reminder + Lottery/Individual Reward + Probability of
Winning',
                        Var1=='3'~'Reminder + Individual Punishment',
                        Var1=='4'~'Reminder + Lottery + Social Benefit',
                        Var1=='5'~'Reminder + Social Punishment',
                        Var1=='6'~'Pure Control'),
         treatment_condition=Var1) %>%
  select(-Var1)

#Table total N
tab_2<-as.data.frame(table(fieldex$type)) %>%
  mutate(treatment_condition='TOTAL N') %>%
  pivot_wider(.,names_from = "Var1",values_from = "Freq")

#Final table
tab<-rbind(tab_1,tab_2);rm(tab_1,tab_2)

tab[c(7,1,2,3,4,5,6,8),c(3,1,2)]
```

```
## # A tibble: 8 x 3
##   treatment_condition                                ‘good taxpayer‘ ‘bad taxpayer‘
##   <chr>                                                        <int>          <int>
## 1 "Pure Control"                                                7243           3412
## 2 "Reminder of Taxes Due (Placebo Control)"                     1532           2080
## 3 "Reminder + Lottery/Individual Reward"                         767           1050
## 4 "Reminder + Lottery/Individual Reward + Probab~                751           1043
## 5 "Reminder + Individual Punishment"                            1465           2109
## 6 "Reminder + Lottery + Social Benefit"                         1519           2057
## 7 "Reminder + Social Punishment"                                1507           2111
## 8 "TOTAL N"                                                    14784          13862
```

```
################################################
message("APPENDIX FIGURE A19")
```

```
## APPENDIX FIGURE A19
```

```
# Field Experiment: Complete Results.

# Filter into eligible and ineligible datasets
fieldexE <- fieldex %>% filter(type=="good taxpayer")
fieldexN <- fieldex %>% filter(type=="bad taxpayer")

## Eligibles
web <- as.data.frame(rbind(
  with(fieldexE[(fieldexE$treatment==6 | fieldexE$treatment==0),],
       ttest(JUL_2014_WEBACCESS,(treatment==0)))[c(3:4,8)],
```

```
  with(fieldexE[(fieldexE$treatment==6 | fieldexE$treatment==1),],
       ttest(JUL_2014_WEBACCESS,(treatment==1)))[c(3:4,8)],
  with(fieldexE[(fieldexE$treatment==6 | fieldexE$treatment==3),],
       ttest(JUL_2014_WEBACCESS,(treatment==3)))[c(3:4,8)],
  with(fieldexE[(fieldexE$treatment==6 | fieldexE$treatment==4),],
       ttest(JUL_2014_WEBACCESS,(treatment==4)))[c(3:4,8)],
  with(fieldexE[(fieldexE$treatment==6 | fieldexE$treatment==5),],
       ttest(JUL_2014_WEBACCESS,(treatment==5)))[c(3:4,8)]))
web$outcome <- "Intended Compliance (Accessed Web Account)"

missed <- as.data.frame(rbind(
  with(fieldexE[(fieldexE$treatment==6 | fieldexE$treatment==0),],
       ttest(JUL_2014_ontime,(treatment==0)))[c(3:4,8)],
  with(fieldexE[(fieldexE$treatment==6 | fieldexE$treatment==1),],
       ttest(JUL_2014_ontime,(treatment==1)))[c(3:4,8)],
  with(fieldexE[(fieldexE$treatment==6 | fieldexE$treatment==3),],
       ttest(JUL_2014_ontime,(treatment==3)))[c(3:4,8)],
  with(fieldexE[(fieldexE$treatment==6 | fieldexE$treatment==4),],
       ttest(JUL_2014_ontime,(treatment==4)))[c(3:4,8)],
  with(fieldexE[(fieldexE$treatment==6 | fieldexE$treatment==5),],
       ttest(JUL_2014_ontime,(treatment==5)))[c(3:4,8)]))
missed$outcome <- "Paid Bill On Time"

fieldex_plotE <- rbind(missed, web)
fieldex_plotE$type <- "Eligible Taxpayers"

## Noneligibles
web <- as.data.frame(rbind(
  with(fieldexN[(fieldexN$treatment==6 | fieldexN$treatment==0),],
       ttest(JUL_2014_WEBACCESS,(treatment==0)))[c(3:4,8)],
  with(fieldexN[(fieldexN$treatment==6 | fieldexN$treatment==1),],
       ttest(JUL_2014_WEBACCESS,(treatment==1)))[c(3:4,8)],
  with(fieldexN[(fieldexN$treatment==6 | fieldexN$treatment==3),],
       ttest(JUL_2014_WEBACCESS,(treatment==3)))[c(3:4,8)],
  with(fieldexN[(fieldexN$treatment==6 | fieldexN$treatment==4),],
       ttest(JUL_2014_WEBACCESS,(treatment==4)))[c(3:4,8)],
  with(fieldexN[(fieldexN$treatment==6 | fieldexN$treatment==5),],
       ttest(JUL_2014_WEBACCESS,(treatment==5)))[c(3:4,8)]))
web$outcome <- "Intended Compliance (Accessed Web Account)"

missed <- as.data.frame(rbind(
  with(fieldexN[(fieldexN$treatment==6 | fieldexN$treatment==0),],
       ttest(JUL_2014_ontime,(treatment==0)))[c(3:4,8)],
  with(fieldexN[(fieldexN$treatment==6 | fieldexN$treatment==1),],
       ttest(JUL_2014_ontime,(treatment==1)))[c(3:4,8)],
  with(fieldexN[(fieldexN$treatment==6 | fieldexN$treatment==3),],
       ttest(JUL_2014_ontime,(treatment==3)))[c(3:4,8)],
  with(fieldexN[(fieldexN$treatment==6 | fieldexN$treatment==4),],
       ttest(JUL_2014_ontime,(treatment==4)))[c(3:4,8)],
  with(fieldexN[(fieldexN$treatment==6 | fieldexN$treatment==5),],
       ttest(JUL_2014_ontime,(treatment==5)))[c(3:4,8)]))
missed$outcome <- "Paid Bill On Time"
```

```
fieldex_plotN <- rbind(missed, web)
fieldex_plotN$type <- "Ineligible Taxpayers"

fieldex_plot <- rbind(fieldex_plotE,fieldex_plotN)

names(fieldex_plot) <- c("mean", "se", "p-value", "outcome", "type")

fieldex_plot$upper <- fieldex_plot$mean + qnorm(.975)*(fieldex_plot$se)
fieldex_plot$lower <- fieldex_plot$mean - qnorm(.975)*(fieldex_plot$se)

fieldex_plot$treatment <- rep(c("Reminder",
                                "Individual\n Reward",
                                "Individual\n Punishment",
                                "Social\n Reward",
                                "Social\n Punishment"), 4)

fieldex_plot$treatment <- as.factor(fieldex_plot$treatment)
fieldex_plot$treatment <- factor(fieldex_plot$treatment,
                                 levels = c("Reminder", "Individual\n Reward",
                                            "Individual\n Punishment",
                                            "Social\n Reward",
                                            "Social\n Punishment"))
class(fieldex_plot$treatment)
```

```
## [1] "factor"
```

```
fieldex_plot$type <- as.factor(fieldex_plot$type)



# FDR and Bonferroni corrections

#Threshold for FDR correction
# get and order the nominal p-values
ordered.ps <- fieldex_plot$`p-value`[order(fieldex_plot$`p-value`,decreasing=F)]
ordered.ps
```

```
##  [1] 0.00000000000043949 0.00000000095673935 0.00001579396908720
##  [4] 0.00002988982333668 0.00004163861098084 0.00177493405890218
##  [7] 0.00560698939283693 0.00649614798294518 0.00992842550242769
## [10] 0.07646818678918403 0.08554740761849493 0.14340588290910314
## [13] 0.14508250319634414 0.21101199882512073 0.23634696267254982
## [16] 0.42456695205421358 0.63467888512288695 0.76737312043598871
## [19] 0.83246937184424630 0.83418565969326919
```

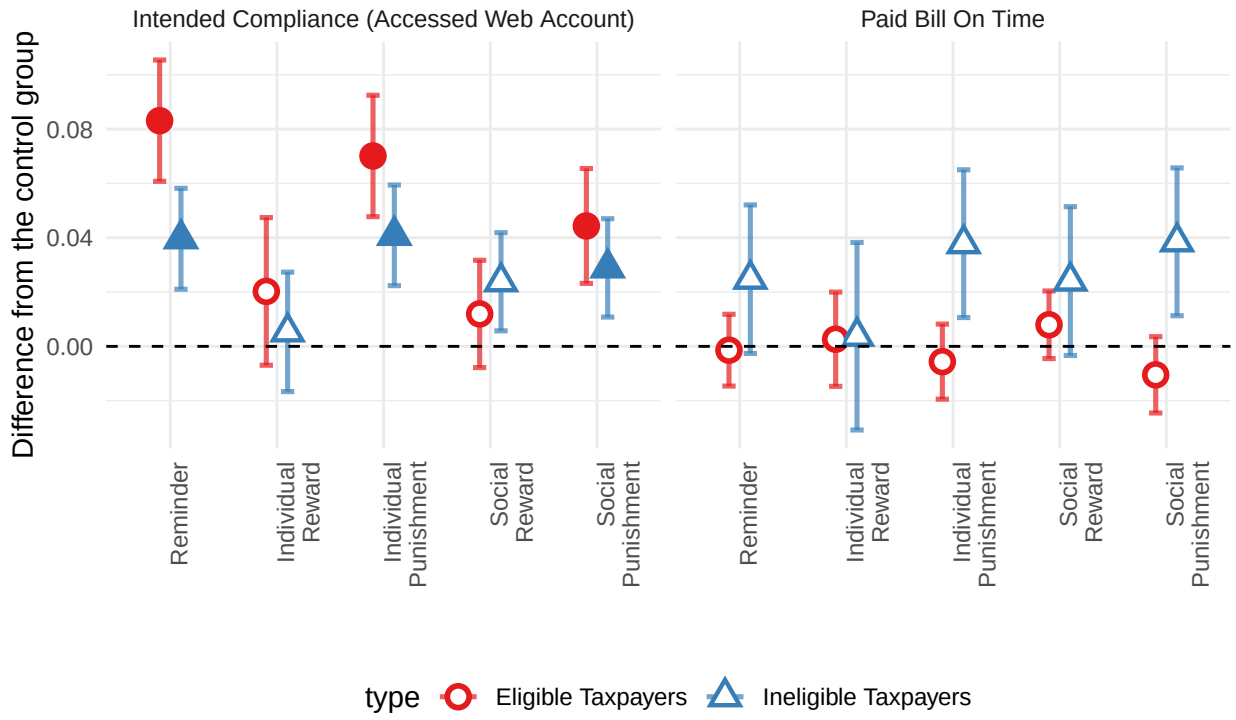
```
comp <- (1:length(ordered.ps)/length(ordered.ps))*(.05)

FDR <- cbind(ordered.ps,comp,ordered.ps<=comp)
fdr <- max(FDR[,1][FDR[,3]==1])
fdr
```

```
## [1] 0.0099284
```

```r
#Threshold for Bonferroni correction
bonf <- 0.05/length(ordered.ps)
bonf
```

```
## [1] 0.0025
```

```r
fieldex_plot$Bonf_reject <- NA
fieldex_plot$FDR_reject <- NA

fieldex_plot$Bonf_reject[fieldex_plot$`p-value`<=bonf] <- "yes"
fieldex_plot$Bonf_reject[fieldex_plot$`p-value`>bonf]  <- "no"

fieldex_plot$FDR_reject[fieldex_plot$`p-value`<=fdr] <- "yes"
fieldex_plot$FDR_reject[fieldex_plot$`p-value`>fdr]  <- "no"

fieldex_plot$bonf_fdr <- as.numeric(fieldex_plot$Bonf_reject=="yes" &
                                      fieldex_plot$FDR_reject=="yes")

pd <- position_dodge(width = 0.4)


#Figure
ggplot(fieldex_plot, aes(x=treatment, y=mean, group=type,
                         color=type, shape=type)) +
  facet_wrap(~outcome) +
  geom_errorbar(aes(x=treatment,
                    ymin=lower, ymax=upper),
                width=.25, size=.9, alpha=.7,
                position=pd) +
  geom_point(size=4.5, position=pd) +
  geom_point(size=2.5, position=pd,
             data=fieldex_plot[fieldex_plot$bonf_fdr==0,],
             aes(x=treatment, y=mean, group=type,
                 shape=type), color = "white") +
  geom_hline(aes(yintercept=0), size=.5, linetype="dashed") +
  xlab(" ") +
  ylab("Difference from the control group") +
  scale_colour_brewer(palette="Set1") +
  theme_minimal() +
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle = 90, hjust = 1))
```

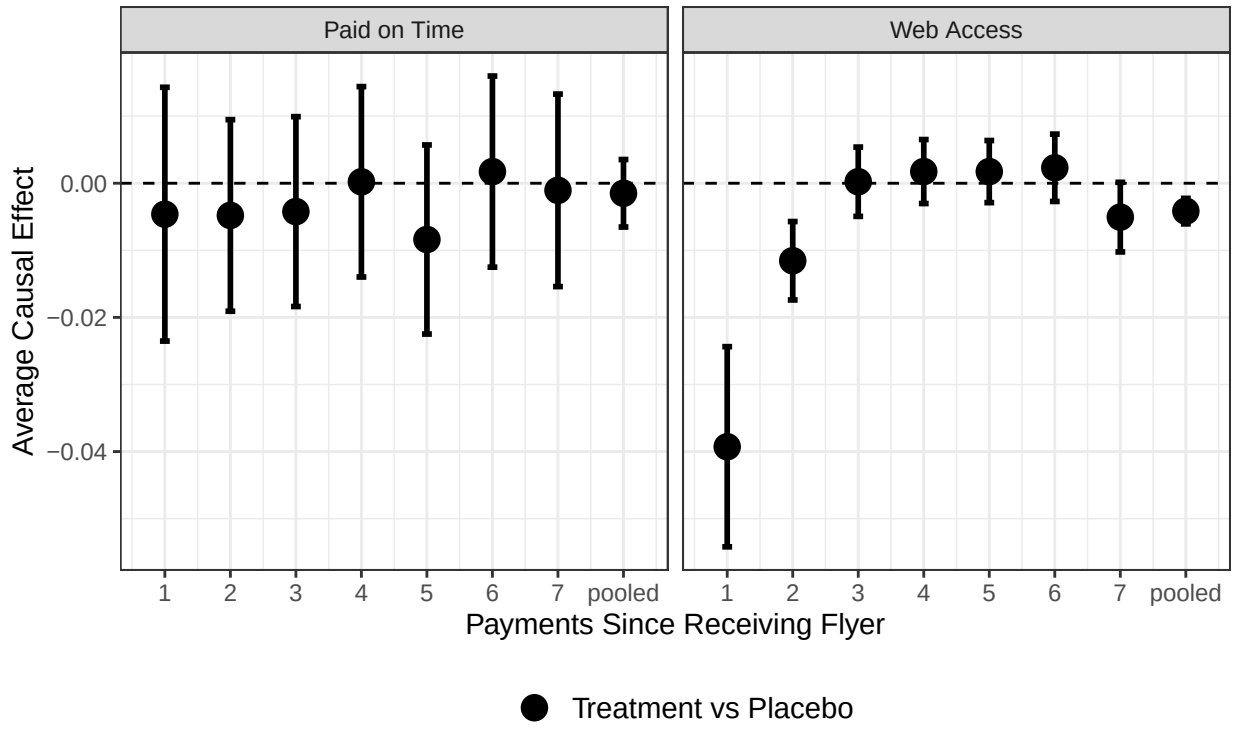



```
################################################
message("APPENDIX FIGURE A20")
```

```
## APPENDIX FIGURE A20
```

```
# Field Experiment: Effects of the Information Treatment
# versus the Placebo Control (no IPW)

# a) Compliance
comp_placebo <- rbind.data.frame(
  difference_in_means(JUL_2014_ontime ~ pooled_124_0, data = fieldex),
  difference_in_means(NOV_2014_ontime ~ pooled_124_0, data = fieldex),
  difference_in_means(MAR_2015_ontime ~ pooled_124_0, data = fieldex),
  difference_in_means(JUL_2015_ontime ~ pooled_124_0, data = fieldex),
  difference_in_means(NOV_2015_ontime ~ pooled_124_0, data = fieldex),
  difference_in_means(MAR_2016_ontime ~ pooled_124_0, data = fieldex),
  difference_in_means(JUL_2016_ontime ~ pooled_124_0, data = fieldex),
  difference_in_means(compliance_1416 ~ pooled_124_0, data = fieldex)
)
comp_placebo$outcome <- "Paid on Time"
comp_placebo$control <- "Treatment vs Placebo"

# b) Intended compliance
intcomp_placebo <- rbind.data.frame(
  difference_in_means(JUL_2014_WEBACCESS ~ pooled_124_0, data = fieldex),
  difference_in_means(NOV_2014_WEBACCESS ~ pooled_124_0, data = fieldex),
  difference_in_means(MAR_2015_WEBACCESS ~ pooled_124_0, data = fieldex),
  difference_in_means(JUL_2015_WEBACCESS ~ pooled_124_0, data = fieldex),
```

```
  difference_in_means(NOV_2015_WEBACCESS ~ pooled_124_0, data = fieldex),
  difference_in_means(MAR_2016_WEBACCESS ~ pooled_124_0, data = fieldex),
  difference_in_means(JUL_2016_WEBACCESS ~ pooled_124_0, data = fieldex),
  difference_in_means(intended_1416 ~ pooled_124_0, data = fieldex)
)
intcomp_placebo$outcome <- "Web Access"
intcomp_placebo$control <- "Treatment vs Placebo"

## Combine and plot
plotdata <- rbind.data.frame(comp_placebo, intcomp_placebo)
rm(comp_placebo, intcomp_placebo)

plotdata$time <- rep(1:8, 2)
#plotdata <- plotdata[plotdata$time!=8,]

plotdata$control <- as.factor(plotdata$control)

pd <- position_dodge(width = 0.6)

#Figure
ggplot(plotdata, aes(x=time, y=coefficients, group = control, shape = control)) +
  facet_wrap( ~ outcome) + #, scales="free"
  geom_point(size=4.5, position=pd) +
  geom_hline(aes(yintercept=0), size=.5, linetype="dashed") +
  geom_errorbar(aes(x=time,
                    ymin=conf.low,
                    ymax=conf.high),
                width=.15, size=1, position=pd) +
  xlab("Payments Since Receiving Flyer") + ylab("Average Causal Effect") +
  theme_bw() +
  scale_colour_manual(values = c("black", "black")) +
  scale_x_continuous(breaks=1:8,
                     labels=c(as.character(1:7), "pooled")) +
  theme(plot.title = element_text(size = rel(1.2)),
        axis.text.x = element_text(size = rel(1)),
        axis.text.y = element_text(size = rel(1)),
        axis.title.y = element_text(size = rel(1)),
        axis.title.x = element_text(size = rel(1)),
        legend.text = element_text(size = rel(1)),
        strip.text.x = element_text(size = rel(1)),
        strip.text.y = element_text(size = rel(1)),
        legend.position = "bottom",
        legend.title=element_blank())
```




```
################################################
message("APPENDIX FIGURE A21")
```

```
## APPENDIX FIGURE A21
```

```
#Field Experiment: Effects of Information About the
# Tax Holiday on Compliance. Heterogeneous Treatment Effects by Taxpayer Type.

#### Eligibles

#### TREATMENT VERSUS PURE CONTROL

# a) compliance
comp_controlE <- rbind.data.frame(
  difference_in_means(JUL_2014_ontime ~ pooled_124_6, data = fieldexE),
  difference_in_means(NOV_2014_ontime ~ pooled_124_6, data = fieldexE),
  difference_in_means(MAR_2015_ontime ~ pooled_124_6, data = fieldexE),
  difference_in_means(JUL_2015_ontime ~ pooled_124_6, data = fieldexE),
  difference_in_means(NOV_2015_ontime ~ pooled_124_6, data = fieldexE),
  difference_in_means(MAR_2016_ontime ~ pooled_124_6, data = fieldexE),
  difference_in_means(JUL_2016_ontime ~ pooled_124_6, data = fieldexE),
  difference_in_means(compliance_1416 ~ pooled_124_6, data = fieldexE)
)
comp_controlE$outcome <- "Paid on Time"
comp_controlE$control <- "Treatment vs Pure Control"
comp_controlE$type <- "Eligibles"

# b) Intended compliance
```

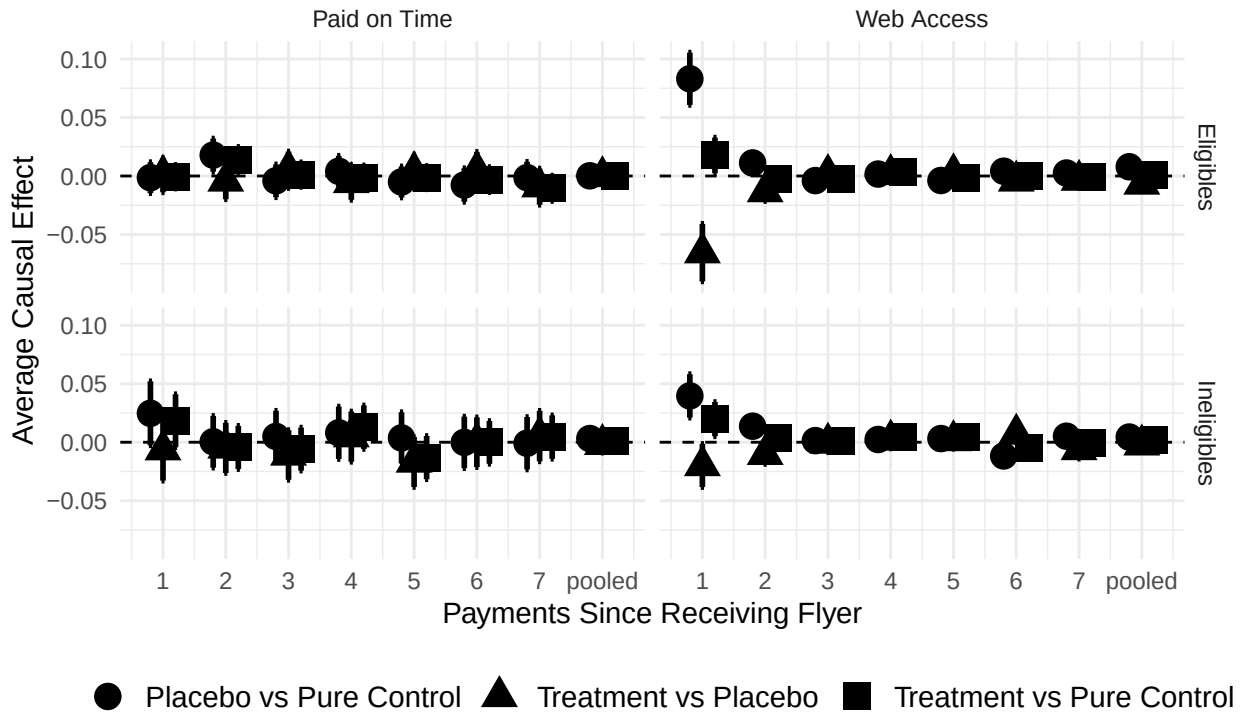
```r
intcomp_controlE <- rbind.data.frame(
  difference_in_means(JUL_2014_WEBACCESS ~ pooled_124_6, data = fieldexE),
  difference_in_means(NOV_2014_WEBACCESS ~ pooled_124_6, data = fieldexE),
  difference_in_means(MAR_2015_WEBACCESS ~ pooled_124_6, data = fieldexE),
  difference_in_means(JUL_2015_WEBACCESS ~ pooled_124_6, data = fieldexE),
  difference_in_means(NOV_2015_WEBACCESS ~ pooled_124_6, data = fieldexE),
  difference_in_means(MAR_2016_WEBACCESS ~ pooled_124_6, data = fieldexE),
  difference_in_means(JUL_2016_WEBACCESS ~ pooled_124_6, data = fieldexE),
  difference_in_means(intended_1416 ~ pooled_124_6, data = fieldexE)
)
intcomp_controlE$control <- "Treatment vs Pure Control"
intcomp_controlE$outcome <- "Web Access"
intcomp_controlE$type <- "Eligibles"

#### TREATMENT VERSUS PLACEBO

# a) Compliance
comp_placeboE <- rbind.data.frame(
  difference_in_means(JUL_2014_ontime ~ pooled_124_0, data = fieldexE),
  difference_in_means(NOV_2014_ontime ~ pooled_124_0, data = fieldexE),
  difference_in_means(MAR_2015_ontime ~ pooled_124_0, data = fieldexE),
  difference_in_means(JUL_2015_ontime ~ pooled_124_0, data = fieldexE),
  difference_in_means(NOV_2015_ontime ~ pooled_124_0, data = fieldexE),
  difference_in_means(MAR_2016_ontime ~ pooled_124_0, data = fieldexE),
  difference_in_means(JUL_2016_ontime ~ pooled_124_0, data = fieldexE),
  difference_in_means(compliance_1416 ~ pooled_124_0, data = fieldexE)
)
comp_placeboE$control <- "Treatment vs Placebo"
comp_placeboE$outcome <- "Paid on Time"
comp_placeboE$type <- "Eligibles"

# b) Intended compliance
intcomp_placeboE <- rbind.data.frame(
  difference_in_means(JUL_2014_WEBACCESS ~ pooled_124_0, data = fieldexE),
  difference_in_means(NOV_2014_WEBACCESS ~ pooled_124_0, data = fieldexE),
  difference_in_means(MAR_2015_WEBACCESS ~ pooled_124_0, data = fieldexE),
  difference_in_means(JUL_2015_WEBACCESS ~ pooled_124_0, data = fieldexE),
  difference_in_means(NOV_2015_WEBACCESS ~ pooled_124_0, data = fieldexE),
  difference_in_means(MAR_2016_WEBACCESS ~ pooled_124_0, data = fieldexE),
  difference_in_means(JUL_2016_WEBACCESS ~ pooled_124_0, data = fieldexE),
  difference_in_means(intended_1416 ~ pooled_124_0, data = fieldexE)
)
intcomp_placeboE$outcome <- "Web Access"
intcomp_placeboE$control <- "Treatment vs Placebo"
intcomp_placeboE$type <- "Eligibles"

#### PURE CONTROL VERSUS PLACEBO

# a) Compliance
comp_control_plaE <- rbind.data.frame(
  difference_in_means(JUL_2014_ontime ~ pooled_0_6, data = fieldexE),
  difference_in_means(NOV_2014_ontime ~ pooled_0_6, data = fieldexE),
  difference_in_means(MAR_2015_ontime ~ pooled_0_6, data = fieldexE),
```

```
  difference_in_means(JUL_2015_ontime ~ pooled_0_6, data = fieldexE),
  difference_in_means(NOV_2015_ontime ~ pooled_0_6, data = fieldexE),
  difference_in_means(MAR_2016_ontime ~ pooled_0_6, data = fieldexE),
  difference_in_means(JUL_2016_ontime ~ pooled_0_6, data = fieldexE),
  difference_in_means(compliance_1416 ~ pooled_0_6, data = fieldexE)
)
comp_control_plaE$outcome <- "Paid on Time"
comp_control_plaE$control <- "Placebo vs Pure Control"
comp_control_plaE$type <- "Eligibles"

# b) Intended compliance
intcomp_control_plaE <- rbind.data.frame(
  difference_in_means(JUL_2014_WEBACCESS ~ pooled_0_6, data = fieldexE),
  difference_in_means(NOV_2014_WEBACCESS ~ pooled_0_6, data = fieldexE),
  difference_in_means(MAR_2015_WEBACCESS ~ pooled_0_6, data = fieldexE),
  difference_in_means(JUL_2015_WEBACCESS ~ pooled_0_6, data = fieldexE),
  difference_in_means(NOV_2015_WEBACCESS ~ pooled_0_6, data = fieldexE),
  difference_in_means(MAR_2016_WEBACCESS ~ pooled_0_6, data = fieldexE),
  difference_in_means(JUL_2016_WEBACCESS ~ pooled_0_6, data = fieldexE),
  difference_in_means(intended_1416 ~ pooled_0_6, data = fieldexE)
)
intcomp_control_plaE$outcome <- "Web Access"
intcomp_control_plaE$control <- "Placebo vs Pure Control"
intcomp_control_plaE$type <- "Eligibles"

#### Ineligibles

#### TREATMENT VERSUS PURE CONTROL

# a) compliance
comp_controlN <- rbind.data.frame(
  difference_in_means(JUL_2014_ontime ~ pooled_124_6, data = fieldexN),
  difference_in_means(NOV_2014_ontime ~ pooled_124_6, data = fieldexN),
  difference_in_means(MAR_2015_ontime ~ pooled_124_6, data = fieldexN),
  difference_in_means(JUL_2015_ontime ~ pooled_124_6, data = fieldexN),
  difference_in_means(NOV_2015_ontime ~ pooled_124_6, data = fieldexN),
  difference_in_means(MAR_2016_ontime ~ pooled_124_6, data = fieldexN),
  difference_in_means(JUL_2016_ontime ~ pooled_124_6, data = fieldexN),
  difference_in_means(compliance_1416 ~ pooled_124_6, data = fieldexN)
)
comp_controlN$outcome <- "Paid on Time"
comp_controlN$control <- "Treatment vs Pure Control"
comp_controlN$type <- "Ineligibles"

# b) Intended compliance
intcomp_controlN <- rbind.data.frame(
  difference_in_means(JUL_2014_WEBACCESS ~ pooled_124_6, data = fieldexN),
  difference_in_means(NOV_2014_WEBACCESS ~ pooled_124_6, data = fieldexN),
  difference_in_means(MAR_2015_WEBACCESS ~ pooled_124_6, data = fieldexN),
  difference_in_means(JUL_2015_WEBACCESS ~ pooled_124_6, data = fieldexN),
  difference_in_means(NOV_2015_WEBACCESS ~ pooled_124_6, data = fieldexN),
  difference_in_means(MAR_2016_WEBACCESS ~ pooled_124_6, data = fieldexN),
  difference_in_means(JUL_2016_WEBACCESS ~ pooled_124_6, data = fieldexN),
```

```
  difference_in_means(intended_1416 ~ pooled_124_6, data = fieldexN)
)
intcomp_controlN$outcome <- "Web Access"
intcomp_controlN$control <- "Treatment vs Pure Control"
intcomp_controlN$type <- "Ineligibles"

#### TREATMENT VERSUS PLACEBO

# a) Compliance
comp_placeboN <- rbind.data.frame(
  difference_in_means(JUL_2014_ontime ~ pooled_124_0, data = fieldexN),
  difference_in_means(NOV_2014_ontime ~ pooled_124_0, data = fieldexN),
  difference_in_means(MAR_2015_ontime ~ pooled_124_0, data = fieldexN),
  difference_in_means(JUL_2015_ontime ~ pooled_124_0, data = fieldexN),
  difference_in_means(NOV_2015_ontime ~ pooled_124_0, data = fieldexN),
  difference_in_means(MAR_2016_ontime ~ pooled_124_0, data = fieldexN),
  difference_in_means(JUL_2016_ontime ~ pooled_124_0, data = fieldexN),
  difference_in_means(compliance_1416 ~ pooled_124_0, data = fieldexN)
)
comp_placeboN$outcome <- "Paid on Time"
comp_placeboN$control <- "Treatment vs Placebo"
comp_placeboN$type <- "Ineligibles"

# b) Intended compliance
intcomp_placeboN <- rbind.data.frame(
  difference_in_means(JUL_2014_WEBACCESS ~ pooled_124_0, data = fieldexN),
  difference_in_means(NOV_2014_WEBACCESS ~ pooled_124_0, data = fieldexN),
  difference_in_means(MAR_2015_WEBACCESS ~ pooled_124_0, data = fieldexN),
  difference_in_means(JUL_2015_WEBACCESS ~ pooled_124_0, data = fieldexN),
  difference_in_means(NOV_2015_WEBACCESS ~ pooled_124_0, data = fieldexN),
  difference_in_means(MAR_2016_WEBACCESS ~ pooled_124_0, data = fieldexN),
  difference_in_means(JUL_2016_WEBACCESS ~ pooled_124_0, data = fieldexN),
  difference_in_means(intended_1416 ~ pooled_124_0, data = fieldexN)
)
intcomp_placeboN$outcome <- "Web Access"
intcomp_placeboN$control <- "Treatment vs Placebo"
intcomp_placeboN$type <- "Ineligibles"

#### PURE CONTROL VERSUS PLACEBO

# a) Compliance
comp_control_plaN <- rbind.data.frame(
  difference_in_means(JUL_2014_ontime ~ pooled_0_6, data = fieldexN),
  difference_in_means(NOV_2014_ontime ~ pooled_0_6, data = fieldexN),
  difference_in_means(MAR_2015_ontime ~ pooled_0_6, data = fieldexN),
  difference_in_means(JUL_2015_ontime ~ pooled_0_6, data = fieldexN),
  difference_in_means(NOV_2015_ontime ~ pooled_0_6, data = fieldexN),
  difference_in_means(MAR_2016_ontime ~ pooled_0_6, data = fieldexN),
  difference_in_means(JUL_2016_ontime ~ pooled_0_6, data = fieldexN),
  difference_in_means(compliance_1416 ~ pooled_0_6, data = fieldexN)
)
comp_control_plaN$outcome <- "Paid on Time"
comp_control_plaN$control <- "Placebo vs Pure Control"
```

```
comp_control_plaN$type <- "Ineligibles"

# b) Intended compliance
intcomp_control_plaN <- rbind.data.frame(
  difference_in_means(JUL_2014_WEBACCESS ~ pooled_0_6, data = fieldexN),
  difference_in_means(NOV_2014_WEBACCESS ~ pooled_0_6, data = fieldexN),
  difference_in_means(MAR_2015_WEBACCESS ~ pooled_0_6, data = fieldexN),
  difference_in_means(JUL_2015_WEBACCESS ~ pooled_0_6, data = fieldexN),
  difference_in_means(NOV_2015_WEBACCESS ~ pooled_0_6, data = fieldexN),
  difference_in_means(MAR_2016_WEBACCESS ~ pooled_0_6, data = fieldexN),
  difference_in_means(JUL_2016_WEBACCESS ~ pooled_0_6, data = fieldexN),
  difference_in_means(intended_1416 ~ pooled_0_6, data = fieldexN)
)
intcomp_control_plaN$outcome <- "Web Access"
intcomp_control_plaN$control <- "Placebo vs Pure Control"
intcomp_control_plaN$type <- "Ineligibles"


## Combine and plot
plotdata <- rbind.data.frame(comp_placeboE, comp_controlE,
                             intcomp_placeboE, intcomp_controlE,
                             comp_control_plaE, intcomp_control_plaE,
                             comp_placeboN, comp_controlN,
                             intcomp_placeboN, intcomp_controlN,
                             comp_control_plaN, intcomp_control_plaN)
rm(comp_placeboE, comp_controlE,
   intcomp_placeboE, intcomp_controlE,
   comp_control_plaE, intcomp_control_plaE,
   comp_placeboN, comp_controlN,
   intcomp_placeboN, intcomp_controlN,
   comp_control_plaN, intcomp_control_plaN)

plotdata$time <- rep(1:8, 6)
#plotdata <- plotdata[plotdata$time!=8,]

plotdata$control <- as.factor(plotdata$control)

pd <- position_dodge(width = 0.6)

#Figure
ggplot(plotdata, aes(x=time, y=coefficients, group = control, shape = control)) +
  facet_grid(type ~ outcome) + #, scales="free"
  geom_point(size=4.5, position=pd) +
  geom_hline(aes(yintercept=0), size=.5, linetype="dashed") +
  geom_errorbar(aes(x=time,
                    ymin=conf.low,
                    ymax=conf.high),
                width=.15, size=1, position=pd) +
  xlab("Payments Since Receiving Flyer") + ylab("Average Causal Effect") +
  theme_minimal() +
  scale_colour_manual(values = c("black", "black")) +
  scale_x_continuous(breaks=1:8,
                     labels=c(as.character(1:7), "pooled")) +
```

```
  theme(plot.title = element_text(size = rel(1.2)),
        axis.text.x = element_text(size = rel(1)),
        axis.text.y = element_text(size = rel(1)),
        axis.title.y = element_text(size = rel(1)),
        axis.title.x = element_text(size = rel(1)),
        legend.text = element_text(size = rel(1)),
        strip.text.x = element_text(size = rel(1)),
        strip.text.y = element_text(size = rel(1)),
        legend.position = "bottom",
        legend.title=element_blank())
```




```
################################################
message("APPENDIX TABLE A6")
```

```
## APPENDIX TABLE A6
```

```
# Field Experiment: Effects of Information About the Tax
# Holiday on Compliance (weighted averages of block-specific
# effects for eligible and ineligible taxpayers)


blockedATE <- rbind.data.frame(
  # Treatment versus placebo: compliance
  difference_in_means(compliance_1416 ~ pooled_124_0, blocks = type,
                      se_type = "default", data = fieldex),
  # Treatment versus pure control: compliance
  difference_in_means(compliance_1416 ~ pooled_124_6, blocks = type,
```

```
                      se_type = "default", data = fieldex),
  # Placebo versus pure control: compliance
  difference_in_means(compliance_1416 ~ pooled_0_6, blocks = type,
                      se_type = "default", data = fieldex),
  # Treatment versus placebo: intended compliance
  difference_in_means(intended_1416 ~ pooled_124_0, blocks = type,
                      se_type = "default", data = fieldex),
  # Treatment versus pure control: intended compliance
  difference_in_means(intended_1416 ~ pooled_124_6, blocks = type,
                      se_type = "default", data = fieldex),
  # Placebo versus pure control: intended compliance
  difference_in_means(intended_1416 ~ pooled_0_6, blocks = type,
                      se_type = "default", data = fieldex)

)
blockedATE$outcome <- rep(c("Paid on Time", "Web Access"), each=3)
blockedATE$control <- rep(c("Treatment vs Placebo","Treatment vs Pure Control",
                            "Placebo vs Pure Control"), 2)

#Table
blockedATE[,c(1,2,4,6,11,18)]
```

```
##   coefficients  std.error  nobs             p.value      outcome
## 1  -0.00119254 0.00231141 10728 0.60591014917715824 Paid on Time
## 2   0.00044087 0.00164977 17750 0.78929236655162760 Paid on Time
## 3   0.00130733 0.00197545 14184 0.50811563755797096 Paid on Time
## 4  -0.00415790 0.00096927 10799 0.00001804460658575   Web Access
## 5   0.00158279 0.00069274 17842 0.02233409051736731   Web Access
## 6   0.00657777 0.00091511 14267 0.00000000000069038   Web Access
##                      control
## 1      Treatment vs Placebo
## 2 Treatment vs Pure Control
## 3   Placebo vs Pure Control
## 4      Treatment vs Placebo
## 5 Treatment vs Pure Control
## 6   Placebo vs Pure Control
```

```
################################################
message("APPENDIX TABLE A7")
```

```
## APPENDIX TABLE A7
```

```
# Survey Experiment: Pooled Lottery vs. Discretionary
# Benefit Conditions

survexp.results <- rbind.data.frame(
  difference_in_means(S1p4 ~ treat_discretion, data = survey_data),
  difference_in_means(S1p1 ~ treat_discretion, data = survey_data),
  difference_in_means(S1p3 ~ treat_discretion, data = survey_data),
  difference_in_means(S1p2 ~ treat_discretion, data = survey_data),
  difference_in_means(S1p5 ~ treat_discretion, data = survey_data)
)
```

```
survexp.results$outcome <- c("Rewards Go To The Same People As Always",
                             "Rewards Are A Waste Of Money",
                             "Worth It To Be Up To Date",
                             "Municipal Government Does A Good Job",
                             "Municipal Taxes Are Just")

survexp.results
```

```
##   coefficients std.error      df nobs statistic      p.value  conf.low
## 1     1.052123   0.21867  691.02 1542   4.81154 0.0000018401  0.622792
## 2    -0.021917   0.17699  979.39 2234  -0.12383 0.9014745169 -0.369246
## 3    -0.477498   0.15120  902.40 2266  -3.15814 0.0016405988 -0.774234
## 4    -0.168227   0.14121 1037.66 2313  -1.19133 0.2337954560 -0.445314
## 5     0.035448   0.03519 1021.98 2291   1.00733 0.3140130560 -0.033605
##   conf.high alpha             term                                outcome
## 1   1.48145  0.05 treat_discretion Rewards Go To The Same People As Always
## 2   0.32541  0.05 treat_discretion            Rewards Are A Waste Of Money
## 3  -0.18076  0.05 treat_discretion               Worth It To Be Up To Date
## 4   0.10886  0.05 treat_discretion    Municipal Government Does A Good Job
## 5   0.10450  0.05 treat_discretion                  Municipal Taxes Are Just
##   condition2 condition1      vcov   design
## 1          1          0 0.0478151 Standard
## 2          1          0 0.0313264 Standard
## 3          1          0 0.0228602 Standard
## 4          1          0 0.0199400 Standard
## 5          1          0 0.0012383 Standard
```

```
#Threshold for FDR correction
# get and order the nominal p-values
ordered.ps <- survexp.results$p.value[order(survexp.results$p.value,decreasing=F)]
ordered.ps
```

```
## [1] 0.0000018401 0.0016405988 0.2337954560 0.3140130560 0.9014745169
```

```
comp <- (1:length(ordered.ps)/length(ordered.ps))*(.05)

FDR <- cbind(ordered.ps,comp,ordered.ps<=comp)
FDR
```

```
##        ordered.ps comp
## [1,] 0.0000018401 0.01 1
## [2,] 0.0016405988 0.02 1
## [3,] 0.2337954560 0.03 0
## [4,] 0.3140130560 0.04 0
## [5,] 0.9014745169 0.05 0
```

```
fdr <- max(FDR[,1][FDR[,3]==1])
fdr
```

```
## [1] 0.0016406
```

```r
#Threshold for Bonferroni correction
bonf <- 0.05/length(ordered.ps)
bonf
```

```
## [1] 0.01
```

```r
survexp.results$Bonf_reject <- NA
survexp.results$FDR_reject <- NA

survexp.results$Bonf_reject[survexp.results$p.value<=bonf] <- "yes"
survexp.results$Bonf_reject[survexp.results$p.value>bonf]  <- "no"

survexp.results$FDR_reject[survexp.results$p.value<=fdr] <- "yes"
survexp.results$FDR_reject[survexp.results$p.value>fdr]  <- "no"

#Table
survexp.results[c(1,3,5,4,2),c(11,1,2,4,6,16,17)]
```

```
##                                 outcome coefficients std.error nobs
## 1 Rewards Go To The Same People As Always     1.052123   0.21867 1542
## 3               Worth It To Be Up To Date    -0.477498   0.15120 2266
## 5                Municipal Taxes Are Just     0.035448   0.03519 2291
## 4    Municipal Government Does A Good Job    -0.168227   0.14121 2313
## 2            Rewards Are A Waste Of Money    -0.021917   0.17699 2234
##        p.value Bonf_reject FDR_reject
## 1 0.0000018401         yes        yes
## 3 0.0016405988         yes        yes
## 5 0.3140130560          no         no
## 4 0.2337954560          no         no
## 2 0.9014745169          no         no
```