Amended Pre-Analysis Plan. Administrative Data.

Positive vs. Negative Incentives for Compliance: Evaluating a Randomized Tax Holiday

T. Dunning, F. Monestier, R. Piñeiro, F. Rosenblatt, and G. Tuñón

October 19, 2014

Contents

Description of the document and setup of the analysis

This document amends our registered pre-analysis plan and adds the code to be used for the analysis of the administrative data. It will be followed up by an additional document amending and providing mock analysis of the survey data.

This document was produced once we got most of the administrative data, but before conducting any analysis. The results presented here were produced using mock data. We generate our mock data by reshuffling the treatment labels in our datasets without replacement (see below). For a few variables for which we still do not have the data we generate mock variables.

Generating mock data for analysis

Natural Experiment

```
load("~/Dropbox/Uruguay state capacity/Analysis/data/final_data/naturalex_data.Rda")

# keeping the good taxpayers
naturalex_gtp <- naturalex # dataset has only good taxpayers

treat_by_tax <- ddply(naturalex_gtp, c("tax", "won_lottery"),
                  summarise,
                  N = length(unique(account)))
treat_by_tax
```

Main data

```
##                         tax won_lottery    N
## 1 Contribucion Inmobiliaria           0 1291
## 2 Contribucion Inmobiliaria           1 1275
## 3        Patente de Rodados           0  344
## 4        Patente de Rodados           1  366
```

```
## 5               Saneamiento           0  386
## 6               Saneamiento           1  439
## 7    Tributos Domiciliarios           0  994
## 8    Tributos Domiciliarios           1  963
```

```r
# Creating vectors for mock treatment and control respecting the type of tax
set.seed(2067)
vector_CI <- sample(
  c(rep(0,treat_by_tax[1,3]),
    rep(1,treat_by_tax[2,3])), (treat_by_tax[1,3]+treat_by_tax[2,3]), replace=F)
vector_PR <- sample(
  c(rep(0,treat_by_tax[3,3]),
    rep(1,treat_by_tax[4,3])), (treat_by_tax[3,3]+treat_by_tax[4,3]), replace=F)
vector_TS <- sample(
  c(rep(0,treat_by_tax[5,3]),
    rep(1,treat_by_tax[6,3])), (treat_by_tax[5,3]+treat_by_tax[6,3]), replace=F)
vector_TD <- sample(
  c(rep(0,treat_by_tax[7,3]),
    rep(1,treat_by_tax[8,3])), (treat_by_tax[7,3]+treat_by_tax[8,3]), replace=F)

# Binding the treatment vectors with the account number for CI
CI <- cbind(unique(naturalex_gtp$account[naturalex_gtp$tax=="Contribucion Inmobiliaria"]),
            vector_CI)
PR <- cbind(unique(naturalex_gtp$account[naturalex_gtp$tax=="Patente de Rodados"]),
            vector_PR)
TD <- cbind(unique(naturalex_gtp$account[naturalex_gtp$tax=="Tributos Domiciliarios"]),
            vector_TD)
TS <- cbind(unique(naturalex_gtp$account[naturalex_gtp$tax=="Saneamiento"]),
            vector_TS)

mock_assignment <- as.data.frame(rbind(CI, PR, TD, TS))
names(mock_assignment) <- c("account", "won_lottery")

# eliminating real treatment variable from the dataset
check <- sum(naturalex_gtp$won_lottery)
naturalex_gtp <- naturalex_gtp[,-which(names(naturalex_gtp)=="won_lottery")]

# Merging to add mock assignment
naturalex_gtp <- merge(naturalex_gtp, mock_assignment, by="account")
stopifnot(check==sum(naturalex_gtp$won_lottery))

# Take Patente de Rodados out
# We exclude this tax fromt the analysis bacause we need to further work with the
# municipality to understand how they implemented the tax holiday in order to
# set up the analysis of the natural experiment correctly. Once we are able to do
# this, we will reincorporate it to the analysis (and the adjustments for multiple
# comparisons)
naturalex_gtp <- naturalex_gtp[naturalex_gtp$tax!="Patente de Rodados",]

# We use a panel dastaset to analyse the natural experiment as it is allows to build
# the plots and recenter all holidays such that they overlap.
table(naturalex_gtp$tax)
```

```
##
```

```
## Contribucion Inmobiliaria        Patente de Rodados
##                     112904                         0
##               Saneamiento    Tributos Domiciliarios
##                     66000                    172216
```

```
levels(naturalex_gtp$tax) <- c("Property", "Vehicle", "Sewage", "Head")

# Take out years within tax holiday
naturalex_gtp <- naturalex_gtp[is.na(naturalex_gtp$plot_time)==F,]

# cleaning environment
rm(list=(ls()[ls()!="naturalex_gtp"]))
```

Debt data For confidentiality reasons, the municipality could not give us debt data of a given account number. We could, however, list the account numbers in each tax type, taxpayer type and treatment (won_lottery) combination and obtain a list of the debt amounts corresponding to such category (here not linked to an account number).

For pragmatic reasons, we add this data to our main dataset following tax type[1], but the link of the debt data to a given account is ficticious.

```
load("~/Dropbox/Uruguay state capacity/Analysis/data/final_data/naturalex_debt_gtp.Rda")

naturalex_debt_gtp$won_lottery <- sample(naturalex_debt_gtp$won_lottery,
                                         length(naturalex_debt_gtp$won_lottery),
                                         replace=F)
```

Field Experiment

```
load("~/Dropbox/Uruguay state capacity/Analysis/data/final_data/fieldex_data.Rda")

# Generating mock treatment assignment
fieldex$treatment[fieldex$type=="eligible"] <-
  sample(fieldex$treatment[fieldex$type=="eligible"],
         length(fieldex$treatment[fieldex$type=="eligible"]), replace=F)

fieldex$treatment[fieldex$type=="noneligible"] <-
  sample(fieldex$treatment[fieldex$type=="noneligible"],
         length(fieldex$treatment[fieldex$type=="noneligible"]), replace=F)

### Adding compliance indicator
fieldex$compliance_july <- ifelse((fieldex$july_nrbills_owed==0 & fieldex$july_ontime==1),1,0)
```

Our field experiment data includes cases that could not be treated by our July experiment, either because they had paid their bill in advance or because they are retired (retired individuals have differente due dates for their bills).

[1] Note that (1) all taxpayers in the main dataset are goodtaxpayers and (2) for the real analysis we will follow tax-treatment combinations.

```r
# Retired individuals in sample
table(fieldex$retired, fieldex$type)
```

```
##
##     eligible noneligible
##   0    14300       14123
##   1        0         177
```

```r
table(fieldex$retired, fieldex$treatment) # Should be balanced by treatment
```

```
##
##        0    1    2    3    4    5
##   0 5666 2883 2881 5670 5668 5655
##   1   34   17   19   30   32   45
```

```r
# Paid in advance in sample
table(fieldex$paid_in_advance, fieldex$type)
```

```
##
##     eligible noneligible
##   0     9461       12483
##   1     4839        1817
```

```r
table(fieldex$paid_in_advance, fieldex$treatment) # Should be balanced by treatment
```

```
##
##        0    1    2    3    4    5
##   0 4395 2255 2235 4353 4360 4346
##   1 1305  645  665 1347 1340 1354
```

We exclude both this types of cases from our analysis.

```r
fieldex <- fieldex[fieldex$retired==0 & fieldex$paid_in_advance==0,]
# We might use them for placebo tests later.
```

This is the new distribution of eligible and ineligible individuals across treatment conditions.

```
##
##                    0    1    2    3    4    5
##   eligible    1901  968  965 1871 1881 1875
##   noneligible 2460 1271 1251 2452 2447 2426
```

Mock data for the variables we do not have yet

```r
# Field experiment
fieldex$current_debt_DiD  <- rnorm(nrow(fieldex),0,100000)
```

Impact of the tax holiday lottery

Bills and tax holidays

CI: Property tax PR: Vehicle tax TS : Sewage TD: Head

New information about how the lottery and the tax holiday work in practice suggest we should review some of the details of our analysis plan. This section summarizes how the tax holiday system works and how it speaks to our research design.

Note that the first part of the process is common to all taxes: the lottery takes place and the IT department filters accounts[2] for which the last four digits match the lottery value. Because the municipality has no way of distinguishing "personas fisicas" from "personas juridicas", all these accounts are notified that they have won and are encouraged to go to the municipality with proof that they are "personas fisicas" to finish the process that effectively concedes the tax holiday.

All winners (regardless of the lottery), go to the same office ("gestion de contribuyentes") to prove they are "personas juridicas". Note that (a) some winners might never do this, in which case they do not get the tax holiday[3] and (b) winners might not do this as soon as they get the note telling them they have won. Although by law they are supposed to contact the municipality in the 90 day window after they receive the note in order to be eligible for they prize, municipal employees have told us they "would never deny anyone the prize", regardless of when it is they go to "cash it". This point is important, as it suggests a first possible source of delay between the time an account wins the lottery and when the tax holiday is effective.

The process then continues in the office in charge of the tax, which has the final power to decide what bills the tax holiday will effectively apply to. One office ("Impuestos inmobiliarios") controls both the CI and TD. PR and TS are under the jurisdiction of two other offices ("Ingresos Vehiculares" and "Administración de Saneamiento") respectively. Despite by law the tax holiday should already be in place for the bill that immediately follows the lottery, each office tends to adjust the policy to their own admistrative/bureaucratic procedure which, together with difference in the number of annual bills and due dates, introduces some between tax variation.

CI consists of 3 bills a year, in March, July and November one every four months. Since lotteries take place every two months, there are two lotteries per CI bill. The office in charge of this tax implemented the policy following closely its design, and thus most tax holidays start for the bill immediatly after the lottery (see table below). Because of the combined timing of bills and lotteries, lotteries always take place between one bill and the next. If $t = 0$ describes the first payment under the tax holiday and $t = 1$ the first payment before the holiday, for CI the lottery always takes place between these two periods. Thus, a good taxpayer by the time of the lottery has complied with the tax for $t = (-3, -2, -1)$. One possible concern with CI is that, because it is an anual tax of a fixed amount, taxpayers can choose to pay the annual rate, paying by march the bills for all the year. However, one can only pay in advance the bills for July and August, but only payments for the current year can be payed in this way. This can also add some delay for the tax holiday to take place, since winners that win after paying the year in advance will not "cash" their prize until the March payment of the following year.

TD consists of 6 bills a year, in February, April, June, August, October and December. Notice that these are also the months when the lotteries take place. These tax holidays were implemented by the same office processing the CI, and its implementation also follows the policy guideline closely: tax holidays start with the first bill that follows the lottery. Unlike CI, however, here the lottery does not take place between payments, they overlap (see table below). Following the notation for CI, here the lottery takes place at $t = -1$. This introduces one specificity to the analysis of the TD holiday lottery. Because they are concurrent, the IT office does not know yet if the bill for that month is correctly paid. Thus, a good taxpayer for TS has, at $t = -1$–the time of the lottery–complied with the tax for $t = (-7, -6, -5, -4, -3, -2)$. By the time the holiday kicks in, it is possible that a winner is no longer a good taxpayer–if she hasn't paid the bill corresponding to $t = -1$. A

[2] We confirmed with the municipality account numbers are unique (ie. they are not repeated across lotteries).

[3] The municipality estimates that around 25% of the people that are registered with real names (and thus likely to be personas fisicas) never engage in the process to get the prize.

natural question here would be if the taxpayer is informed that she has won *before* the due date of the $t = -1$ bill–which overlaps with the lottery. Whereas this is not impossible, it is highly unlikely as the lotteries take place during the last week of the month. TD cannot be paid in advance.

For **TS** there is also 6 bills a year, in January, March, May, July, September and November. The implementation of the tax holiday is close to the policy design, although in practice, the tax holiday does not start with the bill that immediatly follows the lottery but the next. One can think of this case as if the lottery took place between $t = -2$ and $t = -1$. Despite the fact that the lottery takes place *before* the $t = -1$ bill, there is not enough time for the tax holiday to kick in. If someone wins the December lottery, for example, they are judged as good or bad taxpayers looking a the relevant previous six payments, starting by the December bill of the previous year. But by the end of Decembre (when the lottery happes) the municipality is already processing the January bills. Thus, the tax holiday does not kick in until the following bill, in this case March. As for TD, by the time the holiday kicks in, it is possible that a winner is no longer a good taxpayer–if she hasn't paid the bill corresponding to $t = -1$. It is also the case that winner eligibility is judged by the $t = (-7, -6, -5, -4, -3, -2)$ payments. Yet unlike TD, for TS it *is* likely that taxpayers know they have won by the time of their $t = -1$ payment, even if they do not get the tax holiday until $t = 0$. TS cannot be paid in advance.

Finally, **PR** works somewhat different from the other taxes. There are 3 payments—January, May and September—which taxpayers can pay in advance, just as for CI. Despite our close engagement with the municipality, we have not yet been able to undeerstand the concrete process by which the lottery was implemented. For thay reason we have exclude this tax from our analysis, but we will add it as soon as I understan primerotod lo drun auto

	FEB	APR	JUN	AUG	OCT	DEC
CI	MAR	JUL	JUL	NOV	NOV	MAR
PR	JAN	JAN	JAN	JAN	JAN	JAN
TD	APR	JUN	AUG	OCT	DEC	FEB
TS	MAY	JUL	SEP	NOV	JAN	MAR

unique(naturalex_gtp$bill_date[naturalex_gtp$tax==$\text{“Patente de Rodados”}$])

Functions

```
t.test.se <- function(y,x){ # t test with SEs
  diff <- mean(y[x==1],na.rm=T) - mean(y[x==0],na.rm=T)
  sd1a.N <- var(y[x==1],na.rm=T)/length(na.omit(y[x==1]))
  sd1b.N <- var(y[x==0],na.rm=T)/length(na.omit(y[x==0]))
  se.diff <- sqrt(sd1a.N + sd1b.N)
  t <- diff/se.diff
  return(c(diff,se.diff))
}

# Function to test the difference of the differences
comp.eff <- function(dat1_y, dat1_x, dat2_y, dat2_x){

  tt1 <- t.test.se(dat1_y, dat1_x)
  tt2 <- t.test.se(dat2_y, dat2_x)

  dms <- as.data.frame(rbind(tt1, tt2))
  rownames(dms) <- c("DM 1", "DM 2")
  colnames(dms) <- c("DM", "SE")
```

```
  print(dms)

  diff <- tt1[1]-tt2[1]
  se.diff <- sqrt((tt1[2]^2)+(tt2[2]^2))

  res <- c(diff,se.diff)
  names(res) <- c("Diff in effects", "SE")

  return(res)
}
```

Plan for Data Analysis

Hypothesis 1A.

Winning the tax holiday lottery leads to an *increase* in future tax compliance.

Hypothesis 1B.

Winning the lottery leads to a *decrease* in future tax compliance.

Hypothesis 1C.

Winning the lottery leads to *no change* in future tax compliance.

Graphical Analysis We construct a plot in which the horizontal axis is time, measured in tax holiday lotteries (which occur every two or four months, depending on the tax). The vertical axis is one of our three measures of tax payment (compliance, missed payment, or number of payments owed). Note the x axis is not continuous. The left side shows the payment history until an account won or would have won the lottery, centered at 0. The right side of the axis show payments after the tax holiday, starting at 1. The time between 0 and 1 might vary for different taxpayers.[4] Parting the x axis allows for better comparison.

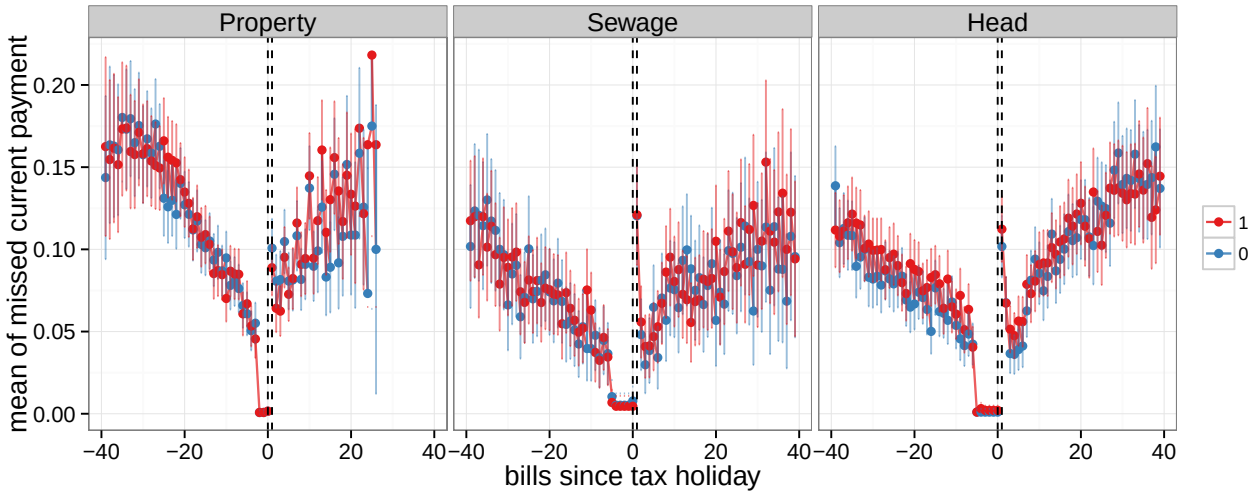
```
# summarizing data to plot
plot_data <- ddply(naturalex_gtp, c("tax", "plot_time", "won_lottery"),
            summarise,
            N  = length(unique(account)),
            nr_missed_mean = mean(nr_missed_payments, na.rm=T),
            se_nr_missed_mean = sd(nr_missed_payments, na.rm=T)/sqrt(N),
            nr_missed_upper= nr_missed_mean + qnorm(.975)*(se_nr_missed_mean),
            nr_missed_lower= nr_missed_mean - qnorm(.975)*(se_nr_missed_mean),
            missed_payment_mean = mean(missed_payment, na.rm=T),
            se_missed_payment_mean = sd(missed_payment, na.rm=T)/sqrt(N),
            missed_payment_upper= missed_payment_mean + qnorm(.975)*(se_missed_payment_mean),
            missed_payment_lower= missed_payment_mean - qnorm(.975)*(se_missed_payment_mean),
            compliance_mean = mean(compliance, na.rm=T),
            se_compliance_mean = sd(compliance, na.rm=T)/sqrt(N),
            compliance_upper= compliance_mean + qnorm(.975)*se_compliance_mean,
```

[4] The final plots will use bill date so that the x axis can be interpreted as time since winning (an the taxes are rescaled such that there are 6 payments of TD and TS for every 3 of CI and PR).

```
            compliance_lower= compliance_mean - qnorm(.975)*se_compliance_mean
            )



plot_data <- na.omit(plot_data[plot_data$plot_time>-40 &
                         plot_data$plot_time<40,])
```

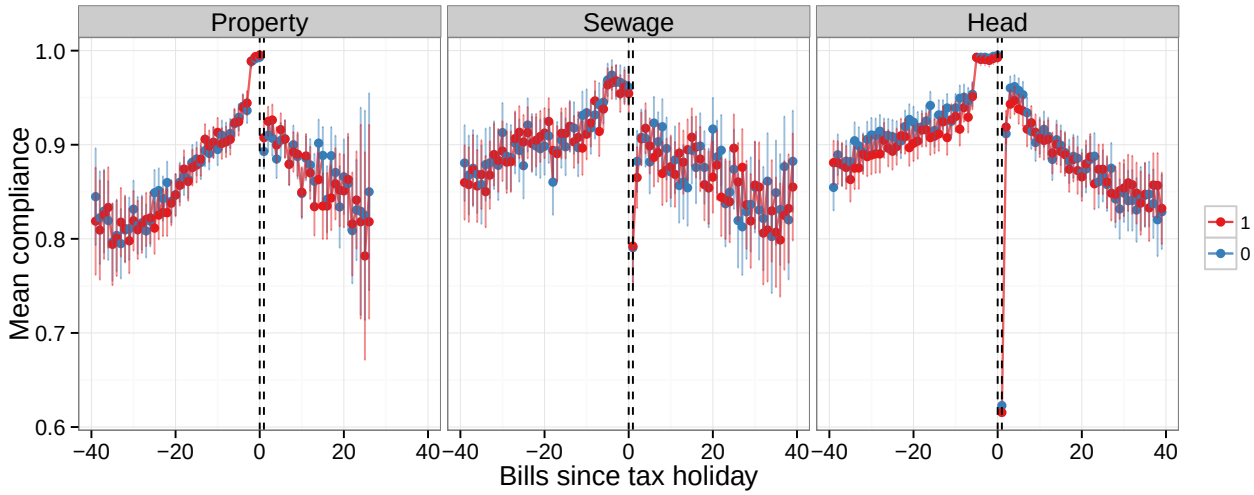
```
### Plots by tax
plot_data$won_lottery <- as.factor(plot_data$won_lottery)
plot_data$won_lottery <- relevel(plot_data$won_lottery, ref="1")
plot_data$after <- as.factor(plot_data$plot_time>0)

# missed payment
p <- ggplot(plot_data, aes(x=plot_time,y=missed_payment_mean,
                           group=c(after), color=won_lottery))
p + geom_line(size=.6, alpha=.7) + facet_wrap(~tax) +
  geom_point(size=2.4)  +
  ylim(min(plot_data$missed_payment_mean), max(plot_data$missed_payment_mean)) +
    xlab("bills since tax holiday") +
  ylab("mean of missed current payment") +
  geom_errorbar(aes(ymin=missed_payment_lower,
                    ymax=missed_payment_upper),
              # colour="blue",
               width=.3, alpha=.5) +
  scale_color_brewer(palette="Set1") +
  #scale_colour_manual(
#  values = c("0" = "navyblue","1" = "red")) +
  #ggtitle("Mean Debt (Property Tax, 2000-2014)") +
  theme_bw() +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text = element_text(size = rel(1)),
        axis.title.y = element_text(size = rel(1.25)),
        axis.title.x = element_text(size = rel(1.25)),
        strip.text.x = element_text(size = rel(1.5)),
        legend.title=element_blank()) +
  geom_vline(aes(xintercept = 1), linetype="dashed") +
  geom_vline(aes(xintercept = 0), linetype="dashed")
```

```
## Warning: Removed 28 rows containing missing values (geom_path).
## Warning: Removed 24 rows containing missing values (geom_path).
## Warning: Removed 48 rows containing missing values (geom_path).
## Warning: Removed 48 rows containing missing values (geom_path).
```



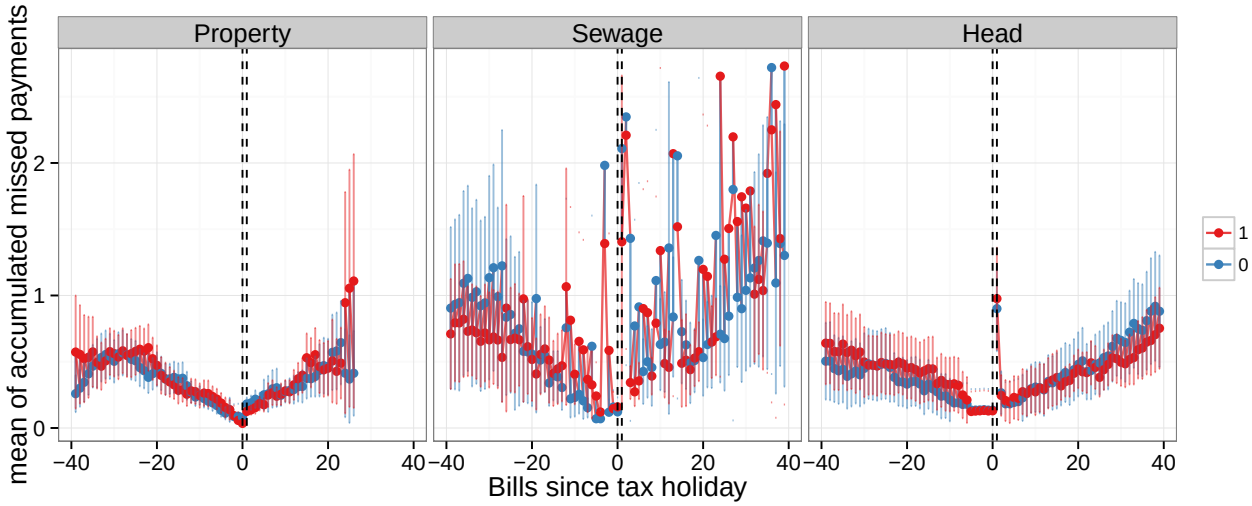

```
# compliance
p <- ggplot(plot_data, aes(x=plot_time,y=compliance_mean,
                           group=c(after), color=won_lottery))
p + geom_line(size=.6, alpha=.7) + facet_wrap(~tax) +
  geom_point(size=2.4)  +
  ylim(min(plot_data$compliance_mean), max(plot_data$compliance_mean)) +
    xlab("Bills since tax holiday") +
  ylab("Mean compliance") +
  geom_errorbar(aes(ymin=compliance_lower,
                    ymax=compliance_upper),
              # colour="blue",
               width=.3, alpha=.5) +
  scale_color_brewer(palette="Set1") +
  #scale_colour_manual(
#  values = c("0" = "navyblue","1" = "red")) +
  #ggtitle("Mean Debt (Property Tax, 2000-2014)") +
  theme_bw() +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text = element_text(size = rel(1)),
        axis.title.y = element_text(size = rel(1.25)),
        axis.title.x = element_text(size = rel(1.25)),
        strip.text.x = element_text(size = rel(1.5)),
        legend.title=element_blank()) +
  geom_vline(aes(xintercept = 1), linetype="dashed") +
  geom_vline(aes(xintercept = 0), linetype="dashed")
```

```
## Warning: Removed 16 rows containing missing values (geom_path).
## Warning: Removed 48 rows containing missing values (geom_path).
## Warning: Removed 8 rows containing missing values (geom_path).
```




```
# nr missed payments
p <- ggplot(plot_data, aes(x=plot_time,y=nr_missed_mean,
                           group=c(after), color=won_lottery))
p + geom_line(size=.6, alpha=.7) + facet_wrap(~tax) +
  geom_point(size=2.4)  +
  ylim(min(plot_data$nr_missed_mean), max(plot_data$nr_missed_mean)) +
    xlab("Bills since tax holiday") +
  ylab("mean of accumulated missed payments") +
  geom_errorbar(aes(ymin=nr_missed_lower,
                    ymax=nr_missed_upper),
              # colour="blue",
               width=.3, alpha=.5) +
  scale_color_brewer(palette="Set1") +
  #scale_colour_manual(
#  values = c("0" = "navyblue","1" = "red")) +
  #ggtitle("Mean Debt (Property Tax, 2000-2014)") +
  theme_bw() +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text = element_text(size = rel(1)),
        axis.title.y = element_text(size = rel(1.25)),
        axis.title.x = element_text(size = rel(1.25)),
        strip.text.x = element_text(size = rel(1.5)),
        legend.title=element_blank()) +
  geom_vline(aes(xintercept = 1), linetype="dashed") +
  geom_vline(aes(xintercept = 0), linetype="dashed")
```

```
## Warning: Removed 24 rows containing missing values (geom_path).
## Warning: Removed 64 rows containing missing values (geom_path).
## Warning: Removed 92 rows containing missing values (geom_path).
## Warning: Removed 64 rows containing missing values (geom_path).
## Warning: Removed 16 rows containing missing values (geom_path).
```

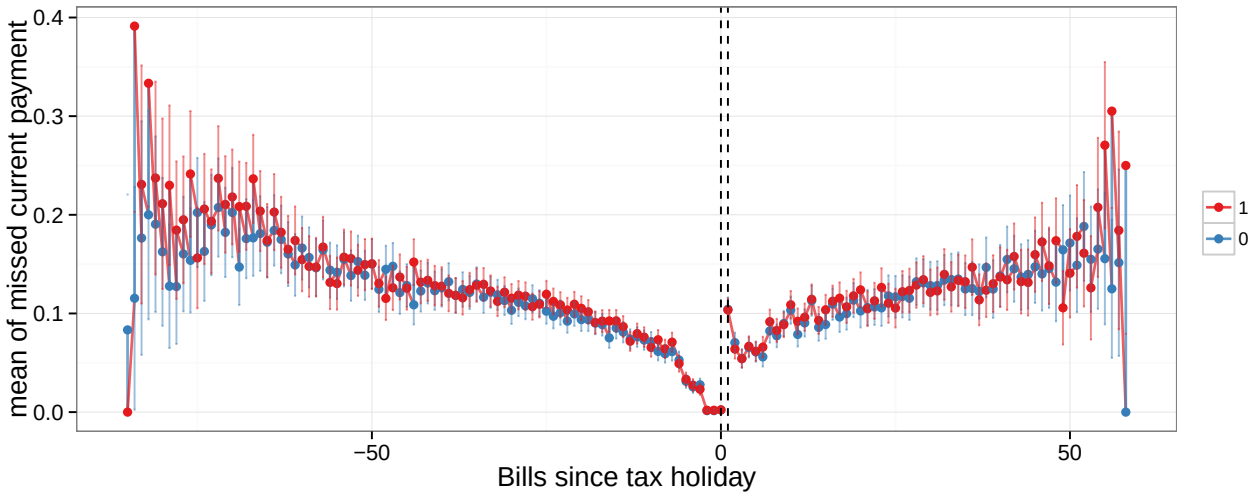



```
#### All taxes in same plot
plot_data2 <- ddply(naturalex_gtp, c("plot_time", "won_lottery"),
                    summarise,
             N  = length(unique(account)),
             nr_missed_mean = mean(nr_missed_payments, na.rm=T),
             se_nr_missed_mean = sd(nr_missed_payments, na.rm=T)/sqrt(N),
             nr_missed_upper= nr_missed_mean + qnorm(.975)*(se_nr_missed_mean),
             nr_missed_lower= nr_missed_mean - qnorm(.975)*(se_nr_missed_mean),
             missed_payment_mean = mean(missed_payment, na.rm=T),
             se_missed_payment_mean = sd(missed_payment, na.rm=T)/sqrt(N),
             missed_payment_upper= missed_payment_mean + qnorm(.975)*(se_missed_payment_mean),
             missed_payment_lower= missed_payment_mean - qnorm(.975)*(se_missed_payment_mean),
             compliance_mean = mean(compliance, na.rm=T),
             se_compliance_mean = sd(compliance, na.rm=T)/sqrt(N),
             compliance_upper= compliance_mean + qnorm(.975)*se_compliance_mean,
             compliance_lower= compliance_mean - qnorm(.975)*se_compliance_mean
             )
```

```
plot_data2$won_lottery <- as.factor(plot_data2$won_lottery)
plot_data2$won_lottery <- relevel(plot_data2$won_lottery, ref="1")
plot_data2$after <- as.factor(plot_data2$plot_time>0)


# missed payment
p <- ggplot(plot_data2, aes(x=plot_time,y=missed_payment_mean,
                            group=c(after), color=won_lottery))
p + geom_line(size=.7, alpha=.7) +
  geom_point(size=2.4)  +
  ylim(min(plot_data2$missed_payment_mean), max(plot_data2$missed_payment_mean)) +
    xlab("Bills since tax holiday") +
  ylab("mean of missed current payment") +
  geom_errorbar(aes(ymin=missed_payment_lower,
                    ymax=missed_payment_upper),
                # colour="blue",
                 width=.3, alpha=.5) +
  scale_color_brewer(palette="Set1") +
  #scale_colour_manual(
```

```
#  values = c("0" = "navyblue","1" = "red")) +
  #ggtitle("Mean Debt (Property Tax, 2000-2014)") +
  theme_bw() +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text = element_text(size = rel(1)),
        axis.title.y = element_text(size = rel(1.25)),
        axis.title.x = element_text(size = rel(1.25)),
        strip.text.x = element_text(size = rel(1.5)),
        legend.title=element_blank()) +
  geom_vline(aes(xintercept = 1), linetype="dashed") +
  geom_vline(aes(xintercept = 0), linetype="dashed")
```

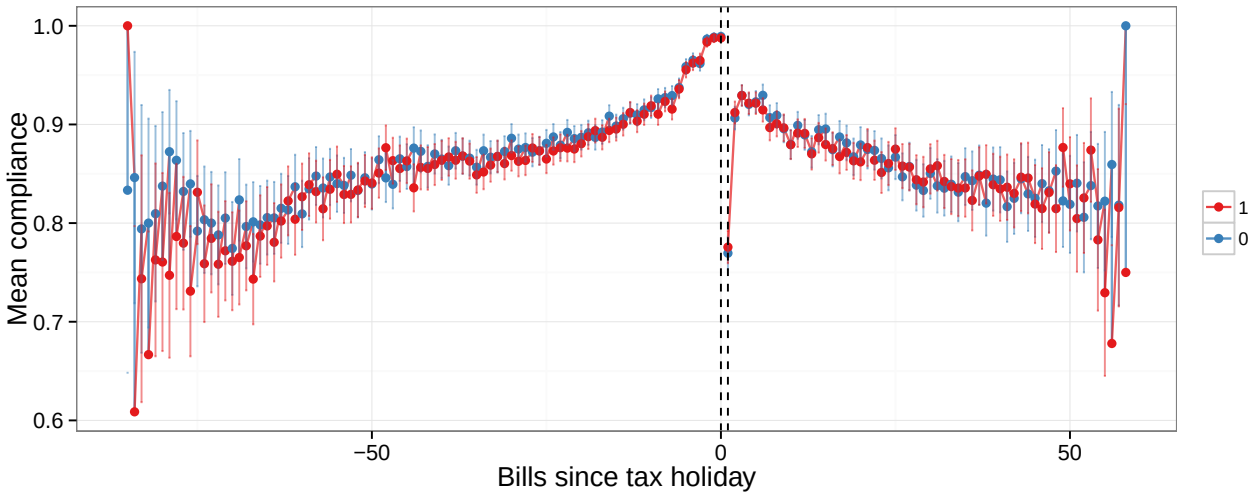
```
## Warning: Removed 12 rows containing missing values (geom_path).
## Warning: Removed 8 rows containing missing values (geom_path).
```




```
# compliance
p <- ggplot(plot_data2, aes(x=plot_time,y=compliance_mean,
                            group=c(after), color=won_lottery))
p + geom_line(size=.6, alpha=.7) +
  geom_point(size=2.4)  +
  ylim(min(plot_data2$compliance_mean), max(plot_data2$compliance_mean)) +
    xlab("Bills since tax holiday") +
  ylab("Mean compliance") +
  geom_errorbar(aes(ymin=compliance_lower,
                    ymax=compliance_upper),
                # colour="blue",
                 width=.3, alpha=.5) +
  scale_color_brewer(palette="Set1") +
  #scale_colour_manual(
#  values = c("0" = "navyblue","1" = "red")) +
  #ggtitle("Mean Debt (Property Tax, 2000-2014)") +
  theme_bw() +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text = element_text(size = rel(1)),
        axis.title.y = element_text(size = rel(1.25)),
        axis.title.x = element_text(size = rel(1.25)),
```

```
        strip.text.x = element_text(size = rel(1.5)),
        legend.title=element_blank()) +
  geom_vline(aes(xintercept = 1), linetype="dashed") +
  geom_vline(aes(xintercept = 0), linetype="dashed")
```

```
## Warning: Removed 12 rows containing missing values (geom_path).
## Warning: Removed 8 rows containing missing values (geom_path).
```




```
# nr missed payments
p <- ggplot(plot_data2, aes(x=plot_time,y=nr_missed_mean,
                            group=c(after), color=won_lottery))
p + geom_line(size=.7, alpha=1) +
  geom_point(size=2.4)  +
  ylim(min(plot_data2$nr_missed_mean), max(plot_data2$nr_missed_mean)) +
    xlab("Bills since tax holiday") +
  ylab("mean of accumulated missed payments") +
  geom_errorbar(aes(ymin=nr_missed_lower,
                    ymax=nr_missed_upper),
                # colour="blue",
                 width=.3, alpha=.5) +
  scale_color_brewer(palette="Set1") +
  #scale_colour_manual(
#  values = c("0" = "navyblue","1" = "red")) +
  #ggtitle("Mean Debt (Property Tax, 2000-2014)") +
  theme_bw() +
  theme(plot.title = element_text(size = rel(1.75)),
        axis.text = element_text(size = rel(1)),
        axis.title.y = element_text(size = rel(1.25)),
        axis.title.x = element_text(size = rel(1.25)),
        strip.text.x = element_text(size = rel(1.5)),
        legend.title=element_blank()) +
  geom_vline(aes(xintercept = 1), linetype="dashed") +
  geom_vline(aes(xintercept = 0), linetype="dashed")
```

```
## Warning: Removed 64 rows containing missing values (geom_path).
## Warning: Removed 16 rows containing missing values (geom_path).
```

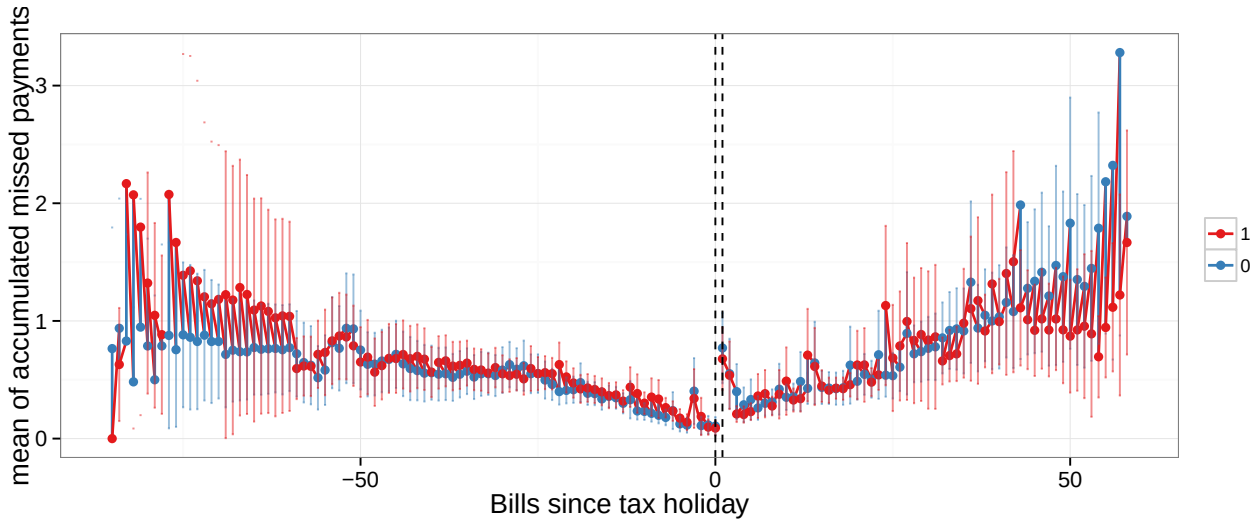


TABLE 1. NATURAL EXPERIMENT. Effects of the tax holiday (difference in differences analysis). Effects of the tax holiday. Comparing winners to non-winners, difference in difference analysis (comparison A=mean of the year before winning vs. mean of the year after the tax holiday; comparison B= mean of three years before winning vs. mean of three years after tax holiday). Tests using compliance as an outcome are conditional on finding effects for either missed payments or number of payments owed for the relevant period. This is because compliance is a stricter test, and if we find effects for neither missed payments of number of payments owed, there will be no effects by construction for compliance.

```
# 1 year diff in diff setup

dd_data <- ddply(naturalex_gtp, c("account","tax", "won_lottery"),
                 summarise,
               compliance_mean_6_DiD_1 = mean(compliance[plot_time>0 & plot_time<=6],
                                              na.rm=T)-
                 mean(compliance[plot_time<0 & plot_time>=(-6)],
                                              na.rm=T),
               compliance_mean_3_DiD_1 = mean(compliance[plot_time>0 & plot_time<=3],
                                              na.rm=T)-
                 mean(compliance[plot_time<0 & plot_time>=(-3)],
                                              na.rm=T),
               missed_payment_mean_6_DiD_1 = mean(missed_payment[plot_time>0 & plot_time<=6],
                                              na.rm=T)-
                 mean(missed_payment[plot_time<0 & plot_time>=(-6)],
                                              na.rm=T),
               missed_payment_mean_3_DiD_1 = mean(missed_payment[plot_time>0 & plot_time<=3],
                                              na.rm=T)-
                 mean(missed_payment[plot_time<0 & plot_time>=(-3)],
                                              na.rm=T),
               nr_missed_payments_mean_6_DiD_1 = mean(nr_missed_payments[plot_time>0 & plot_time<=6],
                                              na.rm=T)-
                 mean(nr_missed_payments[plot_time<0 & plot_time>=(-6)],
                                              na.rm=T),
               nr_missed_payments_mean_3_DiD_1 = mean(nr_missed_payments[plot_time>0 & plot_time<=3],
                                              na.rm=T)-
                 mean(nr_missed_payments[plot_time<0 & plot_time>=(-3)],
                                              na.rm=T))
```

```
dd_data$compliance_mean_DiD_1yr[
  dd_data$tax=="Property"] <- dd_data$compliance_mean_3_DiD_1[dd_data$tax=="Property"]

dd_data$compliance_mean_DiD_1yr[dd_data$tax=="Head" | dd_data$tax=="Sewage"] <-
  dd_data$compliance_mean_6_DiD_1[dd_data$tax=="Head" | dd_data$tax=="Sewage"]

dd_data$missed_payment_mean_DiD_1yr[
  dd_data$tax=="Property"] <- dd_data$missed_payment_mean_3_DiD_1[dd_data$tax=="Property"]

dd_data$missed_payment_mean_DiD_1yr[dd_data$tax=="Head"| dd_data$tax=="Sewage"] <-
  dd_data$missed_payment_mean_6_DiD_1[dd_data$tax=="Head"| dd_data$tax=="Sewage"]

dd_data$nr_missed_payments_mean_DiD_1yr[
  dd_data$tax=="Property"] <- dd_data$nr_missed_payments_mean_3_DiD_1[dd_data$tax=="Property"]

dd_data$nr_missed_payments_mean_DiD_1yr[dd_data$tax=="Head"| dd_data$tax=="Sewage"] <-
  dd_data$nr_missed_payments_mean_6_DiD_1[dd_data$tax=="Head"| dd_data$tax=="Sewage"]

######## 1 year diff in diff

# all taxes, compliance
with(dd_data,
     t.test(compliance_mean_DiD_1yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  compliance_mean_DiD_1yr by won_lottery
## t = -0.3488, df = 4921, p-value = 0.7272
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.013585  0.009481
## sample estimates:
## mean in group 0 mean in group 1
##        -0.08039        -0.07834
```

```
# all taxes, missed payment
with(dd_data,
     t.test(missed_payment_mean_DiD_1yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  missed_payment_mean_DiD_1yr by won_lottery
## t = 0.4444, df = 4914, p-value = 0.6568
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.007641  0.012120
## sample estimates:
## mean in group 0 mean in group 1
##         0.05699         0.05475
```

```r
# all taxes, nr of missed payments
with(dd_data,
     t.test(nr_missed_payments_mean_DiD_1yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  nr_missed_payments_mean_DiD_1yr by won_lottery
## t = 0.8604, df = 4962, p-value = 0.3896
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.08389  0.21512
## sample estimates:
## mean in group 0 mean in group 1
##          0.2382          0.1726
```

```r
# Property, compliance
with(dd_data[dd_data$tax=="Property",],
     t.test(compliance_mean_DiD_1yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  compliance_mean_DiD_1yr by won_lottery
## t = -1.345, df = 2170, p-value = 0.1786
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.033008  0.006146
## sample estimates:
## mean in group 0 mean in group 1
##        -0.06914        -0.05571
```

```r
# Property, missed payment
with(dd_data[dd_data$tax=="Property",],
     t.test(missed_payment_mean_DiD_1yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  missed_payment_mean_DiD_1yr by won_lottery
## t = 1.786, df = 2152, p-value = 0.07427
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.001588  0.033944
## sample estimates:
## mean in group 0 mean in group 1
##         0.07097         0.05479
```

```r
# Property, nr of missed payments
with(dd_data[dd_data$tax=="Property",],
     t.test(nr_missed_payments_mean_DiD_1yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  nr_missed_payments_mean_DiD_1yr by won_lottery
## t = 0.7858, df = 2179, p-value = 0.4321
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.06716  0.15696
## sample estimates:
## mean in group 0 mean in group 1
##         0.07423         0.02932
```

```
# Sewage, compliance
with(dd_data[dd_data$tax=="Sewage",],
     t.test(compliance_mean_DiD_1yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  compliance_mean_DiD_1yr by won_lottery
## t = -0.2429, df = 811.4, p-value = 0.8081
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.03224  0.02514
## sample estimates:
## mean in group 0 mean in group 1
##        -0.08003        -0.07648
```

```
# Sewage, missed payment
with(dd_data[dd_data$tax=="Sewage",],
     t.test(missed_payment_mean_DiD_1yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  missed_payment_mean_DiD_1yr by won_lottery
## t = -0.2318, df = 803.9, p-value = 0.8167
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01937  0.01528
## sample estimates:
## mean in group 0 mean in group 1
##         0.04575         0.04779
```

```
# Sewage, nr of missed payments
with(dd_data[dd_data$tax=="Sewage",],
     t.test(nr_missed_payments_mean_DiD_1yr~won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  nr_missed_payments_mean_DiD_1yr by won_lottery
```

```
## t = 0.9032, df = 778.6, p-value = 0.3667
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.4489  1.2142
## sample estimates:
## mean in group 0 mean in group 1
##          0.8079          0.4252
```

```
# Head, compliance
with(dd_data[dd_data$tax=="Head",],
     t.test(compliance_mean_DiD_1yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  compliance_mean_DiD_1yr by won_lottery
## t = 1.598, df = 1834, p-value = 0.1103
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.002779  0.027218
## sample estimates:
## mean in group 0 mean in group 1
##        -0.09317        -0.10538
```

```
# Head, missed payment
with(dd_data[dd_data$tax=="Head",],
     t.test(missed_payment_mean_DiD_1yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  missed_payment_mean_DiD_1yr by won_lottery
## t = -1.785, df = 1824, p-value = 0.07436
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.025586  0.001201
## sample estimates:
## mean in group 0 mean in group 1
##         0.04573         0.05793
```

```
# Head, nr of missed payments
with(dd_data[dd_data$tax=="Head",],
     t.test(nr_missed_payments_mean_DiD_1yr~won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  nr_missed_payments_mean_DiD_1yr by won_lottery
## t = -0.4974, df = 1834, p-value = 0.619
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1112  0.0662
```

```
## sample estimates:
## mean in group 0 mean in group 1
##          0.1982          0.2207
```

```
# 3 year diff in diff setup

dd_data <- ddply(naturalex_gtp, c("account","tax", "won_lottery"),
                 summarise,
            compliance_mean_18_DiD_3 = mean(compliance[plot_time>0 & plot_time<=18],
                                            na.rm=T) -
              mean(compliance[plot_time<0 & plot_time>=(-18)],
                   na.rm=T),

            compliance_mean_9_DiD_3 = mean(compliance[plot_time>0 & plot_time<=9],
                                           na.rm=T) -
              mean(compliance[plot_time<0 & plot_time>=(-9)],
                   na.rm=T),

            missed_payment_mean_18_DiD_3 = mean(missed_payment[plot_time>0 & plot_time<=18],
                                                na.rm=T) -
              mean(missed_payment[plot_time<0 & plot_time>=(-18)],
                                  na.rm=T),

            missed_payment_mean_9_DiD_3 = mean(missed_payment[plot_time>0 & plot_time<=9],
                                               na.rm=T) -
              mean(missed_payment[plot_time<0 & plot_time>=(-9)],
                   na.rm=T),

            nr_missed_payments_mean_18_DiD_3 = mean(nr_missed_payments[plot_time>0 & plot_time<=18],
                                                    na.rm=T)-
              mean(nr_missed_payments[plot_time<0 & plot_time>=(-18)],
                   na.rm=T),

            nr_missed_payments_mean_9_DiD_3 = mean(nr_missed_payments[plot_time>0 & plot_time<=9],
                                   na.rm=T)-
              mean(nr_missed_payments[plot_time<0 & plot_time>=(-9)],
                                   na.rm=T))


dd_data$compliance_mean_DiD_3yr[
  dd_data$tax=="Property"] <- dd_data$compliance_mean_9_DiD_3[dd_data$tax=="Property"]

dd_data$compliance_mean_DiD_3yr[dd_data$tax=="Head" | dd_data$tax=="Sewage"] <-
  dd_data$compliance_mean_18_DiD_3[dd_data$tax=="Head" | dd_data$tax=="Sewage"]

dd_data$missed_payment_mean_DiD_3yr[dd_data$tax=="Property"] <-
  dd_data$missed_payment_mean_9_DiD_3[dd_data$tax=="Property"]

dd_data$missed_payment_mean_DiD_3yr[dd_data$tax=="Head"| dd_data$tax=="Sewage"] <-
  dd_data$missed_payment_mean_18_DiD_3[dd_data$tax=="Head"| dd_data$tax=="Sewage"]

dd_data$nr_missed_payments_mean_DiD_3yr[
  dd_data$tax=="Property"] <- dd_data$nr_missed_payments_mean_9_DiD_3[dd_data$tax=="Property"]
```

```r
dd_data$nr_missed_payments_mean_DiD_3yr[dd_data$tax=="Head"| dd_data$tax=="Sewage"] <-
  dd_data$nr_missed_payments_mean_18_DiD_3[dd_data$tax=="Head"| dd_data$tax=="Sewage"]

######## 3 year diff in diff

# all taxes, compliance
with(dd_data,
     t.test(compliance_mean_DiD_3yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  compliance_mean_DiD_3yr by won_lottery
## t = -1.341, df = 4916, p-value = 0.1798
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.018420  0.003453
## sample estimates:
## mean in group 0 mean in group 1
##        -0.04459        -0.03711
```

```r
# all taxes, missed payment
with(dd_data,
     t.test(missed_payment_mean_DiD_3yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  missed_payment_mean_DiD_3yr by won_lottery
## t = 1.601, df = 4910, p-value = 0.1095
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.001792  0.017741
## sample estimates:
## mean in group 0 mean in group 1
##         0.03580         0.02783
```

```r
# all taxes, nr of missed payments
with(dd_data,
     t.test(nr_missed_payments_mean_DiD_3yr~won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  nr_missed_payments_mean_DiD_3yr by won_lottery
## t = 1.956, df = 4839, p-value = 0.05054
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.0002338  0.1986982
## sample estimates:
## mean in group 0 mean in group 1
##         0.16082         0.06159
```

```r
# Property, compliance
with(dd_data[dd_data$tax=="Property",],
     t.test(compliance_mean_DiD_3yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  compliance_mean_DiD_3yr by won_lottery
## t = -1.686, df = 2158, p-value = 0.09198
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.034263  0.002586
## sample estimates:
## mean in group 0 mean in group 1
##        -0.04129        -0.02545
```

```r
# Property, missed payment
with(dd_data[dd_data$tax=="Property",],
     t.test(missed_payment_mean_DiD_3yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  missed_payment_mean_DiD_3yr by won_lottery
## t = 2.005, df = 2141, p-value = 0.04504
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.000381 0.034079
## sample estimates:
## mean in group 0 mean in group 1
##         0.04199         0.02476
```

```r
# Property, nr of missed payments
with(dd_data[dd_data$tax=="Property",],
     t.test(nr_missed_payments_mean_DiD_3yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  nr_missed_payments_mean_DiD_3yr by won_lottery
## t = 1.274, df = 2067, p-value = 0.2027
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.03575  0.16840
## sample estimates:
## mean in group 0 mean in group 1
##        0.060352       -0.005971
```

```r
# Sewage, compliance
with(dd_data[dd_data$tax=="Sewage",],
     t.test(compliance_mean_DiD_3yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  compliance_mean_DiD_3yr by won_lottery
## t = -0.0469, df = 766.8, p-value = 0.9626
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.02320  0.02212
## sample estimates:
## mean in group 0 mean in group 1
##        -0.03790        -0.03736
```

```
# Sewage, missed payment
with(dd_data[dd_data$tax=="Sewage",],
     t.test(missed_payment_mean_DiD_3yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  missed_payment_mean_DiD_3yr by won_lottery
## t = 0.2239, df = 772.4, p-value = 0.8229
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01522  0.01914
## sample estimates:
## mean in group 0 mean in group 1
##         0.02827         0.02631
```

```
# Sewage, nr of missed payments
with(dd_data[dd_data$tax=="Sewage",],
     t.test(nr_missed_payments_mean_DiD_3yr~won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  nr_missed_payments_mean_DiD_3yr by won_lottery
## t = 1.322, df = 789.5, p-value = 0.1866
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1496  0.7662
## sample estimates:
## mean in group 0 mean in group 1
##          0.5641          0.2557
```

```
# Head, compliance
with(dd_data[dd_data$tax=="Head",],
     t.test(compliance_mean_DiD_3yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  compliance_mean_DiD_3yr by won_lottery
```

```
## t = -0.0565, df = 1842, p-value = 0.955
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01648  0.01556
## sample estimates:
## mean in group 0 mean in group 1
##        -0.05094        -0.05048
```

```
# Head, missed payment
with(dd_data[dd_data$tax=="Head",],
     t.test(missed_payment_mean_DiD_3yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  missed_payment_mean_DiD_3yr by won_lottery
## t = -0.034, df = 1835, p-value = 0.9729
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.0147  0.0142
## sample estimates:
## mean in group 0 mean in group 1
##         0.03184         0.03209
```

```
# Head, nr of missed payments
with(dd_data[dd_data$tax=="Head",],
     t.test(nr_missed_payments_mean_DiD_3yr~won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  nr_missed_payments_mean_DiD_3yr by won_lottery
## t = 1.067, df = 1341, p-value = 0.2862
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.05465  0.18497
## sample estimates:
## mean in group 0 mean in group 1
##         0.11521         0.05005
```

TABLE 2. NATURAL EXPERIMENT. Effect of the tax holiday (T-test) Comparing winners to non-winners, difference of means test for the total debt as of October, 2014. We limit the comparison to pre-2013 winners as the tax holiday might still apply for more recent winners.

```
with(naturalex_debt_gtp[naturalex_debt_gtp$tax!="pr",],
     t.test(debt_amount ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  debt_amount by won_lottery
## t = 0.9541, df = 3798, p-value = 0.3401
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -313.3  907.2
## sample estimates:
## mean in group 0 mean in group 1
##           709.3           412.4
```

TABLE 3. NATURAL EXPERIMENT. Persistence of the effects of the tax lottery CONDITIONAL TEST). Comparing winners to non-winners, difference in difference analysis looking at the change between the 5 years after the tax holiday and the 5 years before winning the lottery. We conduct this test only if we find effects for missed payments, number of payments owed or total debt for the 3 year window.

```
# 5 year diff in diff setup

dd_data <-
  ddply(naturalex_gtp, c("account","tax", "won_lottery"), summarise,
     compliance_mean_30_DiD_5 = mean(compliance[plot_time>0 & plot_time<=30],
                                          na.rm=T) -
       mean(compliance[plot_time<0 & plot_time>=(-30)], na.rm=T),

      compliance_mean_15_DiD_5 = mean(compliance[plot_time>0 & plot_time<=15],
                                          na.rm=T) -
       mean(compliance[plot_time<0 & plot_time>=(-15)], na.rm=T),

      missed_payment_mean_30_DiD_5 = mean(missed_payment[plot_time>0 & plot_time<=30],
                                          na.rm=T) -
       mean(missed_payment[plot_time<0 & plot_time>=(-30)], na.rm=T),

      missed_payment_mean_15_DiD_5 = mean(missed_payment[plot_time>0 & plot_time<=15],
                                          na.rm=T) -
       mean(missed_payment[plot_time<0 & plot_time>=(-15)], na.rm=T),

        nr_missed_payments_mean_30_DiD_5 = mean(nr_missed_payments[plot_time>0 & plot_time<=30],
                                                na.rm=T)-
       mean(nr_missed_payments[plot_time<0 & plot_time>=(-30)], na.rm=T),

        nr_missed_payments_mean_15_DiD_5 = mean(nr_missed_payments[plot_time>0 & plot_time<=15],
                                  na.rm=T)-
      mean(nr_missed_payments[plot_time<0 & plot_time>=(-15)],
                                  na.rm=T))


dd_data$compliance_mean_DiD_5yr[
  dd_data$tax=="Property"] <- dd_data$compliance_mean_15_DiD_5[dd_data$tax=="Property"]

dd_data$compliance_mean_DiD_5yr[dd_data$tax=="Head" | dd_data$tax=="Sewage"] <-
  dd_data$compliance_mean_30_DiD_5[dd_data$tax=="Head" | dd_data$tax=="Sewage"]

dd_data$missed_payment_mean_DiD_5yr[dd_data$tax=="Property"] <-
  dd_data$missed_payment_mean_15_DiD_5[dd_data$tax=="Property"]

dd_data$missed_payment_mean_DiD_5yr[dd_data$tax=="Head"| dd_data$tax=="Sewage"] <-
  dd_data$missed_payment_mean_30_DiD_5[dd_data$tax=="Head"| dd_data$tax=="Sewage"]
```

```
dd_data$nr_missed_payments_mean_DiD_5yr[
  dd_data$tax=="Property"] <- dd_data$nr_missed_payments_mean_15_DiD_5[dd_data$tax=="Property"]

dd_data$nr_missed_payments_mean_DiD_5yr[dd_data$tax=="Head"| dd_data$tax=="Sewage"] <-
  dd_data$nr_missed_payments_mean_30_DiD_5[dd_data$tax=="Head"| dd_data$tax=="Sewage"]

######## 5 year diff in diff

# all taxes, compliance
with(dd_data,
     t.test(compliance_mean_DiD_5yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  compliance_mean_DiD_5yr by won_lottery
## t = -0.8931, df = 4906, p-value = 0.3718
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.016503  0.006173
## sample estimates:
## mean in group 0 mean in group 1
##        -0.03423        -0.02907
```

```
# all taxes, missed payment
with(dd_data,
     t.test(missed_payment_mean_DiD_5yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  missed_payment_mean_DiD_5yr by won_lottery
## t = 1.117, df = 4897, p-value = 0.2639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.004383  0.015999
## sample estimates:
## mean in group 0 mean in group 1
##         0.02635         0.02054
```

```
# all taxes, nr of missed payments
with(dd_data,
     t.test(nr_missed_payments_mean_DiD_5yr~won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  nr_missed_payments_mean_DiD_5yr by won_lottery
## t = 0.7472, df = 4826, p-value = 0.455
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.07256  0.16194
```

```
## sample estimates:
## mean in group 0 mean in group 1
##         0.10292         0.05823
```

```
# Property, compliance
with(dd_data[dd_data$tax=="Property",],
     t.test(compliance_mean_DiD_5yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  compliance_mean_DiD_5yr by won_lottery
## t = -1.015, df = 2155, p-value = 0.3102
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.028519  0.009067
## sample estimates:
## mean in group 0 mean in group 1
##        -0.02758        -0.01785
```

```
# Property, missed payment
with(dd_data[dd_data$tax=="Property",],
     t.test(missed_payment_mean_DiD_5yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  missed_payment_mean_DiD_5yr by won_lottery
## t = 1.185, df = 2136, p-value = 0.2362
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.006846  0.027749
## sample estimates:
## mean in group 0 mean in group 1
##         0.02761         0.01716
```

```
# Property, nr of missed payments
with(dd_data[dd_data$tax=="Property",],
     t.test(nr_missed_payments_mean_DiD_5yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  nr_missed_payments_mean_DiD_5yr by won_lottery
## t = 0.4836, df = 2114, p-value = 0.6287
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.08083  0.13374
## sample estimates:
## mean in group 0 mean in group 1
##        0.002816       -0.023641
```

```r
# Sewage, compliance
with(dd_data[dd_data$tax=="Sewage",],
     t.test(compliance_mean_DiD_5yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  compliance_mean_DiD_5yr by won_lottery
## t = 0.0521, df = 746.5, p-value = 0.9584
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.02245  0.02368
## sample estimates:
## mean in group 0 mean in group 1
##        -0.03139        -0.03200
```

```r
# Sewage, missed payment
with(dd_data[dd_data$tax=="Sewage",],
     t.test(missed_payment_mean_DiD_5yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  missed_payment_mean_DiD_5yr by won_lottery
## t = -0.0023, df = 778.1, p-value = 0.9982
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01918  0.01913
## sample estimates:
## mean in group 0 mean in group 1
##         0.01742         0.01744
```

```r
# Sewage, nr of missed payments
with(dd_data[dd_data$tax=="Sewage",],
     t.test(nr_missed_payments_mean_DiD_5yr~won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  nr_missed_payments_mean_DiD_5yr by won_lottery
## t = 0.3846, df = 793.4, p-value = 0.7006
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.4039  0.6008
## sample estimates:
## mean in group 0 mean in group 1
##          0.4681          0.3697
```

```r
# Head, compliance
with(dd_data[dd_data$tax=="Head",],
     t.test(compliance_mean_DiD_5yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  compliance_mean_DiD_5yr by won_lottery
## t = -0.2447, df = 1877, p-value = 0.8067
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01931  0.01502
## sample estimates:
## mean in group 0 mean in group 1
##        -0.04283        -0.04069
```

```
# Head, missed payment
with(dd_data[dd_data$tax=="Head",],
     t.test(missed_payment_mean_DiD_5yr ~ won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  missed_payment_mean_DiD_5yr by won_lottery
## t = 0.3309, df = 1877, p-value = 0.7408
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01265  0.01778
## sample estimates:
## mean in group 0 mean in group 1
##         0.02847         0.02590
```

```
# Head, nr of missed payments
with(dd_data[dd_data$tax=="Head",],
     t.test(nr_missed_payments_mean_DiD_5yr~won_lottery))
```

```
##
##  Welch Two Sample t-test
##
## data:  nr_missed_payments_mean_DiD_5yr by won_lottery
## t = 0.6976, df = 1555, p-value = 0.4855
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1126  0.2369
## sample estimates:
## mean in group 0 mean in group 1
##         0.07169         0.00953
```

Mechanisms for H1A

Mechanism 1A.1: ***Informational.***

TABLE 4. FIELD EXPERIMENT. Informational mechanism. Good and bad taxpayers, comparison of treatments 1, 2 and 4 (pooled) vs. the placebo control group (treatment 0). First differences use the value of the dependent variable for the pre-treatment period (March 2014).

```r
### Creating first differences outcomes for the field experiment

fieldex$missed_payment_DiD  <- fieldex$july_ontime - fieldex$march_ontime  # missed payment
fieldex$web_bill_DiD  <- fieldex$july_web_access - fieldex$march_web_access # web access
fieldex$payments_owed_DiD  <- fieldex$july_nrbills_owed - fieldex$adeudadas_2014_MAR # nr payments owed
names(fieldex)
```

```
##  [1] "account"               "due_date"
##  [3] "bill_adddress"         "property_value_2014"
##  [5] "bill_postalcode"       "has_debt_since"
##  [7] "adeudadas_2014_MAR"    "treatment"
##  [9] "type"                  "localidad"
## [11] "phase"                 "correlativo"
## [13] "address_freq"          "bill_padron"
## [15] "property_padron"       "property_address"
## [17] "retired"               "paid_in_advance"
## [19] "march_web_access"      "march_web_access_date"
## [21] "march_ontime"          "july_ontime"
## [23] "july_nrbills_owed"     "july_web_access"
## [25] "july_web_access_date"  "adeudadas_2009_MAR"
## [27] "adeudadas_2009_JUL"    "adeudadas_2009_NOV"
## [29] "adeudadas_2010_MAR"    "adeudadas_2010_JUL"
## [31] "adeudadas_2010_NOV"    "adeudadas_2011_MAR"
## [33] "adeudadas_2011_JUL"    "adeudadas_2011_NOV"
## [35] "adeudadas_2012_MAR"    "adeudadas_2012_JUL"
## [37] "adeudadas_2012_NOV"    "adeudadas_2013_MAR"
## [39] "adeudadas_2013_JUL"    "adeudadas_2013_NOV"
## [41] "compliance_july"       "current_debt_DiD"
## [43] "missed_payment_DiD"    "web_bill_DiD"
## [45] "payments_owed_DiD"
```

```r
# compliance
fieldex$compliance_march <- ifelse(fieldex$march_ontime==1 & fieldex$adeudadas_2014_MAR==0, 1, 0)
fieldex$compliance_july <- ifelse(fieldex$july_ontime==1 & fieldex$july_nrbills_owed==0, 1, 0)
fieldex$compliance_DiD <- fieldex$compliance_july - fieldex$compliance_march

## Pooling treatments 1, 2, and 4 vs 0
fieldex$pooled_124_0 <- NA
fieldex$pooled_124_0[fieldex$treatment==0] <- 0
fieldex$pooled_124_0[fieldex$treatment==1] <- 1
fieldex$pooled_124_0[fieldex$treatment==2] <- 1
fieldex$pooled_124_0[fieldex$treatment==4] <- 1
```

```r
# Missed payment
with(fieldex, t.test(missed_payment_DiD  ~ pooled_124_0))
```

```
##
##  Welch Two Sample t-test
##
## data:  missed_payment_DiD by pooled_124_0
## t = 0.0254, df = 8488, p-value = 0.9797
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -0.01506  0.01545
## sample estimates:
## mean in group 0 mean in group 1
##         0.02707         0.02687
```

```
# Web bill request
with(fieldex, t.test(web_bill_DiD ~ pooled_124_0))
```

```
##
##  Welch Two Sample t-test
##
## data:  web_bill_DiD by pooled_124_0
## t = -0.5018, df = 8392, p-value = 0.6158
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.016399  0.009715
## sample estimates:
## mean in group 0 mean in group 1
##         0.07613         0.07947
```

```
# Payments owed
with(fieldex, t.test(payments_owed_DiD ~ pooled_124_0))
```

```
##
##  Welch Two Sample t-test
##
## data:  payments_owed_DiD by pooled_124_0
## t = 1.561, df = 11034, p-value = 0.1185
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.007046  0.062209
## sample estimates:
## mean in group 0 mean in group 1
##          0.1844          0.1568
```

```
# Current debt
with(fieldex, t.test(current_debt_DiD ~ pooled_124_0))
```

```
##
##  Welch Two Sample t-test
##
## data:  current_debt_DiD by pooled_124_0
## t = -1.755, df = 8730, p-value = 0.07925
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6901.6   380.8
## sample estimates:
## mean in group 0 mean in group 1
##           -1925            1336
```

```r
# Compliance (test conditional on previous outcomes)
with(fieldex, t.test(compliance_DiD ~ pooled_124_0))
```

```
##
##  Welch Two Sample t-test
##
## data:  compliance_DiD by pooled_124_0
## t = -0.1547, df = 8660, p-value = 0.8771
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01584  0.01353
## sample estimates:
## mean in group 0 mean in group 1
##         0.02552         0.02668
```

TABLE 6. FIELD EXPERIMENT: Comparison of effects for good and bad taxpayers: difference of the difference of means for the comparison of treatments 1, 2 and 4 (pooled) vs. the placebo control group (treatment 0).

```r
# Missed payments
comp.eff(fieldex[fieldex$type=="eligible",]$missed_payment_DiD,
         fieldex[fieldex$type=="eligible",]$pooled_124_0,
         fieldex[fieldex$type=="noneligible",]$missed_payment_DiD,
         fieldex[fieldex$type=="noneligible",]$pooled_124_0)
```

```
##             DM       SE
## DM 1  0.005281 0.006721
## DM 2 -0.004853 0.012627

## Diff in effects              SE
##         0.01013         0.01430
```

```r
# Web access
comp.eff(fieldex[fieldex=="eligible",]$web_bill_DiD,
         fieldex[fieldex$type=="eligible",]$pooled_124_0,
         fieldex[fieldex$type=="noneligible",]$web_bill_DiD,
         fieldex[fieldex$type=="noneligible",]$pooled_124_0)
```

```
##             DM       SE
## DM 1       NaN       NA
## DM 2 2.651e-05 0.008193

## Diff in effects              SE
##             NaN              NA
```

Mechanism 1A.2. ***Attitudinal.***

Mechanisms for H1B

Regarding Hypothesis 1B, at least two mechanisms might explain a negative effect of winning the lottery on future compliance:

Mechanism 1B.1: ***Income effects.***

TABLE 7. NATURAL EXPERIMENT. Income effects. Comparison of winners vs. non-winners: heterogeneous effects of winning the lottery by tax bracket.

```
# Coding tax brackets
  a <- 418958
  b <- 1047393
  c <- 2094784
  d <- 41895699

x <- naturalex_gtp$property_value_2014
naturalex_gtp$prop_tax_bracket <- ifelse(x==NA, NA,
                                         naturalex_gtp$prop_tax_bracket)
naturalex_gtp$prop_tax_bracket <- ifelse(x <= a, 1,
                                         naturalex_gtp$prop_tax_bracket)
naturalex_gtp$prop_tax_bracket <- ifelse(((x > a) & (x <= b)), 2,
                                         naturalex_gtp$prop_tax_bracket)
naturalex_gtp$prop_tax_bracket <- ifelse(((x > b) & (x <= c)), 3,
                                         naturalex_gtp$prop_tax_bracket)
naturalex_gtp$prop_tax_bracket <- ifelse(((x > c) & (x <= d)), 4,
                                         naturalex_gtp$prop_tax_bracket)
naturalex_gtp$prop_tax_bracket <- ifelse(x > d, 5,
                                         naturalex_gtp$prop_tax_bracket)

tax_bracket_data <- ddply(naturalex_gtp[is.na(naturalex_gtp$prop_tax_bracket)==F,],
                          c("prop_tax_bracket"),
                 summarise,
                 mean_missed_winners = mean(missed_payment[plot_time==1 & won_lottery==1],
                                            na.rm=T),
                 N_winners = length(na.omit(missed_payment[plot_time==1 & won_lottery==1])),
                 mean_missed_losers = mean(missed_payment[plot_time==1 & won_lottery==0],
                                            na.rm=T),
                 N_losers = length(na.omit(missed_payment[plot_time==1 & won_lottery==0])))

tax_bracket_data
```

```
##   prop_tax_bracket mean_missed_winners N_winners mean_missed_losers
## 1                1              0.1051       923            0.10584
## 2                2              0.1127       772            0.11443
## 3                3              0.0907       441            0.08794
## 4                4              0.1029       311            0.09265
## 5                5              0.0000        22            0.14286
##   N_losers
## 1      907
## 2      804
## 3      398
## 4      313
## 5       21
```

```
chisq.test(tax_bracket_data[,c(3,5)])
```

```
##
```

```
##  Pearson's Chi-squared test
##
## data:  tax_bracket_data[, c(3, 5)]
## X-squared = 2.886, df = 4, p-value = 0.5772
```

Mechanism 1B.2: ***Behavioral/'habit' effects.***

TABLE 8. NATURAL EXPERIMENT. Habit effects. Winners vs. non-winners: heterogeneous treatment effects by time since winning (heterogeneous effects; 1, 2 and 3 years).

```
naturalex_gtp$year_since_win[naturalex_gtp$tax=="Property" &
                          naturalex_gtp$plot_time > 3 & naturalex_gtp$plot_time < 7]  <- 2
naturalex_gtp$year_since_win[naturalex_gtp$tax=="Property" &
                          naturalex_gtp$plot_time > 6 & naturalex_gtp$plot_time < 10]  <- 3
naturalex_gtp$year_since_win[naturalex_gtp$tax=="Property" &
                          naturalex_gtp$plot_time > 9 & naturalex_gtp$plot_time < 13]  <- 4

naturalex_gtp$year_since_win[(naturalex_gtp$tax=="Sewage" | naturalex_gtp$tax=="Head") &
                          naturalex_gtp$plot_time > 6 & naturalex_gtp$plot_time < 13]  <- 2
naturalex_gtp$year_since_win[(naturalex_gtp$tax=="Sewage" | naturalex_gtp$tax=="Head") &
                          naturalex_gtp$plot_time > 12 & naturalex_gtp$plot_time < 19] <- 3
naturalex_gtp$year_since_win[(naturalex_gtp$tax=="Sewage" | naturalex_gtp$tax=="Head") &
                          naturalex_gtp$plot_time > 18 & naturalex_gtp$plot_time < 25]  <- 4


comp.eff(naturalex_gtp[naturalex_gtp$year_since_win==2,]$missed_payment,
         naturalex_gtp[naturalex_gtp$year_since_win==2,]$won_lottery,
         naturalex_gtp[naturalex_gtp$year_since_win==3,]$missed_payment,
         naturalex_gtp[naturalex_gtp$year_since_win==3,]$won_lottery)
```

```
##            DM       SE
## DM 1 0.002196 0.003928
## DM 2 0.001752 0.004475

## Diff in effects              SE
##       0.0004442       0.0059547
```

```
comp.eff(naturalex_gtp[naturalex_gtp$year_since_win==3,]$missed_payment,
         naturalex_gtp[naturalex_gtp$year_since_win==3,]$won_lottery,
         naturalex_gtp[naturalex_gtp$year_since_win==4,]$missed_payment,
         naturalex_gtp[naturalex_gtp$year_since_win==4,]$won_lottery)
```

```
##            DM       SE
## DM 1 0.001752 0.004475
## DM 2 0.007384 0.005126

## Diff in effects              SE
##       -0.005631        0.006805
```

Mechanisms for H1C

Mechanism 1C.1: ***Erroneous beliefs.***

Rewards vs. punishments. (Positive vs. negative incentives)

Hypothesis 2A.

TABLE 9. FIELD EXPERIMENT. Priming knowledge of punishment. Good and bad taxpayers, comparison of treatments 3 and 5 (pooled) vs. the placebo control group (treatment 0).

```
# Creating pooled variable
fieldex$pooled_35_0 <- NA
fieldex$pooled_35_0[fieldex$treatment==0] <- 0
fieldex$pooled_35_0[fieldex$treatment==3] <- 1
fieldex$pooled_35_0[fieldex$treatment==5] <- 1


# Missed payment
with(fieldex, t.test(missed_payment_DiD ~ pooled_35_0))
```

```
##
##  Welch Two Sample t-test
##
## data:  missed_payment_DiD by pooled_35_0
## t = 0.6746, df = 8472, p-value = 0.5
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01001  0.02051
## sample estimates:
## mean in group 0 mean in group 1
##         0.02707         0.02182
```

```
# Web bill request
with(fieldex, t.test(web_bill_DiD ~ pooled_35_0))
```

```
##
##  Welch Two Sample t-test
##
## data:  web_bill_DiD by pooled_35_0
## t = -0.2676, df = 8509, p-value = 0.789
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01492  0.01134
## sample estimates:
## mean in group 0 mean in group 1
##         0.07613         0.07792
```

```
# Payments owed
with(fieldex, t.test(payments_owed_DiD ~ pooled_35_0))
```

```
##
##  Welch Two Sample t-test
##
## data:  payments_owed_DiD by pooled_35_0
## t = -0.3257, df = 11757, p-value = 0.7446
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.04237  0.03029
## sample estimates:
## mean in group 0 mean in group 1
##          0.1844          0.1904
```

```
# Current debt
with(fieldex, t.test(current_debt_DiD ~ pooled_35_0))
```

```
##
##  Welch Two Sample t-test
##
## data:  current_debt_DiD by pooled_35_0
## t = -0.3088, df = 8674, p-value = 0.7575
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4208  3062
## sample estimates:
## mean in group 0 mean in group 1
##           -1925           -1352
```

```
# Compliance (test conditional on results of previous tests)
with(fieldex, t.test(compliance_DiD ~ pooled_35_0))
```

```
##
##  Welch Two Sample t-test
##
## data:  compliance_DiD by pooled_35_0
## t = 0.6624, df = 8569, p-value = 0.5077
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.009692  0.019585
## sample estimates:
## mean in group 0 mean in group 1
##         0.02552         0.02058
```

TABLE 10. FIELD EXPERIMENT: Comparison of effects for good and bad taxpayers: difference of the difference in means for the comparison of treatments 3 and 5 (priming knowledge of sanctions, pooled) vs. the placebo control group (treatment 0).

```
# Missed payments
comp.eff(fieldex[fieldex$type=="eligible",]$missed_payment_DiD,
         fieldex[fieldex$type=="eligible",]$pooled_35_0,
         fieldex[fieldex$type=="noneligible",]$missed_payment_DiD,
         fieldex[fieldex$type=="noneligible",]$pooled_35_0)
```

```
##             DM       SE
## DM 1  0.001792 0.006792
## DM 2 -0.011222 0.012603

## Diff in effects             SE
##         0.01301         0.01432
```

```
# Web access
comp.eff(fieldex[fieldex$type=="eligible",]$web_bill_DiD,
         fieldex[fieldex$type=="eligible",]$pooled_35_0,
         fieldex[fieldex$type=="noneligible",]$web_bill_DiD,
         fieldex[fieldex$type=="noneligible",]$pooled_35_0)
```

```
##             DM      SE
## DM 1 0.0004106 0.01096
## DM 2 0.0029640 0.00830

## Diff in effects             SE
##       -0.002553        0.013751
```

We also assess one possible mechanism:

Mechanism 2A.1: ***Beliefs about probability of punishment***

We are especially interested in comparing the size of the effects of manipulating negative and positive incentives to pay taxes, though we are agnostic about the direction of the effect:

Hypothesis 2B.

TABLE 11. FIELD EXPERIMENT. Positive vs negative incentives. Good and bad taxpayers, comparison of treatments 1, 2 and 4 (positive incentives, pooled) vs 3 and 5 (negative incentives, pooled). Test using compliance conditional on significant effects for missed payment, number of payments owed or total debt.

```
fieldex$pooled_124_35 <- NA
fieldex$pooled_124_35[fieldex$treatment==1] <- 1
fieldex$pooled_124_35[fieldex$treatment==2] <- 1
fieldex$pooled_124_35[fieldex$treatment==4] <- 1
fieldex$pooled_124_35[fieldex$treatment==3] <- 0
fieldex$pooled_124_35[fieldex$treatment==5] <- 0


# Missed payment
with(fieldex, t.test(missed_payment_DiD ~ pooled_124_35))
```

```
##
##  Welch Two Sample t-test
##
## data:  missed_payment_DiD by pooled_124_35
## t = -0.8012, df = 17108, p-value = 0.423
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -0.017413  0.007308
## sample estimates:
## mean in group 0 mean in group 1
##         0.02182         0.02687
```

```
# Web bill request
with(fieldex, t.test(web_bill_DiD ~ pooled_124_35))
```

```
##
##  Welch Two Sample t-test
##
## data:  web_bill_DiD by pooled_124_35
## t = -0.2909, df = 17392, p-value = 0.7711
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01199  0.00889
## sample estimates:
## mean in group 0 mean in group 1
##         0.07792         0.07947
```

```
# Payments owed
 with(fieldex, t.test(payments_owed_DiD ~ pooled_124_35))
```

```
##
##  Welch Two Sample t-test
##
## data:  payments_owed_DiD by pooled_124_35
## t = 1.879, df = 17206, p-value = 0.06025
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.001449  0.068687
## sample estimates:
## mean in group 0 mean in group 1
##          0.1904          0.1568
```

```
# Current debt
with(fieldex, t.test(current_debt_DiD ~ pooled_124_35))
```

```
##
##  Welch Two Sample t-test
##
## data:  current_debt_DiD by pooled_124_35
## t = -1.776, df = 17405, p-value = 0.07573
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5653.9   278.4
## sample estimates:
## mean in group 0 mean in group 1
##           -1352            1336
```

```
# Compliance debt
with(fieldex, t.test(compliance_DiD ~ pooled_124_35))
```

```
##
##  Welch Two Sample t-test
##
## data:  compliance_DiD by pooled_124_35
## t = -1.007, df = 17370, p-value = 0.3141
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.017992  0.005781
## sample estimates:
## mean in group 0 mean in group 1
##         0.02058         0.02668
```

TABLE 12. FIELD EXPERIMENT. Comparison of effects of positive vs negative incentives for good and bad taxpayers: difference of the difference in means for the comparison of treatments 1, 2 and 4 (pooled) and 3 and 5 (pooled).

```
# Missed payments
comp.eff(fieldex[fieldex$type=="eligible",]$missed_payment_DiD,
         fieldex[fieldex$type=="eligible",]$pooled_124_35,
         fieldex[fieldex$type=="noneligible",]$missed_payment_DiD,
         fieldex[fieldex$type=="noneligible",]$pooled_124_35)
```

```
##            DM       SE
## DM 1 0.003489 0.005451
## DM 2 0.006368 0.010235

## Diff in effects             SE
##        -0.00288         0.01160
```

```
# Web access
comp.eff(fieldex[fieldex$type=="eligible",]$web_bill_DiD,
         fieldex[fieldex$type=="eligible",]$pooled_124_35,
         fieldex[fieldex$type=="noneligible",]$web_bill_DiD,
         fieldex[fieldex$type=="noneligible",]$pooled_124_35)
```

```
##             DM       SE
## DM 1  0.007406 0.008713
## DM 2 -0.002938 0.006602

## Diff in effects             SE
##         0.01034         0.01093
```

Hypothesis 2C: *Marginal taxpayers I.*

TABLE 13. FIELD EXPERIMENT. Marginal taxpayers. Good taxpayers. Heterogeneous effects, taxpayers at risk. Comparison of treatment effect of 1, 2 and 4 (pooled) vs control (A-Information about the tax lottery), on one test and 3 and 5 (pooled) vs control on another (B-Information about sanctions).}

```r
# Identifying good taxpayers with history of debt
# Ever owed a bill since March 2009?
names(fieldex[,c(26:40,7)])
```

```
##  [1] "adeudadas_2009_MAR" "adeudadas_2009_JUL" "adeudadas_2009_NOV"
##  [4] "adeudadas_2010_MAR" "adeudadas_2010_JUL" "adeudadas_2010_NOV"
##  [7] "adeudadas_2011_MAR" "adeudadas_2011_JUL" "adeudadas_2011_NOV"
## [10] "adeudadas_2012_MAR" "adeudadas_2012_JUL" "adeudadas_2012_NOV"
## [13] "adeudadas_2013_MAR" "adeudadas_2013_JUL" "adeudadas_2013_NOV"
## [16] "adeudadas_2014_MAR"
```

```r
sum_bills_owed <- apply(fieldex[,c(26:40,7)], 1, sum)

fieldex$goodtp_at_risk <- ifelse(sum_bills_owed>0, 1, 0)
table(fieldex$goodtp_at_risk[fieldex$type=="eligible"])
```

```
##
##    0    1
## 7164 2297
```

```r
### A. Information on the lottery - Heterogeneous effects for taxpayers at risk

# Missed payments
comp.eff(fieldex$missed_payment_DiD[fieldex$type=="eligible" & fieldex$goodtp_at_risk==1],
         fieldex$pooled_124_0[fieldex$type=="eligible" & fieldex$goodtp_at_risk==1],
         fieldex$missed_payment_DiD[fieldex$type=="eligible" & fieldex$goodtp_at_risk==0],
         fieldex$pooled_124_0[fieldex$type=="eligible" & fieldex$goodtp_at_risk==0])
```

```
##            DM      SE
## DM 1 0.013129 0.01708
## DM 2 0.002724 0.00696

## Diff in effects              SE
##         0.01041         0.01845
```

```r
# Web access
comp.eff(fieldex$web_bill_DiD[fieldex$type=="eligible" & fieldex$goodtp_at_risk==1],
         fieldex$pooled_124_0[fieldex$type=="eligible" & fieldex$goodtp_at_risk==1],
         fieldex$web_bill_DiD[fieldex$type=="eligible" & fieldex$goodtp_at_risk==0],
         fieldex$pooled_124_0[fieldex$type=="eligible" & fieldex$goodtp_at_risk==0])
```

```
##            DM      SE
## DM 1 0.028336 0.02132
## DM 2 0.001317 0.01277

## Diff in effects              SE
##         0.02702         0.02485
```

```
### B. Information on sanctions - Heterogeneous effects for taxpayers at risk

# Missed payments
comp.eff(fieldex$missed_payment_DiD[fieldex$type=="eligible" & fieldex$goodtp_at_risk==1],
         fieldex$pooled_35_0[fieldex$type=="eligible" & fieldex$goodtp_at_risk==1],
         fieldex$missed_payment_DiD[fieldex$type=="eligible" & fieldex$goodtp_at_risk==0],
         fieldex$pooled_35_0[fieldex$type=="eligible" & fieldex$goodtp_at_risk==0])
```

```
##             DM       SE
## DM 1  0.012920 0.017062
## DM 2 -0.001441 0.007071

## Diff in effects              SE
##         0.01436         0.01847
```

```
# Web access
comp.eff(fieldex$web_bill_DiD[fieldex$type=="eligible" & fieldex$goodtp_at_risk==1],
         fieldex$pooled_35_0[fieldex$type=="eligible" & fieldex$goodtp_at_risk==1],
         fieldex$web_bill_DiD[fieldex$type=="eligible" & fieldex$goodtp_at_risk==0],
         fieldex$pooled_35_0[fieldex$type=="eligible" & fieldex$goodtp_at_risk==0])
```

```
##             DM      SE
## DM 1  0.008871 0.02113
## DM 2 -0.002059 0.01279

## Diff in effects              SE
##         0.01093         0.02470
```

TABLE 14. FIELD EXPERIMENT. Marginal taxpayers. Bad taxpayers. Heterogeneous effects, salvageable taxpayers. Comparison of treatment effect of 1, 2 and 4 (pooled) vs control (A-Information about the tax lottery), on one test and 3 and 5 (pooled) vs control on another (B-Information about sanctions). Test using compliance conditional on significant effects for missed payment, number of payments owed or total debt.

```
# Identifying bad taxpayers not too in debt

# Ever owed a bill since March 2009?
names(fieldex)[7]
```

```
## [1] "adeudadas_2014_MAR"
```

```
fieldex$salvageable_btp <- ifelse(fieldex$adeudadas_2014_MAR>3, 0, 1)

### A. Information on the lottery - Heterogeneous effects for taxpayers at risk

# Missed payments
comp.eff(fieldex$missed_payment_DiD[fieldex$type=="noneligible" & fieldex$salvageable_btp==1],
         fieldex$pooled_124_0[fieldex$type=="noneligible" & fieldex$salvageable_btp==1],
         fieldex$missed_payment_DiD[fieldex$type=="noneligible" & fieldex$salvageable_btp==0],
         fieldex$pooled_124_0[fieldex$type=="noneligible" & fieldex$salvageable_btp==0])
```

```
##             DM      SE
## DM 1 -0.003946 0.01620
## DM 2 -0.006449 0.01261

## Diff in effects              SE
##        0.002503        0.020529
```

```
# Web access
comp.eff(fieldex$web_bill_DiD[fieldex$type=="noneligible" & fieldex$salvageable_btp==1],
         fieldex$pooled_124_0[fieldex$type=="noneligible" & fieldex$salvageable_btp==1],
         fieldex$web_bill_DiD[fieldex$type=="noneligible" & fieldex$salvageable_btp==0],
         fieldex$pooled_124_0[fieldex$type=="noneligible" & fieldex$salvageable_btp==0])
```

```
##             DM       SE
## DM 1  0.002772 0.010583
## DM 2 -0.007161 0.008238

## Diff in effects              SE
##        0.009933        0.013411
```

```
# Nr of payments owed
comp.eff(fieldex$payments_owed_DiD[fieldex$type=="noneligible" & fieldex$salvageable_btp==1],
         fieldex$pooled_124_0[fieldex$type=="noneligible" & fieldex$salvageable_btp==1],
         fieldex$payments_owed_DiD[fieldex$type=="noneligible" & fieldex$salvageable_btp==0],
         fieldex$pooled_124_0[fieldex$type=="noneligible" & fieldex$salvageable_btp==0])
```

```
##             DM      SE
## DM 1 -0.009227 0.01682
## DM 2 -0.175959 0.10960

## Diff in effects              SE
##          0.1667          0.1109
```

```
# Compliance (conditional)
comp.eff(fieldex$compliance_DiD[fieldex$type=="noneligible" & fieldex$salvageable_btp==1],
         fieldex$pooled_124_0[fieldex$type=="noneligible" & fieldex$salvageable_btp==1],
         fieldex$compliance_DiD[fieldex$type=="noneligible" & fieldex$salvageable_btp==0],
         fieldex$pooled_124_0[fieldex$type=="noneligible" & fieldex$salvageable_btp==0])
```

```
##             DM       SE
## DM 1 -0.002700 0.015822
## DM 2 -0.001144 0.005121

## Diff in effects              SE
##       -0.001557        0.016630
```

```
### B. Information on sanctions - Heterogeneous effects for taxpayers at risk

# Missed payments
comp.eff(fieldex$missed_payment_DiD[fieldex$type=="noneligible" & fieldex$salvageable_btp==1],
         fieldex$pooled_35_0[fieldex$type=="noneligible" & fieldex$salvageable_btp==1],
         fieldex$missed_payment_DiD[fieldex$type=="noneligible" & fieldex$salvageable_btp==0],
         fieldex$pooled_35_0[fieldex$type=="noneligible" & fieldex$salvageable_btp==0])
```

```
##             DM      SE
## DM 1 -0.011346 0.01621
## DM 2 -0.005506 0.01291

## Diff in effects             SE
##        -0.00584         0.02072
```

```
# Web access
comp.eff(fieldex$web_bill_DiD[fieldex$type=="noneligible" & fieldex$salvageable_btp==1],
         fieldex$pooled_35_0[fieldex$type=="noneligible" & fieldex$salvageable_btp==1],
         fieldex$web_bill_DiD[fieldex$type=="noneligible" & fieldex$salvageable_btp==0],
         fieldex$pooled_35_0[fieldex$type=="noneligible" & fieldex$salvageable_btp==0])
```

```
##            DM       SE
## DM 1  0.01219 0.010779
## DM 2 -0.01996 0.008121

## Diff in effects             SE
##         0.03216         0.01350
```

```
# Nr of payments owed
comp.eff(fieldex$payments_owed_DiD[fieldex$type=="noneligible" & fieldex$salvageable_btp==1],
         fieldex$pooled_35_0[fieldex$type=="noneligible" & fieldex$salvageable_btp==1],
         fieldex$payments_owed_DiD[fieldex$type=="noneligible" & fieldex$salvageable_btp==0],
         fieldex$pooled_35_0[fieldex$type=="noneligible" & fieldex$salvageable_btp==0])
```

```
##            DM      SE
## DM 1  0.03338 0.02670
## DM 2 -0.07526 0.09817

## Diff in effects             SE
##          0.1086          0.1017
```

```
# Compliance (conditional)
comp.eff(fieldex$compliance_DiD[fieldex$type=="noneligible" & fieldex$salvageable_btp==1],
         fieldex$pooled_35_0[fieldex$type=="noneligible" & fieldex$salvageable_btp==1],
         fieldex$compliance_DiD[fieldex$type=="noneligible" & fieldex$salvageable_btp==0],
         fieldex$pooled_35_0[fieldex$type=="noneligible" & fieldex$salvageable_btp==0])
```

```
##              DM       SE
## DM 1 -0.0127398 0.015810
## DM 2  0.0009157 0.005252

## Diff in effects             SE
##        -0.01366         0.01666
```

Hypothesis 2D: *Marginal taxpayers II.*

Individual vs. social incentives.

Hypothesis 3A: *Social benefits.*

TABLE 15. FIELD EXPERIMENT. Good and bad taxpayers. Social vs individual rewards. Comparison of treatments 1, 2 (pooled) vs 4. Test using compliance conditional on significant effects for missed payment, number of payments owed or total debt.

```
fieldex$pooled_12_4 <- NA
fieldex$pooled_12_4[fieldex$treatment==1] <- 1
fieldex$pooled_12_4[fieldex$treatment==2] <- 1
fieldex$pooled_12_4[fieldex$treatment==4] <- 0

# Missed payment
with(fieldex, t.test(missed_payment_DiD ~ pooled_12_4))
```

```
##
##  Welch Two Sample t-test
##
## data:  missed_payment_DiD by pooled_12_4
## t = -0.4156, df = 8623, p-value = 0.6777
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.02119  0.01378
## sample estimates:
## mean in group 0 mean in group 1
##         0.02499         0.02870
```

```
# Web bill request
with(fieldex, t.test(web_bill_DiD ~ pooled_12_4))
```

```
##
##  Welch Two Sample t-test
##
## data:  web_bill_DiD by pooled_12_4
## t = -0.6687, df = 8780, p-value = 0.5037
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.019616  0.009637
## sample estimates:
## mean in group 0 mean in group 1
##         0.07694         0.08193
```

```
# Payments owed
with(fieldex, t.test(payments_owed_DiD ~ pooled_12_4))
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  payments_owed_DiD by pooled_12_4
## t = -1.499, df = 7244, p-value = 0.134
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.08363  0.01116
## sample estimates:
## mean in group 0 mean in group 1
##          0.1384          0.1746
```

```
# Current debt
with(fieldex, t.test(current_debt_DiD ~ pooled_12_4))
```

```
##
##  Welch Two Sample t-test
##
## data:  current_debt_DiD by pooled_12_4
## t = -0.7607, df = 8781, p-value = 0.4468
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5836  2573
## sample estimates:
## mean in group 0 mean in group 1
##           508.1          2139.6
```

```
# Compliance (conditional)
with(fieldex, t.test(compliance_DiD ~ pooled_12_4))
```

```
##
##  Welch Two Sample t-test
##
## data:  compliance_DiD by pooled_12_4
## t = -0.4933, df = 8763, p-value = 0.6218
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.02114  0.01264
## sample estimates:
## mean in group 0 mean in group 1
##         0.02453         0.02878
```

TABLE 16. FIELD EXPERIMENT. Social (4) vs individual rewards (1 and 2, pooled), comparison of effect between good and bad taxpayers.

```
# Missed payments
comp.eff(fieldex[fieldex$type=="eligible",]$missed_payment_DiD,
         fieldex[fieldex$type=="eligible",]$pooled_12_4,
         fieldex[fieldex$type=="noneligible",]$missed_payment_DiD,
         fieldex[fieldex$type=="noneligible",]$pooled_12_4)
```

```
##             DM      SE
## DM 1 -0.005270 0.00758
## DM 2  0.009944 0.01452
```

```
## Diff in effects              SE
##        -0.01521         0.01638
```

```
# Web access
comp.eff(fieldex[fieldex$type=="eligible",]$web_bill_DiD,
         fieldex[fieldex$type=="eligible",]$pooled_12_4,
         fieldex[fieldex$type=="noneligible",]$web_bill_DiD,
         fieldex[fieldex$type=="noneligible",]$pooled_12_4)
```

```
##            DM       SE
## DM 1 0.001810 0.012327
## DM 2 0.007491 0.009147

## Diff in effects              SE
##       -0.005681        0.015350
```

Hypothesis 3B: *Social sanctions.*

TABLE 17. FIELD EXPERIMENT. Good and bad taxpayers. Social vs individual sanctions. Comparison of treatments 3 vs 5. Test using compliance conditional on significant effects for missed payment, number of payments owed or total debt.

```
fieldex$pooled_3_5 <- NA
fieldex$pooled_3_5[fieldex$treatment==3] <- 0
fieldex$pooled_3_5[fieldex$treatment==5] <- 1


# Missed payment
with(fieldex, t.test(missed_payment_DiD ~ pooled_3_5))
```

```
##
##  Welch Two Sample t-test
##
## data:  missed_payment_DiD by pooled_3_5
## t = 0.8602, df = 8476, p-value = 0.3897
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.00981  0.02515
## sample estimates:
## mean in group 0 mean in group 1
##         0.02565         0.01798
```

```
# Web bill request
with(fieldex, t.test(web_bill_DiD ~ pooled_3_5))
```

```
##
##  Welch Two Sample t-test
##
## data:  web_bill_DiD by pooled_3_5
## t = -0.4795, df = 8609, p-value = 0.6316
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -0.01854  0.01125
## sample estimates:
## mean in group 0 mean in group 1
##         0.07610         0.07975
```

```
# Payments owed
with(fieldex, t.test(payments_owed_DiD ~ pooled_3_5))
```

```
##
##  Welch Two Sample t-test
##
## data:  payments_owed_DiD by pooled_3_5
## t = -0.2992, df = 8092, p-value = 0.7648
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.05988  0.04402
## sample estimates:
## mean in group 0 mean in group 1
##          0.1864          0.1944
```

```
# Current debt
with(fieldex, t.test(current_debt_DiD ~ pooled_3_5))
```

```
##
##  Welch Two Sample t-test
##
## data:  current_debt_DiD by pooled_3_5
## t = -0.5044, df = 8619, p-value = 0.614
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5261  3108
## sample estimates:
## mean in group 0 mean in group 1
##         -1889.1          -812.4
```

```
# Compliance
with(fieldex, t.test(compliance_DiD ~ pooled_3_5))
```

```
##
##  Welch Two Sample t-test
##
## data:  compliance_DiD by pooled_3_5
## t = 0.503, df = 8600, p-value = 0.615
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01244  0.02102
## sample estimates:
## mean in group 0 mean in group 1
##         0.02272         0.01842
```

TABLE 18. FIELD EXPERIMENT. Social vs individual sanctions. Comparison of effects for good and bad taxpayers. Comparison of treatments 3 vs 5. Test using compliance conditional on significant effects for missed payment, number of payments owed or total debt.

```r
# Missed payments
comp.eff(fieldex[fieldex$type=="eligible",]$missed_payment_DiD,
         fieldex[fieldex$type=="eligible",]$pooled_3_5,
         fieldex[fieldex$type=="noneligible",]$missed_payment_DiD,
         fieldex[fieldex$type=="noneligible",]$pooled_3_5)
```

```
##             DM       SE
## DM 1 -0.004118 0.007833
## DM 2 -0.009563 0.014434

## Diff in effects              SE
##        0.005445        0.016422
```

```r
# Web access
comp.eff(fieldex[fieldex$type=="eligible",]$web_bill_DiD,
         fieldex[fieldex$type=="eligible",]$pooled_3_5,
         fieldex[fieldex$type=="noneligible",]$web_bill_DiD,
         fieldex[fieldex$type=="noneligible",]$pooled_3_5)
```

```
##            DM       SE
## DM 1 0.005127 0.012306
## DM 2 0.002296 0.009526

## Diff in effects              SE
##        0.002831        0.015562
```