Simulate Collusion: Application Examples for Cartel Simulation based on Harrington/Chang (2015)

```
knitr::opts_chunk$set(echo = TRUE, warning = FALSE, tidy=TRUE, tidy.opts =
list(width.cutoff=80))

rm(list = ls())
```

Set Seed for Reproducibility

```
sim_seed <- 123
set.seed(sim_seed)
```

Read in Data. Baseline Model 1 with default parameters.

```
sim_list <- sim_col()
cartels_duration <- get_cartel_duration(sim_list$cartels, sim_list$detection,
sim_list$leniency)
```

Combine data and parameters

```
cartels_duration <- left_join(cartels_duration, sim_list$input_ind, by =
"industry")
```

Simple panel data for time series plot

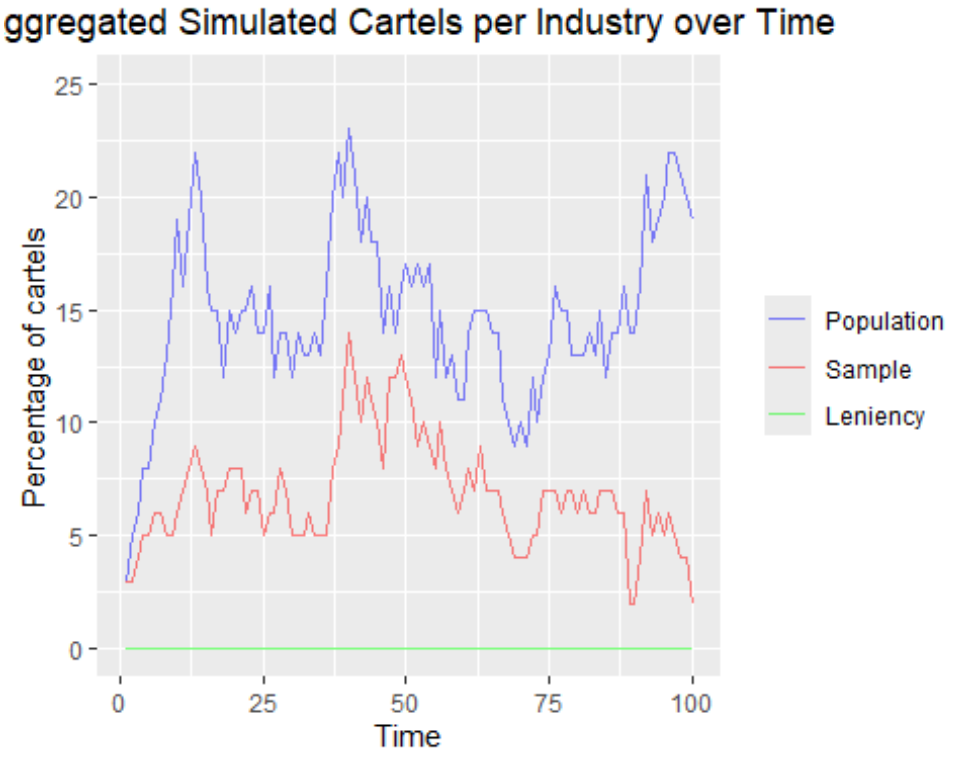
```
num_industries <- nrow(sim_list$input_ind)
periods <- nrow(sim_list$cartels)
cartels_population <- sim_list$cartels
sample_duration <- filter(cartels_duration, detected == 1)
cartels_sample <- get_sample_panel(sample_duration, periods, num_industries)
leniency_duration <- filter(cartels_duration, leniency == 1)
if (nrow(leniency_duration) > 0) {
    cartels_leniency <- get_sample_panel(leniency_duration, periods,
num_industries)
}
```

Plot cartel time series

```
c_det <- rowSums(cartels_sample)/num_industries  # number of detected cartels
/ number of industries
c_pop <- rowSums(cartels_population)/num_industries  # number of cartels /
number of industries
if (nrow(leniency_duration) > 0) {
    c_len <- rowSums(cartels_leniency)/num_industries  # number of cartels /
number of industries
} else {
    c_len <- (rep(0, length(c_pop)))
```

```
}
sim_cartels <- ts(data = cbind(c_pop, c_det, c_len))
colnames(sim_cartels) <- c("Population", "Sample", "Leniency")
pallete <- c("blue", "red", "green")
y_label <- "Percentage of cartels"
x_label <- "Time"

plot <- autoplot(sim_cartels * 100, alpha = 0.5) + ylim(c(0, 25)) +
xlab("Time") +
    ylab("Percentage of cartels") + ggtitle("Aggregated Simulated Cartels per
Industry over Time") +
    scale_colour_manual(values = pallete) + theme(legend.title =
element_blank(),
    plot.title = element_text(hjust = 0.5))
plot
```



Summary Statistics

```
describe(cartels_duration)
```

```
##               vars   n   mean    sd median trimmed   mad    min    max
range
## industry         1 254  49.58 27.97  52.50   49.57 36.32   1.00 100.00
99.00
## start            2 254  51.27 30.19  51.00   51.45 38.55   1.00 100.00
99.00
## end              3 254  56.09 29.90  54.50   56.64 38.55   1.00 100.00
99.00
## duration         4 254   5.81  6.59   3.00    4.59  2.97   1.00  52.00
51.00
```

```
## cartel           5 254   2.62  1.64   2.00    2.40  1.48   1.00   8.00
7.00
## detected         6 254   0.37  0.48   0.00    0.34  0.00   0.00   1.00
1.00
## leniency         7 254   0.00  0.00   0.00    0.00  0.00   0.00   0.00
0.00
## in_sample        8 254   0.37  0.48   0.00    0.34  0.00   0.00   1.00
1.00
## nTc              9 254   1.06  1.07   1.00    0.92  1.48   0.00   6.00
6.00
## rep_off         10 254   0.30  0.46   0.00    0.25  0.00   0.00   1.00
1.00
## delta           11 254   0.85  0.00   0.85    0.85  0.00   0.85   0.85
0.00
## pi_mean         12 254   0.00  0.00   0.00    0.00  0.00   0.00   0.00
0.00
## pi_sd           13 254   1.50  0.00   1.50    1.50  0.00   1.50   1.50
0.00
## MU              14 254   4.08  0.00   4.08    4.08  0.00   4.08   4.08
0.00
## pi_min          15 254   1.00  0.00   1.00    1.00  0.00   1.00   1.00
0.00
## pi_max          16 254 100.00  0.00 100.00  100.00  0.00 100.00 100.00
0.00
## K               17 254   0.06  0.02   0.07    0.06  0.03   0.00   0.10
0.10
## eta             18 254   2.81  1.24   2.57    2.66  1.17   1.13   7.15
6.02
## prob_leniency   19 254   1.00  0.00   1.00    1.00  0.00   1.00   1.00
0.00
##                 skew kurtosis   se
## industry      -0.01    -1.17 1.76
## start         -0.03    -1.27 1.89
## end           -0.08    -1.22 1.88
## duration       3.01    14.03 0.41
## cartel         1.04     0.60 0.10
## detected       0.52    -1.74 0.03
## leniency        NaN      NaN 0.00
## in_sample      0.52    -1.74 0.03
## nTc            1.12     1.69 0.07
## rep_off        0.87    -1.24 0.03
## delta           NaN      NaN 0.00
## pi_mean         NaN      NaN 0.00
## pi_sd           NaN      NaN 0.00
## MU              NaN      NaN 0.00
## pi_min          NaN      NaN 0.00
## pi_max          NaN      NaN 0.00
## K             -0.43    -0.83 0.00
## eta            1.07     0.84 0.08
## prob_leniency   NaN      NaN 0.00
```

Correlations

```
df_cor <- Filter(function(x) sd(x) != 0, cartels_duration)
df_cor <- select(df_cor, -c(industry, cartel))
cor(df_cor)
```

```
##                  start         end    duration    detected   in_sample
## start      1.000000000  0.97598154 -0.15236029 -0.13876544 -0.13876544
## end        0.975981543  1.00000000  0.06660896 -0.10367173 -0.10367173
## duration  -0.152360289  0.06660896  1.00000000  0.16523004  0.16523004
## detected  -0.138765439 -0.10367173  0.16523004  1.00000000  1.00000000
## in_sample -0.138765439 -0.10367173  0.16523004  1.00000000  1.00000000
## nTc        0.361713718  0.38710906  0.09951269  0.46382891  0.46382891
## rep_off    0.389075186  0.40693258  0.06412823  0.27674458  0.27674458
## K         -0.018298593 -0.03606098 -0.07978771 -0.07511223 -0.07511223
## eta       -0.008007306 -0.09194095 -0.38043055 -0.32248174 -0.32248174
##                   nTc     rep_off           K           eta
## start      0.36171372  0.38907519 -0.01829859 -0.008007306
## end        0.38710906  0.40693258 -0.03606098 -0.091940953
## duration   0.09951269  0.06412823 -0.07978771 -0.380430547
## detected   0.46382891  0.27674458 -0.07511223 -0.322481735
## in_sample  0.46382891  0.27674458 -0.07511223 -0.322481735
## nTc        1.00000000  0.82907182  0.08111369 -0.470975299
## rep_off    0.82907182  1.00000000  0.09749461 -0.365633876
## K          0.08111369  0.09749461  1.00000000 -0.030364199
## eta       -0.47097530 -0.36563388 -0.03036420  1.000000000
```

Mean comparison tests for sumstats

```
t.test(duration ~ detected, data = cartels_duration, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  duration by detected
## t = -2.478, df = 157.38, p-value = 0.01427
## alternative hypothesis: true difference in means between group 0 and group
1 is not equal to 0
## 95 percent confidence interval:
##  -4.0365991 -0.4558207
## sample estimates:
## mean in group 0 mean in group 1
##        4.974843        7.221053
```

```
t.test(eta ~ detected, data = cartels_duration, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  eta by detected
## t = 5.7502, df = 234.48, p-value = 2.764e-08
## alternative hypothesis: true difference in means between group 0 and group
1 is not equal to 0
```

```
## 95 percent confidence interval:
##  0.5414683 1.1058835
## sample estimates:
## mean in group 0 mean in group 1
##        3.119798        2.296122
```

```
t.test(K ~ detected, data = cartels_duration, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  K by detected
## t = 1.2372, df = 219.01, p-value = 0.2173
## alternative hypothesis: true difference in means between group 0 and group
1 is not equal to 0
## 95 percent confidence interval:
##  -0.002290924  0.010017599
## sample estimates:
## mean in group 0 mean in group 1
##      0.06439045      0.06052711
```

Linear Regression

```
reg <- lm(log(duration) ~ K + log(eta) + log(nTc + 1), data =
cartels_duration)
summary(reg)
```

```
##
## Call:
## lm(formula = log(duration) ~ K + log(eta) + log(nTc + 1), data =
cartels_duration)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.08797 -0.58846  0.02231  0.61704  2.09432
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.9792     0.2341  12.724  < 2e-16 ***
## K             -4.3904     2.1883  -2.006  0.04590 *
## log(eta)      -1.3042     0.1530  -8.525 1.47e-15 ***
## log(nTc + 1)  -0.3333     0.1263  -2.639  0.00885 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.862 on 250 degrees of freedom
## Multiple R-squared:  0.2494, Adjusted R-squared:  0.2404
## F-statistic: 27.69 on 3 and 250 DF,  p-value: 1.699e-15
```

Further Research: Convert to Paneldata

```
# num_industries <- nrow(sim_list$input_ind) periods <-
nrow(sim_list$cartels)
```

```r
cartels_population <- data.frame(sim_list$cartels)
df_panel <- get_paneldata(cartels_population, cartels_duration, periods,
num_industries)

# Combine df with inputs
df1 <- df_panel %>%
    left_join(sim_list$input_ind, by = c("industry"))
head(df1)
```

```
## # A tibble: 6 × 16
##   industry period in_cartel detected leniency in_sample     R delta
pi_mean
##      <dbl>  <int>     <dbl>    <dbl>    <dbl>     <dbl> <dbl> <dbl>
<dbl>
## 1        1      1         0        0        0         0     0  0.85
0
## 2        1      2         0        0        0         0     0  0.85
0
## 3        1      3         0        0        0         0     0  0.85
0
## 4        1      4         0        0        0         0     0  0.85
0
## 5        1      5         0        0        0         0     0  0.85
0
## 6        1      6         0        0        0         0     0  0.85
0
## # ℹ 7 more variables: pi_sd <dbl>, MU <dbl>, pi_min <dbl>, pi_max <dbl>,
## #   K <dbl>, eta <dbl>, prob_leniency <dbl>
```