BEIS Concentration Data Analysis

- Kyung In Hwang and Anthony Savagar
- Data source: Department for Business, Energy and Industrial Strategy (BEIS) (https://www.gov.uk/government/statistics/business-sectors-indicators-of-concentration-and-churn)

```
# Clear the workspace
rm(list=ls())

dt <- readRDS("dt.rds")
library(tidyverse, quietly = TRUE, warn.conflicts = FALSE)

## -- Attaching packages --------------------------------------------------------------------

## v ggplot2 3.1.0        v purrr   0.3.2
## v tibble  2.1.1        v dplyr   0.8.0.1
## v tidyr   0.8.3        v stringr 1.4.0
## v readr   1.3.1        v forcats 0.4.0

## -- Conflicts ------------------------------------------------------------------------ tidy
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
# set the ggplot2 theme to black and white
theme_set(theme_bw())
library(tikzDevice)

# Set working directory
# setwd("C:/Users/Anthony/Dropbox/BUSINESS/RESEARCH/code/MeasuredSolowResid_IdentifyingTechnology")
```
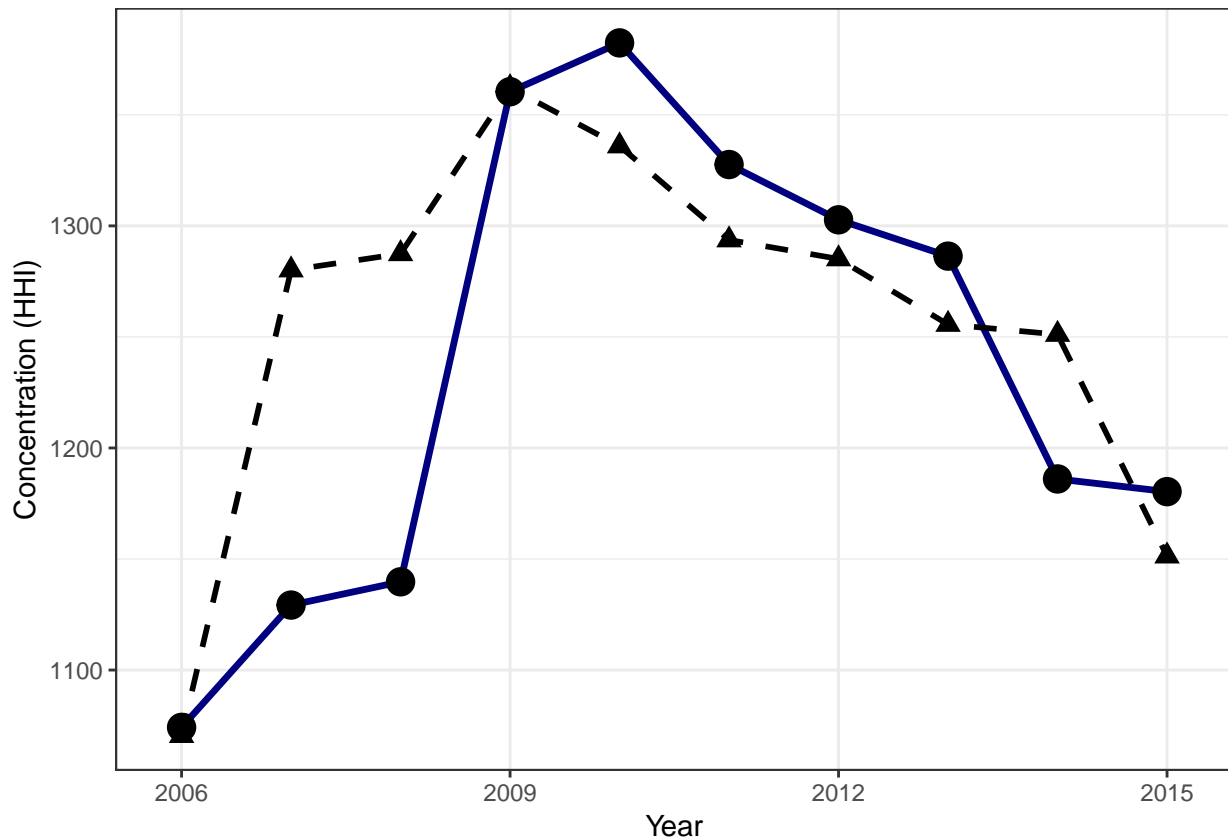
Average Herfindahl index

The figure below demonstrates the evolution of the UK economy's market concentraion from 2006 to 2015. The market concentration measure is the Herfindahl index. Since BEIS data provides sector-level Herfindahl index, I calculate weighted average (solid line, weight: sectoral turnovers) and simple average (dashed line) across sectoral indices. As depited, average Herfindahl index looks like a reverse U-shape. It rises from 2006 to 2010, then decreases after 2010. In spite of decline after 2010, the index (as of 2015) has not returned to the pre-recession level.

```
HHI <- dt %>%
  group_by(year) %>%
  summarise(HHIw = weighted.mean(hhi, turnoverbn, na.rm=TRUE),
            HHIs = mean(hhi, na.rm=TRUE))


HHI_plot <- ggplot(HHI) +
            geom_line(mapping=aes(x=year, y=HHIw), color="navy", size=1.2) +
            geom_point(mapping=aes(x=year, y=HHIw), size=4.5) +
            geom_line(mapping=aes(x=year, y=HHIs), size=1, linetype="dashed") +
            geom_point(mapping=aes(x=year, y=HHIs), size=3, shape=17) +
            scale_x_discrete(name = "Year", limits=c(2006, 2009, 2012, 2015)) +
            scale_y_continuous(name = "Concentration (HHI)") +
            scale_color_manual(values = c("rosybrown", "navy"))
plot(HHI_plot)
```

```
tikz('weight_agg_HHI.tex', width = 3.5, height = 2.5)
  plot(HHI_plot)
dev.off()
```

```
## pdf
##   2
```

Distributions of the Herfindahl index

To get an idea of the evolution of the entire distribution of HI, I draw histograms across three years. They are right-skewed distributions and the middle part and the upper tail have fatten during the period of 2006 to 2010. The shape of 2015 distribution has become similar to 2006. These findings indicate that the increase in average Herfindahl index is not driven by a few sectors. Rather, most sectors have experienced rise in market concentrations from 2006 to 2010.

```
dt %>%
  filter(year == 2006 | year == 2010 |year == 2015) %>%
  ggplot() +
    geom_histogram(aes(x=hhi, group=year), color="navy", fill="white") +
    geom_density(aes(x=hhi, group=year), adjust=2) +
    facet_grid(~year) +
    ylab("Frequency") +
    xlab("Herfindahl Index")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 8 rows containing non-finite values (stat_bin).

## Warning: Removed 8 rows containing non-finite values (stat_density).
```
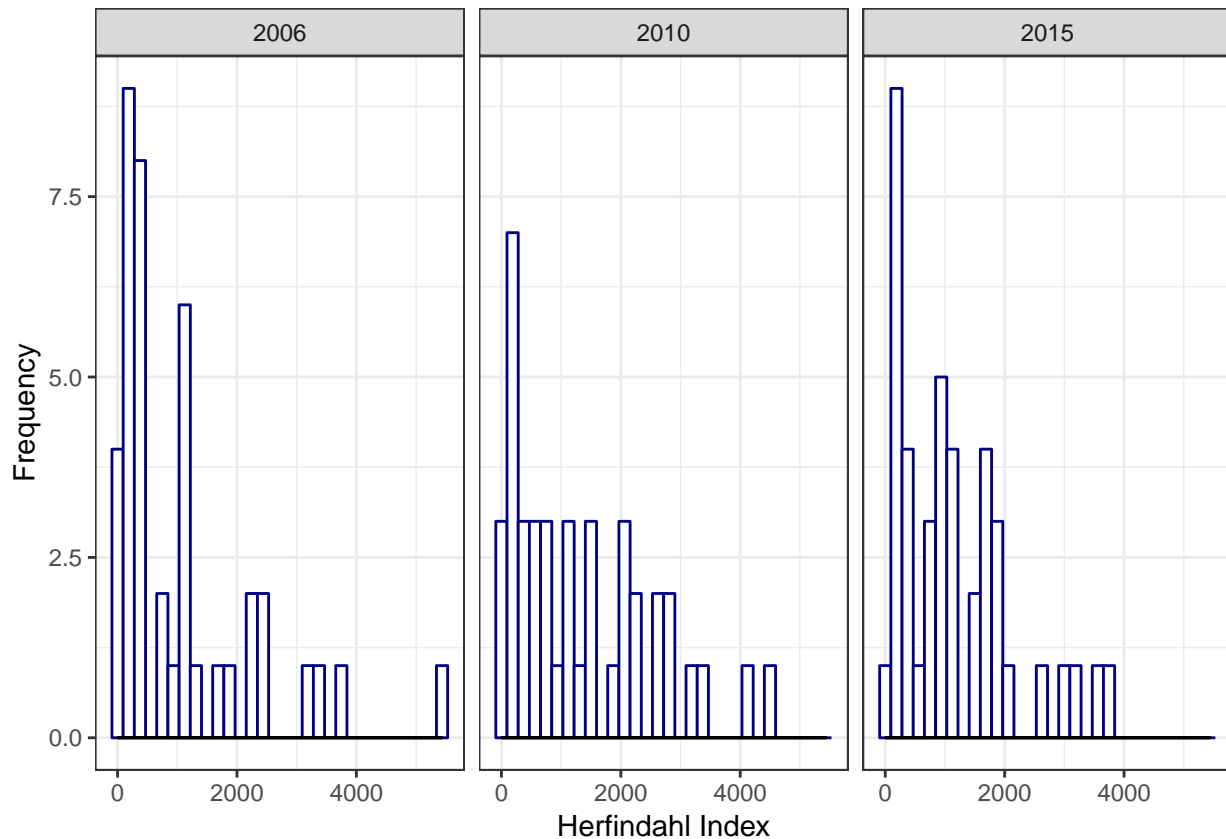
Herfindahl index by 1-digit SIC sectors

I plot time trends in the Herfindahl index by 1-digit sectors. While BEIS data provides information of 44 2-digit SIC sectors, the figures below are drawn by (mean) Herfindahl index calculated at the 1-digit SIC sector level.

The first panel documents that, although level of market concentrations differ across sectors, many of sectors have similar time trends with the overall trend. To see the sectoral comparison about levels of the index, I list 1-digit sectors in the order of the Herfindahl index. The second panel shows that Manufacturing and wholesale/retail sectors mark fourth and fifth largest indices, respectively.

```
dt %>%
  filter(sic1name != "NA") %>%
  group_by(sic1name, year) %>%
  summarise(sic1hhi = weighted.mean(hhi, turnoverbn, na.rm=TRUE)) %>%
  ggplot() +
    geom_line(mapping= aes(x=year, y=sic1hhi), linetype="twodash",
              size=1.2, color = "red") +
    facet_wrap(~sic1name) +
    scale_x_discrete(name = "year", limits=c(2006, 2009, 2012, 2015)) +
    scale_y_continuous(name = "Herfindahl Index")
```
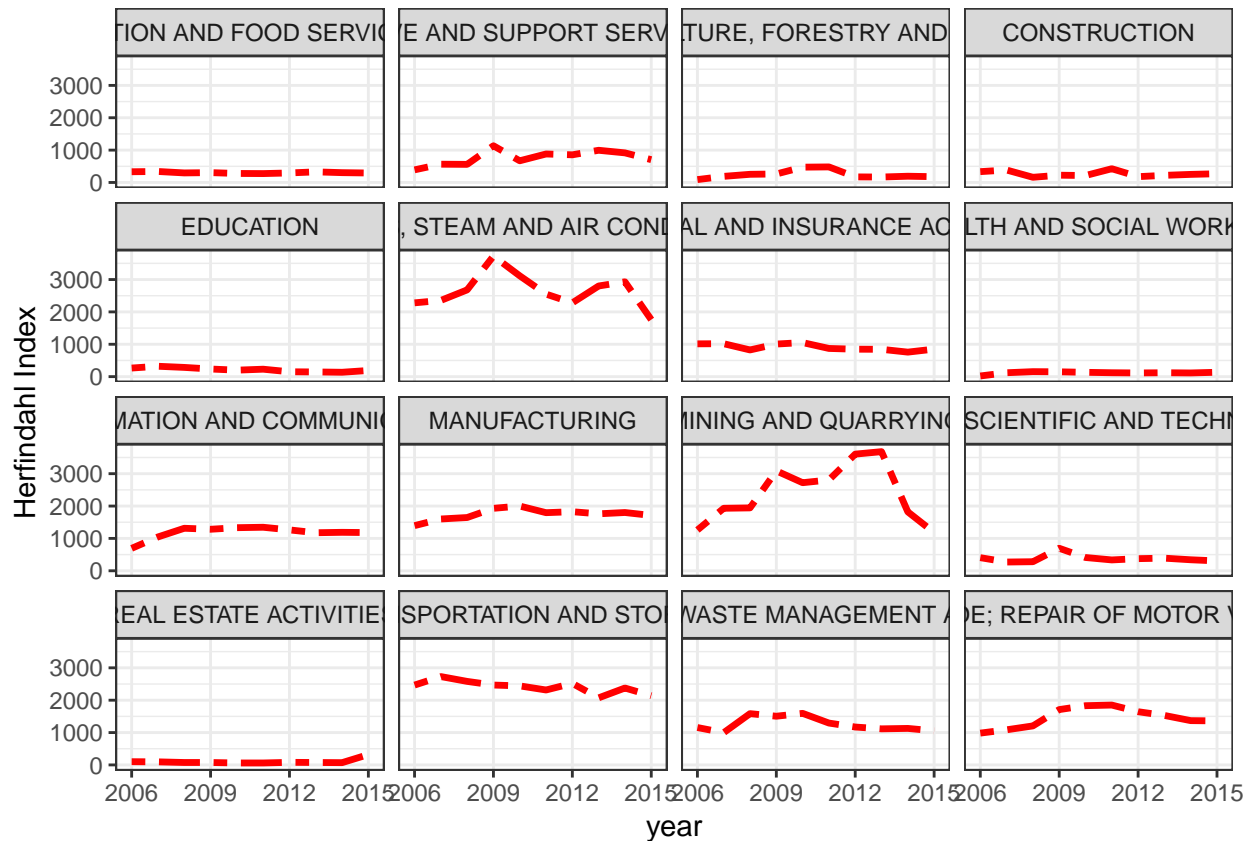
```
dt %>%
  filter(sic1name != "NA") %>%
  group_by(sic1name) %>%
  summarise(sic1hhi = weighted.mean(hhi, turnoverbn, na.rm=TRUE)) %>%
  arrange(sic1hhi) %>%
  mutate(sic1name=factor(sic1name,sic1name)) %>%
  ggplot( aes(x=sic1name, y=sic1hhi)) +
  geom_segment( aes(x=sic1name, xend=sic1name, y=0, yend=sic1hhi,
                    size = I(10)), color="skyblue", size=2) +
  geom_point( color="blue", size=4, alpha=0.6) +
  theme_light() +
  coord_flip() +
  theme(
   panel.grid.major.y = element_blank(),
   panel.border = element_blank(),
   axis.ticks.y = element_blank()
    ) +
  xlab("") +
  ylab("Average Herfindahl Index")
```

ELECTRICITY, GAS, STEAM AND AIR CONDITIONING SUPPLY
MINING AND QUARRYING
TRANSPORTATION AND STORAGE
MANUFACTURING
WHOLESALE AND RETAIL TRADE; REPAIR OF MOTOR VEHICLES AND MOTORCYCLES
WATER SUPPLY; SEWERAGE, WASTE MANAGEMENT AND REMEDIATION ACTIVITIES
INFORMATION AND COMMUNICATION
FINANCIAL AND INSURANCE ACTIVITIES
ADMINISTRATIVE AND SUPPORT SERVICE ACTIVITIES
PROFESSIONAL, SCIENTIFIC AND TECHNICAL ACTIVITIES
ACCOMMODATION AND FOOD SERVICE ACTIVITIES
CONSTRUCTION
AGRICULTURE, FORESTRY AND FISHING
EDUCATION
HUMAN HEALTH AND SOCIAL WORK ACTIVITIES
REAL ESTATE ACTIVITIES

0   1000 2000
Average Herfindahl In

Other measures: CR5, CR10, CR15

CR5 (market shares of 5 largest firms), CR10 and CR15 seem to have similar pattern with the Herfindahl index. However, they look slightly more stable over years than the Herfindahl index. The biggest five firms, on average, account for 30 to 35 %, and market shares of the largest ten firms are 40-45%.

```
dt %>%
  select(sector, year, sic1, top05ms, top10ms, top15ms) %>%
  mutate(CR5 = top05ms, CR10 = top10ms, CR15 = top15ms) %>%
  select(sector, year, sic1, CR5, CR10, CR15) %>%
  gather(key=CR, value = value, c(CR5, CR10, CR15)) %>%
  group_by(year, CR) %>%
  summarise(value = mean(value, na.rm=TRUE)) %>%
  ggplot(aes(x=year, y=value, fill=CR)) +
  geom_bar(stat="identity", position=position_dodge()) +
   scale_fill_brewer(palette="Paired") +
   scale_x_discrete(name = "Year",
                 limits=c(2006, 2007, 2008, 2009, 2010,
                          2011, 2012, 2013, 2014, 2015)) +
   scale_y_continuous(name = "Market Share(%)")
```
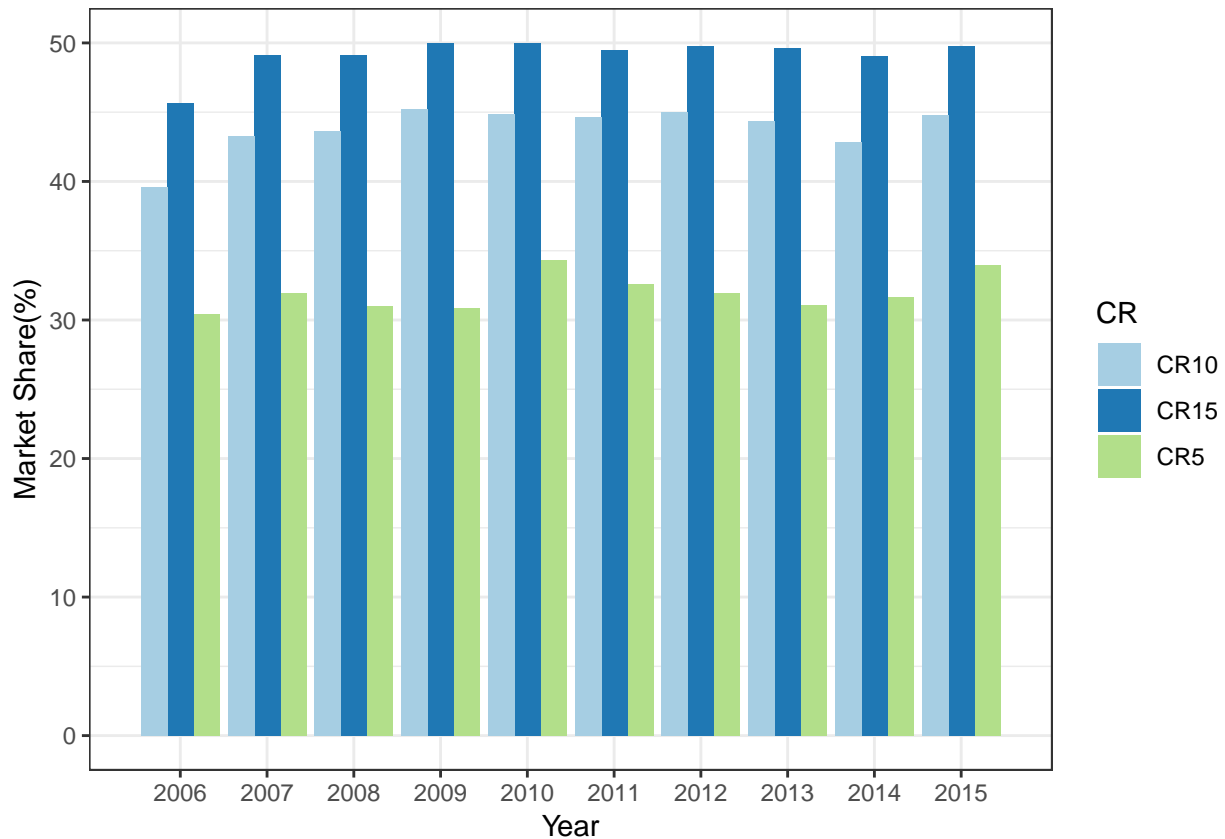
A closer look at three 1-digit SCI sectors

I focus on manufacturing, retail/wholesale and agricultural sectors with relatively higher turnovers in the UK. The figure below reveals that manufacturing, retail/wholesale and agricultural sectors have similar time trends with the overall trend: (1) their Herfindahl indices start to rise, peak in 2010 (or 2011), and decrease up to 2015. (2) In spite of decline after 2010 (or 2011), they have not returned to the pre-recession level.

```
dt %>%
  filter(sic1 == "C") %>%
  ggplot() +
  geom_line(mapping=aes(x=year, y=hhi), color="darkred") +
  geom_point(mapping=aes(x=year, y=hhi), size=2) +
  scale_x_discrete(name = "year", limits=c(2006, 2009, 2012, 2015)) +
  scale_y_continuous(name = "Manufacturing HHI")
```
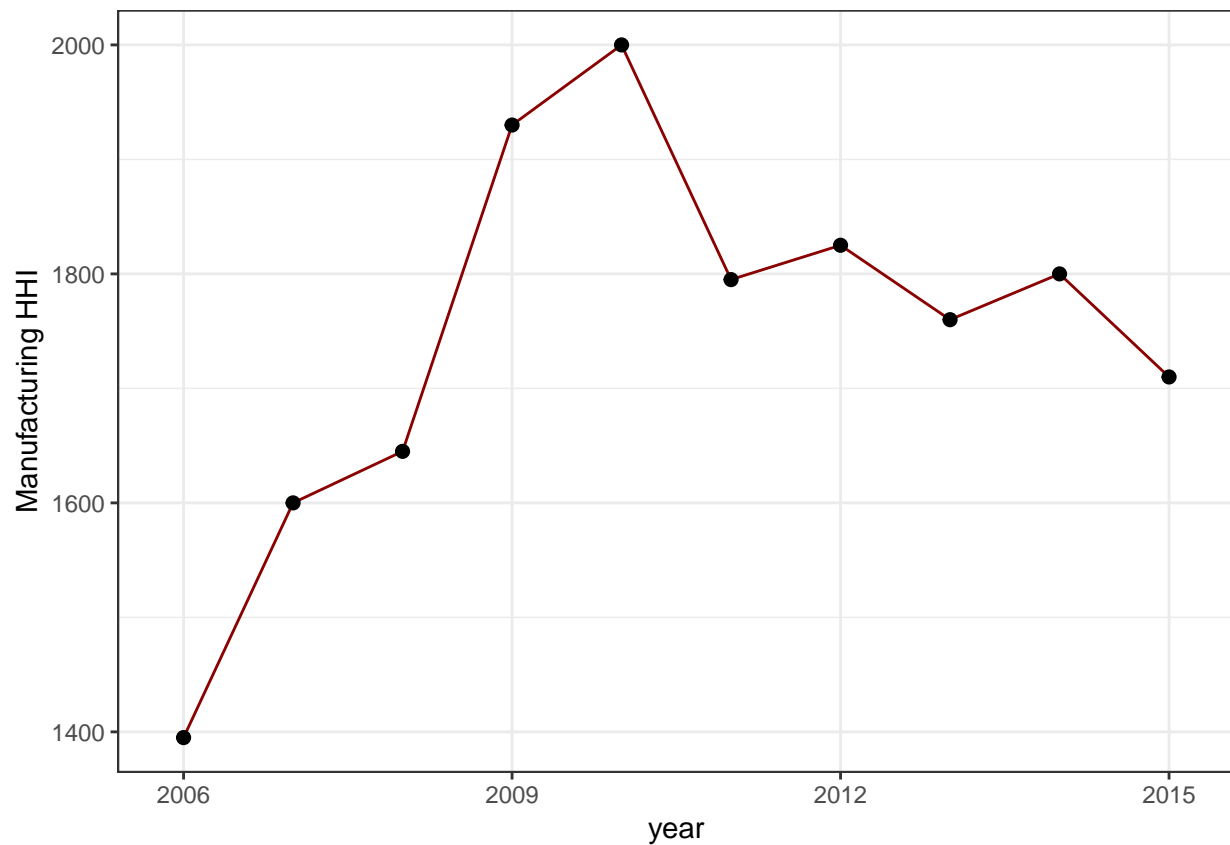
```
dt %>%
  filter(sic1name != "NA") %>%
  group_by(sic1, year) %>%
  summarise(sic1hhi = weighted.mean(hhi, turnoverbn, na.rm=TRUE)) %>%
  filter(sic1 == "G") %>%
  ggplot() +
  geom_line(mapping=aes(x=year, y=sic1hhi), color="darkblue") +
  geom_point(mapping=aes(x=year, y=sic1hhi), size=2) +
  scale_x_discrete(name = "year", limits=c(2006, 2009, 2012, 2015)) +
  scale_y_continuous(name = "Retail/wholesale HHI")
```
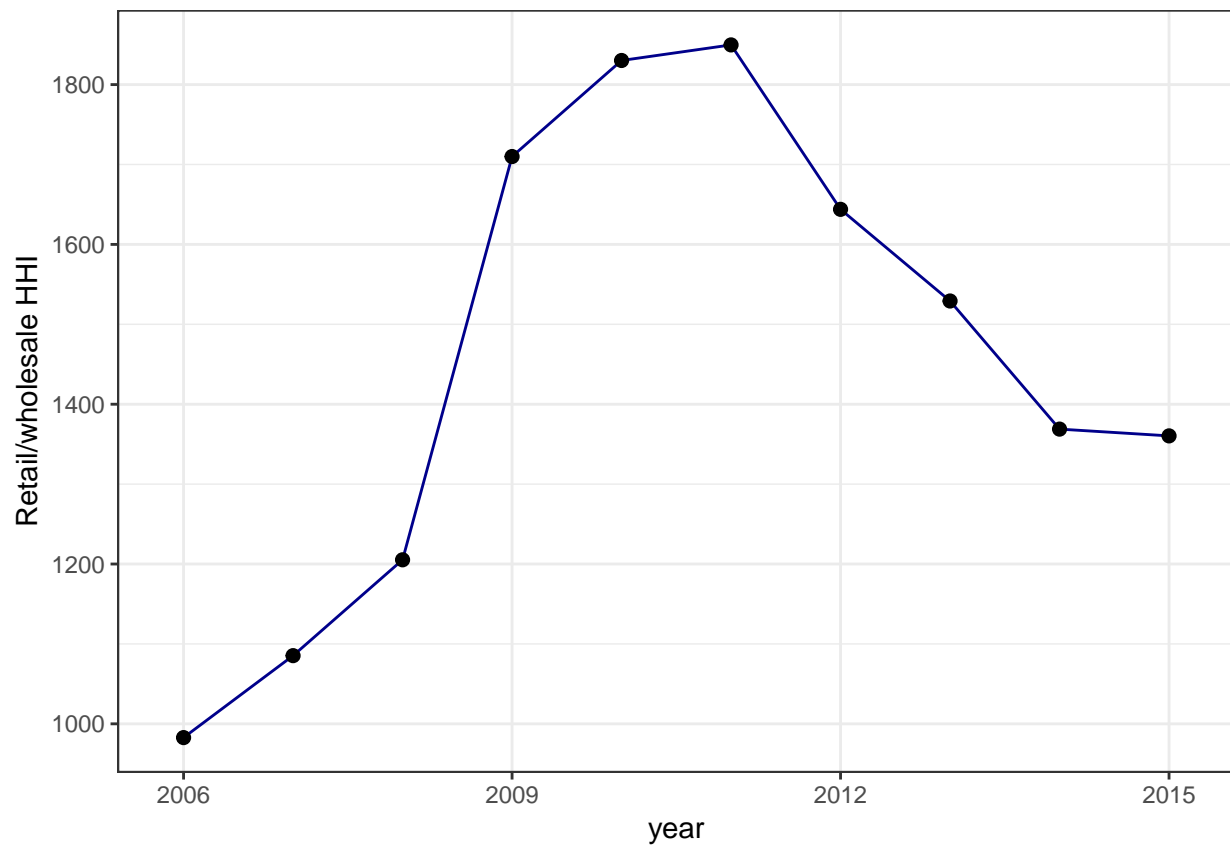
```
dt %>%
  filter(sic1name != "NA") %>%
  group_by(sic1, year) %>%
  summarise(sic1hhi = weighted.mean(hhi, turnoverbn, na.rm=TRUE)) %>%
  filter(sic1 == "A") %>%
  ggplot() +
  geom_line(mapping=aes(x=year, y=sic1hhi), color="darkgreen") +
  geom_point(mapping=aes(x=year, y=sic1hhi), size=2) +
  scale_x_discrete(name = "year", limits=c(2006, 2009, 2012, 2015)) +
  scale_y_continuous(name = "Agricultural HHI")
```
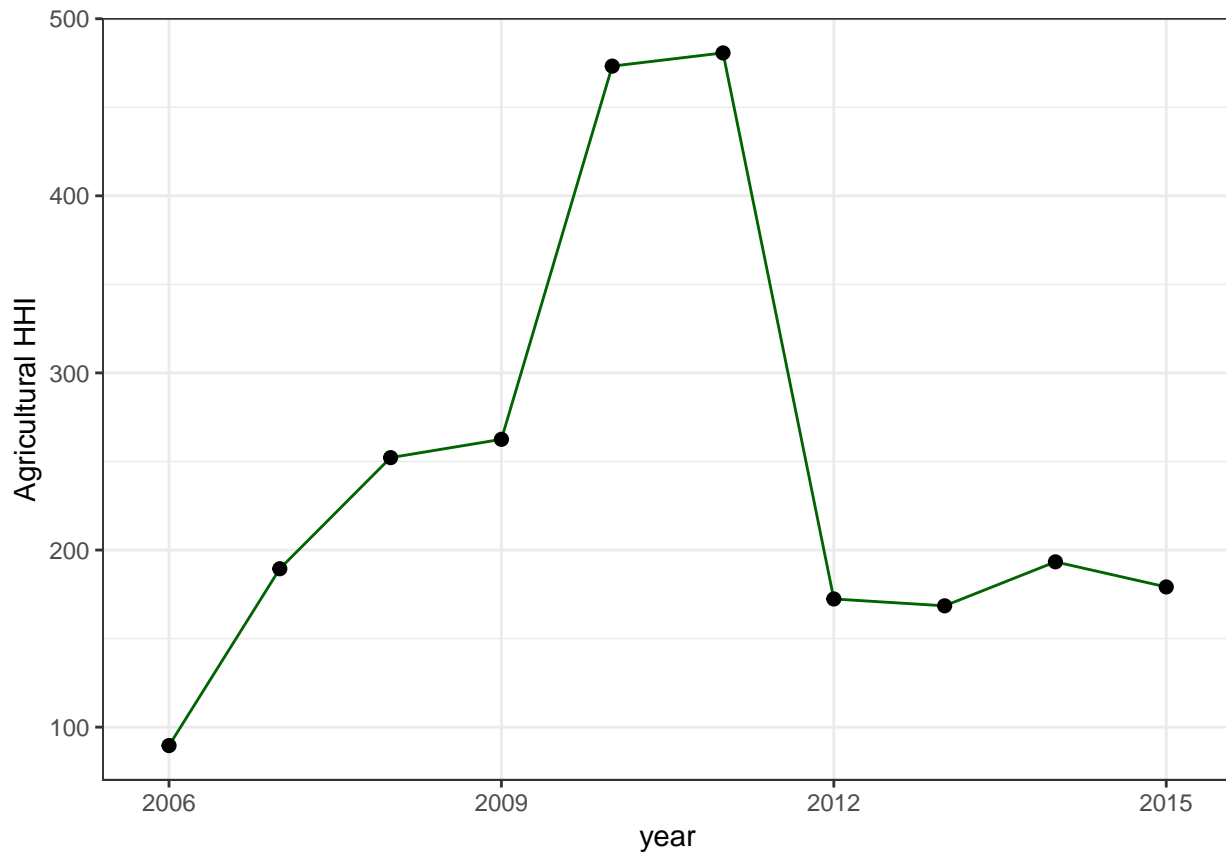
Trends: Herfindahl index vs. Net Entry Rate

Net entry rate, which is defined as $\frac{\text{Entry nbr} - \text{Exit nbr}}{\text{Active firm nbr}} \times 100$, has an opposite trend compared to the Herfindahl index: it decreases from 10% to -5% during th period of 2007 to 2012 whereas increases to 6% in 2014. The following figures illustrate the evolutions of entry rates and exit rates, respectively. According to the figures, the number of firm entries is rapidly diminishing from 2007 to 2009. While the number of firm exits is also decreasing during the same period, the (negative) slope of entry is much bigger than the slope of exit during the period of 2007 to 2009.

```r
dt %>%
  filter(year>2006 & year<2015) %>%
  group_by(year) %>%
  mutate(netentryratio = ((entrynbr-exitnbr)/ firmnbr)*100) %>%
  summarise(netentryratio=weighted.mean(netentryratio, turnoverbn, na.rm=TRUE)) %>%
  ggplot() +
  geom_line(mapping=aes(x=year, y=netentryratio), color="darkblue") +
  geom_point(mapping=aes(x=year, y=netentryratio), size=2) +
  scale_x_discrete(name = "Year",
                   limits=c(2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015)) +
  scale_y_continuous(name = "Net Entry Rate")
```
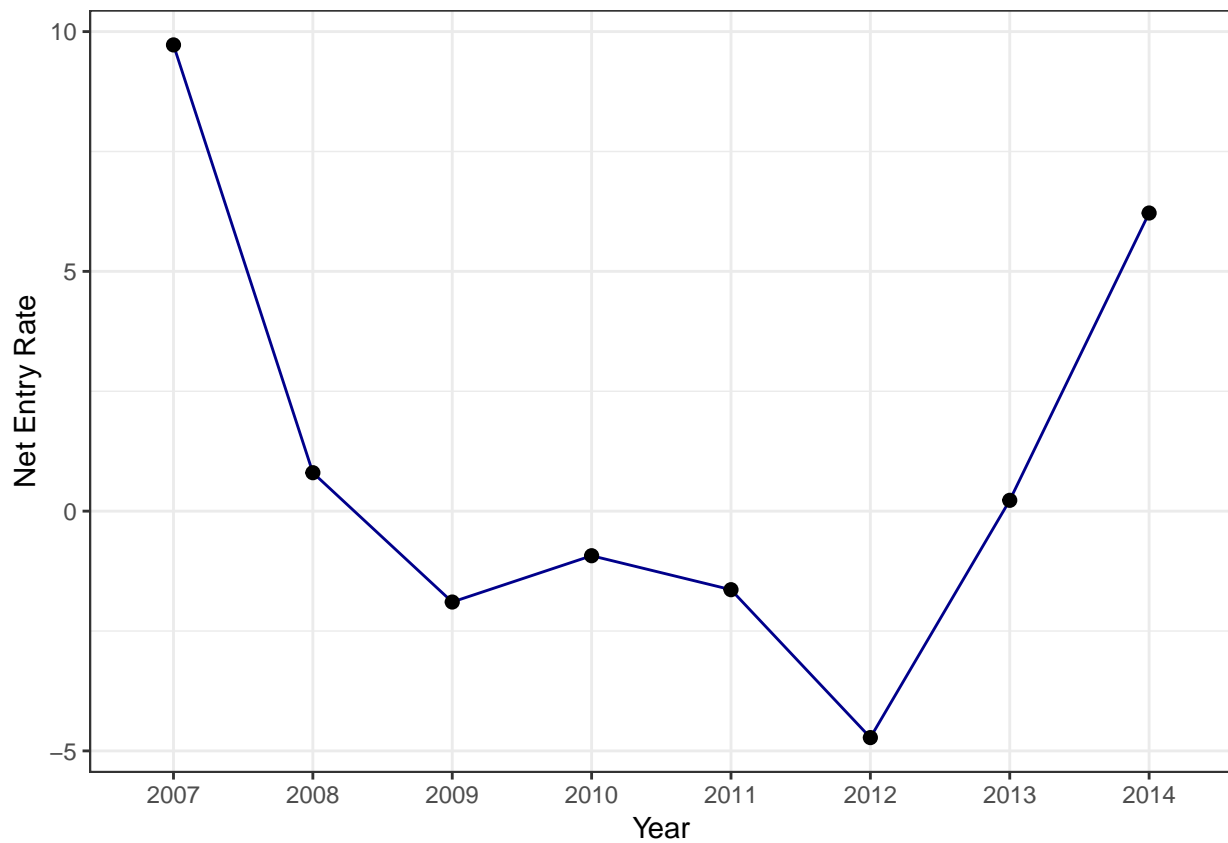
```
dt %>%
  filter(year>2006 & year<2015) %>%
  group_by(year) %>%
  mutate(entryratio = (entrynbr/ firmnbr)*100) %>%
  summarise(entryratio=weighted.mean(entryratio, turnoverbn, na.rm=TRUE)) %>%
  ggplot() +
  geom_line(mapping=aes(x=year, y=entryratio), color="darkgreen") +
  geom_point(mapping=aes(x=year, y=entryratio), size=2) +
  scale_x_discrete(name = "Year",
                limits=c(2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015)) +
  scale_y_continuous(name = "Entry Rate")
```
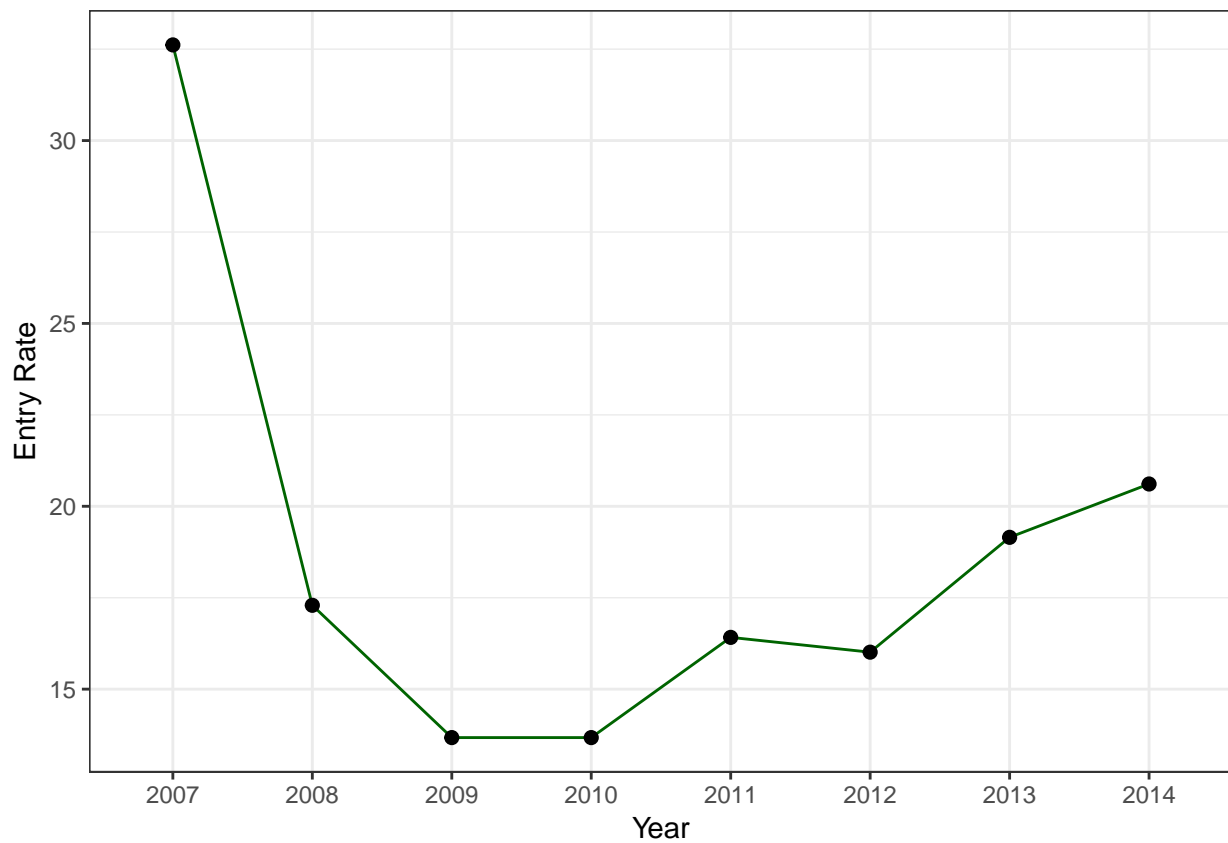
```
dt %>%
  filter(year>2006 & year<2015) %>%
  group_by(year) %>%
  mutate(exitratio = (exitnbr/ firmnbr)*100) %>%
  summarise(exitratio=weighted.mean(exitratio, turnoverbn, na.rm=TRUE)) %>%
  ggplot() +
  geom_line(mapping=aes(x=year, y=exitratio), color="darkgreen") +
  geom_point(mapping=aes(x=year, y=exitratio), size=2) +
  scale_x_discrete(name = "Year",
                   limits=c(2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015)) +
  scale_y_continuous(name = "Exit Rate")
```
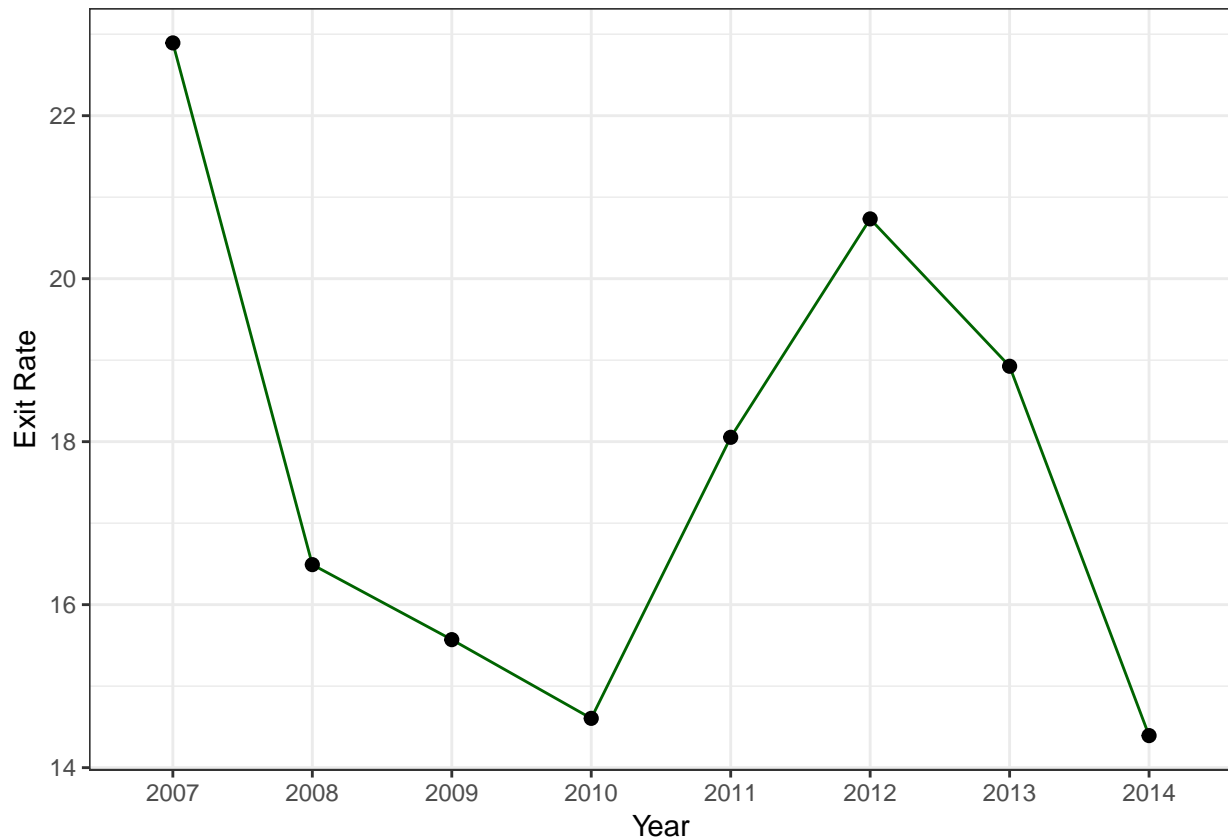
Scatter plot: Herfindahl index vs. Net Entry Rate

To get an idea on the relationship across sectors between the Herfindahl index and Net Entry Rate, I draw a scatter plot between two variables using the pooled cross-section data. The first figure below suggests a positive association. Focuing on only manufacturing, retail/wholesale, agricultural sectors, I further draw a scatter plot between two variables with the subsample in which manufacturing, retail/wholesale, agricultural sectors are included. The following figure indicates that, at least for those three sectors, there is a negative relationship between market cencentration and net entry rate.

```r
dt %>%
  mutate(netentryratio = ((entrynbr-exitnbr)/ firmnbr)*100) %>%
  filter(netentryratio>-20 & netentryratio<50) %>%
  ggplot() +
  geom_point(aes(x=netentryratio, y=hhi), shape=1,  size=3) +
  geom_smooth(aes(x=netentryratio, y=hhi), method='lm', se=FALSE)
```
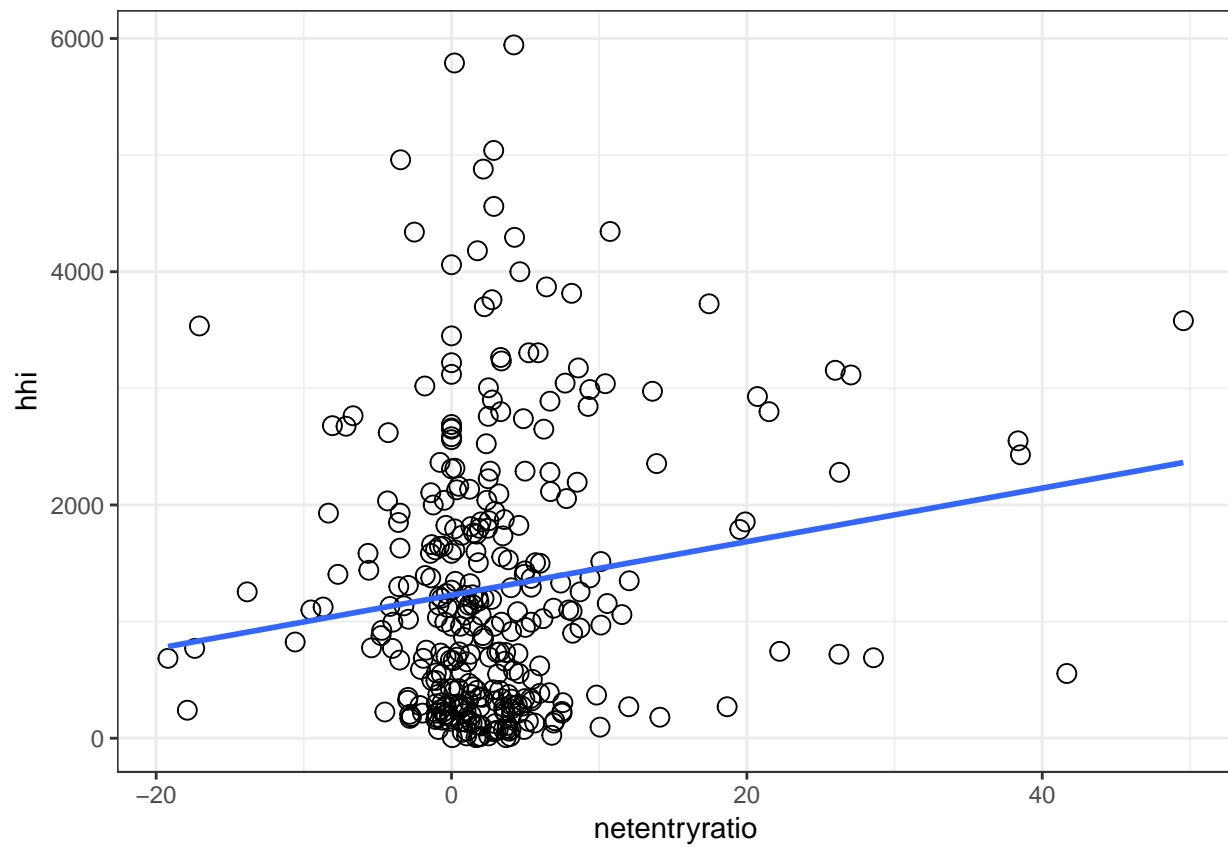
```
dt %>%
  filter(sic1 =="C" | sic1 == "G" | sic1 == "A") %>%
  mutate(netentryratio = ((entrynbr-exitnbr)/ firmnbr)*100) %>%
  filter(netentryratio < 70) %>%
  ggplot() +
  geom_point(aes(x=netentryratio, y=hhi, color=sic1), size=3) +
  geom_smooth(aes(x=netentryratio, y=hhi),  method='lm',  se=FALSE)
```