

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())

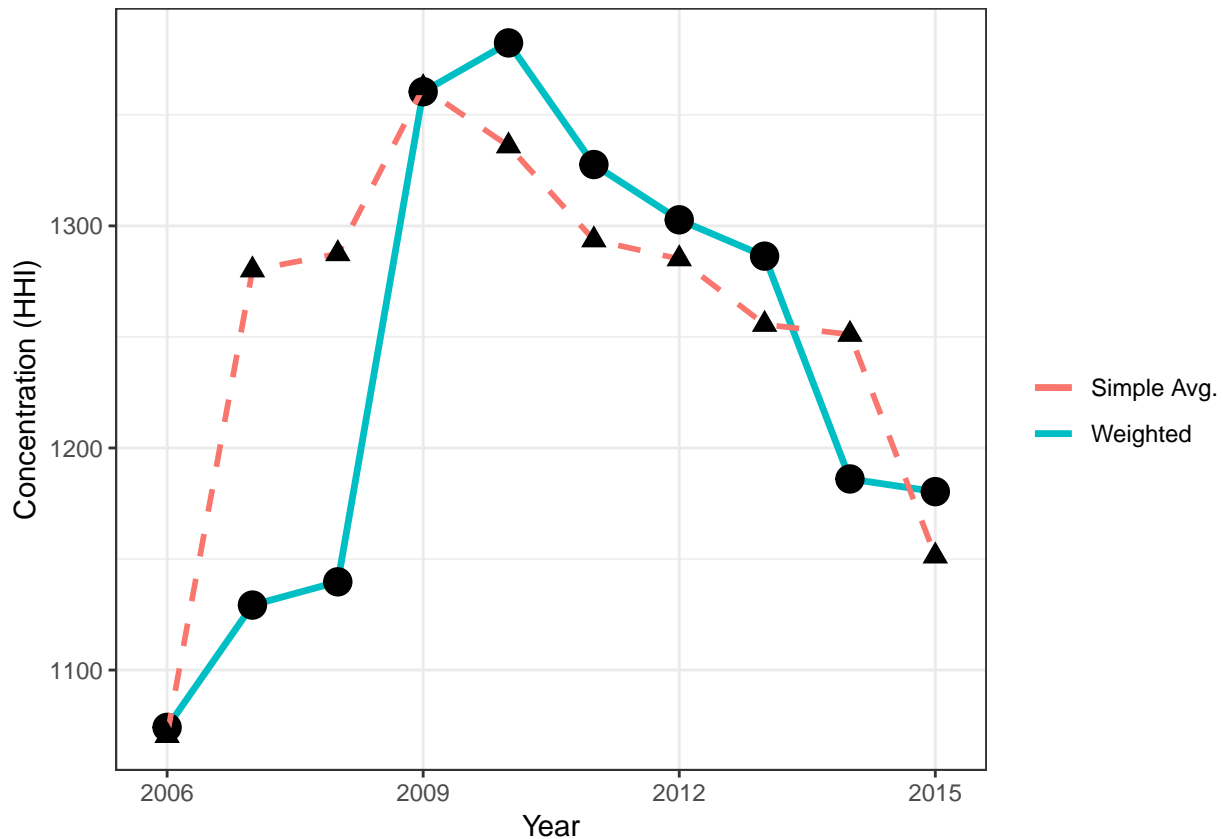
dt <- readRDS("dt.rds")
library(tidyverse, quietly = TRUE, warn.conflicts = FALSE)
theme_set(theme_bw()) # set the ggplot2 theme to black and white
```

Average Herfindahl index

The figure below demonstrates the evolution of the UK economy's market concentration from 2006 to 2015. The market concentration measure is the Herfindahl index. Since BEIS data provides sector-level Herfindahl index, we calculate weighted averages (solid line, weight: sectoral turnovers) and simple averages (dashed line) across sectoral indices. As depicted, average Herfindahl-Hirschman index is hump-shaped. It rises from 2006 to 2010, then decreases after 2010. Despite the decline after 2010, the index (as of 2015) has not returned to the pre-crisis levels, implying a consolidation of market share. The hump-shape is to be expected during a recession as firm exit increases (see below), which raises market share, then as the economy pulls out of recession net entry recovers and share declines. However, the striking feature here is the lack of reversion by 2015: implying structural change and some long-run consolidation.

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

HHI_plot <- ggplot() +
  geom_line(data = HHI, mapping = aes(x = year, y = HHIw, color = "red"), size = 1.2) +
  geom_point(data = HHI, mapping = aes(x = year, y = HHIw), size = 4.5) +
  geom_line(data = HHI, mapping = aes(x = year, y = HHIs, color = "darkblue"), size = 1,
    linetype = "dashed") +
  geom_point(data = HHI, 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_discrete(name = "", labels = c("Simple Avg.", "Weighted"))
plot(HHI_plot)
```



Distributions of the Herfindahl-Hirschman index

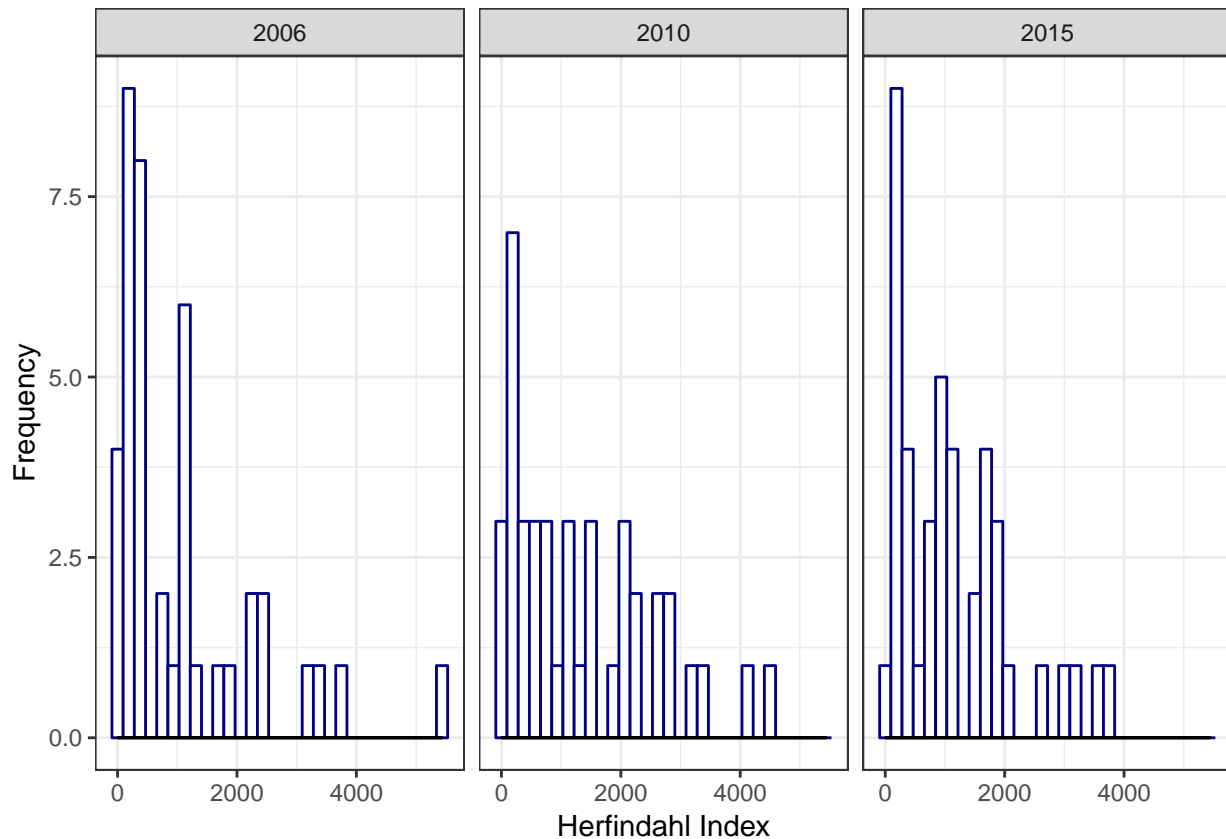
To get an idea of the evolution of the entire distribution of HHI, I draw histograms across three years. They are right-skewed distributions implying most sectors have low concentration. The middle part and the upper tail become thick during the period of 2006 to 2010 indicating a higher concentration in more sectors. By 2015 the shape of the distribution reverts towards 2006, but with a thicker middle representing more sectors with moderate concentration. These findings indicate that the increase in average HHI 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).
```

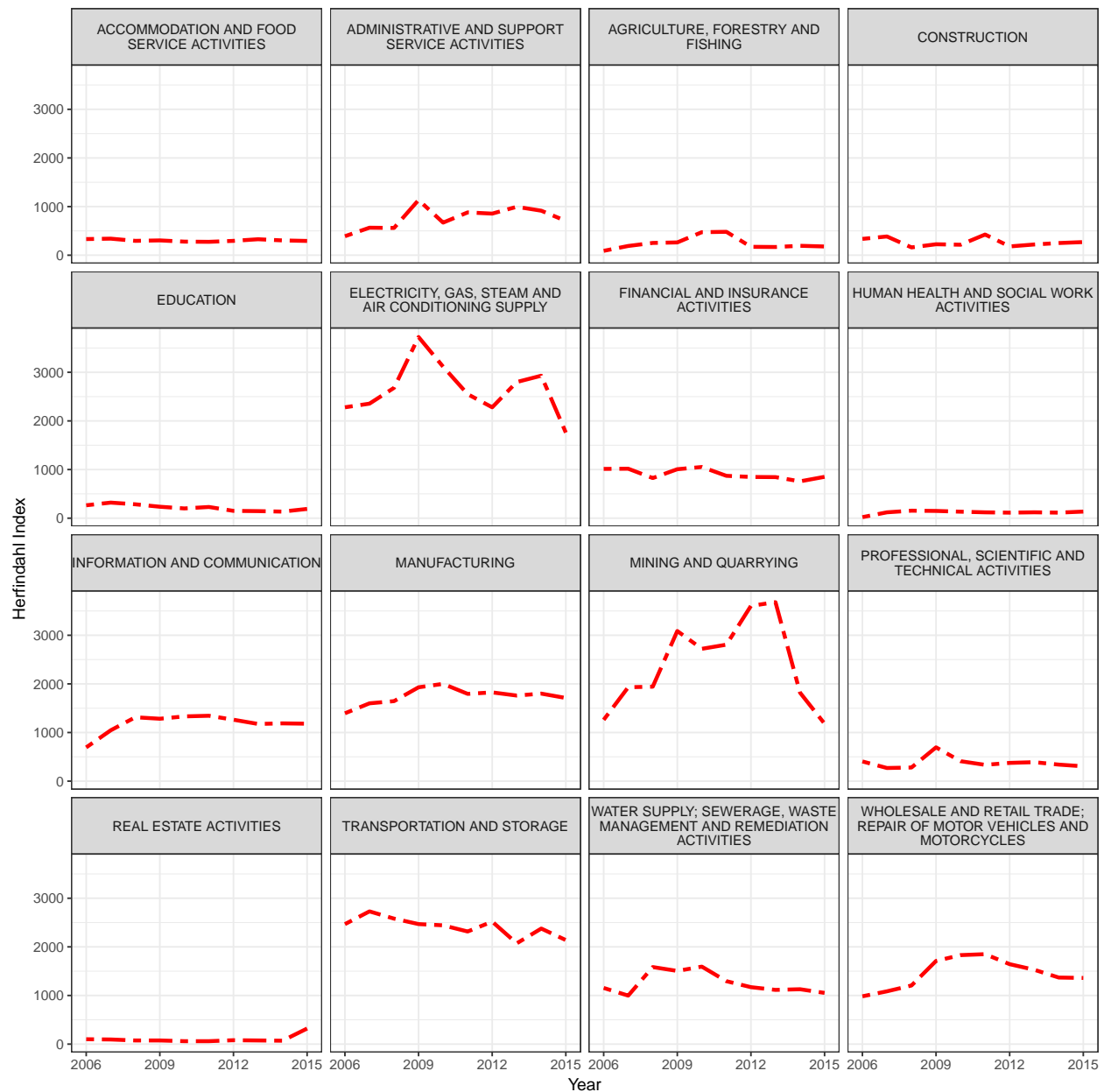


Herfindahl index by 1-digit SIC sectors

We plot time trends in the Herfindahl-Hirschman 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 – a hump-shape.

```
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, ncol=4, labeller = labeller(sic1name = label_wrap_gen(width = 30))) +
  scale_x_discrete(name = "Year", limits = c(2006, 2009, 2012, 2015)) +
  scale_y_continuous(name = "Herfindahl Index")
```

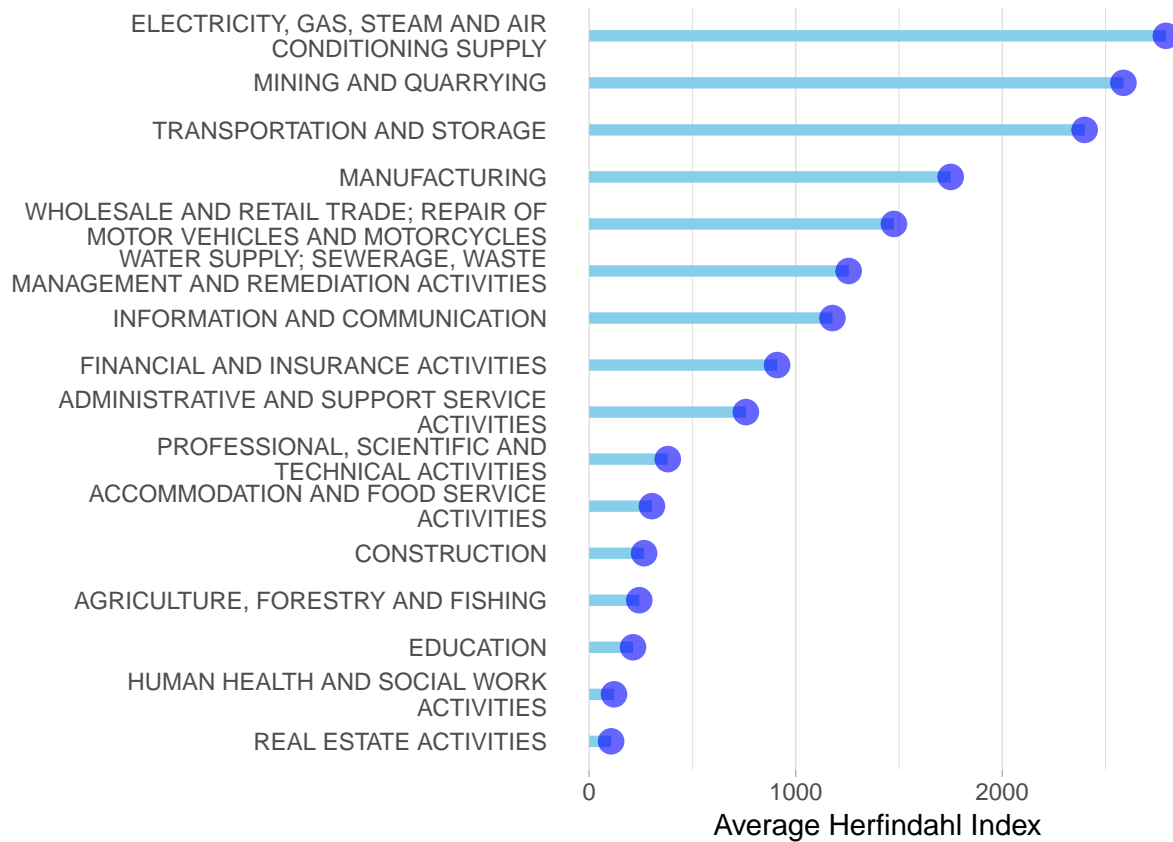


```
ggsave("HHI_by_sector.png", width = 25, height = 25, units = "cm")
```

The second panel ranks sectors by Hirfindahl-Hirschman index. The ranked sectoral comparison shows the average HHI for 1-digit sectors. Manufacturing and wholesale/retail sectors mark fourth and fifth largest indices, respectively.

```
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) +
scale_x_discrete(labels = function(sic1name) str_wrap(sic1name, width = 37)) +
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")
```

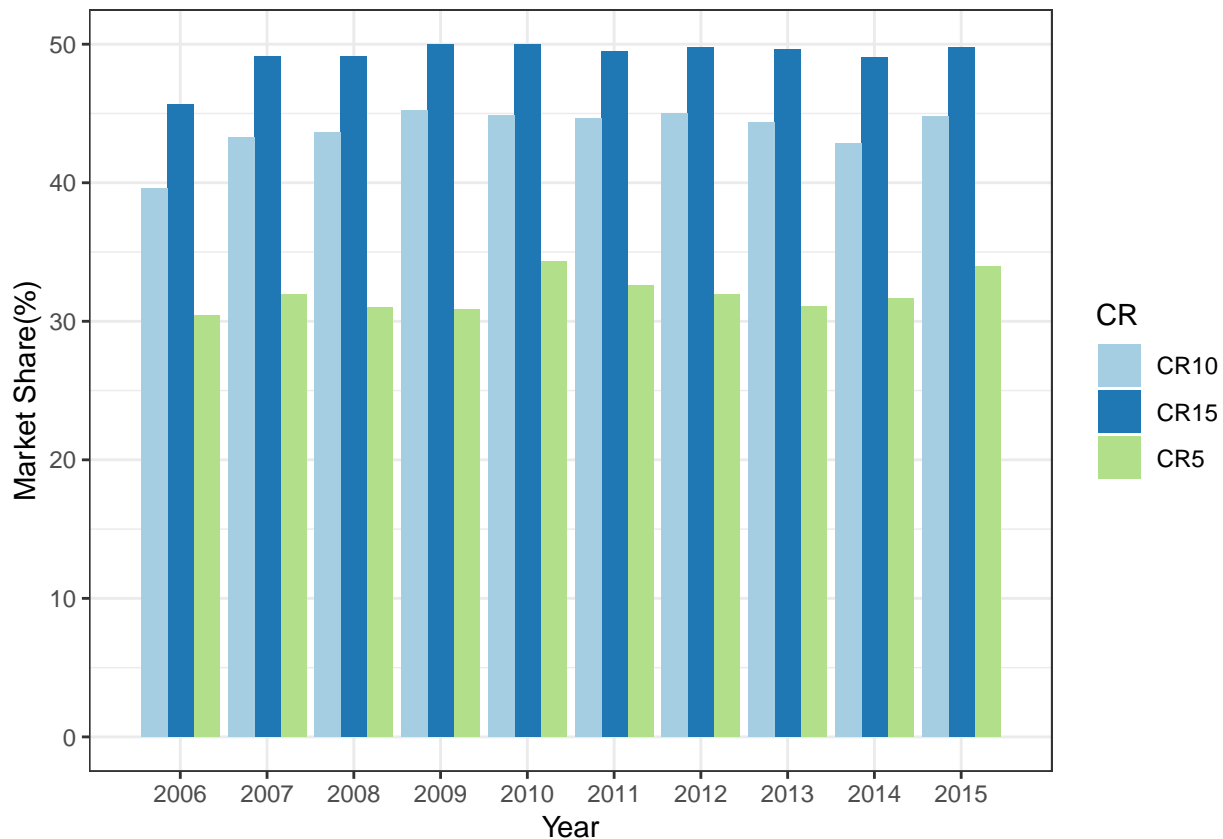


Other measures: Concetration Ratios (CR5, CR10, CR15)

CR5 (market share of 5 largest firms), CR10 and CR15 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)) +
```

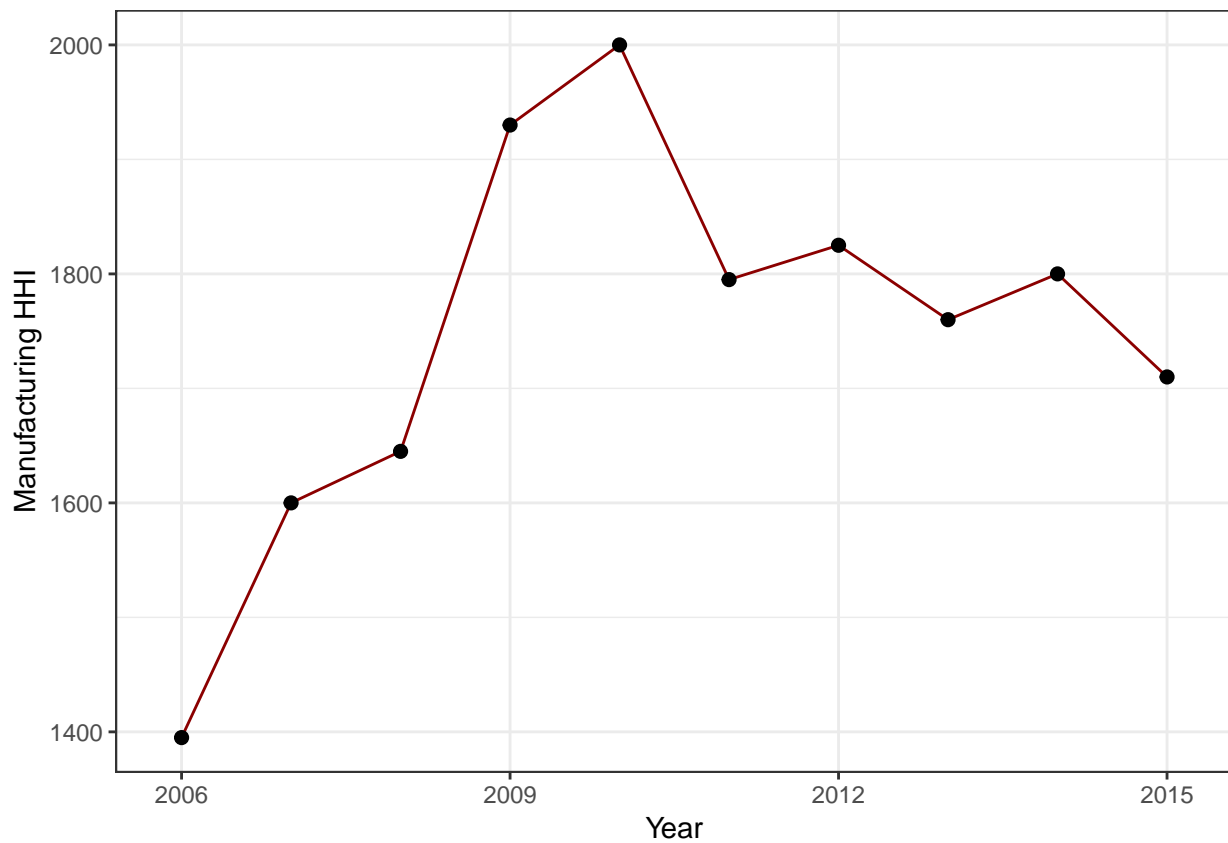
```
geom_bar(stat = "identity", position = position_dodge()) +
scale_fill_brewer(palette = "Paired") +
scale_x_discrete(
  name = "Year",
  limits = c(
    2006, 2007, 2008, 2009, 2010,
    2011, 2012, 2013, 2014, 2015
  )
) +
scale_y_continuous(name = "Market Share(%)" )
```



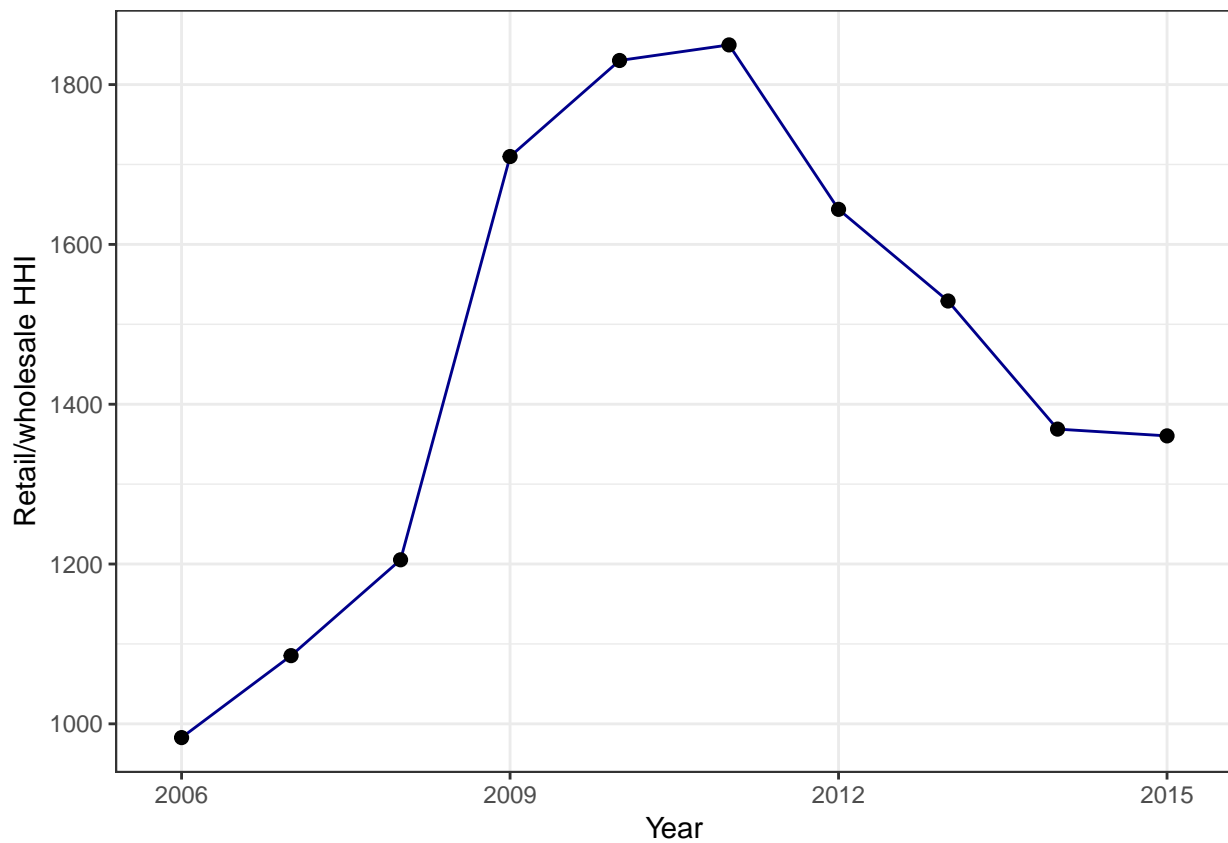
A closer look at three 1-digit SIC sectors

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 decreases up to 2015. (2) In spite of decline after 2010 (or 2011), they have not returned to the pre-recession level.

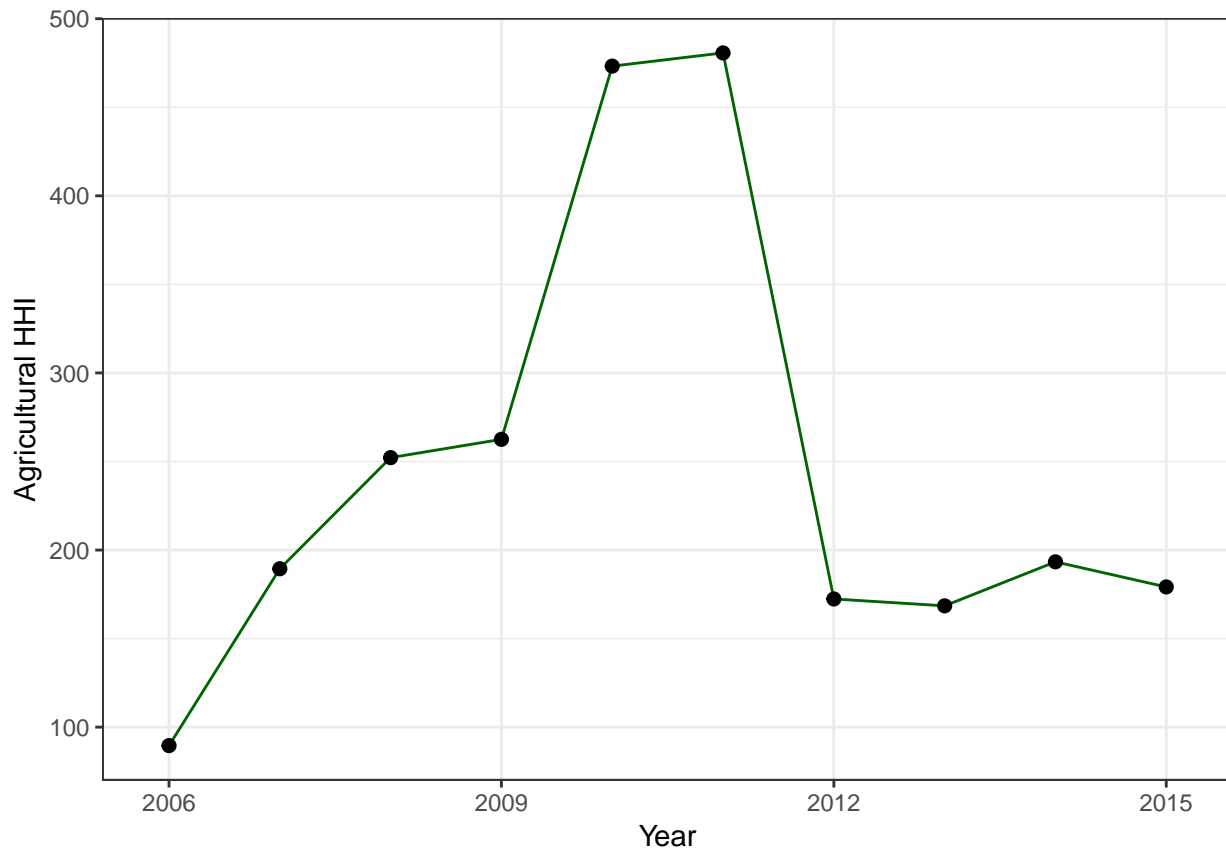
```
dt %>%
  filter(sic1 == "C") %>%
  ggplot() +
  geom_line(mapping = aes(x = year, y = hhi), color = "darkred") +
  geom_point(mapping = aes(x = year, y = hhi), size = 2) +
  scale_x_discrete(name = "Year", limits = c(2006, 2009, 2012, 2015)) +
  scale_y_continuous(name = "Manufacturing HHI")
```



```
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")
```



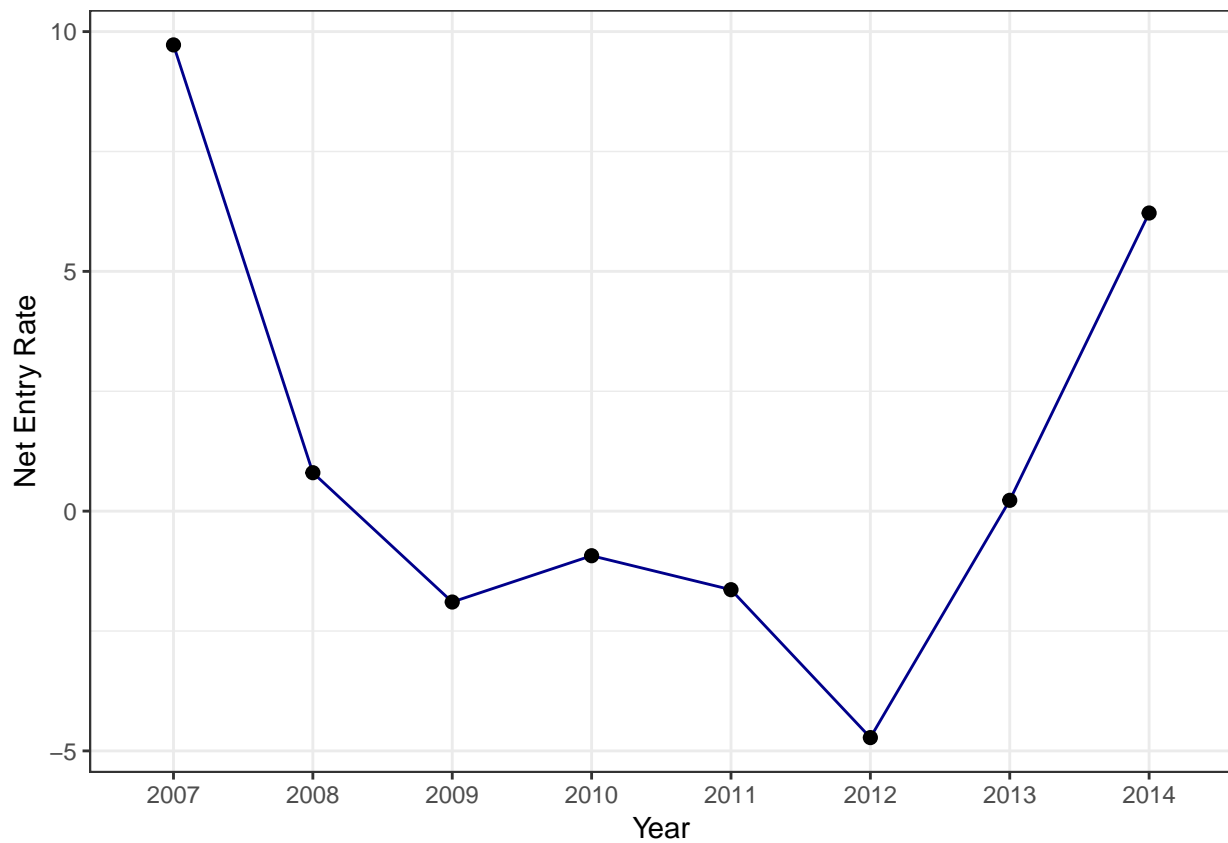
```
dt %>%
  filter(sic1name != "NA") %>%
  group_by(sic1, year) %>%
  summarise(sic1hhi = weighted.mean(hhi, turnoverbn, na.rm = TRUE)) %>%
  filter(sic1 == "A") %>%
  ggplot() +
  geom_line(mapping = aes(x = year, y = sic1hhi), color = "darkgreen") +
  geom_point(mapping = aes(x = year, y = sic1hhi), size = 2) +
  scale_x_discrete(name = "Year", limits = c(2006, 2009, 2012, 2015)) +
  scale_y_continuous(name = "Agricultural HHI")
```

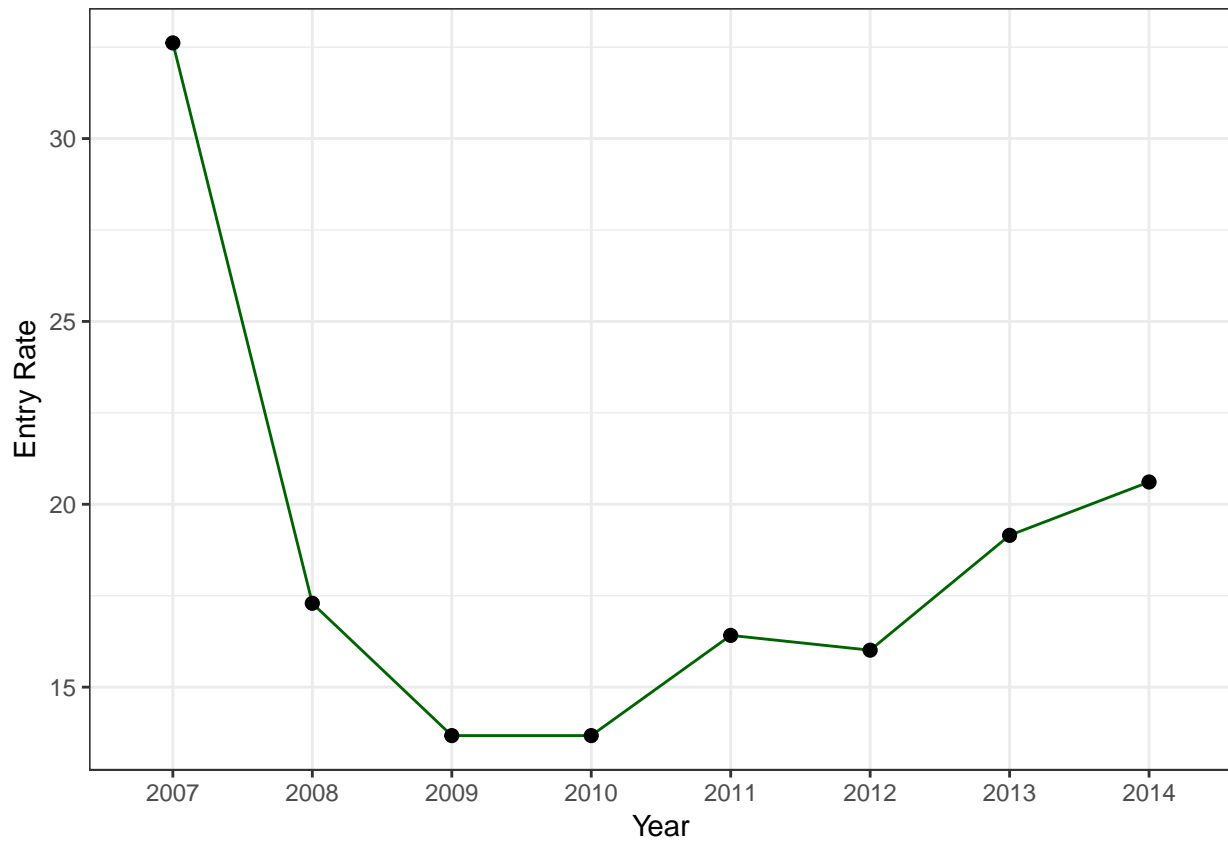
Trends: Herfindahl-Hirschman index vs. Net Entry Rate

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 the period of 2007 to 2012 whereas increases to 6% in 2014. This reflects a weakening of entry with the onset of recession, and subsequent improvement in entry with economic recovery. 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. The potentially puzzling decline in exit is likely attributed to falling entry, removing those startups that are most likely to exit – this counterintuitive dynamic corroborates in Mukoyama and Lee. They emphasize the importance of the firm creation channel over the exit channel.

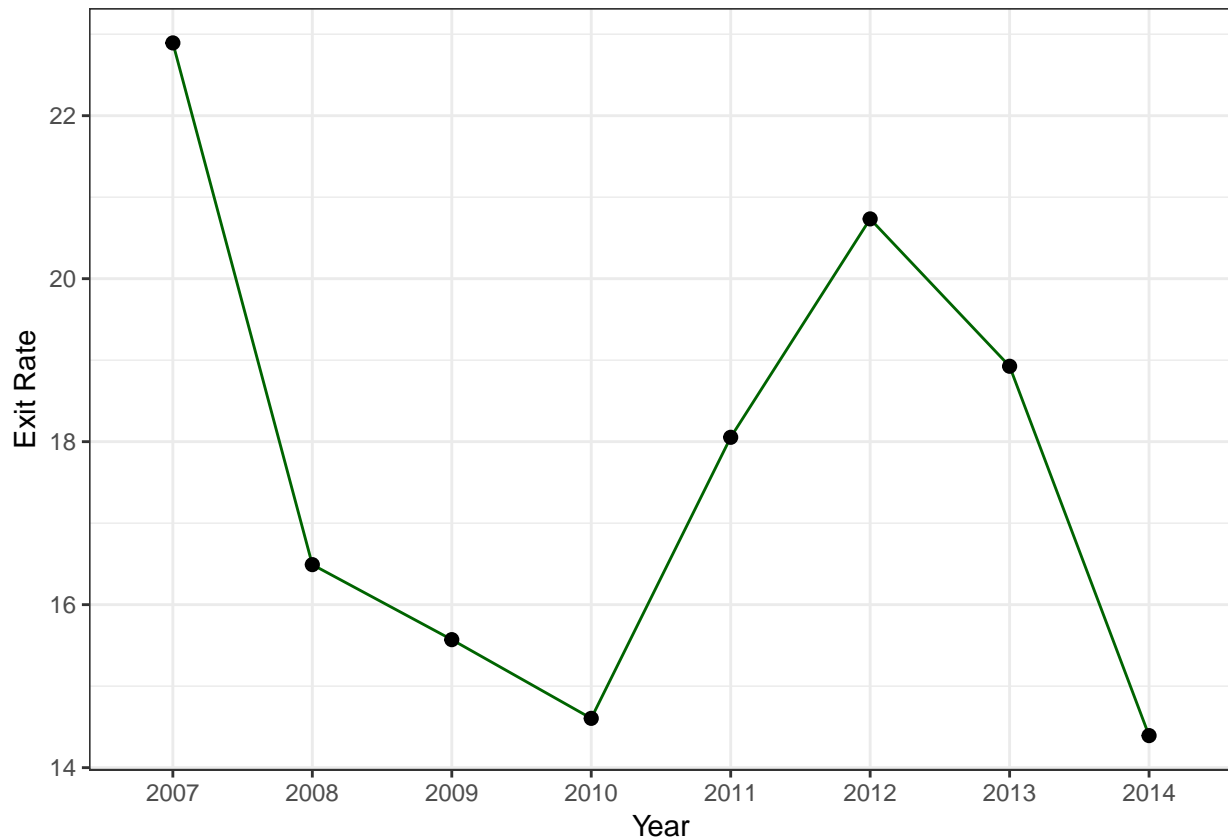
```
dt %>%
  filter(year > 2006 & year < 2015) %>%
  group_by(year) %>%
  mutate(netentryratio = ((entrynbr - exitnbr) / firmnbr) * 100) %>%
  summarise(netentryratio = weighted.mean(netentryratio, turnoverbn, na.rm = TRUE)) %>%
  ggplot() +
  geom_line(mapping = aes(x = year, y = netentryratio), color = "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")
```



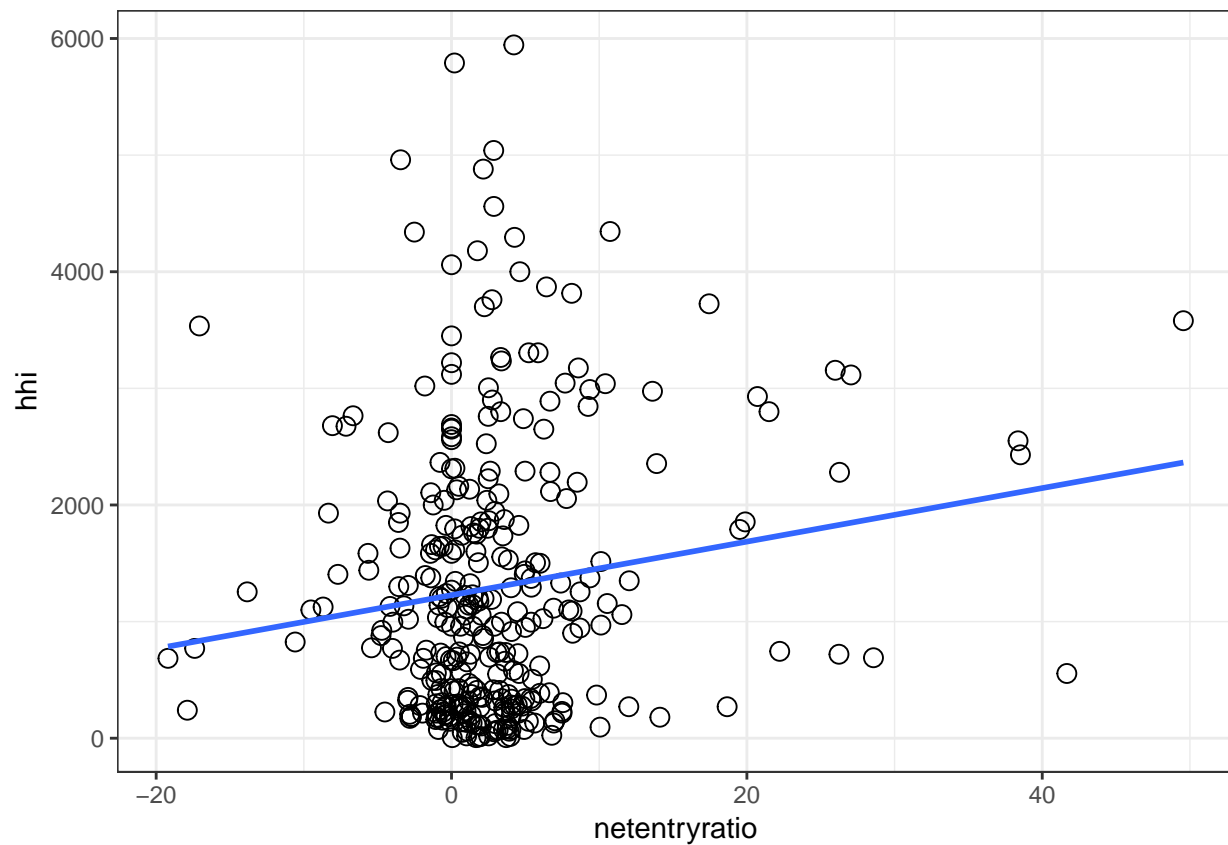
```
dt %>%
  filter(year > 2006 & year < 2015) %>%
  group_by(year) %>%
  mutate(exitratio = (exitnbr / firmnbr) * 100) %>%
  summarise(exitratio = weighted.mean(exitratio, turnoverbn, na.rm = TRUE)) %>%
  ggplot() +
  geom_line(mapping = aes(x = year, y = exitratio), color = "darkgreen") +
  geom_point(mapping = 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-Hirschman 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. Focusing 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 concentration and net entry rate.

```
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)
```

