# concstats: Market Structure, Concentration and Inequality Measures, with R

Andreas Schneider [ORCID] [1]

1 Independent Researcher, Schneider Consulting, Asuncion-Paraguay

Summary

Measures of concentration, competition, and inequality are important key aspects in many disciplines of social sciences. In economics they give a first insight of a given market structure in a particular market. They are important to determine public policies and strategic corporate decisions. However, in research and in practice the most commonly used measure is the Herfindahl Hirschman Index (Herfindahl, 1950) (Hirschman, 1945), calculated as the sum of squared shares retained by individual firms in a given market.

Statement of need

The Herfindahl Hirschman Index, given by $HHI(x) = \sum_{i=1}^{n} \left(\frac{x_i}{\sum x_i} \right)^2$ is still the most widely used measure when it comes to determine concentration or diversity. In economics it usually defines if a given market is concentrated or not. However, relying on one single standard measure may lead to erroneous conclusions and decisions. Various complementary or alternative measures exist, which - used as a set - might reduce uncertainty with respect to a given market situation, and, in consequence, make better and informed decisions. The goal of the concstats package is to offer a set of alternative and/or additional measures for researchers in social sciences and practitioners in institutions concerned with competition on a regular basis to better determine a given market structure and therefore reduce uncertainty with respect to a given market situation. While not limited to this task, various functions or groups of functions are available to achieve the desired goal in a fraction of time. Some of the available functions are scattered in other R packages. The Herfindahl Hirschman Index can be found in hhi (Waggoner, 2018) and the divseg (Kenny, 2022) packages. While the *hhi* package has just one function, neither of both packages offer a normalized version of the measure. The latter offers as well functions for the Entropy, Gini and Simpson measures. The acid (Sohn, 2016) and the ineq (Zeileis & Kleiber, 2014) packages offer functions for inequality and competition measures, e.g. for the Entropy and Gini metric. Some popular measures, e.g. Gini or the Herfindahl Hirschman index have also been implemented in Python.

However, almost none of these packages offer a normalized calculation of the respective measure, with the exception of the ineq package. Other functions are new implementations in R, e.g. Dominance Index, Palma ratio, Stenbacka Index, GRS measure, and the dual of the Herfindahl Hirschman Index. The main advantage for researchers and practitioners of the concstats package is the reach set of available functions to reduce uncertainty in their work.

The package is sustained by a suite of unit tests, continuous integration and maintenance, git version control, and an extensive documentation. The implementation and coding of the package has been guided by statistical software standards of the rOpenSci community to guarantee a user friendly experience.

The R package concstats is designed to filling the above mentioned gap, by offering a

simple-to-use set of 15 individual measures - more or less used by academics or practitioners in a straight forward manner, thereby avoiding dependence on a single metric while reducing uncertainty.

Overview and features

The package offers four different groups with the following functions or functionality:

- **concstats_concstats** calculates eight pre-selected measures in a one-step procedure to provide a first insight of the market.
- **concstats_mstruct** is a wrapper for the following market structure measures: `concstats_firm()` computes the number of firms in a given market with market shares, `concstats_nrs_eq()` (Numbers equivalent), `concstats_top()`, `concstats_top3()` and `concstats_top5()` computes the respective cumulative sums of market shares in a given market, and `concstats_all_mstruct()` calculates all measures as a group.
- **concstats_comp** is a group wrapper for the following competition / concentration measures: `concstats_hhi()` calculates the Herfindahl-Hirschman Index (Herfindahl, 1950; Hirschman, 1945) and `concstats_hhi_d()` its respective dual (Tabak et al., 2009), the `concstats_hhi_min()`, the `concstats_dom()`, the dominance index (Idunate, 1994), the `concstats_sten()` computes the result of the Stenbacka index (Melnik et al., 2008), and finally `concstats_all_comp()` which calculates all measures as a group.
- **concstats_inequ** is a wrapper for inequality and diversity measures and contains: `concstats_entropy()` (Shannon, 1948), `concstats_gini()`, `concstats_simpson()`, the `concstats_palma()` (Palma ratio) (Palma, 2006) and the `concstats_grs()` as an alternative measure proposed by Ginevicius (Ginevičius & Čirba, 2009), `concstats_all_inequ()` calculates the group measures.

For calculating any individual measure users just need to call the function, e.g. concstats_hhi(), its dual concstats_hhi_d(), or any other function of interest, and include the name of the vector with the market shares. The user may also calculate a group of measures, e.g. `concstats_mstruct` (market structure) in a one-step procedure and the results will be printed in a table (see Figure 1). Alternatively the user might just want a quick overview of a given market calculating a predefined set of measures using the `concstats_concstats()` function.

Where it makes sense, measures are bounded [0, 1] to provide a better comparability with respect to other measures. The user will be guided through a number of defensive coding features avoiding erroneous inferences, e.g. a warning message when the vector of shares is not in a decimal format or a stop error message when the specified vector does not sum to 1. These and other features should help the user to reduce uncertainty about a given market situation and finally make better informed decisions in a consistent way and in much shorter time. Last but not least, the package comes with a real world data set of large credit unions in Paraguay.

A practical use case and further details regarding the different measures used in this package can be found in the article *Kreditgenossenschaften: Marktstruktur, Wettbewerb und Verhalten. Das Beispiel Paraguay* (Schneider, 2022), which analyzes the market structure and competition of large credit unions in Paraguay.

Examples

First of all, make sure you have the `concstats` package installed from CRAN or the latest development version from GitHub.

Let us get a quick overview of a given market situation.

```
library(concstats)
# We create some simple data of market shares of each firm in a given market
x <- c(0.35, 0.4, 0.05, 0.1, 0.06, 0.04)
sum(x)
# [1] 1
```

Calculate a selected set of market structure and concentration measures

```
share_con <- concstats_concstats(x) # creates the object share_con
#         Measures Values
# 1           Firms   6.00
# 2 Nrs_equivalent   3.33
# 3         Top (%)  40.00
# 4        Top3 (%)  85.00
# 5        Top5 (%)  96.00
# 6             HHI   0.30
# 7     Entropy(RE)   0.79
# 8     Palma ratio   2.67
```

Now, assume we are just interested in one single measure, e.g. the dominance index.

```
# We use the same vector of market shares, and we will name our object share_dom
x <- c(0.35, 0.4, 0.05, 0.1, 0.06, 0.04)
share_dom <- concstats_dom(x)
share_dom

# [1] 0.4519308

# Alternative way to compute the same measure, and controlling the output
x <- c(0.35, 0.4, 0.05, 0.1, 0.06, 0.04)
share_dom2 <- concstats_comp(x, type = "dom", digits = 2)
share_dom2

# [1] 0.4519308
```

As you can see the functions can be accessed in different ways. Now, let us assume we are interested in a group of measures, e.g. inequality.

```
# Our known vector
x <- c(0.35, 0.4, 0.05, 0.1, 0.06, 0.04)
share_ineq <- concstats_inequ(x, type = "all", digits = 2)
share_ineq

#           Measure Value
# 1         Entropy  2.04
# 2     Gini Index  0.46
# 3 Simpson Index  0.70
# 4     Palma Ratio  2.67
# 5             GRS  0.40
```

Note that for some measures the parameter normalized = FALSE is the default setting.

Visual summaries

The `concstats` package integrates fine with other visualization packages from the R ecosystem, e.g. `kabelExtra` for tables or `ggplot2` for nice looking plots.

This time, we will use our `creditcoops` data set, which comes with the package.

```
data("creditcoops")
head(creditcoops)

# A tibble: 6 x 5
#  coop_id   year   total_loans paired total_loans_log
#    <dbl>   <fct>        <dbl>  <int>           <dbl>
#        1   2016    173892358       1            19.0
#        1   2018    199048199       1            19.1
#        2   2016    323892456       2            19.6
#        2   2018    461609439       2            20.0
#        3   2016    179981404       3            19.0
##       3   2018    227232008       3            19.2
```

You will need the following two packages. Make sure you have these packages installed.

```
library(dplyr)
library(kableExtra)
```

Now, we will filter out data for the year 2016 and create a reusable table.

```
coops_2016 <- creditcoops |> dplyr::filter(year == 2016)
head(coops_2016)

coops_2016 <-  coops_2016[["total_loans"]] # atomic vector of total loans
coops_2016 <- coops_2016/sum(coops_2016)   # convert the vector in decimal
                                           # form of market shares

# We then use the new object `coops_2016` to calculate the market
# structure measures as a group in a one-step-procedure:
coops_2016_mstruct <- concstats_all_mstruct(coops_2016)
coops_2016_mstruct_tab <- coops_2016_mstruct |>
  kableExtra::kbl(caption = "Market structure 2016", digits = 1,
                  booktabs = T, align = "r") |>
  kableExtra::kable_classic(full_width = F, html_font = "Arial")
coops_2016_mstruct_tab
```

Market structure 2016

Measure	Value
Firms	22.0
Nrs_equivalent	8.2
Top (%)	22.1
Top3 (%)	55.0
Top5 (%)	68.8

Figure 1: Market structure summary table

The resulting table is a summary of the available group functions for assessing a given market structure.

Now, let's go a step further. We will make a visual comparison of the two samples for years 2016 and 2018. For this purpose, we will select from our `creditcoops` data set the relevant columns (coop_id, year, paired, and total_loans_log) and make a new data frame.

```
df_shares <- creditcoops |>  dplyr::select(coop_id, year, paired, total_loans_log)
```

Make sure you have the `ggplot2` package installed. Load the package.

```
library(ggplot2) # Create Elegant Data Visualizations Using the Grammar
                 # of Graphics
df_shares_plot <- df_shares |>
  ggplot(aes(year, total_loans_log, fill = year)) +
  geom_boxplot() +
  geom_point() +
  geom_line(aes(group = paired)) +
  labs(title = "Credit cooperatives (type A)", y = "Total loans (log)",
       caption = "Source: Andreas Schneider with data from INCOOP") +
  theme(legend.position = "none")
df_shares_plot
```
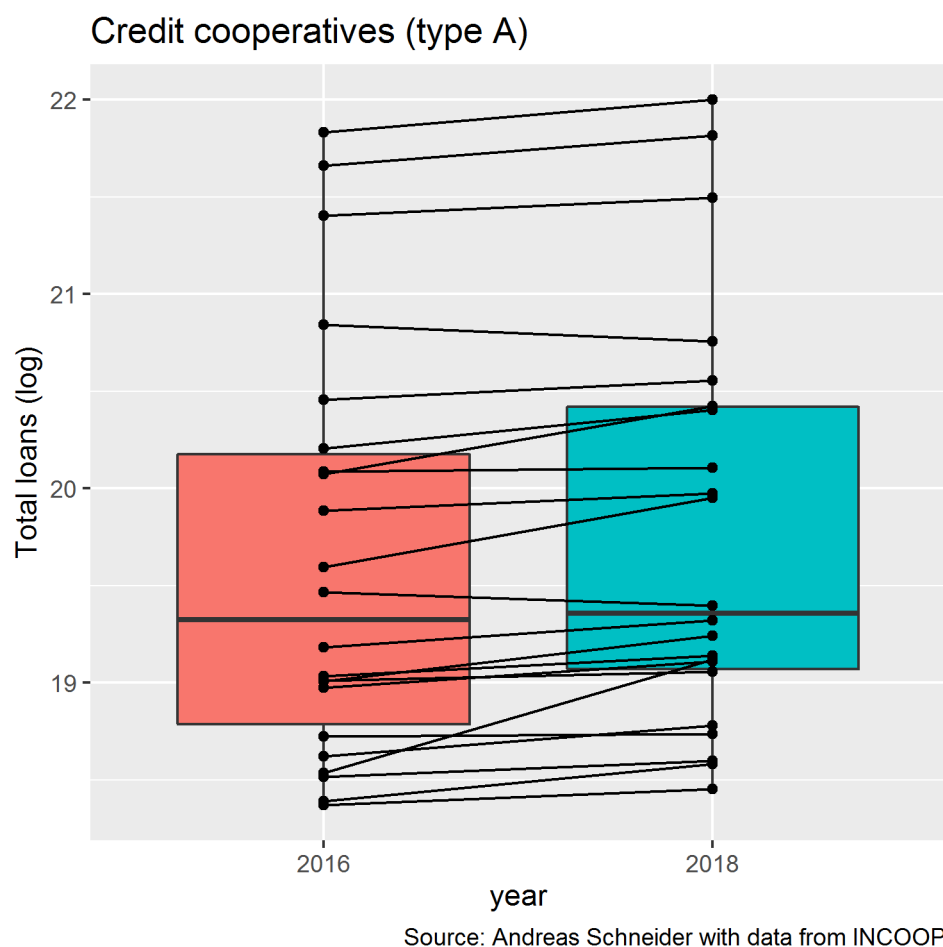

Figure 2: Paired sample of large credit cooperatives

Having a look a the output of Figure 2, we see a box plot with paired values of the cooperatives and the evolution of their respective total loans over time for the two sample years 2016 and 2018.

Availability

A stable version of concstats is available on CRAN and has a dedicated website for a quick start. The development version is available on GitHub

Acknowledgements

References

Ginevičius, R., & Čirba, S. (2009). Additive measurement of market concentration. *Journal of Business Economics and Management*, *10*(3), 191–198. https://doi.org/10.3846/1611-1699.2009.10.191-198

Herfindahl, O. C. (1950). *Concentration in the steel industry* [PhD thesis]. Columbia University, New York.

Hirschman, A. O. (1945). *National power and structure of foreign trade*. University of California Press.

Idunate, P. G. A. (1994). Un indice de dominacion para el analisis de la estructura de los mercados. *El Trimestre Economico*, *61*(243(3)), 499–524. https://www.jstor.org/stable/20856736

Kenny, C. T. (2022). *divseg: Compute diversity and segregation indices* (Version v0.0.5). https://CRAN.R-project.org/package=divseg

Melnik, A., Shy, O., & Stenbacka, R. (2008). Assessing market dominance. *Journal of Economic Behavior & Organization*, *68*(1), 63–72. https://doi.org/10.1016/j.jebo.2008.03.010

Palma, J. (2006). Globalizing inequality: "Centrifugal" and "centripetal" forces at work. *UN Department of Economic and Social Affairs (DESA) Working Papers*, *35*. https://doi.org/10.18356/395c7874-en

Schneider, A. (2022). Kreditgenossenschaften: Marktstruktur, wettbewerb und verhalten. Das beispiel paraguay. *Zeitschrift Fuer Das Gesamte Genossenschaftswesen*, *72*(1), 7–38. https://doi.org/10.1515/zfgg-2022-0002

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sohn, A. (2016). *Acid: Analysing conditional income distributions* (Version v1.1.0). https://CRAN.R-project.org/package=acid

Tabak, B. M., Guerra, S. M., & De Souza Penaloza, R. (2009). Banking concentration and the price-concentration relationship: The case of brazil. *International Journal of Accounting and Finance*, *1*(4), 415–435. https://doi.org/10.1504/IJAF.2009.029147

Waggoner, P. D. (2018). The hhi package: Streamlined calculation and visualization of herfindahl-hirschman index scores. *Journal of Open Source Software*, *3*(28), 828. https://doi.org/10.21105/joss.00828

Zeileis, A., & Kleiber, C. (2014). *Ineq: Measuring inequality, concentration, and poverty* (Version v0.2-13). https://CRAN.R-project.org/package=ineq