LMU
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
INSTITUT FÜR STATISTIK

Jasmin Abedieh & Andreas Groll & Manuel J. A. Eugster

Measuring Concentration in Data with an Exogenous Order

Technical Report Number 140, 2013
Department of Statistics
University of Munich

http://www.stat.uni-muenchen.de

Measuring Concentration in Data with an Exogenous Order

Jasmin Abedieh * Andreas Groll † Manuel J. A. Eugster ‡

February 8, 2013

Abstract
Concentration measures order the statistical units under observation according to their market share. However, there are situations where an order according to an exogenous variable is more appropriate or even required. The present article introduces a generalized definition of market concentration and defines a corresponding concentration measure. It is shown that this generalized concept of market concentration satisfies the common axioms of (classical) concentration measures. In an application example, the proposed approach is compared with classical concentration measures; the data are transfer spendings of German Bundesliga soccer teams, the "obvious" exogenous order of the teams is the league ranking.
Keywords Market concentration, Concentration measure, Exogenous order, German Bundesliga.

1 Introduction

Market concentration is a concept

> "*[...] which measures the relative position of large enterprises in the provision of specific goods or services [...]*",

as defined by the *Glossary of Industrial Organisation Economics and Competition Law* (OECD, 1993). It is a prevalent but highly discussed concept; many attempts have been made to define sound measures of concentration (prominent publications are Hall and Tideman, 1967; Hannah and Kay, 1977; Encaoua and Jacquemin, 1980).

An interesting aspect of the given definition—and of all similar definitions available in the literature—is the term "large enterprises". Well-established concentration measures like the concentration ratio and the Herfindahl index interpret "large" in terms of the "specific goods". In detail, the measures are computed on the enterprises ordered according to their market share. From the economics point of view this modus operandi has nice properties; Saving (1970), for example, shows the relation between

*Department of Statistics, Ludwig-Maximilians-University Munich, 80539 Munich, *jasmin.abedieh@hotmail.de*

†Department of Mathematics, Workgroup Financial Mathematics, Ludwig-Maximilians-University Munich, Theresienstr. 39, 80333 Munich, Germany, *andreas.groll@math.lmu.de*

‡Department of Statistics, Ludwig-Maximilians-University Munich, 80539 Munich, *Manuel.Eugster@stat.uni-muenchen.de*

the Lerner measure of the degree of monopoly and concentration ratios expressed by the market share of the g largest enterprises.

However, in this paper we present a different interpretation of “large enterprises”. In fact, we generalize the definition of market concentration and allow an order of the enterprises according to an exogenous variable. This exogenous order can be defined by any property of the enterprises—for example, the number of employees, a rating agency’s ranking, the geographical position from south to north, or the enterprises’ environmental dues. We introduce an appropriate concentration measure and show that this concentration measure satisfies the conventional axiomatic of (classical) concentration measures.

The paper is organized as follows. In Section 2 we review the formalism of classical concentration measures and show common representatives. In Section 3 we introduce the concentration measure for data with an exogenous order. Following the introduced formalism we define a specific concentration rate, concentration curve and concentration index. Furthermore, we discuss its axiomatic (the formal proofs are available in the Appendix A.2). In Section 4 we present an illustrative application. We investigate the transfer spendings among the soccer teams of the German Bundesliga for the seasons 1992/1993 to 2009/2010. The “obvious” exogenous order of the teams is their league ranking at the end of each season. Finally, in Section 5 the conclusions are given. All data sets and source codes for replicating our analyses are freely available (see Appendix A.1 on computational details).

2 Concentration measures

In this section we review the formalism of classical concentration measures and show common representatives. Given $1, \ldots, n$ statistical units (e.g., enterprises), let X be a specific characteristic of the statistical units (e.g., market share) and $x_1, \ldots, x_n$ positive realizations (observations). The increasing order or decreasing order of the observations is denoted by

$$0 \leq x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$$

and

$$x^{(1)} \geq x^{(2)} \geq \ldots \geq x^{(n)} \geq 0,$$

respectively (with $\sum_{i=1}^{n} x_i > 0$). The corresponding ordered relative sums of observations then are defined by

$$p_i := \frac{x_{(i)}}{\sum\limits_{j=1}^{n} x_j} \qquad \text{and} \qquad c_i := \frac{x^{(i)}}{\sum\limits_{j=1}^{n} x_j}.$$

The vectors $\mathbf{p}^T = (p_1, \ldots, p_n)$ and $\mathbf{c}^T = (c_1, \ldots, c_n)$ represent the corresponding successive sums for $i = 1, \ldots, n$. Using this formalism, we are able to define common measures of concentration; we present three representatives which we use in the application example (Section 4).

Concentration ratio. The concentration ratio is defined as

$$\mathrm{CR}_g := \sum_{i=n-g+1}^{n} p_i = \sum_{i=1}^{g} c_i,$$

and yields values in $[0, 1]$. Concentration ratios show the extend of control of the g largest statistical objects. In terms of our application example, the concentration ratio explains the extent of control of a group of soccer teams with the g highest transfer spendings in the league, illustrating the degree of dominance. Based on the concentration ratios, the inequality can be visualized by the concentration curve (see, e.g., Wagstaff et al., 1991).

Herfindahl index. The Herfindahl-Hirschman index (Hirschman, 1964) is defined as

$$\mathrm{H} := \sum_{i=1}^{n} p_i^2 = \sum_{i=1}^{n} c_i^2,$$

and results in values $\frac{1}{n} \leq H \leq 1$. H is an indicator of the amount of competition among the statistical units, i.e., represents the degree of concentration. In our application example, it can be used as an indicator whether there is a monopoly or a significant competition on the transfer spendings.

Rosenbluth Index. The Rosenbluth index (Rosenbluth, 1955, 1957) is defined as

$$\mathrm{RB} := \frac{1}{2\sum_{i=1}^{n} i\, c_i - 1} = \frac{1}{2A},$$

with

$$A = \sum_{i=1}^{n} i\, c_i - \frac{1}{2} \qquad \text{and} \qquad \frac{1}{2} \leq A \leq \frac{n}{2},$$

and results in values $\frac{1}{n} \leq \mathrm{RB} \leq 1$. RB denotes the area above the concentration curve. It constitutes an alternative measure to investigate the absolute concentration of a particular group of statistical units based on the kurtosis of the concentration curve.

For both Herfindahl and Rosenbluth index, normalized versions are available:

$$\mathrm{H}^* := (\mathrm{H} - \frac{1}{n})/(1 - \frac{1}{n})$$

and

$$\mathrm{RB}^* := (\mathrm{RB} - \frac{1}{n})/(1 - \frac{1}{n}),$$

with $0 \leq \mathrm{H}^* \leq 1$ and $0 \leq \mathrm{RB}^* \leq 1$. Furthermore, the measures' inverse, $n_{\mathrm{H}} = 1/\mathrm{H}$ and $n_{\mathrm{RB}} = 1/\mathrm{RB}$, are of interest as well—they can be interpreted as the "equivalent number of equal sized units".

Other well-known measures of concentration within this formalism are the Lorenz curve, the Gini coefficient, the Entropy (Shannon, 1948) and the Exponential index (for all see, e.g., Curry and George, 1983).

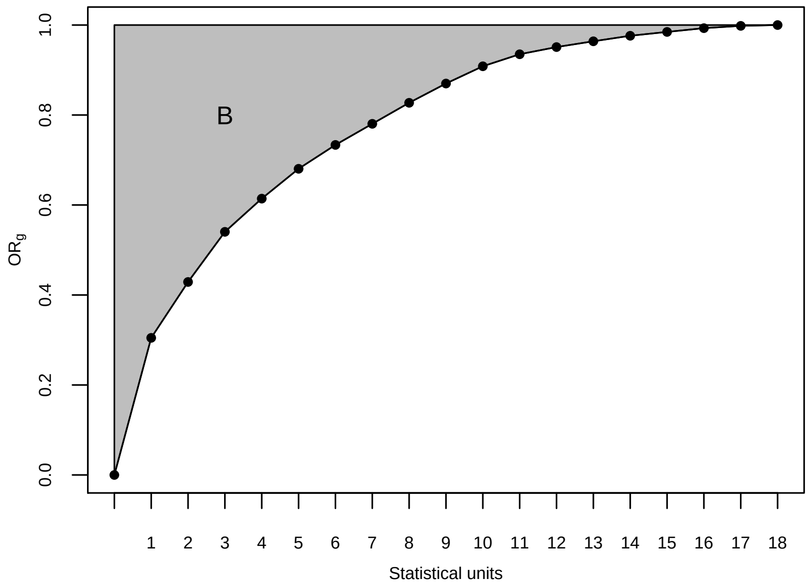

Figure 1: Illustration of a concentration curve for data with an exogenous order.

3 Concentration measure for data with an exogenous order

Following the formalism introduced in the previous section we define a concentration measure based on the concentration rate CR_g for data with an exogenous order. Given the data $x_1, \ldots, x_n$, their order based on an exogenous variable is denoted by

$$x_{[1]}, x_{[2]}, \ldots, x_{[n]}.$$

Analogously, we assume $\sum_{i=1}^{n} x_i > 0$ and define the ordered relative sum of statistical units with respect to the exogenous order:

$$q_i := \frac{x_{[i]}}{\sum\limits_{j=1}^{n} x_j}.$$

The vector $\mathbf{q}^T = (q_1, \ldots, q_n)$ represents the corresponding successive sums for $i = 1, \ldots, n$.

Concentration ratio. In analogy to the concentration ratio CR_g we define the exogenously ordered concentration rate; characterizing which part of the sum of objects lies on the group $x_{[1]}, \ldots, x_{[g]}$. We define

$$\mathrm{OR}_g := \sum_{i=1}^{g} q_i,$$

with values in $[0, 1]$. OR_g explains the proportion of the first g statistical units on the overall sum with respect to the exogenous order. In terms of the application example $g = 3$ and $g = 9$ are of special interest: OR_3 is the concentration ratio of transfer spendings of the teams competing for the qualification of the UEFA Champions League; and OR_9 is the concentration of the better half of the league.

Concentration curve. Based on OR_g we define the exogenously ordered concentration curve. In contrast to the (classical) concentration curve, ordered relative sums of objects according to the exogenous order form a curve which is still monotone increasing but not necessarily concave. As a consequence the frequency polygon can cross the diagonal from $(0,0)$ to $(n,1)$. Figure 1 illustrates a schematic exogenously ordered concentration curve.

Concentration index. We use the schematic exogenously ordered concentration curve in Figure 1 to motivate the definition of an appropriate concentration index. The inequality in data with an exogenous order is illustrated by the area B, which lies above the exogenously ordered concentration curve. The following relation holds: The smaller the surface, the bigger the proportion among the "first few" statistical units. In the extreme case that the whole balance applies on the first statistical unit, one obtains $B_{min} = 0.5$. Note that the uniform distribution does not represent one of the two extreme cases anymore. The inequality among the first statistical units is now minimal, if the whole balance applies to the last statistical unit; in this case we obtain $B_{max} = n - 0.5$. For the uniform distribution we get $q_i = \frac{1}{n}$ for all i and the exogenous ordered concentration curve is the diagonal with the corresponding area $B = \frac{n}{2}$.

In general, B is computed by the formula

$$B = \sum_{i=1}^{n} i q_i - 0.5.$$

Based on this area B, we introduce an index which captures the concentration in data with an exogenous order:

$$\mathrm{OI} := \frac{1}{2B} = \frac{1}{2\sum_{i=1}^{n} i q_i - 1}$$

OI results in $[\frac{1}{2n-1}, 1]$. Note that the uniform distribution with $B = \frac{n}{2}$ results in $\mathrm{OI} = \frac{1}{n}$. For interpretation, the following statements can be proposed:

$OI \in (\frac{1}{n}, 1]$	concentration on anterior statistical units
$OI \in [\frac{1}{2n-1}, \frac{1}{n})$	concentration on posterior statistical units
$OI = \frac{1}{n}$	no concentration, all statistical units have the same proportion of the sum

In analogy to the classical concentration measures we define a normalized version as well as the measures' inverse. The normalized version is ($OI^* \in [0,1]$):

$$\mathrm{OI}^* := \frac{B - c}{1 - c} \qquad \text{with} \qquad c = \frac{1}{2n-1}$$

And the measures' inverse, as the "equivalent number of congenerous units", is $n_{\mathrm{OI}} = \frac{1}{\mathrm{OI}}$, with $1 \leq n_{\mathrm{OI}} \leq 2n - 1$.

Note that now the interpretation of the measures' inverse n_{OI} is subject to the following restrictions. Here, the extreme cases occur if the whole balance applies to "the first" or "the last" statistical unit (in sense of the exogenous order) resulting in $n_{\mathrm{OI}} \in \{1; 2n-1\}$. Consequently, the uniform distribution does not represent an extreme case anymore and can be interpreted as a "medium concentration of the first statistical units" with the corresponding equivalent number of congenerous statistical units being exactly equal to the true number of statistical units, $n_{\mathrm{OI}} = n$. Furthermore, it is possible that the equivalent number of congenerous units exceeds the actual number of statistical units, i.e., $n_{\mathrm{OI}} > n$; if this occurs, the balance applies on the last statistical units. In contrast we obtain $n_{\mathrm{OI}} < n$, if the balance applies on the first statistical units.

Axiomatic of the concentration index OI. Literature discusses and defines a set of characteristics required by concentration measures (see, e.g., Hannah and Kay, 1977; Encaoua and Jacquemin, 1980; Piesch, 1975). The concentration index for exogenous ordered data OI satisfies the following axioms (always with respect to the exogenous order).

Independence: Unchanged concentration when proportionally transforming the data.

Symmetry: Unchanged concentration when permuting the ranking of two statistical units while simultaneously exchanging their value.

Continuity: Continuous concentration with respect to the exogenously ordered relative sum of statistical units g.

Translations: Decreasing (increasing) concentration when reallocating a value from an anterior statistical unit to a posterior statistical unit (and vice versa).

Proportionality: Decreasing concentration of factor $1/c$ when replacing each statistical unit by c equal sized units.

Extention: Unchanged concentration when augmenting the data with new statistical units each with value zero.

Standardization: Transformation of the concentration index into a standardized version with values in $[0, 1]$.

Appendix A.2 provides the formal proofs.

4 Transfer spendings in the German Bundesliga

In this section we illustrate the differences between classical approaches of concentration measurement and our proposed approach. For this purpose, we investigate the inequality with respect to the variable *transfer spendings* among the soccer teams of the German Bundesliga, which are ordered by their league rankings at the end of each season. The data contains the amount of transfer spendings (in Euro), spent at the start of each season, for the years 1992/1993 to 2009/2010. Classical measures of concentration could be used to analyze, if a general financial inequality can be observed in the German Bundesliga or if some soccer teams are financially superior to others, respectively. What is more, it is to be assumed that those teams that spend more money for new players than other teams, could derive a sportive advantage from this and thus could achieve better league ranks. Such a relationship can not be captured by classical measures of concentration. Our proposed concentration measure for exogenous ordered data—the league rankings of the teams—allows to interpret the observed inequality with respect to the sportive success of the soccer teams.

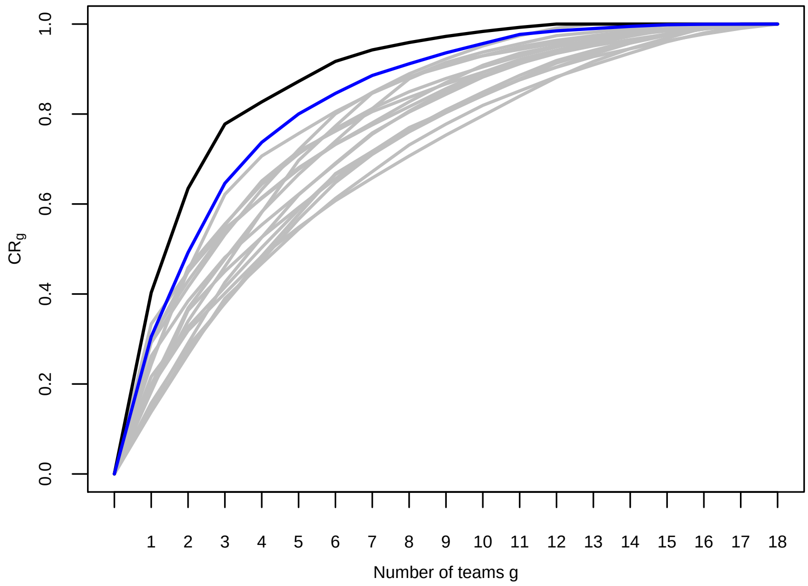

Figure 2: Concentration curves of transfer spendings in the German Bundesliga for the seasons 1992/1993–2009/2010; the black curve reflects the highest concentration (season 1992/1993); the blue curve reflects the second-highest concentration (season 2003/2004).

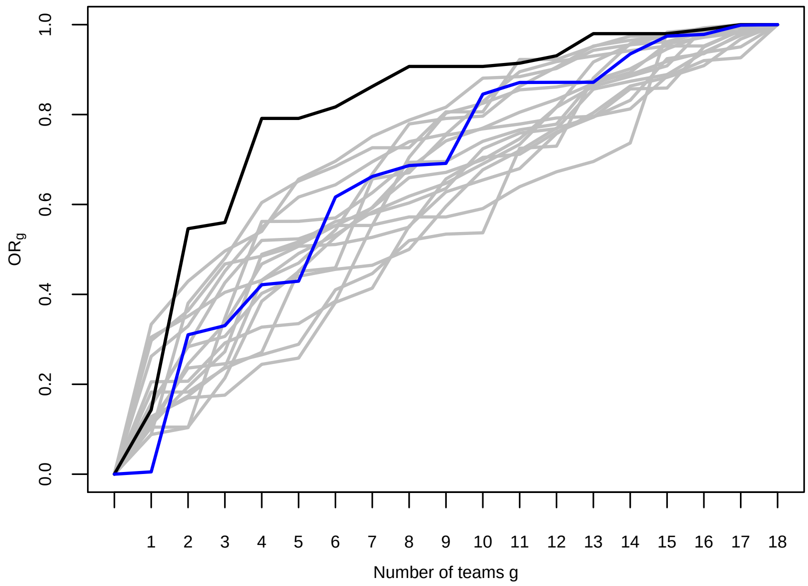

Figure 3: Exogenously ordered concentration curves of transfer spendings in the German Bundesliga for the seasons 1992/1993–2009/2010 (black curve: season 1992/1993; blue curve: season 2003/2004).

First, we show the conventional concentration curves corresponding to the transfer spendings of the 18 considered seasons in Figure 2. It can be seen, that all seasons exhibit a certain level of inequality, which is for some seasons more distinct than for others. The highest concentration is observed for the season 1992/1993 and is represented by the black curve, which lies above all other curves. For example, the three teams with the biggest amount of transfer spendings own a proportion of almost 80% of the league's total spendings. This indicates, that the league was in general very unbalanced in that season with respect to investments into new players. However, there is no information about the sportive success of these three teams available.

For comparison, Figure 3 shows the corresponding exogenously ordered concentration curves. Again, the black curve on top represents the season 1992/1993. As teams are now ordered by their league rankings, the curves contain more information. For example, it can be seen that the *top* four teams own already a proportion of about 80% of the league's total spendings. Hence, in this season the sportive success of the teams seems to be strongly connected with the amount of money that has been spent for new players at the start of the season. What is also remarkable is that in some seasons, as for example in the season 2003/2004 (lower blue curve), the top team's proportion of the league's total spendings is almost zero (here $\mathrm{OR}_1 = 0.005$). This may be due to huge investments in the preceding seasons and results in an ordered concentration curve that partly lies below the bisecting line.

Season	H	$H^\star$	RB	$RB^\star$	OI	$OI^\star$
1992/1993	0.244	0.200	0.225	0.180	0.143	0.118
1993/1994	0.090	0.037	0.091	0.038	0.074	0.047
1994/1995	0.083	0.029	0.090	0.036	0.062	0.035
1995/1996	0.088	0.034	0.096	0.043	0.069	0.041
1996/1997	0.097	0.044	0.106	0.053	0.083	0.056
1997/1998	0.145	0.095	0.142	0.091	0.104	0.078
1998/1999	0.099	0.046	0.099	0.046	0.074	0.047
1999/2000	0.115	0.063	0.124	0.072	0.065	0.037
2000/2001	0.137	0.086	0.129	0.078	0.079	0.051
2001/2002	0.120	0.068	0.112	0.059	0.099	0.072
2002/2003	0.149	0.099	0.146	0.095	0.065	0.037
2003/2004	0.170	0.121	0.165	0.115	0.083	0.056
2004/2005	0.145	0.094	0.128	0.077	0.090	0.064
2005/2006	0.089	0.035	0.097	0.044	0.072	0.045
2006/2007	0.115	0.063	0.122	0.070	0.071	0.044
2007/2008	0.154	0.104	0.124	0.072	0.111	0.085
2008/2009	0.101	0.048	0.101	0.048	0.076	0.050
2009/2010	0.143	0.092	0.125	0.073	0.094	0.067

Table 1: Herfindahl, Rosenbluth and ordered concentration index, together with their normalized versions for transfer spendings in the German Bundesliga, for the seasons 1992/1993–2009/2010.

Additionally, we present the corresponding conventional measures of concentration as well as the novel exogenously ordered concentration index in Table 1, together with their normalized versions. By construction, for all measures the concentration values of the normalized versions are smaller than for the original measures. In the following we focus on the original versions of the indices. Both Herfindahl and Rosenbluth index yield concentrations between 0.08 and 0.25, which indicates a general inequality for transfer spendings in the German Bundesliga. Though, more information is offered by the ordered concentration index. No concentration on better or worse teams would be given, if all teams had the same proportion of the total sum of transfer spendings. In this case, the ordered concentration index would yield OI = 1/18 = 0.056. The results in Table 1 show that this limit is exceeded for each of the considered seasons. Consequently, a concentration of the transfer spendings on better teams is observed, and never on worse teams. This seems to confirm the supposition, that spending more money for new players results in sportive success.

Season	1/H	1/RB	1/OI
1992/1993	4.09	4.44	6.99
1993/1994	11.09	10.98	13.54
1994/1995	12.06	11.14	16.06
1995/1996	11.38	10.45	14.56
1996/1997	10.33	9.46	12.02
1997/1998	6.90	7.06	9.62
1998/1999	10.07	10.14	13.55
1999/2000	8.67	8.08	15.45
2000/2001	7.30	7.73	12.74
2001/2002	8.35	8.96	10.10
2002/2003	6.70	6.87	15.50
2003/2004	5.88	6.08	12.00
2004/2005	6.91	7.80	11.07
2005/2006	11.30	10.33	13.91
2006/2007	8.66	8.19	14.04
2007/2008	6.50	8.07	9.00
2008/2009	9.91	9.92	13.10
2009/2010	7.02	8.02	10.68

Table 2: Equivalent number of equal sized units of Herfindahl and Rosenbluth index as well as equivalent number of congenerous units of ordered concentration index for transfer spendings in the German Bundesliga, for the seasons 1992/1993–2009/2010.

This becomes even more obvious, when considering the corresponding equivalent numbers, see Table 2. The results of the equivalent numbers of equal sized units of Herfindahl and Rosenbluth index represent adequately the general unbalance for transfer spendings in the German Bundesliga, for the years 1992/1993–2009/2010. They correspond to an effective quantity from minimum 4, to maximum 12 equally active teams on the transfer market. In contrast, the equivalent number of congenerous units, 1/OI, offers more information than just recognizing that the league is generally unbalanced with respect to transfer spendings. As we obtain $n_{\mathrm{OI}} < 18$ for all seasons, the balance strictly applies on better teams, with a minimum of 7 and a maximum of 16 congenerous teams. Once again, the supposition that investing into new players can increase the sportive success of soccer teams, at least for the German Bundesliga, is confirmed by these results.

5 Conclusion

The present paper introduces a generalization of market concentration. Instead of ordering statistical units (e.g., enterprises) according to their market share, the generalization allows their ordering according to an exogenous variable. This now enables to see market concentration in relation to further aspects, e.g, number of employees or a rating agency's ranking. We define an appropriate concentration measure for data with an exogenous order and show that this measure satisfies the common axioms for classical concentration measures. The application example shows the proposed method in terms of transfer spendings of German Bundesliga soccer teams—whose "obvious" exogenous orders are their league rankings.

Acknowledgement

The authors would like to thank Thomas Augustin for usefull discussion and help.

A Appendix

A.1 Computational details

All computations and graphics have been done using the statistical software R 2.15.2 (R Development Core Team, 2012), and the SportsAnalytics package (Eugster, 2011). R itself and all packages used are freely available under the terms of the General Public License from the Comprehensive R Archive Network at `http://CRAN.R-project.org/`.

Data sets and source codes for replicating our analyses are available in the SportsAnalytics package. Execute the analysis of the German Bundesliga via:

```
R> demo("concentration-bundesliga", package = "SportsAnalytics")
```

The source code file for a demo is accessible via:

```
R> edit(file = system.file("demo", "concentration-bundesliga.R",
+                          package = "SportsAnalytics"))
```

A.2 Axiomatic properties of the concentration index OI

Independence of the measuring scale. Let $x_{[1]}, \dots x_{[n]}$ be exogenously ordered realizations of a metric object X, with corresponding relative sums of objects $q_i = \frac{x_{[i]}}{\sum_{j=1}^{n} x_j}, i = 1, \dots, n$. If the data $x_{[1]}, \dots, x_{[n]}$ is proportionally transformed to $\tilde{x}_{[i]} = ax_{[i]}, a > 0, i = 1, \dots, n$, we obtain the new relative sums of objects:

$$\tilde{q}_i = \frac{\tilde{x}_{[i]}}{\sum\limits_{j=1}^{n} \tilde{x}_j} = \frac{ax_{[i]}}{\sum\limits_{j=1}^{n} ax_j} = \frac{ax_{[i]}}{a\sum\limits_{j=1}^{n} x_j} = \frac{x_{[i]}}{\sum\limits_{j=1}^{n} x_j} = q_i.$$

So it follows directly from the definition of OI that the data $x_{[1]}, \dots, x_{[n]}$ and $\tilde{x}_{[1]}, \dots, \tilde{x}_{[n]}$ have exactly the same concentration. Consequently, X and aX, $a > 0$, have the same concentration.

Symmetry. If one permutates the ranking of two statistical units i and j and simultaneously transfers their observations x_i and x_j, there will be no change in the vector of sums of objects corresponding the ranking position $\mathbf{q}$. It is trivial that the value of OI remains also unchanged.

Continuity. The following holds:

$$\begin{aligned}
\lim_{\mathbf{q}\to\tilde{\mathbf{q}}} \mathrm{OI}(n;\mathbf{q}) &= \lim_{\mathbf{q}\to\tilde{\mathbf{q}}} \frac{1}{2\sum_{i=1}^{n} q_i - 1} \\
&\stackrel{(*)}{=} \frac{1}{2\sum_{i=1}^{n} \lim\limits_{q_i\to\tilde{q}_i} q_i - 1} \\
&= \frac{1}{2\sum_{i=1}^{n} \tilde{q}_i - 1} \\
&= \mathrm{OI}(n;\tilde{\mathbf{q}}),
\end{aligned}$$

where we used in $(*)$ that the limit of a vector is regarded componentwise, thus $\lim\limits_{\mathbf{q}\to\tilde{\mathbf{q}}} = (\lim\limits_{q_1\to\tilde{q}_1}, \ldots, \lim\limits_{q_n\to\tilde{q}_n})$.

Translations. For given vector of successive sums $\mathbf{q}^T = (q_1, \ldots, q_n)$ and two statistical units i and j with $i < j$, reallocation of an object sum d, $0 < d \leq q_i$, from the anterior unit i to the posterior unit j results in the new vector of sums of objects $\tilde{\mathbf{q}}^T := (q_1, \ldots q_{i-1}, q_i - d, q_{i+1}, \ldots, q_{j-1}, q_j + d, q_{j+1}, \ldots, q_n)$ and we obtain:

$$\begin{aligned}
\sum_{k=1}^{n} k\tilde{q}_k &= \sum_{k=1}^{i-1} kq_k + iq_i - id + \sum_{k=i+1}^{j-1} kq_k + jq_j + jd + \sum_{k=j+1}^{n} kq_k \\
&= \sum_{i=1}^{n} kq_k - id + jd \\
&= \sum_{i=1}^{n} kq_k + d(j-i) \\
&\stackrel{(*)}{>} \sum_{i=1}^{n} kq_k,
\end{aligned}$$

where we used in $(*)$ that $d(j-i) > 0$ holds, as $j > i$. Consequently:

$$\begin{aligned}
& 2\textstyle\sum_{k=1}^{n} k\tilde{q}_k && > && 2\textstyle\sum_{i=1}^{n} kq_k \\
\Longrightarrow\quad & 2\textstyle\sum_{k=1}^{n} k\tilde{q}_k - 1 && > && 2\textstyle\sum_{i=1}^{n} kq_k - 1 \\
\Longrightarrow\quad & \tfrac{1}{2\sum_{k=1}^{n} k\tilde{q}_k - 1} && < && \tfrac{1}{2\sum_{i=1}^{n} kq_k - 1} \\
\Longrightarrow\quad & \mathrm{OI}(n;\tilde{\mathbf{q}}) && < && \mathrm{OI}(n;\mathbf{q}).
\end{aligned}$$

In this case the concentration decreases. Note that for conventional concentration measures usually the additional conditions $q_i > q_j$ and $q_i - h = q_j$ with $0 < d < \frac{h}{2}$ are claimed. This reflects a special case, which is also covered by the more general proof given above.

In contrast, if one transfers instead an arbitrary value $0 < d \leq q_j$ from a posterior unit j to an anterior unit i, one obtains $\tilde{\mathbf{q}}^T := (q_1, \ldots q_{i-1}, q_i + d, q_{i+1}, \ldots, q_{j-1}, q_j - d, q_{j+1}, \ldots, q_n)$ and the corresponding proof remains valid analogously with turned inequalities.

Proportionality. For $n \geq 2$ statistical units let the relative sums of objects be given by $\mathbf{q}^T = (q_1, \ldots, q_n)$. Then the corresponding space over the ordered concentration curve yields $B = \sum_{i=1}^{n} iq_i - 0.5$. If one replaces each unit i with corresponding proportion q_i by c equal sized units with proportions $\frac{q_i}{c}$, one obtains the new relative vector of sums of objects $\tilde{\mathbf{q}} = (\tilde{q}_1, \ldots, \tilde{q}_{cn})^T = (\frac{q_1}{c}, \ldots, \frac{q_1}{c} \ldots, \frac{q_n}{c}, \ldots, \frac{q_n}{c})$ with corresponding space $\tilde{B} = \sum_{i=1}^{cn} i\tilde{q}_i - 0.5$. In the following we show that $\tilde{B} = cB$ is satisfied. We can derive:

$$\begin{aligned}
\tilde{B} &= \sum_{i=1}^{cn} i\tilde{q}_i - 0.5 \\
&= \sum_{i=1}^{c} i\tilde{q}_i + \sum_{i=c+1}^{2c} i\tilde{q}_i + \ldots + \sum_{i=c(n-1)+1}^{cn} i\tilde{q}_i - 0.5 \\
&= \sum_{i=1}^{c} i\frac{q_1}{c} + \sum_{i=c+1}^{2c} i\frac{q_2}{c} + \ldots + \sum_{i=c(n-1)+1}^{cn} i\frac{q_n}{c} - 0.5 \\
&= \frac{q_1}{c}\sum_{i=1}^{c} i + \frac{q_2}{c}\sum_{i=c+1}^{2c} i + \ldots + \frac{q_n}{c}\sum_{i=c(n-1)+1}^{cn} i - 0.5 \\
&= \frac{q_1}{c}\frac{c(c+1)}{2} + \frac{q_2}{c}\left(c \cdot c + \sum_{i=1}^{c} i\right) + \ldots + \frac{q_n}{c}\left(c(n-1) \cdot c + \sum_{i=1}^{c} i\right) - 0.5 \\
&= q_1\frac{(c+1)}{2} + \frac{q_2}{c}\left(c \cdot c + \frac{c(c+1)}{2}\right) + \ldots + \frac{q_n}{c}\left(c(n-1) \cdot c + \frac{c(c+1)}{2}\right) - 0.5 \\
&= q_1\frac{(c+1)}{2} + q_2\left(c + \frac{c+1}{2}\right) + \ldots + q_n\left((n-1) \cdot c + \frac{c+1}{2}\right) - 0.5 \\
&= cq_2 + 2cq_3 + \ldots + (n-1)cq_n + \frac{c}{2}\underbrace{\sum_{i=1}^{n} q_i}_{=1} + 0.5\underbrace{\sum_{i=1}^{n} q_i}_{=1} - 0.5 \\
&= cq_2 + 2cq_3 + \ldots + (n-1)cq_n + \frac{c}{2} + 0.5 - 0.5 \\
&= cq_2 + 2cq_3 + \ldots + (n-1)cq_n - \frac{c}{2} + c \\
&= cq_2 + 2cq_3 + \ldots + (n-1)cq_n - \frac{c}{2} + c\sum_{i=1}^{n} q_i \\
&= cq_1 + 2cq_2 + \ldots + ncq_n - \frac{c}{2} \\
&= c\sum_{i=1}^{n} iq_i - \frac{c}{2} \\
&= cB
\end{aligned}$$

It follows:

$$\mathrm{OI}(cn; \tilde{\mathbf{q}}) = \frac{1}{2\tilde{B}} = \frac{1}{2cB} = \frac{1}{c}\mathrm{OI}(n; \mathbf{q}).$$

Extension. If one augments a distribution $\mathbf{q}^T = (q_1, \ldots, q_n)$ by m new units, each with an object sum equal to zero, one obtains the new vector of ordered relative sums $\tilde{\mathbf{q}} := (q_1, \ldots q_n, 0, \ldots, 0)$ of length $n+m$. Then the following holds:

$$\begin{aligned}
\mathrm{OI}(n+m; \tilde{\mathbf{q}}) &= \frac{1}{2\sum_{i=1}^{(n+m)} iq_i - 1} \\
&= \frac{1}{2\left(\sum_{i=1}^{n} iq_i + \sum_{i=n+1}^{n+m} iq_i\right) - 1} \\
&= \frac{1}{2\left(\sum_{i=1}^{n} iq_i + \sum_{i=n+1}^{n+m} i \cdot 0\right) - 1} \\
&= \frac{1}{2\sum_{i=1}^{n} iq_i - 1} = \mathrm{OI}(n; \mathbf{q}).
\end{aligned}$$

Thus, the concentration remains unchanged.

Standardization. With the standardization proposed in Section 3, OI can be transformed into a version OI* with values in $[0, 1]$.

With $B_{max} = n - 0.5$ and $B_{min} = 0.5$ we get:

$$\begin{aligned}
\max_q \mathrm{OI} &= \frac{1}{2B_{min}} = \frac{1}{2 \cdot 0.5} = 1, \\
\min_q \mathrm{OI} &= \frac{1}{2B_{max}} = \frac{1}{2(n-0.5)} = \frac{1}{2n-1},
\end{aligned}$$

consequently $\mathrm{OI} \in [\frac{1}{2n-1}, 1]$ holds. In the standardized version OI* the lower interval limit is adjusted by the numerator up to 0. The denominator normalizes OI by the corresponding interval width, so that the new interval length is equal to one.

References

Curry, B. and K. D. George (1983). Industrial concentration: A survey. *The Journal of Industrial Economics 31*(3), 203–255.

Encaoua, D. and A. Jacquemin (1980). Degree of monopoly, indices of concentration and threat of entry. *International Economic Review 21*(1), 87–105.

Eugster, M. J. A. (2011). *SportsAnalytics: Sports Analytics.* R package version 0.1.

Hall, M. and N. Tideman (1967). Measures of concentration. *Journal of the American Statistical Association 62*(317), 162–168.

Hannah, L. and J. A. Kay (1977). *Concentration in Modern Industry: Theory, Measurement and the U.K. Experience.* Macmillan.

Hirschman, A. O. (1964). The paternity of an index. *American Economic Review 54*, 761–762.

OECD (1993). *Glossary of Industrial Organisation Economics and Competition Law.* Organization for Economic.

Piesch, W. (1975). *Statistische Konzentrationsmae. Formale Eigenschaften und verteilungstheoretische Zusammenhnge*. Mohr (Siebeck).

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rosenbluth, G. (1955). Measures of concentration. In National Bureau of Economic Research (Ed.), *Business Concentration and Price Policy*. Princeton University Press.

Rosenbluth, G. (1957). *Concentration in Canadian Manufacturing Industries*. Princeton University Press.

Saving, T. S. (1970). Concentration ratios and the degree of monopoly. *International Economic Review 11*(1), 139–146.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal 27*, 379–423 and 623–656.

Wagstaff, A., P. Paci, and E. van Doorslaer (1991). On the measurement of inequalities in health. *Social Science & Medicine 33*(5), 545–557.