Abstract

While the two most widely used measures of market (industrial) concentration, the \( m \)-firm concentration ratio \( CR_m \) and the Herfindahl-Hirschman index \( H \), have no precise functional relationship, they can be related by means of boundary formulations. Such bounds and potential relationships, which have been considered in some earlier reported studies, are being re-examined, corrected, and reformulated in this paper. The underlying analysis uses a different approach based on majorization theory and the results are supported by computer simulation. Such boundary relationships make it possible to determine approximate values of \( H \) from those of \( CR_m \) and vice versa for any given set of market shares. Much more accurate predictions of \( H \)-values can be obtained with knowledge of the individual market shares of the \( m \) largest firms within a market (industry), with or without knowledge of the total number of firms.

Keyword: Economics
1. Introduction

Market concentration, also often referred to as industry concentration, refers to the extent to which the market shares of the largest firms within a market (industry) accounts for a large proportion of economic activity such as sales, assets, or employment. As stated by OECD [1]:

“The rationale underlying the measurement of industry or market concentration is the industrial organization economic theory which suggests that, other things being equal, high levels of market concentration are more conducive to firms engaging in monopolistic practices which leads to misallocation of resources and poor economic performance. Market concentration in this context is used as one possible indicator of market power.”

Increasing market concentration causes decreasing competition and efficiency and increasing market power. Any such trends are being monitored by the business community and by government antitrust authorities such as the U.S. Department of Justice (DOJ) and the Federal Trade Commission (FTC) [2].

As measures of market concentration, the best-known candidates are the \textit{m}-firm concentration ratio \( CR_m \), especially the 4-firm \( CR_4 \), and the \textit{Herfindahl-Hirschman Index} \( H \) after Herfindahl [3] and Hirschman [4] (see, e.g., [5], pp. 116–118; [6], pp. 97–101; [7], Ch. 8). The \( CR_m \) is defined as the combined market share of the \( m \) largest firms within the market whereas \( H \) equals the sum of all the squared market shares. Of the two measures, \( H \) appears to have become the generally preferred one in terms of its properties (e.g., [5], pp. 116–118; [7], Ch. 8, pp. 610–615). In terms of the merger guidelines of the U.S. DOJ and FTC [2], the earliest 1968 \textit{Merger Guidelines} utilized the \( CR_4 \) while the later guidelines, the most recent being the 2010 \textit{Horizontal Merger Guidelines}, have been using the \( H \) index as a screening tool for potential antitrust concerns raised by a proposed merger.

These two measures \( CR_m \) and \( H \) are sufficiently different that no precise functional relationship can exist between them. Nevertheless, it would be informative and potentially lead to approximate relationships if bounds and inequalities between the measures can be derived. Such work was done by Pautler [8], Kwoka [9], and Sleuwaegen et al. [10, 11], obtaining bounds on \( H \) in terms of \( CR_m \). Their work was done, at least in part, in response to the change in the U.S. merger guidelines, replacing the four-firm concentration ratio \( CR_4 \) with the \( H \) index. Results from actual market-share data showed that the absolute variation in values of \( H \) increased greatly with increasing \( CR_4 \).

Since these early explorations of potential \( H-\text{CR}_4 \) relationships, there appears to have been no reported attempt to verify, correct, or expand on these results. It is the purpose of the present paper to take another critical look at those earlier findings.
using a more rigorous and transparent approach, resulting in some corrections or modifications and alternative formulations. The analytic approach used is that of majorization theory [12, 13] supported by data from computer simulation generating random market-share distributions. Some real market-share data are also being used.

If the objective were to simply determine the “best” function to describe the relationship between $H$ and $CR_m$ (or vice versa), then some statistical model could be explored using regression analysis. Such analysis could be performed for real or simulated market share data. Kwoka [9] reported one such effort by relating the logarithm $\log CR_m$ linearly to $\log H$ for $m = 2$ and $m = 4$ and obtained quite a good fit to real market-share data. More recently, Pavic et al. [14] fitted real data to a model in which $CR_4$ is expressed as a power function of $H$. Those authors fitted market-share data at different levels of aggregation and obtained good model fits. By contrast, instead of using a function that aims to relate each value of $H$ to an approximate single value of $CR_m$ or vice versa, the approach used in the present paper uses majorization theory to develop bounds that can in turn be used to approximately relate one measure to another. This approach also provides tolerance or error limits within which the value of $H$ has to lie given any particular value of $CR_m$ and vice versa.

The majorization theory as used in this paper has become a well-established approach to a wide variety of problems and fields of study. Since the celebrated book by Marshall and Olkin [13], there has been a surge of interest in potential applications of majorization theory in a wide variety of fields. The more recent edition [12] provides a more up-to-date account of the broad spectrum of applications. One of the early applications was by economists interested in the measurement of income inequality, notably by the work of Dalton and Lorenz (see [12], Ch. 1). In fact, the notion of economic inequality is closely linked to majorization or the Lorenz order [15]. Other applications have been reviewed by Arnold [16] and some relate to as diverse fields as quantum mechanics [17] and statistical variation [18]. The theory is particularly useful for establishing inequalities and extreme values of functions of discrete distributions. This is precisely the reason for using majorization theory in the present paper involving market-share distributions.

Since $CR_m$, particularly $CR_4$, and $H$ are by far the most popular measures of market concentration and since market concentration is an indicator of competition, efficiency, and market power of firms within a market or industry, it is important to economists, policy makers and anyone with interests in such issues that any relationships between the two measures are accurate and reliable and based on an approach that is rigorous, complete, and explained in sufficient detail for verification. This has been the objective of the present paper.

While any relationship between $H$ and $CR_m$ can only be approximate, a more accurate estimation or prediction for $H$ may be possible from some of the largest individual market shares such as those on which $CR_m$ is based. Such a formulation would
be important since market-share data are frequently reported for some of the largest firms with the smaller firms being combined into an “other” category. This type of formulation is also being considered in this paper.

2. Theory

2.1. Some introductory definitions

In order to appreciate the logic behind majorization theory, some definitions and properties are needed. In particular and conceptually, a vector or distribution $X_n = (x_1, \ldots, x_n)$ is said to be majorized by a second vector (distribution) $Y_n = (y_1, \ldots, y_n)$ if the components of $X_n$ are “more nearly equal”, “more evenly distributed”, or “less concentrated” than are the components of $Y_n$. Formally, with $X_n$ and $Y_n$ ordered decreasingly as

$$x_1 \geq x_2 \geq \cdots \geq x_n, \quad y_1 \geq y_2 \geq \cdots \geq y_n$$

(1)

$X_n$ is majorized by $Y_n$, denoted by $X_n < Y_n$, under the following conditions:

$$X_n < Y_n \text{ if } \begin{cases} \sum_{i=1}^{j} x_i \leq \sum_{i=1}^{j} y_i, \quad j = 1, \ldots, n - 1 \\
\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i. \end{cases}$$

(2)

For example, if $X_4$ is a vector (distribution) such that $x_i \geq 0$ for all $i$ and $\sum_{i=1}^{n} x_i = 1$, then the following majorization applies:

$$\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) < (x_1, \ldots, x_n) < (1, 0, \ldots, 0)$$

([12], p. 9).

The concept of Schur-convexity of a function is a property that preserves the order of majorization. That is, a function $f$ is Schur-convex if its value increases as its arguments become increasingly uneven or concentrated. Formally, $f$ is Schur-convex if

$$X_n < Y_n \text{ implies } f(X_n) \leq f(Y_n)$$

(3)

and $f$ is strictly Schur-convex if the inequality in (3) is strict when $X_n$ is not a permutation of $Y_n$.

As a simple example, consider the two vectors or distributions $X_4 = (0.40, 0.30, 0.20, 0.10)$ and $Y_4 = (0.50, 0.30, 0.19, 0.01)$. It is readily seen from the definition in (2) that $X_4$ is majorized by $Y_4$. Thus, if $X_4$ and $Y_4$ were to represent two market-share distributions and if a measure of market concentration is Schur-convex, then it follows from (3) that $Y_4$ shows a higher degree of...
concentration than does $X_4$. This result would seem entirely reasonable from simply looking at the forms of the two distributions.

### 2.2. Bounds on $H$

Let $S_n = (s_1, \ldots, s_n)$ denote the set or distribution of market shares for some particular industry or market with $n$ firms where $\sum_{i=1}^{n} s_i = 1$ (or 100%). With the market shares decreasingly ordered as in (1), i.e.,

$$s_1 \geq s_2 \geq \cdots \geq s_n$$

the $m$-firm concentration ratio is defined as

$$CR_m = \sum_{i=1}^{m} s_i$$

and the Herfindahl-Hirschman index as

$$H = \sum_{i=1}^{n} s_i^2.$$  \hspace{1cm} (6)

An objective is then to determine bounds on $H$ in terms of $CR_m$, $m$, and $n$. A lower bound on $H$ is relatively simple to derive. In the expression

$$H = \sum_{i=1}^{n} s_i^2 = \sum_{i=1}^{m} s_i^2 + \sum_{i=m+1}^{n} s_i^2$$

each of the terms is strictly Schur-convex in their respective arguments ([12], pp. 138–139). Furthermore, the following majorizations exist:

$$\left(\frac{CR_m}{m}, \ldots, \frac{CR_m}{m}\right) < (s_1, \ldots, s_m), \quad \left(\frac{1 - CR_m}{n-m}, \ldots, \frac{1 - CR_m}{n-m}\right) < (s_{m+1}, \ldots, s_n)$$

which are rather apparent from (2) (see also [12], pp. 9, 21). It then follows from (3), (8), and the strict Schur-convexity of the terms in (7) that

$$H \geq \frac{CR_m^2}{m} + \frac{(1 - CR_m)^2}{n-m}$$

which become a lower bound on $H$.

Derivation of an upper bound on $H$ will be based in part on the following theorem given by Marshall et al. ([12], pp. 192–193) and attributed to Kemperman [19]: For a set of real-valued data $X_n = (x_1, \ldots, x_n)$ with $L \leq x_i \leq U$, $i = 1, \ldots, n$, there exists a unique integer $k \in [0, 1, \ldots, n]$ and a unique $\theta \in [L, U]$ where
\[ K - 1 \leq k < K, \quad K = \frac{\sum_{i=1}^{n} x_i - nL}{U - L} \]  

(10)

and

\[ \theta = \sum_{i=1}^{n} x_i - (n - k - 1)L - kU \]  

(11)

such that

\[ X_n < \left( U, \ldots, U, \theta, L, \ldots, L \right) \text{ k} \]

(12)

It is readily apparent from the proof given by Marshall et al. ([12], pp. 192–193) that this theorem also holds when \( L = \min \{x_1, \ldots, x_n\} \) and \( U = \max \{x_1, \ldots, x_n\} \).

Using this theorem and the fact that

\[ s_{m+1} \leq s_i \leq CR_m - (m - 1)s_{m+1}, \quad i = 1, \ldots, m \]

it follows from (10) and (11) that \( K = 1 \) and hence \( k = 0 \) and

\( \theta = CR_m - (m - 1)s_{m+1} \)

so that

\[ (s_1, \ldots, s_m) < (CR_m - (m - 1)s_{m+1}, s_{m+1}, \ldots, s_{m+1}) \]  

(13)

which is also given in Marshall and Olkin ([13], p. 133). From (13) and the (strict) Schur-convexity of \( \sum_{i=1}^{m} s_i^2 \), it follows that

\[ \sum_{i=1}^{m} s_i^2 \leq \left[ CR_m - (m - 1)s_{m+1} \right]^2 + (m - 1)s_{m+1}^2. \]  

(14)

Furthermore, since \( 0 \leq s_i \leq s_{m+1}, \quad i = m + 1, \ldots, n \), it is seen from (10), (11) and (12) that \( \theta = 1 - CR_m - ks_{m+1} \) and

\[ (s_{m+1}, \ldots, s_n) < \left( s_{m+1}, \ldots, s_{m+1}, 1 - CR_m - ks_{m+1}, 0, \ldots, 0 \right) \]  

(15)

which, together with the (strict) Schur-convexity of \( \sum_{i=m+1}^{n} s_i^2 \), result in

\[ \sum_{i=m+1}^{n} s_i^2 \leq ks_{m+1}^2 + (1 - CR_m - ks_{m+1})^2. \]  

(16)

Treating \( k \) as a continuous variable, it is clear that the bound in (16) is strictly convex in \( k \in [K - 1, K] \) where \( K = (1 - CR_m)/s_{m+1} \) from (10) with \( L = 0 \) and \( U = s_{m+1} \).

For this \( K \) and for both \( k = K - 1 \) and \( k = K \) (although \( k \) is strictly less than \( K \), the bound in (16) is seen to equal \( (1 - CR_m)s_{m+1} \) for both \( k \) - values, i.e.,
\[ \sum_{i=m+1}^{n} s_i^2 \leq (1 - CR_m)s_{m+1}. \] (17)

Thus, from (7), (14), and (17),

\[ H = \sum_{i=1}^{n} s_i^2 \leq U_H = [CR_m - (m - 1)s_{m+1}]^2 + (m - 1)s_{m+1}^2 + (1 - CR_m)s_{m+1}. \] (18)

The second-order partial derivative \( \frac{\partial^2 U_H}{\partial s_{m+1}^2} = 2m(m - 1) \) from (18) so that \( U_H \) is convex in \( s_{m+1} \) (strictly so if \( m > 1 \)). From this convexity and since

\[ \frac{1 - CR_m}{n - m} \leq s_{m+1} \leq \min \left\{ \frac{CR_m}{m}, 1 - CR_m \right\} \] (19)

so that, from (18) and (19),

\[
U_H = \begin{cases} 
U_{H1} = \left[ CR_m - (m - 1) \left( \frac{1 - CR_m}{n - m} \right) \right]^2 + (n - 1) \left( \frac{1 - CR_m}{n - m} \right)^2 & \text{for } s_{m+1} = \frac{1 - CR_m}{n - m} \\
U_{H2} = \frac{CR_m}{m} & \text{for } s_{m+1} = \frac{CR_m}{m} \\
U_{H3} = \left[ 1 - m(1 - CR_m) \right]^2 + m(1 - CR_m)^2 & \text{for } s_{m+1} = 1 - CR_m
\end{cases}
\] (20)

and hence

\[ H \leq \max \{ U_{H1}, \min \{ U_{H2}, U_{H3} \} \} \] (21)

which is then an upper bound on \( H \) in terms of any given \( m, n \), and \( CR_m \).

The bound in (21) is equivalent to one given by [10, 11] with two exceptions. First, their bound did not incorporate the term \( U_{H3} \) in (20). Second, their bound involved a potential fraction term that was indeterminable, but presumably sufficiently small that it could be ignored. Sleuwaegen et al. ([11], p. 628) presented an upper bound as being equivalent to the largest of \( U_{H1} \) and \( U_{H2} \) in (20), “except for a fraction if the maximum is \( C_k/k \) and \( k(1 - C_k)/C_k \) is non-integer”, with their \( C_k \) and \( k \) being equivalent to \( CR_m \) and \( m \) used in the present paper. See also ([10], pp. 206–207).

For a large number of firms \( n \) (strictly for \( n \to \infty \)), it is clear from (9) and (20), (21) that

\[ \frac{CR_m^2}{m} \leq H \leq \max \{ CR_m^2, \min \{ U_{H2}, U_{H3} \} \}. \] (22)
By (a) defining $V = U_{H2} - U_{H3}$, which is strictly concave in $CR_m$ for any given $m$ since $\partial^2 V/\partial CR_m^2 = -2m(m+1) < 0$, (b) setting $V = 0$, and (c) solving the resulting second-order equation for $CR_m$, it is found that
\[ U_{H2} \geq U_{H3} \text{ if, and only if, } \frac{m}{m^2(m+1)} \leq CR_m \leq \frac{m^3 + 1}{m^2(m+1)}. \tag{23} \]

Similarly, for $n \to \infty$ and $W = CR_m^2 - U_{H3}$, with $\partial^2 W/\partial CR_m^2 = 2(1 - m^2 - m) < 0$
\[ CR_m^2 \geq U_{H3} \text{ if, and only if, } \frac{m(m-1)+1}{m(m+1)-1} \leq CR_m \leq 1. \tag{24} \]

It is apparent from (22), (23) and (24) that the upper bound in (22) equals $U_{H3}$ only for $m = 1$ when, from (20), $U_{H3} = CR_1^2 + (1 - CR_1)^2$. This also becomes the bound in (21) for $m = 1$ and for $n$ not large since then, from (20), $U_{H1} = CR_1^2 + (1 - CR_1)^2/(n-1)$. Also, note that, from (20), $U_{H1} = U_{H3}$ when $n = m + 1$ for any $m \geq 1$. Consequently, it follows from (22), (23) and (24) that
\[ \frac{CR_m^2}{m} \leq H \leq \max \left\{ \frac{CR_m^2}{m}, \frac{CR_m}{m} \right\} = \begin{cases} CR_m^2 \text{ for } CR_m \geq \frac{1}{m}, m > 1, n \to \infty \\ CR_m \text{ for } CR_m \leq \frac{1}{m}, m > 1, n \to \infty. \end{cases} \tag{25} \]

These (asymptotic) bounds are equivalent to those given by [10, 11] with one exception: the upper bound in (25) when $CR_m < 1/m$ contains no indeterminate fraction term mentioned by those authors. Furthermore, the upper bound given by [8, 9] for $m = 4$ is $CR_4/4$ irrespective of the value of $CR_4$ and whether $n$ is finite or not. See also Martin ([20], p.337).

### 2.3. Some comments

It may perhaps be tempting to assume that for any given $CR_m$, $m$, and $n$, the upper bound on $H$ corresponds to the market-share distribution
\[ \left( 1 - (n-1) \left( \frac{1-CR_m}{n-m}, \frac{1-CR_m}{n-m}, \ldots, \frac{1-CR_m}{n-m} \right) \right) \tag{26} \]
(see, e.g., [10, 11]). However, this is not necessarily true because of (21) and the fact that, as can easily be verified, the value of $H$ for the distribution in (26) equals the $U_{H1}$ in (20).

As a counterexample, consider the market-share distribution $S_5 = (0.26, 0.25, 0.25, 0.23, 0.01)$ for which $H = 0.2456$. For $m = 2$, $CR_2 = 0.51$, and $n = 5$, the value of
$H$ for the corresponding distribution in (26) becomes 0.2269, which equals $U_{H1}$, but which is less than the $H = 0.2456$. Rather, the upper bound from (21) becomes $U_{H2} = 0.51/2 = 0.2550$.

While in the preceding counterexample, $CR_m > 1/m$, consider next the following example with $CR_m < 1/m$:

$$S_{27} = (0.10, 0.05, \ldots, 0.05, 0.01, \ldots, 0.01)$$

for which $H = 0.0510$. For $m = 3$, $CR_3 = 0.2000$, and $n = 27$, the value of $H$ for the corresponding distribution in (26) becomes 0.0467, which equals $U_{H1}$ in (20), but which is less than the $H = .0510$. In fact, the correct upper bound on $H$ from (21) is $U_{H2} = 0.2000/3 = 0.0667$.

For the case when all market shares are equal, i.e., $s_i = 1/n$ for $i = 1, \ldots, n$, it follows from (6) and (20) that $H = U_{H1} = U_{H2} = 1/n$. However, in this equal-share case, $U_{H3}$ in (20) equals $1/n$ if and only if $m = n - 1$; otherwise, $U_{H3} > 1/n$.

For the particular case when $m = 1$ so that the concentration ratio is simply equal to market share $s_1$ of the largest firm (see the order in (4)), it is readily seen from (20) that $U_{H1} \leq U_{H2}$, $U_{H1} < U_{H3}$, and $U_{H2} \leq U_{H3}$ for $p_1 \leq 1/2$ and $U_{H2} > U_{H3}$ for $p_1 > 1/2$. Therefore, in the $m = 1$ case, the upper bound on $H$ from (21) is either $U_{H2}$ or $U_{H3}$, depending upon which one is the smaller.

3. Results

3.1. Simulation results

Besides the above boundary comparisons between the $H$ and $CR_m$ in (5) and (6), an analysis has also been performed in terms of a scatter diagram of $H$ versus $CR_4$. The four-firm concentration ratio $CR_4$ was chosen since it is by far the most widely used one (e.g., [6], p. 97; [21], p. 255). Rather than using real data of limited sample size as done by [9, 10, 11], computer simulation was used to randomly generate a very large number of market-share distributions $S_n = (s_1, \ldots, s_n)$, with each $s_i$ and $n$ based on random number generation.

The algorithm developed was based on the following steps:

(1) Generate $n$ as a random integer such that $5 \leq n \leq 100$.

(2) For each $n$ - value generated in Step 1, generate $s_1, \ldots, s_{n-1}$ (to the desired number of decimal places) as random numbers within the following intervals:
\[
\frac{1}{n} \leq s_1 \leq 1
\]
\[
\frac{1 - s_1}{n - 1} \leq s_2 \leq \min\left\{s_1, 1 - s_1\right\}
\]
\[
\vdots
\]
\[
\frac{1 - \sum_{j=1}^{n-2} s_j}{n - (n - 2)} \leq s_{n-1} \leq \min\left\{s_{n-2}, 1 - \sum_{j=1}^{n-2} s_j\right\}
\]

that is,
\[
\frac{1 - \sum_{j=1}^{i-1} s_j}{n - (i - 1)} \leq s_i \leq \min\left\{s_{i-1}, 1 - \sum_{j=1}^{i-1} s_j\right\} \text{ for } i = 2, \ldots, n - 1
\]

(3) Compute \( s_n = 1 - \sum_{j=1}^{n-1} s_j \).

A total of 10,000 such \( S_n \) distributions were generated and the corresponding values for \( H \) and \( CR_4 \) were computed. The results are summarized in the scatter diagram in Fig. 1.

The solid curves in Fig. 1 represent the boundary conditions in (25) and appear to be entirely appropriate even though the number of firms \( n \) is finite \((5 \leq n \leq 100)\). Thus, the lower solid curve in Fig. 1 represents the lower bound on \( H \) as

\[
L_H = \frac{CR_4^2}{4}
\]

and upper solid curve represents the upper bound as

\[
U_H = \begin{cases} 
CR_4^2 & \text{for } CR_4 \geq \frac{1}{4} \\
\frac{CR_4}{4} & \text{for } CR_4 \leq \frac{1}{4}
\end{cases}
\]

(28)

It is clear from Fig. 1 that the variation in potential values of \( H \) for any given \( CR_m \) (with \( m = 4 \) in Fig. 1) tends to increase dramatically with increasing \( CR_m \). Most interesting is perhaps the systematic nature of the variation in \( H \)-values as restricted by the bounds \( L_H \) and \( U_H \) in (27) and (28). The form of the scatter graph in Fig. 1 is generally similar to results given by [8, 9, 10, 11], although their results are based on a limited number of data points. See also Pavic et al. [14].
Another interesting observation can be made with respect to the relative variation of \( H \)-values for varying \( CR_m \). Such variation can be measured in terms of the range \( U_H - L_H \) relative to the midrange defined by the estimated (predicted) \( H \) as

\[ H_1 = \frac{1}{2} (L_H + U_H) \]  

(29)

with \( L_H \) and \( U_H \) defined in (27) and (28). Thus, the relative variation \( RV = (U_H - L_H)/H_1 \) becomes \( 6/5 \) for \( CR_4 \geq \frac{1}{4} \) whereas \( RV = 2(1 - CR_4)/(1 + CR_4) \) for \( CR_4 \leq \frac{1}{4} \). That is, in spite of the considerable absolute variation in \( H \) for each value of \( CR_4 \) as seen from Fig. 1 when \( CR_4 \) is not small (\( CR_4 > \frac{1}{4} \)), the relative variation remains constant. While these results are based on \( m = 4 \), there seems no reason to suspect that the results would not hold generally for all \( m \).

Fig. 1. Comparison of values of \( CR_4 \) and \( H \) for 10,000 randomly generated market-share distributions \( S_n = (s_1, ..., s_n) \) with the number of firms \( n \) varying as a random integer between 5 and 100. The solid curves are the bounds in (25).
3.2. $H$ as function of $CR_m$

From the different expressions for $H$ and $CR_m$ in (5) and (6) and from the scatter graph in Fig. 1 and those in [8, 9, 10, 11], there cannot be any precise functional relationship between $H$ and $CR_m$ in general. However, a very approximate relationship can be formulated in terms of the lower bound $L_H$ in (9) and the upper bound $U_H$ in (21). The fact that the true value of $H$ for any given market-share distribution $S_n = (s_1, ..., s_n)$ falls within the interval $[L_H, U_H]$ can be expressed as

$$H = \frac{1}{2} (L_H + U_H) \pm \frac{1}{2} (U_H - L_H).$$

(30)

Note that the first term in (30) is the same as $H1$ in (29) (for $m = 4$) whereas the second one is a tolerance or error term.

If the total number of firms $n$ in a market or industry is known, then $L_H$, $U_H$, and (30) can be computed from (9) and (21). For very large $n$ and for $m > 1$, those computations can be done from (25) or from (27) and (28) for $m = 4$. Furthermore, when $n$ is not large, the bounds in (25) (and (27) and (28)) still apply. That is,

$$L_H = \frac{CR_m^2}{m} \leq H \leq U_H = \begin{cases} CR_m^2 & \text{for } CR_m \geq \frac{1}{m} \\ \frac{CR_m}{m} & \text{for } CR_m \leq \frac{1}{m} \end{cases}$$

(31)

for any $n$ and $m > 1$. This lower bound follows immediately from (9) by ignoring the term involving $n$ whereas the upper bound can be proved as follows. First, the inequality in (18) can be expressed as

$$U_H = CR_m^2 - s_{m+1} [(2m - 1)CR_m - m(m - 1)s_{m+1} - 1]$$

$$= CR_m^2 - s_{m+1}A$$

(32)

where $A$ is strictly decreasing in $s_{m+1}$. Second, for any given $m$ and $CR_m$, the maximum possible value of $s_{m+1}$ will necessarily occur when $s_i = CR_m/m$ for $i = 1, ..., m + 1$ in which case

$$A = mCR_m - 1.$$  

(33)

Then, from (33), $A \geq 0$ for $CR_m \geq 1/m$ which, together with (32), gives $U_H \leq CR_m^2$ for $CR_m \geq 1/m$. Finally, since the $U_{H2}$ in (20) is independent of $n$, this completes the proof of $U_H$ in (31).

Equivalent bounds can also be derived for $CR_m$ in terms of $H$. In fact, it follows immediately from (31) that
As in the case of (31), the bounds in (34) hold for any \( n \) and \( m > 1 \) (and not just when \( n \to \infty \)). Then, as in the case of \( H1 \) in (29) (for \( m = 4 \)) and \( H \) in (30), the following formulation applies to \( CR_m \):

\[
CR_m = CR^l_m \pm \frac{1}{2} (U_{CR_m} - L_{CR_m}), \quad CR^l_m = \frac{1}{2} (L_{CR_m} + U_{CR_m}).
\]  

In order to explore these and other formulations discussed subsequently, various market-share distributions were randomly generated using the computer algorithm described in Subsection 3.1 and with \( m = 4 \) and \( n \in [5, 100] \). The results are summarized in Table 1 for 25 different distributions. As an example of the computation involved for (29), (30) and (35), consider Data Set 1 in Table 1 with \( CR_4 = 0.2881 \) and \( H = 0.0459 \). From (27), (28) and (29), \( L_H = 0.0208, \quad U_H = 0.0830, \quad \) and \( H1 = 0.0519 \) so that, from (30), \( H = 0.0519 \pm 0.0311 \) or \([0.0208, 0.0830]\). That is, the value of \( H \) is roughly equal to 0.0519, but it has to lie in the interval between 0.0208 and 0.0830. The true value of \( H \) is 0.0459 in Table 1. The computations with respect to \( CR_4 \) are similarly done for (34) and (35) with \( m = 4 \).

It is evident from the results in Table 1 that \( H1 \) as a function of \( CR_4 \) in (29) and \( CR^l_4 \) as a function of \( H \) in (35) do provide some respectable indications of the values of the indices \( H \) and \( CR_4 \). Thus, knowing the value of one of these two indices, one can make a rough estimation (prediction) about the corresponding value of the other index.

A statistical approach to determining the relationship between \( H \) and \( CR_m \) would be the use of regression analysis using simulated data or real market-share data. Such real-data analysis has been reported by Kwoka [9] and Pavic et al. [14]. Based on the values of \( H \) and \( CR_4 \) in Table 1, the following regression model is obtained:

\[
H^2 = aCR^b_4, \quad a = 0.4055, \quad b = 1.3861.
\]  

However, the fit of this model to the data is unimpressive as seen from the values of \( H2 \) based on (36) as given in Table 1. The coefficient of determination, when properly computed [22], becomes \( R^2 = 1 - \sum \frac{(H - H2)^2}{\sum (H - \bar{H})^2} = 0.77 \) (where \( \bar{H} \) denotes the mean of the \( H \)-values). That is, only 77% of the variation of \( H \) (about its mean) is explained by the fitted model in (36).

By comparison with the results in (36), Kwoka [9] obtained the parameter estimates \( a = 0.315 \) and \( b = 1.724 \) for some real market-share data. Similarly, Pavic et al. [14] fitted a power function to real market data for different levels of aggregation or degrees of specificity of the commodity. Their results corresponded to \( a \)-values ranging
from about 0.70 to 0.93 and $b$-values ranging from 1.74 to 1.84 for the function in (36).

### 3.3. $H$ as function of $s_1$, ..., $s_m$ and $n$

The developments so far have been concerned with potential relationships between the Herfindahl-Hirschman Index $H$ and the $m$-firm concentration ratio $CR_m$ in (5) and (6). Consider next the case when the individual market shares $s_1$, ..., $s_m$ for the $m$ largest firms are known (and not just their sum $CR_m$) as well as the total number of firms $n$ in the market or industry. In this case, $m$ may be rather small, most typically $m = 4$, but $m$ could also be 3, 8, or 10, for example. Another situation

| Data Set | $n$ | $CR_4$ | $H$ | $H_1$ | $CR_4^1$ | $H_2$ | $H_3$ | $H_4$ |
|----------|-----|--------|-----|-------|----------|-------|-------|-------|
| 1        | 27  | 0.2881 | 0.0459 | 0.0519 | 0.3060 | 0.0723 | 0.0459 | 0.0402 |
| 2        | 80  | 0.0535 | 0.0125 | 0.0070 | 0.1368 | 0.0070 | 0.0125 | 0.0068 |
| 3        | 66  | 0.3628 | 0.0678 | 0.0823 | 0.3906 | 0.0995 | 0.0679 | 0.0689 |
| 4        | 60  | 0.1210 | 0.0181 | 0.0170 | 0.1707 | 0.0217 | 0.0182 | 0.0120 |
| 5        | 81  | 0.0652 | 0.0126 | 0.0092 | 0.1374 | 0.0092 | 0.0125 | 0.0069 |
| 6        | 16  | 0.9666 | 0.5764 | 0.5839 | 1.1388 | 0.3868 | 0.5765 | 0.5767 |
| 7        | 31  | 0.3670 | 0.0805 | 0.0842 | 0.4256 | 0.1011 | 0.0809 | 0.0736 |
| 8        | 28  | 0.2584 | 0.0520 | 0.0417 | 0.3320 | 0.0621 | 0.0523 | 0.0406 |
| 9        | 65  | 0.5115 | 0.1708 | 0.1635 | 0.6199 | 0.1601 | 0.1709 | 0.1693 |
| 10       | 75  | 0.0566 | 0.0133 | 0.0075 | 0.1419 | 0.0076 | 0.0133 | 0.0073 |
| 11       | 22  | 0.1898 | 0.0455 | 0.0282 | 0.3043 | 0.0405 | 0.0455 | 0.0273 |
| 12       | 52  | 0.1051 | 0.0196 | 0.0176 | 0.1792 | 0.0179 | 0.0196 | 0.0118 |
| 13       | 30  | 0.4724 | 0.1553 | 0.1395 | 0.5911 | 0.1434 | 0.1558 | 0.1509 |
| 14       | 84  | 0.1085 | 0.0158 | 0.0150 | 0.1573 | 0.0187 | 0.0157 | 0.0109 |
| 15       | 38  | 0.8513 | 0.4773 | 0.4529 | 1.0363 | 0.3244 | 0.4775 | 0.4779 |
| 16       | 74  | 0.1045 | 0.0146 | 0.0144 | 0.1500 | 0.0177 | 0.0147 | 0.0094 |
| 17       | 25  | 0.2131 | 0.0418 | 0.0323 | 0.2881 | 0.0476 | 0.0419 | 0.0289 |
| 18       | 93  | 0.8406 | 0.6755 | 0.4416 | 1.2328 | 0.3188 | 0.6756 | 0.6754 |
| 19       | 85  | 0.0968 | 0.0140 | 0.0133 | 0.1463 | 0.0159 | 0.0140 | 0.0091 |
| 20       | 98  | 0.1340 | 0.0145 | 0.0190 | 0.1494 | 0.0250 | 0.0146 | 0.0114 |
| 21       | 59  | 0.0751 | 0.0170 | 0.0101 | 0.1644 | 0.0112 | 0.0170 | 0.0093 |
| 22       | 62  | 0.1324 | 0.0184 | 0.0187 | 0.1724 | 0.0246 | 0.0185 | 0.0123 |
| 23       | 91  | 0.0604 | 0.0112 | 0.0080 | 0.1282 | 0.0083 | 0.0112 | 0.0066 |
| 24       | 79  | 0.1646 | 0.0174 | 0.0240 | 0.1667 | 0.0333 | 0.0175 | 0.0141 |
| 25       | 62  | 0.2096 | 0.0271 | 0.0317 | 0.2188 | 0.0465 | 0.0272 | 0.0218 |
may be one in which some of the smaller market shares are simply excluded from the computation of $H$ since their effect on $H$ is relatively low. Whatever the case may be, it would be of interest to determine if $H = \sum_{i=1}^{n} s_i^2$ could reasonably be approximated by some function of $s_1, \ldots, s_m$ and $n$. This was briefly considered by [11].

From the Schur-convexity of $\sum_{i=m+1}^{n} s_i^2$ and the majorization \( \left( \frac{1-CR_m}{n-m}, \ldots, \frac{1-CR_m}{n-m} \right) < (s_{m+1}, \ldots, s_n) \) ([12], pp. 21, 138–139) and from (17), it follows that

$$A = \left( \frac{1 - CR_m}{n-m} \right)^2 \leq \sum_{i=m+1}^{n} s_i^2 \leq \left( 1 - CR_m \right)s_m = B. \quad (37)$$

Then, from (7) and (37), it would seem reasonable to consider a measure approximating $H$ as being $\sum_{i=1}^{m} s_i^2$ plus a mean of the bounds $A$ and $B$. The most obvious mean would be the simple arithmetic mean, resulting in the measure

$$H' = \sum_{i=1}^{m} s_i^2 + \frac{1}{2}(A + B) \quad (38)$$

as being one whose values would approximate those of $H$. However, this measure is not Schur-convex as seen from the fact that

$$\frac{\partial H'}{\partial s_{m-1}} - \frac{\partial H'}{\partial s_m} = 2(s_{m-1} - s_m) - \frac{1}{2}(1 - CR_m)$$

which is not necessarily nonnegative.

An alternative formulation can be considered in terms of $\sum_{i=1}^{m} s_i^2$ plus a weighted arithmetic mean of the bounds $A$ and $B$ in (37) with $s_m \leq CR_m/m$. That is,

$$H_w = \sum_{i=1}^{m} s_i^2 + w \left( \frac{1 - CR_m}{n-m} \right)^2 + \left( 1 - w \right) \left( 1 - CR_m \right) \frac{CR_m}{m}, \quad w \in [0, 1]. \quad (39)$$

In order to explore the Schur-convexity of $H_w$, it follows from (39) that

$$\Delta_i = \frac{\partial H_w}{\partial s_i} - \frac{\partial H_w}{\partial s_{i+1}} = \begin{cases} 2(s_i - s_{i+1}), & i = 1, \ldots, m - 1 \\ \Delta_m, & i = m \\ 0, & i = m + 1, \ldots, n - 1 \end{cases} \quad (40)$$

where

$$\Delta_m = 2s_m + \left( \frac{2 \left( 1 - CR_m \right)}{m(n-m)} \right) \left[ n(1-w) - m \right] \frac{1 - w}{m}.$$ \quad (41)

Since this $\Delta_m$ in (40) and (41) is not necessarily nonnegative, the $H_w$ cannot be Schur-convex ([12], p. 84). From exploratory data analysis, it becomes readily apparent from various market-share distributions that values of $\sum_{i=1}^{m} s_i^2 + (1 - CR_m)^2/(n-m)$ are generally
much closer to the corresponding values of \( H \) than are the values of \( \sum_{i=1}^{m} s_i^2 + \frac{(1 - CR_m)CR_m}{m} \). This observation can be accounted for by choosing a large value of \( w \) in (39) such as \( w = 1 - 1/m(n - m) \), resulting in

\[
H3 = \sum_{i=1}^{m} s_i^2 + \frac{(1 - CR_m)^2}{n - m} + \frac{(1 - CR_m)(nCR_m - m)}{m^2(n - m)^2}.
\]  

(42)

In order to explore how accurately values of \( H3 \) from (42) approximate those of \( H \), the randomly generated market-share distributions described in Subsection 3.1 were used to compute \( H, CR_4 \), and \( H3 \) from (42) with \( m = 4 \). Again, the results are given in Table 1.

It is rather striking from the results in Table 1 how closely the values of \( H3 \) agree with those of \( H \). Their slight differences occur only in the fourth decimal place, with 0.0005 for Data Set 13 being the largest difference. In fact, if values of \( H \) are predicted to equal those of \( H3 \), it is found from the results in Table 1 that the coefficient of determination, when properly computed [22], becomes

\[ R^2 = 1 - \frac{\sum (H - H3)^2}{\sum (H - \overline{H})^2} = 1.0000 \text{ (rounded off to four decimal places)}. \]

That is, nearly all of the variation of \( H \) (about its mean \( \overline{H} \)) is explained (accounted for) by the fitted model stating that the predicted \( H \) equals \( H3 \).

### 3.4. \( H \) as function of \( s_1, \ldots, s_m \)

The \( H3 \) in (42) requires knowledge of the market shares of the \( m \) largest firms within a market or industry as well as the total number of firms \( n \). There may, of course, be cases when \( s_1, \ldots, s_m \) are reported, but \( n \) is not reported or the true value of \( n \) is not available since the market shares of an unknown number of firms have been combined into “others.” It would therefore be desirable to have an alternative measure that approximates \( H \), but that depends only on \( s_1, \ldots, s_m \).

Such a potential measure could again be considered in terms of the bounds in (37) except for setting the lower bound \( A = 0 \) to eliminate the dependence on \( n \). When approximating \( \sum_{i=m+1}^{m} s_i^2 \) with the arithmetic mean of \( A \) and \( B \) in (37), the elimination of \( A \) is partly compensated for by using \( s_m \) instead of \( s_{m+1} \) in the bound in (17).

Therefore, a new measure can be defined as

\[
H4 = \sum_{i=1}^{m} s_i^2 + \frac{1}{2}(1 - CR_m)s_m
\]  

(43)

which is only a function of \( s_1, \ldots, s_m \). Furthermore, in terms of tolerance (error) intervals, the true value of \( H \) can be expressed as

\[
H = H4 \pm \frac{1}{2}(1 - CR_m)s_m.
\]  

(44)
Thus, for example, for the market-share distribution $(0.20, 0.15, 0.15, 0.10, \ldots)$ and for $m = 4$, $CR_4 = 0.40$ and $s_4 = 0.10$ so that, from (43), $H_4 = 0.115$. Then, from (44), $H = 0.115 \pm 0.020$ so that the true value of $H$ has to fall within the interval $[0.095, 0.135]$.

The $H_4$ is not, however, Schur-convex as seen from the following differences between partial derivatives (using the descending order in (1)):

$$\Delta_i = \frac{\partial H_4}{\partial s_{i-1}} - \frac{\partial H_4}{\partial s_i} = \begin{cases} 2(s_{i-1} - s_i), & i = 2, \ldots, m - 1 \\ 2(s_{i-1} - s_i) - (1 - CR_m)/2, & i = m \\ (3/2)s_m + (1 - CR_m)/2, & i = m + 1 \\ 0, & i = m + 2, \ldots, n. \end{cases}$$

(45)

The terms in (45) are nonnegative, indicating Schur-convexity, but with one exception: $i = m$. In this case, $\Delta_m$ can potentially take on negative values. To place this relatively minor limitation of $H_4$ within some context, consider the following implication from the above results: among the $n(n - 1)/2$ possible transfers of small amount of market shares from larger to smaller firms, only $m - 1$ such transfers to the $m$-th largest firm could potentially lead to a slight increase in the value of $H_4$. All other possible transfers could only lead to decreasing or nonincreasing $H_4$. In most real situations, cases with $\Delta_m < 0$ in (45) would probably be rather exceptional so that $H_4$ can still be considered as a reasonable measure for all practical purposes.

For the randomly generated market-share distributions behind the data in Table 1, the values of $H_4$ were also computed as shown in the table. The results show that the values of $H_4$ tend to approximate quite closely those of $H$ and nearly as closely as those of $H_3$. From the data in Table 1, the coefficient of determination is $R^2 = 1 - \sum (H - H_4)^2 / \sum (H - \bar{H})^2 = 0.9985$ for the fitted model $\hat{H} = H_4$. There would seem to be little disadvantage in using $H_4$ instead of $H_3$ in (42), which also incorporates the number of firms $n$ in a market (industry). When using two decimal places, which is clearly adequate for practical purposes, the values of $H_3$ and $H_4$ may differ by only about $\pm 0.01$. These results are based on the most common choice of $m = 4$ and would probably differ somewhat for different $m$.

### 3.5. Results from real market data

The numerical results have so far been based on market-share distributions that have been randomly generated. Such results have the broadest possible implication without any particular bias or restrictions. By comparison, results from using real market-share data may depend on factors such as the market classification system used and the level of aggregation.

Nevertheless, it would be of interest to subject the above developments to some real data and compare the results with those from computer generated random samples.
Therefore, readily accessible market-share data from a wide diversity of markets were used to determine the true values of $CR_4$ and $H$ in (5) and (6) and their approximate values from $H1$ in (29), $CR_4^1$ in (35), and $H3$ and $H4$ in (42) and (43). The results are given in Table 2. While the results in Table 1 are given with four decimal places, primarily to illustrate the accuracy with which $H3$ estimates (predicts) $H$ from (42), two decimal places are used in Table 2 since this is adequate for all practical purposes.

It is clear from the values of $CR_4$, $H$, $H1$, and $CR_4^1$ in Table 2 that the estimation (prediction) of $H$ from any given $CR_4$ and vice versa is subject to substantial inaccuracy since the values of $CR_4^1$ differ considerably from those of $CR_4$ and similarly for $H1$ versus $H$ (by up to a multiplicative factor of 2). These results differ significantly from those of Table 1 probably because the values of $CR_4$ and $H$ are generally much larger in Table 2 than in Table 1. The relatively small values of $CR_4$ and $H$ in Table 1 are partly due to the rather large values of $n$ that affect the generated market-share distributions.

The estimated (predicted) values of $H$ based on $CR_m$, and those of $CR_m$ based on $H$, given in Table 2 could probably be improved upon by considering different weights for the two pairs of bounds. Thus, instead of $H1$ in (29), one could consider the bounds in (31) for $CR_m \geq 1/m$ and define

**Table 2.** Values of $CR_4$, $H$, $H1$, $H1^*$, $CR_4^1$, $CR_4^{1*}$, $H3$, and $H4$ in (5), (6), (29), (47), (35), (48), (42), and (43) for some real market-share data.

| n  | $CR_4$ | $H$  | $H1$ | $H1^*$ | $CR_4^1$ | $CR_4^{1*}$ | $H3$  | $H4$  | Source (Market type) |
|----|--------|------|------|--------|----------|------------|------|------|---------------------|
| 16 | 0.50   | 0.10 | 0.16 | 0.08   | 0.47      | 0.55       | 0.09 | 0.09 | [23] (Airline travel) |
| 16 | 0.60   | 0.12 | 0.23 | 0.12   | 0.52      | 0.61       | 0.12 | 0.12 | [23] (Airline travel) |
| 8  | 0.75   | 0.16 | 0.35 | 0.18   | 0.60      | 0.70       | 0.17 | 0.17 | [24] (U.S. distilled liquor) |
| 10 | 0.64   | 0.14 | 0.26 | 0.13   | 0.56      | 0.65       | 0.14 | 0.13 | [25] (Paints, coatings) |
| 10 | 0.52   | 0.11 | 0.17 | 0.09   | 0.50      | 0.58       | 0.11 | 0.10 | [26] (Pharmaceuticals) |
| 15 | 0.54   | 0.10 | 0.18 | 0.09   | 0.47      | 0.55       | 0.10 | 0.10 | [27] (Insurance companies) |
| 12 | 0.69   | 0.18 | 0.30 | 0.15   | 0.64      | 0.74       | 0.19 | 0.18 | [28] (Weapons exporters) |
| 30 | 0.34   | 0.05 | 0.07 | 0.04   | 0.32      | 0.39       | 0.05 | 0.05 | [29] (Car sales, Britain) |
| 12 | 0.60   | 0.12 | 0.23 | 0.12   | 0.52      | 0.61       | 0.11 | 0.11 | [30] (Auto manufacturers, US) |
| 8  | 0.77   | 0.18 | 0.37 | 0.19   | 0.64      | 0.74       | 0.18 | 0.17 | [25] (Craft beer, US) |
| 9  | 0.75   | 0.16 | 0.35 | 0.18   | 0.60      | 0.70       | 0.16 | 0.16 | [25] (Running shoe sales) |
| 10 | 0.72   | 0.17 | 0.32 | 0.17   | 0.62      | 0.72       | 0.16 | 0.17 | [25] (Top charter airlines) |
| 10 | 0.68   | 0.18 | 0.29 | 0.15   | 0.64      | 0.74       | 0.16 | 0.17 | [25] (Farm machinery, equip.) |
| 20 | 0.48   | 0.08 | 0.14 | 0.07   | 0.42      | 0.49       | 0.07 | 0.09 | [31] (Global car sales) |
| 10 | 0.60   | 0.12 | 0.23 | 0.12   | 0.52      | 0.61       | 0.12 | 0.12 | [25] (Top airlines worldwide) |
\[ H_1^* = wL + (1 - w)U_m = CR_m^2 \left[ 1 - w \left( 1 - \frac{1}{m} \right) \right], \quad 0 \leq w \leq 1. \quad (46) \]

By exploring different values of \( w \) from the data in Table 2 with \( m = 4 \), \( w = 0.9 \) becomes appropriate so that, from (46),

\[ H_1^* = \left( \frac{1.30}{4} \right) CR_4^2 \quad \text{for} \quad CR_4 \geq \frac{1}{4}. \quad (47) \]

Then, setting \( H_1^* = H \) and solving for \( CR_4 \) gives

\[ CR_4^{1*} = 1.75 \sqrt{H} \quad \text{for} \quad H \geq 0.02. \quad (48) \]

From the results in Table 2, it is seen that the values of \( H_1^* \) and \( CR_4^{1*} \) from (47) and (48) are considerably closer to the values of \( H \) and \( CR_4 \) than are those of \( H_1 \) and \( CR_4 \).

However, with respect to the measures \( H3 \) in (42) and \( H4 \) in (43) and the index \( H \), the results in the two tables are quite comparable. Thus, from the results in Table 2, it is seen that the values of \( H3 \) and \( H4 \) provide close approximation to those of \( H \) just as is the case with Table 1. Even though \( H3 \) incorporates the additional variable \( n \) (the total number of firms in a market), the values of \( H4 \) are about as close to those of \( H \) as are the \( H3 \)-values.

**4. Conclusions**

This careful and detailed re-examination of the boundary relationships between the Herfindahl-Hirschman index \( H \) and the \( m \)-firm concentration ratio \( CR_m \) is important in a number of regards. The various derivations and intermediate steps are verifiable and based on a rigorous approach using majorization theory, resulting in some new findings and revisions of earlier results. Since (a) market (industrial) concentration is generally considered to be an indicator of the competitiveness, efficiency, or power of a market (or industry), (b) it is convenient and useful to be able to measure this market property, and (c) the \( H \) and \( CR_m \) (especially \( CR_4 \)) are the most widely used measures, it is essential that the formulations of any \( H-CR_m \) relationships be correct, clear, and reliable. This requirement is clearly important to economists, policy planners, and others using such summary measures for any purpose.

Although there is clearly a relationship between the two indices \( CR_m \) and \( H \), this paper is emphasizing the very approximate nature of the relationship based on bounds using majorization theory. If only the values of \( CR_m \) or \( H \) are known, without knowledge of any of the individual market shares \( s_1, \ldots, s_n \) or the total number of firms in the market or industry \( n \), then their approximate relationships are given by (27), (28) and (29) for \( m = 4 \) and more generally by (30) and (31) as well as...
by (34) and (35), depending upon which is considered the dependent and which the independent (explanatory) variable.

The relationship in (34) and (35) is based on the bounds on CR$_m$ for any $m > 1$. Alternatively, one could derive CR$_m$ as the inverse function of $H = (L_H + U_H)/2$ in (31), with (27), (28) and (29) being the particular, but most common, case of $m = 4$. This inverse procedure yields

$$CR_m = \begin{cases} \sqrt{\frac{2mH}{m + 1}} & \text{for } H \geq \frac{m + 1}{2m^3} \\ \frac{\sqrt{1 + 8mH} - 1}{2} & \text{for } H \leq \frac{m + 1}{2m^3} \end{cases}$$

(49)

However, this relationship does not necessarily produce the same results as (34) and (35) and differ from (48).

As an example of the approximate nature of the conversion from one of the indices $H$ and $CR_m$ to the other, it would be of interest to consider the most recent 2010 Horizontal Merger Guidelines [2] that uses $H$ as the concentration index. Those guidelines use $H = 0.15$ and $H = 0.25$ as two of the benchmarks. From (35) with $m = 4$, the equivalent values of $CR_4$ would be, respectively, $0.58 \pm 0.19$ and $0.75 \pm 0.25$. If individual values of $CR_4$ are computed from (35) and (36) and (48) and (49) for $H = 0.15$ and $H = 0.25$, those values of $CR_4$, which differ considerably, are found to have mean values of about $0.55$ for $H = 0.15$ and $0.75$ for $H = 0.25$. Thus, in terms of those mean values, and from the 2010 Guidelines, the following equivalence applies:

Unconcentrated Markets: $H < 0.15$ or $CR_4 < 0.55$

Moderately Concentrated Markets: $0.15 \leq H \leq 0.25$ or $0.55 \leq CR_4 \leq 0.75$

Highly Concentrated Markets: $H > 0.25$ or $CR_4 > 0.75$.

While any relationship between $H$ and $CR_m$ can only provide a rough approximation, the estimation (prediction) of $H$ becomes rapidly more accurate when given some of the largest market shares $s_1, \ldots, s_m$. An interesting observation is the fact that knowledge of the size of the market (industry) $n$ does not generally have any substantial effect on such estimation accuracy. That is, the measure $H_4$ in (43) tends to be as close to the real measure $H$ as is $H_3$ in (42) as exemplified by the data in Tables 1 and 2 for $m = 4$.

Such support for the $H_4$ in (43) is important since market-share data are often reported with the smaller market shares grouped into an “other” category without any specification of $n$. In some cases, an “other” category may account for more than 50% of the market shares without any indication of the number of firms included in that category. Furthermore, $H_4$ together with its tolerance (error) limits
in (44) provide a complete description of the true value of the index $H$. The $H4$ also has the advantage of having the zero-indifference property, i.e., introducing a firm with zero market share does not affect $H4$. In spite of the fact that $H4$ is not Schur-convex, it does indeed appear to provide good approximation to $H$.

All the proofs and derivations in this paper involving the $m$-firm concentration ratio $CR_m$ are done for any $m$ rather than for some specific $m$-value. Whenever a particular $m$ is used, the discussion involves $m = 4$ since, as pointed out earlier, the 4-firm concentration ratio is the one used most frequently. If any other $m$ were to be of particular interest, such as $m = 2$ or $m = 8$, the analysis would simply require such substitution into the appropriate equations. Some of the numerical results may, of course, differ depending upon $m$.

A concluding comment is also warranted about the comparison between $CR_m$ and $H$. From majorization theory as commented on in Subsection 2.1, it follows that if two market-share distributions $S_n = (s_1, ..., s_n)$ and $R_n = (r_1, ..., r_n)$ are comparable with respect to majorization (i.e., $S_n$ and $R_n$ can be compared as in (2)), then $H$ and $CR_m$ provide the same size (order) comparison. That is, if $H(S_n) < H(R_n)$, it is implied that $CR_m(S_n) \leq CR_m(R_n)$ and vice versa. This follows from the fact that $CR_m$ is Schur-convex and $H$ is strictly Schur-convex. If one has to choose between the two concentration measures, and if the market shares are known for all the firms within a particular market (industry), it should be noted that $H$ has the advantage of strict Schur-convexity and of incorporating all available information about the market shares. If the market shares $s_1, ..., s_m$ of the $m$ largest firms, but not the market size $n$, are known, a good compromise would seem to be $H4$ in (43) and (44), with $H4$ being Schur-convex and strictly Schur-convex in $s_1, ..., s_m$.

**Declarations**

**Author contribution statement**

Tarald O. Kvalseth: Conceived and designed the analysis; Analyzed and interpreted the data; Contributed analysis tools or data; Wrote the paper.

**Funding statement**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Competing interest statement**

The authors declare no conflict of interest.
Additional information

No additional information is available for this paper.

Acknowledgements

The author wants to thank the three reviewers for their helpful and constructive comments.

References

[1] OECD (Organization for Economic Co-operation and Development), OECD Glossary of Statistical Terms: Concentration, 2003. URL https://stats.oecd.org/glossary/detail.asp?ID=3165.

[2] U.S. Department of Justice and the Federal Trade Commission, Horizontal Merger Guidelines, 2010.

[3] O.C. Herfindahl, Concentration in the Steel Industry, Unpublished Ph.D. dissertation, Columbia University, U.S.A., 1950.

[4] A.O. Hirschman, National Power and the Structure of Foreign Trade, University of California Press, Berkeley, CA, 1945.

[5] P.A. Gaughan, Mergers, Acquisitions, and Corporate Restructuring, fifth ed., Wiley, Hoboken, NJ, 2011.

[6] B. Andreosso, D. Jacobson, Industrial Economics and Organization: a European Perspective, second ed., McGraw-Hill, London, 2005.

[7] V.J. Tremblay, C. Horton Tremblay, New Perspectives on Industrial Organization, Springer, New York, 2012.

[8] P.A. Pautler, A guide to the Herfindahl index for antitrust attorneys, Res. Law Econ. 5 (1983) 167–190.

[9] J.E. Kwoka Jr., The Herfindahl in theory and practice, Antitrust Bull. 30 (winter) (1985) 915–947.

[10] L. Sleuwaegen, W. Dehandschutter, The critical choice between the concentration ratio and the $H$-index in assessing performance, J. Ind. Econ. XXXV (1986) 193–208.

[11] L.E. Sleuwaegen, R.R. DeBondt, W.V. Dehandschutter, The Herfindahl index and concentration ratio revisited, Antitrust Bull. 34 (fall) (1989) 625–640.

[12] A.W. Marshall, I. Olkin, B.C. Arnold, Inequalities: Theory of Majorization and its Application, second ed., Springer, New York, 2011.
[13] A.W. Marshall, I. Olkin, Inequalities: Theory of Majorization and its Applications, Academic Press, San Diego, CA, 1979.

[14] I. Pavic, F. Galetic, D. Piplica, Similarities and differences between the CR and HHI as an indicator of market concentration and market power, Br. J. Econ. Manag. Trade 13 (1) (2016) 1–8.

[15] M. Ibragimov, R. Ibragimov, Market demand elasticity and income inequality, Econ. Theor. 32 (2007) 579–587.

[16] B.C. Arnold, Majorization: here, there and everywhere, Stat. Sci. 22 (2007) 407–413.

[17] M.A. Nielsen, G. Vidal, Majorization and the interconversion of bipartite states, Quant. Inf. Comput. 1 (2001) 76–93.

[18] T.O. Kvålseth, Bounds on sample variation measures based on majorization, Commun. Stat. Theor. Methods 44 (2015) 3375–3386.

[19] J.H.B. Kemperman, Moment problems for sampling without replacement. I, II, III, Nederl. Akad. Wetensch. Prac. Ser. A 76 (=Indag. Math. 35) (1973), 149 – 164, 165 – 180, and 181 – 188. [MR 49(1975(9997a,b,c; Zbl. 266(1974) 62006, 62007, 62008] (1973).

[20] S. Martin, Advanced Industrial Economics, second ed., Blackwell, Oxford, U.K, 2002.

[21] D.W. Carlton, J.M. Perloff, Modern Industrial Organization, fourth ed., Addison-Wesley, Boston, MA, 2005.

[22] T.O. Kvålseth, Cautionary note about $R^2$, Am. Statistician 39 (1985) 279–285.

[23] M.G. Lijesen, P. Nijkamp, P. Rietveld, Measuring competition in civil aviation, J. Air Transport. Manag. 8 (2002) 189–197.

[24] J.S. Bain, Industrial Organization, Wiley, New York, 1959.

[25] Market Share Reporter, twenty seventh ed., Gale, 2017.

[26] Editorial, New 2016 data and statistics for global pharmaceutical products and projections through 2017, ACS Chem. Neurosci. 8 (2017) 1635–1636.

[27] Statista, Market Share of the Leading Insurance Companies in Belgium as of 2016, 2018. https://www.statista.com/statistics/780454/market-share-leading-insurance-companies-belgium/. (Accessed 24 June 2018).

[28] Statista, Market Share of the Leading Exporters of Major Weapons between 2013 and 2017, by Country, 2018. https://www.statista.com/statistics/
[29] SMMT- The Society of Motor Manufacturers and Traders, Best-selling Car Marques in Britain in 2018 (Q1), 2018. https://www.best-selling-cars.com/britain-uk/2018-q1-britain-best-selling-car-brands-and-models/. (Accessed 24 June 2018).

[30] Statista, U.S. Market Share of Selected Automobile Manufacturers 2013, 2014. https://www.statista.com/statistics/249375/us-market-share-of-selected-automobile-manufacturers/. (Accessed 12 September 2014).

[31] Carsalesbase.com, Global Car Sales Analysis 2017-Q1, 2017. http://carsalesbase.com/global-car-sales-2017-q1/. (Accessed 25 June 2018).