RESEARCH ARTICLE

Measurement of market (industry) concentration based on value validity

Tarald O. Kvålseth

Department of Mechanical Engineering and Department of Industrial & Systems Engineering, University of Minnesota, Minneapolis, MN, United States of America

kvals001@umn.edu

Abstract

As measures of concentration, especially for market (industry) concentration based on market shares, a variety of different measures or indices have been proposed. However, the various indices, including the two most widely used ones, the concentration ratio and the Herfindahl-Hirschman index (HHI), lack an important property: the value-validity property. An alternative index with this and other desirable properties is introduced. The new index makes it permissible to properly assess the extent of the concentration and make order and difference comparisons between index values as being true representations of the real concentration characteristic (attribute). Computer simulation data and real market-share data are used in the analysis. It is shown that the new index has a close functional relationship with the HHI index and has a firm theoretical relationship with market power as measured by the price-cost margin. Corresponding modifications to existing merger guidelines are presented.

1. Introduction

When dealing with nominal data, as with all types of data, it is often of interest to use some summary measures to represent certain characteristics or attributes associated with the data. One such characteristic is the concentration reflected by the probabilities or proportions \( p_1, \ldots, p_n \) of \( n \) mutually exclusive and exhaustive categories or events. Concentration is then considered high when one or a few of the \( p_i \)'s are relatively large and decreases as the \( p_i \)'s become increasingly equal. The converse characteristic, qualitative variation (evenness), is measured by indices whose values increase as the \( p_i \)'s become increasingly even, equal, or uniform (e.g., \([1, 2]\)).

One area in which the measurement of concentration has a long history is that of measuring market concentration or industrial concentration where the \( p_i \)'s are the market shares of the \( n \) firms within a market or industry (e.g., \([3, \text{Ch. 4}, 4, 5, \text{pp. 221–238}], [6, \text{Ch. 8}, 7–9]\)). Increasing market concentration tends to decrease competition and efficiency and increase market power. Any such trends are of concern to the business community and are being monitored by the U.S. Department of Justice (DOJ) and the Federal Trade Commission (FTC) in the case of antitrust. While such potentially negative consequences from increasing market concentration is the generally accepted proposition, it does not necessarily apply...
A number of different indices or measures of market (industry) concentration have been proposed over the years as outlined in this paper. However, two such indices have become by far the most popular ones: the \textit{m-firm concentration ratio} and the \textit{Herfindahl-Hirschman index HHI} (e.g., [13, pp. 116–118], [14, pp. 97–101], [6, Ch. 8], [15]). The \textit{m-firm concentration ratio (CR}_m) is simply defined as the sum of the market shares of the \textit{m} largest firms in a market, with the 4-firm \textit{CR}_4 being the most commonly used member of the \textit{CR}_m -family [16, p. 255]. The \textit{HHI}, after Herfindahl [17] and Hirschman [18], includes all market shares and is defined as the sum of their squares.

This paper points out a serious limitation of the various concentration indices, including \textit{CR}_m and \textit{HHI}, that is specifically related to the numerical values taken on by the indices for different market-share distributions. While those indices have some highly desirable mathematical properties, no concern seems to have been raised so far about an important question concerning the actual values of an index: does a concentration index \textit{C} take on numerical values throughout its range that can be rigorously explained or justified as providing realistic, true, or valid representations of the concentration characteristic or attribute? That is, does \textit{C} have value validity?

A reason for this concern is the fact that different indices with similar mathematical properties can produce wildly different values for the same data sets. Comparisons between index values can therefore result in potentially misleading and unreliable interpretations, findings, and conclusions. What is needed of a concentration index \textit{C} is an additional property requirement and its basic condition to provide \textit{C} with value validity. This is the focus of this paper.

This paper is organized as follows. First, the set of properties generally required of a concentration index \textit{C} are defined and discussed, followed by a definition and discussion of the \textit{value-validity property} of \textit{C}. Second, a concise historical account is given of the various indices that have been proposed over the years and their lacking properties are identified, with particular emphasis on the two indices \textit{CR}_4 and \textit{HHI}. Third, since existing indices lack the value-validity property, a new index \textit{C}_K with this and other necessary properties is introduced and discussed as an extension of the author’s earlier and preliminary effort to measure the homogeneity of categorical data [20]. Fourth, \textit{C}_K is compared with other indices, notably \textit{CR}_4 and more so \textit{HHI}, using both computer simulation data and real market-share data. Fifth, an approximate functional relationship between \textit{C}_K and \textit{HHI} is established with impressive accuracy and serving various purposes: (i) the behavior and changes in \textit{HHI} can be compared with those of \textit{C}_K, (ii) reported values of \textit{HHI} can be converted to those of \textit{C}_K, (iii) the theoretical relationship between industry profitability and \textit{HHI} can be extended to that of \textit{C}_K, and (iv) the DOJ and FTC merger guidelines based on \textit{HHI} can be equivalently expressed in terms of \textit{C}_K.

Since the new index \textit{C}_K has the value-validity property as proved in this paper, \textit{C}_K is equally sensitive to changes in a market-share distribution throughout its potential range of values, permitting order and difference comparisons between index values. By contrast, other indices such as the most widely used \textit{HHI} lack this property so that different types of comparisons between index values reflect those of an index itself rather than of the concentration attribute being measured. In the case of \textit{HHI}, the effect of lacking the value-validity property implies
that HHI lacks adequate sensitivity to changes in the market-share distribution for smaller values of HHI whereas it has excessive sensitivity for larger values of HHI. This important difference between C\(_k\) other indices, notably HHI, is indeed relevant when considering the anticompetitive effects of mergers and the ongoing debate about whether or not and to what extent concentration and market power have increased over the recent decades (e.g., [21–24]).

2. Required properties

2.1 General requirements

In order for \(C\) to qualify as a concentration index, it has to have a number of specific properties as discussed extensively in the published literature (e.g., [3, pp.47-58] [4, 25, 26]). First, however, a comment on notation: while the strictly mathematically correct notation would be to refer to \(C(P_n)\) as being the value of the index or function \(C\) for the distribution \(P_n = (p_1, \ldots, p_n)\), \(C\) and \(C(P_n)\) will both be used throughout this paper to denote both an index (function) and its value in order to simplify the notation when there is no chance of ambiguity. Required properties of \(C\) may then be outlined as follows:

(P1) **Continuity:** \(C\) is a continuous function of all the \(p_i\) (\(i = 1, \ldots, n\)).

(P2) **Symmetry:** \(C\) is (permutation) symmetric in \(p_1, \ldots, p_n\).

(P3) **Zero-indifference:** \(C\) is unaffected if one or more firms with zero market share enter the market, i.e., \(C(p_1, \ldots, p_n, 0, \ldots, 0) = C(p_1, \ldots, p_n)\).

(P4) **Schur-convexity:** \(C\) is strictly Schur-convex.

(P5) **Value validity:** \(C\) has value validity.

The continuity Property (P1) ensures that small changes in some of the \(p_i\) (\(i = 1, \ldots, n\)) result in only a small change in the value of \(C\). Property (P2) simply states that \(C\) is invariant with respect to the order in which the original \(p_i\) (\(i = 1, \ldots, n\)) are given (e.g., [27], [5, p. 222]). According to Property (P3), the addition or deletion of one or more \(p_i = 0\) components to or from the distribution \(P_n = (p_1, \ldots, p_n)\) has no effect on the value of \(C\). Property (P3) has two important implications. First, with respect to market concentration, (P3) together with (P1) and (P4), has the effect of slightly decreasing the value of \(C\) if one or more additional small firms enter the market (or the reverse effect if the small firms leave the market) (e.g., [28, Ch. 8], [3, p. 53]). Second, Property (P3) implies that \(C\) cannot be an explicit function of \(n\) such as if \(C\) were to be normed to some specific range by adjusting (controlling) for \(n\). Properties (P4)-(P5) are subsequently defined and discussed in detail.

Another property advocated by some is the so-called replication property. First suggested by Hall and Tideman [29], this property means that if each \(p_i\) is split into \(k\) equal parts \(p_i/k, \ldots, p_i/k\) for \(i = 1, \ldots, n\), resulting in the \(kn\) component distribution denoted by \(P_n/k\), then \(C\) should be reduced by a factor of \(1/k\), i.e., \(C(P_n/k) = C(P_n)/k\). Replication, which is different from the homogenous property of a function, may be an interesting and novel mathematical concept, but is entirely unrealistic of any practical situation involving concentration. While favored by some (e.g., [29, 30]), others do not view replication as a necessary property (e.g., [3, pp.47-58]), some do not even mention replication among the properties of a concentration index (e.g., [5, pp.221-223], [27]), and some feel that it is not self-evident that this property is desirable [31, pp. 63–64]. The index \(C\) and its normalized form \(C^* \in [0, 1]\) cannot both have the replication property, but no rigorous explanation has been offered as to which one should have this property and why. It is also worth noting that the (sample) standard deviation \(s_n\) (with devisor \(n\)) of \(p_1, \ldots, p_n\) has the replication property, but \(s_{n-1}\) (with devisor \(n-1\)) does not (nor does the
variance $s_n^2$). However, nobody would argue that $s_n$ is to be preferred over $s_{n-1}$ because of the replication principle even though $s_{n-1}$ has the statistical property of unbiasedness, nor would $s_n$ be preferred over $s_n^2$ for this reason.

Most importantly, however, is the fact that no generally accepted and rigorous basis appears to have been provided for the replication property as being of any importance to a concentration index. Also, from the value-validity property (P5) discussed below, it becomes apparent that replication is inconsistent with this property. It appears that the replication property lacks any real or solid justification.

### 2.2 Schur-convexity

The strict Schur-convexity Property (P4), which was also discussed by Hannah and Kay [3, Ch. 4], means that if the components of $P_n = (p_1, \ldots, p_n)$ are “less unevenly distributed” or “less spread out” than are those of another distribution $Q_n = (q_1, \ldots, q_n)$, then $C(P_n) < C(Q_n)$. Such rather vague statements can be expressed more precisely in terms of majorization theory (e.g., Marshall et al. [32]). Thus, by definition, if the $p_i$'s are ordered such that

$$p_1 \geq p_2 \geq \cdots \geq p_n$$

then $P_n$ is majorized by $Q_n$ (denoted by $P_n \prec Q_n$) under the following condition:

$$P_n \prec Q_n \text{ if } \sum_{i=1}^{j} p_i \leq \sum_{i=1}^{j} q_i, \quad j = 1, \ldots, n - 1$$

(2)

with $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1$. $C$ is then strictly Schur-convex if

$$P_n \prec Q_n \text{ implies } C(P_n) < C(Q_n)$$

(3)

assuming $P_n$ is not a permutation of $Q_n$ [32, pp. 8, 80]. If the inequality in (3) is not strict, then $C$ is Schur-convex.

The condition in (3), which is referred to by economists as the *Dalton condition* or the *Pigou-Dalton condition*, also implies that (a) $C$ has the transfer property and (b) $C$ preserves the Lorenz order [32, pp. 5–8, 560, 712–723]. The principle of “upward” transfers means that if $p_i > p_j$, and an amount $\delta < p_i - p_j$ is transferred from $p_i$ to $p_j$, the value of $C$ increases. The fact that the strictly Schur-convex $C$ preserves the Lorenz order means that if the distribution $P_n$ is majorized by $Q_n$ as defined in (2), then (a) the Lorenz curve (after [33]) for $P_n$ falls nowhere below the Lorenz curve for $Q_n$ and (b) $C(P_n) < C(Q_n)$. By definition, if the $p_i$'s are ordered such that $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(n)}$, with $F_{(0)} = \sum_{j=1}^{0} p_{(j)}$ and $F_{(n)} = 0$, then the Lorenz curve for $P_n$ is obtained from the line segments between the consecutive points $i/n, F_{(i)}$ for $i = 0, \ldots, n$.

### 2.3 Value validity

#### 2.3.1 The importance of the value-validity property.

In order to provide a simple confirmation of the need to impose a restriction on the values of a concentration index, consider first the complete lack of uniformity among the values taken on by different concentration indices for the same data sets. For some of the indices $C_i$ (defined and discussed in Section 3 below) and for the distribution $P_{10}^{15} = (0.55, 0.05, \ldots, 0.05)$, the respective values of the indices $C_1(CR_4)$, $C_6(HHI)$, $C_5$, $C_6$, $C_9$, $C_{13}$, $C_{14}$, and $C_{17}$ become 0.70, 0.33, 0.18, 0.59, 0.48, 0.41, 0.19, 0.14, and 0.45. That is as much as 400% variation in index values for the same data set in
spite of the fact that many of these indices have the same properties. Even the values of $C_3(HHI)$ and $C_{13}$ differ by a factor of nearly 2 although both indices are members of the parameterized family $C_{18}$ (with $\alpha = 1$ and $\alpha \rightarrow 0$).

In fact, one could define a family of indices as the power function $HHI^\alpha$ with arbitrary parameter $\alpha > 0$ and with all members having Properties (P1)-(P4). Of course, the values of different family members could vary greatly for the same data set and difference comparisons between index values for different data sets could vary greatly, depending upon $\alpha$. The need for some additional constraint on index values would seem to be paramount.

Such variation in index values is also reflected in their differing sensitivities to changes in the distribution $P_n = (p_1, \ldots, p_n)$. In the case of the most important index $HHI$, as discussed more extensively later in the paper, $HHI$ is more sensitive to changes in $P_n$ for large index values than for small index values. As a numerical illustration, consider the two distributions $(0.30, 0.30, 0.20, 0.20)$ and $(0.75, 0.15, 0.10, 0.00)$ for which $HHI = 0.26$ and 0.28, respectively, where the first distribution is obtained from the second one by means of a transfer as defined above. Then, consider the same transfer involving the two distributions $(0.70, 0.15, 0.10, 0.05)$ and $(0.75, 0.15, 0.10, 0.00)$ for which $HHI = 0.53$ and 0.60. The increase in $HHI$ values between the last two more concentrated distributions is about three times that between the first two less concentrated distributions (i.e., 0.07 versus 0.02), with relative increases of 13.2% and 7.7%.

Since all proposed concentration indices take on their extremal values for the two distributions

$$P_n^0 = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right), \quad P_n^1 = (1, 0, \ldots, 0),$$

it may be interesting and informative to consider index values for the following mean of $P_n^0$ and $P_n^1$:

$$P_n^{0.5} = \frac{1}{2} P_n^0 + \frac{1}{2} P_n^1 = \left(\frac{1}{2n}, \frac{1}{2n}, \ldots, \frac{1}{2n}\right).$$

In terms of this information alone, the only defensible value of a concentration index $C$ is that its value for $P_n^{0.5}$ in (5) should be the mean of its values for $P_n^0$ and $P_n^1$ in (4), i.e.,

$$C(P_n^{0.5}) = \frac{1}{2} C(P_n^0) + \frac{1}{2} C(P_n^1).$$

For the $P_n^{0.5} = (0.55, 0.05, \ldots, 0.05)$ used in the above example and with $C(P_n^0) = 1/n$ and $C(P_n^1) = 1$ as for many indices, including $HHI$, (6) requires that $C(P_n^{0.5}) = 0.55$. However, by comparison, $HHI(P_n^{0.5}) = 0.33$ i.e., 40% less than the corresponding requirement from (6).

Other indices fare no better, such as $C_{14}(P_n^{0.5}) = 0.14$ < 0.55.

The requirement in (6) can also be expressed in terms of distances or differences as

$$C(P_n^1) - C(P_n^{0.5}) = C(P_n^{0.5}) - C(P_n^0)$$

showing the $C(P_n^{0.5})$ is an equal distance from each of the two extreme values $C(P_n^0)$ and $C(P_n^1)$. Furthermore, by considering the distributions $P_n^0$, $P_n^{0.5}$, and $P_n^1$ as being points or vectors in $n$-dimensional Euclidean space, the following equality between Euclidean distances can be seen to hold:

$$d(P_n^0, P_n^{0.5}) = d(P_n^{0.5}, P_n^1).$$
That is, the distance between $P_{n}^{0.5}$ and $P_{n}^{l}$ is the same as that between $P_{n}^{0.5}$ and $P_{n}^{0}$, supporting the requirement in (7).

### 2.3.2 Generalized requirement.

The development from (4) to (8) can be further generalized by using the lambda distribution

$$P_{n}^{l} = \left( \frac{\lambda}{n}, \frac{\lambda - \lambda}{n}, \ldots, \frac{\lambda - \lambda}{n} \right), \quad 0 \leq \lambda \leq 1 \quad (9)$$

introduced by Kvålseth [34] (the $1-\lambda$ is used here as a convenient form when dealing with concentration). This particular distribution provides an important and general basis for the value-validity property as first discussed in Kvålseth [34]. The present paper adapts those basic concepts to the measurement of concentration.

The real-valued parameter $\lambda$ is basically a concentration parameter of which $\lambda = 0$, $\lambda = 0.5$, and $\lambda = 1$ are the particular cases in (4) and (5). In fact, the following linear function or weighted arithmetic mean of $P_{n}^{0}$ and $P_{n}^{l}$

$$P_{n}^{l} = (1 - \lambda)P_{n}^{0} + \lambda P_{n}^{l}, \quad 0 \leq \lambda \leq 1 \quad (10)$$

includes the mean in (5) as a particular case with $\lambda = 1/2$. Furthermore, the condition in (6) would be a particular case of the general formulation

$$C(P_{n}^{0}) = (1 - \lambda)C(P_{n}^{0}) + \lambda C(P_{n}^{l}). \quad (11)$$

Similarly, it is seen that (7) and (8) generalize as follows:

$$C(P_{n}^{l}) - C(P_{n}^{0}) = \left( \frac{1 - \lambda}{\lambda} \right) [C(P_{n}^{0}) - C(P_{n}^{l})] \quad (12)$$

and

$$d(P_{n}^{l}, P_{n}^{0}) = \left( \frac{1 - \lambda}{\lambda} \right) d(P_{n}^{l}, P_{n}^{0}). \quad (13)$$

Besides considering (11) as a direct extension of (6), although clearly supported by (12)-(13), the general condition for value validity in (11) also follows from the requirement that $C$ should be equally sensitive to small changes in the distribution $P_{n} = (p_{1}, \ldots, p_{n})$ throughout its range from $C(P_{n}^{0})$ to $C(P_{n}^{l})$. This sensitivity requirement can equivalently be expressed in terms of the discriminant ability of $C$, i.e., the ability of $C$ to discriminate between distributions $P_{n}$ that are not unduly different. Thus, $C$ should have a constant or uniform ability to detect small changes in any $P_{n}$. Such a requirement can be imposed on an index $C$ with Properties (P1)-(P4) because of the following equality:

$$C(P_{n}^{0}) = C(P_{n}^{l}) \quad \text{for one unique } \lambda \quad (14)$$

for any $P_{n}$ and the $P_{n}^{l}$ in (10). Consequently, as the $p_{i}$'s become increasingly unequal or uneven so that $C(P_{n})$ increases, the value of the concentration parameter $\lambda$ increases correspondingly for any given $n$. As a simple illustration involving the index HHI, consider $HHI(0.50, 0.30, 0.20) = 0.38$ and $HHI(0.70, 0.20, 0.10) = 0.54$ for which the respective values of $\lambda$ from (14) is found to be $\lambda = 0.26$ and $\lambda = 0.56$.

The requirement or condition that $C(P_{n})$ should be equally sensitive to small changes in the form of $P_{n}$ throughout its range becomes equivalent to requiring the partial derivative $\partial C(P_{n})/\partial \lambda$ to be constant for all $\lambda$ and any given $n$. Consequently, $C(P_{n}^{0})$ has to be a linear function of $\lambda$, which leads immediately to (11).
3. Assessment of indices

A concise historical account of concentration indices proposed to date is summarized in Table 1. Some of these are parameterized families of indices such as $C_{19}$, $C_{11}$, $C_{12}$, $C_{13}$, and $C_{19}$ since they depend on an arbitrary parameter $\alpha$. Two of the individual indices, $C_3(HHI)$ and $C_4$.

| $C_i$ | Formula | Reference | Notes | LP |
|-------|---------|-----------|-------|----|
| $C_1$ (CR_m) | $\sum_{i=1}^{m} p_i$ | Various authors | P1, P4 |
| $C_6$ | $p_i + \sum_{i=1}^{n} p_i^2(2 - p_i)$ | Horvath [35] | P4, P5 |
| $C_7$ | $\sum_{i=1}^{n} p_i(m + 1 - i)/m, m \geq 1$ | Hart [36] | b | P1, P3, P4 |
| $C_9$ | $2\sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n} p_i$ | Hart [37] | c | P5 |
| $C_{10}$ | $\sum_{i=1}^{n} p_i^2 + \{p_i(\sum_{j=1}^{n} p_j - p_i^2)\}^\alpha, \alpha>1$ | Hase [39] | P5 |
| $C_{11}$ | $(\sum_{i=1}^{n} p_i^\alpha)^{1/(1-\alpha)}, \alpha>0$ | Hannah & Kay [3] | d | P5 |
| $C_{12}$ | $(\sum_{i=1}^{n} p_i^\alpha)^{1/(1-\alpha)}, \alpha>0$ | Davies [40] | P1, P3, P5 |
| $C_{13}$ | $\prod_{i=1}^{n} p_i$ | Bruckmann [41], Hani [42] | P5 |
| $C_{14}$ | $\sum_{i=1}^{n} p_i[1 + n(1 - p_i)]^{-1}$ | Ginevičius [43] | P1, P3, P5 |
| $C_{15}$ | $\sum_{i=1}^{n} p_i^2 / \sum_{i=1}^{n} p_i^2, m \geq 1$ | Anbarci & Katzman [44] | P4, P5 |
| $C_{16}$ | $\sum_{i=1}^{n} (\log p_i)^2 / n - (\sum_{i=1}^{n} \log p_i)^2 / n^2$ | Hannah & Kay [3] | P1, P4, P5 |
| $C_{17}$ | $1 - (2\sum_{i=1}^{n} p_i - 1)/n$ | Marfels [45] | P1, P3 |
| $C_{18}$ | $(\sum_{i=1}^{n} p_i^{x+1})^{1/x}, \alpha > -1$ | Bruckmann [41] | P5 |
| $C_{19}$ | $\alpha^{-1}\log\sum_{i=1}^{n} p_i e^{p_i}, \alpha > 0$ | Bruckmann [41] | P5 |

Notes: (a) This is the famous entropy by Shannon [46]; it is strictly Schur-concave; (b) Hart [36] also suggested replacing $m$ in $C_7$ with $n$; (c) The $C_6$ was proposed as an alternative to $C_7$; (d) This family of indices ranges from $C_{11}(p^n_i) = n$ to $C_{11}(p^n_i) = 1$ for the market-share distributions in (4) and measures deconcentration; it is strictly Schur-concave.

https://doi.org/10.1371/journal.pone.0264613.t001
C13, are seen to be members of the Cα-family (for α = 1 and α→0). Note that, although all indices in Table 1 have the Symmetry Property (P2), the indices C1, C4, C6, C7, C9, and C17 are based on the descending order among the p_i's as in (1).

As indicated in the right column of Table 1, the various indices lack some of the Properties (P1)-(P5) as can easily be verified. A number of the potential indices lack the strict Schur-convexity (Property (P4)) while several others lack the zero-indifference Property (P3). Over three of the indices, C1(CR_m), C7, and C12, meet the condition in (11) required by the value-validity Property (P5). The other indices fail to meet even the weaker condition in (6). However, each of those three indices lack other properties as identified in Table 1. In particular, CR_m lacks Property (P1) and the strict Schur-convexity (Property (P4)), although it can be shown to be Schur-convex.

In the case of C4(HHI) and the distributions of P^c_n in (4) and P^i_n in (9), it is seen that

\[ HHI(P^c_n) = \lambda^2 HHI(P^c_1) + (1 - \lambda^2) HHI(P^c_n) = (1 - \frac{1}{n}) \lambda^2 + \frac{1}{n} \]  

(15)

with HHI(P^c_1) = 1/n and HHI(P^c_n) = 1. When compared with (1−1/n)λ+1/n as required by (11), it is clear that HHI understates the true extent of the concentration. This negative bias occurs throughout the range of values of HHI, but is seen to be most pronounced when λ = 1/2. From the partial derivative \( \partial HHI(P^c_n)/\partial \lambda = 2(1 - 1/n) \lambda \) and the equality in (14), the implication is that for any given n, HHI(P_n) is not equally sensitive to small changes in P_n throughout its range. Rather, HHI(P_n) has the bias of being increasingly sensitive to changes in P_n = (p_1, . . . , p_n) as the p_i's become increasingly uneven or variable, i.e., with increasing HHI(P_n)-values. The important point is this: while a concentration index should clearly be sensitive to changes in P_n, such sensitivity should not be biased in a particular direction such as toward uneven (skewed) distributions as in the case of HHI.

4. The new index

4.1 Index formulation

Several of the indices defined in Table 1 can be considered as members of the following class of weighted sums:

\[ C = \sum_{i=1}^{n} w_i p_i \]  

(16)

where w_1, . . . , w_n are positive weights with each w_i being a function of (or depending upon) p_i and possibly other components of the distribution P_n = (p_1, . . . , p_n). Other indices in Table 1 can be viewed as strictly increasing functions of the C in (16). The most obvious members of (16) are perhaps HHI with \( w_i = p_i \) for i = 1, . . . , n and CR_m with \( w_i = 1 \) for i = 1, . . . , m and 0 otherwise (with the p_i's being ordered as in (1)).

A family of concentration indices as in (16) can also be derived from theoretical models. For example, Encaoua and Jacquemin [27] arrived at (16) by means of an axiomatic characterization. In their derivation, w_i is a nondecreasing function of p_i such that \( w_i p_i \) is convex (see also Tirole [5, pp. 221–223]). Dickson [47] showed that a concentration index can be expressed as in (16) with \( w_i \) being the firms’ so-called conjectural variation elasticities, with \( w_i \in [0, 1] \) being the economically meaningful or valid interval for each w_i.

Two of the members of (16) defined in Table 1 (C7 and C17) use weights \( w_i \) based on the ranks from the distributions P_n = (p_1, . . . , p_n) when ordered as in (1). The C_i is also based on ranks, but this index belongs to the reciprocal of the family in (16). With rank 1
for \( p_1 \) (the largest \( p_i \)), rank 2 for \( p_2 \) (the second largest \( p_i \)), etc., each of these three indices are decreasing functions of the weighted mean rank \( \sum_{i=1}^{n} p_{i} i \).

Alternatively, instead of linearly decreasing weighted mean rank or the reciprocal weighted mean rank, consider the following weighted mean reciprocal rank:

\[
C_K = \sum_{i=1}^{n} p_{i} / i
\]

(17)

with the \( p_i \)'s ordered as in (1). This index is proposed as a new concentration index with desirable properties. In the case of market shares \( p_1, \ldots, p_n \), \( C_K \) is simply the market share of the largest firm (divided by 1), plus the market share of the second largest firm divided by 2, \ldots, plus the market share of the \( n \)th largest (i.e., smallest) firm divided by \( n \). The expression in (17) applies whether some or all of the \( p_i \)'s are equal. In the extreme case of \( P_n^0 \) in (4),

\[
C_K(P_n^0) = \sum_{i=1}^{n} (1/n)(1/i).
\]

When there are ties among the \( p_i \)'s so that \( p_1 = p_{i+1} = \cdots = p_{i+k} = p \) for any \( i \) and \( k \), it may appear as if different terms of \( C_K \) in (17) contribute differently to the value of \( C_K \) because their set of weights \( \{1/i\} \) differ even though the \( p_i \)'s are equal. However, the contribution toward \( C_K \) by such tied \( p_i \)'s can be expressed equivalently in terms of the arithmetic mean \( \left( \sum_{j=1}^{i+k} 1/j \right)/k \) of their reciprocal ranks as their common weight such that

\[
p/i + p/(i+1) + \cdots + p/(i+k) = \left( \sum_{j=1}^{i+k} 1/j \right)/k (p + p + \cdots + p).
\]

This expression shows clearly how the terms with equal \( p_i \)'s can all be considered as contributing equally toward the overall value of \( C_K \). Of course, \( C_K(P_n^0) \) represents the extreme case when \( p_1 = \cdots = p_n = 1/n \).

Besides being a weighted mean of reciprocal ranks or a weighted sum of the \( p_i \)'s (ordered as in (1)), it may be of interest to note that \( C_K \) is also the reciprocal of the weighted harmonic mean of the ranks 1,2, \ldots, \( n \). The \( C_K \) can also be given a geometric interpretation in terms of the cumulative \( p_i \)'s and \( 1/i \) as explained and illustrated in Fig 1.

The range of values of \( C_K \) based on the extreme distributions in (4) is given by

\[
C_K(P_n^0) = \left( \frac{1}{n} \right) \sum_{i=1}^{n} 1/i \leq C_K(P_n) \leq C_K(P_n^1) = 1.
\]

(18)

The lower bound on \( C_K \) is simply the reciprocal of the harmonic mean of the first \( n \) positive integers (or the ranks 1,2, \ldots, \( n \)). This is also by itself an interesting mathematical quantity. In particular, \( H_n = \sum_{i=1}^{n} 1/i \) is the partial sum of the harmonic series, also referred to as the \( n \)-th harmonic number. A recurrence relation for \( C_K(P_n^0) \) follows immediately from \( H_{n+1} = H_n + 1/(n+1) \). Although tabulated \( H_n \) values are readily available, a quick way to obtain values of \( C_K(P_n^0) \) is by means of the following formula:

\[
C_K(P_n^0) = \lfloor \log(n + 1/2) + 0.5772 \rfloor / n
\]

(19)
where the 0.5772 is the Euler-Mascheroni constant (to 4 decimal places). This formula is found to be correct to at least 3 decimal places for \( n > 5 \).

### 4.2 Properties of \( C_K \)

It is readily apparent from the expression in (17) that \( C_K \) has each of the Properties (P1)-(P3). The strict Schur-convexity of \( C_K \) (Property (P4)) is rather apparent from the form of (17) and the Schur-convexity definition in (2)-(3). It also follows from the fact that (a) \( C_K \) has the symmetry Property (P2) and the partial derivative \( \partial C_K/\partial p_i = 1/i \) is strictly decreasing in \( i = 1, \ldots, n \) [32, p. 84].

What sets \( C_K \) apart from other concentration indices is the fact that \( C_K \) has the value-validity property (P5). From (10) and (17),

\[
C_K(P_n^*) = \lambda + (1 - \lambda)C_K(P_n^0)
\]

so that, with \( C_K(P_n^0) = 1 \), \( C_K(P_n^*) \) complies with the value-validity condition in (11). The \( C_K(P_n^0) \) is the lower bound on \( C_K \) defined in (18).

Market concentration may be viewed as consisting of two different components: the size of a market or industry \( (n) \) and the inequality between the market shares. The index \( C_K \) can be
formulated so as to reflect those two components separately by expressing \( C_K(P_n) \) as follows:

\[
C_K(P_n) = \left[ C_K(P_n) - C_K(P_n^i) \right] C_K(P_n) + C_K(P_n^i) \quad \text{and} \quad C_K(P_n) = \frac{C_K(P_n) - C_K(P_n^i)}{C_K(P_n) - C_K(P_n^i)} \tag{21}
\]

with \( C_K(P_n^i) = 1 \) and \( C_K(P_n) \in [0, 1] \). The \( C_K(P_n^i) \) is a function of \( n \) only and the normalized form \( C_K(P_n) \), which controls or adjusts for \( n \), is a measure of market-share inequality. It is clear from (21) that for any fixed \( C_K(P_n^i) \), \( C_K(P_n) \) is a strictly increasing function of \( C_K(P_n) \) and strictly decreasing function of \( n \). Similarly, for any fixed \( n \), \( C_K(P_n) \) is strictly increasing in \( C_K(P_n^i) \). That is, \( C_K(P_n) \) increases as the size of the market decreases and as the inequality between the market shares increases.

The sensitivity of \( C_K(P_n) \) to changes in the inequality \( C_K(P_n^i) \) and to changes in \( n \) via \( C_K(P_n) \) can be compared by using partial derivatives and treating \( C_K(P_n^i) \) as a continuous variable for mathematical purpose. Then, it follows from (21) that

\[
\frac{\partial C_K(P_n)}{\partial C_K(P_n^i)} = \frac{1 - C_K(P_n^i)}{1 - C_K(P_n^i)} \cdot \frac{\partial C_K(P_n)}{\partial C_K(P_n^i)} \tag{22}
\]

It appears from (22) that whether \( C_K(P_n) \) is more sensitive to changes in \( C_K(P_n^i) \) than to changes in \( C_K(P_n^i) \) or vice versa depends on whether \( C_K(P_n) > C_K(P_n^i) \) or \( C_K(P_n) < C_K(P_n^i) \).

As an extension of the weighted mean of the two extremal distributions \( P_n \) and \( P_n^i \) in (10), consider the weighted mean of any two distributions \( P_n = (p_1, \ldots, p_n) \) and \( Q_n = (q_1, \ldots, q_n) \), both arranged in descending order as in (1). It then follows immediately from the definition of \( C_K \) in (17) that the value of \( C_K \) for the mixture distribution with weights \( w \) and \( 1-w \) is given by

\[
C_K[wP_n + (1-w)Q_n] = wC_K(P_n) + (1-w)C_K(Q_n), \quad 0 \leq w \leq 1. \tag{23}
\]

This property of \( C_K \), which would seem to be an intuitively reasonable one, is basically an extension of the value-validity requirement in (11). Thus, if an index complies with (23), it will necessarily comply with (11) since \( P_n^0 \) and \( P_n^i \) are simply special forms of \( Q_n \) and \( P_n \), respectively.

The \( C_K \) does not have the replication property, although, as discussed in Section 2.1, this property cannot be considered essential and may rather be a questionable one. In fact, the replication property and the value-validity property (PS) are inconsistent or mutually exclusive. This inconsistency can be proved by first noting that for the distribution \( P_n^i \) in (9) and analogously to (21), the value-validity condition in (11) can also be expressed as

\[
C^i (P_n^i) = \frac{C(P_n^i) - C(P_n^0)}{C(P_n^i) - C(P_n^0)} = \hat{\lambda} = \frac{n}{\sqrt{n} - 1} s_n = s_n^i (P_n^i) \tag{24}
\]

where \( s_n \) is the standard deviation (with divisor \( n \)) of the \( n \) components of \( P_n^0 \) and \( s_n^i (P_n^i) \) is its normalized form. If \( C \) were to have the replication property, then \( C(P_n^i/k) = C(P_n^i)/k \), \( C(P_n^0/k) = C(P_n^0)/k \), and \( C(P_n^i/k) = C(P_n^0)/k \) so that \( C^i (P_n^i/k) = C (P_n^0/k) \). However, since \( s_n^i (P_n^i/k) = \hat{\lambda} s_n (P_n^i/k) \) and hence \( s_n^i (P_n^i/k) \neq s_n^i (P_n^i) \), \( C \) cannot simultaneously have the replication property and meet the condition in (24) for value validity.

Some prefer to use a concentration index that has the property of being a so-called \textit{numbers equivalent} index such as the index family \( C_{11} \) by Hannah and Kay [3] defined in Table 1. The numbers equivalent \( NC_K \) of \( C_K \) in (17) can be obtained by simply setting \( C_K(P_n) = \)

\[
C_K (P_{NC_K}) = \left( \sum_{i=1}^{NC_K} 1/i \right)/NC_K \text{ and solving for the nearest integer } NC_K \text{ for any given } P_n. \text{ Thus, for}
\]
the concentration \( C_K(P_n) \) of the market-share distribution \( P_n = (p_1, \ldots, p_n) \). \( NC_K \) becomes the number of firms in an equivalent market with equal market shares and whose market concentration equals the given \( C_K(P_n) \). However, while \( NC_K \) does provide an alternative interpretation for any index value \( C_K(P_n) \), its utility is limited by its lack of the value-validity property (P5) since \( NC_K \) is not a linear function of \( C_K(P_n) \) as would be required for compliance with (11).

Another aid for interpreting or visualizing values of \( C_K \) may be the use of the equivalent lambda distribution \( P_n^* \) obtained from (10) and (14) and by setting \( C_K(P_n) = C_K(P_n^*) \) for the \( C_K(P_n) \) in (20) and solving for \( \lambda = \lambda_n \) and for any distribution \( P_n = (p_1, \ldots, p_n) \). This solution is seen to equal the normalized \( C_K(P_n) \) in (21), i.e., \( \lambda_n = C_K(P_n) \). Thus, for example, consider \( P_s = (0.40, 0.30, 0.15, 0.10, 0.05) \) for which \( C_K(P_s) = 0.64 \) and \( C_K(P'_s) = 0.46 \) so that \( \lambda_s = C^*_s(P'_s) = 0.33 \). This gives \( P^*_s = (0.48, 0.13, \ldots, 0.13) \) as being the equivalent lambda distribution having (approximately) the same \( C_K \)-value as \( C_K(P_s) \) for the given \( P_s \).

5. Comparison with other indices

The most significant difference between the new index \( C_K \) in (17) and other indices proposed to date is due to the fact that \( C_K \) has the value-validity property (P5) whereas most other indices lack this property. However, the most interesting comparison is between \( C_K \) and the two indices \( CR_m \) and \( HHI \) since those are by far the most popular ones (e.g., [13, pp. 1216–118], [14, pp. 97–101], [6, Ch. 8]). Among the \( CR_m \) members, the most commonly used member is the four-firm concentration ratio \( (CR_4) \) (e.g., [16, p. 255]). Comparisons between \( HHI \) and \( CR_4 \) have been extensively discussed [15].

In order to compare the values of \( C_K \) with those of \( HHI \) and \( CR_4 \) for a wide variety of market-share distributions \( P_n = (p_1, \ldots, p_n) \), a computer algorithm was used to randomly generate \( P_n \) as follows. First, \( n \) was generated as a random integer between \( n = 5 \) and \( n = 100 \) (inclusive). The lower limit \( n = 5 \) was chosen since the four-firm \( CR_4 \) was used and to avoid \( CR_4 = 1 \) values simply because of values \( n<5 \) being generated. Then, for each such generated \( n \), each \( p_i \) (ordered as in (1)) was generated as a random number (to the desired decimals) within the following intervals:

\[
\frac{1}{n} \leq p_1 \leq 1,
\]

\[
\frac{1 - \sum_{j=1}^{i-1} p_j}{n - (i - 1)} \leq p_i \leq \min\left\{ p_{i-1}, \frac{1 - \sum_{j=1}^{i-1} p_j}{n - (i - 1)} \right\} \text{ for } i = 2, \ldots, n - 1
\]

\[
p_n = 1 - \sum_{j=1}^{n-1} p_j.
\]

A total of 1000 such distributions were generated and the corresponding values of \( C_K \), \( CR_4 \), and \( HHI \) were computed. The results for the \( C_K-CR_4 \) comparison is illustrated in Fig 2.

It is clear from their differing expressions that there cannot be any precise functional relationship between \( CR_4 \) and \( C_K \) as is also demonstrated in the result in Fig 2. The rather systematic result is that the variation in \( CR_4 \) values for any \( C_K \)-value tends to increase quite dramatically with increasing \( C_K \). More specifically, the change in the variation of \( CR_4 \)-values tends to increase up to about \( C_K = 0.5 \) beyond which it decreases. The same type of general relationship is also found between \( CR_4 \) and \( HHI \) [15].
From the simulation results for \( HHI \) versus \( C_K \) in Fig 3, it is rather strikingly clear that for any given value of either \( C_K \) or \( HHI \), the variation in the other index is quite limited. In fact, this scatter diagram supports the proposition that a functional relationship exists between the two indices.

In order to explore this potential relationship, three additional sets of distributions \( P_n \) have been analyzed: (1) the lambda distribution in (10), (2) randomly generated \( P_n \) as described above, and (3) real market-share distributions. The real data consist of the market shares of the firms within 20 different markets and were chosen so as to be readily available, include a wide variety of markets, and cover a reasonably wide range of concentration values. Those 20 data sets were simply obtained from internet searches with the source of each being identified. The purpose of using those real data sets was for exploring the potential relationship between \( C_K \) and \( HHI \) rather than providing reliable and representative market concentration results for specific markets. The same data sources were also used in Kvålseth [15] for comparing \( CR_4 \) and \( HHI \).

Those three different types of market-share data were then used to explore potential functions and transformations that would provide a reasonably accurate description of the apparent relationship between \( C_K \) and \( HHI \). The following result has emerged from this analysis:

\[
C_K \approx C_{KH} = \left( \frac{1}{3} \right) \log([e^3 - 1]HHI + 1)
\] (25)
for any distribution \( P_n = (p_1, \ldots, p_n) \) where the logarithm is to base \( e \) (natural logarithm) and the exponential term \( e^{3-1} = 19.0855 \). The approximation between \( C_K \) and the transformation \( C_{KH} \) in (25) results from regression analysis. Specifically, for the regression (through the origin) model \( C_K = \alpha C_{KH} \), the following estimates are obtained: \( \hat{\alpha} = 0.98 \) for the data in Table 2 involving \( P_n \) in (10) and different values of \( n \) and \( \lambda \), \( \hat{\alpha} = 1.00 \) for the data in Table 3 involving randomly generated \( P_n \), and \( \hat{\alpha} = 1.02 \) for the real market results in Table 4.

In fact, for \( \hat{\alpha} = 1.00 \), the fitted model \( \hat{C}_K = C_{KH} \) provides an excellent fit to all data sets. The coefficient of determination, when properly computed [62], is found to be \( R^2 = 1 - \frac{\sum (C_K - C_{KH})^2}{\sum (C_K - \bar{C}_K)^2} = 0.985, 0.990, \) and 0.990 for the data in Tables 2–4, respectively. That is, 99% of the variation of \( C_K \) (about its mean) is explained (accounted for) by the fitted model.

While the excellent fit of the relationship \( C_K \approx C_{KH} \) in (25) is apparent from the data in Tables 2–4, it is more conveniently depicted in Fig 4. It is clear from this scatter diagram that (25) is indeed an appropriate relationship supported by the data. When looking at the residuals \( (C_K - C_{KH}) \) in Fig 4, they tend to fall within a rather uniform band along each side of the curve. There appears to be no other systematic pattern among the residuals that would indicate the need for any alternative expression.
It is clear from these results that the approximation $C_K(P_n) \approx d(P_n, P_0^n)$ holds to a highly respectable degree of accuracy. In fact, the properly computed $R^2$-values for the fitted model $\hat{C}_K(P_n) = d(P_n, P_0^n)$ are found to be $R^2 = 1 - \sum(C_K - \hat{C}_K)^2 / \sum(C_K - \bar{C}_K)^2 = 0.990$ for all the data in each of Tables 3 and 4.

In many practical situations involving potential concerns about mergers and market structure, the number of firms $n$ within a market or industry may be quite small. Although the data in Tables 2–4 include a number of data sets with $n \leq 10$, it may be worthwhile to look at some additional data as in Table 5 obtained for market-share distributions $P_n = (p_1, \ldots, p_n)$ computer generated from the algorithm described above for $2 \leq n \leq 10$.

### Table 2. Values of $C_K$ in (17), HHI defined in Table 1, and $C_{KH}$ in (25) for $P_0^n$ in (9) with varying $\lambda$ and $n$.

| $\lambda$ | $n$ | $C_K$ | HHI | $C_{KH}$ |
|-----------|-----|------|-----|---------|
| 0.10      | 2   | 0.78 | 0.51| 0.79    |
| 0.30      | 2   | 0.83 | 0.55| 0.81    |
| 0.50      | 2   | 0.88 | 0.63| 0.86    |
| 0.70      | 2   | 0.93 | 0.75| 0.91    |
| 0.90      | 2   | 0.98 | 0.91| 0.97    |
| 0.10      | 5   | 0.51 | 0.21| 0.54    |
| 0.30      | 5   | 0.62 | 0.27| 0.61    |
| 0.50      | 5   | 0.73 | 0.40| 0.72    |
| 0.70      | 5   | 0.84 | 0.59| 0.84    |
| 0.90      | 5   | 0.95 | 0.85| 0.95    |
| 0.10      | 10  | 0.36 | 0.11| 0.38    |
| 0.30      | 10  | 0.51 | 0.18| 0.50    |
| 0.50      | 10  | 0.65 | 0.33| 0.66    |
| 0.70      | 10  | 0.79 | 0.54| 0.81    |
| 0.90      | 10  | 0.93 | 0.83| 0.94    |
| 0.10      | 20  | 0.26 | 0.06| 0.25    |
| 0.30      | 20  | 0.43 | 0.14| 0.43    |
| 0.50      | 20  | 0.59 | 0.29| 0.63    |
| 0.70      | 20  | 0.75 | 0.52| 0.80    |
| 0.90      | 20  | 0.92 | 0.82| 0.94    |
| 0.10      | 30  | 0.22 | 0.04| 0.19    |
| 0.30      | 30  | 0.39 | 0.12| 0.40    |
| 0.50      | 30  | 0.57 | 0.28| 0.62    |
| 0.70      | 30  | 0.74 | 0.51| 0.79    |
| 0.90      | 30  | 0.91 | 0.82| 0.94    |
| 0.10      | 50  | 0.18 | 0.03| 0.15    |
| 0.30      | 50  | 0.36 | 0.11| 0.38    |
| 0.50      | 50  | 0.54 | 0.27| 0.61    |
| 0.70      | 50  | 0.73 | 0.50| 0.79    |
| 0.90      | 50  | 0.91 | 0.81| 0.93    |

https://doi.org/10.1371/journal.pone.0264613.t002
These results clearly support those in Table 2 for 2 ≤ n ≤ 50, Table 3 for 3 ≤ n ≤ 29, and Table 4 for 3 ≤ n ≤ 30. Again, the relationship between \( C_K \) and \( HHI \) is quite accurately described by (25), with \( R^2 = 0.986 \) for the fitted model \( \hat{C}_K = C_{KH} \) and the data in Table 5. Similarly, the normalized \( C_K \in [0,1] \) and the normalized Euclidean distance \( d' \in [0,1] \) show an impressive agreement, with \( R^2 = 0.993 \) for the fitted model \( \hat{C}_K = d' \) and the data in Table 5.

This close approximation between \( C_K(P_n) \) and \( d'(P_{n-1}, P_n) \) has important implications. First, it provides \( C_K(P_n) \) with an interesting mathematical property: \( C_K(P_n) \) becomes approximately a normalized distance metric (see, e.g., Chen et al. [63] for such metrics). This is of significance since the metric property is a strong one with far-reaching mathematical consequences. Second, this close approximation between \( C_K(P_n) \) and \( d'(P_{n-1}, P_n) \) also means that \( C_K(P_n) \) is an approximately linear function of the standard deviation \( s_n \) of \( p_1 \ldots p_n \) (with divisor \( n \)).

Table 3. Values of \( C_K \) in (17), \( C_K \) in (21) and (26), \( d' \) in (26), \( HHI \) and \( CR_4 \) defined in Table 1, and \( C_{KH} \) in (25) for randomly generated \( P_n = (p_1 \ldots p_n) \) and \( 2 \leq n \leq 30. \)

| n  | \( C_K \) | \( C'_K \) | \( d' \) | \( HHI \) | \( CR_4 \) | \( C_{KH} \) |
|----|------------|------------|---------|----------|-----------|------------|
| 27 | 0.84       | 0.81       | 0.76    | 0.59     | 0.94      | 0.84       |
| 29 | 0.16       | 0.03       | 0.03    | 0.04     | 0.18      | 0.17       |
| 19 | 0.75       | 0.69       | 0.68    | 0.49     | 0.79      | 0.78       |
| 8  | 0.50       | 0.24       | 0.23    | 0.17     | 0.63      | 0.49       |
| 12 | 0.45       | 0.26       | 0.24    | 0.14     | 0.55      | 0.43       |
| 9  | 0.71       | 0.58       | 0.53    | 0.36     | 0.96      | 0.69       |
| 15 | 0.48       | 0.33       | 0.32    | 0.16     | 0.54      | 0.47       |
| 3  | 0.79       | 0.46       | 0.47    | 0.48     | 1.00      | 0.77       |
| 14 | 0.24       | 0.01       | 0.01    | 0.07     | 0.30      | 0.29       |
| 4  | 0.72       | 0.42       | 0.40    | 0.37     | 1.00      | 0.70       |
| 21 | 0.35       | 0.22       | 0.20    | 0.09     | 0.38      | 0.33       |
| 24 | 0.36       | 0.24       | 0.21    | 0.08     | 0.42      | 0.32       |
| 13 | 0.82       | 0.76       | 0.72    | 0.56     | 0.89      | 0.82       |
| 12 | 0.91       | 0.88       | 0.85    | 0.75     | 0.95      | 0.91       |
| 15 | 0.24       | 0.02       | 0.02    | 0.07     | 0.29      | 0.28       |
| 29 | 0.60       | 0.54       | 0.48    | 0.26     | 0.78      | 0.60       |
| 20 | 0.26       | 0.10       | 0.09    | 0.06     | 0.31      | 0.25       |
| 8  | 0.77       | 0.65       | 0.60    | 0.44     | 1.00      | 0.75       |
| 13 | 0.40       | 0.21       | 0.19    | 0.11     | 0.53      | 0.37       |
| 17 | 0.92       | 0.90       | 0.88    | 0.78     | 0.95      | 0.92       |
| 19 | 0.38       | 0.24       | 0.22    | 0.10     | 0.43      | 0.36       |
| 14 | 0.28       | 0.06       | 0.05    | 0.07     | 0.36      | 0.29       |
| 28 | 0.29       | 0.17       | 0.16    | 0.06     | 0.35      | 0.25       |
| 17 | 0.65       | 0.56       | 0.52    | 0.32     | 0.73      | 0.65       |
| 25 | 0.49       | 0.40       | 0.39    | 0.19     | 0.50      | 0.50       |
| 17 | 0.91       | 0.89       | 0.85    | 0.74     | 0.97      | 0.90       |
| 25 | 0.36       | 0.24       | 0.22    | 0.08     | 0.44      | 0.32       |
| 21 | 0.18       | 0.01       | 0.01    | 0.05     | 0.20      | 0.22       |
| 29 | 0.34       | 0.24       | 0.22    | 0.08     | 0.37      | 0.31       |
| 23 | 0.78       | 0.74       | 0.68    | 0.49     | 0.98      | 0.78       |

https://doi.org/10.1371/journal.pone.0264613.t003
Specifically, it readily follows that

\[ C_K(P_n) \approx \left[ 1 - C_K(P_n^t) \right] \left( \frac{n}{\sqrt{n-1}} \right) \bar{d} + C_K(P_n^t). \] (27)

In the case of \( P_n = P_n^t \) in (10) and since \( C_K \) satisfies (11)–(13), the relationship in (27) becomes an equality as is also implied by (24). Therefore, because of the equality in (14) and from (24), the approximate relationship in (27) is not an unexpected one. It is important because it shows that, similarly to (21), \( C_K(P_n) \) is a function of both the number of firms in a market and their market-share variation as measured by the well-known standard deviation.

For the sake of completeness, the results for \( CR_4 \) are also given in Tables 3 and 4. Even though there is a substantial correlation between \( C_K \) and \( CR_4 \) (Pearson’s \( r = 0.97 \) and 0.92 and Spearman’s \( r_s = 0.89 \) and 0.95 for Tables 3 and 4, respectively), there are also substantial individual differences. Such differences between the values of \( C_K \) and \( CR_4 \) for individual data points seem to increase with increasing values of the two indices as in Fig 2. The \( HHI \) and \( CR_4 \) are also highly correlated with \( r = 0.89 \) and \( r_s = 0.91 \) and \( r = 0.81 \) and \( r_s = 0.96 \) for Tables 3 and 4, respectively. In spite of such high correlations, the different indices can certainly produce very conflicting results.

6. Discussion

6.1 Comments on (25)

The approximate, although quite accurate, relationship between \( C_K \) and \( HHI \) in (25) has important implications. First, with \( HHI \) becoming the increasingly dominant concentration index [7], reported \( HHI \) values can be converted into the corresponding \( C_K \) values for valid comparisons and interpretations of market concentrations.

Table 4. Values of \( C_K \) in (17), \( C_K \) in (21) and (26), \( d^t \) in (26), \( HHI \) defined in Table 1, and \( C_KH \) in (25) for a sample of real market-share data.

| n  | \( C_K \) | \( C_K \) | \( d^t \) | \( HHI \) | \( CR_4 \) | \( C_KH \) | Source | (Market type) |
|----|--------|--------|--------|--------|--------|--------|-------|-------------|
| 16 | 0.37   | 0.20   | 0.20   | 0.10   | 0.50   | 0.36   | [48]  | (Airline travel) |
| 16 | 0.42   | 0.27   | 0.25   | 0.12   | 0.60   | 0.40   | [48]  | (Airline travel) |
| 8  | 0.47   | 0.20   | 0.20   | 0.16   | 0.75   | 0.47   | [49]  | (U.S. distilled liquor) |
| 10 | 0.45   | 0.22   | 0.21   | 0.14   | 0.64   | 0.43   | [50]  | (Paints, coatings) |
| 10 | 0.38   | 0.12   | 0.11   | 0.11   | 0.52   | 0.38   | [51]  | (Pharmaceuticals) |
| 15 | 0.39   | 0.22   | 0.19   | 0.10   | 0.54   | 0.36   | [52]  | (Insurances companies) |
| 12 | 0.52   | 0.35   | 0.32   | 0.18   | 0.69   | 0.50   | [53]  | (Weapons exporters) |
| 30 | 0.26   | 0.15   | 0.13   | 0.05   | 0.34   | 0.22   | [54]  | (Car sales, Britain) |
| 12 | 0.39   | 0.18   | 0.20   | 0.12   | 0.60   | 0.40   | [55]  | (Auto Mnf., US) |
| 8  | 0.49   | 0.23   | 0.25   | 0.18   | 0.77   | 0.50   | [50]  | (Craft beer, US) |
| 9  | 0.46   | 0.21   | 0.23   | 0.16   | 0.75   | 0.47   | [50]  | (Running shoe sales) |
| 5  | 0.97   | 0.94   | 0.94   | 0.91   | 1.00   | 0.97   | [56]  | (Search eng., Norway) |
| 4  | 0.71   | 0.39   | 0.38   | 0.36   | 1.00   | 0.69   | [57]  | (Comm. water heaters) |
| 3  | 0.90   | 0.74   | 0.74   | 0.70   | 1.00   | 0.89   | [58]  | (Microprocessors) |
| 10 | 0.51   | 0.31   | 0.28   | 0.17   | 0.72   | 0.48   | [50]  | (Top charter airlines) |
| 5  | 0.58   | 0.23   | 0.22   | 0.24   | 0.79   | 0.57   | [59]  | (Global cigarettes, 2019) |
| 10 | 0.52   | 0.32   | 0.30   | 0.18   | 0.68   | 0.50   | [50]  | (Farm mach., equip.) |
| 20 | 0.32   | 0.17   | 0.18   | 0.08   | 0.48   | 0.31   | [60]  | (Global car sales) |
| 10 | 0.40   | 0.15   | 0.15   | 0.12   | 0.60   | 0.40   | [50]  | (Top airlines, world) |
| 8  | 0.70   | 0.55   | 0.51   | 0.35   | 0.83   | 0.68   | [61]  | (Consumer products) |

https://doi.org/10.1371/journal.pone.0264613.t004
Second, while the varying sensitivity of HHI to small changes in market-share distributions was discussed above based on (15) and on the particular distribution $P_n$ in (10), this analysis can be done for any distribution $P_n$ by using (25). With $C_K$ and HHI being functions of $P_n$ and taking the derivatives of both sides of (25) with respect to $P_n$ gives the following approximate relationship:

$$\frac{dHHI(P_n)}{dP_n} \approx A \frac{dC_K(P_n)}{dP_n}, \quad A = 3[HHI(P_n) + (e^3 - 1)^{-1}]. \quad (28)$$

Compared with $C_K$, which has constant sensitivity or discriminant power because of its value-validity property (P5), that of HHI is seen from (28) to increase with increasing HHI-values. Since $A = 1$ for $HHI = 0.32$ in (28) and since $C_K$ behaves appropriately, the implication is that HHI lacks adequate sensitivity when $HHI<0.32$ and has excessive and rapidly increasing sensitivity for $HHI>0.32$. This result is not inconsistent with the more general representation by the data in Fig 4.

### 6.2 $C_K$ and economic theory

One of the major reasons for the popularity of HHI is its solid theoretical relationship with market power. Cowling and Waterson [64], for example, showed that for an oligopolistic market (industry) with profit-maximizing firms, the average price-cost margin is directly related...
More specifically, the price-cost margin of Firm \(i\) is defined as

\[
PCM_i = \frac{P_i - MC_i}{P_i}
\]

where \(P_i\) is the market price set by the firm and \(MC_i\) is the firm’s marginal cost. The \(PCM_i\), which is also the well-known Lerner index, is considered as a measure of a firm’s market power. Then, as shown by Cowling and Waterson [64] and under certain economic assumptions (including the assumption that \(P_i = P\) for all \(i\)), the weighted mean \((PCM)\) of all the \(PCM_i\)’s within a market or an industry, using the market shares as the weights (i.e., \(PCM = \sum_{i=1}^{n} p_i PCM_i\)), is related to \(HHI\) as follows:

\[
PCM = \frac{HHI}{|\eta|}
\]

(29)

where \(\eta\) is the market (industry) price elasticity of demand. The price elasticity of demand,
which measures the variation in demand in response to a variation in price, is generally negative so that the right side in (29) may also be expressed as $-\frac{HHI}{\eta}$.

From (29), and by inverting the expression in (25), it follows that

$$PCM \approx \left( \frac{1}{\eta} \right) \left( \frac{e^{\frac{CK}{3}} - 1}{e^i - 1} \right).$$

(30)

That is, while the relationship between $PCM$ and $HHI$ is linear, the $PCM$ is related to $C_K$ in terms of an approximate exponential function. Since the close approximation in (25) has been empirically established, the approximation in (30) can be expected to be similarly close. Consequently, with $PCM$ being a well-known measure of market power, the relation in (30) shows that increasing market concentration as measured by $C_K$ results in increasing market power and hence decreasing competition and efficiency.

### 6.3 $C_K$ and merger implications

The $HHI$ is often computed from market shares as being percentages (rather than probabilities or proportions), which is the most typical form for reporting such data. Thus, the potential values of $HHI$ range from 0 to 10,000 as is also used in the DOJ and FTC Merger Guidelines. For any $HHI \in [0, 10,000]$, the corresponding values of $C_K$ in percentage points can be obtained from (25) as

$$C_K \approx C_{EH} = \left( \frac{100}{3} \right) \log \left( e^{\frac{HHI}{10,000}} - 1 \right) + 1 \in (0, 100].$$

(31)

Based on (31), the most recent (2010) Horizontal Merger Guidelines may then be summarized in terms of $C_K$-values and changes $\Delta C_K$ as follows:

- **Small Change in Concentration**: Mergers involving an increase $\Delta C_K < 2$ percentage points are unlikely to have adverse competitive effects and ordinarily require no further analysis.

- **Unconcentrated Markets** ($C_K < 45$): Mergers resulting in unconcentrated markets are unlikely to have adverse competitive effects and ordinarily require no further analysis.

- **Moderately Concentrated Markets** ($45 < C_K < 58$): Mergers resulting in moderately concentrated markets that involve $\Delta C_K > 2$ potentially raise significant competitive concerns and often warrant scrutiny.

- **Highly Concentrated Markets** ($C_K > 58$): Mergers resulting in highly concentrated markets that involve $2 < \Delta C_K < 3$ potentially raise significant competitive concerns and often warrant scrutiny. Mergers resulting in highly concentrated markets that involve $\Delta C_K > 3$ will be assumed to likely enhance market power. The presumption may be rebutted by persuasive evidence showing that the merger is unlikely to enhance market power.

Similarly, the European Commission (EC) Merger Guidelines (Regulation) also uses the $HHI$ index [13, p. 120] [66]. The EC guidelines may be expressed in terms of $C_K$ in percentage points as follows:

- It is unlikely that a merger will cause horizontal competition concerns if (a) the post-merger $C_K < 36$, (b) the post-merger $C_K$ is between 36 and 52 and $\Delta C_K < 4$, or (c) the post-merger $C_K > 52$ and $\Delta C_K < 2$, except in case of special given circumstances.

While these guidelines are directly transferred from those using $HHI$, it is important to note that since $HHI$ lacks the value-validity property (P5), changes $\Delta HHI$ in $HHI$-values do not truly reflect changes in the extent of the concentration characteristic, but rather reflect
changes in HHI values. As discussed above, the sensitivity of HHI to small changes in the market-share distribution $P_n$, rather than being constant as in the case of $C_K$, increases substantially with increasing HHI values. Thus, for example, an increase of $\Delta HHI = 100$ at $HHI = 1,500$ and at $HHI = 2,500$, as used by the Merger Guidelines, cannot be considered as equivalent changes in concentration. In fact, the corresponding changes in $C_K$ from (25), rather than being equal, are found to be 1.7 and 1.1, respectively.

Such invalid assumptions or interpretations of changes in HHI are of increasing importance, both theoretically and empirically, from studies emphasizing changes in concentration over its actual level. For example, Nocke and Whinston [8] show "that there is both a theoretical and empirical basis for focusing solely on the change in the Herfindahl index, and ignoring its level, in screening mergers for whether their unilateral effect will harm consumers (p.1)." Similarly, Miller et al. [67] used a Monte Carlo simulation study to evaluate price effects of mergers, finding a strong correlation between $\Delta HHI$ and the price change.

Whatever the purpose is of an investigation involving market concentration, whether theoretical or empirical, one proposition seems clear: changes in the values of a concentration index may be as important, if not more so, than its individual index values. Results that truly represent reality require an index with the value-validity property (P5). The $C_K$ is claimed to be such an index.

7. Conclusion

Value validity is introduced as a property required of a concentration index in order for its numerical values and their changes to provide reliable and true representations of the market or industry concentration characteristic. Since other indices proposed to date, including the most popular HHI, lack this value-validity property (P5), the new concentration index $C_K$ is introduced. Besides having this additional property, $C_K$ shares properties (P1-P4) with other indices, including HHI.

One desirable property of any summary measure is its simplicity and meaningfulness. The $C_K$ is certainly simple to compute and, as discussed, has meaningful interpretations. Computationally, $C_K$ simply equals the largest market share (divided by 1), plus the second largest market share divided by 2, plus the third divided by 3, and so forth for all the $n$ firms within the market or industry. The fact that the computation of $C_K$ requires the market shares to be ordered from the largest to the smallest as in (1) is not a real disadvantage since market-share data are typically reported in that format.

From the definition of $C_K$ in (17), it is clear that when the number of firms $n$ is large, the weights $(1/i)$ assigned to the different market shares ($p_i$) become sufficiently small such that the very small $p_i$’s can effectively be ignored from the computation of $C_K$. This is really a practical advantage of $C_K$ (as it also is with HHI) since market-share data are typically reported for the larger firms while the very small firms are grouped into an “others” category.

Reported results from research involving market concentration seem to be increasingly based on the HHI. The empirically determined relationship in (25) can then conveniently be used to derive the corresponding values of $C_K$ to a high degree of accuracy. Thus, valid concentration comparisons can be based on $C_K$.

Acknowledgments

The author expresses his appreciation to the three reviewers for their constructive comments and suggestions. Also, he thanks Andrew Kvålseth for an interesting discussion on the subject.
Author Contributions

Conceptualization: Tarald O. Kvålseth.
Data curation: Tarald O. Kvålseth.
Formal analysis: Tarald O. Kvålseth.
Investigation: Tarald O. Kvålseth.
Methodology: Tarald O. Kvålseth.
Project administration: Tarald O. Kvålseth.
Resources: Tarald O. Kvålseth.
Software: Tarald O. Kvålseth.
Supervision: Tarald O. Kvålseth.
Validation: Tarald O. Kvålseth.
Visualization: Tarald O. Kvålseth.
Writing – original draft: Tarald O. Kvålseth.
Writing – review & editing: Tarald O. Kvålseth.

References

1. Wilcox AR. Indices of qualitative variation and political measurement. Polit Res Q. 1973; 26: 325–343.
2. Kvålseth TO. Evenness indices once again: critical analysis of properties. Springer Plus. 2015; 4: 232.
3. https://doi.org/10.1186/s40064-015-0944-4 PMID: 26020023
4. Hannah L, Kay JA. Concentration in modern industry. London: Macmillan; 1977.
5. Curry B, George KD. Industrial concentration: a survey. J Ind Econ. 1983; XXXI(3): 203–255.
6. Tirole J. The theory of industrial organization. Cambridge, MA: MIT Press; 1988.
7. Tremblay VJ, Tremblay CH. New perspectives on industrial organization. New York: Springer; 2012.
8. Dutra J. Paradigm shifts on mergers efficiencies in antitrust analysis. Kansas Law Rev. 2020; 68: 1017.
9. Nocke V, Whinston MD. Concentration screens for horizontal mergers. Working Paper, 2020.
10. Farrell J, Shapiro C. The 2010 Horizontal Merger Guidelines after 10 years. Rev Ind Organ. 2020; 58:1–11. https://doi.org/10.1007/s11151-020-09807-6.
11. Casu B, Girardone C. Bank competition, concentration and efficiency in the single European market. The Manchester School. 2006; 74: 441–468.
12. Azzam AM. Measuring market power and cost-efficiency effects of industrial concentration. J Ind Econ. 1997; 45(4): 377–388.
13. Sari N. Efficiency outcomes of market concentration and managed care. International J Ind Org. 2003; 21(10): 1571–1589.
14. Gaughan PA. Mergers, acquisitions, and corporate restructuring. 5th ed. Hoboken, NJ: Wiley; 2011.
15. Andreosso B, Jacobson D. Industrial economics and organization: a European perspective. 2nd ed. London: McGraw-Hill; 2005.
16. Kvålseth TO. Relationship between concentration ratio and Herfindahl-Hirschman index: a re-examination based on majorization theory. Heliyon. 2018; 4(10): 1–24. https://doi.org/10.1016/j.heliyon.2018.e00846 PMID: 30338305
17. Carlton DW, Perloff JM. Modern industrial organization. 4th ed. Boston: Addison-Wesley; 2005.
18. Herfindahl OC. Concentration in the steel industry. PhD dissertation. New York City, NY: Columbia University; 1950.
19. Hirschman AO. National power and the structure of foreign trade. Berkeley, California: University of California Press; 1945.
20. U.S. Department of Justice and the Federal Trade Commission. Horizontal Merger Guidelines. 1968, 2010.
20. Kvåleth TO. A measure of homogeneity for nominal categorical data. Percept Mot Skills. 1993; 76: 1129–1130.
21. Gutierrez G, Philipp T. How EU markets became more competitive than US markets: A study of institutional drift. Technical Report, EconPapers. 2018.
22. Grullon G, Larkin Y, Michaely R. Are US industries becoming more concentrated? Review of Finance. 2019; 23(4): 697–743.
23. Autor D, Dorn D, Katz LA, Patterson C, Reenen JV. The fall of the labor share and the rise of superstar firms. The Quarterly Journal of Economics. 2020; 135(2): 645–709.
24. Bajgar M, Berlingieri G, Calligaris S, Criscuolo C, Timmis J. Industry concentration in Europe and North America. OECD Productivity Working Papers, 2019.
25. Rosenbluth G. Round Table-Gespräch über Messung der Industriellen Konzentration. In: Arndt S, editor. Die Konzentration in der Wirtschaft. Berlin, Germany: Duncker & Humbolt; 1961, p. 391.
26. Ginevičius R, Čirba S. Determining market concentration. J Bus Econ Manag. 2007; 8(1): 3–10.
27. Encaoua D, Jacquemin A. Degree of monopoly, indices of concentration and threat to entry. Int Econ Rev. 1980; 21(1): 87–105.
28. Theil H. Economics and information theory. Amsterdam: North-Holland; 1967.
29. Hall M, Tideman N. Measures of concentration. J Am Stat Assoc. 1967; 62: 162–168.
30. Chakravarty SR, Eichhorn W. An axiomatic characterization of a generalized index of concentration. J Prod Anal. 1991; 1:103–112.
31. Cowell FA. Measuring inequality. 3rd ed. Oxford, UK: Oxford University Press; 2011.
32. Marshall AW, Olkin I, Arnold BC. Inequalities: theory of majorization and its applications. 2nd ed. New York: Springer, 2011.
33. Lorenz MO. Methods for measuring concentration of wealth. J Am Stat Assoc. 1905; 9: 209–219.
34. Kvåleth TO. The lambda distribution and its applications to categorical summary measures. Adv Appl Stat. 2011; 24: 83–106.
35. Horvath J. Suggestion for a comprehensive measure of concentration. Southern Econ J. 1970; 36(4): 446–452.
36. Hart PE. Entropy and other measures of concentration. J Royal Stat Soc: Series A. 1971; 134: 73–85.
37. Hart PE. Moment distributions in economics: an exposition. J Royal Stat Soc: Series A. 1975; 138: 423–434.
38. Linda R. Methodology of concentration analysis applied to the study of industries and markets. Brussels, Belgium: Commission of the European Communities; 1976.
39. Hauze JC. The measurement of concentrated industrial structure and the size distribution of firms. Annals Econ Soc Meas. 1977; 6(1): 73–107.
40. Davies S. Measuring industrial concentration: an alternative approach. Rev Econ Stat. 1980; 62: 306–309.
41. Bruckmann G von. Einige bemerkungen zur statistischen messung der konzentration. Metrika. 1969; 14: 183–213.
42. Häni PK. Die Messung der Unternehmenskonzentration: Eine Theoretische und Empirische Evolution von Konzentrationsmassen. Diss. Verlag Rüegger; 1987.
43. Ginevičius R. Some problems of measuring absolute market concentration. In: Ginevičius R, Bivainis J, Melnikas B, editors. Modern business: priorities of development. Vilinius, Lithuania: Technika (in Lithuanian); 2005. pp. 12–36.
44. Anbarci N, Katzman B. A new industry concentration index. Econ Papers. 2015; 34(4): 222–228.
45. Marfels C. Absolute and relative measures of concentration reconsidered. Kyklos. 1971; 24(4): 753–766.
46. Shannon CE. A mathematical theory of communication. Bell Systems Tech J. 1948; 27: 379–423, 623–656.
47. Dickson VA. Conjectural variation elasticities and concentration. Econ Letters. 1981; 7: 281–285.
48. Lijesen MG, Nijkamp P, Rietveld P. Measuring competition in civil aviation. J Air Transport Manag. 2002; 8: 189–197.
49. Bain JS. Industrial Organization. New York: Wiley; 1959.
50. Market Share Reporter. 27th ed. Gale. 2017.
51. Editorial. New 2016 data and statistics for global pharmaceutical products and projections through 2017. ACS Chem Neurosci. 2017; 8: 1635–1636. https://doi.org/10.1021/acscchemneuro.7b00253 PMID: 28810746

52. Statista, Market share of the leading insurance companies in Belgium as of 2016, 2018. https://www.statista.com/statistika/statistics/780454/market-share-leading-insurance-companies-belgium/. (Accessed 24 June 2018).

53. Statista, Market share of the leading exporters of major weapons between 2013 and 2017, by country, 2018. https://www.statista.com/statistika/statistics/267131/market-share-of-the-leading-exporters-of-conventional-weapons/. (Accessed 24 June 2018).

54. 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).

55. Statista, U.S. market share of selected automobile manufacturers 2013, 2014. https://www.statista.com/statistika/statistics/249375/us-market-share-of-selected-automobile-manufacturers/. (Accessed 12 September 2014).

56. Statista, Market share held by the leading search engines in Norway as of July 2020. https://www-statista-com.ezp2.lib.umn.edu/statistika/statistics/621417/most-popular-search-engines-in-Norway/. (Accessed 12 October 2020).

57. Statista, Commercial water heater market share in the United States 2017–2019, by company. https://www-statista-co.ezp2.lib.umn.edu/statistika/statistics/700299/us-commercial-gas-water-heater-market-share/. (Accessed 12 October 2020).

58. Statista, Global microprocessor market share from 1st quarter 2009 to 3rd quarter 2011, by vendor. https://www.statista.com/statistika/statistics/270560/global-microprocessor-market-share-since-2009-by-vendor/. (Accessed 12 October 2020).

59. Statista, Global cigarettes market share as of 2019, by company. https://www.statista.com/statistika/statistics/279873/global-cigarette-market-share-by-group/. (Accessed 12 October 2020).

60. Carsalesbase.com, global car sales analysis 2017-Q1, 2017. http://carsalesbase.com/global-car-sales-2017-q1/. (Accessed 25 June 2018).

61. Statista, Share of the world’s leading consumer products companies 2013, by product sector. https://www-statista-com.ezp2.lib.umn.edu/statistika/statistics/225768/share-of-the-leading-consumer-products-companies-worldwide-by-product-sector/. (Accessed 12 October 2020).

62. Kvålseth TO. Cautionary note about R². The American Statistician. 1985; 39: 279–285.

63. Chen S, Ma B, Zhang K. On the similarity metric and distance metric. Theor Comput Sci. 2009; 410:2365–2376.

64. Cowling K, Waterson M. Price-cost margins and market structure. Economica. 1976; 43: 267–276.

65. Martin S. Advanced industrial economics. 2nd ed. Oxford, U.K.: Blackwell; 2002.

66. Union European. Guidelines on the assessment of horizontal mergers. Official Journal C-031. 2004; 05/02/2004: 0005–0018.

67. Miller NH, Remer M, Ryan C, Sheu G. Upward pricing pressure as a predictor of merger price effects. Int J Ind Organ. 2017; 52: 216–247.