A GENERALIZED CONCEPT OF CONCENTRATION AND ITS MEASUREMENT

by

Gerhart BRUCKMANN

Appendix: Computer Program
by Erich WERNER

Research Memorandum No.43
June 197o

C o n t e n t s

> "Our ultimate subject is not a collection of objects (establishments, firms), but a relationship among them, like gravity or electricity, which can be "seen" only by its effects".
>
> M.A.Adelman, {2}, p.37o

I The Problem

The concepts of concentration are several, not one.

1) First and foremost, one is likely to speak of an industry or a sector to be highly concentrated when a high percentage of total turnover (or labor force, production, etc.) is accounted for by a small number of (large) firms, or of a small percentage of the firms; or, in more general terms, when the size distribution is strongly dispersed.
2) Secondly, concentration may arise from the fact that many of the firms of a certain sector are interlinked in some way or other. On first sight, the size distribution within a given branch might not be too strongly dispersed; but some (many, or all) of the firms might have combined their marketing actions. They might have exchanged part of their stock. Some might be under common management or might be owned by one holding company.
3) Furthermore, "concentration" on the domestic market may stem from the influence of foreign firms who might have acquired a substantial share of domestic sales in their respective branch.
4) Yet another aspect of concentration is given by the fact that large enterprises often are engaged in a wide variety of activities, even within a single plant. The "monopoly power" of such firms, therefore, is much stronger than the share of one of its activities shows. Furthermore, efforts to group companies in order to measure concentration by product or by industry or by industrial segment encounter much uncertainty as to the proper location of the boundaries of the group. In measurements through time, the difficulty is compounded by frequent changes in the activities that companies carry on.

In the vast literatur on concentration, all of those aspects have been analysed and discussed thoroughly in a qualitative manner. Any quantitative investigation, however, and, in particular, all indices of concentration developed so far have restricted themselves to the first aspect of concentration only. The customary approach to the concept of concentration has been the following:

N entities (individuals, firms) T_i (i = 1,2,...,N) are given. A value p_i (share of total income, share of total turnover and the like) is assigned to each of these entities, with

$$\sum_{i=1}^{N} p_i = 1$$

Two types of indices can be distinguished: indices of relative concentration (inequality) or indices of absolute concentration (concentration).

In a recent paper {3}, certain criteria were established that an index must meet to qualify either as an index of relative concentration or as an index of absolute concentration. The most important distinction between them is the following: If a random sample of n entities (0 < n < N) is taken, a coefficient of relative concentration computed from the sample of size n should be independent of n; more precisely, its expected value should equal the population coefficient of relative concentration. The value of a coefficient of absolute concentration, on the other hand, clearly depends on n.

Neither index is superior to the other; in some problems, a coefficient of relative concentration, in other problems a coefficient of absolute concentration gives the appropriate answer.

It was shown that the GINI-coefficient (the ratio of the area between Lorenz-curve and diagonal to the area of the triangle) qualifies as an index of relative concentration, whereas an entire class of coefficients was shown to qualify as indices of absolute concentration (cfr.{3}).

One special case of the latter class of coefficients is the well-known HERFINDAHL-index, (also attributed to HIRSCHMAN): [1]

$$H = \sum_{i=1}^{N} p_i^2$$

Another special case of this class is the exponential of the entropy-measure $\sum p_i \log p_i$,

$$\Pi \; p_i^{p_i} \quad .$$

Of the second, third and fourth aspect of concentration mentioned in the introductory remarks, it may well never be possible to express the fourth one quantitatively, at least not in the form of one coefficient. In this paper, however, a methodological solution will be offered to express quantitatively, in one coefficient, concentration arising both from the first and the second aspect, viz. from a dispersed size distribution and from relations between firms. This index, defined in section IV under (5), is applicable whenever

a) it is possible to assess, at least roughly, the shares of individual firms and the strength of the relation between any two of them;

b) these relations can, to a sufficient degree, be considered symmetric.

On first sight, a relation between two firms will rarely be symmetric. If firm T_i (with a share p_i), e.g., owns the majority of the stock of firm T_j (with a share p_j) the relation r_{ij} between T_i and T_j is one of dominance. For the purpose of measuring concentration in that branch, however, it might often be irrelevant whether T_i dominates T_j or vice versa: the fact remains that the two firms appear on the market not in a separated, but rather in a combined way.

In section II and III this concept of "relation" will be formalized; in section IV the index will be defined; section V gives a brief comment on the third aspect of concentration (influence of foreign firms), sections VI through IX show properties of the index and give some applications.

---

[1] For a recent paper on the H-index, cfr.{1}.

II Relations amongst entities

The concept of a possible interrelation between entities (firms) as outlined above can be formalized as follows:

Between any pair of entities T_i, T_j $(i,j = 1,2,\ldots,N)$ a <u>relation</u> r_{ij} is defined, with the following properties:

1) $0 \leqq r_{ij} \leqq 1$ (1)

The more interdependent T_i and T_j, the larger r_{ij}; if T_i and T_j are entirely independent, $r_{ij} = 0$; if T_i and T_j are totally interdependent, $r_{ij} = 1$.

2) $r_{ij} = r_{ji}$ (2)

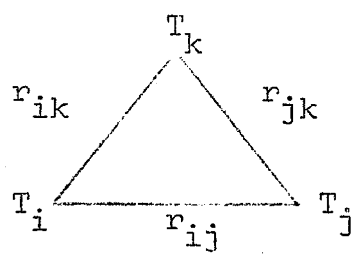
The value of r_{ij} does not express the direction of dependence. Whether firm T_i determines the business policy of T_j, or vice versa, does not affect r_{ij}.

3) $r_{ii} = 1$ (3)

For formal reasons, it is necessary to set the value of the relation of an entity with itself equal to 1 (total dependence).

4) $r_{ij} \geqq r_{ik}\, r_{jk}$ $(i,j,k = 1,2,\ldots,N)$ (4)

This condition, which is essential for the following, requires justification. Suppose both T_i and T_j are totally dependent on T_k; (for example, company T_k is the holding company for both T_i and T_j). We then have $r_{ik} = 1$ and $r_{jk} = 1$. It would, apparently, not make sense to allow $r_{ij} = 0$; it would not even make sense to allow $r_{ij} < 1$, since the total dependence of T_i and T_j on T_k <u>implies</u> total dependence also between T_i and T_j.



Now let us suppose $r_{ik} = .5$ and $r_{jk} = .5$ (whatever the economic meaning of these values may be). Again it would not be justified to let $r_{ij} = 0$; on the other hand, it makes sense to allow $r_{ij} < .5$. [2]

Condition (4), as stated above, provides a useful tool for determining the value of an <u>induced</u> relation r_{ij} that results from the existence of relations r_{ik} and r_{jk}. In other words: r_{ik} and r_{jk} <u>induce</u> together a value $r_{ij} = r_{ik} \cdot r_{jk}$, unless there exists an autonomous value $r_{ij} > r_{ik} \cdot r_{jk}$.

Examples:

a) $r_{ik} = r_{jk} = 0 \rightarrow \quad r_{ij} \geqq 0$

r_{ij} may, in fact, have any value $0 \leqq r_{ij} \leqq 1$.

b) $r_{ik} = r_{jk} = 1 \rightarrow \quad r_{ij} = 1$

c) $r_{ik} = r_{jk} = .5 \rightarrow \quad r_{ij} \geqq .25$

Of course, r_{ij} may have an autonomous value $r_{ij} > .25$.

d) $r_{ik} = .3, \quad r_{jk} = .7 \rightarrow \quad r_{ij} \geqq .21$

If, in this example, r_{ij} had an autonomous value of $r_{ij} < .21$ (for example: $r_{ij} = .15$), the induced value outweighs (and therefore: substitutes) the autonomous value. If, however, the autonomous value is, e.g., $r_{ij} = .8$, this would - in turn - have an effect upon r_{ik}, inasmuch as $r_{ik} \geqq r_{ij} \cdot r_{jk} = .8 \cdot .7 = .56$.

---

2 These properties are not implausible, insofar as the negative logarithm of $r_{ij} > 0$

$$d_{ij} = -\log r_{ij}$$

represents an ecart (a pseudometric) on the set T, fulfilling

1) $d_{ii} = 0$

2) $d_{ij} = d_{ji}$

3) $d_{ij} \leqq d_{ik} + d_{kj}$

III Matrix of relations

Considering the four properties stated above, we can now construct the matrix of relations

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1N} \\ r_{21} & r_{22} & \cdots & r_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ r_{N1} & r_{N2} & \cdots & r_{NN} \end{pmatrix}$$

First step: Insertion of autonomous relations

We insert any given autonomous relations, including the elements in the main diagonal ($r_{ii} = 1$, $i = 1,2,\ldots, N$). As an example, let us consider the case $N = 6$ with $r_{12} = .3$, $r_{34} = .3$, $r_{35} = .5$, $r_{45} = .9$, $r_{46} = .3$, $r_{56} = .5$. The (preliminary) matrix R then looks as follows:

$$R_{(prel.)} = \begin{pmatrix} 1 & .3 & 0 & 0 & 0 & 0 \\ .3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & .3 & .5 & 0 \\ 0 & 0 & .3 & 1 & .9 & .3 \\ 0 & 0 & .5 & .9 & 1 & .5 \\ 0 & 0 & 0 & .3 & .5 & 1 \end{pmatrix}$$

Second step: Revision of the (autonomous) relations

All elements $0 \leq r_{ij} < 1$ must be checked to see whether they satisfy condition (4) with respect to any T_k. Whenever there is a (preliminary) autonomous $r_{ij} < r_{ik} \cdot r_{jk}$, substitute $r_{ij} = r_{ik} \cdot r_{jk}$.

An easy algorithm for detecting and substituting all $r_{ij} < r_{ik} \cdot r_{jk}$ is [3] :

---

3 A different (equivalent) algorithm which lends itself better for computer application is given in the appendix.

Compute all products $r_{ik} \cdot r_{jk}$ $(i \neq k, j \neq k, r_{ik} > 0, r_{jk} > 0)$ and order them by size. Check in descending order, whether condition (4) is satisfied. Whenever it turns out that $r_{ik} \cdot r_{jk} > r_{ij}$, substitute the value of the product $r_{ik} \cdot r_{jk}$

$$r_{ij(rev.)} = r_{ik} \cdot r_{jk} \quad ^{4}$$

Immediately after this substitution, recalculate all products in which r_{ij} is a factor (if we had formerly $r_{ij} = 0$, these products are calculated for the first time) and insert the new values in the descending order of all products in their proper places before proceeding to the next (smaller) product. By this procedure, any newly created induced $r_{ij} > 0$ will, in turn, participate in the (possible) creation of (smaller) induced elements r_{lm} .

In our example, we find

$$r_{35} \cdot r_{45} = .45 > r_{34} \rightarrow r_{34(rev.)} = .45$$

$$r_{45} \cdot r_{56} = .45 > r_{46} \rightarrow r_{46(rev.)} = .45$$

$$r_{35} \cdot r_{56} = .25 > r_{36} \rightarrow r_{36(rev.)} = .25$$

The final matrix R, in our example, is therefore

$$R = \begin{pmatrix} 1 & .3 & 0 & 0 & 0 & 0 \\ .3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & .45 & .5 & .25 \\ 0 & 0 & .45 & 1 & .9 & .45 \\ 0 & 0 & .5 & .9 & 1 & .5 \\ 0 & 0 & .25 & .45 & .5 & 1 \end{pmatrix}$$

[4] The need to compute the value of an induced relation may also arise from the fact that some autonomous relation r_{ij} is unknown. In this case, the induced value $r_{ij(rev.)} = r_{ik} \cdot r_{jk}$ serves as a substitute (in fact: the smallest possible substitute) for the true, but unknown r_{ij}.

As one can see, we have two groups of entities: T_1 and T_2 on one side, T_3, T_4, T_5 and T_6 on the other side. There are relations $r_{ij} > 0$ within each group, but none between the two groups.

To give a second example, let us see how R would be altered if just one relation between the two groups is added, e.g., $r_{23} = .5$:

$$R_{(prel.)} = \begin{pmatrix} 1 & .3 & 0 & 0 & 0 & 0 \\ .3 & 1 & .5 & 0 & 0 & 0 \\ 0 & .5 & 1 & .3 & .5 & 0 \\ 0 & 0 & .3 & 1 & .9 & .3 \\ 0 & 0 & .5 & .9 & 1 & .5 \\ 0 & 0 & 0 & .3 & .5 & 1 \end{pmatrix}$$

This one additional relation $r_{23} = .5$ induces the relations $r_{13} = .15$, $r_{14} = .0675$, $r_{15} = .075$, $r_{16} = .0375$, $r_{24} = .225$, $r_{25} = .25$, $r_{26} = .125$. This yields

$$R = \begin{pmatrix} 1.0000 & .3000 & .1500 & .0675 & .0750 & .0375 \\ .3000 & 1.0000 & .5000 & .2250 & .2500 & .1250 \\ .1500 & .5000 & 1.0000 & .4500 & .5000 & .2500 \\ .0675 & .2250 & .4500 & 1.0000 & .9000 & .4500 \\ .0750 & .2500 & .5000 & .9000 & 1.0000 & .5000 \\ .0375 & .1250 & .2500 & .4500 & .5000 & 1.0000 \end{pmatrix}$$

By introducing the autonomous $r_{23} = .5$ each single entity is now linked to each other one, at least to some small degree. Obviously, any matrix R constructed by the above procedure fulfills the four initial conditions.

IV A generalized coefficient of concentration

In the introduction, it was stated that "concentration" amongst entities may be attributed to two phenomena, namely

a) to an unequal distribution of shares amongst entities,
b) to (stronger or weaker) relations amongst entities.

The H-index measures the first, the matrix of relations R measures the second.

Let us write the vector of shares

$$p = \begin{pmatrix} p_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ p_N \end{pmatrix} \qquad p' = (p_1, \ldots, p_N)$$

A coefficient of concentration that combines both phenomena can now be defined as the quadratic form

$$G = p'R\,p = \sum_{i=1}^{N} \sum_{j=1}^{N} r_{ij}p_ip_j \qquad (5)$$

Let us proceed to investigate the properties of this coefficient.

a) If $r_{ij} = 0$ for all $i \neq j$,

$$G = p'I\,p = p'p = H$$

The H-index, therefore, can be considered a special case of G, the case in which all T_i are independent.

b) Maximum concentration is attained when all $r_{ij} = 1$ $(i,j = 1,2,\ldots,N)$, regardless of the size distribution p_i. In this case all entities are completely dependent and

$$G = \sum \sum p_ip_j = \sum p_i \sum p_j = 1$$

Obviously, maximum concentration is also attained if there is only one entity T_1 with $p_1 = 1$ and $p_2 = \ldots = p_N = 0$. In this case

$$G = r_{11}p_1^2 = 1$$

c) If $r_{ij} = 1$, it follows from condition (4) that

$$r_{ik} \geq r_{ij}r_{jk} = r_{jk} \quad \text{and}$$

$$r_{jk} \geq r_{ij}r_{ik} = r_{ik}, \quad \text{hence:}$$

$$\text{If } r_{ij} = 1, \quad r_{ik} = r_{jk} \; (k = 1,2,\ldots,N) \qquad (6)$$

Let us now consider a subset of n entities $(1 < n < N)$; without loss of generality it may be assumed that this subset consists of the last n entities. For any pair T_i and T_j belonging to the subset, let $r_{ij} = 1$. We then have

$$G = \sum_{i=1}^{N} \sum_{j=1}^{N} r_{ij}p_ip_j = \sum_{i=1}^{N-n} \sum_{j=1}^{N-n} r_{ij}p_ip_j + \sum_{i=1}^{N-n} \sum_{j=N-n+1}^{N} r_{ij}p_ip_j +$$

$$+ \sum_{i=N-n+1}^{N} \sum_{j=1}^{N-n} r_{ij}p_ip_j + \sum_{i=N-n+1}^{N} \sum_{j=N-n+1}^{N} r_{ij}p_ip_j =$$

$$= \sum_{i=1}^{N-n} \sum_{j=1}^{N-n} r_{ij} p_i p_j + \sum_{j=N-n+1}^{N} p_j \sum_{i=1}^{N-n} r_{i,N-n+1} p_i + \sum_{i=N-n+1}^{N} p_i \sum_{j=1}^{N-n} r_{N-n+1,j} p_j +$$

$$+ \sum_{i=N-n+1}^{N} p_i \sum_{j=N-n+1}^{N} p_j$$

We set $\sum_{i=N-n+1}^{N} p_i \equiv \sum_{j=N-n+1}^{N} p_j = p^{\times}_{N-n+1}$ and $p_i = p^{\times}_i$ $(i = 1,2,\ldots,N-n)$

We then have

$$G = \sum_{i=1}^{N-n} \sum_{j=1}^{N-n} r_{ij} p^{\times}_i p^{\times}_j + 2p^{\times}_{N-n+1} \sum_{i=1}^{N-n} r_{i,N-n+1} p^{\times}_i + (p^{\times}_{N-n+1})^2 =$$

$$= \sum_{i=1}^{N-n+1} \sum_{j=1}^{N-n+1} r_{ij} p^{\times}_i p^{\times}_j \qquad (7)$$

In other words: If there is a subset of n entities which are totally dependent, G yields exactly the same value as if the subset of n entities were taken as one entity.

d) We compare two states of concentration, Z and $Z^{\times}$, characterized by p and R, $p^{\times}$ and $R^{\times}$ respectively. Let $p = p^{\times}$, but for one or more elements of R $r^{\times}_{ij} > r_{ij}$. It follows immediately from (5) that $G^{\times} > G$. In other words: Any increase in the strength of the relation between any two entities yields an increase of G.

e) We compare again two states of concentration, Z and $Z^{\times}$, characterized by p and R, $p^{\times}$ and $R^{\times}$, respectively. Let $R = R^{\times}$, but

$$p = \begin{pmatrix} p_1 \\ \vdots \\ \vdots \\ \vdots \\ p_N \end{pmatrix} \qquad p^{\times} = \begin{pmatrix} p_1 \\ \vdots \\ p_{k-1} \\ p_k - \varepsilon \\ p_{k+1} \\ \vdots \\ p_{m-1} \\ p_m + \varepsilon \\ p_{m+1} \\ \vdots \\ p_N \end{pmatrix} \qquad (p_m \geqq p_k, \varepsilon > 0) \qquad (8)$$

In other words, we proceed from Z to $Z^{\times}$ by shifting a (small) amount ε from T_k to T_m ($p_m \geqq p_k$). In {3}, it had been stated as a requirement which any coefficient of (relative or absolute) concentration C must fulfill that $C^{\times} > C$ if such a shift is carried out.

As an example, the H-index gives

$$H^{\times} - H = \sum (p_i^{\times})^2 - \sum p_i^2 = 2\varepsilon(p_m - p_k + \varepsilon) > 0 \qquad (9)$$

It can be shown, however, that this requirement is not always met by the generalized coefficient G, and, furthermore, that this is not a deficiency of G but rather a meaningful property.

For G, we have in this case

$$G^{\times} - G = \sum_{i=1}^{N} \sum_{j=1}^{N} r_{ij} \, (p_i^{\times} p_j^{\times} - p_i p_j)$$

Inserting (8), we obtain after some algebra

$$G^{\times} - G = 2\varepsilon\Big(\sum_{i \neq k,m} (r_{im} - r_{ik}) p_i + (1 - r_{km})(p_m - p_k + \varepsilon)\Big) \qquad (1o)$$

$$= 2\varepsilon \, \Big(\sum_{i=1}^{N} (r_{im} - r_{ik}) p_i + (1 - r_{km}) \varepsilon\Big) \qquad (11)$$

The special case when $r_{km} = 1$ is quickly discussed. From what has been said above under c) we expect the shift of ε from T_k to T_m not to alter the value of G. In fact, from (6) it follows that in (11) every $r_{im} = r_{ik}$ ($i=1,2,\dots,N$) and we obtain immediately $G^{\times} - G = 0$.

Now let $r_{km} < 1$.

Interpreting (1o) we can state: The increase in concentration $G^{\times} - G$ will be the larger

1) the larger ε;
2) the larger the values of r_{im} ($i \neq k,m$), i.e., the stronger the relations between T_m and the entities T_j ($i \neq k,m$);
3) the smaller the values of r_{ik} ($i \neq k,m$), i.e., the weaker the relations between T_k and the entities T_i ($i \neq k,m$);
4) the smaller r_{km};
5) the larger $p_m - p_k$, the difference between the shares of T_m and T_k.

It is, however, possible, that $G^{x} - G < 0$. From (1o) we obtain - because of $\varepsilon > 0$ - as a necessary and sufficient condition for $G^{x} - G < 0$

$$(1-r_{km})(p_m-p_k+\varepsilon) < \sum_{i \neq k,m} (r_{ik}-r_{im})p_i \tag{12}$$

From (4) we have

$$r_{im} \geqq r_{ik} \cdot r_{km}$$

$$r_{ik}-r_{im} \leqq r_{ik}(1-r_{km})$$

Inserting in (12), we obtain an upper limit for the right hand side

$$(1-r_{km})(p_m-p_k+\varepsilon) < \sum_{i \neq k,m} r_{ik}(1-r_{km})p_i$$

Because of $r_{km} < 1$, we finally have

$$p_m-p_k+\varepsilon < \sum_{i \neq k,m} r_{ik}p_i \tag{13}$$

as a necessary (though not sufficient) condition for $G^{x} - G < 0$, or, vice versa,

$$p_m-p_k+\varepsilon > \sum_{i \neq k,m} r_{ik}p_i \tag{14}$$

as a sufficient (though not necessary) condition for $G^{x} - G > 0$.

In other words: If the relations between T_k and the other T_i ($i \neq k,m$) are quite strong as compared to the relations between T_m and the other T_i ($i \neq k,m$) and if ε and p_m-p_k are not too large, the increase in concentration due to the shifting of ε from T_k to T_m -resulting in an increase of inequality among shares - may be more than outweighed by the decrease in concentration due to a lesser effect of the existing relations.

Example: $N = 3$; $p_1 = .25$, $p_2 = .35$, $p_3 = .4o$; $r_{12} = .6$, $r_{13} = r_{23} = 0$; $k = 2$, $m = 3$, $\varepsilon = .o5$; or else

$$p = \begin{bmatrix} .25 \\ .35 \\ .4o \end{bmatrix} \qquad p^{x} = \begin{bmatrix} .25 \\ .3o \\ .45 \end{bmatrix} \qquad R = \begin{bmatrix} 1 & .6 & 0 \\ .6 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

For the H-index we have

$$H = \sum p_i^2 = .345, \quad H^x = \sum (p_i^x)^2 = .355, \quad H^x - H = +.010$$

For the coefficient G, we have

$$G = \sum r_{ij} p_i p_j = .450, \quad G^x = .445, \quad G^x - G = -.005$$

The relation between T_1 and T_2 ($r_{12} = .6$) is so strong that the shift of $\varepsilon = .05$ from T_2 to T_3 results, altogether, in a smaller concentration than before.

V The influence of foreign firms on (domestic) concentration

Measurement of the effect on concentration exerted by foreign firms competing on the domestic market is not a methodological question but requires only a redefinition of the entities T_i (or of the shares p_i, respectively). Instead of using shares of total production or total employment, one could, e.g., define p as the vector of shares of domestic sales [5]. The T_i, then, are all firms competing on the domestic market, whether domestic or foreign.

This definition allows automatically also for any possible relations between domestic and foreign firms: Foreign firms may be controlled by domestic ones, or vice versa, and there may exist all kinds of weaker relations between foreign and domestic firms just as much as among domestic firms.

How to define the T_i and what variable to base p upon in a particular application, however, is not a statistical problem but rather an economic one.

VI The effect on G of neglecting small firms

In practice, it is often the case that market shares p_i are known of the n largest firms of a branch, whereas little is known of the distribution of market shares amongst the remaining N - n small firms. It might even be that N itself is unknown.

If we arrange the firms in descending order and denote the total market share of the N - n small firms as α, we have

$$\sum_{i=1}^{n} p_i = 1-\alpha, \qquad \sum_{i=n+1}^{N} p_i = \alpha, \tag{15}$$

---

5 cfr.{5}, p.295 ff.

In his recent paper, ADELMAN {1} has shown that an upper limit for H can be obtained if only $H' = \sum_{i=1}^{n} p_i^2$ is available: Because $p_i \leq p_n$ $(i = n+1, n+2, \ldots, N)$, the number of small firms must be at least $\frac{\alpha}{p_n}$, which is the case when all small $p_i = p_n$ $(i = n+1, n+2, \ldots)$. In this case, H assumes its maximum value as

$$H_{max} = \sum_{i=1}^{n} p_i^2 + \frac{\alpha}{p_n} \cdot p_n^2 = H' + \alpha p_n \qquad (16)$$

If M is an upper limit of the (unknown) total number of firms we have as a lower limit for the true value of H

$$H_{min} = \sum_{i=1}^{n} p_i^2 + \sum_{i=n+1}^{M} \left(\frac{\alpha}{M-n}\right)^2 = H' + \frac{\alpha^2}{M-n} > H' \qquad (17)$$

The problem is aggravated if we consider G. The r_{ij} existing between any two of the n larger firms might be assessed, but not the r_{ij} between large and small firms or between small firms. Considering (15), we can compute only the concentration measure for the n largest firms:

$$G' = \sum_{i=1}^{n} \sum_{j=1}^{n} r_{ij} p_i p_j$$

How far can G' deviate from the "true" value G, defined for all N firms?

As to the upper limit, it follows immediately from the definition that the value G can steadily be increased until all relations between small firms are equal to one. It has been shown that, in such a case, such firms can be considered as a unit; to obtain G_{max} we can therefore consider the sum of the small firms as one entity T_{n+1} with a value $p_{n+1} = \alpha$.

Hence, the problem reduces to the following question: What values $r_{i,n+1}$ $(i = 1,2,\ldots,n)$ will maximize G ?

We have an obvious additional condition that the values $r_{i,n+1}$ should not alter the given values r_{ij} $(i,j = 1,2,\ldots,n)$; in other words, no (unknown) value $r_{i,n+1}$ shall be of a size that would affect any value of the relations between the n large firms. It can be shown that G attains its maximum value without violating this additional condition if there exists an $r_{i,n+1} = 1$ between T_{n+1} and a particular one of the T_i $(i = 1,2,\ldots,n)$. This particular T_i need not be the one with the largest share. Since the T_i need not be ordered by size, we may - without loss of generality - assume that this particular T_i is T_n. If we write $p_i^{\times} = p_i$ $(i = 1,2,\ldots,n-1)$, $p_n^{\times} = p_n+\alpha$, we have

$$G_{max} = \sum_{i=1}^{n} \sum_{j=1}^{n} r_{ij} p_i^{\times} p_j^{\times}$$

and, after some algebra,

$$G_{max} = G' + 2\alpha \sum_{i=1}^{n} r_{in} p_i + \alpha^2 \qquad (18)$$

Remembering that - except in the case of total concentration - at least one $r_{in} < 1$, we have, because of (15), as a rough upper limit

$$G_{max} < G' + 2\alpha(1-\alpha) + \alpha^2 = G' + 2\alpha - \alpha^2 < G' + 2\alpha \qquad (19)$$

The lower limit of G is attained by analogy with (17). If there are a large number of small firms (not exceeding M - n) and if there is no relation $r_{ij} < 0$ between any of them nor between small and large firms,

$$G_{min} = G' + \frac{\alpha^2}{M-n} > G' \; . \qquad (20)$$

Summarizing, we can state that the "true" value G will lie between the limits

$$G' < G' + \frac{\alpha^2}{M-n} < G < G' + 2\alpha \max_{(k)} \sum_{i=1}^{n} r_{ik} p_k + \alpha^2 < G' + 2\alpha \qquad (21)$$

In practice, however, it will almost always be possible to arrive at a narrower range from what little information on the small T_i is available. As a rough (upper) estimate of G one might assume that each of the large firms dominates (r_{ij} = 1) a number of small firms, the share of which is proportional to the share of the respective large firm ($p_i^{\times} = p_i \cdot \frac{1}{1-\alpha}$, i = 1,2,...,n):

$$G \doteq \sum_{i=1}^{n} \sum_{j=1}^{n} r_{ij} p_i^{\times} p_j^{\times} = \frac{1}{(1-\alpha)^2} G' \doteq G' \ (1+2\alpha) \tag{22}$$

Sometimes it may be possible to estimate, at least approximately, the <u>average</u> relation existing a) between big and small firms and b) between small firms.

Let us denote this average value as ρ. In our notation, this means that all $r_{ij} = \rho$ for $i \neq j$ except for the case in which <u>both</u> T_i and T_j belong to the subset of the n large entities. Proceeding as in (7), we obtain in this case

$$G = G' + \rho\left(1-(1-\alpha)^2\right) + (1-\rho)\sum_{i=n+1}^{N} p_i^2 \tag{23}$$

By analogy with (16),

$$\sum_{i=n+1}^{N} p_i^2 < \alpha p_n < \frac{\alpha}{n}$$

Therefore, the third term on the right hand side is small as compared to the second term and can be neglected. Hence, if ρ can be assessed, a fair estimate of G is given by

$$G = G' + \rho\left(1-(1-\alpha)^2\right) \tag{24}$$

VII <u>Application</u>

In practical application, obviously, the elements of R are not directly available. Whereas the p_i can very often be obtained from official statistics, the strength of the relation between two firms cannot be immediately measured, except in the case r_{ij} = 1 (one firm totally dependent on another).

The autonomous elements of R must, therefore, be assessed to the best of the knowledge of the investigator, from what ever information on relations between firms is available.

Economists will find it easy to point out a number of reasons why an accurate assessment of the r_{ij} is never possible. It is fallacious, however, to reject the use of G for these grounds; using H instead, means setting $r_{ij} = 0$ for any $i \neq j$. The error committed by setting $r_{ij} = 0$ is certainly greater than the error committed by a somewhat faulty assessment of the value of r_{ij}, whenever strong relations between firms actually exist.

For these reasons, the index G, based on even the roughest estimates of the autonomous elements of R, will reflect the concentration among the T_i far better than H, an index that disregards entirely the influence on concentration of relations between firms.

I hope economists will not consider it the typical hubris of a statistician if I call it a challenge to economists to try and find agreed rules for ascribing numerical values to certain kinds of relations existing in economic practice.

VIII An economic application

We analysed a certain branch of the Austrian industry consisting of (at the time of the investigation) 56 small and medium sized firms. For p, we used the number of employees as a share of total employment.

The following relations existed:

a) 46 of the 56 firms were members of a (rather loose) association to promote common interests; we valued this relation at .1 .

b) 32 of the 56 firms were members of a slightly stronger association (value .2).

c) 3 firms (of the top five firms) cooperated to prevent market disruptions, leaving the individual firms' decision power untouched in any other respect (value .3).

..

d) Three groups of firms, consisting of 4, 3 and 2 firms, respectively, had exchanged substantial portions of their shares (between 30% and 45%) and had mutually assumed certain management functions (value .5).

e) An even stronger alliance between 3 other firms was valued at .7 .

f) One firm holds the majority of the stock of two other firms, which we valued as .9 .

Whenever several relations existed between two firms, we took the value of the strongest relation.

After having revised the matrix as outlined in section III, we found

$$G = .2540$$

whereas

$$H = .1174$$

Furthermore, we used this example to check the effect on G of neglecting small firms (section VI). We assumed that the p_i and the r_{ij} were known for the 12 largest firms only; of the 44 remaining firms, the only thing known is their total share

$$\sum_{i=13}^{56} p_i = \alpha = .1673$$

We then have

$$H' = \sum_{i=1}^{12} p_i^2 = .1164 \quad \text{and}$$

$$G' = \sum_{i=1}^{12} \sum_{j=1}^{12} r_{ij} p_i p_j = .2220$$

It is not surprising that H' deviates less from H than G' deviates from G; the difference between H and H' consists of the 44 values p_i^2 $(i = 13, 14, \ldots, 56)$, whereas the difference between G and G' consists of the $56^2 - 12^2 = 2992$ values $r_{ij} p_i p_j$ not contained in G'.

For (16) and (17) we have

$$H_{max} = .1191$$

$$H_{min} = .1171$$

(18) and (19) yield

$$G_{max} = .4147 < .5286$$

Both upper limits gravely overestimate G, as they are based on the assumption of total dependence a) between all small firms and b) between all small firms and one of the large firms; in reality, most relations are valued as .1 or .2 . For the analogous reason, also (22), yielding .32o2, overestimates G.

For (2o) we have

$$G \gtrapprox .2227$$

(23) and (24), depending on the additional knowledge of an "average" relation between the small and big firms, come much closer: For $\rho = .1$, we obtain $G \approxeq .2535$ and $G \approxeq .2527$; for $\rho = .2$ we obtain $G \approxeq .2848$ and $G \approxeq .2841$. Hence, the true value $G = .254o$ is quite well hit inspite of the rough and "faulty" assessment of $\rho = .1$ or $\rho = .2$. In practical applications, therefore, (24) will serve as a useful estimate.

IX <u>An example from political science</u>

Rae {4} has applied the H-index to measure concentration of a legislative body.

Let us consider as an example the French parliament of 1968. Of the 482 seats, the Communist Party (T_1) held 34, the Federation of the Left (T_2) 57, the Center (T_3) 29, the Gaullists (T_4) 296, the Independent Republicans (T_5) 54 and the Conservatives (T_6) 12.

Converting into shares, we have

$$p = \begin{pmatrix} .071 \\ .118 \\ .060 \\ .614 \\ .112 \\ .025 \end{pmatrix}$$

and H = .41 .

If, however, the Gaullists and the Independent Republicans were considered one party, we have

$$p^{\times} = \begin{pmatrix} .071 \\ .118 \\ .060 \\ .726 \\ .025 \end{pmatrix}$$

and H = .55 .

The G-index permits a more precise analysis, considering <u>any</u> relations between the six parties.

a) The Communist Party and the Federation of the Left sometimes cooperate in the National Assembly. Usually, they join together for the second ballot in elections to the National Assembly (r_{12} = .3)

b) Especially in the field of foreign policy, the Federation of the Left and the Center are connected closely in the National Assembly. In many cases the two parties form coalitions for the second ballot, mainly against the Gaullists. Often, the behavior of the two parties in the National Assembly looks like the behavior of a combined moderate left opposition against a moderate right government (r_{23} = .5)

c) To combat the electoral coalitions formed by the Communists and the Federation of the Left, the Center and the Gaullists frequently form a coalition for the second ballot. There are also very close personal links between former MRP-members, who are now Gaullists, and the majority of former MRP-members now in the Center (r_{34} = .3)

d) Between the Center and the Independent Republicans, the Giscard d'Estaing-wing of the government, there is an additional connection. The only important difference between the Independent Republicans and the Gaullists is the question of European policy; the policy of the Center differs from the Gaullist policy and is almost the same as the policy of the Independent Republicans (r_{35} = .5)

e) A rather small common base between the Center and the Conservatives is the position taken towards certain aspects of the Gaullist policy, for example the pro-Arab policy of the government (r_{36} = .1)

f) The Gaullists and the Independent Republicans are united in the UDR (Democratic Union for the Republic) and in the government. On most issues the relation between these two groups is practically 1 (r_{45} = .9)

g) The Gaullists and the Conservatives (mainly the rest of the old Independents Antoine Pinay's) are allied in their opposition against all leftist parties, especially the Communists. The Conservatives, for example, favored the government's policy after the "revolution" of May 1968 (r_{46} = .3)

h) The Independent Republicans and the Conservatives have a common origin, the old Independent Party. Many personal relations still exist between the Pro-Gaullist and the Anti-Gaullist wing of the Independents (r_{56} = .5) .

Constructing the (preliminary) matrix of autonomous relations from these values, we obtain the matrix that was used as a second example in section III. From this matrix, we obtain the final matrix as computed at the end of section III and

$$G = .65$$

Comparing G and H, it can be seen that a substantial part of the concentration existing in the French parliament of 1968 is to be attributed to the relations between parties.

To measure concentration in legislative bodies by G rather than by H has several advantages:

1) When two legislative bodies are compared, one may be more dispersed, yielding a lower H-index; some of its parties, however, may have much stronger relations (like those between Independent Republicans and Gaullists in France). The G-index, measuring total concentration, resembles both features.

2) The same is true when comparing one legislative body at two different periods.

3) To split a (large) party into sub-divisions (factions) greatly effects H. On the other hand, such a split has no influence on the value of G if the relations between the factions are 1; the farther below 1 the value of the relations, the stronger its influence on the value of G.

References

{1} ADELMAN, M.A., "Comment on the "H"Concentration Measure as a Numbers-Equivalent", REV.EC.STAT.,LI (Feb.1969),99-1o1.

{2} ARNDT, H. (editor), "Die Konzentration in der Wirtschaft", (Berlin: Duncker & Humblot, 1961).

{3} BRUCKMANN, G., "Einige Bemerkungen zur statistischen Messung der Konzentration", METRIKA XIV (2-3,1969),183-213.

{4} RAE, D.W., "The Political Consequences of Electoral Laws", (New Haven and London: Yale University Press, 1967).

{5} THEIL, H., "Economics and Information Theory" (Amsterdam: North Holland Publishing Co., 1967).

Appendix: Subroutine MIC

by Erich Werner

This subroutine adjusts values r_{ij} to the conditions

$$r_{ij} \geqq r_{ik} \cdot r_{kj} \quad (i,j, = 1,2,\ldots,n \text{ and } k = 1,2,\ldots,n \text{ and } k \neq i,j) \tag{1}$$

where r_{ij} are the coefficients of a n by n symmetric matrix R:

$$0 \leqq r_{ij} \leqq 1 \tag{2}$$

with

$$r_{ij} = 1 \quad (i = 1,2,\ldots,n; \quad j = i). \tag{3}$$

1) Mathematical background

Explicitly, the given matrix R is of the form:

$$\begin{pmatrix} 1 & r_{12} & \cdot & \cdot & \cdot & r_{1n} \\ r_{21} & 1 & \cdot & \cdot & \cdot & r_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{n1} & r_{n2} & \cdot & \cdot & \cdot & 1 \end{pmatrix} \tag{4}$$

The number of different elements r_{ij} is equal to the binomial coefficient $\binom{n}{2}$. The total number N of inequalities of the form (1) is

$$N = \binom{n}{2} \cdot (n - 2) . \tag{5}$$

If all products P_s, according to (1),

$$P_s = r_{ik} \cdot r_{kj} \quad (s = 1,2,\ldots,N) \tag{6}$$

and the maximal product

$$P_{max} = \max (P_s) \quad (s = 1,2,\ldots,N) \tag{7}$$

are computed and the condition

$$r_{i_m j_m} \geqq r_{i_m k_m} \cdot r_{k_m j_m} \quad (r_{i_m k_m} \cdot r_{k_m j_m} = P_{max}) \tag{8}$$

is satisfied, it can be shown that N_1 inequalities satisfy condition (1). $N_1 = 3\ (n - 2)$ derives from $r_{i_m j_m}$, $r_{i_m k_m}$ and $r_{k_m j_m}$ respectively:

If $$r_{i_m k_m} \geqq r_{k_m j_m} \tag{9}$$

there is

$$r_{i_m j_m} \cdot r_{i_m k_m} \geqq r_{i_m j_m} \cdot r_{k_m j_m} \tag{1o}$$

and considering (2)

$$r_{i_m k_m} \geqq r_{i_m j_m} \cdot r_{k_m j_m} \tag{11}$$

According to (8) and (9) there is

$$r_{j_m k_m} \geqq r_{i_m j_m} \tag{12}$$

and considering (2)

$$r_{j_m k_m} \geqq r_{i_m j_m} \cdot r_{i_m k_m} \tag{13}$$

2) Programming considerations

The given matrix R is assumed stored columnwise in compressed form, i.e., lower triangular part only.

There are two versions of the program. Version one has been tested with a maximal size of the matrix R being $n_{max} = 15$. Version two has been computed with $n_{max} = 56$. It is seven times slower than version one but it needs approximately $2.n^3$ less storage capacity.

In accordance with the usual FORTRAN notation the variables $r_{i_m j_m}$ will be denoted in the following by r (im, jm).

Version 1

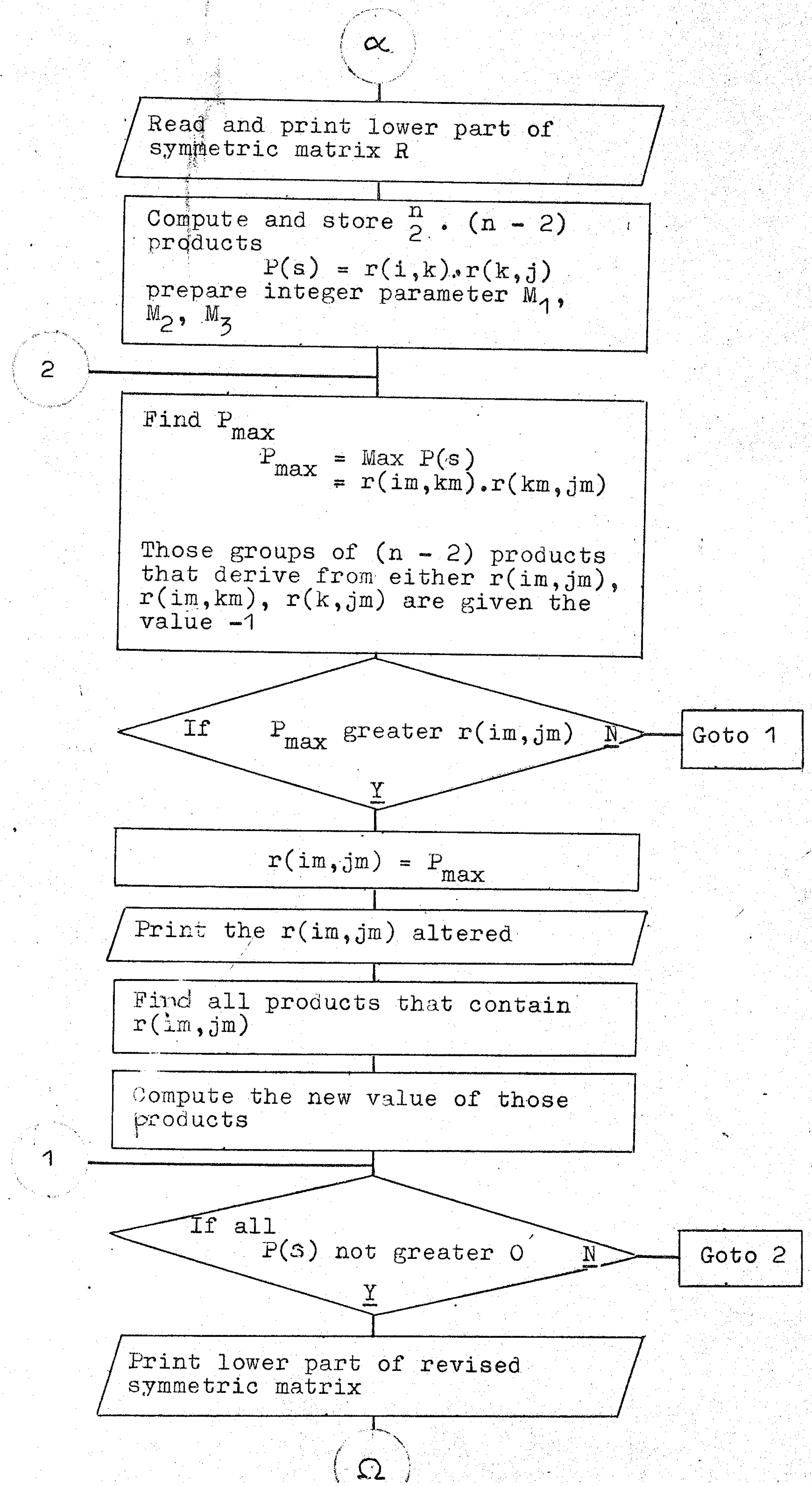
$\binom{n}{2}$. (n - 2) products P(s) = r (i,k) . r (k,j) are computed and stored. M_1, M_2, M_3 are onedimensional arrays that contain i, j, k respectively.

P_{max}P(s) = r (im, km) . r (km, jm). Those (n - 2) products that derive from either r (im, jm), r (im, km) or r (km,jm) are given the value -1 . If P_{max} is greater r (im, jm), then r (im, jm) shall be equal P_{max} and printed. All products that contain r (im, jm) are found and their new value is computed. If all P(s) are not greater zero the lower part of the revised symmetric matrix R is printed and the subroutine ends.

Version 2

Instead of listing all products, only those products are stored in the onedimensional array LEK that determine either r (im, jm), r (im, km) or r (km, jm). If the products for all r (i,j) are listed in LEK, or if P_{max} is not greater than the lower limit for r (i, j) VERG, the subroutine ends.

FLOWCHART VERSION 1



FLOWCHART VERSION 2