

An Entropy Analysis of Areal Concentration for Manufacturing Plants

Peter C. Luisa and B. L. Dillman*

INTRODUCTION

Recent changes in the structure of manufacturing industry in the United States are characterized by strong tendencies toward relocation of firms from the "old industrial belt" to the so-called "sunbelt states," and by decentralization of industry from the metropolitan to the nonmetropolitan counties. Such inclinations may be referred to as industrial dispersion. It is argued that this spreading out process is not a random one in which all manufacturing industry has suddenly become footloose, but rather that industry still seeks out such location factors as access to a suitable labor force, adequate services and amenities, and a sufficiency of natural resources such as water.

Previous research on the location of manufacturing industry in South Carolina has shown a significant relationship between manufacturing wages and employment and population potential (2). The latter is a distance-weighted measure of the accessibility of localities to population derived from the general gravity model.

The overall objective of this analysis was to see if industrial dispersion, as measured by a series of entropy indexes for South Carolina, is related to the population potential of various localities and to see whether this dispersion is different for recently established plants (since 1969). Further, we wished to see if the degree of dispersion is less for heavy water users, on the assumption that these firms are less footloose and thus less able to select areas of high population potential. A change in concentration of heavy water using plants was expected after about 1970 because of the new and anticipated federal and state water pollution control regulations of the late 1960's which, presumably, removed much of the incentive for locating on flowing streams.

THE ENTROPY CONCEPT

The use of a general entropy concept was first made in the field of physics where it was used as a theoretical measure of the ultimate, unavoidable decay of all matter into states of greater disorder and disarray.

*The authors are, respectively, Regional Economist, Corps of Engineers, Savannah District, and Professor, Department of Agricultural Economics and Rural Sociology, Clemson University.

#The work upon which this publication is based was supported in part by funds provided by the Office of Water Resources Research and Technology (Project No. A-039-SC), U.S. Department of the Interior, Washington, D.C., as authorized by the Water Research and Development Act of 1978. Technical Contribution No. 1816, South Carolina Agricultural Experiment Station.

Such a thermodynamic reference to entropy is one with which most students of the term are familiar: the greater the entropy of a system, the greater the state of disorder of its molecules and the lower is its total quantity of free energy.

Georgescu-Roegen related this concept to the economic process in general and to the economics of natural resources in particular (4). His thesis dealt with the concept that mankind is operating on a finite resource base which, in accordance with the entropy law, is being broken down into states of less useable energy.

Entropy is an index of concentration, or rather it is an inverse measure of concentration since, as the measure increases, greater disaggregation (more complete dispersion) is indicated. Entropy has proven to be a useful measure since it possesses a number of unique properties which make it superior to other simpler indexes used for measuring relative concentration, such as the Lorenz curve, the Gini coefficient or X-firm Concentration ratios.

An index of economic entropy as a measure of market disaggregation was first introduced by Henri Theil (9). Such a concept was an extension of the use of an entropy concept in communications theory by Shannon (8) and later by Horowitz (6,7) and has been shown to be capable of measuring the competitiveness as well as the concentration of firms in a region. Hexter and Snow applied the entropy index to measure the market concentration of the 500 largest industrial firms in the United States and how it changed from 1955 to 1966 (5). Garrison and Paulson used this same measure to show the distribution of manufacturing employment among counties classified by the largest city in each county (3). Yandle and Hite used the entropy coefficient to measure changing market shares, this time among cigarette producers (11). Finally, Blalock and Doeksen measured the entropy of new industry in Oklahoma among counties of varying populations and statistically tested the change in entropy over an eleven-year time period (1).

Mathematically, the entropy index is computed as:

$$E = \sum_{i=1}^N y_i \ln 1/y_i$$

where

E = the entropy index;

y_i = percentage share of the i th unit of observation; and

N = total number of observations.

Due to the inherent nature of the index each set being measured has its own particular range of possible entropies, with total concentration being represented by an entropy of zero and total dispersion indicated by an entropy equal to the natural logarithm of N. The entropy measure is a favored tool for concentration analyses because it is sensitive to such factors as: 1) the share of the unit being measured tending toward only a few observations, 2) the number of observations in the data set, and 3) the relative share of each observation. Moreover, the total entropy index can

be disaggregated into between set and within set entropies, where a set is a particular aggregation of the observations being measured. Mathematically this is expressed as:

$$E = \sum_{s=1}^S Y_s \ln 1/Y_s + \sum_{s=1}^S \left[Y_s \sum_{j=1}^{N_s} \left(\frac{y_{sj}}{Y_s} \ln \frac{Y_s}{y_{sj}} \right) \right]$$

where

S = total number of sets;

N_s = number of observations in the s th set;

Y_{sj} = an observation included in the s th set; and

$$Y_s = \sum_{j=1}^{N_s} y_{sj}.$$

The left side of the right hand term is known as between set entropy, while the right side is the weighted sum of the within set entropies.

The between set entropy index measures the amount of dispersion between all separate sets. Such a number is particularly useful in time series studies in that changes in the value reflect a tendency towards concentration of the data in one set, or a few sets, or dispersion between all the various sets or classes.

The within set entropy concept measures actual dispersion of the data within each individual class.

The addition of the two entropy forms yields the total entropy measure for the entire data set, reflecting the tendency towards dispersion for the data considered as a whole.

One other interesting and useful aspect of the entropy index is that, due to the existence of a maximum and minimum value for disaggregation, a percentage of the maximum value can be computed for all classes of observations. Using these percentages, the classes can be compared with each other and any unique tendencies analyzed.

OBJECTIVES

The objectives of this research were to use the entropy index as a tool to examine the dispersion of manufacturing in South Carolina for testing the following null hypotheses:

- 1) that manufacturing employment is randomly dispersed throughout all areas of the state of South Carolina without regard to population potentials;
- 2) that the locational tendencies of high water-using firms do not differ from those of low water-using firms;
- 3) that the degree of dispersion does not differ for manufacturing plants established prior to 1970 and those established from 1970 through 1977.

THE DATA

The data utilized in this study are a part of a generalized data set developed for multiple use in many types of plant location studies in South Carolina. This set contains data for each of the manufacturing firms in the state including its location by longitude and latitude, its employment, value of output, water usage and year of establishment.

Each of these firms is located in one of one hundred thirty two (132) equal sized grid squares which make up a grid network for South Carolina. The system was modified by eliminating seven grids which were not viable choices for location, leaving one hundred twenty five (125) areas used for the analysis.

Population potentials were calculated for the midpoint of each grid. The concept of a population potential and its uses are explained in detail in another paper by Dillman, Twilley, and Lytle (2), but the basic mathematical concept is as follows:

$$V_i = \sum_{j=1}^m \frac{P_j}{D_{ij}}$$

where

- V_i = population potential of the i th locality (grid);
- P_j = population of the j th census enumeration district;
- D_{ij} = distance in miles between the i th grid and the j th census enumeration district.

Distances are calculated from the center of each grid to the center of each enumeration district using an algorithm yielding linear distances from geographic coordinates. Grid squares were classified by population potentials and placed in five separate classes for analysis. Each class contains as equal a number of grids as possible (Table 1). It is to these five classes that the within and between set entropy formulas are applied.

Finally, the existence of water use data for each plant allowed classification by the intensity of water usage. Index numbers were calculated for each firm using water usage per dollar of output and water usage per employee. The firms were then aggregated by classifying the plants contained in the upper quartile of each set as high water users and those in the lower three quartiles as low water users, conforming to the Department of Commerce classification (10). The entropy formula was applied to each set to test for any differences in their tendencies toward aggregation.

Table 1.
Population Potential Class Descriptions

Class	Number of Grids	Potential Range
I	24	0-50,848
II	24	50,849-59,202
III	26	59,203-64,446
IV	26	64,447-71,408
V	25	71,409-216,665

RESULTS

The results of the entropy analysis of South Carolina employment and manufacturing firm spatial location are presented in Tables 2 and 3. The upper figures in each row are the actual entropy values, while the lower figures, in parentheses, are relative entropies calculated as percentages of the maximum dispersion possible (with maximum entropy each grid contains an equal share of employment or firms in the class). Discussion will center around the relative entropy figures since they are directly comparable.

The data have been categorized in a number of ways to test our previously posed hypotheses. It is readily apparent that total relative entropy values for the data comprised of all employment (Table 2) and all firms (Table 3) in South Carolina have decreased between the two time periods. Although the decreases are not as great as those computed from some of the other more finely disaggregated data in Tables 2 and 3, it is clear that there has been a greater tendency toward concentration for total firms and total employment since 1970.

Classifying the data on all employment into five population potential classes, we again see a tendency towards concentration of employment between the two time periods for all of the potentials, particularly class four (Table 2). On the other hand, the manufacturing plant data (Table 3) do not show nearly as consistent a tendency to concentrate over time, but the highest potential classes, four and five, do show a marked decrease in relative entropy. While plant locations became much more concentrated for potential classes four and five only, the employment aggregation is occurring over all classes over time.

The percentage share of employment and plants within each population potential class is tending toward equality over time. This is shown by the increases in between set entropy for all employment and all plants in Tables 2 and 3.

The data were further classified by the high/low water use criterion. Results indicate that there is a very consistent difference between the entropy values for high and low water-using firms (high water users being more concentrated) and that this difference is intensified over time. The employment concentration over time is depicted by the decreased total relative entropy measure in Table 2. For high water-using employment this measure fell by thirteen percent, while for all other employment it fell by less than one percent. This trend is exhibited in the individual classes of population potentials.

For each of the classes except class 5, high water-using employment has a lower relative entropy than that for low water users, signifying greater concentration. Furthermore, the relative entropy is lower for both water classes after 1969 than before, again with the exception of class 5, and such decreases are markedly greater for high water using employment than for all other employment.

The between set entropy figure reflects similar characteristics for the water use classification as for all plants: dispersion of employment over all

Table 2.
Entropy Values for South Carolina Manufacturing Employment, by Water Use and Population Potential
Classes, for Plants Established Prior to 1970 and from 1970 through 1977.

Entropy Type	All Employment		High Employment by Water/Employee		Low Employment by Water/Employee	
	before '70	'70 through '77	before '70	'70 through '77	before '70	'70 through '77
Total	4.01246 (83.1)*	3.92835 (81.4)	3.85057 (79.7)	3.18599 (66.0)	3.86063 (80.0)	3.83500 (79.4)
Between set	1.26670 (78.7)	1.44940 (90.1)	1.27813 (79.4)	1.49772 (93.0)	1.25218 (77.8)	1.42088 (88.3)
Weighted Σ within set	2.74575	2.47895	2.57244	1.68827	2.60846	2.41412
Potential classes						
1.	2.53185 (79.7)	2.16088 (68.0)	1.62139 (51.0)	1.36896 (43.1)	2.33097 (73.3)	1.82608 (57.5)
2.	2.85196 (89.7)	2.35817 (74.2)	2.53271 (79.7)	1.13513 (35.7)	2.81919 (90.7)	2.30174 (72.4)
3.	2.76770 (84.9)	2.31061 (70.9)	2.43502 (74.7)	1.56730 (48.1)	2.64032 (81.0)	2.13498 (65.5)
4.	2.83468 (87.0)	2.51733 (66.2)	2.71959 (83.5)	1.88609 (57.9)	2.74225 (84.2)	2.47687 (76.0)
5.	2.70594 (84.1)	2.63759 (81.9)	2.62796 (81.6)	1.91990 (59.6)	2.54408 (79.0)	2.65668 (82.5)

*Numbers in parentheses are the relative entropies: actual calculated entropy as a percentage of its maximum possible value.

Table 3.
Entropy Values for South Carolina Manufacturing Plants Established Prior to 1970 and from 1970 through 1977, by Water Use and Population Potential Classes

Entropy Type	All Plants		High Plants by Water/Employee		Low Plants by Water/Employee	
	before '70	'70 through '77	before '70	'70 through '77	before '70	'70 through '77
Total	4.03886 (83.6)*	3.79500 (78.6)	3.96473 (82.1)	3.50725 (72.6)	4.00414 (82.9)	3.89838 (80.7)
Between set	1.30015 (80.8)	1.33996 (83.3)	1.28856 (80.1)	1.43644 (89.3)	1.30248 (80.9)	1.31499 (81.7)
Weighted Σ within set	2.73872	2.45504	2.67616	2.07081	2.70165	2.58339
Potential classes						
1.	2.21949 (69.8)	2.22011 (69.9)	2.15796 (67.9)	1.56066 (49.1)	2.78084 (87.5)	2.27837 (74.8)
2.	2.41496 (76.0)	2.50200 (78.7)	2.76783 (87.1)	1.74782 (55.0)	2.98698 (87.7)	2.44123 (76.8)
3.	2.70094 (82.9)	2.77814 (85.3)	2.76924 (85.0)	1.97295 (60.6)	2.81338 (86.4)	2.50217 (76.8)
4.	3.17705 (97.5)	2.75952 (84.7)	2.88773 (88.6)	2.25381 (86.2)	2.82201 (86.6)	2.76641 (84.9)
5.	2.68557 (83.4)	2.30050 (71.5)	2.60527 (80.9)	2.21770 (69.9)	2.57653 (80.0)	2.60357 (80.9)

*Numbers in parentheses are the relative entropies: actual calculated entropy as a percentage of its maximum possible value.

classes of population potentials is increasing over time. Moreover, this dispersion is greater for high water users than low.

When classified according to plants (Table 3) the entropy values again reflected increased concentration over time but not as great as for employment. The total relative entropy for high water-using plants again decreased by about ten percent while all other plants fell by only two percent. The between set relative entropy figures show an increased dispersion of firms among the five potential classes over time as it did for the employment data. Also, a comparison of relative entropies for each class again shows increased concentration after 1970 for both high and low water-using firms. This was much more pronounced for high water users.

A note of departure in the overall trend is that for population potential class five, neither employment nor firms in the low water using category showed a decrease in relative entropy between the two time periods examined. Rather, the index increased slightly indicating, if anything, increased dispersion over time for that class. A possible explanation is that for non water oriented firm locations in urban (high population potential) areas there are many good locations, and dispersion increases over time as city sizes and city boundaries expand.

A similar analysis was also done for water using firms as classified by water per dollar of output. The analysis yielded nearly exactly the same results as those just described where firms were classified by water use per employee.

CONCLUSIONS

While no statistical test is employed in Tables 2 and 3, the consistency of all of the results derived from the entropy analysis argue strongly for rejection of all three of the original null hypotheses. The relative entropy comparisons have shown that manufacturing firms and employment do indeed concentrate along definite spatially defined lines; that high water-using employment and firms have more agglomerative tendencies than their low water-using counterparts; and that while this tendency towards concentration has increased from before to after 1970, the process is much more pronounced for high water users than for the low.

The results lead the authors to believe that further investigations into the degree and causes of spatial dispersion in South Carolina are fruitful areas of research. This analysis gives us an insight into the importance of water as a location factor for some manufacturing plants. While manufacturing plants and employment are becoming more dispersed among all population potential classes of locations, concentration is increasing within all of these classes. This tendency is much stronger for high water using employment and plants, probably indicating the necessity for these plants to locate near good water sources. On the other hand, non water oriented employment and plants show less tendency to concentrate in the 1970-77 period, and for the highest population potential class, no increased concentration at all.

The year chosen to separate our data (1970) was an arbitrary one as were

the five classes of population potentials. Further analyses will attempt to show a year by year change in location and to identify more exact characteristics of the areas of choice.

Population potential class designation has not shown any definite bearing on the entropy values although there is an indication that concentration is highest for classes one and five and that there may be an increase in dispersion with increasing values of potential up to a threshold level and then a decrease. More detailed analysis is required before conclusions can be drawn concerning critical levels of population potential.

REFERENCES

1. Blalock, M. G. and G. A. Doeksen. "An Examination of the Trend in Industrial Dispersion in Oklahoma from 1963 to 1974." *Southern Journal of Agricultural Economics*, Volume 9, No. 1, July, 1977.
2. Dillman, B. L., Wm. Edward Twilley, and J. S. Lytle. "A Population Potential Model for Community Manufacturing Employment." Paper presented at the Southern Regional Science Association Annual Meetings, April, 1978.
3. Garrison, C. B. and Albert Paulson. "Entropy as a Measure of Areal Concentration of Water Oriented Industry." *Water Resources Research*, Volume 9, No. 2, April, 1973.
4. Georgescu-Roegen, N. *The Entropy Law and the Economic Process*. Cambridge, Mass., Harvard University Press, 1971.
5. Hexter, J. L. and J. W. Snow. "An Entropy Measure of Relative Aggregate Concentration." *Southern Economic Journal*, Volume 36, No. 3, January, 1970.
6. Horowitz, A. and J. Horowitz. "Entropy, Markhov Processes and Competition in the Brewing Industry." *Journal of Industrial Economics*, Volume 16, No. 2, 1968.
7. Horowitz, Ira. "Numbers-Equivalents in U.S. Manufacturing Industries." *Southern Economic Journal*, Volume 37, No. 3, 1966.
8. Shannon, C. E. "A Mathematical Theory of Communication." *Bell System of Technology Journal*, Volume 27, 1948.
9. Theil, H. *Economics and Information Theory*. Amsterdam: North-Holland Publishing Company, 1967.
10. U.S. Bureau of the Census. *Census of Manufacturers*, 1972. Special Report Series: Water Use in Manufacturing, MC72(SR)-4. U.S. Government Printing Office, Washington, D.C., 1975.
11. Yandle, B. and A. Hite. "Branded Competition and Concentration Measures." *Southern Economic Journal*, Volume 44, No. 4, April, 1977.