An Information Theoretic Approach to Measurement of Spatial Inequality

J. A. WALSH
Carysfort College, Dublin

M. E. O'KELLY
McMaster University, Hamilton

Précis: In this paper we present some quantitative indices for measuring and comparing levels of inequality in discrete spatial distributions. The indices presented are based on the Kullback Information Measure which we consider to be the most general. In the final section of the paper the indices are applied to measure inequality in *per capita* income and certain categories of employment distribution in Ireland. The results indicate a tendency towards increasing uniformity in the spatial distribution of *per capita* incomes and employment opportunities.

I INTRODUCTION

Over the last two decades or so problems related to socio-economic or spatial inequality have been receiving increased attention. In many countries, governments have launched programmes aimed at reducing the extent of the inequality that exists within and between regions. However, despite the increasing concern with the equity goal, there has been very little research into methods of defining and measuring socio-economic and spatial inequality.

In their recent review of the literature on the theme of spatial inequality, Hinderink and Sterkenburg (1978) distinguished three main types of study, one of which was concerned with the use of space as a formal framework for the analysis of inequality. In their discussion they argued that techniques to measure spatial inequality in levels of development have received much attention and implied that measurement problems were being over-emphasised. While that may be so, it is argued in this paper that most of the commonly used measures suffer from some serious limitations. Thus, the aim of the paper is to establish quantitative indices for comparing discrete spatial

distributions with respect to inequality. Indices will be developed to facilitate cross-sectional and longitudinal analyses of inequality in spatial distributions for different levels of spatial aggregation.

The paper consists of three major sections which are followed by some conclusions. The first section consists of a brief review and assessment of some of the statistical inequality indices that are most frequently used by economists and geographers. This is followed by an outline of an approach to spatial inequality analysis based on measures from Information Theory. In Section III of the paper some procedures for information theoretic hypothesis testing will be discussed. It will be shown that the most general information measure can be decomposed or partitioned so as to facilitate the measurement of inequality and the testing of hypotheses about spatial distributions at different levels of spatial aggregation. In the final section of the paper there is an application of information statistics to the measurement of inequality in the spatial distribution of *per capita* incomes and certain categories of employment as defined by Baker (1966). Indices of spatial inequality have been computed for a number of spatial aggregations from data pertaining to four different points in time. Thus, finally, temporal trends in levels of inequality are identified.

II THE MEASUREMENT OF INEQUALITY

Already there is a wide variety of measures of equality (or its converse, inequality) available in the literature of the social sciences. A number of them have been recently compared and evaluated by Bartels and Nijkamp (1976) and Gaile (1977). Often income and other distributions are simply analysed for equality by various measures of central tendency. The simplest measure is probably the range of the values or some function of it, such as the range divided by the sum of all the values. Other relatively simple measures of central tendency that are often used are the standard deviation or the coefficient of variation or some weighted version of these (see, for example, Williamson, 1965; Martin, 1971; and Ross, 1977). These measures have a number of limitations. The range and standard deviation are scale-dependent. An assumption of normality or a transformation of the data is necessary if the standard deviation is to be useful for inference purposes. Furthermore, the coefficient of variation, as well as being affected by the magnitude of the mean, is, along with the standard deviation, restricted to data in interval or ratio form.

Two more commonly used inequality measures for income distributions are the Lorenz curve and the Gini coefficient. The Lorenz curve is a diagrammatical device which allows for both visual and quantitative comparison of the cumulative relationship between two variables (Smith, 1975, pp. 200-

204). This technique has a number of drawbacks when it is applied to income distributions (Atkinson, 1970). Major difficulties arise if one wishes to compare Lorenz curves that intersect. The Gini coefficient is a direct function of the Lorenz curve (Kendall and Stuart, 1958, p. 49) and seems to be unduly influenced by extreme values at the upper end of the curve. Furthermore, it does not lend itself to within- and between-set decompositions (Gaile, 1977). Some conceptual shortcomings of the measure have been discussed by Dasgupta *et al.* (1973), Rothschild and Stiglitz (1973) and Chipman (1974).

From the foregoing brief review it is clear that most of the commonly used inequality measures suffer from a number of limitations. It seems fair to assume that these limitations could affect the findings of inequality studies. In the remainder of this section of the paper we present an inequality measure based on concepts from information theory. The measure is neither scale- nor mean-dependent, it is distribution free, it is not unduly affected by extreme values and it can be used to analyse inequality at different geographical scales simultaneously.

Over the last decade a number of spatial analysts have become increasingly aware of the usefulness of information statistics for describing and analysing spatial distributions (Medvedkov, 1967 and 1970; Gurevich, 1969; Semple *et al.*, 1970, 1973 and 1977; and Haynes and Enders, 1975). At the same time a number of information measures have emerged and there has been some discussion concerning the situations for which each measure might be most appropriate (Pielou, 1966; Batty, 1974 and 1975; and Walsh and Webber, 1977). The different measures are briefly summarised here and the most appropriate ones for the hypotheses we wish to test are identified.

Most information measures rely on the proposition that the information received from an event is proportional to the improbability of the event happening. Let X be a discrete random variable that can attain any of M values, x_1, x_2, ..., x_M. Measurements on X are made N times. The probability that X = x_i is denoted as p_i, i = 1, 2, ..., M. Then the potential information obtained from the event X = x_i is defined to be $I(X = x_i) = -\log p_i$ (Goldman, 1953; Good, 1956). This measure is also known as the *self information content* of the event X = x_i (Ingels, 1974). A number of other interpretations have been given by Walsh (1976, p. 7). Since self information is also a random variable its mean value can be calculated as

$$H_1 = H_1(p_1, p_2, ..., p_M) = -\sum_{i=1}^{M} p_i \log p_i \qquad (1)$$

where $\sum_i p_i = 1$. This measure is commonly known as the Shannon entropy

or Shannon information measure, following his introduction of it to communication theory (Shannon, 1948). This is the measure that has been most commonly used in spatial analysis (Nutenko, 1970; Chapman, 1970 and 1973; Garrison and Paulson, 1973; and Marchand, 1975).

In recent years, however, there has been considerable controversy over the appropriateness of the Shannon measure for spatial analysis. On the one hand, it has been argued by Walsh and Webber (1977) that even if Shannon's were a valid measure, its range of application is limited to situations where there is only sample evidence about conditions in a large population. This is because the measure relies on a sampling with replacement process. When the data that are available describe conditions in a completely sampled population, then an appropriate information measure is one deduced from a sampling without replacement process as

$$H_2 = \frac{1}{N} \left(\log N! - \sum_{i=1}^{M} \log n_i! \right) \tag{2}$$

where n_i is the number of times the event $X = x_i$ occurs (Walsh and Webber, 1977, p. 402). Thus an application of H_1 assumes a particular sampling process.

Furthermore, among spatial analysts there has been some controversy and confusion concerning the choice of the Shannon measure or an alternative formulation known as the Kullback measure. This controversy arises from some restrictive properties of the Shannon measure. Both H_1 and H_2 are discrete, dimensionless and nonspatial. Their magnitude depends on M, the number of classes or types of events. They do not take into account the size of the classes or, in a spatial context, the size of the areal units over which the measurements are made. To overcome some of these difficulties Batty (1974) proposed a spatial entropy measure, S, which ensures that the size of the classes or areal units is considered explicitly. The measure is

$$S = \sum_{i=1}^{M} p_i \log (p_i/\Delta x_i) \tag{3}$$

where Δx_i is the fraction of the total area within which x_i occurs. Although this measure introduces the effect of the areal partitioning into the analysis, it has some limitations. It can be negative which could lead to difficulties in interpretation, especially since most information theorists insist that a measure of information should be non-negative (Renyi, 1970, p. 561). Further difficulties arise if one wishes to use the measure to compare spatial distributions in different sized systems (Batty, 1976, p. 4).

The limitations associated with spatial entropy, S, can be overcome by using a more general information measure that has been proposed by Kullback (1959). The measure is defined as the difference between a prior

distribution, $\{q_i\}$, and a posterior distribution, $\{p_i\}$ — sometimes called expected information or the information gained by comparing a state of ignorance encoded in the prior distribution with additional knowledge about reality as reflected in the posterior distribution (Batty, 1976, p. 5). Then the information gained from observing that $X = x_i$ is $\log \dfrac{1}{q_i} - \log \dfrac{1}{p_i} = \log \dfrac{p_i}{q_i}$. This mean information gain is

$$I = \sum_{i=1}^{M} p_i \log (p_i/q_i).\qquad(4)$$

Equation (4) is probably the most general of all information measures. Hobson (1969) has proved that under some general conditions the Kullback measure, I, is the expression for the information contained in a message that alters a prior probability distribution to a posterior one. An important property of the measure is that it is always positive for any distributions $\{p_i\}$ and $\{q_i\}$ (Theil, 1972; Tribus, 1969). Also, the order of magnitude of the measure is independent of the number of events. Thus, in a geographical context where the number of events may be the number of areal units (zones, counties), expected information can be used to compare inequality in distributions over systems with different numbers of units.

III INFORMATION THEORETIC HYPOTHESIS TESTING

In this section we discuss some ways in which information measures can be used to test hypotheses about spatial inequality in income distributions. The procedures developed here can also be applied to test hypotheses about inequality in employment patterns.

Let N and Y denote the total population and the total personal income arising in a country. The population and income of county i are denoted by n_i and y_i, respectively, where $\sum_{i=1}^{M} n_i = N$ and $\sum_{i=1}^{M} y_i = Y$ and M is the number of counties. A number of hypotheses may be tested. One relates the distribution of income to the form of the areal partitioning. The null hypothesis is that there is no inequality in the spatial distribution of income. Define prior and posterior probability distributions $\{q_i\}$ and $\{p_i\}$ as

$$q_i = \frac{1}{M} \text{ and } p_i = y_i/Y \text{ for all i.}\qquad(5)$$

On substituting these probabilities into equation (4), the expected information is defined as

$$I_1 = \frac{1}{Y} \sum_i y_i \log y_i - \log \frac{Y}{M} \qquad (6)$$

When there is no information gain in going from the prior to the posterior distribution, the income distribution is uniform — that is, there is no inequality. Thus, if I_1 is close to zero, the hypothesis is confirmed. If, on the other hand, all the income is concentrated in one county, the information gain is maximised. Then $I_1 = \log M$ and there is maximum inequality. Thus, the range of I_1 in equation (6) is from zero to $\log M$, representing, respectively, total equality and maximum inequality (Semple and Gauthier, 1972).

A second hypothesis that may be tested is that the distribution of income is directly proportional to the distribution of the population. In other words, the null hypothesis is that there is no inequality in the distribution of *per capita* incomes. In this case the posterior probabilities are as defined above, but the prior probabilities are redefined as

$$q_i = \frac{n_i}{N}. \qquad (7)$$

Substituting these prior probabilities into equation (4), one obtains

$$I_2 = \sum_i p_i \log \frac{p_i N}{n_i}$$

$$\doteq \log N - \log Y + \frac{1}{Y} \sum_i y_i \log \frac{y_i}{n_i}. \qquad (8)$$

The range of I_2 is from zero to $\log N$. As before, $I_2 = 0$ implies a uniform distribution or no inequality and the hypothesis is accepted. If, on the other hand, all the income arising is given to one person, there is maximum inequality and the value of I_2 is $\log N$.

To compare inequality indices for distributions measured over different sized populations, it is necessary to standardise the inequality measures. This is done by taking the ratio of the actual inequality measure to its maximum value. Then one obtains a degree of inequality index, $RI_2 = I_2/\log N$, that ranges between zero and unity. When the index value is zero, there is no inequality; conversely, when it is unity, inequality is at its maximum. Formulae similar to Equations (6) and (8) have been used by Batty (1975) and Haynes and Storbeck (1978) to test hypotheses about the spatial distribution of urban population and population density.

The procedures outlined so far can be used to test hypotheses about spatial inequality in distributions at the national level. In the remainder of this section we outline some procedures for testing simultaneously hypotheses about levels of inequality at different levels of spatial aggregation. These procedures are made possible by the ease with which information

measures decompose into simple additive forms. The decomposition property arises from Shannon's third axiom (Shannon, 1948). The property has been exploited by economists (Theil, 1967 and 1972; Horowitz, 1968) and geographers (Chapman, 1970 and 1973; Semple *et al.*, 1970, 1972, 1973 and 1977) to measure inequality or concentration at different levels of aggregation, both spatial and non-spatial. The decomposition principle was extended by Batty (1974, pp. 20-21) for the discrete Shannon entropy (Equation (1)) and the spatial entropy measure (Equation (3)). Here the principle is illustrated for the most general of all information measures, Equation (4).

The discrete random variable, X, is distributed over K regions, R_1, R_2, ..., R_K. Each region contains m_k counties such that $\sum_{k=1}^{K} m_k = M$, the total number of counties. The prior and posterior probabilities about the values of X at the regional level are Q_k and P_k, respectively. The corresponding probabilities at the county level are q_i and p_i. The following constraints must be satisfied:

$$\sum_{i \in R_k} p_i = P_k \qquad \qquad \sum_{k=1}^{K} P_k = \sum_{k=1}^{K} \sum_{i \in R_k} p_i = 1$$

$$\sum_{i \in R_k} q_i = Q_k \qquad \qquad \sum_{k=1}^{K} Q_k = \sum_{k=1}^{K} \sum_{i \in R_k} q_i = 1$$

The expected information measure, equation (4), may be written as

$$I = \sum_i p_i \log p_i - \sum_i p_i \log q_i \qquad \qquad (9)$$

The first term on the right hand side of equation (9) may be rewritten as

$$\sum_{k=1}^{K} \sum_{i \in R_k} p_i \log p_i = \sum_k P_k \left(\sum_{i \in R_k} p_i/p_k (\log p_i/P_k + \log P_k) \right)$$

$$= \sum_k P_k \log P_k + \sum_k P_k \left(\sum_{i \in R_k} p_i/P_k \log p_i/P_k \right) \qquad (10)$$

By similar steps the second term on the right hand side of equation (9) can be rewritten as

$$\sum_k P_k \log Q_k + \sum_k P_k \left(\sum_{i \in R_k} p_i/P_k \log q_i/Q_k \right). \qquad (11)$$

Combining equations (10) and (11) one obtains

$$I = \sum_{i=1}^{M} p_i \log p_i/q_i$$

$$= \sum_k P_k \log \frac{P_k}{Q_k} + \sum_k P_k \left(\sum_{i \in R_k} \frac{p_i}{P_k} \log \left(\frac{p_i}{q_i} \cdot \frac{Q_k}{P_k} \right) \right). \qquad (12)$$

The first term on the right hand side of equation (12) is the between-region expected information measure. The bracketed part of the second term is the within-region expected information measure. Thus, the total expected information is the sum of the between-region measure, plus all of the within-region measures weighted by their regional probabilities.

Equation (12) can be used to test hypotheses about inequality in the distribution of random variables at different levels of aggregation. One hypothesis that can be tested is that there is less inequality in the distribution of income at the regional level than at the sub-regional level. To test this hypothesis, define the regional and county prior probabilities as $k = \frac{1}{K}$ and $q_i = \frac{1}{M}$, respectively. The corresponding posterior probabilities may be defined as $P_k = \frac{Y_k}{Y}$ and $p_i = \frac{y_i}{Y}$, where Y_k is the income arising in region k. Substitute these probabilities into equation (12) to obtain

$$I_1 = \sum_k \frac{Y_k}{Y} \log \frac{Y_k \cdot K}{Y} + \sum_k \frac{Y_k}{Y} \sum_{i \in R_k} \frac{y_i}{Y_k} \log \left(\frac{y_i}{Y_k} \cdot \frac{M}{K} \right)$$

$$= \left\{ \frac{1}{Y} \sum_k (Y_k \log Y_k) - \log \frac{Y}{K} \right\}$$

$$+ \left\{ \frac{Y}{Y} (\sum_i y_i \log y_i - \sum_k Y_k \log Y_k) - \log \frac{K}{M} \right\} \qquad (13)$$

The term in the first set of chain brackets in equation (13) measures the between-region inequality in the distribution of income. The second term in equation (13) measures the within-region inequality. Clearly the sum of the between- and within-region inequalities is equal to the total inequality as measured by Equation (6).

The degree of inequality at both the regional and sub-regional levels is obtained by comparing the observed measures with the maximum values that are possible at those levels. An absence of inequality at the regional level arises when an equal amount of income arises in each region. Then $Y_k = \frac{Y}{K}$ for all k and the between-region inequality index is zero. Maximum

inequality occurs when all of the income is concentrated in one region. Then $Y_k = Y$ for one k and zero for all the others and the inequality index is log K. In a similar manner it is easily verified that the within-region inequality index ranges from zero to $\log \dfrac{M}{K}$. Then, the degree of inequality at both the regional and sub-regional levels is measured by the following indices:

$$RI_1' = \frac{\sum\limits_k Y_k \log Y_k + \log K - \log Y}{Y \log K} \quad , \tag{14}$$

and

$$RI_1'' = \frac{\sum(\sum\limits_{k \; i \in R_k} y_i \log y_i - Y_k \log Y_k) + \log M - \log K}{Y(\log M - \log K)} \quad . \tag{15}$$

Both of these indices range between zero and unity. At both levels an index of zero implies a distribution without inequality. Indices of unity occur when there is maximum inequality. The hypothesis that the distribution of income is more uniform at the regional scale than at the sub-regional scale is tested by comparing RI_1' and RI_1''. If $RI_1' < RI_1''$, the hypothesis is accepted.

Following similar procedures, one may use Equation (12) to test the hypothesis that the distribution of incomes is more proportional to the distribution of population at the regional level than at the sub-regional level. Then, the null hypothesis is that there is less inequality in the *per capita* income distribution at the regional level than at the sub-regional level. To test the hypothesis, define the regional and county prior and posterior probabilities as $Q_k = \dfrac{N_k}{N}$, $q_i = \dfrac{n_i}{N}$, $P_k = \dfrac{Y_k}{Y}$ and $p_i = \dfrac{y_i}{Y}$, respectively, where N_k and Y_k are the regional shares of population and income. On substituting these values into equation (12), one obtains

$$I_2 = \left\{ \frac{1}{Y}(\sum\limits_k Y_k \log \frac{Y_k}{N_k}) - \log \frac{Y}{N} \right\}$$

$$+ \left\{ \frac{1}{Y} \sum\limits_k (\sum\limits_{i \in R_k} y_i \log \frac{y_i}{n_i} - Y_k \log \frac{Y_k}{N_k}) \right\} \tag{16}$$

As before, the degree of inequality at both the regional and county scales is obtained by dividing the calculated values at these scales by their respective maximum values. Between-region inequality is at a maximum when all the income is concentrated in the region with the smallest population. In that

case $Y_k = Y$ for one k and zero for all the others. Then from the first half of Equation (16) the between-region inequality index has a value of $\log \dfrac{N}{N_k}$. Minimum inequality occurs when the income and population distributions are identical. Then $\dfrac{Y_k}{Y} = \dfrac{N_k}{N}$ for all k and the inequality index is zero. By similar arguments it follows that the within-region inequality index ranges between zero and $\log N_k$.

To test the hypothesis that the degree of inequality in the distribution of *per capita* incomes at the regional scale is less than at the sub-regional scale, use the following indices that measure degrees of inequality:

$$RI_2' = \frac{\dfrac{1}{Y}(\sum_k Y_k \log \dfrac{Y_k}{N_k}) - \log \dfrac{Y}{N}}{\log N/N_k} \tag{17}$$

and

$$RI_2'' = \frac{\dfrac{1}{Y}\sum_k (\sum_i y_i \log y_i/n_i - Y_k \log Y_k/N_k)}{\log N_k} \tag{18}$$

Both of these indices range between zero and unity; the closer they are to zero, the greater the degree of equality at both scales. If $RI_2' < RI_2''$, the hypothesis is accepted.

IV EMPIRICAL ANALYSIS OF IRISH INCOME AND EMPLOYMENT DISTRIBUTION

In this section we use the methodology that has been developed above to test hypotheses about inequality levels in the spatial distribution of *per capita* incomes and certain categories of employment in Ireland. The data used were taken from various reports prepared by Ross (1969, 1972 and 1977) and Baker (1966 and 1975). The limitations of the data have been discussed by both these authors and, therefore, will not be repeated here. Using these data, indices of inequality have been computed at the national level and at the between- and within-region level for four different regional systems. The four regional systems are shown on Figure 1. In the first system the regions are the four provinces. The second system corresponds with the traditional East-West partitioning of the country, while the third system represents a modification of the second one, following Baker and Ross (1970). The nine planning regions constitute the fourth regional system. In

the tables below these regional systems will be referred to as A, B, C and D, respectively.
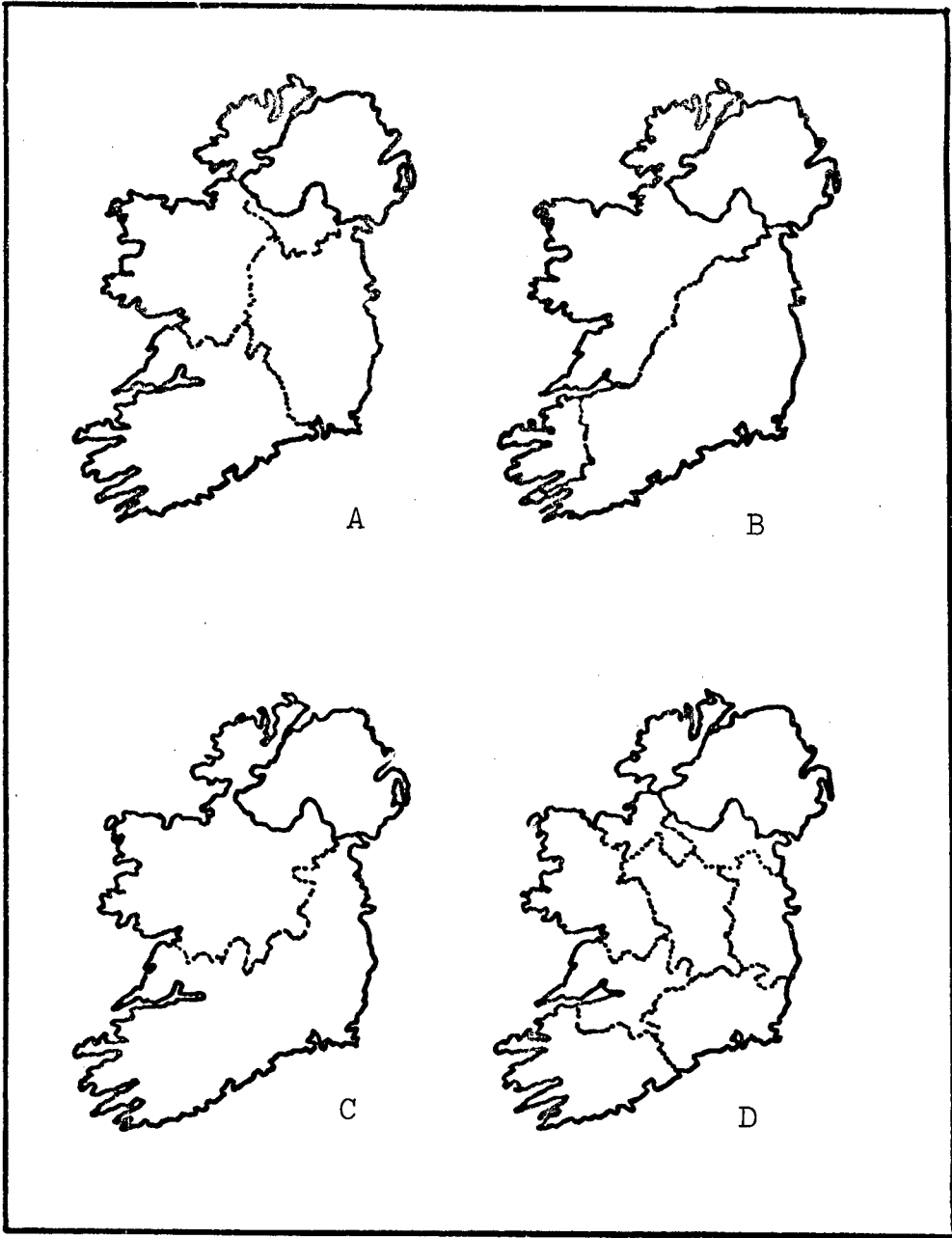

Figure 1: The four regional systems.

Absolute and relative inequality indices for the distribution of *per capita* incomes have been computed for 1960, 1965, 1969 and 1973. These indices are summarised in Table 1. The hypothesis of interest is that there is always some inequality in the spatial distribution of *per capita* incomes. The hypothesis is accepted if the inequality index is different from zero. Since the data base represents a population, confidence limits are not appropriate for this type of analysis. Instead the significance of the results must be judged in terms of the possible ranges of the statistics.

Table 1: *Inequality indices for per capita income distribution.*

Regional system	Inequality index	1960	1965	1969	1973
The Country	I_2	0.02022	0.02347	0.02353	0.01273
	RI_2	0.00136	0.00158	0.00158	0.00085
A	I_2'	0.01060	0.01073	0.01068	0.00574
	RI_2'	0.00415	0.00410	0.00403	0.00214
	I_2''	0.00962	0.01274	0.01285	0.00699
	RI_2''	0.00078	0.00104	0.00105	0.00057
B	I_2'	0.01043	0.01100	0.01005	0.00599
	RI_2'	0.00881	0.00885	0.00785	0.00455
	I_2''	0.00979	0.01247	0.01348	0.00674
	RI_2''	0.00072	0.00091	0.00099	0.00050
C	I_2'	0.00909	0.01001	0.00935	0.00538
	RI_2'	0.00845	0.00865	0.00873	0.00482
	I_2''	0.01113	0.01346	0.01418	0.00735
	RI_2''	0.00083	0.00100	0.00105	0.00054
D	I_2'			0.01758	0.00983
	RI_2'			0.00438	0.00242
	I_2''			0.00595	0.00290
	RI_2''			0.00055	0.00027

None of the inequality indices in Table 1 is equal to zero. A number of trends are apparent from the data. Firstly, the degree of inequality in the total distribution increased considerably between 1960 and 1965, but declined between 1969 and 1973. Thus throughout the growth era of the early 'sixties there was a tendency towards increasing divergence in the spatial distribution of *per capita* incomes. This occurred during the period of most rapid economic expansion in the history of the state. Conversely,

throughout the late 'sixties and early 'seventies as the pace of national economic expansion slowed down and as more emphasis was put on the achievement of balanced regional growth, there was a convergence in the spatial distribution of *per capita* incomes, as indicated by the declining inequality indices.

Taking the four provinces as regional system A, it is observed that at the *inter-provincial* level there has been a constant decline since 1960 in the relative level of inequality (RI_2'). Most of the decline was in the period 1969-73. At the *intra-provincial* level inequality peaked in 1969 and has declined since then. Comparing RI_2' with RI_2'' it is seen that in all years the degree of inequality at the *inter-provincial* level was greater than at the *intra-provincial* level.

Using regional systems B and C, it is clear that the trends in inter-regional inequality have been similar to those for the total inequality index. For both systems inter-regional inequality increased between 1960 and 1965 and thereafter declined steadily. Similar trends are evident at the sub-regional level (that is, an increase from 1960-1969 and a decline thereafter). The indices of relative inequality at the intra-regional level have not varied very much over the years. The inter-regional relative inequality measure has, however, shown a large decrease. Manipulating these measures provides some insight into the Baker and Ross (1970) suggestion that the East-West split has undergone a change. Define the ratio of inter-regional to intra-regional relative inequality as a measure of regional separation. Then, examining Table 1a which shows values of RI_2'/RI_2'' computed over both schemes B and C, it is clear that the measure of regional separation was greater for the traditional East-West divide in 1960 and that this persisted until 1965. In 1969, however, the revised scheme proposed by Baker and Ross (1970) shows a higher ratio, indicating that regional separation *was* more pronounced in the revised scheme. By 1973, the picture was again reversed and East-West contrasts, although less than in 1960, were greater than those in the Baker-Ross scheme. These results indicate the importance of indicating clearly the regional system when discussing convergence or divergence trends in the spatial distribution of income levels. Finally, by examining RI_2' and RI_2'' for systems B and C, it is clear that inequality in the distribution of *per capita* incomes was greater at the inter-regional level than at the intra-regional level in all years.

Finally, examining the results for the IDA regions (Scheme D), both relative and absolute inequality declined between 1969 and 1973 at both inter-regional and intra-regional scales. Note also that the inequality was greater between than within planning regions. The measure of regional separation, however, increased, indicating that contrasts between planning regions have increased despite an overall decline in income inequality.

Table 1a: *Measures of group separation, A, using schemes B and C.*

Scheme	1960	1965	1969	1973
B	12.23611	9.72527	7.92929	9.10000
C	10.18072	8.65000	8.31429	8.92593

In the analysis of employment patterns we have followed the lead of Baker and Ross who have hypothesised that economic activity in an area as small as an Irish county can be divided into autonomous and induced sectors.

> The autonomous sector is that whose product of either goods or services is "exported" from that county to either the remainder of the country or the rest of the world. It also includes that social sector whose provision of services to the county itself is determined by national rather than local standards, and whose financing is at least in part a national responsibility. The induced sector is that whose goods and services are consumed within the county itself, and whose level of activity is determined by the size and prosperity of the autonomous sector (Baker and Ross, 1970, pp. 155-156)..

Their hypothesis continues to the effect that the autonomous sector can itself be divided into agricultural and non-agricultural activities and that the higher the proportion of non-agricultural activities within the sector, the larger will tend to be the induced sector and "the more developed can be regarded the county" (Baker and Ross, 1970, p. 156).

If their hypothesis is accepted, it provides a standard for measuring the relative development of different counties. Baker and Ross ranked counties according to this development criterion and levels of *per capita* income for a number of years to outline the changing regional pattern in the relative development and prosperity of Ireland. Here we have calculated indices of inequality for the spatial distribution of two categories of employment. They are the non-agricultural proportion of autonomous employment and the proportion of the work force in the induced sector. (Inequality indices were also computed for two other categories of employment — the numbers of non-agricultural autonomous and induced employees. Since, however, information of the size of the work force in each county is not used in the calculations of those indices, it is more difficult to draw meaningful conclusions from the results.) It is hypothesised that increasing economic development is associated with an increasing proportion of each county's work force in the non-agricultural autonomous and induced sectors and that inter-county differences in these levels will decline over time. While the hypothesis suggests

that there is a close relationship between the shares of employment in both sectors, there are, of course, some divergences. These usually arise from historical or geographical factors such as situation of towns serving hinterlands which cross county boundaries and employment induced by the spending of people outside of the work force. These divergences have been examined in detail by Baker and Ross (1975). Inequality indices for the two distributions have been calculated by us for four different regional systems for each of the four years, 1951, 1961, 1966 and 1971. The regional systems used here are the same as those used earlier.

To calculate the inequality indices for the distribution of the non-agricultural share of the autonomous sector, we used Equations (4) and (12). The prior and posterior probabilities were defined as $p_i = \dfrac{n_i}{N}$ and $q_i = \dfrac{a_i}{A}$ where n_i, N, a_i and A are non-agricultural autonomous employment in county i, non-agricultural autonomous employment in the country, autonomous employment in county i and autonomous employment in the country, respectively. Similar probabilities were defined at the regional level. The indices presented below are based on the distribution over twenty-five counties since it was necessary to exclude Dublin because of the difficulties of distinguishing between autonomous and induced employment there.

The absolute and relative inequality indices for the total distribution and each of the four disaggregations are summarised in Table 2. All of the indices differ from zero. Thus there is not a total absence of inequality in any of the distributions. At the national level there has been a steady decline in both the absolute and relative levels of inequality in the total distribution. The 1971 levels were only slightly over half those for 1951. After only a gradual decline during the 1950s there was a rapid advance towards uniformity in this spatial distribution during the 1960s.

The inter-regional inequality indices declined over the period for each of the regional systems. As expected, the between-region measures for the relatively simpler systems B and C are greater than those for the larger systems. At the intra-regional level there was a steady decline over the years for regional systems A, C and D. For regional system B there was a slight increase in 1961 beyond the 1951 level.

The hypothesis that the degree of inequality at the regional level is greater than at the sub-regional level is confirmed for the four regional systems for each of the time periods. When the within-region inequality index is expressed as a percentage of the total inequality, it appears that of the four regional systems the one that minimises within-region inequality is the traditional East-West division, system B. That result contrasts strongly with the conclusion of Baker and Ross (1970). It is noteworthy that in 1971 the partitioning of inequality into within- and between-region shares was almost the

same for both the traditional provincial and modern economic planning regional systems. In both cases approximately half of the inequality was at the regional level and half at the sub-regional level.

Table 2: *Inequality indices for the distribution of the non-agricultural share of the autonomous sector.*

Regional system	Inequality index	1951	1961	1966	1971
The Country	I_2	0.0852	0.0742	0.0635	0.0446
	RI_2	0.0063	0.0056	0.0048	0.0034
A	I_2'	0.0457	0.0353	0.0325	0.0222
	RI_2'	0.0218	0.0161	0.0147	0.0099
	I_2''	0.0395	0.0389	0.0310	0.0224
	RI_2''	0.0035	0.0035	0.0028	0.0021
B	I_2'	0.0595	0.0470	0.0389	0.0259
	RI_2'	0.0800	0.0342	0.0470	0.0295
	I_2''	0.0257	0.0272	0.0246	0.0189
	RI_2''	0.0020	0.0022	0.0020	0.0015
C	I_2'	0.0447	0.0371	0.0344	0.0233
	RI_2'	0.0490	0.0384	0.0342	0.0221
	I_2''	0.0405	0.0371	0.0291	0.0213
	RI_2''	0.0032	0.0030	0.0024	0.0018
D	I_2'			0.0301	0.0224
	RI_2'			0.0097	0.0070
	I_2''			0.0334	0.0222
	RI_2''			0.0033	0.0022

Inequality indices for the distribution of the induced share of the work force have also been calculated for the same four regional systems. The results are summarised in Table 3. While none of the indices is exactly zero, at the sub-regional level most of the relative inequality indices are very close to the minimum. At the national level both the absolute and relative inequality indices have declined steadily from 1951 to 1971. The 1971 values were only about 45 per cent of those in 1951. For all four years the degree of inequality in this distribution was considerably less than that for the distribution of the non-agricultural share of the autonomous sector.

For each of the four regional systems both the absolute and relative inequality indices have declined steadily over the years. Furthermore, for each regional system the degree of inequality at the regional level was

consistently larger than at the sub-regional level. As before, the partitioning of inequality into between- and within-range levels shows that the system for which the within-region measure is lowest is the traditional East-West division of the country. For this distribution the within-region inequality is greatest when the nine economic planning regions are used.

Table 3: *Inequality indices for the distribution of the induced share of the total work force.*

Regional system	Inequality index	1951	1961	1966	1971
The Country	I_2	0.0339	0.0269	0.0206	0.0152
	RI_2	0.0025	0.0020	0.0015	0.0011
A	I_2'	0.0186	0.0153	0.0118	0.0078
	RI_2'	0.0086	0.0068	0.0052	0.0034
	I_2''	0.0153	0.0114	0.0088	0.0074
	RI_2''	0.0013	0.0010	0.0003	0.0007
B	I_2'	0.023	0.0184	0.0154	0.0109
	RI_2'	0.0283	0.0210	0.0169	0.0115
	I_2''	0.0105	0.0083	0.0052	0.0043
	RI_2''	0.0008	0.0007	0.0004	0.0003
C	I_2'	0.0182	0.0149	0.0116	0.0080
	RI_2'	0.0184	0.0142	0.0107	0.0070
	I_2''	0.0156	0.0118	0.0090	0.0072
	RI_2''	0.0012	0.0009	0.0007	0.0006
D	I_2'			0.0107	0.0072
	RI_2'			0.0033	0.0022
	I_2''			0.0099	0.0080
	RI_2''			0.0010	0.0008

V CONCLUSIONS

This paper has shown the type of results that can be obtained using quantitative indices derived from information theory. Although the primary concern has been with the methodological question, the results do provide a focus for some general comments. Inequality in the spatial distribution of *per capita* incomes increased up to the mid 'sixties and then declined steadily over time. This trend confirms to some extent the change in emphasis in

economic planning, away from a primary concern with national growth towards a concern with regional balance. In the early 'sixties the rapid expansion during the First Programme and the first half of the Second Programme contributed to inflationary pressures which were manifested in increased spatial inequality in income patterns. The more recent decline in inequality can be seen as part of a more mature form of economic planning in which primary emphasis was placed on greater regional equality.

The trends in employment distributions indicate an increase in spatial equality in both measures over time. This tendency towards more dispersed non-agricultural autonomous employment (in the counties *other than* Dublin) together with the improved spatial income distribution indicates some overall improvement in the regional balance of the country.

REFERENCES

ATKINSON, A.B., 1970. "On the Measurement of Inequality," *Journal of Economic Theory*, Vol. 2, pp. 244-63.

BAKER, T.J., 1966. *Regional Employment Patterns In The Republic of Ireland*, Dublin: Economic and Social Research Institute, Paper No. 32.

BAKER, T. J. and M. ROSS, 1970. "The Changing Regional Pattern In Ireland," *Economic and Social Review*, Vol. 1, No. 2, pp. 155-165.

BAKER, T. J. and M. ROSS, 1975. *Employment Relationships in Irish Counties*, Dublin: Economic and Social Research Institute, Paper No. 81.

BARTELS, C.P.A. and P. NIJKAMP, 1976. "An Empirical Welfare Approach to Regional Income Distribution: Alternative Specifications and Estimates of Income Inequality Measures", *Socio-Economic Planning Sciences*, Vol. 10, pp. 117-128.

BATTY, M., 1974. "Spatial Entropy", *Geographical Analysis*, Vol. 6, pp. 1-32.

BATTY, M., 1975. "Urban Density and Entrophy Functions", *Geographical Papers, No. 36*, Dept. of Geography, University of Reading.

BATTY, M., 1976. "Entropy in Spatial Aggregation", *Geographical Analysis*, Vol. 8, pp. 1-21.

CHAPMAN, G.P., 1970. "The Application of Information Theory to the Analysis of Population Distributions In Space", *Economic Geography*, Vol. 46, pp. 317-331.

CHAPMAN, G.P., 1973. "The Spatial Organization of the Population of the United States and England and Wales", *Economic Geography*, Vol. 69, pp. 325-343.

CHIPMAN, J.S., 1974. "The Welfare Ranking of Pareto Distribution", *Journal of Economic Theory*, Vol. 9, pp. 273-282.

DASGUPTA, P., A. SEN, and D. STARRETT, 1973. "Notes on the Measurement of Inequality", *Journal of Economic Theory*, Vol. 6, pp. 180-187.

GAILE, G.L., 1977. "Effiquity; A Comparison of a Measure of Efficiency with an Entropic Measure of the Equality of Discrete Spatial Distributions", *Economic Geography*, Vol. 53, pp. 265-282.

GARRISON, C.B. and A.S. PAULSON, 1973. "An Entropy Measure of the Geographic Concentration of Economic Activity", *Economic Geography*, Vol. 49, pp. 319-324.

GOLDMAN, S., 1953. *Information Theory*, New York: Prentice Hall.

GOOD, I.J., 1956. "Some Terminology and Notation in Information Theory," *Proceedings of IEE*, Part C, 103, pp. 200-204.

GUREVICH, B.L., 1969. "Geographical Differentiation and its Measures in a Discrete System," *Soviet Geography: Review and Translation*, Vol. 10, pp. 387-413.

HART, P.E., 1971. "Entropy and Other Measures of Concentration", *Journal of the Royal Statistical Society*, Vol. 134, pp. 73-85.

HAYNES, K.E. and W.T. ENDERS, 1975. "Distance, Direction and Entropy in the Evolution of a Settlement Pattern," *Economic Geography*, Vol. 51, pp. 357-365.

HAYNES, K.E. and J.E. STORBECK, 1978. "The Entropy Paradox and the Distribution of Urban Population", *Socio-Economic Planning Sciences*, Vol. 12, pp. 1-6.

HINDERINK, J. and J.J. STERKENBURG, 1978. "Spatial Inequality in Underdeveloped Countries and the Role of Government Policy", *Tijdschrift voor Econ. en Soc. Geografic*, Vol. 69, pp. 5-16.

HOBSON, A., 1969. "A New Theorem of Information Theory", *Journal of Statistical Physics*, Vol. 1, pp. 383-391.

HOROWITZ, A. and J. HOROWITZ, 1968. "Entropy, Markov Processes and Competition In the Brewing Industry," *Journal of Industrial Economics*, Vol. 16, pp. 196-211.

INGELS, F.M., 1974. *Information and Coding Theory*, London: Intext.

KENDALL, M.G. and A. STUART, 1958. *The Advanced Theory of Statistics*, Vol. 1, New York: Hafner Publishing Co.

KULLBACK, S., 1959. *Information Theory and Statistics*, New York: Wiley.

MARCHAND, D., 1975. "On the Information Content of Regional Maps: The Concept of Geographical Redundancy", *Economic Geography*, Vol. 51, pp. 117-128.

MARTIN, J. P., 1971. "Regional Growth and Income Patterns, 1960-65", *Economic and Social Review*, Vol. 2, No. 3, pp. 349-366.

MEDVEDKOV, Y., 1967. "The Concept of Entropy in Settlement Pattern Analysis", *Papers of the Regional Science Association*, Vol. 18, pp. 165-168.

MEDVEDKOV, Y., 1970. "Entropy: An Assessment of Potentialities In Geography," *Economic Geography*, Vol. 46, pp. 304-316.

NUTENKO, L.Y., 1970. "An Information Theory Approach to the Partitioning of an Area", *Soviet Geography: Review and Translation*, Vol. 11, pp. 540-544.

PIELOU, E. C., 1966. "Shannon's Formula as a Measure of Specific Diversity — Its Use and Misuse", *American Naturalist*, Vol. 100, pp. 463-465.

RENYI, A., 1970. *Probability Theory*, New York: American Elsevier.

Review of Progress 1964-1967 on Second Programme for Economic Expansion, Stationery Office, Dublin 1968.

ROSS, M., 1969. *Personal Incomes by County, 1965*, Dublin: Economic and Social Research Institute, Paper No. 49.

ROSS, M., 1972. *Further Data on County Incomes in the Sixties*, Dublin: Economic and Social Research Institute, Paper No. 64.

ROSS, M., R. JONES and E. O'MALLEY, 1977. *Personal Incomes by County in 1973*, Dublin: National Economic and Social Council, Report No. 30.

ROTHSCHILD, M., and J.E. STIGLITZ, 1973. "Some Further Results on the Measurement of Inequality," *Journal of Economic Theory*, Vol. 6, pp. 188-204.

SEMPLE, R.K. and R.G. GOLLEDGE, 1970. "An Analysis of Entropy Changes in a Settlement Pattern Over Time," *Economic Geography*, Vol. 46, pp. 157-160.

SEMPLE, R.K. and H.L. GAUTHIER, 1972. "Spatial-Temporal Trends in Income Inequalities in Brazil", *Geographical Analysis*, Vol. 4, pp. 169-180.

SEMPLE, R.K., 1973. "Recent Trends in the Spatial Concentration of Corporate Headquarters", *Economic Geography*, Vol. 49, pp. 309-318.

SEMPLE, R.K. and G.J. DEMKO, 1977. "An Information Theoretic Analysis; An Application to Soviet-COMECON Trade Flows", *Geographical Analysis,* Vol. 9, No. 1, pp. 51-63.

SHANNON C.F., 1948. "A Mathematical Theory of Communication", *Bell Systems Technical Journal,* Vol. 27, pp. 379-423, 623-656.

SMITH, D.M., 1975. *Patterns in Human Geography,* Harmondsworth: Penguin Books.

THEIL, H., 1967. *Economics and Information Theory,* Amsterdam: North-Holland.

THEIL, H., 1972. *Statistical Decomposition Analysis,* Amsterdam: North-Holland.

TRIBUS, M., 1969. *Rational Descriptions, Decisions and Designs,* Oxford: Pergamon Press.

WALSH, J.A., 1976. *Information Theory and Classification in Geography,* D.P. 4, Department of Geography, McMaster University, Hamilton, Ontario.

WALSH, J.A. and M.J. WEBBER, 1977. "Information Theory: Some Concepts and Measures", *Environment and Planning,* Vol. 9, pp. 395-417.

WILLIAMSON, J.G., 1965. "Regional Inequality and the Process of National Development, A Description of the Patterns", *Economic Development and Cultural Change,* Vol. 13, pp. 3-45.