

A MULTIVARIATE ANALYSIS
OF SPATIAL INEQUALITIES

H.J. Blommestein

P. Nijkamp

P. Rietveld

Research Memorandum 1978-9

Paper presented at the
Fourth Advanced Studies
Institute in Regional
Science, Siegen,
August 6-19, 1978



1. Introduction

In modern economic literature about regional planning theory a substantial part has been devoted to the well known 'efficiency-equity' dilemma (cf. Myrdal [1957], Nijkamp and Verhage [1978], Richardson [1977], Stilwell [1972]). This dilemma arises from the fact that the efficiency objective (such as the achievement of a maximum (regional) welfare with a minimum of factor inputs) does not guarantee an equitable distribution of welfare and of factor inputs.

It is clear, however, that the concepts of efficiency and equity are of a multidimensional nature. Efficiency can be measured inter alia by means of proxy indicators such as average value added or labour participation, while equity can be measured inter alia via the skewness of the income distribution, regional unemployment and the like. In some or other respect both the efficiency and the equity indicators determine the regional welfare levels by means of a set of mutual intricate relationships.

In many recent analyses the attention has been focussed on the spatial or personal distribution of only one of these welfare indicators (see for an extensive study Bartels [1977]). Given the fact that the multidimensional configuration of a whole set of indicators determines regional welfare, it is extremely important to develop an integrated analysis covering the multiple dimensions of welfare (differences) in a multiregional system (cf. Coates et al. [1977]).

A basic notion for an appropriate starting-point of such a multidimensional multiregional analysis of welfare discrepancies is a so-called regional welfare profile, subdivided into a socio-economic profile, a demographic-physical profile and an environmental profile (see section 2). In this section some problems inherent in the use of multidimensional profile methods will be discussed, especially the standardisation problem and the problem of multicollinearity. The multicollinearity problem will be attacked by means of a recently developed multivariate technique, viz. interdependence analysis. A critical discussion and some possible adjustments of the latter technique will also be presented.

The next section will be devoted to the notion of a multidimensional inequality indicator encompassing all elements of a regional welfare profile, while special attention will be paid to a generalized coefficient of variation. Several properties of this inequality measure will be derived, followed by an analysis of the implications of introducing trade-offs among the elements of a regional welfare profile.

(see for a further use of environmental profiles Nijkamp [1977c]).

It is evident that the information contained in s^S , s^D and s^E is necessary to make inferences about discrepancies among regions. Each comprehensive inequality indicator has to be based on the multidimensional aspects of regional welfare.

A first problem in employing the information from (1.) is the fact that each indicator should be measured in appropriate units (for example, income per capita, percentage unemployment etc.) The problems inherent in such standardisation procedures are discussed more thoroughly in Paelinck and Nijkamp[1976].

A slightly different but allied problem is the impact of each of the welfare indicators on welfare itself. For the sake of simplicity in the derivation of analytical results of multidimensional inequality measures, the assumption will be made that all arguments of the regional welfare profiles are defined as 'benefit' indicators ('the higher, the better'). This implies that 'cost' indicators (like congestion or pollution) are provided with a minus sign in the regional welfare profile. As far as certain indicators are included for which a satiation level \hat{s}_{ir} may be assumed, the corresponding indicator in the regional welfare profile may be defined as:

$$(2.) \quad s_{ir} = - \left| s_{ir} - \hat{s}_{ir} \right|$$

This might be the case for unemployment e.g., for which a certain friction level is considered to be necessary to achieve a balanced situation on the labour market.

Another problem is related to the multicollinearity of a set of regional welfare indicators. By including an extensive set of profile elements in an inequality analysis, a large amount of redundant information might be taken into account due to the multicollinearity among the welfare arguments.

A new technique, viz. interdependence analysis, may be used to attack to a certain extent this problem. Interdependence analysis is an optimal subset selection technique, by means of which a subset of variables which best represents an entire variable set can be chosen (see Beale et al. [1967], and Boyce et al. [1974]). In the past several multivariate data-reducing techniques have been developed, such as principal component analysis and factor analysis. A basic shortcoming in the use of these techniques has always been the lack of a clear theoretical interpretation of the statistically calculated components

or factors.

Interdependence analysis attempts to side-step the latter problem by selecting an optimal subset of the original variables, so that a data transformation is not necessary. Suppose a data matrix with N observations on K variables. Suppose next that P variables are to be selected from the K variables such that these P variables reflect an optimal correspondence with respect to the original data set. Consequently, (K - P) variables are to be 'rejected' or eliminated.

Now the interdependence analysis starts with a successive regression analysis between the 'dependent' (K - P) variables to be rejected and the 'independent' P variables to be retained.

Suppose that X_P is the N x P reduced matrix pertaining to variables 1, ..., P. Then the following regression equation is obtained for each variable P + 1, ..., K :

$$(3.) \quad x_l = X_P \beta_l + \epsilon_l \quad , \quad l=P+1, \dots, K$$

where x_l is a (N x 1) vector with observations on the lth variable, β_l is a (P x 1) regression coefficient, and ϵ_l a (N x 1) vector of disturbance terms. The estimated squared multiple correlation coefficient of (3) will be denoted by R_l^2 . It is clear that for $l = P + 1, \dots, K$ (K - P) regression equations have to be calculated, so that there are also (K - P) correlation coefficients. Next, the minimum value of R_l^2 ($l = P + 1, \dots, K$) is selected:

$$(4.) \quad R^2 = \min_l R_l^2$$

A low value of R_l^2 means that the variables 1, ..., P are a bad representation of variable l.

It is clear that the abovementioned regression procedure can be repeated for each variation of P and (K - P) variables, so that theoretically the total number of regressions to be carried out is equal to $\binom{K}{P}$ (K - P). Then the optimal subset is defined as that subset which maximizes over all $\binom{K}{P}$ permutations the values of R_{\min}^2 , i.e.,

$$(5.) \quad R^{*2} = \max R_{\min}^2$$

Essentially this solution can be seen as the equilibrium solution of a game procedure, in which the information contained in a data matrix is reduced such that the selected variables constitute a maximum representation of the information pattern (with a fairly low multicollinearity). This max-min solution might lead to an enormous com-

putational load, but the strength of interdependence analysis is that it finds the optimal subset without a complete enumeration of all possible regressions. Instead, a set of demarcation criteria and bounding rules are introduced to speed up the search for an optimal subset. By means of elimination procedures via critical threshold levels based on statistical properties of the successive correlation coefficients, the computational work can be facilitated significantly, so that in principle an optimal subset can be selected within a reasonable time limit. The reader is referred to Boyce et al. [1974] for further details.

The appealing feature of interdependence analysis is that it selects a subset of rather independent variables which have a maximum correspondence with the original data set without using arbitrary or artificial data transformations.

Hence, the interpretation of the results is straightforward.

Interdependence analysis has been applied inter alia in optimal network algorithms (see Boyce et al. [1974]), in multi-criteria analyses (Nijkamp [1977a]) and in multi-dimensional analyses of human settlements (Nijkamp [1977b]). The experiences with interdependence analysis are rather favourable so far, so that a broader application of this technique may be worth while.

In the analysis of regional inequalities the interdependence analysis will be used to analyse the structure of the data set. We will also consider the question whether interdependence analysis can be used in a meaningful way to eliminate redundant information.

The purpose of interdependence analysis is to select a certain number of variables from a larger set. The choice criterion is that the correlation between the selected and discarded variables is as high as possible in the 'max-min' sense. One should realize, however, that this criterion is to a certain extent arbitrary and that also other, more or less appealing criteria can be proposed. If $R_{p,l}^2$ denotes the correlation coefficient of equation (3.), one might also use

$$(6.) \quad \max! \sum_{l=P+1}^K R_{p,l}^2$$

as a selection criterion. (6.) is not a better or worse criterion than the min-max criterion. It may give rise to more complicated computations, however.

Another criterion can be identified when we consider the question whether the procedures discussed above result in a matrix X_p with as less multicollinearity as possible. It is clear that a) the maximization of the correlation between discarded and selected variables

and b) the minimization of the correlations between the selected variables (in other words, of the multicollinearity in X_p) are related objectives. It can not be proved, however, that a) and b) are equivalent. Therefore it might be worthwhile to introduce the criterion of minimum multicollinearity explicitly.

If $r_{k,l}^2$ denotes the squared simple correlation coefficient between variables k and l , the criterion of a minimum multicollinearity can be formalized as:

$$(7.) \quad \min! \max_{k,l \leq P} r_{k,l}^2$$

Given the discussion above, an obvious alternative for (7.) is

$$(8.) \quad \min! \sum_{k,l \leq P} r_{k,l}^2$$

An appealing advantage of (7.) and (8.) above (5.) and (6.) is that they imply calculations of a negligible effort.

We conclude this section with a final remark about methods to select in a certain way P variables from K original ones. The articles referred to for this subject show that the fixation of P has a strong influence on the variables selected. Therefore, we may conclude that a sensitivity analysis is necessary to avoid wrong interpretations of the results of interdependence analysis. Another conclusion is, that it is worthwhile to seek for an adjusted method satisfying the following relationship: let S_p be the set of P variables selected by means of a selection method. Then

$$(9.) \quad S_p \subset S_{p'}, \quad \text{when } P < P'$$

The methods discussed in this section do not satisfy this requirement, unless this condition is imposed a priori.

3. Multidimensional Inequality Analysis

We consider the question whether it is possible to find indicators for interregional inequalities, taking account of all dimensions of the profile matrix S .

A study of interregional differences by means of S is not new, of course. Grouping procedures for regions e.g., are based on the notion of similarity of regional profiles (cf. Fisher [1977], Johnston [1968] and Paelinck

and Nijkamp [1976]). Groups of relatively homogeneous regions are obtained by minimizing the multidimensional differences. For the economic subprofile e.g., the concept of economic distance between regions has been introduced (cf. Paelinck and Nijkamp [1976]).

A new element of the present study is, that welfare elements are included in the analysis. Fig. 1 shows intuitively that to that end the ordinary distance concept is not an adequate tool.

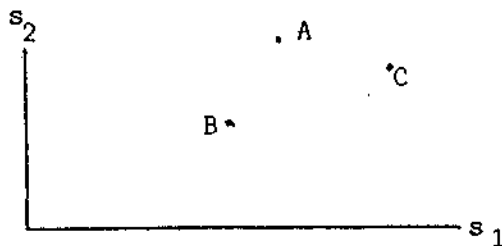


Fig. 1. Profiles for three regions.

An analysis of the profiles of the three regions by means of distances gives among others rise to the conclusion that B is more similar to A than to C and that B and C are equally similar to A. The distance analysis does not enable one to compare the welfare positions of the regions, however. As we assume that for both s_1 and s_2 high values lead to high welfare, one can conclude from Fig.1 that A and C are in a better state than B. The question, whether A's position should be preferred to C's can not be answered a priori. It depends on the weights attached to the elements of the profile, i.e. on the welfare functions for the regions.

An interesting conclusion, which can be drawn is that interregional inequality in welfare can be enlarged by inequalities in the elements of the profile (A compared to B), but also reduced (A compared to C). It is reasonable therefore, to base an analysis of interregional inequality on the welfare function :

$$(10.) \quad \omega_r = \omega_r(\underline{s}_r) \quad r = 1, \dots, R$$

In this study we will assume that the function has the same structure for each region and therefore, (10.) is transformed into:

$$(11.) \quad \omega_r = \omega(\underline{s}_r) \quad r = 1, \dots, R$$

or in matrix notation:

$$(12.) \quad \underline{\omega} = \underline{\omega}(S)$$

where $\underline{\omega} = (\omega_1, \dots, \omega_R)'$

An obvious objection against the use of this welfare function is that it is difficult to assess. Why not simply proceed without the use of such a tedious concept? The answer is, that it appears to be very fruitful when it is used in the same way as in multiple objective programming methods, viz. by leaving the exact specification of the function open as long as possible during the analysis. Then it presents a framework for a very general analysis of multidimensional interregional inequalities.

The framework developed here proceeds as follows. We study the inequality in welfare, rather than in the different dimensions, and use measures from the large set of inequality measures developed thus far for single variables (cf. Bartels [1977]), by making use of the welfare index $\underline{\omega}$ rather than, for example, the income per capita.

The results of this substitution will be shown for two inequality measures; these two measures are selected, because they have favourable analytical properties, namely the coefficient of variation and the mean deviation with respect to the maximum value.

A. Derivation of the coefficient of variation

The coefficient of variation V is equal to the quotient of standard deviation σ and mean μ . The mean μ is equal to

$$(13.) \quad \underline{\mu} = \frac{1}{R} \underline{1}' \underline{\omega}$$

where $\underline{1} = (1, \dots, 1)'$

The variance σ^2 is equal to

$$(14.) \quad \sigma^2 = \frac{1}{R} \left[\underline{\omega} - \frac{1}{R} (\underline{1}' \underline{\omega}) \underline{1} \right]' \left[\underline{\omega} - \frac{1}{R} (\underline{1}' \underline{\omega}) \underline{1} \right]$$

which can be reduced to

$$(15.) \quad \sigma^2 = \frac{1}{R} \underline{\omega}' \left[I - \frac{1}{R} (\underline{1} \underline{1}') \right]' \left[I - \frac{1}{R} (\underline{1} \underline{1}') \right] \underline{\omega}$$

As $\left[I - \frac{1}{R} (\underline{1} \underline{1}') \right]$ is symmetric and idempotent, (15.) can be transformed as:

$$(16.) \quad \sigma^2 = \frac{1}{R} \underline{\omega}' \left[I - \frac{1}{R} (\underline{1} \underline{1}') \right] \underline{\omega}$$

Let us assume a linear welfare function

$$(17.) \quad \underline{w} = S' \underline{\lambda}$$

with non-negative weights $\underline{\lambda} = (\lambda_1, \dots, \lambda_I)'$

Then we find after substitution of (17.) in (13.) and (16.):

$$(18.) \quad v^2 = \frac{\sigma^2}{\mu^2} = \frac{\underline{\lambda}' S \left[I - \frac{1}{R} \underline{1}\underline{1}' \right] S' \underline{\lambda}}{\underline{\lambda}' S \left[\frac{1}{R} \underline{1}\underline{1}' \right] S' \underline{\lambda}}$$

or

$$(19.) \quad v^2 = -1 + \frac{\underline{\lambda}' [S S'] \underline{\lambda}}{\underline{\lambda}' \left[S \left(\frac{1}{R} \underline{1}\underline{1}' \right) S' \right] \underline{\lambda}}$$

The last result indicates how the outcome for the inequality measure V depends on a) the information of the profile matrix S as well as on b) the weights of the welfare function \underline{w} . Especially the second element is very interesting. It implies that inferences about multidimensional inequality can only be drawn with the weights attached to the objectives in mind.

(19.) presents a general framework for the analysis of multidimensional regional inequality. The coefficient of variation for the individual profile elements can be obtained by putting $\underline{\lambda} = (1, 0, \dots, 0)'$, $(0, 1, 0, \dots, 0)'$, ... etc.

Similar to the parametric programming approach, (19.) can be solved for intermediate weight vectors $\underline{\lambda}$.

Of special interest is the vector $\underline{\lambda}^m$ which gives rise to a minimum amount of inequality. The calculation procedure to find $\underline{\lambda}^m$ is almost identical to the one used in discriminant analysis (cf. Bolch and Huang [1970]). In Appendix A we discuss the possibility to derive $\underline{\lambda}^m$.

B. Derivation of the mean deviation with respect to the maximum value

The second measure of multidimensional inequality we discuss here, is based on the notion of a discrepancy with respect to an ideal regional profile. The ideal regional profile $*s_i$ is defined as

$$(20.) \quad *s_i = \max_r s_{ir} \quad i = 1, \dots, I$$

The discrepancy between the ideal profile and the profile matrix S is:

$$(21.) \quad D = [*s \ 1' - S]$$

The corresponding inequality measure MD is - after the introduction of weights $\lambda = (\lambda_1, \dots, \lambda_I)'$ for the discrepancies and after normalization - :

$$(22.) \quad MD = \frac{\lambda' [*s \ 1' - S] \ 1}{\lambda' | S \ 1 |}$$

where $|S \ 1|$ is the vector with the absolute values of $S \ 1$.

MD allows the application of the same method of varying systematically and successively the values of λ as at V^2 .

The problem for which values of λ MD attains its maximum or minimum value is easier to solve than for V. Consider

$$(23.) \quad \begin{cases} \min! & \frac{\lambda' c}{\lambda' d} \\ \lambda \geq 0 \end{cases}$$

We find that MD attains its minimum $\min \left(\frac{c_i}{d_i} \right)$ for $\lambda = \epsilon_i$, where $\epsilon_i = (0, \dots, 0, 1, 0, \dots, 0)'$ with an element 1 in the i -th position.

Of course, these are not the only inequality measures which can be derived. Other measures may reveal interesting correspondences with multiple objective decision methods. The Gini-coefficient in a multi-dimensional setting for example, has close links with decision methods using pairwise comparisons, such as the concordance analysis (cf. Roy [1971] and Van Delft and Nijkamp [1977]).

For a classification of inequality measures it will be useful to define a number of potential properties of those measures. Denote therefore, the welfare inequality measure as j . We know, that j depends on the weights λ :

$$(24.) \quad j = j(\lambda)$$

Denote the corresponding inequality measures for the I single profile elements as $\underline{\delta} = (\delta_1, \dots, \delta_I)$. Then we know:

$$(25.) \quad \delta_i = j(\underline{\epsilon}_i)$$

For the function j the following properties may be relevant

$$(26.) \quad j \text{ is linear when } j(\underline{\lambda}) = \underline{\delta}' \underline{\lambda} \quad \text{for all } \underline{\lambda} \geq 0, \mathbf{1}' \underline{\lambda} = 1$$

$$(27.) \quad j \text{ is concave when } j(\underline{\lambda}) \geq \underline{\delta}' \underline{\lambda} \quad \text{for all } \underline{\lambda} \geq 0, \mathbf{1}' \underline{\lambda} = 1$$

$$(28.) \quad j \text{ is convex when } j(\underline{\lambda}) \leq \underline{\delta}' \underline{\lambda} \quad \text{for all } \underline{\lambda} \geq 0, \mathbf{1}' \underline{\lambda} = 1$$

Another characteristic of j has to do with the question whether a simultaneous analysis of the elements of inequality results in a more equal picture of inequality than a separate analysis of the elements (so-called synergetic effects). This can be formalized as follows: j has the maximum equality property when

$$(29.) \quad \max_{\underline{\lambda}} j(\underline{\lambda}) = \max_i (\delta_1, \dots, \delta_I)$$

and

j has the minimum equality property when

$$(30.) \quad \min_{\underline{\lambda}} j(\underline{\lambda}) = \min_i (\delta_1, \dots, \delta_I)$$

A brief examination of MD and V shows, that MD is linear in a special case (viz. when S is normalized) and that it has the maximum as well as the minimum equality property. V has neither the maximum equality property, nor the minimum equality property.

5. Numerical Applications.

The concepts and methods exposed in the preceding sections will be applied to Dutch regional data from 1970.

For the 11 provinces and 13 profile elements, data have been collected (so $R = 11$ and $I = 13$).

The socio-economic (SE) variables are:

- s_1 : fiscal income per capita (measured in guilders).
- s_2 : unemployment rate.
- s_3 : wealth per capita (measured in guilders).
- s_4 : index of cost of living (especially of housing rents).

The environmental (E) variables are:

- s_5 : population density (measured in persons per square km).
- s_6 : size of natural environments as percentage of total regional area.
- s_7 : rate of industrialization (the quotient of industrial output and total output).
- s_8 : index of the emission of pollutants related to regional area.

The infrastructural (I) variables are:

- s_9 : density of transport network (length of roads measured in kms divided by the size of the regional area measured in square kms).
- s_{10} : cultural index (index of cultural centres and sport accommodations per capita).
- s_{11} : educational index (index of the number of schools of various types per capita).
- s_{12} : distance to the centre of the Netherlands (measured in kms).
- s_{13} : medical index (index of the number of physicians, chemists, hospitals etc. per capita).

In Appendix B these variables have been defined more precise. It also contains the sources of the data. Table 1 presents the profile matrix S of order 13×11 . The variables 2, 4, 5, 7, 8 and 12 have been multiplied with a factor -1 to indicate that for these variables a smaller value is preferred to a larger value. The table clearly shows that provinces with a good socio-economic performance (such as North and South Holland) have a relatively bad natural environment. The opposite holds true for provinces such as Friesland and Drenthe. It is striking that the provinces which are lagging in socio-economic respects have relatively good performances of some infrastructural variables.

Table 1 The profile matrix S.

province variable	Gro- ningen	Fries- land	Drenthe	Over- ijssel	Gelder- land	Utrecht	North Holland	South Holland	Zeeland	North Brabant	Limburg
	1	2	3	4	5	6	7	8	9	10	11
1	4652.60	4352.00	4419.00	4501.00	4752.00	5438.00	5846.00	5859.00	4990.00	4556.00	4422.00
2	-3.20	-3.30	-4.30	-2.30	-4.70	-1.20	-1.80	-1.50	-2.30	-2.20	-2.80
3	5978.00	4825.00	4954.00	4837.00	5355.00	7134.00	6563.00	6559.00	7440.00	4316.00	3370.00
4	-144.00	-140.00	-143.00	-137.00	-154.00	-153.00	-148.00	-153.00	-144.00	-153.00	-147.00
5	-227.00	-159.00	-143.00	-249.00	-311.00	-623.00	-856.00	-1049.00	-176.00	-377.00	-472.00
6	2.30	9.40	16.40	13.50	23.20	15.60	9.30	4.30	3.30	16.40	16.30
7	-0.50	-0.30	-0.45	-0.50	-0.43	-0.35	-0.40	-0.40	-0.50	-0.54	-0.53
8	-7.90	-2.40	-3.50	-7.10	-7.70	-12.30	-31.40	-57.20	-8.00	-9.60	-14.70
9	1.50	1.50	1.70	2.10	2.30	2.50	2.90	3.40	2.60	2.30	3.20
10	223.00	247.00	245.00	179.00	181.00	165.00	196.00	175.00	208.00	194.00	178.00
11	17.50	19.60	17.10	14.50	13.10	13.50	11.30	11.10	15.90	12.20	12.90
12	-195.00	-170.00	-170.00	-90.00	-65.00	0.00	-55.00	-65.00	-170.00	-55.00	-179.00
13	146.20	79.70	104.20	35.30	108.00	172.90	133.90	116.90	80.80	82.70	84.50

The correlations between the variables can be found in Table 2. The mutual correlations between the variables within each subprofile do not seem to be significantly higher than the correlations between variables of different subprofiles. Notice the negative correlation between the emission of pollution (s_8) and the industrialization rate (s_7). The latter appears to be a poor proxy for environmental pollution.

Table 2. The correlation matrix.

1.00	.71	.74	-.51	-.87	-.34	.58	-.81	.64	-.47	-.62	.60	.58
.71	1.00	.47	-.54	-.66	.06	.34	-.51	.64	-.85	-.74	.83	.36
.74	.47	1.00	-.21	-.34	-.53	.52	-.34	.17	-.13	-.08	.29	.57
-.51	-.64	-.21	1.00	.59	-.33	-.25	.45	-.53	.56	.70	-.65	-.42
-.87	-.66	-.34	.59	1.00	.16	-.44	.93	-.79	.59	.78	-.62	-.44
-.34	.06	-.53	-.33	.16	1.00	.02	.36	-.03	-.27	-.25	.38	-.11
.58	.34	.52	-.25	-.44	.02	1.00	-.33	.05	-.04	-.03	.46	.57
-.81	-.51	-.34	.45	.93	.36	-.33	1.00	-.75	.44	.65	-.39	-.25
.64	.64	.17	-.53	-.79	-.03	.05	-.75	1.00	-.73	-.84	.42	.04
-.47	-.85	-.13	.56	.59	-.27	-.04	.44	-.73	1.00	.82	-.71	-.26
-.62	-.74	-.08	.70	.78	-.25	-.03	.66	-.84	.82	1.00	-.73	-.18
.60	.33	.29	-.65	-.62	.38	.46	-.39	.42	-.71	-.73	1.00	.43
.58	.36	.57	-.42	-.44	-.11	.57	-.25	.04	-.26	-.18	.43	1.00

Table 3 presents the results of the application of the interdependence analysis on the data set for 1 to 8 selected variables.

number of selected variables		1	2	3	4	5	6	7	8
correlation coefficient resulting from the regression of each variable with the selected variables as explaining variables. (S indicates a selected variable)	s ₁	.959	.996	.992	.985	.991	.999	S	S
	s ₂	.732	.926	.968	.973	.980	.989	.992	S
	s ₃	.912	.960	.958	.962	.974	.988	S	.998
	s ₄	.945	S	.994	.987	.992	.998	.999	.999
	s ₅	.829	S	S	.984	.988	.989	.993	S
	s ₆	.723	.797	S	S	S	S	S	S
	s ₇	.917	.985	.971	.969	.981	.986	.989	S
	s ₈	.649	.927	.949	S	S	S	S	S
	s ₉	S	.976	.967	.968	S	S	.997	.995
	s ₁₀	.873	.985	S	.980	.980	S	S	.998
	s ₁₁	.836	.982	.993	.979	.980	.995	.995	.995
	s ₁₂	.616	.828	.916	S	S	S	S	S
	s ₁₃	.877	.946	.940	S	S	S	S	S
minimum correlation		.616	.797	.916	.962	.974	.986	.989	.995

Table 3. Results of the interdependence analysis. *

* The authors thank Bas Wiersma for the possibility to use his interdependence analysis computer program.

The set of selected variables appears to become stable when 4 or more variables are selected: s_6 , s_8 , s_{12} and s_{13} are elements of all subsequent selections. The variables of the socio-economic subprofile play a minor role in the selected sets; they are apparently not very representative for the set of variables. When, for example, 6 of the 13 variables are selected no economic variables are chosen, which is striking because in political debates about regional inequality the economic variables dominate the discussions. This points out a weakness of interdependence analysis: it performs a selection on purely numerical and statistical grounds and is not based on theoretical or other a priori considerations concerning the meaning of the variables. Consequently, interdependence analysis should be integrated with such a selection on theoretical grounds. In our case this produces a difficulty because the variables of the data set are not homogeneous: they can be assigned to three subclasses in a meaningful way. Interdependence analysis in its present form ignores this datum, although it is possible to refine interdependence analysis such that a certain number of variables out of all subclasses will certainly be selected.

The coefficient of variation and the mean deviation of the various profile elements are presented in Table 4. It is striking that the interregional inequality for the socio-economic variables is considerably smaller than for the environmental variables. Especially the inequality in s_5 , s_6 , s_8 and s_{12} is substantial.

variable	coefficient of variation (V^2)	mean deviation (MD)
1	.11	2.16
2	.36	5.54
3	.22	3.68
4	.04	.74
5	.68	7.27
6	.53	10.59
7	.13	2.51
8	1.04	9.21
9	.26	4.82
10	.14	2.64
11	.18	3.94
12	.58	11.00
13	.27	6.50

Table 4. Inequality measures for 13 profile elements.

An extension of (19) and (23) by means of parametric programming will not be performed here: it would be very time consuming because of the large number of profile elements used. Instead, a more detailed analysis will be executed with a reduced number of profile elements. As interdependence analysis appeared not very convincing as a firm base to make a selection from many elements, we may construct an index aggregating the variables in each sub-profile. Thus only three indices remain.

As the variables in the sub-profiles have different dimensions, they have to be normalized first. This has been carried out such that the length of the profile vector after normalization is equal to 1. The indices are obtained now by calculating the unweighted average of the normalized variables in each sub-profile.¹⁾ The results of these calculations can be found in Table 5. It shows a positive correlation between the SE- and the I- index ($r = .80$) and a negative cor-

province	SE	E	I
Groningen	.047	-.131	.162
Friesland	.019	-.043	.153
Drenthe	-.014	-.015	.161
Overijssel	.059	-.068	.168
Gelderland	.091	-.014	.192
Utrecht	.151	-.105	.259
North Holland	.129	-.251	.222
South Holland	.139	-.399	.214
Zeeland	.109	-.118	.154
North Brabant	.053	-.086	.183
Limburg	.014	-.117	.150

Table 5. The profile matrix consisting of a socio-economic, an environmental and an infrastructural index.

relation between the E- index on the one hand and the SE- as well as the I- index on the other ($r = -.60$ and $-.43$, respectively). Because of the limited

1) The cost of living index received a smaller weight (.25 instead of 1) because this variable concerns only the housing rents, which is of limited importance.

number of criteria some cases of dominance arise, which are described in Table 6. The provinces Gelderland and Utrecht

	1	2	3	4	5	6	7	8	9	10	11
1	-										
2		-									+
3			-								
4				-							+
5	+	+	+	+	-					+	+
6	+					-	+	+	+		+
7							-				
8								-			
9									-		
10										-	
11											-

Table 6. Dominance relationships among 11 provinces.

appear to dominate several other provinces. Gelderland owes its position to its splendid natural environment and its reasonable performance for the other indices. Utrecht achieves the highest value for the socio-economic as well as the infrastructural index. Relatively poor is the performance of Groningen and Limburg which are dominated by several other provinces.

In Table 7 the results of a parametric programming approach applied to the inequality in the three welfare indices can be found. V^2 and MD indicate that the inequality is at a minimum for the infrastructural index.

As Table 4 shows a considerable inequality in the infrastructural variables, we conclude that in this case the construction of an aggregate index implies the abandonment of the inequalities of the components to a certain extent. The opposite holds true for the socio-economic variables. Here we find that the aggregate index is less equally distributed than the separate components. The reason of this

weights			V ²	MD
λ_1	λ_2	λ_3		
.0	.1	1.0	.033	5.2
.3	.2	.8	.046	6.5
.6	.4	.6	.408	8.1
.3	.5	.4	72717.109	10.4
.0	.3	.2	1.845	13.7
.6	1.0	.9	.770	18.9
.2	.3	.8	.349	6.4
.2	.2	.6	.068	8.2
.2	.4	.4	.989	10.7
.2	.6	.2	6.486	14.5
.2	.3	.0	.922	20.9
.4	.0	.6	.079	8.3
.4	.2	.4	.124	11.0
.4	.4	.2	4.508	15.5
.4	.6	.0	1.505	23.7
.6	.3	.4	.136	11.5
.6	.2	.2	.294	16.9
.6	.4	.3	39.363	27.9
.8	.3	.2	.253	18.8
.8	.2	.0	1.033	35.0
1.0	.3	.0	.526	49.6

Table 7. Outcomes of inequality measures for various combinations of weights.

difference can be found in the correlation matrix (Table 2), which shows positive correlations between the main socio-economic variables, but a mixture of positive and negative correlations between the infrastructural variables.

Table 7 clearly shows that MD has the maximum - as well as the minimum - equality property. Its extreme values are attained for extreme combinations of weights (viz. $\lambda = \underline{\epsilon}_1$ and $\lambda = \underline{\epsilon}_3$). The coefficient of variation does not share this characteristic. It may attain values which are far bigger than the value for $\lambda = \underline{\epsilon}_2$ (.770). In the way suggested in Appendix A, the weights λ^m have been calculated which lead to a minimum value for V. We find that $\lambda^m = (0, .07, .93)$ and that the corresponding value for V² is equal to .032. When the welfare levels of the provinces are calculated for these weights the outcome is that Utrecht has maximum welfare and Limburg minimum welfare.

When the provinces are ranked in descending order of welfare we find the series: Utrecht, North Holland, Gelderland, South Holland, North Brabant, Overijssel, **Drenthe**, Groningen, Friesland, Zeeland, Limburg.

6. Conclusions and Suggestions for Further Research.

Various types of problem regions can be distinguished: underdeveloped, depressed and congested ones (cf. Stilwell [1972]). The elements of the regional profiles used in this study may be used to identify the problem regions (see Table 8.).

		type of problem region		
		underdeveloped	depressed	congested
profile	socio-economic	-	-	+
elements:	environmental	+	-	-
	infrastructural	-	+/-	+/-

Table 8. A Typology of problem regions.

The regional profiles are also relevant for the study of interregional migration of families and industries.

The multidimensional inequality measures developed in this study may contain fruitful information for the evaluation of policy proposals. They are an essential link between integral interregional models and an integral regional policy. The emphasis put on the integral character of regional inequality may prevent one-sided policies that aim at reducing inequality in only some special respects.

Various subjects of further research can be suggested:

- in addition to the two inequality measures other inequality measures might be introduced
- the analysis may be repeated for regions of a smaller scale
- an international comparison of interregional inequalities would be interesting
- the same holds true for an intertemporal comparison
- ordinal information on interregional inequalities might be used via multidimensional scaling techniques.

Appendix A. The Minimization of the Coefficient of Variation.¹⁾

The problem dealt with in this appendix is: how can the vector $\underline{\lambda} = \underline{\lambda}^m$ be determined, for which

$$(A.1) \quad V^2 = -1 + \frac{\underline{\lambda}' K \underline{\lambda}}{\underline{\lambda}' M \underline{\lambda}}$$

attains its minimum. In (19) we find:

$$(A.2) \quad \begin{cases} K = S S' \\ M = S \left(\frac{1}{R} \underline{1} \underline{1}' \right) S' \end{cases}$$

so we may conclude that K and M are symmetric and positive (semi-) definite. Consequently, M has rank 1 and is singular.

In exceptional cases $\underline{\lambda}^m$ can be found in a straightforward way. For instance if S is a square non-singular matrix (I = R), it is clear that

$$(A.3) \quad \underline{\lambda}^m = (S')^{-1} \underline{1}$$

is the solution of the problem, because then $V^2 = 0$.

A more general solution can be found when we use the following theorem about quadratic forms (cf. Franklin [1968] and Gantmacher [1965]): "the solution of

$$(A.4) \quad \max! \frac{\underline{\lambda}' M \underline{\lambda}}{\underline{\lambda}' K \underline{\lambda}}$$

where K and M are symmetric and positive (semi) definite, while K is non-singular, is equal to the largest eigenvalue of the matrix $K^{-1} M$. $\underline{\lambda}^m$ is equal to the corresponding eigenvector".

As the minimization of (A.1) is equivalent to (A.4) we have to concentrate on the eigenvalues of $K^{-1} M$.

1) The authors want to thank Rens Trimp for his advise on the subject of this appendix.

As M has rank 1, also $K^{-1} M$ has rank 1. Consequently, $K^{-1} M$ has only one non-zero eigenvalue which is equal to the trace of $K^{-1} M$.

Another problem arises when one wishes to interpret $\underline{\lambda}^m$. This vector is a series of political weights attached to various criteria in such a way that a minimum amount of inequality results. As all criteria have been defined such that larger values are preferred to smaller values, it is reasonable to add the side-condition that all weights are non-zero. Hence, (A.4) has to be replaced by

$$(A.5) \left\{ \begin{array}{l} \max: \frac{\underline{\lambda}' M \underline{\lambda}}{\underline{\lambda}' K \underline{\lambda}} \\ \text{subject to } \underline{\lambda} \geq \underline{0} \end{array} \right.$$

It is impossible to find an analytical solution of A(5). In addition to the existing numerical methods for non-linear programming problems, one may devise an algorithm for this special case. Especially when the number of criteria is not too large, a repetitive solution of (A.4) with some weights set equal to zero may prove to be efficient.

Another way to deal with (A.5) is to interpret it as a geometric programming problem. The general specification of a geometric programming model is:

$$(A.6) \left\{ \begin{array}{l} \min \varphi = \underline{c}_0' \underline{f}_0 \\ \text{s.t.} \\ \underline{c}_j' \underline{f}_j \leq 1 \quad j = 1, \dots, J \\ \text{with} \\ \underline{c}_j = (c_{j1}, \dots, c_{jI})' \geq \underline{0} \quad j = 0, 1, \dots, J \\ \text{and} \\ \ln \underline{f}_j = A_j \ln \underline{x}, \\ \underline{x} \geq \underline{0} \end{array} \right.$$

where \underline{f}_j is an $(I \times 1)$ vector, A_j an $I \times K$ matrix with typical coefficients a_{ik}^j ($i = 1, \dots, I; k = 1, \dots, K$) and \underline{x} a $K \times 1$ vector of decision variables. The coefficients a_{ik}^j may be positive or negative (see for

a more extensive exposition among others Duffin et al [1967], Nijkamp [1972] [1978]). The latter model can be proved to be a convex programming model, which has a unique solution. Numerical procedures to derive this solution are inter alia gradient techniques and steepest descent methods.

It is easily seen that (A.5) is a geometric programming model:

$$(A.7) \left\{ \begin{array}{l} \min! (\underline{\lambda}' K \underline{\lambda}) z^{-1} \\ \text{s.t.} \\ \underline{\lambda} \geq 0 \\ (\underline{\lambda}' M \underline{\lambda}) z^{-1} = 1 \\ z \geq 0 \end{array} \right.$$

It should be noted that a unique solution is only guaranteed in case of positive coefficients. In all other cases a so-called signomial programming emerges, which may also be solved by means of numerical techniques, but for which no unambiguous solution can be proved to exist.

Anyway, the conclusion of the latter analysis is that a solution for models of type (A.5) can be derived in principle.

Appendix B. Specification of the welfare profile of a province

a. The socio-economic subprofile:

- S1: fiscal income per capita (measured in guilders).
- S2: ratio of the number of unemployed persons (male and female) and total dependent labour force.
- S3: wealth ($\geq 100,000$ guilders) per capita.
- S4: index of cost of living; It is assumed that this is equal for all Dutch provinces, except for housing costs. As housing costs we have taken the so-called CBS-norm i.e. adjusted bruto rent.

b. The environmental subprofile:

- S5: population density (measured in persons per square KM).
- S6: size of natural environments (woods, waste lands, reed and rush) as percentage of total regional area.
- S7: rate of industrialization, i.e. the quotient of industrial output of enterprises and government (factor costs).
- S8: quantity of pollutants related to the surface of a province; We have used a study of the Institute for Environmental Problems of the Free University (IvM-VU, 1977) in which industrial sectors are characterized by 4 pollution criteria:
 - 1) aggregate air-pollution (LUVO E)
 - 2) aggregate water-pollution by heavy metals (ZME)
 - 3) aggregate water-pollution by OXYGEN binding materials (INW E)
 - 4) chemical waste (tons/year).

In this study so-called primary pollution-coefficients (emission of a pollutant per value unit production in a certain sector) have been calculated. By multiplying the production (in value units) of sectors in a certain province with these coefficients we obtain the quantity of pollutants.

c. The infrastructural subprofile:

- S9: density of transport network (length of roads measured in KMS divided by the size of the regional area measured in square KMS).
- S10: a cultural index defined as follows:

$$CI_j = \frac{SC_j + AC_j}{B_j}$$

in which SC_j = number of social-cultural centers and sport-accmodations in province j, AC_j = number of concerts, opera-, theatre- and balletperformances in province j, B_j = total population in province j.

S11: an educational index constructed as follows

$$SVI_j = \frac{\sum_{i=1}^4 \alpha_i S_{ij}}{B_j}$$

in which S_{ij} = number of schools of type i in province j, α_i = weight attached to a school of type i; B_j = total population in province j.

The following school-types were used in the analysis:

- primary education (S_{1j})
- elementary professional education (S_{2j})
- secondary education (S_{3j})
- higher education (S_{4j})

The following set of weights was specified:

$$0 < \alpha_1 < \alpha_2 < \alpha_3 < \alpha_4 < 1$$

S12: distance to the centre of the Netherlands (measured in KMS). This centre was determined by means of a Weber-analysis. The distance has been measured from this centre to each of the 11 countytowns.

S13: the medical index has been constructed in the following way:

$$MV_j = \frac{\sum \beta_i V_{ij}}{B_j}$$

in which V_{ij} = number of medical services of type i in province j, β_i = weight attached to a medical service of type i, B_j = total population of province j.

The following types of medical services were considered: family doctors, medical specialists, social medical doctors, dentists, district nurses, confinement nurses, pharmaceutical chemists, psychical hospitals and hospitals.

Data sources:

- S1: Regionaal Statistisch Zakboek 1974 (CBS*).
- S2: Internal note of the CBS.
- S3: Regionaal Statistisch Zakboek 1974 (CBS).
- S4: Maandstatistiek van de prijzen (CBS, april 1976 and september 1977).
- S5: Regionaal Statistisch Zakboek 1972 (CBS).
- S6: Internal note of the CBS, Hoofdafdeling Landbouwstatistiek, Bodemstatistiek.

- S7: Regionale indicatoren 1970 (CBS).
- S8: Milieuverontreiniging en economische Structuur, Rapport aan de Minister van Volksgezondheid en Milieuhygiëne, Verkenningen van het IvM-VU, 1977.
- S9: Regionaal Statistisch Zakboek 1972 (CBS).
- S10: Sociaal Culturele Centra, 1969 (CBS); Regionaal Statistisch Zakboek 1972 (CBS).
- S11: Internal note of the CBS, Hoofdafdeling Statistieken van Onderwijs en Wetenschappen.
- S13: Regionaal Statistisch Zakboek 1972 (CBS).

* The authors want to thank Mr. Strankinga of the CBS (Central Bureau of Statistics) for making available datamaterial.

References.

- Bartels C.P.A., 1977, Economic Aspects of Regional Welfare (Martinus Nijhoff, Leiden).
- Beale, E.M.L., M.G. Kendall and D.W. Mann, 1967, "The Discarding of Variables in Multivariate Analysis", Biometrika 54 . 357-366.
- Bolch, B.W., and C.J. Huang, 1974, Multivariate Statistical Methods and Economics, (Prentice Hall, Englewood Cliffs).
- Boyce, D.E., A. Farli and R. Weischedel, 1974, Optimal Subject Selection, (Springer Verlag, Berlin).
- Coates, B.E., R.J. Johnston and P.L. Knox, 1977, Geography and Inequality, (Oxford University Press, Oxford).
- Delft, van A., and P. Nijkamp, 1977, Multicriteria Analysis and Regional Decision-Making (Martinus Nijhoff, The Hague).
- Duffin, R.J., E.L. Peterson, and C.M. Zener, 1967, Geometric Programming, (Wiley, New York).
- Fisher, M., 1977, "Numerical Taxonomy and Regionalization", paper presented at the 17th European Conference of the Regional Science Association, Gracow (mimeographed)
- Franklin, J.N., 1968, Matrix Theory, (Prentice Hall, Englewood Cliffs).
- Gantmacher, F.R., 1965, Matrizen Rechnung, (VEB Deutscher Verlag der Wissenschaften, Berlin).
- Johnston, R.J., 1968, Choice in Classification: the Subjectivity of Objective Methods", Annals of the American Association of Geographers, 575-589.
- Myrdal, G., 1957, Economic Theory and Underdeveloped Regions, (Duckworth, London).
- Nijkamp, P., 1972, Planning of Industrial Complexes by Means of Geometric Programming, (Rotterdam University Press, Rotterdam).
- Nijkamp, P., 1977a, "Stochastic Quantitative and Qualitative Multi-Criteria Analysis for Environmental Design", Papers of the Regional Science Association 39, 175 - 199.
- Nijkamp, P., 1977 b, "A Spatial Complex Analysis of Agglomeration and Settlement Patterns", Discussion Paper no. 1977-2, Free University, Dept. of Economics, Amsterdam.

- Nijkamp, P., 1977, Theory and Application of Environmental Economics, (North Holland, Amsterdam).
- Nijkamp, P., 1978, "A Note on Duality Conditions of Quadratic Assignment Models", Discussion Paper no. 1978-5, Department of Economics, Free University, Amsterdam.
- Nijkamp, P., and C. Verhage (eds.), 1978, Regionaal Beleid, (Stenfert Kroese, Leiden).
- Paelinck, J.H.P., and P. Nijkamp, 1976, Operational Theory and Method in Regional Economics, (Saxon House, Lexington).
- Richardson, H.W., 1977, "Efficiency and Equity", Spatial Inequalities and Regional Development (H.Folmer and J.Oosterhaven, eds.), (Martinus Nijhoff, The Hague).
- Roy, B., 1971, "Problems and Methods with Multiple Objective Functions", Mathematical Programming, 239-266.
- Stilwell, J.B., 1972, Regional Economic Policy, (MacMillan Press, London).

Serie Research Memoranda:

- 1977-1 L. Hordijk, P. Nijkamp, Estimation of Spatiotemporal Models
New directions via distributed lags
and Markov schemes.
- 1977-2 P. Nijkamp, Gravity and entropy models: The state
of the art.
- 1978-1 J. Klaassen, Valuta problemen in de Jaarrekening.
- 1978-2 P. Nijkamp, M.W. van Psychometric scaling and preference methods
Veenendaal, in spatial analysis.
- 1978-3 P. Nijkamp, J. Spronk, Interactive Multiple Goal Programming.
- 1978-4 P. Nijkamp, P. Rietveld, New Multi Objective Techniques in
Physical Planning.
- 1978-5 P. Nijkamp, Decision Models for Planning against
Stagnation.
- 1978-6 P. Nijkamp, An Analysis of Interdependent Decision via
Non-Linear Multi-Objective Optimization.
A Theory of Displaced Ideals.
- 1978-7 P. Nijkamp, Conflict Patterns and Compromise Solu-
tions in Fuzzy Choice Theory.
An analysis and application.
- 1978-8 J.D.P. Kasper, De Bijdrage van Detailhandelsmarketing
aan Maatschappelijke Welvaart en Welzijn.