International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

# A New Approach in Energy Demand. Part I: Methodology and Illustrative Examples

Beaujean, J.-M., Chaix, B., Charpentier, J.-P. and Ledolter, J.

IIASA Professional Paper
March 1977

PP-77-4

# A NEW APPROACH IN ENERGY DEMAND
# PART I: METHODOLOGY AND ILLUSTRATIVE EXAMPLES

J.M. Beaujean
B. Chaix
J.P. Charpentier
J. Ledolter

March 1977

ABSTRACT

A great deal of work has been carried out on the relation between per capita GNP and per capita energy consumption.

In this short paper we substitute the structure of GNP to its absolute level. Three sectors were only retained: namely agriculture, industry and services (including transportation). The relation between per capita energy consumption and GNP structure explicitly constructed and adjusted on data is a potential in the space of GNP structures.

TABLE OF CONTENTS

<u>LIST OF FIGURES</u>

## A NEW APPROACH IN ENERGY DEMAND

## PART I. METHODOLOGY AND ILLUSTRATIVE EXAMPLES

To make sound forecasting on world energy consumption, one has to embed the energy systems in the overall economic strata and to take development strategies, especially for developing countries, into account. Many methodological reports have been published on the subject. In the first part, we will review these techniques and point out their weak points. In the second part, we will present an alternative approach. In the third part, we will apply it to a sample of a given year and to historical data, whose results will be used to build strategies for probable scenarios of development. However, this analysis has remained more or less qualitative, and we will suggest some guidelines for formalization.

### I CLASSICAL ANALYTICAL TOOLS

The classical analytical tools used to forecast energy demand are based on:

1. econometric analysis,
2. engineering analysis of systems,
3. energy content and energy basket approach.

Some studies using one of these approaches are very detailed and go down to microeconomic levels; some are broader and use macroeconomic indicators or variables. Therefore, we can build a classification of approaches and degrees of analysis.

Table 1. Classification of possible methodologies which might be used in energy demand study.

| Scope of analysis approaches | World (global analysis) | Country or group of countries (macro analysis) | Sectors (micro analysis) |
|---|---|---|---|
| Econometric | GNP/cap-energy consumption/cap | Estimation of the coefficients of a theoretical formula | Elasticity/ prices |
| Engineering | World average figures | Average energy input per output | Diagram of energy flows |
| Energy content | Scenarios | I - O Herendeen & Bullard | Process analysis "à la Slesser" |

Our aim is to make an energy demand forecast for the world. This table shows that the only tool commonly used is the correlation between GNP/cap and energy consumption/cap. All other methods (column 3) could be employed and their results aggregated; however, the number of equations and variables would become too large.

The global econometric approach is too broad to give accurate results. Mainly because the linear equation holds between upper and lower bounds: e.g. GNP/cap $\leq$ energy consumption $\leq$ 3 GNP/cap.

Therefore, we look for other variables. We found that the GNP structure is more significant for energy demand forecasts than its absolute level. This is mainly due to the fact that industry, services and agriculture have very different energy consumption patterns. So, we divided the GNP in these three sectors:

a) agriculture,
b) industry,
c) services.

This three-dimensional vector is a better indicator of development than the gross value of GNP/cap. It allows us to capture final energy demand over long term in more detail, given development scenarios and explicit relations among energy per capita, GNP structure and development level.

## II. METHODOLOGY

If we define:

A as agriculture share in GNP,
I as industry share in GNP,
S as services and transportation shares in GNP,

then every country is defined by a vector $\begin{pmatrix} A \\ I \\ S \end{pmatrix}$ which can be represented as a point in $\mathbb{R}^3$. Given that $A + I + S = 1$, the set of points is in a sub-domain defined by this equation.

The summits of the triangle are the extremities of the unitary vectors:

$$\vec{e}_1 = (1, 0, 0)$$

$$\vec{e}_2 = (0, 1, 0)$$

$$\vec{e}_3 = (0, 0, 1)$$

The new coordinates (a, s, i) are defined by $(a, s, i) = \lambda(A, S, I)$ and are the usual triangular coordinates.

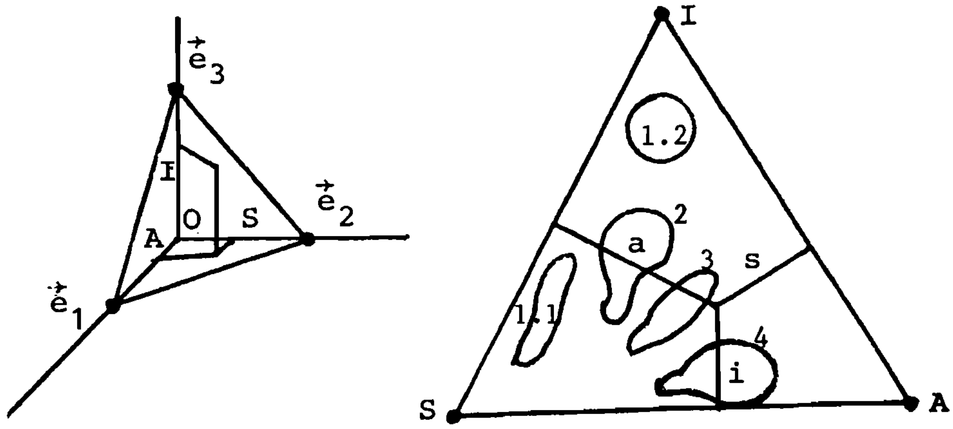



Figure 1. Representation of GNP structure in triangular coordinates.

The plotting program we use directly yields this triangular representation, as is shown in Figure 1. This graph presents five groups of countries:

1.1 Highly developed market economy countries,
1.2 Highly developed planned economy countries,
2. Developed countries,
3. Third world,
4. Fourth world.

To make broad distinctions, the following comments can be made :

1.1 This group gathers all countries which have relatively small shares of agriculture. This means that agriculture here is very efficient and does not pose any problem in contrast to groups 3 and 4. They are more service-oriented than group 1.2, which shows the importance of the service sub-sector composed of banking, financial institutions, consumer services, leisure services, etc.

<u>Highly developed market economy countries</u>:

Austria, Australia, Canada, Chile, Denmark, France, F.R.G., Israel, Japan, Luxemburg, Netherlands, Republic of South Africa, Sweden, United Kingdom, U.S.A.

1.2 Countries in this group exhibit a relatively more important share of industry in the GNP, and a correlated lower level of agriculture and services. This is easily explained by the fact that the development strategy of most of these countries emphasized the heavy industry. Because of their socio-economic particularities, services did not take the place they have in market economy countries.

Highly developed planned economy countries

C.S.S.R., G.D.R., Hungary, Poland, Rumania, U.S.S.R.

2. This group, compared with the group 1.1. has more important shares of agriculture and industry. As J.-P. Charpentier shows in his article[1], these countries are characterized by their growth rates, and are on the verge of attaining the same standard as groups 1.1. and 1.2. in the near future.

Developed countries

Argentina, Bolivia, Finland, Greece, Irak, Iran, Ireland, Mexico, Portugal, Rhodesia, Spain, Yugoslavia.

3. This group is balanced between shares of services and agriculture, in contrast to the industry-intensive countries of group 1.1. The four more service-oriented countries are Jordan, Syria, Panama and Guatemala. Jordan is known to have a chronic deficit in its balance of payments[2]. Syria has many installations for transporting energy products from Irak, thus providing great revenues. Panama is well known for its revenues from both the canal and pavilion facility fees.

Third world countries

Burma, Brazil, Columbia, Ecuador, Egypt, Guatemala, Kenya, South Korea, Malaysia, Malawi, Morocco, Nicaragua, Panama, Paraguay, Peru, Philippines, Sri Lanka, Syria, Thailand.

4. This group is characterized by a share of agriculture higher than 45%. We find countries in this group from South East Asia and Africa, but none from Latin America. Their most frequent and most important problem is the provision of good to the population because of the imbalance between growth of GNP and population growth.

Thus, we rediscover the well-known classification of countries. Apart from this clarification and for the analysis and sound forecasts of their energy demand, we suggest to divide these countries according to social, cultural and climatic factors. For instance, we do not hesitate to group Austria England, France, F.R.G., Japan and South Korea together. Our classification is shown below.

Grouping of countries according to their socio-cultural

Climatic shares

1. C.S.S.R., G.D.R., Hungary, Poland, U.S.S.R.,
2. Australia, Canada, U.S.A.,
3. Austria, England, France, F.R.G., Japan, South Korea,

---

[1] J.-P. Charpentier, Toward a Better Understanding of Energy Consumption II, Factor Analysis: a New Approach to Energy Demand, Energy, Pergamon Press 1976,

[2] And this deficit is counted for in the services share.

4. Finland, Netherlands, Norway, Sweden,
5. Greece, Italy, Portugal, Rumania, Republic of South Africa, Spain, Turkey, Yugoslavia,
6. Argentina, Brazil, Ecuador, Guatemala, Nicaragua, Mexico,
7. Egypt, Jordan, Iran, Morocco,
8. Ethiopia, Kenya, Malawi, Tanzania,
9. Burma, India, Indonesia, Malaysia, Pakistan, Philippines, Thailand, Sri Lanka.

One may notice that this is quite similar to the ten regions of Pestel and Mesarovic, which are essentially based on geographical considerations.

## III RELATION BETWEEN ENERGY/CAP AND GNP STRUCTURE FOR EACH GROUP OF HOMOGENEOUS COUNTRIES

We associate to each country an energy consumption per capita, which seems to be distributed on levels.

Let us take, for example, the Mediterranean countries (group 5) shown in Figure 14; we are looking for a family of curves which could correspond to these levels. There are numerous possibilities for such curves, especially because we have 7 points and 5 curves. But we will demonstrate that two curves cannot intersect and that they have a special convexity.

### Convexity of the curves

Suppose there exists a potential E, i.e. a function

$$E = \mathbb{R}^3 \to \mathbb{R}_+$$

$$(x, y, z) \to E(x, y, z)$$

where

$x$ = % agriculture
$y$ = % services
$z$ = % industries.

If E is differentiable, we can write:

$$dE = \frac{\delta E}{\delta x} dx + \frac{\delta E}{\delta y} dy + \frac{\delta E}{\delta z} dz \ . \qquad (1)$$

As

$$x + y + z = 1 \ , \qquad (2)$$

we have

$$dx + dy + dz = 0 \ . \qquad (3)$$

The partial derivative $\frac{\delta E}{\delta x_i}$ represents the variation of the energy consumption due to a small variation in structure. There are three partial derivatives:

$$\frac{\delta E}{\delta x}(x, y, z) \qquad \frac{\delta E}{\delta y}(x, y, z) \qquad \frac{\delta E}{\delta z}(x, y, z) \quad .$$

We shall examine their traces respectively for x = constant, y = constant or z = constant. We, therefore, have six functions:

$$\left.\frac{\delta E}{\delta x}\right|_{y = \lambda} \quad \left.\frac{\delta E}{\delta x}\right|_{z = \lambda} \quad \left.\frac{\delta E}{\delta y}\right|_{x = \lambda} \quad ,$$

$$\left.\frac{\delta E}{\delta y}\right|_{z = \lambda} \quad \left.\frac{\delta E}{\delta z}\right|_{x = \lambda} \quad \left.\frac{\delta E}{\delta z}\right|_{y = \lambda} \quad .$$

Now we shall make some economic assumptions for the developing and developed countries.

A. Economic assumptions for the developing countries

We rank the six functions according to the following economic hypotheses:

a) industry is much more energy intensive than services,
b) services are a little more energy intensive than agriculture.

Let us examine one partial derivative and its two associated functions:

$\frac{\delta E}{\delta x}$ denotes the variation of energy required by a variation of the agricultural share;

Suppose that the industry share is content ($z = \lambda$). When we substitute services for agriculture, as services are more energy intensive than agriculture, we can write

$$\frac{\delta E}{\delta x}(x, y, z = \lambda) < 0 \ .$$

If now $y = \lambda$, with analogous reasoning, we also have

$$\frac{\delta E}{\delta x}(x, y = \lambda, z) < 0 \ ,$$

and

$$\frac{\delta E}{\delta x}\ (x,\ y = \lambda,\ z) < \frac{\delta E}{\delta x}\ (x,\ y,\ z = \lambda) \quad .$$

Doing so for the two other partial derivatives, we have the following order:

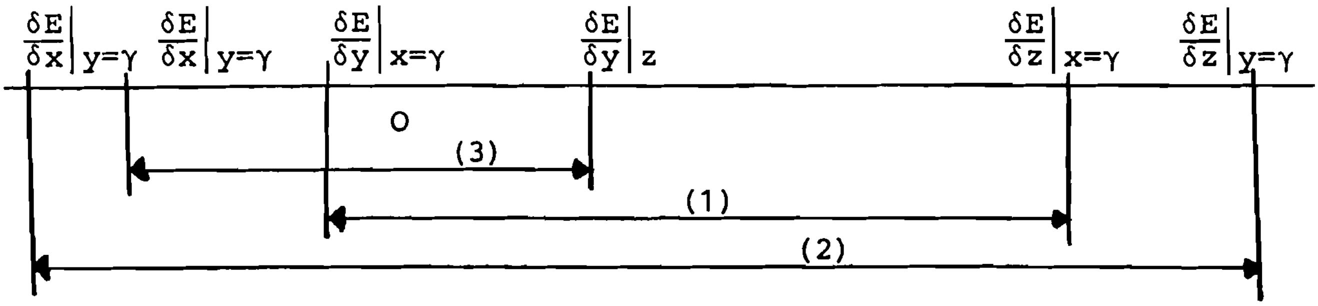



Figure 2. Ranking of partial derivatives of energy consumption to GNP structure for developing countries.

given

$$dE = \frac{\delta E}{\delta x}\, dx + \frac{\delta E}{\delta y}\, dy + \frac{\delta E}{\delta z}\, dz \quad . \qquad (1)$$

We shall examine for an additional share of energy ($dE > 0$), the three possible deplacements along each variable.

If we choose $x$ = constant, we arrive at $dx = 0$; so (1) becomes

$$dE = \frac{\delta E}{\delta y}\ (x{=}\gamma,\ y,\ z)\ dy + \frac{\delta E}{\delta z}\ (x{=}\gamma,\ y,\ z)\ dz \quad , \qquad (2)$$

and (3) yields $dy + dz = 0$.

So we can calculate

$$dy = \frac{dE}{\left.\frac{\delta E}{\delta y}\right|_{x=\gamma} - \left.\frac{\delta E}{\delta z}\right|_{x=\gamma}} \quad ,$$

and

$$dz = \frac{-\ dE}{\frac{\delta E}{\delta y}\ x{=}\gamma - \frac{\delta E}{\delta z}\ x{=}\gamma} \quad ,$$

according to the assumptions dy < 0 and dz > 0

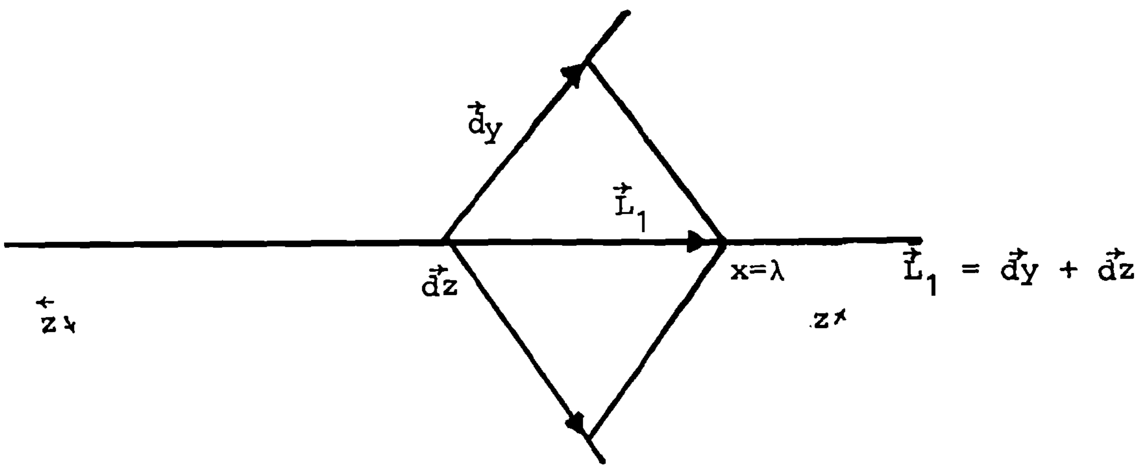



Figure 3. Vector representation of possible variation in services and industry shares (agriculture remaining constant).

if

$$\vec{L}_1 = \vec{d}y + \vec{d}z$$

$$\| \vec{L}_1 \| = \sqrt{3} \frac{dE}{\left.\frac{\delta E}{\delta y}\right|_{x=\lambda} - \left.\frac{\delta E}{\delta z}\right|_{x=\lambda}} ,$$

and

$\vec{L}_1$ is oriented towards z on the line x=λ.

We shall call $\vec{L}_2$ the deplacement along y=λ towards increasing z, and $\vec{L}_3$ the deplacement along z=λ oriented towards increasing y.

Using the same assumptions we have

$$\|\vec{L}_2\| \lesssim \|\vec{L}_1\| \ll \|\vec{L}_3\|$$

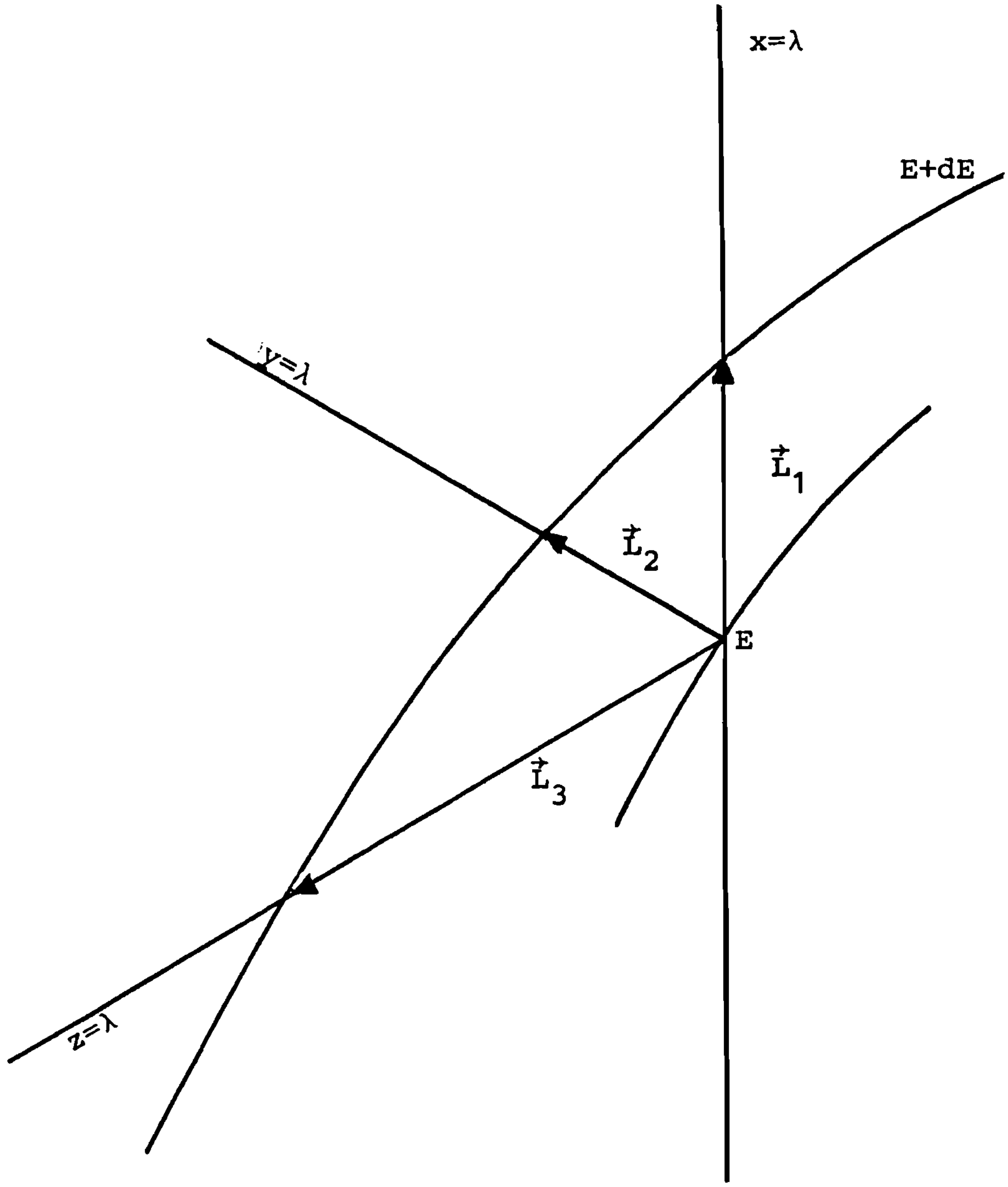



Figure 4. Convexity of iso energy consumption per capita and GNP structure in developing countries.

B. Economic assumptions for developed countries

The basic assumptions are:

a) industry is still more energy intensive than services;
b) but services become far more energy intensive than agriculture and get closer to the industry energy intensiveness.

These could be imaged by the deplacement of the values of the partial derivatives which become as follows:

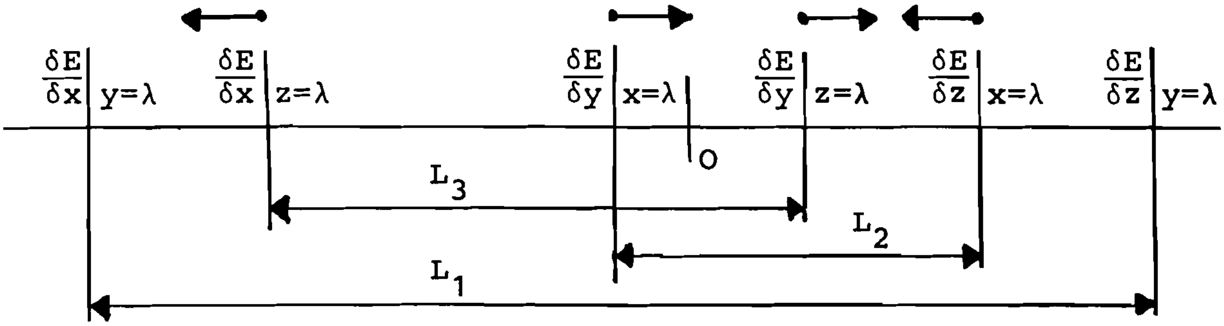



Figure 5. Ranking of partial derivation of energy consumption per capita to GNP structure for developed countries

Making the same calculation one can demonstrate that

$$\|\vec{L}_2\| \lesssim \|\vec{L}_3\| \ll \|\vec{L}_1\|$$

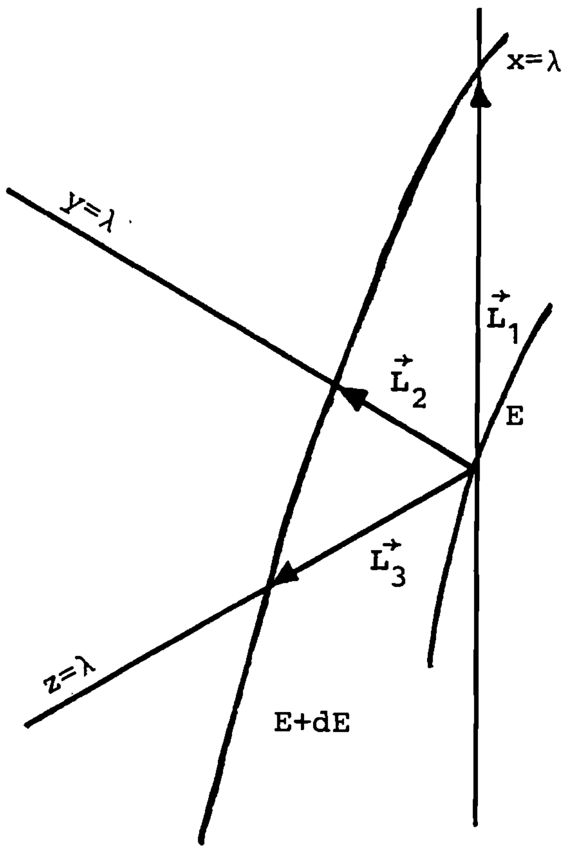



Figure 6. Convexity of iso energy consumption per capita and GNP structure in developed countries.

C. <u>The curves cannot intersect</u>

Suppose that two energy levels $E_1$, $E_2$ intersect. In a first step we choose $E_2=E_1+dE$ with $dE>0$

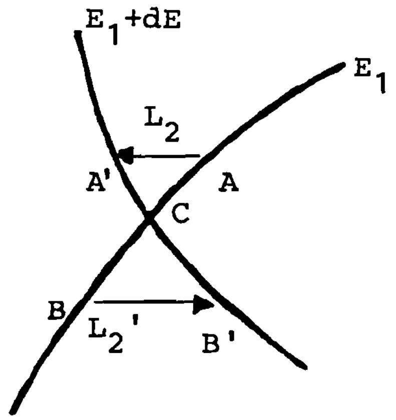



Figure 7. Intersection of two iso curves of per capita energy consumption and possible variation in GNP structure.

We showed earlier that $\vec{L}_2$ must be oriented towards increasing values of z.

This condition is satisfied when the points move from A to A' but is invalid when it goes from B to B'. Therefore, the only possibility is this:

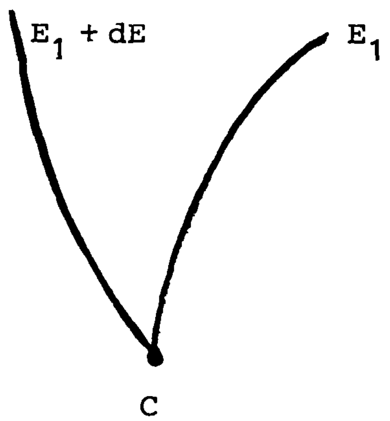



Figure 8. Two isocurves of per capita energy consumption intersecting at only one point.

but in point C as $dE > 0$ we proved before $\| L_2 \| \neq 0$ which is a contradiction. This result is extended to the full space by local continuous propagation.

Therefore, the energy potential can be represented by a family of iso-energy per capita curves which spreads out (Figure 10). The family of curves has been parametrized for each group of countries as shown in Figures 11-14[3].

For each group, one can observe that the energy potential doubles for equi-distance gaps.

## IV THE DYNAMICS OF DEVELOPMENT

Historical data have been obtained for France, U.K., Germany and Italy over the period 1789-1969. Figure 15 shows their development paths.

They all approximately start from the position where the developing countries are nowadays and are all now in the left group of the developed countries; one can notice that the individual paths fluctuate around a trend except for Italy, the path of which till 1914 looks more stochastic. But they got there by different speeds and at different points in time. In their order of arrival, there are U.K., Germany, France and Italy. The region to where all the countries try to go could be called, if we use the resilience theory, an attractor which stands between industry and services with a preference in time for industry rather than services.

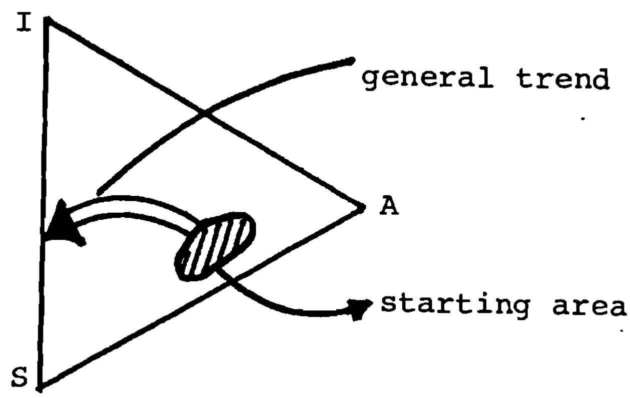



Figure 9. Long trend evolution in GNP structure. (cf. Figure 15 for details)

Let's examine for instance, the effect of the 1929 crisis on the developed countries. Both France and Italy reacted to the crisis in the same way; their attractors remained oriented towards services. The U.K. at that time was in the position where Italy is nowadays. This country also drops towards the

[3]In Figure 2 of the Appendix, the isocurves intersect because two statistical adjustments have been made for groups 4 and 5.

services attractor but with a deep slope. Germany, on the contrary, steps back to its early position it acquired in 1913. The direction of the regression path is opposed to industry; the crisis did not attach relatively more importance to services as in other countries. That means that the attractor is not services, and that the German system was not resilient and could not absorb the shock smoothly. In five years Germany went the same path as in the previous 16 years (1913-1929).

During the same period, France absorbed the shock but was affected longer. Italy, after 1933, went backwards and agriculture became its attractor. The after-war period has been characterized in the five countries by a great attraction towards the industry and after 1950-1955, the attractor changed to services. However, if there is a general trend to the services attractor, it can be shown in the figure that fluctuations occur between industry and services.

If we now look at the iso-energy consumption per capita curves on Figure 15, one can see that for all countries the historical energy consumption data we had* fit very well into our network.

---

*mainly postwar period. We can conjecture that the assumption remains valid for periods without great shock or deep structural modifications such as the periods: 1945-1973, 1929-1939, 1918-1929, etc.

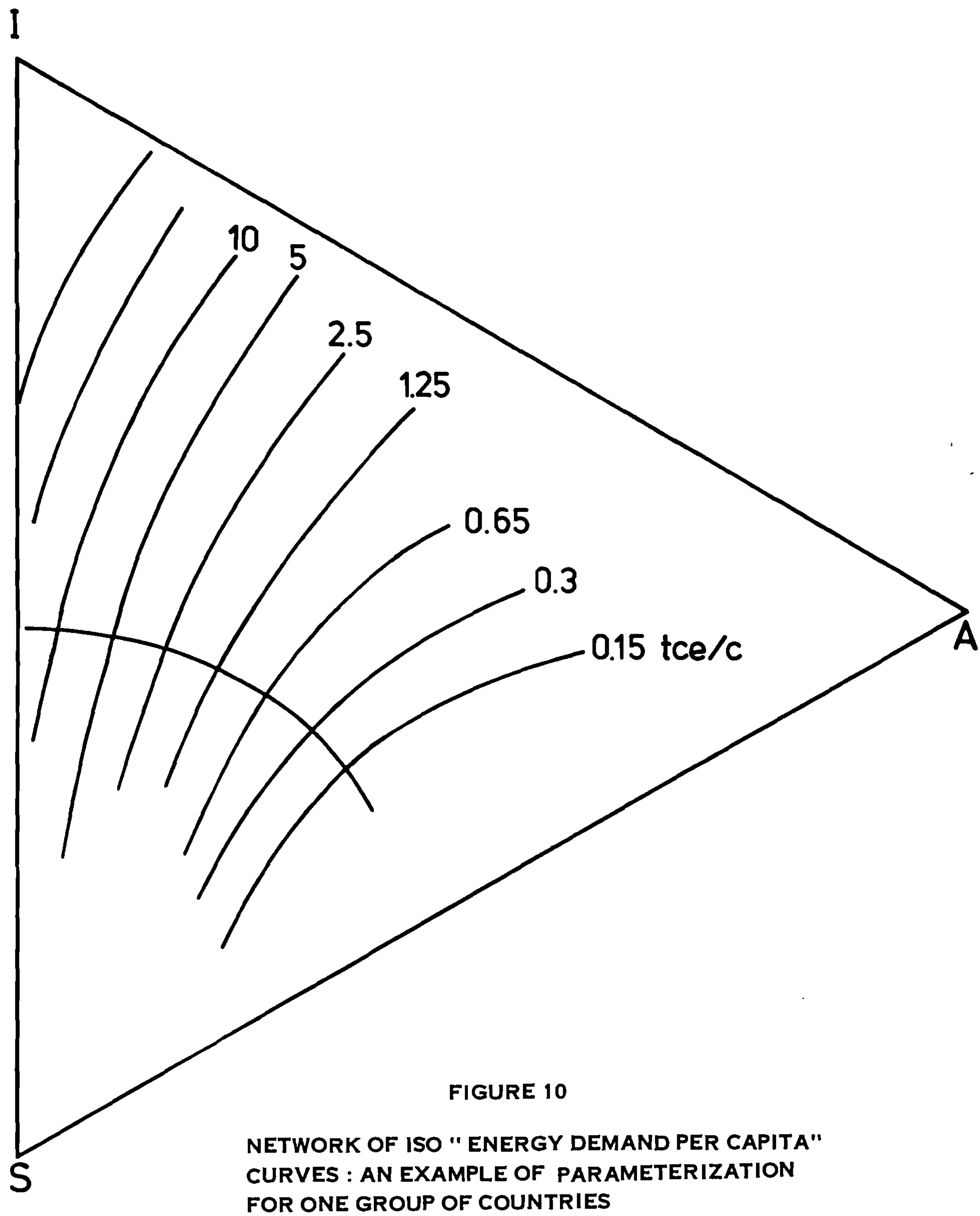



FIGURE 10

NETWORK OF ISO " ENERGY DEMAND PER CAPITA" CURVES : AN EXAMPLE OF PARAMETERIZATION FOR ONE GROUP OF COUNTRIES

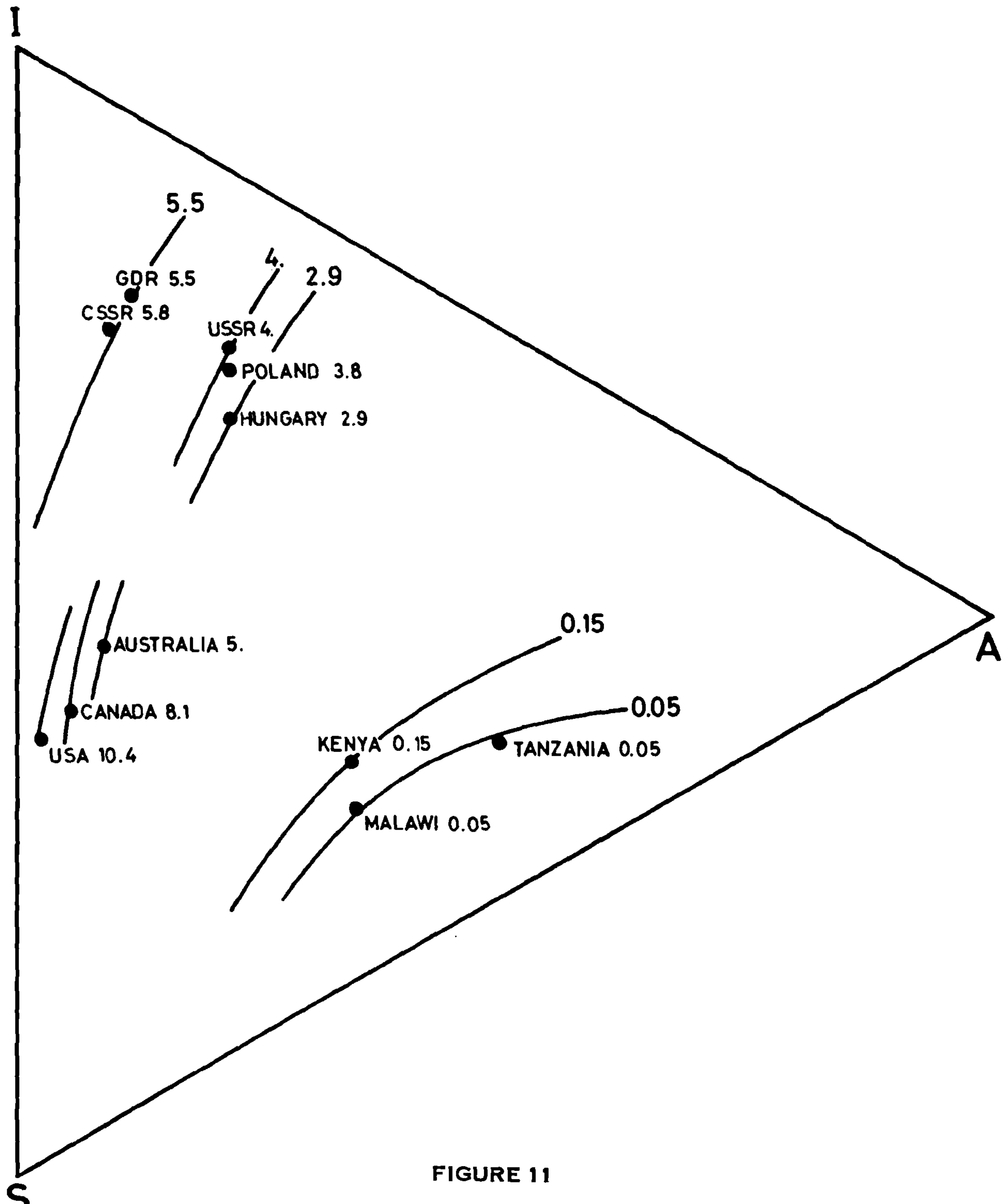



FIGURE 11

ISO CURVES OF PER CAPITA ENERGY CONSUMPTION FOR DIFFERENT GNP STRUCTURES AND FOR GROUPS OF COUNTRIES 1,2,8

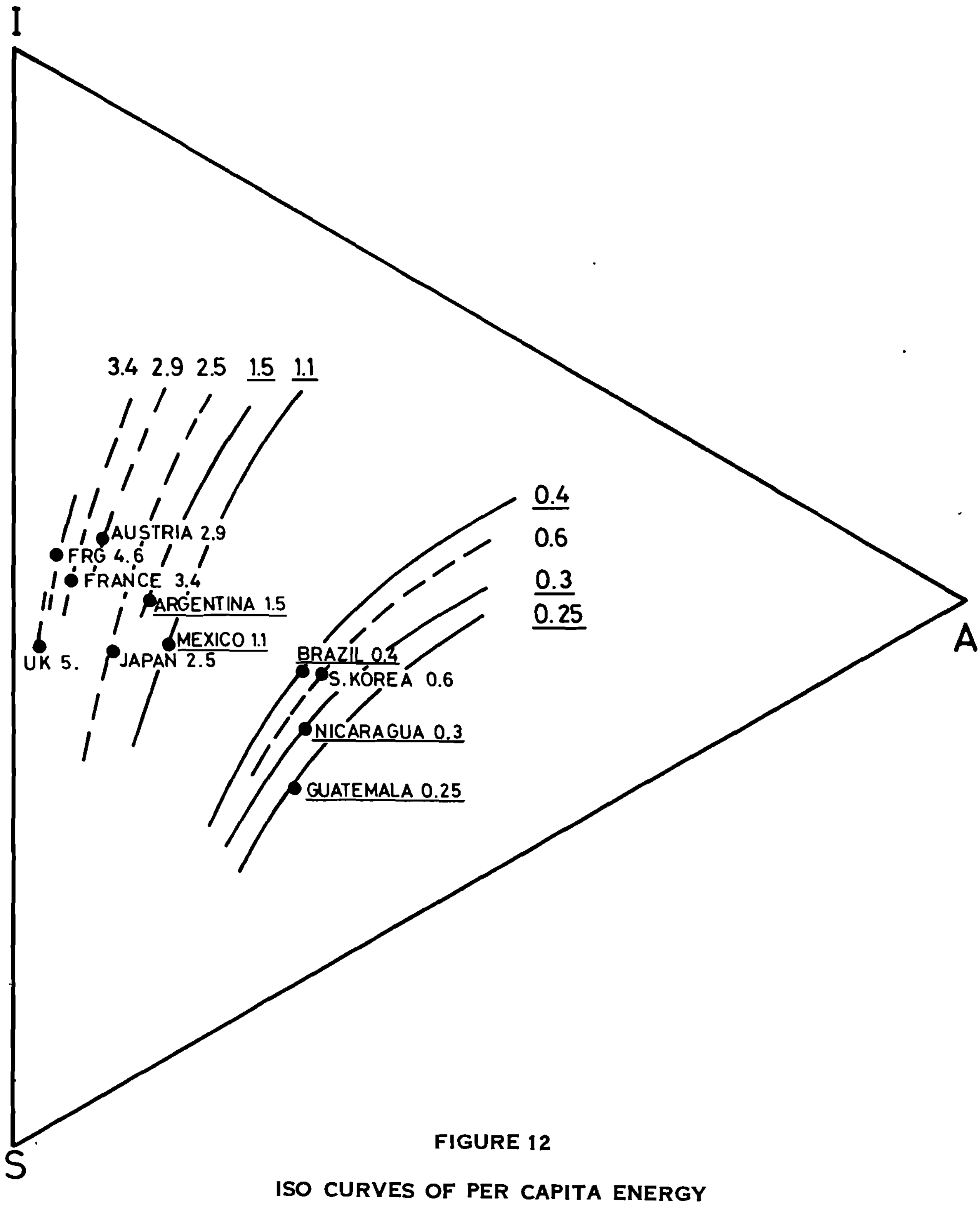



**FIGURE 12**

**ISO CURVES OF PER CAPITA ENERGY CONSUMPTION FOR DIFFERENT GNP STRUCTURES AND FOR GROUPS OF COUNTRIES 3 , 6**

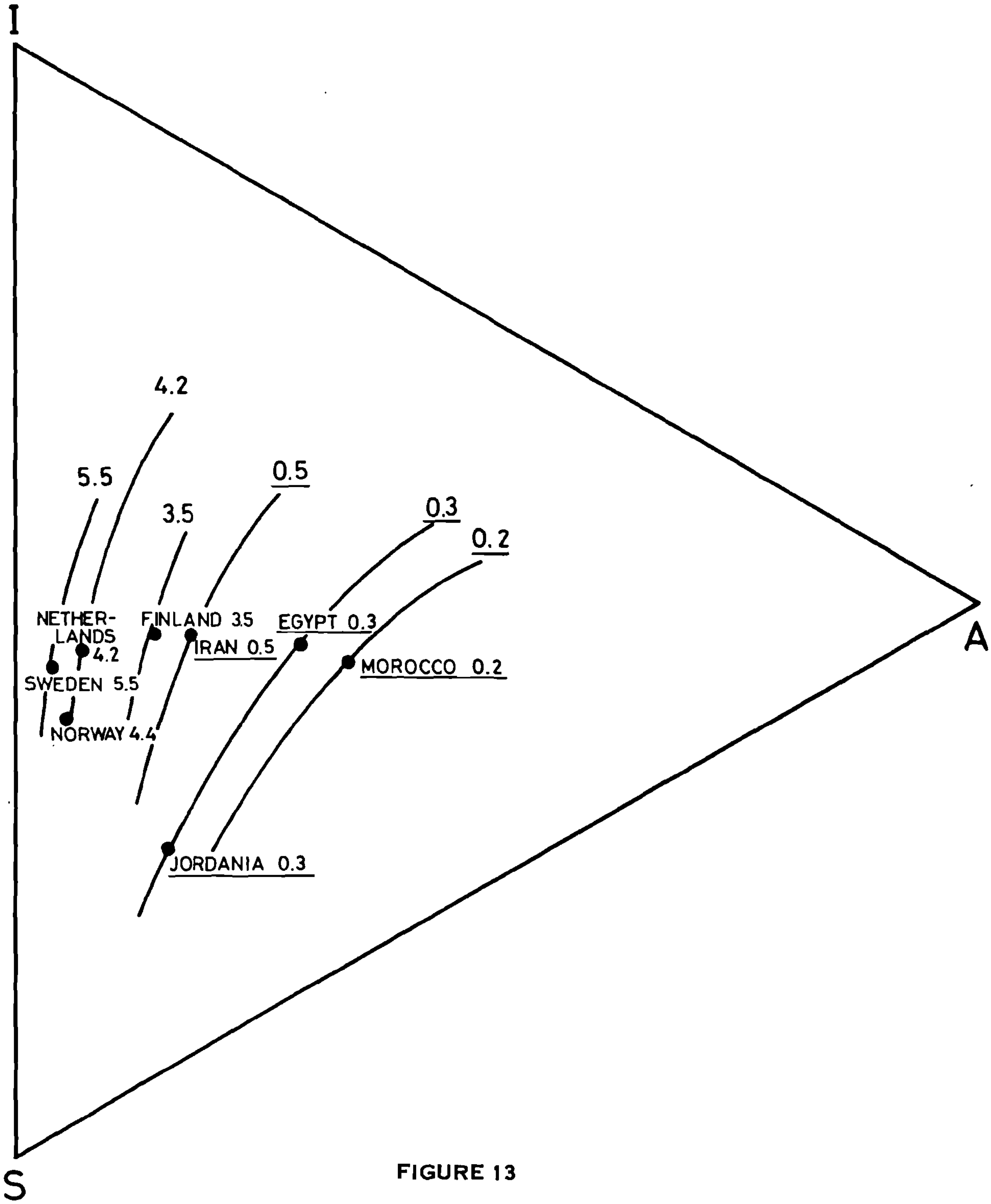



**FIGURE 13**

**ISO CURVES OF PER CAPITA ENERGY CONSUMPTION FOR DIFFERENT GNP STRUCTURES AND FOR GROUPS OF COUNTRIES 4,7**

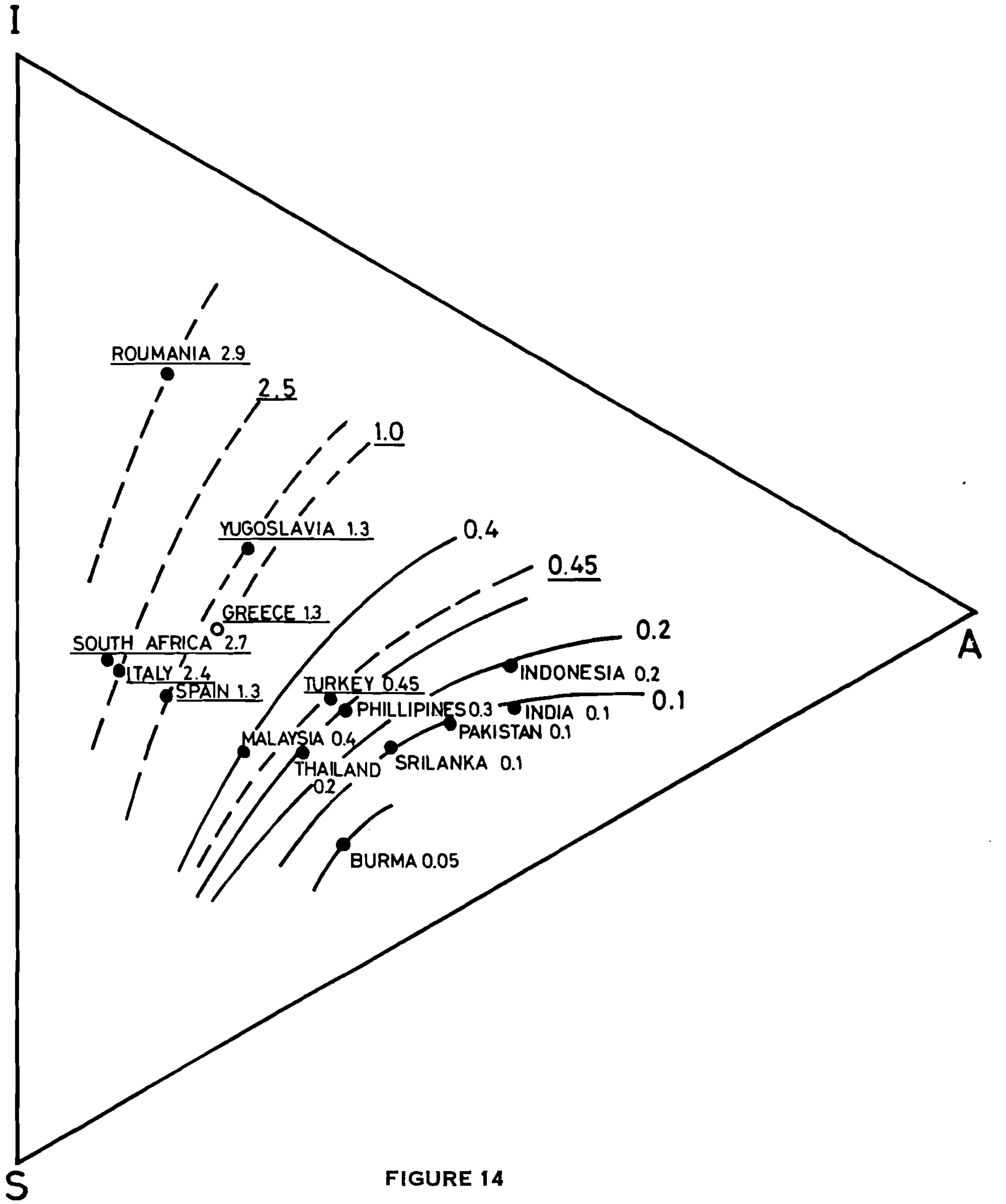
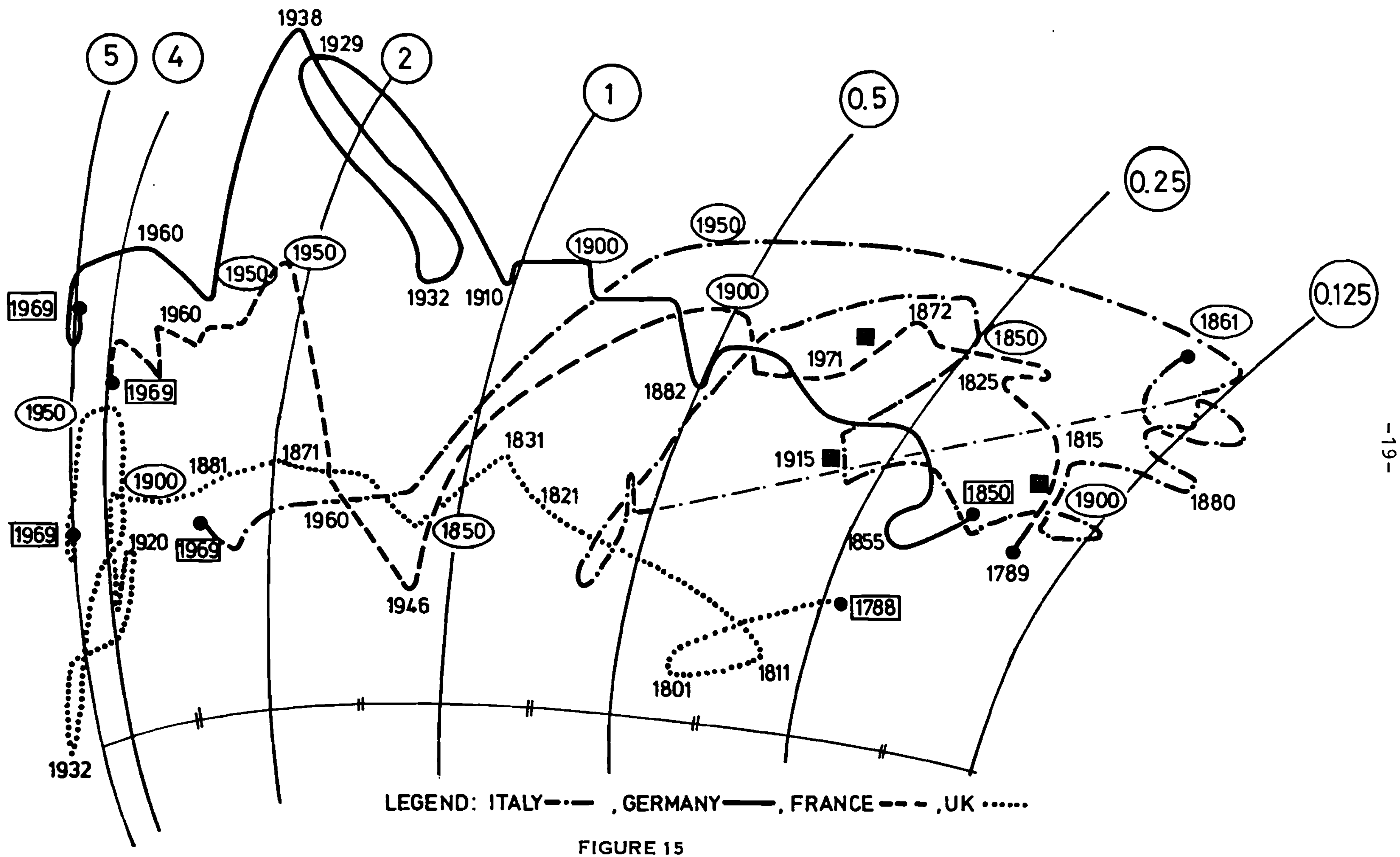



FIGURE 14

ISO CURVES OF PER CAPITA ENERGY CONSUMPTION FOR DIFFERENT GNP STRUCTURES AND FOR GROUPS OF COUNTRIES 5,9




LEGEND: ITALY —·— , GERMANY ——, FRANCE – – – , UK ······

FIGURE 15

EVOLUTION OF GNP STRUCTURE AND ENERGY POTENTIAL FOR UNITED KINGDOM, GERMANY, ITALY AND FRANCE DURING THE LAST TWO CENTURIES.

## APPENDIX

In this appendix (Memo of Ledolter to Häfele, Balinski and Beaujean, of September 13, 1976), a statistical analysis of country specific energy consumption data related to the GNP structure (share of agriculture, industry, services and transportation) is given. Second order models relating the explanatory variables (GNP structure) to the dependent variable (energy consumption) are shown to be adequate. Parameters in this model are estimated using observations on 47 countries.

Furthermore an interpretation of second order models is given and it is shown how they can be used in deriving iso energy consumption per capita curves.

### *Introduction*

Since industry, services and agriculture have different energy consumption patterns, it was pointed out by Beaujean and Chaix that the GNP structure might be more significant in determining the per capita energy consumption than its absolute level.

GNP is thus divided into its share corresponding to

(i) industry
(ii) services and transportation
(iii) agriculture.

In this appendix we investigate the relationship between country specific energy consumption and GNP structure. In the first part of the appendix we give an outline of the used data. The second part deals with statistical model building techniques (response surface analysis) and in the third section we apply these techniques to our data. Parameters are estimated and iso energy consumption curves are drawn.

### *1. Description of the data*

In our analysis we use observations on n = 47 countries, measuring GNP structure and per capita energy consumption. The data, together with grouping into social, cultural and climatic classes, is given in the Appendix.

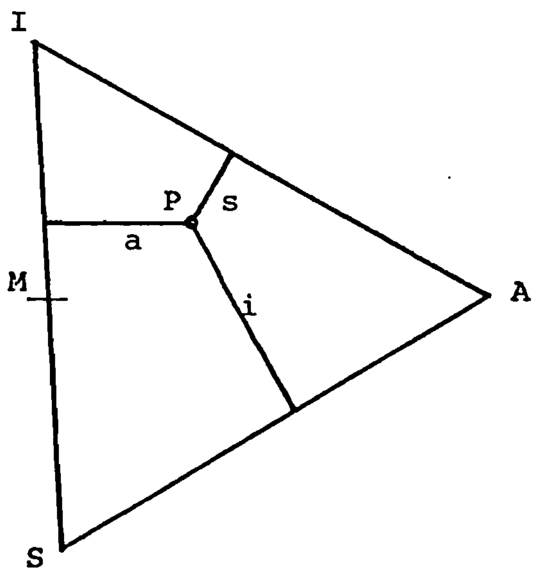
The following notation is used:

$A_j$ share of GNP in agriculture (of country j)
$I_j$ share of GNP in industry (of country j)
$S_j$ share of GNP in services (of country j)
$Y_j$ energy consumption (of country j).

The GNP structure of country j can be represented in different ways:

(i) In the three dimensional space as triple $(A_j, I_j, S_j)$. Since there are only two independent components (due to the restriction that $A_j + I_j + S_j = 1$ $(1 \leq j \leq n)$) the countries are restricted to the triangle whose summits are the endpoints of the unitary vectors.

(ii) In the two dimensional space in terms of three triangular coordinates defined by $(a,i,s) = \lambda(A,I,S)$ where $\lambda$ is given by $\sqrt{\frac{3}{2}}$




(iii) In the two dimensional space in terms of two independent coordinates.

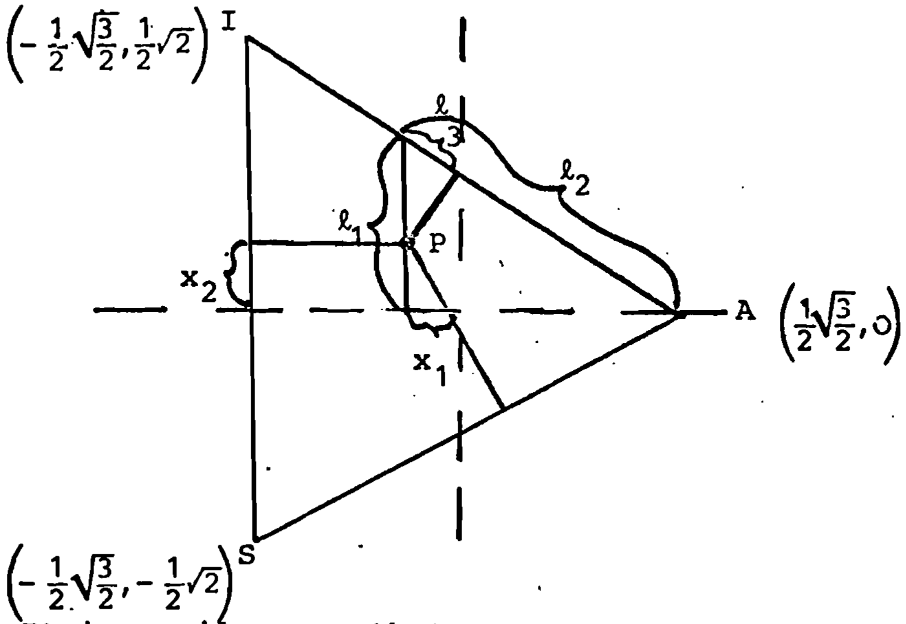
It is shown below that the GNP structure of country j can be described by

$$x_{1j} = \sqrt{\frac{3}{2}}\left[\frac{1}{2} - I_j - S_j\right]$$

and (1.1)

$$x_{2j} = \frac{1}{\sqrt{2}}\left[I_j - S_j\right]$$

<u>Proof</u>: The triangular coordinates of P are given by $\sqrt{\frac{3}{2}}$ (A,I,S)




(i) It is easily seen that

$$x_1 = -\frac{1}{2}\sqrt{\frac{3}{2}} + \sqrt{\frac{3}{2}}\,A = \sqrt{\frac{3}{2}}\left(\frac{1}{2} - I - S\right) \quad .$$

(ii) It can be seen that

$$\left(\sqrt{\frac{3}{2}} - \sqrt{\frac{3}{2}}\,A\right)^2 + \ell_1^2 = \ell_2^2 \quad .$$

Since $\cos 60^{\circ} = \frac{1}{2} = \frac{\ell_1}{\ell_2}$ it follows that

$$\left(\sqrt{\frac{3}{2}} - \sqrt{\frac{3}{2}}\,A\right)^2 + \ell_1^2 = 4\ell_1^2$$

$$\ell_1 = \frac{1}{\sqrt{2}}\,(1-A) \quad .$$

Furthermore

$$(\ell_1 - x_2)^2 = \left(\sqrt{\frac{3}{2}}\,S\right)^2 + \ell_3^2 \quad \text{and} \quad \ell_3 = \frac{\ell_1 - x_2}{2} \quad .$$

Thus

$$(\ell_1 - x_2)^2 = \left(\sqrt{\frac{3}{2}}\,S\right)^2 + \frac{(\ell_1 - x_2)^2}{4}$$

$$\frac{3}{4}\,(\ell_1 - x_2)^2 = \left(\sqrt{\frac{3}{2}}\,S\right)^2$$

$$\ell_1 - x_2 = \sqrt{2}\,S \quad .$$

Substituting $\ell_1$ results in

$$x_2 = \frac{1}{\sqrt{2}}\,(1 - A) - 2\,S = \frac{1}{\sqrt{2}}\,(I - S) \quad . \qquad \text{q.e.d.}$$

## *2. Comments on Statistical Model Building*

In the following section we study the relationship between a set of independent variables $x_1$, $x_2$, ..., $x_k$ and a dependent variable y. We are interested in describing the response function (response surface)

$$\eta_j = f(x_{1j}, x_{2j}, \ldots, x_{kj}) \quad , \qquad (2.1)$$

relating the levels $x_{1j}, x_{2j}, \ldots, x_{kj}$ to its response $\eta_j$. Although a certain amount of prior knowledge as to the nature of the response surface may be available from physical or economic theories, the exact form of the surface will often be unkown. In such cases an exact determination of the response surface is usually impossible for the following reasons:

(i) there is generally an error involved in the measurement of the true response $\eta$. This error is commonly called "sampling error".

(ii) There may be an error in the measurement of the independent variables.

(iii) The exact form of the true response function may be extremely complicated.

Inspite of all these above mentioned difficulties one may be able to find some simplified representation of the response surface, one which would approximate key characteristics of the true surface over a limited region of the space spanned by the independent variables (region of interest R).

A great number of functions can be represented quite closely over a limited region R by some type of polynomial. This comes from the fact that if the true response function is continuous and has continuous derivatives over the region R, then it can be approximated to any degree of accuracy by a finite number of terms of its Taylor series expansion (which of course are polynomials) about some point in R. This approximation would usually involve many terms if we wished to represent the response over a large region of R. In practice, however, we are often concerned with the behavior of the response function over a relatively small region. In such cases it is usually possible to obtain good approximations to the true function by means of a relatively simple polynomial, perhaps one involving just linear (first order) or linear and quadratic (second order) terms.

We hope that the above discussion provides some basis for an approach which attempts to approximate the true response function by a polynomial in the independent variables $(x_1,x_2,\ldots,x_k)$. Polynomials have the added advantage that they are fairly easy to work with due to well known procedures to fit polynomials to data and to analyze the properly fitted polynomials.

We thus suppose to represent the true response surface

$$\eta = f(x_1,x_2,\ldots,x_k)$$

by a polynomial of degree m in the variables $(x_1,x_2,\ldots,x_k)$. Denoting this polynomial by $P_m(x_1,x_2,\ldots,x_k)$, we can write the observed response for the jth observation as

$$y_j = [y_j - \eta_j] + [\eta_j - P_m(x_{1j},\ldots,x_{kj})] + P_m(x_1,x_2,\ldots,x_k). \quad (2.2)$$

(i) $y_j-\eta_j$ represents the difference between observed and true response. This is the sampling error as referred to above. Usually this error is a composite of many small errors; it arises due to factors beyond control and is thus assumed completely random.

(ii) The term $\eta_j - P_m(x_{1j},\ldots,x_{kj})$ represents the difference between the true response function and the polynomial which was chosen to represent it at the point $(x_{1j},\ldots, x_{kj})$. This discrepancy is called lack of fit which may result from the fact that $P_m(x_1,\ldots,x_k)$ is still only an approximation to the true response function which may actually be more complex (e.g., of higher order than m).

Writing down the model we combine the discrepancy due to sampling error and lack of fit into a single error term denoted by $\varepsilon_j$.

$$y_j = P_m(x_{1j},\ldots,x_{kj}) + \varepsilon_j \quad . \quad (2.3)$$

Dropping the subscript j provides the general form

$$y = P_m(x_1,\ldots,x_k) + \varepsilon \quad . \tag{2.4}$$

By making assumptions about the error terms we specify the nature of error randomness. A common initial assumption is that $\varepsilon$ is a random variable with mean zero and constant variance $\sigma^2$; furthermore it is usually assumed that the errors are independent and Normally distributed. Two important kinds of deviations from these assumptions which can occur are the serial dependence of the errors (especially when the observations are ordered in time) and non Normality of the distribution. These situations are discussed in detail in the statistical literature, but are not investigated further at this point.

Statistical model building is necessarily iterative. The original model will often have to be modified as new information about the response function is derived. In the absence of prior information about the response function, the model builder will start out by entertaining a relatively simple model. If the model, however, does not appear consistent with the data (e.g., residual analysis indicates lack of fit), it has to be revised until the data under study seems to confirm the model. Even then, one cannot say that the model is the correct one; one can only say that the data which were investigated have not offered evidence that the model is false.

Models which are using parameters parsimoniously and which have been shown to provide adequate approximations to many common response functions are the first and second order models.

(i) First order model

$$y = P_1(x_1,\ldots,x_k) + \varepsilon = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon \tag{2.5}$$

A model which includes only linear terms in the variables $x_1$, $\ldots,x_k$ is called a first order model. The unknown parameters are estimated from the data so as to give the best fit of the model to the data (best fit in terms of least squares fit). It can be shown that the least square estimates of $\beta_i$, let's call them $b_i$, are minimum variance linear unbiased if the errors $\varepsilon$ are independently distributed with mean zero and constant variance. For further details of least squares estimation see Draper and Smith[4].

[4] Draper, N.R. and Smith, H., "Applied Regression Analysis", Wiley, New York, 1966.

The function of the $x_j$ obtained by replacing the $\beta_i$ in the model by their estimates $b_i$ and disregarding the error term defines the <u>fitted surface</u>

$$\hat{y} = \hat{y}(x_1,\ldots,x_k) = b_0 + b_1x_1 + \ldots + b_kx_k \quad . \qquad (2.6)$$

The residuals are the deviations of the observed and fitted response

$$e_j = y_j - \hat{y}(x_{1j},\ldots,x_{kj}) \quad . \qquad (2.7)$$

In cases where our initial model will not be adequate enough to account for the variation in the data, one might investigate a more complex model, perhaps including quadratic terms in the $x_i$.

(ii) <u>Second order model</u>

$$y = P_2(x_1,\ldots,x_k) + \varepsilon$$

$$= \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum\sum_{i<\ell} \beta_{i\ell} x_i x_\ell + \varepsilon . \qquad (2.8)$$

In total we thus need (k+1)(k+2)/2 parameters to describe the model. The additional second order terms in the model provide considerable flexibility for graduating surfaces. Again, the method of least squares can be used to provide estimates of the coefficients and the fitted equation is given by

$$\hat{y} = \hat{y}(x_1,\ldots,x_k)$$

$$= b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{i=1}^{k} b_{ii} x_i^2 + \sum\sum_{i<\ell} b_{i\ell} x_i x_\ell \quad . \qquad (2.9)$$

"Canonical analysis" enables us to reduce the above equation to an alternative form which can be readily interpreted.

The method of canonical analysis consists of

(a) moving the origin of the measured variables $(x_1,\ldots,x_k) = (0,\ldots,0)$ to the center of the contour system represented by the fitted equation and

(b) rotating the coordinate axes until they coincide with major axes of the contour system.

Then the fitted contours can be expressed as

$$\hat{y} - \hat{y}_c = \lambda_1 x_1^{*2} + \ldots + \lambda_k x_k^{*2} \quad , \qquad (2.10)$$

where the new coordinates $x_i^*$ are expressible as linear combinations of $(x_1, \ldots, x_k)$ and a constant. $\hat{y}_c$ is the fitted response at the center of the contour system. The sizes and the signs of the $\lambda$ can be examined and main features of the fitted surface can be readily understood.

To illustrate this more clearly we consider the case of k = 2 independent variables: Equation (2.10) can represent several types of surfaces (such as elliptical contours, stationary ridge, rising ridge, saddle situations which arise from hyperbolic curves). Which of these types of contours will arise depends on the values of the b's.

For the second order model it can be shown that the center of the new coordinate system is given by

$$x_{1c} = (b_2 b_{12} - 2b_1 b_{22})/4(b_{11}b_{22} - \tfrac{1}{4} b_{12}^2)$$

$$x_{2c} = (b_1 b_{12} - 2b_2 b_{11})/4(b_{11}b_{22} - \tfrac{1}{4} b_{12}^2)$$

and

$$\hat{y}_c = b_0 + \tfrac{1}{2}(b_1 x_{1c} + b_2 x_{2c})$$

$$\lambda_1, \lambda_2 = \tfrac{1}{2}\left[b_{11} + b_{22} \pm \sqrt{(b_{11} - b_{22})^2 + b_{12}^2}\right]$$

the surface then takes the form

$$\hat{y} - \hat{y}_c = \lambda_1 x_1^{*2} + \lambda_2 x_2^{*2}$$

where

$$x_1^* = \frac{1}{\sqrt{(b_{11} - \lambda_1)^2 + b_{12}^2/4}}\left[-\tfrac{1}{2} b_{12}(x_1 - x_{1c}) + (b_{11} - \lambda_1)(x_2 - x_{2c})\right]$$

$$x_2^* = \frac{1}{\sqrt{(b_{11} - \lambda_2)^2 + b_{12}^2/4}}\left[-\tfrac{1}{2} b_{12}(x_1 - x_{1c}) + (b_{11} - \lambda_2)(x_2 - x_{2c})\right].$$

For example if $\lambda_1 > 0$, $\lambda_2 > 0$ the contours (iso curves) are ellipses centered at $x_{1c}$, $x_{2c}$ with semi-axes $\lambda_1$, $\lambda_2$.

*3. Analysis of data*

In this section we report on the analysis of the data given in the Appendix.

$$y_j = P_m(x_{1j}, x_{2j}) + \varepsilon_j \qquad (3.1)$$

where $x_{1j}$ and $x_{2j}$ are measuring the position of country j in the triangle described in Section 1.

Several models were investigated. The details of the various regression runs are not reported here. It was found that the first order model (m = 1) showed significant contribution of the cross product term could be found. Furthermore the 9 chosen groups appeared different in their level and curvature in $x_1^2$. Dummy variables for different levels and curvature were included (for further discussion of the use of dummy variables see Draper and Smith[4]). Some of the estimated coefficients were not significant and could be dropped from the model.

A model which describes the data well (residual analysis could not detect serious inadequacy of the model; multiple correlation coefficient of .99) is given by

$$y = \beta_0^{(1)} z_1 + \beta_0^{(2)} z_2 + \beta_1 x_1 + \beta_2 x_2 + \sum_{i=1}^{9} \beta_{11}^{(i)} x_1^2 v_i + \beta_{22} x_2^2 + \varepsilon$$

where

$$z_1 = \begin{cases} 0 & \text{if country is from 2nd group} \\ 1 & \text{otherwise} \end{cases}$$

$$z_2 = \begin{cases} 1 & \text{if country is from 2nd group} \\ 0 & \text{otherwise} \end{cases}$$

$$v_i = \begin{cases} 1 & \text{if country is from ith group} \\ 0 & \text{otherwise.} \end{cases}$$

The unkonwn parameters are estimated by least squares and iso energy consumption curves are plotted for the different groups in Figures 16-20. Using canonical analysis these curves can be represented as ellipses with from group to group changing center and semi axes.

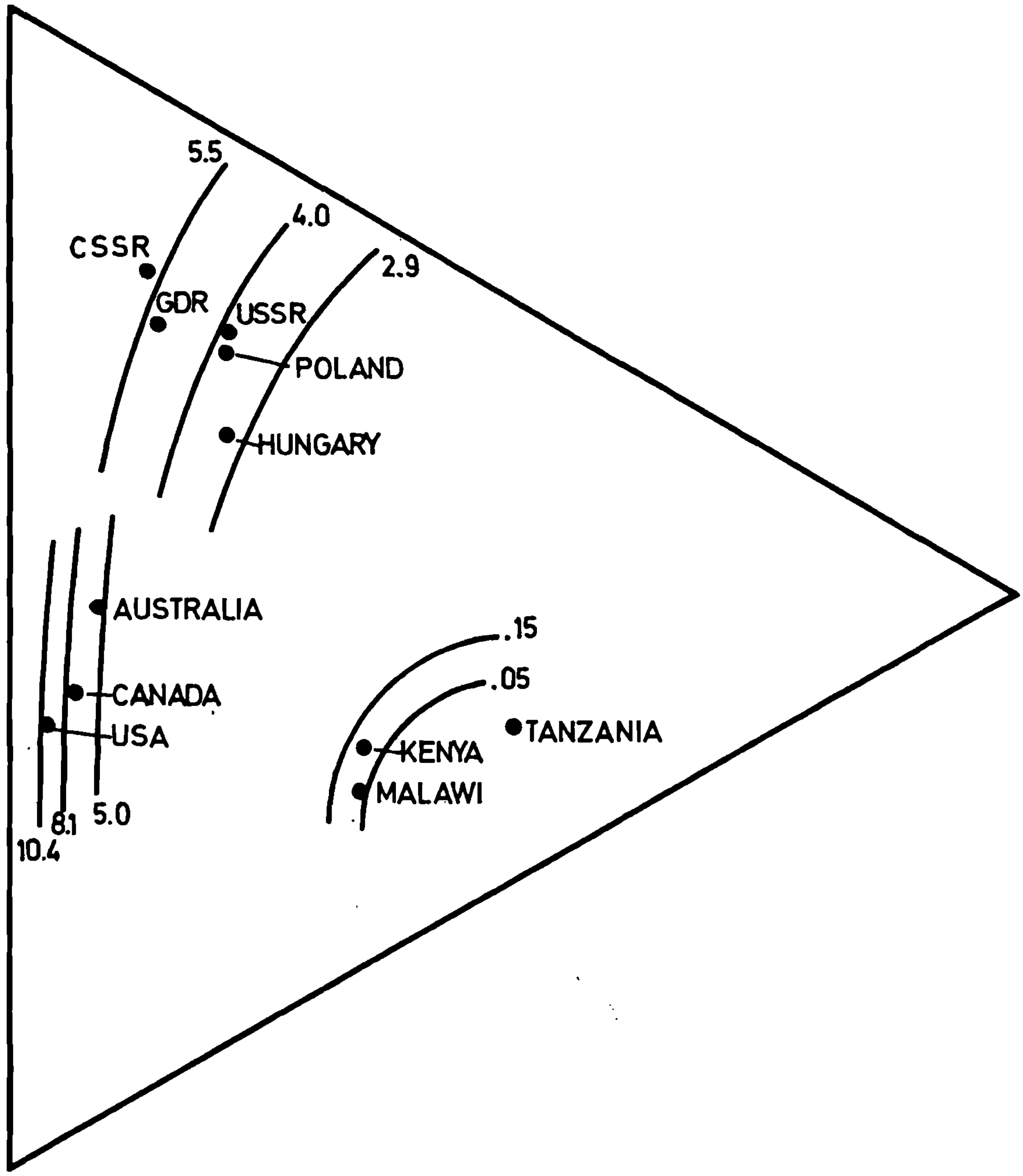



FIGURE 16: PARAMETERIZATION FOR GROUPS 1 , 2 , 3

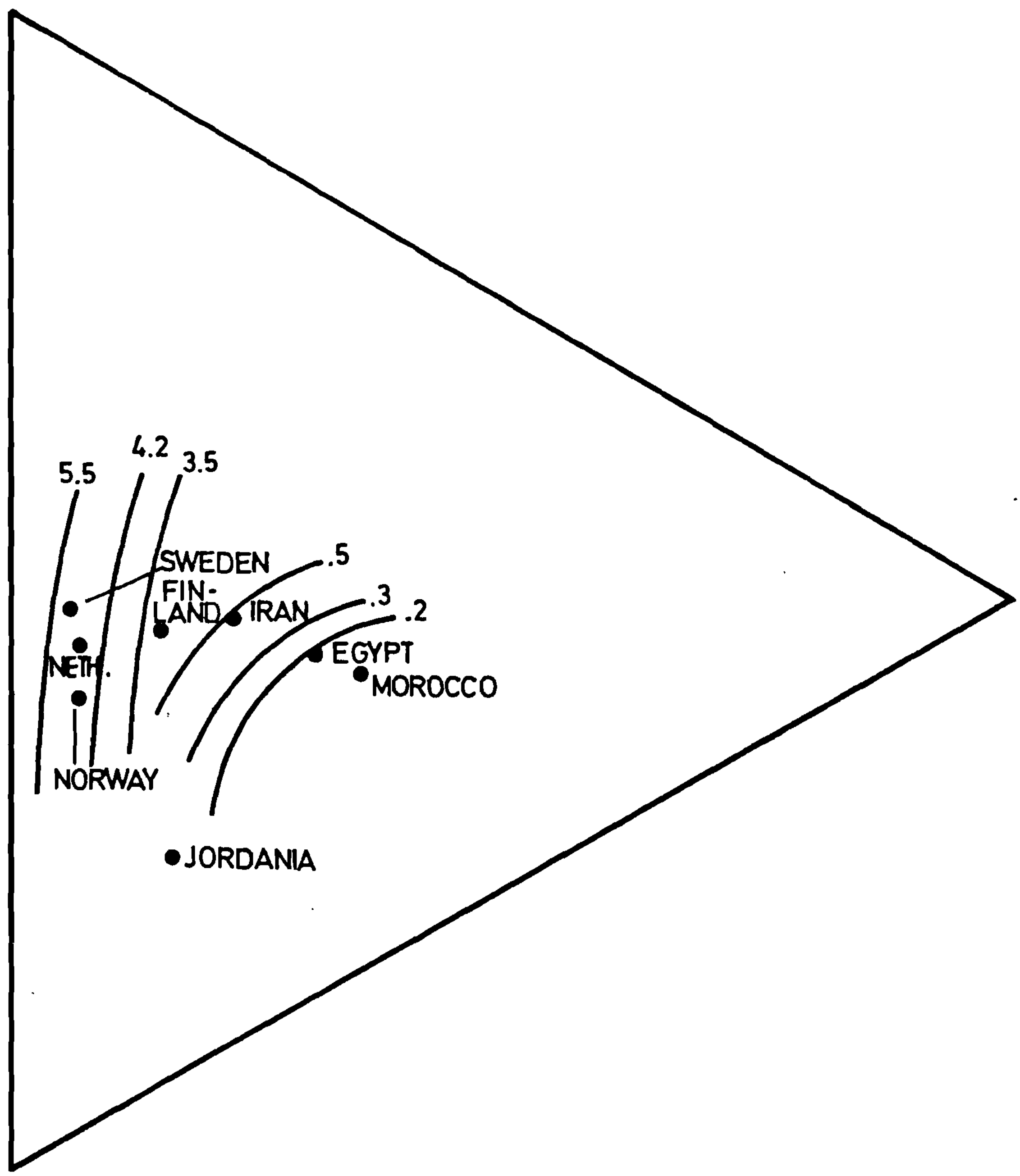



FIGURE 17: PARAMETERIZATION FOR GROUPS 4 AND 5

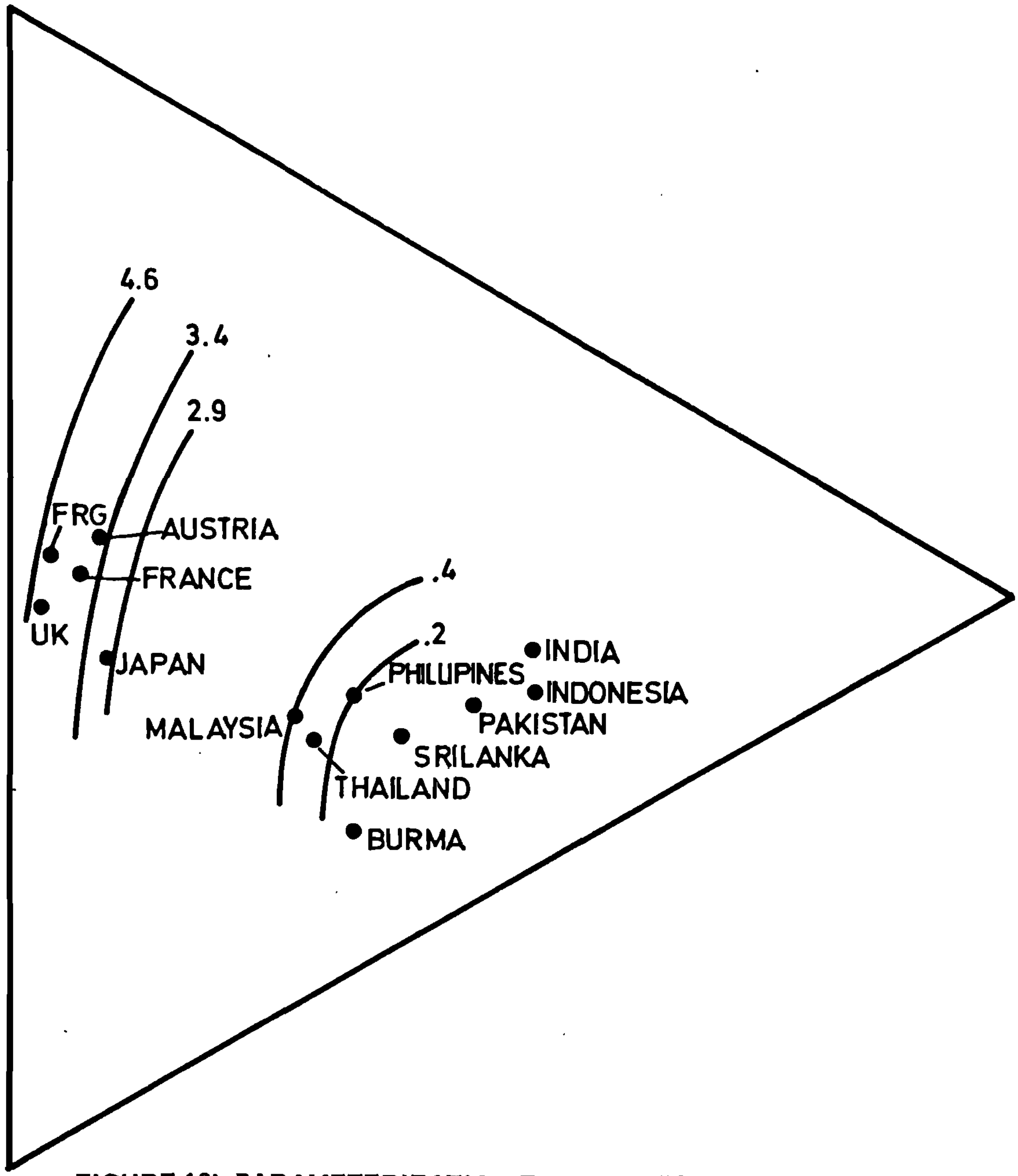



FIGURE 18: PARAMETERIZATION FOR GROUPS 6 AND 8

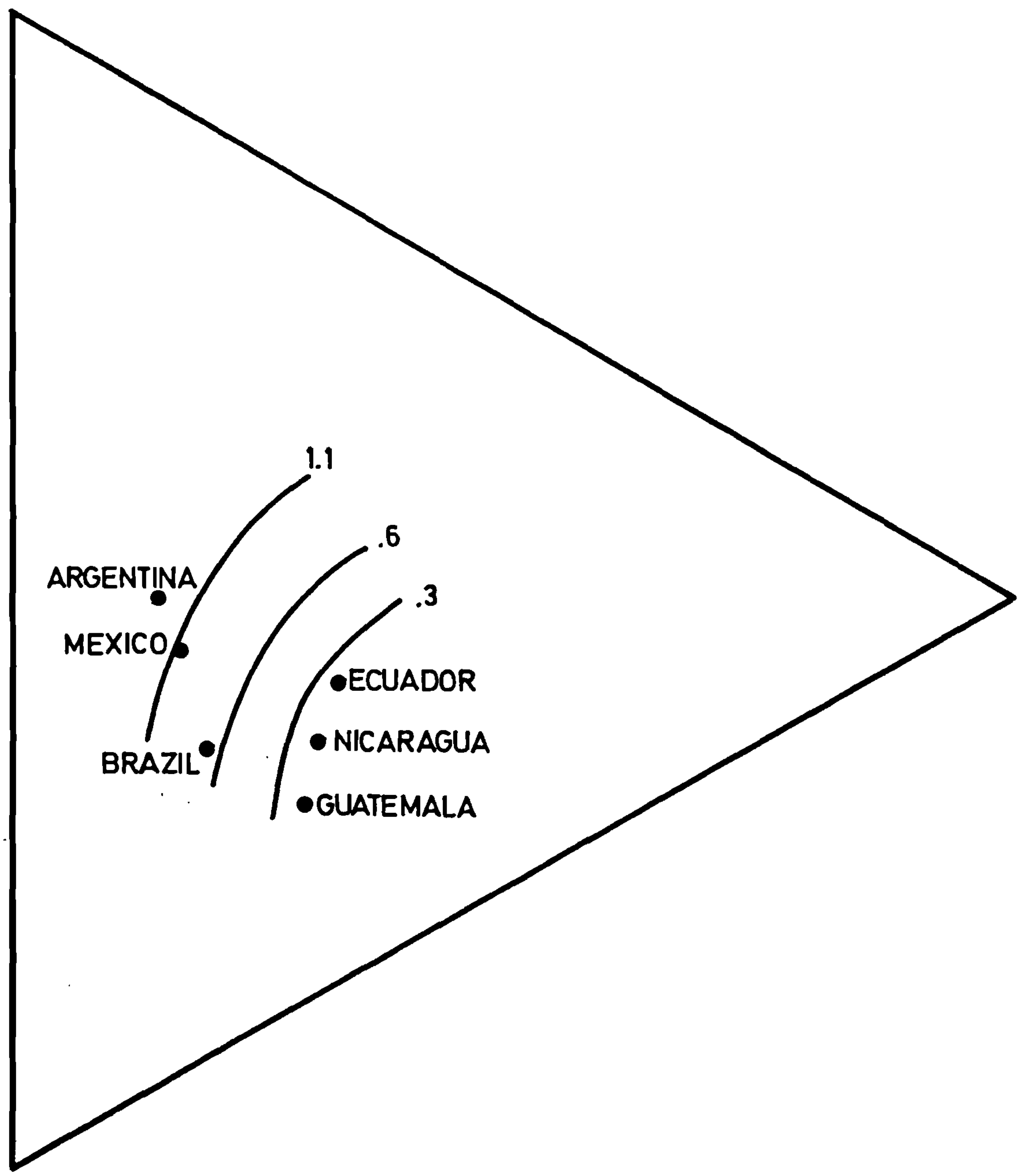



FIGURE 19: PARAMETERIZATION FOR GROUP 7

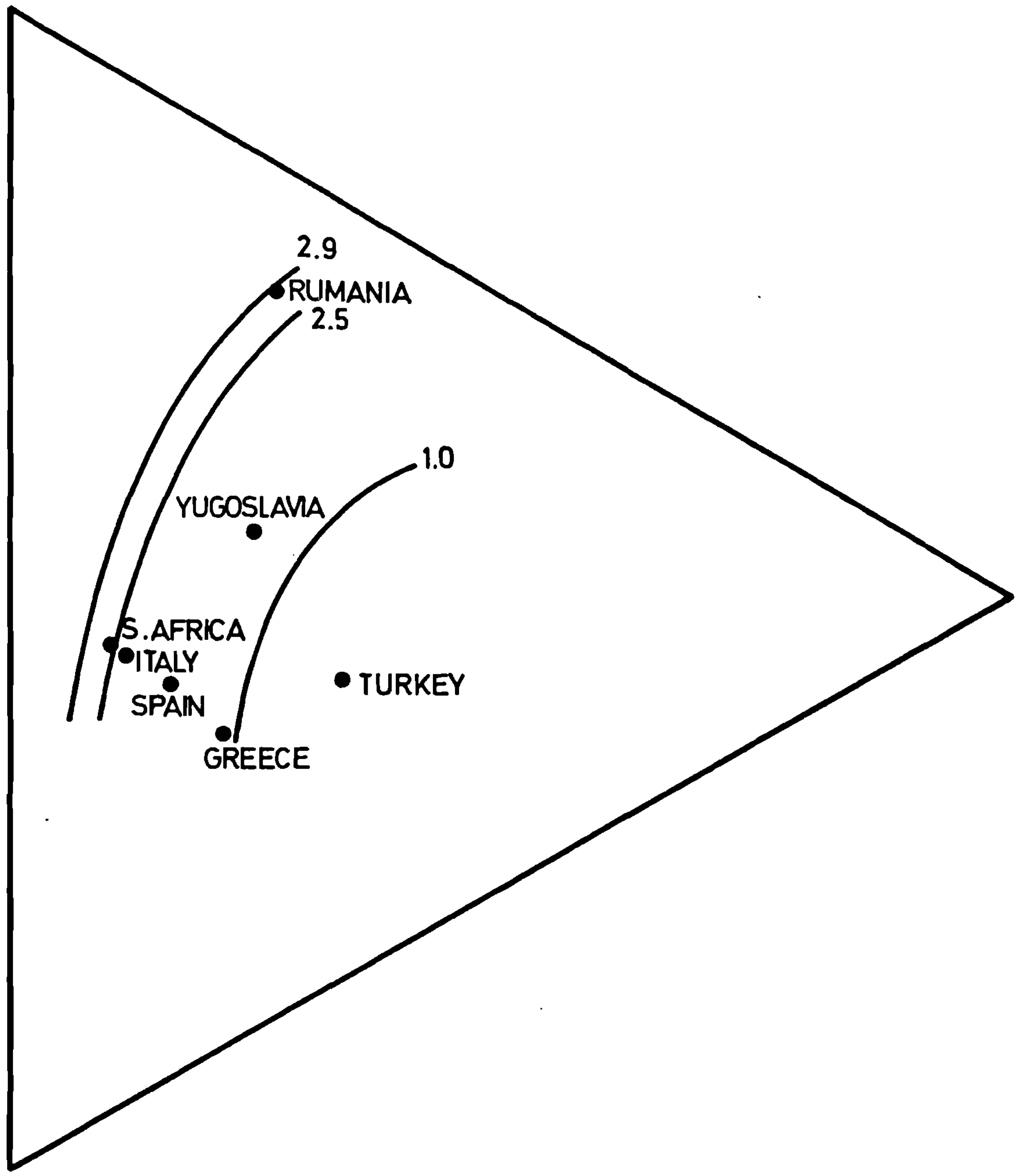



FIGURE 20: PARAMETERIZATION FOR GROUP 9

The following data were used in the analysis. The available 47 countries were divided into 9 groups according to their social, cultural and climatic position. This classification is not questioned in this paper.

| | Share of Agriculture | Industry | Services | Per Capita Energy Consumption |
|---|---|---|---|---|
| Group 1 | | | | |
| CSSR | .13 | .71 | .16 | 5.8 |
| GDR | .14 | .66 | .20 | 5.5 |
| Hungary | .21 | .53 | .26 | 2.9 |
| Poland | .21 | .60 | .19 | 3.8 |
| USSR | .21 | .62 | .17 | 4.0 |
| Group 2 | | | | |
| Australia | .09 | .42 | .49 | 5.0 |
| Canada | .06 | .38 | .56 | 8.1 |
| USA | .03 | .37 | .60 | 10.4 |
| Group 3 | | | | |
| Kenya | .35 | .19 | .46 | .15 |
| Malawi | .35 | .15 | .50 | .05 |
| Tanzania | .50 | .13 | .37 | .05 |
| Ethiopia | .58 | .15 | .27 | .25 |
| Group 4 | | | | |
| Netherlands | .07 | .42 | .51 | 4.2 |
| Sweden | .06 | .46 | .48 | 5.5 |
| Norway | .07 | .37 | .56 | 4.4 |
| Finland | .15 | .39 | .46 | 3.5 |

| | | | | |
|---|---|---|---|---|
| <u>Group 5</u> | | | | |
| Iran | .22 | .37 | .41 | .5 |
| Egypt | .30 | .30 | .40 | .3 |
| Jordania | .16 | .19 | .65 | .3 |
| Morocco | .35 | .26 | .39 | .2 |
| <u>Group 6</u> | | | | |
| UK | .03 | .47 | .50 | 5.0 |
| FRG | .04 | .51 | .45 | 4.6 |
| France | .07 | .48 | .45 | 3.4 |
| Austria | .09 | .50 | .41 | 2.9 |
| Japan | .10 | .39 | .51 | 2.5 |
| South Korea | .32 | .26 | .42 | .6 |
| <u>Group 7</u> | | | | |
| Argentina | .14 | .42 | .44 | 1.5 |
| Mexico | .16 | .37 | .47 | 1.1 |
| Brazil | .19 | .27 | .54 | .4 |
| Nicaragua | .30 | .22 | .48 | .3 |
| Guatemala | .29 | .17 | .54 | .25 |
| Ecuador | .32 | .26 | .42 | .25 |
| <u>Group 8</u> | | | | |
| Malaysia | .28 | .25 | .47 | .4 |
| Thailand | .30 | .22 | .48 | .2 |
| Philippines | .34 | .24 | .42 | .3 |
| Indonesia | .52 | .15 | .33 | .2 |
| Sri Lanka | .39 | .18 | .43 | .1 |
| Pakistan | .46 | .17 | .37 | .1 |
| India | .52 | .19 | .29 | .1 |
| Burma | .34 | .12 | .54 | .05 |
| <u>Group 9</u> | | | | |
| Turkey | .33 | .26 | .41 | .45 |
| Romania | .26 | .63 | .11 | 2.90 |
| S-Africa | .10 | .40 | .50 | 2.70 |
| Italy | .11 | .39 | .50 | 2.40 |
| Spain | .16 | .34 | .50 | 1.30 |
| Yugoslavia | .24 | .43 | .33 | 1.30 |
| Greece | .21 | .27 | .52 | 1.30 |