PROBLEMS

Congruence Among Spanish Economic Regions in the Nineteenth Century

Nicolás Sánchez-Albornoz
New York University

Because of the nature of our data, this essay relates to three different fields: geography, economics and history. The basic data of the problem are numerical values recorded in districts which encompass a territory, in a perfectly contiguous fashion. They refer not just to points but also to surfaces, and thus their geographic connotation is evident. Moreover these values correspond to two agricultural products of great economic importance, and they are presented as time series. Therefore the data have not only a geographic but also an economic and an historical dimension.

Two decades ago there began what Burton has called the quantitative revolution in geography. Ever since then attention in this discipline has been focused increasingly on the methods of elaborating large quantities of numerical information. The most appropriate geographic application of the new techniques appears to lie in the vast field of spatial analysis, whose main facets are spatial distribution, areal association and the classification or clustering of areas into regions.[1]

Some of these methods are applicable to our data. This is specially true of those which determine the degree of correspondence among the values recorded in areas or districts, as well as of those which later classify the results according to the mathematical affinities therein. Having taken these steps in relation to our problem, there appears a regionalization which is implicit in the original figures for each of our two sets of data. From the

[1] O. E. Duncan et al., *Statistical Geography. Problems in Analysing Areal Data*, Glencoe, 1961; D. S. Neft, *Statistical Analysis for Areal Distributions*, Regional Science Research Institute, Monography Series No. 3, Philadelphia, 1965; and B. J. L. Berry and D. F. Marble (eds.), *Spatial Analysis. A Reader in Statistical Geography*, Englewood Cliffs, 1968. It is precisely I. Burton's article *The Quantitative Revolution and Theoretical Geography*, which heads this last collection of essays.

corresponding analysis there emerge individual regional systems whose degree of interrelation it should be interesting to determine.

In geography the use of the quantitative method to determine the degree of association between areas — the correlation between the respective data — has already been properly established; so has the method of reducing to a limited number of categories the description of the various connections observed. This second operation — factor analysis — clusters the areas into a broader unit according to the intensity and direction of the established correlations. Finally a third procedure — factor matching — serves to identify the congruence among the units of two regional systems.[2] In the following sections of this essay we will show how the data concerning our topic have been subjected to these three types of statistical analysis.

Originally designed for use in psychology, factor analysis and factor matching were later adopted by the social sciences and by geography.[3] Their use is recent, and it has expanded owing as much to the greater ease and

---

[2] On the application of correlation and factor analysis in general to geography, see J. P. COLE and C. A. M. KING, *Quantitative Geography. Techniques and Theories in Geography*, London, 1968, chapters 6 and 7, and L. J. KING, *Statistical Analysis in Geography*, Englewood Cliffs, 1969, chapters 6 and 7. The formula for factor similarity or congruence, is developed by H. H. HARMAN in *Modern Factor Analysis*, second edition, Chicago, 1967, pp. 269-272.

[3] The first to cluster spatial data regionally by means of factor analysis were Kendall, Hagood and Price. See M. G. KENDALL, *The Geographical Distribution of Crop Productivity in England*, first published in 1939 and reissued in B. J. L. BERRY and D. F. MARBLE (eds.), *op. cit.*, pp. 387-406; M. J. HAGOOD, *An Examination of the Use of Factor Analysis in the Problem of Subregional Delineation*, « Rural Sociology », 6 (3), 1941, pp. 215-233, and *Statistical Methods for Delineation of Regions Applied to Data on Agriculture and Population*, « Social Forces », 21, 1943, pp. 287-297; and D. O. PRICE, *Factor Analysis in the Study of Metropolitan Centers*, « Social Forces », 20, 1941, pp. 449-455.

The advent years later of electronic computers gave rise to what B. J. L. BERRY has called « an innovation due to a technical impulse » (*Cities as Systems within Systems of Cities*, in J. FRIEDMANN and W. ALONSO (eds.), « Regional Development and Planning. A Reader », Cambridge (Mass.), 1964, p. 125). Multifactor analysis came into wide use owing to the speed and ease which the new tools offered in contrast to the cost and time required by the calculations when performed by purely mechanical means.

The main paths followed by geographers since then were either the international classification of levels of development or regional discrimination, even including problems of urban ecology. For an example of essays dealing with the first theme, see B. J. L. BERRY, *An Inductive Approach to the Regionalization of Economic Development*, in N. GINSBURG (ed.), « Essays in Geography and Economic Development », Chicago, 1960, pp. 78-107; by the same author, *Basic Patterns of Economic Development*, in N. GINSBURG, « Atlas of Economic Development », Chicago, 1961, pp. 110-119; and L. F. SCHNORE, *The Statistical Measurement of Urbanization and Economic Development*, « Land Economics », 37 (3), 1961, pp. 229-245. On the second theme see B. J. L. BERRY, *A Method for Deriving Multi-Factor Uniform Regions*, « Przeglad Geograficzny », 33 (2), 1961, pp. 263-283; C. A. MOSER and W. SCOTT, *British Towns. A Statistical Study of Their Social and Economic Differences*, Edimburgh, 1961; J. A. THOMPSON et al., *Toward a Geography of Economic Health: The Case of New York State*, in J. FRIEDMANN and W. ALONSO (eds.),

economy of means possible in performing the calculations as to requirements of a purely theoretical nature. In this respect history lags far behind geography, sociology and economics.[4]

I.

The two basic sets of data for this investigation are the annual price averages for wheat and barley recorded in forty-eight provinces in Spain from the harvest year (July to June) of 1856-57 to that of 1889-90, a thirty-four year period. Each set is presented as an information matrix laid along the width of a horizontal axis and along the length of a vertical one. The districts, which are the variables with which we operated, are arranged horizontally. Their order is alphabetical. Thus Avila precedes Badajoz and so forth. On the other hand, the annual observations for the districts are lined up from top to bottom along the vertical axis, in chronological order. Hence each column contains the time series of prices for one particular province. In turn each line reveals the distribution of wheat or barley prices for all of Spain during a specific year of the period in question. The horizontal axis expresses the geographic coordinate of the matrix, and the vertical one the temporal coordinate, while the datum itself, a market price, is unmistakeably economic in nature. The information matrices for wheat and barley prices take up forty-eight columns and thirty-four lines, and they have the same order and dimension.[5]

Each set was examined first separately by correlation and factor analysis, and then the results were subjected jointly to factor matching. We will now discuss the steps in this process.

First a second matrix is built by intercorrelating the price series in all the districts for a product — wheat or barley. The computation skips the

---

op. cit., pp. 187-206; M. MEGEE, *Economic Factors and Economic Regionalization in the United States*, « Geografiska Annaler », 47 B (2), 1965, pp. 125-137; B.C. JONES and W. W. GOLDSMITH, *Studies in Regional Development: A Factor Analysis Approach to Sub-Regional Definition in Chenango, Delaware and Otsego Counties*, Cornell University, Ithaca, 1965; and J.C. FISHER, *Yugoslavia. A Multinational State*, San Francisco, 1966. In the classic manual by W. ISARD, *Methods of Regional Analysis, An Introduction to Regional Science*, Cambridge (Mass.), 1960, we find in pages 293-305 and 642-643 a discussion of the application of factor analysis to regionalization.

[4] On the applicability of some of these statistical techniques to price history, see our essay *Determining Economic Regions from Time Series Data. A Factor Analysis of the 19th Century Spanish Wheat Prices*, New York University, Ibero-American Language and Area Center, Occasional Papers no. 1, 1973. A revised edition of this paper: *Las regiones económicas de España en el siglo XIX. Su determinación mediante análisis factorial de los precios del trigo*, « Revista de Occidente », 134, 1974, pp. 212-227.

[5] In our study *A formação de mercado nacional. Espanha e Itália*, « Anais de História » 5, 1973, pp. 55-84 (soon available also in Italian), the source is described in greater detail. The same remarks made there regarding the series for wheat are also valid for the series for barley.

annual observations, which are then contained in the correlation coefficient for each pair of variables. The temporal dimension is thus erased by the synthesis of the covariances that the new indices represent. Now the locational characteristics become more prominent. Indeed, the provinces come to occupy symmetrically the horizontal as well as the vertical axis of the intercorrelation matrix. The coefficients which indicate the degree of association between two districts are located at the point where the columns and the lines intersect. Since the matrix is square and symmetrical, they are set down twice, on both sides of the diagonal axis where the perfect correlations of the variables with themselves are registered. Hence the upper half of this matrix is identical to the lower one, so that we may simply consider only one of them. The dimensions of the intercorrelation matrices for both wheat and barley prices are forty-eight by forty eight, which is the number of provinces in Spain. The number of coefficients computed from each set of data is 1,128; the total for both products is 2,256.

There is little that we can easily deduce and elaborate from a table containing 1,128 measures of the degree of association. The sheer amount of information makes it difficult to interpret the numerical relationships between the provinces. But factor analysis serves precisely this purpose by expressing the connection succinctly and describing the covariance parsimoniously and numerically.

Factor analysis, a branch of multivariance analysis, extracts the numerical categories extant in the correlation between variables. Such parameters, usually called principal components or factors, are derived by means of matrix algebra calculations described in the appropriate manuals.[7] It should be pointed out that the term « factor » lacks the causal connotation which it

---

[6] The size of both matrices makes their publication difficult. Nonetheless we hope to make them available on another occasion. The number of coefficients was calculated according to the formula $N = \frac{K\,(K-1)}{2}$ where K represents the number of variables.

[7] Besides the early works on factor analysis by L. L. THURSTONE, *Multiple-Factor Analysis*, Chicago, 1961 (6th edition); R. B. CATHELL, *Factor Analysis. An Introduction and Manual for the Psychologist and Social Scientist*, New York, 1952, and D. N. LAWLEY and A. E. MAXWELL, *Factor Analysis as a Statistical Method*, London, 1963, see Harman's book, already mentioned, which is more recent and more comprehensive. On pages 4 and 5 he dwells on the descriptive nature and the parsimonious aim of this type of analysis. Throughout parts II and III he expounds the various solutions available and in chapter 14 he examines particularly the orthogonal varimax solution which we used in the computation of our data. It is important to make explicit what was the chosen solution, since, as Harman says (p. 4) « from a particular correlation matrix it is possible to extract factors in an infinite number of ways ». See also R. J. RUMMEL, *Applied Factor Analysis*, Evanston, 1970; S. A. MULAIK, *The Foundations of Factor Analysis*, New York, 1972; J. TORRENS-IBERN, *Modèles et méthodes de l'analyse factorielle*, Paris, 1972 and K. UBERLA, *Faktorenanalyse. Eine systematische Einführung für Psychologen, Mediziner und Wirtschafts- und Sozialwissenschafter*, Berlin, 1968.

may possess in another context. Factor analysis merely identifies and classifies hypothetical numerical components. What their patterns mean is left to the researcher. In the words of Thurstone, « it is really an inductive process demanding some familiarity with the subject matter ».[8] The equivocal nature of the term is due to the purposes which factor analysis was originally intended to serve. The method was developed by psychometricians in an effort to arrive at a definition of intelligence, with the hope that the operations of the mind would reveal the presence of substantive factors. Subsequent research has not borne out this supposition.

The results of factor analysis of wheat and barley prices in Spain are presented in turn as a third pair of matrices along whose horizontal axis lie those factors or components which have been extracted, while the provinces are now lined up along the vertical axis. New coefficients appear at the cells where the columns (factors) and the lines (districts) intersect. They tell us the degree to which the variance of the prices of the original variable, or province, is connected with each factor. Such a measure is not expressed as a percentage. To convert it precisely into such a relative expression of the total variance of the variable it is necessary to square the coefficients. The sum of these squares (including those portions corresponding to the errors or to the specific variations)[9] should then be 1. In order to visualize this more clearly it might be helpful to present the indices as a percentage by merely multiplying the squares by 100. Accordingly, the sum of the new values will be 100 instead of 1.

In presenting tables one and two we have followed the above criterion, that is, we offer there the squares of the coefficients of the factor matrices for wheat and barley prices all multiplied by 100. Both tables show the same disposition as the factor matrices. Since each one of these matrices included 240 coefficients (48 districts by 5 factors), the tables show the same number, which is nearly a fifth of all the indices in the intercorrelation matrix. The factor matrix, as well as the table derived from it, presents, in short, a simpler and more readily intelligible configuration. Not only are there less coefficients to be interpreted but they are also condensed according to mathematical criteria, so that there will be a narrower margin for subjectivity in the analysis of the correlation table.

The number of factors or components of the variance of the variables which have been disclosed is five. This is in line with the magnitude of the

---

[8] L. L. THURSTONE, *Problèmes actuels et méthodes nouvelles en analyse factorielle*, in Colloques Internationaux du Centre National de la Recherche Scientifique, « L'analyse factorielle et ses applications », Paris, 1955, p. 38.

[9] It should be noted that the sum of the partial variances related to the factors constitutes the common variance or communality, represented by the expression h^2. The communality differs from the total possible variance of a variable in that it ignores the specific variances, that is to say, those which are not correlated with any other variable and also the presumed errors of variance.

« characteristic roots » of the respective intercorrelation matrices.[10] Still, the chosen solution is only one among several, all of which are statistically valid and correct. The specific indices are the most relevant in terms of the way the problem was presented. However, were the matter to be approached differently, according to other hypotheses which would require, for instance, the clarification of only two factors, neither the coefficients nor their arrangement would remain the same.[11]

Because the values are thus relative to the parameters of the investigation, factor analysis has been criticized for the fact that, in spite of the complexity of the calculations, it does not identify components of general validity. It is argued that the method lacks the rigour of other procedures. Admittedly most studies of this kind retain a certain arbitrariness regarding several steps in the analysis, mainly in the selection of variables to be included in the information matrix, as well as in the choice of the solution and of the number of factors to be extracted. Finally, the interpretation of the factors themselves gives the investigator even more leeway.

However, in our investigation the condition of the data restricts the area of personal decision. As far as variables are concerned, the number retained is quite exhaustive. There has been no selection or condensation of any sort, since all of the nation's provinces have been included.[12] As to the attributes or observations they are extremely homogeneous: they have been reduced to

---

10 The number of factors to be extracted is given by the magnitude of the « characteristic roots » of the intercorrelations matrix (with units on the diagonal axis). Regarding the roots of the characteristic equation of the intercorrelation matrix, called « latent roots » or eigenvalues, see HARMAN, *op. cit.*, pp. 140-146. According to the criterion of several authors, expressed by COLE and KING, *op. cit.*, p. 297 (see also the definition on page 683), only those factors with eigenvalues over 1 should be considered. The matrix for wheat manifests five of these « latent roots » higher than 1. However, the matrix for barley has a sixth one which denotes the presence of an additional vector. Our reasons for having done without the component related to this vector are that, on the one hand, when extracting only five factors from the intercorrelation matrix of barley prices, the sixth is nonetheless properly glossed rather than omitted; on the other hand, for the purposes of matching the factor matrices of both agricultural products, the number of factors to be compared must perforce be the same. To exclude a weak factor for barley seemed more practical than to include a totally insignificant sixth factor for wheat. This is an example of the sort of arbitrary decision that the researcher must make when using this analytical technique.

11 Access to conventional programmes has determined the choice of methods and computation solutions. What was left up to us was the number of factors to be extracted. Without these programmes and without a computer it would not have been possible to perform the calculations required. The machine at our disposal was the CDC6600 of the Courant Institute of Mathematical Sciences of New York University. The Institute's library provided us with the FANOD 103 programme for the computation of the factor and intercorrelation matrices, and the FACMAT 53 for the matrix of the coefficients of convergence.

12 The only Spanish province omitted in this study are the Canary Islands, whose price series are not included in the official sources.

the same metric and monetary units, have the same number of digits (two whole ones and two decimals), and refer to only one type of operation (sale and purchase) and one product (wheat or barley). The variations among prices are due simply to differences and modifications of levels in time and space. In our data we have not lumped together disparate traits, as happens when such items as area, population and income are thrown into the same computation, items which moreover have often been compiled using different measures.

A more detailed analysis, starting with monthly district (*comarca*) prices rather than annual provincial ones would be welcome. However, greater detail would alter neither the spatial or temporal relevance of the data nor the coherence of the attributes. The total area would remain the same, even though the number of observation points would increase, and the time lapse would be merely broken up rather than lengthened. The metric and monetary units and the product and the transaction would likewise be unchanged. We may conclude that even though the coefficients would be smaller (since the more detailed the data, the lesser the correlation), the meaning and the direction of the correlations would not be altered.[13]

Since what distinguishes the variables is mainly their geographic location, this locational element is preserved when comparing the variation among them. The later factor analysis clustering of the variables tends in turn to condense a group of spatial relationships. Since contiguous zones have greater contact, it is normal that their condensation would result in units of contiguous surface.

Do the spatial units thus defined correspond to economic regions? Two conditions appear to support this contention. In the first place, the association among the provinces is not accidental or momentary but reveals rather a common behaviour through a long period of time. The cohesion is then due to traits which are fairly permanent. In the second place, the commercial character of the data — the price of an agricultural product in a particular market — casts an economic stamp on those entities, even though it represents a limited part of the entire economic system (since the data refer to only one of the system's elements, namely the sale and purchase of goods). A more satisfactory explanation would of course be found within units defined by means of several indicators of that nature. Nonetheless the goods under consideration are, on the one hand, goods of mass production and consumption; moreover, they are very significant for the type of economy still prevalent in nineteenth century Spain. On the other hand, the data refer not just to one but to two products. If the corographic projections derived from

---

13 Our forthcoming book *Materiales para la história económica de España. Los precios agrícolas durante la segunda mitad del siglo XIX*. Vol. I: *Trigo y cebada* (to be published by the Bank of Spain) will include analyses based on monthly price data. Since no district series ara available, no more detailed regional unity could be produced there. This is unfortunate because provinces may include zones socially and economically different.

the data for both products happen to coincide, then the probability is twice as great that the distribution reflects economic conditions.

The very purpose of factor matching is to measure the similarity among the components of two matrices of equal dimensions. Since we are dealing with factors which refer to regions, the greater the coefficient of congruence among the former, the greater also the correspondence among the areas.

The outcome of both factor and congruence analysis does not disclose why the districts group themselves in any particular way. As stated previously, the term « factor » does not imply causality but rather expresses the measure of a relationship. The congruence of the factors does not point out any causes either. Both of these statistical techniques perform a descriptive function. Nonetheless, is it possible to carry our investigation beyond mere description and to interpret the reasons for the particular groupings? This is where, as Thurstone said, familiarity with the subject matter becomes important. Prior knowledge of the economic history of the period suggests some possible explanations. Owing to the fact that some regions were mainly regions of production and others regions of consumption, each reacted in its own way to prices, production and consumption. Their differences in economic structure and behavior suggest the basis on which the relationship in each area presumably developed. In sum, statistical techniques do not offer an alternative to historical investigation but rather a rigorous framework for scientific research and explanation.

II.

The tables on which we will now comment were developed in the manner and with the elements just described. Thus the coefficients appear in cells where the line representing a province intersects the column corresponding to each factor. These indices express the proportion of the total variance of a district related to a factor. The average of each column shows that proportion of the total variance of the forty-eight provinces which corresponds to each factor. The sum of the partial variances for a line or district makes up a communality (h^2). The extent of the common variance for all the variables may be determined in two ways: by adding the partial figures on the bottom line of the table, or by averaging the sum of the communalities into the number of districts.

Table one contains the results of the analysis of barley prices. The five factors derived account for eighty-seven per cent of the variance of the forty-eight provincial series, certainly a very high proportion. The remainder does not exceed thirteen per cent. Examined individually, the first component alone contributes thirty-nine per cent to the general variance; the second, third and fourth provide about one seventh each — fifteen, thirteen and six-

teen per cent respectively. Finally the fifth component accounts for only four per cent. Factor one is way ahead of the other three, which are really of much less importance. The incidence of the last factor is extremely small.

In turn the communalities manifest another aspect of the close connection between the variance and the factors. Indeed, the five common factors account for a very large proportion of the total variance for each variable. Thus twenty-one provinces show communalities of more than ninety per cent: in other words, only a small portion of their variance is not subject to the influence of the five factors. In nineteen districts the value of the respective communalities is lower, from eighty to ninety per cent, and in seven others from seventy to eighty per cent. Only in one province, Vizcaya, do the five components account for less, in this case a little more than half of its variance.

Almost all the twenty-one provinces with high communalities show high indices in the first column. Thus these districts seem to exhaust their connections within the region represented by this column, or, by inverse logic, they only react occasionally to the conditions in the markets outside the region. The factor one group seems then to exert a strong cohesive influence. Let us take the example of Avila, where we find in the first factor eighty-one per cent of the variance of the prices paid throughout the thirty-four year period. Of the remaining nineteen per cent, five per cent is lost among specific unknown components and fourteen per cent is spread out among the other known factors. Therefore in the final distribution there is little that may be assigned to each of these four secondary components. The connections within the region itself are strong, while those between Avila and the other regions in Spain seem weak and ancillary.

The provinces with communalities from seventy to ninety per cent show higher indices in columns two, three and four. Consequently we may assume that the respective regions exhibit less cohesion than the first one.[14]

---

[14] The six-factor solution reveals a very different distribution from the one described. The proportion of the variance of the forty-eight provinces which is explained by the six factors comes to ninety per cent, a reduction of three points. Nonetheless the ditribution among the six is substantially altered. Twenty-eight per cent of the general variance, instead of thirty-nine, is related to the first component; by contrast, twenty-one per cent is related to the new factor, the sixth one. The second, third, fourth and fifth components lie between those two extremes. The respective proportions of the general variance are thirteen, ten, fourteen and four. Thus this solution takes off region six from region one; the former supremacy of region one is broken.

The change in perspective adds to the communality of many provinces specific variances which had been excluded by the five-factor solution. Indeed, in twenty-five provinces the communality surpassed the ninety per cent index, and in twenty-one the eighty per cent. Under these conditions only two districts register percentages under this second limit.

Hence the sixth region may be characterized as a subregion. The particular elements of its identity are fused to those of the more extensive region one when the focus is narrowed. Then we lose sight of some zones located on the edges of region six which

TABLE 1

BARLEY
PROPORTIONS OF THE VARIANCE OF THE PROVINCES
RELATED TO FACTORS

		1	2	3	4	5	h^2
1	Alava	39	17	16	1	12	85
2	Albacete	43	15	0	26	7	91
3	Alicante	13	7	9	49	7	85
4	Almería	2	6	2	72	7	89
5	Avila	81	2	9	2	2	95
6	Badajoz	66	0	3	20	3	90
7	Baleares	3	11	21	49	0	83
8	Barcelona	4	31	8	39	8	90
9	Burgos	61	7	18	2	7	95
10	Cáceres	71	0	6	12	1	89
11	Cádiz	44	0	0	36	1	81
12	Castellón	3	24	3	45	0	75
13	Ciudad Real	76	4	1	12	0	93
14	Córdoba	71	1	0	20	6	97
15	Coruña	20	0	59	4	1	84
16	Cuenca	62	17	1	6	5	90
17	Gerona	13	15	25	20	8	82
18	Granada	31	13	0	40	0	84
19	Guadalajara	68	17	3	4	1	93
20	Guipúzcoa	32	15	9	8	9	73
21	Huelva	39	0	0	45	7	92
22	Huesca	0	79	2	13	0	94
23	Jaén	66	5	1	9	0	81
24	León	67	1	25	0	1	94
25	Lérida	3	60	1	3	10	77
26	Logroño	33	35	11	4	9	91
27	Lugo	8	2	72	2	0	85
28	Madrid	79	9	0	2	1	91
29	Málaga	53	1	0	35	2	92
30	Murcia	7	11	2	65	3	88
31	Navarra	24	55	1	3	1	83
32	Orense	4	0	62	2	5	73
33	Oviedo	22	5	43	1	8	78
34	Palencia	61	6	23	1	2	93
35	Pontevedra	2	2	70	2	1	77
36	Salamanca	72	0	15	2	0	90
37	Santander	30	10	34	1	11	86
38	Segovia	81	5	7	2	2	97
39	Sevilla	56	0	0	29	10	96
40	Soria	54	19	5	1	9	89
41	Tarragona	0	29	17	34	0	81
42	Teruel	6	62	0	11	8	87
43	Toledo	78	7	1	7	0	93
44	Valencia	8	15	7	23	26	79
45	Valladolid	78	4	12	0	1	95
46	Vizcaya	21	6	14	3	9	53
47	Zamora	74	2	16	0	0	92
48	Zaragoza	8	76	0	5	2	92
General proportion of the variance		39	15	13	16	4	87

As an initial approach to the problem we shall, in an effort to simplify as much as possible, reduce the group of indices to two main categories: a high one including indices of fifty per cent and up, and a low one with indices from forty-nine down to twenty-five. With anything lower than this it becomes difficult to distinguish what is significant from what is merely fortuitous in the relationship expressed by the coefficient.

Both categories are projected onto map one, determining the factor to which the indices are related. The vertical lines represent factor one; the horizontals factor two; the leftward oblique ones factor three; the dots factor four; and the rightward oblique ones factor five. The extension of these signs varies according to the category: for indices over fifty per cent they cover the entire surface of the district; for indices of the next category they are incomplete.

In most of the provinces one component prevails over the rest. However, in eleven of them — almost one fourth of the total — there is a certain ambiguity. In these cases a second component usually arises, surpassing the twenty-five per cent lower limit. That is the case, for example, in León. It also happens that two factors exhibit equivalent coefficients in the same district. Thus forty-four per cent of Logroño's variance goes to factor one and almost as much — thirty-five per cent — to factor two. The same thing happens in Santander. This Cantabric province reacts almost identically to factors one and three (thirty and thirty-four per cent respectively). In both cases the signs of the two main factors have been placed in the appropriate districts.

Once we project the indices graphically according to the criteria already set forth, there soon appear continuous surfaces where, during the thirty-four years in question, barley prices fluctuated in a closely related fashion. These regions are, on the one hand, Andalusia and both the Castiles, clustered by factor one, and on the other Aragon (factor two), Galicia (factor three) and the Mediterranean littoral (factor four). The fifth factor is only apparent in one province, Valencia, with a rather low index.

In the factor one area there prevails a nucleus with strong inner cohesion. Its corresponding coefficients surpass the sixty-five per cent limit. The provinces included in this sort of focus are Avila, Badajoz, Cáceres, Ciudad Real, Córdoba, Guadalajara, Jaén, León, Madrid, Salamanca, Segovia, Toledo, Valladolid and Zamora. The list is long and embraces the territory between the Duero River and the upper valley of the Guadalquivir River, basically Extre-

---

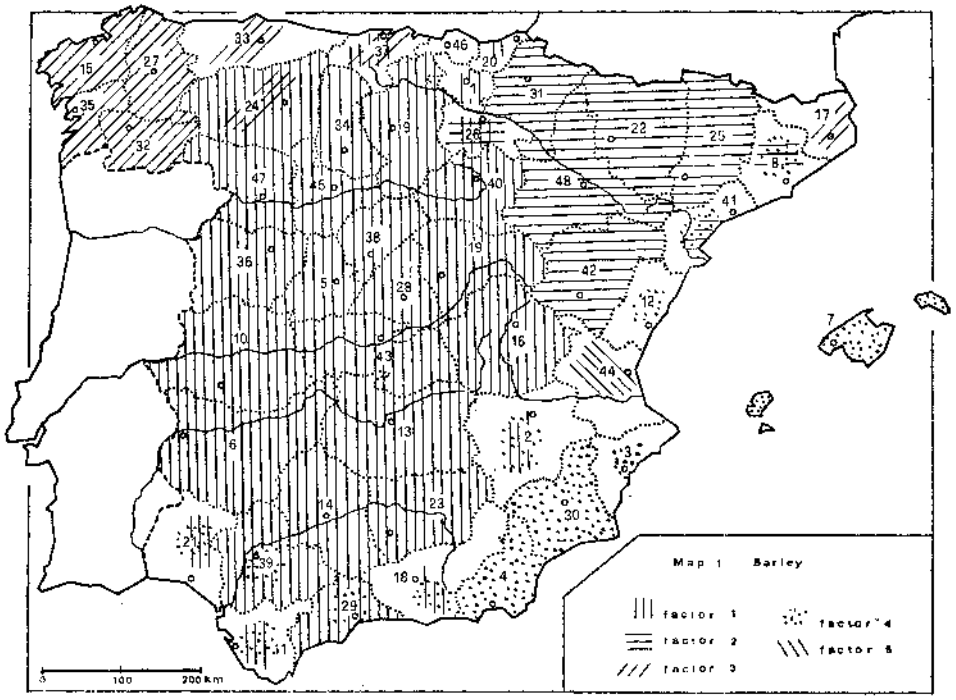
are connected to it but not very obviously to region one. The moment we force regions one and six to fuse, these zones are then left out, without any apparent connection, and the residue of the respective communalities increases. This happens in Vizcaya. In the six component solution the Basque province shows an eighty-three per cent communality; in the five component solution only fifty-three.

madura and New Castile, the southern portion of Old Castile plus Upper Andalusia. Physically, this zone stretches mainly along the central tableland.

Around this nucleus, the region includes likewise districts where the agglutinating factor decreases gradually in intensity. In contiguous zones it even becomes secondary, ranking of course behind a more important component. This happens, for example, towards the South, in Granada. where the index for factor one is thirty-one per cent, while the factor four coefficient comes to forty per cent. The gradual decrease continues still further. Even in areas with different characteristics such as Navarra or Alicante, the same factor one intrudes slightly.

Region two stretches from the Pyrenees to the Maestrazgo ranges (southern Aragón), with the Ebro valley in between. The area includes, besides the three Aragonese provinces (Huesca, Zaragoza and Teruel), Navarra and the Catalonian province of Lérida, so similar in many respects to nearby Aragón. Factor two is again present, in somewhat attenuated form, in the borderland areas of Tarragona and Castellón which open out on to the Mediterranean coastline. Moving up the Ebro it is also discernible in Logroño and Alava.

Factor three spreads out over the four Galician provinces. It penetrates Asturias (Oviedo) less deeply. Although gradually decreasing, it is also pre-



sent in the north and west of Old Castile: Santander, León, Zamora, Palencia and Burgos. Unexpectedly factor three reappears at the opposite end of the Peninsula, in the province of Gerona.

The center of factor four may be found in the Southeast (Murcia and Almería). Its influence stretches along the Mediterranean coast, from Barcelona to Cádiz, and through the Balearic Islands. It then goes up the Guadalquivir River to Seville, reaches the Atlantic coast of Huelva, and also enters the interior in Albacete.

The regions sketched out in this fashion coincide roughly with the historical ones. The contrast between the interior and the periphery, which has been the subject of so much discussion, is hereby reaffirmed. In its own way, the relationship between barley prices proves that this contrast is no mere intellectual fancy but rather a deeply rooted, undeniable reality. Thus we find once again the old theme of the tableland *vis à vis* the coast, although the littoral itself is not uniform, and includes several subdivisions.

In turn, table two shows the indices derived from the wheat prices. Here the five factors account for ninety-two (and a fraction) per cent of the total variance (in place of barley's eighty-seven per cent); with thirty-six and twenty-four per cent for the first and second factors respectively. The three remaining ones follow next, accounting in turn for thirteen, nine and eleven per cent. The first two components together represent sixty per cent of the total variance; the last three represent one third. Unlike the case of barley prices, there is no prime factor in wheat, nor does the fifth factor turn out to be insignificant.

As to the communalities, one third of the provinces appears to be subject to the five factors by more than ninety-five per cent, and another third by ninety to ninety-five per cent. None of them falls below the eighty-one per cent level. Error or specific variances have been reduced to a minimum.

If we juxtapose those provinces where the same factor prevails, several regions, whose cartographic projection has been carried out according to the guidelines for the previous case also become apparent. Some zones stand out clearly, such as Castile (factor one), Andalusia (factor two) and Galicia (factor three). Not so Aragón, which is now identified by factor five and which as a region does not have an identity as clearly defined as it had for barley.

Factor one covers both Castiles — the Old and the New — plus the Basque Provinces and Navarra. The same factor turns up, in a more muted fashion, in Aragón (Zaragoza), Northern Extremadura (Cáceres) and Upper Andalusia (Córdoba and Jaén). The nucleus of this bloc, with indices over sixty-five per cent, is composed of Avila, Segovia, Soria, Valladolid and León, in Old Castile, and of Madrid, Toledo, Guadalajara and Cuenca, in New Castile. This is the central area of the high Castilian tableland.

In the periphery of this nucleus an attenuated factor one predominates hedged in on the north, south and west. On the north it extends up to the coast of the Cantabric Sea, but the maritime provinces begin to show, as expected, connections with factor three, the characteristic one for the northwest of the peninsula (Galicia and Asturias). East of this northern strip (in Navarra and Logroño) we can also discern a subdued influence coming from the area where factor five prevails.

To the south and the west, Albacete and Ciudad Real on the one hand, and Zamora and Salamanca on the other, are linked secondarily with Andalusia (factor two). Reciprocally, prices in the northern provinces of the Andalusian region and in all of Extremadura were susceptible to the fluctuations in Castile. This occurred quite obviously in Jaén, Córdoba and Cáceres, but to some extent also in Badajoz and even as far as Seville and Granada. Thus between both centres, Castile and Andalusia, there is a strip of territory connected to one as well as to the other nucleus in proportion to the distance between each province and the center of the region.

Towards the Levant and the Northwest the gradation is shorter and the transition is more abrupt. The fluctuations of Castilian prices are still noticeable in Zaragoza and Teruel, but not in Alicante, Valencia and Huesca. On the western side they do not go beyond Orense and Oviedo.

Andalusia, as just mentioned, is the epicentre of factor two. More precisely it is found in Málaga and the lower valley of the Guadalquivir (Seville, Cádiz and part of Huelva). Here the indices exceed the sixty-five per cent level. The influence of this factor, represented by horizontal lines, later extends by sea through the Southeast (Granada, Almería and Murcia), moves on to the Levant (Alicante, the Balearic Islands and Castellón), and reaches as far north as the frontier of Gerona and Lérida with France. Throughout the front of the Mediterranean littoral wheat prices are related to those in Andalusia, even in Barcelona and Valencia.

Factor three is typical of Galicia (Lugo, Orense, Coruña and Pontevedra), as well as of Asturias (Oviedo). It spreads with secondary impact through the rest of the Cantabric coast. It is likewise visible west of the Castilian tableland (León, Zamora and Salamanca). The Catalonian coast and the Balearic Islands, at the other end of northern Spain, show a virtual relation with factor three, the same for wheat as for barley.

Valencia is the only province where factor four prevails, and even here not very emphatically. This component affects price fluctuations in Teruel and, to a lesser degree, in the rest of Aragón (Zaragoza and Huesca). To the south it spreads out towards Alicante, Albacete and Murcia, but specially towards Almería. The Balearic Islands do not escape the Valencian influence either.

Finally, the centre of factor five may be found in the lower valley of the Ebro, completely dominating the province of Tarragona (sixty-one per cent

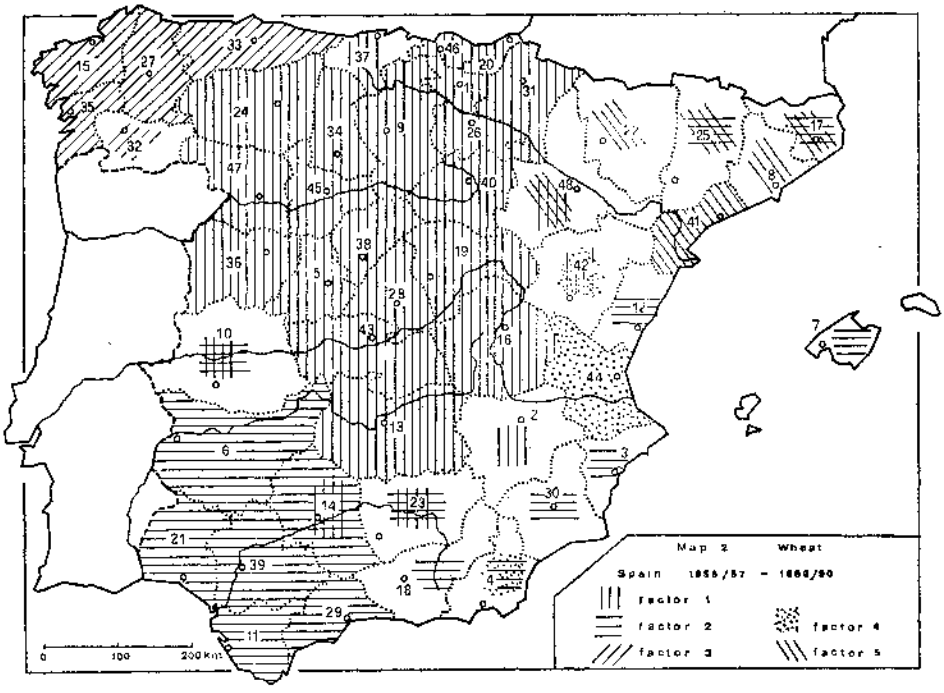
TABLE 2

WHEAT
PROPORTIONS OF THE VARIANCE OF THE PROVINCES
RELATED TO FACTORS

		1	2	3	4	5	h^2
1	Alava	53	4	20	10	9	96
2	Albacete	44	23	2	22	4	95
3	Alicante	19	44	9	15	6	92
4	Almería	7	37	2	37	8	92
5	Avila	69	14	11	2	0	96
6	Badajoz	22	65	6	1	0	94
7	Baleares	2	29	19	15	15	81
8	Barcelona	12	15	12	8	40	86
9	Burgos	60	13	16	2	5	96
10	Cáceres	40	44	9	0	0	93
11	Cádiz	11	71	5	2	2	91
12	Castellón	15	33	9	24	4	85
13	Ciudad Real	59	21	1	8	3	92
14	Córdoba	27	64	0	1	4	97
15	Coruña	8	0	60	6	12	86
16	Cuenca	71	10	1	11	3	95
17	Gerona	11	27	18	1	32	86
18	Granada	17	45	1	14	18	95
19	Guadalajara	68	14	4	6	6	97
20	Guipúzcoa	48	6	20	3	9	86
21	Huelva	8	69	2	7	5	92
22	Huesca	22	8	4	18	40	91
23	Jaén	30	36	1	6	13	86
24	León	71	10	12	0	1	95
25	Lérida	8	31	9	1	38	86
26	Logroño	56	10	11	9	11	96
27	Lugo	13	7	71	4	0	89
28	Madrid	66	13	1	2	11	93
29	Málaga	11	72	1	3	9	97
30	Murcia	9	48	1	19	12	90
31	Navarra	53	9	4	8	19	92
32	Orense	19	17	39	0	12	88
33	Oviedo	21	2	55	1	5	84
34	Palencia	64	13	15	2	5	98
35	Pontevedra	6	9	50	1	17	82
36	Salamanca	58	20	16	1	1	96
37	Santander	45	5	19	12	5	86
38	Segovia	72	11	9	3	2	97
39	Sevilla	17	73	1	1	4	96
40	Soria	66	14	2	11	3	96
41	Tarragona	0	9	17	3	61	90
42	Teruel	26	14	3	38	12	92
43	Toledo	66	21	2	5	2	96
44	Valencia	17	18	3	54	1	92
45	Valladolid	65	12	14	1	4	97
46	Vizcaya	57	2	23	7	2	90
47	Zamora	62	15	15	0	4	97
48	Zaragoza	34	8	5	20	27	95
General proportion of the variance		36	24	13	9	11	92

index) and the south of Lérida and Huesca. From there it flows over to the Northeast (Barcelona and Gerona), upriver to Zaragoza and less intensely to Teruel, Navarra and Logroño. Another connection with factor five appears in the Balearic Islands, whose prices thus seem to be subject to all the factors operating along the littoral of the peninsula. Incidentally, factor five also lightly affects the Galician coast (Coruña and Pontevedra).

In brief, the two self-determining regions are Castile and western Andalusia. There is no sudden cut-off between them, but rather an area where both influences are struggling. Towards the North, the Castilian continental nucleus also exhibits a glacis which is sensitive to other influences. Likewise the Northeast is a self-determining region, although its maritime character makes communication with the other landstrips of the peninsular littoral rather easy. The East, however, stands as a basically receptive and compound area, its particular factors being rather weak and in a sense within the orbit of those prevailing in the contiguous territories. The distinct combination of the principal components here shapes subdivisions. Aragón and Catalonia share factor five, which, depending on the location, becomes mixed with currents from nearby Castile or, in the coast, from faraway Andalusia. In the Levant a peculiar component of limited geographical spread (number four) also intervenes. It enters the Maestrazgo and spreads out on



the southeastern coast. In sum, the Southeast represents essentially an open region. Its particular component is very tenuous and does not quite assert itself. Here Western Andalusia is clearly exerting its influence.

III.

Wheat and barley are not grown on the same soils. The more fertile ones are devoted to wheat, since it commands a higher price and is also an important item for human consumption. Barley is grown in soils of diminishing returns and is used as fodder, even though occasionally it may be part of the rural population's diet. But in spite of these differences they both require similar land and climate. Hence the concordance among economic regions for wheat and barley should be high, as the projections we have just described demonstrate. Nonetheless the coincidence among the areas need not be exact, since the conditions of cultivation and demand are not the same. Thus wheat and barley are two products sufficiently alike so that analysis of their prices should reveal many analogies, but still different enough in regards to their economic role so that the price behaviour of one grain should stress by its contrast the other's peculiarities of production and consumption.

We will next try to measure the degree of accord among regions of the same kind, and then we will comment upon the most glaring discrepancies.

The degrees of similarity among the regions are shown in table three. It exhibits the coefficients of congruence between the columns of the factor matrix for barley and the kindred columns for wheat. The five A lines of the congruence matrix represent the correlations of each factor of barley, in turn, to the five factors of wheat; inversely, the AA columns indicate how each factor of wheat correlates to the five factors of barley. Among these coefficients only the very high ones indicate accord among the factors.

The greatest congruence (0.957) may be found among factors number three for barley as well as for wheat. In either case the region they refer to corresponds roughly to Galicia. Thus the northwest of Spain possesses an undeniable identity.

Component one of both matrices also reveals a close accord. This should come as no surprise. For wheat, factor one clusters basically the entire Castilian tableland and secondarily Andalusia. For barley, both Castiles and Andalusia coincide quite ostensibly, as a result of the factor solution which was used. The index of similarity between 1A and 1AA is no less than 0.948. On the other hand, since the specific behaviour of Andalusian barley prices falls within factor one, as shown above, while in the case of wheat Andalusia belongs mainly to factor two, the accord between the second column and the first line should also be high. It is indeed, with a 0.833 coefficient.

TABLE 3

MATRIX OF THE COEFFICIENTS OF CONGRUENCE

	AA 1	2	3	4	5
A					
1	0.948	0.833	0.654	0.636	0.601
2	0.723	0.636	0.555	0.817	0.876
3	0.660	0.500	0.957	0.457	0.591
4	0.619	0.904	0.567	0.831	0.770
5	0.427	0.083	0.255	0.524	0.153

Next we may consider region four of the prices for barley. Its core is located in the Southeast, which is linked to the entire Mediterranean coast and especially to Lower Andalusia. The 0.904 coefficient at the intersection of line four and column two shows precisely such a connection, with a high concordance for that zone with the factor for wheat prices represented by Lower Andalusia.

Aragón presents a special problem: for barley it is identified with factor two; for wheat it is factor five which is in closest accord with that region, even though the respective indices are not very high. From one product to the other, the labels for the region vary. However, congruence analysis reveals the undeniable consistence between both areas. The coefficient for 2A and 5AA is the highly significant one of 0.876.

Still, Aragón does not show merely internal coherence, but also exhibits strong external links. The coefficient of 4A with 5AA (0.770) and of 2A with 2AA (0.636) suggests that the association between Aragón and Lower Andalusia (or more exactly Cádiz, Málaga, Seville, Granada, Almería and also Murcia) is an important one. At the same time Aragón appears linked to Valencia, whose influence on the fluctuation of Aragonese wheat prices (Teruel) is evident in the factor four column of the corresponding matrix (table two). This fact is confirmed by the high congruence evinced by 4AA and 2A (coefficient 0.817), even though it does not quite explain it. As to wheat prices, Valencia's links with the Southeast (Almería, Murcia) are equally strong. The opposite occurs for barley, since here the Southeast appears to control the connection with Valencia. At any rate, at the intersection of both factors four we have a coefficient of 0.831.

Finally, among the fairly strong connections revealed by the present matrix, there is the one between Old Castile and Galicia (3AA with 1A, or 3A with 1AA; coefficients 0.654 and 0.660 respectively), as well as the one between Old Castile and Aragón (1AA with 2A; 0.723 coefficient). We shall discuss the nature of this last connection right away.

A close look at the indices on table one (based, we repeat, on the factor matrix for barley) reveals first that the Aragonese provinces and those related

to them exhibit high values in column two corresponding to the individual factor which clusters them;[15] second, that on the other hand the same districts show low values in column one, which is related to Castile and Andalusia. On the contrary, in table two (wheat) we can see that the Aragonese provinces only have low values in column five, the equivalent of column two for barley,[16] while most of them carry high indices in column one (Castile).[17]

One possible explanation of that discrepancy would be as follows: since early times Aragón was a grain area which could not produce enough to meet its own market's demand, not so much for barley as for wheat. For barley the region was probably self-sufficient, at least to a high proportion, and prices fluctuated according to domestic conditions, even though its variations could affect prices in the Catalonian coast of Barcelona and Tarragona. For wheat consumption, however, local production did not match local demand, and thus prices in Aragón were strongly influenced by prices in the Castilian tableland, whence wheat used to be imported. We may then assume, on the basis of the indices in table two, column one, that Castile was Aragón's main supplier of wheat during that third of the century from 1856-57 to 1889-90, even though that relationship has not been proven conclusively. That interpretation is abetted by the low correlation found among the series of yearly wheat and barley prices in the Aragonese districts. In Zaragoza, for example, the correlation coefficient (r) between the two series comes to only 0.233. Thus the index of determination (r^2) of the variance of one series by another turns out to be insignificant. Consequently the local price of barley was little affected by variations in the price of wheat, and viceversa. That disjunction may be explained by the Castilian origin of much of the wheat sold in the central Aragonese province.

We can then deduce from tables one and two relationships between the provinces and regions which are not always identical, but which instead vary according to product and circumstances. Thus Santander, as Aragón, reveals frequent connections for barley with the provinces of the Cantabric Gulf and with Galicia, with which it shares the price variations. As for wheat, however, its links with the producing areas of the tableland are quite evident. This is not surprising, since it was the port of exit for Castilian wheat and flour.

In summing up all these impressions, it would appear that provincial barley prices are usually rather susceptible to the region's internal changes, that is to say, that they respond to local market fluctuations. These markets are in turn subject to a large extent to the area's conditions of production. Barley consumption (mainly for fodder), does not give rise to trade currents of the same order as those of wheat. In the case of barley demand tends to

---

[15] Huesca 79, Zaragoza 76, Teruel 62, Lérida 60, Navarra 55.

[16] Huesca 40, Lérida 38, Zaragoza 27, Navarra 18 and Teruel hardly 12.

[17] Navarra 53, Zaragoza 34, Teruel 26, Huesca 22. Lérida however, is of little account here, only 8.

adjust itself to the area's productive capability. By contrast, demand for wheat, the principal bread grain, is conditioned not so much by productive capability as by population size. If wheat is needed, it will be imported.

Incidentally, at this time the littoral preferred to stop growing bread grains altogether in its most productive lands, in order to turn to more profitable crops such as fruit. The coastal zone where this process of agricultural specialization took place was counting on grain produced elsewhere to supply its markets regularly. Table two also reflects this change in Spanish agriculture, which favoured precisely the intrusion of Castilian and Andalusian wheat into the periphery.

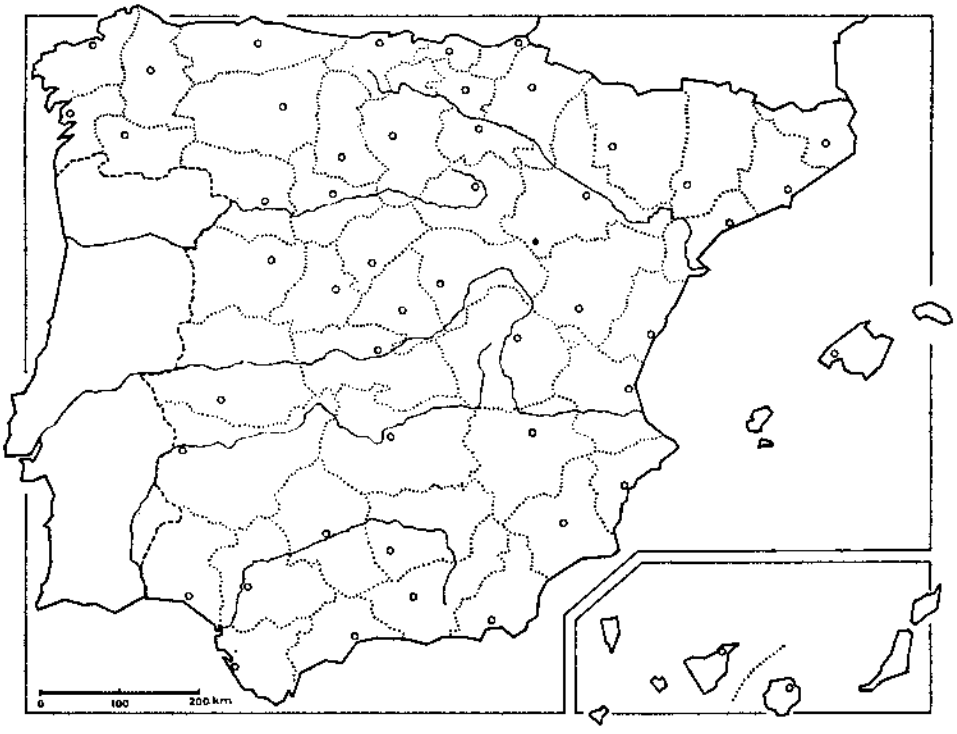
IV.

In concluding this essay, the main point we would like to stress is that our choice of investigative techniques has been fully justified by the outcome of the research. Factor analysis and factor matching have proven significantly applicable to data with the economic, historical and geographic characteristics of ours.

The nature of the statistical procedures used underscores one of the characteristics of the data and glosses over the others. Having taken the provinces as the variables in the problem, it is the locational traits which stand out in the solution. It is possible to imagine another approach where, taking the chronological units as the variables, the same techniques would emphasize the temporal elements and subsume the locational attributes or any others.

The answer to the question at hand is above all a geographic one. Our investigation has shown the behaviour of Spanish provincial prices from which economic regions are shaped. These areas exhibit a great deal of congruence and correspond generally to the historical regions, so that we have not hesitated to employ a traditional terminology in referring to them. The old names of Galicia, Castile, Andalusia, etc., do not distort the relationships which have been revealed. Still, this does not mean that the spatial units refer to surfaces with fixed limits or frontiers, such as the ones for the historical regions. Their outlines shift occasionally; what seems to characterize them is a tropism. Each region resembles rather a fascine of flows which vary individually but which always refer back to a common nexus. The regions are surfaces in continuous mutation, defined *hinc et nunc.* Thus geography links up with history.

There is another way in which history comes to bear on our problem. Factor analysis shows how the provinces cluster, but offers no explanation for that arrangement. To examine the economic background of the price relationships sheds more light on this particular point than the analysis of the geographic distribution of the data. This economic dimension is not, by the



way, independent of its historical context, and it is economic history which best clarifies the reasons for the provincial associations observed in our research. Factor analysis describes a set of relationships, but history suggests an explanation of them to the extent possible with our limited knowledge.

In sum, while our subject has a clear-cut geographic facet, it also possesses several aspects of interest to historians. First, it informs them of the availability of statistical techniques which can be of significant value in outlining historical regions. They can even provide solutions of a purely temporal character as long as the parameters of the problem are inverted for this kind of solution. Secondly, it is obvious that the analysis not only reveals generic relationships but also gives us some intimation of historical eventualities. Finally — and this is basic — it shows that without prior historical knowledge it is very difficult to infer an adequate explanation of the events.