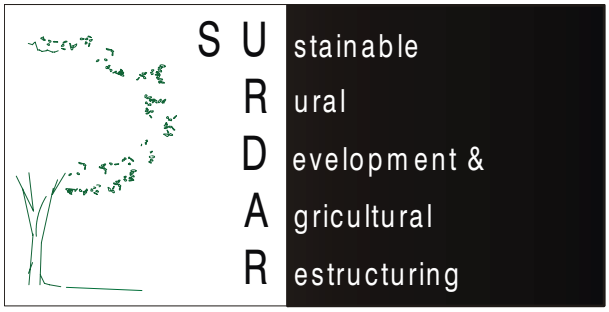

URBAN-RURAL DICHOTOMIES IN POLAND AND THE EU: MULTINATIONAL COMPARISONS AND AN EXPLORATION OF THE POLISH SMALL AREA DATA BASE

Working Paper No. 4

Zbigniew Zolkiewski, Hilary Ingham & Mike Ingham

August 2004

Address for correspondence:

Department of Economics
Lancaster University
Lancaster LA1 4YX

e-mail: h.ingham@lancaster.ac.uk

Introduction

Notwithstanding overall levels of economic growth that have, in general, been most impressive by European standards in the years since the mid nineteen-nineties, Poland remains a country with very uneven levels of development in which the post-communist dividend has been unequally distributed. Indeed, certain spatial areas, as well as certain socio-demographic groups, have even yet to experience many of the benefits of liberalisation. Following on from the beginnings made in Ingham (2003) and in recognition of the vast body of endeavour that has identified rural disadvantage as a particularly serious cause of concern in Poland, both before and after 1989, this paper is concerned with the identification of spatial disparities in development using statistical cluster analysis. The focus of attention in this paper will be on the smallest official territorial units identified in Poland, namely the NUTS 5 level gminas.

The information source employed in the paper is the Small Area Database (SADB), assembled by the Polish Central Statistical Office (GUS), for the year 2001. Although the number of available data series falls as the focus sharpens from higher to lower levels of spatial disaggregation, this is the most important source of small area statistics in Poland. The choice of indicators employed in the analysis to follow has in part been dictated by availability – for example, GDP and labour market data is not available at the gmina level – but also by recognition of the fact that certain local area characteristics are both direct measures of development and determinants of its future likely trajectory. In particular, attention focuses on housing, community infrastructure, business activity and structure, local amenities, the structure of agriculture, gmina finance and, finally, the environment. Having examined the spatial distribution of each of these indicators in turn, an overall typology based upon their aggregation is then compiled. In this way, it is hoped that further insights can be gained into the location and possible causes of successful and unsuccessful transitions in the Polish countryside and that transferable lessons might be drawn from the research. The results of the exercise will also be used to inform the choice the choice of localities for more in-depth case-study analysis in the next stage of the SURDAR project.

The next section of the paper describes the official delineation of Polish space into its NUTS components and into its rural and urban components. This exercise thereby permits an overview of the structure of the SADB. Confronted with the administrative definitions of the rural-urban status of areas employed within the country, comparisons are made with prevailing international conventions. While these are shown to alter the apparent complexion of Polish space, in truth the alternative definitions are equally arbitrary. There then follows, in Section 3, a summary of the statistical clustering procedure that will be employed to categorise local areas along the seven dimensions to be highlighted in Section 4. This is followed by the aggregate typology of development across gminas that follows from these clustering analyses. A summary and conclusions close the paper.

Polish Space

Poland has a three tier local government structure: the sixteen regional voivodships, the county level powiats, of which there were 372 in 2001, and the 2489 community level gminas. One aspect of the pre-accession approximation process required Poland

to disaggregate its territory into units that conform to the EU's NUTS classification. The voivodships are NUTS 2 level entities, the powiats NUTS 4 level and the gminas are NUTS 5 level. In addition, there are 44 subregions, each composed of groups of powiats, which constitute the country's NUTS 3 level tier of spatial classification. This paper concentrates on the gminas, the lowest spatial aggregates for which data is officially made available.

The Polish definition of what constitutes a rural area is administrative and rather circular; it being 'territory situated outside town administrative boundaries' (MARD, 1998). Although the definition is applied at the level of the gmina, three types of unit are in fact defined: pure rural and urban localities, along with mixed urban/rural areas. The first three rows of Table 1 provide the breakdown of the 2489 gminas in existence in 2001, along with that of the total population, according to their location within this hierarchy. Under Polish conventions, the country is clearly mostly rural, with almost two-thirds of communities and twenty-eight per cent of the population classified as purely rural.

Table 1
Gminas and their populations in 2001

	Number (% total)	Population (% total)
Urban	318 (12.8)	19,175,331 (49.6)
Rural	1595 (64.1)	10,868,664 (28.1)
Mixed Urban/Rural	576 (23.1)	8,587,458 (22.2)
< 150 people per km^2	1978 (79.5)	15,654,232 (40.6)
< 100 people per km^2	1724 (69.3)	12,622,672 (32.7)

As discussed in Ingham (2003), international organisations, such as the OECD and Eurostat, adopt definitions of rurality that are based on what, at first sight, might seem to be more objective population density criteria. The OECD measure, which defines a community as rural if it has a population density of less than 150 people per square kilometre is the easiest of these to apply and will form the basis of the comparison effected here. However, it must be recognised that the precise density measure adopted as the delineator is arbitrary: the figures given in the fourth and fifth columns of Table 1 are based on cut-off points of 150 and 100 people per square kilometre of gmina territory, respectively, and they actually increase the apparent size of rural Poland.

Given the focus of the project on sustainable rural development and agricultural restructuring, it was decided that, for the purpose of this and other papers, special attention would be paid to the performance of the eastern voivodships of Lubelskie, Podkarpackie, Podlaskie and Świętokrzyskie. This decision was based on their poor relative economic performance, peripherality and agricultural complexions, as revealed by Table 2. This ranks these regions among Poland's sixteen voivodships on certain measures of interest in the context of the overall study and highlights their under-developed structure. It should be noted that the rankings are direct for three of the measures, but inverse in the case of the percentage of employment accounted for by agriculture. Throughout the remainder of this paper these territories will be referred to by variants off the expression 'the east'.

Table 2
Ranking of Eastern Voivodships: 2001

	GDP per capita	Gross value added per employed person	Agricultural employment (% of total)	Rural Gminas (% of total)
Lubleskie	16	15	16	16
Podkarpackie	15	16	14	14
Podlaskie	13	13	13	12
Świętokrzyskie	12	14	15	13

Cluster Analysis: Methodology

Cluster analysis is a useful suite of non-parametric techniques for the identification of patterns and segments in multivariate data sets. In this particular instance, a non-hierarchical procedure – the *FASTCLUS* programme contained within the SAS statistical package – was used to identify groupings of internally homogeneous gminas between which there is as much heterogeneity as possible. Non-hierarchical procedures require the number of resulting clusters to be pre-specified, although the choice in this case was determined by running the routine nineteen times, allowing the number of final clusters to range from two to twenty. The preferred, final clustering solution was then selected according to certain diagnostic statistics that will be outlined below. With the current data set, a hierarchical approach that allows for the selection of the preferred solution from the results of a continuous process of cluster aggregation ranging from 2489 (the total number of observations) to just one was clearly not feasible.

FASTCLUS uses a method known as centroid sorting. For an 'n' cluster solution an initial set of cluster seeds are selected using the first n observations in the data set that have no missing values. Each observation is then assigned to one of these preliminary clusters and the cluster means are computed. The original seeds are then replaced by these cluster means and the observation assignment exercise is repeated. This process continues until no further changes occur in the membership of the clusters. The notion of distance, which the process seeks to minimise between members of the same cluster and maximise between clusters, is central to the clustering technique. *FASCLUS* allocates observations using squared Euclidean distances and, in the case of missing values, the programme computes an adjusted distance using the non-missing values. In order to overcome the potential problems associated with different units of measurement across variables, all data in the following clustering analyses were standardised to have a mean of 0 and a standard deviation of 1 prior to clustering. It must be noted, however, that the complementary descriptive statistics relate to the original, unstandarised data.

Two diagnostic statistics were employed to select the optimal number of clusters:

1. The pseudo F-statistic developed by Calinski and Harabasz (1974). This involves taking the maximum value of:

$$C(g) = \frac{trace(B)}{g-1} \Big/ \frac{trace(W)}{n-g}$$

where *B* is the between-cluster sum of squares matrix, *W* is the within-cluster sum of squares matrix, g is the number of clusters and n is the number of observations.

2. The cubic clustering criterion (CCC) devised by Sarle (1983). This involves taking the maximum value of :

$$\frac{(1-E(R^2))}{(1-R^2)} x \frac{((np/2)^{0.5})}{((0.001+E(R^2))^{1.2}}$$

where R^2 is the proportion of the variance accounted for by the clusters and its expected value is determined under the assumption that the data have been sampled from a uniform distribution based on a hyperbox. The variable *p* is an estimate of the dimensionality of the between-cluster variation and the constant terms in the statistic were chosen on the basis of extensive simulation results.

In an independent simulation study, Milligan and Cooper (1985) found that the pseudo F-statistic allocated 390 observations to the 'correct' clusters out of a possible total of 426, whereas the corresponding figure for the CCC was 321. In fact, the pseudo F-statistic was the best performer out of the 30 stopping rules examined in their work and where conflicts occur in what follows, the pseudo F-statistic has been afforded primacy.

Prior to undertaking each clustering exercise, the data were initially screened using the SAS *UNIVARIATE* procedure, which generates summary statistics for each of the variables. Also, the results presented below provide a listing of 'outliers' – the five highest values and the five lowest ones for each variable – which may prove helpful in identifying seemingly successful or disadvantaged gminas. These extreme observations are presented in tabular form, with the last column representing the seven digit SADB identifier for the gmina concerned. The first two digits represent the voivodship identifier and these are presented in Table 3.

Table 3
SADB 2-digit voivodship identifiers

Voivodship	Code
Dolnośląskie	02
Kujawsko-Pomorskie	04
Lubelskie	06
Lubuskie	08
Łódzkie	10
Małopolskie	12
Mazowieckie	14
Opolskie	16
Podkarpackie	18
Podlaskie	20
Pomorskie	22
Śląskie	24
Świętokrzyskie	26
Warmińsko-Mazurskie	28
Wielkopolskie	30
Zachodniopomorskie	32

The next two digits identify the powiat, and the next two the gmina. In the interest of brevity, these will not be detailed explicitly here. The final digit of the identifier is, however, central and represents the gmina type, as defined in Table 4 below.

Table 4
SADB gmina type identifier

Gmina type	Identifier
Urban	1
Rural	2
Mixed	3

Clusters of Development: Results

As noted above, this paper focuses on seven gmina development indicators drawn from the 2001 SADB. The results obtained from conducting *FASCLUS* cluster analyses on each of these in turn are presented in the following sub-sections. In each case, special emphasis is afforded to gminas lying in the four eastern regions that were highlighted above.

1. HOUSING

The variables considered initially used for the housing cluster represent a mix of physical stocks and flows, financial statistics and ownership structure and these are presented in Table 1.1:

Table 1.1
Housing Cluster Variables

Variable Number	Variable description	Mnemonic
1	Total number of dwellings per capita	DWELPC
2	Dwellings completed per capita	DCOMPPC
3	Useable floor space per capita (m^2)	FLSPPC
4	Proportion of dwellings owned by the gmina	PRDWGM
5	Proportion of useable floor space owned by the gmina	PRSPGM
6	Modernisation: proportion of dwellings newly fitted with water line	WATPD
7	Modernisation: proportion of dwellings newly fitted with sewerage	SEWPD
8	Modernisation: proportion of dwellings newly fitted with termic line (i.e. central heating)	TERPD
9	Modernisation: proportion of dwellings newly fitted with hot water	HWTRPD
10	Modernisation: proportion of dwellings newly fitted with gas	GASPD

With the exception of PRDWGM, it was assumed in each case that higher scores represented an indicator of greater development. Earlier discussions had indicated the desirability of including some measure of rental arrears in the data set as a proxy for local income levels, but appropriate information has not been included in the SADB since 1994. In the case of DWELPC, FLSPPC and PRDWGM only an aggregate figure was provided for the eleven individual gminas within Warsaw and therefore the variables were defined as missing for these observations.

It is computationally inefficient to include highly collinear variables in a cluster analysis, which in any case do not provide additional information, and therefore the correlation matrix for the indicators included in Table 1.1 was examined as the first step in the analysis. This is reproduced as Table 1.2, with the variables being identified by the numbers assigned to them in Table 1.1.

Table 1.2
Housing Correlation Matrix

	1	2	3	4	5	6	7	8	9	10
1		-0.287	0.956	1.000	0.332	0.009	0.012	0.008	0.005	0.005
2			-0.241	-0.288	0.027	-0.013	-0.007	-0.006	0.005	-0.001
3				0.954	-0.124	0.007	0.008	0.002	0.005	-0.002
4					0.986	0.010	0.015	0.009	0.006	0.006
5						0.077	0.199	0.044	0.041	0.054
6							0.086	0.012	0.063	-0.005
7								0.060	0.048	0.056
8									0.174	0.029
9										0.044
10										

Given that the correlation coefficient between PRDWGM and PRSPGM was 0.986, only the former was retained in the final data set.[1]

Table 1.3 provides the summary statistics for the variables finally included in the housing cluster analysis. It should be recalled that these data are in the original units of measurement.

Table 1.3
Basic Statistics for Housing

Variable	N	Mean	-Standard Deviation	Minimum	Maximum
DWELPC	2478	0.2782	0.0375	0.1620	0.4807
DCOMPPC	2489	0.0015	0.0033	0.0000	0.0720
FLSPPC	2478	19.1548	2.4493	12.2396	36.2616
PRDWGM	2478	0.0520	0.0705	0.0000	0.5656
WATPD	2489	0.0001	0.0005	0.0000	0.0202
SEWPD	2489	0.0001	0.0006	0.0000	0.0171
TERPD	2489	0.0000	0.0002	0.0000	0.0072
HWTRPD	2489	0.0000	0.0001	0.0000	0.0023
GASPD	2489	0.0000	0.0003	0.0000	0.1280

Extreme observations (4 eastern regions in bold)

The following list identifies explicitly the extreme observations for each of the housing variables. Given the interest of the research in the four eastern voivodships of Lubelskie, Podkarpackie, Podlaskie and Świętokrzyskie any of the outliers that lie in these regions are presented in bold typeface.

Variable	Value	Gmina	SADB Code
DWELPC			
Lowest			
	0.1620	Labowa	1210082
	0.1750	Grybów – r	1210042
	0.1769	Chelmiec	1210022
	0.1801	Kamionka Wielka	1210052
	0.1843	Przodkowo	2205032
Highest			
	0.42919	**Narew**	**2005082**
	0.43003	**Milejczyce**	**2010062**
	0.43050	**Dubicze Cerkiewne**	**2005052**
	0.45816	**Bielsk Podlaski – r**	**2003032**
	0.48073	**Orla**	**2003062**

The interesting aspect of these summary statistics is that rural areas have both the most and the fewest dwellings per capita and, just as significantly, the five gminas in the former category all lie within Podlaskie, which is one of the four eastern voivodships that were suspected initially to be development blackspots. On the other hand, four of the five gminas with the fewest dwellings are to be found in Małopolskie, a southern region centred on Kraków.

Variable	Value	Gmina	SADB Code
DCOMPPC			
Lowest			
	1.7142E-05	Swietochlowice	2476011
	4.4930E-05	Grybów – r	1210042
	7.0482E-05	Niemodlin	1609073
	7.1911E-05	Zawadzkie	1611073
	7.2343E-05	Warta	1014093
Highest			
	0.0400	Kolbaskowo	3211022
	0.0443	Jablonna	1408022
	0.0473	Jastarnia	2211021
	0.0537	Warszawa – Bialoleka	1431021
	0.0720	Krynica Morska	2210011

In the case of dwellings completed per capita – a flow as opposed to a stock measure – the picture is more varied than was evident for DWELPC. However, it might be noted that the voivodships of Mazowieckie, which is the capital region and has the highest GDP per capita in Poland, and Pomorskie each account for two of the gminas with the highest dwelling completion rates. No communities in the four eastern voivodships appear in either of the outlier lists.

Variable	Value	Gmina	SADB Code
FLSPPC			
Lowest			
	12.2396	**Tyrawa Woloska**	**1817062**
	12.4710	Grybów – r	1210042
	12.8036	Korzenna	1210062
	12.8345	Gniewino	2215052
	12.9819	Podegrodzie	1210142
Highest			
	29.4573	**Milejczyce**	**2010062**
	30.6366	**Bielsk Podlaski – r**	**2003032**
	30.6423	**Dubiczne Cerkiewne**	**2005052**
	31.9244	**Orla**	**2003062**
	36.2616	Podkowa Lesna	1405021

Evidence of spatial concentration emerges once again in the case of floor space per head. In particular, three of the gminas with the lowest values of FLSPPC are in Małopolskie, which scored badly on dwellings per capita, while four of the highest values are to be found in the eastern voivodship of Podlaskie, which scored highly on dwellings per head. However, the lowest value of all lies in Podkarpackie, which is another of the eastern regions. The picture is further clouded by the fact that nine out of the ten outliers are to be found in rural gminas.

Variable	Value	Gmina	SADB Code
PRDWGM			
Lowest			
	0.00	25 Gminas	
Highest			
	0.4725	Mieroszów	0221063
	0.4853	Kanienna Góra	0207011
	0.4896	Wojcieszów	0226011
	0.5080	Leknica	0811011
	0.5656	Bogoszów-Gorce	0221011

Of the twenty-five communities gminas for which the proportion of the dwelling stock owned by the gmina takes the value zero, all are rural, ten are in the east and five of those in Lubelskie. The highest concentrations of public housing, on the other hand, are to be found in four urban and one mixed locality. Four of these five observations are located in Dolnoślaskie.

Variable	Value	Gmina	SADB Code
WATPD			
Lowest			
	0.00	2297 Gminas	
Highest			
	0.0035	Chelmno – r	0404022
	0.0040	Bogatynia	0225033
	0.0054	Dobromierz	0219032
	0.0077	Denbowa Laka	0417022
	0.0202	Lewin Klodzki	0208092

Over ninety per cent of gminas experienced no water modernisation in 2001. The largest number of the 192 gminas for which the variable is positive were rural areas, which might seem to reflect a disadvantaged starting position. However, both urban and mixed gminas were in fact twice as likely to have experienced such modernisation than countryside communities. Overall, thirty-two of the gminas with modernised water connections lay in the eastern voivodships, while twenty-six (13.5%) were in Mazowieckie alone.

Variable	Value	Gmina	SADB Code
SEWPD			
Lowest			
	0,00	2228 Gminas	
Highest			
	0.0065	Bobrowice	0802022
	0.0075	Pieszyce	0202031
	0.0076	Walim	0221082

0.0077	Jedlina-Zdrój	0221021
0.0171	Miroslawiec	3217033

The vast majority of communities witnessed no sewerage modernisation work in 2001. Mixed communities account for the largest number of the 261 gminas for which the variable was positive, although, in percentage terms, urban areas were equally likely to experience such activity. A total of seventy-seven rural gminas were included in the total. In terms of individual voivodships, Dolnoślaskie and Wielkopolskie had the largest numbers of communities benefiting from modernisation, while Lubleskie accounted for fourteen of thirty-eight localities in the east to have so profited.

Variable	Value	Gmina	SADB Code
TERPD			
Lowest			
	0.00	2365 Gminas	
Highest			
	0.0015	**Klukowo**	**2013042**
	0.0016	Unislaw	0404072
	0.0017	Zareby Koscielne	1416112
	0.0021	Chojnów – u	0209011
	0.0072	Swiekatowo	0414102

Only 124 gminas witnessed any central heating modernisation in 2001 and urban localities account for the largest number of these. While the eastern regions perform relatively badly on this score, it might be noted that a rural gmina in Podlaskie appears amongst the five areas with the best records. Furthermore, notwithstanding the fact that they were least likely to experience such improvements, rural areas accounted for four of these.

Variable	Value	Gmina	SADB Code
HWTRPD			
Lowest			
	0.00	2451 Gminas	
Highest			
	0.0007	Pluznica	0417042
	0.0010	**Hajnówka – u**	**2005011**
	0.0014	Pepowo	3004042
	0.0014	Knurów	2405011
	0.0023	Kobiór	2410022

Only thirty-eight gminas had any dwellings newly fitted with hot water in 2001 and two-thirds were urban. Nevertheless, three out of the best five performers on this measure were rural communities. One member of this list of five was an urban gmina in Podlaskie..

Variable	Value	Gmina	SADB Code
GASPD			
Lowest			
	0.00	2388 Gminas	
Highest			
	0.0017	Borzechów	0609022
	0.0018	Swidwin	3216062
	0.0019	Bytom Odrzanski	0804023
	0.0040	Swieszyno	3209082
	0.0128	Szamocin	3001053

Over ninety-five per cent of communities had no dwellings newly fitted with gas in 2001. Of the 101 localities that did experience such work, thirty-six were urban, thirty-one rural and thirty-four mixed gminas, respectively. Zachodniopomorskie was the voivodship witnessing the heaviest activity and Podkarpackie in the east ranked fourth in the list. None of the five gminas experiencing the most intense activity were urban and none was located in the east.

Cluster Results

Table 1.4 presents the diagnostic statistics obtained from undertaking the discrete cluster analyses for pre-determined numbers of clusters ranging from two to twenty for which, it will be recalled, the data have been standardised to z-scores. Under both the pseudo-F test and the CCC, the best fit to the data is obtained by implementing an eighteen cluster solution. Table 1.5 provides the resulting housing cluster summary and Table 1.6 the associated statistics for the included variables. Table 1.7, which provides the matrix of cluster and variable means, confirms that an eighteen cluster solution is rather a large number of groupings for purposes of interpretation and, as the summary schedule in Table 1.5 attests, it is driven by the presence of eight outlying clusters that have either one or two constituent members. Also, a number of clusters are very similar, as indicated by the short distances between cluster centroids and the similarity of several of the cluster means for individual variables that are reported in Table 1.7.

The cause of this inability to discriminate between observations appears to lie in the identical, zero values of WATPD, SEWPD, TERPD, HWTRPD and GASPD that are reported for most gminas. As these variables are only picking up flows over a one year period, they may not in fact accurately reflect differences in stocks of modernised as opposed to un-modernised housing. In view of this possibility and given the difficulties of interpretation of the results when they are included in the data set, it was decided to re-run the FASTCLUS procedure with them omitted. That is, the cluster analysis was repeated using only the variables DWELPC, DCOMPPC, FLSPPC and PRDWGM.

Table 1.4
Housing Fastclus Results

N	Pseudo F	R^2	Cubic Clustering Criterion
20	421.28	0.4938	188.157
19	413.94	0.4876	180.443
18	**464.21**	**0.4810**	**197.643**
17	446.39	0.4739	185.181
16	443.22	0.4663	178.567
15	346.98	0.4582	127.529
14	357.53	0.4494	126.906
13	365.97	0.4398	124.682
12	267.21	0.4292	64.520
11	316.95	0.4176	85.192
10	349.78	0.4047	94.076
9	358.56	0.3796	92.661
8	381.15	0.3519	95.560
7	316.56	0.3213	60.782
6	283.91	0.2870	40.041
5	254.63	0.2480	21.954
4	254.29	0.2028	16.901
3	248.89	0.1494	10.013
2	231.09	0.0841	0.702

Table 1.5
Housing Cluster Summary

	Frequency	RMS Standard Deviation	Max Distance from Seed to Cluster	Nearest Cluster	Distance Between Cluster Centroids
1	383	0.5183	5.8850	13	1.7581
2	41	0.8680	7.4549	13	3.2789
3	6	1.4279	6.3603	17	9.8151
4	14	1.2210	7.3350	13	6.8069
5	1	.	0	11	15.8319
6	1	.	0	12	26.5354
7	666	0.3458	6.1052	13	1.4081
8	1	.	0	9	20.7788
9	31	1.1657	6.9527	2	5.9640
10	1	.	0	15	32.9520
11	2	1.3269	2.8148	4	14.0397
12	2	1.2753	2.7054	2	9.7389
13	971	0.3623	6.5710	7	1.4081
14	305	0.6426	6.5558	13	2.5529
15	32	1.0393	7.4637	13	4.7655
16	1	.	0	7	13.7150
17	30	0.8977	5.1383	1	4.4811
18	1	.	0	16	29.7323

Table 1.6
Statistics for Housing Variables

Variable	Total STD	Within STD	R-Square	RSQ/(1-RSQ)
DWELPC	1	0.6954	0.5209	1.0874
DCOMPPC	1	0.5209	0.7305	2.7105
FLSPPC	1	0.6148	0.6247	1.6644
PRDWGM	1	0.6069	0.6342	1.7337
WATPD	1	0.3139	0.9021	9.2163
SEWPD	1	0.4868	0.7646	3.2483
TERPD	1	0.3197	0.8985	8.8515
HWTRPD	1	0.2491	0.9384	15.2309
GASPD	1	0.4042	0.8378	5.1635
OVER-ALL	1	0.4900	0.7616	3.1937

Table 1.7
Housing Cluster Means

Cluster	WATPD	SEWPD	TERPD	HWTRPD	GASPD	DWELPC	DCOMPPC	FLSPPC	PRDWGM
1	-0.09	-0.13	-0.10	-0.08	-0.06	0.96	-0.07	1.54	-0.37
2	2.96	0.82	-0.07	-0.08	-0.04	-0.02	-0.18	-0.20	0.37
3	-0.13	-0.18	-0.12	-0.08	-0.08	1.70	14.14	1.08	-0.09
4	0.20	-0.00	1.60	6.45	0.28	0.27	0.42	0.05	0.36
5	4.36	3.43	-0.12	31.75	-0.08	0.42	0.15	0.29	0.55
6	39.08	-0.18	-0.12	-0.08	-0.08	0.77	-0.17	0.11	2.88
7	-0.08	-0.14	-0.10	-0.08	-0.04	-1.01	-0.13	-0.90	-0.35
8	-0.13	26.59	-0.12	-0.08	1.73	0.09	-0.37	-0.85	2.95
9	0.07	5.98	-0.12	0.03	0.01	0.05	-0.12	-0.05	1.11
10	-0.13	3.56	37.35	-0.08	-0.08	-0.98	-0.37	-0.82	-0.26
11	-0.13	-0.13	7.16	19.15	2.25	-0.08	-0.41	-0.08	0.66
12	12.63	0.68	-0.12	-0.08	-0.08	0.08	-0.46	0.58	1.11
13	-0.10	-0.11	-0.10	-0.06	-0.04	0.02	-0.13	0.05	-0.32
14	-0.05	-0.01	-0.02	-0.03	0.05	0.86	0.11	-0.26	2.06
15	0.18	0.09	4.61	-0.08	0.30	-0.10	-0.01	-0.06	0.18
16	-0.13	-0.18	-0.12	-0.08	13.66	-1.33	-0.18	-1.17	0.21
17	-0.04	0.03	-0.12	-0.08	0.06	0.78	4.39	1.61	-0.10
18	-0.13	1.42	-0.12	-0.08	43.31	-0.33	-0.38	-0.16	0.60

Repeated Clustering Results

Table 1.8 presents the diagnostic statistics obtained when the FASTCLUS procedure was invoked repeatedly to generate two to twenty clusters with the restricted data set. Under the pseudo F-statistic, a five cluster would appear to be optimal, although globally the CCC points to rather more. Indeed a five cluster solution is only weakly locally optimal under this criterion. Nevertheless, given the stated preference for the pseudo F-statistic, a five cluster solution was adopted.

Table 1.8
Restricted Housing Fastclus Results

N	Pseudo F	R^2	Cubic Clustering Criterion
20	675.81	0.7798	29.567
19	648.23	0.7739	24.709
18	698.89	0.7675	29.065
17	672.39	0.7606	24.257
16	707.36	0.7530	26.487
15	712.34	0.7447	25.117
14	754.56	0.7355	27.710
13	768.79	0.7253	27.025
12	750.30	0.7139	22.668
11	754.87	0.7009	20.593
10	722.78	0.6806	15.510
9	743.07	0.6688	13.461
8	804.99	0.6484	16.575
7	792.23	0.6238	11.448
6	756.22	0.5933	3.377
5	**889.97**	**0.5541**	**11.644**
4	547.54	0.4536	-15.259
3	763.36	0.3343	13.296
2	442.67	0.1882	-11.614

Table 1.9 provides the new housing cluster summary, with half of all observations falling into Cluster 5. Cluster 1 and, even more, Cluster 2 are very small, but their distances from other cluster centroids suggests that they represent genuine outlier groupings.indicate the statistics for the restricted set of housing variables and the cluster means.

Table 1.9
Restricted Housing Cluster Summary

	Frequency	RMS Standard Deviation	Max Distance from Seed to Cluster	Nearest Cluster	Distance Between Cluster Centroids
1	42	1.3196	10.2748	3	3.9621
2	6	2.1418	7.1640	1	10.4698
3	804	0.6717	6.5035	5	1.9641
4	356	0.8248	5.4547	3	2.6493
5	1281	0.5068	3.1009	3	1.9641

Tables 1.10 and 1.11 provide the statistics for the restricted set of housing variables and the variable by cluster means and their associated rankings. It is here assumed that higher values on all of the included variables except for PRDWGM represent higher development levels. The outlying Cluster 2 stands out as the most developed on the housing measures, with it ranking poorly only on the proportion dwellings owned by the gmina. Cluster 5, on the other hand, has almost diametrically opposite profile.

Table 1.10
Statistics for Restricted Housing Variables

Variable	Total STD	Within STD	R-Square	RSQ/(1-RSQ)
DWELPC	1	0.7305	0.4673	0.8773
DCOMPPC	1	0.5153	0.7349	2.7742
FLSPPC	1	0.7177	0.4857	0.9443
PRDWGM	1	0.5771	0.6675	2.0073
OVER-ALL	1	0.6416	0.5890	1.4331

Table 1.11
Restricted Housing Cluster Means & Ranks

Cluster	DWELPC	DCOMPPC	FLSPPC	PRDWGM
1	0.5674 (4)	3.7582 (2)	1.7593 (1)	-0.1479 (3)
2	1.7033 (1)	14.1442 (1)	1.0845 (2)	-0.0965 (4)
3	0.6521 (3)	-0.1054 (4)	0.9097 (3)	-0.3543 (1)
4	0.8203 (2)	0.0572 (3)	-0.2634 (4)	2.0095 (5)
5	-0.6571 (5)	-0.1392 (5)	-0.5538 (5)	-0.3236 (2)

Under the above assumptions regarding the relationship between the magnitude of the variable means and levels of development, it is possible to construct an aggregate housing score for each cluster by summing across the rankings on the four separate variables. As shown in Table 1.12, Cluster 2 appears as the most developed and Cluster 5 the least.

Table 1.12
Overall Cluster Rankings

Cluster	Score	Rank
1	10	2
2	8	1
3	11	3
4	14	4
5	17	5

Table 1.13 disaggregates the data to indicate the proportion of each voivodship's gminas falling into each of the five clusters. Opolskie and Warmińsko-Mazurskie aside, the eastern regions stand out as having no members of either of the first two clusters. Indeed, Podkarpackie is the worst housing region of all, with almost ninety-four per cent of its gminas falling into the poor profile Clusters 4 and 5. However, the other eastern voivodships perform reasonably well on the housing metric and only Opolskie performs better than Podlaskie. The overall spatial distribution of the membership of the clusters is depicted in Map 1.

Table 1.13
Distribution of Gminas Across Restricted Housing Clusters by Voivodship (%)

Voivodship	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Dolnoslaskie	1 (0.6)	0 (0.0)	44 (26.9)	92 (54.4)	32 (18.9)
Kujawsko-Pomorskie	1 (0.7)	0 (0.0)	10 (6.9)	24 (16.7)	109 (75.7)
Lubelskie	**0 (0.0)**	**0 (0.0)**	**123 (57.7)**	**6 (2.8)**	**84 (39.4)**
Lubuskie	1 (1.2)	0 (0.0)	17 (20.5)	30 (36.1)	35 (42.2)
Lodzkie	2 (1.1)	0 (0.0)	105 (59.3)	18 (10.2)	52 (29.4)
Malopolskie	2 (1.1)	0 (0.0)	41 (22.5)	3 (1.6)	136 (74.7)
Mazowieckie	18 (5.5)	2 (0.6)	113 (34.8)	24 (7.4)	168 (51.7)
Opolskie	0 (0.0)	0 (0.0)	54 (76.1)	12 (16.9)	5 (7.0)
Podkarpackie	**0 (0.0)**	**0 (0.0)**	**10 (6.3)**	**3 (1.9)**	**147 (91.9)**
Podlaskie	**0 (0.0)**	**0 (0.0)**	**88 (74.6)**	**6 (5.1)**	**24 (20.3)**
Pomorskie	7 (5.7)	2 (1.6)	4 (3.3)	22 (17.9)	88 (71.5)
Slaskie	1 (0.6)	0 (0.0)	101 (60.8)	30 (18.1)	34 (20.5)
Swietokrzyskie	**0 (0.0)**	**0 (0.0)**	**44 (43.1)**	**0 (0.0)**	**58 (56.9)**
Warminsko-Mazurskie	0 (0.0)	0 (0.0)	9 (7.8)	26 (22.4)	81 (69.8)
Wielkopolskie	5 (2.2)	0 (0.0)	33 (14.6)	18 (8.0)	170 (75.2)
Zachodniopomorskie	4 (3.5)	2 (1.8)	8 (7.0)	42 (36.8)	58 (50.9)

In Table 1.14, overall housing cluster membership is broken by gmina type. This indicates that, nationwide, rural gminas represent half of Cluster 1's membership and one-third of that of Cluster 2. On the face of it, this might seem to indicate that housing conditions in the countryside are better than those in towns and cities, but it must be noted that three-quarters of the membership of the worst performing Cluster 5 are rural.

Table 1.14
Restricted Housing Cluster Membership by Gmina Type (%)

Gmina Type	1	2	3	4	5
Urban	17 (40.5)	3 (50.0)	71 (8.8)	178 (50.0)	49 (3.8)
Rural	21 (50.0)	2 (33.3)	592 (73.6)	21 (5.9)	959 (74.9)
Mixed Urban/Rural	4 (9.5)	1 (16.7)	141 (17.5)	157 (44.1)	273 (21.3)

2. INFRASTRUCTURE

The general attributes gathered together under the infrastructure banner relate to the local road network, the prevailing communications system and the coverage of public utilities' (gas, sewerage and water) services. Because different problems arise in the context of each of the measures examined under these headings, the preliminary discussion will be presented beneath three sub-headings for the purpose of clarity.

2(a) ROAD NETWORK

Data on the road network were missing in the 2001 SADB and had to be substituted by information from the year 2000. Four gminas that were categorized as 'rural' in 2000 were categorized as 'mixed' in 2001, hence the last digit of their SADB identifiers changed. These changes are recorded in Table 2a.1. For the purposes of the analysis these gminas were given their 2001 identifiers to simplify merging files accurately.

Table 2a.1
Gmina Status Changes 2000-2001

Gmina	SADB 2000 code	SADB 2001 code
Ryglice	1216062	1216063
Hanilów	1412072	1412073
Krzanowice	2411032	2411033
Koprzywnice	2609042	2609043

The SADB contains information on four categories of urban and non-urban roads, although it does not contain any data on non-gmina highways passing through individual communities. In order to simply the analysis, the urban and non-urban elements of the road network were summed and deflated by the total gmina area to give the four variables described in Table 2a.2.

Table 2a.2
Road Network Cluster Variables

Variable description	Mnemonic
Gmina roads, total	RDTOTPSQKM
Gmina roads, hard surface	RDHRDPSQKM
Gmina roads, improved hard surface	RDIHSPSQKM
Gmina roads, repaired	RDREPPSQKM

The correlation matrix for these variables is provided in Table 2a.3 below, which has an obvious simplification of the variable mnemonics in its first column. Given the high values of some of the bivariate correlation coefficients, only two variables were retained for further analysis: RDHRDPSQKM (as used by Czyżewski and Zienkowski, 2000) and RDREPPSQKM were used in the cluster analysis

Table 2a.3
Road Network Correlation Matrix

	RDTOTPSKM	RDHRDPSQKM	RDIHSPSQKM	RDREPPSQKM
RDTOT	1	0.656	0.587	0.122
RDHRD		1	0.889	0.147
RDIHS			1	0.139
RDREP				1

2(b) TELECOMMUNICATIONS

It was necessary to use data from the 1999 SADB for the coverage of telephone lines and installations. Thus, in addition to the changes noted in Table 2a.1 between the 2000 and 2001 SADBs, account also had to be taken of five further changes from rural to mixed status between the 1999 and 2000 SADBs. These latter amendments are noted in Table 2b.1

Table 2b.1
Gmina Status Changes 1999-2000

Gmina	SADB 1999 code	SADB 2000 code
Prsice	0220022	0220023
Tyszowce	0618122	0618123
Kosów Lacki	1429052	1429053
Nekla	3030032	3030033
Jastrowice	3031022	3031023

The telecommunications variables originally considered for inclusion in the infrastructure cluster analysis, all of which are defined per head of population, are presented in Table 2b.2.

Table 2b.2
Telecommunications Cluster Variables

Variable Number	Variable description	Mnemonic
1	Main telephone lines (including ISDN) per capita	MTLPC
2	Main telephone lines (including ISDN) per capita – home	MTLHPC
3	Main telephone lines (including ISDN) per capita – company	MTLCPC
4	Standard telephone lines per capita	STLPC
5	Standard telephone lines per capita – home	STLHPC
6	Standard telephone lines per capita – company	STLCPC
7	Public telephones per capita	PUBPHONESPC

Table 2b.3 provides the correlation matrix for the variables in Table 2b.1, where each of these has been identified by its assigned number. However, as the lowest value of any of the bivariate correlation coefficients was 0.948, only the main telephones lines variable per capita (MTLPC) was retained for the cluster analysis. There were sixteen missing observations for this variable.

Table 2b.3
Communications Correlation Matrix

	1	2	3	4	5	6	7
1		0.998	0.977	1.000	0.998	0.974	0.985
2			0.962	0.998	1.000	0.958	0.986
3				0.977	0.962	0.999	0.953
4					0.998	0.974	0.985
5						0.958	0.986
6							0.948
7							

2(c) GAS, WATER AND SEWERAGE

Data from the 2001 SADB was available for the provision of public utility services and the variables originally considered for inclusion under this head are specified in Table 2c.1. However, GASNET, GASCON and ELECCON each had only 1731 observations and appeared only to be specified for urban areas. These variables were therefore omitted from further consideration.[2]

Table 2c.1
Public Utilities Cluster Variables

Variable description	Mnemonic
Water line distribution network, km^2	WLNET
Water line connections leading to residential buildings and residences for communities	WLCON
Sewerage distribution network, km^2	SEWNET
Sewerage connections leading to residential buildings and residences for communities	SEWCON
Gas line distribution network in metres	GASNET
Gas line connections to buildings in 'pcs'	GASCON

Table 2c.2 presents the correlation matrix for the four variables considered for inclusion in the utility services component of the infrastructure clustering exercise. As the bivariate correlation between WLNET and WLCON was 0.834 and that between SEWNET and SEWCON was 0.945 only WLNET and SEWNET were retained for the cluster analysis and both were deflated by gmina area – WLNETPSQKM and SEWNETPSQKM.

Table 2c.2
Utility Services Correlation Matrix

	WLNET	WLCON	SEWNET	SEWCON
WLNET		0.834	0.698	0.677
WLCON			0.842	0.866
SEWNET				0.945
SEWCON				

As a result of these preliminary screenings, five variables remained for inclusion in the infrastructure clustering analysis. These variables and their basic descriptive statistics are presented in Table 2.1 below.

Table 2.1
Basic Statistics for Infrastructure

Variable	N	Mean	Standard Deviation	Minimum	Maximum
MTLPC	2489	0.1667	0.0810	0	0.6022
RDHRDPSQKM	2489	0.0047	0.0061	0	0.7033
RDREPPSQKM	2489	0.0002	0.0008	0	0.0304
WLNETPSQKM	2478	0.0097	0.0094	0	0.0644
SEWNETPSQKM	2478	0.0036	0.0083	0	0.0780

Extreme observations (4 eastern regions in bold)

For variables with many observations of 0, only the highest values are reported explicitly.

Variable	Value	Gmina	SADB Code
MTLPC			
Lowest			
	0.00	12 Gminas	
Highest			
	0.5515	Rewal	3205072
	0.5792	Warszawa - Bemowo	1431011
	0.5747	Warszawa - Centrum	1431041
	0.5815	Siedlce – r	1426082
	0.6021	Warszawa – Wlochy	1431181

A total of twelve localities had no main telephone lines, of which nine were in Mazowieckie and ten were rural, of which one was in Podkarpackie. At the same time, the capital region also accounted for four out of the five gminas with the highest densities of lines. Perhaps surprisingly, rural areas occupied two of the places on this list.

Variable	Value	Gmina	SADB Code
RDHRDPSQKM			
Lowest			
	0.00	29 Gminas	
Highest			
	0.0424	Swietochlowice	2476011
	0.0474	Glowno - u	1020011
	0.0481	Buczkowice	2402032
	0.0481	Warszawa – Ursus	1431141
	0.7033	Piastów	1421011

Twenty-nine gminas had no hard roads, of which twenty-five were rural, with the other four being mixed localities. Amongst the total, the greatest concentrations were to be found Opolskie (6), Warminsko-Mazurskie (6) and Zachodniopomorskie (7). Three members on the list were located in the eastern territories. Only one rural gmina was present amongst the areas with the highest hard road densities, while Mazowieckie and Śląskie each housed two of the five.

Variable	Value	Gmina	SADB Code
RDREPPSQKM			
Lowest			
	0.00	1200 Gminas	
Highest			
	0.0076	Pilchowice	2405042
	0.0081	Wieslowies	2405082
	0.0087	Nowy Tomysl	3015043
	0.0137	Podkowa Lesna	1405021
	0.0303	Skoczów	2403103

A total of 1200 gminas underwent no road repairs in 1200, of which seventy per cent were rural communities. The most intense activity was evident in Śląskie, which accounted for three of the five gminas with the most road repairs.

Variable	Value	Gmina	SADB Code
WLNETPSQKM			
Lowest			
	0.00	27 Gminas	
Highest			
	0.0581	Rydultowy	2415031
	0.0596	Puck –u	2211031
	0.0618	Aleksandrów Kujawski -u	0401011
	0.0632	Koscian - u	3011011
	0.0644	Czeladz	2401021

Twenty-seven gminas had no water line distribution network and all of these were rural. Six of these were in Podkarpackie and one in Lubelskie. However, Malopolskie accounted for eight of the observations and Mazowieckie for a further six.

Variable	Value	Gmina	SADB Code
SEWNETPSQKM			
Lowest			
	0.0	469 Gminas	
Highest			
	0.0589	Ketrzyn	2808011
	0.0621	Zabki	1434031
	0.0618	Aleksandrów Kujawski -u	0401011
	0.0658	Ilawa - u	2807011
	0.0780	Piastów	1421011

There was no sewerage distribution network in 469 gminas, of which 96.4 per cent were rural communities. The highest five densities on this variable were all to be

found in urban localities, with two being in Mazowieckie and two in Warminsko-Mazurskie.

Cluster Results

Table 2.2 presents the diagnostic statistics from the preliminary cluster screening. In this case, there is a conflict between the optimal number of clusters indicated by the pseudo F-statistic and that indicated by the CCC. In particular, the former points to a five cluster solution while the latter suggests that rather more would be optimal. Given the conflict, appeal was made to the primacy of the pseudo F-statistic and a five cluster solution was adopted.

Table 2.2
Infrastructure FASTCLUS Results

N	Pseudo F	R^2	Cubic Clustering Criterion
20	756.04	0.7029	84.882
19	757.01	0.6965	82.916
18	754.56	0.6896	80.369
17	778.29	0.6821	81.238
16	783.31	0.6741	79.432
15	800.46	0.6653	79.069
14	835.70	0.6556	80.824
13	882.48	0.6450	83.672
12	929.65	0.6331	86.137
11	853.68	0.6198	73.345
10	913.79	0.6047	76.961
9	970.73	0.5874	79.529
8	1043.06	0.5671	83.139
7	908.73	0.5430	62.104
6	1034.64	0.5135	71.017
5	**1122.45**	**0.4439**	**83.064**
4	1118.91	0.3633	82.767
3	239.74	0.2678	-32.661
2	316.33	0.1507	-14.784

Table 2.3 provides the infrastructure cluster summary, which reveals that eighty-four per cent of all gminas have been allocated to Cluster 2, while Clusters 1 and 5 are extremely small. Given their respective distances from other clusters, Cluster 1 and 5 appear to represent genuine outliers.

Table 2.3
Infrastructure Cluster Summary

	Frequency	RMS Standard Deviation	Max Distance from Seed to Cluster	Nearest Cluster	Distance Between Cluster Centroids
1	5	1.8841	7.0966	3	10.5331
2	2094	0.4751	5.4448	3	2.9400
3	272	0.9512	5.9414	2	2.9400
4	117	1.1683	9.6149	3	3.7269
5	1	.	0	1	26.8353

Table 2.4 presents the statistics for the included variables and Table 2.5 the matrix of cluster and variable means for the chosen five cluster solution, along with their associated rankings. The latter reflect the assumption that higher values on each of the variables included in the analysis represent higher levels of development. The extremely high mean values for Clusters 1 and 5 for repaired road distances again points to their true outlier status.

Table 2.4
Statistics for Infrastructure Variables

Variable	Total STD	Within STD	R-Square	RSQ/(1-RSQ)
MTLPC	1	0.8067	0.3504	0.5393
RDHRDPKM	1	0.6286	0.6055	1.5349
RDREPPKM	1	0.4451	0.8022	4.0568
WLNETPKM	1	0.5591	0.6880	2.2047
SENETPKM	1	0.4761	0.7737	3.4197
OVER-ALL	1	0.5973	0.6438	1.8075

Table 2.5
Infrastructure Cluster Means & Ranks

Cluster	MTLPC	RDHRDPKM	RDREPPKM	WLNETPKM	SENETPKM
1	0.7685 (4)	0.0817 (4)	10.6364 (2)	0.2854 (4)	0.2872 (4)
2	-0.2543 (5)	-0.3198 (5)	-0.0918 (5)	-0.3198 (5)	-0.3080 (5)
3	1.2221 (2)	1.3181 (3)	0.2414 (4)	1.1840 (2)	0.8790 (2)
4	1.6705 (1)	2.6253 (2)	0.3085 (3)	3.1808 (1)	3.7053 (1)
5	0.7774 (3)	3.6438 (1)	37.2251 (1)	0.9684 (3)	0.4232 (3)

Summing the cluster ranks across variables yields the overall rankings produced in Table 2.6. Overall, Cluster 4 appears as the most highly developed in terms of its infrastructure, while Cluster 2, which performs worst on each individual component indicator, is clearly the least advanced.

Table 2.6
Overall Cluster Rankings

Cluster	Score	Rank
1	18	4
2	25	5
3	13	3
4	8	1
5	11	2

Table 2.7 presents the percentage of gminas in each voivodship in the five clusters, while Map 2 highlights the spatial distribution of cluster membership. Cluster 2, which is both the largest and the worst performing group, accounts for more than ninety per cent of all gminas in the eastern regions. Ślaskie is the most favoured in this regard, with only just over half of its communities in this under-developed cluster, compared to the national average of seventeen in twenty. Conversely, the eastern voivodships are under-represented in the most developed Cluster 4, while Pomorskie is the most heavily represented. Overall, the impression to be gained from Map 2 is that the area around Warsaw in Mazowieckie and the Ślaskie region in the south are relatively well endowed with infrastructure.

Table 2.7
Distribution of Gminas Across Infrastructure Clusters by Voivodship (%)

Voivodship	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Dolnoslaskie	0 (0.0)	136 (80.5)	23 (13.6)	10 (5.9)	0 (0.0)
Kujawsko-Pomorskie	0 (0.0)	128 (88.9)	8 (5.6)	8 (5.6)	0 (0.0)
Lubelskie	**0** **(0.0)**	**193** **(90.6)**	**14** **(6.6)**	**6** **(2.8)**	**0** **(0.0)**
Lubuskie	0 (0.0)	77 (92.8)	5 (6.0)	1 (1.2)	0 (0.0)
Lodzkie	0 (0.0)	156 (88.1)	12 (6.8)	9 (5.1)	0 (0.0)
Malopolskie	0 (0.0)	136 (74.7)	42 (23.1)	4 (2.2)	0 (0.0)
Mazowieckie	2 (0.6)	256 (78.8)	46 (14.2)	21 (6.5)	0 (0.0)
Opolskie	0 (0.0)	67 (94.4)	3 (4.2)	1 (1.4)	0 (0.0)
Podkarpackie	**0** **(0.0)**	**142** **(88.8)**	**14** **(8.8)**	**4** **(2.5)**	**0** **(0.0)**
Poldlaskie	**0** **(0.0)**	**108** **(91.5)**	**9** **(7.6)**	**1** **(0.8)**	**0** **(0.0)**
Pomorskie	0 (0.0)	101 (82.1)	8 (6.5)	14 (11.4)	0 (0.0)
Slaskie	2 (1.2)	90 (54.2)	56 (33.7)	17 (10.2)	1 (0.6)
Swietokrzyskie	**0** **(0.0)**	**95** **(93.1)**	**5** **(4.9)**	**2** **(2.0)**	**0** **(0.0)**
Warminsko-Mazurskie	0 (0.0)	100 (86.2)	8 (6.9)	8 (6.9)	0 (0.0)
Wielkopolskie	1 (0.4)	206 (91.2)	9 (4.0)	10 (4.4)	0 (0.0)
Zachodniopomorskie	0 (0.0)	103 (90.4)	10 (8.8)	1 (0.9)	0 (0.0)

Finally, Table 2.8 shows that over ninety-five per cent of rural gminas are to be found in the under-developed Cluster 2 and that they account for almost three-quarters of the observations in that cluster. Very nearly the same proportion of mixed gminas are also to be found in that grouping. Urban communities, on the other hand, account for all of the observations in the most highly developed Cluster 4, in which there are no rural gminas from the eastern regions. Nevertheless, half of all urban gminas are to be found in cluster three, which is here regarded as middle ranking in terms of its attained development profile.

Table 2.8
Infrastructure Clusters by Gmina Type

Gmina Type	**1**	**2**	**3**	**4**	**5**
Urban	1 (20.0)	39 (1.9)	161 (59.2)	117 (100.0)	
Rural	3 (60.0)	1524 (72.8)	68 (25.0)		
Mixed Urban/Rural	1 (20.0)	531 (25.4)	43 (15.8)		1 (100.0)

3. BUSINESS

The variables selected for consideration for the business activity clustering analysis relate to the structure of community industry, with all of the data being collected originally for the official REGON register of Polish enterprise. However, given that errors were uncovered in the register for 2001, the data used here are for the year 2000. Table 3.1 presents a listing of the variables extracted originally. It was assumed that high values on each of these measures except PRPUB, PRSOE, PRAG and PRPUBMAN represented higher levels of development

Table 3.1
Business Cluster Variables

Variable Number	Variable description	Mnemonic
1	Total number of REGON entities per capita	TOTREGPC
2	Proportion of REGON entities in the public sector	PRPUB
3	Proportion of REGON entities that are SOEs	PRSOE
4	Proportion of REGON entities in the private sector	PRPRI
5	Proportion of REGON entities with foreign capital participation	PRPFC
6	Proportion of REGON entities in agriculture	PRAG
7	Proportion of REGON entities in manufacturing	PRMAN
8	Proportion of REGON entities in public sector manufacturing	PRPUBMAN
9	Proportion of REGON entities in private sector manufacturing	PRPRIMAN
10	Proportion of REGON entities in financial intermediation	PRFIN
11	Proportion of REGON entities in real estate, renting and business activities	PRREAL

Table 3.2 provides the correlation matrix for these initial eleven variables in which each has been identified by its reference number from Table 3.1. Due to the high degree of correlation between many of the variables, only TOTREGPC, PRPUB, PRAG, PRPFC and PRFIN were retained for the clustering exercise.

Table 3.2
Business Variables: Correlation Matrix

	1	2	3	4	5	6	7	8	9	10	11
1		0.261	0.330	0.308	0.183	0.260	0.307	0.265	0.307	0.291	0.264
2			0.815	0.856	0.611	0.406	0.828	0.779	0.828	0.865	0.879
3				0.942	0.754	0.423	0.906	0.894	0.905	0.945	0.904
4					0.826	0.463	0.971	0.929	0.971	0.992	0.984
5						0.352	0.773	0.750	0.773	0.780	0.865
6							0.452	0.430	0.452	0.450	0.449
7								0.931	1.000	0.965	0.940
8									0.930	0.918	0.898
9										0.965	0.939
10											0.969
11											

Table 3.3 provides the summary statistics for the variables finally included in the business cluster analysis.

Table 3.3
Basic Statistics for Business

Variable	N	Mean	Standard Deviation	Minimum	Maximum
TOTREGPC	2489	0.0587	0.0300	0.0171	0.4482
PRPUB	2489	0.0464	0.0231	0.0057	0.1860
PRAG	2489	0.0814	0.0815	0	0.8041
PRPFC	2489	0.0068	0.0095	0	0.1516
PRFIN	2489	0.0241	0.0123	0	0.1053

Extreme observations (4 eastern regions in bold)

For variables with many observations of 0, only the highest values are identified explicitly.

Variable	Value	Gmina	SADB Code
TOTREGPC			
Lowest			
	0.0171	Górow Ilaweckie – r	2801052
	0.0185	**Bejsce**	**2603012**
	0.0193	**Hrubieszów – r**	**0604042**
	0.0196	**Kraniczyn**	**0606062**
	0.0200	**Potok Górny**	**0602112**

Highest			
	0.2994	Jastarnia	2211021
	0.3210	Leba	2208021
	0.3443	Mielno	3209052
	0.3887	Rewal	3205072
	0.4482	Krynica Morska	2210011

The lowest five values of REGON enterprises per head all occur in rural gminas and three of these are in the eastern voivodship of Lubelskie. On the other hand, three of the five highest occur in urban areas in Pomorskie, although the remaining two are in rural gminas in Zachodniopomorskie, a western border region in the upper half of Polish voivodships by GDP per capita.

Variable	Value	Gmina	SADB Code
PRPUB			
Lowest			
	0.0057	Warszawa – Targówek	1431141
	0.0060	Raszyn	1421062
	0.0061	Warszawa – Ursynów	1431151
	0.0062	Kolobrzeg – r	3208042
	0.0066	Wladyslawowo	2211041
Highest			
	0.1479	Tolkmicko	2804093
	0.1667	Nowy Dwór	2011062
	0.1706	Zloty Stok	0224073
	0.1771	Medrzechów	1204042
	0.1860	**Bejsce**	**2603012**

Recalling that large concentrations of public sector enterprises are here regarded as a negative development indicator, it can be seen that, perhaps surprisingly, the most important of these in per capita terms occur outside urban communities, with the largest of all being found in the eastern voivodship of Świętokrzyskie. Three of the five gminas with the smallest concentrations of public sector enterprises per head appear in the capital region of Mazowieckie, which is by far the wealthiest in the country, as measured by GDP per capita

Variable	Value	Gmina	SADB Code
PRAG			
Lowest			
	0	Jastarnia	2211021
	0	Hel	2211011
	0	**Gawluszowice**	**1811042**
	0	Rusinów	1423072
	0	Leknica	0811011

Highest			
	0.6096	Sypniewo	1411092
	0.6232	Krzynowloga	1422062
	0.6630	Przasnysz – r	1422072
	0.7930	Czernice Borowe	1422032
	0.8041	Dzierzgowo	1413022

Unsurprisingly, the highest values of PRAG occur in rural gminas but, rather less predictably, all of these gminas are located in Mazowieckie. At the opposite end of the spectrum, five gminas had no agricultural enterprises on their REGON registers and two of these were actually in rural communities, with one in the eastern voivodship of Podkarpackie.

Variable	Value	Gmina	SADB Code
PRPFC			
Lowest			
	0.00	770 Gminas	
Highest			
	0.0712	Lesznowola	1418032
	0.0742	Warszawa – Wilanów	1431171
	0.0755	Warszawa – Wlochy	1431181
	0.0791	Raszyn	1421062
	0.1516	Leknica	0811011

Thirty per cent of Polish communities have no enterprises with foreign capital participation and more than four in ten of these gminas are located in the four eastern voivodships. In order to place the second of these statistics in context, it might be noted that the highlighted eastern regions account for one-quarter of the total number of voivodships and gminas, but only eighteen per cent of the total population of the country. It would appear that, in general, foreign capital shuns the countryside, insofar as ninety per cent of the zero values represent rural gminas. However, it needs to be noted that two of the localities with the greatest foreign capital inputs are rural, although four of the five listed above lie in the capital region of Mazowieckie.

Variable	Value	Gmina	SADB Code
PRFIN			
Lowest			
	0.00	47 Gminas	
Highest			
	0.0787	**Szypliszki**	**2012082**
	0.0844	Dabie	3009043
	0.0901	**Terespol – u**	**0601021**
	0.0914	**Cieszanów**	**1809023**
	0.1053	**Slawatycze**	**0601142**

Rather surprisingly, two of the gminas with the highest penetration of financial intermediaries are rural and both are to be found in the east. Indeed, two of the remaining three communities with relatively large financial sectors also lie in the east. Forty-seven localities have no financial intermediation and 95.7 per cent of these are rural. Of the latter, eighteen are in the eastern voivodships, with nine being in Podkarpackie and seven in Podlaskie. More surprisingly, Mazowieckie accounts for a further eight of the areas with no financial enterprises.

Cluster Results

Table 3.4 presents the diagnostic statistics from the preliminary cluster screening. Both the pseudo-F test and the CCC point to the optimality of a two cluster classification of the data.

Table 3.4
Business Fastclus Results

N	Pseudo F	R^2	Cubic Clustering Criterion
20	394.92	0.7029	21.936
19	404.90	0.6965	22.082
18	399.94	0.6896	18.745
17	415.22	0.6821	19.796
16	429.11	0.6741	20.311
15	433.01	0.6653	18.502
14	424.87	0.6556	13.984
13	441.97	0.6450	14.574
12	441.37	0.6331	11.245
11	474.55	0.6198	14.374
10	483.96	0.6047	12.386
9	489.92	0.5874	9.361
8	500.09	0.5671	6.657
7	452.42	0.5430	-7.217
6	406.80	0.5135	-21.386
5	564.48	0.4439	11.142
4	516.73	0.3633	6.832
3	513.26	0.2668	8.209
2	**643.90**	**0.1507**	**22.694**

Table 3.5 presents the business cluster summary, with the observations divided in the ratio 3:1 in favour of the first group.

Table 3.5
Business Cluster Summary

	Frequency	RMS Standard Deviation	Max Distance from Seed to Cluster	Nearest Cluster	Distance Between Cluster Centroids
1	1843	0.8503	9.7006	2	2.3127
2	646	0.9997	14.5062	1	2.3127

Table 3.6 provides the statistics for the included variables, while Table 3.7 gives the matrix of cluster and variable means for the final solution, along with their associated rankings. From the latter of these, it can be seen that the gminas in Cluster 2 are the more developed under current assumptions insofar as they have more REGON companies per capita, a lower incidence of public ownership and agriculture and a higher incidence of companies with foreign capital participation and companies engaged in financial intermediation.

Table 3.6
Statistics for Business Variables

Variable	Total STD	Within STD	R-Square	RSQ/(1-RSQ)
TOTREGPC	1	0.7748	0.3999	0.6665
PRPUB	1	0.8807	0.2247	0.2898
PRAG	1	0.9439	0.1095	0.1229
PRPFC	1	0.8661	0.2501	0.3336
PRFIN	1	0.9779	0.0441	0.0461
OVER-ALL	1	0.8914	0.2057	0.2589

Table 3.7
Business Cluster Means & Ranks

Cluster	TOTREGPC	PRPUB	PRAG	PRPFC	PRFIN
1	-0.3743 (2)	0.2806 (2)	0.1959 (2)	-0.296 (2)	-0.1244 (2)
2	1.0679 (1)	-0.8005 (1)	-0.5588 (1)	0.844 (1)	0.3545 (1)

Table 3.8 presents the proportion of gminas in each voivodship falling into each cluster, while Map 3 depicts the spatial distributions of each cluster's membership. The latter reveals a concentration of Cluster 2 gminas in the North East of the country, around Białystok – Wasilków, Supraśl and Juchowiec Koscielny. The former reveals that the eastern regions have the least well developed business environments of all, with very few of their gminas falling into Cluster 2. The western border territories of Lubuskie and Zachodniopomorskie, on the other hand, perform best on this score.

Table 3.8
Distribution of Gminas Across Business Clusters by Voivodship (%)

Voivodship	Cluster 1	Cluster 2
Dolnoslaskie	87 (51.5)	82 (48.5)
Kujawsko-Pomorskie	122 (84.7)	22 (15.3)
Lubelskie	**202** **(94.8)**	**11** **(5.2)**
Lubuskie	28 (33.7)	55 (66.3)
Lodzkie	141 (79.7)	36 (20.3)
Malopolskie	140 (76.9)	42 (23.1)
Mazowieckie	244 (75.1)	81 (24.9)
Opolskie	52 (73.2)	19 (26.8)
Podkarpackie	**146** **(91.3)**	**14** **(8.8)**
Podlaskie	**105** **(89.0)**	**13** **(11.0)**
Pomorskie	78 (63.4)	45 (36.6)
Slaskie	95 (57.2)	71 (42.8)
Swietokrzyskie	**94** **(92.2)**	**8** **(7.8)**
Warminsko-Mazurskie	103 (88.8)	13 (11.2)
Wielkopolskie	158 (69.9)	68 (30.1)
Zachodniopomorskie	48 (42.1)	66 (57.9)

Finally, Table 3.9 provides the distribution of cluster membership by gmina type. This indicates that whereas eighty per cent of urban gminas are in Cluster 2, only one-third of mixed urban/rural gminas are in this cluster. More telling still, just 12.3% of rural gminas are so allocated. Of the forty-six Cluster 2 gminas in the highlighted eastern voivodships, only two - Juchnowiec Koscielny (SADB code 2002052) and Sitkówka-Nowiny (2604172) - are rural.

Table 3.9
Business clusters by gmina type

Gmina Type	**1**	**2**
Urban	61 (3.3)	257 (39.8)
Rural	1399 (75.9)	196 (30.3)
Mixed Urban/Rural	383 (20.8)	193 (29.9)

4. AMENITIES

The variables selected for the amenities clustering analysis reflect those noted in a recent MARD document (MARD, 2004) on rural development and are presented in Table 4.1 below. It is assumed that higher values on each of these measures are indicative of greater levels of development

Table 4.1
Amenities Cluster Variables

Variable Number	Variable description	Mnemonic
1	Museums per capita	MUSPC
2	Cinemas per capita	CINPC
3	Hospital beds per capita	BEDPC
4	Libraries per capita	LIBPC
5	Shops per capita	SHOPPC
6	Nursery school places per capita	NURSPC
7	Pharmacies per capita	PHARMPC
8	General secondary school places per capita	GENSECPC

Other variables were examined, but they were highly co-linear with variables already included and so were not considered further. For example, thought was given to the inclusion of the relative size of cinema audiences, but this was highly correlated with the number of cinemas and so was not retained for the analysis. The correlation matrix for the eight variables finally included in the work, identified by the numbers allocated to them in Table 4.1, is presented as Table 4.2.

Table 4.2
Correlation Matrix

	1	2	3	4	5	6	7	8
1		0.212	0.144	-0.740	0.186	0.118	0.112	0.142
2			0.238	-0.143	0.449	0.185	0.219	0.260
3				-0.305	0.492	0.282	0.320	0.319
4					-0.374	-0.222	-0.253	-0.314
5						0.393	0.510	0.478
6							0.289	0.2659
7								0.343
8								

Table 4.3 provides the summary statistics for variables included in the amenities cluster analysis.

Table 4.3
Basic Statistics for Amenities

Variable	N	Mean	Standard Deviation	Minimum	Maximum
MUSPC	2489	0.00001	0.00004	0	0.0005
CINPC	2489	0.00001	0.00003	0	0.0006
BEDPC	2489	0.0018	0.0048	0	0.0602
LIBPC	2489	0.0004	0.0002	0	0.0012
SHOPPC	2489	0.0094	0.0042	0.001	0.0542
NURSPC	2489	0.0134	0.0151	0	0.1343
PHARMPC	2489	0.0002	0.0001	0	0.0009
GENSECPC	2489	0.00003	0.0001	0	0.0005

Extreme observations (4 eastern regions in bold)

Given the high incidence of zero values for the amenities variables only the highest values are reported explicitly in what follows

Variable	Value	Gmina	SADB Code
MUSPC			
Lowest			
	0.00	2085 Gminas	
Highest			
	0.0004	Lubowo	3003062
	0.0004	Swidnica	0809072
	0.0004	**Bialowieza**	**2005022**
	0.0004	Kazimierz Dolny	0614043
	0.0005	Smoldzino	2212092

The densest concentrations of museums all occur in non-urban localities and one of these – Bialowieza – is located in the eastern voivodship of Podlaskie. However, 2085 gminas have no museum provision at all and nearly three-quarters are rural territories.

Variable	Value	Gmina	SADB Code
CINPC			
Lowest			
	0.00	2025 Gminas	
Highest			
	0.0003	**Drohiczyn**	**2010023**
	0.0003	Ustronie Morskie	3208072
	0.0004	**Bialowieza**	**2005022**
	0.0004	Mielno	3209052
	0.0006	Rewal	3205072

Once again, the highest values of cinemas per capita occur in non-urban gminas, three of which are in the western border region of Zachodniopomorskie and two in the eastern voivodship of Podlaskie. At the same time, over two thousand localities have no cinema and three-quarters of these are in country areas.

Variable	Value	Gmina	SADB Code
BEDPC			
Lowest			
	0.00	2022 Gminas	
Highest			
	0.0363	Gistynin – u	1404011
	0.0382	Radziejów	0411011
	0.0395	Phszczykowo	3021021
	0.0524	Wilkowice	2402102
	0.0602	Karpacz	0206011

Over eighty per cent of communities have no hospital bed provision and the large majority of these places are rural. With one exception, the greatest numbers of hospital beds per head of population are to be found in urban locatities. This finding reflects a long recognised imblance between levels of health care provision between Polish towns and the countryside.

Variable	Value	Gmina	SADB Code
PHARMPC			
Lowest			
	0.0	462 Gminas	
Highest			
	0.0011	Skórcz – u	2213021
	0.0011	Laskarzew	1403021
	0.0011	Stoczek Lukowski u	0611021
	0.0011	Krynica Morska	2210011
	0.0012	Kosakowo	2211052

Once again, four out of the five highest values of a health related variable – pharmacies per capita – are to be found in urban communities and two of these, along with the sole rural district, are to be found in Pomorskie. Nevertheless, 462 gminas have no pharmacy and 97.6 per cent of these are rural.

Variable	Value	Gmina	SADB Code
LIBPC			
Lowest			
	0.00	16 Gminas	
Highest			
	0.0322	Slawatycze	0601142
	0.0326	Zukowice	0203062
	0.0450	Platerówka	0210062
	0.0462	Siemysl	3208062
	0.0542	**Dynów – r**	**1816052**

In contrast to most of the other amenity measures, only sixteen gminas had no libraries, but all of these were rural, with four being in the east, although Mazowieckie alone accounted for five. The highest values for this variable were all to be found in rural localities, two in Dolnośląskie and one, the highest of all, in the eastern region of Podkarpackie.

Variable	Value	Gmina	SADB Code
SHOPPC			
Lowest			
	0.0012	**Sejny – r**	**2009052**
	0.0025	Boranów	1405032
	0.0025	**Nowy Dwór**	**2011062**
	0.0026	**Jeleniewo**	**2012032**
	0.0028	**Koln0 – r**	**2006032**

Highest			
	0.0484	Tomaszów Lubelski –u	0618011
	0.0484	Mielno	3209052
	0.0485	Rewal	3205072
	0.0485	Krynica Morska	2210011
	0.1343	Leba	2208021

All of the five lowest shop densities are to be found in rural localities, with four of these lying in the eastern region of Podlaskie. On the other hand, the highest values on this indicator are divided between rural and urban localities, although two are to be found in Pomorskie and two in Zachodniopomorskie.

Variable	Value	Gmina	SADB Code
NURSPC			
Lowest			
	0.00	526 Gminas	
Highest			
	0.0006	Lubrza	1610032
	0.0006	**Lutowiska**	**1801052**
	0.0007	Krupski Mlyn	2413052
	0.0008	Górzow Slaski	1608023
	0.0008	Tolkmicko	2804093

Over twenty per cent of communities do not possess any nursery school places, of which 519 are rural gminas and the remaining seven mixed. Perhaps surprisingly, non-urban localities account for all of the five highest values on this indicator and one of these is located in the eastern voivodship of Podkarpackie.

Variable	Value	Gmina	SADB Code
GENSECPC			
Lowest			
	0.00	1643 Gminas	
Highest			
	0.0004	Stoczek Lukowski – u	0611021
	0.0004	Duszniki Zdrój	0208011
	0.0004	Raciaz – u	1420021
	0.0004	Gorzów Ilaweckie – u	2801021
	0.0005	Nieszawa	0401031

Two-thirds of gminas have no general secondary school places within their borders and, once again, the vast majority of these (87.1%) are rural and only twenty-five urban. Those with the greatest number in proportionate terms are all located within urban localities and none of these lies in the eastern regions.

Cluster Results

Table 4.4 contains the diagnostic statistics generated by the analyses pre-specifying from two to twenty eventual clusters of gminas on the amenities measures. Both the pseudo-F test and the CCC point to the optimality of a two cluster solution.

Table 4.4
Amenities Fastclus Results

N	Pseudo F	R^2	Cubic Clustering Criterion
20	269.96	0.5342	76.062
19	273.78	0.5278	74.282
18	249.71	0.5210	57.427
17	288.46	0.5137	73.567
16	278.43	0.5059	64.064
15	273.96	0.4975	57.099
14	312.64	0.4883	70.956
13	299.23	0.4783	59.213
12	307.35	0.4673	57.254
11	277.56	0.4452	36.900
10	300.82	0.4416	41.281
9	292.97	0.4262	30.406
8	317.21	0.3952	38.646
7	357.61	0.3609	51.311
6	343.84	0.3224	41.892
5	495.50	0.2786	84.697
4	518.68	0.2272	81.573
3	550.28	0.1679	77.008
2	**816.06**	**0.0945**	**109.059**

Table 4.5 provides the final amenities cluster summary, with the ratio of observations falling in Cluster 2 relative to Cluster 1 being roughly 4:1.

Table 4.5
Amenities Cluster Summary

	Frequency	RMS Standard Deviation	Max Distance from Seed to Cluster	Nearest Cluster	Distance Between Cluster Centroids
1	543	1.2079	19.7714	2	3.4034
2	1945	0.7460	9.1500	1	3.4034

Table 4.6 provides the statistics for the included variables and Table 4.7 presents the matrix of cluster and variable means and their rankings. The latter shows clearly the superiority of the much smaller Cluster 1 on all counts except libraries per head.

Table 4.6
Statistics for Amenities Variables

Variable	Total STD	Within STD	R-Square	RSQ/(1-RSQ)
MUSPC	1.0000	0.9735	0.0528	0.0557
CINPC	1.0000	0.8775	0.2303	0.2992
BEDPC	1.0000	0.7742	0.4009	0.6692
LIBPC	1.0000	0.9009	0.1888	0.2327
SHOPPC	1.0000	0.7396	0.4532	0.8290
NURSPC	1.0000	0.9227	0.1491	0.1752
PHARMPC	1.0000	0.8837	0.2194	0.2811
GENSECPC	1.0000	0.8476	0.2820	0.3926
OVER-ALL	1.0000	0.8679	0.2471	0.3281

Table 4.7
Amenities Cluster Means & Ranks

Cluster	MUSPC	CINPC	BEDPC	LIBPC	SHOPPC	NURSPC	PHARMPC	GENSECPC
1	0.435 (1)	0.908 (1)	1.198 (1)	-0.822 (2)	1.274 (1)	0.731 (1)	0.887 (1)	1.005 (1)
2	-0.121 (2)	-0.253 (2)	-0.334 (2)	0.229 (1)	-0.356 (2)	-0.204 (2)	-0.247 (2)	-0.280 (2)

The allocation of the communities within voivodships to the final clusters is given in Table 4.8, while Map 4 depicts the membership of the two groups in space. Only Łódzkie has a smaller proportion of its communities in the best performing Cluster 1 than the eastern regions of Lubelskie, Podkarpackie and Podlaskie.

Table 4.8
Distribution of Gminas Across Amenities Clusters by Voivodship (%)

Voivodship	Cluster 1	Cluster 2
Dolnoslaskie	45 (26.6)	124 (73.4)
Kujawsko-Pomorskie	31 (21.5)	113 (78.5)
Lubelskie	**38** **(17.8)**	**175** **(82.2)**
Lubuskie	24 (28.9)	59 (71.1)
Lodzkie	30 (16.9)	147 (83.1)
Malopolskie	39 (21.4)	143 (78.6)
Mazowieckie	60 (18.5)	265 (81.5)
Opolskie	16 (22.5)	55 (77.5)
Podkarpackie	**26** **(16.3)**	**134** **(83.8)**
Poldlaskie	**21** **(17.8)**	**97** **(82.2)**
Pomorskie	31 (25.2)	92 (74.8)
Slaskie	48 (28.9)	118 (71.1)
Swietokrzyskie	**22** **(21.6)**	**80** **(78.4)**
Warminsko-Mazurskie	32 (27.6)	84 (72.4)
Wielkopolskie	48 (21.2)	178 (78.8)
Zachodniopomorskie	32 (28.1)	82 (71.9)

Finally, Table 4.9 provides the distribution of cluster membership by gmina type. Only thirteen per cent of urban gminas fall into Cluster 2 compared to ninety-seven per cent of rural communities. Mixed gminas, on the other hand, are more evenly divided between the two groupings. Of the forty-six rural localities classified in Cluster 1, ten are to be found in the eastern voivodships, with half of them being in Lubelskie.

Table 4.9
Amenities Clusters by Gmina Type

Gmina Type	1	2
Urban	276 (50.8)	42 (2.2)
Rural	46 (8.5)	1549 (79.6)
Mixed Urban/Rural	221 (40.7)	355 (18.2)

5. AGRICULTURE

The variables selected for the agriculture cluster relate to the extent and use of land for agricultural purposes, all expressed per hectare of total gmina area, and these are reported in Table 5.1. For current purposes, it is assumed that less developed regions possess higher concentrations of each of these.

Table 5.1
Agriculture Cluster Variables

Variable description	Mnemonic
Total agricultural land	DTOT
Private agricultural land	DPRI
Total arable land	DARA
Total orchards	DORC
Total meadows	DMEA
Total pastures	DPAS

Table 5.2 presents the correlation matrix for these variables. This highlights the strong inter-relationship between DTOT, DPRI and DARA. In consequence, only DPRI from these three measures was retained for the clustering exercise.

Table 5.2
Correlation Matrix for Agriculture

	DTOT	DPRI	DARA	DORC	DMEA	DPAS
DTOT		0.899	0.920	0.169	0.160	0.061
DPRI			0.801	0.194	0.205	0.072
DARA				0.061	-0.172	-0.201
DORC					-0.099	-0.067
DMEA						0.324
DPAS						

Table 5.3 provides the summary statistics for the four variables included in the final agricultural clustering analysis.

Table 5.3
Basic Statistics for Agriculture

Variable	N	Mean	Standard Deviation	Minimum	Maximum
DPRI	2489	0.5202	0.2032	0	0.9377
DORC	2489	0.0109	0.0284	0	0.5576
DMEA	2489	0.0831	0.0547	0	0.5800
DPAS	2489	0.0456	0.0338	0	0.2705

Extreme observations (4 eastern regions in bold)

In those instances for which many observations take a zero value, only the highest values are reported explicitly.

Variable	Value	Gmina	SADB Code
DPRI			
Lowest			
	0.0000	Jastarnia	2211021
	0.0015	Legionowo	1408011
	0.0033	Hel	2211011
	0.0076	Zielonka	1434041
	0.0084	Krynica Morska	2210011
Highest			
	0.9203	Goszczyn	1406042
	0.9308	Palecznica	1214042
	0.9310	**Czarnocin**	**2603022**
	0.9350	Badkowo	0401052
	0.9377	Dabrowice	1002032

Unsurprisingly, the gminas with the smallest land areas devoted to private agriculture, none of which are in the eastern voivodships, are all urban. Three of these are in Pomorskie and the remaining two Mazowieckie. Likewise, the five highest values occur in rural communities, with one of these being in Świętokrzyskie.

Variable	Value	Gmina	SADB Code
DORC			
Lowest			
	0.00	40 Gminas	
Highest			
	0.3743	Warka	1406113
	0.3813	**Obrazów**	**2609062**
	0.4213	Goszczyn	1406042
	0.4654	Bledów	1406022
	0.5576	Belsk Duzy	1406012

Of the forty gminas having no land devoted to orchards, sixty per cent are urban, although a total of fiteen are actually rural. Out of the total, the greatest concentration – eight gminas – is to be found in Malopolskie, while the eastern account for three. Four of the five highest values are to be found in Mazowieckie and the remaining one in the eastern region of Świętokrzyskie.

Variable	Value	Gmina	SADB Code
DMEA			
Lowest			
	0.0000	Sopot	2264011
	0.0000	Warszawa – Ursus	1431141
	0.0000	Warszawa – Bernowo	1431011
	0.0000	Legionowo	1408011
	0.0005	Hel	2211011
Highest			
	0.3783	Poronin	1217052
	0.4093	Zwierzyn	0806052
	0.4196	Czarny Dunajec	1211032
	0.5209	Szaflary	1211142
	0.5800	Bialy Dunajec	1217022

The five lowest acreages devoted to meadows are to be found in urban areas, three in Mazowieckie and two in Pomorskie. Conversely, the five highest concentrations are located in rural gminas, with four of these being situated in Malopoolskie.

Variable	Value	Gmina	SADB Code
DPAS			
Lowest			
	0.0000	Mszana	2415092
	0.0000	Hel	2211011
	0.0000	Warszawa – Ursus	1431141
	0.0015	Legionowo	1408011
	0.0002	Goszczyn	1406042
Highest			
	0.1914	Stare Bogaczowice	0221072
	0.1916	Walim	0221082
	0.1918	Jedina-Zdrój	0221021
	0.2075	Wilczeta	2802072
	0.2706	Lewin Klodzki	0208092

The picture painted by the distribution of pasture land is a little more varied. In particular, two of the lowest five densities are in rural communities and one of the highest five is in an urban area. The capital region of Mazowieckie houses three of the lowest observations, while Dolnośląskie contains four of the highest.

Cluster Results

Table 5.4 contains the diagnostic statistics obtained from the initial cluster screening exercise. In this case, the pseudo-F statistic and the CCC point globally to rather different solutions, with the former suggesting that four clusters is optimal and the latter pointing to eighteen. Following the rule that the pseudo-F takes precedence leads to the first of these conclusions, which in any event is locally optimal under the CCC.

Table 5.4
Agriculture Fastclus Results

N	Pseudo F	R^2	Cubic Clustering Criterion
20	639.97	0.7798	25
19	644.66	0.7739	24.273
18	662.08	0.7675	24.751
17	665.95	0.7606	23.492
16	654.90	0.7530	20.345
15	639.24	0.7447	16.508
14	712.05	0.7355	23.022
13	706.49	0.7253	20.188
12	747.18	0.7139	22.331
11	696.14	0.7009	14.081
10	690.48	0.6860	10.655
9	744.65	0.6688	13.632
8	749.72	0.6484	10.790
7	868.27	0.6238	19.035
6	898.41	0.5933	17.540
5	668.58	0.5541	-11.008
4	**913.95**	**0.4536**	**21.897**
3	778.81	0.3343	14.717
2	467.40	0.1882	-9.426

Table 5.5 provides the agriculture cluster summary and it will be noted that Cluster 3 has only twenty-four members. However, it would appear that this is a genuine outlier collection insofar as its mean lies a relatively large distance from that of any other group.

Table 5.5
Agriculture Cluster Summary

	Frequency	RMS Standard Deviation	Max Distance from Seed to Cluster	Nearest Cluster	Distance Between Cluster Centroids
1	691	0.8616	8.1661	2	2.0618
2	956	0.5797	4.1714	4	1.8519
3	24	2.1891	11.3151	2	7.9541
4	818	0.5414	4.4177	2	1.8519

Table 5.6 provides the statistics for the included variables and Table 5.7 the matrix of cluster and variable means for the four group solution, along with their respective rankings. This shows Cluster 3 to have the greatest concentrations of private agriculture and orchards, which is possibly indicative of a development problem. Cluster 4, on the other hand, records the lowest values for three of the four variables included in the analysis. The characteristic features of Cluster 1 are the high concentrations of meadow and pasture land, while cluster two represents what might be seen as an average grouping in the current context.

Table 5.6
Statistics for Agriculture Variables

Variable	Total STD	Within STD	R-Square	RSQ/(1-RSQ)
DPRI	1.0000	0.6114	0.6266	1.6782
DORC	1.0000	0.6031	0.6367	1.7523
DMEA	1.0000	0.7857	0.3835	0.6221
DPAS	1.0000	0.7411	0.4515	0.8231
OVER-ALL	1.0000	0.6899	0.5246	1.1034

Table 5.7
Agriculture Cluster Means & Ranks

Cluster	**DPRI**	**DORC**	**DMEA**	**DPAS**
1	0.2489 (2)	-0.1535 (2)	0.9925 (4)	1.0815 (4)
2	0.7333 (3)	0.0641 (3)	-0.3058 (3)	0.4296 (3)
3	0.9447 (4)	8.0034 (4)	-0.3650 (2)	0.0030 (1)
4	-1.0950 (1)	-0.1801 (1)	-0.4703 (1)	0.4116 (2)

Table 5.8 provides the results obtained from summing each cluster's rank across the included variables. Cluster 4 emerges clearly as the most developed grouping of gminas in terms of its agricultural complexion

Table 5.8
Overall Cluster Rankings

Cluster	Score	Rank
1	12	3
2	12	3
3	11	2
4	5	1

Table 5.9 presents the proportion of gminas in each voivodship falling into each of the four clusters, while Map 5 depicts the spatial distributions of each cluster's membership. Lubelskie has the lowest proportion of its gminas in Cluster 4, although it is followed by Łódzkie and Mazowieckie, with the other eastern regions occupying the next three lowest positions. The western border voivodships of Lubuskie and Zachodniopomorskie are by far the most advantaged in this regard. Nevertheless, communities in the eastern regions are not absent from this developed grouping.

Table 5.9
Distribution of Gminas Across Agriculture Clusters by Voivodship (%)

Voivodship	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Dolnoslaskie	50 (29.6)	53 (31.4)	0 (0.0)	66 (39.1)
Kujawsko-Pomorskie	7 (4.9)	100 (69.4)	0 (0.0)	37 (25.7)
Lubelskie	**52** **(24.4)**	**136** **(63.8)**	**3** **(1.4)**	**22** **(10.3)**
Lubuskie	7 (8.4)	1 (1.2)	0 (0.0)	75 (90.4)
Lodzkie	35 (19.8)	117 (66.1)	2 (1.1)	23 (13.0)
Malopolskie	80 (44.0)	58 (31.9)	5 (2.7)	39 (21.4)
Mazowieckie	115 (35.4)	143 (44.0)	12 (3.7)	55 (16.9)
Opolskie	1 (1.4)	25 (35.2)	0 (0.0)	45 (63.4)
Podkarpackie	**84** **(52.5)**	**43** **(26.9)**	**0** **(0.0)**	**33** **(20.6)**
Poldlaskie	**82** **(69.5)**	**18** **(15.3)**	**0** **(0.0)**	**18** **(15.3)**
Pomorskie	16 (13.0)	26 (21.1)	0 (0.0)	81 (65.9)
Slaskie	38 (22.9)	39 (23.5)	0 (0.0)	89 (53.6)
Swietokrzyskie	**20** **(19.6)**	**57** **(55.9)**	**2** **(2.0)**	**23** **(22.5)**
Warminsko-Mazurskie	64 (55.2)	12 (10.3)	0 (0.0)	40 (34.5)
Wielkopolskie	25 (11.1)	122 (54.0)	0 (0.0)	79 (35.0)
Zachodniopomorskie	15 (13.2)	6 (5.3)	0 (0.0)	93 (81.6)

Table 5.10 provides the distribution of cluster membership by gmina type. While seventy per cent of urban communities are concentrated in Cluster 4, only one in five rural localities are members of this group. Nevertheless, ten gminas in the eastern region of Lubelskie, twenty-one in Podkarpackie, twelve in Podlaskie and another twelve in Świętokrzyskie are members of this cluster. Finally, no rural gminas are allocated to cluster three.

Table 5.10
Agriculture Clusters by Gmina Type

Gmina Type	1	2	3	4
Urban	49 (7.1)	52 (5.4)	0 (0.0)	217 (26.5)
Rural	504 (72.9)	709 (74.2)	19 (79.2)	363 (44.4)
Mixed Urban/Rural	138 (20.0)	195 (20.4)	5 (20.8)	238 (29.1)

6. FINANCE

The original variables selected in order to examine the health of gmina finances, both of which were expressed in per capita terms, are detailed in Table 6.1. However, the correlation coefficient between these two variables was 0.95 and therefore only GMREV was retained for analysis.

Table 6.1
Gmina Finance Variables

Variable description	Mnemonic
Revenue of gmina budget	GMREV
Personal income tax revenue	PIT

With only one variable available it is not possible to conduct a meaningful cluster analysis and an alternative procedure was adopted for the purposes of mapping the ensuing distribution. Before detailing the adopted scheme, however, the basic statistics for gmina revenues and the location of outliers will be presented.

Table 6.2 presents the descriptive statistics for gmina revenues.

Table 6.1
Basic Statistics for Finance

Variable	N	Mean	Standard Deviation	Minimum	Maximum
GMREV	2489	1299.6649	761.1785	834.082	35120.440

Extreme observations (4 eastern regions in bold)

Variable	Value	Gmina	SADB Code
GMREV			
Lowest			
	834.082	**Krasnik - r**	**0607052**
	834.700	**Tomaszów Lubelskie – r**	**0608112**
	843.856	Koniecpol	2404063
	845.020	Slaboszów	1208072
	858.713	Plonsk – r	1420092

Highest

4131.70	Jerzmanowa	0203032
4147.44	**Puchaczów**	**0610052**
4764.25	Warszawa – Wilanów	1431171
4908.64	Rewal	3205072
35120.44	Kleszczów	1001042

Four of the five gminas with the lowest revenues are rural, with the remaining observation being a mixed urban/rural locality. Two of these localities lie in the east, both in Lubelskie. On the other hand, four of the five richest communities are also rural and one of these is in Lubelskie. It will be noted that the highest observation of all, that of Kleszczów in Lódzkie, is particularly extreme, insofar as it is seven times greater than the next highest figure.[3]

For the purposes of highlighting the distribution of gmina revenues in Map 6, the scheme described in Table 6.2 was adopted. Two general impressions flow from this exercise. First, there are small agglomerations of rich gminas in the south and west of the country, around Warsaw and in the south-east. Second, poorer gminas are concentrated in the east of the country.

Table 6.2
Gmina Income Categories

Income rank (descending order)	Code
Richest 10%	6
11-25%	5
26-50%	4
51-75%	3
76-90%	2
Poorest 10%	1

The general impressions to be gained from Map 6 findings are confirmed by the tables that breakdown the clusters according to voivodship and gmina type. Thus, Table 6.6 indicates that only fourteen per cent of the richest gminas are to be found in the east and only Łódzkie rivals Lubelskie and Świętokrzyskie in its paucity of well endowed gminas. Likewise, the latter two voivodships contain the highest proportions of the poorest communities.

Table 6.6
Distribution of Gminas Across Finance Clusters by Voivodship (%)

Voivodship	Cluster 1 Poorest 10%	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6 Richest 10%
Dolnoslaskie	13 (7.7)	24 (14.2)	33 (19.5)	33 (19.5)	39 (23.1)	27 (16.0)
Kujawsko-Pomorskie	5 (3.5)	17 (11.8)	35 (24.3)	46 (31.9)	29 (20.1)	12 (8.3)
Lubelskie	**54 (25.4)**	**49 (23.0)**	**48 (22.5)**	**36 (16.9)**	**18 (8.5)**	**8 (3.8)**
Lubuskie	1 (1.2)	4 (4.8)	21 (25.3)	26 (31.3)	18 (21.7)	13 (15.7)
Lodzkie	39 (22.0)	41 (23.2)	51 (28.8)	28 (15.8)	10 (5.6)	8 (4.5)
Malopolskie	7 (3.8)	40 (22.0)	57 (31.3)	41 (22.5)	26 (14.3)	11 (6.0)
Mazowieckie	38 (11.7)	58 (17.8)	89 (27.4)	69 (21.2)	37 (11.4)	34 (10.5)
Opolskie	7 (9.9)	12 (16.9)	24 (33.8)	14 (19.7)	8 (11.3)	6 (8.5)
Podkarpackie	**11 (6.9)**	**27 (16.9)**	**56 (35.0)**	**44 (27.5)**	**12 (7.5)**	**10 (6.3)**
Poldlaskie	**11 (9.3)**	**20 (16.9)**	**33 (28.0)**	**28 (23.7)**	**15 (12.7)**	**11 (9.3)**
Pomorskie	1 (0.8)	3 (2.4)	18 (14.6)	46 (37.4)	37 (30.1)	18 (14.6)
Slaskie	15 (9.0)	18 (10.8)	35 (21.1)	38 (22.9)	23 (13.9)	37 (22.3)
Swietokrzyskie	**27 (26.5)**	**19 (18.6)**	**25 (24.5)**	**16 (15.7)**	**10 (9.8)**	**5 (4.9)**
Warminsko-Mazurskie	4 (3.4)	13 (11.2)	16 (13.8)	44 (37.9)	29 (25.0)	10 (8.6)
Wielkopolskie	15 (6.6)	27 (11.9)	65 (28.8)	69 (30.5)	29 (12.8)	21 (9.3)
Zachodniopomorskie	0 (0.0)	2 (1.8)	17 (14.9)	43 (37.7)	35 (30.7)	17 (14.9)

Table 6.7 indicates that while rural gminas account for nearly half of the richest group of communities, this only amounts to seven per cent of the total number of countryside communities. On the other hand, thirty per cent of urban localities fall into the richest decile. At the same time, rural gminas account for two-thirds of the poorest ten per cent of local governments. The income profile of mixed gminas more nearly approximates that of rural than urban localities.

Table 6.7
Finance Clusters by Gmina Type

Gmina Type	1	2	3	4	5	6
Urban	19 (7.7)	23 (6.1)	42 (6.7)	82 (13.2)	54 (14.4)	98 (39.5)
Rural	165 (66.5)	267 (71.4)	440 (70.6)	370 (59.6)	240 (64.0)	113 (45.6)
Mixed Urban/Rural	64 (25.8)	84 (22.5)	141 (22.6)	169 (27.2)	81 (21.6)	37 (14.9)

7. ENVIRONMENT

Following Czyżewski and Zienkowski (2000), the state of the local environment was considered as a possible indicator of development. In particular, the SADB purportedly contains data on the following environmental indicators:

- Total harmful waste
- Particle emissions
- Gas emissions

Unfortunately, for the first of these variables the SADB for 2001 contains only 612 observations (163 of which are rural) and, for the second and third of them, there are only 683 observations (157 rural). Whatever the cause of these deficiencies, it was adjudged that there is insufficient evidence available to differentiate adequately between gminas. Environmental factors are not therefore pursued as an element of the desired, overall typology.

Towards an Aggregate Typology

With a large number of observations and multiple clusters, an overall, casual summary of the findings will normally be difficult, unless the results obtained are all very similar. As the latter was not the case in the above analyses, a scoring procedure was sought in order to develop an aggregate ranking of gminas. To achieve this, the clusters emerging in each separate exercise were ranked according to the degree of development they were assumed to reflect. This exercise assigned the number one to the best grouping in each case and total cluster number to the worst performing group. For example, in an exercise that resulted in four clusters, the group members would be assigned the integer numbers 1-4 depending on the degree of development indicated by their cluster. The relevant hierarchies were drawn out in each of the preceding sub-sections. There exist no objective *a priori* grounds for assigning weights to the separate indicators and, in order not to introduce an implicit scheme as a result of the differing number of clusters in each of the above analyses, the hierarchies were re-based to lie in the 0-1 interval. Thus, in the four cluster example, gminas would be assigned the scores 0.00, 0.33, 0.66 and 1, respectively. The resulting scores were then summed for each gmina to yield an overall development level indicator that can take values ranging from zero to six and these were then ranked

To ease interpretation, the gmina scores were grouped as in Table 5. The best performing gmina – Jastarnia in Pomorskie – attained a score of zero, while the worst

three – Biala Rawska and Sadkowice in Łódzkie and Obrazów in Świętokrzyskie – amassed totals of 5.80.

Table 5
Aggregate Typology Groupings

Group	Aggregate score
1	$0 < 1$
2	$1 < 2$
3	$2 < 3$
4	$3 < 4$
5	$4 < 5$
6	$5 \leq 6$

Table 6 provides the distribution of gminas across the groups, both in total and by voivodship, which is clearly heavily concentrated in the poor development groups. Map 7 shows the distribution of the scores across Polish space. While the largest number of the best performing gminas is to be found in Mazowieckie, the greatest concentration is actually in Pomorskie. The eastern regions do not, in general, perform well, with Świętokrzyskie having no communities in the group. However, Lubelskie has three gminas with aggregate scores below unity and this places it in the top half of the distribution of the voivodships. Nevertheless, if attention is focused on the two best performing groups – that is gminas with aggregate scores of less than two – then the eastern regions occupy four of the five lowliest positions. Only Malopolskie from the remainder of the country performs slightly worse than Podkarpackie. Kujawsko-Pomorskie represents the region with the highest proportion of the worst performing communities, although Lubelskie and Świętokrzyskie are ranked next. Podlaskie, in particular, performs relatively well on this score.

Conclusion

This paper has exploited the Small Area Database to examine the development profiles of Poland's NUTS 5 level gminas. In order to do this, cluster analysis was employed to categorise localities on six of the more development dimensions for which data is available: housing, infrastructure, amenities, business activity, amenities, agriculture and finance. The resulting hierarchies were then aggregated to produce an overall typology of development. In accordance with prior expectations, rural areas are among the least developed in Poland and the eastern voivodships of Lubelskie, Podkarpackie, Podlaskie and Świętokrzyskie perform poorly. However, there are exceptions and the results of the analysis should be used to identify good performers to be used as comparators for laggards in the eastern regions.

Table 6
Distribution of Gminas Across Typology Groups (% voivodship gminas)

Voivodship	1	2	3	4	5	6
Dolnoslaskie	4 (2.4)	17 (10.1)	27 (16.0)	42 (24.9)	68 (40.2)	11 (6.5)
Kujawsko-Pomorskie	0 (0.0)	12 (8.3)	5 (3.5)	16 (11.1)	50 (34.7)	61 (42.4)
Lubelskie	**3 (1.4)**	**5 (2.3)**	**5 (2.3)**	**19 (8.9)**	**99 (46.5)**	**82 (38.5)**
Lubuskie	1 (1.2)	10 (12.0)	21 (25.3)	29 (34.9)	22 (26.5)	0 (0.0)
Lodzkie	2 (1.1)	9 (5.1)	15 (8.5)	14 (7.9)	83 (46.9)	54 (30.5)
Malopolskie	2 (1.1)	8 (4.4)	18 (9.9)	28 (15.4)	88 (48.4)	38 (20.9)
Mazowieckie	14 (4.3)	27 (8.3)	26 (8.0)	27 (8.3)	127 (39.1)	104 (32.0)
Opolskie	0 (0.0)	5 (7.0)	7 (9.9)	21 (29.6)	36 (50.7)	2 (2.8)
Podkarpackie	**1 (0.6)**	**8 (5.0)**	**4 (2.5)**	**11 (6.9)**	**88 (55.0)**	**48 (30.0)**
Podlaskie	**1 (0.8)**	**3 (2.5)**	**8 (6.8)**	**23 (19.5)**	**74 (62.7)**	**9 (7.6)**
Pomorskie	9 (7.3)	13 (10.6)	9 (7.3)	18 (14.6)	64 (52.0)	10 (8.1)
Slaskie	8 (4.8)	25 (15.1)	27 (16.3)	47 (28.3)	48 (28.9)	11 (6.6)
Swietokrzyskie	**0 (0.0)**	**3 (2.9)**	**5 (4.9)**	**9 (8.8)**	**52 (51.0)**	**33 (32.4)**
Warminsko-Mazurskie	1 (0.9)	6 (5.2)	8 (6.9)	21 (18.1)	65 (56.0)	15 (12.9)
Wielkopolskie	4 (1.8)	14 (6.2)	22 (9.7)	31 (13.7)	82 (36.3)	73 (32.3)
Zachodniopomorskie	1 (0.9)	14 (12.3)	19 (16.7)	47 (41.2)	33 (28.9)	0 (0.0)
TOTAL	51 (2.0)	179 (7.2)	226 (9.1)	403 (16.2)	1079 (43.4)	551 (22.1)

Table 7 provides the distribution of the aggregate groupings by gmina type. Rural gminas are very clearly the least developed in Poland, as judged by community membership in the four best performing categories. No urban gmina appears in the worst performing group. No gmina in the eastern regions appears in the list of the best performing rural localities, although Lublin, Rzeszów, Bialystok, Chelm and Zamosc-u are included in the equivalent segment of the hierarchy for urban areas. However, no eastern gmina appears in the list of the top twenty performing mixed communities. Given tied scores, it is only possible to identify either the worst three of the worst sixty-one gminas in the country. Of the latter, Lubelskie accounts for

thirteen, Podkarpackie for six and Świętokrzyskie for five observations. Eastern gminas account for seven of the twenty worst performing urban areas, four of the worst twenty mixed gminas and twenty-one of the worst forty-eight rural communities.

Table 7
Distribution of Aggregate Groupings by Gmina Type (% of type)

Gmina Type	1	2	3	4	5	6
Urban	48 (15.1)	144 (45.3)	85 (26.7)	31 (9.7)	10 (3.1)	0 (0.0)
Rural	1 (0.1)	16 (1.0)	52 (3.2)	192 (12.0)	867 (54.1)	476 (29.7)
Mixed Urban/Rural	2 (0.4)	19 (3.4)	89 (15.7)	180 (31.7)	202 (35.6)	75 (13.2)
Total	51	179	226	403	1079	551

REFERENCES

Calinski, R.B. and J. Harabasz (1974) 'A dendrite method for cluster analysis', *Communications in Statistics*', 3, 1-27.

Czyżewski, A.B. and L. Zienkowski (2000) 'Differences in the Regional Devlopment in Poland in 1998', *Research Bulletin*, 9, 5-22.

Ingham, M. (2003) 'Urban-Rural Dichotomies in Poland and the EU', Mimeo, November.

MARD (Ministry of Agriculture and Rural Develoment) (2004) 'Restructuring and Modernisation of the Food Sector and Rural Development 2004-2006', Warsaw: MARD.

MARD (Ministry of Agriculture and Rural Develoment) (1978) 'Coherent Structural Policy for Rural Areas and Agriculture Development', Warsaw: MARD.

Milligan, G.W. and M.C. Cooper (1985) 'An examination of procedures for determining the number of clusters in a data set', *Psychometrika*, 50, 159-179.

Sarle, W.S. (1983) *Cubic Clustering Criterion*, SAS Technical Report A-108, Cary, NC:SAS Institute Inc.

[1] The SADB also contains information on 'Lost Dwellings' fitted with water supply, lavatory, bathroom, gas from network and central heating, respectively. The meaning of these variables is, however, unclear.
[2] This raises the question of whether rural areas are connected to the gas and electricity networks.
[3] This figure isdue to the fact that an oil/gas company is located in this gmina.