Catalogue no. 21-601-MIE — No. 067
ISSN: 1707-0368
ISBN: 0-662-36442-2

Research Paper

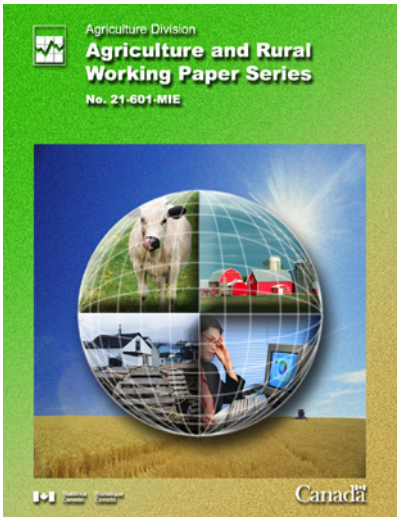
Mapping the socio-economic diversity of rural Canada: A multivariate analysis

1996

by Alessandro Alasia

Agriculture Division
Jean Talon Building, 12th floor, Ottawa, K1A 0T6

Telephone: 1 800-465-1991



Statistics Canada Statistique Canada

Canada

Statistics
Canada
Agriculture Division

Agriculture and Rural Working Paper Series
Working Paper No. 67

Mapping the socio-economic diversity of rural Canada: A multivariate analysis

Prepared by
Alessandro Alasia
Research and Rural Data Section, Agriculture Division

Statistics Canada, Agriculture Division
Jean Talon Building, 12th floor
Tunney's Pasture
Ottawa, Ontario K1A 0T6

March 2004

Statistics Canada
Agriculture Division

Agriculture and Rural Working Paper Series
Working Paper No. 67

Mapping the socio-economic diversity of rural Canada: A multivariate analysis

Published by authority of the Minister responsible for Statistics Canada.

March 2004

Catalogue No. 21-601-MIE

ISSN: 1707-0368
ISBN: 0-662-36442-2

Frequency: Occasional

Ottawa

La version française est disponible sur demande (nº 21-601-MIF au catalogue)

Note of appreciation: Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Executive Summary

The socio-economic performance of localities has increasingly come under scrutiny from a national perspective. Today it is widely recognized that regions have different characteristics that shape their potential path of development and that the policy process should not overlook the diversity of their conditions. One of the crucial questions associated with the policy focus on small geographical units is whether and to what extent it is possible to implement development strategies and policies for each type of region. This in turn has raised a number of questions about the structure and characteristics of homogeneous regions.

This paper assesses the degree of spatial diversity exhibited across Canada by using 1996 Census of Population data, aggregated at the census division (CD) level. The approach taken in this research adopts a broad territorial focus, is exploratory in nature and emphasises territorial performance in a comparative context. The study is based on a range of commonly used and understood demographic, social, and economic variables. A factor analysis was conducted in order to identify underlying dimensions that characterise each CD across Canada. The factor analysis resulted in six factors, each of which provides a profile of the CDs on a number of key attributes. The research is primarily descriptive and is potentially of interest to a broad audience. It can facilitate the diffusion of baseline data to a wide range of stakeholders, stimulate discussion on spatial diversity at the sub-provincial level and inform the debate on potential alternative development paths for each region. It must be acknowledged that this research is in turn constrained by the nature of the data available. The analysis is also static and focused on a cross-section. The causes of the observed diversity are not explicitly accounted for in the study.

Key Findings

- Twenty-seven variables used in the factor analysis can be reduced to 6 factors. The factors capture about 78 percent of the variance in the data set. Some of these factors are more diagnostic, while others are descriptive in nature.

- The first factor, named *labour force and economic attributes,* captures a range of economic and social attributes, including unemployment, income level and educational attainment, that describe the overall economic performance and strength of the CD. This factor appears to describe in particular the regional variation in economic performance (north / east versus south / west). Urban CDs tend to present higher scores on this factor while rural CDs show a greater diversity of conditions.

- The factor named *remote and agro-rural attributes* describes a set of attributes that prevail in remote areas and the part of rural Canada with a higher incidence of agricultural employment. The factor combines demographic and housing characteristics that are common to these areas. Negative scores are overwhelmingly a feature of the remote north (northeast in particular) and the Prairies, while urban CDs present opposite characteristics.

- Two factors are mainly associated with the employment structure. The first, named *complex manufacturing versus primary production attributes,* identifies in particular the non-agricultural resource-based communities whereas the second factor, named *traditional manufacturing versus government employment attributes*, shows a contrasting pattern between traditional manufacturing employment and a higher labour force participation rate for males, on the one hand, and non-market services employment on the other hand.
- Two factors describe in particular the demographic structure and dynamics of the CDs. The first, named *demographic and labour force attributes*, relates in particular to aging population trends, but describes also labour force characteristics (non-agricultural self-employment and part-time work). This factor shows similar conditions for urban CDs and remote CDs on the attribute descriptors. The second factor, labelled *demographic dynamics attributes*, identifies areas with a young and growing population. The distribution by regional type shows the diversity of conditions recorded by both urban and by rural CDs.
- The spatial distribution of the factors reveals regional differences as well as differences between CD regional types. This demonstrates the utility of the broad territorial approach adopted, which allows a better understanding of both regional patterns as well as hierarchical spatial structures (i.e., the groupings of CDs with similar spatial patterns).
- The current definition of CD regional types (i.e. predominantly urban, intermediate, rural metro-adjacent, rural non-metro-adjacent and rural northern regions) captures, relatively well, the variation of some of the identified dimensions. This classification has the major advantage of simplicity and clear-cut definitions. Nonetheless, for specific policy purposes, it would be appropriate to use a more refined typology that focuses on the policy issue under consideration.
- The results indicate the multi-dimensional nature of the performance of regions and the variety of associated demographic, social and economic characteristics (e.g. resource-based regions with a poor economic performance versus resource-based regions with an above average economic performance, etc.).
- With regard to specific indicators, the results appear to trace the two prevailing dimensions of income variation, one across macro-regions and the second between rural and urban regions.

Table of contents

List of tables

List of maps

1. Introduction

In recent years, the socio-economic performance of small territorial units has increasingly come under scrutiny. Several factors have influenced this trend. On the supply side, the growing availability of data for small geographic units has been a precondition for this type of analysis. On the demand side, it has been increasingly recognized that regions have widely different opportunities and constraints which shape their potential path of development – and that the policy process should not overlook this diversity of conditions (Pezzini, 2001; OECD, 2001). Hence, at the provincial and federal level, attention to small territorial units is required to understand how universal policies might affect different areas, as well as to assess the potential for tailor-made local policies. Parallel to that, the continuous process of decentralization and downloading of responsibility toward municipal and county administrations has also stimulated the analysis at a small geographic scale. At the municipal and county levels, the management and design of local development policy motivate the growing interest in the comparative conditions and strengths of a given area.

One of the crucial questions associated with this shift in geographical scale is whether and to what extent it is possible to implement development strategies and policies for each type of region. This in turn has raised a number of questions about the structure and characteristics of homogeneous regions and the way in which these should be identified. Much of the socio-economic data that are typically used for this purpose have become relatively easy to access. For instance, over the last few years, a variety of Internet sites have been established which provide a wide range of data at the municipal and county level, such as employment by sector, unemployment rate, demographic indicators and so on (for example, click on "community profiles" at www.statcan.ca). Similar initiatives have been developed at the provincial level (see for instance, Community Accounts for Newfoundland and Labrador, at www.communityaccounts.ca). The mere availability of these data, however, does not make the interpretation of spatial patterns of socio-economic diversity across the country straightforward. The multitude of indicators and of territorial units makes it difficult to capture the broad picture of the comparative social and economic conditions of different regions. One of the major challenges for researchers and policy makers is to reduce the complexity to a manageable set of indicators that can be used to interpret reality. Moreover, attempts to identify homogeneous areas need to be updated periodically, as well as compared with other prevailing classification schemes to understand how each of them performs in capturing the variety of socio-economic conditions.

The purpose of the research was to establish the degree of spatial diversity existing across Canada on a number of demographic, social and economic indicators and to highlight the implications of the observed spatial variation for rural development policy. The questions answered here are:

- Can we identify a number of underlying dimensions that characterize the social, demographic and economic diversity that exists across the country?
- What is the spatial distribution of these dimensions?
- What is their relationship with the prevailing regional classifications?
- What is their meaning for rural development initiatives?

The research is based on 1996 Census of Population data, aggregated at the census division (CD) level. Twenty-seven indicators were selected that cover a broad range of attributes of each CD. The method used in this study is a factor analysis, which allows one to identify similar CDs on a small and manageable number of dimensions summarizing the range of indicators selected. Three key features characterize the research approach: a comprehensive territorial focus which embraces all the CDs of Canada without *a priori* specification of regional types; a focus on comparative characteristics and performance rather than on absolute thresholds; and an exploratory data analysis. The methods and research approach adopted distinguish this study from previous studies conducted for similar purposes, even though the current analysis builds on some of these experiences and the findings of this literature (Shearmur and Polèse, 2001; Hawkins, 1995; Alasia, 1999; Keddie and Alasia, 1999).

The paper is structured as follows. Section 2 discusses some key concepts used in this study and presents a survey of the literature in this area of research, with a particular focus on empirical analysis in the Canadian context. Section 3 discusses the study approach and the methods used for data analysis. This section gives a brief description of the statistical terms used in the report, explains the principles of factor analysis and discusses the data and the variable selection process. Section 4 presents the results of the factor analysis – each factor is described and its spatial pattern is mapped. Section 5 presents an example of how the results can be further employed by combining some of the dimensions generated in the factor analysis. Finally, the last section summarizes the major findings, discusses their implications and potential use and highlights the limitations of the study.

2. A survey of concepts and empirical research

The problem faced in this study is in essence that of classifying a set of observations. Classification is a fundamental concern in any field of research and policy analysis. In order to make sense of complex realities and phenomena, analysts are often required to organize the observations by 'types', which are not identical, but rather tend to behave alike. In the context of spatial analysis, this classificatory work is defined as regionalisation (Rogers, 1971). Thus, regionalisation is the process of classifying and grouping small territorial units into larger aggregations that share elements of similarity or commonality.

The development of a territorial typology can serve many purposes. Indeed, the appropriateness of a typology itself should be evaluated against the purpose that it is intended to serve. In this regard, two broad alternatives can be identified. Some typologies are intended to address specific policy measures. For instance, it is possible to develop a typology of regions based on unemployment structure. In this case the attention can be restricted to a single or a few indicators. In other cases, the interest is on typologies that have a broad analytical or policy purpose that attempts to embrace a wide range of demographic, social and economic indicators.

When the interest is in a broad range of regional characteristics, there is a major methodological distinction between ways to proceed in defining and analyzing 'regional types'. On one hand, it is possible to define, *a priori* and based on theoretical reasoning, a set of criteria that will be used to assign each area to a regional type. For instance, it is possible to

define population thresholds or thresholds based on distance from major urban areas. Each CD can be assigned to a certain group based on these criteria and then the researcher can compare the various groups on a range of other indicators (e.g. income, unemployment, and so on). Alternatively, it is possible to follow an exploratory type of approach and to use the range of data to construct types of regions. This means that the data set is 'explored' in an attempt to recognize any non-random patterns or structure in the existing set of variables, generally without imposing any pre-determined model of relationship between these variables.

There is a third major distinction in methodologies that derive regional typologies. The applied research on regionalisation has tended to emphasize two alternative aspects: homogeneity and nodality or functionality (Rogers, 1971). When the emphasis is on homogeneity, the researcher aggregates areas that tend to be uniform with respect to a set of characteristics 'contained' within each unit of observation. However, the focus on nodality or functionality implies the aggregation around a relevant pole (generally an urban centre) on the basis of the linkages and functional relationships between areas (for instance, using commuting flows or trade linkages).

The remainder of this section presents a survey of empirical research whose characteristics are defined along the lines of the concepts reviewed above. The research on regional typology that is surveyed tends to be broad in scope, exploratory in nature and focused on the homogeneity aspect. These are also some of the characteristics of the present study, while other features of the approach taken in this analysis are further discussed in section 3. The review presented below starts with a brief summary of the different techniques that have been used in this area of research and then moves to applications in the Canadian context.

2.1. An overview of alternative methods

Attempts to develop regional or rural typologies have generally relied on multivariate statistical techniques and used population census or census-type data for this purpose (see Blunden *et al.*, 1998). In regional applications, the dominant approach has been either factor analysis or a combination of principal component and cluster analysis (see also Rogers, 1971).

The application of factor analysis to the study of community dimensions reached a peak in the 1960s and 1970s. This approach to the study of spatial diversity became known as factorial ecology and had a rapid expansion as well as an equally rapid decline (Rees, 1971; Berry, 1971). Critics of this method indicated as a major weakness the exploratory nature of this research, which resulted in fragile theoretical foundations for the conclusions provided. A weakness that, on the contrary, was seen as a strength by some scholars, because the understanding of a certain situation was "learned" rather than "imposed" by *a priori* theory (Berry, 1971). As will be further discussed below, there is no doubt that factor analysis, like all the exploratory data techniques, is exposed to a certain degree of subjectivity, particularly in the selection of the variables and the interpretation of the factors. Once the limitations of this method are recognized, it is clear that this technique can still provide a useful characterization of territorial units. Indeed, factor analysis has continued to be widely used in applied regional studies.

One classic example in this area of research is the study by Jonassen and Peres (1960) that attempted to identify the dimensions that underlay the loose concept of community. The

authors used 88 counties of Ohio and 82 variables that covered a multitude of social, demographic and economic aspects of community life. From the analysis, seven factors were identified. Thompson *et al.* (1964) used a factor analysis to explore the concept of economic health using county data for 58 counties in New York State. They concluded that a concept like "economic health" is a multi-dimensional construct which cannot be measured or summarized by a single indicator (or factorial dimension). More recently, factor analysis (in combination with principal component analysis) was used in Italy in an analysis that covered the national and the regional level (Cannata, 1989). The objective of the analysis was to map the marginal areas. The study used 29 variables, mainly agricultural-related, at the municipal level for the whole country and for 20 Italian regions. A similar study was conducted in France to develop a typology of agricultural areas (SEGESA 1992). Another study in France by Chapuis and Brossard (1989) used factor analysis to identify homogeneous regional types based on demographic structure and dynamics. Finally, Copus and Crabtree (1992) used a factor analysis, in combination with other statistical procedures, to evaluate how well the delimitation of regions for rural development purposes implemented by the U.K. Government and the EU reflects the spatial reality of Scotland. The authors concluded that the use of large and heterogeneous administrative units as a building block for the delimitation of target areas resulted in a poor overlap with areas that were fragile according to their analysis.

Examples of the application of principal component and cluster analysis are numerous. In these studies, principal component analysis is used as a data reduction method, which allows the extraction of linear combinations of the original variables. In the second phase, cluster analysis is performed using principal component scores to identify grouping of areas with similar profiles in terms of component scores. Using this approach, Shields and Deller (1996) produced a classification of counties in Wisconsin. A set of economic and demographic variables was used, which was largely based on employment by sector. The classification resulted in 8 clusters of counties. Six clusters identify the main (or a combination of the main) economic activity of the county (agriculture, forestry, manufacturing, services, trade, government, and tourism), one cluster is defined as urban centres and the last is a single-county cluster. Quadrado *et al.* (2001) applied a similar approach, combining also other inequality measures, to classify 20 counties in Hungary. Montresor and Mazzocchi (2001) used this two step approach to classify 100 EU regions based on 39 variables, where most of the variables were agricultural-related indicators. Stimson *et al.* (2001) combined discriminant analysis and cluster analysis to analyze a set of opportunity and vulnerability indicators of urban communities in Australia, which yielded a nine-cluster solution.

Recently other data reduction or exploratory methods have found application in regional classificatory exercises. A variant of traditional clustering methods was put forward by Lipshitz and Raveh (1998) in an application to Israel. The method is defined as "co-plot" by the authors and is in essence a graphical display-based technique, where geographic units are grouped on the basis of a measure of dissimilarity between each pair of observations. Blunden *et al.* (1998) presented a particularly interesting methodology for the classification of rural areas in the European context, which relies on a neural network application. Neural networks belong to a set of exploratory data techniques that have received increasing attention in recent years. However, the method appears particularly complex and data intensive. In order to generate optimal outcomes the neural network needs to be "trained", using data from typical examples of typologies. In the application presented by Blunden *et al.* (1998), the network was

trained on the basis of the expert knowledge of practitioners who identified examples of five generic rural categories.

There seems to be limited research about the comparative performance of different techniques and it is beyond the scope of this study to do so. There are, however, a number of reasons that motivated the use of factor analysis in this study. Compared to the other techniques considered above, factor analysis seems to combine theoretical appeal with relative computational simplicity. Although it is generally agreed that factor analysis (in its standard exploratory formulation) and principal component analysis would yield essentially the same results (because of the factor extraction procedures that are usually followed), they remain conceptually different. Principal component analysis simply extracts linear combinations of the variables at hand. In contrast, factor analysis implies the existence of an underlying factorial structure that explains the variability of the observed indicators; thus in this case the analysis implies an underlying conceptual construct that *causes* the observed variation. This approach also allows exploring the possible multi-dimensional nature of socio-economic performance of a locality, which is often a thorny issue in classification exercises. Finally, in its exploratory form, factor analysis remains relatively simple to implement and relatively straightforward to interpret. The next section, which surveys the empirical research in Canadian context, will show that this approach has also been prominent in Canada.

2.2. Empirical analysis in Canada

There are a number of empirical studies that have assessed regional diversity across Canada and have proposed alternative regional typologies. Some of these studies were specifically focused on rural areas. A comparison of these works is possible in particular with respect to the methods, but given the different variables used and year of reference, it is often difficult in terms of results. Nonetheless, this literature provides an important point of reference for the present research. Methods and main findings are summarised below, starting from the most recent study and finishing with some provincial applications.

Shearmur and Polèse (2001) have recently developed a typology of regions in Canada. The analysis is based on 382 *ad hoc* spatial units that are partially based on the census division geography. A first part of this study is conducted using “synthetic regions” defined *a priori* on the basis of population thresholds and distance. The classification results in 6 classes of core urban areas, 4 classes of peripheral-urban areas (all based on population size) and two classes of rural regions (central and peripheral, based on the distance to urban areas). This analysis considers the evolution of employment structure and other key indictors over the 1971 to 1996 period across each synthetic region. In a second phase, the authors use cluster analysis to identify areas with similar industrial structure; hence the focus is on employment by sector and clustering is based on location quotients. Starting from an 18 sector classification, the analysis results in 14 clusters. The authors conclude that “there is a close correspondence between synthetic regions derived from *a priori* theoretical considerations and clusters derived from an analysis of industrial structure” (Shearmur and Polèse, 2001:79). A second cluster analysis is conducted restricting the attention to 18 sub-sectors of the primary and traditional manufacturing sector (once again clustering is based on location quotients), which results in ten clusters. Nonetheless, in both cases, the multivariate analysis focuses on the employment structure and leaves out of the picture most of the other characteristics of a locality. No

comparison of the clusters is made on any of the other variables used in the first part of the analysis – in essence the typology developed with an exploratory approach remains a classification strictly determined by the employment structure of the spatial unit.

A study by Hawkins (1995)[1] is particularly relevant in this context because of the purpose and methods used in this study. Hawkins used principal component analysis and cluster analysis to re-examine the assumption that rural Canada is a homogeneous space. The analysis covered all 266 census divisions (CDs), using 1986 administrative boundaries. The census divisions included the metropolitan centres. The data set included 57 variables, all from the censuses of 1981 and 1991, which covered demographic, labour market, income, human capital and infrastructure attributes. Among the variables selected there were a large number of trend variables, most of them showing percent change, and some showing absolute change between 1981 and 1991.

The analysis was carried out in two stages. In a first run of the principal component analysis, two groups were identified: the major metropolitan areas (named *Primary settlements*) and the northern areas with high Aboriginal population (named *Native north*). These were designated as the first two clusters of CDs. A second run was performed without these CDs to extract five major components (linear combination of variables). Then a cluster analysis was performed on the component scores and this provided five additional clusters of CDs. One cluster was named *Urban frontier* which was similar to the major urban centres (*Primary settlements),* but had less marked characteristics. The other four clusters identify different rural types.

The first type is named *Rural Nirvana* and is characterized by large population totals, high population growth, slightly lower than average numbers of young people and a slightly above average number of elderly. The economic profile of this rural type indicates the lowest rate of unemployment and a high rate of participation in the labour force. Employment in the primary sector is low compared to the other clusters and the main employment sector is indicated as consumer services, hi-tech manufacturing and construction. Also educational level, share of professional and managerial workers, and income are above the national average, while the share of low-income households is below average. The *Rural Nirvana* type is prevalent in southern Ontario and in the vicinity of other major urban centres.

The second type is named *Rural enclave* and presents the characteristics of the economically disadvantaged rural census divisions, i.e. essentially the opposite of the previous type. This type is prevalent in the Atlantic Provinces.

The third type is called *Agro-rural* and is defined by employment structure, rapid population decline and out-migration. This type is prevalent in the Prairies.

Finally, the last cluster is named *Resourced area* – this is typified by employment in the "other" primary sectors (i.e., other than agriculture) and relatively good demographic,

[1] For a colour map of her typology, see Fellegi (1996) or Hawkins and Bollman (1994).

economic, and educational attainments. This type is dominant in the northwest of Canada, particularly Alberta, and in northern Ontario.

Reimer (1997) conducted a factor analysis, using data at the census sub-division (CSD) level for all of Canada, with the purpose of identify leading and lagging CSDs on a number of social and economic dimensions. The study was undertaken as a preliminary analysis for the New Rural Economy project and was intended to classify areas with different characteristics for further in-depth study. The author used 17 variables from the 1991 Census of Population. Four factors are identified which accounted for approximately 60 percent of the total variance in the data set. The first factor is named *Income-related* and groups three variables (median household income, female income and percent of total income from employment). The second factor is defined as *Labour force-related* and groups CSDs on the basis of government transfer income, unemployment rate, labour force participation rate and percent self-employed. The other two factors appear less relevant in term of variance explained and are named *Marriage and housing costs-related* and *Housing and low income-related*, respectively. However, given the purpose of the analysis, the spatial pattern of the factor scores is not further assessed.

Factor analysis was also used by Simmons and Speck (1986) in a study on spatial patterns of change across Canada. This research builds on a previous study by Simmons which used 1971 census data. The same approach was applied to 1981 census data. In both cases, the focus of the analysis was on the structure of the urban system in Canada. The authors used 124 urban-centred regions obtained by aggregating the (266) census divisions around the urban places with population over 10,000 in 1971. A total of 80 variables are included in the factor analysis, which tend to cover a broad range of demographic, social housing and economic characteristics of each area, while about 10 variables describe the ethnic origin of the population. The analysis results in four factors which are named *Frontier community*, with a strong emphasis on demographic characteristics; *Economic achievements*, with a strong emphasis on demographic growth; *Metropolitan characteristics*, with an emphasis on city size; and *Cultural*, contrasting French ethnicity with individuals with other European origins. The factor scores are mapped but the nature of the association of each factor with the 80 variables is not discussed in detail. Rather, the authors focus on a comparison with 1971 results. The conclusion is that the spatial structure of variation is essentially stable over the 1971 to 1981 period, with only the factor named *Economic achievements* showing some degree of spatial relocation.

Finally, at the provincial level, Alasia (1999) applied a factor analysis using 1991 census data at the CD level for Ontario, while Keddie and Alasia (1999), used CSD data for 1991 and 1996 for southern Ontario. This second study in particular adopted a high level of geographical resolution in the analysis of spatial variation. Thirty-four variables were used in the analysis covering a broad range of socio-economic indicators available in the census. The factor analysis was conducted separately for 1991 and 1996 data and in both cases resulted in 8 factors, which presented essentially an identical structure and similar spatial patterns. The factors were named: *Labour force participation and age; Income, educational and occupational status*; *Socio-economic disadvantage; Unemployment and underemployment; Agro-rural attributes; In-migration and population growth; Non-market service versus manufacturing employment;* and *Traditional employment structure*. The results point out that

at this level of geographical scale which accounted for differences among incorporated places (city, town, village, and township), no consistent distinction emerges between space that is conventionally considered rural and urban. Within the study area considered, most of the diversity is either regional (south-eastern versus south-western Ontario) or between municipal types.

3. The research approach and methods

This section discusses the approach taken in this research, the data and the statistical methods used. The objective is to clarify the logic and the statistical terminology used in the paper. For a more technical presentation of factor analysis the interested reader is referred to a specialised text (for an application to regional analysis, see, for instance, Rogers (1971)).

3.1. The research approach

Some elements of the research approach adopted in this study have already been discussed in Section 2. These elements are a focus on homogeneous regions with respect to a set of characteristics of the territorial unit (as opposed to a focus on functional relationships); the broad scope of the typology, which attempts to embrace a wide range of socio-economic indicators; and the exploratory nature of the statistical analysis, which will be further illustrated in the next section. There are, however, two other characteristics of the research approach that should be briefly mentioned.

First, the emphasis of this research is on the *comparative* socio-economic performance of the CDs. This means that the results of this analysis do not set standards about what is an acceptable performance and what is not. Higher and lower values are defined on the observed range of variation of each indicator and the research does not impose predefined thresholds on any of the indicators used. In this sense, it should also be stressed that comparison is made at the county or CD level. Therefore the results are inevitably different from analysis conducted at the provincial level. Moreover, it is expected that measures of performance will be multidimensional, with combinations of social and economic performance that could often vary in opposite directions.

Second, the study adopts a territorial focus, meaning that although the main interest is on rural areas, the study includes all the CDs of Canada. An alternative would be to use an *a priori* definition of rural CDs and select this sub-sample for the analysis. The focus on relative performance, however, has suggested the utility of including all the CDs of Canada in the analysis. Rural regions are increasingly integrated into regional socio-economic systems, which implies that we can gain a better understanding of their interactions and ultimately of rural conditions by considering the system as a whole, and not only a part of it.

3.2. The exploratory factor analysis

The intent of this section is to explain the logic and the essential terminology of factor analysis. Factor analysis is a multivariate statistical technique that helps to answer questions such as, "Can a small number of unobservable factors explain the variability in many observable variables?" The main assumption of this statistical method is that observable outcome indicators can be accounted for by a limited number of underlying factors, which can be used to explain complex phenomena. For instance, conceptual constructs such as economic health or social distress are not directly observable. Nor they can be measured directly. What a researcher can do is to measure a number of outcome indicators, as for instance income level, unemployment rate, number of low-income families, and so on. One then postulates "factors" as latent variables, or underlying dimensions, that are in some way correlated to variables that are directly observable and measurable.

The search for these underlying dimensions can take place essentially in two ways. First, it is possible to determine *a priori* a theoretical model, which defines the causal relationship between unobservable factors and observable outcomes. This model can then be tested (or confirmed) by empirical analysis. This can be done with a confirmatory factor analysis in which the relationships to be tested are specified by the analyst. Second, it is possible to explore the data in an attempt to identify non-random patterns of associations between variables, and to let the factors (if any) emerge from the analysis. This is in essence what an exploratory factor analysis does. Starting with a large set of variables, the analysis yields a restricted number of factors that are correlated with observed variables and summarises their values.

The observed variables must present some degree of correlation for the factor model to be appropriate. The statistical procedure results in grouping those variables that are more strongly correlated among each other and less related to variables in other groups. The mathematical formulation of this general model appears similar to a multiple regression equation. Each variable can be expressed as a linear combination of factors, which are not actually observed, as follows:

$$x_{ik} = a_{i1}\,\xi_{1k} + a_{i2}\,\xi_{2k} + a_{i3}\,\xi_{3k} + \ldots\ldots + a_{i\,n}\,\xi_{nk} + \delta_i$$

where x_{ik} is the value of the variable i for the k^{th} observation (which in our case is the CD) in standardised form; ξ_{jk} is the value of the j^{th} factor (commonly referred to as *factor scores*) for the k^{th} observation, a_{ij} is the standardised regression coefficient of the i^{th} variable on the j^{th} common factor (commonly referred to as *factor loading*), and δ_i is the unique factor for the variable i.

In the model applied in this study, the factors are assumed to be uncorrelated with each other, and each of the factors ξ affects each of the observed variables x_i. These factors are called common factors, since their effect is shared in common with more than one variable. The δ are called unique factors, or errors in variables, since their effect is unique to one variable. For instance, the standardised value of the variable called average household income (HI) for the k^{th} observation can be expressed as a linear combination of the factors:

$$HI_k = a_1 (\text{Factor } 1)_k + a_2 (\text{Factor } 2)_k + \ldots\ldots + a_n (\text{Factor } n)_k + U_{HI}$$

In this expression, Factor 1 to Factor 'n' are the common factors and their values for each observation (CD) are the *factor scores*. These are in essence a summary variable. The factor scores are standardised, with values usually ranging from about -3 to +3. Hence, their values indicate the relative performance on the particular dimension identified by the factor and summarise the behaviour of a group of observable variables associated with it. The U is the unique factor. To it is attributed that part of the variability of HI that can not be explained by the common factors. The proportion of the variance of a variable that is explained by the common factors is called the communality of the variable. The coefficients a_i are the *factor loadings*. In simple terms, the value of this coefficient describes the closeness of the relationship between a variable and the factor. Since the variables are standardised, the factor loadings indicate how much weight is assigned to each factor. In the specific computation method applied in this research to generate the factors (principal component analysis), the factor loadings also show the correlation between the factor and the variable. The higher the value of the factor loading (whether positive or negative), the closer is the relationship.

Ideally, the analysis should result in variables with a high loading on one factor and low loading on all the other factors. In this way, it is possible to identify the variable(s) that are closely related to each of the factors identified, and consequently, to understand the nature of the factor. The proportion of the variance of each variable that is explained by the model is used to assess how well the factor model describes the original variables. In the model used, we assume the factors are uncorrelated with each other and thus the total proportion of the variance explained by the model is just the sum of the variance proportions explained by each factor. Furthermore, the nature of the problem at hand does not imply any inference. The technique is used here strictly for data reduction purposes – no reference is made to statistical significance of the coefficients.

3.3. Data characteristics and variables used in the factor analysis

All the data used in the study are from the 1996 Census of Population[2]. All data are aggregated at the census division (CD) level. The CDs selected for the study are all the 288 CDs of Canada existing in 1996. Table 3.1 lists the 27 variables used in the analysis, which for convenience of exposition are grouped under four major headings: demographic indicators; social indicators; housing characteristics; and economic and labour market indicators. Details about the definition of each indicator are provided in Appendix A. Most of the variables selected are indicators commonly used and understood by professionals and the general public.

It should be stressed that the results of a factor analysis are largely determined by the variables included in the procedure. This is obvious in the sense that whether or not an indicator of a certain social or economic sphere of concern is included will determine the possibility to have this dimension in the results that are generated. However, there is also a less obvious but still

[2] Unfortunately, data from the 2001 Census of Population were not available when this analysis was undertaken. However, we would expect the 2001 patterns to be essentially the same.

important aspect of the variable selection process. Within each area of concern, there is generally a large variety of specific indicators that could be selected. For instance, the census offers a large range of unemployment variables, which include breakdowns by gender and age groups. Similarly, participation rates and demographic cohorts can be expressed in a variety of different ways (including absolute terms). The specific indicator used and the way in which it is expressed can have, in some cases, a considerable effect on the results.

The selection of the variables, therefore, represents a crucial phase of the research. Indeed, it is the phase in which a large part of the theoretical reasoning, that is often said to be lacking in factor analysis, should be focused. For the purpose of this study, the selection of the indicators among those available from the census was guided by findings from the literature, previous research experience and some general criteria. In particular, the following criteria were used in selecting the indicators.

First, all the variables considered are expressed in relative terms (i.e. as percent) or are average values for the CD. There is no variable that measures the absolute dimension of an indicator (i.e. total population, absolute population change, etc.). Hence, the comparison among CDs focuses on the structural characteristics of the territorial unit and not on the absolute size. Given the enormous diversity among CDs in term of size, the introduction of absolute values would force the results in a certain direction, which might not reflect the structural characteristics of a locality. Second, except for the population and income change variables, all the other variables employed in the factor analysis identify conditions in 1996. Thus, the focus is on conditions in 1996. Third, the attempt was to cover a broad range of socio-economic indicators, including demographic structure, employment and labour market characteristics and indicators of potential social distress. But no physical descriptors of the CD were used, such as the distance from major urban centres. Also, compared to previous studies, this analysis does not include any indicator of ethnic origins of the population – notably, a variable to measure the share of the population with an Aboriginal ethnic origin is not included.

Following these criteria, a total of 52 indicators were initially selected for the analysis. Each indicator was mapped individually to gain an understanding of its spatial pattern. From this pool of 52 indicators, the 27 variables used in the analysis were chosen (Table 3.1). Most of the variables that were dropped had a very high correlation with one or more of the other variables and thus they were essentially duplicates of each other. For some spheres of concern, it was decided to include only one indicator rather than a set of indicators. For instance, a single indicator of the overall unemployment rate was used rather than a set of unemployment rates by age and gender. Finally, some of the variables were dropped because of the relatively high number of missing values.[3] In sum, it is undeniable that the choice of the indicators presents a degree of subjectivity and that the final selection of specific indicators employed is partially driven by an empirical rationale. But it should be noted that the changes that would

[3] An exception to this is the share of persons living in low-income economic families. For six CDs in the Territories, no value was reported for this variable. The missing values were replaced with CD average values. This represents an economical solution and given the small number of CDs affected, it is not expected to have any significant effect on the results. This was the only adjustment required to the data set.

occur with alternative specifications of the indicators are not dramatic, since the main structural relationships among variables hold anyway.

Tables 3.2 and 3.3 display some basic descriptive statistics for the variables used in the analysis. The first table shows the average values within each septile[4], the average value for all CDs (note that this is not the national weighted average, but rather is the simple average across CD average values), the value for the median CD, and the range (minimum CD and maximum CD). The figures are not discussed here, but the table will be of particular help when presenting factor analysis results. The second table shows the average values and standard deviation by CD regional types: predominantly urban, intermediate, rural metro-adjacent, rural non-metro-adjacent and rural northern regions (see du Plessis *et al.*, 2002). Some differences between types emerge clearly and details will be discussed when presenting the factor analysis results. However, it should be noted that the standard deviations are generally greater within each rural type than within predominantly urban or intermediate CDs. This points to the overall greater diversity of conditions observed across rural regions which will clearly emerge in the factor analysis results discussed in the next section.

3.4. Caveats

This research identifies socio-economic dimensions for general descriptive purposes and portrays the overall patterns of spatial variation. The use and interpretation of the results of this study cannot be stretched beyond the scope of the research itself. The utility of factor analysis stems from a reduction of the complexity of socio-economic conditions that can be observed. On the other hand, the factors constitute summary variables and thus the procedure itself leads to a loss of information. The analysis of every individual variable provides more information than one can assimilate and makes it difficult to see the forest for the trees. While each variable may be of interest in and of itself, it is not possible to see how they vary together. Similarly, the analysis does not provide the best possible information for any single CD. Specific information on the individual territorial unit is lost in order to gain a broad view on spatial patterns across the country.

Moreover, two caveats should be borne in mind by the reader, one related to the data used in the study and the other related to the nature of the technique applied. With regard to the data used, the results of the factor analysis depend on the nature of the variables used in the computation. Many attributes that give quality of life for individuals and families are not captured by the data on which this analysis is based. Variation in level of access to and satisfaction with primary health care; similar issues regarding primary and secondary education for children; or the available levels of support services for elderly are just a few examples of attributes that are not available from the census.

---

[4] For each variable, the 288 CDs are ranked from the smallest to the largest value for the given variable. Thus, for each row of Table 3.2, the 288 CDs are re-ranked according to the size of the variable in the row and the CDs are grouped into 7 groups with 1/7 of the CDs in each group (column), where "septile" implies 1/7. The reported numbers in Table 3.2 are the average values for the given variable (row) with each septile of the variable (i.e., within each column). Thus, POPCH is the first row and the $1/7^{th}$ of the CDs with the smallest POPCH reported an average POPCH of -4.8 percent between 1991 and 1996. Similarly, the $1/7^{th}$ of CDs with the highest POPCH reported an average POPCH of 16.3 percent.

With regard to the nature of the statistical technique applied in this study, it should be remembered that the set of variables used in the analysis captures both causes and effects of certain phenomena. This research did not discern between the two aspects. The approach taken assumed that all observed variables are directly affected by all common factors and that all common factors are uncorrelated among each other. Hence, the theoretical appropriateness of the patterns that are identified is assessed *ex post*; although all factors provide key insights into the relative position of CDs across grouping of variables and show the regional diversity of the study area. Finally, the focus of the analysis is on the condition prevailing in 1996. Hence, the study provides a static description of regional conditions.

Table 3.1 Variables used in the study: definition

Code	Variable definition (for detailed definitions see Appendix A)
Demographic indicators	
POPCH	Percent population change 1991 to 1996
POPL20	Percent of population less than 20 years of age
POPO65	Percent of population age 65 years of age and over
IMOLD	Senior in-migration rate: Percent of persons 55 to 74 years of age living in a different CSD 5 years ago
FERTIL	Fertility rate (estimated as number of persons under 19 years of age divided by the number of women 25 to 54 years of age)
Social indicators	
EDUAVE	Average years of schooling for population 25 to 54 years of age
HHBLICO	Percent of persons in low-income economic families
INTRSF	Social transfer income as a percent of total income
UNTOT	Total unemployment rate (for the labour force, 15 years of age and older)
Housing characteristics	
RENT30	Percent of households with gross rent equal to or greater than 30 percent of household income
HOWN30	Percent of households with the owner's gross housing costs equal to or greater than 30 percent of household income
POWN	Percent of households owning their home
Economic and labour market indicators	
EMAGR	Percent agricultural employment
EMPRIM	Percent other primary employment (i.e., forestry, fishing, mining, gas and oil)
EMTRM	Percent traditional manufacturing employment
EMCMA	Percent complex manufacturing employment
EMDSE	Percent dynamic services employment
EMSSE	Percent non-market services employment
PARTEC	Participation rate
WKO2	Percent of families with two or more members in the labour force
MFPART	Male participation rate divided by female participation rate
SELF	Percent self-employed (non-agricultural)
WKPT	Percent with part-time employment
AVINCO	Average income per person reporting some income
ERN10	Percent of workers earning less than $10 per hour
OFFF	Off-farm earnings as a percent of total farm family income (for economic families with a census-farm operator present)
INCH	Average income growth between 1991 and 1996 (percent)

Table 3.2 Variables used in the study: descriptive statistics by septile

Variable Code	Septile of CD							Mean CD	Median CD	Min CD	Max CD
	1st	2nd	3rd	4th	5th	6th	7th				
POPCH	-4.7	-1.1	0.9	3.1	5.1	8.4	16.3	3.9	2.9	-35.3	26.2
POPL20	24.7	26.6	27.8	28.8	29.8	31.2	37.1	29.4	28.8	20.9	49.4
POPO65	5.0	9.0	10.8	12.2	13.4	14.7	17.9	11.9	12.2	1.9	23.1
IMOLD	4.5	6.9	8.3	9.4	10.6	12.6	17.8	10.0	9.4	1.7	26.2
FERTIL	1.1	1.2	1.2	1.3	1.4	1.5	1.9	1.4	1.3	0.9	2.9
EDUAVE	11.1	11.8	12.2	12.5	12.7	13.1	13.6	12.4	12.5	9.8	14.7
HHBLICO	10.9	13.6	15.3	16.8	18.1	20.4	25.2	17.1	16.8	4.3	46.6
INTRSF	9.9	13.4	15.6	17.6	19.8	23.5	30.0	18.5	17.6	5.1	39.4
UNTOT	5.4	7.9	9.2	10.7	12.5	15.6	25.7	12.4	10.7	2.7	39.4
RENT30	20.7	31.7	36.1	39.9	42.7	46.1	50.3	38.2	39.7	6.0	60.1
HOWN30	8.3	10.8	12.4	13.6	14.7	16.3	20.0	13.7	13.5	4.7	25.6
POWN	50.2	65.7	70.5	73.2	75.5	77.8	83.4	70.8	73.2	16.0	89.7
EMAGR	0.2	0.9	1.9	3.3	5.6	9.9	23.8	6.4	3.3	0.0	45.8
EMPRIM	0.2	0.4	1.0	2.1	3.8	6.8	13.0	3.9	2.0	0.0	25.8
EMTRM	1.0	3.1	4.6	6.3	8.7	12.1	18.9	7.8	6.3	0.0	31.5
EMCMA	1.0	1.9	2.8	4.1	6.2	9.3	14.7	5.7	4.1	0.0	23.5
EMDSE	4.7	6.1	6.9	7.8	8.8	10.3	13.9	8.3	7.7	2.7	20.6
EMSSE	15.4	18.2	19.9	21.5	23.4	26.5	34.9	22.8	21.5	9.6	55.5
PARTEC	53.1	58.1	61.2	63.4	66.0	69.1	74.0	63.5	63.2	48.3	81.4
WKO2	32.7	41.1	45.5	49.4	52.4	55.9	61.9	48.3	49.5	23.1	72.1
MFPART	1.1	1.2	1.2	1.3	1.3	1.4	1.4	1.3	1.3	1.1	1.5
SELF	5.2	7.1	8.2	9.1	10.1	11.4	13.8	9.3	9.1	3.3	19.8
WKPT	16.1	17.9	19.2	20.5	21.8	23.2	25.4	20.6	20.5	13.1	28.7
AVINCO	12,308	13,791	14,702	15,889	17,081	18,206	20,228	16,005	15,888	6,681	25,934
ERN10	19.9	24.2	26.8	30.1	33.3	37.2	45.2	30.9	29.7	15.5	56.3
OFFF	0.0	0.0	28.7	55.3	58.8	63.2	71.1	39.4	55.1	0.0	79.9
INCH	3.7	6.4	8.1	9.6	11.5	13.6	18.2	10.1	9.6	-4.7	27.3

Note: For each row, the CDs are (re)ranked according to the specific variable – CDs are (re)grouped into 7 groups of equal number of CDs (called "septiles") and the average for the specific variable in each septile is reported. All data are expressed as percentage or ratios, except AVINCO is expressed in dollars and EDUAVE is expressed in years (see Table 3.1 and Appendix A for the definition and computation of each variable).

Source: Author's computation based on Census of Population, 1996.

Table 3.3 Variables used in the study: descriptive statistics by CD regional type

Variable Code	Average					Standard deviation				
	PU	IN	RMA	RNM	RN	PU	IN	RMA	RNM	RN
POPCH	7.2	4.0	5.3	2.1	4.5	5.0	4.6	6.5	6.4	11.5
POPL20	27.7	27.9	29.1	28.9	37.4	2.9	1.8	2.8	3.4	6.5
POPO65	9.9	11.4	12.4	13.3	5.1	3.1	2.1	2.9	3.9	2.8
IMOLD	8.1	9.2	10.9	11.0	4.7	1.4	2.8	3.7	4.5	1.8
FERTIL	1.1	1.2	1.3	1.4	1.8	0.1	0.1	0.2	0.2	0.5
EDUAVE	13.5	13.0	12.5	12.1	11.8	0.6	0.6	0.6	0.7	1.0
HHBLICO	19.6	17.8	15.3	18.0	15.8	5.8	3.7	3.7	5.0	5.4
INTRSF	12.2	15.9	18.0	21.8	14.9	2.8	3.3	5.1	6.4	7.9
UNTOT	9.0	10.5	10.7	14.2	16.1	2.0	2.7	5.6	7.7	7.5
RENT30	41.9	43.0	38.6	38.3	23.9	4.0	4.5	9.1	9.0	11.0
HOWN30	17.5	14.8	14.1	12.9	10.1	3.3	2.8	3.5	3.4	2.9
POWN	64.0	65.8	75.1	74.4	52.4	10.9	7.1	5.8	7.5	19.7
EMAGR	0.9	2.8	8.8	8.1	0.7	0.7	2.0	8.0	9.7	1.3
EMPRIM	0.4	1.1	2.1	5.6	9.8	0.9	1.7	2.6	4.7	6.0
EMTRM	5.0	6.9	7.8	8.9	6.1	2.0	4.8	5.8	6.2	6.7
EMCMA	9.3	10.1	6.6	3.6	2.0	3.7	5.6	4.5	2.8	2.7
EMDSE	13.9	9.8	8.1	7.1	6.8	2.7	2.9	2.0	2.1	2.1
EMSSE	23.3	23.5	21.2	22.0	30.8	6.3	5.4	5.6	5.3	9.4
WKO2	54.3	49.9	50.2	45.3	48.0	5.3	6.0	8.8	9.9	9.8
MFPART	1.2	1.2	1.3	1.3	1.3	0.1	0.1	0.1	0.1	0.1
PARTEC	67.8	64.0	64.5	60.9	67.6	3.9	4.3	6.0	6.8	7.9
SELF	9.5	8.3	9.9	9.2	8.1	2.0	1.7	2.5	3.1	3.0
WKPT	19.6	21.3	20.7	20.9	17.9	2.2	2.3	2.8	3.3	2.5
AVINCO	19459	17357	15978	14857	16026	2204	2068	1973	2208	3577
ERN10	22.4	25.9	32.5	34.5	23.4	3.6	4.0	6.2	8.4	5.0
OFFF	49.7	49.0	47.5	34.7	6.4	29.1	24.4	24.6	29.8	21.3
INCH	6.4	8.1	9.5	11.6	11.9	2.8	3.0	4.3	4.5	6.2

Note: All data are expressed as percentage or ratios, except AVINCO is expressed in dollars and EDUAVE is expressed in years (see Table 3.1 and Appendix A for the definition and computation of each variable). The regional types are those used by Statistics Canada: predominantly urban (PU), intermediate (IN), rural metro-adjacent (RMA), rural non-metro-adjacent (RNM) and rural northern (RN).

Source: Author's computation based on Census of Population, 1996.

4. The identification of socio-economic dimensions

This section presents the results of the factor analysis. Six factors were identified, which account for about 78 percent of the variance in the data set. This means that 78 percent of the variance of the 27 indicators used in the factor analysis is explained by 6 summary variables (factors). Given the wide variability in performance and outcomes across census divisions, our 6 dimensions provide an excellent "fit" with the data. For this type of analysis, we consider we have good results if only 22 percent of variability in the data remains unexplained due to "other" or random effects.

The factors are named on the basis of the variables that are strongly associated with each. Hence, the name attributed to the factors is somewhat subjective, but the name tends to reflect the nature of the variables strongly associated with each. The names used for the six factors are as follows (in brackets is the percent of variance explained by the factor):

- Labour force and economic attributes (26.3 percent)
- Remote and agro-rural attributes (18.7 percent)
- Demographic and labour force attributes (14.7 percent)
- Employment attributes: complex manufacturing versus primary production attributes (7.4 percent)
- Employment attributes: traditional manufacturing versus government employment attributes (5.8 percent)
- Demographic dynamics attributes (5 percent)

Overall, it appears that the nature of the factor, that is the unobservable dimension that is captured, is conceptually consistent, meaningful and suitable for interpretation. Each factor appears to be strongly associated with a few variables, while still having a considerable effect on a set of other variables, generally this is what would be expected. Because of this combination of variables, some of the factors are more ‘diagnostic’ in that the variables explained are indicative of desirable or undesirable attributes or features of a locality (e.g., unemployment rate, high income, etc.). Other factors tend to be more ‘descriptive’ in nature, meaning that the variables explained do not have a strong normative connotation.

The remainder of this section discusses the six factors in detail. Three summary tables are reported for each factor. The first table displays the factor loadings. The second table indicates the average value of the variables that load on this factor by septiles of factor scores. The final table reports the distribution of CDs by size of factor score for each CD regional type, using the classification adopted by Statistics Canada. For the interested reader, Box 1 explains how to interpret the figures displayed in each type of table.

Box 1. How to interpret the results

For each factor, the results are summarized in three tables and a map. Below is an explanation of how to interpret the data reported in these tables and the map.

The *Factor loadings* table provides information on the nature of the factor. Recalling that the factor score "summarizes" the value of a set of variables, the variables listed in this table are those explained by the factor. The variables are ranked on the basis of the absolute value of the factor loading that is the strength of the association regardless of the sign of this association. But it is important to notice that some of the variables present a positive association with the factor score (meaning a high factor score corresponds to a high value of the variable, and vice versa) while others have a negative association (meaning a high factor score corresponds to a low value of the variable, and vice versa). The top part of the table lists the variables that are strongly associated with the factor (loading greater that 0.5 in absolute value). These tend to be more relevant in explaining the nature of the factor. The bottom part of the table reports the variables that show a weaker association with the factor (between 0.15 and 0.49 in absolute terms).

The second table is *Average values by factor score septiles*. This table indicates the average value of the variable associated with the factor by septile of factor score. (We use the term "septile" because the CDs are ranked and grouped into 7 groups with an equal number of CDs in each group). A given CD is assigned to a column based on the size of the factor score for that CD. The rows in the table show the average value of the given variable for CDs assigned to each column. Thus, each column of this table provides a profile of a grouping of CDs (based on the factor score values) with respect to the variables that are associated with the factor itself. Note that the septile categories are the same as those used in the map. As can be expected, the average values of the indicators that have a high loading tend to diverge more than those of the indicators with a small loading. This simply means that the factor score captures in a better way the total variability of the indicator with a high loading, while it does not capture well (or not at all) the variability of indicators that have a low association with it. To gain an appreciation of how much of the variability of each variable is captured by the factor, it is also possible to compare the average values reported in this table with the average by the variable's septiles reported in Table 3.2. For analogous septile categories, the closer are the values of the two tables the better the factor captures the variability of the indicator.

The last table is the *Percent distribution of CDs by size of factor score within each type of region*. The regional type used here are those defined by Statistics Canada (du Plessis *et al.*, 2002). Each row of the table sums to 100 percent. This table indicates whether the distribution of CDs according to their factor scores shows any association with the prevailing definition of rural types. Generally, if this was the case, one would expect to find a higher frequency of observations along the diagonal of the table or, alternatively, a high concentration of observations in a few cells for each regional type.

Finally, for each factor, the factor values score are mapped. The maps use the same groupings (septiles) that are employed for the tables. The maps allow an assessment of spatial patterns, regardless of any predefined regional type, and provides an understanding of the broad regional distribution of the factor scores (and consequently of the values of the indicators associated with the factor).

4.1. Labour force and economic attributes

This factor is strongly associated with 9 variables and it explains about 26 percent of the total variance, which makes it the main factor in terms of total variance explained. Table 4.1 lists the nine variables with a high loading on this factor (top part of the table). Six of these have a positive loading, which are participation rate, percent of families with two or more members in the labour force, average income per person, average years of schooling for population 25 to 54 years of age, percent (non-agricultural) self-employed, and percent with dynamic services employment. The value of these variables tends to be high when the factor score is high. Three

variables have a negative loading – these are social transfer income as a percent of total income, percent of persons in low-income economic families, and total unemployment rate. Thus, the value of these three indicators tends to move in the opposite direction of the previous six indicators. Ten other variables have a moderate but still sizeable loading as listed in the bottom part of the table and these include, in particular, other income-related variables and labour force indicators.

The nature of this factor appears rather clear: it captures a broad range of mainly economic attributes that describe the overall performance and economic strength of the CD. This is evidenced by Table 4.2, which shows the average values of the indicators by septile of factor score. Each column presents a profile of a grouping of CDs, based on the factor score values. CDs falling in the highest septile have on average a participation rate of 73 percent and about 60 percent of the families have two or more members in the labour forces. The average income is about $19,000, only 11 percent of which is government transfer income, the total unemployment rate is 7 percent and 12 percent of the individuals are in low-income families. The population tends also to have a higher educational level (13 years of schooling on average) and both dynamic services employment and non-agricultural self-employment are more relevant than for the other grouping of CDs. At the opposite end, the first column of the table presents a rather contrasting profile. CDs with a low factor score have a participation rate of only 54 percent and only 34 percent of families have two or more members in the labour force. The average income is only $12,800, of which 28 percent is government transfers. About 24 percent of the individuals live in low-income families and the unemployment rate is 22 percent, while the educational level is also lower (11 years of schooling on average). The bottom part of the table shows the average values for the other variables that are moderately associated with the factor. The differences are not captured as well as for the previous variables, but the patterns are still rather clear and consistent.

Table 4.3 shows the percent distribution of CDs across septile groups of factor scores within each type of region. Predominantly urban CDs are more likely to record higher factor score values but for the other regional types, the distribution of CDs is spread across septiles of factor scores, indicating a greater diversity of performance. However, for intermediate and rural metro-adjacent CDs, most are concentrated in the upper part of the distribution (between the 3rd and 6th or 7th septile); while for rural non-metro-adjacent CDs, the lower half of the distribution is considerably heavier (1st to 4th septile). It is interesting to note that rural northern CDs present two peaks at the opposite ends of the distribution. Over 17 percent of the rural northern CDs fall into the first septile of the factor score, which means poor economic performance and labour force attributes, while at the opposite end about 26 percent of the rural northern CDs show a higher than average economic performance. Hence, remoteness does not necessarily imply a lower economic performance.

Finally, Map 4.1 shows the spatial pattern of the factor score for *labour force and economic attributes* in 1996. The mapping uses the same septile breakdown presented in Tables 4.2 and 4.3. There are a number of clusters of CDs with high positive scores. In the eastern part of Canada, where the administrative geography of CDs permits a more detailed spatial representation, the clusters tend to be aggregate around major urban cores, but the urban centre does not necessarily present the highest score in the region. Examples are Québec City, Montréal and Toronto. In southern Ontario, a continuous cluster of high score CDs is found in

the area surrounding Toronto and stretching south-west to Lambton and Essex County. To some extent, this overlaps with the *Rural nirvana* regional-type identified by Hawkins (1995). But a notable difference is represented by the band of CDs extending along a north-south axis between the urban poles of Ottawa and Toronto which, according to the results of the factor analysis, presents markedly lower characteristics in terms of economic and labour market attributes. High scores (i.e., high labour force and economic attributes) are also characteristics of a cluster of CDs surrounding Winnipeg and the southern fringe of Saskatchewan, while they cover a large part of Alberta and British Columbia.

In contrast, low factor scores are concentrated in the Atlantic Provinces, particularly in Newfoundland and Labrador and the northern part of New Brunswick. They also characterize many CDs in Québec (although the pattern is scattered here) outside the urban fringe of Montréal and Québec City. Also the CDs located in the north of Manitoba and Saskatchewan and CDs in the territories without a capital city share a similar low performance on this economic dimension.

Table 4.1 Labour force and economic attributes: Factor loadings

Code	Variable	Loading	
PARTEC	Participation rate	0.901	
WKO2	Percent of families with two or more members in the labour force	0.876	
INTRSF	Social transfer income as a percent of total average income		-0.860
AVINCO	Average income per person	0.792	
HHBLICO	Percent of persons in low-income economic families		-0.773
UNTOT	Total unemployment rate		-0.714
EDUAVE	Average years of schooling for population 25 to 54 years of age	0.682	
SELF	Percent self-employed (non-agricultural)	0.555	
EMDSE	Percent with dynamic services employment	0.505	
ERN10	Percent earning less than $10 per hour		-0.338
OFFF	Off-farm earnings as a percent of total farm family income	0.328	
EMTRM	Percent traditional manufacturing employment		-0.327
MFPART	Male participation rate over female participation rate		-0.288
IMOLD	Senior in-migration rate		-0.257
POPCH	Population growth between 1991 and 1996	0.255	
POPO65	Percent of population age 65 years of age and over		-0.239
RENT30	Percent of households with gross rent equal to or greater than 30 percent of household income		-0.237
WKPT	Percent part-time employment	0.213	

Source: Author's computation.

Table 4.2 Labour force and economic attributes: Average values by factor score septiles

Variable code	Factor score septile							CDs Average
	1st Lowest	2nd	3rd	4th Middle	5th	6th	7th Highest	
PARTEC	54.2	58.2	61.2	64.1	65.0	68.6	73.2	63.5
WKO2	34.1	41.7	46.8	49.6	51.0	54.9	60.2	48.3
INTRSF	28.0	22.6	20.5	17.4	16.0	14.0	10.9	18.5
AVINCO	12852	14259	14875	16167	17086	17743	19053	16005
HHBLICO	23.8	19.2	17.3	17.4	15.2	15.1	12.1	17.1
UNTOT	22.4	15.9	11.8	10.8	9.6	8.9	7.3	12.4
EDUAVE	11.4	11.9	12.2	12.5	12.8	13.0	13.1	12.4
SELF	6.5	7.9	8.9	9.4	10.3	10.5	11.2	9.3
EMDSE	6.6	6.5	7.0	9.0	8.4	9.7	10.9	8.3
ERN10	34.3	33.3	34.5	29.6	28.3	29.0	27.1	30.9
OFFF	13.8	32.1	43.8	43.7	46.1	51.4	44.7	39.4
EMTRM	9.9	10.5	10.1	8.5	5.9	5.0	4.3	7.8
MFPART	1.32	1.31	1.27	1.26	1.22	1.22	1.26	1.27
IMOLD	10.8	11.4	11.0	9.8	10.7	9.2	7.0	10.0
POPCH	1.3	1.0	2.4	5.1	5.6	5.7	6.5	3.9
POPO65	12.3	13.0	13.2	11.8	12.6	11.2	8.9	11.9
RENT30	41.0	39.3	37.9	39.3	40.3	38.5	30.8	38.2
WKPT	18.6	20.0	20.9	20.2	22.5	21.9	19.9	20.6

Source: Author's computation.

Table 4.3 Labour force and economic attributes: Percent distribution of CDs by septile class of factor scores for each regional type

Regional type (CD units)	1st Lowest	2nd	3rd	4th Middle	5th	6th	7th Highest
Predominantly urban regions (25)	4.0	4.0	0	24.0	12.0	32.0	24.0
Intermediate regions (37)	2.7	18.9	21.6	8.1	16.2	21.6	10.8
Rural metro-adjacent regions (86)	10.5	7.0	17.4	16.3	17.4	16.3	15.1
Rural non-metro-adjacent regions (117)	22.2	20.5	13.7	14.5	11.1	7.7	10.3
Rural northern regions (23)	17.4	13.0	8.7	8.7	17.4	8.7	26.1

Source: Author's computation.

4.2. Remote and agro-rural attributes

This factor is strongly associated with 9 variables and explains almost 19 percent of the total variance. The variables are listed in Table 4.4. Five of them show a positive loading – fertility rate, employment in agriculture, percent of population below age 20, and two income indicators measuring the prevalence of population with earning less than $10 per hour and the income growth between 1991 and 1996. Four indicators have a negative loading (i.e. tend to vary in the opposite direction of the previous five). These are two housing indicators measuring the percent of households with rent or owner's gross housing costs above 30 percent of household income, the educational level, and average income. Another seven variables are moderately associated with this factor – their sign appears coherent with the previous indicators and are reported in the bottom part of Table 4.4.

Table 4.5 shows the average value of the variables associated with the factor by septiles of factor scores. The same classification in septiles is used in Map 4.2, where the lowest septile correspond to the dark blue and the highest correspond to the dark red. The highest septile includes CDs with a fertility rate of 1.86 (ratio between persons below age 19 and women between 25 and 54) and a high share (35 percent) of population below age 20 years of age. Agricultural employment is on average 20 percent, with only 25 percent of households renting or owning a house with housing costs higher than 30 percent of their income. But income is low (on average about $14,200), there are on average 40 percent of income earners who earn less than $10 per hour and also educational levels are low. Nevertheless, average income growth between 1991 and 1996 has been higher than average, growing about 14 percent in nominal terms. The opposite profile is shown in the first column of the tables.

The nature of "*remote and agro-rural attributes*", and the reason why the factor was named in this way, is evidenced in Table 4.6, which displays the percent distribution of the CD scores by regional type. A gradient is clearly evident in the table. Predominantly urban CDs are concentrated in the first two septiles. There is no predominantly urban CD beyond the second septile of factor scores. Intermediate CDs are also largely concentrated in the lower half of the distribution. In contrast, as one moves from rural metro-adjacent to rural non-metro-adjacent to rural northern CDs, the distribution shifts steadily toward the highest septiles, i.e. those that denote a prevalence of remote and agricultural attributes.

This pattern is also evidenced by Map 4.2 which shows the mapping of the factor score by septiles. In all of central Canada and a large part of Alberta, the only CDs that fall in the bottom septiles (i.e., a lack of remote and agro-rural attributes) are the urban CDs. The northern part of B.C. presents similar characteristics. A second major cluster of CDs with a high factor score is located in south-east Québec, between the St. Lawrence River and the border with the U.S. While in Ontario, the only CDs with high scores are located in the agricultural heartland of the province, surrounding Huron County. In contrast, is it worth noting that a lack of these attributes (areas that are in blue shades) characterise most of the CDs in the Atlantic Provinces, the CDs located immediately north of the St. Lawrence River in Québec, and a cluster of CDs in southern B.C. It should be stressed that this does not mean that these CDs are in any way "less rural", but rather points to the fact that they share different

characteristics of what seems to be a prevailing type of rural across Canada, which is captured by this dimension.

Some further remarks should be made about the characteristics of this factor. First, the factor appears to explain demographic and housing characteristics and not only agricultural employment. In fact, rural northern CDs have an average share of agricultural employment that is lower than urban CDs (this is evidenced by Table 3.3). Hence, this dimension does not describe *only* a prevailing type of employment structure, even though there are three employment by sector indicators which are associated with this factor to various degrees. Rather the factor captures a combination of demographic, housing, employment and income characteristics that prevails in a large part of what is commonly considered rural Canada. However, another part of rural Canada does not share these "remote and agro-rural" characteristics.

Second, the factor seems to capture a different spatial variation of income than that associated with the previous factor. Looking at the patterns of Maps 4.1 and 4.2 and the variables associated with the two factors, the results could appear somehow incongruent, particularly for the association of the average income indicator (AVINCO). To explain this apparent contradiction it should be recalled that the factor model is in essence a regression model where observed variables are regressed against factor scores. The coefficients (loadings) of each variable equation indicate the effect of the observed variables on the factors. In the case of the average income indicator, the first factor (a dimension that appears related to general economic strength of the CD) explains a substantial part of the income variation across CDs. But after accounting for this factor, another part of the variation of income is explained by this second factor (which appears to be more clearly related to a rural-urban distinction). Therefore, the results suggest that there are two major dimensions of income variability across CDs. The first one is predominantly defined along the boundaries of macro-regions, which to some extent overlap with the provinces. The most striking divides are between Eastern Canada and the northern parts of central Canada, on the one hand, and southern Ontario and Western Canada, on the other hand. The second dimension is more clearly defined along an urban-rural divide and highlights some of the main characteristics of remote and agro-rural Canada.

Finally, rural income trends appear to be properly described by the *remote and agro-rural* factor also for another reason. It is interesting to note the association between the lower level of income and a higher growth between 1991 and 1996. Although this analysis did not focus on changes across time, a study conducted in parallel to this one showed that the economic recession period of 1991-1994 did not hit rural areas as much as the urban core (Alasia, 2003). Hence, rural regions recorded income growth rates generally above average during that period, and this trend appears to be captured by this factor.

Table 4.4 Remote and agro-rural attributes: Factor loadings

Code	Variable	Loading	
FERTIL	Fertility rate	0.850	
RENT30	Percent of households with gross rent equal to or greater than 30 percent of household income		-0.733
EMAGR	Percent agricultural employment	0.694	
HOWN30	Percent of households with owner's gross housing costs equal to or greater than 30 percent of household income		-0.693
POPL20	Percent of population less than 20 years of age	0.688	
ERN10	Percent earning less than $10 per hour	0.590	
INCH	Average income growth between 1991 and 1996	0.521	
EDUAVE	Average years of schooling for population 25 to 54 years of age		-0.515
AVINCO	Average income per person		-0.504
EMDSE	Percent dynamic services employment		-0.480
MFPART	Male participation rate over female participation rate	0.423	
EMCMA	Percent complex manufacturing employment		-0.397
WKO2	Percent of families with two or more members in the labour force	0.248	
IMOLD	Senior in-migration rate		-0.241
PARTEC	Participation rate	0.168	
UNTOT	Total unemployment rate		-0.165

Source: Author's computation.

Table 4.5 Remote and agro-rural attributes: Average values by factor score septiles

Variable code	Factor score septile							CDs Average
	1st Lowest	2nd	3rd	4th Middle	5th	6th	**7th Highest**	
FERTIL	1.14	1.21	1.23	1.29	1.37	1.45	1.86	1.36
RENT30	45.9	43.6	43.3	40.3	37.0	32.2	24.7	38.2
EMAGR	1.2	2.3	2.9	3.4	5.8	9.8	19.6	6.4
HOWN30	18.1	15.7	14.8	12.7	13.0	11.9	9.5	13.7
POPL20	26.3	27.7	27.7	28.5	29.6	30.7	35.0	29.4
ERN10	25.1	25.6	29.9	30.6	30.4	34.3	40.1	30.9
INCH	6.2	8.6	9.2	10.5	11.2	10.9	14.2	10.1
EDUAVE	13.3	12.9	12.4	12.2	12.2	12.2	11.9	12.4
AVINCO	18425	17435	15772	15481	15503	15234	14201	16005
EMDSE	11.8	9.7	7.8	7.2	7.4	7.2	7.0	8.3
MFPART	1.17	1.21	1.25	1.28	1.32	1.31	1.31	1.27
EMCMA	8.4	7.9	5.6	5.7	5.1	4.9	2.2	5.7
WKO2	48.4	48.6	44.3	43.4	46.2	52.3	55.3	48.3
IMOLD	11.2	10.9	11.0	9.8	10.3	8.5	8.3	10.0
PARTEC	63.1	63.9	61.3	61.3	62.1	66.1	66.7	63.5
UNTOT	11.3	11.7	14.7	16.6	13.9	9.7	8.6	12.4

Source: Author's computation.

Table 4.6 Remote and agro-rural attributes: Percent distribution of CDs by the septile class of factor scores for each regional type

Regional type (CD units)	1st Lowest	2nd	3rd	4th Middle	5th	6th	7th Highest
Predominantly urban regions (25)	64.0	36.0	0	0	0	0	0
Intermediate regions (37)	35.1	29.7	13.5	13.5	8.1	0	0
Rural metro-adjacent regions (86)	9.3	11.6	15.1	18.6	11.6	22.1	11.6
Rural non-metro-adjacent regions (117)	3.4	9.4	18.8	15.4	19.7	13.7	19.7
Rural northern regions (23)	0	0	4.3	13.0	21.7	26.1	34.8

Source: Author's computation.

4.3. Demographic and labour force attributes

The factor named *demographic and labour force attributes* explains 14.7 percent of the total variability in the data set and presents 6 variables strongly associated with it, all with positive loading. The variables are listed in Table 4.7. High factor scores are associated with a high proportion of population age 65 years of age and over, a high percent of seniors moving into the CD over the previous five years, high shares of part-time and self-employment, high shares of households owning the house of residence, and high shares of individuals with earnings below $10 per hour.

The profile of the CDs by septile of factor score is displayed in Table 4.8. Also in this case, comparing these figures with the average values by the variable's septiles reported in Table 3.2 shows that the factor picks up most of the variability of the 6 variables with a high loading. For instance, the 1st septile has 6 percent of its population above age 65 (versus 17 percent for the 7th septile) and the share of senior in-migrants is 6 percent (versus 15 percent for the 7th septile). Part-time employment is 17 percent and non-agricultural self-employment is 8 percent (versus 25 and 12 percent respectively for the 7th septile). Low wage earners represent 24 percent of the 1st septile (versus 37 percent), but house owners represent 58 percent of the 1st septile (versus 76 percent for the top septile).

Table 4.9 shows the distribution of CDs across septile groups of factor scores within each type of region. The table presents an interesting pattern. Low factor scores (i.e., low share of elderly, part-time, self-employment, and low wage earners) are a dominant feature of predominantly urban CDs but also of the rural northern CDs, at the opposite side of the regional type spectrum. Intermediate CDs show also a distribution skewed toward low scores, even though the range covers all the septiles. In contrast, rural metro-adjacent and rural non-metro-adjacent CDs present distributions skewed toward high factor scores. But also in this case the variation within the regional type is noticeable, as groupings of these types of CDs are found in each category of factor scores. The dual connotation (urban-remote) of the low factor scores distribution is evident when the factor scores are mapped (Map 4.3).

Table 4.7 Demographic and labour force attributes: Factor loadings

Code	Variable	Loading	
POPO65	Percent of population age 65 years of age and over	0.826	
IMOLD	Senior in-migration rate	0.765	
WKPT	Percent part-time employment	0.755	
SELF	Percent self-employed (non-agricultural)	0.582	
POWN	Percent of households owning their home	0.511	
ERN10	Percent earning less than $10 per hour	0.500	
POPL20	Percent of population less than 20 years of age		-0.434
EMAGR	Percent agricultural employment	0.394	
RENT30	Percent of households with gross rent equal to or greater than 30 percent of household income	0.365	
OFFF	Off farm earnings as a percent of farm family income	0.319	
INTRSF	Social transfer income as a percent of total average income	0.253	
EMSSE	Percent non-market services employment		-0.249
MFPART	Male participation rate over female participation rate		-0.243
PARTEC	Participation rate		-0.190
UNTOT	Total unemployment rate		-0.184
EDUAVE	Average years of schooling for population 25 to 54 years of age	0.182	

Source: Author's computation.

Table 4.8 Demographic and labour force attributes: Average values by factor score septiles

Variable code	Factor score septile							CDs Average
	1st Lowest	2nd	3rd	4th Middle	5th	6th	**7th Highest**	
POPO65	6.0	10.2	10.5	12.1	13.1	14.0	17.1	11.9
IMOLD	5.6	8.2	8.2	9.6	11.0	11.9	15.4	10.0
WKPT	17.3	18.8	19.6	20.1	21.1	22.3	24.7	20.6
SELF	7.9	7.5	8.7	8.5	9.3	10.5	12.3	9.3
POWN	57.7	69.1	70.2	72.2	74.5	75.9	76.2	70.8
ERN10	24.3	26.7	28.4	31.4	32.9	35.1	37.3	30.9
POPL20	34.4	29.0	28.8	28.4	28.2	29.1	27.9	29.4
EMAGR	1.9	3.3	4.5	5.3	7.0	10.3	12.8	6.4
RENT30	29.1	37.6	39.4	39.7	41.2	39.5	40.6	38.2
OFFF	18.2	33.6	42.0	39.9	44.8	45.7	51.6	39.4
INTRSF	13.7	19.0	17.8	19.7	19.8	19.3	20.2	18.5
EMSSE	27.6	21.0	23.1	21.9	22.5	21.3	21.9	22.8
MFPART	1.29	1.31	1.27	1.26	1.25	1.26	1.23	1.27
PARTEC	67.8	62.2	63.9	62.2	62.6	63.9	62.0	63.5
UNTOT	11.9	15.0	13.1	13.3	12.4	11.6	9.1	12.4
EDUAVE	12.2	12.2	12.6	12.4	12.4	12.4	12.6	12.4

Source: Author's computation.

Table 4.9 Demographic and labour force attributes: Percent distribution of CDs by septile class of factor scores for each regional type

Regional type (CD units)	1st Lowest	2nd	3rd	4th Middle	5th	6th	7th Highest
Predominantly urban regions (25)	40.0	16.0	20.0	16.0	4.0	0	4.0
Intermediate regions (37)	8.1	29.7	21.6	24.3	10.8	2.7	2.7
Rural metro-adjacent regions (86)	7.0	7.0	11.6	14.0	25.6	25.6	9.3
Rural non metro-adjacent regions (117)	5.1	12.0	14.5	14.5	12.0	15.4	26.5
Rural northern regions (23)	69.6	26.1	4.3	0	0	0	0

Source: Author's computation.

4.4. Employment attributes: complex manufacturing versus non-agricultural primary production attributes

This factor is strongly associated with four variables and explains about 7.4 percent of the total variability in the data set. Table 4.10 lists these variables. They primarily describe the employment structure of the CD and in particular the opposing structure between employment in complex manufacturing, with a positive loading, and employment in the "other" primary sector (i.e., other than agriculture), with a negative loading. The total unemployment rate is also explained by the factor, and the association is negative, while off-farm employment by farm family members shows a positive loading. Another ten variables show a moderate relationship with this factor (bottom of Table 4.10); among these are income growth between 1991 and 1996 (negative loading), employment in agriculture and families with two or more members in the labour force (both with positive loading). It should be pointed out that, because of the loading with opposite signs of the two main employment variables ("other" primary and complex manufacturing), the score outcomes are more than just a reflection of the proportional levels of employment in the two sectors. The scores are likely to reflect also the ratio between the two variables. Hence, CDs with extremely low levels of "other" primary employment and only modest levels of employment in complex manufacturing are likely to record a positive score on this factor, and vice versa.

The average values of these variables by septiles of factor score are reported in Table 4.11. The CDs falling in the bottom septile (dark blue on the map) have on average 12 percent of their employment in the "other" primary sector (i.e., excluding agriculture), while only 2 percent are employed in complex manufacturing (and 2 percent in agriculture). They also have an unemployment rate of 20 percent and only 12 percent of the farm family income is generated by off-farm employment. Thus, the area showing these characteristics can be properly described as resource-based communities. On the other hand, the top septile includes CDs with a low share of employment in the "other" primary sector (1 percent) and about 12 percent of employment in complex manufacturing. Also employment in agriculture is higher

than average (9 percent) but is far below the average value of the top septile computed on the variable range, which is 23 percent (see Table 3.2).

Table 4.12 displays the distribution of CDs across septiles of factor scores within each type of region. Not surprisingly, this dimension also presents an urban to rural gradient. Predominantly urban CDs have factor scores concentrated mainly in the top three septiles and no predominantly urban CD falls in the two lowest septiles. In contrast, rural northern CDs have factor scores concentrated in the bottom two septiles. The intermediate, rural metro-adjacent, and rural non-metro-adjacent CDs indicate a gradual shift of the factor score distribution from the top septiles to the bottom ones; yet the range of variation for these three regional types is substantially larger than for the predominantly urban regions and rural northern regions.

Map 4.4 reveals the spatial distribution of factor scores by septiles. The CDs reported in dark blue are those with low factor scores, these are the resource-based regions (relatively intensive in "other" primary employment). As one could expect, many of the CDs with low factor scores are clustered along the Atlantic coast. Another area of relatively low scores stretches from northern Québec to northern Manitoba. In Alberta low scores are recorded for north-eastern Alberta (which includes the tar sands project at Fort McMurray) while blue shades cover a cluster of CDs in northern B.C. from the inland to the coast. In contrast, high factor scores (i.e., relatively high employment in complex manufacturing and a lack of "other" primary employment) are located in most of southern Ontario and Québec. Also the entire province of Saskatchewan, southern Manitoba, and a large part of south-eastern Alberta record low scores. However, for these CDs, the results are also due to the relatively high ratio between the employment in the two sectors, determined by a lack of "other" primary employment more than by high levels of complex manufacturing employment. Relatively high employment in agriculture also affects this result.

Finally, it is interesting to note that the areas that score high on this dimension are not necessarily in the same range for the first factor (labour force and economic attributes). Thus, a resource-based economy does not necessarily imply low performance as measured by the labour and economic dimension discussed above. In this respect, the set of socio-economic dimensions generated by the factor analysis allows a clearer identification of economic structures and performance than appears to be possible with other classification procedures.

Table 4.10 Employment attributes, complex manufacturing versus non-agricultural primary production: Factor loadings

Code	Variable	Loading	
EMPRIM	Percent other primary employment		-0.802
EMCMA	Percent complex manufacturing employment	0.627	
UNTOT	Total unemployment rate		-0.540
OFFF	Off-farm earnings a percent of total farm family income	0.525	
INCH	Average income growth between 1991 and 1996		-0.373
EMAGR	Percent agricultural employment	0.352	
WKO2	Percent of families with two or more members in the labour force	0.345	
HOWN30	Percent of households with owner's gross housing costs equal to or greater than 30 percent of household income	0.300	
POPO65	Percent of population age 65 years of age and over	0.233	
EMSSE	Percent non-market services employment		-0.231
EMDSE	Percent dynamic services employment	0.215	
EDUAVE	Average years of schooling for population 25 to 54 years of age	0.202	
ERN10	Percent earning less than $10 per hour	0.195	
INTRSF	Social transfer income as a percent of total average income		-0.190

Source: Author's computation.

Table 4.11 Employment attributes, complex manufacturing versus non-agricultural primary production: Average values by factor score septiles

Variable code	Factor score septile							CDs Average
	1st Lowest	2nd	3rd	4th Middle	5th	6th	7th Highest	
EMPRIM	12.0	5.8	3.4	2.7	1.7	0.9	0.6	3.9
EMCMA	2.1	2.7	3.6	5.6	5.5	8.6	11.8	5.7
UNTOT	20.2	14.5	12.9	10.7	9.6	9.1	9.5	12.4
OFFF	11.8	26.3	29.6	50.1	48.4	54.2	54.9	39.4
INCH	12.3	12.3	10.5	10.0	10.2	8.1	7.3	10.1
EMAGR	1.6	2.7	4.6	7.5	10.7	9.3	8.6	6.4
WKO2	41.5	46.7	46.2	49.8	52.0	52.0	50.3	48.3
HOWN30	11.9	12.6	13.9	14.2	12.8	14.7	15.6	13.7
POPO65	10.1	10.2	12.0	12.9	12.4	13.3	12.0	11.9
EMSSE	22.7	25.2	25.1	23.3	24.1	21.3	17.6	22.8
EMDSE	6.4	8.3	8.7	8.9	8.5	8.9	8.4	8.3
EDUAVE	11.9	12.4	12.4	12.6	12.5	12.7	12.3	12.4
ERN10	28.7	27.5	30.0	31.8	34.4	32.0	31.7	30.9
INTRSF	21.9	17.8	18.5	18.0	18.1	17.3	17.7	18.5

Source: Author's computation.

Table 4.12 Employment attributes, complex manufacturing versus non-agricultural primary production: Percent distribution of CDs by septile class of factor scores for each regional type

Regional type (CD units)	1st Lowest	2nd	3rd	4th Middle	5th	6th	7th Highest
Predominantly urban regions (25)	0	0	4.0	12.0	24.0	32.0	28.0
Intermediate regions (37)	0	8.1	5.4	16.2	13.5	24.3	32.4
Rural metro-adjacent regions (86)	5.8	10.5	17.4	12.8	18.6	16.3	18.6
Rural non metro-adjacent regions (117)	22.2	17.1	17.1	18.8	11.1	8.5	5.1
Rural northern regions (23)	43.5	39.1	13.0	0	4.3	0	0

Source: Author's computation.

4.5. Employment attributes: traditional manufacturing versus government employment attributes

This factor explains almost 6 percent of the total variance and it is strongly associated with four variables that describe primarily the employment structure of the locality. In particular, this factor points to the way in which non-market services (i.e., government services, educational services and health services) tend not to be in the same CDs as traditional manufacturing employment (which includes processing of agricultural products, fish, wood, minerals, gas and oil). Another ten variables are moderately associated with the factor. All the variables are listed in Table 4.13.

Table 4.14 displays the average value of the variables associated with the factor, by factor score septiles. CDs with high positive factor scores (7th septile) show a lower than average employment share in non-market services (17 percent versus 32 percent for the bottom septile) and a higher than average employment in traditional manufacturing (17 percent versus 3 percent for the bottom septile). High factor loadings are also reported for the participation shares of males relative to females (1.4 ratio versus 1.18 ratio for the bottom septile) and the average share of households owning their home (75 percent versus 54 percent in the 1st septile). In addition to this, educational attainment is lower than average and so is the employment in dynamic service employment. As was indicated for the previous factor, however, because of the loading with opposite signs of two employment variables (non-market services and traditional manufacturing), the score outcomes are likely to reflect also the ratio between the two variables. CDs with extremely low levels of non-market services employment and only modest levels of employment in traditional manufacturing may record a positive score on this factor, and vice versa.

Nevertheless, the nature of the factor seems to describe more than simply the structure of employment by sector. Positive scores are indicative of a more "traditional" employment structure not only because of the traditional manufacturing employment, but also because of

the relative participation rate of the two genders. Female participation in the formal economy, relative to male participation, appears particularly low where traditional manufacturing employment prevails. Educational attainment is also generally lower, while the fertility rate tends to be higher.

Table 4.15 shows the distribution of CDs across groups of factor scores for each type of region. As it was also observed for the factor labelled *demographic and labour force attributes*, the two extreme categories (predominantly urban and rural northern regions) present a somewhat similar pattern. For both regional types, the distribution of CDs by size class of factor scores shows a concentration in the lowest septile (characterised by a relatively higher share of employment in non-market or government services). Yet, the characteristics of the rural northern CDs are more polarised than for predominantly urban CDs. Intermediate CDs present a factor score distribution skewed toward the bottom septiles, while rural metro-adjacent and rural non-metro-adjacent CDs have a distribution skewed toward the top septiles. Nevertheless, for these regional types the diversity of conditions appears considerable.

The spatial pattern of the factor score distribution is showed in Map 4.5. Across eastern Canada, the pattern appears somewhat more scattered than for the previous maps. Given the nature of the factor and the fact that eastern CDs tend to be smaller, this does not appear surprising. The factor scores reflect the rapidly shifting employment structure of regions that are relatively close. However, a major cluster of CDs with high factor scores (traditional employment structure) are found in southern Québec outside the major urban cores; while a small grouping of CDs with these characteristics is also found in south-west Ontario. In contrast, CDs with a prevalence of non-market employment and a less traditional employment structure (dark blue) are found across all the Territories, central Saskatchewan, and particularly the CDs surrounding Regina and Saskatoon. In Manitoba this type of CD is located north and west of Winnipeg, including the city of Winnipeg, but excluding the CDs immediately surrounding Winnipeg. Finally, part of northern Ontario, most of south-east of Ontario, and the core metropolitan areas along the Highway 401 corridor also present low factor scores.

Note that CDs with a national, provincial or territorial capital city (i.e. a strong presence of government employment) are dark blue – that is, they have low factor scores on the *traditional manufacturing versus government employment* dimension. Conversely, CDs that are relatively intensive in fish processing (north and south shores of Newfoundland and a few other counties in the other Atlantic Provinces) and wood processing (Atlantic Provinces, south-eastern Québec, north-western Alberta and the interior of British Columbia) indicate high factor scores.

Table 4.13 Employment attributes, traditional manufacturing versus government employment attributes: Factor loadings

Code	Variable	Loading	
EMSSE	Percent non-market services employment		-0.792
EMTRM	Percent traditional manufacturing employment	0.725	
POWN	Percent of households owning their home	0.574	
MFPART	Male participation rate over female participation rate	0.566	
EDUAVE	Average years of schooling for population 25 to 54 years of age		-0.314
HHBLICO	Percent persons in low-income economic families		-0.279
EMDSE	Percent dynamic services employment		-0.269
FERTIL	Fertility rate		-0.221
EMCMA	Percent complex manufacturing employment	0.207	
WKPT	Percent part-time employment		-0.202
POPL20	Percent of population less than 20 years of age		-0.192
IMOLD	Senior in-migration rate	0.172	
ERN10	Percent earning less than $10 per hour	0.165	
AVINCO	Average income per person		-0.163

Source: Author's computation.

Table 4.14 Employment attributes, traditional manufacturing versus government employment attributes: Average values by factor score septiles

Variable code	Factor score septile							CDs Average
	1st Lowest	2nd	3rd	4th Middle	5th	6th	**7th Highest**	
EMSSE	32.4	25.6	23.4	20.8	21.0	19.2	16.8	22.8
EMTRM	2.8	5.2	4.8	6.6	7.2	11.3	16.5	7.8
POWN	54.4	70.9	72.9	73.5	75.0	74.0	75.1	70.8
MFPART	1.2	1.2	1.2	1.2	1.3	1.3	1.4	1.3
EDUAVE	12.8	12.7	12.6	12.7	12.3	12.2	11.7	12.4
HHBLICO	20.3	18.9	15.9	15.7	15.7	16.7	16.8	17.1
EMDSE	9.9	8.6	8.1	8.9	8.5	7.6	6.6	8.3
FERTIL	1.48	1.34	1.38	1.34	1.35	1.32	1.35	1.36
EMCMA	4.0	4.7	4.6	7.1	6.7	6.0	6.6	5.7
WKPT	20.2	22.1	21.5	21.7	20.6	19.6	18.3	20.6
POPL20	31.4	28.4	28.9	29.1	29.3	29.3	29.2	29.4
IMOLD	7.0	10.7	11.4	10.6	10.3	9.8	10.1	10.0
ERN10	27.5	31.6	33.6	29.6	30.0	30.4	33.4	30.9
AVINCO	16660	15848	16074	17013	16086	15834	14499	16005

Source: Author's computation.

Table 4.15 Employment attributes, traditional manufacturing versus government employment attributes: Percent distribution of CDs by septile class of factor scores for each regional type

Regional type (CD units)	1st Lowest	2nd	3rd	4th Middle	5th	6th	7th Highest
Predominantly urban regions (25)	44.0	8.0	8.0	12.0	16.0	12.0	0
Intermediate regions (37)	27.0	21.6	8.1	21.6	8.1	10.8	2.7
Rural metro-adjacent regions (86)	8.1	11.6	15.1	14.0	16.3	15.1	19.8
Rural non metro-adjacent regions (117)	2.6	17.1	17.1	14.5	15.4	16.2	17.1
Rural northern regions (23)	43.5	4.3	13.0	8.7	8.7	8.7	13.0

Source: Author's computation.

4.6. Demographic dynamics attributes

The last factor to emerge in the analysis explains about 5 percent of the total variance. As indicated in Table 4.16, four variables are associated with this factor, all with a positive sign. These are population growth, which shows the stronger loading, percent of households with owner's gross housing costs greater than 30 percent of their income, percent of population below age 20, and percent with (non-agricultural) self-employment. Another 7 variables present a moderate loading and among these are the percent of population above age 65 (negative loading), and the fertility rate (positive loading). Hence, the factor appears to describe primarily the demographic dynamics of the CD.

Table 4.17 reports the average value of the variables related to the factor by septile of factor score. CDs in the lowest septile recorded on average a population loss of 3 percent between 1991 and 1996 and only 28 percent of the population was below 20 years of age (while the share of population above 64 was 14 percent and the fertility rate was 1.36). At the opposite end of the distribution, the top septile CDs present a population growth of 15 percent on average.

Interestingly, there seems to be almost no relationship between the distribution of CDs by size of factor scores and the regional type of CD. As displayed in Table 4.18, the distribution of CDs within each regional type is almost evenly distributed across septiles. Predominantly urban and intermediate CDs tend to present slightly higher shares in the central septiles. While the only pattern that emerges more neatly is the polarization of scores in the two extreme categories for the rural northern CDs, indicating that these tend to show either very positive or highly negative population dynamics.

Nevertheless, Map 4.6 shows that the spatial distribution of the factor scores presents a rather clear pattern. The CDs with low factor scores (negative population dynamics) cover most of the Atlantic Provinces and eastern Québec. Low factor scores are also found in southern Ontario and across a large part of northern Ontario. A second major cluster of CDs with negative population dynamics covers most of the prairies, except the northern CDs. In contrast, in eastern Canada the CDs with high factor scores (denoting population growth and a high share of youth) are clustered in the area north of Ottawa and Montréal, and in a broad zone extending north of Toronto, but not including it. In central Canada, high scores are predominant across the northern CDs and extend to the Territories. Higher scores are also found in most of Alberta, particularly west of the Calgary-Edmonton corridor and in most of British Columbia.

Table 4.16 Demographic dynamics attributes: Factor loadings

Code	Variable	Loading	
POPCH	Population growth between 1991 and 1996	0.830	
HOWN30	Percent of households with owner's gross housing costs equal to or greater than 30 percent of household income	0.428	
POPL20	Percent of population less than 20 years of age	0.424	
SELF	Percent self-employed (non-agricultural)	0.422	
POPO65	Percent of population age 65 years of age and over		-0.352
FERTIL	Fertility rate	0.338	
POWN	Percent of households owning their home		-0.320
EMAGR	Percent agricultural employment		-0.287
ERN10	Percent earning less than $10 per year		-0.274
INTRSF	Social transfer income as a percent of total average income		-0.246
EMDSE	Percent dynamic services employment	0.187	

Source: Author's computation.

Table 4.17 Demographic dynamics attribute: Average values by factor score septiles

Variable code	Factor score septile							CDs Average
	1st Lowest	2nd	3rd	4th Middle	5th	6th	7th Highest	
POPCH	-2.8	-0.8	1.5	3.2	3.7	7.9	14.9	3.9
HOWN30	10.0	12.8	13.6	13.7	14.4	15.3	15.9	13.7
POPL20	28.4	27.3	27.9	29.1	30.0	30.5	32.5	29.4
SELF	7.9	7.3	8.1	9.6	9.5	10.9	11.5	9.3
POPO65	13.8	13.1	12.6	11.9	11.7	10.0	9.9	11.9
FERTIL	1.36	1.25	1.27	1.33	1.40	1.38	1.56	1.36
POWN	74.9	72.0	70.6	72.1	70.8	71.7	63.7	70.8
EMAGR	12.6	4.4	5.5	6.6	7.2	5.4	3.3	6.4
ERN10	35.8	32.0	30.7	30.1	30.9	28.6	27.9	30.9
INTRSF	20.5	22.7	19.5	17.4	17.9	15.1	16.3	18.5
EMDSE	6.9	7.8	8.3	8.6	8.1	9.2	9.2	8.3

Source: Author's computation.

Table 4.18 Demographic dynamics attribute: Percent distribution of CDs by septile class of factor scores for each regional type

Regional type (CD units)	1st Lowest	2nd	3rd	4th Middle	5th	6th	7th Highest
Predominantly urban regions (25)	4.0	12.0	16.0	24.0	16.0	20.0	8.0
Intermediate regions (37)	10.8	24.3	21.6	13.5	18.9	8.1	2.7
Rural metro-adjacent regions (86)	11.6	7.0	16.3	19.8	12.8	17.4	15.1
Rural non-metro-adjacent regions (117)	19.7	17.9	12.0	11.1	13.7	12.0	13.7
Rural northern regions (23)	13.0	8.7	4.3	4.3	13.0	17.4	39.1

Source: Author's computation.

5. Combining dimensions: an example

Each dimension (or factor) that resulted from the factor analysis allows us to classify the CDs according to the size of the factor score for the given CD. (This was used to generate the maps in Maps 4.1 to 4.6) This section presents an example of how the results generated by the factor analysis can be further used to create regional types that combine the values of more than one dimension (or factor). Specifically, two of the dimensions identified are combined to generate a more articulate classification of regional types: these dimensions are the factor for *labour force and economic attributes* and the factor for *remote and agro-rural attributes*.

For this purpose, CDs are classified into three groups representing the bottom third, the middle third and the top third of the scores of the *labour force and economic attributes* factor. Similarly, we classify three groups of CDs using the factor scores of the *remote and agro-rural attributes* factor. The three groups for each factor are then combined to generate a nine-class typology. The nine-class typology is summarised in Table 5.1 and the results are reported in Map 5.1. The three classes reported in red colours indicate CDs that fall into the bottom 33 percent for the labour force and economic dimension, the yellow colours are the middle 33 percent and the green colours are the top 33 percent, that is the CDs with better labour market and economic performances. For each colour, the light shade indicates that the CD falls into the bottom 33 percent for the remote and agro-rural dimension, meaning that it does not present the typical characteristics of "remote and agro-rural". The dark shade indicate that the CDs falls in the top 33 percent of the distribution, meaning that they tend to present strong "*remote and agro-rural*" attributes.

Table 5.1 Typology from cross-classifying two factors

CDs ranked by size of *"remote and agro-rural attribute "*	CDs ranked by size of *"labour force and economic attributes"*		
	Lower one-third of CDs: red colour	Middle one-third of CDs: yellow colour	Higher one-third of CDs: green colour
Lower one-third of CDs: lighter shade			
Middle one-third of CDs: medium shade			
Higher one-third of CDs: darker shade			

The map shows the spatial patterns of these nine regional types defined on the two major dimensions of the factor analysis. The red shades are concentrated in the Atlantic Provinces, in southern Québec, and in the area extending from the north of Saskatchewan and Manitoba to the Arctic regions. But "remote" characteristics (i.e., dark red) are high in particular for this latter area and for a small cluster of CDs along the Québec and New Brunswick border. Southern Québec presents a more scattered pattern, but except for the areas surrounding the major urban cores, most of the CDs fall into the intermediate categories or present low economic attributes. Southern Ontario also presents rather differentiated characteristics. In general, most of the regions are in lighter shades (indicating low remote and agro-rural attributes), except for two counties in the provincial agricultural heartland. Furthermore, the easternmost and westernmost ends of the area generally present high economic attributes. The central CDs, stretching to the north, present either intermediate or low economic attributes. The extreme north shows high economic performance (i.e., green in colour). A belt of CDs with relatively high economic performance and high remote and agro-rural attributes (i.e., dark green) stretches from northern Ontario, through the south fringe of the Prairies extending north through Alberta to reach central and northern B.C. Another band of CDs constitute almost a continuum from the area south of Winnipeg to the CDs north of Regina and Saskatoon to the north west areas of Alberta; this area presents high remote and agro-rural attributes but lower economic performance (i.e., dark yellow). Similar intermediate conditions, but with a prevalence of low economic attributes are also found in a grouping of CDs in southern B.C. In sum, this nine-group classification offers a perspective on the variety of socio-economic conditions across Canada by combining two of the dimension identified in the factor analysis.

If we select the CDs that rank high on the "*remote and agro-rural*" dimension (i.e., dark shades), we can compare the CDs with low "*labour force and economic*" attributes (i.e., dark red) versus CDs with high "*labour force and economic*" attributes (i.e., dark green). Dark green CDs (i.e., high "*remote and agro-rural*" attributes with high "*labour force and economic*" attributes) are concentrated in a band running from northern Ontario through southern Manitoba and southern Saskatchewan to most of Alberta (on either side of the Calgary – Edmonton corridor) and into northern British Columbia. Dark red CDs are also high on the "*remote and agro-rural*" dimension but with relatively low economic performance. In western Canada, this includes northern Manitoba, northern Saskatchewan and the eastern Arctic. In eastern Canada, this includes scattered CDs north of the St. Lawrence River (plus a

few scattered CDs south of the St. Lawrence River), the Gaspé region of Québec, northern New Brunswick, most of Nova Scotia outside of Halifax and all of Newfoundland outside of St. John's.

6. Summary and conclusions

The concept of rural diversity is today well established. The challenge is to gain an understanding of the nature of this diversity and its spatial distribution. The analysis presented in this report helps to achieve this objective. This research was undertaken to assess the degree of spatial diversity existing across Canada on a range of demographic, social and economic indicators. The study adopted a broad territorial approach, which allows assessing the relative condition of rural areas in the national context without imposing an *a priori* definition of rural.

Twenty-seven variables, which reflect demographic, social and economic characteristics of each CD, are used in the study. All variables are from the 1996 Census of Population. A factor analysis was applied to the data. The 27 variables used in the analysis were reduced to 6 factors and these factors capture about 78 percent of the variance in the data set. The factors summarise how the selected indicators tend to vary across space and in which way each of them is associated with the other social and economic indicators considered. The results provide a simplified but meaningful picture of a complex reality. Each factor reveals the patterns of associations among the variables used and provides a perspective on the relative position of each CD across groupings of variables. The factors provide a profile of the CDs on a number of attributes. Although the distinction is not always clear-cut, some of the factors are more "diagnostic" in that the variables loading on them are indicators of socio-economic conditions deemed to be of desirable or undesirable attributes, such as unemployment, and percent of individuals in low-income families. Others tend to be more "descriptive", in that the variables that load on them are not necessarily indicative of unfavourable social or economic conditions.

The nature and spatial distribution of the factors can be summarised as follows:

- The first factor, named *labour force and economic attributes*, captures a range of economic and social attributes, including unemployment, income level and educational attainment, that describe the overall economic performance and strength of the CD. This factor appears to describe in particular the regional variation in economic performance (north/east versus south/west). Predominantly urban CDs tend to present higher scores on this factor, while rural CDs show a greater diversity of conditions.
- The factor named *remote and agro-rural attributes* describes a set of attributes that prevail in remote areas and parts of rural Canada with a higher incidence of agricultural employment. The factor combines demographic and housing characteristics that are common to these areas. Negative scores are overwhelmingly a feature of the remote north (northeast in particular) and the Prairies, while urban CDs present opposite characteristics.
- Two factors are mainly associated with the employment structure. The first, named *complex manufacturing versus primary production attributes*, identifies in particular the resource-based communities. The second, named *traditional manufacturing versus*

government employment attributes, shows contrasting trends between traditional manufacturing employment and greater participation rate for males, on the one hand, and non-market services employment, on the other hand.

- Two factors describe in particular the demographic structure and dynamics of the CDs. The first named *demographic and labour force attributes* relates in particular to aging population trends, but also describes labour force characteristics (non-agricultural self-employment and part-time work). This factor shows similar conditions for predominantly urban CDs and rural northern CDs. The second factor, labelled *demographic dynamics attributes*, identifies areas with a young and growing population. The distribution by regional type show the diversity of conditions recorded by both urban and by rural CDs.

More generally the following conclusions can be drawn form this study:

- The spatial distribution of the factors reveals regional differences as well as differences between CD regional types. This demonstrates the utility of a broad territorial approach, which allows a better understanding of both regional patterns as well as hierarchical spatial structures (i.e., the groupings of CDs with similar spatial patterns).
- The current definition of regional types of CDs seems to capture relatively well the variation of some of the dimensions identified. This regional classification has the major advantage of simplicity and clear-cut definition. Nonetheless, for specific policy purposes, it would appear appropriate to use more refined regional types, which could be more sensitive to the spatial variation of conditions that prevail in particular among rural areas.
- The results indicate the multi-dimensional nature of performance and the variety of possible combinations of demographic, social and economic characteristics (e.g. resource-based regions with poor economic performance versus resource-based regions with high economic performance; various combinations of demographic, rural attributes and economic performance; and so on). This illustrates the usefulness of the factor analysis approach which allows a better understanding of this combination of conditions.
- With regard to specific indicators, the results appear to trace the two prevailing dimensions of income variation, one across macro-regions and the second between rural and urban types.

References

Alasia, Alessandro. 2003. *Sub-provincial Income Disparity in Canada: Evidence from 1992 to 1999.* Agriculture and Rural Working Paper Series No. 63. Cat. No. 21-601-MIE. Ottawa: Statistics Canada. Available at: www.statcan.ca/english/research/21-601-MIE/21-601-MIE2003063.pdf

Alasia, Alessandro. 1999. Leading and Lagging Areas in Ontario: Huron County in the Provincial Context, in Tony Fuller and Paul Nichol (editors), Dynamics *of the New Rural Economy: An Exploration of Community-sponsored Research form Huron County*, Guelph: University of Guelph.

Berry, Brian J. L. 1971. Introduction: The Logic and Limitation of Comparative Factorial Ecology, *Economic Geography* 47(2): 209-19.

Blunden, John R., W.T.R. Pryce, and Peter Dreyer. 1998. The Classification of Rural Areas in the European Context: An Exploration of a Typology Using Neural Network Applications. *Regional Studies* 32(2):149-160.

Bollman, Ray. 1994. A Preliminary Typology of Rural Canada, in J. M. Bryden (editor) *Towards Sustainable Communities*. Guelph: University School of Rural Planning and Development.

Cannata, Giovanni. 1989. *I sistemi agricoli territoriali italiani*. Milano: Franco Angeli.

Chapuis, Robert, and Thierry Brossard. 1989. The Demographic Evolution of Rural France (1968-1982). *Journal of Rural Studies* 5(4):357-365.

Copus, Andrew, and Bob Crabtree. 1992. Mapping Economic Fragility: an Assessment of the Objective 5b Boundaries in Scotland. *Journal of Rural Studies* 8(3):309-322.

du Plessis, Valerie, Roland Beshiri, Ray D. Bollman and Heather Clemenson. (2002) *Definitions of Rural*. Agriculture and Rural Working Paper No. 61. Cat. No. 21-601-MIE. Ottawa: Statistics Canada. Available at: www.statcan.ca/cgi-bin/downpub/listpub.cgi?catno=21-601-MIE).

Fellegi, I.P. 1996. *Understanding Rural Canada*. Ottawa: Statistics Canada, Cat. No. 21F0016XIE. Available at: www.statcan.ca/english/freepub/21F0016XIE/free.htm

Hawkins, Liz, 1995, *Mapping the Diversity of Rural Economies: A Preliminary Typology of Rural Canada*. (Ottawa: Statistics Canada, Agriculture Division, Agriculture and Rural Working Paper No. 29).

Hawkins, Liz and Ray D. Bollman. 1994. "Revisiting Rural Canada – It's Not All the Same," *Canadian Agriculture at a Glance* (Ottawa: Statistics Canada, Cat. No. 96-301), pp. 78-80.

Jonassen, C. T., and S. H. Peres. 1960. *Interrelationship of Dimensions of Community Systems: A Factor Analysis of Eighty-two Variables*. Columbus: Ohio State University Press.

Keddie, Philip D. 1997. *The Demographic, Social and Economic Diversity of Rural and Small* Town Southern Ontario, A Descriptive Atlas of a Six County Area. Occasional Paper No. 23. University of Guelph, Department of Geography. Guelph.

Keddie, Philip, and Alessandro Alasia. 1999. *The Socio-Economic Diversity of Rural and Small Town Southern Ontario: A Comparative Profile*. Sustainable Rural Communities Research Program, University of Guelph - Ontario Ministry of Agriculture Food and Rural Affairs. Guelph.

King, Leslie J. 1966. Cross-sectional Analysis of Canadian Urban Dimensions: 1951 and 1961. *Canadian Geographer* no. 4: 205-24.

Lipshitz, Gabriel, and Adi Raveh. 1998. Socio-economic Differences among Localities: A New Method of Multivariate Analysis. *Regional Studies* 32(8):747-757.

Montresor, Elisa, and Maio Mazzocchi. 2001. *The Agricultural and Rural Development: An Analytical Approach*. Paper presented to the 73rd Seminar of the European Association of Agricultural Economics, Ancona, June.

OECD. 2001. *Territorial Outlook 2001*. Paris: OECD.

Pezzini, Mario. 2001. Rural Policy Lesson from OECD Countries. *International Regional Science Review* 24(1): 134-145.

Quadrado, L., W. Heijman, and H. Folmer. 2001. "Multidimensional analysis of regional inequality: The case of Hungary". *Social Indicator Research* 56: 21-42.

Rees, Philip H. 1971. Factorial Ecology: An Extended Definition, Survey, and Critique of the Field, *Economic Geography* 47(2, Supplement): 221-33.

Reimer, Bill. 1997. *Analysis of Leading and Lagging CSDs: Notes for a Report to the NER Project*. Draft paper, Montreal: Concordia University.

Rogers, Andrei. 1971. *Matrix methods in urban and regional analysis*. San Francisco: Holden-Day.

SEGESA – Direction de l'Espace Rural et de la Foret. 1992. *Nouvelle Approche de la Diversité des Agricultures Régionales*. Paris: Ministère de l'Agriculture et de la Foret.

Shearmur, Richard, and Mario Polèse. 2001. *Economic Development in Canadian Peripheral Regions, 1971 to 1999. A Statistical Overview*. Research report. Montreal: INRS.

Shields, Martin, and Steven C. Deller.1996. *Clustering Wisconsin Counties for Analytical Comparisons*. CCED Staff Paper 96.7. University of Wisconsin.

Simmons, James W., and Brian Speck. 1986. *Spatial Patterns of Social Change: The Return of the Great Canadian Factor Analysis*. Research Paper No. 160. Centre for Urban and Community Studies. University of Toronto.

Stimson, Robert, Scott Baum, Patrick Mullins, and Kevin O'Connor. 2001. A typology of community opportunities and vulnerability in metropolitan Australia. *Papers in Regional Science* 80, p. 45-66.

Statistics Canada. 1997. *1996 Census Dictionary*. Ottawa: Industry Canada, Catalogue no. 92-351-XPE.

Thompson, John H., Sidney C. Sufrin, Peter R. Gould, and Marion A. Buck. 1964. Toward a Geography of Economic Health: The Case of New York State. In John Friedmann, and William Alonso (editors). *Regional Development and Planning*. Cambridge, Massachusetts: The MIT Press.

Appendix A: Definition of variables

The operational definition of all the variables used in this study is given below. The data source is the 1996 Census of Population. The following list explains how the variables used in this study were computed. In some cases the definition of the census is presented. For a detailed definition of the original variables refer to Statistics Canada (1997). For ease of use the variables are grouped in four categories.

Demographic indicators

Percent population change 1991 to1996. This variable is taken from the 1996 Census of Population database without further computation.

Percent of population less than age 20. This and the following variable are computed by aggregating the corresponding age cohorts available in the census.

Percent of population age 65 years of age and over. As previous variable.

Senior in-migration rate: Percent of persons age 55 to 74 living in different CSD 5 years ago. This variable is computed as the number of person 55 to 74 years of age living in a different CSD (census sub-division) five years ago divided by total number of individuals 55 to 74 years of age now living in the CSD. Note that movement from one CSD to another CSD within the same CD (census division) will contribute to the senior in-migration rate at the CD level, which is the level of analysis in this study.

Fertility rate. This variable is not available in the 1996 census database. A proxy variable was used, which was computed as the number persons below 19 years of age divided by the number of women 25 to 54 years of age.

Social indicators

Average years of schooling for population 25 to 54 years of age. This variable is taken from the Census of Population 1996 database without further computation.

Percent of persons in low-income economic families. This variable is taken from the census. The incidence of low-income is the proportion of individuals in economic families or unattached individuals below the low-income cut-off. The threshold values used by Statistics Canada for the determination of the incidence of low-income economic families and low-income unattached individuals vary by four urban size categories and for rural (five categories in all) and for families by family size (seven categories). For example, the low-income cut-off for a family of four in 1996 ranges from $31,753 for an urban place of 500,000 or more (e.g. Toronto) to $21,944 for rural areas. This represents an attempt to capture “cost of living differences” by type location in making a determination of the incidence of low-income. For more detail on the low-income cut-off values, see Statistics Canada (1997).

Social transfer income as a percent of total income. This variable is computed as average social transfer income for persons age 15 and over divided by total average income for persons age 15 and over.

Total unemployment rate. The number of individuals, 15 years of age and over, unemployed in the week prior to the census divided by the number in the labour force (i.e., the number employed plus the number unemployed in the week prior to the census).

Housing characteristics

Percent of households with gross rent equal to or greater than 30 percent of household income. This variable is computed as the number of households with gross rent equal to or greater than 30 percent of household income divided by the total number of households living in rented accommodations.

Percent of households with owner’s gross housing costs equal to or greater than 30 percent of household income. This variable is computed as the number of private households with owner’s gross housing costs equal to or greater than 30 percent of household income divided by the total number of households living in an owned dwelling.

Percent of household owning their home. This variable is computed as the number of households living in an owned dwelling divided by the total number of households.

Economic and labour market indicators

Percent agricultural employment. This is calculated as the experienced employment in agriculture and services related to agriculture divided by the total experienced labour force 15 years and over.

Percent other primary employment. This is calculated as the experienced employment in other primary sectors (fishing and trapping industries, logging and forestry industries, mining, quarrying & oil well industries) divided by the total experienced labour force 15 years and over.

Percent traditional manufacturing employment. This is calculated as the experienced employment in traditional manufacturing industries divided by the total experienced labour force 15 years and over. Traditional manufacturing sectors include employment in food processing, beverages, tobacco, rubber, plastic, leather, primary textile, clothing, wood, furniture and fixtures, and paper manufacturing sectors.

Percent complex manufacturing employment. This is calculated as the experienced employment in complex manufacturing industries divided by the total experienced labour force 15 years and over. Complex manufacturing sectors include employment in printing, primary metals, fabricated metal, machinery, transportation equipment, electrical and electronic, non-metallic metal, refined petroleum and coal, chemical and "other" manufacturing sectors.

Percent dynamic services employment. This is calculated as the experienced employment in dynamic service industries divided by the total experienced labour force 15 years and over. Dynamic services employment includes employment in transportation and storage industries, communication and other utility industries, wholesale trade industries, finance and insurance industries, real estate operator and insurance agent industries and business service industries.

Percent non-market services employment. This is calculated as the experienced employment in non-market services divided by the total experienced labour force 15 years and over. The non-market services employment includes employment in government service industries, educational service industries and health and social service industries.

Participation rate. This variable is calculated as total labour force (employed and unemployed persons age 15 and over) divided by the population 15 years of age and over.

Male participation rate over female participation rate. The variable is computed as the ratio between the two participation rates (males and females age 15 and over).

Percent of families (married and common-law couples) with two or more members in the labour force. This variable is computed as the number of families with two or more members in the labour force divided by the total number of families in private households.

Percent with non-agricultural self-employment activity. For individuals whose main job is not in the agricultural industry, we identify self-employment activity as the class of worker being "self-employed" (including both working in an unincorporated enterprise and in an incorporated enterprise) *or* reporting some non-farm self-employment income in the year previous to the census. We calculated the percent with some non-agricultural self-employment activity as the number of individuals, 25 to 54 years of age, with some non-agricultural self-employment activity divided by the total number of individuals 25 to 54 years of age.

Percent with part-time employment. The variable is computed as persons who worked part time (less than 30 hours per week) divided by total employment.

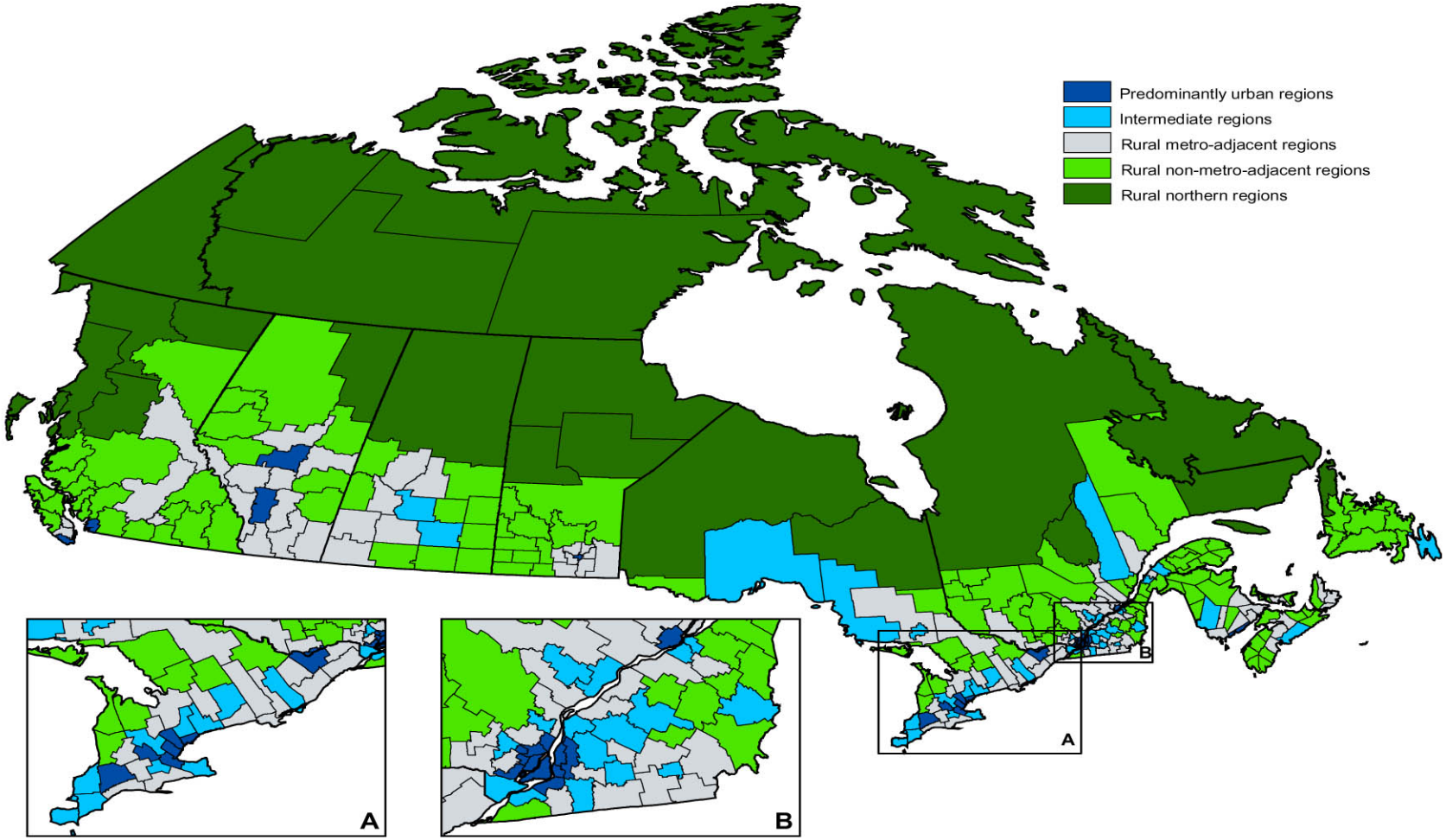
Average income per person. Average income from all sources, per person reporting some income.

Percent earning less than $10 per hour. For persons with some earned income (i.e. wages and salaries and / or self-employment income is not equal to zero) and with some hours worked last week and with some weeks worked last year, average hourly earnings is calculated as earned income last year divided by estimated hours worked last year (calculated as hours worked last week multiplied by weeks worked last year). We then calculate the percent of individuals with earnings less than $10 per hour.

Off-farm earnings of census-farm operator families as a percent of total family income. This variable is computed as off-farm earnings (i.e., wage and salary income plus non-farm self-employment income) of economic families with a census-farm operator divided by total income of economic families with a census-farm operator.

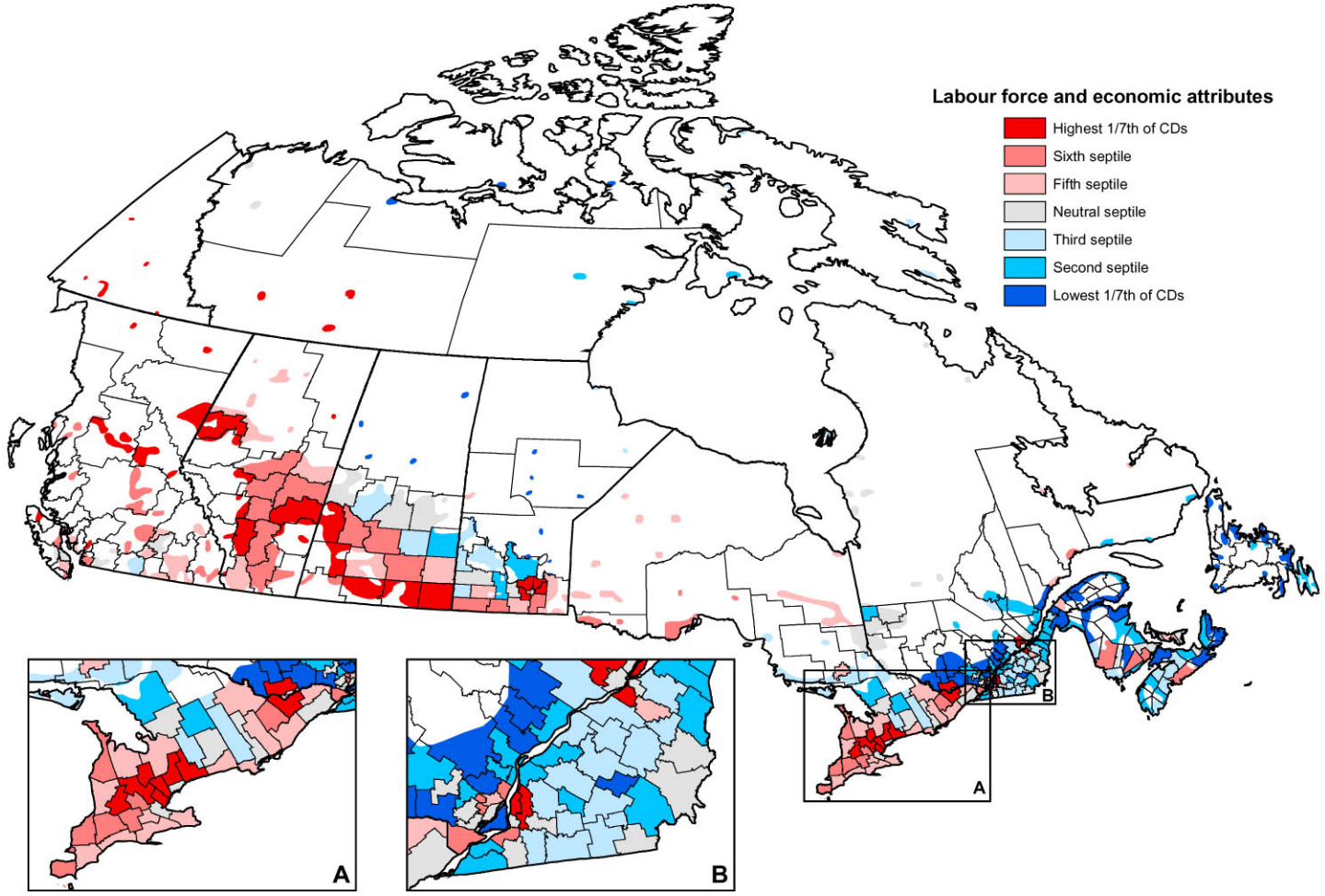
Average income growth between 1991 and 1996. This variable is computed as percent change *average income per* person, in nominal terms (i.e. with no adjustment for inflation).

Figure 3.1. Census division regional types



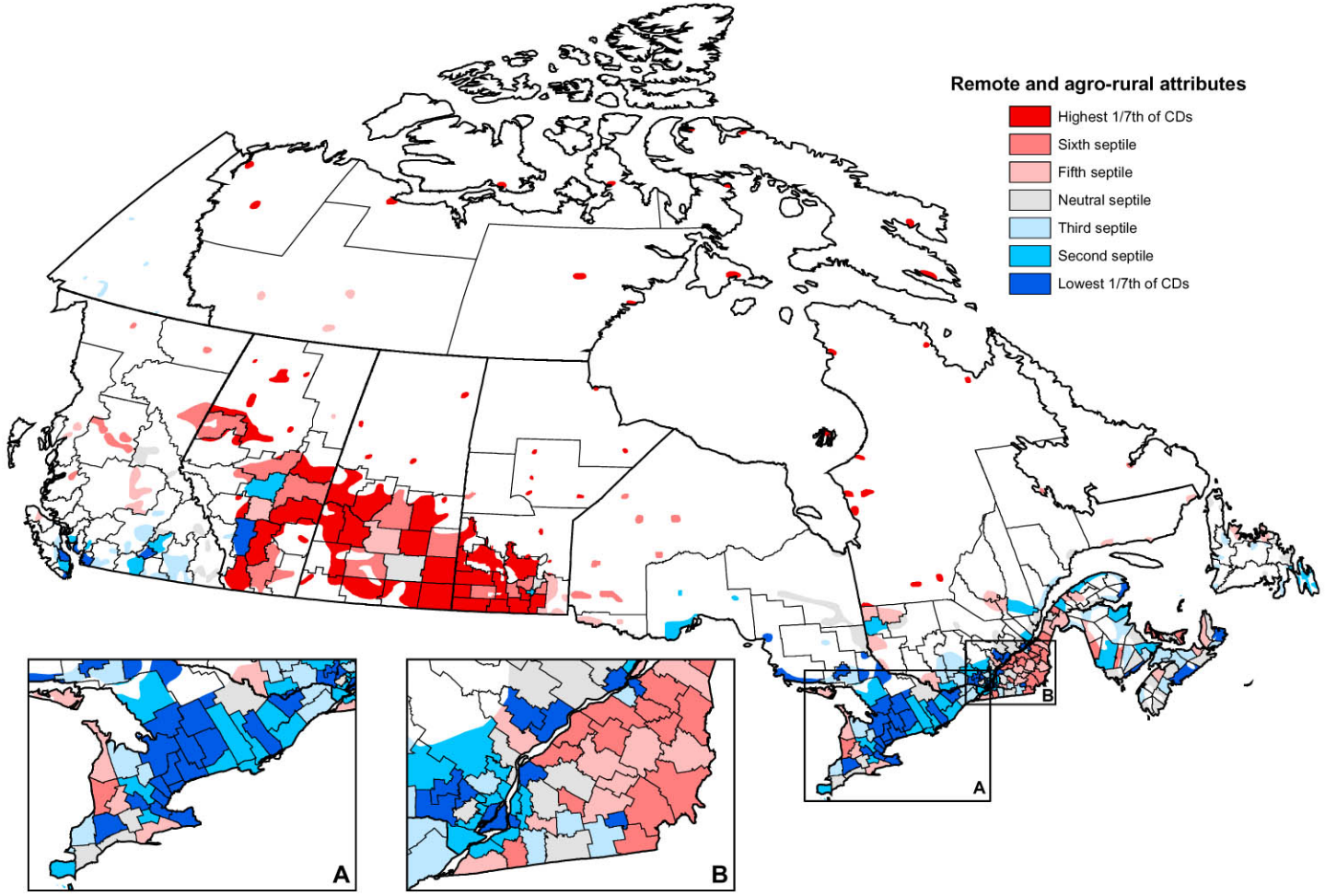
Source: Statistics Canada, Census of Population, 1996.
Map produced by Spatial Analysis and Geomatics Applications (SAGA), Agriculture Division, Statistics Canada, 2003.

Map 4.1. Labour force and economic attributes



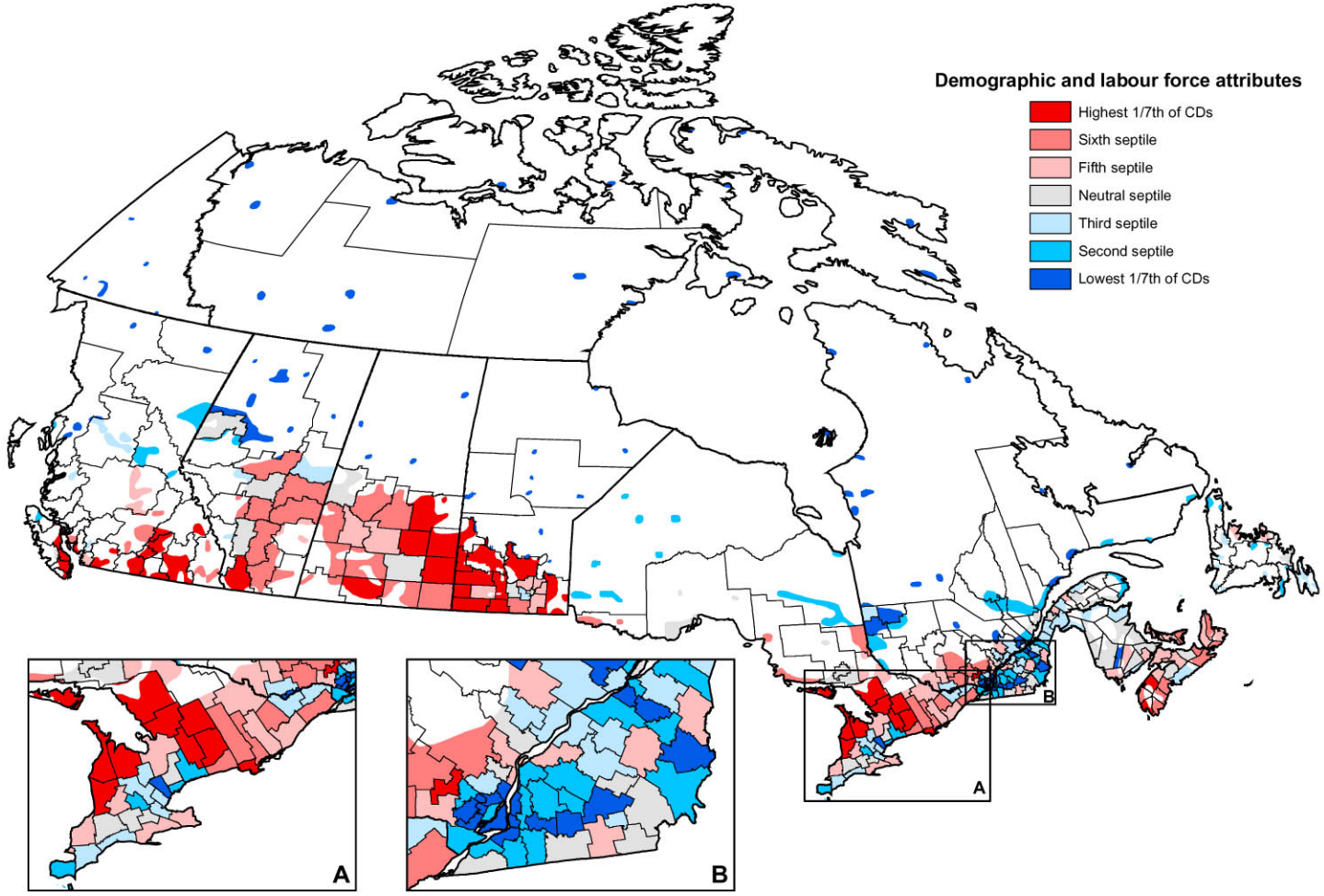
Source: Author's computation based on Census of Population, 1996.
Map produced by Spatial Analysis and Geomatics Applications (SAGA), Agriculture Division, Statistics Canada, 2003.

Map 4.2. Remote and agro-rural attributes



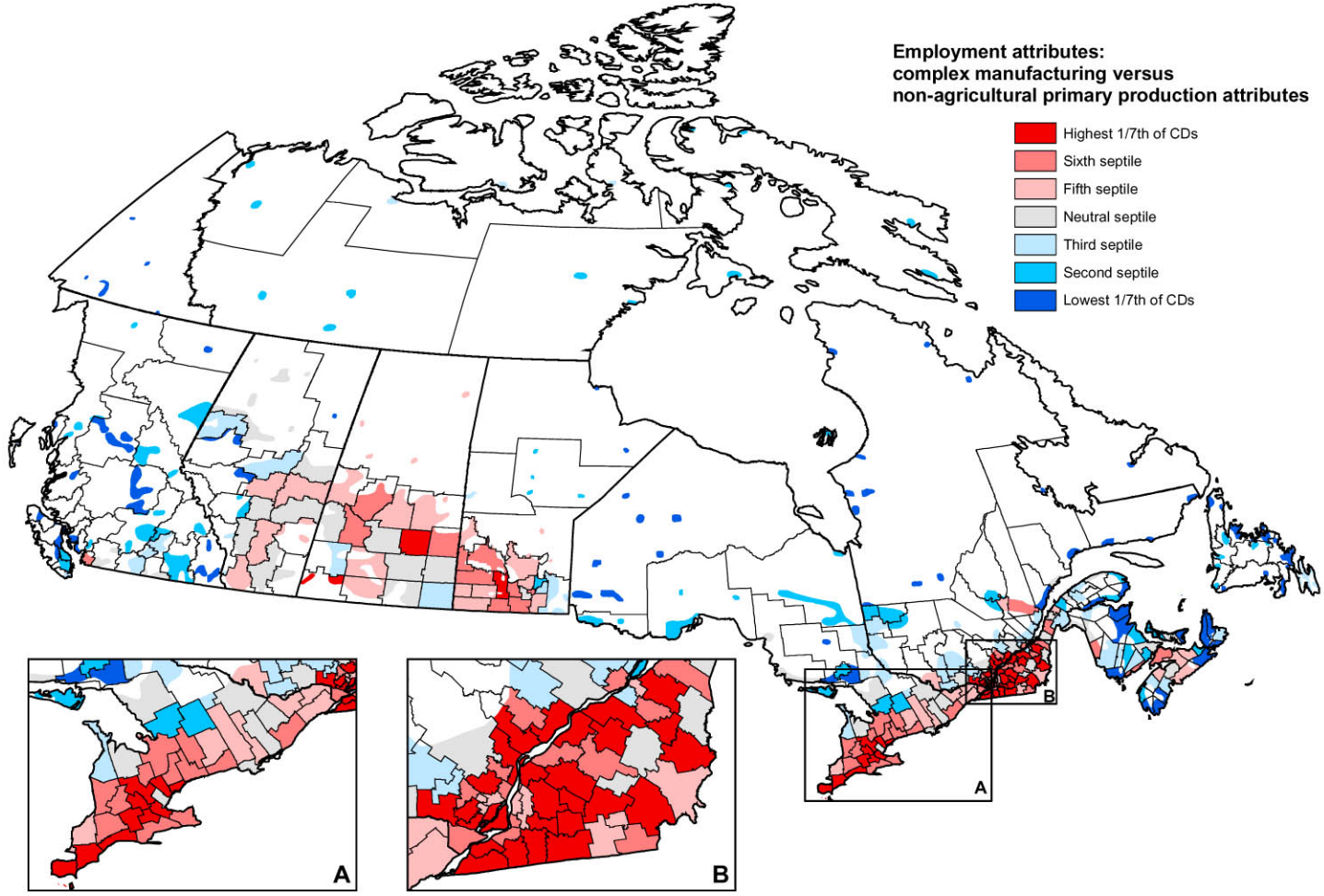
Source: Author's computation based on Census of Population, 1996.
Map produced by Spatial Analysis and Geomatics Applications (SAGA), Agriculture Division, Statistics Canada, 2003.

Map 4.3. Demographic and labour force attributes



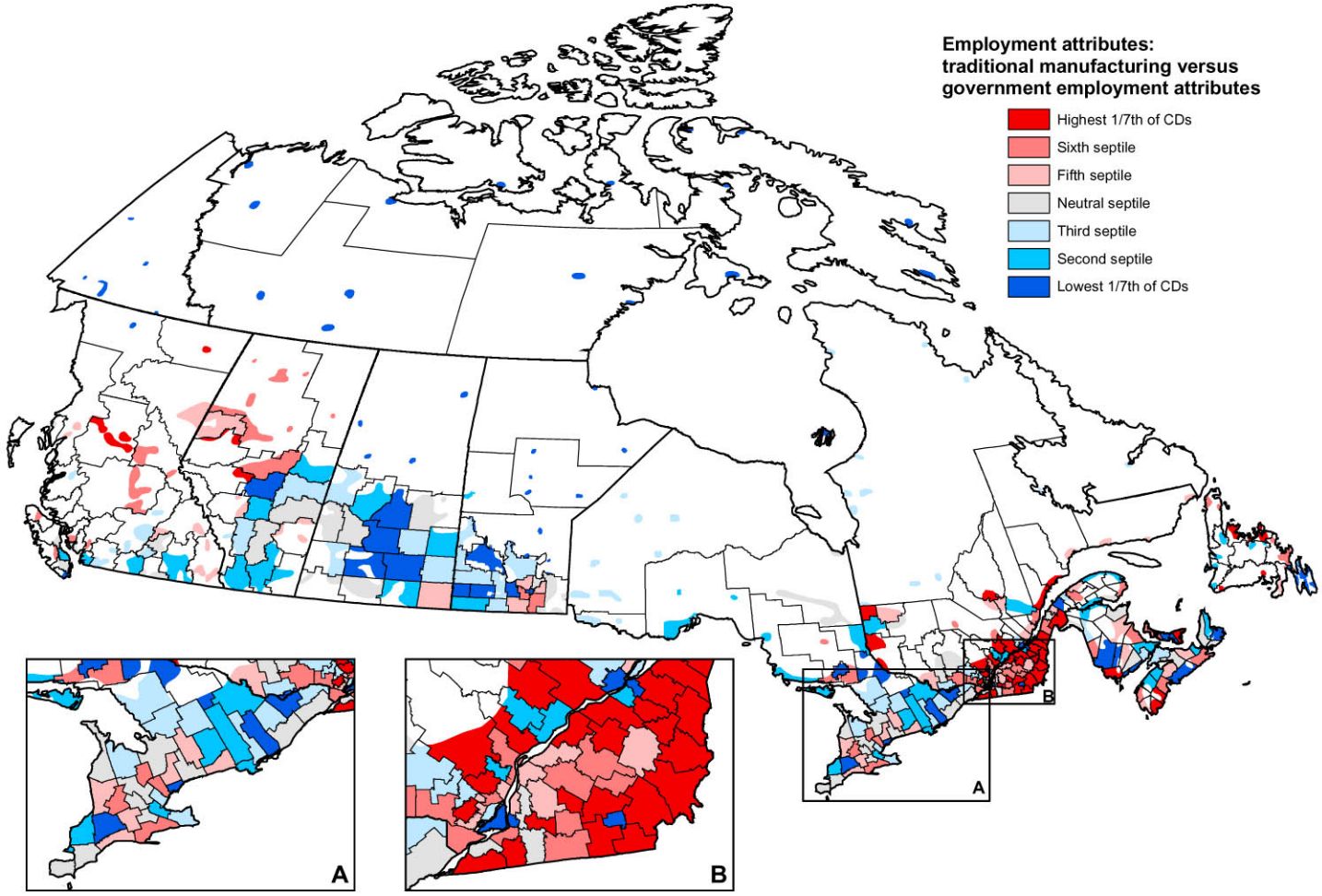
Source: Author's computation based on Census of Population, 1996.
Map produced by Spatial Analysis and Geomatics Applications (SAGA), Agriculture Division, Statistics Canada, 2003.

Map 4.4. Employment attributes: complex manufacturing versus non-agricultural primary production attributes



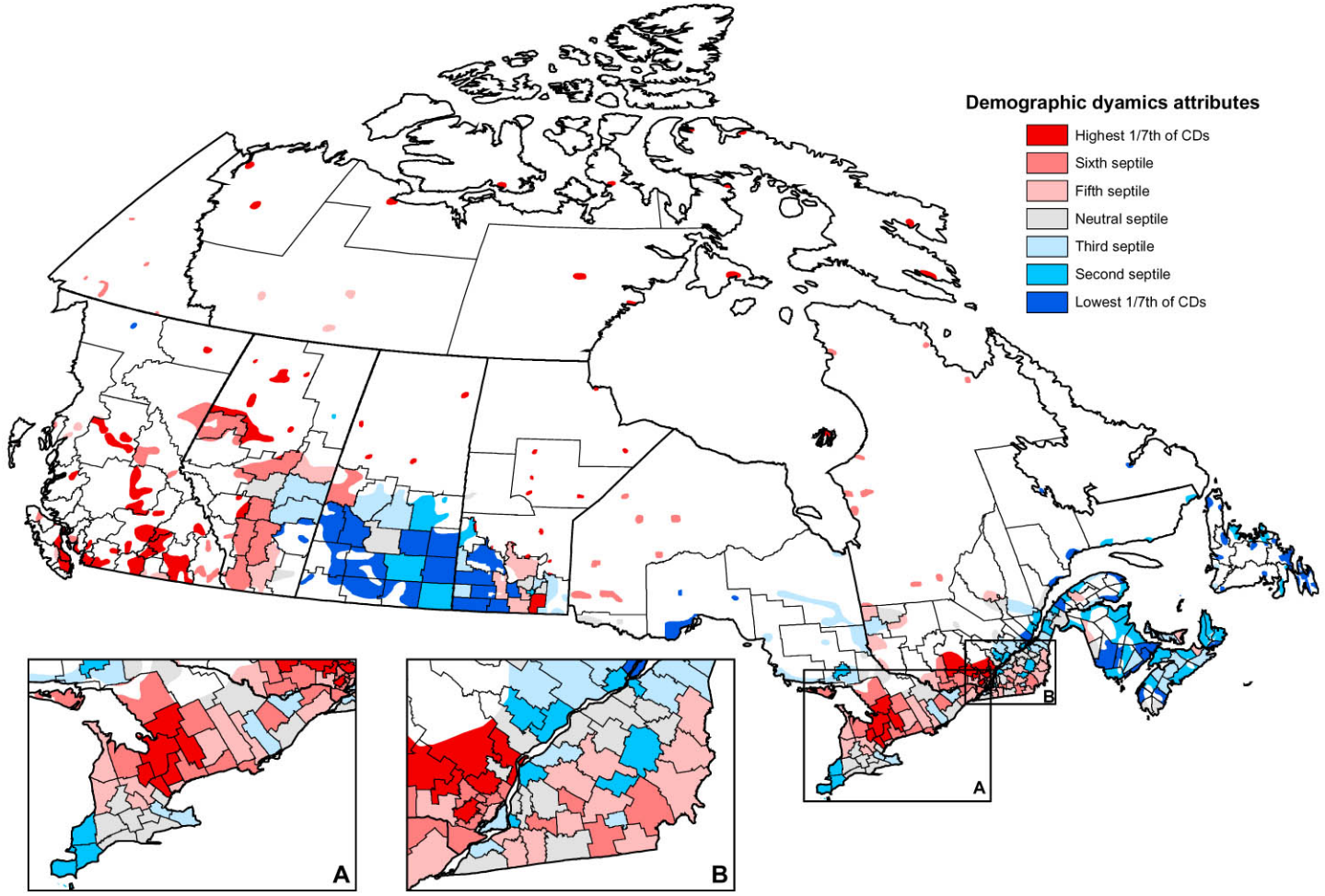
Source: Author's computation based on Census of Population, 1996.
Map produced by Spatial Analysis and Geomatics Applications (SAGA), Agriculture Division, Statistics Canada, 2003.

Map 4.5. Employment attributes: traditional manufacturing versus government employment attributes



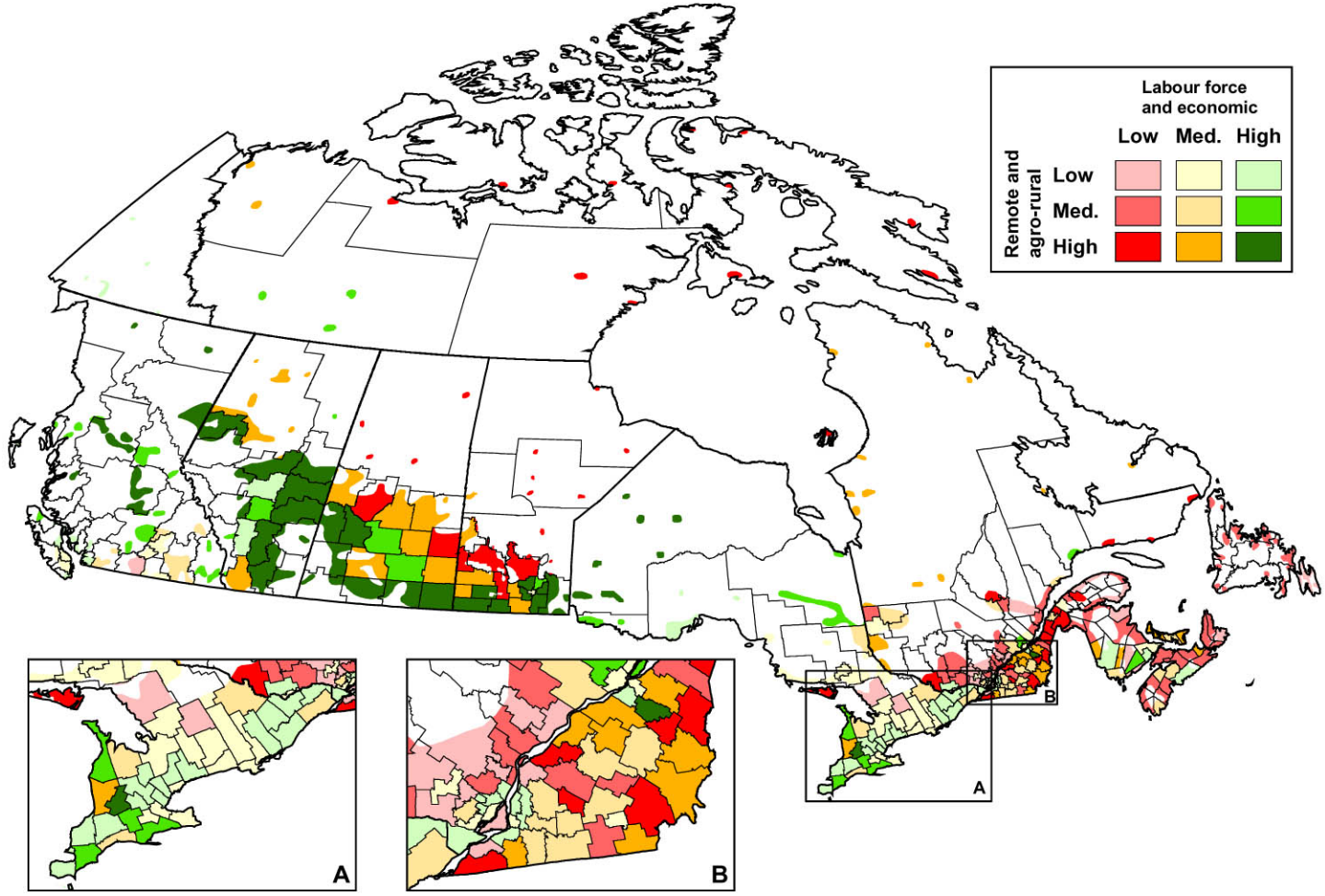
Source: Author's computation based on Census of Population, 1996.
Map produced by Spatial Analysis and Geomatics Applications (SAGA), Agriculture Division, Statistics Canada, 2003.

Map 4.6. Demographic dyamics attributes



Source: Author's computation based on Census of Population, 1996.
Map produced by Spatial Analysis and Geomatics Applications (SAGA), Agriculture Division, Statistics Canada, 2003.

Map 5.1. Overlap of two dimensions: "labour force and economic" and "remote and agro-rural" attributes



Source: Author's computation based on Census of Population, 1996.
Map produced by Spatial Analysis and Geomatics Applications (SAGA), Agriculture Division, Statistics Canada, 2003.

Agriculture and Rural Working Paper Series

(* The *Agriculture and Rural Working Paper Series* is now available on Statistics Canada's Web Site (www.statcan.ca). From the *Our products and services* page, under *Browse our Internet publications (PDF or HTML)*, choose *Free*.)

No.1	(21-601-MPE1980001)	**A Description of Theil's RMPSE Method in Agricultural Statistical Forecasts (1980)**, Stuart Pursey
No.3	(21-601-MPE1981003)	**A Review of the Livestock Estimating Project with Recommendations for the Future (1981)**, Bernard Rosien and Elizabeth Leckie
No.4	(21-601-MPE1984004)	**An Overview of the Canadian Oilseed Industry (1984)**, Glenn Lennox
No.5	(21-601-MPE1984005)	**Preliminary Analysis of the Contribution of Direct Government Payments to Realized Net Farm Income (1984)**, Lambert Gauthier
No.6	(21-601-MPE1984006)	**Characteristics of Farm Entrants and their Enterprises in Southern Ontario for the Years 1966 to 1976 (1984)**, Jean B. Down
No.7	(21-601-MPE1984007)	**A Summary of Commodity Programs in the United States (1984)**, Allister Hickson
No.8	(21-601-MPE1984008)	**Prairie Summerfallow Intensity: An Analysis of 1981 Census Data (1984)**, Les Macartney
No.9	(21-601-MPE1985009)	**The Changing Profile of the Canadian Pig Sector (1985)**, Mike Shumsky
No.10	(21-601-MPE1986010)	**Revisions to the Treatment of Imputed House Rents in the Canadian Farm Accounts, 1926-1979 (1986)**, Mike Trant
No.11	(21-601-MPE1992011)	**The Ratio Estimator: an Intuitive Explanation and Its Use in Estimating Agriculture Variables (1992)**, François maranda and Stuart Pursey
No.12	(21-601-MPE1991012)	**The Impact of Geographic Distortion Due to the Headquarters Rule (1991)**, Rick Burroughs
No.13	(21-601-MPE1991013)	**The Quality of Agriculture Data - Strengths and Weaknesses (1991)**, Stuart Pursey
No.14	(21-601-MPE1992014)	**Alternative Frameworks for Rural Data (1992)**, A.M. Fuller, Derek Cook and Dr. John Fitzsimons
No.15	(21-601-MPE1993015)	**Trends and Characteristics of Rural and Small Town Canada (1993)**, Brian Bigs, Ray Bollman and Michael McNames
No.16	(21-601-MPE1992016)	**The Microdynamics and Farm Family Economics of Structural Change in Agriculture (1992)**, Phil Ehrensaft and Ray Bollman
No.17	(21-601-MPE1993017)	**Grains and Oilseeds Consumption by Livestock and Poultry, Canada and Provinces 1992**, Livestock and Animal Products Section
No.18	(21-601-MPE1994018)	**Trends and Patterns of Agricultural Structural Change: Canada / US Comparison**, Ray Bollman, Leslie A. Whitener and Fu Lai Tung
No.19	(21-601-MPE1994019)	**Farm Family Total Income by Farm Type, Region and Size for 1990 (1994)**, Saiyed Rizvi, David Culver, Lina Di Piétro and Kim O'Connor
No.20	(21-601-MPE1991020)	**Adjustment in Canadian Agriculture (1994)**, George McLaughlin
No.21	(21-601-MPE1993021)	**Microdynamics of Farm Size Growth and Decline: A Canada-United States Comparison**, Fred Gale and Stuart Pursey
No.22	(21-601-MPE1992022)	**The Structures of Agricultural Household Earnings in North America: Positioning for Trade Liberalization**, Leonard Apedaile, Charles Barnard, Ray Bollman and Blaine Calkins
No.23	(21-601-MPE1992023)	**Potatoes: A Comparison of Canada/USA Structure**, Glenn Zepp, Charles Plummer and Barbara McLaughlin
No.24	(21-601-MPE1994024)	**Farm Structure Data: A US-Canadian Comparative Review**, Victor J. Oliveira, Leslie A. Whitener and Ray Bollman
No.25	(21-601-MPE1994025)	**Grain Marketing Statistics Statistical Methods Working Paper Version 2**, Karen Gray

Agriculture and Rural Working Paper Series (continued)

(* The *Agriculture and Rural Working Paper Series* is now available on Statistics Canada's Web Site (www.statcan.ca). From the *Our products and services* page, under *Browse our Internet publications (PDF or HTML)*, choose *Free*.)

No.26	(21-601-MPE1994026)	**Farm Business Performance: Estimates from the Whole Farm Database**, W. Steven Danford
No.27	(21-601-MPE1994027)	**An Attempt to Measure Rural Tourism Employment**, Brian Biggs
No.28*	(21-601-MIE1995028)	**Delineation of the Canadian Agricultural Ecumene for 1991**, Timothy J. Werschler
No.29	(21-601-MPE1995029)	**Mapping the Diversity of Rural Economies: A preliminary Typology of Rural Canada**, Liz Hawkins
No.30*	(21-601-MIE1996030)	**Structure and Trends of Rural Employment: Canada in the Context of OECD Countries**, Ron Cunningham and Ray D. Bollman
No.31*	(21-601-MIE1996031)	**A New Approach to Non-CMA/CA Areas**, Linda Howatson-Leo and Louise Earl
No.32	(21-601-MPE1996032)	**Employment in Agriculture and Closely Related Industries in Rural Areas: Structure and Change 1981-1991**, Sylvain Cloutier
No.33*	(21-601-MIE1998033)	**Hobby Farming - For Pleasure or Profit?**, Stephen Boyd
No.34*	(21-601-MIE1998034)	**Utilization of Document Imaging Technology by the 1996 Canadian Census of Agriculture**, Mel Jones and Ivan Green
No.35*	(21-601-MIE1998035)	**Employment Patterns in the Non-Metro Workforce**, Robert Mendelson
No.36*	(21-601-MIE1998036)	**Rural and Small Town Population is Growing in the 1990s**, Robert Mendelson and Ray D. Bollman
No.37*	(21-601-MIE1998037)	**The Composition of Business Establishments in Smaller and Larger Communities in Canada**, Robert Mendelson
No.38*	(21-601-MIE1998038)	**Off-farm Work by Census-farm Operators: An Overview of Structure and Mobility Patterns**, Michael Swidinsky, Wayne Howard and Alfons Weersink
No.39*	(21-601-MIE1999039)	**Human Capital and Rural Development: What Are the Linkages?**, Ray D. Bollman
No.40*	(21-601-MIE1999040)	**Computer Use and Internet Use by Members of Rural Households**, Margaret Thompson-James
No.41*	(21-601-MIE1999041)	**RRSP Contributions by Canadian Farm Producers in 1994**, Marco Morin
No.42*	(21-601-MIE1999042)	**Integration of Administrative Data with Survey and Census Data**, Michael Trant and Patricia Whitridge
No.43*	(21-601-MIE2001043)	**The Dynamics of Income and Employment in Rural Canada: The Risk of Poverty and Exclusion**, Esperanza Vera-Toscano, Euan Phimister and Alfons Weersink
No.44*	(21-601-MIE2001044)	**Rural Youth Migration Between 1971 and 1996**, Juno Tremblay
No.45*	(21-601-MIE2001045)	**Measuring Economic Well-Being of Rural Canadians Using Income Indicators**, Carlo Rupnik, Margaret Thompson-James and Ray D. Bollman
No.46*	(21-601-MIE2001046)	**The Geographical Patterns of Socio-Economic Well-Being of First Nations Communities in Canada**, Robin P. Armstrong
No.47*	(21-601-MIE2001047)	**Distribution and Concentration of Canadian Livestock**, Martin S. Beaulieu
No.48*	(21-601-MIE2001048)	**Intensive Livestock Farming: Does Farm Size Matter?**, Martin S. Beaulieu
No.49*	(21-601-MIE2001049)	**Agriculture Statistics for Rural Development**, Ray D. Bollman
No.50*	(21-601-MIE2001050)	**Rural and Small Town Employment: Structure by Industry**, Roland Beshiri and Ray D. Bollman

Agriculture and Rural Working Paper Series **(continued)**

(* The *Agriculture and Rural Working Paper Series* is now available on Statistics Canada's Web Site (www.statcan.ca). From the *Our products and services* page, under *Browse our Internet publications (PDF or HTML)*, choose *Free*.)

No.51*	(21-601-MIE2001051)	**Working Time: How do Farmers Juggle with it and How has it Impacted Their Family Total Income**, Sylvain Cloutier
No.52*	(21-601-MIE2001052)	**Growers of Genetically Modified Grain Corn and Soybeans in Quebec and Ontario: A Profile**, Bernard Hategekimana
No.53*	(21-601-MIE2002053)	**Integration of Canadian and U.S. Cattle Markets**, Rita Athwal
No.54*	(21-601-MIE2002054)	**Genetically Modified Grain Corn and Soybeans in Quebec and Ontario in 2000 and 2001,** Bernard Hategekimana
No.55*	(21-601-MIE2002055)	**Recent Migration Patterns in Rural and Small Town Canada,** Neil Rothwell et al
No.56*	(21-601-MIE2002056)	**Performance in the Food Retailing Segment of the Agri-Food Chain**, David Smith and Michael Trant
No.57*	(21-601-MIE2002057)	**Financial Characteristics of Acquired Firms in the Canadian Food Industry**, Martin S. Beaulieu
No.58*	(21-601-MIE2002058)	**Provincial Trade Patterns**, Marjorie Page
No.59*	(21-601-MIE2002059)	**An Analysis of Profits within the Canadian Food Processing Sector**, Rick Burroughs and Deborah Harper
No.60*	(21-601-MIE2002060)	**Rural Diversification**, Marjorie L. Page
No.61*	(21-601-MIE2002061)	**Definitions of « Rural »**, Valerie du Plessie et al
No.62*	(21-601-MIE2003062)	**A Geographic Profile of Canadian Livestock**, Martin S. Beaulieu et Frédéric Bédard
No.63*	(21-601-MIE2003063)	**Sub-provincial Income Disparity in Canada: Evidence from 1992 to 1999**, Alessandro Alasia
No.64*	(21-601-MIE2003064)	**Canada – Mexico Agricultural Economies and Trade Under Closer North American Relations**, Verna Mitura et al
No.65*	(21-601-MIE2003065)	**Computer Technology Adoption by Canadian Farm Businesses: An Analysis Based on the 2001 Census of Agriculture**, Jean Bosco Sabuhoro and Patti Wunsch
No.66*	(21-601-MIE2004066)	**Factors Associated with Household Internet Use in Canada**, Vik Singh