SOCIO-ECONOMIC INDEXES FOR AREAS

PAPER

Australian
Bureau of
Statistics

1991 CENSUS

INFORMATION

SEIFA

five measures of Australia's population

INFORMATION PAPER

1991 CENSUS
SOCIO-ECONOMIC INDEXES FOR AREAS

IAN CASTLES
Australian Statistician

CONTENTS

INQUIRIES   • *for further information about statistics in this publication and the availability of related unpublished statistics,* contact the SEIFA support officer on Canberra (06) 252 5954 or any ABS State Office.

                  • *for information about other ABS statistics and services* please refer to the back page of this publication.

1. INTRODUCTION

This publication presents five summary measures, or indexes, derived from the 1991 Population Census to measure different aspects of socio-economic conditions by geographic areas.

The 1991 Population Census provides information on a broad range of social and economic aspects of the Australian population. Over forty questions of social and economic interest are asked in the Census. People using census data are often interested not just in these items taken one at a time, but in an overview or summary of a number of related items. Statistical techniques can be used to provide such summaries and the index presented in this publication is one such measure.

A measure of socio-economic disadvantage was first produced by the ABS from the 1971 Census. The Socio-economic Indexes for Areas (SEIFA), in their present form, were first produced in 1990 and consist of five indexes formed from the 1986 Census data. The indexes have now been constructed from the 1991 Census using the same methodology.

This information paper describes the indexes and illustrates their possible uses. The approach for the construction of the indexes is outlined with a discussion of the limitations concerning their use. The socio-economic indexes which are available are described, and information is provided on how they can be obtained.

2. THE INDEXES

2.1 Description Of The Indexes

There are five indexes described in this information paper. They relate to socio-economic aspects of geographic areas. Each index summarises a different aspect of the socio-economic conditions in an area. The indexes have been obtained by summarising the information from a variety of underlying social and economic variables, each index using a different set of underlying variables.

The five indexes are:

- Urban Index of Relative Socio-Economic Advantage

- Rural Index of Relative Socio-Economic Advantage

- Index of Relative Socio-Economic Disadvantage

- Index of Economic Resources

- Index of Education and Occupation

All the indexes have been constructed so that relatively advantaged areas, for example areas with many high income earners, have high index values.

The Indexes of Relative Socio-Economic Advantage and the Index of Relative Socio-Economic Disadvantage are general socio-economic indexes. They summarise variables related to the economic resources of households, education and occupation. The variables underlying the Indexes of Relative Socio-Economic Advantage are indicators of relative socio-economic well being (eg high income, tertiary education, skilled occupations). In contrast, the variables used to create the Index of Relative Socio-Economic Disadvantage focus on attributes such as low income, low educational attainment and high unemployment.

A higher score on one of the Indexes of Relative Advantage means that an area has attributes such as a relatively large proportion of households with high incomes or a trained workforce. Conversely, a lower score on the index represents a smaller proportion of households with high incomes, employees in skilled occupation, etc.

A higher score on the Index of Relative Disadvantage suggests that the area has characteristics such as fewer families of low income and fewer people with little training and in unskilled occupations. Conversely, a lower score on the index suggests the area has more low income families and more people with little training and in unskilled occupations.

In most cases, an area that has a high score on the Index of Relative Socio-Economic Advantage will also have a high score on the Index of Relative Socio-Economic Disadvantage. However, it is possible for an area to contain two quite extreme groups, for example, inner city areas in the process of redevelopment. Such an area could have a low score on the Index of Disadvantage (because of the large proportion of low income households) and a high score on the Index of Advantage(because of the high income households).

The Index of Relative Socio-Economic Disadvantage covers all areas in Australia, whereas the Index of Relative Advantage is split into an urban index and a rural index. The urban index is based on areas in urban centres with a population of 1,000 and over, and the rural index is based on the remaining areas of Australia. The urban/rural split was necessitated by the major structural differences found in the relationships between socio-economic variables related to advantage for the urban and rural areas.

The *Index of Economic Resources* reflects the profile of the economic resources of families within the areas. The Census variables which are summarised by this index reflect the income and expenditure of families, such as income and rent and home ownership. Additionally, variables which reflect non-income assets, such as dwelling size and number of cars are also included. The income variables are specified by family structure since this affects disposable income.

The index excludes education and occupation variables, because they are not directly related to economic resources. It also misses some assets such as savings or equities which, although relevant, could not be included because the information was not collected in the 1991 Census.

A higher score on the Index of Economic Resources indicates that the area has a higher proportion of families on high income, a lower proportion of low income families, more households purchasing or owning dwellings, and living in large houses. A low score indicates the area has relatively large proportions of households on small incomes, and living in small dwellings.

The *Index of Education and Occupation* is designed to reflect the educational and occupational structure of communities. The education variables in the index show either the level of qualification achieved or whether further education is being undertaken. The occupation variables classify the workforce into the ASCO (Australian Standard Classification of Occupations) major groups and the unemployed. This index does not include any income variables.

An area with a high score on this index would have a high concentration of persons with higher education or undergoing further education, with people being employed in the higher skilled occupations, rather than being labourers

or unemployed. A low score indicates an area with concentrations of either persons with low educational attainment or unskilled or unemployed people.

Appendix A lists the variables summarised by the indexes. The method for deriving the indexes is briefly described in Section 3, 'Derivation of the Indexes'. See Appendix B for a more detailed description of this method. Factors to be taken into account when interpreting the indexes are discussed in Section 4, 'Comments and Caveats on the Interpretation of the Indexes'.

2.2 Available Geographic Areas

The five index scores are available for a number of different geographic areas, namely:

- Collection District (CD)

- Statistical local area (SLA)

- Legal Local Government Area (LGA)

- Statistical Sub-division (SSD)

- Statistical Division (SD)

- (CD derived) Postcode (PC)

The smallest area for which the indexes are available is the Collection District (CD). A CD is roughly equivalent to a small group of suburban blocks in urban areas. In urban areas it comprises on average about 250 dwellings, while in rural areas it usually contains fewer. In 1991 there were about 31,400 CDs throughout Australia.

Based on the scores for CDs, scores have also been calculated for aggregated geographical areas (SDs, SSDs, SLAs, LGAs, PCs). The index scores for these aggregated areas were formed by taking the weighted average, using population counts from the 1991 Census, across all CDs in the larger geographic area.

Legal Local Government Area (LGA) is a type of spatial unit which represents the whole geographical area of responsibility of an incorporated Local Government Council. There are over 830 LGAs in Australia which collectively cover only a part of Australia.

Statistical local areas are for the most part legal LGA based. In special cases, where a legal LGA is much larger and more populous than the general run of legal LGAs (as in the City of Brisbane), or where there are no legal local government authorities (as in the ACT), the administrative areas have been subdivided to form areas roughly equivalent in extent and population. SLAs cover, in aggregate, the whole of Australia without gaps or overlaps.

Statistical Subdivisions consist of one or more SLAs and Statistical Divisions consist of one or more SSDs. SDs and SSDs do not cross State or Territory boundaries.

CD derived postcodes are approximated by aggregating CDs that lie wholly or partly within the postcode area. A new boundary is then drawn to encompass all the CDs. The data for the CDs within this boundary are aggregated to produce CD derived postcode data. Postcodes that cross over State or Territory boundaries have been totally assigned to the State or Territory that contains the highest proportion of the population in that postcode.

Index values for regions other than postcode, SD, SSD, LGA or SLA, may also be derived. These values are based on the index score of the CDs which make up the region. Each CD score is multiplied by its Census population count and divided by the total regional population count. The overall region score is then the addition of each of these (population) adjusted CD values. Population counts by CD have been provided with the index scores to enable weighted index scores to be calculated for user-defined regions.

Indexes for CDs are provided to construct indexes for larger geographic areas, and are not intended to be used for comparison of individual CDs. The indexes themselves are subject to some imprecision for reasons outlined in section 4.5.

Also, the index values can be distorted by unusual characteristics for that CD. A discussion of some of the characteristics which can lead to CD's having unusually high or low index values is contained in section 4.3.

2.3 Distribution of Index Values

Each index has been designed to have an average across all CDs in Australia of 1000 and a standard deviation of 100 index points. (For normally distributed variables, 95% of units have values within two standard deviations of the average, ie. in this case, 800-1200). In order to provide an intuitive understanding of the indexes, that is, what is a high value and what is a low value, several tables of summary statistics are provided below (see Tables 1-3).

These tables give the average index values for the geographic areas CD, SLA and Postcode in each state and in Australia, as well as a range of quantiles. A quantile is a value at or below which a given fraction of the data must lie.

The 10% quantile for a State gives the index cutoff below which 10% of the index values for that State lie. For example, for CDs in NSW the 10% quantile for the Index of Relative Socio-Economic Disadvantage is 876. This value has the interpretation that one-tenth of CD's in NSW have a score on the Index of Relative Socio-Economic Disadvantage below 876. Similarly, for SLAs in NSW the 10% quantile for the Index of Relative Socio-Economic Disadvantage is 939. This value has the interpretation that one-tenth of SLA's in NSW have a score on the Index of Relative Socio-Economic Disadvantage below 939.

The distributions of index scores are similar across the states. Most noticeable are the different distributions observed for the Northern Territory and ACT. For example, see Table 1b which summarises the CD index scores for the Urban Index of Relative Socio-Economic Advantage. Greater differences would be observed between geographic areas smaller than the state level, for example, postcode or local government area.

The distribution of index values in Tables 1 - 3 refer to different types of spatial unit. Because index scores of SLAs and Postcodes are formed by taking the weighted average of index values of the CDs in the area, their values depend on the distribution of population weights across the CDs.

TABLE 1 SUMMARY DATA FOR CD LEVEL INDEXES

State	Average	*Quantile*				
		10%	25%	50%	75%	90%
INDEX OF RELATIVE SOCIO-ECONOMIC DISADVANTAGE						
New South Wales	1002	876	949	1013	1073	1121
Victoria	1017	909	967	1030	1079	1115
Queensland	988	879	940	997	1046	1086
South Australia	986	841	934	1004	1062	1108
Western Australia	992	866	944	1009	1062	1102
Tasmania	980	861	931	992	1047	1093
Northern Territory	918	747	828	948	1019	1056
Australian Capital Territory	1071	962	1042	1094	1134	1167
Australia	**1000**	**879**	**949**	**1012**	**1068**	**1111**
URBAN INDEX OF RELATIVE SOCIO-ECONOMIC ADVANTAGE						
New South Wales	1002	891	935	982	1052	1147
Victoria	1010	907	947	996	1061	1142
Queensland	983	888	928	969	1028	1099
South Australia	980	866	923	972	1027	1104
Western Australia	1014	892	943	996	1077	1165
Tasmania	970	871	917	961	1024	1083
Northern Territory	929	821	875	936	976	1026
Australian Capital Territory	1091	952	1017	1074	1161	1243
Australia	**1000**	**892**	**936**	**984**	**1052**	**1136**
RURAL INDEX OF RELATIVE SOCIO-ECONOMIC ADVANTAGE						
New South Wales	1007	887	932	1005	1075	1134
Victoria	1025	915	955	1018	1086	1150
Queensland	985	856	907	977	1059	1128
South Australia	992	873	917	983	1055	1128
Western Australia	972	858	896	960	1045	1120
Tasmania	1025	917	962	1019	1082	1149
Northern Territory	913	794	827	881	997	1098
Australian Capital Territory	1035	879	914	991	1101	1267
Australia	**1000**	**876**	**924**	**994**	**1070**	**1135**
INDEX OF ECONOMIC RESOURCES						
New South Wales	1014	890	937	1003	1082	1163
Victoria	1012	906	950	1007	1073	1128
Queensland	979	880	923	972	1031	1088
South Australia	970	846	905	968	1038	1097
Western Australia	997	879	936	996	1062	1120
Tasmania	954	859	902	951	1010	1057
Northern Territory	934	737	822	953	1036	1078
Australian Capital Territory	1100	958	1049	1109	1160	1227
Australia	**1000**	**884**	**933**	**993**	**1063**	**1129**
INDEX OF EDUCATION AND OCCUPATION						
New South Wales	1010	888	938	1002	1076	1152
Victoria	1020	905	953	1014	1083	1147
Queensland	965	859	903	955	1020	1086
South Australia	985	867	918	976	1054	1126
Western Australia	985	870	918	974	1050	1123
Tasmania	970	844	901	955	1039	1120
Northern Territory	999	868	947	1014	1064	1098
Australian Capital Territory	1144	1077	1111	1142	1184	1226
Australia	**1000**	**880**	**929**	**991**	**1067**	**1140**

TABLE 2 SUMMARY AREA DATA FOR SLA LEVEL INDEXES

State	Average	Quantile				
		10%	25%	50%	75%	90%
INDEX OF RELATIVE SOCIO-ECONOMIC DISADVANTAGE						
New South Wales	999	939	966	993	1028	1078
Victoria	1019	956	985	1022	1053	1081
Queensland	1000	916	965	999	1043	1087
South Australia	984	917	955	987	1017	1057
WesternAustralia	989	929	964	993	1017	1042
Tasmania	981	938	952	978	1013	1041
Northern Territory	929	772	874	957	1003	1045
Australian Capital Territory	1075	965	1046	1096	1121	1157
Australia	**1002**	**929**	**966**	**1001**	**1042**	**1089**
URBAN INDEX OF RELATIVE SOCIO-ECONOMIC ADVANTAGE						
New South Wales	983	934	954	968	987	1051
Victoria	990	939	958	979	1007	1059
Queensland	994	914	946	977	1032	1100
South Australia	973	906	931	966	996	1077
Western Australia	990	898	941	977	1012	1087
Tasmania	943	886	929	947	962	1023
Northern Territory	925	801	874	925	984	1025
Australian Capital Territory	1097	1004	1040	1078	1161	1237
Australia	**994**	**920**	**949**	**975**	**1029**	**1096**
RURAL INDEX OF RELATIVE SOCIO-ECONOMIC ADVANTAGE						
New South Wales	1008	914	946	998	1062	1115
Victoria	1037	936	978	1031	1099	1137
Queensland	1017	885	926	1012	1088	1168
South Australia	990	902	934	969	1031	1110
Western Australia	960	893	916	940	990	1087
Tasmania	1032	938	975	1033	1077	1148
Norther Teritory	933	834	844	902	1009	1090
Australian Capital Territory	1066	903	959	1032	1111	1351
Australia	**1008**	**901**	**938**	**995**	**1070**	**1133**
INDEX OF ECONOMIC RESOURCES						
New South Wales	973	904	920	951	1008	1079
Victoria	979	918	937	965	1014	1058
Queensland	994	904	937	986	1042	1095
South Australia	933	867	888	911	973	1032
Western Australia	959	903	921	940	990	1042
Tasmania	936	880	909	933	955	999
Northern Territory	948	775	848	989	1042	1078
Australian Capital Territory	1102	974	1051	1117	1157	1208
Australia	**982**	**896**	**924**	**966**	**1029**	**1099**
INDEX OF EDUCATION AND OCCUPATION						
New South Wales	982	912	935	957	1012	1091
Victoria	994	934	955	980	1022	1081
Queensland	980	883	915	967	1038	1090
South Australia	952	882	904	939	971	1068
Western Australia	953	893	909	937	965	1042
Tasmania	938	876	905	932	964	1018
Northern Territory	1013	898	960	1027	1068	1100
Australian Capital Territory	1141	1082	1117	1138	1176	1215
Australia	**988**	**896**	**928**	**965**	**1040**	**1123**

TABLE 3 SUMMARY AREA DATA FOR POSTCODE LEVEL INDEXES

| State | Average | *Quantile* | | | | |
		10%	25%	50%	75%	90%
INDEX OF RELATIVE SOCIO-ECONOMIC DISADVANTAGE						
New South Wales	1001	919	962	1002	1040	1094
Victoria	1027	956	992	1032	1067	1099
Queensland	989	917	960	992	1022	1051
South Australia	994	905	952	994	1036	1094
Western Australia	992	907	964	1000	1031	1063
Tasmania	987	904	951	983	1031	1083
Northern Territory	931	829	862	967	990	1010
Australian Capital Territory	1071	983	1024	1088	1109	1136
Australia	**1003**	**923**	**966**	**1005**	**1045**	**1086**
URBAN INDEX OF RELATIVE SOCIO-ECONOMIC ADVANTAGE						
New South Wales	998	923	949	975	1031	1119
Victoria	1003	930	958	987	1037	1113
Queensland	974	904	938	967	994	1058
South Australia	978	884	928	967	1016	1087
Western Australia	995	899	942	976	1046	1112
Tasmania	957	886	929	948	1000	1047
Northern Territory	899	815	850	901	953	992
Australian Capital Territory	1073	904	1051	1065	1144	1171
Australia	**991**	**914**	**945**	**975**	**1027**	**1103**
RURAL INDEX OF RELATIVE SOCIO-ECONOMIC ADVANTAGE						
New South Wales	1013	911	946	1004	1073	1132
Victoria	1033	928	963	1022	1095	1164
Queensland	997	890	919	978	1070	1131
South Australia	1008	902	938	984	1075	1155
Western Australia	978	884	913	949	1040	1123
Tasmania	1035	941	971	1027	1089	1162
Northern Territory	962	808	870	961	1009	1181
Australian Capital Territory	1107	909	961	1061	1247	1480
Australia	**1011**	**903**	**940**	**997**	**1076**	**1139**
INDEX OF ECONOMIC RESOURCES						
New South Wales	995	895	925	978	1047	1134
Victoria	991	910	936	973	1050	1093
Queensland	961	885	913	954	1000	1052
South Australia	952	867	893	935	1007	1080
Western Australia	966	897	919	950	1019	1072
Tasmania	942	860	901	935	980	1036
Northern Territory	933	792	867	949	1007	1039
Australian Capital Territory	1089	981	1024	1116	1141	1164
Australia	**977**	**889**	**919**	**962**	**1031**	**1092**
INDEX OF EDUCATION AND OCCUPATION						
New South Wales	992	895	928	968	1049	1125
Victoria	998	918	949	984	1038	1106
Queensland	942	863	891	930	982	1039
South Australia	967	884	909	944	1014	1088
Western Australia	954	880	909	940	979	1057
Tasmania	950	853	896	934	992	1081
Northern Territory	994	894	951	989	1047	1082
ACT	1136	1056	1121	1134	1176	1197
Australia	**977**	**886**	**921**	**960**	**1024**	**1098**

8

2.4 Applications of the Indexes

There are a number of ways the indexes can be used; such as targeting areas for business or services, demographic profiling, strategic planning, allocation of funds, sample surveys, and social or economic research. Some of these uses are described below.

Uses in Research/Data Analysis

The indexes may be useful for modelling or explaining behaviour in other variables. In some studies it is desirable to determine if socio-economic factors are influencing a variable of interest. The researcher may also be interested in reducing the number of variables in the analysis. In such cases, one or more of the indexes can be used as a summary of a range of socio-economic factors.

Example

* The risk of children dying of sudden infant death syndrome (SIDS) in Australia may be associated with the socio-economic environment of parents. This environment could be measured by using a summary measure such as the Index of Education and Occupation.

Initial investigations into the effect of education and occupation could use simple tables or graphs which show values of the variable of interest by index category range. For example, a sample of birth records can be coded into decile groups using the Index of Education and Occupation (1=lowest status to 10=highest status). The number of SIDS deaths can be graphed against these deciles, showing that a significant relationship does exist between SIDS incidence and lower socio-economic status of areas in Australia.

**SIDS DEATHS, AUSTRALIA 1990-92,
BY INDEX OF EDUCATION AND OCCUPATION**

Index EO birth deciles
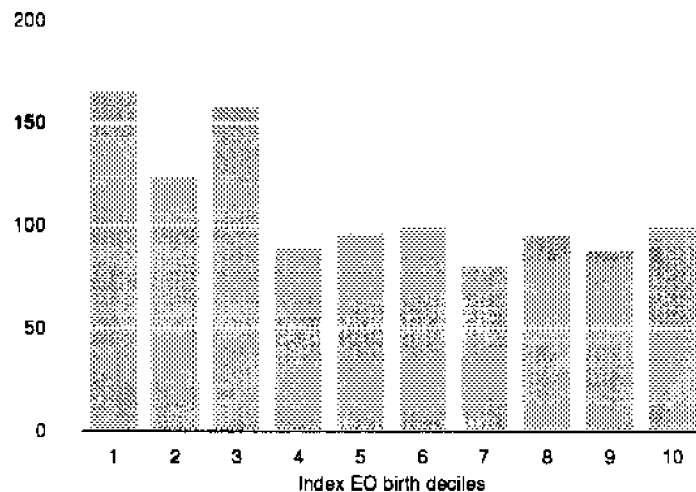
Targeting Areas for Services

The indexes are of interest in their own right as summaries of area characteristics. Areas with different index values will have different socio-economic characteristics. This information can be used by itself or in conjunction with other information, to assist in determining the allocation of services. Those intending to use the indexes in this way are strongly advised to be aware of the limitations of the indexes, described in Section 4, 'Comments and Caveats on the Interpretation of the Indexes'.

Example

● A health organisation has developed a fund allocation formula and wants to ensure resources go to those localities which need them most. Specifically, it wants to check whether disadvantaged suburbs are actually being allocated a sensible proportion of resources using that formula. The rankings of the Index of Relative Socio-Economic Disadvantage scores could be compared to those obtained from using the formula.

Targeting Areas for Business

Businesses might use the indexes to assist with marketing and strategic planning. The indexes might be used simply as summaries of area characteristics. Information from the geographical indexes will be useful for making business decisions, such as siting outlets and targeting promotion campaigns. The indexes are also useful for consumer research.

Example

● A retail organisation wants to establish a chain of boutiques to sell up-market women's clothing in Melbourne and needs to know where to locate the shops. The Urban Index of Relative Socio-Economic Advantage could be sorted to list the most highly ranked Statistical Local Areas (SLA's) within the Melbourne metropolitan Area. Also, mapping the location of the ranges of these index values within Melbourne will help pinpoint the most suitable locality for the boutiques.

Urban Index of Advantage, Melbourne 1991

Melbourne Statistical Division by Statistical Local Area

Index

■ More than 1070
▓ 1020 to 1070
░ 960 to 1020
☐ Less than 960
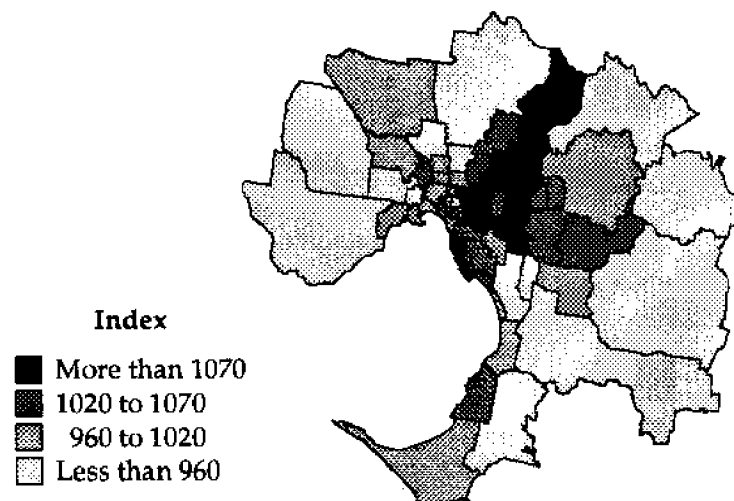
DERIVATION OF THE INDEXES

3.1 Background

The socio-economic indexes, in their present form, were first produced in 1990 from the 1986 Census data, using a statistical technique known as principal component analysis. The same methodology was used to form similar indexes for the 1991 Census data, with minor changes made to some variables underlying the indexes.

3.2 Methodology – Principal Components Analysis

Many aspects of the socio-economic profile of a community cannot be measured directly but there may be several variables which are recognised as contributing to a particular dimension. Often a single composite of these variables, or index, which reflects the population profile of these variables is required to aid social and economic investigations. Principal Component Analysis (PCA) is a technique which is often used to summarise a large number of related variables. By conducting a PCA on a range of variables related to the socio-economic factor of interest, a socio-economic index can be derived. This socio-economic index measures what is common to the variables included in the analysis.

While PCA was the multivariate technique used to summarise variables, there were several other equally important stages in the production of these indexes. The first stage involved choosing and refining the list of variables to be analysed using PCA. Then, after conducting the PCA, the indexes were checked to ensure that they were indeed reflecting the desired socio-economic aspects of each area. Each of these stages is discussed briefly in the following sections. For more details see Appendix B.

3.3 Methodology – Choosing the Variables

In deriving an index, PCA is being used to find a combination of variables to act as a summary measure, drawing out what is common to the variables. It is therefore important to ensure that sufficient variables are used to represent all aspects of a particular socio-economic indicator.

The initial variables for each index were selected from the information available from the 1991 Census. The variables were chosen subjectively, based on experience with the 1986 indexes. In deciding the variables for the 1986 indexes, comment on the lists of variables were sought, and obtained from several academics and research institutions around Australia. Their advice resulted in the removal of some variables, and the addition (where data was available) of others. Similar initial variables were adopted for the 1991 Census, with some minor changes to the variables, especially those underlying the Rural Index of Socio-Economic Advantage and the Index of Economic Resources. See Appendix B for further detials on these changes.

The finest level at which complete Census data is disseminated is the Census Collection District (CD), which corresponds to the workload of one census collector. In 1991 there were about 31,400 CDs throughout Australia. By calculating an index at the CD level, an index at any broader level can be obtained by combining the constituent CD index scores . Thus the CD was chosen as the appropriate level for analysis.

The data on the initial input variables was analysed to ensure that particular socio-economic aspects were not over-represented in the analysis, as this would lead to an index weighting unreasonably highly on this aspect. Also, only variables which were well-related to the general thrust of the index were included. Analysis was performed on the initial results of the PCA to exclude variables which correlated poorly with the index.

3.4 Producing the Indexes from the Raw Scores

To allow for easy recognition of high and low scores, the CD level index scores have been standardized to have a mean of 1000 and a standard deviation of 100. The scores themselves are not in direct proportion to the relative socio-economic conditions of the various geographic areas (CDs, SDs, SSDs, LCAs, SLAs and Postcodes). That is, a CD with a score of 1200 is not necessarily twice as advantaged as a CD with a score of 600. Therefore, the indexes should be used only for ordering CDs and not for analyses which purport to somehow quantify socio-economic conditions.

Scores for areas larger than CDs have been calculated by weighting together constituent CD scores, using the CD population size for weighting. It must be noted that these scores are CD weighted averages and are not quite the same as those that would have been produced if the PCA had been carried out separately on the larger geographic areas.

3.5 Validation

Socio-economic well-being is not a simple, nor well defined concept. Given the need to choose which variables to include and exclude, and the need to interpret the meaning of the summary variables resulting from the analysis, it was clearly necessary to scrutinise the final indexes carefully, to ensure that they provided a valid measure and behaved as expected. The main validation exercises carried out on the final indexes were to:

• check that variables and their weights make sense;

• compare the indexes with the 1986 indexes;

• use local subjective knowledge to rank some CDs, and to study extreme CDs; and

• compare the indexes with data from other sources.

These validation exercises resulted in improvements to the methodology and helped to ensure the final scores are valid and consistent with other sources of information. See Appendix B for more detail on these validation exercises.

4. COMMENTS AND CAVEATS ON THE INTERPRETATION OF THE INDEXES

4.1 The Choice of Variables for each Index

The Indexes which have been produced depend upon the variables that were analysed using Principal Components Analysis. Different underlying variables would result in different final indexes; the indexes presented in this paper are just five of the many that could be derived from the Census variables. Other indexes could be developed which focus on particular social conditions. The choice of an index depends on the socio-economic aspect of interest, and the underlying variables which represent those aspects most precisely.

Indexes produced using PCA can be affected if some socio-economic aspects are over- or under-represented in the variables being analysed. Over-representation was dealt with using the procedures described in Appendix B. However, if variables relating to an important aspect of a socio-economic dimension under consideration are absent from a particular index, users should be aware that the index cannot completely represent that aspect of socio-economic advantage or disadvantage. Consequently, the indexes described in this paper do not provide good measures for all social conditions. They were derived as good overall indexes, but should be used in

conjunction with other information that relates to the topic of interest. For example, the age structure of the population is not used directly in any of the indexes. Thus, if the topic of interest relates to the health needs for infant care, or for the aged, specific data on that segment of the population should be used in addition to the indexes.

4.2 Socio-Economic Dimensions not Represented in the Indexes

Users of the Indexes should examine the constituents of the indexes (see Appendix A) to ascertain whether they are appropriate to their problem or analysis. However there are three factors in particular which the Indexes do not represent well, and which should be noted.

Firstly, the Indexes contain only limited aspects of wealth. While income and expenditure are represented, aspects such as inherited wealth, savings, indebtedness, and property values are not included. These aspects were not included as details on them were not collected by the Census. This affects the Index of Economic Resources more seriously than the other indexes.

Secondly, family structure (number of income earners, number of parents, number of dependents etc) is not strongly represented in the indexes though it does appear to some extent in the Index of Economic Resources. As a consequence, the Indexes will perform relatively poorly at distinguishing between different family types directly.

Thirdly, access to infrastructure such as schools, community services, shops and transport are not represented by the indexes. These variables are considered to be integral to the concept of advantage or disadvantage. For example, rapidly growing outer suburban areas may suffer from a locational disadvantage situation rather than a socio-economic disadvantage.

4.3 Understanding the Indexes

The indexes produced by Principal Component Analysis depend solely on the linear relationship between variables. It is quite probable however, that the socio-economic conditions in an area are not related to the variables in the analysis in a purely linear fashion. As a result, the indexes are 'ordinal measures' and not 'interval measures'. That is, using the indexes to order the CDs is meaningful but other arithmetic relationships between index values are not meaningful. For example, a CD with an index value of 1200 does not have twice the well being of a CD with an index value of 600. Similarly, the socio-economic difference between two CDs with index values of 800 and 900, is not necessarily the same as the difference between two CDs with index values of 1050 and 1150.

The indexes reflect the socio-economic well being of an area, rather than that of individuals. They were calculated at the CD level, and therefore reflect CD characteristics. Because all people within a CD are not identical, the index scores for a CD do not directly apply to individuals within that CD but rather the scores reflect the way that people group together in CDs. Hence it is possible for a relatively advantaged person to be resident in a CD which may have a low score on some or all of the indexes. Thus it is not appropriate to make inferences regarding a particular individual on the basis of the index scores.

The degree of heterogeneity within a CD influences the index score of that CD; the more homogeneous CDs tend towards the extreme index scores. That is, those CDs which have large proportions of households with similar characteristics, will tend to have the lowest or highest index scores.

Partly because of this, the interpretation of the index values is more straightforward for areas which have extreme values (ie very high or very low index values). For example, it is usually easy to see why a CD which is in the top (or bottom) 5% of index values has that status. In contrast, areas with mid-range index values tend to contain a broader mix of people and households. As a consequence, it is more difficult to draw strong comparisons between two mid-ranked CDs (for example, between a CD just below the average on the index (just below 1000), and another CD just above the average).

4.4 Areas With No Index Values

For confidentiality reasons and to ensure the indexes are meaningful, the data for some CDs has been withheld and is not available at any of the locality levels. These excluded CDs have one or more of the following characteristics:

- populations smaller than or equal to 10;

- 5 people or fewer employed;

- more than or equal to 70% of people not responding to any of the Census questions on family income, occupation, labour force status, type educational institution, and qualifications;

- more than 20% of dwellings are non-private; or

- off-shore and migratory CDs.

In total about 1% of Australian CDs fell in one of the above. The Australian mean and standard deviation were calculated without using these CDs.

4.5 Limitations of Census Data

There are a number of features of the Census data used to construct the indexes, which can affect the usefulness of the indexes. Users should therefore be aware of the following:

(a) The variables included in analysis are limited to those for which data is collected by the Census. Ideally, an indicator of a socio-economic factor should include all measures of relevance to that factor. However, the Census does not obtain any information relating, for example, to wealth, and access to infrastructure such as schools, community services, transport and shops . The indexes cannot therefore purport to summarize these facets of socio-economic well being.

(b) Missing data is a further impediment to index construction. Although non-response to individual Census items is overall quite low, it does vary between CDs. It is possible that item non-response rates correlate directly with socio-economic disadvantage. Where possible, non-response for a variable has been dealt with by redefining the population associated with the variable, to include only those persons who answered the relevant questions. This approach implicitly assumes that non-respondents within a CD resemble respondents within that area, with respect to the characteristics measured by the variables.

(c) All variables pertaining to families and dwellings, in contrast to persons, are based on data from "private" dwellings and caravans in parks only. Persons in non-private dwellings (eg motels, boarding houses, hospitals, refuges) are therefore "under-represented".

(d) The Census tables on social and economic aspects of the population is based on people's place of enumeration and not their usual residence, ie the population is classified to CDs according to where they were spending the night at the time of the Census. Although the census is timed to attempt to capture the typical situation, holiday resort areas, such as the Gold Coast, may show a large enumeration count compared with the usual residence count.

4.6 Comparison with the 1986 Census

It is important that the index scores from the 1991 Census should not be compared directly to the indexes based on the 1986 Census. The index values for CDs are standardised to have a mean of 1000 across Australia and the difference between the scores of an area in 1986 and 1991 does not represent the change of socio-economic conditions in the area. Also, since the indexes are not interval measures (see section 4.3 above) the difference between the index scores of two areas in 1991 cannot be compared to the difference in 1986 to show whether the gap between the socio-economic conditions of the two areas is narrowing or widening.

Boundaries of CDs in some areas have changed between censuses. The actual number of CDs in Australia has increased from 29,632 in 1986 to 31,401 in 1991. Consequently, the boundaries of the higher geographic areas such as SLA and LGA may not be comparable. Approximately 77% of 1991 CDs in Australia are comparable within a 10% dwelling limit to the 1986 CDs.

As well, there are some minor difference between the 1986 indexes and the 1991 indexes. However, these are not of significance in terms of their impact on the rankings of CDs, and for all practical purposes in terms of using the indexes to rank CDs, the 1991 indexes can be viewed as an update of the 1986 indexes.

5. DATA AVAILABILITY

Index values for the five indexes are available at the CD, SD, SSD, LGA, SLA and CD-derived Postcode level for Australia. State and Australia wide indexes are available on floppy disk. The first disk product will be a 'stand alone' product, complete with a software package which provides a fully-documented user-friendly interface to the indexes. The second product, is a 'CDATA91 add-on' option suitable for use by those who have access to CDATA91. This option enables users to access and manipulate the indexes through the powerful functions within CDATA91 for area selection, mapping, graphing and display of information.

An order form for the floppy disk products is given at the back of this publication. Alternatively, special index data sets for particular areas can be obtained on floppy disk or as hard copy, from the ABS Information Consultancy or Statistical Consultancy Services in your state capital city office. Our Statistical Consultants will also be able to provide assistance with using the indexes for various applications or, if necessary, designing other indexes to meet specific needs.

For users who are interested in data on specific Census variables, CDATA91 contains tables on community profile for all geographic areas and covering most topics on the census form. Further small area data are available from GeoLink, which enables users to get access to ABS data from a wide range of social and economic data collections, including Population Census, Estimated Resident Population, Monthly Population Survey, Business Register, Agricultural Census, Manufacturing Census, Retail Census, Building Activity Survey, etc. Two sets of data are available, one at SD level and above, the other

at SLA level and above. GeoLink lets users export data directly into CDATA and thematically display the data alongside Census information.

If you are interested in the above products or services, your first point of contact in all circumstances should be the inquiry staff of your state office of the Australian Bureau of Statistics through the numbers given below.

SEIFA CONTACT INFORMATION

NEW SOUTH WALES
3rd Floor, St Andrews House
Sydney Square
SYDNEY NSW 2000
(Box 796, GPO Sydney 2001)
Phone (02) 268 4611
Fax (02) 268 4668

VICTORIA
5th Floor, Rialto North Tower
525 Collins Street
MELBOURNE VIC 3000
(Box 2796Y, GPO Melbourne 3001)
Phone (03) 615 7000
Fax (03) 615 7798

QUEENSLAND
13th Floor, 313 Adelaide Street
BRISBANE QLD 4000
(Box 9817, GPO Brisbane 4001)
Phone (07) 222 6351
Fax (07) 229 6042

SOUTH AUSTRALIA
Commonwealth Centre
55 Currie Street
ADELAIDE SA 5000
(Box 2272, GPO Adelaide 5001)
Phone (08) 237 7100
Fax (08) 237 7566

WESTERN AUSTRALIA
Level 1, Hyatt Centre
30 Terrace Road
EAST PERTH WA 6004
(Box K881, GPO Perth 6001)
Phone (09) 323 5140
Fax (09) 221 2374

TASMANIA
1st Floor, 175 Collins Street
HOBART TAS 7000
(Box 66A, GPO Hobart 7001)
Phone (002) 20 5800
Fax (002) 20 5995

NORTHERN TERRITORY
6th Floor, MLC Building
81 Smith Street
DARWIN NT 0800
(Box 3796, PO Darwin 0801)
Phone (089) 81 3456
Fax (089) 81 1218

AUSTRALIAN CAPITAL TERRITORY
Wing 5, Cameron Offices
Chandler Street
BELCONNEN ACT 2617
(Box 10, PO Belconnen 2616)
Phone (06) 252 6627
Fax (06) 253 1404

APPENDIX A

VARIABLES UNDERLYING SOCIO-ECONOMIC INDEXES

In this Appendix, we list the initial variables considered for inclusion in the various indexes. We also group the variables by the value of their weight to indicate the contribution of each variable to the index. Not all the initial variables are included in the final indexes. (See Section 3.3 and Appendix B). We include below also those variables which have been excluded from the indexes as a result of the analysis.

Index of Relative Socio-Economic Disadvantage

weight between 0.3 and 0.4
Families with income less than $16,000 (%)
Persons aged 15 and over with no qualifications (%)
Employed Persons classified as 'Labourer & Related Workers' (%)
Males (in labour force) unemployed (%)
Females (in labour force) unemployed (%)

weight between 0.2 and 0.3
Dwellings with no motor cars at dwelling (%)
Households renting (government authority) (%)
Persons aged 15 and over who left school at or under 15 years of age (%)
One parent families with dependent offspring only (%)
Persons aged 15 and over separated or divorced (%)

weight between 0.1 and 0.2
Dwellings with 1 or no bedrooms (%)
Households renting (non-government authority) (%)
Persons aged 15 and over who did not go to school (%)
Employed Males classified as 'Tradespersons' (%)
Aboriginals or Torres Strait Islanders (%)
Lacking fluency in English (%)

dropped initial variables
Families with offspring having parental income less than $16,000 (%)
Employed Females classified as 'Tradespersons' (%)
Employed Females classified as 'Salespersons & Personal Service Workers' (%)
Employed Males classified as 'Clerks' (%)
Households in improvised dwellings (%)
Occupied private dwellings with 2 or more families (%)
Recent migrants from non-English speaking countries (%)

Urban Index of Socio-Economic Relative Advantage

weight between 0.3 and 0.5
Families with income greater than $70,000 (%)
Persons aged 15 and over with degree or higher (%)
Employed Persons classified as 'Managers or Administrators' (%)
Employed Persons classified as 'Professionals' (%)

weight between 0.2 and 0.3
Average number of bedrooms per person
Dwellings with 4 or more bedrooms (%)
Dwellings with 3 or more motor cars (%)
Persons aged 15 and over at CAE or university (%)
Employed Males classified as 'Salespersons & Personal Service Workers' (%)

weight between 0.1 and 0.2
Households owning dwelling (%)
Households owning or purchasing dwelling (%)
Persons aged 15 and over who are still at school (%)

dropped initial variables
Families with offspring having parental income greater than $60,000 (%)
Households purchasing dwelling (%)
Persons aged 15 and over with trade or 'other' qualification (%)
Persons aged 15 and over at TAFE (%)
Employed Males classified as 'Managers or Administrators' (%)
Employed Females classified as 'Managers or Administrators' (%)
Employed Males classified as 'Professionals' (%)
Employed Females classified as 'Professionals' (%)
Employed Males classified as 'Para-Professionals' (%)
Employed Females classified as 'Para-Professionals' (%)
Employed Males classified as 'Clerks' (%)
Employed Females classified as 'Clerks' (%)

Rural Index of Socio-Economic Relative Advantage

weight between 0.3 and 0.4
Households purchasing dwelling (%)
Persons aged 15 and over with trade or 'other' qualification (%)
Employed Persons classified as 'Professionals' (%)
Employed Males classified as 'Salespersons & Personal Service Workers' (%)

weight between 0.2 and 0.3
Families with income greater than $70,000 (%)
Families with offspring having parental income greater than $60,000 (%)
Households owning or purchasing dwelling (%)
Persons aged 15 and over with degree or higher (%)
Persons aged 15 and over at CAE or university (%)
Employed Males classified as 'Tradespersons' (%)
Employed Males classified as 'Clerks' (%)
Employed Females classified as 'Clerks' (%)
Employed Females classified as 'Salespersons & Personal Service Workers' (%)

weight between 0.1 and 0.2
Persons aged 15 and over at TAFE (%)
Employed Females classified as 'Para-Professionals' (%)

dropped initial variables
Households owning dwelling (%)
Dwellings with 3 or more motor cars (%)
Dwellings with 4 or more bedrooms (%)
Average number of bedrooms per person
Persons aged 15 and over who are still at school (%)
Employed Males classified as 'Managers or Administrators' (%)
Employed Females classified as 'Managers or Administrators' (%)
Employed Persons classified as 'Managers or Administrators' (%)
Employed Males classified as 'Professionals' (%)
Employed Females classified as 'Professionals' (%)
Recent migrants from non-English speaking countries (%)
Lacking fluency in English (%)

Index of Economic Resources

weight between 0.2 and 0.4

Families with other family structure, and income greater than $70,000 (%)

Dwellings with 4 or more bedrooms (%)

Households owning or purchasing dwelling (%)

Rent greater than $227 per week (%)

Families consisting of a single parent with dependent offspring, with income greater than $25,000 (%)

Families consisting of a two parent family with dependent offspring, and income greater than $60,000 (%)

Families consisting of a couple only, and with income greater than $50,000 (%)

weight between 0 and 0.2

Households purchasing dwelling (%)

Dwellings with 3 or more motor cars (%)

Mortgage greater than $1000 per month (%)

Households owning dwelling (%)

weight between -0.2 and 0

Households in improvised dwellings (%)

Dwellings with 1 or no bedrooms (%)

Households renting (government authority) (%)

Families consisting of a single parent with dependent offspring, with income less than $12,000 (%)

weight between -0.3 and -0.2

Dwellings with no motor cars (%)

Rent less than $78 per week (%)

Families consisting of a two parent family with dependent offspring, and income less than $25,000 (%)

% families consisting of a couple only, and with income less than $16,000 (%)

% families with other family structure, and income less than $25,000 (%)

dropped initial variables

Average number of bedrooms per person

Households renting (non-government authority) (%)

Households who are group households (%)

Index of Education and Occupation

weight between 0.2 and 0.4

Employed Persons classified as 'Professionals' (%)

Persons aged 15 and over at CAE or university (%)

Employed Males classified as 'Salespersons & Personal Service Workers' (%)

weight between 0 and 0.2

Persons aged 15 and over at TAFE (%)

Employed Males classified as 'Para-Professionals' (%)

Employed Females classified as 'Para-Professionals' (%)

Employed Males classified as 'Clerks' (%)

Employed Females classified as 'Clerks' (%)

Persons aged 15 and over who are still at school (%)

weight between -0.3 and 0

Employed Males classified as 'Tradespersons' (%)

Employed Females classified as 'Tradespersons' (%)

Males (in labour force) unemployed (%)

Females (in labour force) unemployed (%)

weight between -0.5 and -0.3

Persons aged 15 and over who left school at or under 15 years of age (%)
Employed Persons classified as 'Labourer & Related Workers' (%)
Persons aged 15 and over with no qualifications (%)

dropped initial variables

Persons aged 15 and over with degree or higher (%)
Persons aged 15 and over with trade or 'other' qualification (%)
Persons aged 15 and over who did not go to school (%)
Employed Males classified as 'Managers or Administrators' (%)
Employed Females classified as 'Managers or Administrators' (%)
Employed Persons classified as 'Managers or Administrators' (%)
Employed Males classified as 'Professionals' (%)
Employed Females classified as 'Professionals' (%)
Employed Females classified as 'Salespersons & Personal Service Workers' (%)

APPENDIX B

DERIVATION OF THE SOCIO-ECONOMIC INDEXES FOR AREAS

1 Background

A measure of socio-economic disadvantage was first produced by the ABS from the 1971 Census, using a statistical technique known as principal components analysis (PCA). Since then indexes have been produced from the 1976 and 1981 Censuses specifically to reflect educational disadvantage and therefore were based on Census variables which had been found to be correlated with educational achievement. However, evaluation indicated that although these indexes were derived to measure educational disadvantage, they could also be useful as a general socio-economic index.

In many circumstances, a single socio-economic index may be too broad to be useful in a particular analysis or survey design, and for this reason the ABS developed a group of indexes for the 1986 Census. This group was made up of two general indexes (an Index of Relative Advantage and an Index of Relative Disadvantage) and two more specific indexes. Separate indexes for urban and rural areas were derived for the general Index of Relative Socio-Economic Advantage, when it became apparent there were major structural differences in the relationships between socio-economic variables for these two areas. The specific indexes were an index of economic resources and an index of education and occupation. The same five indexes were constructed in 1991 with some minor changes in variables and procedures. A description of the derivation of the indexes follows.

2 Methodology – Principal Components Analysis

Many aspects of the socio-economic profile of a community cannot be measured directly but there may be several variables which are recognised as contributing to a particular dimension. Often a combination of these variables, or an index, which reflects the population profile of these variables is required to aid social and economic investigations. Principal Component Analysis (PCA) is a technique which is often used to summarise a large number of related variables. By conducting a PCA on a range of variables related to the socio-economic factor of interest, a socio-economic index can be derived.

The main aim of PCA is to reduce a large number of related variables to a new set of (uncorrelated) components, which are ordered so that the first few components explain most of the variation present in the original variables. To use the PCA technique, a number of underlying variables are measured on each of the population units. The correlation or covariance matrix of these variables is then analysed, in order to extract the underlying factors or components from them.

Each principal component is a linear combination of the original variables, and is independent of the other components. As well as producing a set of principal components, the PCA technique sorts them. Thus it is possible to talk of the first principal component, the second principal component, and so forth. The first principal component is usually the most important one. It is the linear combination of the original variables which best summarizes the variance in the original data. Subsequent principal components are linear combinations of the original variables, which form the best summary of the variance remaining in the data, after allowing for the previous principal components.

For each of the five indexes, a set of relevant original variables was chosen and in each case, the first principal component has provided the socio-economic measure. The first component was appropriate in each case because the variable loadings and the correlations between the variables and the component, identified a clear socio-economic dimension.

Once the linear combination of variables that compose a principal component is known, a score can be calculated for each CD. The linear combination, in effect, specifies a weight for each of the original variables. The component score for a CD can be calculated by applying the appropriate weight to the value of each variable for the CD, and then adding up the weighted values. These scores can then be used to distinguish between CDs and to rank them.

There were several other equally important stages in the production of these indexes. The first stage involved choosing and refining the list of variables to be analysed using PCA. Then, after conducting the PCA, the indexes were scaled to have an average value of 1000 and were checked to ensure that they were indeed reflecting the desired socio-economic aspects of each area. Each of these stages is discussed in the following sections.

3   Methodology – Choosing the Variables

In deriving an index, PCA is being used to find a combination of variables to act as a summary measure. It is drawing out what is common to these variables. It is therefore important to ensure that sufficient variables to represent all aspects of a particular socio-economic indicator are included.

The initial variables for each index were selected from the information available from the 1991 Census. While the majority of variables were formed from the 1991 Census Basic Community Profiles, several variables had to be derived from the 1991 Final Unit Record File. These include the family income by family type variables and recent migrants to Australia.

The variables were chosen subjectively, based on experience with the earlier indexes. Variables pertaining to family income, educational attainment, unemployment, occupation, marital status, household occupancy, Aboriginality and migrant status were included in the analysis. All variables which were chosen had to satisfy the criterion of face validity (i.e. they seem to conform to what intuition might dictate). When the variables were considered for the 1986 indexes, comment on the lists of variables was obtained from several academics and research institutions around Australia. Their advice resulted in the removal of some variables, and the addition (where data was available) of others.

Most of the variables for the 1986 indexes have been retained for the 1991 indexes. Changes were made to the initial variables for the Index of Economic Resources and Indexes of Advantage to improve the family income by family structure variables , to add new variables on households owning or purchasing their dwelling, and to include family as well as parental incomes.(Family income and parental income were highly related in the urban areas of Australia, but not so highly in the rural areas.) These variables were considered relevant to economic resources and relative socio-economic advantage.

It is equally important to ensure that particular socio-economic aspects are not over-represented in the analysis, as this would lead to an index weighting unreasonably highly on this aspect. An extreme example would be the inclusion of the same variable twice in the analysis. To avoid such over-representation, the correlations of the initial input variables were

examined. If any two of the variables had a very high correlation only one of the pair was retained for the PCA.

It is also important that only variables which are well-related to the general thrust of the index are included. Variables which correlate poorly with the index do little but add to the variability of the index. These variables are not related to the main thrust of the Index, and can make the index unnecessarily sensitive to small changes in the population over time. Therefore, after the first PCA, those variables which had low correlations with the index were excluded. The PCA was then repeated to produce the final index.

All the variables used in the analysis were expressed as ratios or percentages (eg as a percentage of persons aged 15 years or more, as a percentage of males in the employed labour force etc) to make the measurements comparable between CDs. When deriving the principal components, the correlation matrix for the variables was used. Using correlations rather than covariances in the PCA, gives equal prominence to all variables, so that variables with a large range of values do not dominate the indexes.

4 Selecting Variables for the Indexes of Relative Socio-Economic Advantage and Disadvantage

A single general socio-economic index will often fail to identify an area with two quite extreme groups. For example, an inner city area might contain both pensioner households and double income households with no children. Thus it might have high proportions of both relatively advantaged and disadvantaged households. It was therefore decided to produce two indexes: an Index of Relative Socio-Economic Advantage and an Index of Relative Socio-Economic Disadvantage. In this way groups of concentrated need can be identified even if they are mixed with a group of wealthy individuals, or vice versa.

Because major structural difference had been found in the relationships between socio-economic variables related to advantage for the urban area and the rural area, the Index of Relative Advantage was split into two: an urban index and a rural index. CDs were defined as urban or rural using the 1991 Population Census definition of 'Section of State'. Thus the Urban Index of Relative Advantage was based on all CDs in urban centres with a population of 1000 and over. The Rural Index of Relative Socio-Economic Advantage was based on all remaining CDs.

A preliminary PCA was conducted on the variables chosen to be in the Indexes of Advantage and Disadvantage, producing an overall general socio-economic index. Variables which correlated negatively with the overall index were assigned to the Index of Relative Disadvantage. The Index was then derived by carrying out PCA for this set of variables only, excluding any unnecessary variables using the procedures described in the previous section.

The Urban and Rural Indexes of Advantage were derived by first conducting a preliminary PCA in urban and rural areas separately on the variables chosen to be in the Indexes of Advantage and Disadvantage, producing general urban and general rural socio-economic indexes. Variables which correlated positively with these overall indexes were assigned to the Urban Index of Advantage and Rural Index of Advantage. PCA was then carried out on these sets of separate variables, excluding any unnecessary variables using the procedures described in the previous section.

5 Selecting variables for the Specific Indexes

The Index of Economic Resources was derived by doing the PCA on variables which reflect the profile of the economic resources of families within CDs, such as income and rent and mortgage variables. Additionally, variables which

reflected non-income assets, such as dwelling size and number of cars were also included. The income variables were specified by family structure since this affects disposable income. Unnecessary variables were excluded from the initial list using the same procedures as other indexes.

The Index of Education and Occupation includes variables which are related to the educational and occupational structure of communities. This index is derived by the same procedures as the other indexes.

6 Producing the Indexes from the Raw Scores

To allow for easy recognition of high and low scores, the CD level index scores have been standardized to have a mean of 1000 and a standard deviation of 100.

Scores for areas larger than CDs can be calculated by weighting together constituent CD scores. Scores have been produced for Statistical Divisions (SDs), Statistical Subdivisions (SSDs), Legal Local Government Areas (LGAs), Statistical Local Areas (SLAs) and Postcodes, using the CD population size for weighting. It must be noted that these scores are CD weighted averages and are not quite the same as those that would have been produced if the PCA had been carried out separately on the larger geographic areas.

7 Validation

The derivation of socio-economic indexes is subjective in nature, as socio-economic well-being is not a simple, nor well defined concept. Given the need to choose which variables to include and exclude, and the need to interpret the meaning of the summary variables resulting from the analysis, it was clearly necessary to scrutinise the final indexes carefully, to ensure that they provided a valid measure and behaved as expected. Validation has been a very important part of the derivation of the indexes.

One important check on the indexes is whether the variables and their weights make sense. After each PCA, the first principal component was examined to see if it was summarising the input variables adequately. In all cases the final indexes were explaining about 30% of the variability in the underlying input variables - a good indication that some common underlying factor was being identified and summarised. The weights of the variables in each index also displayed face validity, ie they made intuitive sense (high income has a high weight, while low income or unemployment have low or negative weights; purchasing a dwelling has a higher weight than renting a dwelling; high rent has a higher weight than low rents; tertiary education has a higher weight than leaving school at 15 and so forth).

The indexes were further validated using local subjective knowledge. In each state, several CDs were identified which covered the full range of index values (from low to high status). These CDs were independently assessed and graded using local subjective knowledge and direct observation of the CDs. With the exception of the Rural Index of Advantage, the gradings assigned subjectively corresponded closely with the rankings provided by the index values.

One of the problems associated with the rankings of the Rural Index of Advantage was that a local knowledge based approach ranked some CDs with high percentage of Aboriginals/Torres Strait Islanders much lower than the index ranking suggested. Investigation showed that this was a result of one or more of the following:

• a high percentage of people not stating their income, education and/or occupation;

- a small percentage of people employed; and/or

- a large percentage of people not in the labour force (variables used in the indexes exclude those not in the labour force).

The solution to this problem was to exclude CDs which have highly unstable variables due to the sparsity of responses to the Census. It was therefore decided to exclude from the PCA those CDs with one or more of the following characteristics:

- populations smaller than or equal to 10;

- 5 people or fewer employed;

- more than or equal to 70% of people not responding to any of the Census questions on family income, occupation, labour force status, type educational institution, and qualifications;

- more than 20% of dwellings are non-private; or

- off-shore and migratory CDs.

In total this led to the exclusion of about one per cent of Australian CDs. The Australian mean and standard deviation were calculated without using these CDs. When aggregating areas to a higher level the index values for these CDs are excluded from the weighted average.

The validation also suggested a problem with the inclusion of the variable "Percentage of Persons classified as Managers/Administrators" in the Rural Index of Relative Socio-Economic Advantage. This variable was at first considered as an advantage variable when PCA was performed for all CDs in Australia, but in the rural areas (where many managers/administrators were farmers) the variable, along with another three variables, was found to be negatively correlated with the Rural Index of Advantage. This suggested the PCA for determining advantage variables should be performed separately for the urban and rural areas. The validation of the Rural Index of Advantage also suggested that there was too much emphasis placed on the occupation related variables. To counteract this problem two income variables were added to the index. The derivation method was amended and the analysis was rerun and subsequently, the negatively correlated variables were dropped from the index.

The top and bottom ranked CDs in each state were also examined using local subjective knowledge. The characteristics which caused them to have such extreme index values all made intuitive sense. They tended to be homogeneous CDs, ie ones with uniform characteristics throughout the CD. High values related to high income, home ownership and professional employment or trade qualification. CDs with low index values tended to be characterised by low car ownership, unemployment or unskilled jobs, relatively lower educational achievements and low incomes.

The indexes were then compared to the indexes obtained from the 1986 Census. The top and bottom ranked CDs and SLAs in each state compared favourably. After matching up comparable CDs, correlations between the 1986 and 1991 indexes were found to be quite high at CD, SLA, LGA and Postcode levels, with the exception of the Rural Index of Advantage - this index has changed the most from 1986.

Finally, additional data sources were used to validate the index scores. Data from the current 1988 ABS Household Expenditure Survey (HES) was used to compare (HES) income and mortgage expenditure information, with index scores from the Index of Economic Resources. As expected, the data showed slight increases in income and mortgage payments as the index values increased. 1993 Labour Force Survey data was used to try to find relationships between the Index of Education and Occupation and various fine level occupation groups. As expected, low index values were correlated with areas with higher proportions of the unskilled occupation groups and above average values with areas with higher proportions of the skilled occupation groups.

Australian Bureau of Statistics

1991 Census
Socio-Economic Indexes For Areas

ORDER FORM

I wish to purchase the following:

(Please tick the appropriate box below)

Catalogue Number 2725	Disk Size 3.5"/5.25"	Price $	Quantity	Cost $
Stand Alone				
☐ Australia		2,200		
☐ New South Wales		1,500		
☐ Victoria		1,500		
☐ Queensland		1,500		
☐ South Australia		900		
☐ Western Australia		900		
☐ Tasmania		500		
☐ Northern Territory		500		
☐ Australian Capital Territory		500		
CDATA add-on (for CDATA91 Users only)				
☐ Australia		2,200		
☐ New South Wales		1,500		
☐ Victoria		1,500		
☐ Queensland		1,500		
☐ South Australia		900		
☐ Western Australia		900		
☐ Tasmania		500		
☐ Northern Territory		500		
☐ Australian Capital Territory		500		
Total Cost				

Note: This product includes an Information Paper and User's Manual, with demonstration disks being available on request. Discounts are available for bulk purchases. If you have previously purchased the 1986 Socio-Economic Indexes, you are entitled to a 20% discount.

Method of payment

☐ My Crossed Cheque/Purchase Order for $ _____ payable to the **Australian Bureau of Statistics** is attached

☐ Bankcard ☐ Mastercard ☐ Visa

Card Number

Amount $ _____

Name on Card _____   Expiry date _____ / _____

Signature _____

Please send the package to: *(Block letters please)*

Name _____

Position _____

Organisation _____

Address _____

_____ Postcode _____

Telephone ( ) _____   Fax ( ) _____

Catalogue No. 2725._ _ _ _ _	For Internal Use		Project 5241
Sales Person	Name	Sales Office	
	Phone ( )	Date ____ / ____ / ____	
Comments			

I have read and understood the conditions of sale

Signature _____   Date ____ / ____ / ____

CONDITIONS OF SALE FOR ABS PRODUCTS

Australian Bureau of Statistics (ABS) products are sold by the Commonwealth of Australia ("the Commonwealth") through the ABS. Prices are subject to change without notice.

COPYRIGHT

The Commonwealth holds the copyright of ABS products. The Client agrees not to copy or otherwise reproduce the product or any information contained within it for the benefit of third parties, other than allowed by these Conditions of Sale, without the prior written consent of the Commonwealth.

Where such consent is sought the Commonwealth reserves the right to set an appropriate charge or to require a revenue sharing arrangement.

The Client is permitted to quote selected statistical data contained in the products, providing that:
- the ABS is cited as the source of the data used;
- the terminology used is that used by the ABS for describing data; and
- any analysis or transformation of the data is not attributed to the ABS.

In respect of any data in computer readable form, or software, the Commonwealth authorises the Client to use the data or software on a non-transferable and non-exclusive basis. Copying of data or software for purposes other than back-up is prohibited.

WARRANTY

Except where required by law, no conditions or warranties, either express or implied, are given or offered for the products sold pursuant to the Client's order nor does the Commonwealth warrant the suitability of these products for any purpose. The Commonwealth shall not be liable for any damage, injury or loss arising or resulting, directly or indirectly from the Client's use of any of these products.

Neither the Commonwealth or the ABS nor their servants or agents shall be liable to the Client for any loss or damage arising or resulting directly or indirectly from any statement, information or advice made or given, whether negligently or otherwise, in relation to any product, its compilation or production, or the use to which it is intended to be put and without limiting the generality of the foregoing, any liability for faulty material or inaccuracies in the basic data or the physical condition of the product is expressly negated.

Where the data is is provided by the ABS on magnetic media, due care will be taken in preparation and handling. However, deterioration may occur between the time of copy and the time of receipt by the Client. Accordingly, if the data or software is unreadable on receipt and this is reported to the ABS within 28 days, it will be recopied free of charge.

LIMITATION OF THE LIABILITY OF THE COMMONWEALTH

The liability of the Commonwealth in respect of any damage, injury or loss suffered by the use of any of the products shall be limited to: The replacement of the product or the supply of an equivalent product; or An amount not exceeding $100 for any claim arising out of or in connection with the relationship established by this Agreement.

APPLICABLE LAW

This Agreement shall be construed in accordance with the law of the Australian Capital Territory and the parties submit to the jurisdiction of the courts of that Territory.

FURTHER INFORMATION

If the Client wishes to deviate from these Conditions of Sale in any way, or has any further questions relating to these Conditions, they should contact:

The Director
Business Development Unit
Australian Bureau of Statistics
PO Box 10
BELCONNEN ACT 2616
Telephone: (06) 252 7422   Fax: (06) 253 1404

The Client is reminded that additional Conditions of Sale may apply to certain ABS products.

2291200008912