

Information Paper



Australian
Bureau of
Statistics

NEW ISSUE

**SOCIO-ECONOMIC INDEXES FOR
AREAS**

Catalogue No. 1356.0

EMBARGOED UNTIL 11.30 A.M. 22 OCTOBER 1990

SOCIO-ECONOMIC INDEXES FOR AREAS

W. McLENNAN
Acting Australian Statistician

AUSTRALIAN BUREAU OF STATISTICS

CATALOGUE NO. 1356.0

© Commonwealth of Australia 1990

Printed by R. D. RUBIE, Commonwealth Government Printer, Canberra

PHONE INQUIRIES

MAIL INQUIRIES

ELECTRONIC
SERVICES

- *about these statistics and the availability of related unpublished statistics* — contact Edward Szoldra on Canberra (06) 252 7325 or any ABS State Office.
- *about other statistics and ABS services* — contact Information Services on Canberra (06) 252 6627, 252 5402, 252 6007 or any ABS State Office.
- *write to Information Services*, ABS, PO Box 10, Belconnen, ACT 2616 or any ABS State Office.
- on DISCOVERY — key *656#.
- on TELESTATS — phone (06) 252 5404 Foreign Trade Statistics, (06) 252 5405 Main Economic Indicator inquiries
- on AUSSTATS — phone (06) 252 6017.

CONTENTS

Section	Page
INTRODUCTION	1
THE INDEXES	1
2.1 Description of the Indexes	1
2.2 Available Geographic Areas	1
2.3 Summary of the Index Values	2
2.4 Applications of the Indexes	2
DERIVATION OF THE INDEXES	5
3.1 Background	5
3.2 Methodology - Principal Components Analysis	5
3.3 Methodology - Choosing the variables	5
3.4 Selecting Variables for the Indexes of Relative Advantage and Disadvantage	6
3.5 Variables used for the Index of Economic Resources	6
3.6 Variables used for the Index of Education and Occupation	7
3.7 Producing the Indexes from the Raw Scores	7
3.8 Validation	7
COMMENTS AND CAVEATS ON THE INTERPRETATION OF THE INDEXES	8
4.1 The Choice of Variables for each Index	8
4.2 Socio-Economic Dimensions not Represented in the Indexes	8
4.3 Understanding the Indexes	8
4.4 Limitations of Census Data	8
DATA AVAILABILITY	9
APPENDIX	10-11
ORDER FORM AND CONDITIONS OF SALE	13-14

INTRODUCTION

The Population Census provides information, on a five yearly basis, on a broad range of social and economic aspects of the Australian population. Over thirty questions of social and economic interest are asked in the Census. Those 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 this publication presents five summary measures, or indexes, derived from the 1986 Population Census to measure different aspects of socio-economic conditions by geographic areas.

This information paper describes the indexes and illustrates their possible uses. The technical background on 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 relating 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 Disadvantage are general socio-economic indexes. They summarise variables related to the economic resources of households, education, occupation, family structure and ethnicity. 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, relatively lower educational attainment and high unemployment. An area would have a large index value on an Index of Relative Socio-Economic Advantage if it contained a large proportion of relatively well-off households. Similarly, an area would have a low value on the Index of Relative Socio-Economic Disadvantage if it contained a large proportion of comparatively disadvantaged households.

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 Dis-

advantage. 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 Relative Socio-Economic Disadvantage (because of the large proportion of low income households) and a high score on the Higher Indexes (because of the high income households).

The Indexes of Economic Resources and of Education and Occupation are more specific. The first of these concentrates on a number of aspects of the economic resources of a household, including income, housing and car ownership. An area will score highly on the Index of Economic Resources if it contains a large proportion of households with high incomes, large dwellings, and several cars. The Index of Education and Occupation summarises the education and occupation aspects of socio-economic status. An area will have a high score on the Index of Education and Occupation if it contains a relatively high proportion of people with higher education qualifications, or employed in skilled occupations.

The method for deriving the indexes is described in Chapter 3, 'Derivation of the Indexes'. Factors to be taken into account when interpreting the indexes are discussed in Chapter 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. 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. Based on the scores for CDs, scores have also been calculated for legal local government areas (LGAs), statistical local areas (SLAs) and postcodes. The index scores for these aggregated areas were formed by taking the weighted average, using 1986 Population Census counts, across all CDs in the larger geographic area.

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).

Index values for regions other than postcode, LGA or SLA, may also be derived. To calculate an index score at a regional level, it is necessary to know the CDs which make up the region. Each CD score is then 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

THE INDEXES — *continued*

are subject to some imprecision for reasons outlined in section 4.4.

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 very high or low index values is contained in section 4.3.

2.3 *Summary of the Index Values and Variables*

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 percent of units have values within two standard deviations of the average, i.e. 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 Table 1).

These tables give the average index values for the CDs in each State as well as a range of quantiles. The 10 percent quantile gives the index cutoff below which 10 percent of the index values for that State lie. For example, in NSW the 10 percent quantile for the Index of Relative Socio-Economic Disadvantage is 871. 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 871.

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 1a which summarises the 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.

2.4 *Applications of the Indexes*

There are a number of ways the indexes can be used. Some important uses are described below.

Designing a Survey

For sample surveys, it is important that the sample that is selected is representative of the population that is of interest. Usually this involves making sure that the sample is spread across the whole population. The indexes can be used to achieve this, by choosing a sample of collector's districts with a range of different index values. Several individuals or households would then be chosen from each selected CD. Two common techniques for ensuring that a sample of CDs covers a range of index values are known as 'stratification' and 'systematic sampling'.

In some cases, the population of interest is only part of the whole population. In this case it makes sense to limit the scope of the survey to that subpopulation. If the subpopulation can be identified by socio-economic characteristics (for example, low educational attainment and unskilled occupations) then the Indexes can be used to target the areas to be surveyed.

Example: Spreading the Sample - Stratification

- Stratification involves grouping the CDs into categories. These categories are called strata. For example, the index range may be divided into five categories, each containing 20 percent of the population. A separate sample of CDs could then be selected from each of the five categories. It is possible to select larger samples in the survey from the categories that are of greater interest and then adjust the sample weights appropriately. If mapping software is available (for example, the CDATA 86 package), a map of the area can be produced using different colours (or shadings) for each range. Such a map shows the concentration of the five index ranges across the city and highlights the areas which are in each category.

Example: Spreading the Sample - Systematic Selection

- Systematic selection involves sorting the CDs in the region of interest by their index values. Then every tenth CD (for example) in the sorted list could be selected in the sample, thereby ensuring a socio-economic spread of the sample.

Example: Targeting Survey Areas - Limiting Scope

- In some cases, it may be efficient to limit the scope of a survey to a particular socio-economic group. For example, an agency may wish to determine client's attitudes to a particular issue. Suppose the people affected most by the issue are those with few economic resources. It would then make sense to concentrate the survey on such people. The index of economic resources can be used to do this. Within a city each CD has an index value, and can be categorised according to those values. For example, the index range may be divided into four categories each containing 25 percent of the population. The survey could be restricted to those CDs in the bottom range. Of course, inferences from the survey would then be restricted to this group.

There are many decisions to make in designing a survey. Issues such as: the total sample size needed; whether to use stratification or systematic selection; how many strata to use and how many CDs to take from each stratum; how many households to sample per CD, and how to select those households; how to weight the sampled households correctly; how to design a questionnaire and so forth must all be considered. The ABS's Statistical Consultancy service, or other competent statistical consultants can provide further advice on survey design.

Checking on the Socio-Economic Balance of a Sample

In many surveys, non-response and non-contact are important issues. The indexes can be used to check the 'socio-economic balance' of achieved samples to see whether an appropriate range of areas has been covered by the sample.

TABLE 1 SUMMARY DATA FOR CD LEVEL INDEXES

TABLE 4. SUMMARY OF RESULTS FOR THE 1990s						
Quantile						
State	Average	10%	25%	50%	75%	90%
INDEX OF RELATIVE SOCIO-ECONOMIC DISADVANTAGE						
NSW	994	871	938	1002	1061	1115
VIC	1029	924	982	1041	1089	1300
QLD	979	868	933	987	1036	1125
SA	996	867	947	1011	1062	1082
WA	993	868	954	1014	1065	1107
TAS	992	883	945	1003	1053	1099
NT	879	541	756	953	1027	1057
ACT	1085	978	1062	1104	1135	1164
AUST	1000	880	950	1011	1067	1113
RURAL INDEX OF RELATIVE SOCIO-ECONOMIC ADVANTAGE						
NSW	1000	868	921	1012	1076	1113
VIC	999	864	930	1011	1075	1113
QLD	995	849	911	1010	1078	1117
SA	992	860	906	1005	1070	1113
WA	978	819	871	1000	1073	1113
TAS	1049	947	998	1053	1099	1138
NT	1096	1032	1061	1099	1138	1165
ACT	1066	972	1046	1065	1103	1105
AUST	1000	858	921	1017	1078	1117
URBAN INDEX OF RELATIVE SOCIO-ECONOMIC ADVANTAGE						
NSW	998	893	929	975	1041	1140
VIC	1006	903	938	990	1058	1138
QLD	985	905	935	971	1019	1086
SA	981	882	921	969	1027	1097
WA	1019	908	949	997	1076	1168
TAS	977	890	922	965	1025	1085
NT	993	894	945	996	1039	1080
ACT	1133	996	1076	1127	1187	1281
AUST	1000	898	934	980	1047	1136
INDEX OF ECONOMIC RESOURCES						
NSW	1002	887	935	991	1059	1140
VIC	1016	909	955	1011	1078	1133
QLD	987	883	929	983	1044	1103
SA	977	859	918	973	1040	1102
WA	1000	880	938	999	1073	1132
TAS	970	875	918	968	1023	1074
NT	921	686	787	952	1050	1091
ACT	1108	934	1049	1121	1184	1244
AUST	1000	887	936	994	1062	1127
INDEX OF EDUCATION AND OCCUPATION						
NSW	1009	890	938	998	1076	1150
VIC	1021	912	954	1014	1086	1146
QLD	959	850	898	951	1014	1076
SA	986	871	919	977	1053	1122
WA	990	881	926	981	1052	1122
TAS	971	849	900	960	1043	1112
NT	1010	861	958	1040	1076	1115
ACT	1160	1087	1127	1161	1200	1239
AUST	1000	882	930	990	1067	1138

THE INDEXES — continued

Example: Checking a Non-Probability Sample

In quota surveys or those where respondents select themselves, as in magazine surveys, it may be desirable to check that the respondents come from an appropriate range of socio-economic areas. Such a check will add confidence to the applicability of the results of the survey. For example, a magazine survey may ask all respondents to state their postcode along with answering other questions of interest. Each postcode can then be categorised according to the index of interest. The distribution of respondents across these categories can be compared with the distribution of the total population of interest, to assess whether survey respondents were representative. For example, if an organisation conducting a magazine survey wanted an indication of whether it had representative numbers of respondents with different educational levels and occupations, the following table showing sample spread could be derived to use the Index of Education and Occupation to measure balance.

TABLE 2

Range of Index Values for Postcodes	Percentage of Population 15+ (%)	Percentage of Respondents (%)
INDEX OF EDUCATION AND OCCUPATION		
920 or less	9	5
921 - 1000	39	30
1001 - 1150	44	50
1151 or more	8	15

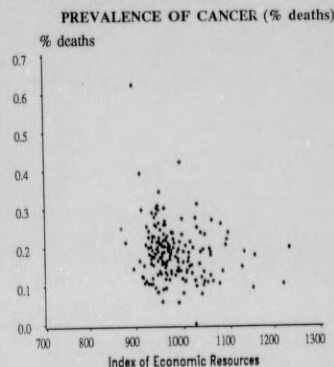
In this example, the percentage of respondents with index values less than 1000 is 35 percent, compared to 48 percent present in the population. Observing the above table would suggest that the survey had an under-representation of people from postcodes with low values for the Index of Education and Occupation.

Uses in 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. In such cases, one or more of the indexes can be used as a summary of a range of socio-economic factors. For example, susceptibility to a particular disease (as measured by prevalence, say) may depend both on the age of the person and on their socio-economic environment. This environment could be measured by using a summary measure such as the Index of Relative Socio-Economic Disadvantage.

Example:

- Initial investigations into the effect of economic resources could use simple tables or graphs which show values of the variable of interest by index category range. For example, a plot of the form illustrated below could be used.

*Targeting Areas for Services or Business*

The indexes are of interest in their own right as summaries of area characteristics. They could be used to assist in determining the allocation of services or in making business decisions. Those intending to use the indexes in this way, are strongly advised to be aware of the limitations of the indexes, described in Chapter 4, 'Comments and Caveats on the Interpretation of the Indexes'.

Example:

- The indexes may be used simply 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 making business decisions. For example, using the Index of Relative Socio-Economic Disadvantage, social welfare agencies might check the availability of services to disadvantaged areas. Similarly, businesses might site outlets, or target promotion campaigns, by using information from the geographic indexes.

An example of such a use is given by the promotional needs of a business launching a new product. The product may be a luxury item. Such a product might be expected to sell larger volumes per capita in relatively advantaged areas. The Indexes of Relative Socio-Economic Advantage could be sorted to list the top ten percent of areas within each State. This will provide a list of both Urban and Rural locations which will provide maximum returns for the promotional outlay.

3. DERIVATION OF THE INDEXES

3.1 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 each of the following two Censuses, with slight changes of method being applied. The indexes produced from these Censuses were derived specifically to reflect educational disadvantage and therefore were based on Census variables which had been found to be correlated with educational achievement. However, the correlations between the variables included and the indexes, indicated that although they were derived to measure educational disadvantage, they could also be useful as a general socio-economic index.

Since 1984, an index of educational disadvantage based on the 1981 Census has been available from the ABS Statistical Consultancy service on request. Because this index has proved to be useful, socio-economic indexes have been derived from the 1986 Census using the same method as that used in 1981, with minor changes. Since in many circumstances, a single socio-economic index may be too broad to be useful in a particular survey design or analysis, two general indexes (an Index of Relative Advantage and an Index of Relative Disadvantage) and two more specific indexes have been derived. 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 are an index of economic resources and an index of education and occupation. A description of the derivation of the indexes follows.

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 variable, or index, which reflects the population profile of these variables is required to aid social and economic investigations. Multivariate statistical techniques can be used to summarise the information contained in several variables into one variable. Principal Component Analysis (PCA) is one such 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.

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 all five indexes, the first principal component has been used as a socio-economic measure. The first component was chosen 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 comprise 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.

While Principal Components Analysis 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 that resulted were checked to ensure that they were indeed reflecting the desired socio-economic aspects of each area. Finally, after this validation stage, the indexes were scaled to have an average value of 1000. Each of these stages is discussed in more detail in the following sections of this chapter.

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 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 1986 Census. The variables were chosen subjectively, based on experience with the earlier indexes. Comment on these lists of variables was 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.

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 high on this aspect. An extreme example would be the inclusion of the same variable twice in the

DERIVATION OF THE INDEXES — *continued*

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.

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 urban areas it typically comprises 200-300 dwellings, while in rural areas it usually contains fewer. In 1986 there were about 29,500 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 in some predetermined way. Thus the CD was chosen as the appropriate level for analysis.

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.

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

When deriving the indexes from previous Censuses, PCA was applied to a selection of variables which were considered to be related in some way to educational disadvantage. Variables pertaining to family income, educational attainment, unemployment, occupation, marital status, household occupancy, residential mobility, Aboriginality and migrant status were included in the analysis. The specific variables chosen for the 1981 index were selected either because they were socio-economic variables which had been shown by educational researchers to be correlated with educational attainment, or because they displayed face validity (i.e. they seemed to conform to what intuition might dictate). However, the variables included were usually fairly general and the resulting index seemed to summarize general socio-economic conditions rather than solely educational disadvantage. Variables similar to those used for the 1981

index were thus selected 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 area could contain both pensioner households and double income households with no children. Thus it might have high proportions of both relatively advantaged and disadvantaged households. Consequently, it was 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.

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

At this stage, it became apparent that it was not appropriate to produce an Index of Relative Advantage that covered both urban and rural areas. Hence the Index of Relative Advantage was split into two: an urban index and a rural index. These urban and rural indexes were derived by returning to the originally identified set of variables for the Index of Relative Advantage and then carrying out separate principal components analyses for CDs from urban areas, and for CDs from rural areas. CDs were defined as urban or rural using the 1986 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. Urban centres include known holiday resorts with smaller population provided that they contain 250 or more dwellings, of which at least 100 were occupied on Census night 1986. The Rural Index of Relative Socio-Economic Advantage was based on all remaining CDs.

A higher score on one of the Indexes of Relative Advantage means that a CD has attributes such as a large proportion of households with high incomes or a trained workforce. A low score on the Index of Relative Disadvantage suggests that the CD has characteristics such as many households of low income and many people with little training, in corresponding occupations.

3.5 Variables used for the Index of Economic Resources

The purpose of the Index of Economic Resources is to reflect the profile of the economic resources of families within CDs. Therefore Census variables were included in the analysis which reflected the income and expenditure

DERIVATION OF THE INDEXES — *continued*

of families, 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.

This index does not include any measure of inherited wealth, and also misses some assets, such as savings. While it would have been desirable to include such variables, it was not possible because the information was not collected in the Population Census. The index also excludes education and occupation variables, because they are not directly related to economic resources.

A high score on this index indicates that the area has a high proportion of families on high incomes, and/or with mortgages and/or with large houses. A low score indicates that the area has relatively large proportions of households on small incomes living in small dwellings.

3.6 Variables used for the Index of Education and Occupation

In many studies it is desirable to have a variable which indicates the educational and occupational structure of communities. The index of education and occupation has been developed for this purpose. 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 occupations requiring training, rather than being labourers or unemployed. A low score indicates an area with concentrations of either lowly educated people or unemployed people.

3.7 Producing the Indexes from the Raw Scores

To allow for easy recognition of high and low scores, the index scores have been standardized to have a mean of 1000 and a standard deviation of 100. The scores themselves are not a measure of the relative socio-economic conditions of the various geographic areas (CDs, LGAs, 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 status on a relative scale.

Scores for areas larger than CDs can be calculated by weighting together constituent CD scores. Scores have been produced for 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.

3.8 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 an important part of the derivation of the indexes.

One important check on the indexes is whether the variables and their weights make sense. This process began with comments on the variables to be included in the initial analysis being sought from, and provided by, several academics and researchers. Following this, 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 more than 20 percent 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. The gradings assigned subjectively corresponded closely with the rankings provided by Index values.

The top and bottom ranked CDs in each State were also examined. 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.

Finally, additional data sources were used to validate the index scores. Data from the 1988 ABS Household Expenditure Survey (HES) was used to compare (HES) income and mortgage expenditure information, with index scores from the index of economic resource. Comparative

DERIVATION OF THE INDEXES — *continued*

analysis of CD rankings within specific indexes, indicated no significant differences in the distribution of the ranks.

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 Chapter 3. 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*

All users of the Indexes should critically examine the construction of the indexes (see Appendix A) to ascertain whether they are appropriate to their problem or analysis. However there are two 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. More detail is given in Section 3.6.

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.

4.3 *Understanding the Indexes*

The indexes produced by Principal Component Analysis depend solely on the correlation matrix, which is a measure of 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 indexes' 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 percent of index values has that rank. 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 *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. The indexes cannot therefore purport to summarize this facet of socio-economic well being.

COMMENTS AND CAVEATS ON THE INTERPRETATION OF THE INDEXES — *continued*

(b) Missing data is a further impediment to index construction. Although non-response to individual Census items is overall quite low (well below 10 percent across Australia for most variables), 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, refugees) are therefore "under-represented".

(d) The Census does not aggregate data to CDs according to their usual residence but according to where they were spending the night at the time of the Census. Clearly, the indicator should ideally use usual residence data. While this affected all States, Western Australia, Queensland and the Northern Territory appear to have been more adversely affected by persons not being at home on Census night in 1986.

5. DATA AVAILABILITY

Index values for the five indexes are available at the CD, LGA, SLA and postcode level for Australia. The Australia wide indexes are available on floppy disc. The indexes are accompanied by a software package which provides a fully-documented user-friendly interface to the indexes. A particularly useful feature of the software is the ability to create user-defined areas in the event that the standard areas are not adequate for the user. An order form for the index products is given at the back of this publication.

Alternatively, special index data sets for particular areas can be obtained on floppy disc or as hard copy, from the ABS Information Consultancy or Statistical Consultancy Services in your State capital city office. 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 Andrew's House
Cnr Kent and Bathurst Streets
SYDNEY NSW 2000
(Box 796, GPO Sydney 2001)
Phone (02) 268 4634
Fax (02) 268 4668

VICTORIA

5th Floor, Rialto North Tower
525 Collins Street
MELBOURNE VIC 3000
(Box 2796Y, GPO Melbourne 3001)
Phone (03) 615 7864
Fax (03) 615 7798

QUEENSLAND

20th Floor, 313 Adelaide Street
BRISBANE QLD 4000
(Box 9817, GPO Brisbane 4001)
Phone (07) 222 6298
Fax (07) 229 6042

WESTERN AUSTRALIA

1st Floor, Hyatt Centre
30 Terrace Road
PERTH WA 6000
(Box K881, GPO Perth 6001)
Phone (09) 323 5385
Fax (09) 325 7376

SOUTH AUSTRALIA

Ground Floor, 41 Currie Street
ADELAIDE SA 5000
(Box 2272, GPO Adelaide 5001)
Phone (08) 237 7577
Fax (08) 237 7566

TASMANIA

1st Floor, 175 Collins Street
HOBART TAS 7000
(Box 66A, GPO Hobart 7001)
Phone (002) 20 5800
Fax (002) 34 6237

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

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

APPENDIX A DEVELOPMENT OF SOCIO-ECONOMIC INDEXES

INITIAL LIST OF VARIABLES

Urban and Rural Indexes of Relative Advantage

Family Income greater than \$50000 (%)

Households purchasing dwelling (%)

Households with 3 or more motor cars (%)

Households with 4 or more bedrooms (%)

Average number of bedrooms per person

With degree or higher (%)

With Trade or OTHER qualification (%)

15+ at school (%)

15+ at TAFE (%)

15+ at CAE or University (%)

Employed Males (ILF) classified as Managers or Administrators (%)

Employed Females (ILF) classified as Managers or Administrators (%)

Employed Persons (ILF) classified as Managers or Administrators (%)

Employed Males (ILF) classified as Professionals (%)

Employed Females (ILF) classified as Professionals (%)

Employed Persons (ILF) classified as Professionals (%)

Employed Males (ILF) classified as para-Professionals (%)

Employed Females (ILF) classified as para-Professionals (%)

Employed Males (ILF) employed as Clerks (%)

Employed Females (ILF) employed as Clerks (%)

Employed Males (ILF) classified as Sales (%)

Employed Females (ILF) classified as Sales or Personal (%)

Index of Relative Disadvantage

Family Income less than \$12000 (%)

Households with no motor cars (%)

Households with 1 or no bedrooms (%)

Households renting (non-government) (%)

Households renting (government authority) (%)

Households in improvised dwellings (%)

No qualifications (%)

Left school less than 15 years of age (%)

Never at school (%)

Employed Males (ILF) classified in Trades (%)

Employed Females (ILF) classified in Trades (%)

Employed Females (ILF) classified in Sales/Personal (%)

Employed Persons (ILF) classified as Labourer or related (%)

Males (ILF) unemployed (%)

Females (ILF) unemployed (%)

Families consisting of Head and dependents (%)

15+ separated or divorced (%)

Occupied private dwellings with 2 or more families (%)

Aboriginal or Torres Strait Islanders (%)

Recent migrant from non-English speaking country (%)

Lacking English fluency (%)

INITIAL LIST OF VARIABLES — *continued*

Index of Economic Resources

- Households purchasing dwelling (%)
- Households with no motor cars (%)
- Households with 3 or more motor cars (%)
- Households with 1 or no bedrooms (%)
- Households with 4 or more bedrooms (%)
- Av. number of bedrooms per person
- Households renting (non-government) (%)
- Households renting (government authority) (%)
- Households in improvised dwelling (%)
- Families consisting of single parent with dependents, with income less than \$15000 (%)
- Families consisting of single parent with dependents, with income greater than \$40000 (%)
- Families consisting of couple with dependents, and income less than \$15000 (%)
- Families consisting of couple with dependents, and income greater than \$40000 (%)
- Families with other family structure, and income less than \$15000 (%)
- Families with other family structure, and income greater than \$40000 (%)
- Mortgage greater than \$600 per month (%)
- Rent less than \$49 per week (%)
- Rent greater than \$50 per week (%)
- Group households (%)

Index of Education and Occupation

- With degree or higher (%)
- With trade or other qualifications (%)
- With no qualifications (%)
- Who left school 15 years of age (%)
- Never at school (%)
- 15+ at school (%)
- 15+ at TAFE (%)
- 15+ at CAE or University (%)
- Employed Males (ILF) classified as Managers or Administrators (%)
- Employed Females (ILF) classified as Managers or Administrators (%)
- Employed Persons in the Labour Force (ILF) classified as Managers or Administrators (%)
- Employed Males (ILF) classified as Professionals (%)
- Employed Females (ILF) classified as Professionals (%)
- Employed Persons (ILF) classified as Professionals (%)
- Employed Males (ILF) classified as para-Professionals (%)
- Employed Females (ILF) classified as para-Professionals (%)
- Employed Males (ILF) classified in Trades (%)
- Employed Females (ILF) classified in Trades (%)
- Employed Males (ILF) classified as Clerks (%)
- Employed Females (ILF) classified as Clerks (%)
- Employed Males (ILF) classified in Sales/Personal (%)
- Employed Females (ILF) classified in Sales/Personal (%)
- Persons (ILF) classified as Labourer or related (%)
- Males (ILF) classified as unemployed (%)
- Females (ILF) classified as unemployed (%)

ORDER FORM FOR SOCIO-ECONOMIC INDEXES

CATALOGUE NUMBERS 1355.0—1355.8

Send completed form to 'The Collector of Public Moneys' in your State office of the ABS, at the addresses specified in the 'Data Availability' section of this document.

Socio-Economic Indexes for Areas

- including index scores for the geographic areas of CD, LGA, SLA and Postcode
- including the Information Paper and the User's Manual
- indexes for States attract a base charge of \$1000 plus a component for the volume of data (i.e. no. of CDs).

I wish to purchase the following:

		State if 5.25" or 3.5" disks required	No. copies	Cost (\$)
	Base Charge			\$1000
plus	AUSTRALIA at	_____	_____	_____
or individual States at				
	New South Wales	_____	_____	_____
	Victoria	_____	_____	_____
	Queensland	_____	_____	_____
	South Australia	_____	_____	_____
	Western Australia	_____	_____	_____
	Tasmania	_____	_____	_____
	Northern Territory	_____	_____	_____
	Australian Capital Territory	_____	_____	_____
	Information Paper at \$15 per copy (additional copies)		_____	_____
	User's Manual at \$40 per copy (additional copies)		_____	_____
TOTAL COST				\$ _____

Consultation Services

- details of our services are available by contacting the Information Services area of your State office of the ABS at the numbers given in the 'Data Availability' section of this document.

Secondary Distributors

A secondary distributor is any person or organization who obtains data from the ABS and distributes to a third party where the major part of the product consists of ABS copyright data. You are not a secondary distributor if you are preparing a consultancy or research report that uses ABS data or where analysis is the major part of the report. If you are a secondary distributor, please contact the ABS to negotiate a licence fee or royalty payment for the commercial use of the data. Refer to part 2(e) of Conditions of Sale over the page for the appropriate contact.

Please send the above products to :

Name: _____
 Position: _____
 Company: _____
 Address: _____
 Postcode: _____
 Telephone No : _____
 Facsimile No : _____

I wish to pay by:

Cheque, Cheques or money orders should be made payable to 'The Collector of Public Moneys'

Card No :
 Bankcard
 Mastercard Expiry Date : _____
 Visa Name on Card : _____
 Signature : _____

I have read and agree to the conditions of sale printed on the reverse side of this form.

Signed: _____

Conditions of Sale of ABS Proprietary Products

1. ABS Proprietary Products

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

2. Copyright of ABS Proprietary Products

(a) The Customer acknowledges that ABS proprietary products are special, valuable and unique products in which the Commonwealth holds the copyright. The Customer agrees not to disclose to any third party any information, including specifications, drawings, designs, know-how, or details of software (whether of a commercial or a technical nature) in the products, without the prior written consent of the Commonwealth or payment of a secondary provider licence fee.

(b) In respect of any data in computer readable form or software comprising the products the Commonwealth authorises the Customer to use the said data or software on a non-transferable and non-exclusive basis and to copy any such data or software only for back-up purposes.

(c) Nothing in these conditions should prevent the Customer from quoting statistical data contained in the product providing

(i) that the ABS is acknowledged as the source of the data;

(ii) any analysis or transformation of the data is not attributed to the ABS;

(iii) the terminology used is that used by the ABS for describing data; and

(iv) the reproduced material is provided free-of-charge to the end-user and there is no likely impact on ABS publication sales. Where permission is sought to utilise Commonwealth copyright material to generate a financial return, the Commonwealth reserves the right to set an appropriate charge or to require an equitable revenue sharing arrangement.

(d) Copying of data or software for purposes other than backup is prohibited. Discounts for multiple copies and site licences apply.

(e) If the Customer wishes to deviate from these conditions in any way, the Customer must first contact the Director, Publishing and Electronic Dissemination, Australian Bureau of Statistics, Cameron Offices, Chandler Street, Belconnen, ACT 2616 (Phone (06) 252 6101).

3. Warranty

Except as provided in Clause 4 -

(a) No conditions or warranties, either express or implied, are given or offered for the proprietary products sold pursuant to the Customer'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 Customer's use of any of these products.

(b) Neither the Commonwealth or the ABS nor their servants or agents shall be liable to the Customer 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 of the products, 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.

(c) Where the data is provided by the ABS on magnetic tape or disk, due care will be taken in preparing and handling each tape or disk. However, deterioration may occur between the time of copy and the Customer's receipt of the tape or disk. Accordingly, if the tape or disk is unreadable on receipt and this is reported to the ABS within 28 days of despatch, it will be recopied free of charge. As an added precaution you are advised to take at least one security copy of the tape or disk on receipt.

4. Limitation of Liability

The liability of the Commonwealth in respect of any damage, injury or loss suffered by the use of any proprietary products shall be limited to -

(a) the replacement of the product or the supply of an equivalent product; or

(b) an amount not exceeding \$100 for any claim arising out of or in connection with the relationship established by this Agreement.

5. 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 the Territory.

Printed by R.D. RUBIE, Commonwealth Government Printer, Canberra
© Commonwealth of Australia 1990

Recommended retail price: \$15.00



2013560012862

ISBN 0 642 15593 3