THE AUSTRALIAN NATURAL DISASTER RESILIENCE INDEX

VOLUME II – INDEX DESIGN AND COMPUTATION

Chapter 3 – Computation of the Australian Natural Disaster Resilience Index

Version	Release history	Date
1.0	Report submitted to BNHCRC	30/08/2019
1.1	Initial release of document	29/07/2020

Australian Government
Department of Industry, Science, Energy and Resources

Business
Cooperative Research
Centres Program

Disclaimer:

CHAPTER 3 – COMPUTATION OF THE AUSTRALIAN NATURAL DISASTER RESILIENCE INDEX

In this chapter

Section 3.1   Reviews the development and use of composite indexes and methods for computing composite indexes.

Section 3.2   Describes the rationale for, and the statistical computation of, the Australian Natural Disaster Resilience Index.

Section 3.3   Describes the methods used to compute the typology of groups of SA2s with similar disaster resilience profiles.

TABLE OF CONTENTS

FIGURES

TABLES

3.1 REVIEW OF COMPOSITE INDEX USE AND METHODS

3.1.1 History of composite indices

Santos et al. (2014) locate the origins of the concept of a single number representing the state of a complex system to the period after World War 2, when national accounting was developed to enable evaluation of the performance of the economy. It was believed at this time that social well-being would be improved by economic growth, and this could be measured by GNP per capita. By the 1970s, and with persistent poverty despite the period of postwar prosperity and economic growth, there was growing acceptance that additional indicators of unemployment, poverty and inequality would be needed to characterise levels of economic development and social well-being.

From this time, some researchers and policy makers began to propose ways of combining these multiple indicators into a single index. Indices included: an index of socio-economic development proposed by the United Nations Research Institute for Social Development (McGranahan et al.1972), a Physical Quality of Life Index proposed by Morris (1978), the Index of Social Progress proposed by Estes (1984) and the Human Suffering Index proposed by Camp and Speidel (1987). While these indices received some scrutiny, it was not until the publication of the Human Development Index (HDI) by the United Nations in 1990 that the methods of composite index construction began to be studied in detail. It appears that the HDI's role in significant funding allocation decisions, such as in United Nations aid and economic development programs, resulted in the methods of computation of the index coming under intense scrutiny, revealing many methodological flaws (see Periera and Mota 2016). By the end of the second decade after the publication of the HDI, the methodological issues inherent in the construction of composite indicators were well understood (OECD 2008; Kovacevic 2010). These issues included the selection of indicators, normalisation or standardisation of indicators and weighting and aggregation of indicators to form a composite index. A number of changes to the HDI methodology were made in response to the issues raised, most notably a change from arithmetic to geometric aggregation of indicators in 2010.

Studies of the vulnerability of communities to climate change emerged in the mid-1990s and, within a decade, indicators of vulnerability and adaptive capacity were being proposed as a means of informing funding allocation decisions, such as under the United Nations Framework Convention on Climate Change (Agder et al. 2004). In the following decade, reviews and critiques of the methods of calculating climate change vulnerability indices were published that ranged from scathing (Füssel 2009) to mildly cautionary (Baptista 2014). The former author referred to mathematically flawed procedures in the calculation of composite indices and argued that composite indicators should

not be used as a basis for allocating program funding. Barnett et al. (2008) drew similar conclusions about the Environmental Vulnerability Index. Rowley and Peters (2009) showed that the weighting and aggregation issues in composite index construction also applied in Life Cycle Assessment.

From 2000 onwards, increasing numbers of proposals for composite indices as measures of social vulnerability to natural hazards were proposed (see, for example, Pelling and Uitto 2001; Cutter et al. 2000; Cutter et al. 2003 and Chakraborty et al. 2005). During this period composite indices of resilience to natural hazards or disasters were also proposed, so that by 2016, Beccari (2016) was able to locate 106 distinct frameworks for calculating composite indices of generic vulnerability or resilience to natural hazards, published between January 1990 and March 2015. New composite indices also continued to be proposed in the human development field during this period, with Yang (2014) cataloguing some 101 indices that have been proposed since the HDI in 1990.

Beccari (2016) found that the great majority of resilience and vulnerability frameworks (some 87 out of 106) were hierarchical, with simple arithmetic or geometric averaging of indicators to form a composite index. Indexes using Principal Components Analysis (PCA) are included in the 87, since the scores on principal components are also constructed additively. The body of work on simple composite indices for vulnerability or resilience to natural hazards has largely ignored the methodological issues raised over a period of several decades in the human development and climate change literature. Gall (2007), in her comparative evaluation of indices of social vulnerability to natural hazards, concluded that compensability (one of the main methodological issues) was a hidden and underestimated problem. Only 19 per cent of the proposals for composite indices of disaster resilience or vulnerability catalogued by Beccari (2016) acknowledged the existence of these methodological issues, and only one proposal undertook a comprehensive sensitivity analysis.

3.1.2 Methodological issues

Methodological issues in the construction of composite indices have been extensively discussed and/or analysed in a number of different disciplines, including:

- environmental condition (Ebert and Welsch 2004; Rowley et al. 2012);
- sustainability (Hudrlikova and Kramulova 2103);
- life cycle analysis (Rowley and Peters 2009);
- climate change vulnerability and resilience (Füssel 2009; Baptista 2014);
- environmental vulnerability (Barnett et al. 2008);
- international development and/or progress comparisons (Cherchye et al. 2007; Munda and Nardo 2009; Natoli and Zuhair 2011; OECD 2008; Salzman 2003; Tarabusi and Guarini 2013; Zhou and Ang 2009);
- human development, quality of life (Kovacevic 2010; Mazziotta and Pareto 2013a; 2013b; 2016a; Pereira and Mota 2016);

- poverty, economic vulnerability, economic development (Guillaumont 2009; De Muro et al. 2011);
- social vulnerability (Rygel et al. 2006);
- health economics (Jacobs et al. 2004; Vidoli et al. 2015);
- financial analysis (Marozzi and Santamaria 2007);
- multi-criteria decision analysis (Munda 2012a; 2012b); and,
- GIS based multi-criteria decision analysis (Malczewski 2000; 2006a; 2006b; Heinrich et al. 2016).

A number of methodological studies of composite indices are relevant to two or more of the disciplines listed above (e.g. Gall 2007; Munda 2012a; Paruolo et al. 2013). There are methodological issues in the construction of composite indices associated with each of the steps in the construction process, with much of the methodological literature being concerned with weighting and aggregation.

3.1.2.1 Functional form, construct validity and content validity

Functional form refers to the nature of the relationship between an indicator and the conceptual entity it is believed to represent. For example, resilience within a geographical area might be affected by median weekly family income in that area. If the functional form is a simple linear relationship then an area with a median weekly family income of $1,000 is believed to have twice the resilience of an area where the median weekly family income is $500. Functional form is rarely discussed in composite index studies (but see Salzman 2003). In most composite indicator studies, a simple linear functional form is assumed, given the absence of either theory or empirical evidence to the contrary.

Where a composite index incorporates reflective measurement models (see Section 3.1.2.6), the concept of construct validity used in psychological and sociological research is transferable to composite indicators. An indicator that does not have construct validity will have no relationship with resilience. The concept of content validity is also transferable. A composite index with content validity will be comprised of indicators for all the factors that might affect resilience.

3.1.2.2 Populations, samples and outliers

Many composite index methods proposed in the literature, or in use, involve some discussion or treatment of outliers. Outliers are values taken by a variable that lie far from the majority of values for that variable. Formal definitions of outliers implicitly (e.g. Osborne and Overbay 2004) or explicitly (e.g. Marriott 1990) involve a sample drawn from a population. In this situation, the presence of outlying values beyond the tails of the sampling distribution signals the possibility of errors in the data collection process, from mis-specified sample frames, faulty drawing of samples, through to key stroke errors in data entry.

The source of error should be investigated and, if possible, corrections made to the data. Remaining outliers may still be a cause for concern. They may violate distributional assumptions upon which parametric inferential statistics depend. However, if removed, this may inflate Type 1 error rates (Bakker and Wicherts 2014). Transformations, such as taking logs of values, and substitution of parametric methods with robust methods, are the preferred options for dealing with retained outliers (Osborne 2002; Osborne and Overbay 2004).

In the composite index literature, there have been detailed discussions of outlier issues by Foa and Tanner (2012), Heinrich et al. (2016), Hudrlikova and Kramulova (2013), Hudrlikova (2013), Jacobs et al. (2004), Mishra (2008), OECD (2008), and Vidoli et al. (2015). Of frequent concern in much of this work is the impact of the treatment of outliers on composite indices. However, very little attention is paid to how outliers are to be understood in data that is used to construct composite indices. Crucial to this is the distinction between samples and the population from which samples are obtained. The treatment of outliers in composite indices based on sample data draws on inferential statistics, where outliers are regarded as values that lie outside the tails of the sampling distribution and therefore are suspected to be measurement or data processing errors. In this case, there are valid grounds for deleting the suspected erroneous data from the composite index construction process.

However, in many cases, such as indices comparing the performance of countries in some area of interest, or spatial indices based on Census data, there is no sample. The data from which the index is being constructed represents the total population. Further, where the data is from official sources such as national accounts or an official census of residents, there are many precautions and checks undertaken in the gathering of data, so that the probability of erroneous data is very low. In this situation, the application of outlier thresholds from inferential statistics makes little sense. Extreme values beyond these thresholds are real values and should be incorporated in some way in calculating composite indices. Rather than abandoning valid data with the misguided application of outlier procedures from inferential statistics, it is preferable to devise robust indices that are resistant to the effects of extreme values among the indicators that comprise the index.

This is particularly the case where the object is to arrive at the spatial distribution of a composite index. For example, there are 2104 Statistical Area Level 2 (SA2) areas in Australia for which the Australian Bureau of Statistics publishes values of the Socio-Economic Indexes for Areas (SEIFA) index, and these give an almost complete spatial coverage of the country (ABS 2011; 2013). If a data set for these 2104 SA2s is constructed using 33 common demographic indicators, and these are scanned for outliers using as a threshold absolute values of the z-scores greater than 3.29 (a common univariate outlier threshold; Tabachnick and Fidell 2007), then some 237 SA2s, or 11.3 per cent of the total number of SA2s in Australia would be discarded as outliers. This does not include

additional SA2s that might be identified as multivariate outliers. The example shows that the misguided application of outlier thresholds from inferential statistics could seriously impair the spatial coverage of a composite index calculated from these 33 demographic indicators.

The alternative is to use aggregation methods that are robust to malformed indicator distributions and extreme indicator values. A number of robust transformations and aggregation methods have been proposed, including:

- standardising indicators by subtracting the median and dividing by the median absolute deviation (Aeillo and Attanasio 2004);
- for composite indices based on Principal Components Analysis (PCA), using robust versions of PCA (Mishra 2008; Vidoli et al. 2015);
- summation of rankings (OECD 2008);
- number of indicators above and below a benchmark (OECD 2008);
- winsorisation and log transformations of indicators (Saisana and Phillips 2012); and,
- modifications to the Benefit of Doubt approach (Vidoli et al. 2015).

When the dataset for composite index calculation is a complete population, these robust methods offer the possibility of avoiding the unwarranted deletion of considerable numbers of records as "outliers".

A further example of the failure to appreciate the difference between analysis of samples and analysis of the complete population is to be found in a number of natural hazards composite index studies. In discussing the question of selection of an appropriate number of components in principal components analysis, Schmidtlein (2008), Tate (2012) and Baptista (2014) all refer to parallel analysis as an appropriate method. However, this is a Monte Carlo method to obtain the sampling distributions of the eigenvalues of a PCA, and to retain components with eigenvalues greater than some chosen inferential threshold. The approach is meaningless if applied to data that represents the whole population, being appropriate only where the object is to generalise from a sample of geographical units to the complete population as to the number of retained components.

3.1.2.3   Transformations of indicators – normalising

The indicators that comprise a composite index are frequently transformed for two reasons: to obtain an indicator distribution that meets the assumptions required by a statistical procedure such as PCA; and/or to give indicators equal influence in a simple additive composite index (the most common aggregation method for composite indices).

Before proceeding, it is necessary to clarify a number of terms that are used inconsistently in the indicator literature.

Normalise: this can be used to refer to transformations that bring a non-normal distribution closer to a normal distribution (e.g. von Hippel 2003), or it can refer to rescaling a variable such that it has a range of 0 – 1 (OECD 2008).

Standardise: this can refer to converting the values of a variable to z-scores (OECD 2008; Schmidtlein et al. 2008), or to rescaling to a range of 0 – 1 (Gall 2007).

There appears to be a belief among some authors (e.g. Jacobs et al. 2004; Hudrlikova and Kramulova 2013), that converting the values of a variable to z-scores:

- " … imposes a standard normal distribution onto each indicator…" (Jacobs et al. 2004, p.37), or
- "…converts all indicators to a common scale in which they are assumed to have a normal distribution" (Jacobs et al. 2004, p.37), or
- "Standardisation (or z-score method) converts data in order to get normal distribution." (Hudrlikova and Kramulova 2013, p.38).

This is not the case: converting an indicator to z-scores simply rescales it to have a mean of 0 and a standard deviation of 1. A skewed indicator will have exactly the same skewness, and a similar departure from normality, after conversion to z-scores. Similar inconsistencies in terminology in the composite index literature have been noted by Heinrich et al. (2016).

In this report, *normalise* means any transformation of an indicator that aims to bring its distribution closer to a normal distribution. *Rescaling* means a change to the range of an indicator, and/or its mean and standard deviation, without altering the shape of its distribution.

Normalising to reduce excessive skewness and kurtosis is a step in many published composite indices (e.g. the Global Innovation Index and the Environmental Sustainability Index; Yang 2014), and is recommended in methodological guides (e.g. OECD 2008; Kovacevic 2010; Hudrlikova 2013). There are two reasons for normalising maldistributed indicator descriptions. Firstly, if an indicator distribution is highly skewed, then this has serious consequences when simple additive aggregation is used to form composite indices. This can be readily demonstrated with the simple example shown in Figure 3.1. An unskewed and a skewed indicator are rescaled to range 0 – 1, and added to form a composite index which is also rescaled to range 0 - 1. This is a very common approach to composite index construction. The bottom panel in Figure 3.1 shows the impact on the composite index of shifts in the values of each of the indicators from their medians to their first and ninth decile values. It can be seen that a shift in the value of skewed indicator from its median to either its first or ninth decile value results in a much smaller shift in the

composite index value. With extreme skewness, an indicator can have almost no impact on the composite index.

Similar problems will occur if the bulk of raw indicator values lie in a narrow range, with a few cases in both left and right tails of the distribution – a leptokurtic distribution. This is similarly illustrated in Figure 3.2. A change from the median to the third or seventh decile value for the normally distributed indicator has a much greater impact on the composite index value than does an equivalent change in the leptokurtic indicator.

When long tailed distributions are rescaled to a range of 0 to 1 (min-max rescaling – see Section 3.2.7), the indicator values for geographical units at one end of the distribution (skewed distribution), or those for the geographical units in the middle of the distribution (leptokurtic distribution), are compressed much more than would be the case in the absence of the long tails.

It is for these reasons that, if wholly or partially additive aggregation methods are used to calculate composite indices, then raw indicators need to be transformed to reduce skewness, overly positive kurtosis and excessive numbers of outliers. If not done, then the relative contribution of indicators to the composite index is being affected by the combination of indicator distributions and min-max rescaling.
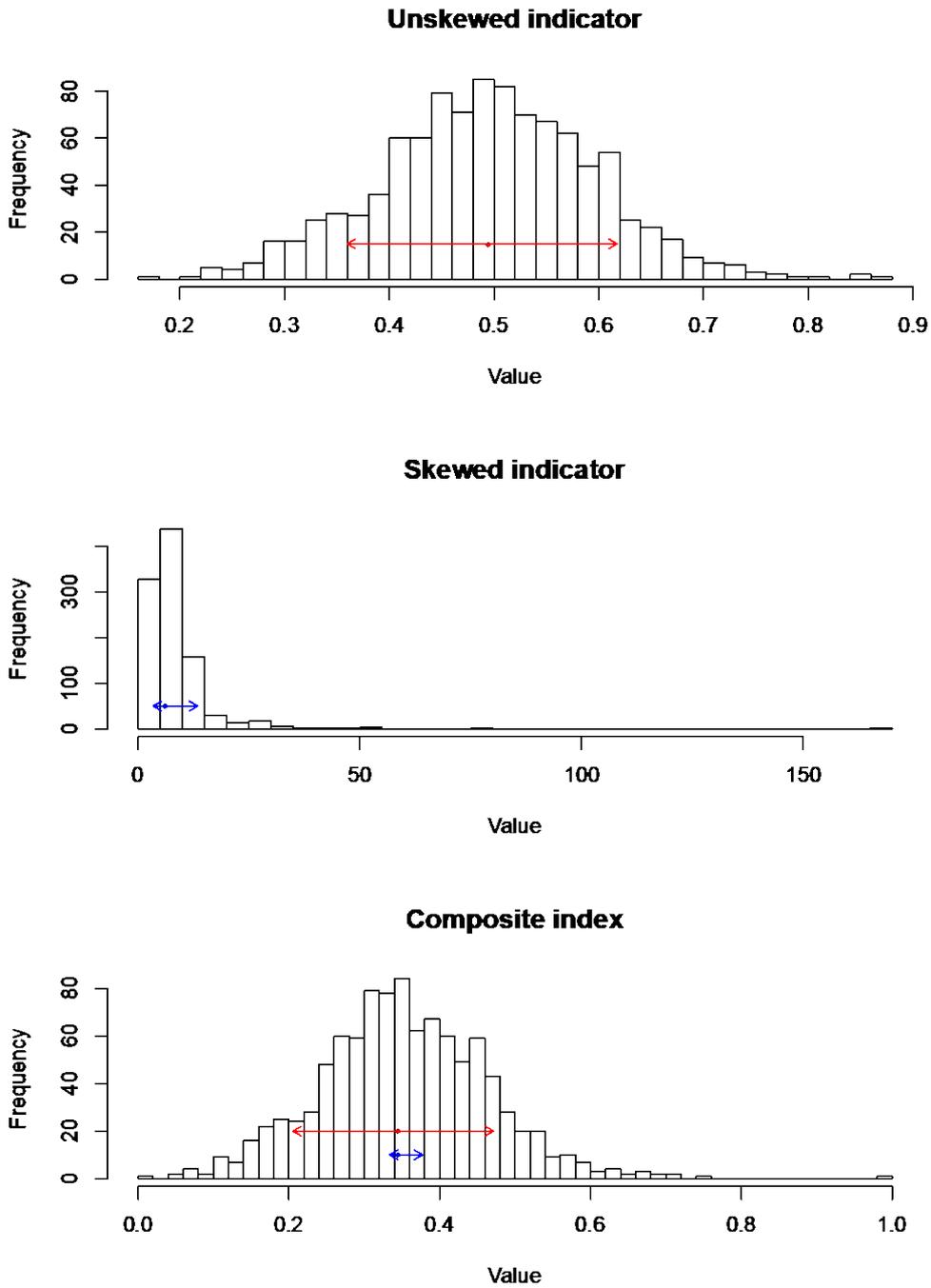
Figure 3.1: Histograms of an unskewed indicator, a skewed indicator and the resulting composite index. For the unskewed indicator, the red dot and arrows show the positions of the median and first and ninth decile values. For the skewed indicator, the positions of the same quantiles are shown similarly in blue. For the composite index, the dots and arrows show the changes in the index as each indicator in turn is changed from its median value to its first decile or ninth decile value.
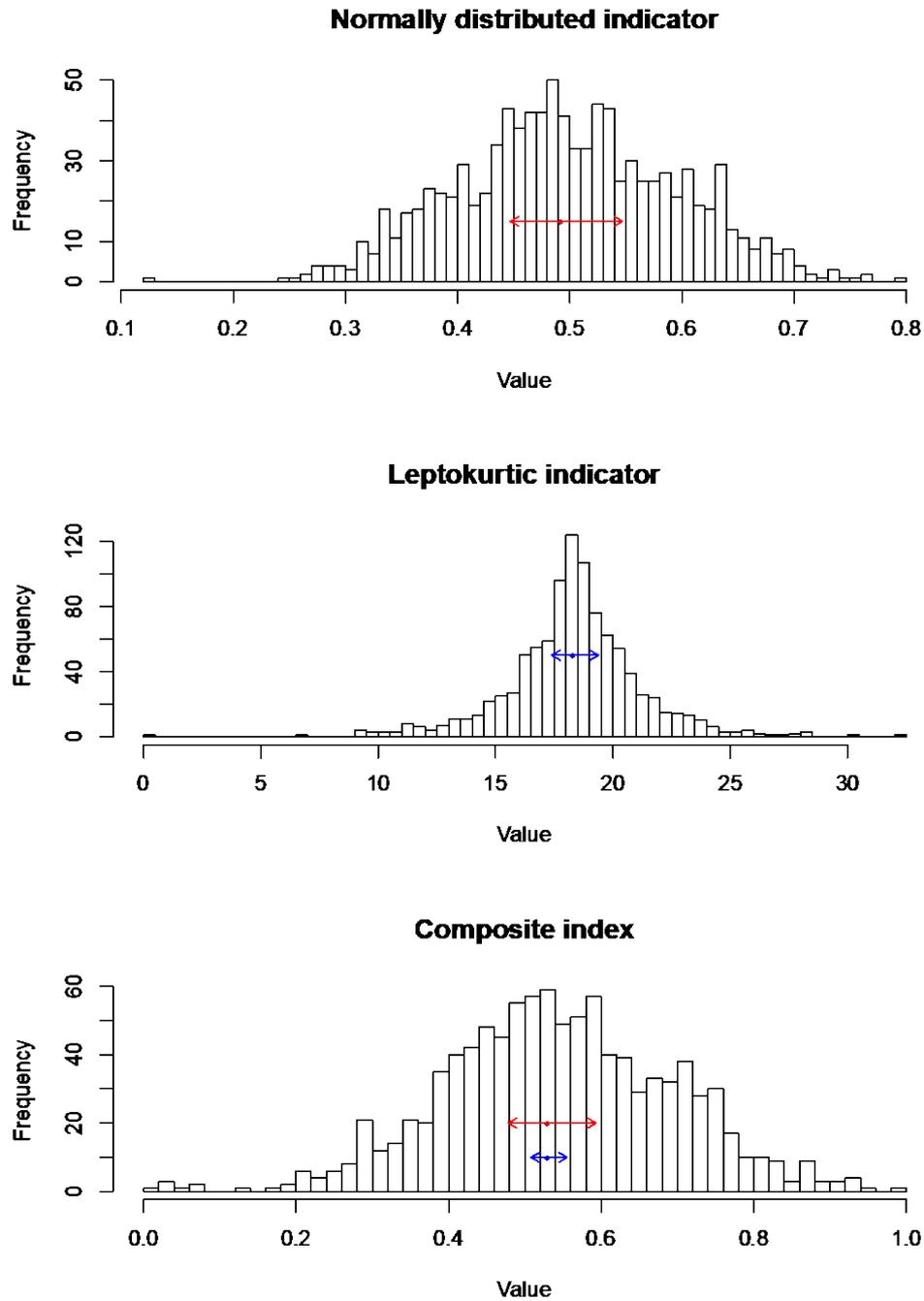
Figure 3.2: Histograms of a normally distributed indicator, a leptokurtic indicator and the resulting composite index. For the normally distributed indicator, the red dot and arrows show the positions of the median and third and seventh decile values. For the leptokurtic indicator, the positions of the same quantiles are shown similarly in blue. For the composite index, the dots and arrows show the changes in the index as each indicator in turn is changed from its median value to its third decile or seventh decile value.

The second reason for the use of normalising transformations of indicators is when PCA is to be used either as a means of dimensional reduction or as a means of deriving weights when a composite index is calculated by weighted addition. PCA is sensitive to substantial departures from normality and the presence of outliers. Depending on the shape of the original indicator distribution, transformation with log or power functions may bring the distribution closer to a normal distribution.

3.1.2.4   Transformations of indicators – rescaling

The need for rescaling of indicators originates in the intuition when simple additive aggregation is being used, that it would be undesirable to add indicators that vary greatly in magnitude. For example, if one indicator is the percentage of people with a rare disease, this might vary between 0.001 and 0.005 per cent. An accompanying indicator for forming a composite index by addition, might be the annual expenditure on the detection and treatment of this disease. This indicator might vary between $500,000 and $1.5million. Obviously, an additive composite index from these two indicators will be dominated by annual expenditure. Rescaling of indicators followed by additive aggregation is a widely used method of index construction, although there are some aggregation methods that do not require rescaling (see Section 3.1.2.7).

Two forms of rescaling that are widely used are min-max rescaling and what, in the interests of clarity, will be termed mean and standard deviation (MSD) rescaling. In the former, the minimum value is subtracted from each value of an indicator, and the result divided by the range of the indicator. In the latter, the mean is subtracted from each value of an indicator, and the result divided by the standard deviation. MSD rescaled values are equivalent to z-scores. As noted in Section 3.1.2.3, MSD rescaling does not bring an indicator distribution closer to a normal distribution.

A number of authors have drawn attention to the fact that the method of rescaling used will affect the relative positions of countries or regions as scored on a composite index. For example, Hudrlikova and Kramulova (2013) tested a range of rescaling methods, including min-max rescaling and MSD rescaling and found the relative position of Czechoslovakian regions on a composite index of sustainable development varied according to the rescaling method chosen. Cherchye et al. (2007) argued that this aspect of composite indices essentially rendered them meaningless:

> In a well-defined mathematical sense, a composite indicator is **not meaningful** when the resulting ordering changes if the original data are transformed in such a way that their informational content is not fundamentally altered. In practice, however, most composite indicators are prone to precisely this deficiency. [emphasis in original]

Similar criticism was levelled at environmental composite indices by Ebert and Welsch (2004):

> *The popular procedure of normalizing* [i.e. rescaling] *data before aggregating them does not provide a solution to the non-comparability of the data and the ensuing ambiguity of orderings. Rather, the arbitrariness of the normalization rules introduces additional ambiguities.*

A number of sensitivity or uncertainty analyses examining the effect of methodological uncertainty upon composite indices have concluded that the effect of choice of rescaling method is relatively small, compared to choices around weighting and aggregation (Tate 2012; Saisana et al. 2005). Nevertheless, the fact that choice of rescaling method is capable of changing a composite indicator suggests that some care in making this choice is warranted (Baptista 2014; Tate 2012; Cherchye et al. 2007).

Of the 105 disaster risk, vulnerability and resilience composite indices reviewed by Beccari (2016), where the rescaling method was described:

- 23 did not use rescaling as the indicators in their raw form were expressed in similar scales, or the aggregation method did not require rescaling;
- 23 used min-max rescaling;
- 19 used MSD rescaling;
- 17 used assignment to categories;
- 2 used ranking; and,
- 21 used other or mixed methods (Beccari 2016).

Beccari's review shows that, of the rescaling methods in use that retain ratio indicators (rather than converting them to ordinal or categorical indicators) min-max rescaling and MSD rescaling are the most commonly used.

If the main goal of rescaling is to ensure indicators are within the same range of values, and indicators have been transformed to remove excessive skewness and leptokurtosis and in the process outliers have been reduced or removed, then then the argument that MSD rescaling is preferable when there are outliers (Baptista 2014) does not hold. Furthermore, MSD rescaling, because it entails the division by the standard deviation of the indicator, introduces unnecessary variation into the rescaled indicator that is a function of the distributional qualities of the indicator. For example, if two indicators had exactly the same range, then min-max rescaling would not alter their relative contributions to a composite index. If at the same time, the indicators had different standard deviations, then MSD rescaling would inflate or deflate the contribution of one or other indicator to a composite index, something that is manifestly undesirable for indicators that have exactly the same range.

//

It should be noted that some care has to be taken with any form of rescaling for composite indices that are, or are expected to be, measured over time. For min-max rescaling, indicators should not be rescaled in each time period using the minimum and maximum values that pertained at that time period, since this distorts the relativity of the indicator between each pair of consecutive time periods. Rather, the maximum and minimum values across the whole time span of the indicator should be used in the rescaling calculation. Necessarily, this means that the published rescaled values of the indicator for previous periods will change if the value of the indicator in the latest time period has moved outside of the previous range of values. This issue is examined in some detail by Heinrich et al. (2016).

3.1.2.5 Indicator reversal

Where an indicator is believed to have a negative relationship with the concept it is intended to capture, it is necessary to reverse the direction of the indicator. In some studies (Salzman 2003; Cutter et al. 2010; Baptista 2014; Sessa 2016), indicator reversal is accomplished by a reciprocal transformation that changes the negative association to a positive association. However, this form of reversal is not linear and can change the skewness of the indicator distribution. As discussed in Section 3.2.7, skewed indicator distributions have undesirable effects on composite indices constructed by additive aggregation. To illustrate with a simple example, a random sample of 1,000 values from a normal distribution with a mean of 10 and standard deviation of 2 has a skewness close to zero. Taking the reciprocal of these values produces a distribution with a skewness of 3.198. Min-max rescaling of the reciprocal values yields a distribution that still has a skewness of 3.198.

A preferable (and linear) method of indicator reversal is to subtract each value of the min-max rescaled indicator from 1. This maintains the absolute value of the skewness of the distribution. If working with an un-rescaled indicator, then the indicators should be subtracted from their maximum value.

3.1.2.6 Correlation between indicators

The correlation between indicators is frequently used in the construction of composite indices as a criterion for choosing among indicators for inclusion, i.e. indicators that are highly correlated with other indicators are regarded as redundant and can be excluded from the calculation of the composite index. The approach taken ranges from simple subjective thresholds for exclusion (e.g. Cutter et al. 2010; Malczewski 2000; Salzman 2003; OECD 2008), to more detailed correlation analysis (e.g. Sherrieb 2010), to PCA. Beccari (2016) lists some 17 studies that used PCA, mostly following the approach established by Cutter et al. (2003).

There are a number of issues around indicator correlation that receive relatively little attention in the composite index literature, but which deserve careful consideration in the development of composite indices. At the most basic level, there is the question of the intended or implied direction of causation between indicators and a composite index. Structural equation modelling distinguishes between reflective and formative measurement models (Fornell and Bookstein 1982; Edwards and Bagozzi 2000). In reflective measurement models, causation flows from a hypothesised theoretical construct to a number of indicators, e.g. locus of control is a hypothesised psychological trait which affects people's beliefs about the extent to which they can control the course of their lives. It also affects how they respond to certain questionnaire items about their control. In formative measurement models, causation flows from a chosen set of indicators to a hypothesised construct, e.g. various demographic indicators such as proportion of people over 65 are believed to influence the level of a hypothetical construct such as disaster resilience.

In terms of measured indicators that are aggregated in some way to give a measure of a hypothesised construct, formative and reflective measurement models are very similar (Table 3.1). However, the difference in direction of causation has important implications for the transfer of statistical techniques between fields of study. This applies particularly to the interpretation of correlations between indicators and techniques such as factor analysis and Cronbach's alpha (Cronbach 1951). Cronbach's alpha varies between 0 and 1, with higher values for higher levels of correlations among items. It is widely used as a measure of reliability and internal consistency in psychometric scales (i.e. with reflective measurement models), although it is by no means universally accepted as a good measure (see, for example, Tavakol and Dennick 2011). Higher correlations among items suggests that their values are all being affected by the latent hypothetical construct of interest. As pointed out by Christopherson and Konradt (2008), while factor analysis and Cronbach's alpha are important tools in interpreting the correlation among indicators in reflective measurement models, they are essentially meaningless when applied to formative measurement models (i.e. all composite indices of resilience or vulnerability).

Table 3.1: Examples from different fields showing the equivalence of scales, indices and rankings, together with the differences in direction of causation.

Field	Units of analysis	Assessment variables	Aggregate measure	Direction of causation
Disaster resilience	Geographic regions, census tracts	Demographic, economic, infrastructural indicators	Score on disaster resilience index	From indicators to index
Psychology	Individuals	Questionnaire item responses	Score on psychometric scale representing theoretical construct	From construct to item response
Health management	Hospitals	Key performance indicators	Score on hospital performance index	From indicators to index
Human development	Countries	Demographic, economic indicators	Score or ranking on development index	From indicators to index
Multiple criteria decision analysis	Suite of decision options	Criteria for rating options	Score or ranking of options	From criteria to score or ranking
Life cycle analysis	Consumer products	Environmental and health impacts	Score on life cycle impact index	From impacts to index

In formative measurement models, such as composite indices of disaster resilience, higher correlations among items cannot be interpreted in this way. At best, the higher correlations show that the indicators believed to influence resilience just happen to be correlated with each other. Some indicators may in fact be redundant. However, instances can be found in the composite index literature where Cronbach's alpha is used, or even recommended, as a means of determining the internal consistency of indicators, despite these being formative measurement models (e.g. OECD 2008; Ross 2014).

Factor analysis assumes a latent set of factors that are responsible for the values of the observed variables (i.e. a reflective measurement model with causation flowing from the latent constructs to the measured indicators). Factor analysis applied to a formative measurement model, such as a disaster resilience index, can yield no meaningful interpretation. Any groups of correlated indicators are due to the analyst's choice of indicators and cannot say anything about the multidimensionality or otherwise of the hypothetical construct, disaster resilience. Nevertheless, a number of composite index studies refer to factor analysis as an option for dealing with correlation between indicators. These include: Bao et al. (2015); Jacobs et al. (2004); Kovacevic (2010); and, OECD (2008).

PCA is a multivariate statistical technique related to factor analysis. However, it does not treat the values of variables as a manifestation of a number of latent constructs. Rather, it is simply a method to reduce a large number of variously correlated indicators to a smaller number of uncorrelated components, each of which is a variously weighted sum of the original indicators. As Beccari (2016) notes, this approach has been widely used in the construction of composite indices for disaster resilience and vulnerability.

However, it appears to be rarely appreciated in such studies, if at all, that the components obtained from PCA are weighted sums of the (standardised) original indicators and, as such, suffer from the compensability issue. It is implicit in these weighted sums that two units (SA2s in the current study) can end up with the same score on a component, due to low values on some indicators compensating for high values on other indicators. The compensability implicit in composite indices has attracted much criticism, and the same criticism can be applied to the components from PCA that are used as a parsimonious set of indicators to replace a much larger set of original indicators. So a composite index constructed hierarchically of sub-indices that are derived from PCA will be exposed to compensability issues at several levels.

An alternative to the use of PCA is simply to exclude indicators that have high correlations with other indicators (e.g. Cutter et al. 2010; Malczewski 2000; Salzman 2003; OECD 2008; Sherrieb 2010). However, it is important to note that correlation between indicators does not necessarily imply redundancy. Mazziotta and Pareto (2016b) give the example of two indicators of health provision: hospital beds per 1,000 persons and hospital doctors per 1,000 persons. These two indicators are likely to be correlated, but one cannot substitute for the other. A high level of hospital doctors would not compensate for a low level of hospital beds, since beds are needed for doctors to provide care. Each indicator is, in fact, an enabling indicator for the other. If, for example, hospital beds per 1,000 persons was dropped as an indicator, hospital doctors per 1,000 persons would be an inaccurate substitute, overestimating health provision whenever doctors was at a high level, but beds at a low level. Similarly, if a pair of indicators have a high negative correlation, then they are likely to compensate for each other if the aggregation is partly or fully additive. If one indicator is deleted, the remaining indicator will be an inaccurate substitute. In this situation, it would be preferable to retain both indicators and use an aggregation function that allows a realistic degree of compensation.

These examples demonstrate that suites of potential indicators cannot be simply culled with some chosen correlation threshold. The nature of the hypothesised relationships between indicators and the concept represented by the composite index has to be considered. As Hudrlikova (2013) notes, some correlations will be indicative of redundancy, and other correlations may involve non-compensatory indicators, as in the example above. For this reason, indicator selection will always involve a trade off between redundancy if some correlated indicators are retained and loss of information if they are not (Mazziotta and Pareto 2013a). On the other hand, if a suite of chosen indicators is largely uncorrelated, and there is good evidence to believe each indicator has an independent influence on the latent construct represented by the composite index, then the lack of correlations gives some confidence in the content validity of the index (see Section 3.2.2).

A further consideration with correlated indicators is their impact on the composite index and their apparent importance. Paruolo et al. (2013) demonstrate that, if the correlation between an indicator and the composite index is taken as a measure of the importance of that indicator, then the apparent importance of an indicator that has correlations with other indicators is inflated. This problem can be corrected for by weighting (e.g. OECD 2008), although Saisana et al. (2005) point out that in the mathematically equivalent field of multi-criteria decision analysis these correlations would be regarded as a feature of the decision problem that may involve non-compensatory relationships between criteria, and would not be regarded as problematical.

3.1.2.7   Weighting and aggregation

Weighting and aggregation is the area of composite index methodology that has received the most attention in the literature (Hudrlikova 2013). Weighting issues revolve around giving expression in some way to the importance that individual indicators are believed to have in affecting the latent construct that is gauged by the composite index. At the same time, it is necessary to avoid, or at least be aware of, any unintended weighting effects that are implicit in the aggregation methodology itself.

Aggregation issues, in addition to unintended weighting effects, are mostly concerned with arriving at an index that somehow gives expression to the pattern of indicator values, without being unstable or misleading. The central issue, widely discussed in the literature, is compensability between indicators, i.e. whether or not low values of some indicators can be compensated for in the aggregation process by high values of other indicators. A further consideration in weighting and aggregation methodology, that has become relevant in recent times with the use of aggregation operators that allow for detailed prescription of levels of compensability between indicators, is the level of expert input required to model the compensability. In general, methods that require extensive efforts by (possibly volunteer) experts are unlikely to be practicable.

Finally, an enduring issue, despite great improvements in computer processing speeds, is the length of time required for aggregation calculations. While the scoring of options in Multiple Criteria Decision Analysis (MCDA) is mathematically equivalent to constructing indices of resilience to hazards (Table 3.1), and so allows MCDA methods to be applied to the task of constructing resilience indices, the latter are far more computationally intensive than the former. MCDA generally involves a small number of options to be evaluated and up to some tens of criteria, whereas mapping resilience on a national scale can involve several thousand geographical units and up to one hundred indicators. Sophisticated non-compensable aggregation methods from MCDA may have prohibitively long calculation times when applied to a large number of geographical units and indicators.

Equal weighting

Weighting can be carried out as a calculation stage separate from aggregation, or it can be implicit within an aggregation method. Where weighting occurs as a separate operation, a simple and widely used approach is to use equal weights for each indicator (equivalent to not weighting at all). This is considered to be justified if there is no information as to the relative importance of indicators or sub-indices (e.g. Cutter et al. 2010). Salzman (2003) considered that equal weights were justified since, despite their shortcomings, they were preferable to other options for setting weights, such as expert deliberation and PCA, which were even more problematical. However, De Muro et al. (2011) point out that if equal weighting is followed by simple arithmetic averaging (a common procedure), then the apparently "neutral" assumption of equal weights disguises a strong assumption of perfect substitutability between indicators, i.e. indicators are fully compensable and low values of some indicators can be compensated for by high values of other indicators.

Elicited weights

Among weighting methods that involve the specification of weights, there are a number of types. Following Hudrlikova (2013) and Baptista (2014), there are two broad groups of weighting methods: those that derive weights from the indicator data itself, and those that involve elicitation of additional information from experts or members of the public. Among the latter are consultation with experts, discussions with stakeholders, surveys of public opinion, the budget allocation process, conjoint analysis and the analysis hierarchy process (Baptista 2014; Hudrlikova 2013; OECD 2008). Despite the intuition that weights reflect the importance of individual indicators, whenever weights are used with simple additive aggregation of rescaled indicators, they amount to nothing more than the marginal rates of substitution between the indicators (Kovacevic 2010).

Weights from Principal Components Analysis

The main method where weighting is separate from aggregation, and where weights are derived from the indicator data itself is PCA. This has already been discussed in Section 3.1.2.6 in relation to reducing a large set of variously correlated indicators to a parsimonious set of uncorrelated components. However, PCA also features in composite index studies as a means of providing indicator weights to calculate scores on components (also termed pillars or dimensions). The percentage of variance explained by each component is then used as a weight in aggregating the chosen number of components to form the composite index (OECD 2008). Alternatively, the indicator weights in the first principal component are used in constructing a composite index (Salzman 2003). However, as noted in Section 3.1.2.6, implicit in weighting by

PCA is the assumption that indicators are fully compensable in calculating component scores, and that the components themselves are fully compensable in calculating the composite index. Once again, these weights amount to marginal rates of substitution between indicators or components and say little about their relative importance.

Aggregation post-weighting

Where aggregation is carried out as a procedure separate to, and following, the assignment of weights to indicators, simple summation or averaging is the most common approach. However, as mentioned above, this method assumes full compensability between indicators. Furthermore, as noted by Munda and Nardo (2009) and others, the trade-off ratio between indicators is assumed to be constant, regardless of the values of other indicators (i.e. preferential independence). This amounts to assuming that there are no synergistic or antagonistic interactions between indicators (or between sub-indices in a hierarchical composite index). The assumption of full compensability and preferential independence has been seen as problematic by many authors, including Baptista (2014), Tate (2013), Hudrlikova (2013), Mazziotta and Pareto (2016a) and Natoli and Zuhair (2011). Munda and Nardo (2009) reject simple additive aggregation and argue that non-compensatory aggregation must be used whenever weights are intended as importance measures and/or preferential independence cannot be assumed.

While many composite index studies have ignored these undesirable aspects of simple additive aggregation, the problem has nonetheless driven the search for non-compensatory aggregation techniques. These techniques have been widely used in MCDA and many studies involve testing the application of MCDA non-compensatory or partially compensatory aggregation techniques with composite indices in other fields. Examples include:

- water resource management (ELECTRE III method; Chitsaz and Banihabib 2015);
- finance (ELECTRE and preference disaggregation analysis; Doumpis and Zopounidis 2014);
- farm forestry evaluation (ELECTRE II method; Jeffreys 2004);
- materials science (MOORA method; Karande and Chakraborty 2012);
- environmental management (ordered weighted averaging method; Malczewski 2006b);
- industrial process engineering (MOORA method; Mandal and Sarkar 2012);
- regional economic development comparisons (Adjusted Mazziotta-Pareto index and the Mean-min function; Mazziotta and Pareto 2015);
- regional well-being comparisons (Mazziotta-Pareto index and weighted product method; Mazziotta and Pareto 2016a);
- regional well-being comparisons (Choquet integral; Bertin et al. 2018);

- socio-economic development comparisons between countries (Condorcet method; Natoli and Zuhair 2011);
- regional mapping of human development (ELECTRE TRI-C method; Pereira and Mota 2016);
- international development comparisons (Adjusted Mazziotta-Pareto index; Sessa 2016);
- international sustainability comparisons (Choquet integral; Cruciano et al. 2012);
- vulnerability to climate change (Weighted Ordered Weighted Average method; Runfola et al. 2015);
- governance comparisons in Africa (Mean-min function; Tarabusi and Guarini 2013); and,
- regional poverty comparisons (generalised mean; Weziak-Bialowolska and Dijkstra 2014).

It appears that the uptake of non-compensatory or partially non-compensatory aggregation methods in the natural hazards and disaster resilience field has been minimal. Beccari (2016), in his review of 106 composite index methodologies in this field, found that almost all involved some form of additive aggregation. This implies that most methodologies assumed full compensability and preferential independence.

Aggregation and weighting combined

The Benefit of Doubt (BOD) method derives weights endogenously from the indicators themselves, as part of the aggregation process. The method is fully compensatory and the derivation of weights is built on the assumption that the values of indicators reflects how important the corresponding areas are considered to be (Cherchye et al. 2007). For example, high values for health provision in a particular country is an indication that health provision is regarded as important in that country and public policy in health matters has created the high values. Obviously, this assumption is unlikely to hold for small geographic units that do not have the authority to implement public policy, nor does it hold for domains that are generally outside the scope of public policy, such as age and sex distribution.

3.1.3 Uncertainty and sensitivity analysis

Uncertainty and sensitivity analysis treats composite index calculations as a model, where inputs are indicator values and the output is a value for a composite index. The specification of the model (i.e. the method of composite index calculation) may include alternative options for various parts of the calculation, such as aggregation by arithmetic mean or by geometric mean. The values of indicators may be uncertain for a range of reasons: data derived from surveys with sampling uncertainty, random adjustments of small cells in Census data for confidentiality reasons, or spatial disaggregation from data at

one geographical scale to obtain indicators at another scale. Likewise, if theoretical considerations give no guidance, then there may be uncertainties as to which index construction methods should be used. Uncertainty analysis seeks to quantify how the uncertainties in indicators and index construction methods flow through the calculation to the composite index. The results of uncertainty analysis enable the distributional characteristics of the index to be expressed with descriptive statistics such as the mean, standard deviation and quantiles.

Sensitivity analysis aims to apportion the variance in the calculated index among the indicators and methodological choices. Efforts to improve the robustness of the index can be directed to those indicators or choices that have the most effect on the uncertainty of the index, while indicators or methodological choices that have no effect on the index may be unnecessary, their omission leading to a more parsimonious index.

Uncertainty and sensitivity analysis can build confidence that a composite index and its conceptual model are a valid reflection of reality, but this requires that the uncertainties in the indicators and methodological choices are honestly and plausibly specified, and the resultant uncertainty in the composite index is not so large as to render it useless. As Leamer (1990) noted:

> *Conclusions are judged to be sturdy only if the neighborhood of assumptions is wide enough to be credible and the corresponding interval of inferences is narrow enough to be useful.*

Beccari's (2016) review of the 106 distinct frameworks for calculating composite indices of generic vulnerability or resilience to natural hazards, published between January 1990 and March 2015, found that only 20 studies contained any explicit analysis of uncertainty and sensitivity. Just two studies estimated errors in the composite index scores and only one study employed global sensitivity analysis. This latter form of sensitivity analysis systematically traverses the space of all uncertain inputs and was recommended in 2008 as good practice in composite index construction (OECD 2008). The majority of the abovementioned 20 studies only examined the impact of changes to weights on the index. Beccari (2016) concluded that the calls for greater use of sensitivity analysis as a counterbalance to the obvious flaws in composite index construction had largely gone unheeded in the field of disaster vulnerability and resilience.

3.1.4 Conclusions

The history of composite index development shows that the representation of a complex system with a single number has an irresistible allure. The addition or averaging of rescaled indicators has had an intuitive appeal that has made it the most widespread of aggregation methods. However, as composite index

construction has received increasing scrutiny, and an increasing number of fields have found applications for composite indices, their shortcomings are becoming better understood. Issues of indicator rescaling and compensability have received increasing attention. These problems have driven the search for non- or partially-compensatory aggregation methodologies where weights can validly be interpreted as measures of importance.

Taking all the fields where composite indices are used, there is a proliferation of aggregation methods, although the range of methods currently used in natural disaster vulnerability or resilience is more restricted. It is widely recognised that the choice of indicators and their rescaling, weighting and aggregation into an index carries a considerable degree of subjectivity (Baptista 2014, Barnett et al. 2008, Cherchye et al. 2007, Cutter et al. 2010, Hudrlikova 2013, Mazziotta and Pareto 2015, OECD 2008, Schmidtlein et al. 2008, Sessa 2016, Tate 2013, Vidoli et al. 2015). Recognising the inherently subjective nature of composite index construction, many authors emphasise that methodological choices should be made transparent and include the reasoning and context that led to these choices (Agder et al. 2004, Baptista 2014, Beccari 2016, Hudrlikova 2013, OECD 2008, Tate 2012). If possible, this should be formalised in uncertainty and sensitivity analysis.

3.2 INDEX CALCULATION

This section provides an overview of the method of construction of the Australian Natural Disaster Resilience Index, describing the general logic behind the methods used and the approach taken in dealing with the methodological issues raised in the literature review in Section 3.1.

3.2.1 Context

The conceptual basis, design, structure and indictors used in the Australian Natural Disaster Resilience Index are described in detail in Chapter 1 and 2. In brief:

- the purpose is to construct and index that can be used to assess the state of disaster resilience in Australia, with a view to informing policy, planning and community engagement at all levels of government;
- the conceptual basis of the Australian Natural Disaster Resilience Index is grounded in the disaster resilience literature, from which is drawn the idea that resilience is a function of coping capacity and adaptive capacity;
- the literature further suggests that coping capacity is a function of social character, economic capital, emergency services, planning and the built environment, community capital and information access, while adaptive capacity is a function of governance and leadership, and community and social engagement;
- the Australian Natural Disaster Resilience Index uses a top-down approach based on secondary data sources;
- the geographic unit used in the Australian Natural Disaster Resilience Index is the Australian Bureau of Statistics Statistical Area Level 2 (SA2), a choice dictated by the national scale and data availability;
- this conceptual basis requires a hierarchical structural design for composite index calculation, with two coping and adaptive capacity sub-indices and eight theme sub-indices, and,
- the theme sub-indices are calculated from 77 indicators sourced from readily available secondary data.

A number of methodological choices flow from this context, and from the issues discussed in Section 3.1.

3.2.2 Functional form

The functional form of relationships between indicators and sub-indices, and between sub-indices and Australian Natural Disaster Resilience Index, assumes monotonically increasing or monotonically decreasing relationships. For example, an increase in unemployment always results in a decrease in resilience, regardless of the level from which the increase takes place. A fall in unemployment from 60 to 50 per cent will result in an increase in resilience, as

will a fall from 20 to 10 per cent. Generally, the state of knowledge in the current literature that guided the choice of indicators (see Chapter 2) is insufficient to specify functional form in any greater detail.

3.2.3 Spatial coverage

The data acquisition process aimed for maximum national coverage at the SA2 level. Overall, there are 2,214 SA2s across Australia. The Australian Natural Disaster Resilience Index was computed for 2084 of these SA2s: 130 SA2s (6%) were excluded because they were areas of no or low population (e.g. national parks, ports, airports, industrial estates). The theory and dimensions of disaster resilience are unlikely to apply to these 130 SA2s and any index calculation could be misleading. Jervis Bay, Christmas Island, the Cocos-Keeling Islands, Lord Howe Island and French Island were also excluded from the index because the availability of indicator data for these areas was inconsistent. Details of the SA2s included and excluded from the index are provided in Chapter 1.

3.2.4 Missing values

Where a small number of SA2s were missing data for an indicator, the missing values were imputed using the bootstrap EM algorithm available in the 'Amelia' contributed R package (Honaker et al. 2011). Each missing value was taken as the mean of ten iterations of the algorithm. Six out of the eight themes in the Australian Natural Disaster Resilience Index had no missing values and this data was used in the imputation of missing values in the remaining themes. The themes with missing values were governance and leadership and planning and the built environment. Where an SA2 was missing data for a large number of indicators, the SA2 was omitted from the Australian Natural Disaster Resilience Index (see Section 3.2.3).

3.2.5 Indicators and concordance to SA2s

Across the eight themes, 77 indicators were used to compute the Australian Natural Disaster Resilience Index. A detailed description of the indicators used in the index, and the source and treatment of each indicator is provided in Chapter 2.

While a key principle of the index was to obtain data collected at the SA2 resolution, some indicator data are only available for geographic units other than SA2. These units included: SA3, SA4, Local Government Area, police districts and States/Territories. The adjustments made to disaggregate data are outlined in Chapter 2.

3.2.6 Normalisation

The development of Australian Natural Disaster Resilience Index does not allow for the extensive consultation that would be needed to establish the compensatory relationships between 77 indicators. To account for these compensatory relationships in aggregation, it is expected that aggregation methods will be at best partially non-compensatory, and so partially additive. Under these circumstances, it was still necessary to reduce, where possible, the undesirable influence of highly skewed and/or leptokurtic indicator distributions on the composite sub-indices and indices where there is an additive component to the aggregation method.

The indicator transformation process was carried out in two stages. First, skewed indicators were adjusted to zero skewness with a power transform. This was done by finding the power to which indicator values have to be raised to result in a distribution with zero skewness. The process is illustrated using the indicator '% of labour force unemployed' (Figure 3.3). Untransformed, '% of labour force unemployed' has a skewness of 5.10. If each value of this indicator is raised to the power of 0.47, the skewness of the distribution of the transformed indicator is very close to zero.
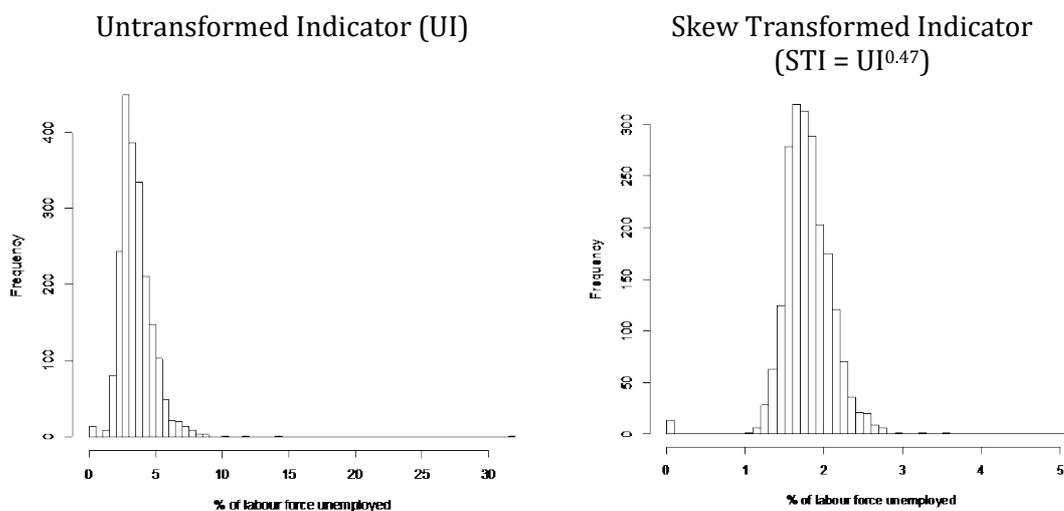
Figure 3.3: Histograms of the '% of labour force unemployed' indicator before and after transformation.

Untransformed, '% of labour force unemployed' has a kurtosis of 89.20, which is strongly leptokurtic. The transformed distribution is not as leptokurtic, having a kurtosis of 11.92. No procedure for reducing leptokurtosis could be found in the literature so a method was developed based on the intuition that the distribution of indicator ranks is strongly platykurtic, so that further transforming indicator values by some linear combination of each skew transformed indicator (STI) value and its rank, could reduce kurtosis to near zero, while

retaining the relative positions of the indicator values. This amounts to solving the equation:

$$KSTI = minmax(STI) + coef * minmax(R),$$

where: *KSTI* is the kurtosis and skew transformed indicator,

STI is a vector of skew transformed indicator values,

R is the rank of individual *STI* values,

coef is a multiplicative coefficient, and

minmax is a function that rescales a vector to range 0 to 1.

If STI has negative kurtosis, then the transformation is not required, since a platykurtic distribution does not carry the same risk of introducing methodological artefacts under rescaling and aggregation.

The resulting distribution of '% of labour force unemployed' after finding the value of *coef* that reduces the kurtosis to zero is shown in Figure 3.4. This distribution has a skewness of 0.04 – still close to zero – and a kurtosis of zero.

Transformed to correct skewness and kurtosis
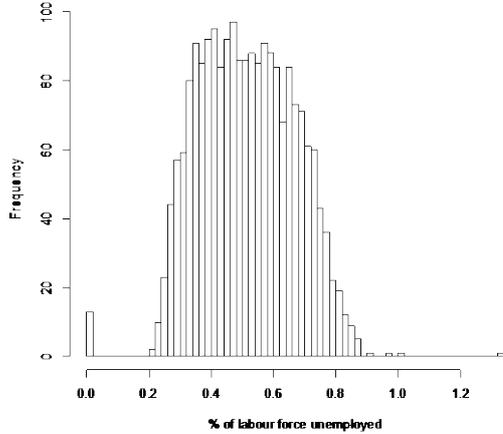
Scatter plot of transformed and untransformed indicator values
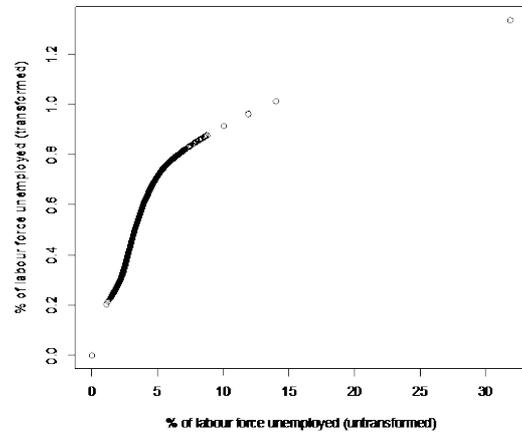
Figure 3.4: Histogram of the indicator '% of labour force unemployed' after transformation for skewness and excessive leptokurtosis, and scatter plot of transformed and untransformed values.

All the Australian Natural Disaster Resilience Index indicators showed that the kurtosis transformation generally resulted in little change to the skewness. The scatter plot in Figure 3.4, shows that the transformation of the indicator is a mildly non-linear transformation. Comparing the ranks of SA2s for the untransformed and transformed indicator showed that the transformation preserves the rank order of SA2s.

As discussed in Section 3.1.2.2, the outlier values of indicators based on the whole population should not be interpreted within an inferential statistics framework. Nonetheless, outliers serve to warn of the presence or otherwise of long tailed distributions. Using the threshold of absolute values of the z-scores greater than 3.29 (a common univariate outlier threshold – Tabachnick and Fidell 2007), the untransformed indicator '% of labour force unemployed' has 10 outliers so defined, while the transformed indicator has just one outlier.

3.2.7   Rescaling

It was anticipated that aggregation of indicators will be at least partially additive. Under these circumstances, it was necessary to reduce where possible the undesirable influence of indicator distribution and scaling effects on the composite sub-indices and indices where there was an additive component to the aggregation method. Consistent with this, and the discussion in Section 3.1.2.4, all indicators were rescaled to a range of 0 to 1.

3.2.8   Indicator reversals

Where the literature suggested that the relationship between an indicator and resilience was negative, the normalised and rescaled values of the indicator were subtracted from 1. Full details of the reversal of indicators are provided in Chapter 2.

3.2.9   Indicator redundancy, correlation and compensability

While most resilience indicator studies, and indeed the composite index construction guidelines of OECD (2008), separate the examination of indicator correlations and redundancy from the aggregation phase, it was found useful in the construction of the Australian Natural Disaster Resilience Index to combine these two analysis stages. The correlation between indicators determines the extent to which compensability issues have to be considered when they are aggregated. For example, when two indicators are highly correlated, each geographical unit will tend to have similar values on those indicators (i.e. both high values or both low values). In this situation, aggregation will not result in low values compensating for high values or vice versa, and simple addition or averaging is an acceptable method of aggregation, as is the case in reflective measurement models.

On the other hand, if a pair of indicators are uncorrelated then many geographical units will have a high value for one indicator and a low value for the other. In a formative measurement model this lack of correlation may be desirable as being indicative of two independent factors that determine resilience. However, the disparate values will compensate for each other if the indicators are aggregated by simple summation or averaging (the most common methods used), and these uncontrolled compensatory effects may or may not be acceptable in the context of the physical reality represented by the indicators.

The steps followed in considering correlation, redundancy and compensability among Australian Natural Disaster Resilience Index indicators are described in the following sub-sections.

3.2.9.1 Structural redundancy

Structural redundancy occurs when an indicator has a direct linear relationship with another indicator, e.g. '% of persons over 75 years of age' and '% of persons 75 years of age or less'. Where structural redundancies occurred, generally as a result of a thorough first round of indicator collection, one or other of the structurally redundant pair of indicators was omitted. The choice of which indicator to omit was guided by the resilience context of the indicators. For example, with the two illustrative indicators above, '% of persons over 75 years of age' might be considered to better represent the aspect of disaster resilience associated with older age, such as mobility and resources (see Chapter 2). The complementary indicator would not capture these resilience relationships. For this reason, the disaster resilience context would suggest the first indicator should be retained.

3.2.9.2 Correlation and compensability

The next step after eliminating any obvious structural redundancy was to examine the overall correlation structure. This was done with a level plot with the indicators in the same order as in the sorted PCA loadings table. An example is given in Figure 3.5.
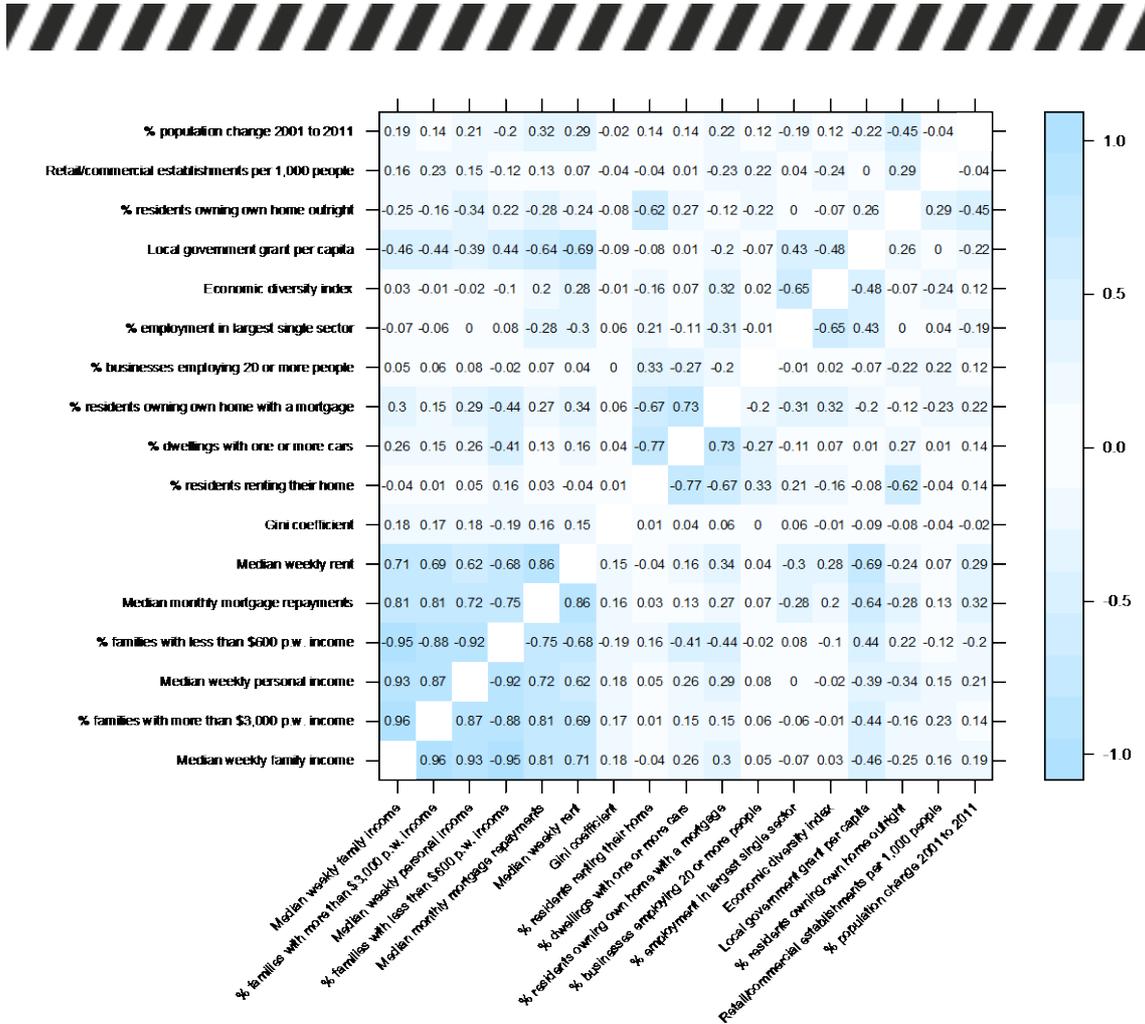
Figure 3.5: Example correlation level plot for the economic capital theme indicators.

The level plot serves two purposes. First, it flags any high correlations between indicators. These correlations require the indicator context be considered to decide whether or not one or other of the indicators should be omitted. For example, if two indicators are structurally redundant one indicator can be omitted. On the other hand, if two correlated indicators have an enabling relationship, as described in Section 3.1.2.6, then they are retained.

Second, the level plot summarises the overall correlation structure of the indicator set and is an important input to the design of the aggregation strategy for forming a sub-index or index. The aggregation strategy has to deal with an unavoidable trade-off in the aggregation process. On the one hand, aggregation functions that allow a well-considered and nuanced accounting for the degree of compensability between indicators, are knowledge demanding and, for many aggregation functions, prohibitively computation intensive as the number of indicators increase. On the other hand, aggregation functions that will handle larger numbers of indicators can only, at best, deal with compensability between indicators in an approximate way. Because of the exhaustive compilation of potential indicators from secondary data sources, most of the sub-indexes in the Australian Natural Disaster

▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨

Resilience Index comprised more than just a few indicators. For this reason, it was necessary to choose a set of aggregation functions suited to the number of indicators to be aggregated and the extent to which it was feasible to specify compensatory effects between indicators.

A second approach to examining the pattern of correlations among indicators was to regress each indicator in turn as a dependent variable on the remaining indicators as independent variables. A high R^2 for an indicator showed that it was well predicted by the remaining indicators and possible omission from the indicator set should be considered.

3.2.10 Indicator aggregation

3.2.10.1 Evaluation of aggregation functions

The handling of compensatory effects when aggregating indicators can be divided into two basic approaches: either attempt to deal with all the compensatory effects in detail, or apply some generic adjustment to the aggregation that is commensurate with the extent to which compensatory effects are believed to be occurring.

The amount of information needed to specify compensatory effects between indicators increases rapidly with the number of indicators. For example, for two indicators A and B it is only necessary to specify the interaction between A and B. However, for three indicators, A, B and C, it necessary to specify the interactions between A and B, A and C and B and C. For n indicators, it is necessary to specify $_nC_2$ two way interactions. Given the Australian Natural Disaster Resilience Index is a top-down index derived from available secondary data, with a limited budget that precludes eliciting information about indicator interactions from experts, aggregation with full specification of compensatory effects was confined to groups of three or two indicators, where it was possible to make plausible assumptions about interactions. The aggregation of groups of four or more indicators, being prohibitively demanding of information about interactions, required a generic adjustment.

The constraints on the choice of aggregation functions for the Australian Natural Disaster Resilience Index are as follows:

- the index should retain as much information as possible from the indicators being aggregated;
- the aggregation function should produce an index and not a ranking or categorisation;
- the aggregation function should allow control of compensatory effects between indicators, either by detailed specification of indicator interactions or by a method of generic adjustment;
- for generic adjustments, the aggregation function should allow the amount of adjustment to be varied; and,

- the aggregation has to be able to be computed within a reasonable time.

The results of an evaluation of aggregation functions for the Australian Natural Disaster Resilience Index are shown in Table 3.2. Each column in the table refers to one of the five constraints listed above. As the Australian Natural Disaster Resilience Index is computed using the R package, the range of aggregation methods tested and evaluated were restricted to those that could be calculated from first principles in base R, and those available as functions in contributed packages.

The Benefit of Doubt (BOD) method was omitted from the aggregation functions to be evaluated as it is premised on the idea that high values of an indicator for an individual country suggests that the associated policy area is considered important in that country. Policy importance can be extended to relevant indicators and this can be used in deriving endogenous weights. This concept is not relevant to disaster resilience at SA2 level since an SA2 does not have the authority to implement policy, and disaster resilience as a goal of public policy is a very recent phenomenon in Australia, so insufficient time has elapsed for the value of indicators to be related to policy initiatives.

Three aggregation functions were rejected because of their long computation times: both forms of Kemeny Optimal Aggregation and ELECTRE 3 (Table 3.2). Eight aggregation functions were rejected as they provide a ranking or ordinal index, rather than a ratio index: sum rankings, above and below a benchmark, Borda's rule, ELECTRE TRI, TOPSIS, RIM, MOORA and VIKOR (Table 3.2). The linear sum or mean was rejected as an aggregation function because of its assumption of unlimited compensability. The Mazziotta-Pareto Index was rejected since the algorithm that penalises the composite index for unbalance has no parameter to control the magnitude of the penalty.

After rejecting these aggregation functions, five functions remain: the generalised mean (together with its special case, the geometric mean), the mean-min function, the WASPAS function, the ordered weighted average (OWA) and the discrete Choquet integral (Table 3.2). Only one of the five functions, the discrete Choquet integral, allows for detailed specification of indicator interactions, and this was selected for use in the Australian Natural Disaster Resilience Index wherever groups of two or three indicators were to be aggregated and it was possible to make plausible estimates of the indicator interactions.

Table 3.2: Evaluation of potential aggregation methods to be used in the Australian Natural Disaster Resilience Index.

Aggregation method	Discards information?	Produces a ratio index?	Controls compensatory effects?	Compensation adjustable?	Computation time
Aggregation methods from the composite index tradition (e.g. OECD 2008)					
Linear sum or mean	No	Yes	No	NA	Negligible
Sum rankings	Yes	No	No	NA	Negligible
Above and below a benchmark	Yes	No	No	NA	Negligible
Generalised mean (includes geometric mean)	No	Yes	Generically	Yes, via parameter β	Negligible
Mean-minimum function	No	Yes	Generically	Yes, via parameters α and β	Negligible
Mazziotta-Pareto Index	No	Yes	Generically	No	Negligible
Borda's rule	Yes	No.	No	NA	Acceptable
Aggregation methods from information science (e.g. Dwork et al. 2001)					
Kemeny Optimal Aggregation with cross entropy Monte Carlo algorithm	Yes	No	Generically	No	Prohibitively long
Kemeny Optimal Aggregation with genetic algorithm	Yes	No	Generically	No	Prohibitively long
Aggregation methods from Multi-Criteria Decision Analysis (e.g. Figueira et al. 2005)					
ELECTRE 3	No	No	Specifically	Yes, in prohibitively extensive detail	Inconveniently long
ELECTRE TRI	No	No	Specifically	Yes, in prohibitively extensive detail	Acceptable
WASPAS	No	Yes	Generically	Yes, via parameter λ	Negligible
TOPSIS	No	No	Generically	Yes, via indicator weights	Negligible
RIM	No	No	Generically	Yes, via indicator weights	Negligible
MOORA	No	No	Generically	Yes, via indicator weights	Negligible
VIKOR	No	No	Generically	Yes, via parameter ν	Negligible
Aggregation methods from the theory of aggregation functions (e.g. Grabisch et al. 2011)					
Ordered weighted average (OWA)	No	Yes	Generically	Yes, via parameter *orness*	Negligible
Discrete Choquet integral	No	Yes	Specifically	Yes, via fuzzy measure	Negligible

The remaining four functions all allow the control of compensatory effects by a generic adjustment and are suitable for aggregating four or more indicators. The functions also have a parameter or parameters that can be varied to control the extent to which compensatory effects are allowed in aggregation.

An aggregation function must be able to accommodate the full range of compensatory effects, from allowing uncontrolled compensation (as occurs with the arithmetic mean as an aggregation function), through to allowing no compensation (as occurs with the minimum as an aggregation function).

The generalised mean, for a value of parameter β equal to 1, is equivalent to the arithmetic mean, and so allows for uncontrolled compensatory effects if required. However, for equivalence to the minimum function, the parameter β has to be set to -∞ which, for aggregation with limited compensatory effects, makes the choice of suitable values of β difficult. Further, the relationship between β and the position of the generalised mean between the arithmetic mean and the minimum is strongly non-linear. For this reason, the generalised mean was rejected as a generic aggregation function for the Australian Natural Disaster Resilience Index.

The WASPAS function is equivalent to the geometric mean when λ = 0 and is equivalent to the arithmetic mean when λ = 1. This means it is not capable of approaching the minimum function for an aggregation situation where no compensatory effects are to be allowed. For this reason, the WASPAS function was rejected as a generic aggregation function for the Australian Natural Disaster Resilience Index.

The remaining two generic aggregation functions, the ordered weighted average (OWA) and the mean-min function both allow the control of compensatory effects from unlimited compensation (the arithmetic mean function) through to no compensation (the minimum function). However, the mean-min function as proposed by Tarabusi and Guarini (2013) uses two parameters, α and β, where α is an unbalance penalty and β sets the extent to which indicators may substitute for each other in aggregation. The concept of unbalance among indicators originates in the economic development literature, where it is accepted that a mixture of very high and very low values of indicators representing various factors driving economic development may be deleterious to development – hence the application of an unbalance penalty in the compution of a composite index of economic development (see, for example, Tarabusi and Guarini 2013).

It would appear that the concept of indicator unbalance is either not relevant to, or has not yet been considered for vulnerability or resilience to natural disasters. Unbalance is not referred to in any of the 106 composite index methodologies surveyed by Beccari in 2016. For this reason, the mean-min function was rejected as a generic aggregation function for the Australian Natural Disaster Resilience Index.

The outcome of the evaluation above was that three aggregation functions were chosen for use in the Australian Natural Disaster Resilience Index:

- the arithmetic mean where the aggregation strategy involved a reflective measurement model;
- the ordered weighted average (OWA) where the aggregation strategy involved a formative measurement model with four or more indicators to be aggregated; and,
- the discrete Choquet integral where the aggregation strategy involved a formative measurement model with two or three indicators or sub-indices to be aggregated.

The use of the discrete Choquet integral depended on the plausible specification of compensatory interactions between indicators. Where this was not possible, OWA was used instead.

3.2.10.2 Choice of aggregation strategy

Aggregation strategy refers to the approach taken in the aggregation calculation for a composite index or sub-index, and is defined by:

- the type of measurement model assumed – formative or reflective;
- the number of stages or levels of aggregation; and,
- the aggregation functions used.

There are two steps to selecting the aggregation strategy used in the Australian Natural Disaster Resilience Index. First, a sequence of considerations and decisions was used to select a relevant measurement model (Figure 3.6).

Second, the aggregation strategy was identified. The Australian Natural Disaster Resilience Index has a hierarchical structure (see Chapter 1). There are eight theme sub-indices, which are aggregated to form the coping and adaptive capacity sub-indices, which are aggregated to form the Australian Natural Disaster Resilience Index. For each of the Australian Natural Disaster Resilience Index sub-indices, an aggregation strategy was chosen from among the four models shown in Figure 3.7.
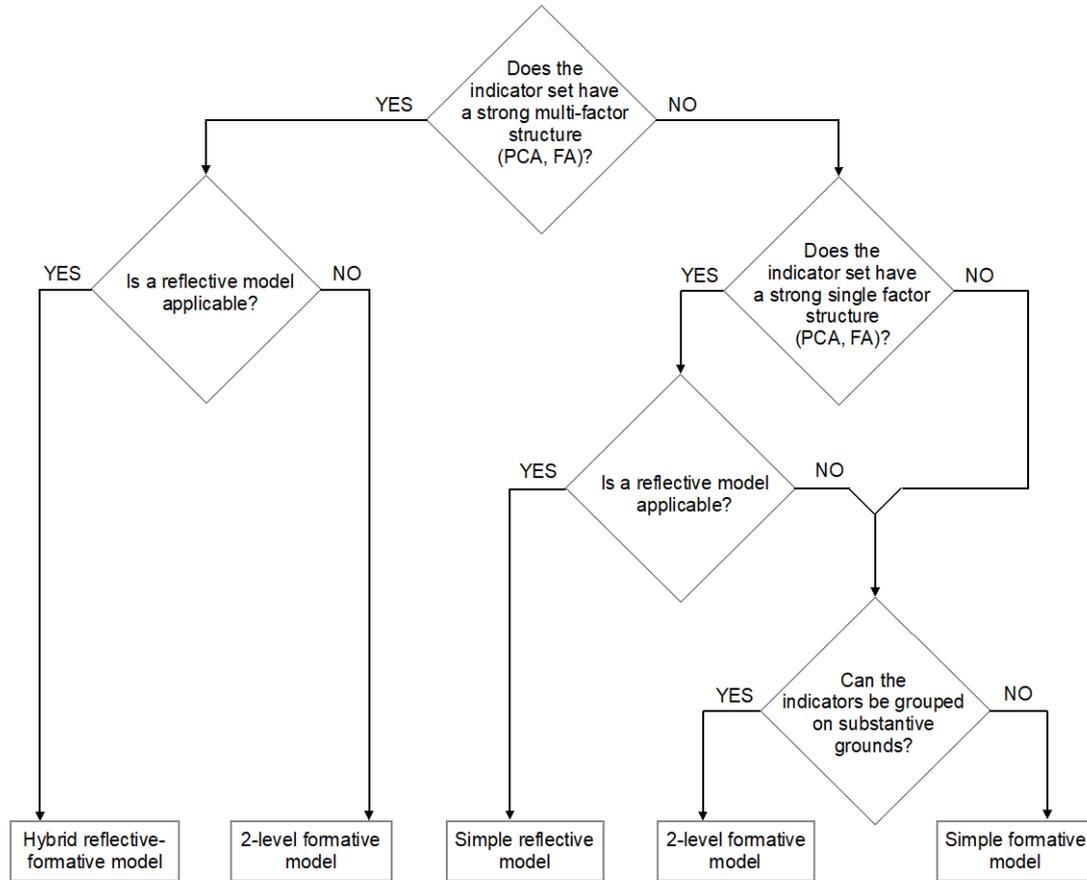
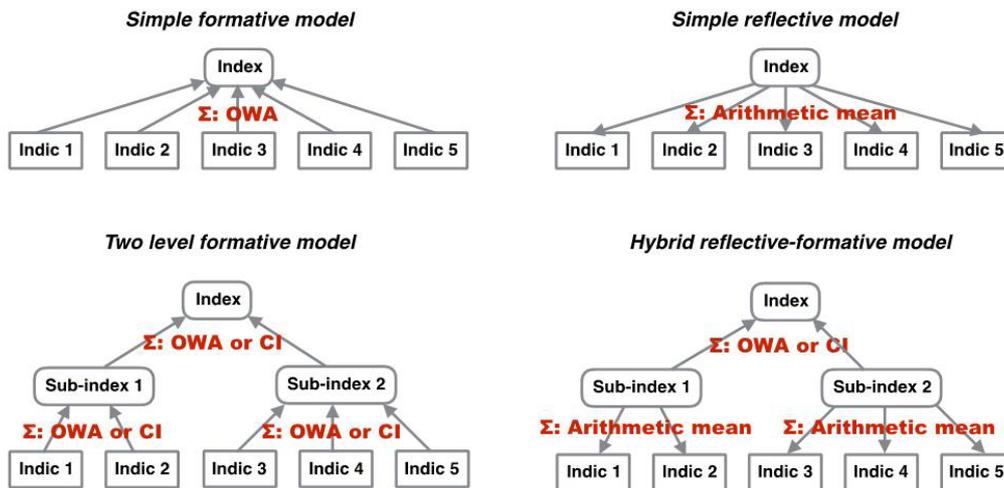**Figure 3.6:** Decision tree for choice of measurement model.

Figure 3.7: Aggregation strategies considered and/or tested in the computation of the Australian Natural Disaster Resilience Index. Arrows denote the direction of causation implicit in the model. Options for the aggregation method within each model are shown in red. OWA = Ordered Weighted Average, CI = Choquet Integral.

The starting point for the choice of measurement model is the correlation level plot (Figure 3.5). If there are one or more blocks of highly correlated indicators along the diagonal, and low correlations between indicators elsewhere (strong factor structure), then the next consideration is whether a reflective measurement model might be applicable. This model can be applied when it can plausibly be argued that the block of correlated indicators have values that are caused by some latent and directly immeasurable characteristic of SA2s. For example, it might be hypothesised that community cohesion is a latent characteristic of a community that will determine the level of volunteering, participation in local working bees and the amount of bartering of goods and services. If the three indicators are highly inter-correlated, then their arithmetic mean is a measure of the latent characteristic of community cohesion. Note that, with the high inter-correlation, there are minimal compensatory effects, so the arithmetic mean is an appropriate aggregation function.

If there is no causal justification for a reflective model, then a formative model, in which causation flows from the indicators to the index, is considered. The formative model corresponding to the example above would be an index of community cooperative behaviour, the value of which is influenced by the incidence of volunteering, working bee participation and bartering. Compensatory effects are now relevant, particularly if inter-correlations between indicators are not high. Can bartering substitute for volunteering, can working bee participation substitute for either?

As shown in Figure 3.6, consideration has to be given to sets of indicators that relate to a single index (either in a formative or reflective model), and sets of indicators that capture two or more distinct dimensions in the correlation structure. The presence of multiple dimensions is signaled by two or more blocks of high inter-indicator correlations in the correlation level plot. For example, in Figure 3.5, there are three dimensions: one related to income, one related to house and car ownership and one related to the local economy. The dimensional structure can be confirmed with factor analysis (if a reflective model is intended) or principal components analysis (if a formative model is intended). The higher the proportion of variance captured by the factors (components), the stronger the case for a two-level or hybrid model where each factor (component) is represented by a sub-index and the sub-indices are aggregated to form a single index. (Figure 3.7). The main advantage of structuring an aggregation of a large number of indicators as a two-level or hybrid model is that it allows for more nuanced control of compensatory effects, provided these are sufficiently well understood, rather than using a single aggregation with some generic adjustment for compensability.

Having selected an aggregation strategy, the final stage is to select aggregation functions appropriate to the context. The selection for the Australian Natural Disaster Resilience Index was from the discrete Choquet

integral, OWA or the arithmetic mean, depending on the measurement model, the number of indicators or sub-indices to be aggregated and the level of knowledge about compensatory effects among them.

3.2.10.3 Parameters for control of compensatory effects in aggregation - OWA

For the Australian Natural Disaster Resilience Index, aggregation of four or more indicators in formative models was done with ordered weighted averaging (OWA).

The parameter by which compensatory effects are adjusted in OWA is a weighting vector, of length equal to the number of indicators and sum equal to 1. Suppose that four indicators, A, B, C and D are being aggregated, using a weighting vector of {**0.25, 0.25, 0.25, 0.25**}. For a particular SA2, the indicator values (rescaled to range 0 – 1) for A, B, C and D are **0.6, 0.1, 0.9, 0.4**, respectively. The ordered weighted average (OWA) is obtained by ordering the indicator values from smallest to largest, viz. B, D, A, C, multiplying each value by the corresponding element of weighting vector and summing. This is:

$$0.25 \text{ x } 0.1 \text{ } + \text{ } 0.25 \text{ x } 0.4 \text{ } + \text{ } 0.25 \text{ x } 0.6 \text{ } + \text{ } 0.25 \text{ x } 0.9 = 0.5$$

So OWA with a weighting vector of {0.25, 0.25, 0.25, 0.25} is equivalent to the arithmetic mean, which allows unconstrained compensatory effects.

Consider a weighting vector of {1, 0, 0, 0}. The OWA for the four indicators is now 0.1 – equivalent to the minimum function, which allows no compensatory effects at all.

The extent of the constraint on compensatory effects is summarised in a single parameter, known as the orness, which is defined as:

$$\text{orness}(\mathbf{w}) = \sum_{i=1}^{n} w_i \frac{i\text{-}1}{n\text{-}1}$$

where $\mathbf{w}$ is a weighting vector of length n.

For the weighting vector of {0.25, 0.25, 0.25, 0.25}, the orness is:

$$0.25 \text{ x } 0/3 \text{ } + \text{ } 0.25 \text{ x } 1/3 \text{ } + \text{ } 0.25 \text{ x } 2/3 \text{ } + \text{ } 0.25 \text{ x } 3/3 \text{ } = \text{ } 0.5$$

For the weighting vector of {1, 0, 0, 0} the orness is zero.

In general, OWA with a weighting vector that has an orness of 0.5 is equivalent to the arithmetic mean (unrestrained compensation), while OWA with a weighting vector with an orness of zero is equivalent to the minimum function

(no compensation). Weighting vectors with orness values between zero and 0.5 give aggregation with partial constraint of compensatory effects.

For the Australian Natural Disaster Resilience Index, the extent to which high values of some indicators could be allowed to compensate for low values of other indicators was known only approximately, or not at all. Consequently, just two orness values were used in aggregations using OWA: 0.125 for situations where there was some certainty that only minimal compensatory effects should be allowed, and an orness of 0.375 for situations where it was reasonable to assume that substantial amounts of compensation were permissible in aggregating indicators. An example of the former is indicators relating to emergency services provision – it is unlikely that high numbers of fire service volunteers could substitute for low numbers of police. An example of the latter is indicators relating to communications – high levels of mobile phone coverage could, in greater part, substitute for low levels of ADSL connectivity, given the widespread ownership of smartphones.

A weighting vector has a unique orness value, but one orness value does not define a unique weighting vector. Many weighting vectors can have the same orness value. For example {0.5, 0.3, 0.2, 0} and {0.55, 0.26, 0.13, 0.06} both have an orness of 0.23. For this reason, for repeatability in computation of the Australian Natural Disaster Resilience Index, it is not enough to specify that the two orness values of 0.125 and 0.375 were used in OWA aggregations. One way to ensure repeatability is to also specify a function that shapes the pattern of elements in the weighting vector. After experimenting with several functions, it was found that an exponential function provided a repeatable and intuitively satisfying way of specifying a unique weighting vector from an orness value. For a weighting vector of length n, and a desired orness of DO, the weighting vector is:

$$\{C^n, C^{(n-1)}, C^{(n-2)}, \dots C^2, C\}$$

where C is a constant such that the orness of $\{C^n, C^{(n-1)}, C^{(n-2)}, \dots C^2, C\}$ = DO. In this expression, superscripts denote exponents. Constant C can readily calculated using a root finding algorithm (such as *uniroot* in R) with:

$$DO - orness(\{C^n, C^{(n-1)}, C^{(n-2)}, \dots C^2, C\})$$

For example, the weighting vector for OWA of 4 indicators, where an orness of 0.125 has been chosen to limit compensatory effects in the aggregation is {0.72, 0.21, 0.06, 0.02}. This weighting vector is obtained when C in the expression above takes the value 3.49.

For an orness of 0.375, the weighting vector is {0.37, 0.28, 0.20, 0.15}. This weighting vector is obtained when C in the expression above takes the value 1.36. These weighting vectors are uniquely defined by the orness value and the value of C.

3.2.10.4   Parameters for control of compensatory effects in aggregation – discrete Choquet integral

For the Australian Natural Disaster Resilience Index, where two or three indicators or sub-indices in a formative model were to be aggregated, consideration was first given to using the discrete Choquet integral. If knowledge of the compensatory effects between, or among, these was insufficient then OWA was used instead.

The parameter by which compensatory affects are adjusted in aggregating with the discrete Choquet integral is called the fuzzy measure and is a set of weighting values. In the context of indicator aggregation, the fuzzy measure weights can be determined by the consideration of bounding cases. With rescaled indicators, the bounding cases are those sets of indicator values where one or more indicators take the values 0 or 1.

Table 3.3 shows the full range of bounding cases for three indicators: ADSL coverage, mobile phone coverage and information availability, that are to be aggregated to give an index of information access. The indicators have been rescaled to range 0 – 1, where 0 represents the minimum value of the indicator and 1 represents the maximum value of the indicator.

Table 3.3: Example of specifying a fuzzy measure for aggregation by discrete Choquet integral of a set of three indicators.

Boundary indicator values			Desired value of index
ADSL	Mobile	Information	
0	0	0	0.0
1	0	0	0.1
0	1	0	0.2
0	0	1	0.1
1	1	0	0.2
1	0	1	0.5
0	1	1	0.7
1	1	1	1.0

Consideration is given to what the index should be for each of these bounding cases. Obviously, the first row, (0, 0, 0), should aggregate to an index of 0, and the last row, (1,1,1), should aggregate to an index of 1. The second to fourth rows of the table show what it is believed the index should be for each of the indicators when they have value 1 while the remaining two have value 0. The values in the table assume that mobile phone coverage by itself gives much better information access than either ADSL coverage or information availability by themselves.

The fifth to seventh rows of Table 3.3 shows what the index should be for pairs of indicators with the value 1 and the remaining indicator with value 0. The values assigned to the index assume that good mobile coverage and good information availability alone should give better information access than good ADSL coverage and good information availability. This, in turn, is assumed to be better than just good mobile and ADSL coverage, with poor information availability.

Using the notation of James (2016), the fuzzy measure weights, *v*, from Table 3.3 are:

> *v*{ADSL, Mobile, Information} = 1
>
> *v*{Mobile, Information} = 0.7; *v*{ADSL, Information} = 0.5; *v*{ADSL, Mobile} = 0.2
>
> *v*{Information} = 0.1; *v*{Mobile} = 0.2; *v*{ADSL} = 0.1
>
> *v*{} = 0

The fuzzy measure weights can equivalently be interpreted as measures of importance. For example, among the single factors affecting information access, mobile phone coverage is more important than either ADSL coverage or information availability. This might reflect mobile's portability, immediacy and ability to receive SMS messages from emergency services. In addition, compensatory effects can be specified by the fuzzy measure weights. In the example above, *v*{ADSL, Mobile} has the same value as *v*{Mobile}, indicating that ADSL and Mobile can substitute for each other.

For a set **x** of *n* indicator values for a particular SA2, and the fuzzy measure weights *v*, above the discrete Choquet integral is given by:

$$C_v(\mathbf{x}) = \sum_{i=1}^{n} x_{(i)} \Big(v(\{(i){:}(n)\}) - v(\{(i{+}1){:}(n)\}) \Big)$$

where *x(i)* denotes the *i*th indicator when the indicator values are ordered from the smallest to the largest.

To illustrate, suppose the set of indicator values is ADSL=0.2, mobile=0.8 and information=0.7, representing an SA2 with poor ADSL coverage, very good mobile coverage and good information availability. The indicator order with the values ordered from smallest to largest is ADSL, Information, Mobile. The first term in the summation formula above will be:

> 0.2 x (*v*{ADSL, Mobile, Information} – *v*{Mobile, Information}) = 0.2 x (1-0.7)

The second term will be:

0.7 x (v{Mobile, Information} – v{Mobile}) = 0.7 x (0.7-0.2).

The third term will be:

0.8 x (v{Mobile} – v{}) = 0.8 x (0.2–0).

The discrete Choquet integral is the sum of these three terms, viz.:

0.06 + 0.35 + 0.16 = 0.57

The fine control over compensatory effects given by the discrete Choquet integral can be illustrated with the aggregation of various combinations of indicator values using the fuzzy measure weights, v, above. The indicator values are ordered ADSL, Mobile, Information.

Cv(0.9, 0.9, 0.1) = 0.26

The low value of the composite index reflects the fact that, regardless of good ADSL and mobile connectivity, information accessibility overall is poor if there is little information available to access.

Cv(0.1, 0.9, 0.9) = 0.66

The high value of the composite index reflects the good information accessibility if mobile coverage is good and there is good information availability. The low ADSL coverage does not affect the overall good information accessibility because ADSL and Mobile have been specified as substitutes in the fuzzy measure weights.

Cv(0.9, 0.1, 0.9) = 0.50

Although mobile coverage is poor, the information availability index is still fairly high, because ADSL can substitute for mobile coverage. However, it is not as high as the previous example, since the fuzzy measure weights specify Mobile by itself as better than ADSL by itself. Note that the arithmetic mean for all three examples is 0.63, completely missing the importance and compensatory relationships among the three indicators.

3.2.10.5   Validation of aggregation choices

The final stage in the aggregation of indicators and sub-indices for the Australian Natural Disaster Resilience Index was to compare the composite index obtained with the chosen aggregation method with the index obtained with the arithmetic mean (unrestrained compensability) and with several methods that have been proposed as partly non-compensable alternatives to the arithmetic mean, viz. the geometric mean and the Mariotta-Pareto index.

If the discrete Choquet integral was used for aggregation, then aggregation by OWA was added to the list of comparison methods.

The purpose of the comparison was to ensure that the composite index values calculated with the chosen method were not substantially inconsistent with the values with other methods, and that any departures in the values with the chosen method from the values with other methods were consistent with the mathematics of the aggregation functions involved. Individual SA2s with large differences in composite index values across the various aggregation methods were checked and it was invariably found that the large differences were associated with SA2s with substantial variation in values across the indicators being aggregated. The variation in indicator values meant that compensatory effects were in full play, so the different ways the aggregation methods dealt with these effects led to markedly different values for the composite index.

An example of the comparison of aggregation methods (the emergency services index) shows the results for a two level formative model (2 level with aggregation by OWA and discrete Choquet integral) and single level models with aggregation by OWA, geometric mean, Mazziotta-Pareto Index and arithmetic mean (Figure 3.8). As expected, the use of OWA with a low value of orness results in considerable separation between the various aggregation methods, with the two level formative model using OWA and the discrete Choquet integral producing the lowest values of the sub-index, apart from the geometric mean. The latter takes the value zero whenever one or more of the constituent indicators has the value zero. The arithmetic mean, because it allows unrestrained compensation when indicators are a mixture of high and low values, has the highest values of the sub-index. The Mazziotta-Pareto Index, with its fixed unbalance penalisation, severely reduces the value of the sub-index when the coefficient of variation for the indicators is high.
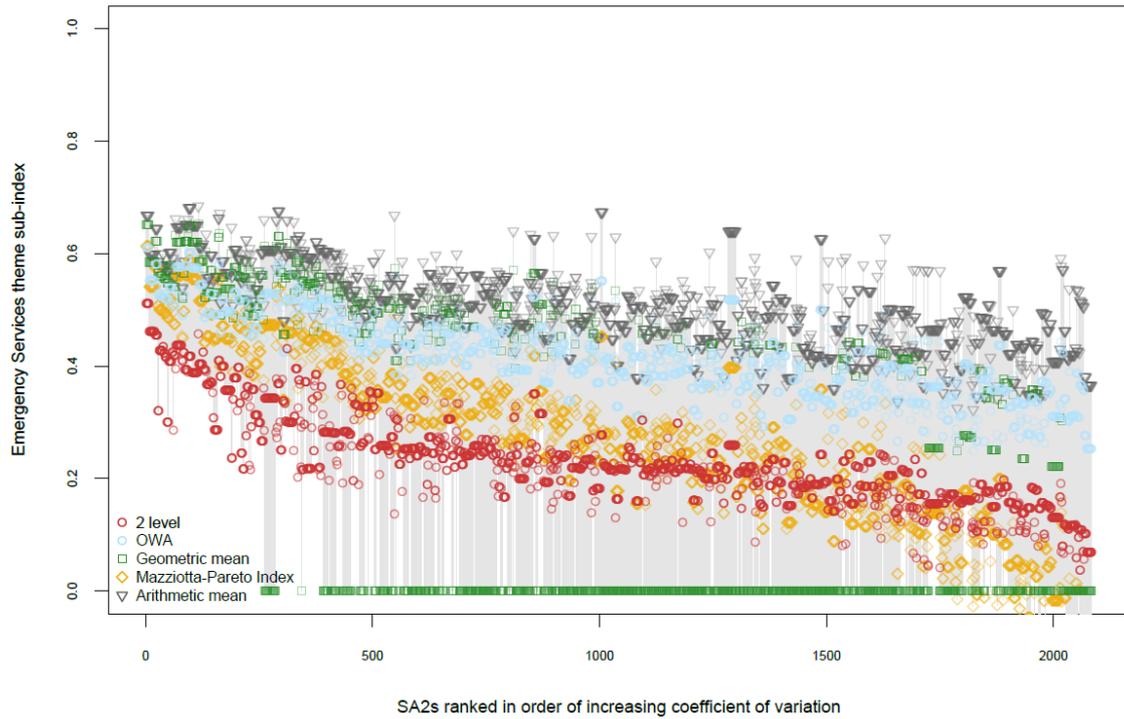
Figure 3.8: Example of comparison of the chosen aggregation method (a two level formative model with aggregation by OWA and discrete Choquet integral (denoted as 2 level in the key) with alternative methods of aggregation. SA2s are ordered according to the coefficient of variation of their indicators, lowest to the left and highest to the right.

3.2.10.6   Weighting for correlation in aggregating sub-indices

When aggregating the theme sub-indices to form the coping capacity and adaptive capacity sub-indices, the possibility arises that some theme sub-indices may share similar or correlated indicators. If this is the case, then the dimension represented by these indicators will have an unduly increased influence on the sub-index obtained by the aggregation of themes and (implicitly) their constituent indicators. One approach for ameliorating this undesired influence that has been reported in the literature is to use simple weighting of the indicators (or sub-indices) in the aggregation step, where weights are set to be the inverse of the correlations (Hudrlikova 2013; OECD 2008; Saisana et al. 2005; Tate 2013).

This issue only arose for one of the aggregation steps in the construction of the Australian Natural Disaster Resilience Index, viz. the aggregation of six theme sub-indices to produce the coping capacity sub-index. Two sub-indices had a correlation of 0.65 – social character and community capital. The remaining sub-index inter-correlations were relatively low. The influence of this social dimension on the coping capacity sub-index was reduced by the simple approach of replacing the two sub-indices with their mean. Given the relatively high correlation between them, compensability issues will be minimal,

and the overall effect is similar to that achieved with weighting as described above.

3.3   DERIVATION OF TYPOLOGY GROUPS

The goal of the typology was to extract a coherent grouping of SA2s with similar disaster resilience profiles. Any particular Australian Natural Disaster Resilience Index value can result from many different combinations of theme sub-index values. Patterns in the sub-index values can provide a meaningful context within which Australian Natural Disaster Resilience Index values can be interpreted and the implications for disaster resilience explored. Cluster analysis was used to extract groups of SA2s with similar disaster resilience profiles. The use of cluster analysis assumes that there are not infinitely many patterns of sub-index values, rather that there are a limited number of such patterns and these fall into groups.

3.3.1   Cluster analysis

The Australian Natural Disaster Resilience Index is constructed hierarchically (see Chapter 1). Each level in the hierarchy could potentially be used as variables in cluster analysis. Several of the levels can be ruled out immediately. First, the indicator level is unsatisfactory for cluster analysis due to the large number of variables (77). Further, the goal of the cluster analysis was to provide a context to the geographical variation in the Australian Natural Disaster Resilience Index in terms of the broad conceptual factors that are known to influence disaster resilience, rather than highly specific characteristics that are captured by a single indicator. Second, the coping and adaptive capacity level in the Australian Natural Disaster Resilience Index hierarchy was ruled out for its lack of cluster structure arising from the two dimensions (Figure 3.9).
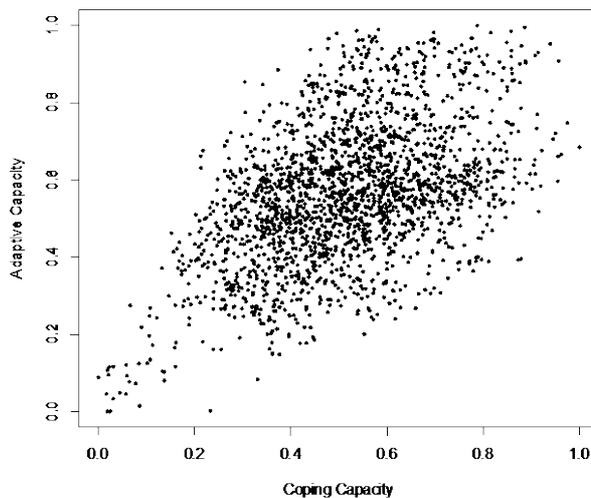
Figure 3.9: Scatterplot of coping and adaptive capacity index values demonstrating lack of cluster structure.

///

Partitioning around medoids confirms this with a maximum silhouette coefficient of 0.19 for the three cluster solution. This is well below the threshold of 0.5, above which fair to good cluster structure is indicated. The low silhouette coefficient is supported by a low cophenetic correlation for hierarchical agglomerative clustering. Values of 0.53 (Ward's method) and 0.40 (complete linkage) are well below the threshold of 0.75, above which fair to good cluster structure is indicated.

If the indicator level and the second top level of the Australian Natural Disaster Resilience Index hierarchy are unsuitable for cluster analysis, this leaves the eight theme sub-indices. A pairs plot for these sub-indices suggests some cluster structure is present (Figure 3.10), although some of this is an artefact of the disaggregation of State level indicators to SA2 level, e.g. information access.
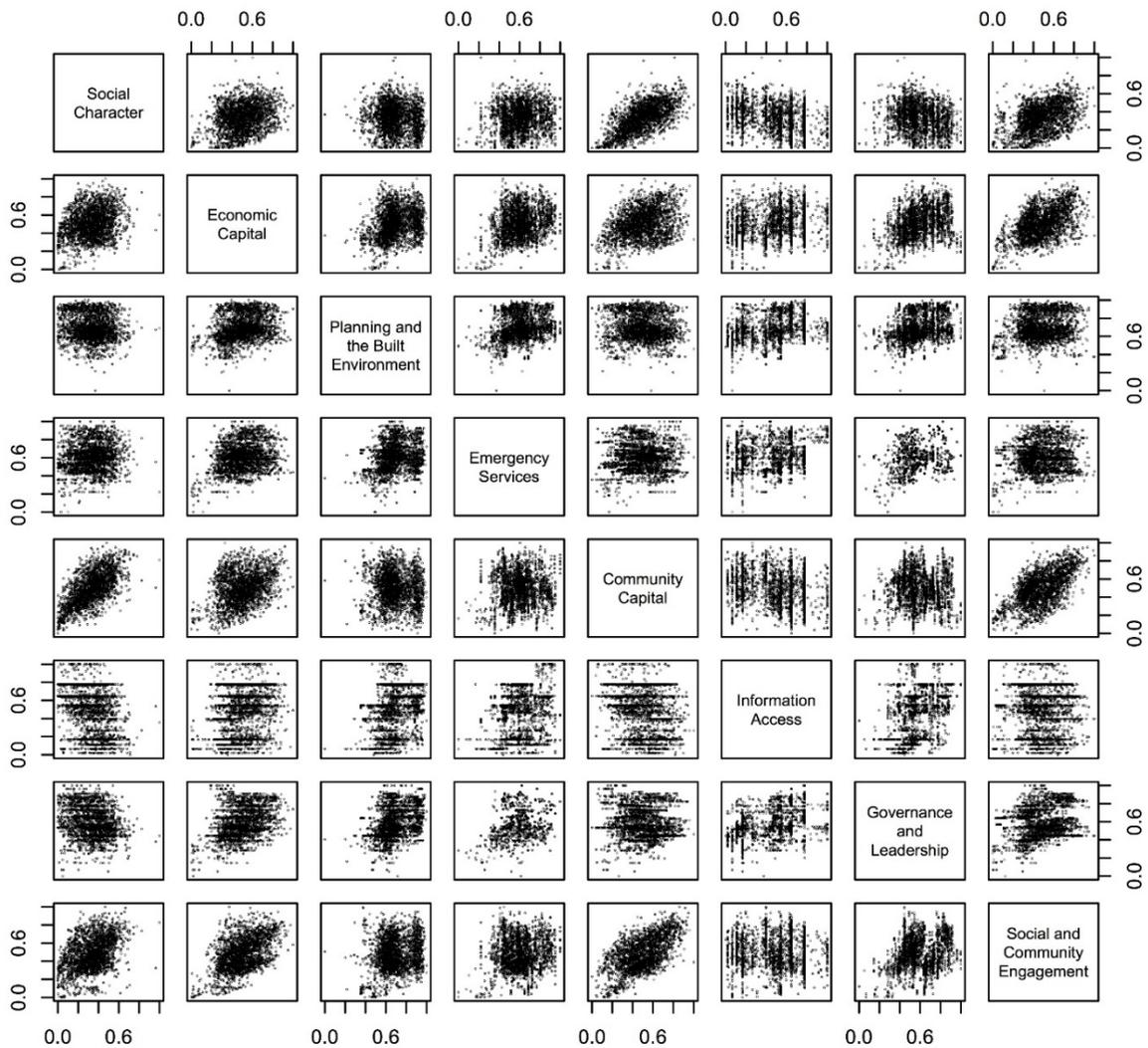

Figure 3.10: Theme level paired scatterplots showing potential cluster structure.

///

Exploratory cluster analysis using several different methods confirmed that the cluster structure is weak. Partitioning around medoids (a variant of the k-means method) gave a silhouette coefficient less than 0.2 for any number of clusters between 2 and 12 (Figure 3.11). The cluster structure was best represented by a three cluster solution (maximum silhouette coefficient).

Two hierarchical agglomerative cluster functions available in R (hclust and agnes) were trialled with two clustering methods, complete linkage and Ward's method. These gave cophenetic correlation coefficients ranging from 0.38 to 0.53, well below the threshold of 0.75, above which fair to good cluster structure is indicated. In view of the weakness of the cluster structure, the scree plots and profiles for a number of clustering techniques were compared (Figure 3.11).

There is some suggestion in the silhouette plot and the scree plots (Figure 3.11) that the cluster structure might be represented by three, five or nine cluster solutions. Gaussian mixed modelling did not give any indication that a model could be fitted to the data. The five cluster solution was chosen for further investigation simply on communication grounds. A nine cluster solution overloads audience capacity for information. The three cluster solution is unnecessarily parsimonious and the five cluster solution strikes a balance between the two.
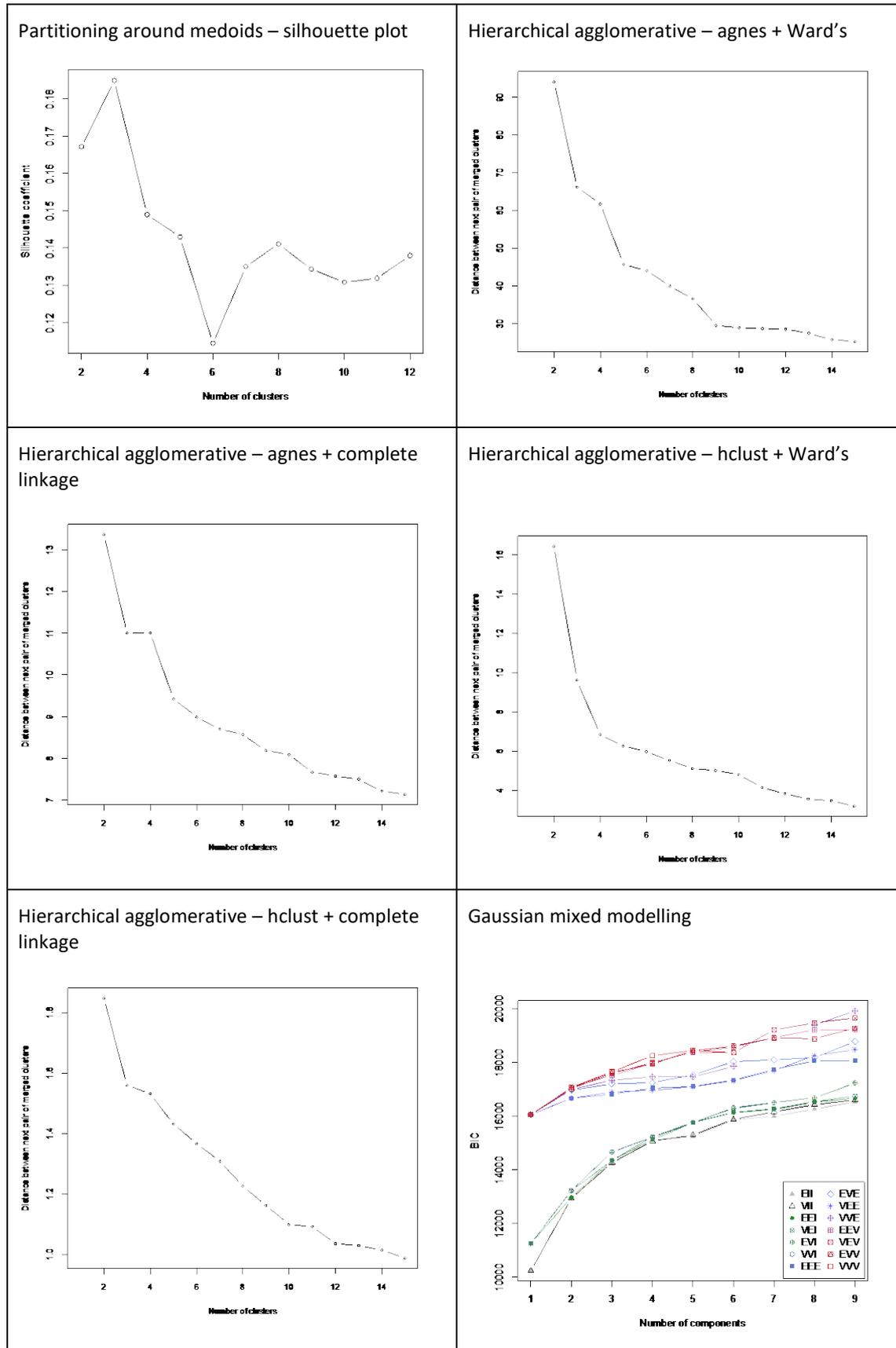
Figure 3.11: Comparison of cluster outcomes from different clustering methods.

3.3.2 Choice of clustering method

Given the weak cluster structure, it is likely that different cluster methods will lead to substantially different cluster solutions. To aid in the choice of a solution that would meet the goal of the cluster analysis, viz. to identify a coherent grouping of SA2s that would give some context to the geographical variation in disaster resilience in the Australian Natural Disaster Resilience Index, three methods of external validation were employed.

The remoteness score is available as a validation variable, i.e. one that was not used in the cluster analysis but is relevant to the assessment of cluster solutions against the goal of the cluster analysis. One method of external validation was to test the difference in the mean remoteness scores for each of the possible clustering methods. The logic behind this approach is that larger differences in mean remoteness scores will imply more geographically coherent clusters. The Kruskal-Wallis rank sum test was used to test the hypothesis that there was no difference in the mean remoteness score among the five clusters (in all cases Bartlett's test indicated significant inhomogeneity of variance among the clusters). The results show that the greatest difference between mean remoteness scores among the five clusters occurs with partitioning around medoids (Table 3.4).

Table 3.4: Kruskal Wallis tests of differences in remoteness scores among clusters, for different clustering techniques.

Clustering method (5 cluster solution)	Kruskal-Wallis χ^2	D.f.	P value
Hierarchical agglomerative – hclust + Ward's	768.9	4	4.2e-165
Hierarchical agglomerative – hclust + complete linkage	644.8	4	3.1e-138
Hierarchical agglomerative – agnes + Ward's	680.5	4	5.8e-146
Hierarchical agglomerative – agnes + complete linkage	555.8	4	5.6e-119
Partitioning around medoids	925.2	4	5.9e-199

A second, more subjective, method of external validation is to plot augmented heatmaps (Figure 3.12). In assessing the extent to which the SA2s form homogenous groups of themes corresponding to the clusters, there is variation among the five clustering methods. The heatmap has relatively few homogenous blocks of purple or green corresponding to particular clusters, which is a reflection of the weak cluster structure discussed above. Hierarchical agglomerative, R:agnes complete linkage (Figure 3.12) is probably the worst with regard to the segmentation by colour and height in the bar plot, and hierarchical agglomerative, R:hclust complete linkage (Figure 3.12)is not far behind. Partitioning around medoids is certainly no worse than the remaining two panels, and possibly slightly better (Figure 3.12).

Hierarchical agglomerative, R:hclust complete linkage

Hierarchical agglomerative, R:hclust Ward's

Hierarchical agglomerative, R:agnes complete linkage

Hierarchical agglomerative, R:agnes Ward's
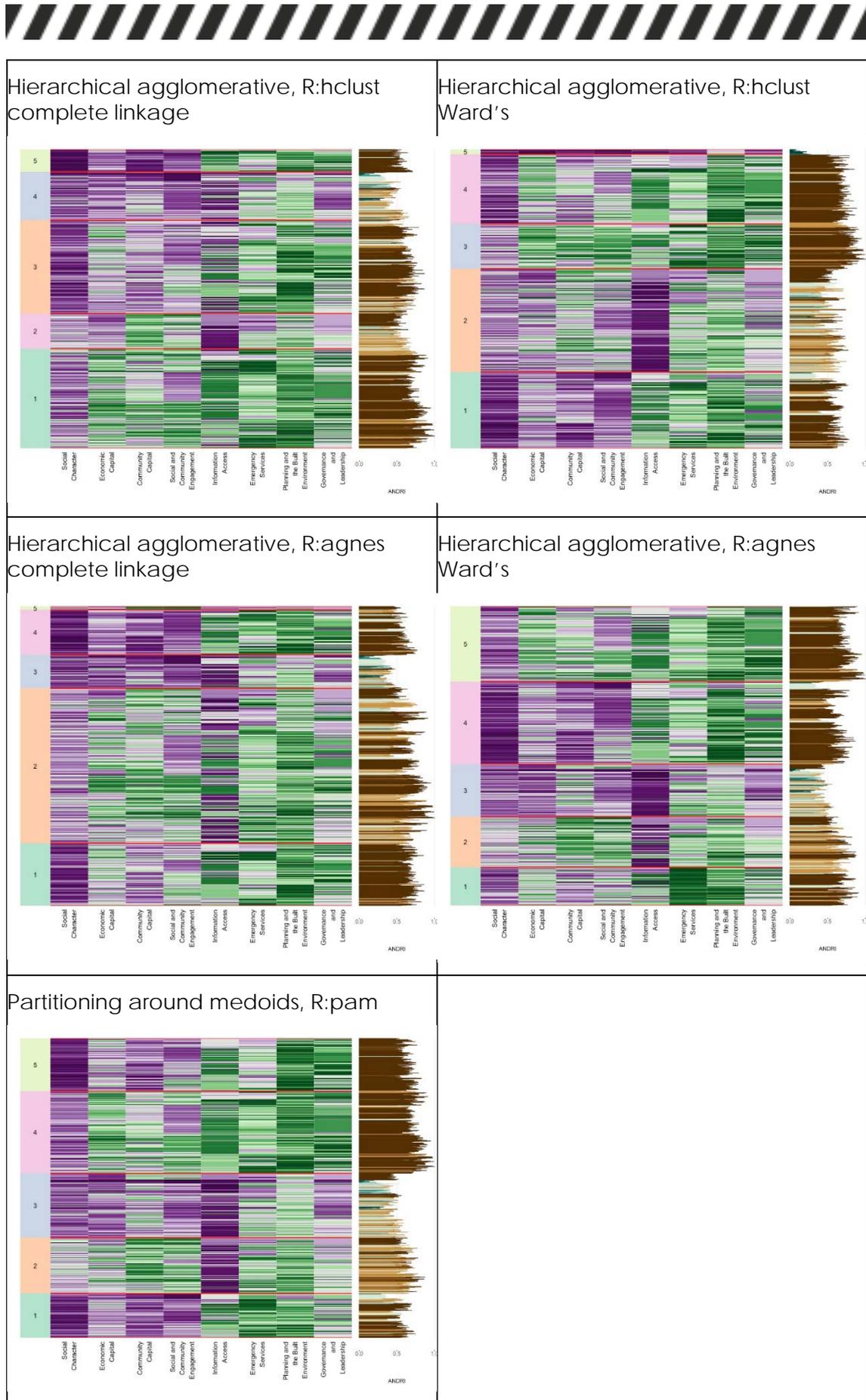
Partitioning around medoids, R:pam

Figure 3.12: Caption on following page

Figure 3.12: Heatmaps of different cluster solutions. The y axis corresponds to cluster groups formed in each method. The x axis corresponds to the eight disaster resilience themes, colour coded by index value, where purple is a low value on the theme sub-index, and green is a high value. The bar plot on the right hand y axis corresponds to the overall Australian Natural Disaster Resilience Index value for each SA2, colour coded by remoteness score, where brown is metropolitan, through to green being very remote.

For these reasons, the cluster assignment of SA2s by partitioning around medoids was chosen as the preferred method for achieving the goal of identifying a coherent grouping of SA2s that would give some context to the geographical variation in the Australian Natural Disaster Resilience Index.

The third method of external validation was applied to the partitioning around medoids cluster solution. Cluster memberships were mapped across Australia. If SA2s in the different clusters are randomly scattered across the country, this is an indication that the cluster assignment is of little value in providing a geographical context within which the Australian Natural Disaster Resilience Index values and constituent theme sub-indices can be interpreted. However, the mapping revealed that SA2s in each of the five clusters tended to form geographically cohesive regions, some of which were mostly confined to metropolitan areas, and some of which were confined to regional Australia. Maps of cluster membership are provided in Volume I.

3.4   REFERENCES

Aiello, F. and Attanasio, M. 2004. How to transform a batch of simple indicators to make up a unique one? In: Atti della XLII Riunione Scientifica della Società Italiana di Statistica, Sessioni Plenarie e Specializzate. CLEUP: Padova. Pages 327-338.

Agder, W.N., Brooks, N., Bentham, G., Agnew, M. and Eriksen, S. 2004. New indicators of vulnerability and adaptive capacity. Tyndall Centre for Climate Change Research Technical Report 7. University of East Anglia, Norwich.

Australian Bureau of Statistics. 2011. Australian Statistical Geography Standard (ASGS): Volume 1 – Main Structure and Greater Capital City Statistical Areas. Australian Bureau of Statistics, July 2011 (1270.0.55.001).

Australian Bureau of Statistics. 2013. Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2011. Australian Bureau of Statistics, Canberra (2033.0.55.001).

Bakker, M. and Wicherts, J.M. 2014. Outlier removal, sum scores, and the inflation of the Type I error rate in independent samples t tests: the power of alternatives and recommendations. Psychology Methods, 19: 409-427.

Bao, J., Li, X. and Yu, C. 2015. The construction and validation of the heat vulnerability index: A review. International Journal of Environmental Research and Public Health, 12: 7220-7234.

Baptista, S.R. 2014. Design and use of composite indices in assessments of climate change vulnerability and resilience. Center for International Earth Science Information Network, Columbia University: New York.

Barnett, J., Lambert, S. and Fry, I. 2008. The hazards of indicators: Insights from the Environmental Vulnerability Index. Annals of the Association of American Geographers, 98:102-119.

Beccari, B.A. 2016. Comparative analysis of disaster risk, vulnerability and resilience composite Indicators. PLOS Currents Disasters, 2016, March 14. doi: 10.1371/currents.dis.453df025e34b682e9737f95070f9b970

Bertin, G., Carrino, L. and Giove, S. 2018. The Italian regional well-being in a multi-expert non-additive perspective. Social Indicators Research, 135: 15-51.

Camp, S.L. and Speidel, J. 1987. The International Human Suffering Index. Washington DC: Population Crisis Committee.

Chakraborty, J., Montz, B.E. and Tobin, G.A. 2005. Population evacuation: Assessing spatial variability in geophysical risk and social vulnerability to natural hazards. Natural Hazards Review, 6: 23-33.

Cherchye, L., Moesen, W., Rogge, N. and Van Puyenbroeck, T. 2007. An introduction to 'benefit of the doubt' composite indicators. Social Indicators Research, 82: 111-145.

Christopherson, T. and Konradt, U. 2008. The Development of a formative and a reflective scale for the assessment of on-line store usability. Systematics, Cybernetics and Informatics, 6: 36-41.

Chitsaz, N. and Banihabib, M.E. 2015. Comparison of different multi criteria decision-making models in prioritizing flood management alternatives. Water Resources Management, 29: 2503-2525.

Cronbach, L.J. 1951. Coefficient alpha and the internal structure of tests. Psychometrika, 16: 297-334.

Cruciano, C., Giove, S., Pinar, M. and Sostero, M. 2012. Constructing the FEEM sustainability index: A Choquet Integral application. Nota di Lavoro 50. Fondazione Eni Enrico Mattei, Milan, Italy.

Cutter, S.L., Boruff, B.J. and Shirley, W.L. 2003. Social vulnerability to environmental hazards. Social Science Quarterly, 84: 242-261.

Cutter, S.L., Burton, C.G. and Emrich, C.T. 2010. Disaster resilience indicators for benchmarking baseline conditions. Journal of Homeland Security and Emergency Management, 7: 1-22.

Cutter, S.L., Mitchell, J.T. and Scott, M.S. 2000. Revealing the vulnerability of people and places: A case study of Georgetown County, South Carolina. Annals of the Association of American Geographers, 90: 713-737.

De Muro, P., Mazziotta, M. and Pareto, A. 2011. Composite indices of development and poverty: An application to MDGs. Social Indicators Research, 104: 1-18.

Doumpis, M. and Zoupounidis, C. 2014. Multicriteria analysis in finance. Springer: Berlin.

Dwork, C., Kumar, R., Naor, M. and Sivakumar D. 2001. Rank aggregation methods for the web. In: Proceedings of the 10th International World Wide Web Conference, Hong Kong. Pages 613-622.

Ebert, U. and Welsch, H. 2004. Meaningful environmental indices: a social choice approach. Journal of Environmental Economics and Management, 47: 270-283.

Edwards, J.R. and Bagozzi, R.P. 2000. On the nature and direction of relationships between constructs and measures. Psychological Methods, 5: 155-174.

Estes, R.J. 1984. The Social Progress of Nations. Praeger: New York.

Figueira, J., Greco, S. and Ehrogott, M. 2005. Multiple Criteria Decision Analysis: State of the Art Surveys. Springer Science and Business Media: New York.

Foa, R. and Tanner, J.C. 2012. Methodology of the Indices of Social Development (No. 2012-04). ISD Working Paper Series. Retrieved from http://hdl.handle.net/1765/50510

Fornell, C. and Bookstein, F.L. 1982. Two structural equation models: LISREL and PLS applied to consumer exit theory. Journal of Marketing Research, 19: 440-452.

Füssel, H-M. 2009. Review and quantitative analysis of indices of climate change exposure, adaptive capacity, sensitivity, and impacts. Background note to World Development Report 2010. Potsdam Institute for Climate Impact Research: Potsdam.

Gall, M. 2007. Indices of social vulnerability to natural hazards: a comparative evaluation. PhD Thesis, University of South Carolina.

Grabisch, M., Marichal, J-L., Mesiar, R. and Pap, E. 2011. Aggregation functions: construction methods, conjunctive, disjunctive and mixed classes. Information Sciences, 181: 23-43.

Guillaumont, P. 2009. An Economic Vulnerability Index: Its design and use for international development policy. Oxford Development Studies, 37: 193-228.

Heinrich, K., Huber, C. and Rinner, C. 2016. Making area-based composite indices comparable across time: the role of standardization and outlier removal. Journal of the Urban and Regional Information Systems Association, 27:1-20.

Honaker, J., King, G. and Blackwell, M. 2011. Amelia II: A program for missing data. Journal of Statistical Software, 45: 1-47.

Hudrlikova, L. 2013. Composite indicators as a useful tool for international comparison: the Europe 2020 example. Prague Economic Papers, 4: 459-473.

Hudrlikova, L. and Kramulova, J. 2103 . Do transformations matter? The case of sustainability indicators in Czech regions. Metodološki zvezki, 10: 31-48.

Jacobs, R., Smith, P. and Goddard, M. 2004. Measuring performance: an examination of composite performance indicators. CHE Technical Paper Series 29. University of York Centre for Health Economics: York, UK.

James, S. 2016. An introduction to data analysis using aggregation functions in R. Springer: Switzerland.

Jeffreys, I. 2004. The use of compensatory and non-compensatory multi-criteria analysis for small-scale forestry. Small-scale Forest Economics, Management and Policy, 3: 99-117.

Karande, P. and Chakraborty, S. 2012. Application of multi-objective optimization on the basis of ratio analysis (MOORA) method for materials selection. Materials and Design, 37: 317-324.

Kovacevic, M. 2010. Review of HDI critiques and potential improvements. Human Development Research Paper 210/33. United Nations Development Programme: New York.

Leamer, E.E. 1983. Let's take the con out of econometrics. American Economics Review, 73: 31-43.

Malczewski, J. 2000. On the use of weighted linear combination method in GIS: common and best practice approaches. Transactions in GIS, 4: 5-22.

Malczewski, J. 2006a. GIS-based multicriteria decision analysis: a survey of the literature. International Journal of Geographical Information Science, 20: 703-726.

Malczewski, J. 2006b. Integrating multicriteria analysis and geographic information systems: the ordered weighted averaging (OWA) approach. International Journal of Environmental Technology and Management, 6: 7-19.

Mandal, U.K. and Sarkar, B. 2012. Selection of best intelligent manufacturing system (IMS) under fuzzy Moora conflicting MCDM environment. International Journal of Emerging Technology and Advanced Engineering, 2: 301-310.

Marriott, F. 1990. A dictionary of statistical terms. Longman Scientific & Technical: New York.

Marozzi, M. and Santamaria, L. 2007. Composite indicators for finance. Statistica Applicata, 19: 271-278.

Mazziota, M. and Pareto, A. 2013a. Methods for constructing composite indices: one for all or all for one? Rivista Italiana di Economia Demografia e Statistica, 67: 67-80.

Mazziota, M. and Pareto, A. 2013b. Composite indices building: A comparison of unbalance adjustment methods. Advances in Latent Variables - Methods, Models and Applications. Brescia - Department of Economics and Management, June 19, 2013 – June 21, 2013.

Mazziotta, M. and Pareto, A. 2015. Comparing two non-compensatory composite indices to measure changes over time: a case study. Statistika: Statistics and Economy Journal, 2: 44-53.

Mazziota, M. and Pareto, A. 2016a. Methods for constructing non-compensatory composite indices: A comparative study. Forum for Social Economics, 45: 213-229. DOI: 10.1080/07360932.2014.996912.

Mazziotta, M. and Pareto, A. 2016b. On a generalized non-compensatory composite index for measuring socio-economic phenomena. Social Indicators Research, 127: 983-1003.

McGranahan, D.V., .Richard-Proust, C., Sovani, N.V. and Subramanian, M. 1972. Contents and measurement of socioeconomic development: A staff study of the United Nations Research Institute for Social Development (UNRISD). New York: Praeger.

Mishra, S.K. 2008. Construction of composite indices in the presence of outliers. MPRA Paper No. 8874, Munich University.

Morris D. 1978. A physical quality of life index. Urban Ecology, 3: 225-40.

Munda, G. 2012a. Choosing aggregation rules for composite indicators. Social Indicators Research, 109: 337-354.

Munda, G. 2012b. Intensity of preference and related uncertainty in non-compensatory aggregation rules. Theory and Decision, 73: 649-669.

Munda, G. and Nardo, M. 2009. Noncompensatory/nonlinear composite indicators for ranking countries: a defensible setting. Applied Economics, 41: 1513-1523.

Natoli, R. and Zuhair, S. 2011. Rediscovering the Condorcet approach as an aggregation technique for progress measures. Economic Papers: A Journal of Applied Economics and Policy, 30: 368–376.

Organisation for Economic Cooperation and Development (OECD). 2008. Handbook on constructing composite indicators: methodology and users guide. OECD: Paris, France.

Osborne, J.E. 2002. Notes on the use of data transformations. Practical Assessment, Research and Evaluation, 8: 1-10.

Osborne, J.W. and Overbay, A. 2004. The power of outliers (and why researchers should ALWAYS check for them). Practical Assessment, Research and Evaluation, 9(6): 1-8 [Online].

Paruolo, P., Saisana, M. and Saltelli, A. 2013. Ratings and rankings: voodoo or science? Journal of the Royal Statistical Society, 176: 609-634.

Pelling, M. and Uitto, J.I. 2001. Small island developing states: natural disaster vulnerability and global change. Environmental Hazards, 3: 49–62.

Pereira, D.V.S. and Mota, C.M.M. 2016. Human Development Index based on ELECTRE TRI- C multicriteria method: An application in the City of Recife. Social Indicators Research, 125: 19-45.

Ross, A.D. 2014. Local disaster resilience: Administrative and political perspectives. Routledge: New York.

Rowley, H.V. and Peters, G. 2009. Multi-criteria methods for the aggregation of life cycle impacts. In: Proceedings of the Sixth Australian Conference on Life Cycle Assessment, ALCAS, Melbourne, 17-18 February 2009.

Rowley, H.V., Peters, G.M., Lundie, S. and Moore, S.J. 2012. Aggregating sustainability indicators: Beyond the weighted sum. Journal of Environmental Management, 111: 24-33.

Runfola, D. M., Ratick, S., Blue, J., Machado, E.A., Hiremath, N., Giner, N., White, K. and Arnold, J. 2015. A multi-criteria geographic information systems approach for the measurement of vulnerability to climate change. Mitigation and Adaption Strategies for Global Change, 22: 349-368.

Rygel, L., O'Sullivan, D. and Yarnal, B. 2006. A method for constructing a social vulnerability index: an application to hurricane storm surges in a developed country. Mitigation and Adaptation Strategies for Global Change, 11: 741–764.

Saisana M. and Philipas, D. 2012. The Global Innovation Index 2012: Annex 3- Statistical tests on the GII. European Commission/JRC, INSEAD & WIPO.

Saisana, M., Saltelli, A. and Tarantola, S. 2005. Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. Journal of the Royal Statistical Society A, 168: 307-323.

Salzman, J. 2003. Methodological choices encountered in the construction of composite indices of economic and social well-being. Paper presented at the Annual Meeting of the Canadian Economics Association, Carleton University, Ottawa, Ontario, 30 May – 1 June, 2003.

Santos, M. E. and Santos, G. 2014. Composite indices of development. In: Currie-Alder, B., Kanbur, R., Malone, D.M. and Medhora, R. (Editors). International Development: Ideas, Experience, and Prospects. Oxford University Press: Oxford. Pages 133-150.

Sessa, M. 2016. Measuring and monitoring the achievements of the millennium development goals through dynamic composite indices. Social Indicators Research, 127: 469-503.

Sherrieb, K., Norris, F.H. and Galea, S. 2010. Measuring capacities for community resilience. Social Indicators Research, 99: 227-247.

Schmidtlein, M.C., Deutsch, R.C., Piegorsch, W.W. and Cutter, S.L. 2008. A sensitivity analysis of the social vulnerability index. Risk Analysis, 28: 1099-1114.

Sessa, M. 2016. Measuring and monitoring the achievements of the Millennium Development Goals through dynamic composite indices. Social Indicators Research, 127: 469-503.

Sokal, R.R. and Rohlf, J.F. 1981. Biometry: the principles and practice of statistics in biological research. W.H. Freeman and Company: San Francisco.

Tabachnick, B. and Fidell, L. 2007. Using multivariate statistics. 5th edition. Pearson, Allyn and Bacon: Boston, USA.

Tarabusi, E.C. and Guarini, G. 2013. An aggregation method for development indicators. Social Indicators Research, 112: 19-45.

Tate, E. 2012. Social vulnerability indices: a comparative assessment using uncertainty and sensitivity analysis. Natural Hazards, 63: 325-347.

Tate, E. 2013. Uncertainty analysis for a social vulnerability index. Annals of the Association of American Geographers, 103: 526-543.

Tavakol, M. and Dennick, R. 2011. Making sense of Cronbach's alpha. International Journal of Medical Education, 2: 53-55.

Vidoli, F., Fusco, E. and Mazziotta, C. 2015. Non-compensability in composite indicators: A robust directional frontier method. Social Indicators Research, 122: 635–652.

von Hippel, P. 2003. Normalization. In: Lewis-Beck, M.S., Bryman, A. and Liao, T.F. (Editors). The SAGE Encyclopedia of Social Science Research Methods. Sage: Thousand Oaks, California.

Weziak-Bialowolska, D. and Dijkstra, L. 2014. Monitoring multidimensional poverty in the regions of the European Union. JRC Science and Policy Reports, European Commission, Number 26627.

Yang, L. 2014. An inventory of composite measures of human progress. UNDP Human Development Report Office Occasional Paper on Methodology. UNDP: New York.

Zhou, P. and Ang, B.W. 2009. Comparing MCDA aggregation methods in constructing composite indicators using the Shannon-Spearman measure. Social Indicators Research, 94: 83-96.