Evidence-Based Modelling of Organizational Social Capital with Incomplete Data: An NCaRBS Analysis

Malcolm J. Beynon and Rhys Andrews

Cardiff Business School, Cardiff University,
Colum Drive, Cardiff, CF10 3EU, Wales, UK
{BeynonMJ, AndrewsR4}@cardiff.ac.uk

Abstract. Organizational social capital is critical to effective organizational functioning. Yet, different aspects of social capital are likely to be present to varying degrees within any given organization. In this study, alternative blends of structural, relational and cognitive social capital are modelled using a range of key organizational variables drawn from an incomplete dataset. A novel evidence-based approach to the ambiguous classification of objects (N-state Classification and Ranking Belief Simplex or NCaRBS) is used for the analysis. NCaRBS is uniquely able to capture the full range of ambiguity in the antecedents and effects of social capital, and to do so by incorporating incomplete data without recourse to the external management of the missing values. The study therefore illustrates the multi-faceted potential of analytical techniques based on uncertain reasoning, using the Dempster-Shafer theory of evidence methodology.

Keywords: Dempster-Shafer theory · Incomplete data · NCaRBS · Social Capital · Validation

1 Introduction

Positive relationships amongst organization members are essential for efficient knowledge transfer and creation [8]. Nevertheless, the social capital within organizations may be contingent upon internal structural characteristics, such as size, decentralization and staffing cutbacks. In this study, NCaRBS [4], a development on the original CaRBS technique introduced in [1, 2], is used to model alternative combinations of three key dimensions of social capital: structural (connections among actors); relational (trust among actors); and cognitive (shared goals and values) [9].

As a technique whose rudiments are based on the Dempster-Shafer theory of evidence [5, 11], NCaRBS undertakes *n*-state classification analysis based on uncertain reasoning. One of the strengths of NCaRBS (and CaRBS in general) is that it can be applied directly to incomplete data without having to manipulate or exclude cases with missing values. Using a large-scale survey dataset with a sizeable number of missing values, the results of two NCaRBS models of alternative social capital data sets are compared, namely when all missing values are included and when using a case deletion approach to the management of missing values.

Graphical analysis of the contribution of size, decentralization and staffing cutbacks towards social capital blends affirms the value of including missing values and the ability of NCaRBS to undertake such types of analysis. Added confidence in these results comes from a re-sampling procedure that identifies near-identical relationships.

2 NCARBS

NCaRBS [4] models the ambiguous classification of n_{O} objects (o_1, o_2, ..), to n_{D} decision outcomes (d_1, d_2, ..), based on their description by n_{C} characteristics (c_1, c_2, ..). The characteristics' evidence is expressed through the construction of *constituent* BOEs (bodies of evidence) from a characteristic value $v_{i,j}$ (i^{th} object, j^{th} characteristic), to discern between an object's association to a decision outcome (say d_h), its complement ($\neg d_h$) and a level of concomitant ignorance ($\{d_h, \neg d_h\}$).

The construction of a constituent BOE, defined $m_{i,j,h}(\cdot)$ (i^{th} object, j^{th} characteristic, h^{th} outcome), discerning between $\{d_h\}$ and $\{\neg d_h\}$, is described Fig. 1.

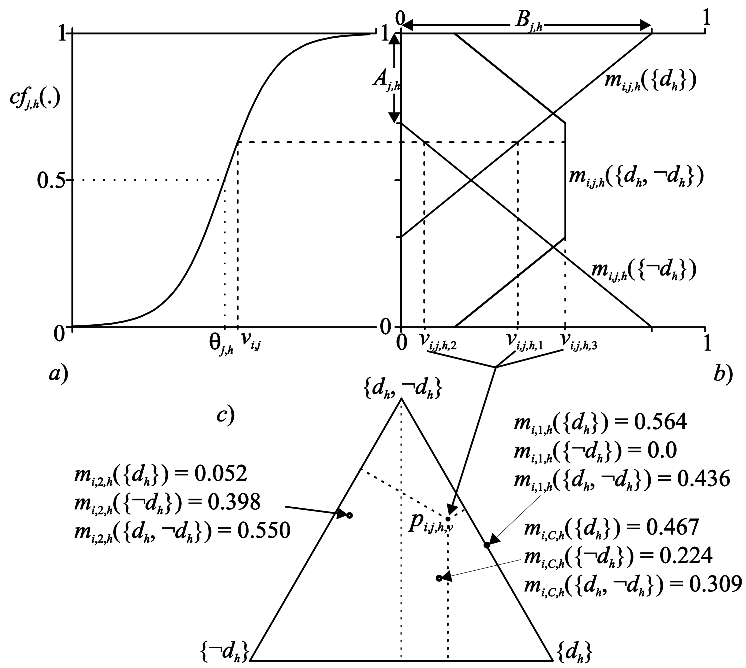

Fig. 1. Stages within the NCaRBS technique

In Fig. 1, stage *a*) shows the transformation of a characteristic value $v_{i,j}$ into a confidence value $cf_{j,h}(v_{i,j})$, using $cf_{j,h}(v_{i,j}) = 1/(1 + \exp(-k_{j,h}(v_{i,j} - \theta_{j,h}))$, with control parameters $k_{j,h}$ and $\theta_{j,h}$. Stage *b*) transforms a $cf_{j,h}(v_{i,j})$ into a constituent BOE $m_{i,j,h}(\cdot)$, made up of the three mass values (see [10]);

$$m_{i,j,h}(\{d_h\}) = \max\left(0, \frac{B_{j,h}}{1-A_{j,h}} cf_{j,h}(v_{i,j}) - \frac{A_{j,h}B_{j,h}}{1-A_{j,h}}\right),$$

$$m_{i,j,h}(\{\neg d_h\}) = \max\left(0, \frac{-B_{j,h}}{1-A_{j,h}} cf_{j,h}(v_{i,j}) + B_{j,h}\right),$$

and $m_{i,j,h}(\{d_h, \neg d_h\}) = 1 - m_{i,j,h}(\{d_h\}) - m_{i,j,h}(\{\neg d_h\})$,

where $A_{j,h}$ and $B_{j,h}$ are two further control parameters. Stage *c*) shows a BOE $m_{i,j,h}(\cdot)$; $m_{i,j,h}(\{d_h\}) = v_{i,j,h,1}$, $m_{i,j,h}(\{\neg d_h\}) = v_{i,j,h,2}$ and $m_{i,j,h}(\{d_h, \neg d_h\}) = v_{i,j,h,3}$, can be represented as a simplex coordinate ($p_{i,j,h,v}$) in a simplex plot (equilateral triangle).

Dempster's rule of combination is used to combine these BOEs. To illustrate, the combination of two constituent BOEs, $m_{i,j_1,h}(\cdot)$ and $m_{i,j_2,h}(\cdot)$, for the same object (o_i) and single outcome (d_h), defined $(m_{i,j_1,h} \oplus m_{i,j_2,h})(\cdot)$, results in a combined BOE with mass values (and focal elements) given by:

$$(m_{i,j_1,h} \oplus m_{i,j_2,h})(\{d_h\}) = \frac{\begin{array}{r} m_{i,j_1,h}(\{d_h\})m_{i,j_2,h}(\{d_h\}) + m_{i,j_2,h}(\{d_h\})m_{i,j_1,h}(\{d_h,\neg d_h\}) \\ + m_{i,j_1,h}(\{d_h\})m_{i,j_2,h}(\{d_h,\neg d_h\}) \end{array}}{1-(m_{i,j_1,h}(\{\neg d_h\})m_{i,j_2,h}(\{d_h\}) + m_{i,j_1,h}(\{d_h\})m_{i,j_2,h}(\{\neg d_h\}))}$$

$$(m_{i,j_1,h} \oplus m_{i,j_2,h})(\{\neg d_h\}) = \frac{\begin{array}{r} m_{i,j_1,h}(\{\neg d_h\})m_{i,j_2,h}(\{\neg d_h\}) \\ + m_{i,j_2,h}(\{d_h,\neg d_h\})m_{i,j_1,h}(\{\neg d_h\}) \\ + m_{i,j_2,h}(\{\neg d_h\})m_{i,j_1,h}(\{d_h,\neg d_h\}) \end{array}}{1-(m_{i,j_1,h}(\{\neg d_h\})m_{i,j_2,h}(\{d_h\}) + m_{i,j_1,h}(\{d_h\})m_{i,j_2,h}(\{\neg d_h\}))}$$

$$(m_{i,j_1,h} \oplus m_{i,j_2,h})(\{d_h,\neg d_h\}) = 1-(m_{i,j_1,h} \oplus m_{i,j_2,h})(\{d_h\}) - (m_{i,j_1,h} \oplus m_{i,j_2,h})(\{\neg d_h\}).$$

This combination process is graphically demonstrated for two example BOEs, $m_{i,1,h}(\cdot)$ and $m_{i,2,h}(\cdot)$, see Fig. 1*c*.

The combination process can be performed iteratively to combine the characteristic based evidence, constituent BOEs $m_{i,j,h}(\cdot)$ $j = 1, .., n_C$, for an object o_i to a single outcome d_h, producing a *outcome* BOE, defined $m_{i,-,h}(\cdot)$ (other ways of combining the evidence can be considered). The respective outcome BOEs can also be combined to bring together the evidence contained in them, the result termed an *object* BOE, for object o_i it is defined $m_{i,-,-}(\cdot)$ (reduced to $m_i(\cdot)$), contains the evidence on the associations of the object to the n_D decision outcomes.

The object BOEs are made up of mass values associated with focal elements, which are the power set of $\{d_1, d_2, ..\}$ (minus the empty set). To enable the assignment of values to individual outcomes, the pignistic probability function $BetP_i(d_h) = \sum_{\substack{s_j \subseteq \{d_1,d_2,..\} \\ s_j \cap \{d_h\} \neq \varnothing}} m_i(s_j)/|s_j|$ for object o_i, it represents the level of pignistic probability associated with the outcome d_h from the object BOE $m_i(\cdot)$. The series of pignistic probability values $BetP_i(d_h)$ $h = 1, .., n_D$ (see [6]), dictates the association of the object o_i to each of the outcomes d_h $h = 1, .., n_D$.

The effectiveness of the NCaRBS technique, is governed by the values assigned to the incumbent control parameters $k_{j,h}$, $\theta_{j,h}$, $A_{j,h}$ and $B_{j,h}$, $j = 1, .., n_C$ and $h = 1, .., n_D$. This necessary configuration is considered as a constrained optimization problem, solved here using trigonometric differential evolution (TDE) [7]. The configured NCaRBS system can be measured by a defined objective function ($\mathrm{OB}^{\mathrm{NCaRBS}}$), the $\mathrm{OB}^{\mathrm{NCaRBS}}$ defined is given as:

$$\mathrm{OB}^{\mathrm{NCaRBS}} = \frac{1}{3n_O}\sum_{i=1}^{n_O}\sqrt{\sum\nolimits_{h=1}^{n_D}\left(BetP_i(d_h) - v_{d_h,i}\right)^2}\ ,$$

in the limit, $0 \leq \mathrm{OB}^{\mathrm{NCaRBS}} \leq 1$ (see [3, 4]).

3 Social Capital

The social capital analysis considered here utilises data from a comparative large-N survey of senior public sector executives conducted in ten European countries (Austria, Estonia, France, Germany, Hungary, Italy, Netherlands, Norway, Spain, United Kingdom) in 2012. The survey was sent to over 21,000 executives via post and email. There were 4,814 valid answers, with a response rate of 22.6%. Missing values are present for a range of questions that some respondents chose not to answer.

Table 1. Organizational social capital dimensions (and items)

People in my organization....	
Structural (S_socap)	Engage in open and honest communication with one another Share and accept constructive criticisms Willingly share information with one another
Relational (R_socap)	Have confidence in one another Have a good team spirit Are trustworthy
Cognitive (C_socap)	Share the same ambitions and vision for the organization Enthusiastically pursue collective goals and mission View themselves as partners in charting the organization's direction

Within the survey (see Table 1 and in text), the *structural* dimension of social capital (S_socap) was gauged by asking informants to score on seven-point scales, ranging from 1 (strongly disagree) to 7 (strongly agree), three questions about the exchange of information between organization members. Three further questions dealing with the strength of working relationships were used to assess *relational* social capital (R_socap). The *cognitive* dimension (C_socap) was then evaluated by posing three questions about the extent to which values and objectives are shared by all staff within the organization.

Alternative combinations, or blends, of the different dimensions of social capital may be the product of key internal organizational characteristics, such as organization size, decentralization of decisions and staffing cutbacks. The *size* of the organizations for which executives worked is measured using a survey question asking respondents to indicate the approximate overall number of employees within the organization in which they worked. Executives were also asked about the presence of ‘decentralization of financial decisions’ and ‘decentralization of staffing decisions’ within their organizations on a 7-point scale, and an index of *decentralization* was then constructed from the responses. *Staffing cutbacks* is measured using a question asking respondents to indicate from 1 (not at all) to 7 (to a great extent) to what extent their organization had applied staff layoffs in response to the fiscal crisis.

As mentioned earlier, the original data set is incomplete, see Table 2, which shows the number of cases which have a certain number of missing characteristic values for the analysis that is undertaken.

Table 2. Levels of incompleteness across considered 4,814 cases

Number Missing	0	1	2	3
Number Cases	3144	1017	112	541

From Table 2, 4,273 respondents have at least 1 value present amongst the characteristics, then using case deletion to deal with the missing values (for example), would mean that only 3,144 respondents would be considered in the analysis of organizational social capital. By contrast, NCaRBS is able to analyse fully the incomplete dataset, thereby permitting the inclusion of over 1,000 further cases in the modelling process.

Prior to undertaking comparative analysis of the incomplete and managed datsets, the separate S_socap, R_socap and C_socap values are transformed into a hybrid vector, which accounts for the distribution of the three values (to reduce the effects of social desirability bias for relational social capital for instance), see Table 3 (following the approach in [4]).

Table 3. Example of social capital blend vector construction

Details	S_socap	R_socap	C_socap
Mean	4.855	5.013	4.532
Standard deviation	1.209	1.20	1.302
Original Capital values (o_{16})	5.667	5.333	5.000
Transformed Capital values (o_{16})	0.354	0.319	0.327

In Table 3 the mean and standard deviation values associated with the three social capital dimensions are presented, showing the general differences in their scores. An example transformation case is shown, for o_{16}, where consideration of the R_socap and C_socap value demonstrates the mitigation of social desirability bias. As the individual social capital blend vectors are made up of three values, which add up to one, they can be represented as points in a simplex plot, see Fig. 2.

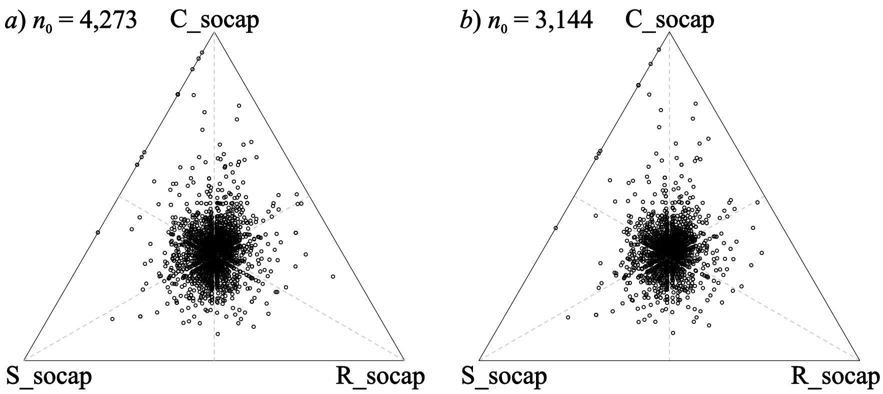

Fig. 2. Social capital blend vectors of cases (a – incomplete, b – managed)

Fig. 2 shows the social capital blend vectors in their simplex coordinate format, for the incomplete (*a* - n_O = 4,273) and managed (*b* - n_O = 3,144) data sets. There is slight variation in the simplex coordinate positions shown across the two different sets of hybrid social capital values (coming from the different numbers of cases considered in each version of the data set).

4 NCaRBS analyses of Social Capital Data Set

This section presents the comparative NCaRBS analyses of the original incomplete social capital data set and an alternative version that is managed through case deletion. To analyse the incomplete data set there has to be a process to model a missing characteristic value, say $v_{i,j}$. Within NCaRBS, and CaRBS in general, from [2], the associated constituent BOE describing a missing value is defined as:

$$m_{i,j,h}(\{d_h\}) = 0,\ m_{i,j,h}(\{\neg d_h\}) = 0 \text{ and } m_{i,j,h}(\{d_h, \neg d_h\}) = 1.$$

This constituent BOE is fixed, and does not change depending on the control parameters found when configuring NCARBS (see for example Fig 1).

The results from the two NCaRBS analyses are restricted here to the level of model fit (based on respective $\text{OB}^{\text{NCaRBS}}$ values) and contributions of organizational characteristics to the objects' social capital blend vectors. Each model was run 10 times, with best fit for the incomplete data being $\text{OB}^{\text{NCaRBS}} = 0.070779$ and for the managed data set $\text{OB}^{\text{NCaRBS}} = 0.070336$, indicating that the model fit for the incomplete data set exhibits the slightly worse predictive fit. To understand the variation in fit values, we should consider the actual numbers of available organizational characteristics to configure on. For the incomplete and managed data sets there are 11,578 and 9,432 characteristic values respectively to model social capital. Hence with 81.465% of the data to work with, it is not entirely surprising that the $\text{OB}^{\text{NCaRBS}}$ value is lower for the managed data set.

The results in terms of characteristics' contributions are explored here through their graphical representation; see Fig. 3 and Fig. 4.

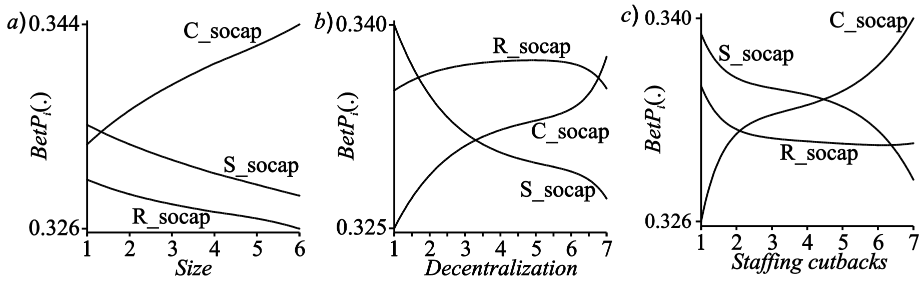

Fig. 3. Characteristic contribution based on incomplete data set

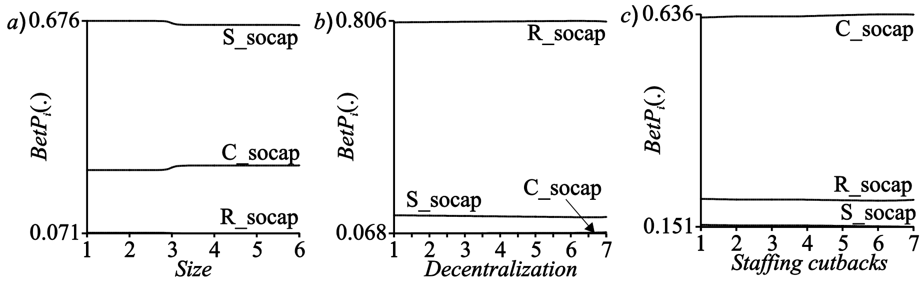

Fig. 4. Characteristic contribution based on managed data set

It can be clearly seen that alternative forms of information are gained from the NCaRBS analysis of the incomplete data (Fig. 3) and the managed data (Fig. 4). Generally, NCaRBS is able to fully demonstrate the nonlinearity in the associations between objects (respondents) and outcomes (social capital blend). Concentrating on the results from Fig. 3 (incomplete data set), Fig. 3a indicates that as organization size increases structural and relational capital decline, but cognitive social capital becomes stronger. Fig. 3b illustrates that as decentralization increases structural social capital declines, but relational and (especially) cognitive social capital grow. Finally, Fig. 3c highlights that staff cutbacks are associated with declining intra-organizational communication and interpersonal trust, but higher levels of shared mission.

5 Validation Analysis of NCaRBS results (Using Re-sampling)

The results presented in section 4 are from a one-off analysis using all the available data (3,144 cases for incomplete data set and 4,273 for managed data set). To add further confidence in the validity of the results from this analysis, a re-sampling procedure is undertaken and the models recalculated (see for example [12]). Due to page limitation this validation exercise is undertaken on the incomplete data set only.

The re-sampling undertaken here was based on performing multiple runs of the NCaRBS technique using identified in-samples and out-of-samples of cases. Here, 10 runs were performed, in each run 90% of cases (3,846) were used as the in-sample on which the NCaRBS was run to configure a model, and 10% of cases (427) were used an out-of-sample.

For each pair of in-sample and out-of-sample sets of data, levels of fit can be found based on the objective function (OB^{NCaRBS}), see Fig. 5.

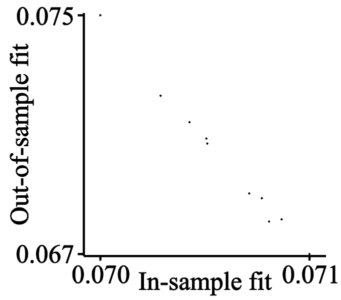

Fig. 5. Scatter-plot of in-sample and out-of-sample fit values (based on incomplete data set)

In Fig. 5, the two axes depict the OB^{NCaRBS} fit value for in-sample (horizontal) and out-of-sample (vertical). Clearly, there is a relatively consistent inverse relationship between the pairs of fit values, namely as the level of in-sample fit increases so the level of out-of-sample fit decreases. Beyond this relationship, whether there is significant difference between the in-sample and out-of-sample fit values are considered using a paired-sample t-test. From the test there was not a significant difference between the fit values for in-sample (M = 0.0708, SD = 0.000270) and out-of-sample (M = 0.0705, SD = 0.00236) sets of data; $t(9) = 0.372$, $p = 0.718$. Briefly, for the managed data set, similar analysis also suggested not a significant different between the fit values for in-sample (M = 0.0703, SD = 0.000270) and out-of-sample (M = 0.0711, SD = 0.000270) sets of data; $t(9) = -0.682$, $p = 0.512$. The results suggest the configured NCaRBS models in each of the 10 runs fit the out-of-sample cases.

The contribution of the individual variables to the social capital blends following the re-sampling procedure can be illustrated graphically as for the one-off analysis using all of the data (see Fig. 6).

In Fig. 6, the ten contribution lines associated with each separate social capital dimension derived from the re-sampling runs are plotted together to illustrate the general trends found through the re-sampling process. Comparison of these graphs with those for the one-off analysis presented in Fig. 3, reveals very similar patterns in the relationship between the structural characteristics and each dimension of social capital.

For example, comparing Fig. 3*a* with Fig. 6*a*, 6*b* and 6*c*, as organization size increases, structural and relational capital decline, but cognitive social capital becomes stronger; as decentralization increases, structural social capital declines, but relational and (especially) cognitive social capital grow; staff cutbacks are associated with declining intra-organizational communication and interpersonal trust, but higher levels of shared mission.

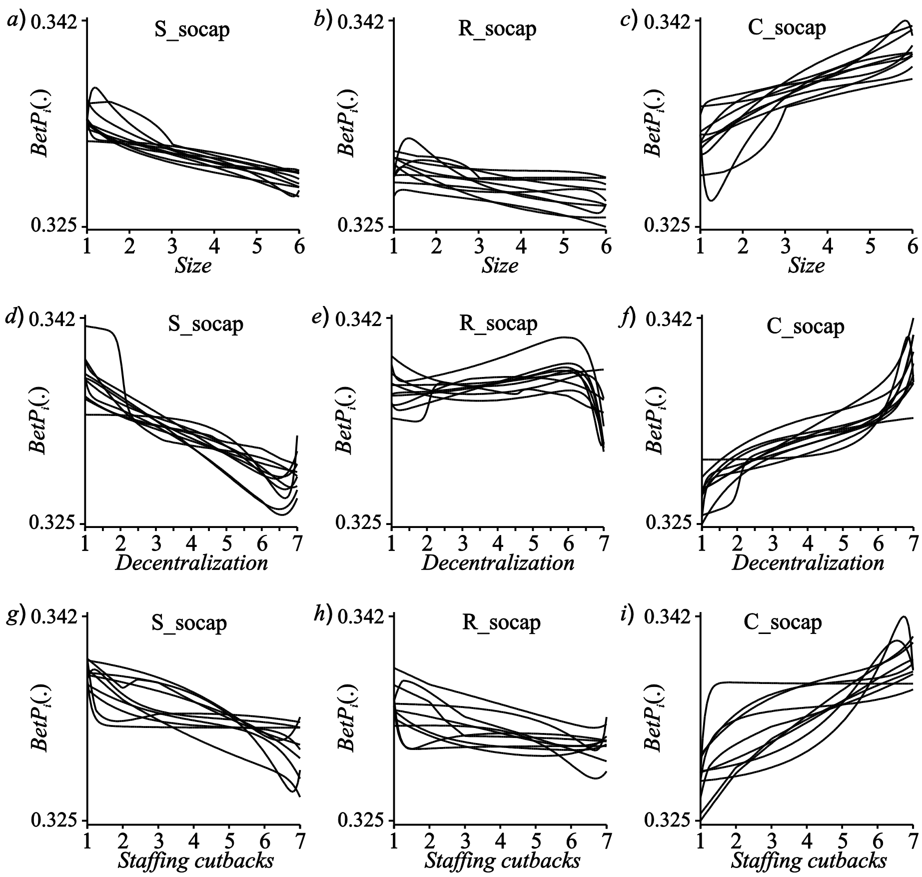

Fig. 6. Characteristic contribution in 10 runs (based on incomplete data)

6 Conclusions

The NCaRBS technique, along with all the family of CaRBS based techniques, offer an almost unique opportunity to analyze incomplete data without the need to manipulate or exclude cases with missing values. The results for the evidence-based modelling of organizational social capital blends presented here dramatically illustrate the impact of this facility for incorporating incompleteness. Nevertheless, the full potential of the technique has yet to be explored. Further research could investigate the sensitivity of the technique to alternative ways of capturing the impact of ignorance in the data. In particular, it would be interesting to evaluate the effect of weighting the impact of cases in the objective functions depending on how incomplete the information associated with them is.

For now though, this paper concludes by observing that NCaRBS offers organizational analysts and scientists in other fields a powerful means for incorporating data with missing values in their research. In this respect, we concur with others who call for more and better work developing and demonstrating novel applications of Dempster-Shafer theory of evidence based analysis techniques.

7 References

1. Beynon, M.J.: A novel technique of object ranking and classification under ignorance: An application to the corporate failure risk problem. European Journal of Operational Research. **167**, 493-517 (2005a)
2. Beynon, M.J.: Optimizing object classification under ambiguity/ignorance: Application to the credit rating problem. International Journal of Intelligent Systems in Accounting, Finance and Management. **13**, 113-130 (2005b)
3. Beynon, M.J., Andrews, R.A., Boyne, G.: Evidence-based Modelling of Strategic Fit: An Introduction to RCaRBS. European Journal of Operational Research. **207**(2), 886-896 (2010)
4. Beynon, M.J., Andrews, R.A., Boyne, G.: Evidence-based Modelling of Hybrid organizational strategies. Computational and Mathematical Organization Theory Journal. (2014) doi: 10.1007/s1058801491745
5. Dempster A.P.: Upper and lower probabilities induced by a multiple valued mapping. *Ann. Math. Statistics*. **38**, 325-339 (1967)
6. Denœux, T., Zouhal, L.M.: Handling possibilistic labels in pattern classification using evidential reasoning. Fuzzy Sets and Systems. **122**, 409-424 (2001)
7. Fan, H.-Y., Lampinen, J.: A trigonometric mutation operation to differential evolution. Journal of Global Optimization. **27**, 105-129 (2003)
8. Kogut, B., Zander, U.: What do firms do? Coordination, identity and learning. Organization Science. **7**, 502-518 (1996)
9. Nahapiet, J., Ghoshal, S.: Social capital, intellectual capital and the organizational advantage. Academy of Management Review. **23**, 242-266 (1998)
10. Safranek, R.J., Gottschlich, S., Kak, A.C.: Evidence accumulation using binary frames of discernment for verification vision. IEEE Transactions on Robotics and Automation. **6**, 405-417 (1990)
11. Shafer, G.A.: Mathematical theory of evidence. Princeton, Princeton University Press (1976)
12. Twomey, J.M., Smith, A.E.: Bias and Variance of Validation Methods for Function Approximation Neural Networks Under Conditions of Sparse Data. IEEE Transaction on Systems, Man and Cybernetics – Part C: Applications and Reviews. **28**(3), 417-430 (1998)