Survey Research Methods (2011)
Vol.5, No.2, pp. 63-74
ISSN 1864-3361
http://www.surveymethods.org

Measurement Quality in Indicators of Compositions. A Compositional Multitrait-Multimethod Approach

Germà Coenders
University of Girona

Valentina Hlebec
University of Ljubljana

Tina Kogovšek
University of Ljubljana

Compositional data, also called multiplicative ipsative data, are common in survey research instruments in areas such as time use, budget expenditure and social networks. Compositional data are usually expressed as proportions of a total, whose sum can only be 1. Owing to their constrained nature, statistical analysis in general, and estimation of measurement quality with a confirmatory factor analysis model for multitrait-multimethod (MTMM) designs in particular are challenging tasks. Compositional data are highly non-normal, as they range within the 0-1 interval. One component can only increase if some other(s) decrease, which results in spurious negative correlations among components which cannot be accounted for by the MTMM model parameters.

In this article we show how researchers can use the correlated uniqueness model for MTMM designs in order to evaluate measurement quality of compositional indicators. We suggest using the additive log ratio transformation of the data, discuss several approaches to deal with zero components and explain how the interpretation of MTMM designs differs from the application to standard unconstrained data.

We show an illustration of the method on data of social network composition expressed in percentages of partner, family, friends and other members in which we conclude that the face-to-face collection mode is generally superior to the telephone mode, although primacy effects are higher in the face-to-face mode. Compositions of strong ties (such as partner) are measured with higher quality than those of weaker ties (such as other network members).

Keywords: compositional data, ipsative data, multitrait-multimethod, correlated uniqueness model, social networks

Introduction

Statistical compositions consist of positive data arrays with a fixed sum. The commonest examples are proportions or percentages of the set of components of a total, whose sum can only be 1 or 100%. Compositional data are thus severely constrained.

Composition indicators are more frequent in social science data collected by surveys than one might anticipate. Among many examples given by Aitchison (1986), compositional data from household budget surveys and time-use surveys (e.g. ATUS 2010; Bittman and Rice 2002; Lader et al. 2006; Pentland et al. 1999) are typical. Budget surveys have the choice of using either amounts or percentages spent on the given good or service categories. In time-use surveys, however, data tend to be compositional by nature, as the total amount of available time which is divided into activities is usually constant (e.g. 24h a day). Time use has been found to be a key predictor variable in many social issues and time-use surveys have become so widespread that even a specialised journal is devoted to the issue (electronic International Journal of Time Use Research). Another field of application are compositional indicators used in network surveys (e.g. Burt 1984; Kogovšek and Hlebec 2008; Marsden 1987; Müller et al. 1999), usually expressed in percentages of family members, friends, neighbours, co-workers and others. Relative estimates are preferred to sums of network members (number of relatives, number of friends) as people have local networks of different sizes and percentages make it possible to compare networks of different sizes. This article will include an empirical example in this field.

Compositional data do not lend themselves easily to standard statistical analyses for unconstrained data:

- On the one hand, specialised techniques for compositional data are starting to appear (e.g. Thió-Henestrosa and Martín-Fernández 2005).
- On the other hand, compositional data can be transformed so that they can be subject to standard statistical techniques as they are, or with minor modifications (Aitchison 1986).

When it comes to assessing measurement quality of questions, standard statistical techniques for unconstrained data such as confirmatory factor analysis (CFA) are commonly understood and applied by social scientists, and the

Contact information: Germà Coenders, Department of Economics, University of Girona, e-mail: germa.coenders@udg.edu

approach of transforming the data and keeping analyses as standard as possible shows greater promise. Up to now, the assessment of measurement quality of compositional survey data by means of CFA models has been hindered by a lack of useful guidelines regarding how to take into account the compositional nature of the data. The aim of this article is to present and illustrate appropriate choices regarding data preparation and model specification and interpretation.

The first sections of the article consist in the explanation and derivation of the statistical procedure which can be used to evaluate the measurement quality of compositional indicators with a CFA model for multitrait-multimethod (MTMM) designs. Whenever we had the choice, we have opted for the simplest options, in order to narrow the gap between methodological and applied knowledge. In the second to last section we present an illustration applied to indicators of composition of egocentred social support networks. The last section concludes.

Correlated uniqueness model for multitrait-multimethod (MTMM) designs

MTMM designs (Campbell and Fiske 1959) are a well established approach to assess measurement quality of survey questions (see Saris and Gallhofer 2007 and references therein). These designs consist of multiple measures of at least three factors (traits) with the same set of at least three measurement procedures (methods).[1] So, these designs include *DM* measures, that is the number of methods (*M*) times the number of traits (*D*). MTMM designs are usually analysed by means of CFA models, a particular case of structural equation models (-SEM-, see Schumacker and Lomax 1996 for an overview). A number of CFA models for MTMM data have been formulated and tested in the literature (see Coenders and Saris 2000, for a review). These authors showed the great flexibility of the so-called correlated uniqueness (CU) model (Marsh 1989), of which many other MTMM models constitute particular cases. The CU model is a CFA model specified as follows.

Let x_{idm} be the measurement of individual *i*, for trait *d* with method *m*:

$$x_{idm} = \tau_{dm} + \lambda_{dm} t_{id} + e_{idm} \tag{1}$$

where t_{id} is the latent variable score of individual *i* corresponding to trait *d* and e_{idm} is the measurement error term of individual *i*, for trait *d* with method *m*, with the assumptions:

$$\begin{aligned} E(t_{id}) = E(e_{idm}) &= 0 \\ cov(e_{idm}, e_{id'm}) &= \theta_{dd'm} \\ cov(t_{id}, t_{id'}) &= \varphi_{dd'} \\ cov(e_{idm}, e_{id'm'}) &= 0 \end{aligned} \tag{2}$$

The model parameters are:

- τ_{dm}: intercept term of x_{idm}. If the assumptions in (2) are fulfilled, it can be interpreted as the expected value of x_{idm}. The comparison of these parameters across methods for the same trait makes it possible to assess differences in method bias.
- λ_{dm}: loading of x_{idm} on trait t_{id}. It relates the scales of x_{idm} and t_{id}. One loading for each trait (i.e. when m =1) has to be constrained to unity for latent variable identification purposes (λ_{d1} =1).
- θ_{dm}: measurement error variance of e_{idm}.
- $\theta_{dd'm}$: covariance between two measurement error terms sharing a common method e_{idm} and $e_{id'm}$. In an MTMM design it is expected that the use of the same method involves common errors. These covariances are called method effects for this reason.
- φ_{dd}: variance of the trait latent variable t_{id}.
- $\varphi_{dd'}$: covariance between two trait latent variables t_{id} and $t_{id'}$.

Two main measurement quality indicators can be obtained by the model:

- Standardized λ_{dm} trait loadings measure the strength of the relationship between observed scores and trait latent scores. Other measures of measurement quality can be obtained by re-expressing the standardized trait loadings. The squared standardized loading is the percentage of variance of x_{idm} explained by t_{id}. The standardized error variance is one minus the squared standardized loading. Of course, these sets of measures are mutually redundant and just one of them is enough.
- Non-standardized τ_{dm}, $\tau_{dm'}$,... intercepts make it possible to assess the differences in the bias of several methods m, m',... when measuring trait *d*. If $\tau_{dm} = \tau_{dm'}$, then there is no difference in the biases of *m* and *m'* when measuring trait *d*. If $\tau_{dm} > \tau_{dm'}$, then *m* yields systematically larger scores than *m'*, when measuring trait *d*.

A path diagram of the CU model is displayed in Figure 1.

Statistical data analysis of compositions

Challenges in the analysis of compositional data

Compositional data concern the relative size of *D* components within a total, usually expressed in proportions over 1 or over 100%. We refer, for instance, to percentages of friends, family and other types of members of a personal network, to percentages of time used for different types of activities, or to percentages of budget spent on a set of goods and services.

The study of the measurement quality of compositional data cannot be undertaken by just fitting the proportions or percentages to a SEM (e.g. to a CU model). Compared to unconstrained data (e.g. number of friends, family and so on in the network, expenditure in euro on different goods and

[1] The 3 trait x 3 method guideline is a general one. In some instances, two traits or two methods may suffice. The correlated uniqueness model presented in this section is identified with two or more traits and three or more methods.

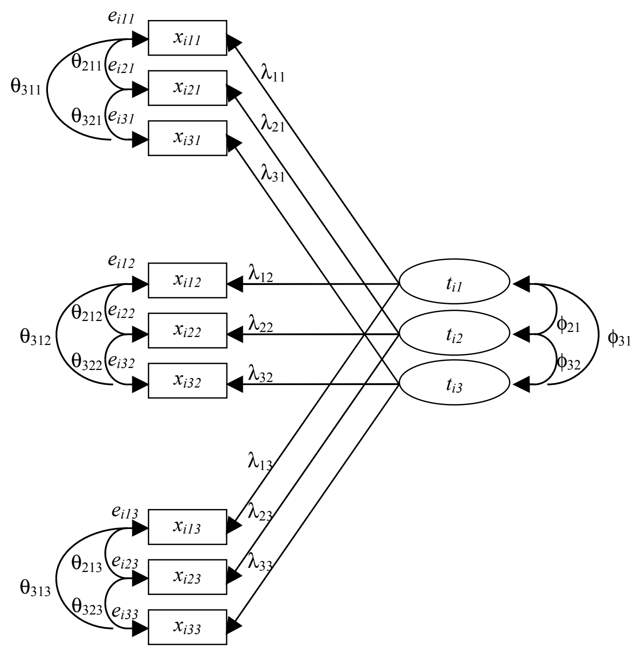

Figure 1. Path diagram of the CU model for MTMM designs. 3 traits and 3 methods. τ_{dm} parameters are omitted for simplicity.

services), compositional data are expressed in relative terms and lie in a heavily constrained space. A D-term composition measured on individual i with a given method m takes the form $x_{i1m}, x_{i2m}, ..., x_{iDm}$ with the constraints:

$$0 \leq x_{idm} \leq 1 \; and \sum_{d=1}^{D} x_{idm} = 1 \tag{3}$$

The unconstrained data in absolute terms, which are often unknown, are $s_{im}x_{i1m}, s_{im}x_{i2m}, ..., s_{im}x_{iDm}$ where s_{im} is size for individual i as given with method m.

Aitchison (1986) warns against the severe problems arising when using standard statistical analysis tools on compositional data:

- Compositional data are highly non-normal, as they range within the 0-1 interval and are often highly skewed, especially for very small or very large components.
- Compositional data have a constrained sum: one component can only increase if some other(s) decrease. This results in spurious negative correlations among components, even if the absolute $s_{im}x_{idm}$ data are independent. The sum of covariances in a compositional data matrix can only be negative even if some particular covariances can still be positive (Aitchison 1986). Drawing from the fact that a constant has zero variance and from the standard properties of the variance of a sum of variables:

$$\begin{aligned} var\left(\sum_d x_{idm}\right) &= var(1) = 0 \\ \sum_d var\,(x_{idm}) + 2\sum_{d<d'} cov\,(x_{idm}, x_{id'm}) &= 0 \\ \sum_{d<d'} cov\,(x_{idm}, x_{id'm}) &= -(1/2)\sum_d var\,(x_{idm}) \end{aligned} \tag{4}$$

- Because of the unit sum constraint, the true dimensionality of a set of compositional variables measured with a given method m is D-1. Analysis of all D dimensions leads to non-positive definite covariance matrices, perfect collinearity and the like. This is probably the least serious problem of all, as one can usually get by when dropping one component from the analysis.

In the context of SEM, constant sum data are referred to as ipsative data (Chan 2003). Zero sum data are called additive ipsative data and unit sum data (which is the case in compositional data) are called multiplicative ipsative data. While additive ipsative data have successfully been dealt with in the SEM context (Chan 2003; Cheung 2004), this is not the case for multiplicative ipsative data.

The general problems reported by Aitchison (1986) also apply to the CU model with compositional data, with an important addition. Even if the absolute $s_{im}x_{i1m}$ data fit a CU model, the compositional x_{idm} data do not. Let us consider the model for D = 4 components measured with method 1:

$$\begin{aligned} x_{i11} &= \tau_{11} + \lambda_{11} t_{i1} + e_{i11} \\ x_{i21} &= \tau_{21} + \lambda_{21} t_{i2} + e_{i21} \\ x_{i31} &= \tau_{31} + \lambda_{31} t_{i3} + e_{i31} \\ x_{i41} &= \tau_{41} + \lambda_{41} t_{i4} + e_{i41} \end{aligned} \tag{5}$$

Both the expected and the individual compositions must add up to 1:

$$\begin{aligned} \tau_{11} + \tau_{21} + \tau_{31} + \tau_{41} &= 1 \\ x_{i11} + x_{i21} + x_{i31} + x_{i41} &= 1 \end{aligned} \tag{6}$$

From (5) and (6) we can express any component as a function of the rest:

$$\begin{aligned} x_{i11} = \tau_{11} + \lambda_{11} t_{i1} + e_{i11} &= 1 - \tau_{21} - \lambda_{21} t_{i2} - e_{i21} \\ &\quad -\tau_{31} - \lambda_{31} t_{i3} - e_{i31} - \tau_{41} - \lambda_{41} t_{i4} - e_{i41} \\ e_{i11} &= -\lambda_{11} t_{i1} - \lambda_{21} t_{i2} - e_{i21} - \lambda_{31} t_{i3} - e_{i31} \\ &\quad -\lambda_{41} t_{i4} - e_{i41} \end{aligned} \tag{7}$$

Any error is thus dependent on all traits and on all remaining errors, within a method. Intuitively, for a given set of true t_{idm} compositions the observed x_{i1m} component can only increase if some other components decrease. x_{i1m} is thus not only dependent on t_{i1m} but on all t_{idm}. The CU model assuming each observed variable to load only on a trait is miss-specified, which can lead to bias in any parameter estimate. Likewise, in order to preserve the unit sum, an error term can be positive only if some other error term(s) within the same method are negative, and vice-versa. Fortunately, error covariances between two measurement error terms sharing a common method do not constitute a misspecification of the CU model because they are accounted for by the $\theta_{dd'm}$ parameters. However, this negative bias in $\theta_{dd'm}$ has several important consequences. First, it prevents us from interpreting $\theta_{dd'm}$ as method effects in the usual manner. Secondly, it prevents us from using any SEM in which error correlations among components are omitted as model parameters.

Compositional data transformations

In compositional data the absolute size of components is lost. Only the relative size of some components to the others is maintained. Thus, ratios are the only meaningful way of expressing the data. The analysis of compositional data with standard statistical methods is only possible after some kind of ratio transformation has been applied.

Several ratio transformations have been suggested in the literature. Among them are the additive logratio transformation (alr), the centred logratio transformation (clr), both suggested by Aitchison (1986) and the isometric logratio transformation (ilr) suggested by Egozcue et al. (2003). The clr transformation leads to additive ipsative data and, thus, there is not much to be gained from it in the SEM context. While alr and ilr are both feasible, they are not equally convenient for all purposes. The ilr transformation leads to correct angles and distances and is therefore a must for such applications as cluster analysis or graphical displays (Egozcue et al. 2003). On the negative side, its computation is complex without resorting to specialised compositional software and its interpretation is far from intuitive to non specialists in the field. The alr transformation is by far the easiest to compute and interpret, as it is simply computed as the log ratio of each component to the last, and can be successfully applied for modelling purposes:

$$\begin{aligned} y_{idm} &= ln(x_{idm}/x_{iDm}) = \\ &ln(x_{idm}) - ln(x_{iDm}) \; with \; d = 1, 2, ..., D-1 \end{aligned} \tag{8}$$

Of course, any component may be situated in the last position at will. It must be clear by now that the alr transformed composition has one fewer dimension than the original composition.

An attractive property of the alr transformation is that it yields the same result when computed from components or from the original unconstrained data:

$$\begin{aligned} ln(x_{idm}/x_{iDm}) &= ln(s_{im}x_{idm}/s_{im}x_{iDm}) = \\ &\quad ln(s_{im}x_{idm}) - ln(s_{im}x_{iDm}) \end{aligned} \tag{9}$$

From this point onwards we focus on the alr transformation.

The alr transformed y_{idm} variables recover the full $-\infty$ to ∞ range. Whether the data thus transformed follow a normal distribution or fail to do so will, of course, depend on the particular case at hand. However, plenty of cases of approximately normal alr-transformed data are reported in the literature (Aitchison 1986).

As regards interpretation, the alr-transformed data are equal to zero if the given component is equal to the last: $y_{idm}=0$ if $x_{idm} = x_{iDm}$. Similarly, $y_{idm}>0$ if $x_{idm}>x_{iDm}$; for instance, $y_{idm}=1$ if $x_{idm}=2.72x_{iDm}$ and $y_{idm}<0$ if $x_{idm}<x_{iDm}$; for instance, $y_{idm}=-1$ if $x_{iDm}=2.72x_{idm}$.

While the original dimensionality in an MTMM design is DM, the transformed data set has $(D\text{-}1)M$ dimensions. One dimension (trait) per method is lost. The CU model is simply estimated on the alr-transformed y_{idm} data on the $(D\text{-}1)M$-dimensional data set by using conventional methods for SEM estimation.[2]

The CU model on alr-transformed data still has some limitations regarding parameter interpretation:

- Trait correlations tend to be positive because alr data have a common denominator. The correlations among

[2] For certain SEM it may be feasible to define D $\ln(x_{idm})$ first order latent variables and D-1 y_{idm} observed variables with null error variances and loadings trivially constrained to 1 and -1 following the alternative expression of the alr transformation as $y_{idm}=\ln(x_{idm})$-$\ln(x_{iDm})$. This is similar to the approach of Cheung (2004) for additive ipsative data. When only one dimension is lost, this is likely to lead to identified SEM. In MTMM models, as many as M dimensions are lost and this is not feasible as it leads to model under-identification.

ratios cannot be interpreted as if they were correlations among the original absolute $s_{im}x_{idm}$data. It can be shown that the covariance between any two components contains the variance of the Dth component, which can only be positive. Drawing from (9):

$$\begin{aligned}
&cov(y_{idm}, y_{id'm})\\
&= cov\Big(ln(s_{im}x_{idm}) - ln(s_{im}x_{iDm}),\\
&\quad ln(s_{im}x_{id'm}) - ln(s_{im}x_{iDm})\Big)\\
&= cov\Big(ln(s_{im}x_{idm}), ln(s_{im}x_{id'm})\Big)\\
&\quad +var\Big(ln(s_{im}x_{iDm})\Big)\\
&\quad -cov\Big(ln(s_{im}x_{idm}), ln(s_{im}x_{iDm})\Big)\\
&\quad -cov\Big(ln(s_{im}x_{id'm}), ln(s_{im}x_{iDm})\Big) \qquad (10)
\end{aligned}$$

- For the same reason, $\theta_{dd'm}$ error term covariances are spurious and positive, as they contain the variance of the Dth error term. Therefore, they cannot be interpreted as method effects or measurement invalidity in the classical sense. The appropriateness of the CU model for compositional data lies in the fact that it includes these error covariance parameters for all pairs of components measured with a given method. In this case these error covariances play a methodological rather than a substantive role and are not interpreted. Using alternative MTMM models including method factors instead of error covariances would introduce unreasonable constraints to the error covariances and might make the researcher fall into the temptation of interpreting the loadings on method factors as method effects or invalidity. The CU model with unconstrained error covariances is indeed less parsimonious than certain alternative models with method factors (Coenders and Saris 2000) but still leads to identified models for two or more traits and three or more methods.

The alr transformation solves the model misspecification in (7). Once we have taken into account the proper interpretation of trait covariances and error covariances, the main parameters of interest are interpretable in the usual manner. Standardized trait loadings indicate measurement quality and raw (i.e. non-standardized) intercepts indicate differences in method bias. Standardized error variances can also be interpreted, although the information they provide is redundant with that provided by standardized trait loadings.

Dealing with zero components

If either x_{idm} or x_{iDm} equal zero, then y_{idm} cannot be computed. Zeroes have thus to be dealt with prior to analyzing compositional data. An obvious first procedure is to amalgamate small components with many zeroes into larger ones with fewer zeroes. For instance, in a budget survey, infrequent product types could be merged into an "other" category. This is feasible if the number of components is large and the amalgamated components have some degree of theoretical similarity.

In certain instances, some zero components result from individual characteristics. For instance, people who have never been employed cannot have co-workers in their social network, non smokers cannot have any proportion of their budget spent on tobacco, and so on. Aitchison (1986) refers to this situation as essential zeroes. It can be understood that the established components do not make sense to a portion of the population. When essential zeroes are present and external variables are available to identify them, it may be advisable to narrow the definition of the target population and remove individuals with essential zeroes from the sample.

Amalgamation of components and redefinition of the population, when feasible, are the most theoretically sound means to handle zeroes. If their application is not possible or if zeroes remain after their application, zeroes must be replaced as explained below.

Zeroes may be understood as components which are too small to be detected with the measurement method used. For instance:

- In a time-use diary with half hour intervals, a small amount of time devoted to an activity will likely not be recorded and components smaller than 1/48 will likely not be detected.
- In a social network questionnaire in which respondents are allowed to mention up to s_m members, x_{idm} components in smaller proportions than $1/s_m$ will not be detected.

Both examples differ in one important respect. The first is based on numeric variables (in time units, albeit rounded to the nearest half hour) and the second on multinomial variables (counts of network members with given characteristics).

Zeroes are usually replaced by a small amount which is likely to be undetected. For numeric variables, we define the smallest detectable proportion as δ_{idm}. Martín-Fernández et al. (2003) suggest replacing x_{idm} =0 with:

$$x'_{idm} = k\delta_{idm} \; with \; 0 < k < 1 \qquad (11)$$

The authors suggest using k =0.65, although a sensitivity analysis of the results on the choice of k is always advisable.

For multinomial variables, Pierotti et al. (2009)'s Bayesian approach implies replacing x_{idm} =0 with:

$$x'_{idm} = \frac{1}{D(s_{im}+1)} \qquad (12)$$

Non-zero x_{idm} values have to be reduced in order to preserve the unit sum. The sum of quantities replacing zeroes is subtracted proportionally from non-zero values, as suggested by Martín-Fernández et al. (2003) in what they call multiplicative replacement strategy. Both in the numeric and multinomial case, this implies replacing $x_{idm} > 0$ with:

$$x'_{idm} = x_{idm}(1 - \sum_{x_{idm}=0} x'_{idm}) \qquad (13)$$

CU model estimation

The CU model on alr-transformed data can be estimated with the usual available methods for SEM, like maximum likelihood (ML). Satorra and Bentler (1994) developed corrections to standard errors and test statistics to render them robust to non normal data, which are nowadays common place in software packages and which we recommend practitioners to use.

A severe drawback of the standard MTMM approach is that at least three methods must be included. If this is done within a single data collection, memory effects can only be avoided by a generous spacing of questions, resulting in long, costly questionnaires and a high respondent burden. Panel designs increase costs even further and tend to lower response rates at later waves. Saris et al. (2004) suggest solving this problem by the split-ballot MTMM design. The split-ballot MTMM design employs various random samples of the same population, but in each of the samples only two methods are used. In the authors' three group design, the respondent group A gets methods 1 and 2, group B gets methods 2 and 3 and group C gets methods 3 and 1. Although no group provides data on all three methods, the covariances between any two variables can be estimated with data from at least one group. The only statistical implication is that this design requires estimation methods which can handle missing data.

Direct ML estimation of SEM with missing data was discussed by Arbuckle (1996), Finkbeiner (1979) and Lee (1986) assuming that the data are normally distributed and missing at random (the probability of a datum being missing depends only on observed individual characteristics, i.e. on variables which are in the model). Schafer and Graham (2002) discuss the state of the art and alternatives.

A variant of the direct ML estimator with missing data described by Yuan and Bentler (2000) is robust to non-normality but assumes data to be missing completely at random (the probability of a datum being missing does not depend on any individual characteristic). The robust test statistics are described in Yuan and Bentler (2000) and in Arminger and Sobel (1990). A carefully designed split-ballot MTMM experiment in which individuals are randomly assigned to groups ensures that the data are missing completely at random and thus this is the estimation method of choice. The literature reports this method to perform quite well even when data are just missing at random (Enders 2001) though a small overestimation of standard errors is reported in Gold et al. (2003).

Illustration

Data

The focus of this example are indicators of network composition obtained from egocentred networks. As opposed to complete networks (which consist of a group of individuals with one or more relations defined among them), egocentred networks consist of a single individual (usually named ego) with one or more relations defined between her/him and a number of individuals who are members of her/his personal network. The network members are often called alters. Once the names of alters are obtained with the so-called name generator questions, several additional questions (name interpreters) are posed to find out about characteristics of network members (e.g. age, gender) and ties connecting ego to her/his alters (e.g. type of relation between ego and alters, frequency of contacts, geographical distance). The characteristics measured through name interpreters can be used to classify network members into a set of network components. In fact, network composition, in other words, percentages of partner, kin, friends and other members within the network, is among the most often calculated and interpreted egocentred network characteristics (Burt 1984; Kogovšek and Hlebec 2008; Marsden 1987; Müller et al. 1999).

Very often data about egocentred networks are collected in a survey setting. Several studies have addressed measurement quality of egocentred networks measured with surveys (e.g. Kogovšek 2006; Kogovšek et al. 2002; Kogovšek and Ferligoj 2003; 2004; 2005; Lozar-Manfreda et al. 2004; Vehovar et al. 2008). We will focus our example on the same data as Kogovšek et al. (2002).

Egocentred networks in the Kogovšek et al. (2002) study were defined as personal social support networks understood as a multidimensional construct (Cohen and Wills 1985; Hirsch 1980; Vaux 1988; Weiss 1974; Wills 1985). Kogovšek et al. (2002) used the typology of Cohen and Wills (1985), which distinguishes among instrumental support, informational support, emotional support and social companionship (see name generators in the appendix). Kogovšek et al. (2002) used two widely used measures of tie strength (the frequency of contact of ego with each alter and feelings of closeness of ego towards each alter) and a measure of negative aspects of social relationships (the frequency of each alter upsetting ego). The units of analysis were egocentred networks as a whole and the traits were averages (e.g. the average frequency of contact of ego with all his/her alters). These traits were thus not compositional but unconstrained. The purpose of the example in this article is to evaluate measurement quality of indicators of network composition. The components (traits) used in this example are defined as the percentages of network members represented by:

1: partner
2: friends
3: others
4: family (reference component for the alr transformation)

Kogovšek et al. (2002) report about data quality of telephone versus face-to-face data collection modes, and two data collection techniques, "by alters" and "by questions". The specificities of telephone and face-to-face data collection modes are well documented in the general literature on survey methodology (e.g. de Leeuw 1992). As regards the two mentioned techniques, once we have the list of alters obtained by name generator questions, name interpreter questions can be presented to respondents in two ways. The first, "by alters", is to take each alter and to ask all name interpreter questions about him/her before moving to the next alter. The second way, "by questions", is to take each question

Table 1: The split-ballot MTMM design of the study

Group	n	First measurement with method:	Second measurement with method:
A	317*	Face to face ordered by alters	Telephone ordered by alters
B	311	Face to face ordered by alters	Telephone ordered by questions
C	402	Telephone ordered by alters	Telephone ordered by questions

* Original n was 320. Three cases with network size equal to 1 were deleted for our analyses.

(e.g. on the alters' relationship to ego) and ask this question for all alters before moving to the next question. The three different methods used in the study of Kogovšek et al. (2002) and in our example are a combination of the aforementioned choices regarding mode and technique.

1: Face-to-face mode with by-alters technique

2: Telephone mode with by-questions technique

3: Telephone with by-alters technique

We expected that the quality of indices of composition should be highest when using the face-to-face data collection mode. This expectation derives from studies comparing the two data collection modes, which conclude that, for cognitively demanding questions, face-to-face interviews are preferred and give data of higher quality (Kogovšek et al. 2002; Kogovšek and Ferligoj 2004; 2005).

Kogovšek et al. (2002) designed their study as a split-ballot MTMM design (Table 1).

The data were collected between March and June 2000 by computer-assisted telephone interview (CATI) and computer-assisted personal interview (CAPI) for a representative sample of 1033 inhabitants of the city of Ljubljana, Slovenia. The sampling frame was the telephone directory of Ljubljana. Respondents were randomly assigned to the three groups specified in the design. These respondents produced 7223 alters. The response rate was 25% which is at least partly owing to the fact that respondents were informed upon recruitment that they would be interviewed twice. 59% were women and 41% men. 16% were in single households, 44% in two-member households and about 33% in households with three or more members. More than half (56%) of respondents had no children, 25% had one, 17% two and about 3% three or more children. With regard to education, less than 10% of respondents had primary school or less, 43% had 4-year secondary school and more than a third had higher education. 36% of respondents were younger than 35 years and more than a quarter (28%) were older than 55. The sample consisted of 30% single respondents, 8% of widowers, 7% divorces and 55% married or living as married.

Results

Pierotti et al. (2009)'s approach in (12) and (13) was used for replacing zero components. Estimation was made by the direct ML method with the robust test statistics described in Yuan and Bentler (2000) and in Arminger and Sobel (1990). The Mplus6.1 program (Muthén and Muthén 2007) was used. Table 2 shows the descriptive results. The x'_{dm} scores are reported after zeroes have been substituted.

The x'_{dm} scores show roughly similar means for the different methods. All methods give the partner as the smallest component, others as the second to smallest, family as the second to largest and friends as the largest. Nearly all components have significant skewness and kurtosis coefficients. The smallest components (family and others) have rather extreme coefficients.

The y scores are relative to the 4th component (family). The mean values show friends to be a somewhat larger component than family and partner and others to be much smaller components. The y scores do not follow a normal distribution, some skewness coefficients being still significant, but the degree of non-normality is much reduced.

Table 3 shows the so-called MTMM matrix, which is the ordered correlation matrix among all measures. Correlations between the same trait using two methods are especially interesting. They are highest between methods 1 and 2 (.738, .634 and .586) and lowest between methods 2 and 3 (.615, .634 and .553). The diagonal sub-matrices show correlations among different traits using a common method and are all positive, as is often the case with alr transformed scores, which have a common denominator.

The CU model yielded a robust Yuan and Bentler χ^2 statistic 14.95 with 15 degrees of freedom and p-value=0.455. The 90 Percent C.I. for RMSEA (Root Mean Square Error of Approximation) was 0.000 to 0.029. Other usual goodness of fit measures also revealed an excellent fit. CFI (Comparative Fit Index) and TLI (Tucker and Lewis Index) were both 1.000. The SRMR (Standardized Root Mean Square Residual) was 0.022.

Measurement quality, as indicated by the standardized trait loadings in Table 4, is highest for method 1 (face to face by alters) for all three components (.883, .829, .786). The second method (telephone by questions) is better than the third (telephone by alters) regarding the first trait (ratio of partner over family: .816). Regarding the remaining two traits, the second and the third method behave about equally well (.771 vs. .782 and .750 vs. .737). In any case, differences among methods are not dramatic when compared to sampling variability, as confidence intervals overlap to a great extent. We expected that data quality of indices of composition should be highest when using the face-to-face data collection mode, which is a slower means of communication than telephone and makes it possible for respondents to give better answers to cognitively complex questions. This has been confirmed as the measurement quality estimates were highest for the face-to-face method. The main difference in method quality arises from the data collection mode, the data collection technique (order by questions or by alters) being a side issue.

In general, the partner to family ratio (trait 1) has the highest measurement quality, only method three deviating

Table 2: Descriptive statistics of the raw x'_{dm} compositional scores and the additive log ratio y_{dm} scores[a]

	Min	Max	Mean	St.dev.	Skewness	Kurtosis
x'_{11}	.013	.875	.124	.116	1.95*	6.92*
x'_{21}	.015	.938	.402	.235	0.09	-0.82*
x'_{31}	.015	.917	.145	.170	1.95*	3.96*
x'_{41}	.015	.906	.329	.210	0.42*	-0.38*
x'_{12}	.010	.458	.111	.093	1.13*	0.94*
x'_{22}	.016	.946	.405	.237	0.05	-0.83*
x'_{32}	.015	.917	.143	.167	1.88*	3.55*
x'_{42}	.013	.917	.340	.215	0.43*	-0.37*
x'_{13}	.010	.850	.112	.099	1.65*	5.00*
x'_{23}	.018	.953	.415	.223	-0.02	-0.64*
x'_{33}	.010	.850	.136	.148	1.65*	2.53*
x'_{43}	.013	.938	.338	.204	0.53*	-0.09
y_{11}	-3.84	3.04	-1.07	1.40	0.49*	-0.07
y_{21}	-3.77	4.12	0.21	1.60	-0.11	-0.06
y_{31}	-3.84	3.50	-1.08	1.60	0.72*	-0.05
y_{12}	-3.62	2.76	-1.17	1.29	0.39*	-0.24
y_{22}	-3.60	4.11	0.18	1.66	-0.09	-0.17
y_{32}	-3.78	3.50	-1.13	1.56	0.64*	-0.30
y_{13}	-3.81	2.83	-1.22	1.26	0.41*	-0.18
y_{23}	-3.83	4.11	0.22	1.53	-0.22*	0.28
y_{33}	-3.83	4.02	-1.18	1.49	0.62*	-0.12

[a] pairwise deletion

First subindex shows trait (1: partner; 2: friends; 3 others; 4: family).

Family is the common denominator of the alr y_{dm} scores and is therefore absent.

Second subindex shows method (1: Face to face by alters; 2: Telephone by questions; 3: Telephone by alters).

* Significant skewness or kurtosis (α=5%).

Table 3: Correlation matrix[a]

	y_{11}	y_{21}	y_{31}	y_{12}	y_{22}	y_{32}	y_{13}	y_{23}	y_{33}
y_{11}	1.000								
y_{21}	.504	1.000							
y_{31}	.519	.453	1.000						
y_{12}	.738	.382	.382	1.000					
y_{22}	.364	.634	.384	.564	1.000				
y_{32}	.332	.413	.586	.479	.474	1.000			
y_{13}	.651	.321	.282	.615	.357	.314	1.000		
y_{23}	.277	.626	.320	.340	.634	.376	.497	1.000	
y_{33}	.342	.339	.578	.307	.337	.553	.462	.489	1.000

[a] direct maximum likelihood with missing data

First subindex shows trait (1: partner; 2: friends; 3 others).

Second subindex shows method (1: Face to face by alters; 2: Telephone by questions; 3: Telephone by alters).

slightly from this pattern. This could result from the fact that the partner is the single most prominent provider of social support. The ratio involving other network members (trait 3) has the lowest measurement quality for all methods. This category likely comprises the respondents' weakest ties.

As regards the expected ratios (Table 4), the largest differences among methods are encountered in the relative size of the partner network to the family network (trait 1). Method 3 gives the smallest ratio (-1.221) and method 1 the largest (-1.047), method 2 being close to method 3. The differences among methods regarding relative size of the others' network to the family network (trait 3) go in the same direction, although they are of a lower magnitude. It would seem that method 1 tends to increase the relative size of the smallest components, compared to method 3. As regards the friends to family ratio, all methods behave about equally.

It must be considered that data about the type of relationship were collected in both the face-to-face and the telephone interviews in such a way that answer categories (11 categories, see name interpreter in the appendix) were read aloud to respondents. Evidence from other surveys suggests that reading of lists of answers can lead to a recency effect (the last answer or the few last answers are more often selected, de Leeuw 1992:68-69; Schwarz et al. 1992). This effect is

Table 4: Measurement quality estimates and 95% C.I. from the CU model[a]

	Standardized λ_{dm}: loadings			τ_{dm}: expected values		
	lower 95% limit	estimate	upper 95% limit	lower 95% limit	estimate	upper 95% limit
y_{11}	.811	.883	.955	-1.144	-1.047	-0.950
y_{21}	.757	.829	.901	0.103	0.216	0.330
y_{31}	.710	.786	.862	-1.188	-1.072	-0.957
y_{12}	.745	.816	.886	-1.265	-1.177	-1.090
y_{22}	.698	.771	.844	0.078	0.190	0.303
y_{32}	.671	.750	.830	-1.242	-1.135	-1.029
y_{13}	.687	.751	.815	-1.307	-1.221	-1.136
y_{23}	.716	.782	.847	0.099	0.203	0.308
y_{33}	.667	.737	.807	-1.290	-1.186	-1.083

[a]Loading and intercept parameters in equation (1)

First subindex shows trait (1: partner; 2: friends; 3 others).

Second subindex shows method (1: Face to face by alters; 2: Telephone by questions; 3: Telephone by alters).

Negative expected values show proportions of partner, friends and others to be lower than the proportion of family.

Positive expected values show proportions of partner, friends and others to be higher than the proportion of family.

more pronounced in telephone surveys. However, since the list of answers was composed in such a way that the most important social support providers were listed at the beginning (the first being the partner, followed by co-workers, co-members, neighbours and other categories), this may have caused a primacy effect for the face-to-face data collection mode. We can assume that partner is among the strongest ties and, as shown above, is measured with the highest measurement quality, but also as the first category on the list is perhaps most vulnerable to primacy measurement bias.

Discussion

Indicators of network composition measured with surveys have not yet been properly evaluated with MTMM models with regard to their measurement quality owing to the difficulties involved by the compositional nature of these indicators. In this article we highlight the necessary data transformations, the appropriate type of MTMM model and the required changes in the interpretation of model parameters. The CU model is preferable to alternative MTMM models including method factors and only trait loadings and intercepts are interpretable. Whenever possible, we have chosen the simplest options in order to narrow the gap between methodology and application. Both standard and split-ballot MTMM designs can be analysed, although the presented example is that of a split-ballot design. The conclusions which follow must be understood within the particular illustration and used as a guideline and example for the interpretation of the results.

The question evaluating the relationship between ego and members of his/her network was assessed with a rather long list (11 categories) of possible relations. We compiled a shorter list of amalgamated categories for the analysis (% of partner, % of family, % of friends, % of others), which follows the substantive logic of combining categories of network members of similar characteristics and fulfils the statistical requirement of avoiding categories with large proportions of 0 values.

A regards the data collection mode, we compared only the face-to-face and telephone surveys, with the univocal finding that the face-to-face survey gives higher measurement quality of network composition indicators. This is in line with previous research, however optimistic about telephone surveys, still giving priority to face-to-face surveys for cognitively demanding questions. Nevertheless, the original question used in both the face-to-face and telephone surveys, was read to respondents aloud without show-cards. In this setting, the long list of nominal answers can cause recency effects, so that categories read the last are given more attention by respondents and therefore are more likely to be selected. However, in our case, the first category was partner, which is the most prominent provider of social support and usually most close to respondents, causing the primacy effect instead (the first category receives the highest attention and choice), especially for the face-to-face method.

Several studies have addressed measurement quality of ego-centred networks measured with surveys (e.g. Kogovšek 2006; Kogovšek et al. 2002; Kogovšek and Ferligoj 2003; 2004; 2005). It would be important to include also web data collection as it is being used for network research in recent years (e.g. Kogovšek 2006; Lozar-Manfreda et al. 2004; Vehovar et al. 2008). Data collection technique (by questions or by alters) had no great effect on measurement quality, therefore indicating that for the two considered data collection modes we can use either. For web surveys, the by-question technique is more appropriate, owing to the specific structure of web questionnaires (Coromina and Coenders 2006).

Furthermore, in this example we have used data obtained by name generator questions which did not limit the number of mentioned alters. The approach described in this article is even more appropriate for evaluating measurement quality of network compositions with constrained name generator questions (imposing a certain number of alters for all respondents), with questions using the role relation format or with questions using the specific format of stressful events (Kogovšek and Hlebec 2008). Under these formats, total network size becomes constant and thus the researcher has no choice but to use compositional data approaches.

Further research in the area will include:

- extending the CU model to a full SEM in which components are predicted by a set of covariates or in which components are covariates predicting a set of outcome variables,
- evaluating the relative merits of other more complex data transformations such as the ilr,
- meta analyzing estimates of measurement quality of compositional data obtained in several studies and
- the web survey mode.

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability*. London: Chapman & Hall. (Reprinted in 2003 with additional material by The Blackburn Press)

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. Marcoulides & R. E. Schumacker (Eds.), *Advanced Structural Equation Modeling* (p. 243-277). Mahwah, NJ: Erlbaum.

Arminger, G., & Sobel, M. E. (1990). Pseudo maximum likelihood estimation of mean and covariance structures with missing data. *Journal of the American Statistical Association*, *85*(409), 195-203.

ATUS. (2010). *American Time Use Survey.* U.S. Bureau of Labour Statistics. www.bls.gov/tus.

Bittman, M., & Rice, J. M. (2002). The spectre of overwork: An analysis of trends between 1974 and 1997 using Australian time-use diaries. *Labour & Industry*, *12*(3), 5-25.

Burt, R. S. (1984). Network items in the general social survey. *Social Networks*, *6*(4), 293-339.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81-105.

Chan, W. (2003). Analyzing ipsative data in psychological research. *Behaviormetrika*, *30*(1), 99-121.

Cheung, M. W. L. (2004). A direct estimation method on analyzing ipsative data with Chan and Bentler's (1993) method. *Structural Equation Modeling*, *11*(2), 217-243.

Coenders, G., & Saris, W. E. (2000). Testing nested additive, multiplicative and general multitrait-multimethod models. *Structural Equation Modeling*, *7*(2), 219-250.

Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin*, *98*(2), 310-357.

Coromina, L., & Coenders, G. (2006). Reliability and validity of egocentered network data collected via Web. A meta-analysis of multilevel multitrait multimethod studies. *Social Networks*, *28*(3), 209-231.

De Leeuw, E. D. (1992). *Data Quality in Mail, Telephone, and Face to Face Surveys*. Amsterdam: TT-Publikaties.

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., & Barceló-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, *35*(3), 279-300.

Enders, C. K. (2001). The impact of non-normality on full information maximum likelihood estimation for structural equation models with missing data. *Psychological Methods*, *6*(4), 352-370.

Finkbeiner, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika*, *44*(4), 409-420.

Gold, M. S., Bentler, P. M., & Kim, K. H. (2003). A comparison of maximum likelihood and asymptotically distribution free methods of treating incomplete non-normal data. *Structural Equation Modeling*, *10*(1), 47-79.

Hirsch, B. J. (1980). Natural support systems and coping with major life changes. *American Journal of Community Psychology*, *8*(2), 159-172.

Kogovšek, T. (2006). Reliability and validity of measuring social support networks by web and telephone. *Metodološki zvezki*, *3*(2), 239-252.

Kogovšek, T., & Ferligoj, A. (2003). Merjenje egocentričnih omrežij socialne opore. *Družboslovne razprave*, *19*(43), 127-148.

Kogovšek, T., & Ferligoj, A. (2004). The quality of measurement of personal support subnetworks. *Quality & Quantity*, *38*(5), 517-532.

Kogovšek, T., & Ferligoj, A. (2005). Effects on reliability and validity of egocentered network measurements. *Social networks*, *27*(3), 205-229.

Kogovšek, T., Ferligoj, A., Saris, W. E., & Coenders, G. (2002). Estimating the reliability and validity of personal support measures: full information ML estimation with planned incomplete data. *Social networks*, *24*(1), 1-20.

Kogovšek, T., & Hlebec, V. (2008). Measuring ego-centered social networks: do cheaper methods with low respondent burden provide good estimates of network composition? *Metodološki zvezki*, *5*(2), 127-143.

Lader, D., Short, S., & Gershuny, J. (2006). *The Time Use Survey, 2005*. London: The national Amendment, Office for National Statistics.

Lee, S. Y. (1986). Estimation for structural equation models with missing data. *Psychometrika*, *51*(1), 93-99.

Lozar-Manfreda, K., Vehovar, V., & Hlebec, V. (2004). Collecting ego-centred network data via the Web. *Metodološki zvezki*, *1*(2), 295-321.

Marsden, P. V. (1987). Core discussion networks of Americans. *American Sociological Review*, *52*(1), 122-131.

Marsh, H. W. (1989). Confirmatory factor analysis of multitrait-multimethod data: many problems and few solutions. *Applied Psychological Measurement*, *13*(4), 335-361.

Martín-Fernández, J. A., Barceló-Vidal, C., & Pawlowsky-Glahn, V. (2003). Dealing with zeros and missing values in compositional data sets using non-parametric imputation. *Mathematical Geology*, *35*(3), 253-278.

Müller, C., Wellman, B., & Marin, A. (1999). How to use SPSS to study ego-centered networks. *BMS: Bulletin of Sociological Methodology*, *64*(1), 83-100.

Muthén, L. K., & Muthén, B. (2007). *Mplus User's Guide*. Los Angeles, CA: Muthén and Muthén.

Pentland, W. E., Harvey, A. S., Powell-Lawton, M., & McCoil, M. A. (1999). *Time Use Research in the Social Sciences*. New York: Kluwer Academic / Plenum Publishers.

Pierotti, M. E. R., Martín-Fernández, J. A., & Seehausen, O. (2009). Mapping individual variation in male mating preference space: multiple choice in a colour polymorphic cichlid fish. *Evolution*, *63*(9), 2372-2388.

Saris, W. E., & Gallhofer, I. N. (2007). *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. New York: Wiley.

Saris, W. E., Satorra, A., & Coenders, G. (2004). A new approach to evaluating the quality of measurement instruments: the split-ballot MTMM design. *Sociological Methodology*, *34*(1), 311-347.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. Clogg (Eds.), *Latent Variables Analysis: Applications to Developmental Research* (p. 399-419). Thousand Oaks, CA: Sage.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147-177.

Schumacker, R. E., & Lomax, R. G. (1996). *A Beginner's Guide to Structural Equation Modeling*. Mahwah, NJ: Lawrence Erlbaum.

Schwarz, N., Hippler, H. J., & Noelle-Neumann, E. (1992). A cognitive model of response-order effects in survey measurement. In N. Schwarz & S. Sudman (Eds.), *Context Effects in Social and Psychological Research* (p. 187-201). New York: Springer-Verlag.

Thió-Henestrosa, S., & Martín-Fernández, J. A. (2005). Dealing with compositional data: the freeware codapack. *Mathematical Geology*, *37*(7), 773-793.

Vaux, A. (1988). *Social Support, Theory, Research, and Intervention*. New York: Praeger.

Vehovar, V., Lozar-Manfreda, K., Koren, G., & Hlebec, V. (2008). Measuring ego-centered social networks on the web: questionnaire design issues. *Social networks*, *30*(3), 213-222.

Weiss, R. S. (1974). The provisions of social relations. In Z. Rubin (Ed.), *Doing unto Others* (p. 17-26). Englewood Cliffs: Prentice Hall.

Wills, T. A. (1985). Supportive functions of interpersonal relationships. In S. Cohen & S. L. Syme (Eds.), *Social Support and Health* (p. 61-82). Orlando: Academic Press.

Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, *30*(1), 165-200.

Appendix

Name generators used in this illustration

1. From time to time people borrow something from other people, for instance a piece of equipment, or ask for help with small jobs in or around the house. Who are the people you usually ask for this kind of help? (material support)

2. From time to time people ask other people for advice when a major change occurs in their life, for instance a job change or a serious accident. Who are the people you usually ask for advice when such a major change occurs in your life? (informational support)

3. From time to time people socialize with other people, for instance they visit each other, go together on a trip or to a dinner. Who are the people with whom you usually do these things? (social companionship)

4. From time to time, most people discuss important personal matters with other people, for instance if they quarrel with someone close to them, when they have problems at work, or other similar situations. Who are the people with whom you discuss personal matters that are important to you? (emotional support)

5. Suppose you would find yourself in a situation, when you would need a larger sum of money, but not have it yourself at the moment, for instance five average monthly wages (approximately 500.000 tolars[1]). Whom would you ask to lend you the money (a person, not an institution, e.g. a bank)? (financial support)

Name interpreter used in this illustration

In what type of a relationship are you with this person (for instance, friend, sibling, co-worker etc.)? If your relationship with this person can be described in more than one of the mentioned types of relations, you can specify more types of relations (for instance, the same person can be a co-worker and a friend to you).

1. partner
2. father or mother
3. brother or sister
4. child
5. other relative
6. co-worker
7. co-member
8. neighbor
9. friend
10. advisor
11. other

[1] ca. 2000 EUR.