

## **A Bayesian Approach to Social Relations Modelling**

Bachelor Thesis

27-06-2021

Alex Hartmann

s2139049

Supervisor: Dr. J.-P. Fox

(Second) assessor: Dr. R. Feskens

B.Sc. Psychology

Department of Research Methodology, Measurement, and Data Analysis (OMD)

University of Twente, Enschede, Netherlands

# Abstract

In this paper, a Bayesian social relations model (BSRM) for the analysis of round robin data is presented and compared to the well-known social relations model (SRM) of Kenny and La Voie (1984). The BSRM and SRM are evaluated regarding their performance in recovering the true parameters in a simulation study. Special attention is given to identify inter-individual relationship effects, which represent interaction effects between actors and partners. Three different types of inter-individual effects are considered: (1) approximately no effects, (2) varying effects across partners, and (3) constant moderate effects. In the end, an empirical application to child bullying data is provided.

It is shown that the BSRM point estimates for the inter-individual relationship effect variances are slightly overestimated, since the posterior distributions are skewed to the right. While both models are precise in their point estimates, the SRM is much more accurate in this regard. In the SRM, the standard errors are overestimated, leading to over-coverage of the 95% confidence intervals. The BSRM performs better with respect to the precision of the estimates. The BSRM comes with a powerful Bayes factor test for hypotheses about inter-individual relationship effects, which performs well. The power of the Bayes factor depends on the sample size, but relationship effects could be identified with small data sets with two outcome variables and 8-20 actors. It is also shown that the BSRM is more flexible than the SRM in modeling relationship effects.

# 1 Introduction

The study of attitudes and opinions is a growing topic both in the broader, political-societal sense as well as in lower-level social relations. This trend is illustrated by the fact that supply of attitude-related content on the internet is increasing. Thus, the general demand for attitude-related information appears to be high as well. It is possible to define attitudes as interpersonal relationship evaluations (Albarracín et al., 2018), showing how the study of social relations data is of interest now and likely in the future. Such studies are already commonplace and often consist of in-person group meetings where each person rates every other person on some scale. Therefore, one obtains dyadic data, where judgements happen in pairs, or dyads as defined in Becker and Useem (1942). There are also a variety of different approaches like simply administering questionnaires about peers as well as more complex approaches like the one of Carlson et al. (2018), who analyze dancing behaviour in groups. In another example, one could collect data on classroom relations and obtain a data set representing things such as the lovingness or distance in a relationship between each pair of students. Here, patterns could emerge such as that one person is particularly likeable and makes all other people score them high on the lovingness scale while they themselves are not very outgoing and, thus score all others high on distance. Then, we obtain data on two outcome variables for each dyad.

We call the type of data structure used here round-robin data, as first introduced by Gleason and Halperin (1975). A round-robin design (RR) is generally a design in which all possible pairs are formed from a group of people, having them interact with one another (Warner et al., 1979). The corresponding data set contains an observation (e.g., rating, judgement) about the relationship of person  $i$  to person  $j$ ,  $X_{ij}$  and vice versa. Person  $i$  is referred to as an actor when rating person  $j$ , who is referred to as the partner. A dyad made up of one actor and one partner constitutes the smallest type of clustering possible in social relations data. Each actor rates all partners, excluding themselves, in the RR. This is illustrated in Table 1 for a RR with  $n$  people. The main diagonal of the data matrix does not contain observations, denoted by *NA*, since actors do not rate themselves. The dyadic observation  $X_{ij}$  is the unit of analysis representing an interaction or a relation between two people.

It is possible to have multiple round-robin designs. For instance, for multiple groups a data matrix represents the dyadic measurements for the people in each group. There are

Table 1. Round robin data structure

Actor	Partner			
	1	2	...	$n$
1	NA	$X_{12}$	...	$X_{1n}$
2	$X_{21}$	NA	...	$X_{2n}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
$n$	$X_{n1}$	$X_{n2}$	...	NA

as many data matrices as there are groups. We represent such sets in a block diagonal matrix notation. Let  $A = \{A_1, A_2, \dots, A_G\}$  be a set of  $G$  data matrices as described in Table 1, then the complete data matrix is given by

$$\begin{pmatrix} A_1 & NA & \dots & NA \\ NA & A_2 & \dots & NA \\ \vdots & \vdots & & \vdots \\ NA & NA & \dots & A_G \end{pmatrix}. \quad (1)$$

The design is said to be balanced if each group has the same number of people and there are the same amount of outcomes per person. One can illustrate this using a data set of child relations (van Noorden et al., 2016). In this example data set, there are 272 dutch children rating each other in groups of 8 on 2 outcome variables each in a balanced RR. The topic is social distance and the items referred to (1) how child  $i$  likes to play with child  $j$  and (2) how inclined they were to help the other child if they needed it. All responses were collected on a scale from 0 to 100 (completely disagree - completely agree). This yields two observations per dyad and data matrices

$$\mathbf{A}_k = \begin{pmatrix} NA & X_{121} & \dots & X_{1j1} \\ NA & X_{122} & \dots & X_{1j2} \\ X_{211} & NA & \dots & X_{2j1} \\ X_{212} & NA & \dots & X_{2j2} \\ \cdot & \cdot & & \cdot \\ X_{i11} & X_{i21} & \dots & NA \\ X_{i12} & X_{i22} & \dots & NA \end{pmatrix} \quad (2)$$

for each group  $k$ , where  $X_{ijl}$  is person  $i$ 's dyadic judgement of person  $j$  on item  $l$ . If we

look closer at the first four children from group 1, we obtain the data matrix shown in Table 2.

*Table 2.* Case study data matrix for first 4 children in group 1

Actor	Outcome	Partner Index			
		1	2	3	4
1	1	<i>NA</i>	49	70	53
	2	<i>NA</i>	31	31	30
2	1	83	<i>NA</i>	3	47
	2	66	<i>NA</i>	6	51
3	1	97	18	<i>NA</i>	18
	2	96	95	<i>NA</i>	94
4	1	5	78	4	<i>NA</i>
	2	4	4	3	<i>NA</i>

Here, one observes that, for example, child one is rated quite extremely by their fellow students. Child two and three rate them quite highly on both outcomes (83, 66 and 97, 96), while child four rates them low (5, 4). Therefore, there seem to be some differences between actors in the rating of child one. Such inter-individual relationship effects are often of interest in social psychology.

Statistical models have been developed for the analysis of interpersonal relationships. A commonly used model (e.g., see Carlson et al. (2018), Kluger et al. (2020), Liao et al. (2018)) is the social relations model (SRM), developed by Kenny and La Voie (1984). The SRM describes each dyadic observation by disentangling three major components, an actor effect, a partner effect, and an interpersonal relationship effect. Hypothesis testing within the SRM is traditionally based on ANOVA procedures. As a result, testing multiple relationship effects requires multiple hypothesis testing. Furthermore, the relationship effects are represented by interaction parameters. Therefore, the number of model parameters increases quickly when increasing the number of actors/partners and the number of outcome variables, which also complicates the statistical testing of hypotheses concerning interaction effects.

In a Bayesian modeling approach, a more parsimonious SRM is represented, where the interaction effects are disentangled in a contribution of the intra-individual relationship correlation and of the inter-individual relationship correlation. In a marginal Bayesian modeling approach, the effects are modeled as dependencies in the covariance matrix of the observations and not as interaction parameters in the mean component. There-

fore, the number of intra-individual and inter-individual correlation parameters do not increase when increasing the number of outcome variables. The testing of homogeneity in relationship effects for an actor or partner simply leads to testing a single correlation parameter.

The model is referred to as the Bayesian Social Relations Model (BSRM). Bayesian hypothesis testing is employed to test the presence of relationship effects.

## 2 The SRM

The Social Relations Model (SRM) by Kenny and La Voie (1984) states that

$$X_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}, \quad (3)$$

where  $i$  is the actor index and  $j$  the partner index. (Goldring, 2020). The  $\mu$  represents the grand mean,  $\alpha_i$  the actor effect,  $\beta_j$  the partner effect and  $\gamma_{ij}$  the interaction (relationship) effect. For data for more than one round-robin design, the SRM becomes

$$X_{ij} = \mu + g_k + \alpha_i + \beta_j + \gamma_{ij}, \quad (4)$$

where  $k$  is the group index and  $g_k$  the specific group effect. Considering multiple outcome variables to increase the number of observations per dyad, indicating outcome variables by  $l$ , we obtain the equation

$$X_{ijl} = \mu + g_k + \alpha_i + \beta_j + \gamma_{ij} + e_{ijl}. \quad (5)$$

Here, we add an error term  $e_{ijl}$  to pick up random error variation for each outcome variable. The random components  $\alpha_i$ ,  $\beta_j$ ,  $\gamma_{ij}$  and  $e_{ij}$  are each independently normally distributed.

The model parameters can be estimated using an ANOVA approach. In the simulation study, this is done through the *TripleR* package (Schönbrodt et al., 2012) for the programming language R (R Core Team, 2020). The *TripleR* package estimates parameter values as well as variances for all parameters in Equation (4), except for the error term.

### 3 The BSRM

In the conditional BSRM, observations for outcome  $l$  are described by actor, partner and interaction effect parameters

$$X_{ijl} = \alpha_i + \beta_j + \gamma_{bij} + \gamma_{aji} + e_{ijl}. \quad (6)$$

Note that the group effect  $g_k$  is missing from this equation. In the BSRM, this group effect is modeled as a factorized actor effect. Thus, it is a collective actor effect that is shared among all members of a group. The random components have the prior distributions

$$\begin{aligned} \alpha_i &\sim N(\mu_a, \sigma_a^2) \text{ (actor effect),} \\ \beta_j &\sim N(\mu_b, \sigma_b^2) \text{ (partner effect),} \\ \gamma_{bij} &\sim N(\beta_j, \tau_{bj}) \text{ (inter-individual relationship effect),} \\ \gamma_{aji} &\sim N(\alpha_i, \tau_{ai}) \text{ (intra-individual relationship effect),} \\ e_{ijl} &\sim N(0, \sigma^2) \text{ (measurement error).} \end{aligned}$$

In this conditional BSRM, the interaction effects are explicitly parameterized. In our (marginal) BSRM approach, the interaction parameters are integrated out, which leads to a multivariate modeling approach, where the covariance matrix represents dependencies due to relationship effects. Let  $\mathbf{X}_j$  represent the observations of  $n$  actors and partner  $j$  on the  $m$  outcome variables and denote by  $\mathbf{J}_m$  an  $m \times m$  matrix of ones. The BSRM assumes a multivariate normal distribution for the observations for partner  $j$  with mean term

$$\boldsymbol{\mu}_j = (\boldsymbol{\alpha} \otimes \mathbf{1}_m) + \mathbf{1}_{nm}\beta_j \quad (7)$$

and covariance matrix

$$\boldsymbol{\Sigma}_j = (\mathbf{I}_n \otimes \{\sigma^2 + \tau_{ai}\}\mathbf{I}_m) + (\mathbf{I}_n \otimes \mathbf{J}_m\tau_{bj}). \quad (8)$$

Here, the symbol  $\otimes$  refers to the Kronecker product, first described by Zehfuss in 1858 (Henderson et al., 1983). The error variances on the diagonal consist of components  $\sigma^2 + \tau_i$  for  $i = 1, \dots, n$ , which is denoted with the curly brackets. The covariance components are

represented by a block diagonal matrix with blocks  $\mathbf{J}_m\tau_{bj}$ . They represent the common covariance among observations from the same actor to partner  $j$ . The parameter  $\tau_{bj}$  represents the variation in inter-individual relationship effects. When  $\tau_{bj}$  is equal to zero, there are no inter-individual relationship effects for partner  $j$ . When  $\tau_{bj}$  is greater than zero, the inter-individual relationship effects differ across actors, and differences increase for increasing  $\tau_{bj}$ .

Consider the observations from actor  $i$ . They are also assumed to be multivariate normally distributed with mean

$$\boldsymbol{\mu}_i = (\boldsymbol{\beta} \otimes \mathbf{1}_m) + \mathbf{1}_{nm}\alpha_i \quad (9)$$

and covariance matrix

$$\boldsymbol{\Sigma}_i = (\mathbf{I}_n \otimes \{\sigma^2 + \tau_{bj}\}\mathbf{I}_m) + (\mathbf{I}_n \otimes \mathbf{J}_m\tau_{ai}). \quad (10)$$

The parameter  $\tau_{bj}$  represents the variation in intra-individual relationship effects. When  $\tau_{ai}$  is equal to zero, there are no intra-individual relationship effects for actor  $i$ . When  $\tau_{ai}$  is greater than zero, the intra-individual relationship effects differ across partners, and differences increase for increasing  $\tau_{ai}$ .

By modeling dependencies induced by relationship effects, dyad-specific model parameters are not needed. So, increasing the number of dyads or the number of observations per dyad, does not lead to more relationship effect parameters. This reduces the computational efforts that are needed to estimate the model parameters. Furthermore, testing the presence of relationship effects can be done by examining hypotheses concerning  $\tau_{ai}$  and/or  $\tau_{bj}$ , which is much easier than testing the non-equivalence of multiple interaction effects.

The covariance parameters  $\tau_{ai}$  and  $\tau_{bj}$  are allowed to be negative. Negative relationship effect correlations cannot be modeled with interaction effects. So, these kind of relationships cannot be captured by the SRM. A negative covariance  $\tau_{ai}$  states that if actor  $i$  scores high on one outcome a low score is expected on another outcome for the same partner  $j$ . For instance, for the example of Back and Kenny (2010), uniquely smiling at a particular person is assumed to be positively related to uniquely choosing this person. However, it is possible that there is a negative relationship between smiling and uniquely

choosing this person.

A Bayes factor test is developed for evaluating the inter-individual relationship effects. If  $\tau_{bj} > 0$ , it indicates that there are unique relationship effects regarding partner  $j$ . Actors have generally different opinions about partner  $j$ . If  $\tau_{bj} = 0$ , it means that there are no inter-individual relationship effects, and actors rate partner  $j$  exactly according to their corresponding actor effects on each outcome  $l$ . The final case,  $\tau_{bj} < 0$ , means that there are outcome-specific (inter-individual) relationship effects for partner  $j$ .

Testing for this kind of structure is done by computing the logarithm of the Bayes factor. In order to do this, three hypotheses are defined as

$$H0 : \tau_{bj} = 0,$$

$$H1 : \tau_{bj} < 0,$$

$$H2 : \tau_{bj} > 0.$$

One obtains a Bayes factor for each partner  $j$ , expressed as

$$BF_{02j} = \log \left( \frac{p(\tau_{bj} = 0 \mid \mathbf{X})}{p(\tau_{bj} > 0 \mid \mathbf{X})} \right), \quad (11)$$

where  $BF_{02j}$  refers to testing hypothesis 0 against hypothesis 2 for partner  $j$ . Thus, high  $BFs$  represent (overwhelming) evidence for the hypothesis  $\tau_{bj} = 0$ , while negative  $BFs$  represent (overwhelming) evidence for  $\tau_{bj} > 0$ . The BSRM can be applied to both binary and continuous data. If we use continuous data, the model assumes a multivariate normal distribution. Therefore, this must be kept in mind when analyzing data sets.

## 4 Simulation Program

### 4.1 Technical details

We now move on to testing the performance of BSRM using simulated data. In order to do this, some of the simulation groundwork has to be laid. First, simulated data is generated according to the SRM model. By doing this, we ensure that the BSRM is not

given an advantage through fitting data simulation. Thus, an observation  $X_{ijl}$  is given by

$$X_{ijl} = g_k + \alpha_i + \beta_j + \gamma_{ij} + e_{ijkl}. \quad (12)$$

The error term  $e_{ijkl}$  represents the random error for the dyad  $ij$  in group  $k$  on outcome  $l$ . Here, actor, partner, and relationship effects are sampled from normal distributions. The group effects are sampled without taking any covariances into account. The actor and partner effects are drawn from a multivariate normal distribution. Their variance-covariance matrix is defined as

$$\Sigma_{\alpha\beta} = \begin{pmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 \end{pmatrix}. \quad (13)$$

The relationship effects in the SRM,  $\gamma_{ij} \sim N(0, \tau_{bj})$ , are assumed to be independently normally distributed with a partner specific variance parameter to describe inter-individual relationship effects. The SRM also includes a relationship effect covariance  $\sigma_{\gamma\gamma}$ , which represents the covariance between opposite elements in the relationship effect matrix. Note that for this simulation study, only inter-individual relationship effects are considered. They play a significant role further on. Measurement errors are simulated from a normal distribution with mean zero and variance  $\sigma^2$ .

For the BSRM, consider the case of  $m = 2$  outcome variables. The BSRM is represented by

$$\mathbf{X}_{ij} = \alpha_i + \beta_j + \mathbf{E}_{ij},$$

where the actor and partner effects are multivariate normally distributed with a covariance matrix displayed in Equation (13). Group effects can be modeled by adding a matrix of explanatory variables to the BSRM which can specify random and fixed effects, such as shared group effects across actors. The error terms in BSRM are simulated using a multivariate normal distribution. They are simulated for each person  $i$  on  $m$  outcome variables, where  $m = 2$  for each actor  $i$ . The inter-individual relationship effects are simulated for each partner  $j$ . The covariance matrix for dependencies implied by inter-

individual relationship effects and the random errors is represented by

$$\Sigma_{bj} = \begin{pmatrix} \sigma^2 + \tau_{bj} & \tau_{bj} \\ \tau_{bj} & \sigma^2 + \tau_{bj} \end{pmatrix}. \quad (14)$$

Here, the model for simulating data is the SRM, while models of analyses are both SRM and BSRM. The BSRM performance can be compared to the SRM under, presumably, optimal conditions for the SRM. In this simulation study, data is simulated with groups of up to 20 members and analysed using *TripleR* and *BSRM* R packages. All conditions are replicated  $r = 500$  times. Each BSRM replication is conducted with 5000 MCMC iterations and a burn-in of 1000. The true values for the variance parameters for all sampled effects are given in Table 3. The values were similar to the simulation study of Nestler et al. (2020).

Table 3. True parameter values for the validation study.

$\sigma_g^2$	$\sigma_\alpha^2$	$\sigma_\beta^2$	$\sigma_{\alpha\beta}$	$\tau_{bj}$	$\sigma_{\gamma\gamma}$
.1	.25	.25	.075	.4	0

*Note.* For the SRM,  $\tau_{bj}$  represents the partner-specific relationship effect variance and  $\sigma_{\gamma\gamma}$  the relationship covariance.

## 4.2 Simulation study validation

The validity of the simulation program was tested. To achieve this, the SRM was applied to simulated data sets with 10 groups of 10 members. For all simulation parameters, the RMSE, bias as well as the standard error values were recorded using *TripleR*. For example, consider the actor variance  $\sigma_\alpha^2$ . Then, the RMSE was calculated as

$$RMSE = \sqrt{\frac{\sum_{i=1}^r (\sigma_{\alpha i}^2 - \hat{\sigma}_\alpha^2)^2}{r}}, \quad (15)$$

for  $r$  replications, where  $\sigma_{\alpha i}^2$  are estimated values and  $\hat{\sigma}_\alpha^2$  represent the population values. The mean bias was computed as

$$Bias = \frac{\sum_{i=1}^r \sigma_{\alpha i}^2 - \hat{\sigma}_\alpha^2}{r}, \quad (16)$$

and reported with RMSE and parameter estimates in Table 4. The reported parameter estimates refer to mean estimates over all simulation runs provided by *TripleR*. One can

Table 4. RMSE and bias values for *TripleR* data estimates

Measure	$\sigma_\alpha^2$	$\sigma_\beta^2$	$\sigma_{\alpha\beta}$	$\tau_{bj}$	$\sigma_{\gamma\gamma}$
<i>RMSE</i>	.0461	.0443	.0302	.0234	.0191
<i>Bias</i>	-.0082	-.0007	-.0033	.0021	.0028
<i>Estimate</i>	.2418	.2493	.0717	.4021	.0028

Table 5. Standard errors for *TripleR* data estimates

Measure	$\sigma_\alpha^2$	$\sigma_\beta^2$	$\sigma_{\alpha\beta}$	$\tau_{bj}$	$\sigma_{\gamma\gamma}$
SE	.0845	.0932	.0676	.0344	.0344

observe here, that bias values were quite close to zero without obvious patterns, meaning that *TripleR* neither over- nor underestimated parameters. Thus, it was concluded that the simulation data matched the SRM model specifications. RMSE values were generally quite low, regarding the overall magnitude of parameter values discussed here. The estimates were also very close to their true values, deviating at most (leaving out the error variance  $\sigma^2$ ) by 4.60% of the true value in case of the actor-partner covariance parameter  $\sigma_{\alpha\beta}$ . Furthermore, in Table 5, standard errors (SEs) for all parameter estimates are reported.

Note that since the *TripleR* package does not report the group variance for bivariate analyses, it is missing from the Table 4. *TripleR* also does not report the error variance  $\sigma^2$ , so it is missing as well. In general, it can be concluded that the simulate data function operated well, given the very low RSME values and the small biases.

## 5 Simulation Study

In this section, BSRM and SRM analyses were run on simulated datasets, to evaluate and compare their performances. Different scenarios were examined with varying types of the inter-individual relationship effects. All analyses were run according to the procedure defined in Section 4.1, except that simulation parameters were varied. Since *TripleR* is unable to analyze data with more than two outcomes per dyad, we ran all analyses with two outcome variables for now. We analyzed three scenarios, each of which was analyzed

using groups of either  $m = 8$ ,  $m = 12$  or  $m = 20$  people, yielding nine situations total. The first scenario was that there were very little inter-individual relationship effects, with  $\tau_{bj} = .05$ . Second, we increased  $\tau_{bj}$  over partners, from 0 up to 1, denoted by  $\tau_{bj} \in (0, 1)$ . Third, we analyzed data where  $\tau_{bj}$  did not differ across partners and was fixed at 0.5. For all scenarios the error variance was  $\sigma^2 = 0.1$ . To summarize, we obtain Table 6, representing all analyzed basic scenarios. First, we examined the estimation performance

*Table 6.* Simulation study scenarios

Scenario	1.1	1.2	1.3	2.1	2.2	2.3	3.1	3.2	3.3
$\tau_{bj}$	.05	.05	.05	(0,1)	(0,1)	(0,1)	.5	.5	.5
$m$	8	12	20	8	12	20	8	12	20

of both models. To that end, we listed RMSE and mean Bias values, calculated according to Equations 15 and 16, for all relevant parameters over all replications. For small values of  $\tau_{bj}$ , the simulated parameter distribution was quite skewed since data sets were not simulated with negative values for  $\tau_{bj}$ . Therefore, the posterior mean estimate for  $\tau_{bj}$  has proven not to be very accurate, especially for very skewed posterior distributions. In that case, the posterior mean can deviate quite far from the posterior mode, which would be the ideal estimator for the true values but is very difficult to compute accurately. To investigate whether the posterior distribution still serves as a good basis for Bayes factor testing, we compute 95% coverage rates (CR). Furthermore, we note that for scenario 2, the SRM was not expected to do very well. This is because the SRM assumes a normal distribution with a constant variance  $\tau_{bj}$  over all partners, thus making the identification of increasing  $\tau_{bj}$  values impossible.

Since data were simulated under the SRM, where  $\tau_{bj}$  is a variance parameter,  $\tau_{bj}$  must be positive. This means that simulations for scenarios 1 and 2, where  $\tau_{bj}$  can be quite close to 0, yielded skewed simulated parameter distributions over all replications. Since  $\tau_{bj}$  is not limited in this way in the BSRM (because it represents a covariance parameter), the BSRM assumes a more symmetric distribution. CRs were expected to be relatively high for the BSRM in cases where  $\tau_{bj}$  is very small, due to the aforementioned skewness problem with small  $\tau_{bj}$  values.

In Table 7, one can see that scenario 1 lead to relatively low RMSE and bias values under the BSRM. Since the true value was .05 here, RMSEs ranging from .0251 to .0548 are still quite substantial. Generally, bias as well as RMSE decreased with increasing group

size though, to the point where RMSE and bias indicated good estimation performance (for  $m = 20$ ). The CRs were relatively high, ranging from .9879 to .9982, indicating serious overcoverage. Therefore, we concluded that the sampled posterior distribution of  $\tau_{bj}$  was quite wide. This was because of the skewness of the simulated parameter distribution, which does not include negative values, while the BSRM expects some. Furthermore, high  $\tau_{bj}$  values in scenario 1 are rare, which means that most data sets had  $\tau_{bj}$  values close to zero. So, the extent to which the CR shows overcoverage could get smaller when more replications were done, thereby increasing the number of data sets with high  $\tau_{bj}$  values.

Here, CRs also got closer to the nominal 95% with increasing group size. For scenario 2, RMSE and bias values were high, indicating mediocre estimation performance. The CRs were good (ranging from .9475 to .9530), indicating that the posterior distribution fit the simulated parameter distribution well. We also need to take into account that a non-informative prior stretches the posterior more. In the replication of the data this prior uncertainty is not taken into account. Data is generated for fixed  $\tau_{bj}$  values without using any prior distribution. However, the posterior distributions for  $\tau_{bj}$  include this prior uncertainty which leads to slightly wider posteriors and overcoverage when the sample size is small.

For scenario 3, RMSE and bias values were quite similar to scenario 2, while CRs were a little lower, indicating slight undercoverage. Overall, we observed that the BSRM did not do very well in estimating  $\tau_{bj}$  with the mean of the posterior distribution. Furthermore, all biases were positive, indicating that the posterior mean under the BSRM tended to slightly overestimate  $\tau_{bj}$ , which is attributable to the skewed distribution of  $\tau_{bj}$  in the simulations. Since this distribution was often skewed to the right, the mean estimate landed to the right of the mode, leading to overestimation. In terms of CRs, the BSRM

Table 7. BSRM and SRM estimation performance for  $\tau_{bj}$  with two outcomes, by scenario

Model	Measure	Scenario								
		1.1	1.2	1.3	2.1	2.2	2.3	3.1	3.2	3.3
BSRM	<i>RMSE</i>	.0548	.0357	.0251	.6457	.3930	.2402	.5826	.3394	.2324
	<i>Bias</i>	.0113	.0035	.0012	.2367	.1059	.0516	.2235	.1077	.0554
	<i>Coverage</i>	.9982	.9937	.9879	.9475	.9483	.9530	.9250	.9395	.9353
SRM	<i>RMSE</i>	.0144	.0094	.0057	.3535	.3255	.3074	.1210	.0727	.0464
	<i>Bias</i>	.0018	-.0002	-.0000	.0042	-.0025	-.0008	-.0054	-.0012	.0050
	<i>Coverage</i>	1	1	1	.9423	.8810	.7010	1	1	1

did quite well, yielding overcoverage only in the case of scenario 1, when the sample size was small and the prior influence was noticeable. Overall, since CRs were quite good and the point estimate is not of direct relevance for the BSRM analysis, we concluded that the model did well under these scenarios.

In scenario 1, the SRM did very well estimating  $\tau_{bj}$ , with very low RMSE (ranging from .0057 to .0144) as well as bias values (ranging from  $-.0002$  to .0018). CRs were all 1, indicating vast overcoverage. In scenario 2, the SRM did worse. As mentioned before, this was expected, since the scenario violates the SRM assumption of equal  $\tau_{bj}$  across partners. Here, we observed substantial RMSEs (ranging from .3074 to .3535). For small groups, this better than the BSRM estimation, although the BSRM estimator worked better for  $m = 20$ . CRs were quite low (ranging from .7010 to .9423), indicating undercoverage. For scenario 3, the SRM estimation worked very well again, with RMSEs ranging from .0464 to .1210 and biases from  $-.0054$  to .0050. Here, we again observed CRs of exactly 1 for all three group sizes. Overall, the SRM did very well for scenarios 1 and 3, yielding very low RMSE and bias values. For scenarios 1 and 3, the CRs were exceptionally high, meaning that there likely was something wrong with the 95% confidence interval computations. Since *TripleR* only reports a standard error (from which standard deviations for CRs were calculated), which in itself is hard to estimate for variance components, we assumed that the issue here was overestimation of standard errors through *TripleR*.

Apart from  $\tau_{bj}$ , the other parameters were also estimated and estimation performances were analyzed. Both partner and actor variance biases were very close to zero for both models (absolute values less than .1 in all cases). RMSEs were generally quite substantial, ranging from .1 to .19 for the BSRM and from .01 to .11 for the SRM, but also not extraordinarily high. Thus, the models performed well in this regard, with the SRM clearly performing better in terms of point estimation. Furthermore, regarding the other parameters, it also became clear that the sample size had a significant effect on estimation performance. For both models and all scenarios, RMSEs decreased with increasing group sizes. Furthermore, another important parameter was  $\sigma_{\alpha\beta}$ , which the model estimated quite accurately and precisely, with all mean values around the specified .075 with biases of  $\pm .015$ .

When wanting to test a hypothesis about  $\tau_{bj}$  values with the SRM, one must rely on point estimates. Since the parameter space for  $\tau_{bj}$  is  $(0, \infty)$  under the SRM, the point

estimate can never be exactly 0. Therefore, the SRM does not provide a way to test for the existence of inter-individual relationship effects (where  $H_0$  is  $\tau_{bj} = 0$ ). For larger  $\tau_{bj}$ , this is not an issue, since the true value is then not on the lower bound of the parameter space, allowing for a symmetric sampling distribution, meaning that the true value can be better estimated. Since the BSRM treats  $\tau_{bj}$  as a covariance, negative values are allowed, making estimations around 0 easier. In the BSRM, testing zero  $\tau_{bj}$  is done with the Bayes factor.

To illustrate this type of BSRM hypothesis testing, we also analyze  $BF_{02j}$  obtained in this simulation study.  $BF$ s are plotted against partner indices for all scenarios in Figure 1. For the first scenario (1.1 through 1.3), we found that the Bayes factor did well to identify the absence of inter-individual relationship effects. We concluded this because, for all three group sizes, Bayes factors were greater than 2, indicating substantial evidence for  $H_0$ . For the second scenario, we observed that the Bayes factor was able to pick up the changes in  $\tau_{bj}$ , illustrated in Figure 1. We saw that with increasing partner index,  $BF_{02j}$  decreased. This indicated increasing evidence for  $H_2$ . This matched the specified structure for these scenarios, thus indicating that the BSRM does well under conditions of scenario 2. In scenario 3, we observed that the BSRM did not pick up the presence of  $\tau_{bj}$  ( $BF$ s were non-negative), although the sample mean  $\tau_{bj}$  estimate was not very far off (see Table 7). Therefore, we concluded that the issue here was likely the lack of data evidence with only two outcome variables, leading to a more flat  $\tau_{bj}$  posterior with a less pronounced mode.

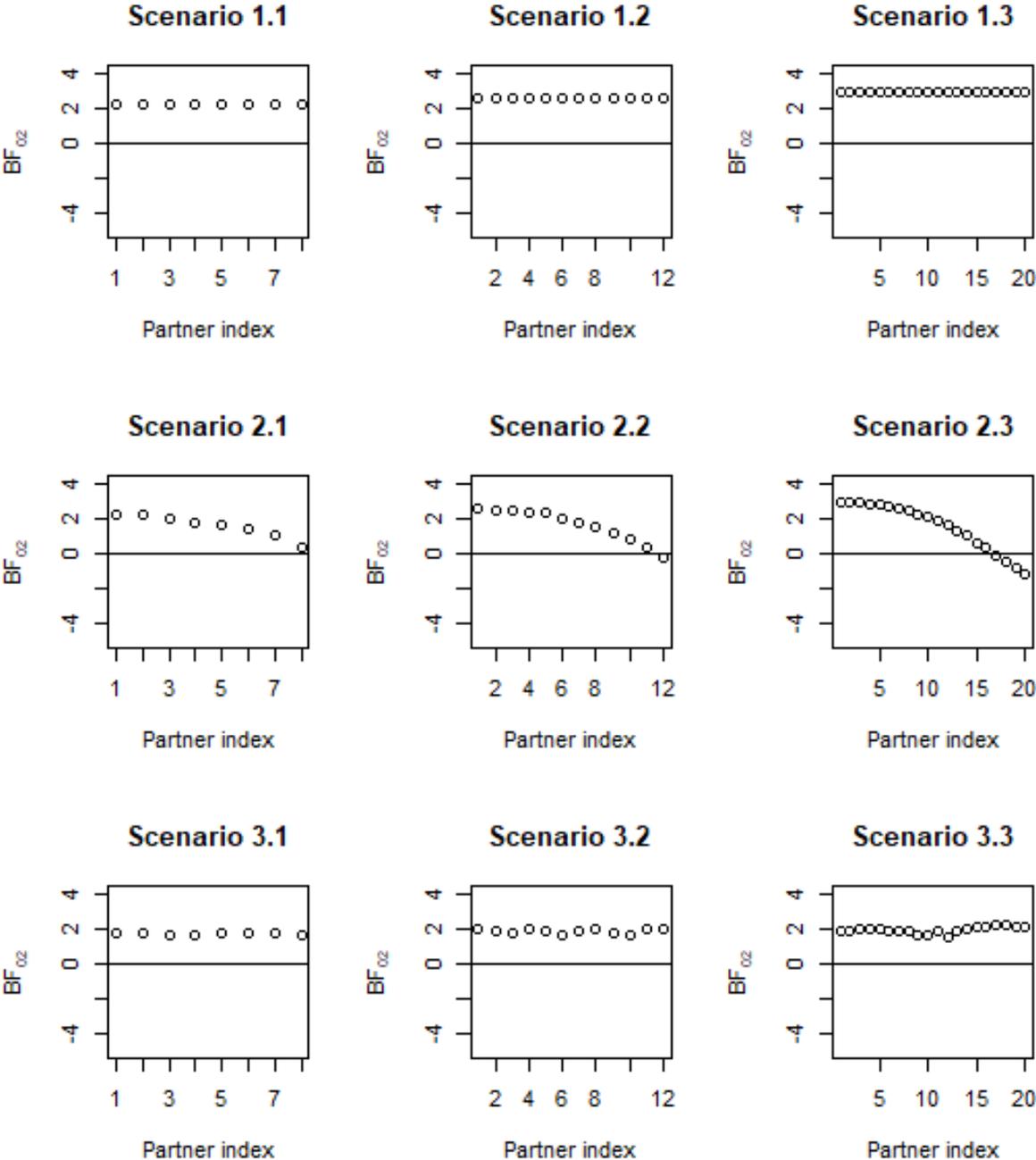


Figure 1.  $BF$ s plotted against partner indices for all scenarios and two outcomes per dyad.  
Note. The solid line at  $\log(BF_{02}) = 0$  represents change in evidence from  $H_0$  to  $H_2$ .

To explore the model’s performance further, we ran the analysis again, this time with five outcome variables. This is a particular strength of the BSRM, as the SRM, using *TripleR*, is unable to perform these analyses. Apart from the outcome count, all scenarios were kept as in the previous analysis. In Table 8, obtained  $\tau_{bj}$  estimation performance measures are listed.

Table 8. BSRM estimation performance for  $\tau_{bj}$  with five outcomes, by scenario

Model	Measure	Scenario								
		1.1	1.2	1.3	2.1	2.2	2.3	3.1	3.2	3.3
<i>BSRM</i>	<i>RMSE</i>	.0197	.0119	.0085	.6668	.3818	.2582	.5675	.3274	.2018
	<i>Bias</i>	.0097	.0042	.0029	.2214	.1020	.0656	.2035	.0906	.0378
	<i>Coverage</i>	.9975	.9892	.9877	.9110	.9153	.9522	.8800	.9010	.9216

Overall, RMSE and bias values were very similar to the analysis with two outcome variables. We also observed that biases were strictly positive, as in the previous analysis, indicating that the BSRM also tends to overestimate  $\tau_{bj}$  slightly for five outcomes. In terms of CRs, the first scenario did not change significantly. The CRs still indicate serious overcoverage. This is no surprise since we already found an explanation for this earlier in this section. The second scenario yielded quite decent CRs, while the third indicated again slight undercoverage. In general, they improved with increasing group size for all three scenarios.

As mentioned earlier, what the BSRM adds to the traditional SRM is the *BF* hypothesis test. We present discussing these results beginning with scenario 1. For scenario one, we saw that  $BF_{02j}$  was greater than 2 for all three group sizes and all partners, indicating strong evidence for  $H_0$ . Therefore, we concluded that the model identified the absence of inter-individual relationship effects successfully. What can also be observed, is that evidence for the hypothesis  $\tau_{bj} = 0$  increased with group size, as the absolute value of *BFs* increased with group size, indicating increasing evidence. For scenario two, we observed a clear downward trend of *BFs* as partner indices increased. Therefore, we also conclude here, that the BSRM is able to identify these increasing variances in inter-individual relationship effects. Here, again, and also in the next scenario, we can also observe that the magnitude of *BFs* increases with group size. This is simply because the more people there are in a group, the more evidence can be found for a certain hypothesis. For the third scenario, we observe that *BFs* do not change significantly with partner index. For

the group of eight, the statistical evidence for either hypothesis was quite low, with  $BFs$  below the threshold for substantial evidence, established by (Jeffreys, 1998). With the other group sizes, the model correctly identified the presence of  $\tau_{bj}$  with all the  $BFs$  below 0, thus favouring the hypothesis  $\tau_{bj} > 0$ . We concluded that the model correctly identified the presence of constant inter-individual relationship effect variance. In general, we can say that the model performed very well under the simulated circumstances, although we observed that the statistical evidence carried by reported  $BFs$  depended heavily on the sample (or group) size.

From this simulation study, we concluded that, as expected, in terms of point estimation, the SRM clearly outperforms the BSRM. Still, the BSRM was able to identify all patterns of inter-individual relationship effects, given a large enough sample. Furthermore, the BSRM has some advantages in terms of flexibility. First, the BSRM performed better when  $\tau_{bj}$  varies across partners, since the SRM is not designed for that. Second, the BSRM has the advantage of being able to handle (technically) unlimited amounts of outcome variables per dyad, whereas the SRM through *TripleR* can only handle two. Generally, both models performed well under the circumstances they were expected to.

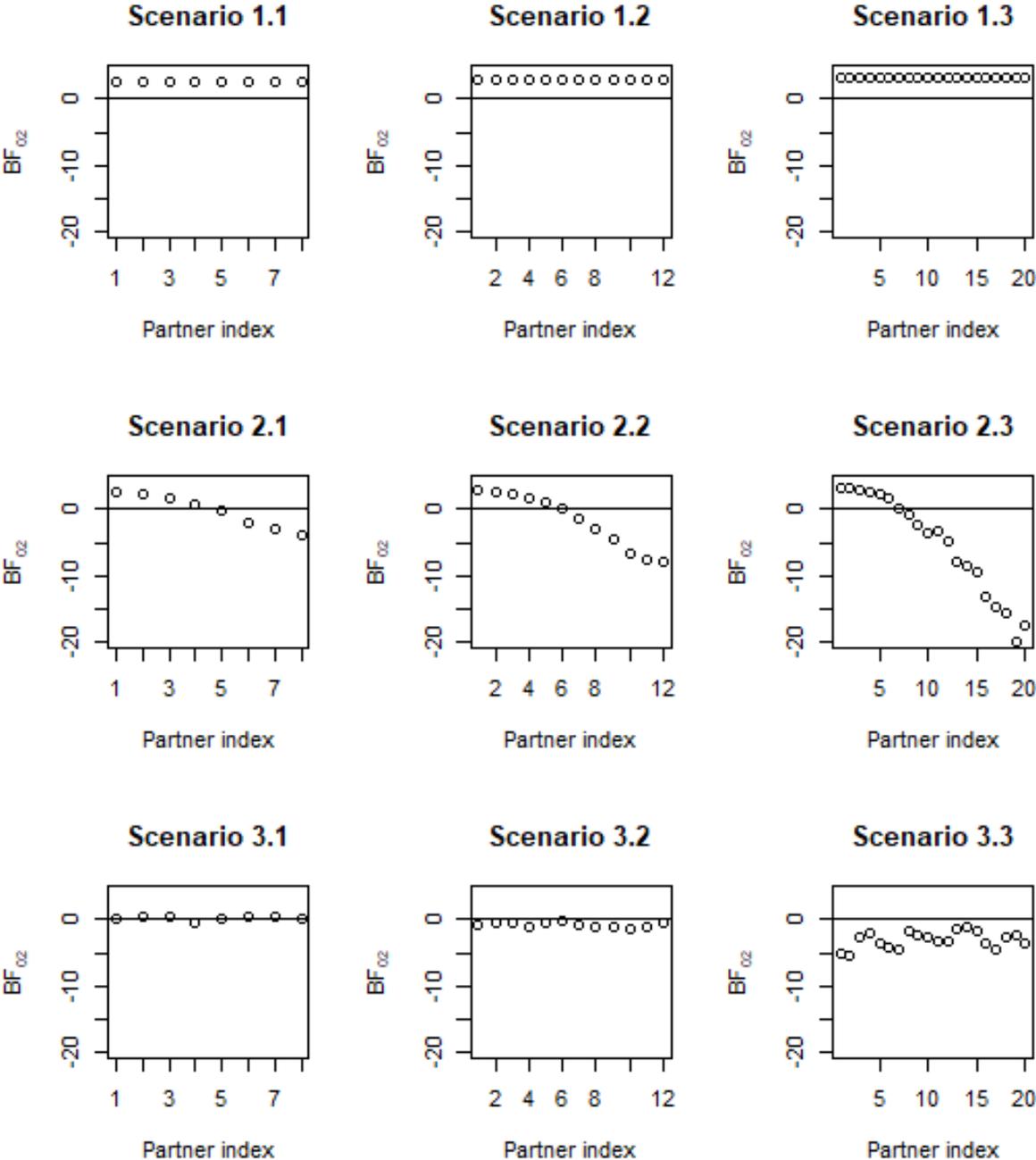


Figure 2. Bayes factors plotted against partner indices for each scenario and five outcomes per dyad.  
Note. Line represents change in evidence from  $H_0$  to  $H_2$ .

## 6 Case Study

In this case study, we apply the BSRM to a real world scenario. We use the aforementioned data set collected by van Noorden et al. (2016). To reiterate, it is a data set of 34 groups of each 8 children. This data set was obtained by admitting a test to many more children in order to obtain bullying and victimization scores. These were then used to sample the final groups of participants. Each group of eight was selectively sampled from their class out of 34 Dutch schools, see Table 9. Children were separated into four different roles, by deviation from mean scores. So a child that scored 1 SD higher on the bullying scale than the average child, got the role of bully. This happened for each group, so that each group had exactly two children (one male, one female) for each of the four categories.

Table 9. Child indices by gender and role

Gender	Role			
	Bully	Victim	Bully & Victim	Noninvolved
Male	1	3	5	7
Female	2	4	6	8

Then, in a follow-up session, each child scored each other child on two outcome variables each from 0 to 100. The first one was regarding how likely they (actors) were to play with the other (partners) and the second one measured how likely actors were to help their respective partners. Therefore, the data set had 34 groups of eight, scored on two outcome variables per dyad in a RR experiment. Note that in the original data set, there was one missing case, thus making the design unbalanced. Therefore, the smaller group was extended with a fictional child that had missing values on all outcome variables.

Inspection of the raw data distribution showed that it was not normally distributed with a lot of clustering around the lower and upper limits of the scale. Therefore, to improve the fit of the BSRM, the data was log-transformed. We then analyzed this data set using the *BSRM* R package and obtained Bayes factor results for the three hypothesis tests described in section 3. As in the simulation study, this analysis was conducted with 5000 MCMC iterations and a burn-in of 1000.

In order to analyze gender differences in this data set, gender specific effects were added to the model through a grouping variable. Therefore, actors were not only separated into the original 34 groups but also into groups for both genders. In the model, this then

yielded group effects for each of the 34 groups, but also for both genders. Then, the  $BF$ s were plotted against partner indices in Figure 3, where blue dots represent male children and red crosses represent female children.

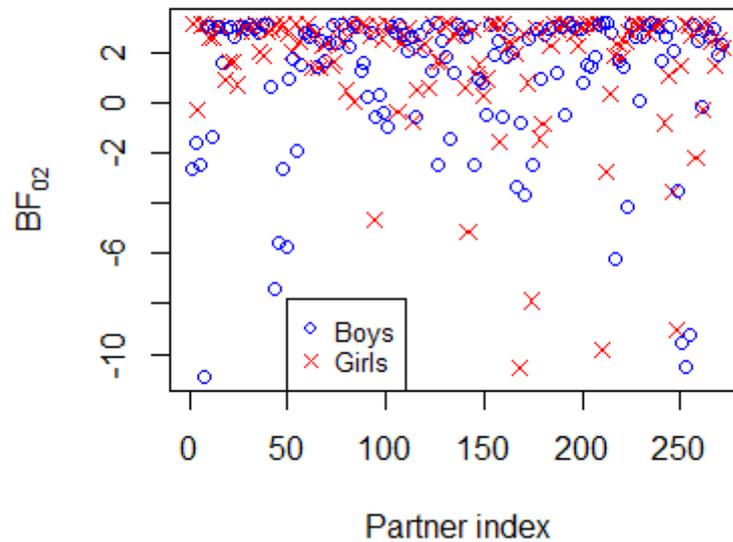


Figure 3. Bayes factors plotted against partner indices for case study.

One can observe here that, for the vast majority of partners,  $BF_{02j} > 0$ , indicating that  $\tau_{bj}$  is likely to be 0. Therefore, we concluded that there were relatively few children, where inter-individual relationship effects could be identified. Therefore, children were mostly of similar opinions regarding these partners. Still, there were also a few that showed quite substantial evidence, with some of them reaching  $BF$ s of  $-10$ , meaning that, regarding these partners, children had vastly different opinions. In terms of comparing genders, we observed that more boys than girls were below the threshold for "very strong evidence" (11 boys, 7 girls), as given by Jeffreys (1998), cited in Dittrich et al. (2019), for hypothesis  $H2$ .

In Figure 4, we plotted  $BF$ s for group 1 against partner indices. Here, we inspected a quite extreme excerpt from Figure 3, yielding a closer look at what the BSRM identified here.

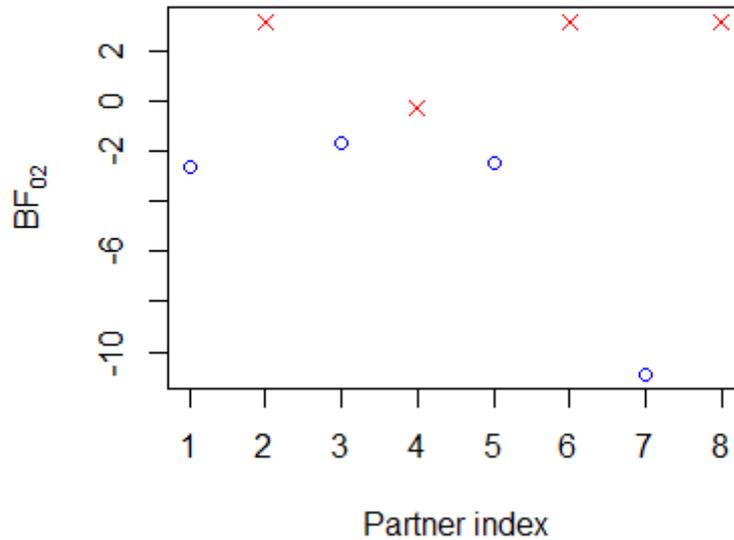


Figure 4. Bayes factors plotted against partner indices for group 1 of case study.

In Figure 4, child (partner) 7 is easily identifiable as an outlier. He had a  $BF$  of less than  $-10$ , indicating very decisive evidence for the hypothesis  $\tau_{b7} > 0$ , meaning that children differed in their relationships to him. To investigate how this was represented in the data matrix, we inspected it in Table 10.

To try to identify this pattern in the data matrix, one must first take into account the actor effects  $\alpha_i$  and partner effects  $\beta_j$  to isolate possible inter-individual relationship effects. So, for example, actor 4 rated the other children quite low overall. This would be an indicator for low  $\alpha_4$ . On the other hand, actor 3 rated quite high, indicating high  $\alpha_3$ . Similarly, partner 3 was rated relatively low overall, indicating low  $\beta_3$ , while partner 8 was rated relatively high, indicating high  $\beta_8$ . Now, the before identified outlier (partner 7) can be examined. First, child 7 was rated very similarly on different outcomes by all but one actors. Therefore, differences in scoring of child 7 are less likely due to random error (than for the other children) but rather due to inter-individual relationship effects, increasing the degree to which differences get represented through  $\tau_{b7}$ . Overall, there was also quite large variance in scoring of partner 7, which can be interpreted as supporting the idea of large  $\tau_{b7}$ . In general, this is very difficult to analyze by investigating the data matrix, but we tried to illustrate in this section, what the BSRM identified through the  $BF$  test.

Table 10. Data matrix for group 1 of the case study data set

Actor index	Outcome	Partner index							
		1	2	3	4	5	6	7	8
1	1	NA	49	70	53	42	48	30	66
	2	NA	31	31	30	30	30	31	31
2	1	83	NA	3	47	2	98	3	88
	2	66	NA	6	51	5	86	2	88
3	1	97	18	NA	18	46	21	80	45
	2	96	95	NA	94	94	93	79	96
4	1	5	78	4	NA	6	77	5	79
	2	4	4	3	NA	3	2	2	5
5	1	99	3	5	98	NA	3	95	3
	2	94	4	3	98	NA	3	99	6
6	1	19	88	5	64	6	NA	10	100
	2	54	83	2	72	1	NA	63	99
7	1	18	17	14	18	20	19	NA	18
	2	11	10	9	10	9	18	NA	16
8	1	22	45	42	61	5	89	16	NA
	2	50	83	91	64	43	93	23	NA

Moving on to the analysis of gender effects on  $BF$ s, we observed that more boys (31) than girls (17) showed significant evidence for  $H2$ . To test whether the two groups came from nonidentical populations, a one-sided Mann-Whitney-Wilcoxon test was run. At .05 significance level ( $p = .0128$ ), we concluded that the sets of  $BF$ s came from nonidentical populations, where the median for boys was lower. For children who had  $BF_{02j} > 0$ , we concluded that there were likely no inter-individual relationship effects. For those who had  $BF_{02j}$ , this judgement was not as simple. To find out to what degree the amount of evidence differed between the genders for these cases, another one-sided Mann-Whitney-Wilcoxon test was run, this time only on the children with  $BF_{02j} < 0$ . At .05 significance level, we did not find enough evidence to reject the null hypothesis (that both gender groups came from identical populations) with  $p = .4745$ . Therefore, we concluded that the difference in genders was only in the frequency with which inter-individual relationship effects could be detected and not in their magnitude.

## 7 Discussion

Results confirmed what we already assumed before analyzing the models. We established that the SRM provided better point estimates than the BSRM. There are a few points to underline why this should not be over-emphasized. First, the point estimation is not the primary objective of the BSRM, but rather a by-product of sampling the posterior distribution. Second, the CRs were overall quite good for the BSRM, especially for larger group sizes. We concluded that the posterior distribution fit the simulated parameter distribution quite well. Therefore, the information about  $\tau_{bj}$  was included in the posterior distribution, it simply was not extracted well. Because of the before discussed lower-bound issue with  $\tau_{bj}$  simulation under the SRM, the simulated parameter distribution was quite skewed. Therefore, the posterior mean is simply not a good estimator for a point estimate of  $\tau_{bj}$ , thus leading to the estimation problems discussed earlier. In future research, it will likely be possible to add a point estimator to the BSRM (e.g., a MAP estimation), although this is not the model's primary objective.

The BSRMs primary objective is hypothesis testing through *BFs*. The model provides a *BF* for each partner in the data set for multiple hypothesis tests. While the SRM assumes equal variance of inter-individual relationship effects across partners, the BSRM discriminates between partners. This leads to multiple novel possibilities for analyzing relationship structures. Under the SRM, it is impossible to meaningfully analyze data where inter-individual relationship effect variances differ across partners. We showed in the simulation study, that the BSRM, through its *BF* test, is capable of identifying such inter-individual relationship structures. Another benefit we identified is that the BSRM is able to test the hypothesis that there are no or very little inter-individual relationship effects. This sort of testing for existence is quite difficult using point estimates, which is where the BSRM adds new functionality.

In the simulation study, we found that the BSRM struggled to identify relatively small inter-individual relationship effects where we had only two outcome variables per dyad. In the case study, we then observed that the BSRM was able to identify inter-individual relationship effects where differences in scoring were very large. Since the case study included groups of eight with two outcomes per dyad and the BSRM worked well for that example, we can say that the BSRM is well suited to operating on small data sets, although the statistical power the *BF* test carries depends on the sample size, as well as,

of course, the population  $\tau_{bj}$  magnitude. Therefore, the model provides an alternative to traditional hypothesis testing that often requires quite large data distributions to function properly (e.g., ANOVA procedures).

In this paper, we focused solely on testing for the presence of  $\tau_{bj}$ , so  $H0 : \tau_{bj} = 0$  vs.  $H2 : \tau_{bj} > 0$ . In the paper, we presented one additional hypothesis already, that is  $H1 : \tau_{bj} < 0$ . Therefore, we established three hypotheses that can all be tested through  $BF$  tests. So, for example, one could test for  $H1$  vs.  $H2$  to investigate whether  $\tau_{bj}$  was more likely to be positive or negative. Therefore, in the future, there is reason to analyze this model with regard to its capability in testing different relationship structures, such as negative  $\tau_{bj}$ . This is only possible since, under the BSRM,  $\tau_{bj}$  is a covariance parameter for  $\gamma_{aij}$  (representing intra-individual relationship effects) and can therefore be negative.

Throughout this paper, we restricted our BSRM analyses to inter-individual relationship effects. As defined in Section 3, the model can also include intra-individual relationship effects. For example, one could analyze here, to what degree actors differed in their rating of partners, controlling for actor and partner effects. Since this was not yet implemented in the *BSRM* R package, this functionality could not be tested. Since this is a further novel and interesting contribution to the field, it definitely warrants future research.

In terms of applicability, the model is quite flexible. As mentioned before, the BSRM can be applied to small sample size data sets but is also not limited in terms of how large an input can be. Therefore, one could not only comfortably analyze large RR groups but also relationships between groups. Here, we would extend the RR design to one where groups of people say something about their own group, but also about other groups' members. This could potentially yield very interesting results when applied to, for example, different nationality groups. Then one could have sampled people from nation A say something about their peers but also about people from nation B. Then, by looking at  $BF$  results per partner, we could make inference about for example (1) how much variation there is in judgements of people from different nations and (2) whether variation is more likely due to individuals differing in their judgement of others or due to people differing in their judgement of an individual (inter- vs. intra-individual relationship effects). An interesting factor for inter-individual relationship effects might be cultural homogeneity here. So, for example, if the groups are themselves somewhat culturally homogeneous (and

we assume for now that cultures have somewhat stable opinions about other cultures), one might expect that actors disagree little about partners, leading to low  $\tau_{bj}$ . If, on the other hand, a group is rather heterogeneous, we might expect that its members have some disagreement about partners, leading to high  $\tau_{bj}$ . The BSRM would then be able to pick up these differences, possibly enabling the researcher to make inference about relationships between nations. Furthermore, the other way around, one might be able to make inferences about homogeneity of countries by looking at how similar  $\tau_{bj}$  are for each country. This idea would also be applicable to intra-individual relationship effects. The extent to which an actor varies in their relationship to others could then be measured using the BSRM. This is a very brief sketch of an idea for what the BSRM can technically be applied to and, of course, leaves many details as well as other possibilities on the table.

## References

- Albarracín, D., Sunderrajan, A., Lohmann, S., Chan, M.-P. S., and Jiang, D. (2018). The psychology of attitudes, motivation, and persuasion. In *The handbook of attitudes*, pages 3–44. Routledge.
- Becker, H. and Useem, R. H. (1942). Sociological analysis of the dyad. *American Sociological Review*, 7(1):13–26.
- Carlson, E., Burger, B., and Toiviainen, P. (2018). Dance like someone is watching: A social relations model study of music-induced movement. *Music & Science*, 1.
- Dittrich, D., Leenders, R. T. A., and Mulder, J. (2019). Network autocorrelation modeling: A bayes factor approach for testing (multiple) precise and interval hypotheses. *Sociological Methods & Research*, 48(3):642–676.
- Gleason, J. R. and Halperin, S. (1975). A paired compositions model for round-robin experiments. *Psychometrika*, 40(4):433–454.
- Goldring, M. (2020). The extended social relations model: Understanding dissimulation and dissensus in the judgment of others.
- Henderson, H. V., Pukelsheim, F., and Searle, S. R. (1983). On the history of the kronecker product. *Linear and Multilinear Algebra*, 14(2):113–120.
- Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.
- Kenny, D. and La Voie, L. (1984). The social relations model. In *Advances in experimental social psychology*, volume 18, pages 141–182. Elsevier.
- Kluger, A. N., Malloy, T. E., Pery, S., Itzhakov, G., Castro, D. R., Lipetz, L., Sela, Y., Turjeman-Levi, Y., Lehmann, M., New, M., et al. (2020). Dyadic listening in teams: social relations model. *Applied Psychology*.
- Liao, W., Bazarova, N. N., and Yuan, Y. C. (2018). Unpacking medium effects on social psychological processes in computer-mediated communication using the social relations model. *Journal of Computer-Mediated Communication*, 23(2):90–106.

- Nestler, S., Lüdtke, O., and Robitzsch, A. (2020). Maximum likelihood estimation of a social relations structural equation model. *Psychometrika*, pages 1–20.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Schönbrodt, F. D., Back, M. D., and Schmukle, S. C. (2012). Tripler: An r package for social relations analyses based on round-robin designs. *Behavior research methods*, 44(2):455–470.
- van Noorden, T. H., Bukowski, W. M., Haselager, G. J., Lansu, T. A., and Cillessen, A. H. (2016). Disentangling the frequency and severity of bullying and victimization in the association with empathy. *Social Development*, 25(1):176–192.
- Warner, R. M., Kenny, D. A., and Stoto, M. (1979). A new round robin analysis of variance for social interaction data. *Journal of Personality and Social Psychology*, 37(10):1742.
- Zehfuss, G. (1858). Über eine gewisse determinante. *Zeitschrift für Mathematik und Physik*, 3(1858):298–301.

## Appendix

### A Simulation algorithm pseudocode

```

INPUT true simulation parameters
STORE true parameters in true_p
FOR replication 1 to max_replications by 1
  GENERATE simulated data set X
  INPUT X in TripleR yielding output_SRM
  INPUT X in BSRM yielding output_BSRM
  RETRIEVE seed from X
  STORE seed in seed_vector[replication]
  STORE relevant posterior means in output_matrix_BSRM[replication,]
  STORE relevant point estimates in output_matrix_SRM[replication,]
  CALCULATE BSRM_Bias[replication] as output_matrix_BSRM[replication,x] - true_p[x]
  CALCULATE SRM_Bias[replication] as output_matrix_SRM[replication,x] - true_p[x]
  CALCULATE HDI for BSRM tau posterior
  CALCULATE CI for TripleR tau estimate
  FOR EACH partner per group
    IF real_tau is in HDI
      STORE 1 in coverage_vector_BSRM[replication]
    ELSE
      STORE 0 in coverage_vector_BSRM[replication]
    IF real_tau is in CI
      STORE 1 in coverage_vector_SRM[replication]
    ELSE
      STORE 0 in coverage_vector_SRM[replication]
  END FOR
OUTPUT coverage_vectors, output_matrices, seed_vector

```

Here, "GENERATE" refers to using the function from Appendix C, while "RETRIEVE" refers to selecting an element from a function's output. The above pseudocode illustrates the simulation study process, which can be found in detailed R code in Appendix B.

### B Simulation study R code

This script analyzes *times* data sets with specified parameters, using the `simRRdata` function from Appendix C to simulate individual data sets.

```

BSRM_SRM_multi <- function(groups = 1, partners, Tau, increasingTau, outcomes,times
= 500, simRRData){

```

```

BSRM <- function(X){
  out <- BSRM::MCMC_BSRM(X$stackMBDYData$data,ncol(X$stackMBDYData$data),XG = 5000,X$stack
= X$stackMBDYData$MBDY,Icovtb = TRUE)
  return(out)
}

actorBias <- matrix(nrow = times,ncol = partners)
partnerBias <- matrix(nrow = times, ncol = partners)
actorBias_RR <- matrix(nrow = times,ncol = partners)
partnerBias_RR <- matrix(nrow = times, ncol = partners)
actorpartnercovBias_RR <- vector(length = times)
seed <- vector(length = times)
varactorBias <- vector(length = times)
varactorBias_RR <- vector(length = times)
varpartnerBias <- vector(length = times)
varpartnerBias_RR <- vector(length = times)
Msigma <- matrix(nrow = times, ncol = 3)
tau <- matrix(nrow = times, ncol = 3*partners)
tauBias <- matrix(nrow = times, ncol = partners)
tauBias_RR <- matrix(nrow = times, ncol = partners)
Msigma2 <- matrix(nrow = times, ncol = 2*partners)
sigma2Bias <- matrix(nrow = times, ncol = partners)
BF <- array(dim = c(times,partners,3))
tauCoverage <- matrix(0,nrow = times,ncol = partners)
tauCoverage_RR <- matrix(0,nrow = times,ncol = partners)
for(i in 1:times){
  print(i)
  dat <- simRRData(groups,partners,outcomes,tau = Tau, increasing = increasingTau)
  seed[i] <- dat$analysisInformation$seed
  out <- BSRM(dat)
  outR <- TripleR::RR(A+B~act*part|grp,dat$DataFrameNoSelf)
  #coverage rate:
  #BSRM
  for(j in 1:partners){
    low <- HDInterval::hdi(out$Mtau[1000:5000,j])[1]
    hi <- HDInterval::hdi(out$Mtau[1000:5000,j])[2]
    if(dplyr::between(dat$tau[j],low,hi)){
      tauCoverage[i,j] <- 1
    }
  }
}
#TripleR

```

```

for(j in 1:partners){
  mn <- outR$bivariate$estimate[5]
  se <- outR$bivariate$se[5]
  sd <- sqrt(partners) * se
  low <- mn - 1.96*sd
  hi <- mn + 1.96*sd
  if(dplyr::between(dat$tau[j],low,hi)){
    tauCoverage_RR[i,j] <- 1
  }
}
tauBias_RR[i,] <- rep(outR$bivariate$estimate[5],partners) - dat$tau
#actor/partner bias:
BSRMactor <- vector(length = partners)
for(j in 1:partners){
  BSRMactor[j] <- mean(out$EAPtheta[j:((j-1) + outcomes)])
}
#actor-partner covariances
actorpartnercovBias_RR[i] <- outR$bivariate$estimate[3] - 0.075
Msigma[i,] <- apply(out$Msigma[,],2,mean)
#0.25 = actor variance and partner variance
varactorBias[i] <- var(BSRMactor) - 0.25
varactorBias_RR[i] <- outR$bivariate$estimate[1] - 0.25
varpartnerBias[i] <- var(out$EAPBetaR) - 0.25
varpartnerBias_RR[i] <- outR$bivariate$estimate[2] - 0.25
tau[i,] <- apply(out$Mtau[1000:5000,],2,mean)
tauBias[i,] <- tau[i,1:partners] - dat$tau#this is the population tau
Msigma2[i,] <- apply(out$Msigma2[1000:5000,],2,mean)
sigma2Bias[i,] <- Msigma2[i,1:partners] - 0.1#population error variance
BF[i,,] <- apply(out$MBF0[1000:5000,],2,mean)
}
return(list(seed = seed, BSRM = list(Msigma = Msigma, Mtau = tau, MBF0 = BF, Msigma2
= Msigma2, tauBias = tauBias, sigmaBias = sigma2Bias, tauCoverage = tauCoverage,
actorVarianceBias = varactorBias, partnerVarianceBias = varpartnerBias), TripleR
= list(actorpartnercovBias = actorpartnercovBias_RR,tauCoverage = tauCoverage_RR,
tauBias = tauBias_RR, actorVarianceBias = varactorBias_RR, partnerVarianceBias =
varpartnerBias_RR)))
}

```

## C Data simulation R code

This function simulates round robin-type data with specified parameters *group number*, *group size*, *outcome count* and *all aforementioned simulation parameters*. It outputs a data frame which can be fed to the *TripleR* function `RR()`. Also, it outputs a data matrix for *BSRM* as well as all necessary design and factor elements.

```
simRRData <- function(groups, groupSize, items = 2, varianceVector = c(0.1,0.25,0.25,0.4), A
= 0.075, increasing = FALSE, uniqueIndices = TRUE, seed=sample(1:10000,1), binary
= FALSE, tau = 0, errorVariance = 0.1){
  #save seed
  set.seed(seed)
  J <- groups
  m <- groupSize
  N <- J*m
  K <- items
  v <- varianceVector
  XF <- matrix(0,ncol = J, nrow = N*K)
  apcov <- ActorPartnerCovariance
  outTau <- vector(length = m)
  outSigma2 <- vector(length = m)
  #group effects
  g <- rnorm(J,mean = 0, sd = sqrt(v[1]))
  #actor & partner effects:
  sigma <- diag(c(v[2]-apcov,v[3]-apcov)) + apcov
  actorpartner <- mvtnorm::rmvnorm(N,mean=rep(0,2),sigma=sigma)
  #extract actor and partner effects
  a <- actorpartner[,1]
  b <- actorpartner[,2]

  #####interaction effects#####
  relationshipEffects <- function(N, K, inc = FALSE){
    R <- array(dim = c(N,N,K))
    sigma2 <- rep(errorVariance,N)
    if(inc == TRUE){
      taubj <- seq(from = 0, to = tau, length.out = N)
    }
    else{
      taubj <- rep(tau,N)
    }
    sim <- matrix(nrow = N,ncol = K)
```

```

for(i in 1:N){
  sigma <- diag(rep(sigma2,K),K) + taubj[i]
  sim <- mvtnorm::rmvnorm(N,sigma = sigma)
  R[,i,] <- sim
  outTau[i] <- cov(R[,i,1],R[,i,2])
  outSigma2[i] <- mean(c(var(R[,i,1]) - outTau[i]),var(R[,i,2] - outTau[i]))
}
return(list(R = R, tau = outTau, sigma = outSigma2, populationTau = taubj))
}
E_temp <- relationshipEffects(N,K,increasing)
E <- E_temp$R
outTau <- E_temp$tau
outSigma2 <- E_temp$sigma

#structure vectors
grp <- vector(length = N*m)
act <- vector(length = N*m)
part <- vector(length = N*m)
#calculate observed values
Y <- vector(length = N*m*K)
for(G in 1:J){#groups
  for(A in 1:m){#actors
    for(P in 1:m){#partners
      for(kk in 1:K){#items
        Y[(G-1)*m*m*K + (A-1)*m*K + (P-1)*K + kk] <- g[G] + a[(G-1)*m + A] + b[(G-1)*m
+ P] + E[(G-1)*m + A,P,kk]# + Ieff[kk] + e[(G-1)*m*m + (A-1)*m + (P-1)*K + kk]
        grp[(G-1)*m*m + (A-1)*m + P] <- G
        if(uniqueIndices == FALSE){
          part[(G-1)*m*m + (A-1)*m + P] <- P
          act[(G-1)*m*m + (A-1)*m + P] <- A
        }
      }
    }
  }
}
}
}
}

```

```

}
if(binary == TRUE){
  Y[Y < 0] = 0
  Y[Y > 0] = 1
}
#structure data matrix with different columns for different items
if(K == 1){
  YI <- as.matrix(Y)
}
else if(K==2){
  YI <- matrix(ncol = K,nrow = N*m)
  YI[,1] <- Y[row(as.matrix(Y))%%2 != 0]
  YI[,2] <- Y[row(as.matrix(Y))%%2 == 0]
}
else{
  YI <- matrix(ncol = K,nrow = N*m)
  for(i in 1:(N*m)){
    for(j in 1:K){
      YI[i,j] <- Y[(i-1)*K + j]
    }
  }
}
#complete data frame for TripleR
cmp <- as.data.frame(cbind(grp,act,part))
for(i in 1:K){
  cmp <- as.data.frame(cbind(cmp,YI[,i]))
}
colnames(cmp) <- c("grp","act","part",LETTERS[1:K])

#data frame without self-judgements
Z <- cmp[!act==part,]

#####MBDY data#####
for(i in 1:J){
  for(j in 1:(m*K)){
    XF[(i-1)*(m*K) + j, i] <- 1
  }
}
#mbdy design matrix

```

```

MBDY <- matrix(data = 1, ncol = N, nrow = N*K)

for(i in 1:N){
  for(j in 1:K){
    MBDY[(i-1)*K + j, i] <- 0
  }
}
#mbdy data matrix
YY <- matrix(data = NA, nrow = N*K, ncol = N)
for(i in 1:J){
  for(j in 1:m){
    for(k in 1:m){
      for(kk in 1:K){
        YY[(i-1)*m*K + (j-1)*K + kk, (i-1)*m + k] <- Y[(i-1)*m*m*K + (k-1)*K + (j-1)*m*K
+ kk]
      }
    }
  }
}
MBDY[is.na(YY)] <- 0
YY[MBDY == 0] <- NA
#nll vector
nll <- vector(length = N*K)
for(i in 1:J){
  for(k in 1:m){
    for(j in 1:K){
      nll[(i-1)*m*K + (k-1)*K + j] <- (i-1)*m + k
    }
  }
}
#####stack MBDY data#####
sMBDY <- matrix(data = 1, ncol = m, nrow = N*K)

for(i in 1:J){
  for(j in 1:m){
    for(k in 1:K){
      sMBDY[(i-1)*K*m + (j-1)*K + k, j] <- 0
    }
  }
}

```

```

#mbdy data matrix
sYY <- matrix(data = NA, nrow = N*K, ncol = m)
for(i in 1:J){
  for(j in 1:m){
    for(k in 1:m){
      for(kk in 1:K){
        sYY[(i-1)*m*K + (j-1)*K + kk,k] <- Y[(i-1)*m*m*K + (k-1)*K + (j-1)*m*K
+ kk]
      }
    }
  }
}
sMBDY[is.na(sYY)] <- 0
sYY[sMBDY == 0] <- NA
#nll vector
snll <- vector(length = N*K)
for(i in 1:J){
  for(k in 1:m){
    for(j in 1:K){
      snll[(i-1)*m*K + (k-1)*K + j] <- (i-1)*m + k
    }
  }
}

return(list(tau = E_temp$populationTau, Y = Y, MBDYData = list(MBDY = MBDY, data
= YY, nll = nll,XF=XF), stackMBDYData = list(MBDY = sMBDY, data = sYY, nll = snll,XF=XF),com
= cmp, DataFrameNoSelf = Z, effects = list(group = g, actor = a, partner= b, relationship
= E), variances = list(group = var(g) ,actor = var(a), partner = var(b), tau = outTau,
sigma2 = outSigma2, ActorPartnerCov = cov(a,b)),structure = list(group = grp, actor
= act, partner = part),analysisInformation = list(groups = groups, groupSize = groupSize,
variances = varianceVector, ActorPartnerCovariance = ActorPartnerCovariance, errorVariance=e
increasing = increasing, seed = seed)))
}

```