

## Dealing with Multiple Solutions in Structural Vector Autoregressive Models

Adriene M. Beltz & Peter C. M. Molenaar

**To cite this article:** Adriene M. Beltz & Peter C. M. Molenaar (2016) Dealing with Multiple Solutions in Structural Vector Autoregressive Models, *Multivariate Behavioral Research*, 51:2-3, 357-373, DOI: [10.1080/00273171.2016.1151333](https://doi.org/10.1080/00273171.2016.1151333)

**To link to this article:** <https://doi.org/10.1080/00273171.2016.1151333>



Published online: 19 Apr 2016.



Submit your article to this journal [↗](#)



Article views: 810



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 14 View citing articles [↗](#)

## Dealing with Multiple Solutions in Structural Vector Autoregressive Models

Adriene M. Beltz and Peter C. M. Molenaar

Department of Human Development and Family Studies, The Pennsylvania State University

### ABSTRACT

Structural vector autoregressive models (VARs) hold great potential for psychological science, particularly for time series data analysis. They capture the magnitude, direction of influence, and temporal (lagged and contemporaneous) nature of relations among variables. Unified structural equation modeling (uSEM) is an optimal structural VAR instantiation, according to large-scale simulation studies, and it is implemented within an SEM framework. However, little is known about the uniqueness of uSEM results. Thus, the goal of this study was to investigate whether multiple solutions result from uSEM analysis and, if so, to demonstrate ways to select an optimal solution. This was accomplished with two simulated data sets, an empirical data set concerning children's dyadic play, and modifications to the group iterative multiple model estimation (GIMME) program, which implements uSEMs with group- and individual-level relations in a data-driven manner. Results revealed multiple solutions when there were large contemporaneous relations among variables. Results also verified several ways to select the correct solution when the complete solution set was generated, such as the use of cross-validation, maximum standardized residuals, and information criteria. This work has immediate and direct implications for the analysis of time series data and for the inferences drawn from those data concerning human behavior.

### KEYWORDS

Equivalent solutions; group iterative multiple model estimation; plausible alternatives; structural equation models; vector autoregression

Time series data have significant potential for providing new insight into human behavior. Although these data have traditionally been underutilized in psychology, the tide is turning, perhaps due to researchers' decreased willingness to accept limitations associated with group-level analysis approaches (Molenaar, 2004) or due to the field's increased appreciation for time-indexed functional neuroimaging and telemetric (e.g., wearable computers and mobile phones) data (Goodwin, Velicer, & Intille, 2008; Poldrack & Wagner, 2004). Thus, there is great interest in the development and refinement of analysis procedures for time series data.

Structural vector autoregression (VAR) is a valuable approach to the analysis of time series data. Structural VARs have been developed and applied extensively in econometrics (Lütkepohl, 2005; Martin, Hurn, & Harris, 2013), and their value for answering questions related to psychological, including brain, science is beginning to be realized (Beltz, Beekman, Molenaar, & Buss, 2013; Gates & Molenaar, 2012). Structural VARs are based on a transformation of standard VARs. Consider a weakly stationary  $p$ -variate time series  $y(t)$ . Weak stationarity implies that  $y(t)$  has constant mean level (conveniently fixed at zero in what follows) and covariance function  $\text{cov}[y(t), y(t-u)]$ , which only depends on the lag  $u$ . A standard

VAR( $a$ ) for  $y(t)$  is specified as

$$y(t) = \Xi_1 y(t-1) + \Xi_2 y(t-2) + \dots + \Xi_a y(t-a) + \varepsilon(t), \quad (1)$$

where  $y(t)$  is the  $p$ -variate weakly stationary time series to be explained at time  $t = 1, 2, \dots, T$ , with  $T$  the length of the time series;  $\Xi_k$  is the  $(p, p)$ -dimensional matrices of regression coefficients at lag  $k = 1, 2, \dots, a$ ; and  $\varepsilon$  is the  $p$ -variate process innovations, lacking sequential dependencies and having a zero mean and full covariance matrix. In standard VARs, the full covariance matrix of  $\varepsilon$  is the source of contemporaneous relations among the component series of  $y(t)$ .

A structural VAR( $a$ ) is specified as

$$y(t) = \Gamma y(t) + \Lambda_1 y(t-1) + \Lambda_2 y(t-2) + \dots + \Lambda_a y(t-a) + v(t), \quad (2)$$

where  $y(t)$  is defined as in Equation (1);  $\Gamma$  is the  $(p, p)$ -dimensional matrix of contemporaneous regression coefficients;  $\Lambda_k$  is the  $(p, p)$ -dimensional matrices of regression coefficients at lag  $k = 1, 2, \dots, a$ ; and  $v$  is the  $p$ -variate process innovations, lacking sequential dependencies and having a zero mean and a diagonal covariance matrix. In the structural VAR in Equation (2),  $\Gamma$  is

the cause of the contemporaneous relations among the component series of  $y(t)$ . The structural VAR typically is obtained by transforming a standard VAR by means of Cholesky decomposition of the full covariance matrix of the process innovations  $\varepsilon(t)$  in Equation (1) (Lütkepohl, 2005). Structural VARs estimated in this way (by means of Cholesky decomposition) are not unique but depend upon the ordering of the univariate component series in  $y(t)$  (Loehlin, 1996; Lütkepohl, 2005; Molenaar & Lo, 2016).

The nonuniqueness of structural VARs obtained by Cholesky decomposition can, however, be avoided by using the unified structural equation model (uSEM; Gates, Molenaar, Hillary, Ram, & Rovine, 2010; Kim, Zhu, Chang, Bentler, & Ernst, 2007). The model is optimal, due in large part to the way in which it is usually fit to the data: Instead of first fitting a standard VAR to the data then carrying out the Cholesky decomposition of the covariance matrix of the process innovations and finally transforming the standard VAR into the structural VAR, Equation (2) is fitted directly to the data. In what follows, the alternative approach is called fitting uSEMs. The fit of a uSEM starts with fitting a null model (i.e., an instance of Equation [2] in which all coefficient matrices are zero). Within an SEM framework, Lagrange multiplier tests (i.e., modification indices that mark how much model fit as assessed by the likelihood ratio will improve if a particular fixed parameter is freed; Sörbom, 1989) are then used to sequentially free parameters in the null model until it provides an excellent fit to the observed data; nonsignificant parameters as assessed by Wald tests are then trimmed from the model. Although the data-driven modification index-based building of covariance structure models has been criticized (Kaplan, 1989; MacCallum, Roznowski, & Necowitz, 1992), hierarchical Lagrange multiplier tests combined with model trimming (as implemented in uSEM) have been shown to perform as well as the likelihood ratio test (Chou & Bentler, 1990). uSEM was developed for use with functional neuroimaging data to map relations among the activity of brain regions of interest during a task or resting state (see also Gates, Molenaar, Hillary, & Slobounov, 2011). In this domain, simulation studies have shown that uSEM accurately recovers lagged and contemporaneous connections between brain regions and accurately reflects the effects of external input (i.e., experimental conditions) on the connections (Gates et al., 2010; Gates et al., 2011). In addition, uSEM has been successfully applied to behavior—for example, revealing sex differences in the temporal nature of vigor of activity and positive affect during children's unstructured group play (Beltz, Beekman, et al., 2013).

A key advantage of uSEM is that it can be employed across possibly heterogeneous individuals using group

iterative multiple model estimation (GIMME; Gates & Molenaar, 2012). GIMME implements uSEMs in two steps, first fitting lagged and contemporaneous variable relations that apply to the full sample and then fitting additional variable relations that only apply to individuals. In this way, GIMME facilitates group-level inferences and maintains the integrity of the person-specific time series. Simulation studies in the neuroimaging domain show that GIMME outperforms all other modeling techniques, including standard VARs and SEMs and dynamic causal models (Gates & Molenaar, 2012; Smith et al., 2011). GIMME shares some features with and performs similarly when analyzing homogeneous data to an approach employed by Ramsey, Hanson, and Glymour (2011); both use multisubject data and automated model fitting to create group-level solutions with causal (i.e., directed) relations. However, GIMME is unique in its ability to handle heterogeneous data by adding individual-level relations to the group-level solutions and to model experimental conditions. GIMME is, therefore, the optimal data analysis approach for identifying contemporaneous and lagged directed relations among variables in a heterogeneous multisubject data set.

The potential for multiple solutions to result from uSEM analysis has not been investigated, but there is reason to suspect that competing solutions exist. Multiple solutions occur when different sets of parameters provide similar fit to the data: Equivalent solutions have identical implied covariance structure and degrees of freedom, whereas plausible alternative solutions have similar (but not identical) fit and provide substantively different interpretations; the substantive differences are large compared to the statistical differences in fit (cf. Markus, 2002). Multiple solutions are obviously problematic because a given result may provide a different interpretation of the data than alternative results that were not considered but that had similar empirical support (see, e.g., Henley, Shook, & Peterson, 2006). Multiple solutions have been most extensively considered in the SEM literature as each SEM potentially has an infinite number of them (MacCallum, Wegener, Uchino, & Fabrigar, 1993; Raykov & Marcoulides, 2001). It is, therefore, likely that uSEMs also have multiple solutions because they are a type of structural VAR implemented within an SEM framework. As SEMs only consider contemporaneous relations, multiple solutions in uSEMs will almost always stem from the contemporaneous relations, particularly during the initial stages of the determination of the direction of influence between two variables. This is illustrated by the simple bivariate case, wherein the directed  $\beta$  from a regression of  $x$  on  $y$  equals the directed  $\beta$  from a regression of  $y$  on  $x$ . All statistical indices are the same for the two analyses, but the interpretations greatly differ. Which result is correct?

When time series data analysis approaches produce multiple solutions, there are some practical and precise ways to select an optimal result. First, eliminate solutions that are conceptually improbable (MacCallum et al., 1993; Markus, 2002). Second, eliminate solutions that contain statistically unlikely or nonsignificant relations—for example, through the use of penalized regression (Valdés-Sosa et al., 2005). Third, gather more information; that is, collect data from additional individuals, variables and task manipulations, or measurement occasions (Nesselroade & Ram, 2004) in order to eliminate solutions or replicate previous findings (MacCallum et al., 1993). Fourth, use fit indices to determine whether the estimated solutions provide excellent fit to the observed data; combinations of indices that have been found to be reliable in past work are particularly meaningful, such as the root mean squared error of approximation (RMSEA), nonnormed fit index (NNFI), comparative fit index (CFI), and standardized root mean square residual (SRMR) (Brown, 2006; Gates & Molenaar, 2012; Hu & Bentler, 1998). Fifth, evaluate the standardized residuals for the estimated relations, for example, using the largest residual as a marker of discrepancy between the estimated solution and observed data (MacCallum & Austin, 2000; e.g., Zhuang, LaConte, Peltier, Zhang, & Hu, 2005). Sixth, use information criteria, such as the Akaike information criterion (AIC; Akaike, 1974) to compare solutions. This is especially relevant when models are not nested and differ in degrees of freedom (see de Marco et al., 2009). A key point is that selection of an optimal solution often requires all possible solutions to be generated; the approaches reviewed thus far have the most utility if they are applied across the complete set of solutions.

### Current study

The aim of this work was to determine whether multiple solutions result from uSEMs implemented with GIMME and to showcase ways to select an optimal solution when several alternatives are generated. Multiple solutions were expected to result when fitting uSEMs if there were iterations in which maximum modification indices were equal. (This is qualitatively different from the multiple solutions that result when estimating structural VARs by fitting standard VARs, then transforming them through Cholesky decomposition of the covariance matrix of process innovations because the ordering of the univariate raw data series does not influence uSEM results.) Aims were accomplished by modifying the GIMME program (<http://www.nitrc.org/projects/gimme/>) to output the set of all possible solutions, resulting in the new program detailed in the following section: GIMME for multiple

solutions (GIMME-MS).<sup>1</sup> GIMME-MS was then applied to three data sets.

The first two data sets were simulated according to a uSEM. The first set contained larger (in magnitude) lagged than contemporaneous relations, and the second set contained smaller lagged than contemporaneous relations. Multiple solutions were not expected for the first simulation since modification indices should indicate to free (relatively large) lagged parameters before (relatively small) contemporaneous parameters. Multiple solutions were expected in the second simulation, however, since modification indices should indicate to free (relatively large) contemporaneous parameters before (relatively small) lagged parameters, and at the initial phases of the sequential application of Lagrange multiplier tests the values of contemporaneous directed relations between each pair of univariate component series are equal. For example, if no other freed relations concern variables  $m$  and  $n$ ,  $\gamma_{m,n}$  fits the observed data as well as  $\gamma_{n,m}$ . Without additional statistical context (e.g., a freed relation that concerns variable  $m$  or  $n$ ) or conceptual guidance, the only way to generate an accurate solution is to free each contemporaneous parameter (e.g.,  $\gamma_{m,n}$  and  $\gamma_{n,m}$ ) in its own solution path so that separate models can be built and evaluated, as was done here. When multiple solutions were found, an optimal result was selected from the set of equivalent and plausible alternative solutions by gathering additional information (e.g., cross-validation in newly simulated data sets), evaluating combined fit indices, examining maximum residuals, and considering an information criterion.

The third data set was empirical, containing vigor of activity and positive affect ratings for dyads of children aged 4–7 years engaged in a play session designed to elicit rough-and-tumble interactions. Data from eight dyads were selected to demonstrate that multiple solutions can occur in GIMME analysis of empirical data. This particular data set was optimal for such a demonstration because past work with uSEM has shown that contemporaneous relations are important for characterizing the vigor of activity exhibited by girls and boys during free play (Beltz, Beekman, et al., 2013), and we expected multiple solutions to result from data sets with large contemporaneous relations. When multiple solutions were found, an optimal result was selected by implementing model selection metrics that have been validated in previous research—namely, by evaluating combined fit indices, examining maximum residuals, and considering an information criterion.

<sup>1</sup> GIMME-MS is a Matlab-based program that interfaces with LISREL. A copy of the program is available from the authors, and a final version will be made publicly available at [http://quantdev.ssri.psu.edu/avada\\_resources/gimme/](http://quantdev.ssri.psu.edu/avada_resources/gimme/).



## Method

### GIMME

#### Standard GIMME

GIMME (Gates & Molenaar, 2012) is an automated Matlab program that calls LISREL (Jöreskog & Sörbom, 1992) to implement uSEMs on person-mean-centered data at the group and individual levels using an estimation method best classified as pseudomaximum likelihood because it concerns the analysis of a block-Toeplitz covariance matrix containing dependent observations (cf. Hamaker, Dolan, & Molenaar, 2005; Molenaar & Nesselroade, 1998); this estimation approach yields results comparable to raw data likelihood and Bayesian methods (Zhang, Hamaker, & Nesselroade, 2008). GIMME for a lag of one is specified as

$$\eta_i(t) = (A_i + A^g)\eta_i(t) + (\Phi_{1,i} + \Phi_1^g)\eta_i(t-1) + \zeta_i(t), \quad (3)$$

where  $\eta_i(t)$  is the  $p$ -variate weakly stationary time series to be explained at time  $t = 1, 2, \dots, T$ , with  $T$  the time series length;  $A$  is the  $(p,p)$ -dimensional matrix of contemporaneous regression coefficients;  $\Phi_1$  is the  $(p,p)$ -dimensional matrix of regression coefficients at lag 1;  $\zeta_i$  is the  $p$ -variate process innovations, lacking sequential dependences and having zero mean and a diagonal covariance matrix; superscript  $g$  indicates group-level relations; and subscript  $i$  indicates individual-level relations. Thus, there are no bidirectional relations in GIMME models as the goal of the program is to define behavioral networks by the endogenous interplay among variables in the network and not by exogenous influences on the network (e.g., as would be reflected by correlated process innovation terms; see Molenaar & Lo, 2016).

GIMME performs several steps in order to estimate parameters for  $A$  and  $\Phi_1$ , that is, to identify directed contemporaneous and lagged relations among variables. First, null models are estimated for all individuals, fit to  $(2p, 2p)$ -dimensional block Toeplitz covariance matrices generated from the behavioral time series (Box, Jenkins, & Reinsel, 2008; Molenaar, 1985), with only the diagonal covariance matrix of the process innovation being freely estimated. Second, a group-level solution is generated. The parameter in  $A^g$  or  $\Phi_1^g$  most improving model fit across individuals is determined with Lagrange multiplier tests (modification indices; Sörbom, 1989), then the model is reestimated with this relation freed. This process is repeated until no relation significantly improves model fit for the group. Third, the group-level solution

is trimmed, ensuring that all freed relations are significant for a critical mass of the individuals, as determined by a researcher-specified criterion. The criterion is the percent of individuals for whom a relation must be significant in order for the relation to be included in the group model for all individuals. Fourth, individual-level solutions are generated. After estimating the group-level solution, parameters in  $A_i$  and  $\Phi_{1,i}$  for each individual are iteratively freed using Lagrange multiplier testing until no relation, if freed, significantly improves model fit. Fifth, individual-level solutions are trimmed, ensuring that all variable relations are either significant for the individual or satisfy the group criterion. Sixth, a confirmatory model is run. Solutions are accepted if two of four indices indicate excellent fit (according to Brown, 2006):  $RMSEA \leq .05$ ;  $NNFI \geq .95$ ;  $CFI \geq .95$ ;  $SRMR \leq .05$ .

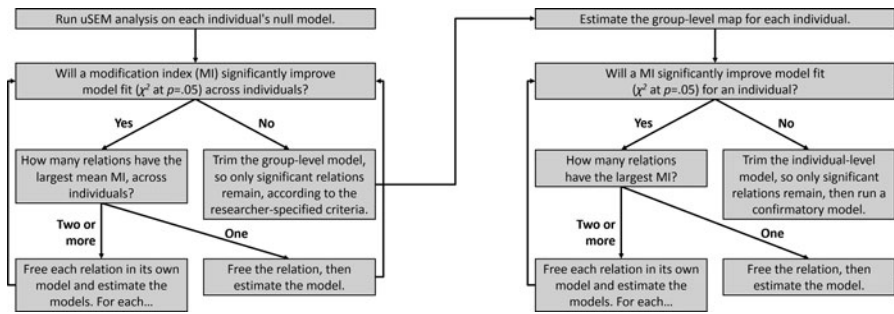
#### GIMME for multiple solutions (GIMME-MS)

GIMME was modified in order to generate multiple solutions (when they are present). A schematic representation of the GIMME-MS program is shown in Figure 1. The GIMME equation in (3) also underlies GIMME-MS, but the procedure for building model solutions has been modified: For iterations in which the maximum modification index is not singular (i.e., it is not singular when two or more parameters have the same, maximal Lagrange multiplier test result), each parameter containing the maximum modification index is freed in its own solution “path.” After freeing parameters in separate solution paths, each path independently iterates, with the potential for additional solution paths to emerge as model fitting progresses (if maximum modification indices of future iterations are not singular). Thus, new paths can be generated at any iteration and at the group or individual level.

Group-level equality of Lagrange multiplier testing is defined as the average maximum modification index across all individuals being equal for two or more parameters:

$$\begin{aligned} \frac{\sum_{i=1}^N MI_{j,i}}{N} &= \frac{\sum_{i=1}^N MI_{k,i}}{N} = \frac{\sum_{i=1}^N MI_{l,i}}{N} \dots > \frac{\sum_{i=1}^N MI_{d,i}}{N} \\ &\geq \frac{\sum_{i=1}^N MI_{e,i}}{N} \geq \frac{\sum_{i=1}^N MI_{f,i}}{N} \dots, \end{aligned} \quad (4)$$

where subscript  $i$  indicates one of  $N$  individuals (with the summation across  $N$  individuals indicating that the modification indices apply at the group level);  $MI$  indicates the modification index; subscripts  $j, k, l$ , and so on, indicate fixed parameters in  $A$  or  $\Phi_1$  (that will be freed in separate paths in the next iteration); and  $d, e, f$ , and so on, indicate fixed parameters in  $A$  or  $\Phi_1$  (that will not be freed in the next iteration). Individual-level equality of Lagrange



**Figure 1.** Schematic representation of the model generation and fitting procedures for group iterative multiple model estimation for multiple solutions (GIMME-MS), an extension of GIMME (Gates & Molenaar, 2012). uSEM = unified structural equation model.

multiplier testing is defined as the maximum modification index being equal for two or more parameters:

$$MI_{j,i} = MI_{k,i} = MI_{l,i} \dots > MI_{d,i} \geq MI_{e,i} \geq MI_{f,i} \dots, \quad (5)$$

where  $MI$ ,  $i$ ,  $j$ ,  $k$ ,  $l$ , and  $d$ ,  $e$ ,  $f$ , are defined as in Equation (4).

Because of the inevitable finite precision of computing with real-valued variables, the GIMME-MS program allows users to define  $MI$  equality. In analyses reported here, (average) modification indices were considered non-singular if they were equal to three decimal places (based on the assumption that differences beyond three decimal places are not substantively meaningful). Prior to GIMME-MS, GIMME selected parameters to be freed based on infinitesimal differences in modification indices, potentially producing unique model fits that were more dependent on machine precision than on psychological relevance.

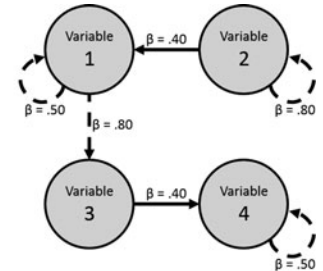
### Simulated data

Two initial structural VAR data sets were simulated according to uSEM. Each simulation contained data for 75 individuals on four variables, and each variable contained 300 measurements (i.e., time series were of length 300). An additional data set used in cross-validation was also simulated, using the same simulation parameters as the initial data set, when an initial data set produced multiple solutions.

Figure 2 shows the parameters used in the simulations in matrix and diagram form, with the zero mean process innovation having a covariance matrix equal to the identity matrix. The same standardized parameters were used for each individual and, therefore, apply at the group level. The first simulation (Figure 2A) contained relatively large lagged parameters in  $\Phi_1$  and relatively small contemporaneous parameters in  $A$ . Appendix A shows an alternate representation of this figure, a representation that can be generalized to all other network figures in the present

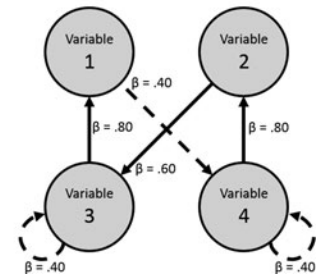
#### A. Simulation 1

$$\Phi_1 = \begin{bmatrix} .5 & 0 & 0 & 0 \\ 0 & .8 & 0 & 0 \\ .8 & 0 & 0 & 0 \\ 0 & 0 & 0 & .5 \end{bmatrix} \quad A = \begin{bmatrix} 0 & .4 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & .4 & 0 \end{bmatrix}$$



#### B. Simulation 2

$$\Phi_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & .4 & 0 & 0 \\ .4 & 0 & 0 & .4 \end{bmatrix} \quad A = \begin{bmatrix} 0 & 0 & .8 & 0 \\ 0 & 0 & 0 & .8 \\ 0 & .6 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$



**Figure 2.** Parameters for two data sets simulated according to a unified structural equation model (uSEM), shown in lagged ( $\Phi_1$ ) and contemporaneous ( $A$ ) matrices and as network diagrams. The same standardized parameters were implemented for all individuals and, thus, apply at the group level. Dashed lines reflect lagged variable relations; solid lines reflect contemporaneous variable relations;  $\beta$ -weights reflect the magnitude of the relations. (A) First simulation, with larger lagged than contemporaneous parameters (replicated for 75 individuals). (B) Second simulation, with smaller lagged than contemporaneous parameters (replicated for 75 individuals).

work. The second simulation (Figure 2B) contained relatively small lagged parameters in  $\Phi_1$  and relatively large contemporaneous parameters in  $A$ .

### Empirical data

Participants were 16 children aged 4–7 years, divided into 8 same-sex dyads (4 female, 4 male). Dyads consisted of a target child and an unrelated same-sex peer of the target's choosing. Dyads participated in video-recorded 14-to-15-minute play sessions designed to elicit rough-and-tumble interactions. In a semiencloded area, children were

instructed to play with a jump-o-leen, pillow, ball, and bobo-type doll (introduced midway through the session). Each child's behavior was coded in 6-second epochs for *vigor of activity* on a 1 (*no activity*) to 5 (*multiple vigorous movements throughout the majority of the epoch*) scale and for *positive affect* on a 1 (*no expression*) to 5 (*continuous positive vocal expression for the majority of the epoch*) scale, resulting in 4-variate time series of length 140 to 150. Two raters coded each dyad, and they were highly reliable, with single measures intraclass correlation coefficients averaging .80 ( $SD = .03$ ) for *vigor of activity* and .79 ( $SD = .08$ ) for *positive affect*; only the codes from the first rater were used in subsequent analyses.

### Data analysis plan

Each data set was submitted to GIMME-MS with a group-level criterion of 50%. When multiple solutions were found, a series of steps were conducted in order to select an optimal result; they are drawn from past research on model selection and summarized in the decision schematic presented in Figure 3. These criteria were applied to multiple solutions generated at the group level (prior to individual-level model fitting) because multiple solutions were not expected at the individual level due to the nature of the data sets; however, individual-level multiple solutions are further considered in the Discussion. Note that not all selection criteria used in past work were applicable here. For example, GIMME ensures that all identified relations are significant for the group or individual, so there is no need to remove solutions with statistically nonsignificant relations.

First and when possible, additional information was gathered in order to eliminate nonreplicated solutions (MacCallum et al., 1993). For the simulated data sets, this was accomplished through cross-validation in additional simulations with the same parameters. Only solutions that were structural replicates of initial solutions (i.e., had exactly the same pattern of relations) were maintained for further evaluation; solutions that occurred only once (between the initial and cross-validation analyses) were eliminated. For the empirical data set, this step was not possible because dyads were only assessed once, and time series were too short to split.

Second, fit indices were used to evaluate how well each (cross-validated) solution fit the observed (i.e., simulated or empirical) data. The four indices used in GIMME (and informed by Brown, 2006; Gates & Molenaar, 2012) were considered:  $RMSEA \leq .05$ ;  $NNFI \geq .95$ ;  $CFI \geq .95$ ;  $SRMR \leq .05$ . The average (across replications or dyads) was calculated for each index, and then the number of indices indicating excellent fit was used to evaluate group-level

solutions, identifying a favored solution (i.e., the one with the most indices indicating excellent fit).

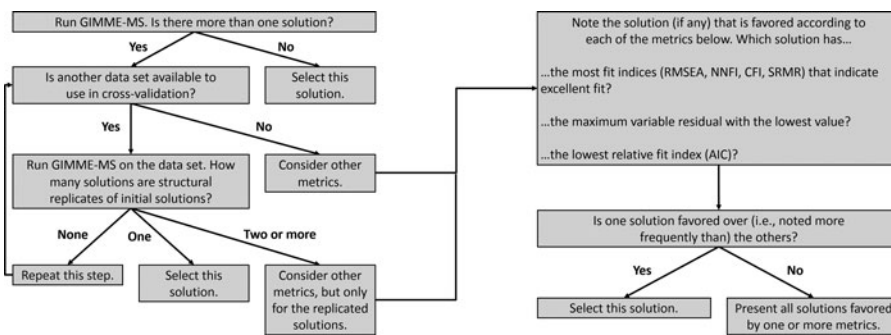
Third, standardized residuals of the variables were also used to evaluate how well each (cross-validated) solution fit the observed data. The largest residual was of primary consideration as it provides an index for the maximum degree of discrepancy between an estimated solution and the observed data (MacCallum & Austin, 2000). The average (across replications or dyads) maximum residual was used to evaluate group-level solutions, identifying a favored solution (i.e., the one with the smallest maximum residual).

Fourth, the AIC (Akaike, 1974), a log-likelihood-based information criterion, was used to compare the (cross-validated) solutions. The AIC is a nonnormed relative fit index that penalizes complexity, so models with few parameters and small AICs are favored over models with many parameters and high AICs. The average AIC (across replications or dyads) was used to evaluate group-level solutions, identifying a favored solution (i.e., the one with the smallest AIC).

The (cross-validated) solution that was most frequently favored by model selection metrics (i.e., fit indices, maximum variable residual, and an information criterion) was selected as optimal. If no solution was consistently favored, then all solutions with equal support were presented.

### Results

Expected results were found for the first simulation that contained large lagged versus contemporaneous relations: GIMME-MS identified a single group-level solution. The single solution accurately recovered the true model; the six parameters that were simulated for all replications were identified as group-level relations. This is shown in Figure 4, with the mean magnitude (across 75 replications) of the relations and their 95% confidence intervals, all of which included the true value (displayed in Figure 2A). Moreover, the parameters were identified according to their simulated order of magnitude, with the large lagged parameters freed first and the small contemporaneous parameters freed last; iteratively, group-level parameters  $\varphi_{3,1}$ ,  $\varphi_{2,2}$ ,  $\varphi_{1,1}$ ,  $\varphi_{4,4}$ ,  $\alpha_{4,3}$ , and  $\alpha_{1,2}$  were freed. Group-level relations fit the data well, indicated by mean fit indices:  $RMSEA = .01$ ;  $NNFI = 1.00$ ;  $CFI = 1.00$ ;  $SRMR = .02$ . There were no multiple solutions at the individual level, but for one replication, an additional contemporaneous individual-level relation ( $\alpha_{3,4}$ ) was found. This relation was not in the true model, and although significant, it was not essential to achieve excellent model fit, as three of four fit indices indicated that a model excluding this relation fit well.



**Figure 3.** Schematic representation of the process used to select an optimal solution from the set of possible solutions output by GIMME-MS. GIMME-MS = group-iterative multiple model estimation for multiple solutions; RMSEA = root mean squared error of approximation; NNFI = nonnormed fit index; CFI = comparative fit index; SRMR = standardized root mean square residual; AIC = Akaike information criterion.

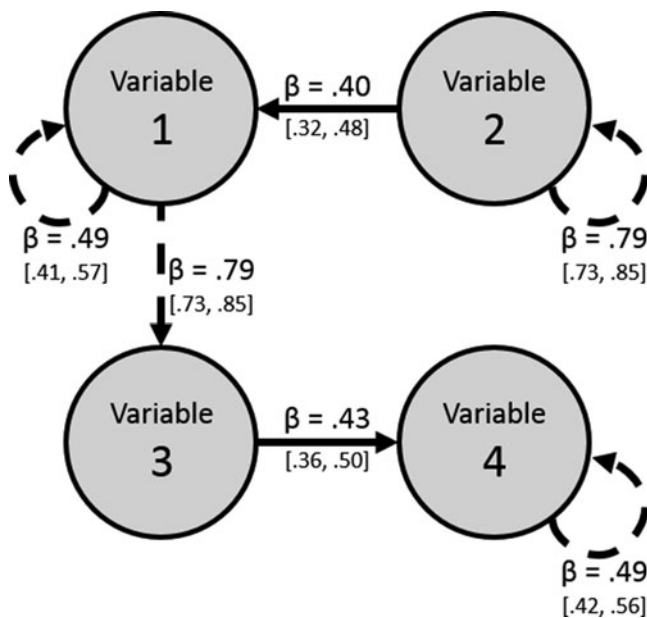
Expected results were also found for the second simulation that contained large contemporaneous versus lagged relations: GIMME-MS identified multiple group-level solutions. Specifically, five solutions were found (a, b, c, d, e); these are displayed in Figure 5A as branching columns depicting the order in which parameters were freed. Four iterations resulted in equal modification indices for two parameters; these are the bifurcating solution paths in the figure. In each of these iterations, the equal indices occurred early in the model-fitting process and concerned contemporaneous relations between the

same two variables:  $\alpha_{3,1}$  and  $\alpha_{1,3}$ ;  $\alpha_{4,2}$  and  $\alpha_{2,4}$ ;  $\alpha_{3,2}$  and  $\alpha_{2,3}$ . Parameters trimmed from the group model due to nonsignificance are crossed out in the figure. After trimming, solution d was not unique; it replicated the structure of solution c and was therefore excluded from the solution set.

To select an optimal solution, additional analyses were conducted following the steps in the decision schematic depicted in Figure 3. First, additional information was used to cross-validate the findings. Specifically, GIMME-MS was conducted on another set of data simulated according to the same parameters (i.e., the Simulation 2 parameters shown in Figure 2B). Results were similar to those of the initial analysis: As shown in Figure 5B, five group-level solutions were identified (f, g, h, i, j), and four iterations resulted in equal modification indices for two contemporaneous parameters ( $\alpha_{3,1}$  and  $\alpha_{1,3}$ ;  $\alpha_{4,2}$  and  $\alpha_{2,4}$ ;  $\alpha_{3,2}$  and  $\alpha_{2,3}$ ). The three solutions highlighted in gray (g, h, j) were replicates as they had the same structure (i.e., pattern of freed relations) but not necessarily the same sequence (i.e., relations may have been freed in different orders) as solutions from the initial analysis. The cross-validated solutions are shown as diagrams in Figure 6, and they were the only solutions to be further evaluated.

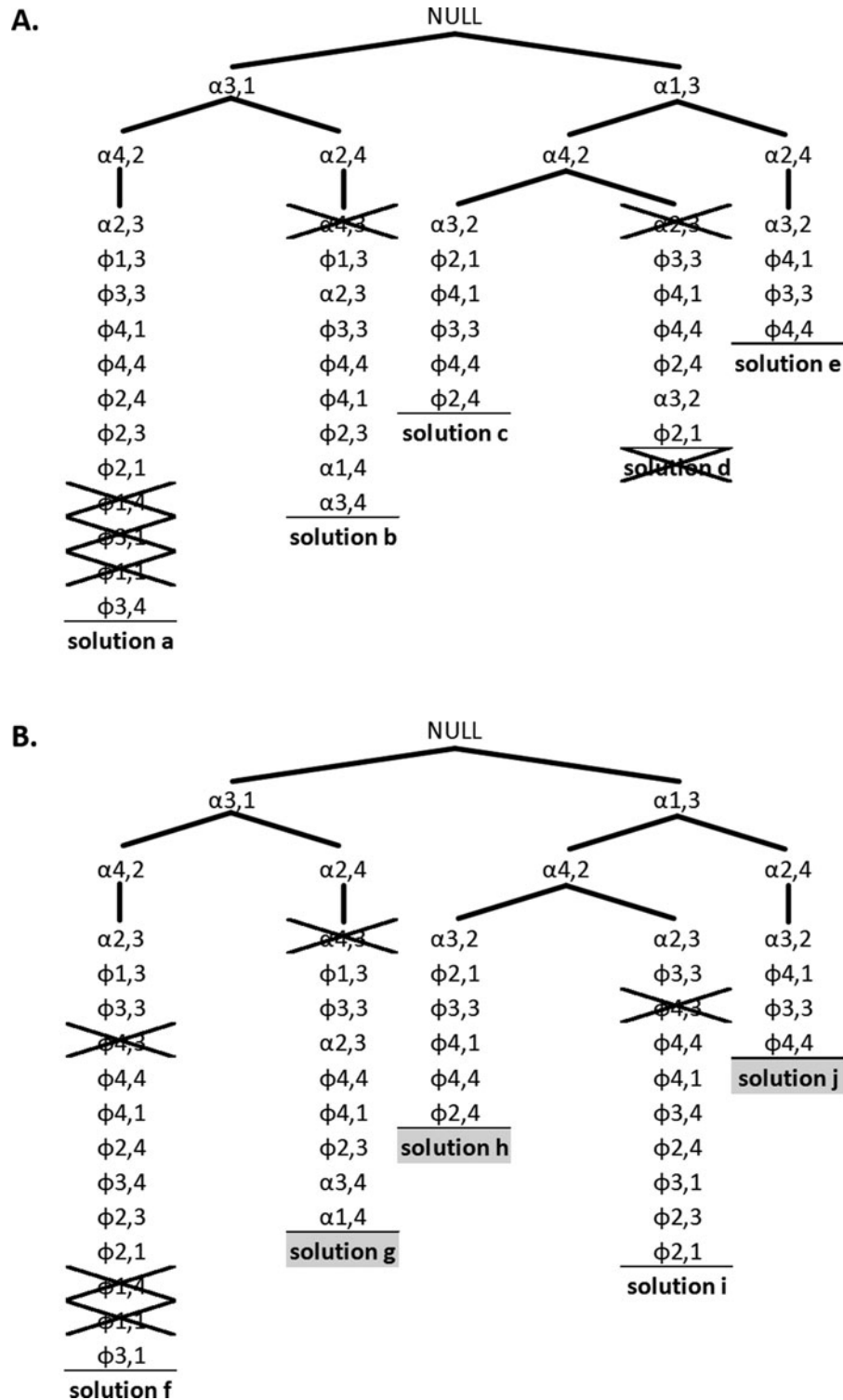
Second, fit indices were used to evaluate how well cross-validated solutions b, c, and e fit the simulated data. Table 1 shows the average RMSEA, NNFI, CFI, and SRMR for each group-level solution. According to these indices, all solutions provided an excellent fit to the data, so there was not a favored solution.

Third, standardized variable residuals were used to evaluate how well the cross-validated solutions fit the simulated data. Table 1 shows the average standardized residuals for each variable in each solution, with the largest residual in bold. Solution c had the largest residual, followed by solution b, then solution e. These results favor solution e, suggesting that its estimates minimize misfit with the simulated data.

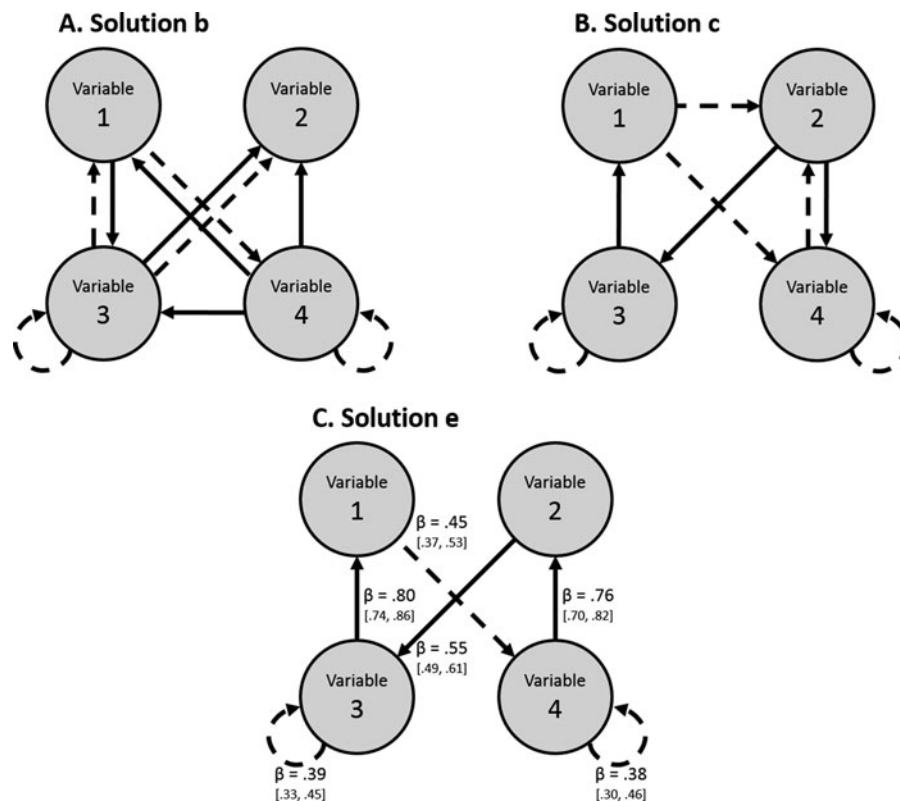


**Figure 4.** Diagram showing the single group-level solution from GIMME-MS analysis of data from Simulation 1, with relatively large lagged and small contemporaneous relations. The simulated parameters were identified as group-level relations among variables. Dashed lines reflect lagged variable relations; solid lines reflect contemporaneous variable relations;  $\beta$ -weights are the mean magnitude of the relations (averaged across 75 replications); bracketed values are 95% confidence intervals of the estimated relations (see Figure 2A).





**Figure 5.** Diagrams showing the group-level solutions from GIMME-MS analysis of data from Simulation 2 (columns), listing the parameter freed in each iteration of each solution (rows). Branches, or bifurcations, reflect iterations in which equivalent modification indices were found. Parameters that are crossed out were removed from the solution because they were not significant at the group criterion when freed in the context of all other relations. The crossed-out solution was not unique. (A) Results from analysis of the initial simulated data set. (B) Results from analysis of the cross-validation data set, with replicated solutions highlighted in gray; these solutions were identical in structure, but not necessarily in sequence, to a solution presented in (A).



**Figure 6.** Diagrams of the three cross-validated group-level solutions (shown in Figure 5A) from GIMME-MS analysis of data from Simulation 2, with relatively small lagged and large contemporaneous relations. Dashed lines reflect lagged variable relations; solid lines reflect contemporaneous variable relations. (A) Solution b is suboptimal. (B) Solution c is suboptimal. (C) Solution e is optimal according to model comparison metrics (see Table 1) and contains  $\beta$ -weights showing the mean magnitude of the relations (averaged across 75 replications); the pattern of group-level relations matches the simulated parameters, which are within the bracketed 95% confidence intervals of the estimated relations (see Figure 2B).

Fourth, the AIC was used to compare the cross-validated solutions. Table 1 shows the average AIC for each group-level solution. Solution e had the lowest AIC, followed by solution c, then solution b. Thus, results again favor solution e, suggesting that it provides the most information given the number of parameters.

Solution e was selected as optimal. It was one of three cross-validated solutions in an additional GIMME-MS analysis, and it was favored over the other two solutions

by model comparison metrics, having the lowest maximum standardized residual and the lowest AIC. Solution e is also the correct solution. It accurately recovered the true model; the six parameters that were simulated for all replications were identified as group-level relations. This is shown in Figure 6C, with the mean magnitude (across 75 replications) of the relations and their 95% confidence intervals, all of which included the true value (displayed in Figure 2B). There were no multiple solutions at the

**Table 1.** Model comparison metrics for cross-validated group-level solutions from Simulation 2 data.

	Solution b (Figure 6A)	Solution c (Figure 6B)	Solution e (Figure 6C)
$\chi^2$ -based fit indices			
RMSEA	.01	.01	.01
NNFI	1.00	1.00	1.00
CFI	1.00	1.00	1.00
SRMR	.02	.02	.02
Standardized residuals			
Variable 1	<b>.65</b>	.35	.35
Variable 2	.30	<b>.68</b>	.41
Variable 3	.25	.34	.34
Variable 4	.46	.27	<b>.46</b>
Information criterion			
AIC	454.43	452.08	450.33

*Note.* Maximum residuals for each solution are boldface. RMSEA = root mean squared error of approximation; NNFI = nonnormed fit index; CFI = comparative fit index; SRMR = standardized root mean square residual; AIC = Akaike information criterion.

individual level, but two replications in solution e contained additional individual-level relations that were not in the true model: a contemporaneous parameter ( $\alpha_{4,1}$ ) was freed in one replication and a lagged parameter ( $\varphi_{4,2}$ ) was freed in another replication. The individual-level relations, however, were not essential to achieve excellent model fit, as three of four fit indices indicated that models excluding the relations fit well.

Moreover, GIMME-MS was essential for identifying the correct solution as solution e was not the result of standard GIMME analysis (i.e., without multiple solutions) of Simulation 2 data. Additional exploratory analyses in which the Simulation 2 data were submitted to standard GIMME revealed a group-level model identical to the first seven parameters in solution a (i.e.,  $\alpha_{3,1}$ ,  $\varphi_{4,2}$ ,  $\alpha_{2,3}$ ,  $\varphi_{1,3}$ ,  $\varphi_{3,3}$ ,  $\varphi_{4,1}$ , and  $\varphi_{4,4}$ ).

Finally, additional simulation results are reported in Appendix B, demonstrating that GIMME-MS is robust to variations in sample size, time series length, and relative weights of relations. The first two simulations show that GIMME-MS recovers accurate parameter estimates for Simulation 2 data even when sample size ( $N = 15$ ) and the number of timepoints ( $T = 75$ ) are small. The third simulation shows that GIMME-MS recovers accurate parameter estimates even in a sparse model with large lagged and contemporaneous relations.

Expected results were found for the empirical data set: GIMME-MS identified multiple group-level solutions. Specifically, two solutions were found; there was a bifurcation early in the analysis that was due to a large contemporaneous relation between the *positive affect* of the target and peer. The solutions are shown as diagrams in Figure 7. The first (Figure 7A) contained three autoregressive relations, which were freed after the contemporaneous relation from peer to target *positive affect*. The second (Figure 7B) contained the same three autoregressive relations, which were freed after the contemporaneous relation from target to peer *positive affect*. Thus, the only difference between the two group-level solutions was the direction of the contemporaneous relation.

Additional analyses were conducted in order to select an optimal group-level model, following the steps in the decision schematic depicted in Figure 3. No additional information was available for cross-validation, so model comparison metrics were exclusively used; the results are shown in Table 2. No fit indices indicated that a group-level solution fit the empirical data well, likely because dyad-specific relations were not yet included in the solutions. Standardized variable residuals favored the first solution as it had a smaller maximal residual than the second solution. The information criterion also favored the first solution as it had a lower AIC than the

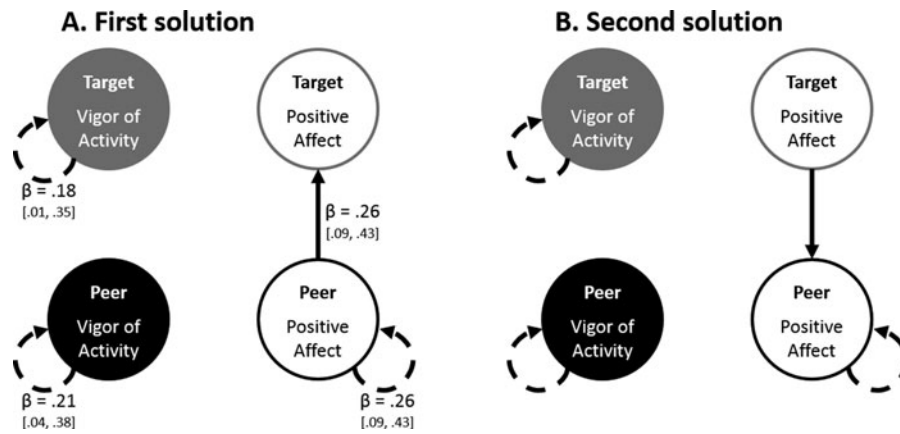
second solution. Thus, the first solution was selected as optimal because it was favored over the second solution by two model selection criteria. Figure 7A shows the mean magnitude (across dyads) of the group-level relations and their 95% confidence intervals for this solution. Individual-level relations were then added to the selected group-level model. No individual-level multiple solutions were generated, but one dyad received two individual-level relations; five dyads received one individual-level relation; and two dyads did not receive any individual-level relations. The resulting models all had excellent fit, according to average fit indices: RMSEA = .01, NNFI = 1.00, CFI = .99, and SRMR = .06.

## Discussion

Multiple solutions resulted from group-level uSEM analysis of simulated and empirical time series data that contained large contemporaneous relations. A new version of GIMME, which reveals relations among variables at different time scales (lagged and contemporaneous) and at different levels of analysis (group and individual)—dubbed GIMME for multiple solutions (GIMME-MS)—was developed to identify the complete set of possible solutions. Selection procedures validated in past research (Akaike, 1974; Brown, 2006; Gates & Molenaar, 2012; MacCallum & Austin, 2000; MacCallum et al., 1993) were then implemented in order to choose the optimal solution from the set of possibilities. These procedures included cross-validation in a new data set (when one was available) and using metrics, such as maximum variable residuals and the AIC.

This research is important, filling a knowledge gap in the literature. Past work has shown that uSEM implemented in GIMME is an optimal structural VAR instantiation because it reveals lagged and contemporaneous relations among variables with greater accuracy than other approaches (Gates & Molenaar, 2012). Nothing was known, however, about the uniqueness of GIMME results, even though there was reason to think that multiple, competing solutions could be present, since uSEM is conducted within an SEM framework and multiple solutions are a characteristic of cross-sectional path models (MacCallum et al., 1993; Raykov & Marcoulides, 2001).

Expected results were found during GIMME-MS analysis of three data sets. The first data set was simulated with four variables and contained four moderate or large lagged relations and two small contemporaneous relations. GIMME-MS recovered a single solution, which accurately recovered the pattern of true group-level relations, with the true parameter values falling within the 95% confidence intervals of the mean estimated parameters. The lagged relations (with large



**Figure 7.** Diagrams of the two group-level solutions from GIMME-MS analysis of empirical data. Dashed lines reflect lagged variable relations; solid lines reflect contemporaneous variable relations. (A) The first solution is optimal according to model comparison metrics (see Table 2) and contains  $\beta$ -weights showing the mean magnitude of the relations (averaged across 8 dyads) with bracketed 95% confidence intervals. (B) The second solution is suboptimal.

modification indices) were freed before the contemporaneous relations (with small modification indices). The lagged relations estimated in uSEM resemble relations in standard VAR models, and it is known that the latter generally yield unique solutions (Lütkepohl, 2005).

The second data set was simulated with four variables and contained three moderate or large contemporaneous relations and three small lagged relations. GIMME-MS recovered multiple solutions because the contemporaneous relations (with large modification indices) were freed before the lagged relations (with small modification indices); the contemporaneous relations estimated in uSEM resemble relations in directed path models as obtained in standard cross-sectional research, and it is known that the latter yield multiple solutions (MacCallum et al., 1993). Specifically, GIMME-MS found five (but only four unique) solutions. All were plausible and instigated substantively different interpretations.

Together, results from the simulated data confirmed expectations that multiple solutions were more likely to result from a data set containing large contempora-

neous relations than from one containing large lagged relations. Expectations were based on knowledge of iterative model building via modification indices, which indicate how much a model's  $\chi^2$  will be reduced if a particular parameter is freed in the subsequent iteration.

When contemporaneous relations are large, modification indices indicate that a contemporaneous relation should be freed in the first iteration. At this early point in the sequence of Lagrange multiplier tests, however, there is no information available to determine the directionality of the contemporaneous relation. For example, assume that modification indices show that a contemporaneous relation between variables  $m$  and  $n$  should be freed in the first iteration; since no variance has been explained in either variable, freeing  $\gamma_{m,n}$  or  $\gamma_{n,m}$  will improve model fit by the same amount (just as the same amount of variance is explained in a regression of  $x$  on  $y$  or of  $y$  on  $x$ ). Instead of using arbitrary decision criteria (e.g., infinitesimal differences in Lagrange multiplier tests reflecting machine precision as occurs during the automatic search in LISREL) to fit a directed relation or instead of fitting

**Table 2.** Model comparison metrics for group-level solutions from empirical data.

	First solution (Figure 7A)	Second solution (Figure 7B)
$\chi^2$ -based fit indices		
RMSEA	.07	.07
NNFI	.68	.65
CFI	.77	.76
SRMR	.08	.08
Standardized residuals		
Target activity	<b>.95</b>	.95
Target affect	.91	<b>1.00</b>
Peer activity	.94	.94
Peer affect	.92	.84
Information criterion		
AIC	880.04	800.91

Note. Maximum residuals for each solution are boldface. RMSEA = root mean squared error of approximation; NNFI = nonnormed fit index; CFI = comparative fit index; SRMR = standardized root mean square residual; AIC = Akaike information criterion.



a bidirectional relation between the variables, GIMME-MS combined with model selection criteria determines the directionality of the contemporaneous relation by creating different solution paths, with each path containing the freed relation in an opposite direction, before commencing iterative model generation. (GIMME-MS does not, however, restrict variables from reciprocally influencing one another as it is possible for contemporaneous and lagged relations to be freed in both directions between the variables.) However, when lagged relations are large, and thus freed in the first iteration, the temporal nature of the time series data provides statistical context that can be used to infer directionality. For example, if the autoregressive component of  $m$  was freed in the first iteration, then the second iteration modification indices for the directed contemporaneous relations between  $m$  and  $n$  will not be equal or lead to multiple solutions (just as the amount of variance explained differs in a regression of  $x$  on  $y$  or of  $y$  on  $x$  when a covariate is included in the models);  $m$  will be able to explain more variance in  $n$  than  $n$  will be able to explain in  $m$  because some variation in  $m$  has already been accounted for by the autoregressive component.

Informed by past work on multiple solutions, decision criteria to be used when selecting a solution from a set of possibilities generated by GIMME-MS were implemented. The criteria are cross-validation in another data set, evaluation of model fit indices, evaluation of model fit using the maximum standardized variable residuals as an indicator of model misfit, and comparison of models using an information criterion. Implementation of these criteria resulted in selection of the correct solution (i.e., solution e) for the data set generated by Simulation 2; this is notable because standard GIMME analysis of the same data set resulted in a different solution. The selected solution had the fewest parameters, so it is not surprising that it also had the lowest AIC since the AIC penalizes complexity. For the same reason, however, it is noteworthy that the selected solution had the lowest maximum residual as complexity is implicitly rewarded by this metric (i.e., a solution with many parameters has the potential to explain more variance than a solution with few parameters). This solution accurately recovered the pattern of true group-level relations, with the true parameter values falling within the 95% confidence intervals of the mean estimated parameters. This convincingly showcases how selection metrics can be used to identify an optimal solution—when the set of all possible solutions has been generated by GIMME-MS.

An important test of the utility of GIMME-MS was conducted when a third, empirical data set was submitted to the program. The data set contained four variables, reflecting ratings of the *vigor of activity* and *positive affect* for each child participating in a dyadic play session

intended to elicit rough-and-tumble interactions. Data from eight dyads were analyzed. GIMME-MS recovered two solutions, with a contemporaneous relation between the *positive affect* of the two children freed before three autoregressive relations; the direction of the contemporaneous relations was opposite in the two solutions. Cross-validation could not be conducted as no additional data were available, so the optimal solution was selected via fit indices, maximum standardized residuals, and an information criterion. The first solution was selected; it was favored by two metrics, having the smallest mean maximum residual and the lowest AIC. It is important to note, however, that the margin of difference between the two solutions was small. Nonetheless, application of GIMME-MS to this empirical data set highlighted the value of the program for substantive research. Due to the similarity in the group-level solutions, it also highlighted the importance of explicitly selecting a solution based upon a set of predetermined criteria over allowing seemingly random differences in modification indices (e.g., based on machine precision) to determine the selection.

There are some important points to consider when interpreting results from the current study and when applying GIMME-MS and the model selection procedures used here to other data sets. First, GIMME-MS currently implements lagged processes of the first order. This has been successful in uSEM applications to brain and behavioral data sets (Beltz, Beekman, et al., 2013; Beltz, Gates, et al., 2013). However, it might not be appropriate for all data sets. A posteriori model testing procedures can be used to determine whether a first-order uSEM is sufficient or whether a higher-order uSEM is needed in applications of GIMME (Beltz & Molenaar, 2015). There is a trade-off between the presence of multiple solutions and the temporal order of a data set. In general, time series with long measurement intervals have low temporal orders (e.g., lag one is sufficient), large contemporaneous relations, and small lagged relations compared to those with short measurement intervals, which have high temporal orders (e.g., lags greater than one are required), small contemporaneous relations, and large lagged relations. Multiple solutions are more prevalent in the former than in the latter and, therefore, are expected in a behavioral daily diary data set with measurements every 24 hours but not in a task-related functional magnetic resonance imaging data set with measurements every 2 seconds (on which past GIMME applications have focused).

Second, the decision schematic used here can be applied at both the group and individual levels. At the group level, the selection metrics (fit indices, variable residuals, information criteria) must be averaged across individuals before comparing solutions, but the metrics can be assessed directly (i.e., no averaging is

necessary) at the individual level. Only group-level evaluations were conducted here because the grouping procedures utilized by GIMME are unique and partially responsible for its success over other structural VAR and time series analysis approaches and because there were no multiple solutions at the individual level. Had there been multiple solutions at the individual level, the generalization of the presented multiple solutions procedures from the group level to the individual level would have been straightforward. Individual-level multiple solutions may occur in person-specific data or in group-level data that are very heterogeneous or that contain a sparse group-level structure. When working with person-specific data, the AIC is a promising metric for selecting a solution from a set of possibilities (see, e.g., Wright, Beltz, Gates, Molenaar, & Simms, 2015). It does not require solutions to be nested and is meaningful even when variables in the model are unexplained (e.g., person-specific solutions are often sparse and have maximum standardized residuals of 1.00). Further exploration of this topic is necessary.

Third, the model selection procedures can be modified to meet the needs of a specific data set or the preferences of a researcher. For example, the results presented here suggest that fit indices may not be a discriminating criterion for all data sets as they twice failed to favor a solution. In the simulated data set, fit indices indicated that all solutions had excellent fit, likely due to homogeneity in the data and the use of the RMSEA, NNFI, CFI, and SRMR by GIMME-MS in model building. In the empirical data set, fit indices indicated that no solution had an excellent fit, likely due to heterogeneity in the data, which is accounted for by dyad-specific relations that had not yet been added when group-level solutions were evaluated. Researchers could elect to use different fit indices as selection criteria if they are extensively validated in simulation experiments, as has been done for the indices used here (Gates & Molenaar, 2012). Researchers could also elect to use theoretical rationale as a selection criterion.

Fourth, there are additional features of GIMME-MS that can be explored in future investigations. One example concerns the criterion value: It can be modulated, allowing users to determine the percent of participants for whom a relation must be significant in order for that relation to be freed at the group level. The criterion value was 50% for the current study; this is lower than the suggested 75% (see Gates & Molenaar, 2012) in order to accommodate the small, heterogeneous behavioral data set analyzed here (compared to the large, relatively homogeneous neuroimaging data sets used in prior applications of GIMME). Researchers using GIMME-MS for other behavioral applications should similarly consider altering the criterion value to suit their data set and research question as the optimal value may vary. Another example

concerns data with external input (i.e., experimental conditions): They can be analyzed with an extended uSEM (euSEM) in GIMME-MS (see Gates & Molenaar, 2012; Gates et al., 2011). In these analyses, lagged and contemporaneous direct effects of the conditions on the variables are modeled. Bilinear effects, or lagged relations among variables that are only present during a condition, are also modeled. Because of the time series length required to run euSEM models, they are currently more prevalent in the brain sciences (e.g., Hillary, Medaglia, Gates, Molenaar, & Good, 2014) than in the behavioral sciences.

## Conclusion

The value of person-specific time series analysis approaches is beginning to be recognized in psychological science, and structural VARs, which map lagged and contemporaneous relations among time-indexed variables, hold great potential in this regard. uSEM is an accurate structural VAR instantiation built on an SEM framework that—when implemented in a data-driven fashion with GIMME—models lagged and contemporaneous relations at the group and individual levels. Although its accuracy has been established, until now, the issue of multiple GIMME solutions had gone unaddressed. A new version of GIMME, dubbed GIMME-MS, was developed to rectify this, generating separate solution paths when Lagrange multiplier tests indicate that multiple relations equally and maximally improve model fit at a given iteration. Results from simulated and empirical data showed that multiple solutions were present when there were large contemporaneous relations in a data set, an expected finding because multiple solutions are a characteristic of cross-sectional path models. A series of selection procedures, including cross-validation, fit indices, variable residuals, and an information criterion, demonstrated how the optimal solution could be selected from the possibilities offered by GIMME-MS. Thus, GIMME-MS is a novel and accurate analysis approach, reflecting the state of the science for mapping personalized behavioral time series data.

## Article information

**Conflict of Interest Disclosures:** Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

**Ethical Principles:** The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment

and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

**Funding:** This work was supported by grant 1157220 from the National Science Foundation to Peter Molenaar. Funding was also provided by a Superior Teaching and Research Award to Adriene M. Beltz from the Pennsylvania State University College of the Liberal Arts.

**Role of the Funders/Sponsors:** None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

**Acknowledgments:** The authors thank Joseph Stitt for lending his programming expertise to the GIMME-MS program, and Sheri Berenbaum for generously providing the empirical data presented in this manuscript. They also thank Matthew Lum and Cassondra Eng for coding the empirical data, and Elizabeth Beckerman and Todd Cly for managing the empirical data. Portions of this work were presented at the Organization for Human Brain Mapping 2014 Annual Meeting. The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors' institution or the National Science Foundation is not intended and should not be inferred.

## References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *AC-19*(6), 716–723. doi:10.1109/tac.1974.1100705
- Beltz, A. M., Beekman, C., Molenaar, P. C. M., & Buss, K. A. (2013). Mapping temporal dynamics in social interactions with unified structural equation modeling: A description and demonstration revealing time-dependent sex differences in play behavior. *Applied Developmental Science*, *17*(3), 152–168. doi:10.1080/10888691.2013.805953
- Beltz, A. M., Gates, K. M., Engels, A. S., Molenaar, P. C. M., Pulido, C., Turrissi, R., et al. (2013). Changes in alcohol-related brain networks across the first year of college: A prospective pilot study using fMRI connectivity mapping. *Addictive Behaviors*, *38*(4), 2052–2059. doi:10.1016/j.addbeh.2012.12.023
- Beltz, A. M., & Molenaar, P. C. M. (2015). A posteriori model validation for the temporal order of directed functional connectivity maps. *Frontiers in Neuroscience: Brain Imaging Methods*, 9, article 304. doi:10.3389/fnins.2015.00304
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time series analysis: Forecasting and control* (4th ed.). Hoboken, NJ: John Wiley & Sons, Inc.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- Chou, C. P., & Bentler, P. M. (1990). Model modification in covariance structure modeling: A comparison among likelihood ratio, Lagrange Multiplier, and Wald tests. *Multivariate Behavioral Research*, *25*(1), 115–136. doi:10.1207/s15327906mbr2501\_13
- de Marco, G., Vrignaud, P., Destrieux, C., de Marco, D., Testelin, S., Devauchelle, B., et al. (2009). Principle of structural equation modeling for exploring functional interactivity within a putative network of interconnected brain areas. *Magnetic Resonance Imaging*, *27*(1), 1–12. doi:10.1016/j.mri.2008.05.003
- Gates, K. M., Molenaar, P. C., Hillary, F. G., Ram, N., & Rovine, M. J. (2010). Automatic search for fMRI connectivity mapping: an alternative to Granger causality testing using formal equivalences among SEM path modeling, VAR, and unified SEM. *NeuroImage*, *50*(3), 1118–1125. doi:10.1016/j.neuroimage.2009.12.117
- Gates, K. M., & Molenaar, P. C. M. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage*, *63*(1), 310–319. doi:10.1016/j.neuroimage.2012.06.026
- Gates, K. M., Molenaar, P. C. M., Hillary, F. G., & Slobounov, S. (2011). Extended unified SEM approach for modeling event-related fMRI data. *NeuroImage*, *54*(2), 1151–1158. doi:10.1016/j.neuroimage.2010.08.051
- Goodwin, M. S., Velicer, W. F., & Intille, S. S. (2008). Telemetric monitoring in the behavior sciences. *Behavior Research Methods*, *40*(1), 328–341. doi:10.3758/BRM.40.1.328
- Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2005). Statistical modeling of the individual: Rationale and application of multivariate stationary time series analysis. *Multivariate Behavioral Research*, *40*(2), 207–233. doi:10.1207/s15327906mbr4002\_3
- Henley, A. B., Shook, C. L., & Peterson, M. (2006). The presence of equivalent models in strategic management research using structural equation modeling: Assessing and addressing the problem. *Organizational Research Methods*, *9*(4), 516–535. doi:10.1177/1094428106290195
- Hillary, F. G., Medaglia, J. D., Gates, K. M., Molenaar, P. C., & Good, D. C. (2014). Examining network dynamics after traumatic brain injury using the extended unified SEM approach. *Brain Imaging and Behavior*, *8*(3), 435–445. doi:10.1007/s11682-012-9205-0
- Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453. doi:10.1037/1082-989x.3.4.424
- Jöreskog, K. G., & Sörbom, D. (1992). *LISREL*. Skokie, IL: Scientific Software International, Inc.
- Kaplan, D. (1989). Model modification in covariance structure analysis: Application of the expected parameter change statistic. *Multivariate Behavioral Research*, *24*(3), 285–305. doi:10.1207/s15327906mbr2403\_2
- Kim, J., Zhu, W., Chang, L., Bentler, P. M., & Ernst, T. (2007). Unified structural equation modeling approach

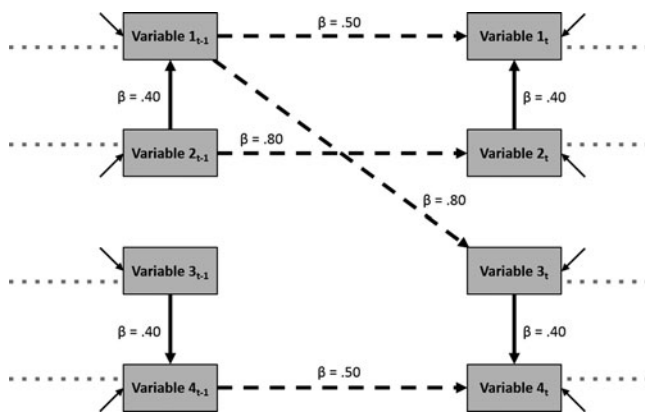


- for the analysis of multisubject, multivariate functional MRI data. *Human Brain Mapping*, 28(2), 85–93. doi:10.1002/hbm.20259
- Loehlin, J. C. (1996). The Cholesky approach: A cautionary note. *Behavior Genetics*, 26(1), 65–69. doi:10.1007/BF02361160
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Berlin, Germany: Springer.
- MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, 51(1), 201–226. doi:10.1146/annurev.psych.51.1.201
- MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, 111(3), 490–504. doi:10.1037/0033-2909.111.3.490
- MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, 114(1), 185–199. doi:10.1037/0033-2909.114.1.185
- Markus, K. A. (2002). Statistical equivalence, semantic equivalence, eliminative induction, and the Raykov-Marcoulides proof of infinite equivalence. *Structural Equation Modeling*, 9(4), 503–522. doi:10.1207/s15328007sem0904\_3
- Martin, V., Hurn, S., & Harris, D. (2013). *Econometric modelling with time series: Specification, estimation and testing*. New York, NY: Cambridge University Press.
- Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, 50(2), 181–202. doi:10.1007/bf02294246
- Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, 2(4), 201–218. doi:10.1207/s15366359mea0204\_1
- Molenaar, P. C. M., & Lo, L. L. (2016). Alternative forms of Granger causality, heterogeneity and non-stationarity. In A. von Eye & W. Wiedermann (Eds.), *Statistics and causality*. Hoboken, NJ: Wiley.
- Molenaar, P. C. M., & Nesselroade, J. R. (1998). A comparison of pseudo-maximum likelihood and asymptotically distribution-free dynamic factor analysis parameter estimation in fitting covariance-structure models to block-Toeplitz matrices representing single-subject multivariate time-series. *Multivariate Behavioral Research*, 33(3), 313–342. doi:10.1207/s15327906mbr3303\_1
- Nesselroade, J. R., & Ram, N. (2004). Studying intraindividual variability: What we have learned that will help us understand lives in context. *Research in Human Development*, 1(1&2), 9–29. doi:10.1080/15427609.2004.9683328
- Poldrack, R. A., & Wagner, A. D. (2004). What can neuroimaging tell us about the mind? Insights from prefrontal cortex. *Current Directions in Psychological Science*, 13(5), 177–181. doi:10.1111/j.0963-7214.2004.00302.x
- Ramsey, J. D., Hanson, S. J., & Glymour, C. (2011). Multi-subject search correctly identifies causal connections and most causal directions in the DCM models of the Smith et al. simulation study. *NeuroImage*, 58(3), 838–848. doi:10.1016/j.neuroimage.2011.06.068
- Raykov, T., & Marcoulides, G. A. (2001). Can there be infinitely many models equivalent to a given covariance structure model?. *Structural Equation Modeling*, 8(1), 142–149. doi:10.1207/s15328007sem0801\_8
- Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., et al. (2011). Network modelling methods for FMRI. *NeuroImage*, 54(2), 875–891. doi:10.1016/j.neuroimage.2010.08.063
- Sörbom, D. (1989). Model modification. *Psychometrika*, 54(3), 371–384. doi:10.1007/bf02294623
- Valdés-Sosa, P. A., Sánchez-Bornot, J. M., Lage-Castellanos, A., Vega-Hernández, M., Bosch-Bayard, J., Melie-García, L., et al. (2005). Estimating brain functional connectivity with sparse multivariate autoregression. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 360(1457), 969–981. doi:10.1098/rstb.2005.1654
- Wright, A. G. C., Beltz, A. M., Gates, K. M., Molenaar, P. C. M., & Simms, L. J. (2015). Examining the dynamic structure of daily internalizing and externalizing behavior at multiple levels of analysis. *Frontiers in Psychology*, 6, article 1914. doi:10.3389/fpsyg.2015.01914
- Zhang, Z., Hamaker, E. L., & Nesselroade, J. R. (2008). Comparisons of four methods for estimating a dynamic factor model. *Structural Equation Modeling*, 15(3), 377–402. doi:10.1080/10705510802154281
- Zhuang, J. C., LaConte, S., Peltier, S., Zhang, K., & Hu, X. P. (2005). Connectivity exploration with structural equation modeling: An fMRI study of bimanual motor coordination. *NeuroImage*, 25(2), 462–470. doi:10.1016/j.neuroimage.2004.11.007

## Appendix A

Figure A1 is an alternate representation of Figure 2A. It is a diagram showing the parameters used in the first simulation. Dashed lines between variables reflect lagged relations (i.e., from  $t - 1$  to  $t$ ); solid lines between variables reflect contemporaneous relations (e.g., at the same  $t$ );  $\beta$ -weights reflect the magnitude of the relations; small solid lines reflect variances standardized at 1; and ellipses reflect the continuation of the variables in time. All other network figures presented in the current work (Figures 2, 4, 6, 7, and those in Appendix B) can be depicted using this alternate representation.





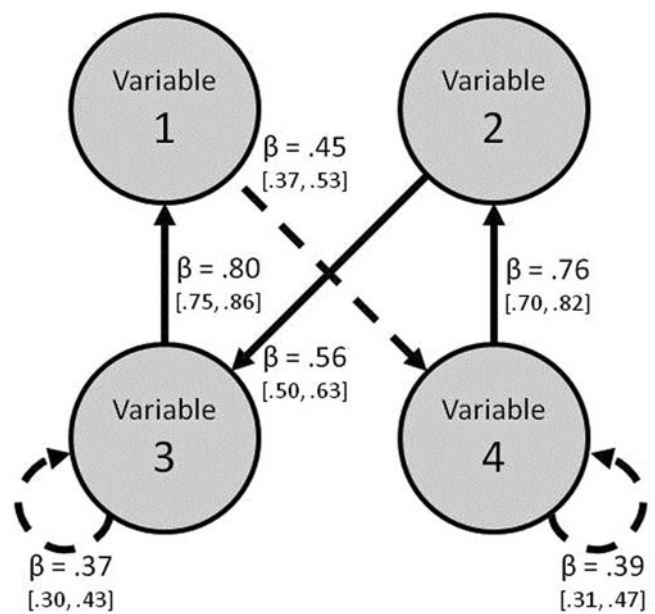
**Figure A1.** Alternate representation of the network shown in Figure 2A. See Appendix A text for description.

## Appendix B

Additional simulations were conducted to examine the performance of GIMME-MS when sample size ( $N$ ), number of timepoints ( $T$ ), and strength of relations ( $\beta$ ) might challenge the generation and selection of the correct map. Three simulations are reported, with the first reducing  $N$  with respect to initial simulations, the second reducing  $N$  and  $T$  with respect to initial simulations, and the third containing sparse, large lagged and contemporaneous relations.

### Simulation with small $N$

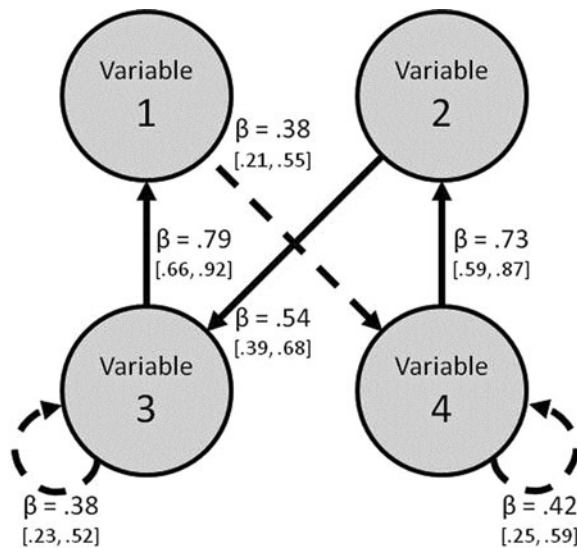
A structural VAR data set with the same parameters as Simulation 2 ( $T = 300$ ; see Figure 2B for  $\beta$  weights) was simulated, but data were generated for  $N = 15$  individuals (instead of 75). The simulated data were submitted to GIMME-MS. Five group-level solutions were generated, two of which were unique and cross-validated in a second data set simulated with identical parameters. Both cross-validated solutions fit the data well according to absolute fit indices as group-level means for all four indices indicated excellent fit. One solution had a lower mean standardized maximum residual (.44 vs. .68) and AIC (411 vs. 414) than the other, so it was selected as optimal. It is shown in Figure B1 with the mean magnitude (across 15 individuals) of the relations and their 95% confidence intervals. Comparison of these results to the simulated parameters in Figure 2B shows that GIMME-MS recovered the true model, with the  $\beta$  weights of the simulated relations falling within the confidence intervals of the estimates. Thus, GIMME-MS is accurate even with a relatively small sample size.



**Figure B1.** Optimal group-level solution selected from GIMME-MS analysis of data generated using the same parameters as simulation 2, but with a small sample ( $N = 15$ ). Dashed lines reflect lagged variable relations, solid lines reflect contemporaneous variable relations,  $\beta$ -weights are the mean magnitude of the relations (averaged across 15 replications), and bracketed values are 95% confidence intervals of the estimated relations.

### Simulation with small $N$ and $T$

A structural VAR data set with the same parameters as Simulation 2 (see Figure 2B for  $\beta$  weights) was simulated, but data were generated for  $N = 15$  individuals (instead of 75) and  $T = 75$  (instead of 300). The simulated data were submitted to GIMME-MS. Four group-level solutions were generated, two of which were unique and cross-validated in a second data set simulated with identical parameters. Both cross-validated solutions fit the data well according to absolute fit indices as group-level means for at least two of four indices indicated excellent fit. One solution had a lower mean standardized maximum residual (.49 vs. .68) and AIC (170 vs. 177) than the other, so it was selected as optimal. It is shown in Figure B2 with the mean magnitude (across 15 individuals) of the relations and their 95% confidence intervals. Comparison of these results to the simulated parameters in Figure 2B shows that GIMME-MS recovered the true model, with the  $\beta$  weights of the simulated relations falling within the confidence intervals of the estimates; note that the size of the confidence intervals increased with the decrease in timepoints. Thus, GIMME-MS is accurate even with a relatively small sample size and time series length.

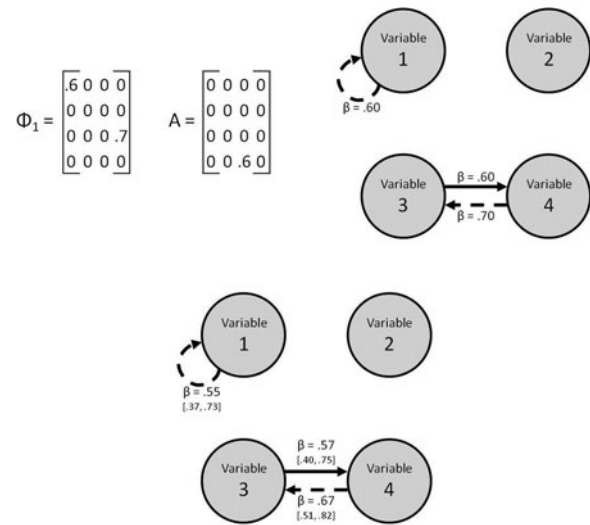


**Figure B2.** Optimal group-level solution selected from GIMME-MS analysis of data generated using the same parameters as simulation 2, but with a small sample ( $N = 15$ ) and number of time points ( $T = 75$ ). Dashed lines reflect lagged variable relations, solid lines reflect contemporaneous variable relations,  $\beta$ -weights are the mean magnitude of the relations (averaged across 15 replications), and bracketed values are 95% confidence intervals of the estimated relations.

### Simulation with sparse, large lagged and contemporaneous relations

A structural VAR data set was simulated according to the standardized uSEM parameters (top portion of Figure B3); note that the map is sparse, containing only three relatively large standardized relations. The data set consisted of a relatively small sample ( $N = 15$ ) and relatively short time series ( $T = 75$ ). The simulated data were submitted to GIMME-MS. A single group-level solution was generated, with  $\varphi_{3,4}$  first estimated, followed by  $\alpha_{4,3}$ , and then  $\varphi_{1,1}$ . The solution fit the data well according to absolute fit indices as group-level means for three of four indices indicated excellent fit: RMSEA = .03; NNFI = .98; CFI = .98; SRMR = .08. The solution

is shown (bottom portion of Figure B3) with the mean magnitude (across 15 individuals) of the relations and their 95% confidence intervals. Comparison of these results to the simulated parameters shows that GIMME-MS recovered the true model, with the  $\beta$  weights of the simulated relations falling within the confidence intervals of the estimates. Even though all simulated relations were relatively large, the largest lagged relation was estimated first, providing statistical information for the modification indices to accurately determine the direction of the contemporaneous relation. Thus, GIMME-MS is accurate even when relative relations are large and the sample and number of timepoints are small.



**Figure B3.** Parameters for a sparse data set simulated for a small sample ( $N = 15$ ) and number of time points ( $T = 75$ ) according to a uSEM, shown in lagged ( $\Phi_1$ ) and contemporaneous ( $A$ ) matrices and as a network (top), and results of GIMME-MS analysis of the simulated data (bottom). In the networks, dashed lines reflect lagged variable relations, solid lines reflect contemporaneous variable relations,  $\beta$ -weights are the true (top) and estimated mean (bottom) magnitude of the relations (averaged across 15 replications), and bracketed values are 95% confidence intervals of the estimated relations.