Original Article

GIMME's Ability to Recover Group-Level Path Coefficients and Individual-Level Path Coefficients

Steffen Nestler [a], Sarah Humberg [a]

[a] *Faculty of Psychology & Sports Science, University of Münster, Münster, Germany.*

Abstract

The growing availability of intensive longitudinal data has increased psychological researchers' interest in ideographic-statistical methods that, for example, reveal the contemporaneous or lagged associations between different variables for a specific individual. However, when researchers assess several individuals, the results of such models are difficult to generalize across individuals. Researchers recently suggested an algorithm called GIMME, which allows for the identification of coefficients that exist across all individuals (group-level coefficients) or are specific to one or a subgroup of individuals (individual-level coefficients). In three simulation studies we investigated GIMME's performance in recovering group-level and individual-level coefficients. For the former, we found that GIMME performed well when the magnitude of the parameters was moderate to high and when the number of measurements was sufficiently large. However, GIMME had problems detecting individual-level coefficients or coefficients that occurred for a subset of individuals from the whole sample.

In recent years, psychology has witnessed an explosion of research using ambulatory assessment techniques such as experience sampling, daily diary studies, and mobile sensing (e.g., Wrzus & Mehl, 2015). These techniques allow researchers to intensively measure the thoughts, feelings, or behaviors of individuals in their natural environments. They are therefore assumed to yield more valid information about a person in comparison with classical assessment techniques such as questionnaires. Also, the resulting longitu-

dinal data permit the investigation of exciting psychological research questions that refer to the processes that are assumed to underlie a number of psychological phenomena (Hamaker & Wichers, 2017; Wright & Zimmermann, 2019).

From a statistical-methodological point of view, ambulatory assessment techniques have led to an increase in the interest in and the use of ideographic statistical methods (Hamaker & Wichers, 2017; Nesselroade & Ram, 2004). These methods use intensive longitudinal data from a single participant to model variations and associations *within* this participant. Most ideographic statistical techniques are based on (or are equivalent to) time-series models. This class of models is very popular in economics (Box-Steffensmeier, Freeman, Hitt, & Pevehouse, 2014) and also has a long history in psychology (e.g., Cattell & Luborsky, 1950). It includes techniques such as univariate and multivariate autoregressive models, exploratory factor analysis techniques for single individuals (P-Technique), dynamic factor analyses, white noise factor models, or simple individual path models (e.g., Hamaker, Dolan, & Molenaar, 2002, 2003, 2005; Little, 2013; Nesselroade, McArdle, Aggen, & Meyers, 2002; Voelkle, Oud, von Oertzen, & Lindenberger, 2012; Zhang, Hamaker, & Nesselroade, 2008).

Idiographic models have been routinely contrasted with nomothetic statistical approaches in which the focus lies in explaining and predicting interindividual rather than intraindividual variation, that is, the determinants and consequences of differences *between* persons. Typically, the two approaches have been placed in opposition to each other because nomothetic models are based on (and permit) averaging across individual participants (Beltz & Gates, 2016). However, studies have shown that when the processes depicted in the statistical model are heterogeneous across individuals, results that are based on nomothetic models are usually neither equivalent to nor consistent with results from ideographic models (Molenaar, 2004; Hamaker et al., 2005). This observation suggests that when researchers expect heterogeneous processes to occur for different people, they should use ideographic statistical models.

One limitation of a pure ideographic approach to data analysis is that it does not permit generalizations to be made across individuals (Spencer & Schöner, 2003), one of the fundamental goals of psychological science. Therefore, in addition to the increase in interest in idiographic statistical models, interest has also increased in methods that allow researchers to identify subsets of individuals in a sample that can be combined because their ideographic results are similar to each other. One method that can be used for this aim is GIMME, which was originally developed for the analysis of neuroscientific data (e.g., fMRI data; Gates, Lane, Varangis, Giovanello, & Guiskewicz, 2017; Gates & Molenaar, 2012). In a series of recent papers (Beltz & Gates, 2017; Gates & Molenaar, 2012; Lane & Gates, 2017; Lane, Gates, Pike, Beltz, & Wright, 2018), it was argued that GIMME can also applied for the analysis of ambulatory assessment data. Given the wide-spread and growing relevance of such data, we believe that this call will lead to increased interest in and use of the GIMME algorithm. Therefore, the goal of the present

paper is to investigate GIMME's performance in the detection of nomothetic relations (i.e., coefficients that are the same across individuals) and idiographic relations (i.e., coefficients that are unique to single individuals) when this algorithm is applied to such data. In the following, we first describe the GIMME algorithm, its statistical background, and its ability to detect the two type of relations. Thereafter, we report the results of three simulation studies. This is followed by a discussion of our main findings and a description of topics for future research.

Statistical Background Behind GIMME

GIMME is based on the unified structural equation model (uSEM; Gates & Molenaar, 2012; Kim, Zhu, Chang, Bentler, & Ernst, 2007; Lane & Gates, 2017), which combines a traditional cross-sectional path model (Bollen, 1989; Mulaik, 2009) and a multivariate autoregressive time-series model (Hamaker et al., 2002, 2003; Lütkepohl, 2005) for the time-series data from a *single* individual. The uSEM is defined by

$$y_t = Ay_t + \Phi y_{t-1} + \epsilon_t \tag{1}$$

Here, y_t is an $m \times 1$ vector of responses to the m variables assessed at time point t for the individual under consideration (e.g., the person's answers to $m = 6$ items on day $t = 4$), and ϵ_t is an $m \times 1$ vector of residuals terms. These residuals are assumed to be white noise errors, that is, they have an expectation of zero and a time-point-independent $m \times m$ covariance matrix Σ_ϵ. The $m \times m$ matrix A constitutes the path model part of the uSEM. It contains the contemporaneous path coefficients (i.e., a_{ij}) of the variables under investigation (e.g., a_{12} indicates whether greater agreement to Item 1 goes along with greater agreement with Item 2 at the same time point). The $m \times m$ matrix Φ describes the lag 1 relationships between the variables. The diagonal elements of Φ, ϕ_{ii}, are the autoregressive (AR) effects between the same variables across subsequent points in time (e.g., ϕ_{11} indicates whether greater agreement with Item 1 at time point t-1 goes along with greater agreement with Item 1 at time point t), whereas the off-diagonal elements, ϕ_{ij}, describe the lagged relationships between different variables across subsequent time points (e.g., ϕ_{12} denotes whether greater agreement with Item 1 at time point t-1 goes along with greater agreement with Item 2 at time point t).

The uSEM and GIMME both assume that the time-series data for the individual under consideration is weakly stationary. That is, the expected values of the variables, their variances, and their lag-1 covariances are each assumed to be constant over time. Let x be a vector that is compromised of the variables at time point t and the lag 1 variables at t-1, that is $x = (y_{t-1}, y_t)$. Building upon the stationarity assumption, the variance-covariance matrix of x implied by Equation 1 (see Kim et al., 2007; Lütkepohl, 2005) is given by

$$\Sigma = \begin{pmatrix} C_1 & I_A \Phi C_1 \\ C_1 \Phi^T I_A^T & I_A(\Phi C_1 \Phi^T + \Sigma_\epsilon)I_A^T \end{pmatrix} \tag{2}$$

where $I_A = (I - A)^{-1}$, and C_1 is the covariance matrix between the m lagged variables y_{t-1} that has to be estimated with the data. Equation 2 shows that the uSEM is a structural equation model (SEM; see Bollen, 1989; Mulaik, 2009) for the data from a single individual, so standard SEM software can be used to estimate the parameters of the model. GIMME is implemented in the R framework and uses the R package `lavaan` for parameter estimation.

Specifically, GIMME first computes a sample covariance matrix S using an individual's time series data contained in x

$$S = \left(\begin{array}{c|c} C_1 & \\ \hline C_{01} & C_0 \end{array} \right) \tag{3}$$

Here, C_1 is the covariance matrix for the lag 1 variables y_{t-1}, C_0 is the estimated covariance matrix of the lag 0 variables y_t, and C_{01} denotes the covariance matrix between the lag 0 and lag 1 variables (i.e., between y_t and y_{t-1}). The model parameters contained in A and Φ are then obtained by using the model-implied covariance matrix Σ, and the empirical covariance matrix S in the maximum likelihood fitting function for SEMs (see again Bollen, 1989).

Of note, GIMME's approach is similar but not identical to the Block-Toeplitz matrix approach, the standard estimation method for univariate and multivariate time-series models, the dynamic factor analysis model, or the white noise factor model using SEM software (Molenaar, 1985; Zhang et al., 2008). The difference between the two approaches is that the Block-Toeplitz approach would use

$$S^{BT} = \left(\begin{array}{c|c} C_0 & \\ \hline C_{01} & C_0 \end{array} \right) \tag{4}$$

as the sample covariance matrix to estimate the uSEM parameters. However, when the assumption of stationarity is met, the two approaches are equivalent (i.e., C_1 converges to C_0 when t approaches infinity) and yield pseudo-maximum-likelihood estimates (Hamaker et al., 2002).

The uSEM described so far models the contemporaneous and temporal associations for a single individual i. Thus, the uSEM is an ideographic statistical model that cannot be used to answer nomothetic research question[1]. For example, it cannot be used to

---

1) It is also possible to estimate the model for the time-series data aggregated across individuals. However, this approach does not allow idiographic relations to be differentiated from nomothetic relations. It also rests on the assumptions that individual processes are homogeneous Molenaar (2004). We will therefore not discuss this approach here.

determine whether there are contemporaneous or lagged associations that occur for all individuals in a sample. To tackle this question, GIMME extends the uSEM by adding the assumption that A and Φ are composed of coefficients that are the same for all individuals and coefficients that are unique for single individuals. The model for a single individual i is thus

$$y_{ti} = (A^i + A^g)y_{ti} + (\Phi^i + \Phi^g)y_{t-1,i} + \epsilon_{ti} \tag{5}$$

where the A and Φ matrices have the same dimensions as above. The superscript g is used for matrices containing paths that are non-zero for all individuals. In articles on GIMME, the coefficients in A^g and Φ^g have been termed group-level effects. The superscript i, by contrast, is used for matrices containing paths that are specific to an individual i; they are termed individual-level effects. Paths that are non-zero in A^i and Φ^i are set to zero in A^g and Φ^g. Thus, the sum of A^i and A^g, for example, is an additive decomposition of person i's matrix A.

To estimate the model parameters, GIMME implements a data-driven approach to obtain the coefficients in A and Φ, but not the coefficients in Σ_ϵ, that occur across all participants or that exist only for single individuals. Specifically, GIMME applies a type of forward-selection procedure that proceeds as follows: First, the algorithm is applied to identify the group level paths. To this end, an *empty* model is estimated for each individual in the sample (i.e., the A and Φ matrices are zero). Then, modification indices (MI; Sörbom, 1989) are computed for each parameter of each single individual's model. The MI of a parameter denotes the amount that the chi-square statistic is expected to decrease if the corresponding parameter is included in the model. Hence, the MI shows which path coefficients one could include to improve model fit. To identify group-level paths, GIMME inspects the MI of a parameter for each person and counts the number of participants for whom the MI of a parameter is significant (after a Bonferroni correction that divides .05 by the size of the sample). Path coefficient that are significant for more than 75% of the individuals in the sample, are included in the model of *each* individual (i.e., the path is added to the to-be-estimated model). If more than one path satisfies this criterion, the path that has the largest average MI across participants is selected. GIMME terminates the search for group-level paths when no path approaches significance for the cutoff of 75%. When this stage is reached, a group-level pruning procedure is conducted by estimating a model in which all of the identified group level paths are included in the models. When the respective path coefficients are significant at a Bonferroni corrected alpha level for 75% of all individuals, the respective paths are retained in the final model.

Second, once the group-level matrices A^g and Φ^g have been found, the search for paths occurring for the single individual begins. To this end, a model is estimated for *each* single individual and the group-level paths are defined as paths that need to be estimated. MIs are again obtained for all remaining path coefficients. For each individual, the coefficients that are significant at $\alpha = .01$ are added to the individual's model. This

procedure is repeated until an excellent fitting model is obtained for each individual, whereby a model is considered to have an excellent fit to the data when two of four standard SEM fit indices indicate good fit. Specifically, GIMME uses the standard cut-off values of RMSEA ≤ 0.08, NNFI ≥ 0.95, CFI ≥ 0.95, and SRMR ≤ 0.08 (see Hu & Bentler, 1999).

In summary, GIMME is a forward-selection procedure that is based on fitting a uSEM for each individual in the sample. MIs are used to decide whether paths are non-zero for most individuals in the sample and to decide whether paths have to be added to the individual models. Two things are worth mentioning because they justify how we conducted our simulations: First, GIMME does not assume that the contemporaneous coefficient matrix A is symmetric. That is, the path coefficient from variable h to variable j can differ from the path coefficient from variable j to variable h. This is important because the MI of two such paths are equal. As a consequence, multiple solutions for the final coefficients (e.g., in A^g and A^i) can occur as a result of the equal MIs of the contemporaneous path coefficients. Research using simulated data examples has shown that the probability of such multiple solutions is nearly removed when the model search procedure begins with an empty model that already includes the AR effects (Lane et al., 2019b) or when the true contemporaneous effects are small (Beltz & Gates, 2016). Second, GIMME can be extended to search for subgroups that are defined as groups of individuals who share contemporaneous and lagged coefficients. To identify subgroups, GIMME uses an extension of the model described in Equation 5 and implements a community detection algorithm called Walktrap (see Gates et al., 2017; Lane & Gates, 2017). We do not describe this algorithm in greater detail, since we will not use the subgroup feature of GIMME in the following simulations.

Before we describe the goal and the contribution of our research, we note that GIMME, and the underlying uSEM approach, differ in important aspects from other models that can be used to analyze experience sampling data such as a multilevel model (MLM; Hedeker & Gibbons, 2006; Mund & Nestler, 2019) or dynamic structural equation models (DSEM; Asparouhov, Hamaker, & Muthén, 2018; Nestler, 2020, 2021). As described above, GIMME starts with the results of a set of individuals and tries to identify those individuals in the set that can be combined because of their similar individual results. The MLM and the DSEM, by contrast, start with assuming a model for all persons in the set (e.g., a linear growth model) and then determines the extent to which the persons differ with respect to the parameters postulated in the model (e.g., the variance of the intercept). The two approaches can, but do not have to produce the same results. For instance, none of the single subjects' parameters would be detected when one uses the DSEM, because the DSEM estimates parameters for all subjects in the sample. Furthermore, and provided that both use the same underlying model (e.g., a growth model for each individual), then the DSEM but not GIMME would produce an estimate concerning how strong a parameter differs between persons (e.g., an intercept variance

estimate). The different modeling philosophies also affect how trends, day effects and so on are handled: While GIMME assumes that the data is weakly stationary (i.e., that there are no trends etc. and if so, they should be removed before the analyses), the DSEM would try to explicitly incorporate these attributes of the data in the model. We refer the reader to Wright et al. (2019) for an illustration of GIMME using a real-data example.

The Present Research: Contribution and General Approach

To the best of our knowledge, there are only two articles that used a Monte Carlo simulation to investigate the performance of GIMME (in all other articles a single simulated or real data example was used; see Beltz & Gates, 2016; Lane & Gates, 2017).[2] Gates et al. (2017), to begin with, examined the influence of the size of the sample (i.e., 25, 75 and 150) and the number of subgroups on GIMME's performance. The model coefficients were set to moderate to high values (i.e., the AR effects ϕ_{ii} were 0.60, the off-diagonal elements ϕ_{ij} were -0.50, and the contemporaneous coefficients were a_{ij} = 0.50) but they did not differ between simulation conditions. The number of time points was large (i.e., T = 250) and it was also kept constant across conditions. The results showed that the number of false discoveries was low and that the sample size had no decisive influence on GIMME's performance. In Lane et al. (2019b), the authors varied the sample size (i.e., 25, 75, and 150), the number of time points (i.e., 30, 60, 90, and 120), and the number of variables k (i.e., 5 and 10). The model coefficients were again kept constant across condition (i.e., ϕ_{ii} = 0.60, ϕ_{ij} = -0.40, and a_{ij} = 0.40). The results showed that the number of false positives was low and that sample size was not important for GIMME's performance. Furthermore, group-level paths were detected at higher rates than individual-level paths. Finally, the larger the number of time points, the better the detection rates of GIMME.

Overall, the two studies provide interesting initial results on the performance of GIMME. Nevertheless, we believe that further simulation research is needed to answer some open questions that remain, because the model parameters were kept constant across the simulation conditions in both articles. For instance, it is unclear whether the ability to detect coefficients depends on the magnitude of the AR effect and whether the number of time points moderates these differences. Similarly, it is also unclear how the magnitude of an AR effect and the magnitude of a single-person coefficient affect the detection of this coefficients and whether the number of time points affect the detection rates. To answer these questions, and to extend earlier research, we conducted three further simulation studies in which we examined the sensitivity and the specificity of

2) Gates and Molenaar (2012) also report the results of a simulation in which they examined the performance of GIMME in three sample size conditions (n = 10, n = 25, and n = 50; see Appendix A of the article). From the description of the simulation in the article, however, it is unclear how the population parameters were chosen and whether more than one replication per sample size condition was examined (i.e., whether a 'true' Monte Carlo simulation was actually conducted). Anyhow, the results of the study are comparable with the results reported in Lane et al. (2019b) in that the size of the sample had no influence on GIMME's performance.

GIMME in the recovery of group-level paths and individual level paths depending on the magnitudes of the respective coefficients. We also tested whether the performance depended on the number of time points and the magnitude of the contemporaneous or the lagged coefficients. In the first study, we considered a rather extreme situation by examining the accuracy of GIMME in the case that *all* nonzero paths exist for *all* individuals. This study served as a kind of benchmark for showing whether GIMME is able to detect group-level paths when no single-individual paths are present. In the second and third studies, we examined GIMME's accuracy when both nonzero group-level and individual-level paths were present.

Study 1

In the first simulation, we randomly drew time-series data for $m = 5$ items and $n = 25$ or $n = 75$ participants. The *same* matrices A and Φ were used for all participants to generate their respective data. Study 1 thus allowed us to determine GIMME's performance when the data-generating model was the same for all individuals.

Method

Simulation Conditions

The 5 × 5 matrix Φ of lagged relationships contained eight parameters

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} & 0 & 0 & 0 \\ 0 & \phi_{22} & 0 & 0 & \phi_{25} \\ 0 & 0 & \phi_{33} & \phi_{34} & 0 \\ 0 & 0 & 0 & \phi_{44} & 0 \\ 0 & 0 & 0 & 0 & \phi_{55} \end{pmatrix} \tag{6}$$

with $\phi_{11} = \cdots = \phi_{55}$ and $\phi_{12} = \phi_{25} = \phi_{34}$. The matrix A was a 5 × 5 matrix that contained three contemporaneous relationship parameters

$$A = \begin{pmatrix} 0 & 0 & 0 & a_{14} & 0 \\ 0 & 0 & a_{23} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{7}$$

with $a_{14} = a_{23} = a_{45}$. A and Φ were the same for all individuals, that is, $A^g = A$ (with $A^i = 0$) and $\Phi^g = \Phi$ (with $\Phi^i = 0$) for all i. Finally, Σ_ϵ was set to be diagonal with variance terms of 1.

We varied the magnitude of the AR effects ϕ_{ii} (i.e., the diagonal elements of Φ, 0.2 vs. 0.4 vs. 0.6), the magnitude of the contemporaneous coefficients a_{ij} (0.2 vs. 0.4 vs.

0.6), and the magnitude of the lagged off-diagonal path coefficients ϕ_{ij} (0.2 vs. 0.4). The selection of these parameters was based on the published simulation data examples for GIMME (see e.g., Beltz & Gates, 2016; Gates & Molenaar, 2012; Lane & Gates, 2017) and the subgroup simulation study by Lane et al. (2019b). Furthermore, these choices reflect estimates that seemed realistic for experience sampling data. Finally, we also varied the length T of the time series with 25, 50, 75, 100, 125, or 250 time points.

Data Generation and Data Analysis

We used R (R Core Team, 2019) to generate 100 replications in each of the 2 × 3 × 3 × 2 × 6 = 216 simulation conditions. In each replication, we used the gimmeSEM function from the GIMME package to recover group-level and individual level paths (Lane et al., 2019a). We used the function with most of its default specifications. Among other things, this means that a unique error variance term is estimated for each observed variable (i.e., Σ_ϵ is a diagonal matrix). The only exception was that we specified that the AR paths should be estimated in each replication to reduce the probability of multiple solutions (see Beltz & Gates, 2016). Finally, in all replications we also used lavaan to estimate the parameters of the model to ensure that potential problems in path recovery were not due to false data generation.

Dependent Measures

In each replication, we saved the number of participants for whom GIMME recovered a specific path in either Φ or A. This allowed us to compute two indices per matrix to evaluate GIMME's accuracy in path recovery. First, for each matrix, we determined the number of true nonzero paths that were recovered. To this end, for each of the three paths ($\phi_{12} = \phi_{25} = \phi_{34}$ or $a_{14} = a_{23} = a_{45}$), we counted the number of individuals for whom the three paths were recovered and added these three numbers together. When n = 25, perfect recovery would be reflected by 75 detected paths (3 paths times 25 participants) for each matrix. Second, we counted the number of falsely recovered paths. We did this by counting the paths that were zero in the population model but that were nevertheless detected by the algorithm. When n = 25, the maximum number was 425 (17 paths times 25 participants) for each of the two matrices.

   We note that the two indices that we calculated correspond to what Gates et al. (2017) and Lane et al. (2019b) call direction recall (i.e., detected true paths) and direction precision (i.e., detected false paths), respectively. The only difference is that we do not divide by the overall number of paths (i.e., 75 and 425). In these articles two other measures, called path recall and path precision, are reported to evaluate the performance of GIMME. These two 'path' indices differ from the two 'direction' indices in that the direction of the paths is not considered in the calculation of the performance measure. For instance, whereas a nonzero ϕ_{21} is not counted as a true detection for direction recall, it would be counted as a true detection in case of path recall. We do not use the 'path'

indices in our studies, because both model matrices are per definition directional in the case of experience sampling data (e.g., the lagged coefficient a_{12} is conceptually different from the lagged coefficient a_{21}). Hence, we think that the two indices are not meaningful for this type of data. Furthermore, when we calculated these indices, we found that their pattern of results is very similar to the result pattern of the 'direction' indices. Thus, the information gain from reporting results about the 'path' indices is small.

Results and Discussion

Preliminary analyses showed that the data followed the intended population models (see Supplementary Materials). We first present the results concerning the number of true recoveries, followed by the results concerning the number of false recoveries. As the results for the $n = 75$ sample size condition were largely identical to the results for the $n = 25$ sample size condition, we do report the results of the latter condition in the main text only. For the $n = 75$ sample size condition, the reader is referred to Appendix A.

Detection of Nonzero Paths

We examined GIMME's performance in identifying true nonzero paths separately for the off-diagonal lagged coefficients ϕ_{ij} and the contemporaneous coefficients a_{ij}. With regard to the lagged coefficients, we first computed an Analysis of Variance (ANOVA) to identify the relevant factors that determined the performance with regard to this type of coefficient. The ANOVA included main effects and interactions for time-series length T, the magnitude of ϕ_{ii}, the magnitude of a_{ij}, and the magnitude of ϕ_{ij}. The results showed that main effects were present for series length T ($\eta_p^2 = 0.85$) and magnitude of ϕ_{ij} ($\eta_p^2 = 0.94$). These main effects were qualified by a significant two-way interaction between series length T and magnitude of ϕ_{ij} ($\eta_p^2 = 0.50$) and a significant three-way interaction of length T, magnitude of ϕ_{ii}, and magnitude of ϕ_{ij} ($\eta_p^2 = 0.13$). For all other effects, η_p^2 was smaller than 0.10.

The left-hand side of Table 1 displays the number of recovered true nonzero paths depending on the number of time points T, the magnitude of ϕ_{ii}, and the magnitude of ϕ_{ij}. Ideally, the number within each cell would be 75. As can be seen in Table 1, the observed number was closer to 75 the longer the time series and the larger the magnitude of ϕ_{ij} (i.e., the magnitude of the to-be recovered coefficient). When the coefficient ϕ_{ij} was 0.20, the number of recovered true path coefficients was higher the larger the AR effect ϕ_{ii}. This influence of the magnitude of ϕ_{ii} was stronger the longer the time series. Finally, when ϕ_{ij} was 0.40, the positive influence of the magnitude of ϕ_{ii} was smaller (or not evident) the longer the time series.

Table 1

GIMME's Performance in Detecting Off-Diagonal Lagged Coefficients ϕ_{ij} or Contemporaneous Coefficients a_{ij} Depending on Time-Series Length T, the Magnitude of the AR Coefficient ϕ_{ii}, and the Magnitude of the Off-Diagonal Lagged Coefficient ϕ_{ij} or the Magnitude of the Contemporaneous Coefficient a_{ij}

		ϕ_{ij}		a_{ij}		
T	ϕ_{ii}	**0.20**	**0.40**	**0.20**	**0.40**	**0.60**
25	0.20	14.09	36.55	9.96	23.30	36.24
25	0.40	15.65	38.73	11.61	28.73	46.70
25	0.60	17.99	42.24	14.01	35.02	54.83
50	0.20	17.63	52.67	12.54	35.80	64.11
50	0.40	18.75	54.57	14.19	44.39	72.97
50	0.60	20.52	57.62	18.07	52.30	74.73
75	0.20	22.34	64.37	16.14	46.92	70.38
75	0.40	22.82	65.82	18.96	61.93	74.03
75	0.60	23.14	67.11	21.50	71.44	75.00
100	0.20	26.90	71.97	19.93	63.81	73.95
100	0.40	26.61	72.91	22.43	73.48	74.71
100	0.60	25.42	74.57	24.32	74.90	75.00
125	0.20	30.63	74.59	23.00	70.10	75.00
125	0.40	29.33	74.83	26.59	74.83	75.00
125	0.60	32.70	75.00	26.18	75.00	75.00
250	0.20	42.00	75.00	31.41	74.56	75.00
250	0.40	49.53	75.00	38.15	75.00	75.00
250	0.60	62.90	75.00	60.02	75.00	75.00

Note. 75 recoveries were possible within each simulation condition.

A very similar result pattern emerged with regard to GIMME's ability to recover the true nonzero contemporaneous coefficients a_{ij}. The ANOVA yielded main effects of time series length T ($\eta_p^2 = 0.83$), magnitude of a_{ij} ($\eta_p^2 = 0.94$), and magnitude of ϕ_{ii} ($\eta_p^2 = 0.32$). These main effects were qualified by a $T \times a_{ij}$ interaction ($\eta_p^2 = 0.56$) and a three-way interaction of length T, magnitude of a_{ij}, and magnitude of ϕ_{ii} ($\eta_p^2 = 0.25$). The effect sizes for the other main effects or interactions were below $\eta_p^2 = 0.08$. The right-hand side of Table 1 displays the average number of detected paths depending on the factors appearing in the three-way interaction. Again, the number of detected true nonzero paths was higher the longer the time series and the larger the magnitude of a_{ij}. When a_{ij} was 0.20, the number of recoveries was greater the larger the AR effect ϕ_{ii}. However, the effect of the $T \times$ magnitude of ϕ_{ii} interaction was smaller (and even not evident) when a_{ij} was 0.40 or 0.60.

False Discoveries

For the lagged coefficient matrix $\mathbf{\Phi}$, we found that the number of false recoveries across simulation conditions was very small ($M = 12.91$, $SD = 14.04$; 425 false path counts are possible within each simulation condition). The ANOVA yielded main effects of time series length T ($\eta_p^2 = 0.91$), magnitude of a_{ij} ($\eta_p^2 = 0.25$), and magnitude of ϕ_{ii} ($\eta_p^2 = 0.14$). All other main effects and interactions had effect sizes smaller than 0.10. A closer examination of the main effects showed that the longer the series, the lower the number of false positive recoveries: $M_{25} = 39.4$, $M_{50} = 16.9$, $M_{75} = 9.6$, $M_{100} = 6.4$, $M_{125} = 4.5$, $M_{250} = 0.5$. Furthermore, a larger AR effect ϕ_{ii} ($M_{0.2} = 14.7$, $M_{0.4} = 13.1$, $M_{0.6} = 10.8$) and a larger contemporaneous coefficient a_{ij} ($M_{0.2} = 15.8$, $M_{0.4} = 12.8$, $M_{0.6} = 10.1$) went along with a smaller number of false lagged coefficient recoveries.

With regard to the contemporaneous coefficients in $\mathbf{A}$, we found that the average proportion of false recoveries across simulation conditions was small ($M = 13.7$, $SD = 13.9$, again, 425 false positives were possible within each replication). The ANOVA showed main effects of time-series length T ($\eta_p^2 = 0.83$), magnitude of a_{ij} ($\eta_p^2 = 0.20$), and magnitude of ϕ_{ii} ($\eta_p^2 = 0.29$). Furthermore, the $T \times a_{ij}$ interaction ($\eta_p^2 = 0.23$) and the $T \times a_{ij} \times \phi_{ii}$ interaction were also sizable ($\eta_p^2 = 0.15$). As can be seen in Table 2, the three-way interaction was nearly exclusive to conditions were T was 25 or 50. Here, the number of false positives was lower when ϕ_{ii} was larger (the only exception was the condition in which $T = 25$ and $a_{ij} = 0.20$). About 50% of the false positives in a simulation condition were due to the false recovery of the opponent path to the true path, for instance, when a true path from Variable 4 to Variable 1 was falsely detected as a path from Variable 1 to Variable 4.

Altogether, the first simulation provided a number of interesting findings. First, the simulation showed that the recovery of true nonzero group-level paths strongly depended on the magnitude of the to-be-selected coefficient: The smaller the coefficient, the smaller the number of correct discoveries. Time-series length and the magnitude of the AR coefficient had a positive influence on recovery, although their effect did not seem to be "omnipotent": When the to-be-recovered coefficient was small, the number of recoveries was small even in the "perfect" conditions. Second, the number of false positive recoveries was small. Interestingly, when false positive recoveries occurred for the contemporaneous paths, a large fraction of them were due to the selection of the opposite compared with the true nonzero paths. In sum, `GIMME` can detect true nonzero group-level paths, but its performance depends on the number of time points, the magnitudes of the respective coefficients, and the magnitude of the AR coefficient.

Study 2

In the first simulation, the population model was the same for all individuals. In the second simulation, we took a more realistic approach by simulating data from a population

Table 2

Number of False Positive Contemporaneous Coefficients a_{ij} Depending on Time-Series Length T, the Magnitude of the AR Coefficient ϕ_{ii}, and the Magnitude of the Contemporaneous Coefficient a_{ij}

		a_{ij}		
T	ϕ_{ii}	**0.20**	**0.40**	**0.60**
25	0.20	30.68	39.88	45.55
25	0.40	31.97	38.95	38.39
25	0.60	34.12	36.23	31.68
50	0.20	21.10	30.88	20.37
50	0.40	20.22	24.09	11.14
50	0.60	17.39	15.18	5.51
75	0.20	19.32	25.61	11.17
75	0.40	16.13	12.81	5.66
75	0.60	10.96	4.22	1.90
100	0.20	18.25	15.05	4.81
100	0.40	13.55	5.09	2.77
100	0.60	7.71	2.02	0.96
125	0.20	17.89	9.07	2.37
125	0.40	12.20	2.93	1.25
125	0.60	5.66	1.04	0.20
250	0.20	12.08	0.98	0.16
250	0.40	4.24	0.17	0.02
250	0.60	0.06	0.00	0.00

Note. 425 false discoveries were possible within each condition.

model in which some coefficients were the same for all individuals, and some coefficients occurred for only a few individual people. Thus, the simulation allowed us to replicate the findings from the first simulation concerning the group-level coefficients. It also allowed us to examine GIMME's ability to detect individual-level effects and to determine whether the detection of these coefficients depends on the magnitude of the group-level coefficients.

Method

Simulation Conditions

Similar to Study 1, we simulated time-series data for $n = 25$ participants and $m = 5$ items. The $n = 75$ participant conditions were skipped, because of the highly similar results compared to the $n = 25$ participant condition in Study 1. The matrices for the group-level path were the same as the matrices in simulation Study 1 (Equation 6 and

Equation 7). For four of the 25 participants, the matrix of lagged effects Φ, but not the matrix of contemporaneous effects, additionally included one further single coefficient. For instance, for one of the four individuals, Φ^i was

$$\Phi^i = \begin{pmatrix} 0 & 0 & 0 & s_1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{8}$$

so that the matrix of lagged effects Φ for this person was

$$\Phi = \Phi^i + \Phi^g = \begin{pmatrix} \phi_{11} & \phi_{12} & 0 & s_1 & 0 \\ 0 & \phi_{22} & 0 & 0 & \phi_{25} \\ 0 & 0 & \phi_{33} & \phi_{34} & 0 \\ 0 & 0 & 0 & \phi_{44} & 0 \\ 0 & 0 & 0 & 0 & \phi_{55} \end{pmatrix}. \tag{9}$$

For the other three individuals, the respective matrices contained the individual-level paths, s_2 to s_4, in the following entries

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} & 0 & 0 & 0 \\ 0 & \phi_{22} & s_2 & 0 & \phi_{25} \\ 0 & 0 & \phi_{33} & \phi_{34} & s_3 \\ 0 & 0 & 0 & \phi_{44} & s_4 \\ 0 & 0 & 0 & 0 & \phi_{55} \end{pmatrix} \tag{10}$$

where $s_1 = s_2 = s_3 = s_4$. Furthermore, for another four individuals of the 25 subjects, the matrix of contemporaneous effects A additionally included another single coefficient. The positions of the individual-level paths, s_5 to s_8, with $s_5 = s_6 = s_7 = s_8$, were

$$A = \begin{pmatrix} 0 & 0 & 0 & a_{14} & s_5 \\ 0 & 0 & a_{23} & s_6 & s_7 \\ 0 & 0 & 0 & 0 & s_8 \\ 0 & 0 & 0 & 0 & a_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{11}$$

Altogether, thus, six coefficients existed for all 25 participants (excluding the AR effects; see Study 1). For eight participants, the matrices additionally included another nonzero individual-level coefficient. Finally, for all participants Σ_ϵ was again set to be a diagonal matrix with variance terms of 1.

Simulation Conditions, Data Generation, and Data Analysis

As in Study 1, we varied the length of the time series. Given that in Study 1, recovery rates were very low in the 25 time point condition and given that the results in the 125 time point condition were very similar to the 100 time point condition, we decided to not include the 25 and the 125 time point conditions in Study 2. However, we added a $T =$ 500 condition to ensure that we included a condition in which perfect recovery would occur. Thus, we simulated five time-series length conditions: 50 vs. 75 vs. 100 vs. 250 vs. 500 time points. As in Study 1, we varied the magnitude of the AR effects ϕ_{ii} (0.2 vs. 0.4 vs. 0.6), the magnitude of the contemporaneous group-level coefficients a_{ij}^g (0.2 vs. 0.4 vs. 0.6), and the magnitude of the lagged off-diagonal group-level coefficients ϕ_{ij}^g (0.2 vs. 0.4). We also varied the magnitude of the individual-level coefficients s (0.2 vs. 0.4). As in Study 1, we generated 100 replications in each of the $5 \times 3 \times 3 \times 2 \times 2 = 180$ simulation conditions. Finally, `gimmeSEM` was used to recover group-level and individual-level paths (using the same specification as described in Study 1), and `lavaan` was used to estimate the parameters of the model (for the results concerning the parameter estimates, see Supplementary Materials).

Dependent Measures

In each replication, we saved the paths that were recovered by `GIMME` for each participant. This allowed us to evaluate `GIMME`'s accuracy in recovering the true nonzero group-level and the true nonzero individual-level path coefficients. The performance measures for the group-level paths were computed as described in Study 1. For the individual-level paths we used essentially the same approach by computing the sum of the individuals for which the true nonzero single path was recovered (four individuals for the lagged coefficients and four individuals for the contemporaneous coefficients).

Results and Discussion

Preliminary analyses showed that the data followed the population models (see Supplementary Materials). Again, we first describe the simulation results for the number of true recovered paths. Thereafter, we present the results concerning the false discoveries.

Detection of Nonzero Paths

We first examined `GIMME`' ability to identify the true nonzero group-level paths. An ANOVA with the five simulation design factors showed that neither the magnitude of the individual-level paths s nor any of the interactions involving s were sizable ($\eta_p^2 < 0.05$). Further analyses showed that essentially the same results as in Study 1 emerged for both types of group-level coefficients. Therefore, we focus here on the description of the

results for the individual-level coefficients and refer the reader to Table B1 in Appendix B for the results concerning the group-level coefficients.

We examined the GIMME's ability to detect true nonzero individual-level paths separately for the off-diagonal lagged path coefficients and the contemporaneous coefficients. For the individual-level *off-diagonal lagged* coefficients (s_1 to s_4), the ANOVA yielded main effects of series length T ($\eta_p^2 = 0.18$), magnitude of ϕ_{ii} ($\eta_p^2 = 0.15$), magnitude of a_{ij}^g ($\eta_p^2 = 0.12$), magnitude of ϕ_{ij}^g ($\eta_p^2 = 0.15$), and magnitude of s ($\eta_p^2 = 0.53$). There was also a $T \times \phi_{ii}$ interaction ($\eta_p^2 = 0.18$). All other main effects or interactions had η_p^2 values smaller than 0.10. For the individual-level *contemporaneous* coefficients (s_5 to s_8), we obtained main effects of length T ($\eta_p^2 = 0.25$), magnitude of ϕ_{ii} ($\eta_p^2 = 0.23$), magnitude of a_{ij}^g ($\eta_p^2 = 0.35$), and magnitude of s ($\eta_p^2 = 0.49$). There was also a $T \times \phi_{ii}$ interaction ($\eta_p^2 = 0.11$). All other main effects or interactions had η_p^2 smaller than 0.10.

Table 3 shows the number of recoveries for the two types of individual-level coefficients. The maximum number of recoveries within a simulation condition was 4. As can be seen, when the magnitude of the individual-level coefficient s was 0.20, recovery rates were very small (between 0% and 50%) irrespective of time-series length and magnitude of the AR effect. Furthermore, when s was 0.40, we found that for *smaller* ϕ_{ii} and longer T, the higher the number of detected true nonzero individual-level paths. This pattern was less pronounced and even reversed when a_{ij}^g was smaller. In fact, path recovery was best when *all* group-level coefficients were 0.20, and the magnitude of s was large (see Table B2 in the Appendix B). That is, the recovery of true nonzero individual-level coefficients was (negatively) influenced by the magnitude of the group-level relations.

False Discoveries

For the lagged coefficient matrix $\mathbf{\Phi}$, similar to Study 1, we found that the number of false recoveries across simulation conditions was small ($M = 8.14$, $SD = 10.11$). The ANOVA showed main effects for series length T ($\eta_p^2 = 0.66$), magnitude of ϕ_{ii} ($\eta_p^2 = 0.33$), and magnitude of a_{ij}^g ($\eta_p^2 = 0.26$). Furthermore, the $T \times \phi_{ii}$ interaction ($\eta_p^2 = 0.14$), the $T \times a_{ij}^g$ interaction ($\eta_p^2 = 0.17$), and the three-way interaction of T, ϕ_{ii}, and a_{ij}^g were also sizable ($\eta_p^2 = 0.14$). Subsequent analyses showed that most false recoveries occurred when the time-series length was 50, ϕ_{ii} was 0.20, and a_{ij}^g was also 0.20. However, even in these conditions the number of false discoveries was rather small ($M = 21.4$). For the contemporaneous coefficient matrix $\mathbf{A}$, the ANOVA yielded main effects of series length T ($\eta_p^2 = 0.80$), magnitude of ϕ_{ii} ($\eta_p^2 = 0.27$), magnitude of a_{ij}^g ($\eta_p^2 = 0.26$), and magnitude of ϕ_{ij}^g ($\eta_p^2 = 0.13$). Also, the $T \times \phi_{ii}$ interaction ($\eta_p^2 = 0.17$), and the $T \times a_{ij}^g$ interaction were sizable ($\eta_p^2 = 0.18$). Again, subsequent analyses showed that most false recoveries occurred when T was 50, ϕ_{ii} was 0.20, and ϕ_{ij}^g was 0.20, but the average number of false recoveries was small even in these conditions ($M = 15.4$). Finally, about half the false positive recoveries were due to a false recovery of the opponent path to the true path.

PsychOpen GOLD

Table 3

GIMME's Performance in Detecting Individual-Level Coefficients s Depending on the Magnitude of the Individual-Level Coefficient, Time-Series Length T, the Magnitude of the AR Coefficient ϕ_{ii}, and the Magnitude of the Contemporaneous Group-Level Coefficient a_{ij}^g

			Lagged s			Contemporaneous s		
s	T	ϕ_{ii}	$a_{ij}^g = 0.2$	$a_{ij}^g = 0.4$	$a_{ij}^g = 0.6$	$a_{ij}^g = 0.2$	$a_{ij}^g = 0.4$	$a_{ij}^g = 0.6$
0.20	50	0.20	0.92	0.68	0.60	1.22	0.93	0.83
0.20	50	0.40	0.93	0.88	0.76	1.25	1.00	0.58
0.20	50	0.60	1.09	1.00	0.69	1.43	1.00	0.30
0.20	75	0.20	0.97	0.74	0.78	1.27	1.04	0.80
0.20	75	0.40	1.18	0.82	0.81	1.61	0.90	0.57
0.20	75	0.60	1.15	0.75	0.62	1.53	0.51	0.25
0.20	100	0.20	1.11	0.88	0.88	1.49	1.02	0.72
0.20	100	0.40	1.31	0.90	0.75	1.59	0.74	0.50
0.20	100	0.60	1.19	0.68	0.48	1.56	0.37	0.16
0.20	250	0.20	1.79	0.94	0.70	2.05	0.95	0.26
0.20	250	0.40	1.56	0.54	0.28	1.63	0.22	0.04
0.20	250	0.60	0.16	0.04	0.01	0.17	0.00	0.00
0.20	500	0.20	0.50	0.12	0.04	0.71	0.12	0.00
0.20	500	0.40	0.09	0.00	0.00	0.07	0.00	0.00
0.20	500	0.60	0.00	0.00	0.00	0.00	0.00	0.00
0.40	50	0.20	2.23	2.02	2.04	2.80	2.37	2.19
0.40	50	0.40	2.50	2.39	2.30	2.96	2.60	1.97
0.40	50	0.60	2.70	2.59	1.92	2.92	2.09	1.33
0.40	75	0.20	2.53	2.64	2.74	3.40	2.85	2.52
0.40	75	0.40	3.09	2.74	2.52	3.42	2.83	2.19
0.40	75	0.60	3.14	2.22	1.87	3.36	1.80	1.01
0.40	100	0.20	2.91	2.87	2.83	3.56	3.07	2.62
0.40	100	0.40	3.33	2.94	2.39	3.62	2.56	1.83
0.40	100	0.60	3.23	1.92	1.56	3.46	1.35	0.80
0.40	250	0.20	3.46	3.25	2.90	3.94	3.35	2.04
0.40	250	0.40	3.69	2.29	1.56	3.87	1.66	0.49
0.40	250	0.60	1.33	0.34	0.21	1.14	0.05	0.00
0.40	500	0.20	3.75	3.53	2.22	3.98	3.15	0.56
0.40	500	0.40	3.02	1.11	0.40	2.83	0.42	0.00
0.40	500	0.60	0.26	0.00	0.00	0.14	0.00	0.00

Note. Four recoveries were possible within each simulation condition.

To summarize, the second simulation replicated the results of the first simulation for the group-level paths: Again, the length of the time series and the magnitude of the

PsychOpen GOLD

AR coefficient had positive influences, and the recovery proportions were low when the two-be-selected coefficient was small. Importantly, the detection of true nonzero group-level paths was not affected by the magnitude of the true nonzero individual-level coefficients. However, for these coefficients, we found that the best recovery was achieved when the individual-level coefficient was high and the group-level coefficients were *low*. Finally, the number of false positive recoveries was rather low with most of them occurring when the length of the time series was small.

Study 3

One goal of Study 3 was to replicate our findings concerning the individual-level coefficients obtained in the second simulation. Furthermore, the second simulation used a population model that contained group-level paths and individual-level paths. However, it has been our experience—and theory also suggests—that when `GIMME` is applied to real data, usually the paths that are detected occur for all individuals in the sample, for single people only, and for a subset of individuals (e.g., six people out of 25). Hence, a second goal of Study 3 was to investigate `GIMME`'s performance for this type of situation.

Method

Simulation Conditions

Again, we simulated time-series data for $n = 25$ participants and $m = 5$ items. The matrices for the group-level paths were the same as the matrices in the first and the second simulations. For three of the 25 subjects, the individual matrix of lagged effects, Φ, additionally included one further single coefficient. For another three of the 25 persons, the individual matrix of contemporaneous effects A, included one further single coefficient. The positions of the individual-level paths in the matrices were:

$$\Phi^i = \begin{pmatrix} \phi_{11} & \phi_{12} & 0 & s_1 & 0 \\ 0 & \phi_{22} & s_2 & 0 & \phi_{25} \\ 0 & 0 & \phi_{33} & \phi_{34} & s_3 \\ 0 & 0 & 0 & \phi_{44} & 0 \\ 0 & 0 & 0 & 0 & \phi_{55} \end{pmatrix} \text{ and } A^i = \begin{pmatrix} 0 & 0 & 0 & a_{14} & s_4 \\ 0 & 0 & a_{23} & s_5 & 0 \\ 0 & 0 & 0 & s_6 & 0 \\ 0 & 0 & 0 & 0 & a_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (12)$$

with $s_1 = s_2 = \cdots = s_6$. Finally, for a subset of six people out of the 25, Φ contained two path coefficients, p_1 and p_2, and for a subset of another six people A contained two path coefficients, p_3 and p_4. The positions in the respective matrices were

$$\Phi^i = \begin{pmatrix} \phi_{11} & \phi_{12} & 0 & 0 & 0 \\ 0 & \phi_{22} & 0 & p_1 & \phi_{25} \\ 0 & 0 & \phi_{33} & \phi_{34} & 0 \\ 0 & 0 & 0 & \phi_{44} & p_2 \\ 0 & 0 & 0 & 0 & \phi_{55} \end{pmatrix} \text{ and } A^i = \begin{pmatrix} 0 & 0 & p_3 & a_{14} & 0 \\ 0 & 0 & a_{23} & 0 & p_4 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{13}$$

with $p_1 = p_2 = p_3 = p_4$. Thus, altogether, there were six path coefficients that existed for all 25 participants. For six participants, the matrices contained one additional single coefficient (either in A or Φ). For 12 participants (of the 25), the matrices contained the same two path coefficients in addition to the group-level paths. For all participants, finally, Σ_ϵ was defined to diagonal with variance terms of 1.

Simulation Conditions, Data Generation, and Data Analysis

We varied the length of the time series T (50 vs. 75 vs. 100 vs. 250 vs. 500), the magnitude of the AR effects ϕ_{ii} (0.2 vs. 0.4 vs. 0.6), the magnitude of the contemporaneous group-level coefficients a_{ij}^g (0.2 vs. 0.4 vs. 0.6), and the magnitude of the lagged off-diagonal group-level path coefficients ϕ_{ij}^g (0.2 vs. 0.4). We also varied the magnitude of the individual-level coefficients s (0.2 vs. 0.4), and the magnitude of the subset coefficients p (0.2 vs. 0.4). In each of the $5 \times 3 \times 3 \times 2 \times 2 \times 2 = 360$ simulation condition, we generated 100 replications, we used `gimmeSEM` for path detection, and we used `lavaan` to estimate the model parameters (for the results see Supplementary Materials).

Dependent Measures

In each replication, we saved the paths that were recovered by GIMME for each participant. This allowed us to compute the correct path recovery measure and the false discovery measure as described in Studies 1 and 2, respectively.

Results and Discussion

Preliminary analyses showed that the data followed the population models (see Supplementary Materials). We begin by describing the simulation results for the recovery of the true nonzero path and then go on to present the results concerning the false discoveries.

Detection of Nonzero Paths

With regard to the true nonzero group-level paths, an ANOVA involving all six simulation design factors showed that the detection of nonzero paths was not affected by the magnitude of s, the magnitude of p, or by any interaction involving the two factors. Similarly, the detection of true nonzero individual-level coefficients was unrelated to the magnitude of p or any interaction involving p. Further analyses showed that the results for the detection of true nonzero group-level and true nonzero individual-level

coefficients were largely identical to the results reported in Studies 1 and 2. Therefore, we focus on `GIMME`'s ability to detect true nonzero subset paths and refer the reader to Appendix C for more detailed results concerning the group-level and individual-level paths.

For the subset paths, the data were simulated in such a way that for six individuals (out of 25), the same two path coefficients appeared in $\mathbf{\Phi}$. For another subset of six individuals, the same two path coefficients appeared in $\mathbf{A}$. Thus, the maximum possible number of nonzero subset paths within each simulation condition was 12 for each of the two matrices. For the *lagged* subset coefficients, the ANOVA yielded main effects of time series length T ($\eta_p^2 = 0.49$), magnitude of ϕ_{ii} ($\eta_p^2 = 0.48$), magnitude of a_{ij}^g ($\eta_p^2 = 0.52$), magnitude of ϕ_{ij}^g ($\eta_p^2 = 0.33$), and magnitude of p ($\eta_p^2 = 0.71$). Furthermore, the $T \times \phi_{ii}$ interaction ($\eta_p^2 = 0.23$), the $\phi_{ii} \times p$ interaction ($\eta_p^2 = 0.18$), and the interaction of $T \times \phi_{ii} \times a_{ij}^g$ was also sizable ($\eta_p^2 = 0.16$). For the subset *contemporaneous* coefficients, we found main effects of ϕ_{ii} ($\eta_p^2 = 0.22$), a_{ij}^g ($\eta_p^2 = 0.48$), and p ($\eta_p^2 = 0.75$). Also, a $T \times \phi_{ii} \times p$ interaction emerged ($\eta_p^2 = 0.10$).

As can be seen in Table 4, the number of correct detections was low when the subset coefficient p was small. When p was large, the results were very similar to the results obtained for the individual-level coefficients (in Studies 2 and 3): There were a larger number of recoveries when the magnitude of the group-level path coefficients was *smaller*. In fact, the largest number of path recoveries occurred when all group-level coefficients were 0.20. In this case (see Table C3 in Appendix C), and independent of the magnitude of the individual-level coefficients s, the number of path recoveries was higher the larger the subset coefficient p and the longer the time series T.

False Discoveries

For the lagged coefficient matrix $\mathbf{\Phi}$, the number of false recoveries across simulation conditions was again small ($M = 5.62$, $SD = 6.81$). The ANOVA yielded main effects for series length T ($\eta_p^2 = 0.80$), magnitude of ϕ_{ii} ($\eta_p^2 = 0.30$), magnitude of a_{ij}^g ($\eta_p^2 = 0.26$), and magnitude of ϕ_{ij}^g ($\eta_p^2 = 0.14$). Furthermore, the $T \times \phi_{ii}$ interaction ($\eta_p^2 = 0.20$) and the $T \times a_{ij}^g$ interaction were also sizable ($\eta_p^2 = 0.18$). As in Studies 1 and 2, most false recoveries occurred when T was 50 and when ϕ_{ii} as well as a_{ij}^g were small. However, even in these conditions, the number of false recoveries was small ($M = 17.4$). A similar result pattern emerged for the contemporaneous coefficient matrix $\mathbf{A}$. The ANOVA yielded main effects for series length T ($\eta_p^2 = 0.64$), magnitude of ϕ_{ii} ($\eta_p^2 = 0.35$), magnitude of a_{ij}^g ($\eta_p^2 = 0.21$), and magnitude of ϕ_{ij}^g ($\eta_p^2 = 0.10$). Also, the $T \times \phi_{ii}$ interaction ($\eta_p^2 = 0.14$) and the $T \times a_{ij}^g$ interaction were sizable ($\eta_p^2 = 0.14$). Again, most false recoveries occurred when time-series length was 50 ($M = 21.4$), and about half of them were due to the false detection of the opponent path rather than the true path.

Replicating Studies 1 and 2, Study 3 showed that the number of true recoveries of group-level paths was higher for larger respective group-level coefficients, larger AR ef-

Table 4

GIMME's Performance in Detecting Subset Coefficients p Depending on the Magnitude of the Subset Coefficient, Time-Series Length T, the Magnitude of the AR Coefficient ϕ_{ii}, and the Magnitude of the Contemporaneous Group-Level Coefficient a_{ij}^g

			Lagged *p*			Contemporaneous *p*		
p	*T*	ϕ_{ii}	$a_{ij}^g = 0.2$	$a_{ij}^g = 0.4$	$a_{ij}^g = 0.6$	$a_{ij}^g = 0.2$	$a_{ij}^g = 0.4$	$a_{ij}^g = 0.6$
0.20	50	0.20	3.84	3.77	2.88	2.14	1.95	1.43
0.20	50	0.40	4.43	3.91	2.46	2.50	2.19	1.46
0.20	50	0.60	4.67	3.52	1.69	3.08	2.02	1.01
0.20	75	0.20	4.98	3.81	3.04	2.71	2.23	1.71
0.20	75	0.40	5.02	3.13	2.51	3.27	2.15	1.51
0.20	75	0.60	4.96	1.82	1.37	3.71	1.48	0.68
0.20	100	0.20	5.50	3.57	2.80	3.48	2.46	1.81
0.20	100	0.40	5.59	2.79	2.09	4.07	2.41	1.37
0.20	100	0.60	5.18	1.55	1.13	4.24	1.16	0.42
0.20	250	0.20	6.78	3.13	1.93	5.46	2.71	1.51
0.20	250	0.40	4.41	1.61	1.29	4.79	1.50	0.58
0.20	250	0.60	0.56	0.15	0.04	0.66	0.06	0.00
0.20	500	0.20	4.19	1.05	0.01	4.39	1.54	0.54
0.20	500	0.40	1.02	0.01	0.00	1.82	0.13	0.05
0.20	500	0.60	0.00	0.00	0.00	0.01	0.00	0.00
0.40	50	0.20	9.28	8.78	7.25	6.05	5.37	3.88
0.40	50	0.40	9.42	8.64	6.68	7.34	6.47	5.06
0.40	50	0.60	9.32	7.16	4.83	8.74	6.83	4.70
0.40	75	0.20	10.55	9.27	8.23	7.30	6.44	5.11
0.40	75	0.40	10.55	8.31	6.79	9.28	7.83	5.97
0.40	75	0.60	9.52	5.33	4.33	10.10	6.54	4.14
0.40	100	0.20	11.12	9.74	8.43	8.61	7.19	5.67
0.40	100	0.40	10.91	7.54	5.99	10.41	7.89	6.12
0.40	100	0.60	9.79	4.29	3.29	10.84	5.89	3.14
0.40	250	0.20	11.84	9.32	6.51	10.54	8.27	6.19
0.40	250	0.40	10.09	4.89	3.14	11.13	8.34	5.86
0.40	250	0.60	3.09	1.54	1.35	6.89	2.50	0.52
0.40	500	0.20	11.46	8.05	5.17	11.04	9.19	6.15
0.40	500	0.40	7.09	3.87	1.16	9.62	7.38	5.47
0.40	500	0.60	1.58	0.05	0.00	5.88	1.41	0.04

Note. 12 recoveries are possible within each simulation condition.

fects, and longer time series. When the to-be-selected coefficient was small, the number of true recoveries was low. Importantly, the detection of nonzero group-level paths did

not depend on the magnitude of the individual level or on the magnitude of the subset coefficients. For the individual-level and the subset coefficients, our results showed that the highest number of true recoveries occurred when the respective coefficient was large and the group-level coefficients were small. Finally, the number of false positives was negligible and most of them occurred when the number of measurement points was small.

General Discussion

The present simulation studies were conducted to investigate GIMME's performance in the recovery of group-level paths, individual-level paths, and paths that exist for a subset of individuals from the whole sample. We examined how the length of the time series and the magnitude of the different path coefficients affected path recovery performance. Finally, we also investigated GIMME's bias in selecting non-existent paths and factors that influence this bias.

A number of interesting findings emerged across the three studies. First, and replicating the findings of Gates et al. (2017) and Lane et al. (2019b), we found that the number of false positive recoveries was very low in all three studies. Most false positive paths were detected when the number of measurements was small ($T \leq 50$). These results suggest that applied researchers should assess a moderate number of measurement points (i.e., $T > 50$) to circumvent false recoveries. Furthermore, in the case of the contemporaneous paths, a large proportion of false positive recoveries were due to the false selection of the opposite paths. That is, rather than selecting the path from variable i to variable j, for example, the algorithm sometimes selected the path from variable j to variable i. We think that a potential remedy for these false recoveries would be *not* to allow the contemporaneous coefficient matrix, A, to be asymmetric during estimation. However, at present, it is not possible to apply GIMME with this specification (at least to our knowledge).

Second, with regard to GIMME's performance in recovering true nonzero group-level paths, we consistently found that its performance was better the longer the time series (Gates et al., 2017 and Lane et al., 2019b for similar results) and the greater the magnitude of the to-be selected of contemporaneous or off-diagonal lagged coefficients. Importantly, GIMME could only detect small to moderate coefficients (i.e., with values of 0.20 to 0.40) with acceptable precision, when the number of time points was sufficiently large. For example, when the contemporaneous coefficients is of moderate size, then one needs more than 100 points. This recommendation stays in contrast to the recommendations of Lane et al., 2019b who suggest that 60 time points are sufficient for a good performance of GIMME. Of note, we belief that this finding (e.g., higher detection rates with a higher magnitude of the to-be selected coefficient) is a direct consequence of the adequacy of the parameter estimates: A larger sample is needed to estimate small coefficients

PsychOpen GOLD

with a small bias. Thus, the longer the time series (i.e., the larger the sample), the smaller the bias in the underlying Maximum Likelihood estimator. This in turn increases the probability that the nonzero path will be discovered by the modification index approach implemented in GIMME. Consequently, when applied researchers assume that their group-level coefficients are small, they should assess a very large number of time points (see Table 1 for an orientation).

Third, earlier research found that group-level paths are detected at higher rates compared to individual-level paths. The results of Study 2 and 3 qualify these results considerably, as they show that the recovery of true nonzero individual-level-and subset paths is stronger, the *smaller* the magnitude of the group-level paths. In fact, when the magnitude of the group-level paths was small, the path recovery performance for the two types of paths was good. However, once one of the group-level paths was moderate in size, the number of detected paths was not sufficient.[3] This finding cannot be explained by parameter estimation bias because the results also occurred when the true nonzero individual-level or subset path coefficient was large and the time series was long. A potential explanation is that these results are due to the cutoffs GIMME is using when testing whether model fit is good enough to terminate the search for paths. Specifically, once the group-level paths have been selected by GIMME, a path model is estimated for each person in the sample to search for paths in this person's model. The search is terminated after two of four standard SEM fit indices indicate good fit. As the individual-level or subset path is only *one* path coefficient that occurs in a single person's model, our result may thus simply be due to using cutoff values for the fit indices that are too moderate. Again, future simulation research is needed to replicate this finding using different cutoff values and to determine cutoffs that ensure sufficient performance in detecting individual-level and subset paths.[4]

Besides further simulation research to better understand the pattern of results of the reported study, we believe that another interesting task for future research is to compare the performance of the GIMME algorithm with other approaches that could be used to identify individuals that can be combined because their ideographic results are similar to other persons in the sample. For instance, Nesselroade and Molenaar (1999) suggested to compare the Bloeck-Toeplitz matrices of the subjects using a chi-square test for the decision of whether their matrices can be pooled. They also suggested that, when the test is significant, one can create subgroups of 'poolable' individuals using the individuals' contributions to the chi-square test statistic. Alternatively, a multiple-group structural equation modeling approach could be used in which one first estimates a saturated

---

3) We checked this by inspecting the number of nonzero path recoveries for conditions in which one of the other group-level paths was 0.4 (and the other coefficients were 0.2). The number of true recoveries was very small in all of these cases.

4) At present it is not possible to specify cutoff-values for the SEM fit indices.

model and then deletes those paths that are not significantly different from zero (i.e., backward selection).

At present, GIMME uses MIs and other fit measures for path detection. It is well known that many of these measures can perform poorly in small samples (or applied to the present context, with only a few time points). Hence, it would be an interesting task for future research to examine whether alternative approaches (e.g., such as bootstrapping or Bayesian model selection approaches) can be used to improve GIMME's performance in these circumstances. Furthermore, we also believe that another interesting task for future research would be to extend the algorithm to cases in which responses were not measured on continuous scales. We think that this is relevant because most experience sampling studies use ordinal scales (e.g., a Likert-scale) and simulation research shows that treating ordinal outcomes as continuous can have negative effects on the parameter estimates and the performance of the statistical tests of a SEM (Rhemtulla, Brosseau-Liard, & Savalei, 2012).

Another interesting avenue for future research would be to implement a continuous-time version of the uSEM in the GIMME algorithm (de Haan-Rietdijk, Voelkle, Keijsers, & Hamaker, 2017). At present, the uSEM assumes a constant time-lag between measurements. However, experience sampling data is often assessed with study designs that cannot ensure that this assumption is met. Research on longitudinal path models showed that parameter estimates can be biased and also hard to interpret in such a case (Voelkle, Oud, Davidov, & Schmidt, 2012). Hence, we consider the development of a continuous-time version of the GIMME algorithm and a comparison of its performance in path detection with the currently implemented discrete-time version a challenging but interesting objective for future research. Finally, the current focus of the GIMME algorithm is on path detection. However, another important usage of experience sampling data (as a type of time series data) is to predict people's future scores on the variables of interest, that is, at time-points not contained in the data (Hyndman & Athanasopoulos, 2018). We believe that investigating GIMME's forecasting performance and comparing it to the predictions made by other forecasting models (e.g., vector autoregressive models, DSEM, etc.) is also an interesting task for future research.[5]

---

5) In a preliminary simulation study, we compared the forecasting performance of GIMME with the performance of a vector autoregressive model (VAM) computed across a pooled sample for a one-step ahead forecast (see Hyndman & Athanasopoulos, 2018). In this simulation study, we used the same population model and the same simulation conditions as in Study 2. The only difference was that we compared only two time series length conditions: 50 vs. 100 time points. To compute the forecast for GIMME, we saved a single participant's path coefficients that were detected by GIMME, and then used these coefficients to compute the forecast for each single participant in a replication. For VAM, we first averaged the values of the participants in a replication, estimated a VAM across the pooled sample, and used the resulting coefficients to compute a forecast for each single participant in a replication. The quality of the GIMME-forecasts and the VAM-forecasts were then compared by computing the root mean square error of the forecasts across the single individuals of a replication and then averaging these values across the replications in a simulation condition. The results showed that the predictive performance was better the larger the number of

In short, then, the present work shows that GIMME produces a very small number of false discoveries. The number of true recoveries depends on the length of the time series, the magnitude of the true nonzero group-level path coefficient, and the magnitude of the to-be-selected contemporaneous or off-diagonal lagged coefficients. An unexpected results emerged with regard to true nonzero individual-level and subset paths where the number of recoveries depended on the magnitude of the group-level paths. However, the present simulation was just a first step in investigating GIMME' performance. More simulation research is needed to investigate its performance with regard to time-series path models. We believe that it would also be a worthwhile task for future research to examine GIMME for other models, such as dynamic factor analysis or exploratory factor analysis.

Supplementary Materials

For this article the following Supplementary Materials for Studies 1, 2, and 3 are available via the OSF repository (for access see Index of Supplementary Materials below):

- Data files containing the parameter estimates.
- lavaan code to estimate the uSEMs.
- R code to determine the (relative) bias of the parameter estimates.

Index of Supplementary Materials

Nestler, S., & Humberg, S. (2021). *Additional Materials for "GIMME's ability to recover group-level path coefficients and individual-level path coefficients"* [Data, code]. OSF. https://osf.io/hrjdw/

---

time-points (RMSE-100 = 1.01 vs. RMSE-50 = 1.06), when GIMME was used compared to VAM (RMSE-GIMME = 1.01 vs. RMSE-VAM = 1.05), and the lower the magnitude of the contemporaneous coefficients (RMSE-0.2 = 1.01, RMSE-0.4 = 1.03, and RMSE-0.6 = 1.07). The latter effect only occurred in case of VAM (VAM: RMSE-0.2 = 1.01, RMSE-0.4 = 1.05, and RMSE-0.6 = 1.10; GIMME: RMSE-0.2 = 1.00, RMSE-0.4 = 1.01, and RMSE-0.6 = 1.02).

References

Asparouhov, T., Hamaker, E. L., & Muthén, B. O. (2018). Dynamic structural equation models. *Structural Equation Modeling, 25*(3), 359-388. https://doi.org/10.1080/10705511.2017.1406803

Beltz, A. M., & Gates, K. M. (2016). Dealing with multiple solutions in structural vector autoregressive models. *Multivariate Behavioral Research, 51*(2-3), 357-373. https://doi.org/10.1080/00273171.2016.1151333

Beltz, A. M., & Gates, K. M. (2017). Network mapping with gimme. *Multivariate Behavioral Research, 52*(6), 789-804. https://doi.org/10.1080/00273171.2017.1373014

Bollen, K. A. (1989). *Structural equations with latent variables*. Chichester, United Kingdom: John Wiley & Sons.

Box-Steffensmeier, J. M., Freeman, J. R., Hitt, M. P., & Pevehouse, J. C. W. (2014). *Time series analysis for the social sciences*. New York, NY, USA: Cambridge University Press.

Cattell, R. B., & Luborsky, L. B. (1950). P-technique demonstrated as a new clinical method for determining personality and symptom structure. *The Journal of General Psychology, 42*(1), 3-24. https://doi.org/10.1080/00221309.1950.9920145

de Haan-Rietdijk, S., Voelkle, M. C., Keijsers, L., & Hamaker, E. L. (2017). Discrete- vs. continuous-time modeling of unequally spaced experience sampling method data. *Frontiers in Psychology, 8*, Article 1849. https://doi.org/10.3389/fpsyg.2017.01849

Gates, K. M., Lane, S. T., Varangis, E., Giovanello, K., & Guiskewicz, K. (2017). Unsupervised classification during time-series model building. *Multivariate Behavioral Research, 52*(2), 129-148. https://doi.org/10.1080/00273171.2016.1256187

Gates, K. M., & Molenaar, P. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage, 63*(1), 310-319. https://doi.org/10.1016/j.neuroimage.2012.06.026

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2002). On the nature of sem estimates of arma parameters. *Structural Equation Modeling, 9*(3), 347-368. https://doi.org/10.1207/S15328007SEM0903_3

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2003). Arma-based sem when the number of time points t exceeds the number of cases n: Raw data maximum likelihood. *Structural Equation Modeling, 10*(3), 352-379. https://doi.org/10.1207/S15328007SEM1003_2

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. https://doi.org/10.1207/s15327906mbr4002_3

Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science, 26*(1), 10-15. https://doi.org/10.1177/0963721416666518

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. Hoboken, NJ, USA: John Wiley & Sons.

PsychOpen GOLD

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice.* OTexts.

Kim, J., Zhu, W., Chang, L., Bentler, P., & Ernst, T. (2007). Unified structural equation modeling approach for the analysis of multisubject, multivariate functional mri data. *Human Brain Mapping, 28*(2), 85-93. https://doi.org/10.1002/hbm.20259

Lane, S. T., & Gates, K. M. (2017). Automated selection of robust individual-level structural equation models for time series data. *Structural Equation Modeling, 24*(5), 768-782. https://doi.org/10.1080/10705511.2017.1309978

Lane, S. T., Gates, K. M., Fisher, Z., Molenaar, P., Hallquist, M., Pike, H., . . . & Luo, L. (2019a). gimme: Group iterative multiple model estimation [Computer software manual] (R package version 0.6-0). Retrieved from https://CRAN.R-project.org/package=gimme

Lane, S. T., Gates, K., Pike, H., Beltz, A., & Wright, A. (2019b). Uncovering general, shared, and unique temporal patterns in ambulatory assessment data. *Psychological Methods, 24*(1), 54-69. https://doi.org/10.1037/met0000192

Little, T. D. (2013). *Longitudinal structural equation modeling.* New York, NY, USA: Guilford Press.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis.* New York, NY, USA: Springer.

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika, 50*(2), 181-202. https://doi.org/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: Interdisciplinary Research and Perspectives, 2*(4), 201-218. https://doi.org/10.1207/s15366359mea0204_1

Mulaik, S. A. (2009). *Linear causal modeling with structural equations.* New York, NY, USA: Chapman and Hall, CRC Press.

Mund, M., & Nestler, S. (2019). Beyond the cross-lagged panel model: Next-generation statistical tools for analyzing interdependencies across the life course. *Advances in Life Course Research, 41*, Article 100249. https://doi.org/10.1016/j.alcr.2018.10.002

Nesselroade, J., McArdle, J., Aggen, S., & Meyers, J. (2002). Dynamic factor analysis models for representing process in multivariate time-series. In D. M. Moskowitz & S. L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measure data: Advances and techniques* (pp. 235-365). Mahwah, NJ, USA: Lawrence Erlbaum Associates.

Nesselroade, J., & Molenaar, P. (1999). Pooling lagged covariance structures based on short, multivariate time series for dynamic factor analysis. In R. H. Hoyle (Ed.), *Statistical techniques for small sample research* (pp. 223-250). Thousand Oaks, CA, USA: Erlbaum.

Nesselroade, J., & 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. https://doi.org/10.1207/s15427617rhd0101&2_3

Nestler, S. (2020). Modeling interindividual differences in latent within-person variation: The confirmatory factor level variability model. *British Journal of Mathematical & Statistical Psychology, 73*(3), 452-473. https://doi.org/10.1111/bmsp.12196

Nestler, S. (2021). Modeling intraindividual variability in growth with measurement burst designs. *Structural Equation Modeling, 28*(1), 28-39. https://doi.org/10.1080/10705511.2020.1757455

R Core Team. (2019). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? a comparison of robust continuous and categorical sem estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354-373. https://doi.org/10.1037/a0029315

Sörbom, D. (1989). Model modification. *Psychometrika, 54*(3), 371-384. https://doi.org/10.1007/BF02294623

Spencer, J. P., & Schöner, G. (2003). Bridging the representational gap in the dynamic systems approach to development. *Developmental Science, 6*(4), 392-412. https://doi.org/10.1111/1467-7687.00295

Voelkle, M. C., Oud, J. H. L., Davidov, E., & Schmidt, P. (2012). An sem approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods, 17*(2), 176-192. https://doi.org/10.1037/a0027543

Voelkle, M. C., Oud, J. H. L., von Oertzen, T., & Lindenberger, U. (2012). Maximum likelihood dynamic factor modeling for arbitrary n and t using sem. *Structural Equation Modeling, 19*(3), 329-350. https://doi.org/10.1080/10705511.2012.687656

Wright, A. G. C., Gates, K. M., Arizmendi, C., Lane, S. T., Woods, W. C., & Edershile, E. A. (2019). Focusing personality assessment on the person: Modeling general, shared, and person specific processes in personality and psychopathology. *Psychological Assessment, 31*(4), 502-515. https://doi.org/10.1037/pas0000617

Wright, A. G. C., & Zimmermann, J. (2019). Applied ambulatory assessment: Integrating idiographic and nomothetic principles of measurement. *Psychological Assessment, 31*(12), 1467-1480. https://doi.org/10.1037/pas0000685

Wrzus, C., & Mehl, M. R. (2015). Lab and/or field? measuring personality processes and their social consequences. *European Journal of Personality, 29*(2), 250-271. https://doi.org/10.1002/per.1986

Zhang, Z., Hamaker, E., & Nesselroade, J. (2008). Comparisons of four methods for estimating a dynamic factor model. *Structural Equation Modeling, 15*(3), 377-402. https://doi.org/10.1080/10705510802154281

Appendices

Appendix A

In this Appendix we present the results for the *n* = 75 persons conditions in Study 1. We begin with the results for the number of true recoveries and then present the results for the false recoveries.

PsychOpen GOLD

Table A1

GIMME's Performance in Detecting the Group-Level Coefficients ϕ_{ij}^g and a_{ij}^g Depending on the Length of the Time Series T, the Magnitude of the AR Coefficient ϕ_{ii}, and the Magnitude of the Off-Diagonal Coefficient ϕ_{ij}^g or the Magnitude of the Contemporaneous Coefficient a_{ij}^g for n = 75 persons

T	ϕ_{ii}	ϕ_{ij} 0.20	0.40	a_{ij} 0.20	0.40	0.60
25	0.20	43.18	108.86	30.39	70.79	109.88
25	0.40	47.07	115.99	35.02	87.42	137.79
25	0.60	53.89	126.02	41.81	104.16	162.12
50	0.20	52.00	157.56	37.71	108.28	167.00
50	0.40	55.40	162.86	44.45	134.35	203.43
50	0.60	61.50	164.69	53.65	152.87	222.03
75	0.20	67.23	187.42	49.08	131.54	213.62
75	0.40	68.49	190.11	57.05	166.12	225.00
75	0.60	68.94	186.21	64.36	197.47	225.00
100	0.20	80.43	206.92	59.52	168.66	215.12
100	0.40	79.41	205.83	68.55	209.10	224.47
100	0.60	74.69	216.75	72.73	224.42	225.00
125	0.20	92.01	221.31	68.64	214.21	222.15
125	0.40	88.32	224.46	79.22	225.00	224.79
125	0.60	78.66	225.00	78.19	225.00	225.00
250	0.20	114.57	225.00	89.36	223.21	225.00
250	0.40	127.26	225.00	98.03	225.00	225.00
250	0.60	151.53	225.00	111.56	225.00	225.00

Note. 225 recoveries were possible within each simulation condition.

Detection of Non-zero Paths

Similar to the n = 25 person condition, the ANOVA for the lagged group-level coefficients yielded main effects of series length T ($\eta_p^2 = 0.94$) and magnitude of ϕ_{ij} ($\eta_p^2 = 0.98$). These main effects were qualified by a significant two-way interaction between $T \times \phi_{ij}$ ($\eta_p^2 = 0.69$) and a significant three-way interaction of T, ϕ_{ii}, and ϕ_{ij} ($\eta_p^2 = 0.19$). As can be seen in Table A1, the results are very similar to the n = 25 person conditions as the observed number was closer to 225 (3 parameters × 75 persons) the longer the time series and the larger the magnitude of ϕ_{ij}. Also, when ϕ_{ij} was 0.20, a larger AR effect lead to a higher number of recovered true path coefficients, whereby this influence of the magnitude of the AR effect was stronger the longer the time series. This effect was smaller or not evident when ϕ_{ij} was 0.40.

For the contemporaneous group-level coefficients, main effects of series length T ($\eta_p^2 = 0.91$), magnitude of a_{ij} ($\eta_p^2 = 0.97$), and magnitude of ϕ_{ii} ($\eta_p^2 = 0.47$) were sizable. Also, the $T \times a_{ij}$ interaction ($\eta_p^2 = 0.68$) and the $T \times a_{ij} \times \phi_{ii}$ interaction was important ($\eta_p^2 = 0.25$). The number of detected true nonzero paths (see 5 again) was higher the longer the time series and the larger the magnitude of a_{ij}. When a_{ij} was 0.20, a larger AR effect went along with more recoveries. This effects was smaller, the larger a_{ij}.

False Recoveries

As in the case of n = 25 persons, the average number of falsely recovered autoregressive paths was small (M = 38.9, SD = 40.8) compared to the 1,275 path that are possible. The ANOVA yielded main effects of time series length T (η_p^2 = 0.96), magnitude of a_{ij} (η_p^2 = 0.31), and magnitude of ϕ_{ii} (η_p^2 = 0.20). Also, and in contrast to n = 25 persons, a main effect of the magnitude of phi_{ij} appeared. Further analyses showed that the longer the series, the lower the number of false positive recoveries: M_{25} = 118.3, M_{50} = 50.3, M_{75} = 30.1, M_{100} = 19.3, M_{125} = 13.5, M_{250} = 2.1. Furthermore, a larger AR effect ϕ_{ii} ($M_{0.2}$ = 44.4, $M_{0.4}$ = 39.4, $M_{0.6}$ = 30.5), a larger contemporaneous coefficient a_{ij} ($M_{0.2}$ = 47.9, $M_{0.4}$ = 38.8, $M_{0.6}$ = 30.5), and a larger autoregressive coefficients phi_{ij} ($M_{0.2}$ = 42.5, $M_{0.4}$ = 35.4) was related to a smaller number of false lagged coefficient recoveries.

With regard to the contemporaneous coefficients in A, we found that the average proportion of false recoveries across simulation conditions was small (M = 42.9, SD = 40.6, again 1,275 false positives were possible within each replication). The ANOVA showed main effects of time-series length T (η_p^2 = 0.91), magnitude of a_{ij} (η_p^2 = 0.30), magnitude of ϕ_{ii} (η_p^2 = 0.49), and the magnitude of phi_{ij} ((η_p^2 = 0.15). Furthermore, the $T \times a_{ij}$ interaction (η_p^2 = 0.15) and the $T \times a_{ij} \times \phi_{ii}$ interaction were also sizable (η_p^2 = 0.25). Again, most false positives occurred when T was 25 or 50. Finally, about 50% of the false positives in a simulation condition were due to the false recovery of the opponent path to the true path.

Appendix B

In this Appendix we present additional results for Study 2. We begin with the results for the number of recoveries of the group-level coefficients. Thereafter, we present additional results for the individual-level coefficients.

Table B1

GIMME's Performance in Detecting the Group-Level Coefficients ϕ_{ij}^g and a_{ij}^g Depending on the Length of the Time Series T, the Magnitude of the AR Coefficient ϕ_{ii}, and the Magnitude of the Off-Diagonal Coefficient ϕ_{ij}^g or the Magnitude of the Contemporaneous Coefficient a_{ij}^g

T	ϕ_{ii}	ϕ_{ij}^g		a_{ij}^g		
		0.20	**0.40**	**0.20**	**0.40**	**0.60**
50	0.20	17.46	52.97	12.68	35.97	64.11
50	0.40	19.21	54.97	15.32	44.59	73.59
50	0.60	20.66	57.54	17.60	52.87	74.85
75	0.20	22.71	64.39	16.16	47.58	71.19
75	0.40	23.00	66.20	18.84	61.72	74.36
75	0.60	22.68	66.33	21.12	71.12	75.00
100	0.20	26.91	71.88	19.96	64.95	74.05
100	0.40	26.50	72.80	22.75	73.83	74.65
100	0.60	25.64	74.52	23.52	74.95	75.00
250	0.20	42.88	75.00	30.93	74.83	75.00
250	0.40	49.83	75.00	39.88	74.96	75.00

PsychOpen GOLD

T	ϕ_{ii}	ϕ_{ij}^{g}		a_{ij}^{g}		
		0.20	**0.40**	**0.20**	**0.40**	**0.60**
250	0.60	63.48	75.00	63.42	75.00	75.00
500	0.20	74.94	75.00	71.31	75.00	75.00
500	0.40	75.00	75.00	74.94	75.00	75.00
500	0.60	75.00	75.00	75.00	75.00	75.00

Note. 75 recoveries were possible within each simulation condition.

Group-Level Coefficients

For the lagged group-level coefficients, the ANOVA yielded main effects of series length T (η_p^2 = 0.89) and magnitude of ϕ_{ij}^{g} (η_p^2 = 0.90). These main effects were qualified by a significant two-way interaction between $T \times \phi_{ij}^{g}$ (η_p^2 = 0.75) and a significant three-way interaction of T, ϕ_{ii}, and ϕ_{ij}^{g} (η_p^2 = 0.11). For the contemporaneous group-level coefficients, the main effects of series length T (η_p^2 = 0.82), magnitude of a_{ij}^{g} (η_p^2 = 0.91), and magnitude of ϕ_{ii} (η_p^2 = 0.25) were sizable. Also, the $T \times a_{ij}^{g}$ interaction (η_p^2 = 0.79) and the $T \times a_{ij}^{g} \times \phi_{ii}$ interaction was important (η_p^2 = 0.28). Table B1 presents concrete results for the number of recoveries for the two types of parameters.

Further Results for Individual-Level Coefficients

Table B2 presents the number of path recoveries when all group-level coefficients were small (i.e., $\phi_{g,ij} = a_{g,ij} = \phi_{ii} = 0.2$):

Table B2

GIMME's Performance in Detecting Individual-Level Coefficients s Depending on the Magnitude of the Individual-Level Coefficient s and the Time-Series Length T

s	T	s_L	s_C
0.20	50	0.95	1.32
	75	1.04	1.40
	100	1.37	1.70
	250	2.08	2.44
	500	0.55	0.72
0.40	50	2.15	2.86
	75	2.58	3.54
	100	3.12	3.64
	250	3.65	4.00
	500	3.72	4.00

Note. Four recoveries were possible within each simulation condition. L = lagged; C = contemporaneous.

Appendix C

In this Appendix we present additional results from Study 3. We start with the results for the detection of the group-level coefficients. Then we present the results for the detection of the individual-level coefficients.

Table C1

GIMME's Performance in Detecting the Group-Level Coefficients ϕ_{ij}^g and a_{ij}^g Depending on the Length of the Time Series T, the Magnitude of the AR Coefficient ϕ_{ii}, and the Magnitude of the Off-Diagonal Coefficient ϕ_{ij}^g or the Magnitude of the Contemporaneous Coefficient a_{ij}^g

		ϕ_{ij}^g		a_{ij}^g		
T	ϕ_{ii}	**0.20**	**0.40**	**0.20**	**0.40**	**0.60**
50	0.20	17.81	52.08	13.16	36.80	66.09
50	0.40	18.67	54.13	15.56	45.48	73.83
50	0.60	19.80	57.27	17.93	55.55	74.95
75	0.20	22.23	64.51	16.95	52.39	71.99
75	0.40	22.64	66.61	19.63	65.67	74.46
75	0.60	21.48	66.68	20.01	72.53	75.00
100	0.20	26.61	72.49	20.24	68.24	73.93
100	0.40	25.84	73.35	23.02	74.06	74.81
100	0.60	25.93	73.71	21.58	74.69	75.00
250	0.20	45.18	75.00	35.71	74.86	75.00
250	0.40	53.71	75.00	50.27	75.00	75.00
250	0.60	64.22	75.00	71.24	75.00	75.00
500	0.20	74.64	75.00	72.42	75.00	75.00
500	0.40	74.93	75.00	74.82	75.00	75.00
500	0.60	75.00	75.00	75.00	75.00	75.00

Note. 75 recoveries were possible within each condition.

Group-Level Coefficients

We computed an ANOVA with the factors time-series length T, magnitude of ϕ_{ii}, magnitude of ϕ_{ij}^g, magnitude of a_{ij}^g, magnitude of s, and magnitude of p. For the lagged group-level coefficients this ANOVA yielded main effects of T ($\eta_p^2 = 0.89$) and ϕ_{ij}^g ($\eta_p^2 = 0.90$). These main effects were qualified by a significant two-way interaction between T × magnitude of ϕ_{ij}^g ($\eta_p^2 = 0.75$) and a significant three-way interaction of T, magnitude of ϕ_{ii}, and magnitude of ϕ_{ij}^g ($\eta_p^2 = 0.11$). For the contemporaneous group-level coefficients, we found main effects of T ($\eta_p^2 = 0.82$), a_{ij}^g ($\eta_p^2 = 0.91$), and ϕ_{ii} ($\eta_p^2 = 0.25$). Also, the T × a_{ij}^g magnitude interaction ($\eta_p^2 = 0.79$) and the three-way interaction of T × magnitude of a_{ij}^g × magnitude of ϕ_{ii} were sizable ($\eta_p^2 = 0.31$). Table C1 provides the average number of recovered paths for the two types of coefficients.

Individual-Level and Subset Coefficients

With regard to the individual-level off-diagonal lagged path coefficients (s_1 to s_3), the ANOVA yielded main effects of time-series length T ($\eta_p^2 = 0.13$), the magnitude of ϕ_{ii} ($\eta_p^2 = 0.13$), the magnitude of a_{ij}^g ($\eta_p^2 = 0.10$), the magnitude of ϕ_{ij}^g ($\eta_p^2 = 0.10$), and the magnitude of the individual-level coefficient s ($\eta_p^2 = 0.47$). There was also a $T \times \phi_{ii}$ interaction ($\eta_p^2 = 0.14$). For the individual-level contemporaneous coefficient, the ANOVA yielded main effects of T ($\eta_p^2 = 0.21$), ϕ_{ii} ($\eta_p^2 = 0.20$), a_{ij}^g ($\eta_p^2 = 0.26$), and s ($\eta_p^2 = 0.43$). Table C2 presents the average number of recoveries for the two types of individual-level coefficients.

Finally, Table C3 shows the number of individual-level and subset coefficient recoveries when all group-level coefficients were small (i.e., $\phi_{g,ij} = a_{g,ij} = \phi_{ii} = 0.2$). Note that the magnitude of the subset coefficient p did not affect the recovery of individual-level path coefficients.

Table C2

GIMME' Performance in Detecting Individual-Level Coefficients s Depending on the Magnitude of the Individual-Level Coefficient, Time-Series Length T, the Magnitude of the AR Coefficient ϕ_{ii}, and the Magnitude of the Contemporaneous Group-Level Coefficient a_{ij}^g

			s_L			s_C		
s	T	ϕ_{ii}	$a_{ij}^g = 0.2$	$a_{ij}^g = 0.4$	$a_{ij}^g = 0.6$	$a_{ij}^g = 0.2$	$a_{ij}^g = 0.4$	$a_{ij}^g = 0.6$
0.20	50	0.20	0.64	0.54	0.52	0.78	0.65	0.59
0.20	50	0.40	0.76	0.70	0.56	0.90	0.74	0.47
0.20	50	0.60	0.79	0.64	0.51	1.04	0.56	0.26
0.20	75	0.20	0.82	0.60	0.57	1.04	0.76	0.68
0.20	75	0.40	0.84	0.59	0.60	1.10	0.76	0.44
0.20	75	0.60	0.86	0.53	0.40	1.08	0.39	0.21
0.20	100	0.20	0.92	0.66	0.60	1.15	0.78	0.61
0.20	100	0.40	0.96	0.69	0.54	1.22	0.60	0.38
0.20	100	0.60	0.93	0.50	0.35	1.26	0.23	0.12
0.20	250	0.20	1.17	0.74	0.49	1.53	0.62	0.21
0.20	250	0.40	0.90	0.31	0.20	0.93	0.15	0.01
0.20	250	0.60	0.06	0.03	0.01	0.03	0.00	0.00
0.20	500	0.20	0.48	0.09	0.06	0.52	0.08	0.00
0.20	500	0.40	0.07	0.00	0.00	0.07	0.00	0.00
0.20	500	0.60	0.00	0.00	0.00	0.00	0.00	0.00
0.40	50	0.20	1.57	1.62	1.56	2.17	1.85	1.64
0.40	50	0.40	1.92	1.77	1.71	2.14	1.77	1.51
0.40	50	0.60	2.01	1.82	1.51	2.16	1.55	1.04
0.40	75	0.20	2.03	1.81	1.97	2.48	2.24	1.96
0.40	75	0.40	2.21	2.09	1.99	2.45	2.09	1.62
0.40	75	0.60	2.40	1.66	1.42	2.49	1.33	0.85
0.40	100	0.20	2.15	2.09	2.05	2.66	2.40	2.10
0.40	100	0.40	2.48	2.16	1.87	2.69	2.06	1.45

s	T	ϕ_{ii}	s_L			s_C		
			$a_{ij}^g = 0.2$	$a_{ij}^g = 0.4$	$a_{ij}^g = 0.6$	$a_{ij}^g = 0.2$	$a_{ij}^g = 0.4$	$a_{ij}^g = 0.6$
0.40	100	0.60	2.48	1.38	1.01	2.60	1.06	0.55
0.40	250	0.20	2.61	2.54	2.16	2.96	2.56	1.46
0.40	250	0.40	2.67	1.60	1.05	2.73	1.24	0.38
0.40	250	0.60	0.78	0.24	0.13	0.52	0.03	0.00
0.40	500	0.20	2.81	2.62	1.84	2.99	2.48	0.47
0.40	500	0.40	2.42	0.83	0.39	2.27	0.40	0.01
0.40	500	0.60	0.26	0.00	0.00	0.12	0.00	0.00

Note. Three recoveries were possible within each simulation condition. L = lagged; C = contemporaneous.

Table C3

GIMME' Performance in Detecting Individual-Level Coefficients s or Subset Coefficients p Depending on the Magnitude of the Respective Coefficient (Called Size) and Time-Series Length T

Size	T	s_L	s_C	p_L	p_C
0.20	50	0.64	0.85	4.04	2.04
0.20	75	0.82	1.14	5.30	2.58
0.20	100	1.00	1.36	6.24	3.38
0.20	250	1.39	1.73	7.84	5.45
0.20	500	0.56	0.54	5.06	4.36
0.40	50	1.61	2.25	9.42	5.76
0.40	75	2.02	2.56	10.75	6.85
0.40	100	2.21	2.77	11.35	7.90
0.40	250	2.64	2.99	11.96	9.84
0.40	500	2.75	3.00	11.96	10.44

Note. Three or twelve recoveries were possible within each simulation condition. L = lagged; C = contemporaneous.