



Specifying exogeneity and bilinear effects in data-driven model searches

Cara Arizmendi¹ · Kathleen Gates¹ · Barbara Fredrickson¹ · Aidan Wright²

Published online: 9 October 2020
© The Psychonomic Society, Inc. 2020

Abstract

Data-driven model searches provide the opportunity to quantify person-specific processes using ambulatory assessment data. Here, the search space typically includes all potential relations among variables, meaning that all variables can potentially explain variability in all other variables. Oftentimes, this is unrealistic. For example, weather is unlikely to be predicted by someone's emotional state, whereas the reverse might be true. Allowing for specification of exogenous variables, or variables that are not predicted within the system, permits more realistic models and allows the researcher to model contextual change processes via the use of moderation variables. We use two sets of daily diary data to demonstrate the capabilities of allowing for the specification of exogenous variables in GIMME (Group Iterative Multiple Model Estimation), a model search algorithm that allows for models with idiographic, individual-level as well as subgroup- and group-level processes with intensive longitudinal data. First, using data collected from individuals diagnosed with personality disorders, we show results where weather-related and temporal basis variables are specified as exogenous, and reports on affect and behavior are endogenous. Next, we demonstrate the modeling of treatment effects in an intervention study, looking at data from a 6-week meditation workshop in midlife adults. Finally, we use the meditation intervention data to demonstrate modeling moderation effects, where relationships between two endogenous variables are dependent on the current stage of the study for a given participant (i.e., currently attending meditation classes or not). We end by presenting adaptive LASSO as a method for probing results obtained from GIMME.

Keywords Data-driven · Gimme · Exogenous · Moderator · Sem · Time series · Daily diary · Model search

Introduction

Data-driven model searches are an increasingly popular approach for detecting relations among psychological, behavioral, and physiological variables. In particular, network searches are of interest to psychologists aiming to model relationships between variables as a grouped latent construct as is done in factor analysis (Cramer et al., 2016; Schmittmann et al., 2013; Bringmann et al., 2015; Costantini et al., 2015; Wright et al., 2018). Typically, network search approaches used in psychology treat all variables as potentially exogenous and endogenous. That is,

all variables in the model search are handled as potentially predicting or being predicted by other variables in the model. Few methods allow for *a priori* specification of exogenous variables, or variables that can predict but not be predicted. A recent extension of the Gaussian Graphical Model (GGM) as a Moderated Network Model (MNM) allows for exogenous and moderating effects in the network framework (Haslbeck et al., 2019). Here, we discuss a similar and novel extension of the Group Iterative Multiple Model Estimation (GIMME; Gates & Molenaar, 2012) algorithm, in which model searches can now include *a priori* specified exogenous variables and moderating effects. We discuss various options for utilizing exogenous variables in addition to discussing interpretation of results.

Allowing for specification of exogenous variables permits more realistic models. For example, in a model search that includes variables for rainfall and self-reported emotions, indicating that rainfall is exogenous does not allow for emotions to predict rainfall but does allow for the reverse. Additionally, allowing for variables to be specified

✉ Cara Arizmendi
carizmendi@unc.edu

¹ The University of North Carolina Chapel Hill, CB #3270, Davie Hall, Chapel Hill, NC 27599-3270, USA

² The University of Pittsburgh, Pittsburgh, PA 15260, USA

as exogenous opens up the possibility for including temporal basis vectors. Temporal basis vectors allow for the modeling aspects of linear detrending and explorations of the influence of time or day of week on variables to occur within the model search procedure. Specification of exogenous variables also allows for the modeling of contextual change processes present in crossover design studies, such as level changes (e.g., no treatment versus currently receiving treatment), as well as moderating effects (e.g., how being in treatment moderates the relationship between positive emotions and some other variable). In this paper, we present the new capabilities of allowing for the specification of exogenous variables in GIMME (Group Iterative Multiple Model Estimation; Gates & Molenaar 2012), a search algorithm that allows for the modeling of idiographic individual level processes as well as subgroup level (Gates et al., 2017) and group level processes (Gates & Molenaar, 2012). GIMME conducts searches for relationships among variables from within a unified structural equation modeling (uSEM) framework (Kim et al., 2006; Gates et al., 2011), which include directed lagged (i.e., autoregressive and cross-regressive) relations as well as directed contemporaneous relations among variables.

This paper first discusses the benefits of allowing for exogenous variables in the network model search, including: more realistic model search spaces, handling non-stationarity (i.e., changing mean or variance level across time), modeling changes in crossover design studies, and modeling interaction effects. During these discussions we present equations in uSEM form. Given the extant literature on uSEM (see Gates et al., 2010, 2017, Beltz & Gates, 2017) we only present general concepts here necessary for interpretation of the empirical findings. We then present findings from three empirical daily diary examples. First, using data collected from individuals diagnosed with personality disorders, we demonstrate a network model search on exogenous, weather-related variables with self-reports of affect and behavior. A linear trend across time is included in the model to account for potential non-stationarity in mean for some subjects. Second, we complete a search on treatment effects in an intervention study looking at the effects of a 6-week meditation workshop on health behaviors in midlife adults. Finally, we use the meditation intervention data and include interaction effects between current stage of study (i.e., currently attending meditation classes or not) and affect in the search space.

In addition, a challenge in this area of research is that in some cases it can be difficult to arrive at inferences and generalizable interpretations from the person-specific patterns of relations and estimates. This is particularly true if there is a lot of heterogeneity in the patterns of relations when looking across individuals. We supplement the

data-driven searches by exploring how the presence and estimates of specific paths, which quantify relations across time, relate to trait relations when looking across people. For model selection we utilize LASSO (least absolute shrinkage and selection operator) for uncovering relationships between individual-level dynamic processes and trait-level variables. Specifically, we conducted LASSO regression using the GIMME-estimated individual-level paths as predictors, exploring potential relationships between path strength (the beta coefficient obtained in the model search) of each individual and the trait-level variables (Woods et al., 2020a).

Benefits of allowing for prespecified exogeneity

Exogenous variables are defined as variables “generated from outside the system” (Yin, 2010). Exogenous variables can only ever be independent variables, meaning that they can only predict and cannot be dependent variables predicted by other variables in the model. The concept of exogeneity has been especially prevalent in the field of econometrics (Engle et al., 1983), where exogenous variables are used to model shocks to economic systems, such as the impact of disruption in oil production on inflation (Kilian, 2008) or customer income on supply and demand models. The notion of income as an exogenous variable has also been conceptualized in the field of social science, where family income is an exogenous variable predicting child’s school attendance and grades (Yin, 2010). Family income cannot be endogenous in this system since school attendance and grades cannot causally predict family income. Another example of exogeneity is time: by most accounts it is linear and not influenced by activity of humans. We expand on the use of exogenous variables across different contexts below.

More plausible models

Allowing for the specification of exogenous variables in the model search allows us to model effects of variables that cannot plausibly be influenced by other variables in the system. The notion of a system is not limited to large economic and social structures described above. A system can also exist within an individual. For example, if we model a system of biological processes in the body (Yin, 2010), a variable generated from outside the body, such as introduction of vitamin D, is understood to be an exogenous variable (Horst, 2010). Similarly, in psychology, we can understand an individual’s emotions and behavior as a system, whereby we can model external influences observationally (e.g., weather-related variables) or experimentally (e.g., experimental group). An individual’s emotions cannot have a causal influence on

the weather or in randomized-controlled experiments, their treatment status. While most data-driven approaches ignore the concept of exogeneity, it is clearly an important concept that will aid in model search procedures.

An extension to the uSEM framework, referred to as extended unified-SEM (Gates et al., 2011), allowed for modeling exogenous variables and moderating effects. In the extended unified structural equation modeling framework, exogenous variables are represented as u_t in Eq. 1.:

$$\eta_t = \mathbf{A}\eta_t + \Phi\eta_{t-1} + \Gamma\mathbf{u}_t + \zeta_t \quad (1)$$

where η is a $N \times 1$ vector of endogenous variables at time (t) or at a lag ($t - 1$), N is the number of endogenous variables, \mathbf{A} is an $N \times N$ matrix of contemporaneous coefficient estimates, Φ an $N \times N$ matrix of lagged coefficient estimates, Γ is an $N \times P$ matrix of regression coefficient estimates where P is the number of exogenous variables, \mathbf{u}_t is an $P \times 1$ vector of exogenous variables at time (which is a scalar when $P=1$), and ζ_t is a vector of error (or innovation) terms. The errors are assumed to be uncorrelated with each other and across time.

Continuing with the example of modeling an individual's emotions as a system impacted by exogenous weather variables, we can understand how exogenous variables are represented in matrix form by setting $\eta_1 = PA$, $\eta_2 = NA$, and $u = temperature$:

$$\begin{bmatrix} PA_t \\ NA_t \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 \\ a_{21} & 0 \end{bmatrix}}_{\text{endogenous}} \begin{bmatrix} PA_t \\ NA_t \end{bmatrix} + \underbrace{\begin{bmatrix} \phi_{11} & 0 \\ 0 & \phi_{22} \end{bmatrix}}_{\text{lagged}} \begin{bmatrix} PA_{t-1} \\ NA_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}}_{\text{exogenous}} Temp_t + \begin{bmatrix} \zeta_{PA,t} \\ \zeta_{NA,t} \end{bmatrix} \quad (2)$$

In Eq. 2, only positive affect (PA) and negative affect (NA) are dependent (endogenous) variables on the left-hand side of the equation. The first term of the equation includes all endogenous variables and their associated weights, in which there is potential for both positive affect and negative affect to influence each other (note that the diagonal must be set to zero since a variable cannot predict itself contemporaneously). Here, only positive affect is predicting negative affect at time t as indicated by the presence of A_{21} . The second term represents the lagged influences of the endogenous variables on themselves. The diagonal of the Φ matrix contains the autoregressive effect estimates, or how an exogenous variable predicts itself at the next time point. Finally, the exogenous variable of temperature and associated coefficient weights estimating the prediction of PA and NA based on temperature values (controlling for the other variables in the equation). This relationship cannot be reversed; temperature may be found to significantly relate to

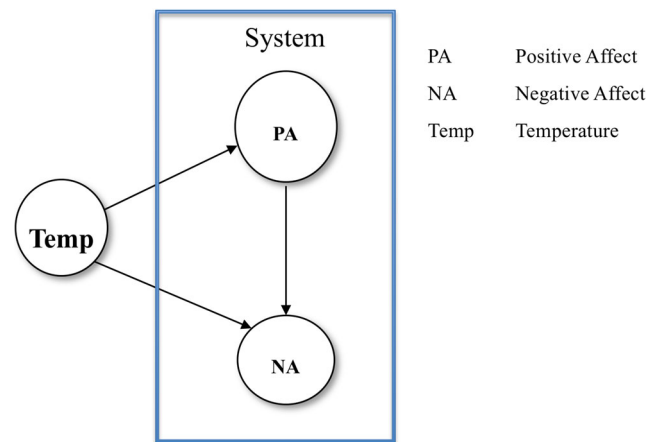


Fig. 1 Example of representation of an exogenous variable in a path diagram. Temperature is “outside the system” and can predict positive and negative affects but cannot be predicted by any variables inside the system. Autoregressive relations for the endogenous variables and error/variance depictions omitted for clarity.

PA and NA after controlling for other variables but it is not possible here for PA or NA to predict the temperature.¹ The final term contains the equation errors for that time point. Figure 1 represents (2) as a path diagram, where temperature exists outside the system of positive and negative affect (shown inside a rectangle).

Modeling treatment effects

Just as modeling weather as an exogenous variable achieves a more realistic model search space, we can also model treatment status as an exogenous variable. In most contexts of randomized controlled experiments, participant data does not affect what treatment they are receiving or whether they are currently receiving treatment or not. For this reason, we would not want treatment status to be treated as endogenous. Some approach this problem by modeling pre-treatment, treatment, and post-treatment separately and examine structural differences in the resulting network patterns of relations. Specifying treatment status as an exogenous variable in the model allows for an alternative approach where we can examine an individual's dynamic processes while controlling for treatment status. Treatment status would be represented in Eq. 2, where \mathbf{u} would be a variable representing treatment status, dummy coded to represent whether the participant was receiving treatment at each time point.

¹Although not displayed here, we may also decide to include autoregressive effects of exogenous variables so that they may predict themselves at subsequent time points. We explore this option in our first empirical example.

Accounting for non-stationarity

Allowing for specification of exogeneity prior to model search also allows for inclusion of time-related trends in the search space. Trend stationarity, or constant mean over time, is an assumption of time series analysis (Shumway, 2003). One can correct for violations of this assumption by including a “time” variable in \mathbf{u} . This time variable can be represented as a linear or non-linear process (Molenaar et al., 1992). In its simplest form, time is represented as a vector numbered from 1 to T , where T is the number of time points. However, a range of non-linear processes could be modeled, such as weekly effects or other anticipated cyclical effects. It is possible to remove trends prior to analysis but as others have recommended (e.g., Molenaar et al., 1992), it may not be desirable to remove these trends if the purpose of the research is to look at change over time. By controlling for possible trends directly in the analysis we are able to identify their influence on specific variables above and beyond the influence of other variables, which may be of interest in itself. In Eq. 2, u would be the exogenous time vector.

Including interaction effects

Finally, allowing for exogeneity in model search allows us to model the moderating effect of exogenous variables on endogenous processes (i.e., interaction effects). Including these terms in the uSEM model results in extended uSEM (euSEM) (Gates et al., 2011). Extending our ongoing example, we could model the moderating effect of endogenous variable positive affect on the relationship between temperature and negative affect. We can update (1) with a term representing this interaction effect for a single interaction effect, shown in Eq. 3, where $\mathbf{u}_t * \eta_{n,t}$ indicates multiplied terms for a given endogenous variable η_n with the exogenous variable u and τ is an $N \times 1$ vector of estimates for the moderated influence of an exogenous variable and another variable.

$$\eta_t = \mathbf{A}\eta_t + \Phi\eta_{t-1} + \gamma\mathbf{u}_t + \tau\mathbf{u}_t\eta_{n,t} + \zeta_t \tag{3}$$

Adding an interaction term to our continuing example, provides: Eq. 4.

$$\begin{bmatrix} PA_t \\ NA_t \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 \\ a_{21} & 0 \end{bmatrix}}_{\text{endogenous}} \begin{bmatrix} PA_t \\ NA_t \end{bmatrix} + \underbrace{\begin{bmatrix} \phi_{11} & 0 \\ 0 & \phi_{22} \end{bmatrix}}_{\text{lagged}} \begin{bmatrix} PA_{t-1} \\ NA_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}}_{\text{exogenous}} Temp_t \tag{4}$$

$$+ \underbrace{\begin{bmatrix} 0 \\ \tau_2 \end{bmatrix} Temp_t * PA_t}_{\text{bilinear}} + \begin{bmatrix} \zeta_{PA,t} \\ \zeta_{NA,t} \end{bmatrix}$$

Note that we selected one variable, PA, to see how its interaction with temperature changed the relationship between NA and temperature. Of course, we could have selected NA as the moderator instead or in addition to PA as a moderator; this decision is left for the user when conducting the analysis within the *gimme* package. Given that the number of variables can quickly increase due to multiplying terms, it is recommended that the researcher indicate the features that they think would best be moderators in their data based on theory and relevant questions.

Another decision point for researchers is whether or not to have the interaction effects be lagged or contemporaneous. In this example the interaction terms are contemporaneous. This is in line with work on daily diary data. The low temporal resolution of daily diary data often results in significant contemporaneous relationships but not lagged relationships (e.g., Lane et al., 2019a). Much like the constraint described in the \mathbf{A} matrix, a variable in a contemporaneous moderation term cannot predict itself at the same time. Hence $Temp_{.t} * PA_t$ cannot predict PA_t .

Method

Using three empirical examples, we demonstrate the utility of allowing for exogeneity and moderating effects using the *gimme* R package. Our aims are: 1. To demonstrate model searches with *a priori* specification of exogenous variables, trend-stationarity, treatment status, and interaction effects within the *gimme* framework; and 2. complete between-person exploratory analysis using LASSO, demonstrating how we can generate hypotheses about the relationship between obtained beta coefficients and trait-level characteristics.

GIMME

GIMME is a data-driven model search algorithm that hitherto has only offered the uSEM framework in the freely distributed R package (R Development Core Team 3.0.1. 2013) *gimme* (Lane et al., 2019a). That is, it did not allow for exogenous variables or euSEMs. GIMME has many benefits over other algorithms as it allows for dynamic modeling of idiographic, individual-level processes as well as subgroup and group-level processes. Additionally, it allows for the search of both contemporaneous and lagged relationships, allowing for directed relationships between variables, making it advantageous for specifying exogenous relationships (Gates et al., 2011, 2012). Details of the

process of the algorithm can be found in Gates and Molenaar (2012) and have also been presented elsewhere (Lane et al., 2019b; Beltz & Gates, 2017). Finally, *gimme* has been found to work well for a range of data, such as daily diary (Wright et al., 2018) and fMRI data (Mumford & Ramsey, 2014; Gates & Molenaar, 2012). Recent extensions allow for the option for unsupervised classification of individuals into subgroups comprised of people who have similar dynamic processes (Gates et al., 2017; Lane et al., 2019b). We use this option in the present examples, as it has been demonstrated to improve the reliability of the path search procedure (Lane et al., 2019a). Figure 2 provides an example summary output of the original *gimme* package demonstrating the result of the model search. The summary plot shows recovered individual, subgroup, and group level processes, as well as directed contemporaneous and lagged relationships. The summary plot serves as a visual starting point for the researcher to explore recovered relationships in the data. Greater detail is provided in additional output, as described in the *gimme* documentation and in tutorials on the *gimme* website <http://gimme.web.unc.edu/63-2/output/>. Tutorials include instruction on preparing the data and environment, running GIMME, GIMME basics, and interpreting output.

Here, we present for the first time, an extension of the model search procedure within *gimme* so that the user can specify variables in their data set that are exogenous. *gimme* then removes the potential relations from the model search space where exogenous variables are being predicted by other variables. Additionally, the user is allowed to indicate the lag order of the exogenous variables. Returning to the

running example, one might hypothesize that weather has a lagged influence with the endogenous variables in addition to a contemporaneous relation. We note that allowing for an autoregressive effect would mean that the variable is conceptually exogenous but not statistically exogenous in the framework of uSEM. By conceptually exogenous, we mean that the variable at time $t-1$ can predict itself at time t , but the variable cannot be predicted by other variables. We, however, allow for this option in *gimme* in case modeling of the AR process is of interest to researchers. Importantly, in some cases, such as having a time vector from $t = 1, 2, \dots, T$, including the lagged variable will introduce high multicollinearity as they will be perfectly correlated. We do not recommend allowing for an autoregressive effect on variables such as these. Regarding the interaction terms, although *gimme* is a data-driven algorithm it is advised that the selection of variables to be multiplied be informed by prior research or hypotheses. If the researcher believes weather may have a lagged influence on NA and PA, they may wish to have the interaction term for temperature times PA to be at $t - 1$ rather than t . These options are available to the user and should depend on prior theory and present questions that are interesting to the researcher.

Data

We demonstrate three examples of exogenous capabilities in *gimme* using two datasets. Example 1 comes from daily diary data collected on individuals ($N=112$) diagnosed with personality disorders (Wright & Simms, 2016). We removed participants whose individual-level models could

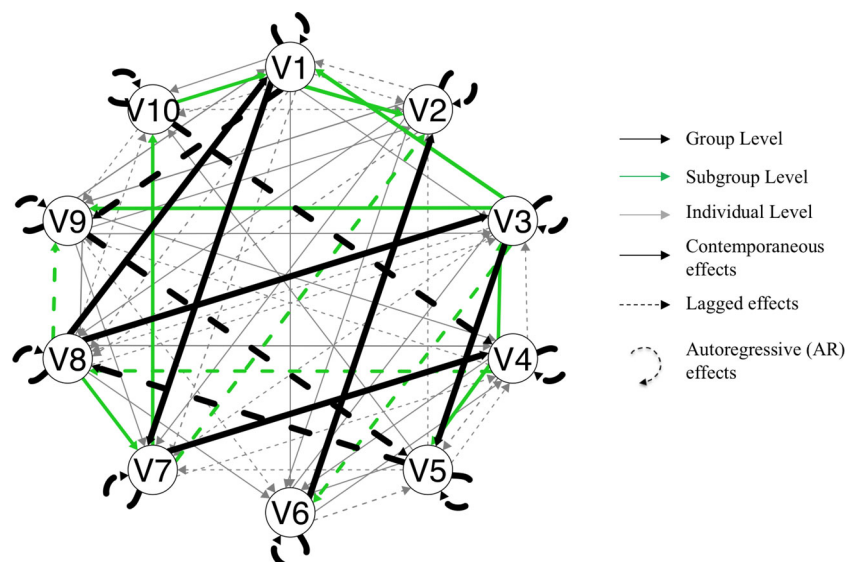


Fig. 2 Example summary plot from *gimme* model search. Group-level paths (i.e., paths that improve model fit for a majority of individuals) are in *black*, subgroup-level paths in *green*, and individual-level paths in *gray*. Lagged paths are indicated by a *dashed line*, and contemporaneous paths are indicated by a *solid line*.

not be estimated due to zero variance in one or more variables or an inadequate number of observations ($T < 60$; Lane et al., 2019a). These restrictions are common for any time series SEM model. The final sample size was $N = 94$ individuals. Participants completed daily reports on emotions and behaviors which were converted to factor scores for the following dimensions: negative affect (NA), positive affect (PA), dominance, and love. *gimme* now allows for modeling of latent variables by estimating factor scores directly within the package prior to conducting the model search (Gates et al., 2019). A breadth of research has explored the relationships between weather and affect (Sarran et al., 2010; Barnston, 1988; Geoffroy et al., 2014). In following this work, we obtained historical weather data from the National Weather Service, choosing to focus on daily temperature recordings, measured in Fahrenheit. Participants also completed trait-level reports at the beginning of the study. Here, we focus on depressivity and investigate which aspects of the dynamic process among NA, PA, dominance, and love for individuals related to their baseline depressivity levels.

Examples 2 and 3 come from daily diary data collected on middle-aged individuals ($N=226$; ages 35-64) completing a 6-week meditation workshop in which participants were assigned to either learn loving-kindness meditation (LKM) or an active control group, mindfulness meditation (MM) (Fredrickson et al., 2017). We, again, removed participants whose individual-level models could not be estimated due to zero variance in one or more variables or an inadequate number of observations and ultimately completed analysis on $N = 92$ individuals. Participants reported daily on affect and health behaviors pre-, during, and post-workshop for a total of 11 weeks of daily reports. We specifically looked at summed scores of NA and PA from the modified Differential Emotions Scale (mDES) (Fredrickson, 2013), amount of time engaged in physical exercise, amount of time engaged in meditation, and number of fruits and vegetables consumed. Summed scores have been shown to be as reliable as factor scores in the *gimme* framework (Gates et al., 2019). The exogenous variable here was whether or not they were actively in the workshop. The 2 weeks of pre-workshop, as well as the 3 weeks post-workshop were coded as zero, while the 6 weeks of workshop were coded as one.

Analysis

Three separate *gimme* analyses were conducted for each empirical example in R 3.5.0 (R Development Core Team 3.0.1. 2013), using the latest version of *gimme*, available at <https://github.com/GatesLab/gimme> (Lane et al., 2019a). In all examples, autoregressive effects among the endogenous variables were set to be estimated for all individuals. In the first example, we also allow the exogenous variable,

temperature, to have an autoregressive effect. The group level cutoff argument was set to 0.51. That is, for a path to be recovered for the entire group, addition of the path to an individuals' model must result in model improvement for a majority of individuals defined as at least 51% here.

In each example, after obtaining each individuals' model, LASSO was performed using the GLMNET package (Friedman et al., 2010) in R. The predictor variables were the individuals' beta estimates for each path (i.e., estimates in the \mathbf{A} , Φ , Γ , and τ matrices). The dependent variable was a trait-level variable (level of depressivity in the case of Example 1) or experimental group (likelihood that individual is in LKM or MM in the case of Examples 2 and 3). A similar approach was used with daily diary data (Woods et al., 2020b). Due to a binary outcome in Examples 2 and 3, logistic LASSO regression was performed. We present this as an exploratory only, hypothesis generating mechanism for researchers. By searching for relationships between individual-level dynamics and trait-level outcomes, researchers can use the GIMME-obtained individual-level models to guide future research.

Results

Example 1

Here, we demonstrate the handling of trend non-stationarity by including a time variable for each person in the model search. We also demonstrate specification of an exogenous variable as part of the model search by including outdoor temperature, measured in Fahrenheit, as a variable corresponding to the temperature across the days in which the participant completed the survey. In *gimme*, these variables are designated as exogenous by naming them in the argument, "exogenous" (see Appendix for example code). Doing so ensures that paths treating temperature or time as a dependent variable will not be included in the model. Lagged and autoregressive effects were included in the search space for temperature but not time. This was because we included the lagged variable for temperature to account for its possible influence on other variables, whereas the lagged time linear variable would be highly collinear with the original one resulting in an unidentifiable model. An example of each exogenous variable for one individual is shown below.

$$Temp = \begin{bmatrix} 56 \\ 58 \\ 60 \\ \vdots \\ 60 \end{bmatrix}, Time = \begin{bmatrix} 1 \\ 2 \\ 3 \\ \vdots \\ 98 \end{bmatrix}$$

No group level paths were found (except autoregressive effects which were set to be estimated for everyone). Twenty subgroups were found, 15 of which were singletons, meaning that these individuals were placed in subgroups by themselves. This tends to indicate they are dissimilar to those in the sample in terms of their dynamic processes. A summary plot of results is displayed in Fig. 3. A variety of subgroup level paths were found across the five subgroups that had more than one individual. Plots broken down by subgroup are displayed in Fig. 4.

Subgroup 1 ($n = 8$) had one subgroup-level path: negative affect predicting love (mean $\beta = 0.59$, $SD = 2.56$). Subgroup 2 ($n = 20$) had two subgroup-level paths: positive affect predicting love (mean $\beta = 0.37$, $SD = 0.22$) and dominance predicting positive affect (mean $\beta = 0.46$, $SD = 0.18$). Subgroup 3 ($n = 22$) had two subgroup-level paths: negative affect predicting love (mean $\beta = -0.59$, $SD = 0.22$) and positive affect predicting love (mean $\beta = 0.23$, $SD = 0.29$). Subgroup 5 ($n = 6$) had one subgroup-level path: love predicting positive affect (mean $\beta = 0.43$, $SD = 0.39$). Subgroup 4 ($n = 23$) had no subgroup-level paths.

Coefficients from LASSO are displayed in Table 1. Errors were minimized at a penalization of 0.09. The R^2 of the penalized model was 0.18. Lagged paths (beta coefficients) found to contribute to the variance of trait-level depressivity were the autoregressive effect for negative affect, the autoregressive effect for positive affect, the lagged effect of dominance predicting positive affect, and the lagged effect of temperature predicting positive affect. Contemporaneous paths (beta

coefficients) found to contribute to the variance of trait-level depressivity were negative affect predicting dominance, positive affect predicting dominance, dominance predicting positive affect, dominance predicting love, and love predicting negative affect. More positive beta weights for the autoregressive effect of negative affect, the lagged effect of dominance predicting positive affect, the lagged effect of temperature predicting positive affect, negative affect predicting dominance, dominance predicting positive affect, and dominance predicting love were associated with higher depressivity. More negative beta weights for the autoregressive effect of positive affect, positive affect predicting dominance, and love predicting negative affect were associated with higher depressivity.

Example 2

In Example 2, treatment status, coded as one for currently active in the meditation workshop and zero for pre- or post-workshop was specified as an exogenous variable. Again, this variable is specified as exogenous in *gimme* through the argument “exogenous.” An example of the variable is below:

$$Treatment = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

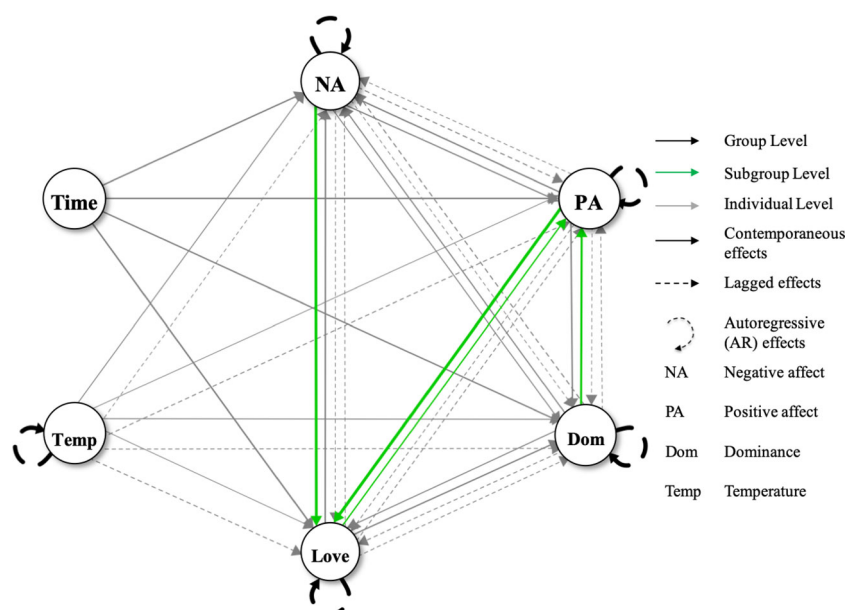


Fig. 3 Summary plot showing results from Example 1. No group-level paths were recovered. Subgroup level paths are indicated in *green*, and individual level paths are indicated in *gray*. *Dashed lines* represent a lagged effect, and *solid lines* represent a contemporaneous effect.

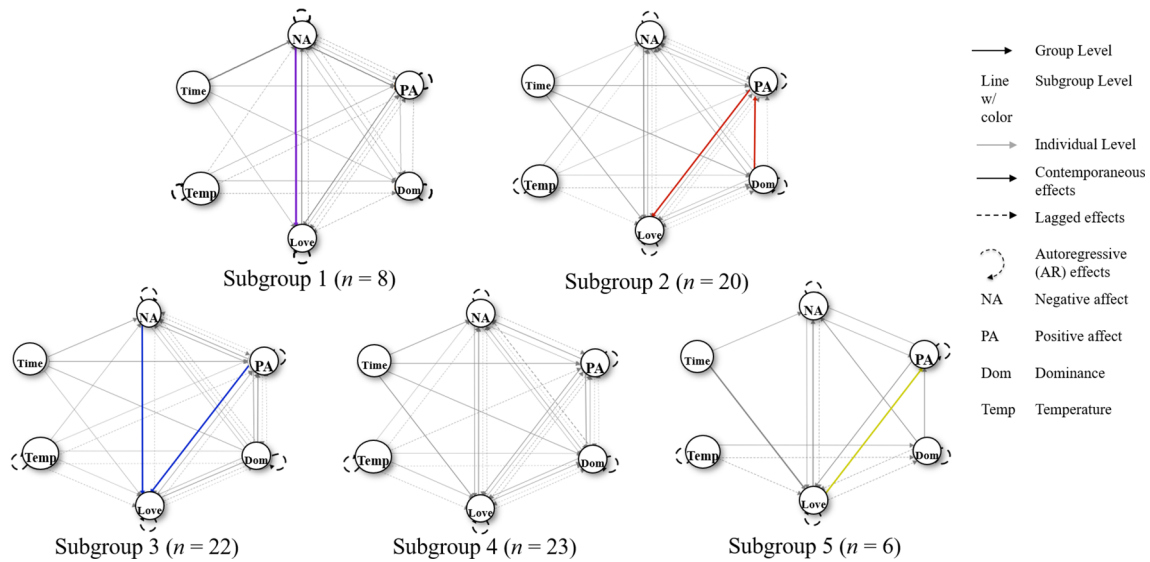


Fig. 4 Summary plot for Example 1 broken down by subgroup. Colored lines indicate paths found at the subgroup level for each subgroup. *n* indicates the number of individuals assigned to each subgroup

A group level path was found where negative affect predicts positive affect. The average path weight for this group level path was $-0.41 (SD = 0.26)$, meaning that on average, increases in negative affect predicted decreases in positive affect, contemporaneously. Seven subgroups were found, five of which were singletons. No subgroup level paths were found for either of the two subgroups. A summary of results is displayed in Fig. 5.

Non-zero coefficients from LASSO are displayed in Table 1, identifying potential paths of interest for distinguishing the two conditions in the study, loving-kindness meditation (LKM) and mindfulness meditation (MM). Errors were minimized at a penalization (λ) of 0.09. Paths found to contribute to the likelihood of being in the LKM group were the autoregressive path for exercise and

the contemporaneous paths of time spent exercising predicting number of fruits and vegetables consumed, treatment status predicting positive affect, and treatment status predicting time spent exercising. For all paths found to be relevant except treatment status predicting time spent exercising, more positive beta weights were related to a higher likelihood of being in the LKM group. For treatment status predicting time spent exercising, more negative beta weights were related to being in the LKM group.

Example 3

In Example 3, the interaction of treatment status and positive affect was additionally specified as an exogenous variable. In *gimme*, this is provided to the argument “mult_vars.”

Table 1 Results from adaptive LASSO. PA = positive affect, NA = Negative affect, DOM = dominance, LOV = love, TEMP = temperature, FV = fruits and vegetables, EX = exercise, MED = time spent meditating, ACT = active in workshop

Example 1	Depressivity <i>b</i>	Example 2	LKM <i>b</i>
NA _{lag} ->NA	1.07	EX _{lag} ->EX	-0.26
PA _{lag} ->PA	0.66	EX- >FV	1.08
DOM _{lag} ->PA	-0.14	ACT- >PA	0.15
TEMP _{lag} ->PA	0.93	ACT- >EX	0.04
NA->DOM	0.29		-1.77
PA->DOM	0.14		
DOM->PA	-0.43		
DOM->LOV	0.06		
LOV->NA	0.16		
	-0.12		

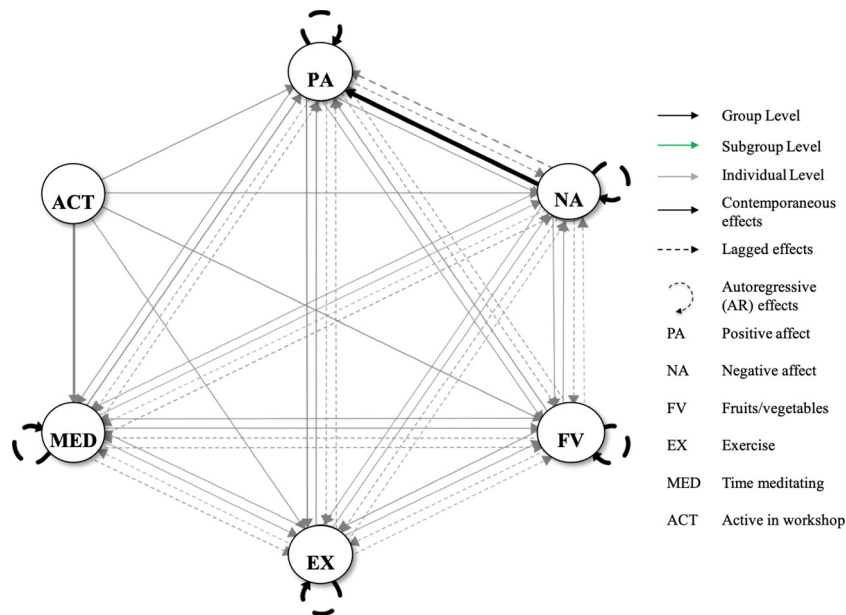


Fig. 5 Summary plot showing results from Example 2

We again found one group-level path, negative affect predicting positive affect contemporaneously. The average path weight for this path remained negative ($mean = -0.64, SD = 1.12$). Eight subgroups were found, five of which were singletons. A summary of results is displayed in Fig. 6.

There were no non-zero coefficients using adaptive LASSO. Errors were minimized at a lambda of 0.12. Thus there is not strong evidence that specific dynamic relations relate to the trait-level qualities measured in this study.

Discussion

This novel extension of the *gimme* algorithm now allows for flexibility to specify exogenous variables, including treatment effects and interaction effects prior to model search. This paper also presents a novel use of extended uSEM (Gates et al., 2011) by employing it in a model search algorithm. *Gimme* is particularly suited for this as the uSEM framework allows for directed contemporaneous and lagged effects. Additionally, *gimme* allows for the

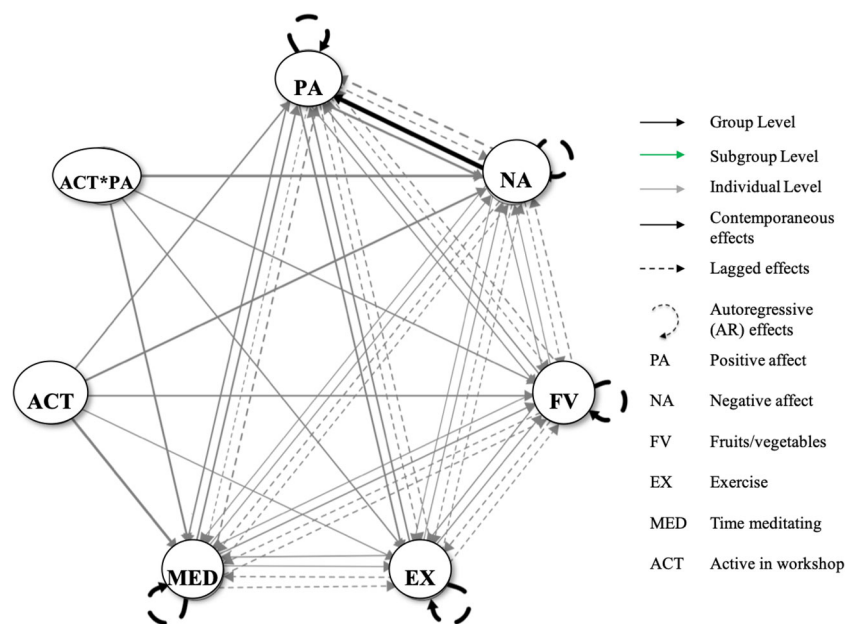


Fig. 6 Summary plot showing results from Example 3

modeling of idiographic processes, finding paths at the individual, subgroup, and group level. In the three examples presented, we successfully performed model searches where variables identified as exogenous or interaction effects could only predict but not be predicted by other variables in the model. Finally, results from *gimme* can be used to generate hypotheses for testing. Here, we demonstrated using LASSO as an exploratory method for finding what paths found in the model search process were related to trait-level outcomes or experimental condition.

Importance of considering exogeneity in model searches

Model searches are a powerful tool for hypothesis generation and especially useful for finding idiographic processes in data where many individuals have many time points. However, many model search procedures are agnostic to the plausible roles that a variable may play in a system of relationships. In particular, exogenous variables which are a priori known to be exogenous present a problem in model searches if the model search algorithm cannot model that variable realistically. Considering exogeneity in model searches in the uSEM framework allows not just for more realistic models, but also allows for handling of non-stationarity within the model search procedure, modeling of crossover treatment designs, and modeling of bilinear effects.

Bridging idiographic and nomothetic debate

In addition to emphasizing the importance of model searches that allow for exogeneity and demonstrating how we can successfully exclude implausible paths in *gimme*, we also demonstrate how intensive longitudinal data can be both used to understand idiosyncrasies in individuals and find commonalities among individuals. *Gimme* allows for this by allowing individuals to have idiosyncratic models in terms of existence of a path, direction of a path, and weight of a path, while also finding which paths support the best models for a majority of individuals. This allows us to avoid averaging out of effects.

We also demonstrate how we can generate nomothetic hypotheses from *gimme* results by performing LASSO where the outcome is some trait-level variable and the predictors are the beta weights found for each individual in *gimme*. While we do not encourage arriving at inferences from these results, we do see them as a way of generating hypotheses for later testing on newly collected data and understanding nuances of individual processes. For example, we may want to see if a lagged relationship between temperature and positive affect is related to

depressivity or we may want to test if greater connectivity between health behaviors like healthy eating and exercise are more likely to occur when participants receive training in LKM versus MM.

Future directions

We note that, in some cases, there may not be enough statistical power to determine that an exogenous variable belongs in the final model. This is especially true for interaction affects (Aguinis, 1995), where a greater sample size is needed for detection of significant moderators than for predictors. It is possible that a moderating effect exists but does not have enough power to be found significant and thus, included in the final model. Future work providing power analysis for the user may be useful in helping researchers make better informed conclusions from the final model. Additionally, the model framework used is sometimes prone to overfitting. Bulteel et al. (2018) The GIMME algorithm aims to avoid overfitting by only modeling relationships that are significant for a majority of individuals. Future work could investigate whether this is goal is achieved.

In the examples presented here, we model time as a linear process in Example 1 and treatment effects as binary in Example 2. It is possible that detrending, as performed in Example 1, is not best modeled as linear for some individuals but is for others. Similarly, some individuals may have a binary response to whether they are active in a treatment, while others may show a for example, a linear increase in some variable during the treatment with a linear decrease in that variable once treatment has ended. In the future, allowing for these differing patterns will be useful for better modeling idiosyncratic processes and comparing individuals in their processes. Additionally, many individuals could not be included in analysis because one or more of their variables did not have enough variability to have an estimable model. Future work on utilizing some aspect of these individuals' data would be informative.

None of the data or materials for the experiments reported here is available, and none of the experiments was preregistered.

Acknowledgments We gratefully acknowledge support from NIH National Institute of Biomedical Imaging and Bioengineering (R01EB022904), the National Cancer Institute (R01CA170128), and the National Institute of Health (MH097325).

Appendix: Example Code

Example 1 code where there are two exogenous variables (time and temperature)

```
myresults <- gimme(data = mydata,  
  out = "filename",  
  sep = ",",  
  header = TRUE,  
  ar = TRUE,  
  plot = TRUE,  
  subgroup = TRUE  
  exogenous = c("temperature", "time"))
```

Example 2 code where there is one exogenous variables (treatment status)

```
myresults <- gimme(data = mydata,  
  out = "filename",  
  sep = ",",  
  header = TRUE,  
  ar = TRUE,  
  plot = TRUE,  
  subgroup = TRUE  
  exogenous = c("active"))
```

```
### Example 3 code where there are two exogenous variables (treatment status),
one of which is bilinear (treatment status*positive affect)
```

```
myresults <- gimme(data = mydata,
  out = "filename",
  sep = ",",
  header = TRUE,
  ar = TRUE,
  plot = TRUE,
  subgroup = TRUE
  exogenous = c("active"),
  mult_vars = c("active*pa"))
```

References

- Aguinis, H. (1995). Statistical power with moderated multiple regression in management research. *Journal of Management*, 21(6), 1141–1158. <https://doi.org/10.1177/014920639502100607>
- Barnston, A. G. (1988). The effect of weather on mood, productivity, and frequency of emotional crisis in a temperate continental climate. *International Journal of Biometeorology*, 32(2), 134–143. <https://doi.org/10.1007/BF01044907>
- 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>
- Bringmann, L. F., Lemmens, L. H. J. M., Huibers, M. J. H., Borsboom, D., & Tuerlinckx, F. (2015). Revealing the dynamic network structure of the Beck Depression Inventory-II. *Psychological medicine*, 45, 747–757. <https://doi.org/10.1017/S0033291714001809>
- Bulteel, K., Mestdagh, M., Tuerlinckx, F., & Ceulemans, E. (2018). VAR (1) based models do not outpredict AR (1) models in current psychological applications. *Psychological Methods*, 23(4), 740–756.
- Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Mõttus, R., Waldorp, L. J., & Cramer, A. O. (2015). State of the aRt personality research: A tutorial on network analysis of personality data in R. *Journal of Research in Personality*, 54, 13–29. <https://doi.org/10.1016/j.jrp.2014.07.003>
- Cramer, O. J., Borkulo, C. D. V., Giltay, E. J., Han, L., Maas, J. V. D., Kendler, K. S., & Scheffer, M. (2016). Major depression as a complex dynamical system. (1500). [arXiv:1606.00416](https://arxiv.org/abs/1606.00416)
- Engle, R. F., Hendry, D. F., & Richard, J.-F. (1983). Exogeneity. *Econometrica*, 51(2), 277–304.
- Fredrickson, B. L. (2013). Positive Emotions Broaden and Build. *Advances in Experimental Social Psychology*, 47, 1–53. Retrieved from <https://doi.org/10.1016/B978-0-12-407236-7.00001-2%7B%5C%7D0A>
- Fredrickson, B. L., Boulton, A. J., Firestone, A. M., Van Cappellen, P., Algoe, S. B., Brantley, M. M., & Salzberg, S. (2017). Positive Emotion Correlates of Meditation Practice: a Comparison of Mindfulness Meditation and Loving-Kindness Meditation. *Mindfulness*. <https://doi.org/10.1007/s12671-017-0735-9>
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1–22. Retrieved from <http://www.jstatsoft.org/v33/i01/>
- 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. <https://doi.org/10.1016/j.neuroimage.2009.12.117>
- Gates, K. M., & Molenaar, P. C. (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>
- Gates, K. M., Molenaar, P. C., Hillary, F. G., & Slobounov, S. (2011). Extended unified SEM approach for modeling event-related fMRI data. *NeuroImage*, 54(2), 1151–1158. <https://doi.org/10.1016/j.neuroimage.2010.08.051>
- Gates, K. M., Lane, S., 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., Fisher, Z. F., & Bollen, K. (2019). A Latent Variable GIMME Using Model Implied Instrumental Variables (MIIVs). *Psychological Methods*, 1–50. <https://doi.org/10.1037/met0000229>
- Geoffroy, P. A., Bellivier, F., Scott, J., & Etain, B. (2014). Seasonality and bipolar disorder: A systematic review, from admission rates to seasonality of symptoms. *Journal of Affective Disorders*, 168, 210–223. <https://doi.org/10.1016/j.jad.2014.07.002>

- Haslbeck, J., Borsboom, D., & Waldorp, L. (2019). Moderated Network Models. *Multivariate Behavioral Research*, 1–32. <https://doi.org/10.1080/00273171.2019.1677207>, arXiv:1807.02877
- Horst, R. L. (2010). Exogenous versus endogenous recovery of 25-hydroxyvitamins D2 and D3 in human samples using high-performance liquid chromatography and the DiaSorin LIAISON Total-D Assay. *Journal of Steroid Biochemistry and Molecular Biology*, 121(1–2), 180–182. <https://doi.org/10.1016/j.jsbmb.2010.03.010>
- Kilian, L. (2008). A comparison of the effects of exogenous oil supply shocks on output and inflation in the G7 countries. *Journal of the European Economic Association*, 6(1), 78–121.
- Kim, J., Zhu, W., Chang, L., Bentler, P. M., & Ernst, T. (2006). 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., Gates, K. M., Fisher, Z., Arizmendi, C., Molenaar, P., Hallquist, M., & Beltz, A. (2019a). *gimme*: Group Iterative Multiple Model Estimation. Retrieved from <https://github.com/GatesLab/gimme/>
- Lane, S., Gates, K. M., Pike, H. K., Beltz, A. M., & Wright, A. (2019b). Uncovering General, Shared, and Unique Temporal Patterns in Ambulatory Assessment Data. *Psychological Methods*, 7(4). <https://doi.org/10.1037/met0000192>
- Molenaar, P. C., De Gooijer, J. G., & Schmitz, B. (1992). Dynamic factor analysis of nonstationary multivariate time series. *Psychometrika*, 57(3), 333–349. <https://doi.org/10.1007/BF02295422>
- Mumford, J. A., & Ramsey, J. D. (2014). Bayesian networks for fMRI: A primer. *NeuroImage*, 86, 573–582. <https://doi.org/10.1016/j.neuroimage.2013.10.020>
- R Development Core Team 3.0.1. (2013). A language and environment for statistical computing, Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.r-project.org>
- Sarran, C., Sachon, P., & Meesters, Y. (2010). Meteorological analysis of symptom data for people affected with SAD.pdf. *Psychiatry Research*, 257, 501–505. <https://doi.org/10.1016/j.psychres.2017.08.019>
- Schmittmann, V. D., Cramer, A. O. J., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology*, 31(1), 43–53. <https://doi.org/10.1016/j.newideapsych.2011.02.007>
- Shumway, R. H. (2003). Time-frequency clustering and discriminant analysis. *Statistics and Probability Letters*, 63(3), 307–314. [https://doi.org/10.1016/S0167-7152\(03\)00095-6](https://doi.org/10.1016/S0167-7152(03)00095-6)
- Woods, W. C., Arizmendi, C., Gates, K. M., Stepp, S., Pilkonis, P., & Wright, A. (2020). Personalized Models of Psychopathology as Contextualized Dynamic Processes: An Example from Individuals with Borderline Personality Disorder. *Journal of Consulting and Clinical Psychology*. Retrieved from <https://europepmc.org/article/med/32068425>
- Woods, W. C., Arizmendi, C., Gates, K. M., Stepp, S. D., Pilkonis, P. A., & Wright, A. G. (2020). Personalized models of psychopathology as contextualized dynamic processes: An example from individuals with borderline personality disorder. *Journal of Consulting and Clinical Psychology*, 88(3), 240–254. <https://doi.org/10.1037/ccp0000472>
- Wright, A. G. C., & Simms, L. J. (2016). Stability and Fluctuation of Personality Disorder Features in Daily Life Stability and Fluctuation of Personality Disorder Features in Daily Life. *Journal of Abnormal Psychology*, 125(5), 641–656.
- Wright, A., Gates, K. M., Arizmendi, C., Lane, S., Woods, W. C., & Edershile, E. A. (2018). Focusing personality assessment on the person: Modeling General, Shared, and Person Specific Processes in Personality and Psychopathology. *Psychological Assessment*. <https://doi.org/10.17605/OSF.IO/NF5ME>
- Yin, Y. (2010). Exogenous variables. In Salkind, N. J. (Ed.), *Encyclopedia of research design* (pp. 1499–1503). <https://doi.org/10.4135/9781412961288>. arXiv:1011.1669v3

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.