

Running head: LV-GIMME

Latent Variable GIMME Using Model Implied Instrumental Variables (MIIVs)

### Abstract

Researchers across many domains of psychology increasingly wish to arrive at personalized *and* generalizable dynamic models of individuals' processes. This is seen in psychophysiological, behavioral, and emotional research paradigms, across a range of data types. Errors of measurement are inherent in most data. For this reason, researchers typically gather multiple indicators of the same latent construct and use methods, such as factor analysis, to arrive at scores from these indices. In addition to accurately measuring individuals, researchers also need to find the model that best describes the relations among the latent constructs. Most currently available data-driven searches do not include latent variables. We present an approach that builds from the strong foundations of Group Iterative Multiple Model Estimation (GIMME), the idiographic filter, and model implied instrumental variables with two-stage least squares estimation (MIIV-2SLS) to provide researchers with the option to include latent variables in their data-driven model searches. The resulting approach is called *Latent Variable GIMME* (LV-GIMME). GIMME is utilized for the data-driven search for relations that exist among latent variables. Unlike other approaches such as the idiographic filter, LV-GIMME does not require that the latent variable model to be constant across individuals. This requirement is loosened by utilizing MIIV-2SLS for estimation. Simulated data studies demonstrate that the method can reliably detect relations among latent constructs, and that latent constructs provide more power to detect effects than using observed variables directly. We use empirical data examples drawn from functional MRI and daily self-report data.

*Keywords:* SEM, GIMME, latent variables, idiographic filter, MIIV-2SLS **Latent-Variable**

**GIMME**

The frequent and increasing use of time series data in psychological research, such as psychophysiological, in-vivo self-report, and passive data collection from digital devices, provides researchers with newfound possibilities to model behavioral, emotional, and cognitive processes as they unfold uniquely for each individual. With these data researchers can quantify the interplay of individual-level attributes (e.g., moods; behaviors; brain processes) in real time to better understand if, for a given individual, they co-occur or if one precedes the other. Time series data have been used to demonstrate heterogeneity in many processes when looking across individuals, such as brain functioning (Laumann et al., 2015; Price, Gates, Kraynak, Thase, & Siegle, 2017), social behaviors (Beltz, Beekman, Molenaar, & Buss, 2013), and symptoms revealed via self-reports (Bringmann, Ferrer, Hamaker, Borsboom, & Tuerlinckx, 2018; Rubel, Fisher, Husen, & Lutz, 2018). Importantly, these data help to bridge the gap between idiographic and nomothetic research (Beltz, Wright, Sprague, & Molenaar, 2016; Wright & Hopwood, 2016), while also moving the field closer towards the use of individual-level analysis to inform clinical psychological practice (Fernandez, Fisher, & Chi, 2017; Fisher & Boswell, 2016).

At this nascent stage of quantitative discovery into human processes researchers often do not have clear hypotheses driving what might be expected to hold true for all individuals (i.e., the nomothetic level) let alone the appropriate model for a given individual (i.e., idiographic or personalized level). Further hindering progress is the lack of robust methods for analyzing this new type of data. There has been success in developing data-driven methods for arriving at individual-level models of processes among observed variables (Epskamp et al., 2018; Gates & Molenaar, 2012; Henry & Gates, 2017; Ramsey, Hanson, & Glymour, 2011). A remaining analytic hurdle to the model-building process is that the actual measurement of underlying latent constructs may differ across individuals (Hamaker, Dolan, & Molenaar, 2005; Molenaar &

Campbell, 2009; Nesselroade, Gerstorf, Hardy, & Ram, 2007). Methods are needed that allow for variation in the measurement model (also referred to elsewhere as “factor model”) as well as the temporal relations among these latent variables (i.e., the “latent variable model”). The present paper introduces a person-specific model building approach, Latent-Variable Group Iterative Multiple Model Estimation (LV-GIMME) that is optimal for use on time series data such as those often obtained in psychological studies. LV-GIMME operates from within a structural equation modeling (SEM) framework for dynamic factor analysis (DFA; Molenaar, 1985) and uses model-implied instrumental variables with two-stage least squares (MIIV-2SLS; Bollen, 1996) for estimation. The synthesis of these two lines of research has been provided in the development of the MIIV-2SLS for estimating dynamic factor models at the individual level has already been provided by Fisher, Bollen and Gates (2019).

GIMME (Gates & Molenaar, 2012) is a data-driven method for arriving at patterns of directed relations among variables across time. Developed initially for use with functional MRI data, the algorithm is increasingly used with behavioral and psychological data gathered across time. It is an appropriate approach for time series data of at least 60 observations per person (Lane, Gates, Pike, Beltz, & Wright, 2017) and fewer than 25 variables. At the core of the GIMME algorithm is the notion that individuals may have some common and unique aspects of their dynamic processes. To avoid the risk of overfitting by conducting solely individual-level model searches, GIMME begins by first identifying which relations among variables, if any, exist for the majority of participants as indicated by significance tests of the modification indices (or Lagrangian multiplier diagnostics, see Sörbom, 1975?) for candidate relations. Importantly, GIMME does not simply aggregate across people; instead it only considers a path as existing for the majority of individuals if it is significant for the majority of individuals. Following the group-

level search, which serves to detect signal from noise, the algorithm uses these relations that were typically found across individuals in the sample as the starting point for individual-level searches. Again, modification indices guide this search. The model search stops for each individual when fit indices indicate adequate model fit, thus favoring parsimony. This method of first detecting signal from noise to obtain starting models closer to the individuals' true models has been shown to provide more reliable results than methods which conduct searches on the individual level (Gates & Molenaar, 2012). Similar to many directed graph searches, GIMME only has been evaluated for use with observed variables.

Integrating a measurement model component into GIMME requires sensitivity to the possibility that individuals may differ in the structure and/or estimates. At its most extreme, some analytic approaches require that all individuals have the same measurement and latent variable models when modeling dynamic processes. This is seen in the vertical concatenation of individuals' time series data or averaging covariance matrices prior to analysis. For instance, in functional MRI (fMRI) brain research it is common to concatenate individuals' data (or equivalently average covariance matrices) to arrive at one sample-level data set.<sup>1</sup> Researchers here are often interested in assessing how brain regions covary together across time. For this reason, a standard approach is to conduct Principal Components Analysis (PCA) on time series data that have been concatenated across individuals to arrive at latent components that reflect aspects of thought. These latent constructs are referred to as "brain networks". Brain networks represent disparate brain regions that tend to covary across time together. As an example, the Fronto-Parietal Network is a collection of brain regions that tend to covary during tasks that

---

<sup>1</sup> It must be noted that this approach is largely unnecessary since typically each individual in these studies provides ample numbers of time points to be considered at the individual level (i.e.,  $T = 200-1000$  observations for each time window of interest).

require attention. The same brain networks have been consistently found (to some extent) across individuals (Wisner, Atluri, Lim, & MacDonald, 2013), suggesting that perhaps aggregating across individuals to arrive at a model of latent components may be appropriate here.

Still, individuals vary in the estimates of the relations comprising these components (as seen in Gonzalez-Castillo et al., 2015), and analytic methods that incorporate this level of personalization are needed. Similarly, psychological studies using daily self-reports typically use scales that were originally developed for cross-sectional data. From this, they arrive at measures of latent constructs for individuals across time and assume that individuals can be measured the same. Since the inception of the use of time series data in psychology (such as daily self-reports) it has been demonstrated that individuals can vary in their measurement model structures (Lamiell, 1981; Molenaar, 2004; Molenaar & Campbell, 2009). More recent work highlights that reliability for scales across people (i.e., cross-sectionally) does not inform the reliability of scales across time for one individual such as in time series data (Fuller-Tyszkiewicz et al., 2017; Hu et al., 2016). All this suggests that personalized assessments via individual-level measurement models are needed.

There is a precedent for allowing personalized measurement models. Nesselroade and colleagues (2007) developed the concept of an Idiographic Filter. The idiographic filter allows for the measurement model structure and factor loadings to vary across individuals. Molenaar and Nesselroade (2012) extended the idiographic filter for use with dynamic factor models (DFM; Molenaar, 1985), which allows for directed relations among the latent variables. Critically, these efforts require that the latent variable model, defined here as the structure of relations among the latent variables, remain invariant across individuals both in terms of the pattern and estimates. However, it is possible (and indeed likely) that individuals vary both in the

measurement of constructs as well as the relations among constructs. As a motivating example, depression has been found to relate to anxiety in cross-sectional studies (Sartorius, Üstün, Lecrubier, & Wittchen, 1996). However, how depression relates to anxiety may differ across individuals when looking across time. For one person depression may be related to feeling anxiety the day before, suggesting that their depressive symptoms are a consequence of anxiety. For another person it may be opposite, with their depression levels predicting anxiety on subsequent days. Hence for both people, the latent variable models for “depression” differ. Yet another person could have no temporal relation between depression and anxiety and have their depression caused by an entirely different mechanism. Allowing for models where the presence and directions of relations among latent variables can differ will better assist researchers and practitioners in identifying models that truly describe the individuals.

This paper demonstrates that the requirement that the latent variable model must be constant across individuals, dictated by the idiographic filter can be loosened when utilizing MIIV-2SLS for estimation. MIIV-2SLS estimation has a number of benefits that are outlined in detail below. The most noteworthy advantage for this problem is that when using MIIV-2SLS the structure and estimates obtained in the latent variable model have no impact on the parameter estimates in the measurement model (Bollen, Gates, & Fisher, 2018). Thus, analytically it is irrelevant if individuals differ in their latent variable models. Specifically, we could hold all latent variable model structures and estimates to be constant across individuals (as is done in the idiographic filter) or not; either way, the measurement model parameter estimates for each individual remains unchanged. This is a stark contrast to full information estimators such as Maximum Likelihood (ML), wherein values at the latent variable model level may influence the estimates of the measurement model coefficients. In utilizing MIIV-2SLS as one possibility for

estimating latent scores across time, LV-GIMME allows for individual-level relations in all aspects of the DFM without sacrificing meaning and interpretability of the models.

The structure of the paper is as follows. Throughout the paper we utilize a fMRI data example. Following a brief introduction to this data set we describe the integration of latent variable models into GIMME using multiple approaches for quantifying the latent construct to create LV-GIMME. LV-GIMME is a data-driven method that arrives at group, subgroup (if desired), and individual-level relations among latent variables. Importantly, this approach allows for individuals to have unique measurement models, both in terms of structure (qualitative) and weights (quantitative). We then utilize simulations to demonstrate acceptable recovery rate of group- and individual-level paths. These simulations are informed by the fMRI results as well as additional conditions. Since MIIV-2SLS offers a number of benefits in addition to those already noted, the final step is to estimate the individual-level models with this estimator. We provide details relevant for the present purpose as well as outline the benefits. Finally, we demonstrate the method with a previously used empirical data example described in Lebo & Nesselrode (1978). Throughout this paper we refer to other references for the more technical details on the GIMME procedure and the MIIV-2SLS estimator in order to make this paper more accessible to readers.

### **Functional MRI (fMRI) Data Example**

Human fMRI brain data serves as a proxy for neuronal activity. Technically, it is a measure of oxygenation levels in the blood that change in response to neuronal activity (Logothetis, 2008). A large number of time points are typically collected for each individual (e.g., > 150), thus enabling investigation into how the brain regions relate to each other across time.



These data serve as an optimal example since the structure of how brain regions relate to latent aspects of thought processes (i.e., the measurement model) is largely considered to be consistent across individuals (Wisner et al., 2013). However, the values of the loadings, or how the observed fMRI data for a brain region relates to a latent brain network, is known to vary across individuals (Laumann et al., 2015). Hence for this example we consider the individuals to be *qualitatively* homogenous by having the same measurement model structure yet *quantitatively* heterogeneous by allowing the factor loadings to differ (although having different structures is permitted in the LV-GIMME algorithm).

For exemplar purposes preprocessed data were obtained from the Autism Brain Imaging Data Exchange (ABIDE; Di Martino et al., 2014). We selected a sample individuals from one of the sites that provide data to the ABIDE project ( $N=34$ ). As noted above, in the fMRI literature a brain network is defined as a cluster of brain regions that tend to covary across time and thus reflects some latent aspect of thought, cognition, or emotion<sup>2</sup>. Considering brain brain networks as latent variables has been done previously with cross-sectional brain data (Bolt et al., 2018). We extend this line of thought by considering each brain network (i.e., latent construct) as comprising brain regions that tend to correlate across time. The observed indicator variables are “brain regions of interest.” To define which brain regions belonged to which brain network we used the Dosenbach brain atlas (Dosenbach et al., 2010). Each brain region loads exclusively on one brain network. Brain region to network membership was spread across ten networks, of which we used the (1) default mode, (2) fronto-parietal, (3) cingulo-opercular, (4) sensorimotor, (5) cerebellum, and (6) occipital networks. The number of observed variables (i.e., brain

---

<sup>2</sup> The use of the word “network” when describing brain networks is not the same as how the word network is used in network theory.

regions) varied for each brain network and ranged from 18 to 34 ( $M=26.7$ ,  $SD = 7.1$ ) for a total of 160 brain regions. The number of observations for each individual was  $T = 246$ . Across participants, after linear detrending the average value for the data was 0 ( $SD = 3.5$ ). Due to preprocessing steps that are common in fMRI (see Appendix A) the data are normally distributed and weakly stationary (i.e., the mean and variance is constant) across time with cyclical trends removed.

For this running example we use the same measurement model structure (i.e., which brain regions load onto which networks) for all individuals but allow the estimates to be unique for each individual. The mapping of brain regions to networks used here is provided by the creators of the brain atlas and is generally accepted (Dosenbach et al., 2010). In these definitions of networks each observed variable (brain region) loads only onto one latent variable (brain network) as is conventional in fMRI literature and outlined in the atlas description of brain regions. This convention could be relaxed as needed for individuals by having some brain regions load on more than one network. For our second empirical example, we use daily self-report and allow for the measurement model structures to vary across individuals to offer the readers a demonstration of this type of structural invariance.

### **LV-GIMME with Latent Variables**

In what follows we show that LV-GIMME performs well whether or not there is a common measurement model structure for individuals. The LV-GIMME approach involves using data in the form of latent variables, such as sum scores, components, or factor score series, as inputs to GIMME. Data are aggregated across brain networks to arrive at scores of “latent” constructs across time as a first step. Subsequently these latent score time series are used as inputs, or variables, in the traditional GIMME algorithm. There are a number of options for

arriving at latent series. One is to conduct factor analysis via pseudo-ML<sup>3</sup> or MIIV-2SLS and then use the resulting estimates to obtain factor series across time for each individual in the sample. Another approach is to simply add the values of indicators at each time point, thus obtaining sum scores for the latent constructs. Using PCA, one can arrive at a weighted sum score series which can then be used as a proxy for the latent construct. The LV-GIMME algorithm with these options available is publicly available within the *gimme* package (Lane, Gates, Fisher, & Molenaar, 2018) for R (R Core Team, 2018). As a final step, parameter estimates are obtained using MIIV-2SLS via the R package *MIIVsem* (Fisher, Bollen, Gates, & Rönkkö, 2018) for reasons described in the next section. But first, we explain the GIMME algorithm.

### **Original GIMME with Observed Variables**

Before moving forward on the approaches for estimating latent constructs it is important to provide basic information regarding the search procedure for LV-GIMME. The relations provided by GIMME are temporal in nature. Contemporaneous (or instantaneous) and lagged relations<sup>4</sup> can occur among variables (be they observed or latent). Contemporaneous relations are necessary for processes that are faster than the rate of measurement. For instance, in fMRI the signal (blood oxygenation levels) is a proxy for neuronal activity that occurs on a much faster scale (Huettel, Song, & McCarthy, n.d.; Logothetis, 2008). For this reason, temporal relations among brain regions tend to be best captured contemporaneously (Smith et al., 2011). Lagged relations provide insight into temporal precedence of constructs. Rationale for including both

---

<sup>3</sup> “pseudo” because the assumption in ML that rows are independent is violated when obtaining factor model parameter estimates on time series data.

<sup>4</sup> At present only lagged relations of order 1 are included. It is suggested that researchers with larger orders regress out the greater lag order prior to GIMME analysis.

comes from work demonstrating that failure to include one type of relation (e.g., lagged) when they truly exist will result in spurious paths and a failure to arrive at the data-generating model (Gates, Molenaar, Hillary, Ram, & Rovine, 2010)<sup>5</sup>. The importance of including temporal relations for personalized models of psychological processes continues to be highlighted in the field (see Epskamp et al., 2018).

Formally, the model found in GIMME (and thus LV-GIMME) is called a unified SEM (uSEM; Kim et al., 2007) and can be written as follows:

$$\boldsymbol{\eta}_t = \boldsymbol{\alpha}_\eta + \mathbf{A}\boldsymbol{\eta}_t + \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1} + \boldsymbol{\zeta}_t \quad (1)$$

where  $t$  indicates time and  $t-1$  indicating variables at a previous time point. The  $P \times P$  matrix  $\boldsymbol{\Phi}$  contains the vector autoregressive (VAR) effects for the  $\boldsymbol{\xi}_{t-1}$  variables predicting the endogenous  $\boldsymbol{\eta}_t$  values and  $P$  is the number of variables in  $\boldsymbol{\xi}_{t-1}$ . Returning to the fMRI example,  $\boldsymbol{\Phi}$  contains the VAR(1) coefficients or quantitative estimates of how values for brain regions at a prior time point relate to those at the next time point. Note that  $\boldsymbol{\xi}_{t-1}$  and  $\boldsymbol{\eta}_t$  are observed variables in the original GIMME algorithm and will be latent variables in the LV-GIMME algorithm. Here, the diagonal provides the autoregressive (AR) effects (or how a given variable predicts itself at later time points) and the off-diagonal effects are the cross-lagged effects, or how a given variable (e.g., brain region) is related to other variables at later time points. The  $P \times P$ -dimensioned  $\mathbf{A}$  matrix contains the contemporaneous relations among the endogenous  $\boldsymbol{\eta}_t$  variables. The  $P \times 1$  vector  $\boldsymbol{\zeta}_t$  contains the errors or disturbances in the equation and are assumed to have a mean of zero and not be correlated with the unique factors [ $C(\boldsymbol{\varepsilon}, \boldsymbol{\zeta}) = 0$ ,  $C(\boldsymbol{\delta}, \boldsymbol{\zeta}) = 0$ ] and  $C(\boldsymbol{\xi}_{t-1}, \boldsymbol{\zeta}_t) = 0$

---

<sup>5</sup> Continuous time models provide an alternative to those that capture discrete processes with contemporaneous relations.

or lagged variables (i.e., errors are not autocorrelated). Figure 1 depicts the uSEM, where contemporaneous relations in the  $\mathbf{A}$  matrix are depicted as solid lines and the lagged relations ( $\Phi$ ) are dashed.

The search procedure for LV-GIMME builds directly from the original observed-variable GIMME. Given that extant literature exists elsewhere on the use (Beltz & Gates, 2017; Lane & Gates, 2017) and technical details (Gates, Lane, Varangis, Giovanello, & Guiskewicz, 2017; Gates & Molenaar, 2012) of GIMME, we provide only brief description of the algorithm and refer the reader to these papers for more information. At its core GIMME is a data-driven approach for arriving at structures of directed relations (i.e., paths) among variables in time series data. These relations are elements of the  $\Phi$  and  $\mathbf{A}$  matrices. The addition of paths is guided by modification indices, a diagnostic Lagrangian multiplier statistic that researchers can use to assess if a given path, if added, would confer significant improvement in the model. The algorithm begins by first detecting which relations are significant for the majority of individuals in the sample. These “group-level” paths or relations are those that demonstrate high replicability across participants.

In finding these paths, GIMME differentiates signal and noise at rates higher than conducting analysis only at the individual-level (Gates & Molenaar, 2012; Smith et al., 2011). GIMME also provides individual-level paths. These are paths that, after accounting for those paths found at the group-level, exist for specific individuals. By starting with models (in this case, group-level paths) that are closer to the individuals’ true models GIMME improves upon recovery of individual-level paths (Gates et al., 2017)<sup>6</sup>.

---

<sup>6</sup> GIMME can also provide subgroup-level paths in addition to group- and individual-level paths. This is outside the scope of the present paper but is a possibility with LV-GIMME.

The LV-GIMME algorithm merges the benefits of measurement models with the accuracy of GIMME for arriving at dynamic model structures. Here, rather than search for paths among observed variables, it is the directed paths among latent variable proxies which are found in a data-driven manner. The  $\boldsymbol{\eta}_t$  vectors in Equation 1 are now predicted versions of the latent variables rather than observed variables. Importantly, the measurement model itself for any given individual is taken to be confirmatory. The LV-GIMME algorithm involves searching for relations among latent constructs and does not include exploratory factor analysis to arrive at person-specific measurement models. In the case where the structure of the measurement model varies across individuals, *a priori* analysis should be conducted to arrive at individual-level models, and these structures can then be used (see exemplar code in the Appendix for doing so with the *gimme* package). It is also possible for the researcher to impose the constraint that all individuals have the same measurement model (as is typically done in fMRI studies). The option is up to the researcher. The purpose of LV-GIMME is to arrive at the temporal dynamics among latent proxies (however they are defined) and provide asymptotically unbiased final estimates via MIIV-2SLS.

### **Integration of Latent Variables with GIMME**

One important requirement of LV-GIMME is that all individuals must have the same number of latent variables which represent essentially the same constructs. Due to the potential for a high level of heterogeneity, allowing for variation in measurement models is best conducted on measures where there are observed variables that are hypothesized to be uniquely related to each latent construct and thought to be the strongest indicator of the latent construct. The user can then define these as the scaling indicator for the factor models in the case of factor analytic options. Additionally, by having the same scaling indicator for each latent variable for each

individual, we enable comparisons of effects of the latent variables because they share a common scale. This would not be possible if the scaling indicator for the same latent variable differed across individual. In the latter case, we would not know if the differences in coefficients or factor loadings were due to differences in the effects or differences in scaling. Standardized coefficients/loadings would enable comparisons of expected differences in standard deviation units, but this adds another complication in comparisons where both an unstandardized coefficient and the standard deviations of variables affect the standardized coefficient (Blalock, 1968).

The general steps for LV-GIMME are as follows:

1. *Arrive at latent scores:* Factor scores (obtained via pseudo-ML or MIIV-2SLS estimates), component series via PCA, or sum scores are calculated.
2. *Obtain latent variable model:* The GIMME algorithm is conducted on these latent (or aggregate) scores to arrive at group- and individual-level patterns of relations among the latent variables.
3. *Estimate final model:* Conduct final estimation of the latent variable and measurement model parameters via MIIV-2SLS.

A number of options exist for arriving at scores that provide latent construct estimates across time (Step 1). For the factor score option, once measurement model parameter estimates are obtained via either pseudo-ML or MIIV-2SLS, there are options for generating the factor score series. The regression method for estimating factor scores from the *lavPredict* function in *lavaan* (Rosseel, 2012) is used in the present paper and is also an option when conducting LV-GIMME in the *gimme* package. These estimates of  $\eta_t$  are then time-embedded to create the lagged version ( $\xi_{t-1}$ ). The search for contemporaneous paths (the estimates of which are in  $\mathbf{A}$ ) and lagged paths

( $\Phi$ ) is then conducted using the original GIMME algorithm on these estimates of latent time series data.

As with the original GIMME, pseudo-ML is used to arrive at estimates during the search procedure. Unfortunately, these estimates are known to have issues for a number of reasons. For one, the latent variable scores that we use in the GIMME search procedure contain measurement error regardless of the method used, and thus bias the estimates of effects and standard errors. Two, even when measurement errors are not a concern it is known that conducting dynamic factor analysis within SEM results in inflated standard errors (Chow, Ho, Hamaker, & Dolan, 2010). Third, the properties of the fit statistics used for evaluating model fit are unknown when used in conjunction with factor scores rather than the observed data. Given these reasons, it is possible that some relations among latent variables may be missed in the present context. The signal to noise ratio may be lower than seen in the study (Smith et al., 2011) that informed *gimme*'s default setting of requiring that 75% of individuals have a significant path before adding it to the group-level model. We suggest using a looser criterion of 51% and back up this recommendation in the simulation studies to follow. As a final step, MIIV-2SLS is used to arrive at final estimates. The MIIV-2SLS estimator corrects for measurement errors and has a number of additional benefits that make it ideal for this context, which we explain in the next section.

Figure 1 provides the results obtained in the fMRI data example when using the sum scores to represent latent brain network activity. The GIMME algorithm was used on these scores to arrive at the structure of relations for the latent variable model. Shown here are the average estimates across individuals for estimates of the GIMME-derived paths obtained from MIIV-2SLS. We used the 160 brain regions across 34 individuals from the fMRI data example described previously. There were a number of group-level paths. When averaged across



individuals the contemporaneous path estimates were positive and the lagged path estimates were negative. Typically in fMRI contemporaneous paths are interpreted. The mean number of individual paths (not shown) across the sample was 3.25 (SD=2.00).

## Output

Taken together, the LV-GIMME algorithm provides group- and individual-level latent variable model path structures (as well as subgroup-level if invoked) directly from the *gimme* package. Typical *gimme* output is provided, such as the path estimates, fit indices, and plots (see Lane & Gates, 2017, for details). As the final estimates for both the measurement and latent variable model coefficients are conducted separately for each individual using MIIV-2SLS via the package *MIIVsem* (Fisher, Bollen, Gates, & Rönkkö, 2018), this additional output is also provided to the user. One difference with the *MIIVsem* output is that the same model fit indices will not be produced with the MIIV-2SLS procedure<sup>7</sup>. However, users can utilize the provided Sargan tests to explore the fit of overidentified equations and identify which individuals may not have appropriate model structures (Sargan, 1958).

### The MIIV-2SLS estimator

Regardless the aggregation method chose, we highly recommend that the final estimates for both the measurement model and latent variable model are obtained using MIIV-2SLS (Bollen, 1996). Fisher, Bollen, & Gates (2019) show that the MIIV-2SLS estimator has a number of advantages in the estimation of dynamic factor models within the SEM framework that are relevant for use with time series data. For one, model misspecifications are better isolated

---

<sup>7</sup> The GIMME procedure in LV-GIMME still uses fit indices for stopping criterion for the search procedure, and this is provided for the user. However these fit indices do not include the measurement component or the final estimates obtained in *MIIVsem*.

(Bollen, 2001). Model misspecification is a pervasive problem when looking at confirmatory models (Browne & Cudeck, 1993) and is even more likely if one does not take into account individual-level nuances (Molenaar & Campbell, 2009). Hence in the presence of a confirmatory or exploratory model, MIIV-2SLS estimation provides greater robustness to misspecifications (Bollen et al., 2018).

A second and related benefit is that the latent variable model parameter estimation is conducted separately from measurement model parameter estimation. As noted above, much work has shown that the strict assumption of homogeneity in processes across individuals is often not met in neuroimaging (Finn et al., 2015; Gates, Molenaar, Iyer, Nigg, & Fair, 2014; Laumann et al., 2015; Miller et al., 2002), behavioral (Anzman-Frasca et al., 2013; Beltz et al., 2013) and psychological (A. J. Fisher, Medaglia, & Jeronimus, 2018; Molenaar, 2004; Ram & Grimm, 2009; Wright et al., 2013) studies. MIIV-2SLS accommodates both the idiographic filter and need for individual-level models by (1) estimating measurement models separately for each individual and (2) estimating the latent variable model in such a way that the measurement model parameter estimates are not influenced by the latent variable paths added, thus enabling straightforward interpretation of both the measurement and latent variable coefficient components. This second point is a crucial feature of LV-GIMME estimated with MIIV-2SLS. It has been shown previously that relations (even misspecifications) occurring for the latent variable model do not influence the measurement model parameter estimates (Bollen et al., 2018) and that MIIV-2SLS estimates of dynamic factor model parameters can be more robust to model misspecification when compared to traditional system-wide estimators (pseudo-ML, Kalman filter) (Fisher, Bollen, & Gates, 2019). Hence in this framework, estimating the entire final model using MIIV-2SLS will provide the same measurement model parameter estimates as

those one would obtain by estimating the model with no latent variable paths. This quality is referred to as “robust unchanged” (Bollen, Gates, & Fisher, 2018).

A third benefit of MIIV-2SLS estimation is that it is conducted one equation at a time for each of the left hand side equations rather than simultaneously as is done in full information estimators. Because there are fewer parameters in a single equation than in a full model, this reduces the burden on the estimation in that there are more observations per parameter in a single equation than when all parameters are estimated simultaneously in the full model. The often-used pseudo-ML estimation encounters convergence problems for shorter time series. This poses a problem across various domains of psychological inquiry. In brain data acquired via fMRI, there is often a large number of brain regions relative to the number of time points. Daily diary studies also encounter this problem if responses on numerous questions are provided for only one or two months. In both situations, researchers may be interested in the measurement model as well as the relations among latent variables at the individual level. MIIV-2SLS is noniterative (i.e., does not attempt multiple solutions) and does not face the problem of nonconvergence that may occur with the pseudo-ML estimator. Another desirable feature of the MIIV-2SLS estimator is that it is possible to estimate identified equations even when the model as a whole is not identified. In some situations, this permits estimation of a key equation even if we cannot estimate all equations in the full model.

In sum, the MIIV-2SLS opens up the new opportunities in that it: 1) is more robust to structural misspecification than full information estimators, 2) does not require equal parameters or structures across individuals, 3) is a noniterative estimator of coefficients and factor loadings that avoids issues of nonconvergence, 4) is computationally quick, and 5) enables estimation of identified equations in underidentified models. Importantly, more recent investigations indicate

that MIIV-2SLS also has unique advantages over pseudo-ML and ML approaches for estimating dynamic factor models (Fisher, Bollen, & Gates, 2019). LV-GIMME takes advantage of these properties in our time series analyses.

### Technical Details of MIIV-2SLS

Details on the MIIV-2SLS are available in several sources (Bollen, 1996; 2001; Bollen, Kirby, Curran, Paxton, & Chen, 2007). Here we give an overview of the estimator with additional details provided in the Supplemental Materials. The MIIV-2SLS estimator starts by transforming the latent variable model into one that consists only of observed variables by replacing each latent variable by its scaling indicator minus its error. This is the latent to observed variable transformation or L2O. Rather than using an estimator that estimates all parameters simultaneously, the MIIV-2SLS estimator can apply one equation at a time. The resulting observed variable equations resulting from the L2O transformation typically have a composite error that correlates with one or more of the covariates of an equation (see Supplemental Material for details). Instrumental variable methods (e.g., Bollen, 2012) are helpful in situations like this. Instrumental variables are uncorrelated with the composite error and correlate with the covariates that are associated with these errors. Unlike traditional applications of instrumental variables where instruments are identified from outside the model, the MIIV approach draws instruments from within the system itself based on the model specification. For this reason, no special steps are required for identifying instruments in the context of empirical data (e.g., the exemplar fMRI data). If for a given equation a variable is uncorrelated with the equation disturbance and correlated with the covariates (both of which are implied by the model) it can be considered a valid instrument. Finding variables that meet these conditions for instruments is done automatically within *MIIVsem* by using an algorithm from

Bollen (1996) that uses the model structure to determine which variables are suitable instruments, hence the term "model implied" instrument.

The equation and MIIVs in hand, a 2SLS estimator is used. We use a modification of the LISREL notation (Jöreskog, Sörbom, Aigner, & Goldberger, 1977; Bollen, 2001) to represent the measurement model component for dynamic factor models of lag one:

$$\mathbf{y}_t = \boldsymbol{\alpha}_y + \boldsymbol{\Lambda}_y \boldsymbol{\eta}_t + \boldsymbol{\varepsilon}_t \quad (2)$$

$$\mathbf{x}_{t-1} = \boldsymbol{\alpha}_x + \boldsymbol{\Lambda}_x \boldsymbol{\xi}_{t-1} + \boldsymbol{\delta}_{t-1} \quad (3)$$

where  $\mathbf{y}_t$  is the vector of observed variables,  $\boldsymbol{\eta}_t$  is the latent endogenous variables at a given time  $t$ ,  $\mathbf{x}_{t-1}$  is a vector of the same observed variables but at one prior time point,  $\boldsymbol{\xi}_{t-1}$  the latent variables at the prior time point,  $\boldsymbol{\alpha}_y$  and  $\boldsymbol{\alpha}_x$  are the vectors of intercepts,  $\boldsymbol{\Lambda}_y$  and  $\boldsymbol{\Lambda}_x$  are the matrices of factor loadings of dimension  $O$  observed variables by  $P$  latent variables, and  $\boldsymbol{\varepsilon}_t$  and  $\boldsymbol{\delta}_t$  contain the measurement errors (or uniqueness) of the indicators. Note that the estimated parameters do not vary across time. The data are assumed to be multivariate stationary with constant cross-correlations across time. We assume that the errors have means of zero [ $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ ,  $E(\boldsymbol{\delta}) = \mathbf{0}$ ] and are uncorrelated with their respective latent variables [ $C(\boldsymbol{\varepsilon}, \boldsymbol{\eta}) = 0$ ,  $C(\boldsymbol{\varepsilon}, \boldsymbol{\xi}) = \mathbf{0}$ ,  $C(\boldsymbol{\delta}, \boldsymbol{\xi}) = 0$ ] and each other [ $C(\boldsymbol{\varepsilon}, \boldsymbol{\delta}) = 0$ ].

The latent variable uSEM (see Equation 1) provides the structure of relationships among these latent variables. The difference is that here,  $\boldsymbol{\xi}_{t-1}$  and  $\boldsymbol{\eta}_t$  variables are latent instead of observed. An important feature of the model as written here is that the latent  $\boldsymbol{\xi}_{t-1}$  variables are by definition predetermined ("exogenous"). This is because a variable at  $t$  cannot cause events backwards in time. As explained in the Supplemental Material, this special property of the

lagged variables provides a set of MIIVs that will be appropriate for the latent variable model parameter estimates.

### **Empirical Example of Measurement Model parameter estimates from MIIVsem from fMRI data**

Figure 2 provides a depiction of the estimates for each of the 18 brain regions of interest (ROIs) that load on the cerebellum network. Here we provide the average estimates across individuals. We can immediately see that most brain regions seem to correspond relatively well with the cerebellum network with almost all estimates being at or above 0.40. Additionally, the standard deviations of the estimates (in parentheses) suggest that there is some variability across individuals in these estimates.

## **Monte Carlo Simulation Studies**

### **Data generation**

The simulation studies to follow serve three purposes. First, we seek to evaluate the accuracy of the LV-GIMME path search across varied methods for arriving at latent scores. (In any case, final estimation using MIIV-2SLS would be recommended for reasons noted above.) Second, due to the presence of measurement errors in the latent scores (however derived), we seek to investigate for the first time if GIMME performs better in terms of path recovery when the threshold for what constitutes the “majority” of individuals in the group-level search is loosened. Three, we seek to evaluate when the algorithm performs well and when it does not. Towards aim 1, we test five methods for arriving at scores across time: (1) scaling indicator only (i.e., observed-variable model); (2) sum scores; (3) pseudo-ML; (4) MIIV-2SLS; and (5) first component series from PCA. These methods are utilized across all conditions to follow.

To investigate the second aim, we run GIMME with the threshold for majority set at 75% and with the threshold set at the most lenient definition of majority, that of being greater than half the individuals (51%). As a reminder, the GIMME search for paths begins by detecting which paths exist for the majority. These become what are considered the group-level paths. GIMME has previously only been evaluated at the 75% threshold for what defines the “majority” (Gates et al., 2017; Gates & Molenaar, 2012; S. Lane et al., 2018). This value was informed by the expected power to detect dynamic effects among brain regions in fMRI studies. Much like other approaches (e.g., Ramsey et al., 2011), GIMME favors parsimony and may stop the model search procedure prior to detecting all true paths. Hence it may be appropriate to allow more paths to be added early in the model search, and this can be achieved by lowering the threshold for what constitutes the ‘majority’ in the group-level search so that more paths are added here. Additionally, true paths may be difficult to detect since we know that measurement errors will be included in the scores for the latent variables regardless the approach used to estimate them. We investigate this by running the algorithm at the default group-level path threshold of 75% and altering this option for what constitutes the majority in the *gimme* function arguments to 51% for one condition.

For the third aim, we varied a number of conditions known to exist across psychological studies. First, we varied the length of the time series:  $T = 60, 90,$  and  $200$ . The first two are more likely to be seen in daily diary studies whereas having  $T = 200$  observations may be more likely in psychophysiological studies such as functional MRI. For these investigations, we simulated data using the parameter estimates for the directed paths obtained from the latent variable models of the empirical fMRI data example. In this way, we could be sure that the degree of individual-level nuances was captured.

Specifically, for each individual the  $\mathbf{A}$ ,  $\Phi$ , and  $\text{Var}(\zeta_t)$  matrices from the results obtained when GIMME was run on the sum scores of the observed brain variables (group-level results depicted in Figure 2) were used to generate factor scores (latent time series) for each of the desired time series lengths. Both group- and individual-level paths among latent variables were used as model structures for data simulations. We opted to use the sum scores rather than the factor scores reconstructed via MIIV-2SLS so that we were not generating the data in a way that was biased towards the factor analytic or PCA approaches.

We then created observed time series according to two possible measurement model scenarios. In the first scenario, all 160 free elements of  $\Lambda$  were set to unity. This represents an ideal situation where each observed variable loads approximately equally onto the construct. We expected each of the methods for arriving at the aggregate scores to work approximately equally well. For the second, free elements of  $\Lambda$  were generated according to a truncated normal distribution with a mean of 0.50, a standard deviation of 0.25, and a range of 0.10 and 1.50. These values were informed by the empirical fMRI example results. In both conditions  $\text{Var}(\epsilon_t)$  was set to identity.

Finally, for each of the described measurement model conditions we simulated a second set of observed time series with randomly assigned cross-loadings that were subsequently omitted during analysis. Hence the measurement models had misspecifications. For data generation we randomly assigned non-zero values to 20% of the elements in  $\Lambda$  that had previously been fixed to zero. Conceptually, this means we allowed some observed variables to be influenced by more than one latent variable. In practice the omission of cross-loadings is likely and it is important to examine if performance of our method generalizes across commonly encountered scenarios. In our simulation design all conditions were fully crossed.

### **Evaluation metrics**



MIIV-2SLS has previously been shown to provide more robust estimates than pseudo-ML in a number of conditions (Bollen, et al., 2007), one of which being the dynamic models considered here (Fisher, Bollen, & Gates, 2019). For this reason we restrict our inquiry to the ability to accurately recover the latent variable paths and omit investigations into bias of estimates. LV-GIMME does not conduct exploratory factor analysis and considers measurement model structures to be known. As such, only the ability for GIMME to recover the data-generating relations among latent constructs (i.e., paths) is assessed here. Sensitivity is used to assess the ability to recover paths that existed in the data-generating models and specificity is used to evaluate the presence of false positives. Sensitivity is the ratio of true positives (i.e., accurately recovered paths) over the total number of paths in the data-generating model. High values here indicate a high rate of recovery of true paths. Specificity was calculated as the ratio of true negatives (i.e., paths appropriately left out of the model) over the total number of possible paths not in the data-generating model. High values would indicate a low rate of false positives. These values are calculated separately for each individual, and consider the recovery of all paths (both group- and individual-level) in arriving at the specificity and sensitivity values.

## Results

The main finding is that recovery of paths via GIMME (i.e., sensitivity) did not differ substantially regardless what measurement approach was used for the latent variables or the distribution of lambda values in the data-generative process (see Figure 3). As expected, rates of recovering the true paths increased as the length of time series increased. These results suggest that GIMME may not reliably detect paths when the time series are shorter than  $T = 90$ . When using the 75% cutoff for the majority, rates become acceptable (the median above 70%) in terms of sensitivity for  $T$  greater than or equal to 90. For long time series of  $T = 200$  the median recovery

rate was approximately 72% across all methods for recovering the true latent variable patterns. False positives, as quantified using specificity, are not a problem for any length of time as seen in the high specificity rates.

Setting the group cutoff to 75% is conventional and performs well. We can see from the top panel of Figure 3 that false positives are not a problem and recovery of true paths is lower than 75% on average. Hence we may be sacrificing the ability to recover the true paths by using a strict definition of majority (i.e., a group cutoff of 75%) in the effort to favor parsimony, thereby simultaneously reducing the risk for false positives. Given that measurement error may be clouding results we also ran GIMME with 51% as the threshold for what constitutes the majority of individuals in the group-level search path. This means that the technical majority – just over half of individuals – would need to have a path be significant in order for it to be added to the group-level results. The bottom panel of Figure 3 depicts the results with this threshold. As expected, the recovery of true paths noticeably increased. For instance, at  $T = 200$  the mean recovery for the truncated normal condition was 82%, compared to 77% for the same condition when 75% set as the group cutoff. Gains were also seen when  $T = 60$ , however this seemed more sensitive to the distribution of lambdas.

Finally, we evaluated results when the models were misspecified. A similar pattern emerged with only a few noteworthy differences. Figure 4 depicts the results from the simulation where data were generated to have 20% of the observed variables cross-load, with the top panel reflecting results with the 75% threshold for the majority and the bottom panel being when 51% was used. In each case for arriving at aggregate latent scores (i.e., sum score, MIIV-2SLS, Pseudo-ML, and PCA), the measurement model was misspecified by omitting these additional paths. Interestingly, latent scores estimated using pseudo-ML, MIIV-2SLS, and PCA were robust to these

misspecifications. The scaling indicator (i.e., using only one variable) and sum score approaches demonstrated a decline in recovery of paths across all lengths of time with sum scores additionally evidencing an increase in false positives as indicated by the specificity plot. As seen in the correctly specified models, decreasing the group-level cutoff for what constitutes the majority to 51% improved model recovery (bottom panel, Figure 4). Again, this was not at the cost of an increased false positive rate; specificity also remained high.

### **Empirical Example: Allowing for Qualitative and Quantitative Differences**

The Lebo & Nesselroade (1978) data have been used previously to explore measurement invariance across individuals. Self-reported data on 75 adjectives assessed with a 5-point scale were collected for five individuals (all female) across 120 days. Lebo and Nesselroade (1978) found some consistency across individuals, particularly for the factors of Well Being (Wb), Energy (En), and Fatigue (Fat). Social Affiliation (SocA) separated from other factors for 2 of the 5 participants, and for the others was subsumed in Well Being. However, they note (as well as others; see Gayles & Molenaar, 2013) that some variability also exists in these factor structures. Here, we capitalize on this previously found consistency across individuals and use the items from all four factors and explore modifications to a four factor solution.

To begin, we remove items for each individual that had low variance (defined here as a standard deviation below .10). Hence already there are qualitative differences in the model structure as individuals have different sets of observed variables. We then conducted confirmatory factor analysis to identify the best scaling indicator to use when looking across individuals. The loadings were rank-ordered and the observed variable per factor with the highest average rank across individuals was selected to be the scaling indicator for all individuals. The scaling indicators obtained were as follows: “cheer” was selected for Well Being, “energy” for Energy, “tired” for

Fatigue, and “warmhearted” for Social Affiliation. With the optimal scaling indicators in hand, we conducted an exploratory factor analysis separately for each individual whereby all non-scaling indicators were allowed to load on any of the four factors. Items whose standardized loadings were greater than or equal to 0.50 in absolute value were retained. Note that cross-loadings were allowed, as well as individual-level variability in the factor structure. The only constraints were that they have the same scaling indicators and the same number of factors. This process results in individual-level measurement models that share a common set of scaling indicators across individuals without imposing further homogeneity constraints on the specific composition of each factor.

The full table of results are in the supplemental materials as arrival at factor solutions is not the primary focus of this paper. Nonetheless, it is important to identify the extent to which the factors can be said to be proxies of the same constructs across individuals. Evidence suggests they are, to varying degrees. For instance, “carefree”, “cheer”, “happy”, “contented”, “at ease”, and “calm” loaded onto the Well Being factor for all individuals. All of these were hypothesized to be on the Well Being factor in Lebo & Nesselrode (1978). The Energy and Fatigue factors recovered here also showed a similar structure to previous results, namely all individuals having “energy”, “active”, “peppy”, and “lively” as indicators of Energy and “tired” as an indicator of Fatigue. For all individuals Social Affiliation contained a mix of hypothesized indicators as well as some from Well Being and Energy: “enthusiastic” (En), “comfortable” (Wb), “contented” (Wb), “relaxed” (Wb), “at ease” (Wb), “calm” (Wb), “warmhearted” (SocA), “affectionate (SocA), and “kindly” (SocA). Given these base sets of shared indicators, it seems that the factors having meaningful consistencies across individuals.

There were also qualitative differences in the factor structures across individuals. For example, only 28 of the recorded variables for Person 1 exhibited sufficient variance to be included in the analysis. This influenced the results: Fatigue had only three indicators for this individual: “sluggish”, “weary”, and “tired”. None of the other hypothesized indicators for this individual had enough variance to be included in the search space. Person 2 had more variables that met this criterion (35 variables of the total set had enough variance to be included) and five of these loaded on Fatigue. Three of the variables were the same as for Person 1. Hence some differences were expected due to the qualities of the data themselves.

In all cases, estimates were uniquely obtained for each individual using the MIIV-2SLS option within *gimme*. To provide quantitative evidence of model consistency we used the congruency coefficient (Lorenzo-Seva & ten Berge, 2006). Details can be found in the Supplemental Material, but overall the congruency would not be considered high, with an average of 0.74 across the factors. Some individuals for some factors scored higher (e.g., 0.92 for persons 1 and 2 on Fatigue) whereas others were low (e.g., 0.32 for Persons 1 and 3 on Fatigue). Much of this difference can be explained by the presence of more variables in the search space as well as poor differentiation for some individuals as evidenced by numerous cross-loadings. Nonetheless, subjective evaluation suggests consistency across individuals and differences were expected.

This example provides one example of how individual nuances can be accommodated via the LV-GIMME approach. It is not required that individuals have the same model, yet it is recommended that researchers evaluate the differing solutions to ensure fidelity and determine whether or not the same constructs are being measured across individuals. The person-specific measurement model structures were used in LV-GIMME to arrive at a set of relations among the latent proxies (see Appendix for example code). Figure 5 depicts the results. As shown previously

(Lebo & Nesselroade, 1978), Energy had a negative relation with Fatigue for all individuals. Similarly, all individuals had strong and positive relations of Social Affiliation with Well Being and Energy.

### **Discussion**

LV-GIMME detects ubiquitous as well as personalized patterns of relations among latent constructs. The algorithm described herein merges the benefits of two previously generated approaches – GIMME and MIIV-2SLS – that each have unique contributions. GIMME provides a data-driven search among latent constructs in a manner that allows for individual-level heterogeneity. It first detects signal from noise by finding paths that exist for the majority of individuals. Importantly, this is done in a way that cannot be swayed by just a few individuals. Prior work had demonstrated its ability to recover patterns of directed relations among observed variables; here we show that it performs equally well on data aggregated in some manner to represent latent constructs. The second important aspect of LV-GIMME is the use of MIIV-2SLS for estimation. Using MIIV-2SLS for final estimation of the measurement and latent variable models ensures greater robustness when estimating the final models even if there are misspecifications.

The simulation results suggest that using factor score estimates or the first component series from PCA outperforms using sum scores (as is commonly done in practice). Additionally, when the observations are imperfect measures of a latent construct, conducting observed-variable analysis performs poorly in terms of recovering relations among the constructs. Finally, gains in recovering paths were seen by decreasing the group-level threshold to 51%. Importantly the rate of false positives remained very low even when allowing for more paths to potentially emerge at the group level. These results point to the flexibility of options when conducting LV-GIMME. The

*gimme* package provides options for the approaches that performed well - factor analysis with MIIV-2SLS or pseudo-ML and PCA - for developing latent construct time series.

### **Allowing for measurement and latent variable models to vary across individuals**

A key benefit of using LV-GIMME is that both the measurement models (how the observed variables relate to the latent construct) and latent variable models (how the latent constructs relate to each other) are allowed to vary. This is critical since much work has shown that individuals vary in how constructs are measured (Hamaker et al., 2005; Molenaar & Campbell, 2009) as well as how constructs relate to each other (Epskamp et al., 2018; Fisher & Boswell, 2016; Fisher et al., 2018; Lane et al., 2018). The use of MIIV-2SLS for estimation separates the estimation of the measurement model from the latent variable model; in this way, the measurement model parameter estimates remain unchanged regardless of which paths are omitted or added to the latent variable model (Bollen et al., 2018). Allowing for both components to vary represents an extension of the Idiographic Filter (Nesselroade et al., 2007), which allows the measurement component to vary across individuals but maintains that the relations among constructs remain the same. Enabling personalized variability in the pattern of relations among latent constructs will allow for understanding of individual-level processes as they unfold over time. LV-GIMME does not require that temporal dynamics be the same across individuals but it does find these similarities should they exist.

### **Comparisons to the Idiographic Filter**

A few clarifications regarding the ideographic filter proposed by Nesselroade et al. (2007) and aspects of the LV-GIMME algorithm are in order. First, in regards to the

measurement model, both approaches propose allowing the measurement model to vary at the individual level. In the case of the LV-GIMME algorithm discussed here an additional assumption is made. Here, we assume that for any given construct at least one indicator is shared across all subjects. This indicator is then set to be the scaling-indicator for the construct in question for all subjects in the sample. Specialized statistical diagnostics to see if this is true require further development, but common guidelines regarding the selection of scaling indicators are appropriate here. Namely, it is important that the scaling indicator chosen is a strong measure of the construct of interest given prior research or current estimates.

Second, we describe differences among the two approaches in terms of the relations between the person-specific factors. Nesselroade et al. (2007) propose the correlations among the factors are equivalent across all subjects as a means of equating the theoretical variables across individuals. By restricting these correlations to be quantitatively equivalent across all individuals, Nesselroade et al. (2007) argue that researchers can be confident that the constructs, although measured by different indicators, are equivalent. While conceptually one might argue that there should be *some* processes that are common across all individuals (Molenaar & Nesselroade, 2012), clinical work indicates that individuals also vary in important ways (Fisher & Boswell, 2016). At present it is unclear what processes are truly nomothetic and hold for all individuals and which ones have personalized nuances, but results obtained on observed variables suggest individuals may vary greatly in their processes (Fisher et al., 2018).

In the approach described here this restriction is relaxed. We do not require that relations among different latent constructs be equivalent across individuals, and in fact we expect them to differ both quantitatively and qualitatively. This is because the relations among constructs of interest in the study of psychology may vary across individuals. As an example, the relationship



between worry and depression may vary across individuals. For some individuals, they may not be related and have a relation estimate close to zero. For other individuals, depression may precede worry whereas for other individuals worry may precede depression. That a relationship between worry and depression exists when looking across individuals (as seen in cross-sectional analysis) does not necessarily inform that the relation might occur when looking across time within individuals (Molenaar, 2004). A recent paper found that much variability exists in the estimates for relations among these constructs when looking across individuals and caution against using assuming the same relations across time exist for all individuals (Fisher, Medaglia, and Jeronimus, 2018).

### **Degree of congruence among individuals' factor models**

Now that we have outlined the major differences in these approaches we address an important set of questions that is pertinent to both approaches: how similar need the factor models be to say they are capturing the same underlying construct? How do we know we are measuring the same construct across individuals if invariance is not imposed at the measurement or level? This question is a bit circular since the premise behind the Idiographic Filter and LV-GIMME is that individuals do vary in their measurement model. However there must be a point at which individual models are too different to be assessing the same construct. Currently best practices seem to be subjective evaluation by the researcher or clinician to identify if the latent constructs indicated by the factor solution patterns seem to represent the same constructs across individuals.

One analytic option might be to investigate the congruence coefficient. Cutoffs are available for identifying when measurement models can be considered “fairly similar” (Lorenzo-Seva & ten Berge, 2006). However, these values were derived from the perspective of evaluating

measurement invariance across individuals, which we argue is not a realistic assumption in all cases. It is certainly possible to have a congruence factor that is lower than the “fairly similar” threshold of 0.80 and still be assessing the same construct. One example would be if one individual has 5 items that load on a factor, and another individual has the same 5 items along with 8 additional items. Are these still the same construct, even though one person’s model was more parsimonious than the second? Perhaps. Clearly more work is needed to identify the lower bound, or a threshold for congruence coefficients, that indicates when measurement models cannot be said to be assessing the same construct. In the meantime, if a researcher hopes to draw conclusions about the consistency of factors across individuals it is recommended that factor compositions are compared across individuals, paying specific attention to the co-occurrence of specific items.

### **Choice of proxy for data-driven detection of relations among latent variables**

An interesting finding is that GIMME performs well regardless of the method used to arrive at latent variable scores. Using pseudo-ML (the current standard) or MIIV-2SLS to estimate factor solutions and subsequent factor scores or using PCA to obtain the first component scores across time all worked approximately equally well as inputs to LV-GIMME. The oft-seen approach of summing across observed items did not consistently perform well. Importantly, the use of a single observed variable in simulations where the constructs were designed to be measured with multiple observed variables performed poorly. This has important implications for the use of observed variables (i.e., single indicators) in data-driven searches at the individual level when measurement errors may be present. Our paper suggests that in some cases observed variable analysis may not be appropriate. Specifically, when the observed variable is better utilized as a measure of an overarching construct then latent variables provide more reliable results in data search procedures

such as this one. Recent efforts have been made to merge observed variable and latent variable analysis (Epskamp, Rhemtulla, & Borsboom, 2017). This could easily be done in the LV-GIMME context by providing a mix of variables with some being factor scores and others being observed variables.

### **Choice of what constitutes *majority* in data-driven search**

A key part of the GIMME algorithm (and therefore, the LV-GIMME approach) is the detection of group-level paths, or paths that exist for the majority of individuals. These paths are used to make inferences that generalize to the population from which the sample came and also provide a strong foundation for the search for individual-level paths (Gates & Molenaar, 2012). Like many search algorithms (Mumford & Ramsey, 2014), the LV-GIMME search algorithm favors parsimony in the model selection approach. This sometimes comes at the cost of missing true relations that should be in the model, as was seen here - false positives or spurious paths were rare in LV-GIMME regardless the method for arriving at the latent variable. Hence it seems that we could be a bit more lenient in the addition of paths.

Within the GIMME algorithm, there is an option to loosen a criterion, called *group cutoff*, which dictates what is considered the threshold for majority when detecting which paths exist for the majority of individuals. GIMME has previously been evaluated with a group cutoff of 75%. This was informed by fMRI research and prior simulation studies (Smith et al., 2011). It again performed well here when there are at least 90 time points per person by accurately detecting both the presence of a relation and its direction. While acceptable, there is still room for improvement in these rates. We thus tested how well paths were recovered when the threshold for what is considered the majority is loosened. When the group cutoff threshold was decreased to 51% the recovery rates improved. Importantly, this was not at the cost of increased numbers of false

positive paths. Hence it seems there is benefit to decreasing the group cutoff threshold and this comes with few drawbacks in the data simulated here.

## **Limitations**

At present, researchers must provide a confirmatory measurement model for each individual (or the same one for all individuals). Ideally the search algorithm would also find personalized measurement models. In light of the potential for latent constructs to vary greatly across individuals, more work is needed towards arriving at reliable models from exploratory factor analysis conducted at the individual level. Another limitation is that LV-GIMME does not perform well for under 90 observations. While the final estimation approach, MIIV-2SLS, can provide unbiased results with fewer time points, the search procedure in GIMME requires at least 90. This may prevent the use of the algorithm with some data, such as daily self-reports collected over only the span of 30 days. However, the approach does capitalize on the MIIV-2SLS ability to arrive at robust results when the number of observed variables is relatively large compared to the number of time points.

## **Closing remarks**

LV-GIMME provides researchers with a method for arriving at truly personalized dynamic models. This will aid in getting the field even closer to targeted treatment, intervention, and prevention plans. The results presented here demonstrate that it performs well when individuals have the same measurement structure but different estimates for those relations. Importantly, LV-GIMME can recover the true process among latent constructs even when individuals vary in how their latent constructs are measured.

## References

- Anzman-Frasca, S., Liu, S., Gates, K. M., Paul, I. M., Rovine, M. J., & Birch, L. L. (2013). Infants' Transitions out of a Fussing/Crying State Are Modifiable and Are Related to Weight Status. *Infancy: The Official Journal of the International Society on Infant Studies*, *18*(5), 662–686.  
<https://doi.org/10.1111/infa.12002>
- Beltz, A. M., Beekman, C., Molenaar, P. C. M., & Buss, K. A. (2013). Mapping Temporal Dynamics in Social Interactions With Unified Structural Equation Modeling: A Description and Demonstration Revealing Time-Dependent Sex Differences in Play Behavior. *Applied Developmental Science*, *17*(3), 152–168. <https://doi.org/10.1080/10888691.2013.805953>
- 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>
- Beltz, A. M., Wright, A. G. C., Sprague, B. N., & Molenaar, P. C. M. (2016). Bridging the Nomothetic and Idiographic Approaches to the Analysis of Clinical Data. *Assessment*, *23*(4), 447–458.  
<https://doi.org/10.1177/1073191116648209>
- Bollen, K. A. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika*, *61*(1), 109–121. <https://doi.org/10.1007/BF02296961>
- Bollen, K. A., Gates, K. M., & Fisher, Z. (2018). Robustness Conditions for MIIV-2SLS When the Latent Variable or Measurement Model is Structurally Misspecified. *Structural Equation Modeling: A Multidisciplinary Journal*, *0*(0), 1–12. <https://doi.org/10.1080/10705511.2018.1456341>
- Bolt, T., Prince, E. B., Nomi, J. S., Messinger, D., Llabre, M. M., & Uddin, L. Q. (2018). Combining region- and network-level brain-behavior relationships in a structural equation model. *NeuroImage*, *165*, 158–169. <https://doi.org/10.1016/j.neuroimage.2017.10.007>
- Bringmann, L. F., Ferrer, E., Hamaker, E. L., Borsboom, D., & Tuerlinckx, F. (2018). Modeling Nonstationary Emotion Dynamics in Dyads using a Time-Varying Vector-Autoregressive Model.

*Multivariate Behavioral Research*, 53(3), 293–314.

<https://doi.org/10.1080/00273171.2018.1439722>

Chow, S.-M., Ho, M. R., Hamaker, E. L., & Dolan, C. V. (2010). Equivalence and Differences Between Structural Equation Modeling and State-Space Modeling Techniques. *Structural Equation Modeling: A Multidisciplinary Journal*, 17(2), 303–332.

<https://doi.org/10.1080/10705511003661553>

Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., ... Milham, M. P. (2014). The Autism Brain Imaging Data Exchange: Towards Large-Scale Evaluation of the Intrinsic Brain Architecture in Autism. *Molecular Psychiatry*, 19(6), 659–667.

<https://doi.org/10.1038/mp.2013.78>

Dosenbach, N. U. F., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., ... Schlaggar, B. L. (2010). Prediction of Individual Brain Maturity Using fMRI. *Science*, 329(5997), 1358–1361.

<https://doi.org/10.1126/science.1194144>

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized Network Psychometrics: Combining Network and Latent Variable Models. *Psychometrika*, 82(4), 904–927.

<https://doi.org/10.1007/s11336-017-9557-x>

Epskamp, S., van Borkulo, C. D., van der Veen, D. C., Servaas, M. N., Isvoranu, A.-M., Riese, H., & Cramer, A. O. J. (2018). Personalized Network Modeling in Psychopathology: The Importance of Contemporaneous and Temporal Connections. *Clinical Psychological Science*, 6(3), 416–427.

<https://doi.org/10.1177/2167702617744325>

Fernandez, K. C., Fisher, A. J., & Chi, C. (2017). Development and initial implementation of the Dynamic Assessment Treatment Algorithm (DATA). *PLOS ONE*, 12(6), e0178806.

<https://doi.org/10.1371/journal.pone.0178806>

- Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., ... Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals based on patterns of brain connectivity. *Nature Neuroscience*, *18*(11), 1664–1671. <https://doi.org/10.1038/nn.4135>
- Fisher, A. J., & Boswell, J. F. (2016). Enhancing the Personalization of Psychotherapy With Dynamic Assessment and Modeling. *Assessment*, *23*(4), 496–506. <https://doi.org/10.1177/1073191116638735>
- Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, *201711978*. <https://doi.org/10.1073/pnas.1711978115>
- Fisher, Z. F., Bollen, K. A. & Gates, K. M. (2019). A Limited Information Estimator for Dynamic Factor Models, *Multivariate Behavioral Research*, DOI: 10.1080/00273171.2018.1519406
- Fisher, Z. F., Bollen, K. A., Gates, K. M. & Rönkkö, M. (2018). MIIVsem: Model Implied Instrumental Variable (MIIV) Estimation of Structural Equation Models. R package version 0.5.3. <https://github.com/zackfisher/MIIVsem>
- Fuller-Tyszkiewicz, M., Hartley-Clark, L., Cummins, R. A., Tomy, A. J., Weinberg, M. K., & Richardson, B. (2017). Using dynamic factor analysis to provide insights into data reliability in experience sampling studies. *Psychological Assessment*, *29*(9), 1120–1128. <https://doi.org/10.1037/pas0000411>
- 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. C. M. (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. M., 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. M., Iyer, S. P., Nigg, J. T., & Fair, D. A. (2014). Organizing Heterogeneous Samples Using Community Detection of GIMME-Derived Resting State Functional Networks. *PLOS ONE*, *9*(3), e91322. <https://doi.org/10.1371/journal.pone.0091322>
- Gonzalez-Castillo, J., Hoy, C. W., Handwerker, D. A., Robinson, M. E., Buchanan, L. C., Saad, Z. S., & Bandettini, P. A. (2015). Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. *Proceedings of the National Academy of Sciences*, *112*(28), 8762–8767. <https://doi.org/10.1073/pnas.1501242112>
- 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](https://doi.org/10.1207/s15327906mbr4002_3)
- Henry, T., & Gates, K. (2017). Causal search procedures for fMRI: review and suggestions. *Behaviormetrika*, *44*(1), 193–225. <https://doi.org/10.1007/s41237-016-0010-8>
- Hu, Y., Nesselroade, J. R., Erbacher, M. K., Boker, S. M., Burt, S. A., Keel, P. K., ... Klump, K. (2016). Test Reliability at the Individual Level. *Structural Equation Modeling : A Multidisciplinary Journal*, *23*(4), 532–543. <https://doi.org/10.1080/10705511.2016.1148605>
- Huettel, S. A., Song, A. W., & McCarthy, G. (n.d.). *Functional Magnetic Resonance Imaging, Second Edition*. 8.
- Jöreskog, K. G., Sörbom, D., Aigner, D. J., & Goldberger, A. S. (1977). *Latent variables in socioeconomic models*.



- Lamiell, J. T. (1981). Toward an idiothetic psychology of personality. *American Psychologist*, *36*(3), 276–289. <https://doi.org/10.1037/0003-066X.36.3.276>
- Lane, S., Gates, K., Pike, H., Beltz, A., & Wright, A. G. C. (2018). Uncovering general, shared, and unique temporal patterns in ambulatory assessment data. *PsyArXiv*.  
<https://doi.org/10.17605/OSF.IO/D3NK9>
- Lane, S. T., & Gates, K. M. (2017). Automated Selection of Robust Individual-Level Structural Equation Models for Time Series Data. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(5), 768–782. <https://doi.org/10.1080/10705511.2017.1309978>
- Laumann, T. O., Gordon, E. M., Adeyemo, B., Snyder, A. Z., Joo, S. J., Chen, M.-Y., ... Petersen, S. E. (2015). Functional System and Areal Organization of a Highly Sampled Individual Human Brain. *Neuron*, *87*(3), 657–670. <https://doi.org/10.1016/j.neuron.2015.06.037>
- lavaan: An R Package for Structural Equation Modeling | Rosseel | Journal of Statistical Software*. (n.d.).  
<https://doi.org/10.18637/jss.v048.i02>
- Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*(7197), 869–878. <https://doi.org/10.1038/nature06976>
- Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's Congruence Coefficient as a Meaningful Index of Factor Similarity. *Methodology*, *2*(2), 57–64. <https://doi.org/10.1027/1614-2241.2.2.57>
- Miller, M. B., Van Horn, J. D., Wolford, G. L., Handy, T. C., Valsangkar-Smyth, M., Inati, S., ... Gazzaniga, M. S. (2002). Extensive Individual Differences in Brain Activations Associated with Episodic Retrieval are Reliable Over Time. *Journal of Cognitive Neuroscience*, *14*(8), 1200–1214.  
<https://doi.org/10.1162/089892902760807203>
- 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 & Perspective*, 2(4), 201–218. [https://doi.org/10.1207/s15366359mea0204\\_1](https://doi.org/10.1207/s15366359mea0204_1)
- Molenaar, P. C. M., & Campbell, C. G. (2009). The New Person-Specific Paradigm in Psychology. *Current Directions in Psychological Science*, 18(2), 112–117. <https://doi.org/10.1111/j.1467-8721.2009.01619.x>
- Molenaar, P. C. M., & Nesselroade, J. R. (2012). Merging the Idiographic Filter With Dynamic Factor Analysis to Model Process. *Applied Developmental Science*, 16(4), 210–219. <https://doi.org/10.1080/10888691.2012.722884>
- 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>
- Nesselroade, J. R., Gerstorf, D., Hardy, S. A., & Ram, N. (2007). Focus Article: Idiographic Filters for Psychological Constructs. *Measurement: Interdisciplinary Research & Perspective*, 5(4), 217–235. <https://doi.org/10.1080/15366360701741807>
- Price, R. B., Gates, K., Kraynak, T. E., Thase, M. E., & Siegle, G. J. (2017). Data-Driven Subgroups in Depression Derived from Directed Functional Connectivity Paths at Rest. *Neuropsychopharmacology*, 42(13), 2623–2632. <https://doi.org/10.1038/npp.2017.97>
- Ram, N., & Grimm, K. J. (2009). Growth Mixture Modeling: A Method for Identifying Differences in Longitudinal Change Among Unobserved Groups. *International Journal of Behavioral Development*, 33(6), 565–576. <https://doi.org/10.1177/0165025409343765>
- Ramsey, J. D., Hanson, S. J., & Glymour, C. (2011). Multi-subject search correctly identifies causal connections and most causal directions in the DCM models of the Smith et al. simulation study. *NeuroImage*, 58(3), 838–848. <https://doi.org/10.1016/j.neuroimage.2011.06.068>

- Rosseel, Y. (2012). lavaan : An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2). Retrieved from <https://doaj.org>
- Rubel, J. A., Fisher, A. J., Husen, K., & Lutz, W. (2018). Translating Person-Specific Network Models into Personalized Treatments: Development and Demonstration of the Dynamic Assessment Treatment Algorithm for Individual Networks (DATA-IN). *Psychotherapy and Psychosomatics*. <https://doi.org/10.1159/000487769>
- Sargan, J. D. (1958). The Estimation of Economic Relationships using Instrumental Variables. *Econometrica, 26*(3), 393–415. <https://doi.org/10.2307/1907619>
- Sartorius, N., Üstün, T. B., Lecrubier, Y., & Wittchen, H.-U. (1996). Depression comorbid with anxiety: Results from the WHO study on “Psychological disorders in primary health care.” *The British Journal of Psychiatry, 168*(Suppl 30), 38–43.
- Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., ... Woolrich, M. W. (2011). Network modelling methods for FMRI. *NeuroImage, 54*(2), 875–891. <https://doi.org/10.1016/j.neuroimage.2010.08.063>
- Wisner, K. M., Atluri, G., Lim, K. O., & MacDonald, A. W. (2013). Neurometrics of intrinsic connectivity networks at rest using fMRI: Retest reliability and cross-validation using a meta-level method. *NeuroImage, 76*, 236–251. <https://doi.org/10.1016/j.neuroimage.2013.02.066>
- Wright, A. G. C., Hallquist, M. N., Morse, J. Q., Scott, L. N., Stepp, S. D., Nolf, K. A., & Pilkonis, P. A. (2013). Clarifying Interpersonal Heterogeneity in Borderline Personality Disorder Using Latent Mixture Modeling. *Journal of Personality Disorders, 27*(2), 125–143. <https://doi.org/10.1521/pedi.2013.27.2.125>
- Wright, A. G. C., & Hopwood, C. J. (2016). Advancing the Assessment of Dynamic Psychological Processes. *Assessment, 23*(4), 399–403. <https://doi.org/10.1177/1073191116654760>



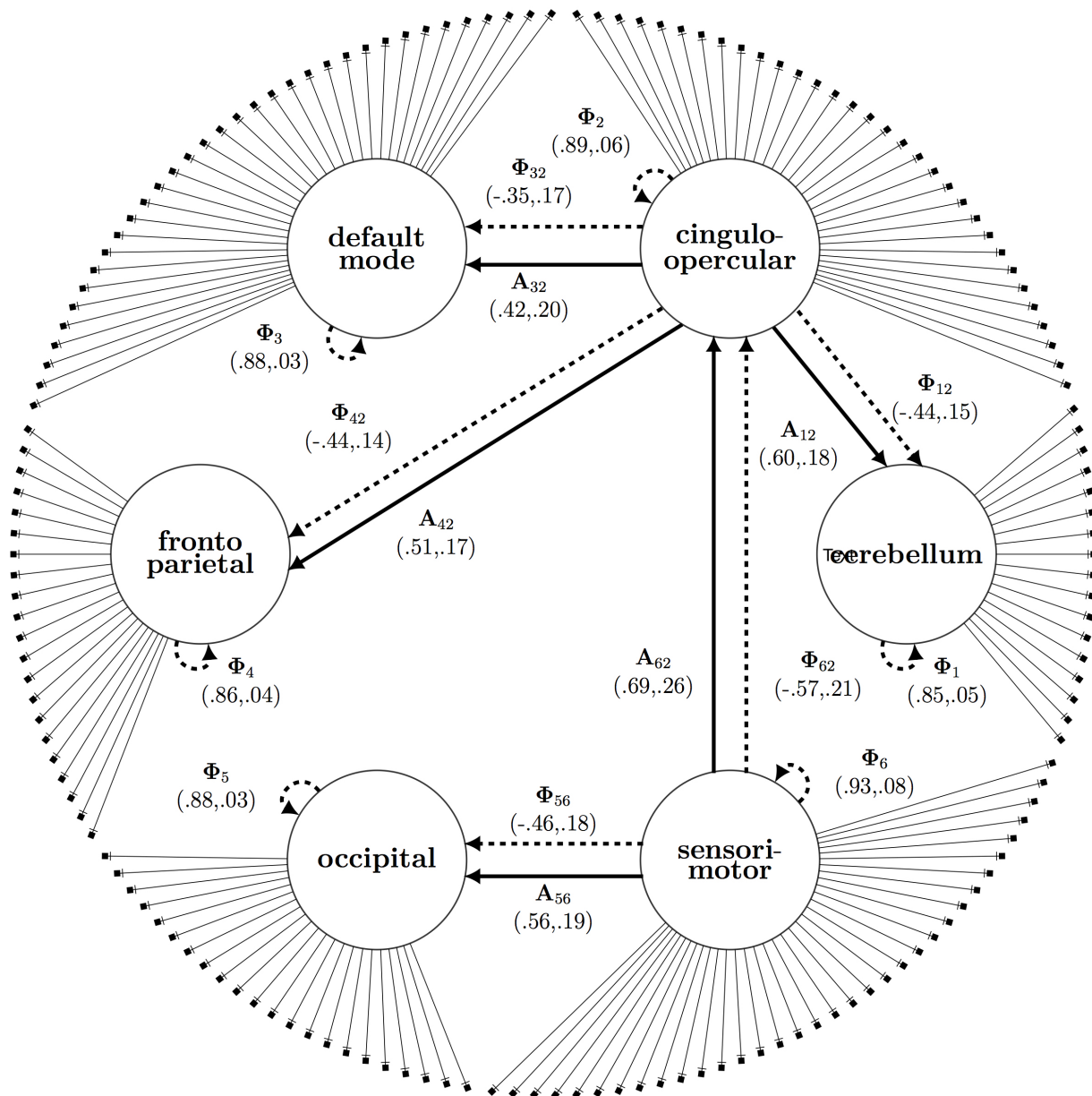
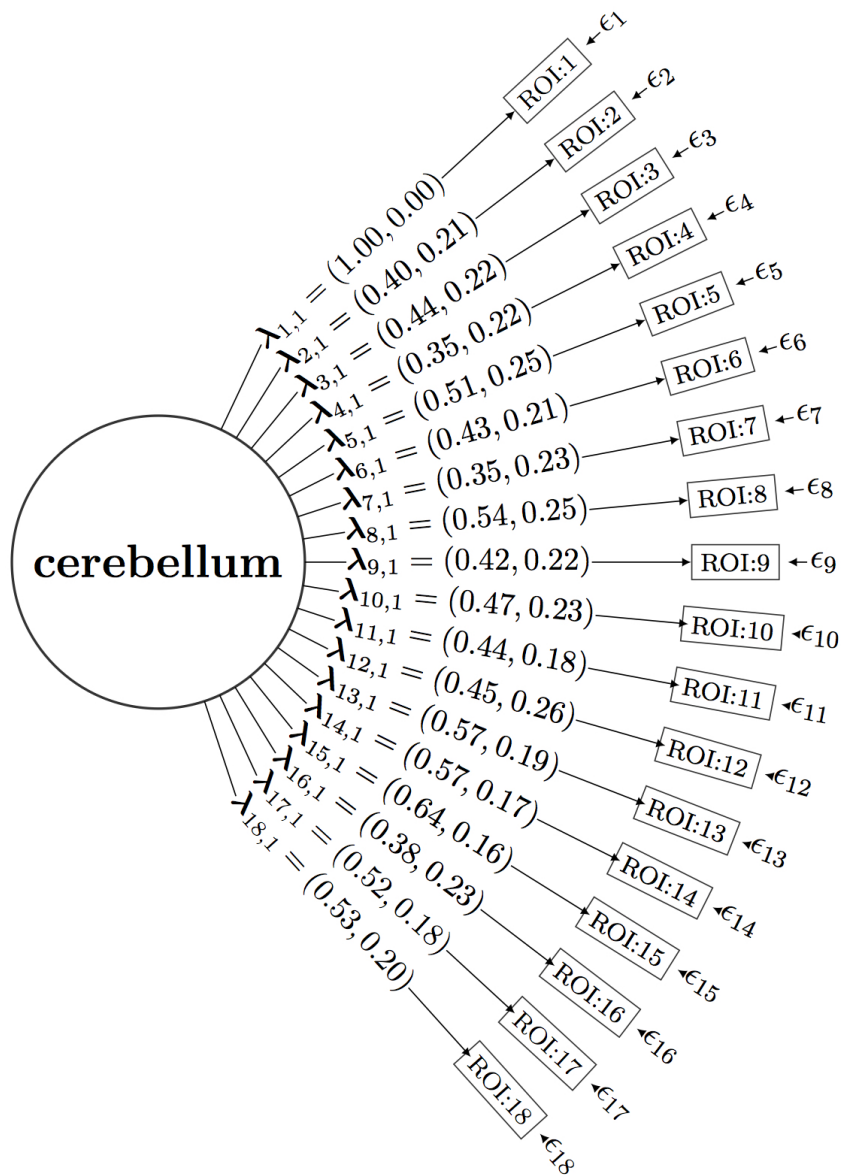
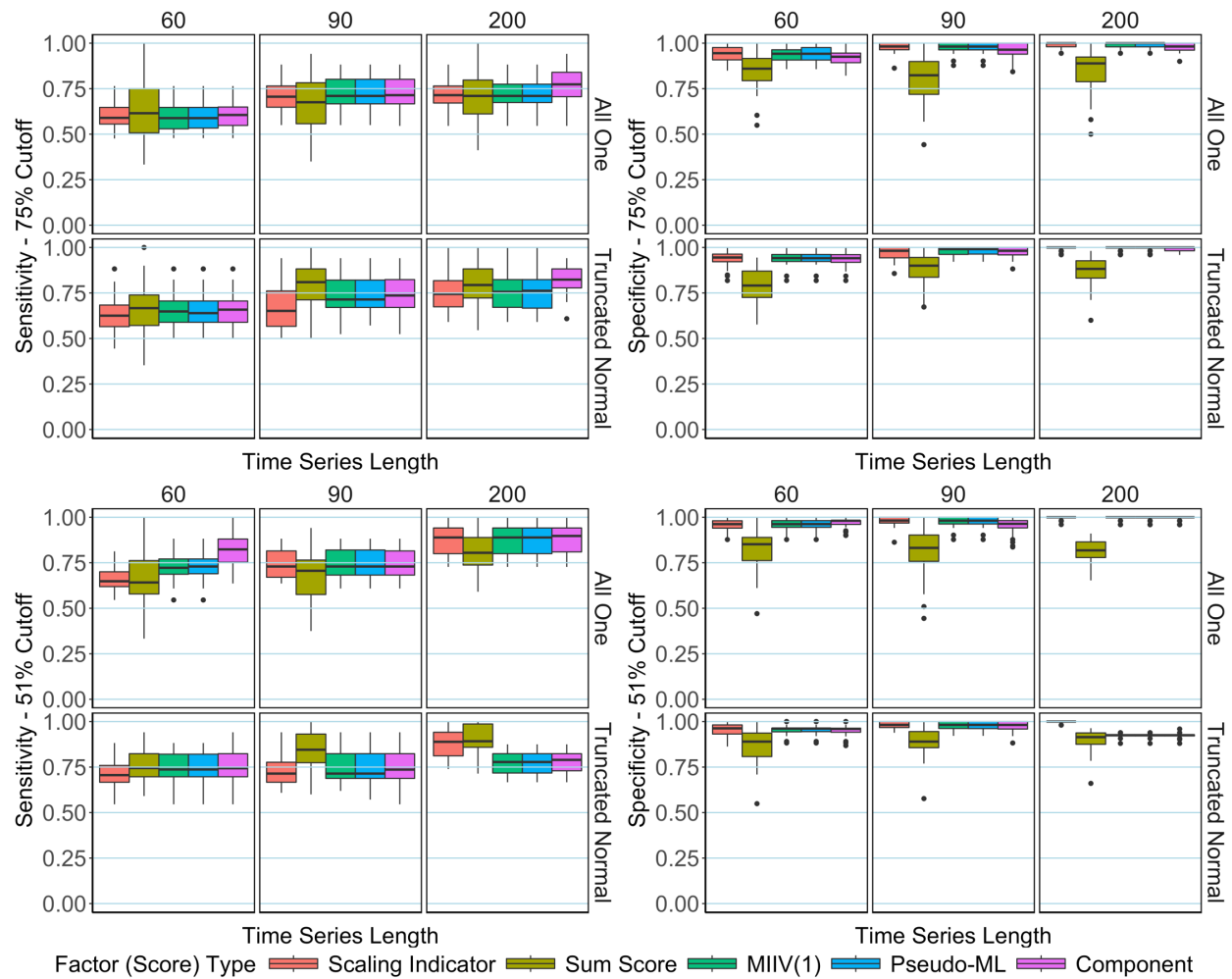


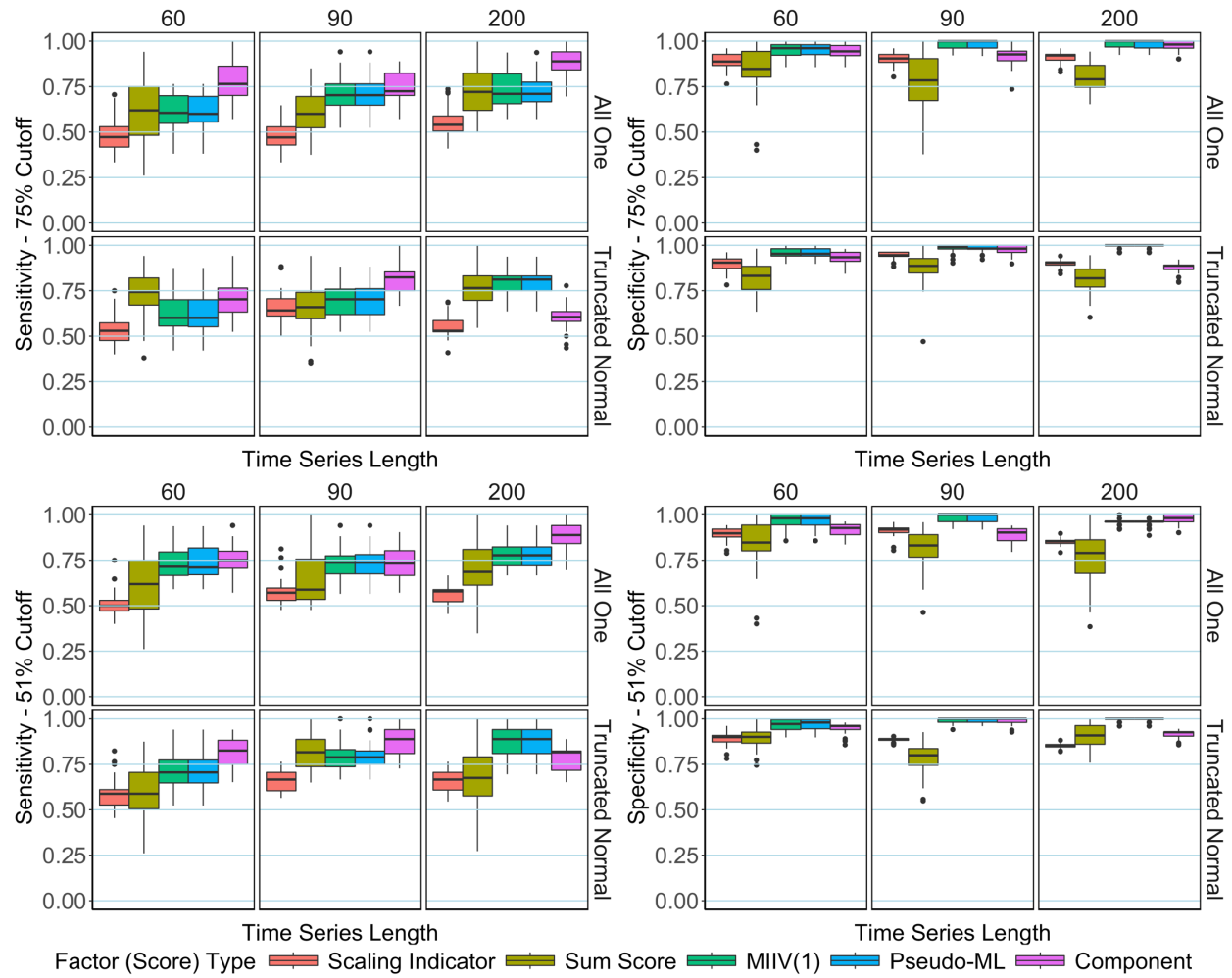
Figure 1. Depiction of GIMME results obtained from sum scores of network brain regions. Values in parentheses represent means and standard deviations for parameter estimates across individuals. Dashed lines indicate lagged relations, solid lines depict contemporaneous relations. Individual-level paths not depicted here. Squares at the end of the arrows relating to circles reflect the number of indicators used for each latent construct (i.e., observed brain regions that load onto a given latent network).



*Figure 2.* Measurement Model parameter estimates for Cerebellum. The cerebellum network (i.e., latent variable) has 18 brain regions of interest (indicated as ROI) that load on it. This figure shows the average lambda estimates obtained via MIIV-2SLS and standard deviations in the parentheses.



*Figure 3.* Sensitivity and specificity results for correctly specified model on simulated data when the cutoff for the proportion representing the majority during group-level path selection = 75% (top panel) and 51% (bottom panel).



**Figure 4.** Sensitivity and specificity results for misspecified model on simulated data when the cutoff for the proportion representing the majority during group-level path selection = 75% (top panel) and 51% (bottom panel).



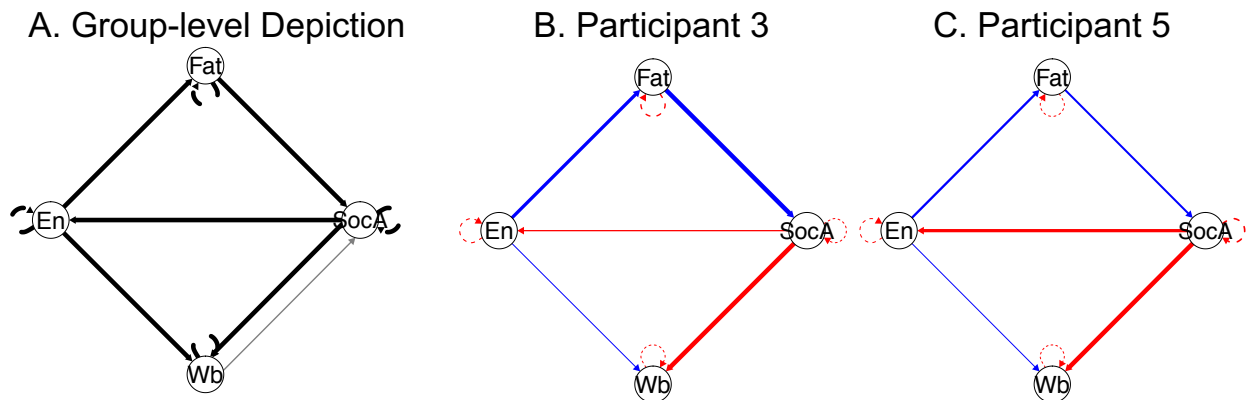


Figure 5. Depiction of (A) group level and two exemplar individual's latent variable model results (B,C). All figures provided directly from the *gimme* package output. While the pattern of relations among latent variables is consistent, the weights differ. The measurement models (available in the Supplemental Material) reveal much variability in the measurement of these constructs for the individuals. Red indicates positive values (hot), blue indicates negative values (cold). Dashed lines are lagged, solid contemporaneous.

## Appendix

Example of code for running LV-GIMME within the *gimme* R package. Here, there are 3 latent constructs (factors) and 9 observed variables.

```
#Define measurement model structure
lv_model <- '
  L1 =~ V1 + V2 + V3
  L2 =~ V4 + V5 + V6
  L3 =~ V7 + V8 + V9
'

# Run LV-GIMME by providing the measurement model structure
fit <- gimmeSEM(
  data      = data_list, # list of data
  lv_model  = replicate(length(data_list),lv_model),
  lv_scores = "regression",
  lv_estimator = "miiv",
  lv_miiv_scaling = "first.indicator",
  lv_final_estimator = "miiv"
)

# Examine estimates for factor loadings and paths among latent
# variables:
fit$lvgimme$miiv_est_table
```