MULTIVARIATE BEHAVIORAL RESEARCH
2024, VOL. 59, NO. 5, 1019–1042

OPEN ACCESS

Path and Direction Discovery in Individual Dynamic Factor Models: A Regularized Hybrid Unified Structural Equation Modeling with Latent Variable

Ai Ye[a] and Kenneth A. Bollen[b]

[a]Lehrstuhl für Psychologische Methodenlehre & Diagnostik, Department Psychologie, Ludwig-Maximilians-Universität München, Munich, Germany; [b]Department of Psychology and Neuroscience, Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

ABSTRACT
There has been an increasing call to model multivariate time series data with measurement error. The combination of latent factors with a vector autoregressive (VAR) model leads to the dynamic factor model (DFM), in which dynamic relations are derived within factor series, among factors and observed time series, or both. However, a few limitations exist in the current DFM representatives and estimation: (1) the dynamic component contains either directed or undirected contemporaneous relations, but not both, (2) selecting the optimal model in exploratory DFM is a challenge, (3) the consequences of structural misspecifications from model selection is barely studied. Our paper serves to advance DFM with a hybrid VAR representations and the utilization of LASSO regularization to select dynamic implied instrumental variable, two-stage least squares (MIIV-2SLS) estimation. Our proposed method highlights the flexibility in modeling the directions of dynamic relations with a robust estimation. We aim to offer researchers guidance on model selection and estimation in person-centered dynamic assessments.

KEYWORDS
Time series data; hybrid unified SEM; regularized SEM; dynamic factor model; model implied instrumental variable two-stage least square

With the development of technology to collect intensive longitudinal or time series data (TSD) in a fast and economical pace, recent decades have witnessed a surge of psychological and neurological research at the individual level (i.e., $N = 1$). Studies focused on person-specific dynamic assessment, i.e., the so-called idiographic approaches, emphasize intra-individual characteristics and development over time. This contrasts to the typical nomothetic approach that draws general inferences from a sample of individuals representing the population of interest using their interindividual differences (Hamaker, 2012; Molenaar, 2004)[1]. Along with the shift of research interest is the need for advancing statistical methods to facilitate person-specific dynamic assessments using TSD. Within the psychometric field, researchers have developed modeling frameworks to fit traditional time series models, such as the Vector Autoregressive (VAR), deeply rooted in statistics (Hamilton, 1994; Lütkepohl, 2005).

The psychometric literature on modeling individual dynamic models on a manifest level has grown (e.g., Epskamp et al., 2018; Gates et al., 2020; Ye et al., 2021). These models estimate the dynamic relations (e.g., temporal and contemporaneous) unpacked in a multivariate TSD. For instance, two representative approaches are the unified Structural Equation Model (uSEM; Gates et al., 2010; Kim et al., 2007), as a time series extension of the SEM, and the graphical VAR (gVAR; Epskamp et al., 2018) model, as a time series extension of the general Gaussian Graphical Model framework. Recently, researchers have discussed the extended VAR model with hybrid representations that can handle both the direct causal effects and undirected contemporaneous associations (Molenaar & Lo, 2016; Ye et al., 2021). These approaches differ in the variant of VAR representation and in the estimation framework to select and identify the optimal model. For instance, uSEM is usually identified by stepwise model search algorithms (Gates & Molenaar, 2012), while gVAR (Epskamp et al., 2018) or hybrid uSEM (Ye et al., 2021) adopt some machine learning methods (e.g., regularization) to identify and estimate the optimal sparse model.

Typically, these modeling methods incorporate a small number of manifest variables without accounting for measurement error. In practice, it is common that more than one indicators measure the same underlying dynamic latent variable. With multiple indicators, latent time series variables could be formed to adjust for measurement error and to reduce the dimensions of observed variables. The combination of a factor model and a time series model results in what is called the dynamic factor model (DFM; Browne & Nesselroade, 2005; Molenaar, 1985). In DFM, dynamic relations (e.g., lagged and contemporaneous relations) are allowed either within the factor series or amongst the factor and the observed time series. In fact, current dynamic modeling approaches have been extended to include a factor model within their restricted VAR version. For example, the uSEM model has been intergrated with latent variables (i.e., LV-uSEM; Gates et al., 2020). In addition, the gVAR model has been combined with a factor model to form the latent variable gVAR (or LV-gVAR; Epskamp, 2020).

However, the specification of the more flexible VAR representation of DFM remains to be developed. Therefore, the primary purpose of our paper is to extend the hybrid uSEM with regularization in Ye et al. (2021) to the regularized hybrid uSEM with latent variables, so that we can estimate a sparse DFM that allows for hybrid contemporaneous dynamic relations between the latent factors themselves. Three steps address this overarching goal: the first is to reform the structural model of the latent variable uSEM (LV-uSEM) to its hybrid version, which the authors refer to as the latent variable hybrid uSEM (or LV-huSEM); the second is to perform model selection using the LASSO regularization in the search for the optimal sparse LV-huSEM; lastly, post-model selection estimation will be implemented to obtain robust parameter estimates of the final optimal sparse LV-huSEM. To evaluate the proposed method with existing ones, a simulation study will be conducted to compare both the model recovery performance (sensitivity and specificity) of different model build methods when they are applied under the LV-huSEM context, as well as the biasedness and robustness of parameter estimates obtained by several estimation methods.

Similar to the alternative modeling framework we investigate, including uSEM, gVAR and their latent variable extensions, we remind the readers the assumptions underlying the type of data and process appropriate for the hybrid uSEM and its latent variable version. First, the current investigation focuses on individual dynamic modeling analysis in the time domain, we acknowledge that there are also analytical methods that are carried out in the frequency domain (e.g., Macaro & Prado, 2014; Molenaar, 1987). In addition, these methods apply to multivariate TSD that is: (1) weakly stationary (i.e., constant mean and variance component); and (2) discrete with equally spaced measurements (the equal distance assumption is less important when the interval is extremely short). These assumptions are necessary for the VAR-based model specification and for achieving asymptotically consistent estimation and also asymptotically unbiased estimation for some parameters. Although extensions of DFMs with nonstationary time series (e.g., Chow et al., 2011; Molenaar et al., 1992) or unequal measurement intervals (e.g., Driver et al., 2015; Ryan et al., 2018) have been developed, the current investigation focuses on the standard case. Later, we discuss the implications, applications, and extensions of these assumptions.

Below, we first introduce the basic concepts, specification, and estimation of DFM, followed by a review of a SEM-based framework to estimate an individual DFM as the LV-uSEM. Next, we point out the issues and limitations of the current approaches, and how the proposed method addresses these issues. This leads to a simulation study to evaluate and compare the proposed method with the existing ones. Finally, conclusions and discussion are drawn from the results of the simulation study.

Dynamic factor models

Dynamic Factor Models (DFMs) represent a class of models that includes lagged relations within the latent variable approach (Browne & Nesselroade, 2005; Molenaar, 1985). It can also be seen as a factor analysis extension to the family of VAR models in the sense that latent variables (or, factor series) are defined in a measurement model and that permits lagged relations either in the measurement model, in the structural (i.e., latent variable) model, or both (Molenaar, 1985). Indeed, DFM is a synthesis of factor analysis and VAR. Such a synthesis is ideal for many psychological studies that aim to unravel hybrid relations in unobserved dynamic processes. A substantive question that can be investigated by a DFM approach is to what extent an increase in a latent dynamic construct (e.g., anxiety) predicts changes in another latent dynamic construct (e.g., depression) as well as changes in the indicators of the other latent construct. As another example, neuroscientists often aim to study functional connectivity in human brains. The latent constructs in this context could be some unknown "brain networks" formed by a cluster of disparate brain regions that tend to interact across time when performing a task (Gates et al., 2020). Integrating a factor model component in the dynamic model opens up the possibility to explore dynamics among latent constructs underlying what we could observe.

A general DFM for a single-subject multivariate TSD is defined by two components, the measurement model and the structural or latent variable model (Molenaar, 1985; Zhang et al., 2008). Recall that traditional DFM typically applies to weakly stationary time series measured at equidistant intervals. Under these assumptions, model parameters are constrained to be time-invariant. Let $Y_t = [y_{1t}, y_{2t}, ..., y_{pt}]^T$ denote a vector of a p-variate time series at a given time point t, with $t = 1, ..., T$. Assuming Y_t represents a weakly stationary linear time series (i.e., with a constant mean, variance and covariance function). To ease the presentation, it is assumed that all the time series have zero mean function (i.e., no intercept term):

$$Y_t = \sum_{u=0} \Lambda(u)\eta(t-u) + \epsilon_t, \epsilon_t \sim N(\mathbf{0}, \mathbf{\Theta}). \quad (1)$$

$$\eta_t = \sum_{u=1} \Phi(u)\eta(t-u) + \zeta_t, \zeta_t \sim N(\mathbf{0}, \mathbf{\Psi}). \quad (2)$$

In the measurement model, the $\eta(t-u)$ is a q-variant set of latent factor series, with the (p, q)-dimensional time-invariant factor loading matrix $\mathbf{\Lambda}(u), u = 0, 1, ..., l$ that denotes the linear relations between the original p-variant time series Y_t and the q-variant factor series η_t at the lag order of u. The ϵ_t is a p-variate measurement error process for the p-variate observed variables in Y_t. We assume that the unique factors ("errors") are independent over time and no cross-loadings. The structural (latent variable) equation in the DFM is a dynamic process of VARMA(m, n), i.e., a VAR of order n with a MA of order m. The $\mathbf{\Phi}(u), u = 1, 2, ..., n$ is a sequence of (q, q)-dimensional matrices of AR and cross-lagged effect of the latent factors at the lag order of u. In Model (2), the parameters are time invariant. In addition, we assume that the errors $\boldsymbol{\epsilon_t}$ and $\boldsymbol{\zeta_t}$ are uncorrelated with their respective latent variables $\eta(t-u)$, the errors are uncorrelated over time, i.e., $cov(\boldsymbol{\zeta_t}, \boldsymbol{\zeta_{t-1}}) = 0$. Here, moving average (MA) coefficient matrices (i.e., current latent variable values predicted by errors from the prediction of previous latent variable values) are not not considered ($m = 0$). In this way, the structural (latent variable) equation in the DFM is a dynamic process of VAR(n), i.e., a VAR of order n. Taken together with the measurement model, this returns a DFM (p, q, l, m, n). The general DFM implies that lagged values of the latent variable can have loadings on future values of the indicators beyond the indirect effect *via* the factor at that concurrent time point.

For model identification and substantive purposes, however, analysts often impose restrictions to allow only one type of the lagged relations. For example, a general DFM is reduced to a simpler, more restrictive version containing only lag relations among the factors, i.e., DFM $(p, q, 0, m, n)$, called the process factor analysis or direct autoregressive factor models (PFA or DAFM; Browne & Nesselroade, 2005); alternatively, it reduces to what is called the shock factor analysis or white noise factor model when lagged relations are only in the measurement model, i.e., DFM $(p, q, l, 0, 0)$ (Molenaar, 1985).

Estimation framework and approaches for DFM

Substantial research in psychometrics and other fields has focused on how to estimate variants of DFMs. Zhang et al. (2008) provided a comprehensive review of four major estimation methods for DFMs: (1) a Kalman filter (KF) algorithm based on a state space model (SSM) representation; (2) the pseudo-ML method based on the construction of a block-Toeplitz covariance matrix in the SEM framework; (3) a least squares (LS) method that also employs the block-Toeplitz matrix; and 4) a Bayesian framework using the Markov Chain Monte Carlo (MCMC) Gibbs sampling under the SSM specification. Their simulation study has shown that all four methods reach appropriate parameter estimates with comparable precision. In addition, many extensions and alternative variations of each estimation method have been proposed. For instance, SSM with KF has been extended to handle nonlinear and nonstationary DFMs (Chow & Zhang, 2013). LS estimators (Browne & Zhang, 2005) and asymptotically distribution-free methods (Molenaar & Nesselroade, 1998) based on the block Toeplitz matrix have been adopted. Bayesian approaches using Gibbs sampling for categorical (Zhang & Nesselroade, 2007) variables have also been developed. As ML and pseudo-ML estimations are still the dominant approaches for single-subject DFM, we highlight some characteristics of each below.

State space model and Kalman filter. The SSM framework encompasses the KF and the Kalman smoothers, which have been common tools in econometrics and engineering to track changes and make predictions in dynamic systems (Dolan, 2002; Kalman, 1960; Shumway & Stoffer, 2004). KF predicts current or future states (i.e., factor scores) given information up to the current time point by minimizing prediction errors (Zarchan & Musoff, 2000) and provides true ML estimates for DFM (Chow et al., 2011). In other words, KF can be regarded as a factor score in uSEMs, or more broadly, a "latent variable" estimation procedure. Chow et al. (2010) provides a comprehensive review of the equivalence and difference between the SSM estimates and those of SEM under various model specifications. Molenaar (1985) has shown that any general DFM (p, q, l, m, n) can be rewritten in the standard SSM form. SSM with time-invariant coefficients corresponds to the special case of DFM that lacks the factor loadings, i.e., DFM $(p, q, 0, m, n)$, or PFA or DAFM.

Block Toeplitz matrix under the SEM framework. Another common method is the pseudo-maximum likelihood (i.e., pseudo-ML[2]; Molenaar & Nesselroade, 1998) or least square estimation using the block Toeplitz matrix under the SEM framework (Molenaar, 1985). The block Toeplitz matrix can be thought of as a moment estimator with lagged autocovariance of the observed variables. A downside of the block Toeplitz matrix is a large number of redundant parameters, because the diagonals of a block Toeplitz matrix are the same but they are estimated as unique (i.e., SEM programs by Zhang & Browne, 2010). Another method to obtain the pseudo-ML estimator under the SEM regime is to use time-embedded raw data (see Equation (4)). Nonetheless, the block Toeplitz remains a popular method for estimating DFM, because of the availability of SEM programs which facilitate the specification of measurement models with contemporaneous relations in the structural model.

[2]The term "pseudo-ML" in the current work specifically refers to the application of using maximum likelihood estimation for TSD where the independence assumption is violated due to temporal dependence between repeated observations, as used in time series literature such as Molenaar and Nesselroade (1998).

In comparison, the SEM estimators based on block Toeplitz or the time-embedded data are good practice for estimating lag-1 and contemporaneous relations, but they are inefficient for complex dynamics such as higher-order lags in the observed or in the error term due to the data structure (very large and sparse covariance matrix to be computed). In contrast, KF are more flexible when modeling time-specific coefficients or higher-order complex dynamic relations. Therefore, while uSEMs emphasize constant dynamic relations taking place instantaneously or at consecutive timepoints, SSMs allow for changing dynamic relations that may also last longer. Another benefit of KF is that subject-specific longitudinal factor scores are a by-product, which provides a reasonable approximation to the true factor scores (Chow et al., 2010). While in the traditional SEM methods, an additional step is needed to compute the factor scores following traditional methods such as Bartlett (Bollen, 1989). This is an issue in some current SEM methods that we aim to resolve in our proposed method.

Differences are also seen in the statistical properties of the parameter estimates obtained by the different approaches. The SSMs with KF produce unbiased ML and standard error estimates, because the within-person time dependency in the data is accounted for explicitly. In contrast, pseudo-ML estimates obtained under the SEM are consistent estimators under the weakly stationary assumptions, but the unbiasedness property is more mixed. Studies by Hamaker et al. (2002), van Buuren (1997), and Zhang et al. (2008) found a consistent pattern in the investigation of SEM estimates for TSD with large T: the pseudo-ML estimates were asymptotically unbiased for VAR parameters associated with observable variables; however, the estimates associated with the MA part of a VARMA model or a pure MA model were biased. Since our investigation will focus on only the VAR parameters, this downside of pseudo-ML estimators from SEM does not matter to the current case.

In summary, SEMs and SSMs are very general modeling approaches for DFMs, representing the dynamic relations of a set of latent and manifest variables. For the current work, SEM is a convenient choice to represent simultaneous structural relations among observed and latent variables and it can be generalized to define the measurement model structures. In contrast, SSMs are better suited when the purpose is to represent more complex intraindividual dynamics with restricted contemporaneous relations. Because our intention is to maximize the flexibility in the directionality of contemporaneous relations, rather than having complex dynamics with longer lags, we limit our current extension and investigation to SEM-based approaches.

Latent variable group iterative multiple model estimation

Latent Variable Group Iterative Multiple Model Estimation (LV-GIMME; Gates et al., 2020) is one of the most flexible model selection and estimation methods for exploratory DFM that operates under the SEM framework with pseudo-ML estimation. LV-GIMME adopts the first-order DAFM model or the DFM $(p, q, 0, 0, 1)$ with lag-1 relations only between latent factors in the structural or latent variable model. The DAFM makes the interpretation and the implementation of LV-uSEM simpler, because the dynamic relations are between factors, which can be separate from the measurement model where factors are extracted from indicators at the same time. The dynamic relations between the latent variables in the LV-uSEM are comparable to those among the observed variables in the uSEM model, with the exception that these dynamic relations take account of measurement errors in the observed variables. Because our aim is to compare approaches of single-subject dynamic models, we focus on individual TSD, that is, no group-level modeling or aggregation of individual models will be performed. In the remaining part of this paper, we refer to this model as LV-uSEM to distinguish from the LV-GIMME that can also apply to multiple subjects TSD.

In the measurement model, LV-uSEM adopts the general form of a first-order DAFM, i.e., DFM $(p, q, 0, 0, 1)$. This part uses a confirmatory factor approach to obtain latent variables with the same qualitative meaning, i.e., $\mathbf{\Lambda}$ has the same structure but individual-specific parameter values across individuals. For each individual TSD:

$$Y_t = \mathbf{\Lambda}\eta_t + \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{\Theta}). \tag{3}$$

In the structural or latent variable model, LV-uSEM inherits a uSEM structure (Gates et al., 2011; Kim et al., 2007) that "unifies" temporal ordering dependency and contemporaneous associations among the latent factors. This part of the model assumes a sparse structure (although unknown), and the estimation requires an exploratory search for the optimal sparse pattern. The following model specification and matrix notations follow a SEM convention that is used in Gates et al. (2017) and Ye et al. (2021).

$$\boldsymbol{\eta} = \boldsymbol{B}\eta + \zeta, \zeta \sim N(\mathbf{0}, \mathbf{\Psi}). \tag{4}$$

where $\boldsymbol{\eta} = [\boldsymbol{\eta}_{t-1}, \boldsymbol{\eta}_t]$ is a $2q \times T$ matrix. That is, the factor variables are time-embedded by appending the data at $t-1$ to the data at t. The data are thus expanded to two consecutive time points $t-1$ and t, so time series vector of lagged (exogenous) factor variables and those of contemporaneous (endogenous) factor variables are appended horizontally. This requires that the error vector ζ also be extended, as well as each of the corresponding matrices. Note that the estimator from the time-embedded raw data is also pseudo-ML due to temporal dependency between observations of TSD, which has essentially the same asymptotic properties as those obtained by the block Toeplitz matrix of the lagged auto-covariances (Molenaar, 1985).

The contemporaneous and lagged regression coefficients collapse into a single $2q \times 2q$ $\boldsymbol{B}$ regression coefficient matrix:

$$\boldsymbol{B} = \begin{pmatrix}
0 & \cdots & \cdots & \cdots & 0 & 0 & \cdots & \cdots & \cdots & 0 \\
\vdots & \ddots & & & \vdots & \vdots & \ddots & & & \vdots \\
\vdots & & \ddots & & \vdots & \vdots & & \ddots & & \vdots \\
\vdots & & & \ddots & \vdots & \vdots & & & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & 0 & 0 & \cdots & \cdots & \cdots & 0 \\
\phi_{11} & \cdots & \cdots & \cdots & \phi_{1q} & 0 & a_{12} & \cdots & \cdots & a_{1q} \\
\vdots & \ddots & & & \vdots & a_{21} & \ddots & \ddots & & \vdots \\
\vdots & & \ddots & & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
\vdots & & & \ddots & \vdots & \vdots & & \ddots & \ddots & a_{(q-1)q} \\
\phi_{q1} & \cdots & \cdots & \cdots & \phi_{qq} & a_{q1} & \cdots & \cdots & a_{q(q-1)} & 0
\end{pmatrix}_{2q \times 2q}$$

in which the upper left $q \times q$ and the upper right $q \times q$ matrix block coefficients are set to zero for lagged factor variables. That is, no factor variable at time t directly predicts a factor variable at time $t-1$, and factor variables at time $t-1$ cannot predict each other. The lower left and right blocks are $\boldsymbol{\Phi}$ parameters for lag-1 relations and $\boldsymbol{A}$ for contemporaneous factor variables at time, respectively.

The error matrix $\boldsymbol{\Psi}$ is a diagonal matrix representing an independent white noises:

$$\boldsymbol{\Psi} = \begin{pmatrix} \psi_{11} & & & & & & & & & \\ \psi_{21} & \ddots & & & & & & & & \\ \vdots & \ddots & \ddots & & & & & & & \\ \vdots & & \ddots & \ddots & & & & & & \\ \psi_{q1} & \cdots & \cdots & \psi_{q(q-1)} & \ddots & & & & & \\ 0 & \cdots & \cdots & \cdots & 0 & \ddots & & & & \\ \vdots & \ddots & & & \vdots & 0 & \ddots & & & \\ \vdots & & \ddots & & \vdots & \vdots & \ddots & \ddots & & \\ \vdots & & & \ddots & \vdots & \vdots & & \ddots & \ddots & \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & \cdots & \cdots & 0 & \psi_{(2q)(2q)} \end{pmatrix}_{2q \times 2q}$$

where the upper left triangle ($\boldsymbol{\Psi_{t-1}}$) contains freely estimated variance and covariances among the lagged factor variables, and the lower right triangle ($\boldsymbol{\Psi_t}$) is the variance and covariance matrix of factor residuals. Note that the covariances among factor residuals are set to zero to represent conditional independency. These matrices provide the foundation from which the extension to the latent variable hybrid uSEM is developed.[3]

A note on the level of heterogeneity across individuals

It is possible that individuals vary both in the latent constructs as well as the relations among latent factors, a scenario where ergodicity does not hold. But the interpretation of the dynamic relations is hard to generalize across people if the latent factors stand for completely different constructs. For this reason, usually some sort of partial ergodicity assumption is imposed. For instance, a typical practice in the GIMME framework is to have a qualitatively homogeneous and quantitatively heterogeneous measurement models (i.e., same nonzero pattern yet individual-specific estimates in $\boldsymbol{\Lambda}$) to ensure the same substantive meaning of the latent factor with some levels of individual variability (Gates et al., 2020). To maximize heterogeneity in the dynamic patterns, the structural models represent entirely idiographic dynamic processes (i.e., person-specific $\boldsymbol{B}$), suggesting partial ergodicity only in the measurement model. Note that this setting is opposite to the practice in idiographic filter DFM (Molenaar, 1985; Nesselroade et al., 2007), where the measurement model structure and factor loading are allowed to vary across individuals on the condition that the structural patterns among the latent variables remain invariant across people, a case of partial ergodicity in the dynamic (structural) model. This is a critical distinction since previous literature has pointed to great variability across individuals both in how constructs are measured and how they relate to each other (Epskamp et al., 2018; Gates et al., 2020; Hamaker et al., 2005). GIMME represents a very flexible approach that maximizes heterogeneity in dynamic patterns across individual DFMs. For this reason, our method stands on the same idiographic ground as does GIMME.

[3]We note that the full LV-uSEM in the above specification is an over-parameterized model. That is, if all the free parameters are nonzero, the model has negative degree of freedom and is under-identified. An underlying assumption is that the true DGM is a sparse model where many parameters in $\boldsymbol{B}$ and $\boldsymbol{\Psi}$ are zero elements. This is why model selection are implemented in these modeling approaches to identify the sparse pattern. This is also where the advantage of the stepwise searching in LV-GIMME (Gates et al., 2020) and the LASSO-regularization in hybrid uSEM (Ye et al., 2021) lies: that is, the model search algorithm can recover the optimal sparse model when the full model is under-identified.

The pseudo-ML model selection in GIMME

The LV-uSEM operates under the SEM framework for DFM that uses a pseudo-ML based model building algorithm and the model-implied instrumental variables with two-stage least squares (MIIV-2SLS; Fisher & Bollen, 1996; Fisher et al., 2019) for parameter estimation. Ideally, the specification for the structural (latent variable) model 3 should be guided by a priori theory. Unfortunately, very little is known about the individual dynamic pattern, besides that research has shown that the patterns vary across people (e.g., Nichols et al., 2014; Wright et al., 2015). As we noted before, model selection for a sparse model is needed to ensure identification of LV-uSEM. GIMME uses a data-driven forward selection algorithm where for every individual, it starts with a null model, and one path with the highest and significant modification indices is added iteratively until the model arrives at an acceptable fit (Gates et al., 2010). This model building procedure is automatic in the free open-source R package gimme (Lane et al., 2019).

The recommended estimation approach by LV-GIMME (Gates et al., 2020) is a sequential, three-step procedure: in the first stage, the measurement model is estimated with MIIV-2SLS, pseudo-ML, or PCA, and latent scores of factor series are derived through a regression method; the second stage is the general GIMME approach which involves a pseudo-ML based stepwise search to identify the sparse individual structural models (i.e., uSEMs) with individualized patterns of dynamic relations amongst the latent factor series obtained from the previous step; finally, parameter estimates of the measurement and the sparse structural model are obtained *via* MIIV-2SLS. The MIIV-2SLS has been shown to be more robust to model misspecification than the system wise ML estimations for traditional SEM (Bollen et al., 2007) and particularly for SEM under DFM framework (Fisher et al., 2019).

The MIIV-2SLS estimation

Technical details of the MIIV-2SLS are available in Fisher and Bollen (1996); Bollen et al. (2021, 2007). Here we give a brief description of the MIIV-2SLS. Unlike system-wise full-information estimators such as maximum likelihood that estimates all parameters simultaneously, the MIIV-2SLS estimates one equation at a time. To start, the MIIV-2SLS transforms the latent variable model into one that contains only the observed variables (i.e., L2O) by substituting each latent variable with its scaling indicator minus its error as specified in the measurement model (Bollen et al., 2021). One assumption for the OLS estimator is that the composite error cannot correlate with any covariates of that equation, which is typically violated after the L2O transformation. A common way to solve the correlated error issue is to use instrumental variables, i.e., variables that are uncorrelated with the composite error but correlate with the covariates that are associated with these errors. The special advantage of the MIIV-2SLS is that qualified instruments were drawn from other equation(s) within the system itself based on the model structure, hence it is called model-implied instrumental variables (MIIVs; Bollen et al., 2021). Finding qualified MIIVs is done automatically using the algorithm from Fisher & Bollen (1996) and is implemented in the R package *MIIVsem* (Fisher et al., 2020). Another advantage of the equation-by-equation estimation is that, besides the χ^2 tests of goodness-of-fit for the whole model, equation-wise overidentification test (e.g., Sargan's χ^2 test; Sargan, 1958) is available when the equation has more than the minimum number of MIIVs. A rejection to the null hypothesis of the overidentification test suggests that at least one of the MIIVs fail to meet the conditions based on the current model specification, which is an evidence that some specifications of the model is in error (Bollen et al., 2021).

To estimate a LV-uSEM using MIIV-2SLS estimator under the GIMME approach, one convenience is that the latent factor variables are obtained prior to the estimation of the structural model and are treated as observed variables in the model selection and estimation for structural relations. In addition, the lagged latent factor variables are predetermined (exogenous) variables because no backward predictions from factor scores at time T to those at the previous time are allowed (recall the zero elements in the $\boldsymbol{B}$ matrix in Equation (4)). This setting of the lagged factor variables guaranteed a minimum set of qualified MIIVs for the parameter estimates in the structural model.

The use of equation-by-equation estimation in MIIV-2SLS is particularly advantageous because it separates the estimation of the measurement model from the structural (latent variable) model so that the measurement model parameter estimates remain unaffected by heterogeneous structures in the latent variable model (Gates et al., 2020). Enabling individual variability in the pattern of contemporaneous or dynamic relations among latent constructs allows for a better understanding of individual-centered processes as they unfold over time. However, MIIV-2SLS is not a model selection tool, hence the MIIV-2SLS estimation of the final sparse LV-uSEM is performed after the stepwise model selection procedure.

The current study

Model specification for hybrid relations

The uSEM framework is used to estimate single-subject dynamic models with latent factors such as the DAFM. However, there are several areas that we propose to extend the single-subject modeling under the GIMME framework. First, it is imperative to move from the restrictive VAR representation in the DAFM from a uSEM to the more flexible hybrid uSEM. That is, directed regressions and undirected error covariances among contemporaneous latent factor variables should be incorporated simultaneously. Because not only can they co-exist, they can carry different causal interpretations as well as practical implications. Therefore, the first goal is to extend the structural model in LV-uSEM to the hybrid representation, i.e., LV-huSEM, by altering the residual covariance matrix $\boldsymbol{\Psi}$ in Equation (4) to $\boldsymbol{\Psi}^*$ below:

$$\boldsymbol{\Psi}^* = \begin{pmatrix} \psi_{11} & & & & & & & & & \\ \psi_{21} & \ddots & & & & & & & & \\ \vdots & \ddots & \ddots & & & & & & & \\ \vdots & & \ddots & \ddots & & & & & & \\ \psi_{q1} & \cdots & \cdots & \psi_{q(q-1)} & \ddots & & & & & \\ 0 & \cdots & \cdots & \cdots & 0 & \ddots & & & & \\ \vdots & \ddots & & & \vdots & \psi_{(q+1)(q)} & & & & \\ \vdots & & \ddots & & \vdots & \vdots & \ddots & \ddots & & \\ \vdots & & & \ddots & \vdots & \vdots & & \ddots & \ddots & \\ 0 & \cdots & \cdots & \cdots & 0 & \psi_{(2q)(q)} & \cdots & \cdots & \psi_{(2q)(2q-1)} & \psi_{(2q)(2q)} \end{pmatrix}_{2q\times 2q}$$

Note that the lower-right matrix block, $\boldsymbol{\Psi}_{\mathbf{t}}^*$, is now a symmetric matrix with contemporaneous variances $\psi_i^*, i = q+1\ldots 2q$ of the factor residuals on the diagonal (as seen in $\boldsymbol{\Psi}_{\mathbf{t}}$). By substituting the off-diagonal 0's in $\boldsymbol{\Psi}_{\mathbf{t}}$ with parameters of contemporaneous residual covariance $\psi_{ij}^*, i, j = q+1\ldots 2q, i \neq j$, $\boldsymbol{\Psi}$ is turned to $\boldsymbol{\Psi}^*$. Each element in $\boldsymbol{\Psi}_{\mathbf{t}}^*$ are now candidates in the model search procedure. This relaxes the conditional independence assumption on the contemporaneous errors of uSEM and allow the errors to be correlated. Additionally, the regression coefficient matrix $\boldsymbol{B}$ in Equation (4) is turned to $\boldsymbol{B}^*$ with the same non-zero pattern yet perhaps different estimates, and the model-implied covariance matrix is now $\boldsymbol{\Sigma}^*$ derived by:

$$\boldsymbol{\Sigma}^* = (\boldsymbol{I} - \boldsymbol{B}^*)^{-1}\boldsymbol{\Psi}^*(\boldsymbol{I} - \boldsymbol{B}^*)^{-1\prime} \tag{5}$$

Model search using LASSO regularization under SEM

Second, the forward selection method of model building is highly dependent on the starting model and the intermediate steps, and can arrive at an arbitrary final model. Results from the simulation study in Ye et al. (2021) also showed that this approach tends to miss relations with moderate to medium strengths under the uSEM or hybrid uSEM with observed variables, even with the correct starting model and a large sample size. Another critical downside unique to the current LV-GIMME is the sequential analysis that involves calculating factor scores. That is, factor scores of the latent variable series are obtained from the measurement model through the traditional Bartlett method (Bartlett, 2011) or regression methods (Thurstone, 1935) in a separate step prior to the model building and are treated as observed variables in the estimation of the structural model. For one, it has been shown analytically and numerically that a naïve use of factor scores as observed variables without correction leads to inconsistent and biased parameter estimates in the context of linear regressions (Skrondal & Laake, 2001) or simultaneous equations (Croon, 2002). For two, this is essentially a P-technique model (Cattell et al., 1947) that ignores the temporal dependency of the TSD, that is, a reduced DFM model without considering any lagged relations between the latent factors. Hence, the factor scores are systematically biased as they are drawn from a model divergent from the true data-generating model, in addition to the sampling error of calculating factor scores. In addition, there is a lack of knowledge about the impact of the measurement errors and random errors from the factor scores, regardless of the method of calculation, on model selection and estimation in simultaneous equations.

Regularization, in contrast, is a global, continuous model selection and a simultaneous estimation method. Regularization introduces sparsity by imposing a penalty term, the level of which is gauged by searching across a prespecified range of λ values until the optimal λ (hence the sparsity level) is reached such that the model has the least mean square error or the lowest BIC (Jacobucci, 2017; Ye et al., 2021). When using the least absolute shrinkage and selection operator (LASSO, aka the $L1$-norm penalty; Tibshirani, 1996), the sum of the absolute values of the parameters are shrunken toward zero as λ increases, and they can eventually reach exactly zero. Hence, LASSO is often used in favor of a sparse model and to perform model selection. Previous simulations (Ye et al., 2021) demonstrated success in adopting the LASSO regularization to identify a sparse huSEM with a high sensitivity (identifying true paths) regardless of the magnitude as well as a high specificity (eliminating zero relations). However, to the authors' knowledge, LASSO regularization has not been implemented under the LV-uSEM context. Therefore, the current method seeks to replace the pseudo-ML stepwise searching and sequential estimation with the LASSO regularization for a simultaneous identification and estimation of the extended LV-huSEM.

To obtain the solution, a ML regularized cost function is derived by adding the user-defined penalty function to the unregularized ML cost function:

$$F_{Reg}(\theta) = F_{ML}(\theta) + \lambda P(\theta^*) \tag{6}$$

in which $F_{ML}(\theta)$ is the unregularized cost function computed from the model implied covariance $\mathbf{\Sigma}^*$ in Equation (5). The set θ includes all the parameters estimated in the model, while θ^* is the subset containing user-specified parameters under penalization.

Post model selection estimation using MIIV-2SLS

Lastly, there is a lack of evaluation and comparison of these methods for the parameter estimation under data-driven model building procedures. Previous researchers have found that pseudo-ML estimates of individual DFMs obtained by the SEM approach showed higher biases and a tendency for inaccurate statistical conclusions compared with true ML estimates obtained from methods such as the SSM approach with KF estimator (Chow et al., 2010). But such evaluation was done on the correctly specified model, without potential biases associated with the model selection procedure. In addition, as an alternative model selection and estimation method, the property of LASSO regularized estimates in the LV-uSEM context has not been studied. In theory, regularization methods have sacrificed some level of unbiasedness for efficiency, because all the parameters under penalty (including the unknown true ones) are shrunk at the same time. But this does not mean that the LASSO estimates are always more biased than unregularized pseudo-ML estimates, because penalization eliminates unnecessary variables and false relations that can also bias the estimates of the correct parameters.

Importantly, both the pseudo-ML and the LASSO regularization are system-wide estimator that are not robust to structural misspecifications because errors in one place can spread out to other parts of the model including those that are correctly specified (Bollen et al., 2021). Using a non-robust estimator to select the structural (latent variable) model would potentially impact final estimates in both the measurement model and the structural model. In contrast, the MIIV-2SLS is a limited-information equation-by-equation estimator that is more robust to structural misspecifications (Bollen et al., 2007). Indeed, the MIIV-2SLS has been shown to be more robust than the pseudo-ML for the estimation of a DFM when estimated under the LV-uSEM framework (Fisher et al., 2019; Gates et al., 2020). Further, Bollen et al. (2018) has illustrated the analytic robustness conditions of the MIIV-2SLS estimator in SEM. Specifically, they found that misspecification errors from the structural (latent variable) model should not contaminate MIIV-2SLS estimates in the measurement model, whereas the impact of misspecifications in the measurement model on the structural model depends on its location (see Table 7, page 858 in Bollen et al., 2018).

This robustness property makes MIIV-2SLS an excellent choice for the estimation of sparse LV-huSEM when the heterogeneous, exploratory individual structural models are conditioned on a common, confirmatory measurement model. Because under the assumption of the correctly specified measurement model, the only source of misspecification comes from the structural model selection, which according to the robustness condition, will not bias the MIIV-2SLS estimates of the measurement model. This property, however, will not be guaranteed when system-wide estimators were used. That is, even though the measurement model is correctly specified, the errors from the misspecified structural model will likely bias pseudo-ML and LASSO regularization estimates of both the structural as well as the measurement model. Hence, the LV-GIMME research group adopted the MIIV-2SLS approach for the final parameter estimation post to the pseudo-ML based stepwise model search.

Therefore, to investigate consistent and robust estimation under possible heterogeneous misspecified structural models, we included the MIIV-2SLS as a post model selection estimation approach as well (i.e., parameter estimation after the optimal sparse LV-huSEM model is selected). Besides model recovery property, we also seek to compare unbiasedness and robustness behaviors of the pseudo-ML, the LASSO regularized, and the MIIV-2SLS estimators under the context of data-driven dynamic modeling. The current investigation will include the evaluations of these properties under the LV-huSEM with entirely idiographic structural models conditioning on a unified, confirmatory measurement model. Having an identical factor structure in the final LV-huSEM models across these methods ensures that the comparison of model recovery and parameter estimation for the dynamic relations are not contaminated by differences in the structure of latent factors.

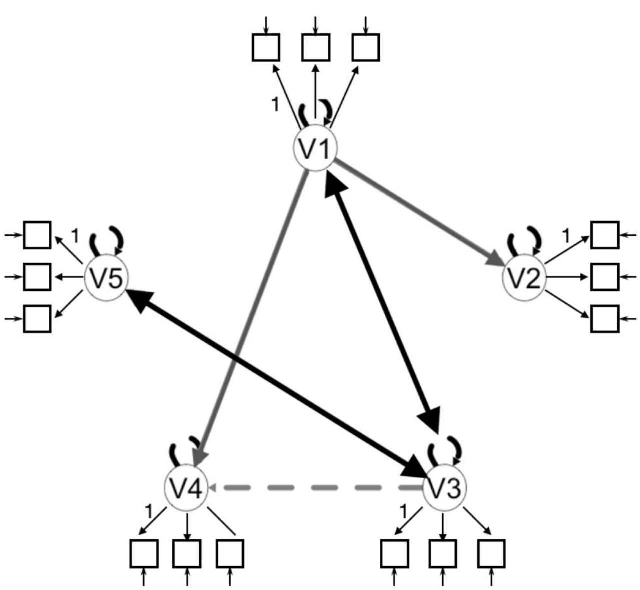

Figure 1. DFM: A time-invariant five-factor DAFM with a hybrid uSEM structure.

The simulation study

In sum, the primary goal of the current study is to evaluate the different model building methods (pseudo-ML vs. regularization) with respect to model recovery as well as estimation approaches (e.g., pseudo-ML, regularization, and MIIV-2SLS) for unbiased and robust parameter estimations for a DAFM with a hybrid VAR representations (i.e., LV-huSEM). We designed a Monte Carlo simulation study to evaluate LASSO regularization and pseudo-ML approach with respect to model recovery as well as their properties for parameter estimation compared to those of MIIV-2SLS under the LV-huSEM context. The goal is to investigate the extent to which building LV-huSEM models with LASSO regularization and MIIV-2SLS estimation is superior to the pseudo-ML approach in terms of (1) sensitivity of finding the true dynamic relations in the structural model, (2) the specificity of excluding the false dynamic relations, and (3) the robustness of parameter estimates to structural misspecifications.

The Data Generating Model (DGM). With our focus on single-subject DFMs, the DGM is a five-factor DAFM with lag-1 hybrid VAR, i.e., hybrid types of contemporaneous relations among the latent factors for all individual TSD (see Figure 1). Following the practice in LV-GIMME, we adopt a homogeneous measurement model as specified by Equation (3). Specifically, each factor has three unique indicators with no cross loadings or lagged relations. In the factor loading matrix $\boldsymbol{\Lambda}$, the scaling indicator of each factor equals 1 with the other two loadings set to 0.9 and 0.7, respectively. The error variance $\boldsymbol{\Theta}$ is a standardized form (i.e., an identity matrix with variance restricted to 1). In the structural model, we include paths of different types and magnitudes to investigate the path recovery for hybrid dynamic relations in $\boldsymbol{B}$ and $\boldsymbol{\Psi}^*$. That is, the contemporaneous relations amongst latent factors include both direct regression path (nonzero elements in $\boldsymbol{A}$) as well as covariance between factors (nonzero elements off the diagonal of $\boldsymbol{\Psi_t}$); Besides AR process within each factor, cross-lag relations are also incorporated (nonzero elements off the diagonal of $\boldsymbol{\Phi}$). We varied the magnitude of coefficients and covariances to examine whether these impact the recovery of the true DGM (the following parameter matrix applies to all simulation conditions).[4]

[4]The authors do not claim that the TSD generated by the LV-huSEM model with such combination of these parameter values returns a typical dynamic process in practice. In fact, we adopt this simplified model structure with sparse relations for the illustration purposes. In the Discussion section, we discuss the generalization of our results to more complicated situations that could be found in empirical data.

$$
\boldsymbol{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},
$$

$$
\boldsymbol{\Phi} = \begin{pmatrix} 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.7 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0.5 \end{pmatrix},
$$

$$
\boldsymbol{\Psi}_{\mathbf{t}} = \begin{pmatrix} 1 & 0 & 0.25 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0.25 & 0 & 1 & 0 & 0.7 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0.7 & 0 & 1 \end{pmatrix}
$$

To investigate the influence of sample size on the performance, data is generated from the same DGM using time lengths varying from 60, 200, to 1,000, representing a range from small to large in practice. This is to be consistent with the simulation design in a previous evaluation on regularized huSEM (Ye et al., 2021). That is, the choice of these design factors are decided such that they represent data structure and characteristics of time series data in psychological and psychophysiological research. For example, although 60 might appear small in panel or cross-sectional data, it would be moderately large in time series such as daily dairy. Note that only the number of timepoints is crossed design, the other factors are investigated within one model. All the DGMs will be replicated 1,000 times, resulting in 3,000 datasets. The weak stationary test on the factor series was performed in the data generating process, i.e., we tested that all eigenvalues of the AR weight matrix, i.e., $(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{\Phi}$, have modulus less than one (Lütkepohl, 2005). All analyses will be performed in R, codes are released and made publicly available on the Open Science Framework site.

Analytic procedure

For the pseudo-ML approach, confirmatory five-factor measurement models are estimated by pseudo-ML in *lavaan* or by MIIV-2SLS in *MIIVsem*, and factor scores are obtained by the default regression method of the 'lavPredict' function in *lavaan*. These factor score series will enter the subsequent structural model for model selection using pseudo-ML forward search in the GIMME package, function *indSEM*. The difference from the original setting in LV-GIMME is that here the starting structural model is a huSEM (with the covariance matrix $\boldsymbol{\psi}^*$) instead of the more restricted uSEM (with $\boldsymbol{\psi}$). Additionally, we focus on individual models and no group level model is considered. For this reason, we refer to this method "pseudoML-FS-huSEM" to indicate it uses modification indexes for the search of sparse huSEM model using the factor scores.

For the proposed method, the LV-huSEM under LASSO regularization (i.e., LASSO-LV-huSEM) will be implemented under the regularized SEM framework. After the LV-huSEM model structure is specified in *lavvan*, *regsem* can import the *lavvan* output, i.e., the unregularized ML cost function $F_{ML}(\theta)$ derived from Equation (6), and perform LASSO regularization with the user-defined list of parameters in the penalty function $\lambda P(\theta^*)$. Note that in LASSO-LV-huSEM the model selection and estimation are performed simultaneously on both the measurement and the structural (latent variable) model. To ensure that factor series represent latent constructs that are consistent with those of "pseudoML-FS-huSEM", the same confirmatory factor structure is estimated without penalty. Parameters in the measurement model (e.g., factor loadings) belong to the freely estimated set in θ but not in set θ^*. Parameters in the set θ^* are regression coefficients for cross-lagged effects and contemporaneous effects (coefficients in the $\boldsymbol{B}$ matrix except the diagonal elements of the lower left block matrix $\boldsymbol{\Phi}$ to free up AR coefficients) as well as the error covariance among contemporaneous latent factors (i.e., off-diagonal elements in the lower right block matrix $\boldsymbol{\Psi}^*$. Ideally, the optimal λ (with the lowest BIC) penalizes all unnecessary parameter(s) to zero and estimates the remaining parameter(s), unraveling the true type of relation between any two latent factors from five possibilities: two cross-lagged effects, two directed contemporaneous regression coefficients, and one undirected contemporaneous error covariance.

Two additional methods were included in the analysis to account for the confounding factor from the use of factor scores in the "pseudoML-FS-huSEM" method. The first one is to repeat the huSEM model search and estimation using factor scores obtained from the DGM (i.e., LV-huSEM) which we call "pseudoML-DGM-FS". By using the population parameters from the DGM, it could reduce the biases because the population parameters remove sampling error of the parameter estimates that are part of generating the factor scores. But note that factor score estimates would still differ from the latent variables – the problem is the measurement errors that are part

Table 1. Two framework for LV-huSEM: Model build and estimation approaches.

Modeling framework	LASSO regularization		pseudo-ML	
Method name	LASSO-LV-huSEM	LASSO-FS-huSEM	pseudoML-FS-huSEM	pseudoML-DGM-FS
Analysis procedure	Simultaneous	Sequential	Sequential	Sequential
Measurement model	confirmatory	factor scores	factor scores	factor scores (DGM)
VAR model build	LASSO penalty	LASSO penalty	Forward stepwise	Forward stepwise
Parameter estimate	LASSO/MIIV-2SLS	n/a	Pseudo-ML/MIIV-2SLS	n/a

of the indicators that form the factor scores. In reality, however, the true model is unknown and so it is impossible to implement this method without error. It is included in the simulation to determine if the use of factor scores improves when population parameters are part of their calculations. Another confounding factor lies in the comparison of "pseudoML-FS-huSEM" (i.e., using pseudo-ML) and "LASSO-LV-huSEM" in their ability to select and estimate true relations and eliminate false ones in the structural model is the fact that the pseudo-ML is subject to measurement errors and random errors in the factor scores, while LASSO regularization simultaneously estimates the measurement and structural (latent variable) models without calculating factor scores. To account for that difference which confounds the comparison between pseudo-ML and LASSO regularization, we included an additional analysis to apply LASSO regularization on a huSEM using the same factor scores from the five-factor measurement model as we did in the pseudo-ML approach, i.e., "LASSO-FS-huSEM". In that case, "LASSO-FS-huSEM" and "pseudoML-FS-huSEM" only differ in their model selection and estimation methods but not how the measurement model and starting structural model are constructed. While the difference between "LASSO-FS-huSEM" and the generic proposed method "LASSO-LV-huSEM" informs the impact of measurement errors from factor scores instead of latent variables.

Finally, recall that both pseudo-ML and LASSO regularization are for both model selection and estimation. But given the potential biases introduced by the model selection procedure and the robustness property of the MIIV-2SLS, the latter is also included as a post model selection estimation. That is, after the final sparse LV-huSEM is selected, it will be estimated again using the MIIV-2SLS to obtain the final parameter estimates of both the measurement and the structural (latent variable) models. In terms of the selection of MIIVs for each equation when the number of MIIVs exceed the minimum number required for model identification, previous simulations studies (e.g., Bollen et al., 2007) found that using one additional MIIV than the minimum number produces the least biased estimation at small sample size conditions, but matters less in large samples. We chose to adopt this approach for the MIIV-2SLS estimation in the current simulation, given that the examination includes small to moderate sample sizes. We refer to this approach MIIV-2SLS-DF1 to indicate the one degree of freedom in the overidentification Sargan's test. With these investigations, we could examine which combination of model selection and estimation regime is the overall optimal practice, accounting for the treatment of the measurement model. The analytical steps and differences for the four methods are summarized in Table 1.

Evaluation measures

We use sensitivity and specificity to evaluate the accuracy of recovering relations with the correct direction. Sensitivity and specificity are common outcome measures in network research (e.g., Abegaz & Wit, 2013; Epskamp & Fried, 2016). Sensitivity is calculated by the ratio of the true positive count discovered in the search over the sum of all true relations in the DGM (i.e., true positives and false negatives). Sensitivity represents the power to detect true relationships. In this paper, because the starting model is the more flexible LV-huSEM with all the free parameters in the extended $\boldsymbol{\Psi}^*$ and $\boldsymbol{B}^*$, we do not distinguish path sensitivity from direction sensitivity. That is, only the relations that are recovered with the correct direction are recorded. Essentially, the sensitivity concept here is equivalent to the direction sensitivity in Ye et al. (2021). Specificity, in comparison, is calculated by the ratio of true negative count over the sum of negatives in the DGM (i.e., the sum of true negative count and false positive count). This represents the percentage of non-existing paths in the DGM that the search procedure accurately omitted in the final model. These measures allow for a global evaluation of a model's ability to detect true recovery and to reject false ones. In both sensitivity and specificity measures, higher values indicate better performance in the selection of true data-generating relations.

However, from previous observations, sometimes a relation between two variables will be recovered with a misspecified direction. For example, at the presence of a directed relation between two contemporaneous factors (e.g., $Y_{t,1} \rightarrow Y_{t,2}$), an alternative relation such as a reversed sign ($Y_{t,1} \leftarrow Y_{t,2}$) or as lagged ($Y_{t-1,1} \rightarrow Y_{t,2}$ or $Y_{t-1,2} \rightarrow Y_{t,1}$) or as undirected covariance ($Y_{t,1} \leftrightarrow Y_{t,2}$) might be selected by the model. In some scenarios, it is better to have another form of relation from the true form than completely missing the relations, but not always. For example, if a direct path between two variables is missed, having a covariance or a lagged path with the correct direction in the selected model is more informative than having no relation; however, if the directed path with the reversed sign is recovered, the information is misleading (i.e., a wrong causal implication). Sometimes, a relation with one or two wrong directions could be selected *instead* when the sample size is too small and the sampling error is large, but other times one or more could also be selected *in addition to* the true direction when the sample size is large. To investigate this behavior in relation to model selection methods and across sample sizes, we calculate relation-specific "direction false positive": the percentage of time where there are at least one misspecified directions in relation to a given true relation being selected in the final model. Accordingly, we distinguish the overall model specificity with the one that eliminate the "direction false positive" related to true relations, the authors refer to them as direction specificity (more stringent) and path specificity (more lenient), respectively.

To examine unbiasedness and robustness, we will exam the relative bias measure. We calculate the mean relative bias (RB) for each parameter as the difference of the actual estimate and the true value divided by the true value, averaged across the cases when the path is recovered by the model (i.e., nonzero). Hence, this is a RB rate conditional on the path recovery. When using MIIV-2SLS, the equation level over-identification test (i.e., Sargan's χ^2 test) is available. The Sargan's χ^2 test informs whether the MIIVs are uncorrelated with the error term of the corresponding equation. Rejection of the null hypothesis suggests that one or more of the MIIVs are inappropriate. This could occur if the equation is misspecified or if another part of the model is misspecified and this leads to one or more incorrect MIIVs for the equation. A significant Sargan test cannot definitely tell which of these is true, but it does alert the researcher to the potential of inconsistent coefficients estimators. Following previous literature (Fisher et al., 2019), we are interested in two properites: (1) the statistical power of the Sargan's χ^2 test when at least one of the MIIVs for an equation is wrong, and (2) the Type I error rate of the Sargan's χ^2 testwhen all MIIVs are valid. Consistent with literature, we will use an α of 0.05 for both circumstances. The aim is to evaluate the RB and robustness behaviors under different misspecified structural models as recovered by pseudo-ML, regularization, versus MIIV-2SLS. From a practical point of view, the convergence behaviors for each method will be recorded and compared.

Results

Model convergence

Some datasets caused nonconvergence when the LASSO regularization or the pseudo-ML approach estimates a LV-huSEM. These datasets were dropped from the analysis of the outcome measures below. It was observed that out of the 1000 datasets, there were 10.9%, 7.3%, and 7.1% that did not converge for the one-step LASSO-LV-huSEM method [5] at sample sizes $N = 60$, 200, 1000, respectively. These rates were increased to 21.9%, 14.5%, 8.3% (respectively) when we used the two-step LASSO-FS-huSEM method. The pseudo-ML method using factor scores from a five-factor measurement model did not converge for 2% datasets at $N = 60$. All the other conditions converged. It is clear that the model using LASSO regularization using a starting model of a full LV-huSEM (with all the parameters included) has a higher chance of model nonconvergence compared to those using the pseudo-ML method that starts with a null model and a much small model specification (as the estimation of the measurement model is separate from that of the structural model).

Sensitivity and specificity

Let us first turn to the sensitivity for recovering true relations of the DGM from the starting LV-huSEM (i.e., confirmatory measurement model and an exploratory structural model with all the free parameters denoted in $\boldsymbol{B}$ and $\boldsymbol{\psi}^*$). All the methods showed an excellent sensitivity for lag-1 effects regardless of the sample size (see Figure 2). Besides lag-1 relations, the probability to recover another true path by any method depends largely on the sample size: the recovery rates were low when the sample size was small ($N = 60$), overall acceptable at a medium sample size

[5]Running the LASSO-LV-huSEM for a single TSD from the DGM (N = 200) in R using the regsem package, following the procedure as described in this paper, takes approximately 10 minutes. This estimate is based on the following hardware configuration: Apple M1 Pro @ 3.06 GHz 1 Processor with 8 Cores with 16Gb SSD.

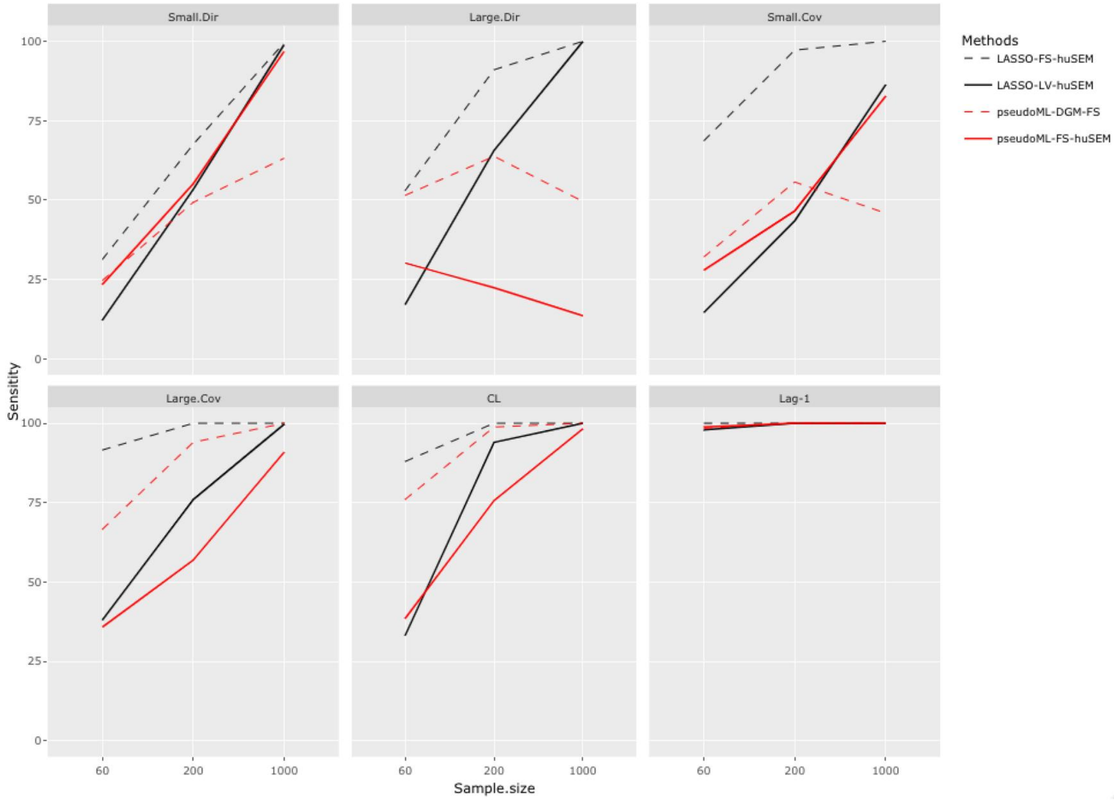

Figure 2. Sensitivity of path recovery by path type and strength across sample size.
Note: Small.Dir = small directed path, Large.Dir = large directed path, Small.Cov = small covariance relation, Large.Cov = large covariance relation, CL = cross-lag effect, Lag-1 = lag-1 effect.

($N = 200$) and satisfactory given a large sample size ($N = 1000$). Specifically, between the two generic methods of our interest, i.e., the pseudo-ML using factor scores (pseudoML-FS-huSEM) and the proposed LV-huSEM under LASSO regularization (LASSO-LV-huSEM), the performance of recovering a small directed path or a covariance relation were similar; however, LASSO-LV-huSEM showed an overall higher sensitivity to strong relations (i.e., directed, covariance, and cross-lagged relations) when given a medium or large sample size. Surprisingly, pseudoML-FS-huSEM performed poorly in recovering the strong directed path even with a large sample size. A closer examination revealed that the majority of time the model tended to recover a true strong directed path as a covariance relation and sometimes as a reversed sign directed path (hence a high rate of direction false positive, see Figure 3). This is a scenario of a recovery that counted as a "path presence recovery" but not as a "direction recovery" in the simulation of Ye et al. (2021). Note that the distinction was emphasized there because the investigation involved more restricted starting models such that some types of relation were misspecified one way or another, their presence can only be recovered by an alternative form between the two variables. It is not the case here where both methods used the true starting structural model (i.e., huSEM) for the search.

Turning to the impact of the type of factor score, it seemed that when a true DGM model (i.e., LV-huSEM) was used to obtain the factor scores for the subsequent pseudo-ML analysis (i.e., pseudoML-DGM-FS), the overall model recovery performance of the strong relations was much better than those from using the factor scores of a confirmatory measurement model alone. This suggests using estimated parameter values for the factor scores affect the recovery of the structural relations amongst the factors. Overall, LASSO-FS-huSEM had the best sensitivity performance of all the methods in Figure 2. This suggests that separating the measurement model from the structural (latent variable) model using factor scores as observed actually increased the chance of recovering the structural relations amongst the latent factors.

Next, we examined the path-specific "direction false positive rate", defined as the chance of recovering a true association yet with a wrong direction (refer to the Evaluation Measures section for the more detailed description). Not surprisingly, it was observed that some relations were recovered with a wrong direction

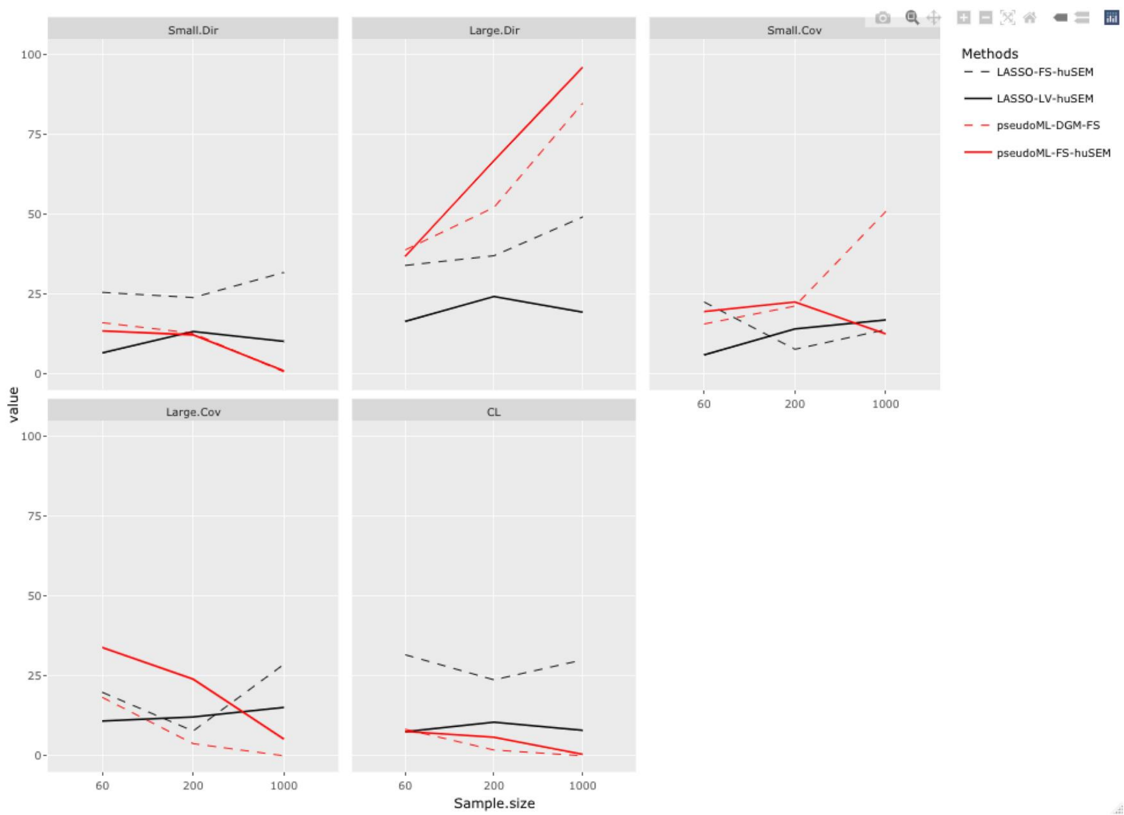

Figure 3. Direction false positive by path type and strength across sample size.
Note: Small.Dir = small directed path, Large.Dir = large directed path, Small.Cov = small covariance relation, Large.Cov = large covariance relation, CL = cross-lag effect.

when the sample size was small. However, even when the sample size was sufficient and in many cases the true path was recovered, sometimes additional paths might still be selected when there existed a strong correlation between the two variables. Hence, direction false positive rates did not necessarily go down with the increase of sample size (Figure 3). Overall, except for cross-lagged relations, pseudo-ML methods had higher direction false positive rates in relation to the true paths in the DGM than did LASSO methods. This is partly the reason that pseudoML-FS-huSEM had very poor sensitivity under some conditions. That is, some relations were recovered only with a wrong direction or type of relation. For instance, at sample size of 200 and 1000, both the LASSO-LV-huSEM (around 12–17%) and pseudoML-FS-huSEM (around 20–34%) methods had some tendency to recover a directed path at the presence of a true covariance relations between two contemporaneous factors. More problematically, pseudo-ML showed a high chance (67% at $N = 200$ or 96% at $N = 1000$) of recovering a reversed signed directed path or a covariance when there existing a strong directed path. Using the factor scores from the DGM (i.e., pseudoML-DGM-FS) did not decrease the chance of false positive directions. In fact, using LASSO on the factors scores to select the structural model seemed to also introduce more direction false positive than the one-step LASSO-LV-huSEM model. These consistent observations that all methods using the factor scores showed a higher rate of direction false positive than their counterparts suggested that the issue of a wrong direction recovery of true relations is very likely tied to the use of factor scores in place of the latent variables.

Both generic methods reached a path specificity above 90% (see Figure 4), suggesting they are reliable in rejecting false paths that were unrelated with those pairs of variables that have a true relation of another form or direction. However, the direction specificity (i.e., the odds of ruling out any path when it is truly false) dropped quite a bit for pseudoML-FS-huSEM (to around 72–77%) or any method that used factor scores. This is again because there was quite an amount of direction false positive paths in relation to true paths.

Conditional relative bias

Overall, the post model selection MIIV-2SLS estimation was the least biased for parameter estimates of contemporaneous relations in the structural model,

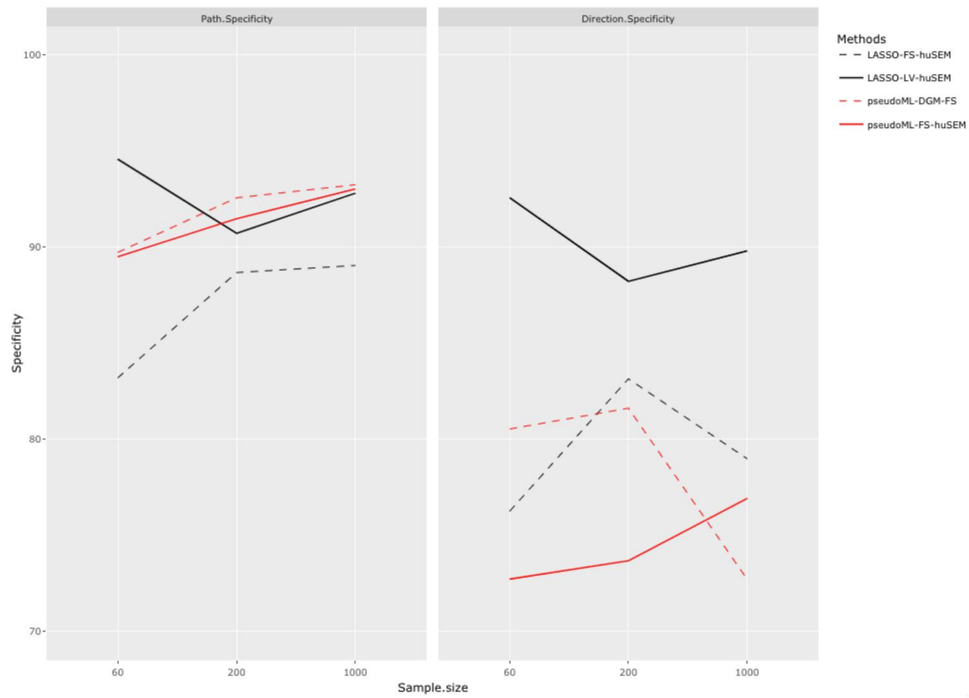

Figure 4. Path and direction specificity in the model by sample size.

even though the estimates of small relations from any method were highly biased at a sample size of 60. In comparison, the LASSO penalized LV-huSEM methods produced the most unbiased estimates for freely estimated parameters, i.e., lag-1 autoregressive effects within each latent factor as well as factor loading in the measurement model, even at a small sample size as low as 60. Two methods produced similar results for cross-lagged effect among latent factors. Particularly, at a sample size of 200 and 1000, the conditional mean relative biases of lag-1 from LASSO-LV-huSEM were only 4.6% and 1.6%, respectively. The mean RBs of factor loading estimates were as low as under 1.8% at $N = 200$ or under 0.4% at $N = 1000$. When sample size was large, MIIV-2SLS-DF1 also produced estimates with small biases on average for factor loading estimates (under 5%), but the mean RBs were higher than those from LASSO-LV-huSEM at a small to moderate sample size (e.g., around 20% for MIIV-2SLS-DF1 compared to around 10% for LASSO-LV-huSEM at $N = 60$, or 12% versus 2% at $N = 200$, respectively). However, for the parameters under penalty in LASSO-LV-huSEM or LASSO-FS-huSEM (i.e., contemporaneous effects in the structural model), estimates were on average more biased (e.g., ranged from 20% to 40% at $N = 1000$) than those from MIIV-2SLS-DF1 (e.g., ranged from 5% to 18% at $N = 1000$), which was as expected given they were under penalty in LASSO methods while being freely estimated by the MIIV-2SLS after the model selection (Figure 5).

The comparisons between LASSO methods and pseudo-ML methods were mixed across different types and strengths of relations. For example, for lag-1 and cross-lagged estimates, pseudoML-FS-huSEM estimates produced larger biases (e.g., mean RBs was between 50 to 60% even at $N = 200$ or 1000) than did LASSO estimates (e.g., mean RBs ranged from 2% to 13% at $N = 200$ or 1000). Evidence that some of the biases come from errors in factor scores is that pseudoML-DGM-FS using factors scores from the DGM produced much less biased estimates than did pseudoML-FS-huSEM (using factor scores from a five-factor measurement model). The true DGM is not known so the pseudoML-DGM-FS is never available, but we include it here to illustrate the impact of forming factor scores from estimated rather than true parameter values. However, this pattern did not apply to all types of parameters. For instance, the two pseudoML methods produced similar and slightly less biased estimates than did the two LASSO for small directed and large covariance relations at each sample size level, although all of them were still more biased than those from MIIV-2SLS. This suggests that errors in factor scores could introduce additional biases in some (e.g., lagged relations) but not all types of relations. However, all the pseudoML and LASSO estimates were largely biased for small covariance

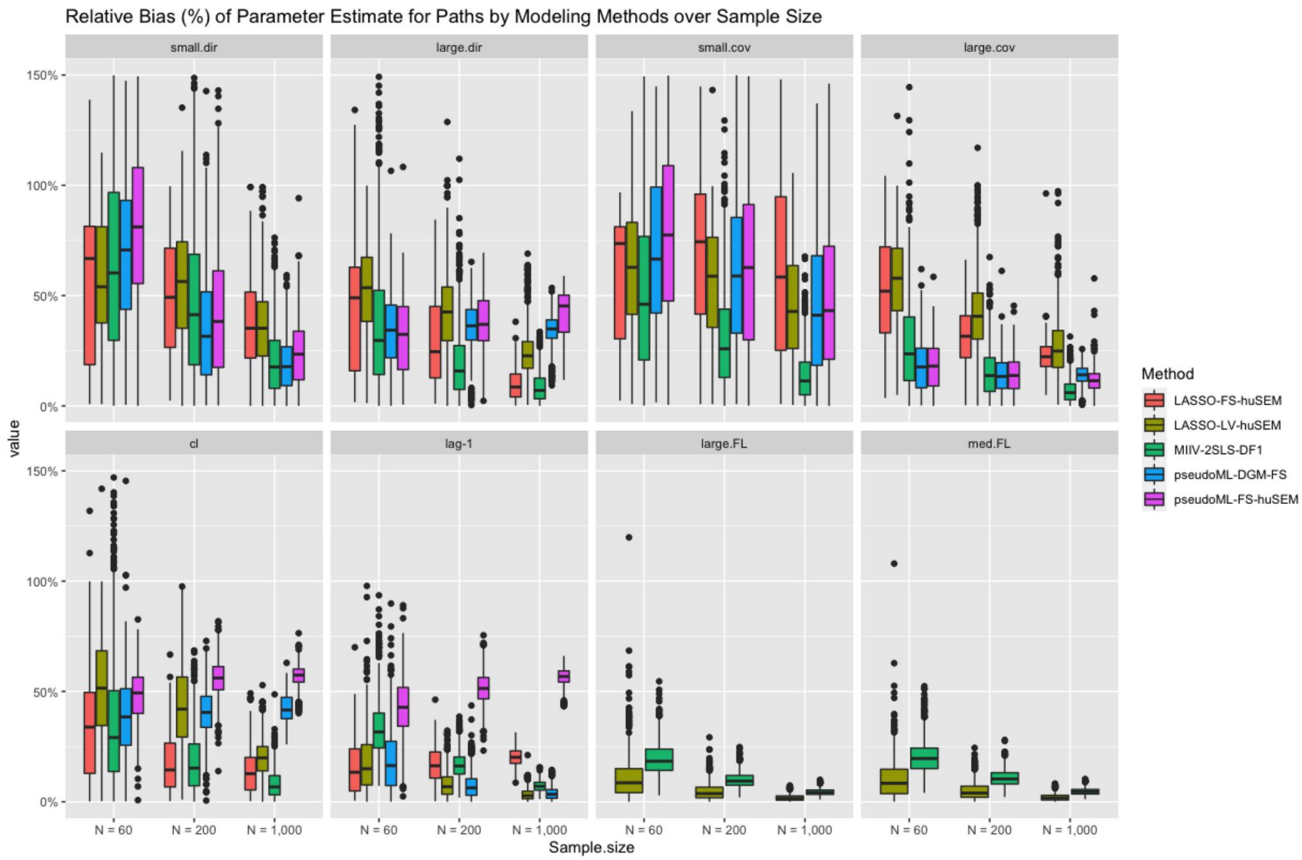

Figure 5. Conditional relative bias (100%) by path type and strength across sample size.
Note: Y-Axis "value" refers to "Conditional Relative Bias (100%)". Small.Dir = small directed path, Large.Dir = large directed path, Small.Cov = small covariance relation, Large.Cov = large covariance relation, CL = cross-lag effect, Lag-1 = lag-1 effect, Large.FL = factor loadings of .9, Med.FL = factor loadings of .7.

relations even with a large sample size (mean RBs ranged from 40% to 70%). The RBs of these estimates were also associated with a large variability across replications, indicating a substantial amount of influence from the sampling error. Therefore, MIIV-2SLS-DF1 is particularly useful to obtain less biased estimates for moderate or small structural relations.

One interesting observation is that mean RBs of pseudo-ML estimates did not show the same asymptotic trend as did LASSO estimates or the MIIV-2SLS estimates. Increased sample size was associated with a decrease in the mean RBs in LASSO estimates and the MIIV-2SLS estimates across all parameters, but the mean RBs of pseudo-ML estimates were not so consistent. For example, the mean RBs of lag-1 and cross-lagged even went up as sample size increased. This suggests that the source of biases in pseudo-ML estimators do not just come from sampling error. This is more evidence of impact of systematic errors in the factor scores. As another way to investigate the shortcoming of using the factor scores instead of latent variables, we need to examine results between the two LASSO methods in which the only difference is that one uses factor scores (LASSO-FS-huSEM) while the other estimates a latent variable model simultaneously with a measurement model (LASSO-LV-huSEM). It seemed that the use of factor scores did not necessarily affect biases in the parameter estimates, as the overall mean RBs were similar between the two LASSO methods; however, in some cases the variability of RBs was larger in LASSO-FS-huSEM than that from LASSO-LV-huSEM, particularly when sample size was small. This seems to suggest that sampling errors and measurement errors in factor scores might not affect the average accuracy of the parameter estimates of the structural model in a systematic way, but it affects the consistency of the estimates such that they are less consistent at an increasing sampling fluctuation than methods using latent variables.

The overidentification test

We evaluated the finite sample properties of Sargan's χ^2 overidentification test of MIIVs. The current examination involves the specificity (i.e., true negative) as well as sensitivity to wrong MIIVs (i.e., the statistical power of the Sargan's χ^2 test when at least one of the MIIVs for an equation is wrong). We investigated the case where a wrong MIIV was included due to an omitted true relation in the model. For instance, the omission of the directed contemporaneous relation from factor one to factor four will render a wrong inclusion of the scale indicator of factor one in the structural equation of factor four. Even though a significant Sargan's Test does not suggest the equation per se is incorrect, it offers evidence that one or more MIIVs of that equation are incorrect. This in turn indicates errors in the model specification that led to these MIIVs. In contrast, passing the Sargan test is consistent with a correctly specified equation and valid MIIVs for that equation. We found that the specificity was around 95–97%. In other words, as expected there was less than 5% of rejection rate in which Sargan's χ^2 test incorrectly identified a wrong selection of MIIVs when in reality the model was correctly specified and all MIIVs were correct. This suggests accurate Type I error across the sample sizes considered here. For the test sensitivity, i.e., when a true relation was omitted from the model and hence a wrong MIIV set would be included in the corresponding equation, the Sargan's test rejected the problematic equation at the rates of 59% at $N = 60$, 66% at $N = 200$, and 92% at $N = 1000$, respectively. This suggests that the test has a moderate power to detect a wrong MIIV at a small to medium sample size, but can do so at a very high rate when the sample size is large.

Discussion

The current study serves to advance and investigate the model search and estimation for a single-person DFM with a hybrid VAR representations. Three goals were achieved in the proposed framework. First, we extended the structural model of the latent variable uSEM (LV-uSEM) to its hybrid uSEM version, i.e., the latent variable hybrid uSEM (or LV-huSEM). In this way, the extended LV-huSEM estimates a DFM with hybrid contemporaneous relations in the structural model. When restrictions on the contemporaneous relations between the latent factors are relaxed, structural and covariance relations can be simultaneously estimated. Second, LASSO regularization in replace of previous pseudo-ML-based stepwise model search is used to perform both model selection for the optimal sparse latent variable hybrid uSEM, and a simultaneous estimation for a freely estimated confirmatory measurement model and an exploratory structural model (with LASSO penalty on the structural paths and covariances between the contemporaneous latent factors). Compared to previous approaches, where measurement model and structural model are estimated sequentially with a stepwise model search procedure using factor scores obtained prior to the model selection (e.g., LV-GIMME[6]; Gates et al., 2020), the current method provides a model search on a continuum and a simultaneous estimation without calculating factor scores. Finally, to obtain final parameter estimates, the selected sparse LV-huSEM is estimated *via* a limited-information estimator, the MIIV-2SLS (Fisher & Bollen, 1996; Fisher et al., 2019). The post model selection MIIV-2SLS estimation is chosen for its robustness property under model structural misspecification (Bollen et al., 2007), which is particularly advantageous for the estimation of sparse LV-huSEM selected in a data-driven manner. The source of biases in the parameter estimates under the current context includes (1) model structural misspecifications result from model selection, (2) the LASSO penalty in the regularized LV-huSEM, or (3) errors of factor scores (if under the LV-GIMME framewrork). The goal is to obtain less biased final parameter estimates of the selected sparse latent variable hybrid uSEM.

A simulation study was conducted to investigate to what extent the novel estimation method for the LV-huSEM models, i.e., a LASSO regularization model build and post model selection MIIV-2SLS estimation, is superior to the pseudo-ML approach similar to the single-subject model in the LV-GIMME framework. For model recovery, the simulation results revealed that the pseudo-ML and the LASSO regularization have comparable recovery rates for some relations such as lagged effects and small contemporaneous effects among factors, and they both are reliable in recovering a close-to-true structural model when the sample size is medium to large. The impact of factor scores on model selection under the SEM context was largely unknown. Even though that the LV-GIMME study (Gates et al., 2020) found that the path recovery performance does not seem to be related to what approach was used to derive the factor scores, there is no comparison with a simultaneous estimation method without the use of factor scores. In addition, the evaluation was on a restricted DFM, i.e., LV-uSEM with only directed contemporaneous relations among factors, which might not apply to the recovery of the more complicated model, LV-huSEM in which

[6]Since the current focus is on individual models without using the group level modeling in the LV-GIMME framework (Gates et al., 2020), the authors referred to the compared method as pseudo-ML method.

the recovery of a true relation between two variables involves a selection among five possible parameters. Indeed, we found that the pseudo-ML methods using factor scores have a higher chance to commit a direction false positive on strong directed relations, that is, a tendency to recover a strong directed as one with a reversed direction or as an undirected covariance relation. This low direction specificity seems tied to the estimation of the factor scores. And it is shown that the performance is improved when with factor scores from the DGM, although the use of such factor scores is still different from the latent variables (due to the measurement errors from the part of the indicators that form the factor scores). Further, in the additional analysis of applying LASSO regularized hybrid uSEM on factor scores, the likelihood to commit a direction false positive is also higher than that of the simultaneous LASSO regularization hybrid uSEM with latent variables. The result suggests that the use of factor scores instead of the latent variable approach is subjected to a higher false positive rate.

This tendency of recovering a relation with a wrong direction undermines the purpose of adopting the more flexible VAR representation, i.e., to accurately represent the hybrid forms of contemporaneous relations that might coexist in practice. Causal implications represented by a model with only directed paths or only undirected covariance could be very different. However, if we choose to use a more flexible model representation, but the chance of selecting false positive relations with a wrong direction is high by the model selection method, we still end up with misleading causal interpretations. In this sense, when there might exist some strong contemporaneous relations, LASSO regularization seems to have a higher tendency to eliminate false positive relations and avoid misleading causal interpretations than does the pseudo-ML method using factor scores. To our knowledge, this is the first evaluation on the impact of using factor scores for model selection under the uSEM context.

In terms of parameter estimation, as expected, the post model selection MIIV-2SLS estimator is the least biased for parameters in the exploratory structural model. But slightly surprisingly, the proposed LASSO regularized hybrid uSEM with latent variable approach produced the least biased estimates for free parameters including factor loading coefficients in the confirmatory measurement model and the lag-1 autoregressive effect between factors. This suggests that the biases for other regularized parameters are mainly from the LASSO penalty. In practice, if we are using regularized SEM method, we should only penalize the uncertain (i.e., exploratory) part of the model. However, both the pseudo-ML estimator and the LASSO regularized estimator introduce quite some biases in the parameter estimates of the structural model, especially for small relations. Hence, post model selection estimation might indeed be a more practical choice to obtain final estimates.

We also observe at least two major downsides of using factor scores in parameter estimation, besides its impact on model selection (i.e., the tendency to select a directed path with a wrong direction). First, the estimation of the lagged effects seem to be particularly largely biased, regardless of sample size. This is not surprising because no matter what estimation method is used to obtain the factor scores, the sample variance matrix of estimated factor scores is an inconsistent and biased estimate of the true variance matrix of factors (Croon, 2002; Skrondal & Laake, 2001). In addition, the time embedding process of the data to obtain lagged factor variables introduces more random error on top of the measurement error within the contemporaneous factors variables themselves, thus causing additional biases in the estimates of relations between the lagged factor variables and the contemporaneous factors variables. Second, besides a high average level of biases in estimates of small contemporaneous relations regardless of which factor score methods are used, methods using factor scores also tend to produce less consistent estimates that is more subject to sampling fluctuations. The substantial amount of variability in the biases across samples especially with a small to medium sample size is very likely a consequence of the sampling error in the sample variance matrix of estimated factor scores.

Limitations and future directions

Limitation in the two approaches under investigation

The simultaneous analysis using LASSO regularization under the LV-huSEM is easy to implement and can avoid biases from the use of factor scores, but it might be more limited in the size and complexity of the model (e.g., number of variables, factors, density of the structural paths, etc.) than is sequential analysis like the GIMME approach. This is because use of factor scores reduces the dimension of the parameter space - the number of parameters is higher in the simultaneous model as it includes estimates for the measurement model besides the structural model. Optimizing the covariance matrix of observed variables with a higher dimension is more difficult than that of the latent factors. Another advantage of a sequential analysis is that it is at a better chance to avoid improper solutions or nonconvergence issues that is not uncommon in simultaneous estimation methods. Nevertheless, the issue with a high direction of false positives using factor scores makes it a less appealing choice for model selection.

Although the use of MIIV-2SLS estimation reduces biases from model misspecification and model selection, the naïve post model selection statistical inference is subject to another source of bias due to the duplicate use of the data, that is, the same dataset is used for model selection and model estimation. Since statistical inference is established under the assumption that the fitted model is known in advance, which is clearly violated here, the naïve inference of a regularized LV-huSEM after the data driven selection process for variables and relations is no longer valid. Because both the randomness introduced by the selection process and the sample space restriction implied by the chosen model influence the sampling distribution of the estimator and need to be accounted for (Huang, 2020). Indeed, it has been found that the naïve method after model selection using regularized SEM tends to obtain significant results for selected zero parameters, resulting in numerous false positive findings in psychology (Huang, 2020). However, post model selection inference methods are extremely difficult to perform, and few has been developed for regularized SEM or been implemented in the *regsem* package at the time of this work. This is probably because the primary goal of statistical learning method such as LASSO regularization is to achieve the least biased prediction, while the aim of traditional psychology and of the current study is to identify the optimal model from which to obtain statistical inference for individual parameters. This is one of the biggest gap yet to be filled in the future development of statistical learning under SEM framework. It would be useful to make these methods available in the software and packages for regularized SEM, so that future studies could be conducted to evaluate and validate their properties under the regularized hybrid uSEM with latent variables.

One aspect that is out of the scope of the current study is DFMs for multiple subject time series. The focus of the study is idiographic, single-subject DFMs on the ground of unifying factor structure, for which there is no consideration of between-person effects or attempt to aggregate individual models. However, the use of group-level or between-person information (i.e., similarities and variances across individuals) has been shown as an effective way to extract true effects from noise information so that it avoids the risk of over-fitting individual dynamic models (Asparouhov et al., 2018; Gates et al., 2020). In fact, one of the strengths in the LV-GIMME algorithm is to construct a group-level model with the most shared information that forms the starting model for each individual model. Such a strategy might address to some extent the high direction false positive rates. This should be a possible extension to the current LASSO regularized LV-huSEM modeling, for which an additional step for forming the group-level model needs to be incorporated. Alternatively, one recent method called "multi-VAR" is proposed (Fisher et al., 2022) that uses LASSO penalization on multiple subjects multivariate TSD for the forecast of dynamic processes at the individual level. Although the goal of multi-VAR is to identify an optimal sparse VAR that achieves the best prediction, rather than recovering the true model or statistical inferences of the selected model.

An alternative framework for multi-subject DFM is to aggregate individual dynamic results using a multilevel structure. One recent promising method is the Dynamic Structural Equation Modeling (DSEM; Asparouhov et al., 2018), where multiple subjects time series are modeled simultaneously to estimate population mean and individual differences (deviations from the mean) in the parameters governing a dynamic process. DSEM is very flexible as it decomposes the observed TS variables into three model components: individual-specific, time-specific, and the deviation of each individual at each timepoint, with fixed and random effects flexibly incorporated in each component. DSEM implemented in Mplus Version 8 is estimated with a Bayesian method using the Gibbs sampler and the Metropolis–Hastings sampler, which has many advantages such as handling missing data, measurement invariance, etc. However, the Bayesian DSEM or other multilevel modeling options is not flexible in specifying idiographic structural models, e.g., model selection for individual contemporaneous structure is currently unavailable in DSEM. This is because individuals cannot differ in their model structure but only in their parameterization under the assumption of normally distributed parameters. Adopting a multilevel structure for the LV-huSEM will sacrifice the flexibility in incorporating individually heterogeneous structural models.

Limitation in the simulation design

Like all simulations, the simulation factors do not represent a comprehensive list of empirical situations. To keep the scope manageable, we did not adopt a cross design of factors such as path strength, model size (i.e., number of variables) and sparsity, level of measurement error in the latent factors, etc. The DGM might also represent an over-simplified, sparse DFM, with a very standard measurement structure i.e., without cross-loadings or local dependency structures. In practice, the latent variable relations in a DFM could be much denser with many weak to moderate relations. In addition, the measurement model could also be extended both in the sense of containing some error correlations as well as to include lagged effects between the observed indicators and latent factors, i.e., a white noise factor model or a hybrid of WNFM and DAFM. These added complexities pose a bigger challenge for the model selection and parameter estimation, because a denser model means that the model search and optimization for solution are performed at a larger parameter space with higher dimensions. This might require a harsher penalty term, which in turn might lead to larger biases in the estimates of correct parameters, although the current proposal to use the post model selection MIIV-2SLS estimation could partially compensate for the biases. Future development in the optimization algorithm for regularized SEM in general is the key to estimate a LV-huSEM with more complex structures and higher dimensions.

Nevertheless, our results of the LASSO regularized hybrid uSEM with latent variables highlights the flexibility of the LASSO regularized SEM in estimating individual DFMs. The data-driven LASSO penalty opens up a variety of possibilities in the development and appraisal of individual dynamic theories. The penalization structure relies on which part of the model is more supported by theory, and which part is more uncertain and needs to be explored by the data. For instance, when the latent factor structure among the candidate indicators is not fully determined by the theory, partially exploratory model selection could be implemented on the measurement model. Specifically, if some indicators of a factor are confirmed by the theory, but the rest of the indicators are not, we can adopt the "semi-confirmatory" factor model from Huang (2020) in the way such that the factor loadings from the uncertain indicators are included as parameters under penalization while the factor loadings of the confirmed indicators are free parameters. This way, only the uncertain part of the measurement model is under the data-driven LASSO selection. This idea could also be used to explore the hybrid form of WNFM and DAFM when the theory is not enough to determine which DFM is more appropriate. That is, in the case where the measurement model may contain either the lagged factor loading, or the contemporaneous factor loading, or some combination thereof. One way is to penalize the lagged factor loading and the contemporaneous factor loading simultaneously, as an automatic search between a WNFM and a DAFM with respect to the factor loading structure across time. Alternatively, we can retain the confirmed structure of the contemporaneous factor loading and only penalize the lagged factor loading if the latter is optional. Depending on the knowledge of the theory available to us, we can choose which relations we feel confident enough to be included as free estimates, and which we are less so and thus allow the data to decide by imposing a LASSO penalty on the corresponding parameters. Therefore, the flexibility of regularization with user-defined estimation and penalization structure lifts the dichotomous boundary between the exploratory approach and the confirmatory one and allows for an expansion and refining of theory on a continuum.

In conclusion, the major contribution of the current work is to propose a flexible framework for individual dynamic factor models, i.e., the regularized hybrid unified SEM with latent variables, that offers an effective model search and estimation framework with flexible directions of dynamic relations. In terms of the application to substantive research in psychology, the authors remind researchers to keep in mind the strengths and shortcomings of the proposed approach. First and foremost, the adoption of the hybrid uSEM representation is particularly useful for cases where there are a handful of variables or latent constructs and causal implications are of interest, especially since only some causal assumptions can be made from the available literature. In other words, (with some uncertainty) one would expect some relations to be causal by nature (e.g., mediation, common cause, etc.), some are pure associations, and the overall structure is subject to some levels of exploration. Second, this type of model requires a certain amount of sample size. It might be challenging for some studies such as daily diaries to have a few hundred to thousands of time points, while in neuroimaging or biometric studies this might be less of an issue. Third, because the property we found is conditional on the weakly stationary assumption in the time series data and time-invariant parameters are estimated from the model, it might not be very useful for studies such as developmental psychology with meaningful nonstationary trends or time-varying relations. For instance, if the trend of growth (or decrease) carries crucial developmental meanings, the removal of the trend takes away the focus of interest; instead, the stage-specific time-varying dynamic relations might be more informative for developmental studies. In these cases, DFMs that can handle nonstationary and time-varying parameters (e.g., Chow et al., 2011; Molenaar et al., 1992) should be used. Lastly, as the time component is treated as discrete, the current method is appropriate when an equal time interval is achieved or with very short measurement windows such that the unequal distance is negligible. However, when an unequal time interval is an issue, an option is to use the continuous time structural equation modeling, which estimates an underlying continuous process by using stochastic differential equations to accurately account for differences in time intervals between measurements (Driver et al., 2015; Ryan et al., 2018). In contrast, this limit has minimal impact on functional connectivity studies where both the time intervals and the entire duration of the brain scan are very short and they do not carry developmental meanings. In summary, dynamic researchers need to be aware of their research purposes and the characteristics of the time series data they have at hand when making the modeling choice.

Article information

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

Ethical principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: This work was not supported.

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

References

Abegaz, F., & Wit, E. (2013). Sparse time series chain graphical models for reconstructing genetic networks. *Biostatistics*, *14*(3), 586–599. https://doi.org/10.1093/biostatistics/kxt005

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

Bartlett, M. (2011). The statistical concept of mental factors. *British Journal of Psychology. General Section*, *28*(1), 97–104. https://doi.org/10.1111/j.2044-8295.1937.tb00863.x

Bollen, K. A., Fisher, Z. F., Giordano, M. L., Lilly, A. G., Luo, L., & Ye, A. (2021). An introduction to model implied instrumental variables using two stage least squares (miiv-2sls) in structural equation models (sems). *Psychological Methods*, *27*(5), 752–772. https://doi.org/10.1037/met0000297

Bollen, K. A., Gates, K. M., & Fisher, Z. F. (2018). Robustness conditions for miiv-2sls when the latent variable or measurement model is structurally misspecified. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(6), 848–859. (PMID: 30573943) https://doi.org/10.1080/10705511.2018.1456341

Bollen, K. A., Kirby, J. B., Curran, P. J., Paxton, P. M., & Chen, F. (2007). Latent variable models under misspecification: Two-stage least squares (2sls) and maximum likelihood (ml) estimators. *Sociological Methods & Research*, *36*(1), 48–86. https://doi.org/10.1177/0049124107301947

Browne, M., & Nesselroade, J. (2005). Representing psychological processes with dynamic factor models: Some promising uses and extensions of autoregressive moving average time series models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), Contemporary psychometrics: A festschrift for Roderick P. McDonald (pp. 415–452). Lawrence Erlbaum Associates Publishers.

Cattell, R. B., Cattell, A., & Rhymer, R. M. (1947). P-technique demonstrated in determining psychophysiological source traits in a normal individual. *Psychometrika*, *12*(4), 267–288. https://doi.org/10.1007/BF02288941

Chow, S.-M., Ho, R., Ho, M., 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

Chow, S.-M., Zu, J., Shifren, K., & Zhang, G. (2011). Dynamic factor analysis models with time-varying parameters. *Multivariate Behavioral Research*, *46*(2), 303–339. (PMID: 26741330) https://doi.org/10.1080/00273171.2011.563697

Croon, M. A. (2002). Using predicted latent scores in general latent structure models. In G. A. Marcoulides, & I. Moustaki (Eds.), *Latent Variable and Latent Structure Models* (pp. 195–224). Erlbaum.

Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous Time Structural Equation Modeling with R Package ctsem. *Journal of Statistical Software, 77*(5), 1–35. https://doi.org/10.18637/jss.v077.i05

Epskamp, S. (2020). Psychometric network models from time-series and panel data. *Psychometrika*, *85*(1), 206–231. https://doi.org/10.1007/s11336-020-09697-3

Epskamp, S., & Fried, E. I. (2016). A primer on estimating regularized psychological networks. arXiv Preprint arXiv: 1607.01367.

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, *53*(4), 453–480. https://doi.org/10.1080/00273171.2018.1454823

Fisher, Z. F., & Bollen, K. A. (1996). An alternative two stage least squares (2sls) estimator for latent variable equations. *Psychometrika*, *85*(3), 660–683. https://EconPapers.repec.org/RePEc:spr:psycho:v:61:y:1996:i:1:p:109-121 https://doi.org/10.1007/BF02296961

Fisher, Z. F., Bollen, K. A., & Gates, K. M. (2019). A limited information estimator for dynamic factor models. *Multivariate Behavioral Research*, *54*(2), 246–263. (PMID: 30829065) https://doi.org/10.1080/00273171.2018.1519406

Fisher, Z. F., Bollen, K. A., Gates, K. M., & Rönkkö, M. (2020). Model implied instrumental variable (miiv) estimation of structural equation models. [r Package Miivsem Version 0.5.5].

Fisher, Z. F., Kim, Y., Fredrickson, B., & Pipiras, V. (2022). Penalized estimation and forecasting of multiple subject intensive longitudinal data. *Psychometrika*, *87*(2), 1–29.

Gates, K. M., Fisher, Z. F., & Bollen, K. A. (2020). Latent variable gimme using model implied instrumental variables (miivs). *Psychological Methods, 25*(2), 227–242. Retrieved from https://doi.org/10.1037/met0000229

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., 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., 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

Hamaker, E. (2012). Why researchers should think" within-person": A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). The Guilford Press.

Hamaker, E., Dolan, C., & Molenaar, P. C. (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

Hamilton, J. D. (1994). *Time series analysis* (Vol. 2). Princeton University Press Princeton.

Huang, P.-H. (2020). Postselection inference in structural equation modeling. *Multivariate Behavioral Research*, *55*(3), 344–360. (PMID: 31304796) https://doi.org/10.1080/00273171.2019.1634996

Jacobucci, R. (2017). *Regsem: Regularized structural equation modeling*.

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

Lane, S., Gates, K. M., Fisher, Z. F., Arizmendi, C., Molenaar, P. C., Hallquist, M., … Beltz, A. (2019). *Gimme: Group iterative multiple model estimation [Computer software manual]*. https://github.com/GatesLab/gimme/

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Macaro, C., & Prado, R. (2014). Spectral decompositions of multiple time series: A bayesian non-parametric approach. *Psychometrika*, *79*(1), 105–129. https://doi.org/10.1007/s11336-013-9354-0

Molenaar, P. C. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*(2), 181–202. https://ideas.repec.org/a/spr/psycho/v50y1985i2p181-202.html https://doi.org/10.1007/BF02294246

Molenaar, P. C. (1987). Dynamic factor analysis in the frequency domain: Causal modeling of multivariate psychophysiological time series. *Multivariate Behavioral Research*, *22*(3), 329–353. (PMID: 26776380) https://doi.org/10.1207/s15327906mbr2203_5

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

Molenaar, P. C., De Gooijer, J. G., & Schmitz, B. (1992). Dynamic factor model analysis of nonstationary multivariate time series. *Psychometrika*, *57*(3), 333–349. https://doi.org/10.1007/BF02295422

Molenaar, P. C., & Lo, L. L. (2016). Alternative forms of granger causality, heterogeneity and non-stationarity. W. Wiedermann, & A. von Eye (Eds.), *Statistics and causality: Methods for applied empirical research* (pp. 205–230). John Wiley & Sons.

Molenaar, P. C., & Nesselroade, J. R. (1998). A comparison of pseudo-maximum likelihood and asymptotically distribution-free dynamic factor analysis parameter estimation in fitting covariance-structure models to block-toeplitz matrices representing single-subject multivariate time-series. *Multivariate Behavioral Research*, *33*(3), 313–342. (PMID: 26782717) https://doi.org/10.1207/s15327906mbr3303_1

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

Nichols, T. T., Gates, K. M., Molenaar, P. C., & Wilson, S. J. (2014). Greater bold activity but more efficient connectivity is associated with better cognitive performance within a sample of nicotine-deprived smokers. *Addiction Biology*, *19*(5), 931–940. https://doi.org/10.1111/adb.12060

Ryan, O., Kuiper, R., & Hamaker, E. (2018). A continuous-time approach to intensive longitudinal data: What, why, and how? *Continuous time modeling in the behavioral and related sciences*, 27–54.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica*, *26*(3), 393–415. http://www.jstor.org/stable/1907619 https://doi.org/10.2307/1907619

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, *66*(4), 563–575. https://EconPapers.repec.org/RePEc:spr:psycho:v:66:y:2001:i:4:p:563-575 https://doi.org/10.1007/BF02296196

Thurstone, L. L. (1935). *The vectors of mind*. University of Chicago Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Wright, A. G., Beltz, A. M., Gates, K. M., Molenaar, P. C., & Simms, L. J. (2015). Examining the dynamic structure of daily internalizing and externalizing behavior at multiple levels of analysis. *Frontiers in Psychology*, *6*, 1914. https://doi.org/10.3389/fpsyg.2015.01914

Ye, A., Gates, K. M., Henry, T. R., & Luo, L. (2021). Path and directionality discovery in individual dynamic models: A regularized unified structural equation modeling approach for hybrid vector autoregression. *Psychometrika*, *86*(2), 404–441. https://doi.org/10.1007/s11336-021-09753-6

Zhang, Z., Hamaker, E., & Nesselroade, J. (2008). Comparisons of four methods for estimating a dynamic factor model. *Structural Equation Modeling a Multidisciplinary Journal*, *15*, 377–402.