STRUCTURAL EQUATION MODELING
2025, VOL. 32, NO. 1
ISSN: 1070-5511 | E-ISSN: 1532-8007

Latent Vector Autoregressive Modeling: A Stepwise Estimation Approach

Manuel T. Rein[a,b], Jeroen K. Vermunt[a], Kim De Roover[a,b] and Leonie V. D. E. Vogelsmeier[a]

[a]Tilburg University; [b]Katholieke Universiteit Leuven

ABSTRACT

Researchers often study dynamic processes of latent variables in everyday life, such as the interplay of positive and negative affect over time. An intuitive approach is to first estimate the measurement model of the latent variables, then compute factor scores, and finally use these factor scores as observed scores in vector autoregressive modeling. However, this approach neglects the uncertainty in the factor scores, leading to biased parameter estimates and threatening the validity of conclusions about the dynamic process. We propose Three-Step Latent Vector Autoregression that adheres to this stepwise procedure while correcting for the factor scores' uncertainty. Stepwise approaches offer various advantages, for example the ability to visualize and inspect the factor scores. A simulation study demonstrates that the method performs well in obtaining correct parameter estimates of a dynamic process. We also provide an empirical example and scripts for implementation in the open-source software R using the *lavaan* package.

KEYWORDS

Factor score regression; intensive longitudinal data; measurement error; reliability adjustment; vector autoregressive model

1. Introduction

In recent years, psychologists have increasingly focused on the research of within-person processes by collecting and analyzing intensive longitudinal data (ILD) instead of cross-sectional data (Hamaker & Wichers, 2017). ILD are characterized by a large number of repeated measures per individual over a relatively short period, for instance, six measures each day for two weeks (Ariens et al., 2020). This allows researchers to obtain insights into the participants' daily lives and study psychological processes, (e.g., how emotions carry over and interact with one another from one moment to the next; Kuppens & Verduyn, 2017).

To study such processes, researchers often use vector autoregressive (VAR) modeling by regressing variables at one time-point on those at the previous time point (e.g., Lütkepohl, 2005). These models assume that the variables are observed, but many constructs in psychological research (e.g., positive affect) are latent. This means that they are not directly observable and, instead, are measured indirectly through one or more items (e.g., positive affect can be assessed by asking to what extent a participant experiences a number of positive emotions, such as happiness and enthusiasm). The so-called measurement model (MM) describes which items measure which latent variable and to what extent (Millsap, 2011). It is commonly evaluated with item response theory (in case of categorical items; De Ayala, 2022) or factor analysis (in case of continuous items; Lawley & Maxwell, 1962). In this article, we focus on factor analysis, where the so-called factors correspond to the latent variables. VAR models can be extended to accommodate latent variables by including an MM, resulting in latent vector autoregressive (LVAR) models.[1] However, LVAR models are more intricate to estimate than regular VAR models, because the MM needs to be estimated in addition to the relations among the factors at subsequent time-points (the so-called structural model; SM).

One way to estimate LVAR models is using structural equation modeling (SEM), a one-step approach where the MM is estimated simultaneously with the SM. However, applied researchers frequently deviate from it and adopt a more intuitive, stepwise approach (Vogelsmeier et al., 2024): First, they solely estimate the MM using factor analysis while disregarding the SM. Then, the researchers compute factor scores for all individuals on all measurement occasions. The factor scores represent the positions on the underlying latent variables identified through the factor analysis. Finally, the factor scores are used as observed scores in regular VAR models or Dynamic Structural Equation Modeling (DSEM; Asparouhov et al., 2018). However, this naïve stepwise estimation ignores the measurement error in the data and the resulting inherent uncertainty in the factor scores (Grice, 2001), leading to biased estimates of the SM (Devlieger et al., 2016; Devlieger & Rosseel, 2017).

To address this issue, several stepwise approaches to SEM that separate the estimation of the MM and the SM while accounting for measurement error have been developed in recent years (for an overview, see Vermunt, 2024). Examples are factor score regression and path analysis (Devlieger et al., 2016; Devlieger & Rosseel, 2017), two-stage path analysis (Lai & Hsiao, 2022), or the structural-after-measurement (SAM) approach to SEM (Rosseel & Loh, 2022). They have been shown to obtain unbiased estimates of regression parameters and to outperform one-step SEM in small sample sizes (e.g., Kelcey, 2019; Savalei, 2019) as well as in the presence of misspecifications of the MM (Devlieger & Rosseel, 2017; Rosseel & Loh, 2022). However, these methods have been developed for cross-sectional data and may not readily accommodate LVAR models. For instance, using standard factor analysis on the stacked data (i.e., on all time-points and persons simultaneously) treats the observations across all persons as independent (i.e., the dependence of observations within a person is ignored). Moreover, VAR models are typically estimated via pairwise regression; that is, pairs of adjacent time-points are entered into the regression equation. This assumes that the scores at a particular time-point are affected only by the scores from the previous time-point.

To tailor the stepwise estimation to LVAR models, we present and evaluate Three-Step Latent Vector Autoregression (3S-LVAR) that extends two-stage path analysis (Lai & Hsiao, 2022) to ILD. Specifically, in the first step, the MM of each latent variable is evaluated using factor analysis (Lawley & Maxwell, 1962). In the second step, the factor scores are computed. In the third step, the SM (i.e., a VAR model) is estimated by regressing the factor scores on those from the previous time-point while correcting for their inherent uncertainty. 3S-LVAR thus effectively combines the strengths of SEM (i.e., accounting for the measurement of latent variables) with the intuitiveness of a stepwise approach, which offers various advantages. The method is designed to closely adhere to the intuitive procedure often used in applied research, making it user-friendly and accessible. The stepwise procedure allows researchers to scrutinize (and potentially adjust) the MM before estimating the parameters of the SM (Bakk et al., 2013; Lai et al., 2023; McNeish et al., 2021; Vermunt, 2010), and also increases robustness against local model misspecifications (Rosseel & Loh, 2022). Furthermore, the factor scores generated in the second step can be visualized and inspected for outliers (Hallgren et al., 2019) or trends and seasonality in the time series (Lütkepohl, 2005). Moreover, the factor scores can be reused in different analyses or by other researchers without needing to redo the measurement modeling. This distinguishes 3S-LVAR from the related SAM approach, which employs a stepwise estimation but does not provide factor scores and requires specifying the MM and the SM simultaneously. Lastly, while we focus on a simple LVAR model in the current article, the proposed stepwise estimation is highly flexible and can be extended to more complex models (e.g., DSEM). Here, using factor scores while correcting for their uncertainty can facilitate model estimation by reducing the dimensionality of the model.

The current article aims to evaluate how well 3S-LVAR performs in obtaining accurate parameter estimates of an LVAR model. The paper is organized as follows: Section 2 describes 3S-LVAR in detail. Section 3 presents a simulation study investigating 3S-LVAR's ability to obtain correct parameter and standard error (SE) estimates of a dynamic process under varying conditions. Section 4 illustrates 3S-LVAR with an application using the open-source software R (R Core Team, 2024). The final section discusses limitations and future directions for research.

2. Method

2.1. Data Structure

ILD pertain to repeated measures nested in individuals. Observed scores are indicated by y_{ijt}, where $i = 1, \ldots, I$ refers to the individuals, $j = 1, \ldots, J$ to the items (observed variables), and $t = 1, \ldots, T$ to the time-points. The number of time-points may differ across individuals (i.e., T_i), but the subscript i is omitted in the following for simplicity of notation. The items measure $q = 1, \ldots Q$ factors (latent variables). The responses of individual i at time-point t are captured in the $J \times 1$ vector $\mathbf{y}_{it}$, which are gathered into the $T \times J$ matrix $\mathbf{Y}_i$ per individual. We assume that the data are organized in the so-called long format, where each variable is represented in a single column, and each row represents one time-point per subject. For instance, a data set comprising seven variables measured 30 times for 20 individuals has 600 rows and seven columns.

2.2. The Latent Vector Autoregressive Model

The specification of an LVAR model comprises two parts. First, the MM describes which items measure which factor and captures the strength of the relationship between the observed items and the underlying factor. Second, the SM describes to what extent the factors predict themselves and each other at consecutive measurement occasions. Figure 1 shows an example of an LVAR model with two latent factors that are measured by three items each.[2]

The MM for a single observation is given by Lawley & Maxwell, (1962)

$$\mathbf{y}_{it} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}_{it} + \boldsymbol{\varepsilon}_{it}, \tag{1}$$

where the $J \times 1$ vector $\boldsymbol{\tau}$ represents the item intercepts and $\boldsymbol{\Lambda}$ indicates the $J \times Q$ matrix of factor loadings. The $Q \times 1$ vector $\boldsymbol{\eta}_{it}$ comprises the (true) latent variable scores[3] of individual i at time-point t, and $\boldsymbol{\varepsilon}_{it}$ is a $J \times 1$ vector of residuals, which are assumed to be independent of $\boldsymbol{\eta}_{it}$.

[2]The model posits that the observations at time-point t only depend on those at the previous time-point $t-1$. In other words, there is no direct effect of, for example, $\boldsymbol{\eta}_{it-2}$ on $\boldsymbol{\eta}_{it}$, and the association between these two observations is fully mediated through $\boldsymbol{\eta}_{it-1}$. In this article, we focus on this so-called lag-1 model as it is most commonly used in psychological research. The model can be extended by including more previous time-points as predictors.

[3]In the literature, these scores are sometimes also referred to as factor scores. To avoid confusion, we use the term "true scores" for the unobserved scores on the latent variable (i.e., $\boldsymbol{\eta}_{it}$), and "factor scores" for the estimated factor scores as proxies for the true scores (i.e., $\hat{\boldsymbol{\eta}}_{it}$). See the following section for an explanation of how these estimated factor scores are obtained.

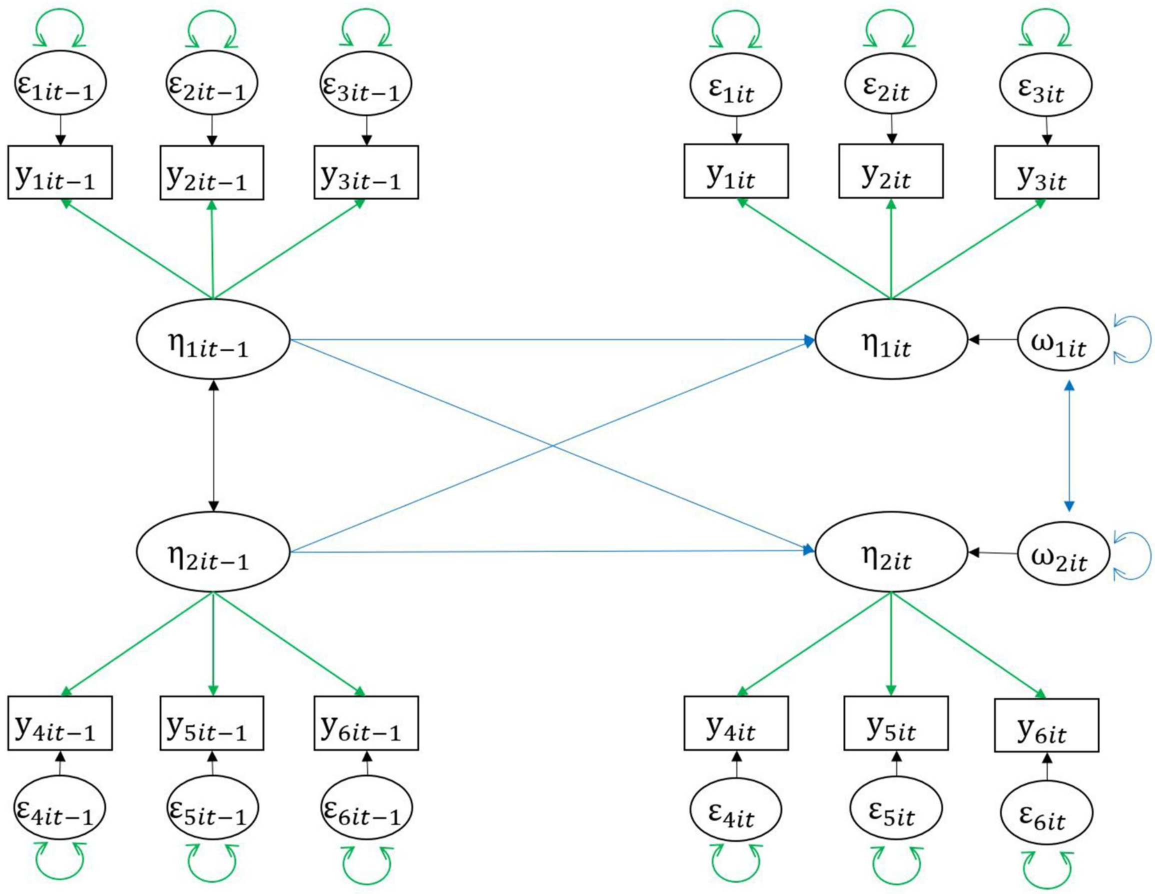

Figure 1. Visualizing a latent vector autoregressive model with two latent constructs and lag one.
Note. The green arrows represent parameters of the measurement model. The blue arrows represent parameters of the structural model. Observed scores are denoted by *y*, latent (true) scores by η, residuals by ε, and innovations by ω

The SM regresses the true scores at time-point t on those at the previous time-point:

$$\boldsymbol{\eta}_{it} = \boldsymbol{\delta} + \boldsymbol{\Phi}\boldsymbol{\eta}_{it-1} + \boldsymbol{\omega}_{it}, \qquad (2)$$

where $\boldsymbol{\eta}_{it}$ and $\boldsymbol{\eta}_{it-1}$ refer to the $Q \times 1$ vectors that contain the true scores of individual i at time-point t and $t-1$, respectively. The $Q \times 1$ vector $\boldsymbol{\delta}$ contains the intercepts (i.e., the predicted true scores if $\boldsymbol{\eta}_{it-1}$ is $\mathbf{0}$). The $Q \times Q$ matrix $\boldsymbol{\Phi}$ comprises the autoregressive (AR) parameters on the diagonal and the crossregressive (CR) parameters on the off-diagonal.[4] The AR parameters indicate to which extent the current value of a variable depends on its value at the previous measurement occasion. For example, in emotion research the AR parameter has been conceptualized as emotional inertia (Kuppens & Verduyn, 2017). A value close to zero means that a person quickly returns to their baseline after deviating from it, while someone with a larger coefficient will take longer (Hamaker, 2012; Jongerling et al., 2015). In contrast, CR parameters indicate to what extent a variable predicts other variables at subsequent time-points. For instance, a negative CR parameter may indicate that a high positive affect value predicts a low negative affect value at the next measurement occasion.

The AR and CR parameters thus pertain to the part of the current observation that is carried over from (i.e., can be predicted by) the previous one. The part that cannot be predicted is referred to as the innovation. The innovations comprise all internal and external events that affect the individual's process but were not part of the previous measurement (Hamaker, 2012). They are gathered in the $Q \times 1$ vector $\boldsymbol{\omega}_{it}$ and assumed to be distributed as $MVN(0, \mathrm{Z})$, where Z indicates the innovation (co)variance matrix. Note that the innovations $\boldsymbol{\omega}_{it}$ affect the true scores $\boldsymbol{\eta}_{it}$ and are thus carried over to the next observations through the AR and CR parameters. On the other hand, the measurement error $\boldsymbol{\varepsilon}_{it}$ affects the scores on the items and is thus not carried over. This is the single characteristic that distinguishes measurement error from innovations (Schuurman & Hamaker, 2019; Schuurman et al., 2015). For example, walking in the rain on the way to work affects an individual's overall mood, and the effect of this innovation persists at subsequent measurement occasions (with the strength of the carry-over being determined by the AR and CR parameters). In contrast, erroneously choosing value "6" over "5" on a mobile phone screen due to sunlight glare is error that will not propagate forward.

[4]Note that the model relies on two assumptions: First, the process is assumed to be stationary, which means that its means and (co)variances are constant across time (Lütkepohl, 2005). Second, the distance between all measurement occasions is assumed to be equal, for example one measurement every evening at 8pm.

2.3. Estimating the Latent Vector Autoregressive Model in a Stepwise Manner

A popular approach among applied researchers is to first estimate the parameters of the MM, then compute factor scores as proxies of the true latent variable scores, and finally use the factor scores to estimate the SM. However, this approach has been shown to lead to biased regression parameters, because the (co)variances of the estimated factor scores ($\boldsymbol{\Sigma}_{\hat{\eta}}$) are not equal to the (co)variances of the latent variables ($\boldsymbol{\Sigma}_{\eta}$) (Devlieger et al., 2016; Devlieger & Rosseel, 2017). The reason for this is that there is an inherent uncertainty in the factor score estimates, which arises from measurement error in the observed variables (Devlieger et al., 2016; Grice, 2001). This uncertainty increases when the number of items measuring a construct is small or when the factor loadings are small relative to the unique item variances (i.e., low reliability; Acito & Anderson, 1986). To avoid biased estimates of the AR and CR parameters, this uncertainty needs to be accounted for, which is the primary objective of 3S-LVAR. Its three steps will now be described in detail.

2.3.1. Step 1: Estimating the Measurement Model

In the first step, the MM (see Equation (1)) is evaluated with confirmatory factor analysis. The model parameters (intercepts, factor loadings, factor variances, and residual variances) are estimated using Maximum Likelihood estimation (Raykov & Marcoulides, 2006). Each of the Q constructs is evaluated separately to improve the robustness against misspecifications in the MM (see the "measurement blocks" in SAM; Rosseel & Loh, 2022). Note that in ILD, researchers usually try to obtain the average within-person factor structure. However, since 3S-LVAR uses regular rather than multilevel factor analysis, the resulting factor structure could be a mixture of within- and between-person factor structures (Hamaker et al., 2017). To avoid this potential conflation without switching to multilevel modeling, researchers should apply regular factor analysis to person-mean centered data because this type of centering removes any between-person variation (Bolger & Laurenceau, 2013). However, centering on the observed means creates a correlation between the predictors (i.e., $\boldsymbol{\eta}_{it-1}$) and the error terms, leading to an underestimation of the AR parameters (Nickell's bias; Nickell, 1981). Latent-mean centering (Asparouhov & Muthén, 2018) avoids this bias by accounting for the error in the sample mean estimates and is thus recommended. It is also possible to approximate this bias and correct for it, as we propose in Online Supplemental Material OSM-B. Further note that standard factor analysis is used, implying that all observations are treated as independent (e.g., 30 subjects with 50 observations each are treated as 1500 independent observations). Although the point estimates of the MM parameters are still unbiased when the temporal dependence of observations is ignored (Molenaar & Nesselroade, 2009), it is essential to use cluster-robust SEs that account for these dependencies (Abadie et al., 2023).

2.3.2. Step 2: Obtaining Factor Scores

In the second step, we compute the (predicted) factor scores $\hat{\boldsymbol{\eta}}_{it}$ with

$$\hat{\boldsymbol{\eta}}_{it} = \mathbf{A}(\mathbf{y}_{it} - \boldsymbol{\mu}), \tag{3}$$

where $\mathbf{A}$ pertains to the factor scoring matrix and $\boldsymbol{\mu}$ to the vector of indicator means. For example, in the case of regression factor scores $\mathbf{A}$ is given by Skrondal and Laake, (2001)

$$\mathbf{A}_R = \boldsymbol{\Psi}\boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}, \tag{4}$$

where $\boldsymbol{\Psi}$ refers to the $Q \times Q$ matrix of factor (co)variances, $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$ to the model-implied covariance matrix of the indicators, and $\boldsymbol{\Theta}$ to the $J \times J$ matrix with the unique item variances on the diagonal and (typically) zeros on the off-diagonal.

The factor scores are used as single indicators for the respective latent variables. To account for their uncertainty, we fix their loadings to $\boldsymbol{\Lambda}^* = \mathbf{A}\boldsymbol{\Lambda}$ and their residual variances to $\boldsymbol{\Theta}^* = \mathbf{A}\boldsymbol{\Theta}\mathbf{A}'$ (Vermunt, 2024). Under this specification, the model-implied variances of the factor scores equal their observed values. A convenient feature of regression factor scores is that λ_q^* for a single factor q equals the model-based reliability ρ_q, which in turn is equal to the ratio of the variance of the factor scores to the total factor variance ψ_q (which was estimated in step 1). Moreover, θ_q^* can be computed with $\psi_q\rho_q(1-\rho_q)$ (Vermunt, 2024).

2.3.3. Step 3: Estimating the Structural Model

In the third step, the parameters of the SM (see Equation 2) are estimated by regressing $\boldsymbol{\eta}_{it}$ on $\boldsymbol{\eta}_{it-1}$, using the factor scores $\hat{\boldsymbol{\eta}}_{it}$ and $\hat{\boldsymbol{\eta}}_{it-1}$ as single indicators with fixed loadings $\boldsymbol{\Lambda}^*$ and residual variances $\boldsymbol{\Theta}^*$. To obtain $\hat{\boldsymbol{\eta}}_{it-1}$, lagged versions of the factor score variables are created by duplicating the data and "shifting" the values by one row (see Figure C1 in the Online Supplemental Material OSM-C).[5]

The model is estimated using a Maximum Likelihood estimator (Raykov & Marcoulides, 2006). Due to the pairwise regression, this can be considered a pseudo Maximum Likelihood estimation since it does not consider all data points of an individual simultaneously. As in step 1, cluster-

[5]Note that this procedure leads to a difference in the values between the lagged variables (the predictors in the VAR model) and non-lagged variables (the outcomes). Since the scores prior to the first observation are unknown, the values of the predictors are missing for the first observation. Consequently, the predictors have one less known value in every individual (for example, with 50 observations, they will have 49 values and one NA each), and the variances will differ slightly between the lagged and non-lagged variables of the same construct. To counteract this, it is necessary to create an additional row for every individual in the data set, which represents the (unobserved) data after the last measurement occasion. For this observation, the outcomes are missing, but the predictors are not. After adding the additional row, both the predictor and outcome variables of the same construct include the same data, and thus their variances are equal. Note that the Maximum Likelihood estimation in *lavaan* can handle missing values.

robust SEs must be computed to account for the nested structure of the data (Abadie et al., 2023). However, these SEs may be inaccurate because they do not consider additional variance that is carried over from the previous steps. This occurs because the step 2 parameters $\mathbf{\Lambda}^*$ and $\mathbf{\Theta}^*$ are treated as known in step 3, despite being computed from the estimated step 1 parameters. In the OSM-A, we describe a standard way of computing SEs in stepwise estimation, where (functions of) estimated parameters are treated as fixed in later steps (Bakk et al., 2014; Gong & Samaniego, 1981).

The three steps can be implemented in standard statistical software that allows estimating latent variables, such as the R package *lavaan* (Rosseel, 2012). Wrapper functions for *lavaan* that automate these steps are provided on GitHub (https://github.com/mt-rein/3S-LVAR), and their application is illustrated in Section 4.

3. Simulation Study

3.1. Problem

We conducted a simulation study to evaluate how well 3S-LVAR performs in obtaining correct estimates of the AR and CR parameters and the SEs. We expected 3S-LVAR to outperform the naïve stepwise approach in which factor scores are used in a VAR model without correction for the factor scores' inherent uncertainty, in the following referred to as naïve factor scores (NFS). As reference points, we also compared 3S-LVAR to stepwise estimation with the (local) SAM approach (Rosseel & Loh, 2022), which is available in *lavaan*, and simultaneous (one-step) estimation with standard SEM. The models specified and estimated in these analyses are visualized in Figures C2 to C4 in OSM-C. For 3S-LVAR, cluster-robust SEs were computed with and without the proposed correction for stepwise estimation (see Table C1 in OSM-C). For NFS we computed cluster-robust SEs in step 3 but did not apply the SE correction (since researchers who would not correct the parameter estimates for measurement error are unlikely to adjust the SEs). Since SEM is a one-step estimation, the SE correction is not applicable, but cluster-robust standard errors are calculated. The implementation of SAM in *lavaan* also adjusts the SEs for stepwise estimation (Rosseel & Loh, 2022), but does not offer cluster-robust SEs. Note that the corrections for stepwise estimation in 3S-LVAR and SAM are based on the same standard procedure (Bakk et al., 2014; Gong & Samaniego, 1981), but may be implemented differently and thus lead to slightly different results.

Performance was rated with respect to three criteria: bias and variability of the estimated AR and CR coefficients, and SE recovery. Overall, we expected 3S-LVAR, SAM, and SEM to yield unbiased estimates of the regression parameters, while NFS was predicted to underestimate them. 3S-LVAR was expected to underestimate the SEs before correction but yield accurate estimates after correction. SEM was predicted to obtain correct SE estimates. Since cluster-robust SEs were not available for SAM, we expected these SEs to be underestimated. The NFS approach was also expected to underestimate the SEs.

To evaluate the performance under different conditions, seven aspects were manipulated. The first aspect concerned the SM. Specifically, we manipulated the size of the AR and CR effects. In line with earlier findings (Devlieger et al., 2016), the bias in the AR and CR estimates of NFS and the SEs was expected to be stronger for larger effect sizes.

The second and third aspects involved the sample size. We manipulated the number of individuals (I) and the number of observations per individual (T). Larger sample sizes improve the precision of estimates. Consequently, we expected the variability of the parameter estimates to decrease. Moreover, the SE recovery was expected to improve across all four analysis methods when the overall sample size increases (Devlieger et al., 2016).

The next three aspects introduced the presence or absence of between-person variation in the latent means, the innovation variances, and the regression coefficients. We expected 3S-LVAR, SAM, and SEM to obtain unbiased results regardless of the presence of differences in innovation variances or AR and CR effects. However, we predicted to obtain biased estimates of the AR effects for these methods in conditions with variation in the latent means due to Nickell's bias. We expected this bias to become smaller when T increases (Nickell, 1981) and to be eliminated by the correction described in OSM-B.[6] For NFS, we expected the bias to remain the same in conditions with and without differences in innovation variances or AR and CR effects, and to become larger in conditions with variation in the latent means. Similarly to the other three methods, this additional bias was expected to become smaller for larger values of T and to be eliminated with the proposed correction.

The seventh and most crucial aspect refers to the degree of uncertainty in the factor scores, which was manipulated by adjusting the model-based reliabilities (ρ). Ignoring the factor scores' inherent uncertainty has been shown to bias the estimates of regression parameters (Devlieger et al., 2016; Devlieger & Rosseel, 2017). For low values of ρ, we thus expected 3S-LVAR, SAM, and SEM to outperform NFS in obtaining correct point estimates. This difference should diminish and disappear as ρ approaches 1. Similarly, the SE recovery was expected to improve when ρ increases (Devlieger et al., 2016).

3.2. Design and Procedure

The seven manipulated aspects included the following levels:

a. effect sizes in $\mathbf{\Phi}$: "small effects" $\mathbf{\Phi} = \begin{pmatrix} .3 & .15 \\ .15 & .3 \end{pmatrix}$, "large effects" $\mathbf{\Phi} = \begin{pmatrix} .6 & .3 \\ .3 & .6 \end{pmatrix}$;

[6]Note that Nickell's bias can be avoided entirely using latent-mean centering, which will be discussed in the discussion section. However, this is beyond the scope of this paper as this is not straightforward to implement in the SEM framework in R, which is the focus of our article.

b. person-level sample size I: 25, 50;
c. observation-level sample size T: 25, 50;
d. between-person differences in latent means: no, yes;
e. between-person differences in regression coefficients: no, yes;
f. between-person differences in innovation variance: no, yes;
g. model-based reliability ρ: .5, .7., .9., .999;

The design resulted in $2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 4 = 256$ conditions. The size of the regression coefficients in $\mathbf{\Phi}$ was chosen such that it still satisfies the stationarity requirement in the "large effects" condition, and was then halved for the "small effects" condition. The chosen values for I and T are commonly found in studies that collect ILD (Can, 2020; van Berkel et al., 2017). For example, 25 observations roughly mirror an ESM design with four observations on seven days. For each cell of the design, 500 data sets (128,000 in total) were generated in R Version 4.4.1 (R Core Team, 2024) and analyzed with the four methods described above using *lavaan* 0.6-18 (Rosseel, 2012). Every data set comprised I time series of length T with two latent constructs measured by four items each and with equidistant observations. The simulation and analysis code can be found on https://osf.io/d6hs7.

3.3. Results

In the following, we evaluate the performance of 3S-LVAR and compare it to NFS, SEM, and SAM. There were no errors or non-convergences for 3S-LVAR, NFS, and SEM, but SAM led to an error in 58 data sets. These were re-estimated once, but the error persisted. We thus removed these data sets and report the results for 127,942 data sets below.

The results are reported separately for conditions without and with between-person differences in the latent means because we expected Nickell's bias in the latter. Additionally, since the two AR and two CR parameters showed nearly identical performance across all measures, we report estimates for only one AR (φ_{11}) and one CR (φ_{12}) parameter each. The results for φ_{22} and φ_{21} can be found in OSM-C, Tables C3 and C4.

To examine the goodness of the regression parameter recovery, we computed the absolute bias as the difference between the mean of the parameter estimates within a condition and the true parameter value. The relative bias scales the absolute bias by the true parameter value. To assess the estimates' variability, we calculated the Root Mean Square Error (RMSE), which is the average squared distance between estimated and true parameter values within conditions. SE recovery was assessed by dividing the average estimated SE for all replications within a condition by the standard deviation of the estimates across these replications. A value of 1 indicates that the SE estimator performs well. Values smaller than 1 indicate that (on average) the SEs are underestimated, while values greater than 1 indicate that they are overestimated. Equations for these criteria are displayed in OSM-C (Table C2).

3.3.1. No Differences in Latent Means

As expected, 3S-LVAR, SAM, and SEM obtained unbiased parameter estimates across all conditions (see Table 1).[7] Further in line with our expectations, NFS underestimated the regression parameters across all conditions. This effect was more pronounced in the AR parameters (mean absolute and relative bias were −.07 and −0.19, respectively) than in the CR parameters (mean absolute and relative bias of −.01 and −0.08, respectively). For NFS, larger effect sizes led to greater absolute bias in the AR parameters as hypothesized, but its relative bias slightly decreased. Moreover, for the CR parameter the absolute bias remained unchanged when the effect sizes increased, while the relative bias increased. Increasing the number of individuals or observations did not affect the bias for any method. It also did not reduce the RMSE substantially, which contradicted our expectations. Finally, as expected the parameter recovery of NFS improved when ρ increased until this approach performed equally well as the other methods when ρ approached 1.

The SE recovery results are also displayed in Table 1. Note that for 3S-LVAR, we report results with and without the correction for stepwise estimation. In line with our expectations, SEM estimated the SEs accurately in all conditions (mean ratio was 1 and 1.01 for AR and CR parameters, respectively). Contrary to expectation, the SEs obtained with 3S-LVAR were recovered accurately without the SE correction for stepwise estimation (average ratio = .99 for both parameters), but SEs were overestimated with the correction (average ratio = 1.14 and 1.1 for AR and CR parameters, respectively). This overestimation was higher when ρ was small. When ρ approached 1, the difference between the adjusted and unadjusted SEs disappeared. Moreover, larger effect sizes also increased the overestimation, while neither the sample size nor between-person differences in innovation variances or regression parameters affected the SE recovery substantially. SAM also overestimated the SEs (mean SE recovery was 1.09 and 1.04 for AR and CR parameters, respectively) and showed the same pattern with respect to the manipulated aspects. Finally, NFS also obtained accurate estimates for the SEs of the AR parameters (average ratio = .98) but underestimated those of the CR parameters (average ratio = .96). This underestimation is more pronounced when effect sizes are large or ρ is small.

3.3.2. With Differences in Latent Means

As expected, all methods obtained biased estimates of the AR effect in the conditions with between-person differences in the latent means (see Table 2). This bias becomes smaller when T increases. The proposed correction reduced the bias and performed better as ρ increases, fully eliminating the bias when ρ approached 1. Moreover, the CR parameter is slightly overestimated when ρ is small, but underestimated when ρ is large. The effects of manipulating the remaining

[7]Note that, in the presence of between-person differences in the innovation variances or AR and CR effects (aspects e and f), this means that the average or fixed effect is estimated without bias.

Table 1. Bias, root mean square error, and standard error recovery for conditions without latent mean differences.

			Parameter							
			AR (φ_{11})				CR (φ_{12})			
Manipulated aspect	Level	Method	AB	RB	RMSE	SER	AB	RB	RMSE	SER
Overall		3S-LVAR	0	−0.01	0.04	0.99 (1.14)	0	0	0.04	0.99 (1.1)
		NFS	−0.07	−0.19	0.08	0.98	−0.01	−0.08	0.03	0.96
		SAM	0	−0.01	0.04	1.09	0	0	0.04	1.04
		SEM	0	0	0.04	1	0	0	0.04	1.01
Effect sizes (Φ)	small	3S-LVAR	0	−0.01	0.04	0.98 (1.06)	0	0	0.04	0.99 (1.03)
		NFS	−0.05	−0.2	0.06	0.98	−0.01	−0.11	0.03	0.98
		SAM	0	−0.01	0.04	1.04	0	0	0.04	1.01
		SEM	0	0	0.04	0.99	0	0	0.04	1
	large	3S-LVAR	0	0	0.04	1.01 (1.23)	0	0	0.04	1 (1.16)
		NFS	−0.09	−0.18	0.1	0.98	−0.01	−0.04	0.03	0.94
		SAM	0	0	0.04	1.14	0	0	0.04	1.06
		SEM	0	0	0.04	1.02	0	0	0.04	1.03
Individuals (*I*)	25	3S-LVAR	0	−0.01	0.04	0.99 (1.14)	0	0	0.04	0.99 (1.1)
		NFS	−0.07	−0.2	0.08	0.97	−0.01	−0.08	0.03	0.96
		SAM	0	−0.01	0.04	1.09	0	0	0.04	1.04
		SEM	0	0	0.04	1	0	0	0.04	1.02
	50	3S-LVAR	0	0	0.03	1 (1.15)	0	0	0.03	0.99 (1.09)
		NFS	−0.07	−0.19	0.08	0.99	−0.01	−0.07	0.03	0.96
		SAM	0	0	0.03	1.09	0	0	0.03	1.03
		SEM	0	0	0.03	1.01	0	0	0.03	1.01
Observations (*T*)	25	3S-LVAR	0	−0.01	0.04	1 (1.15)	0	0	0.04	0.99 (1.1)
		NFS	−0.07	−0.19	0.08	0.98	−0.01	−0.08	0.03	0.96
		SAM	0	−0.01	0.04	1.1	0	0	0.04	1.04
		SEM	0	0	0.04	1.01	0	0	0.04	1.02
	50	3S-LVAR	0	0	0.03	0.99 (1.14)	0	0	0.03	0.99 (1.09)
		NFS	−0.07	−0.19	0.08	0.98	−0.01	−0.07	0.03	0.96
		SAM	0	0	0.03	1.08	0	0	0.03	1.03
		SEM	0	0	0.03	1	0	0	0.03	1.01
Variation in $\mathbf{\Phi}$	yes	3S-LVAR	0	0	0.04	1 (1.15)	0	0	0.04	0.99 (1.1)
		NFS	−0.07	−0.19	0.08	0.99	−0.01	−0.07	0.03	0.96
		SAM	0	0	0.04	1.09	0	0	0.04	1.03
		SEM	0	0	0.04	1.01	0	0	0.04	1.01
	no	3S-LVAR	0	−0.01	0.04	0.99 (1.14)	0	0	0.04	0.99 (1.09)
		NFS	−0.07	−0.2	0.08	0.98	−0.01	−0.08	0.03	0.96
		SAM	0	−0.01	0.04	1.09	0	0	0.04	1.04
		SEM	0	0	0.04	1	0	0	0.04	1.01
Variation in $\mathcal{Z}$	yes	3S-LVAR	0	−0.01	0.04	0.99 (1.14)	0	0	0.04	0.99 (1.1)
		NFS	−0.07	−0.19	0.08	0.98	−0.01	−0.08	0.03	0.96
		SAM	0	−0.01	0.04	1.09	0	0	0.04	1.04
		SEM	0	0	0.04	1	0	0	0.04	1.01
	no	3S-LVAR	0	−0.01	0.04	1 (1.14)	0	0	0.04	0.99 (1.09)
		NFS	−0.07	−0.19	0.08	0.98	−0.01	−0.07	0.03	0.96
		SAM	0	−0.01	0.04	1.09	0	0	0.04	1.04
		SEM	0	0	0.04	1	0	0	0.04	1.01
Reliability (ρ)	.5	3S-LVAR	0	−0.01	0.06	1.02 (1.55)	0	0	0.06	1.01 (1.37)
		NFS	−0.16	−0.43	0.16	0.98	−0.04	−0.22	0.04	0.91
		SAM	0	−0.01	0.06	1.22	0	0	0.06	1.09
		SEM	0	0	0.06	1.04	0	0.01	0.06	1.05
	.7	3S-LVAR	0	−0.01	0.04	1 (1.06)	0	0	0.04	0.99 (1.03)
		NFS	−0.09	−0.25	0.09	0.99	−0.01	−0.07	0.03	0.95
		SAM	0	−0.01	0.04	1.13	0	0	0.04	1.04
		SEM	0	0	0.04	1.01	0	0	0.04	1.01
	.9	3S-LVAR	0	0	0.03	0.98 (0.98)	0	0	0.03	0.99 (0.99)
		NFS	−0.03	−0.08	0.04	0.98	0	−0.01	0.02	0.98
		SAM	0	0	0.03	1.03	0	0	0.03	1.01
		SEM	0	0	0.03	0.98	0	0	0.03	1
	.999	3S-LVAR	0	0	0.02	0.98 (0.98)	0	0	0.02	0.99 (0.99)
		NFS	0	−0.01	0.02	0.98	0	0	0.02	0.99
		SAM	0	0	0.02	0.98	0	0	0.02	1
		SEM	0	0	0.02	0.98	0	0	0.02	0.99

Note. Numbers in brackets pertain to the standard errors that were adjusted for stepwise estimation.
AB: absolute bias; RB: relative bias; RMSE: Root Mean Square Error; SER: Standard Error Recovery; AR: autoregressive parameter (construct 1 regressed on itself); CR: crossregressive parameter (construct 1 regressed on construct 2); 3S-LVAR: Three-Step Latent Vector Autoregression; NFS: naïve factor scores; SAM: structural-after-measurement approach; SEM: structural equation modeling.

aspects were the same as in the previous section. For brevity, we thus only included the overall results and those of the aspects reliability (ρ) and number of observations (T) in Table 2. The full table can be found in OSM-C, Table C5.

3.4. Conclusion

The results of the simulation study demonstrate that 3S-LVAR obtains unbiased point estimates of the AR and CR

Table 2. Bias and RMSE for selected conditions with latent mean differences.

Manipulated aspect	Level	Method	AR (φ_{11}) AB	AR (φ_{11}) RB	AR (φ_{11}) RMSE	CR (φ_{12}) AB	CR (φ_{12}) RB	CR (φ_{12}) RMSE
Overall		3S-LVAR	−0.06	−0.17	0.06	0	0	0.04
		3S-LVAR (adj)	−0.02	−0.05	0.04			
		NFS	−0.12	−0.32	0.12	−0.02	−0.13	0.03
		NFS (adj)	−0.08	−0.2	0.08			
		SAM	−0.06	−0.17	0.06	0	0	0.04
		SAM (adj)	−0.02	−0.05	0.04			
		SEM	−0.06	−0.17	0.06	0	0	0.04
		SEM (adj)	−0.02	−0.05	0.04			
Observations (*T*)	25	3S-LVAR	−0.08	−0.23	0.08	0	0	0.05
		3S-LVAR (adj)	−0.02	−0.07	0.05			
		NFS	−0.13	−0.37	0.13	−0.03	−0.14	0.04
		NFS (adj)	−0.08	−0.21	0.09			
		SAM	−0.08	−0.23	0.08	0	0	0.05
		SAM (adj)	−0.02	−0.07	0.05			
		SEM	−0.08	−0.23	0.08	0	0.01	0.05
		SEM (adj)	−0.02	−0.07	0.05			
	50	3S-LVAR	−0.04	−0.11	0.04	0	0	0.03
		3S-LVAR (adj)	−0.01	−0.03	0.03			
		NFS	−0.1	−0.27	0.1	−0.02	−0.11	0.03
		NFS (adj)	−0.07	−0.2	0.08			
		SAM	−0.04	−0.11	0.04	0	0	0.03
		SAM (adj)	−0.01	−0.03	0.03			
		SEM	−0.04	−0.11	0.04	0	0	0.03
		SEM (adj)	−0.01	−0.03	0.03			
Reliability (ρ)	.5	3S-LVAR	−0.08	−0.24	0.09	0	0	0.05
		3S-LVAR (adj)	−0.04	−0.12	0.07			
		NFS	−0.21	−0.56	0.21	−0.05	−0.28	0.05
		NFS (adj)	−0.17	−0.45	0.17			
		SAM	−0.08	−0.24	0.09	0.01	0.07	0.06
		SAM (adj)	−0.04	−0.12	0.07			
		SEM	−0.08	−0.24	0.09	0.01	0.08	0.06
		SEM (adj)	−0.04	−0.12	0.07			
	.7	3S-LVAR	−0.06	−0.17	0.07	0	0.01	0.04
		3S-LVAR (adj)	−0.02	−0.05	0.04			
		NFS	−0.14	−0.38	0.14	−0.02	−0.13	0.03
		NFS (adj)	−0.1	−0.26	0.1			
		SAM	−0.06	−0.17	0.07	0	0.01	0.04
		SAM (adj)	−0.02	−0.05	0.04			
		SEM	−0.06	−0.17	0.06	0	0.01	0.04
		SEM (adj)	−0.02	−0.05	0.04			
	.9	3S-LVAR	−0.05	−0.14	0.05	−0.01	−0.03	0.03
		3S-LVAR (adj)	−0.01	−0.02	0.03			
		NFS	−0.07	−0.21	0.07	−0.01	−0.05	0.03
		NFS (adj)	−0.03	−0.09	0.04			
		SAM	−0.05	−0.14	0.05	−0.01	−0.03	0.03
		SAM (adj)	−0.01	−0.02	0.03			
		SEM	−0.05	−0.14	0.05	−0.01	−0.03	0.03
		SEM (adj)	−0.01	−0.02	0.03			
	.999	3S-LVAR	−0.04	−0.13	0.05	−0.01	−0.04	0.03
		3S-LVAR (adj)	0	0	0.03			
		NFS	−0.05	−0.14	0.05	−0.01	−0.04	0.03
		NFS (adj)	−0.01	−0.01	0.03			
		SAM	−0.04	−0.13	0.05	−0.01	−0.04	0.03
		SAM (adj)	0	0	0.03			
		SEM	−0.04	−0.13	0.05	−0.01	−0.04	0.03
		SEM (adj)	0	0	0.03			

Note. AB: absolute bias; RB: relative bias; RMSE: Root Mean Square Error; AR: autoregressive parameter (construct 1 regressed on itself); CR: crossregressive parameter (construct 1 regressed on construct 2); 3S-LVAR: Three-Step Latent Vector Autoregression; NFS: naïve factor scores; SAM: structural-after-measurement approach; SEM: structural equation modeling; (adj): adjusted for Nickell's bias.

parameters under various conditions in the absence of between-person differences in the latent means. In the presence of such differences, the bias resulting from the observed-mean centering can at least partially be corrected for. Overall, the method obtains the same point estimates as stepwise estimation with SAM, with only marginal differences to one-step SEM. Importantly, using factor scores without correcting for their uncertainty (NFS) yields biased parameter estimates. Interestingly, 3S-LVAR accurately estimated the SEs even without adjusting them for stepwise estimation, whereas the adjustment overestimates the SEs, particularly for lower values of ρ. This means that the adjusted SEs may be too conservative, which can reduce the power of the hypothesis test.

4. Empirical Example

In the following, we apply the 3S-LVAR method to an empirical data set (Nezlek & Kuppens, 2008). The data set contains self-report measures of emotion from 68 male and 85 female undergraduates that were collected across four weeks. Every evening, participants rated to what extent they experienced positive emotions (e.g., to what extent they felt enthusiastic or joyful) and negative emotions (e.g., guilty or upset) on a 7-point Likert Scale. The participants provided between 10 and 28 daily measures each ($M = 20.1$, $SD = 2.95$, 3072 observations in total).

Specifically, we explore to what extent four constructs related to emotional experience interact with themselves and one another from one day to the next. Our constructs of interest were *positive activated affect*, *positive deactivated affect*, *negative activated affect*, and *negative deactivated affect*.[8] *Positive activated affect* was measured with *enthusiastic*, *happy*, *active*, *energetic*, *alert*, *proud*, and *joyful*. *Positive deactivated affect* was measured with *calm*, *satisfied*, and *relaxed*. *Negative activated affect* was measured with the items *guilty*, *nervous*, *afraid*, *angry*, *ashamed*, *embarrassed*, *upset*, and *disgusted*. Lastly, *negative deactivated affect* was measured with *sluggish*, *sad*, *tired*, *bored*, and *sleepy*.

The functions are available on https://github.com/mt-rein/3S-LVAR. The repository also includes instructions on how to install and load the functions. Before beginning with the analysis, the user needs to ensure that the data are in the long format and sorted by participant and time-point. Moreover, the time interval between observations of any individual must be roughly equivalent. Thus, missing observations (i.e., a participant did not fill in the survey when prompted) should not be removed when cleaning the data. Finally, the observations should be within-person mean centered to disaggregate within- and between-person effects (Bolger & Laurenceau, 2013). A helper function for this is also provided on GitHub.

4.1. Step 1: Estimating the Measurement Model

The first step is performed using the function `step1()`, which has three arguments. The first argument (`data`) indicates the data object. The second argument (`measurementmodel`) requires specifying the MM using the *lavaan* syntax (see https://lavaan.ugent.be/tutorial). The third argument (`id`) provides the name of the variable that indicates which observation belongs to which individual. The following code performs this step for the example data:

```
model <- "
PA_act =~ ENTHUS + HAPPY + ACTIVE + ENERG + ALERT + PROUD + JOY
PA_deact =~ CALM + SATIS + RELAX
NA_act =~
GUILTY + NERVE + AFRAID + ANGRY + ASHAME + EMBAR + UPSET + DISG
NA_deact =~ SLUG + SAD + TIRED + BORED + SLEEP
"
output_step1 <- step1(data = data, measurementmodel = model,
id = "numid")
```

The first command creates an object with *lavaan* syntax to designate which latent construct is measured by which variable in the data set. The operator =~ represents factor loadings. The second command estimates the MM and saves the output in the object `output_step1`. The output comprises two elements: `fit_step1` is the *lavaan* fit object (which can be inspected with the `summary()` function), and `data` is the data set that has been used to estimate the model (which is needed again in step 2).

4.2. Step 2: Obtaining Factor Scores

Next, the output of `step1()` is entered into the `step2()` function to perform the second step. This function has one argument (`step1output`), as demonstrated in the following code:

```
output_step2 <- step2(step1output = output_step1)
```

The output comprises four elements: `data` is the original data set with appended regression factor scores, `lambda_star` and `theta_star` are vectors containing the diagonals of $\mathbf{\Lambda}^*$ (whose values equal the reliabilities ρ) and $\mathbf{\Theta}^*$, respectively, and `fit_step1` is the *lavaan* fit object from step 1.

4.3. Step 3: Estimating the Structural Model

The third step is performed using `step3()`. This function has two arguments: `step2output` provides the output of `step2()`, and `structuralmodel` specifies the SM using *lavaan* syntax[9]. The latter is optional. If it is omitted, the function automatically specifies a SM that includes the auto- and crossregressive effects between all constructs. The following code performs this step with the default VAR model and prints the model summary:

```
output_step3 <- step3(step2output = output_step2)
summary(output_step3$fit)
```

The output comprises three elements. First, `fit_step3` is the *lavaan* object of estimating the SM. It can be used in the `summary()` function to obtain an overview of the fitted model. Second, `data` contains the data set used to estimate the model. Note that `step3()` automatically creates lagged variables and an additional row per participant. Third, `phi` is the matrix of estimated regression parameters.

4.4. Standard Error Adjustment

The SEs can be adjusted to account for the stepwise procedure (see OSM-A) using the function `stepwiseSE()`

[8]These constructs refer to a circumplex model of affect with the dimensions valence (positive/negative) and arousal (activated/deactivated, Feldman Barrett & Russel, 1998).

[9]See the README on GitHub for more information on how to specify the SM manually.

shown below. Its two arguments (`step2output` and `step3output`) require the output of `step2()` and `step3()`, respectively.

```
adjustedSE <- stepwiseSE(step2output = output_step2,
step3output = output_step3)
```

The output provides three vectors that contain adjusted values: `SE`, `z_values`, and `p_values`. Note that the function is computationally expensive and can take several minutes for models with multiple latent variables. For instance, using the SE adjustment in our example (with four factors) took 8 minutes. Moreover, the simulation study results indicate that the correction may overestimate the SEs, which may reduce the power to detect existing effects.

4.5. Results

For the constructs in the example, the values of ρ were .86 for *positive activated affect*, .65 for *positive deactivated affect*, .81 for *negative activated affect*, and .84 for *negative deactivated affect*. Thus, *positive deactivated affect* had the lowest model-based reliability, likely because it was measured by the fewest items (Grice, 2001). Table 3 displays the point estimates and the adjusted SEs of the AR and CR parameters for 3S-LVAR and naïve factor scores (i.e., using factor scores in the VAR model without correcting for their inherent uncertainty).

Using 3S-LVAR, all AR effects were significant ($p < .05$) and ranged from .3 to .43, which means that emotional experience carried over substantially from one day to the next. However, only one out of 12 CR effects was significant: a high score on *positive activated affect* predicted a high score on *negative deactivated affect* on the next day ($\varphi = .16$). When using naïve factor scores, the estimates for the AR parameters were generally lower compared to using 3S-LVAR. This was most noticeable for *positive deactivated affect*, which had the lowest reliability ($\rho = .65$): The parameter estimate was twice as large when using 3S-LVAR. For the other parameters with rather high reliabilities, the differences were less pronounced. Overall, the conclusions regarding the AR parameters remained the same. However, there were was a notable difference with respect to the CR parameters. The effect of *positive activated affect* on *positive deactivated affect* was significant and positive when using naïve factor scores, but non-significant and negative for 3S-LVAR. This demonstrates that neglecting to account for the factor scores' uncertainty can change the statistical inferences drawn from the data.

Table 3. Auto- and crossregressive effects using 3S-LVAR and naïve factor scores.

	Predictor			
Outcome	PA act	PA deact	NA act	NA deact
Using 3S-LVAR				
PA act	0.31 (0.07)*	0.07 (0.1)	0.08 (0.05)	0.07 (0.04)
PA deact	−0.01 (0.07)	0.43 (0.11)*	0.07 (0.05)	0.06 (0.04)
NA act	0.07 (0.04)	−0.09 (0.07)	0.3 (0.05)*	−0.02 (0.03)
NA deact	0.16 (0.06)*	−0.17 (0.09)	0 (0.05)	0.36 (0.04)*
Using Naïve Factor Scores				
PA act	0.26 (0.03)*	0.08 (0.04)	0.05 (0.04)	0.05 (0.03)
PA deact	0.06 (0.02)*	0.21 (0.03)*	0 (0.02)	0.05 (0.02)*
NA act	0.03 (0.02)	−0.05 (0.03)	0.26 (0.04)*	−0.02 (0.02)
NA deact	0.06 (0.02)*	−0.06 (0.04)	0.03 (0.03)	0.29 (0.03)*

Note. Numbers in brackets represent the standard errors. The auto-regressive effects have been adjusted for Nickell's bias. PA act: positive activated affect; PA deact: positive deactivated affect; NA act: negative activated affect; NA deact: negative deactivated affect. *indicates a *p*-value lower than .05.

5. Discussion

In this article, we introduced 3S-LVAR, a three-step approach for estimating VAR models with latent variables. The proposed method adheres as closely as possible to the intuitive stepwise approach applied researchers use to study dynamics in psychological constructs (i.e., first obtaining the MM of the latent variables, then computing factor scores, and finally using these in VAR modeling). In 3S-LVAR, this approach is modified by considering the uncertainty of the factor scores (which is caused by measurement error in the data) in the final step. The simulation study showed that 3S-LVAR greatly outperforms the naïve approach and is on par with the SEM and SAM methods in recovering the parameters of a bivariate dynamic process. The novel modification of the conventional stepwise approach can be seen as a first building block that offers potential for future extensions.

One such extension could be related to heterogeneity in the SM. While 3S-LVAR models intra-individual variation (i.e., within-person differences) in the constructs of interest, it currently does not account for between-person differences in the psychological processes (i.e., the parameters of the SM). The proposed pairwise estimation with SEM obtains unbiased estimates of the average AR and CR parameters when there is between-person variation in the regression parameters or innovation variances. However, the estimates become biased when observed-mean centering. This bias can be avoided with latent-mean centering, which is implemented in the DSEM framework in Mplus (Asparouhov & Muthén, 2018). In R, latent-mean centering can be achieved by including a random intercept in the model. This is theoretically possible in SEM (Usami et al., 2019) but the estimation becomes computationally expensive for models with many time-points, as the model-implied covariance matrix grows quadratically with the number of time points. A better alternative is to use State Space Modeling, which is a latent variable model that is specifically designed to incorporate dynamics (Hunter, 2017).

Furthermore, 3S-LVAR could be extended to model heterogeneity in AR and CR parameters or innovation variances instead of simply estimating an average effect: First, covariates can be used to find between-group differences (e.g., whether an intervention and a control group differ in their emotional inertia). Second, with multi-level modeling the parameters can be allowed to vary across individuals (random effects). Third, clustering (such as mixture modeling) can be used to find latent classes based on the SM parameters (i.e., groups of individuals that share similar AR and CR effects).

Besides heterogeneity in the SM, there can also be between-person differences in the MM (i.e., measurement non-invariance), such as differences in the factor loadings. Ignoring these can bias the estimates of the SM (Chen, 2008; Guenole & Brown, 2014). 3S-LVAR could consider measurement non-invariance by estimating (partially) person-or group-specific MMs in step 1 and person- or group-specific fixed loadings $\boldsymbol{\Lambda}^*$ and residual variances $\boldsymbol{\Theta}^*$ in step 2 (see Lai et al., 2023).

Moreover, the measurement intervals are assumed to be equally spaced (e.g., one measurement every evening at 8 pm). However, ILD are typically collected at (semi)random time-points throughout the day (Scollon et al., 2009). Additionally, participants are usually not prompted during the night, leading to considerably longer night intervals than day intervals. To account for this, 3S-LVAR could be extended to continuous time modeling (Driver et al., 2017).

In addition, 3S-LVAR uses Maximum Likelihood estimation and thus assumes that the indicator variables are continuous and normally distributed. Psychologists commonly use categorical and/or non-normally distributed items (Cain et al., 2017). However, with a sufficient number of response options (e.g., five or more) these variables can be treated as continuous if the non-normality is not too severe (Norman, 2010). In addition, 3S-LVAR could be extended to use item response theory instead of confirmatory factor analysis in the first step (Lai & Hsiao, 2022).

Lastly, contrary to our expectations the results of the simulation study revealed that the proposed SE correction for 3S-LVAR overestimated the sampling variance of the AR and CR parameters (i.e., the SEs may be overcorrected) for small or medium values of ρ. The uncorrected SEs were accurate and similar to those obtained by one-step SEM in all conditions. SAM (which uses a correction based on the same standard procedure) also overestimated the SEs. At least in the conditions explored in the current article's simulation study, using cluster-robust SEs in steps 1 and 3 was sufficient to obtain accurate SEs. Bakk et al. (2014) also found that correcting for stepwise estimation is not always necessary and may lead to overly conservative SEs. However, it is important to be cognizant of this adjustment. Researchers analyzing ILD may not be concerned about the loss of power from overly conservative SEs due to the typically large sample sizes. They may thus prefer to use the correction for a more conservative approach.

Disclosure Statement

The authors report there are no competing interests to declare.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, *138*, 1–35. https://doi.org/10.1093/qje/qjac038

Acito, F., & Anderson, R. D. (1986). A simulation study of factor score indeterminacy. *Journal of Marketing Research*, *23*, 111–118. https://doi.org/10.2307/3151658

Ariens, S., Ceulemans, E., & Adolf, J. K. (2020). Time series analysis of intensive longitudinal data in psychosomatic research: A methodological overview. *Journal of Psychosomatic Research*, *137*, 110191. https://doi.org/10.1016/j.jpsychores.2020.110191

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

Asparouhov, T., & Muthén, B. (2018). Latent variable centering of predictors and mediators in multilevel and time-series models. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*, 119–142. https://doi.org/10.1080/10705511.2018.1511375

Bakk, Z., Oberski, D. L., & Vermunt, J. K. (2014). Relating latent class assignments to external variables: standard errors for correct inference. *Political Analysis*, *22*, 520–540. https://doi.org/10.1093/pan/mpu003

Bakk, Z., Tekle, F. B., & Vermunt, J. K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology*, *43*, 272–311. https://doi.org/10.1177/0081175012470644

Bolger, N., & Laurenceau, J. P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. Guilford Press.

Cain, M. K., Zhang, Z., & Yuan, K. H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods*, *49*, 1716–1735. https://doi.org/10.3758/s13428-016-0814-1

Can, O. (2020). Experience sampling methodology: a systematic review and discussion for organizational research. *Electronic Journal of Business Research Methods*, *18*, 129–141. https://doi.org/10.34190/JBRM.18.2.005

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*, 1005–1018. https://doi.org/10.1037/a0013193

De Ayala, R. J. (2022). *The theory and practice of item response theory*. (2nd ed.). Guilford Press.

Devlieger, I., Mayer, A., & Rosseel, Y. (2016). Hypothesis testing using factor score regression: A comparison of four methods. *Educational and Psychological Measurement*, *76*, 741–770. https://doi.org/10.1177/0013164415607618

Devlieger, I., & Rosseel, Y. (2017). Factor score path analysis. *Methodology*, *13*, 31–38. https://doi.org/10.1027/1614-2241/a000130

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, *18*, 450–469. https://doi.org/10.1093/pan/mpq025

Vermunt, J. K. (2024). Stepwise latent variable modeling: An overview of approaches. https://jeroenvermunt.nl/stepwiseLVM2024.pdf

Vogelsmeier, L. V. D. E., Jongerling, J., & Maassen, E. (2024). Assessing and accounting for measurement in intensive longitudinal studies: Current practices, considerations, and avenues for improvement. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *33*, 2107–2118. https://doi.org/10.1007/s11136-024-03678-0

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*, 1–35. https://doi.org/10.18637/jss.v077.i05

Feldman Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, *74*, 967–984. https://doi.org/10.1037/0022-3514.74.4.967

Gong, G., & Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: Theory and applications. *The Annals of Statistics*, *9*, 861–869. https://doi.org/10.1214/aos/1176345526

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, *6*, 430–450. https://doi.org/10.1037/1082-989X.6.4.430

Guenole, N., & Brown, A. (2014). The consequences of ignoring measurement invariance for path coefficients in structural equation models. *Frontiers in Psychology*, *5*, 980. https://doi.org/10.3389/fpsyg.2014.00980

Hallgren, K. A., McCabe, C. J., King, K. M., & Atkins, D. C. (2019). Beyond path diagrams: Enhancing applied structural equation modeling research through data visualization. *Addictive Behaviors*, *94*, 74–82. https://doi.org/10.1016/j.addbeh.2018.08.030

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). Guilford.

Hamaker, E. L., Schuurman, N. K., & Zijlmans, E. A. (2017). Using a few snapshots to distinguish mountains from waves: weak factorial invariance in the context of trait-state research. *Multivariate Behavioral Research*, *52*, 47–60. https://doi.org/10.1080/00273171.2016.1251299

Hamaker, E. L., & Wichers, M. (2017). No time like the present. *Current Directions in Psychological Science*, *26*, 10–15. https://doi.org/10.1177/0963721416666518

Hunter, M. D. (2017). State space modeling in an open source, modular, structural equation modeling environment. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 307–324. https://doi.org/10.1080/10705511.2017.1369354

Jongerling, J., Laurenceau, J. P., & Hamaker, E. L. (2015). A multilevel AR(1) model: Allowing for inter-individual differences in trait-scores, inertia, and innovation variance. *Multivariate Behavioral Research*, *50*, 334–349. https://doi.org/10.1080/00273171.2014.1003772

Kelcey, B. (2019). A robust alternative estimator for small to moderate sample SEM: Bias-corrected factor score path analysis. *Addictive Behaviors*, *94*, 83–98. https://doi.org/10.1016/j.addbeh.2018.10.032

Kuppens, P., & Verduyn, P. (2017). Emotion dynamics. *Current Opinion in Psychology*, *17*, 22–26. https://doi.org/10.1016/j.copsyc.2017.06.004

Lai, M. H. C., & Hsiao, Y. Y. (2022). Two-stage path analysis with definition variables: An alternative framework to account for measurement error. *Psychological Methods*, *27*, 568–588. https://doi.org/10.1037/met0000410

Lai, M. H. C., Tse, W. W.-Y., Zhang, G., Li, Y., & Hsiao, Y.-Y. (2023). Correcting for unreliability and partial invariance: A two-stage path analysis approach. *Structural Equation Modeling: A Multidisciplinary Journal*, *30*, 258–271. https://doi.org/10.1080/10705511.2022.2125397

Lawley, D. N., & Maxwell, A. E. (1962). Factor analysis as a statistical method. *The Statistician*, *12*, 209. https://doi.org/10.2307/2986915

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. https://doi.org/10.1007/978-3-540-27752-1

McNeish, D., Mackinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2021). Measurement in intensive longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*, 807–822. https://doi.org/10.1080/10705511.2021.1915788

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, *52*, 2287–2305. https://doi.org/10.3758/s13428-020-01398-0

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge.

Molenaar, P. C., & Nesselroade, J. R. (2009). The recoverability of P-technique factor analysis. *Multivariate Behavioral Research*, *44*, 130–141. https://doi.org/10.1080/00273170802620204

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*, 181–202. https://doi.org/10.1007/BF02294246

Nezlek, J. B., & Kuppens, P. (2008). Regulating positive and negative emotions in daily life. *Journal of Personality*, *76*, 561–580. https://doi.org/10.1111/j.1467-6494.2008.00496.x

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, *49*, 1417. https://doi.org/10.2307/1911408

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education: theory and Practice*, *15*, 625–632. https://doi.org/10.1007/s10459-010-9222-y

R Core Team. (2024). *R: A language and environment for statistical computing*. In (Version 4.4.1) R Foundation for Statistical Computing.

Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling*. (2nd ed.). Lawrence Erlbaum.

Rosseel, Y. (2012). lavaan: An R Package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Rosseel, Y., & Loh, W. W. (2022). A structural after measurement (SAM) approach to structural equation modeling. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000503

Savalei, V. (2019). A comparison of several approaches for controlling measurement error in small samples. *Psychological Methods*, *24*, 352–370. https://doi.org/10.1037/met0000181

Schuurman, N. K., & Hamaker, E. L. (2019). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological Methods*, *24*, 70–91. https://doi.org/10.1037/met0000188

Schuurman, N. K., Houtveen, J. H., & Hamaker, E. L. (2015). Incorporating measurement error in n = 1 psychological autoregressive modeling. *Frontiers in Psychology*, *6*, 1038. https://doi.org/10.3389/fpsyg.2015.01038

Scollon, C. N., Kim-Prieto, C., & Diener, E. (2009). Experience sampling: Promises and pitfalls, strengths and weaknesses. In E. Diener (Ed.), *Assessing Well-Being. The Collected Works of Ed Diener*. Springer. https://doi.org/10.1007/978-90-481-2354-4

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, *66*, 563–575. https://doi.org/10.1007/BF02296196

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, *24*, 637–657. https://doi.org/10.1037/met0000210

van Berkel, N., Ferreira, D., & Kostakos, V. (2017). The experience sampling method on mobile devices. *ACM Computing Surveys*, *50*, 1–40. https://doi.org/10.1145/3123988