Supplementary Analyses

This analysis adds details to Exact Discrete Model (EDM) estimation and to the results presented in the main manuscript. Because naïve application of the proposed nonlinear constraints for EDM model (Oud & Delsing, 2010) resulted in variable parameter estimates, the estimation issues were first resolved via a short simulation study that is presented below. Then, additional details, such as individual parameter estimates, are provided in the subsequent Real-Data Results section.

Simulation Study

In the simulation study, 100 data sets were created by simulating a continuous-time stochastic process that underlies the EDM (Appendix); the simulation method was Euler-Maryama method, which corresponds to more accurate Milstein method in this case (Higham, 2001). Simulated data sets were equal with our real data in number of observations (n = 1740), and the sampling resolution was always lowered to correspond the time intervals in the real data (4, 7, and 4 years) after a time-series was simulated and before the model estimation. Estimation difficulties were most prominent regarding the drift matrix A in the model, as it can have multiple solutions leading to same (exact) discrete time representation when time-resolution of data sampling is low (Oud & Jansen, 2000). Only results for the four parameters in this two-by-two matrix are therefore shown, the rest of the results being available upon request.

First, it was clear in simulations that significant number of estimation attempts did not lead to satisfactory convergence status for OpenMx's nonlinear solver (Boker et al., 2011). Median over the 100 solutions did not lead to accurate estimate of the true underlying matrix *A* that drove the simulated data (Table S1, naïve estimate). When the estimation was repeated on a bootstrapped data set each time the OpenMx solver returned other status than 0 (converged) or 1 (status ”green”, i.e., no further improvement in objective function achievable), significantly more accurate median estimate was obtained (Table S1, third column). Despite accurate median estimate, variability of individual estimates showed a non-satisfactory patterning.

Figure S1 shows scatter plots of the ‘status-green’ estimates for the 100 simulated data sets. Clearly, there are heavy dependencies (trade-offs) between the four parameter estimates for *A* matrix. Furthermore, there were indications of clustering, which may reveal existence of multiple solutions for continuous time drift matrix *A* that led to same discrete-time cross-lagged- and auto-effects (Oud & Jansen, 2000). In order to obtain interpretable parameter estimates some extra condition or assumption that precludes some of the solutions were needed. One such assumption is that absolute values of the cross-lagged effects do not exceed one suggesting that strong causality was implausible. Furthermore, the cross-lagged instantaneous effects that exceeded absolute-value one are often associated with oscillating solutions where the *signs* of discrete cross-lagged effects vary with respect to measurement-time interval.

Due to above concerns, the absolute values of the cross-terms in drift matrix, A_{21} and A_{12}, were constrained to be less than or equal to 0.99. Figure S2 shows that while few estimates still stray far from true values of auto-effects, the heavy dependencies observed in Figure S1 are eliminated by the added constraint. A zoom-in to majority of estimates with absolute auto-effects less than 2 (Figure S3) confirmed that estimate variance now appeared more stochastic than deterministic (as desired for a valid re-sampling strategy). Medians of the estimates with (at least) green status for the nonlinear solver output and with cross-effects constrained were very close to true values of the parameters in simulation (Table S1, last column). Hence, a re-sampling approach satisfying these conditions appeared as best solution for real-data analyses.

Real-Data Results

This section represents individual parameter estimates for the EDM models of the main manuscript. The above simulation study showed that some estimates can deviate much from true values, and that median of several estimates is a reliable estimate for the true parameter values. Therefore, we applied a re-sampling strategy where medians of the parameter estimates for 1000 bootstrap data sets were interpreted. Bootstrap data sets are drawn with replacement from the original data, and their distribution approximates the true data distribution as the sample size grows (Efron & Tibshirani, 1993). Table S2 displays the estimates for individual parameters when modeling the co-evolution of depressive symptoms and negative affectivity and when modeling the co-evolution of depressive symptoms and sociability, respectively. Also 95% Bootstrap-Percentile confidence intervals are shown in parentheses, but these are not very accurate confidence-interval estimates (Efron & Tibshirani, 1993) and they incorporate solver/estimation errors as well as the sampling variability. In addition, simulations indicated that median of estimates across the re-samples should be close to the true parameter values, even though minority of estimates are far from truth. Hence, the confidence-interval estimates should be interpreted merely as suggestive values.

Although the estimation procedure produced enough large deviations to render confidence intervals of some parameters very wide, there are also other than simulation-based reasons to believe that the median estimates are, in fact, quite accurate. First, the estimated initial-state variance for depressive symptoms (0.334) and negative affectivity (0.349) and their covariance (0.230) were very close to standard sample estimates (0.335, 0.350, and 0.231, respectively); sample variance of sociability at the initial follow-up was 0.486 where the EDM estimate was 0.483, and sample covariance of sociability and depressive symptoms was -0.058 where the EDM estimate was -0.059. Second, parameters that should be independent of negative affectivity and sociability (e.g., G_{11}) were almost the same in both models. Third, the parameters were reasonable in general. For example, the solutions were stable instead exponentially growing (two distinct negative eigenvalues for matrix *A*), and they did not involve oscillating drift (complex eigenvalues for matrix *A*). Hence, we are reasonably confident that the median estimates are close to truth in what comes to the described EDM representation of the observed data.

Supplementary Appendix: R-language Function for SDE Simulation

```
library(MASS)

SDEser <- function(Tim=15,N=1500,R=10) {
  A=matrix(c(-0.8,0,0.8,-1.1),2,2)
  G=matrix(c(0.5,0,0,0.3),2,2)
  b=matrix(c(0.1,0.1),2,1)
  k=as.matrix(mvrnorm(n=1,c(0,0),matrix(c(0.1,0.05,0.05,0.1),2,2)))
  dt <- Tim/(R*N)
  dW <- sqrt(dt)*rbind(rnorm(R*N),rnorm(R*N))
  Dt <- R*dt
  Xt <- rbind(rep(0,N),rep(0,N));
  Xt[,1] <- as.matrix(mvrnorm(n=1,c(0.01,-0.01),matrix(c(1,0.2,0.2,1),2,2)))
  for (i in 1:(N-1)) {
    Winc <- rowSums(dW[,(R*(i-1)+1):(R*i)])
    Xt[,i+1] <- Xt[,i] + (A%*%Xt[,i]+b+k)*Dt + G%*%Winc
```

```
  }
  Xt
}
```

References

Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, Spies J, Estabrook R, Kenny S, Bates T, Mehta P, Fox J (2011). OpenMx: an open source extended structural equation modeling framework. *Psychometrika 76*, 306–317.

Efron B, Tibshirani RJ. (1993). *An Introduction to the Bootstrap*. Chapman & Hall/CRC: Boca Raton.

Higham, DJ (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, *43*, 525–546.

Oud JHL, Delsing MJMH (2010). Continuous time modeling of panel data by means of SEM. In *Longitudinal Research with Latent Variables* (Eds. K. Montfort, J. Oud and A. Satorra), pp. 201–244. Springer-Verlag : Berlin.

Oud JHL, Jansen RARG (2000). Continuous time state space modeling of panel data by means of SEM. *Psychometrika 65*, 199–215.

Supplementary Tables

Table S1. Parameter Estimates in Simulated Data Sets

Parameter interpretation	parameter	True value	Naïve estimate	Status Green	Green & bounded
Variable 1 auto-effect	A_{11}	-0.80	-0.74	-0.79	-0.80
Effect of variable 1 on variable 2	A_{21}	0.00	0.46	-0.02	0.01
Effect of variable 2 on variable 1	A_{12}	0.80	0.80	0.87	0.79
Variable 2 auto-effect	A_{22}	-1.10	-1.08	-1.12	-1.13

Table S2. All Parameter Estimates in Observed Data

Parameter interpretation	parameter	Negative Aff. (NA)	Sociability (S)
Depressive symptoms auto-effect	A_{11}	-0.635 (-1.551, -0.461)	-0.380 (-35.53, -0.30)
Effect of depressive symptoms on NA/S	A_{21}	0.349 (0.191, 0.990)	-0.101 (-0.990, -0.042)
Effect of the trait on depressive symptoms	A_{12}	0.441 (0.265, 0.990)	-0.108 (-0.990, -0.064)
NA/S auto-effect	A_{22}	-0.725 (-2.594, -0.483)	-0.408 (-154.79, -0.326)
Population average-related parameter	b_1	-0.001 (-0.020, 0.019)	0.000 (-0.135, 0.111)
Population average-related parameter	b_2	-0.000 (-0.035, 0.024)	0.000 (-0.050, 0.764)
Initial-state variance of depressive symptoms	Φ_{x11}	0.334 (0.300, 0.369)	0.334 (0.300, 0.371)
Initial-state covariance	Φ_{x21}	0.230 (0.203, 0.259)	-0.059 (-0.088, -0.031)
Initial-state variance of the NA/S	Φ_{x22}	0.349 (0.317, 0.384)	0.483 (0.430, 0.529)
Trait variance of depressive symptoms	Φ_{k11}	0.047 (0.023, 0.273)	0.031 (1.762, 275.272)
Trait-covariance	Φ_{k21}	-0.022 (-0.111, -0.009)	0.013 (-1212.655, 1.070)
Trait variance of NA/S	Φ_{k22}	0.057 (0.025, 0.848)	0.048 (0.027, 308.539)
Cross-covariance of x1 (initial-state1st coordinate) and k1	$\Phi_{x1,k1}$	0.048 (0.018, 0.134)	0.065 (0.048, 8.340)
Cross-covariance of x1 and k2	$\Phi_{x1,k2}$	0.047 (0.024, 0.219)	0.010 (-1.049, 0.077)
Cross-covariance of x2 and k1	$\Phi_{x2,k1}$	-0.002 (-0.036, 0.031)	0.011 (-0.582, 0.036)
Cross-covariance of x2 and k2	$\Phi_{x2,k2}$	0.093 (0.060, 0.388)	0.098 (0.074, 41.805)
Depressive symptoms diffusion coefficient	G_{11}	0.382 (0.331, 0.581)	0.349 (0.316, 3.298)
NA/S diffusion coefficient	G_{22}	0.354 (0.307, 0.673)	0.380 (0.350, 7.356)

Note: Parentheses show gross/suggestive 95% confidence-interval estimates. In addition to above parameters, also initial-state mean (often denoted μ_x) was estimated but is practically zero in both EDM estimates and observed data, because the data were residuals from the adjusting regression analysis.

Supplementary Figures

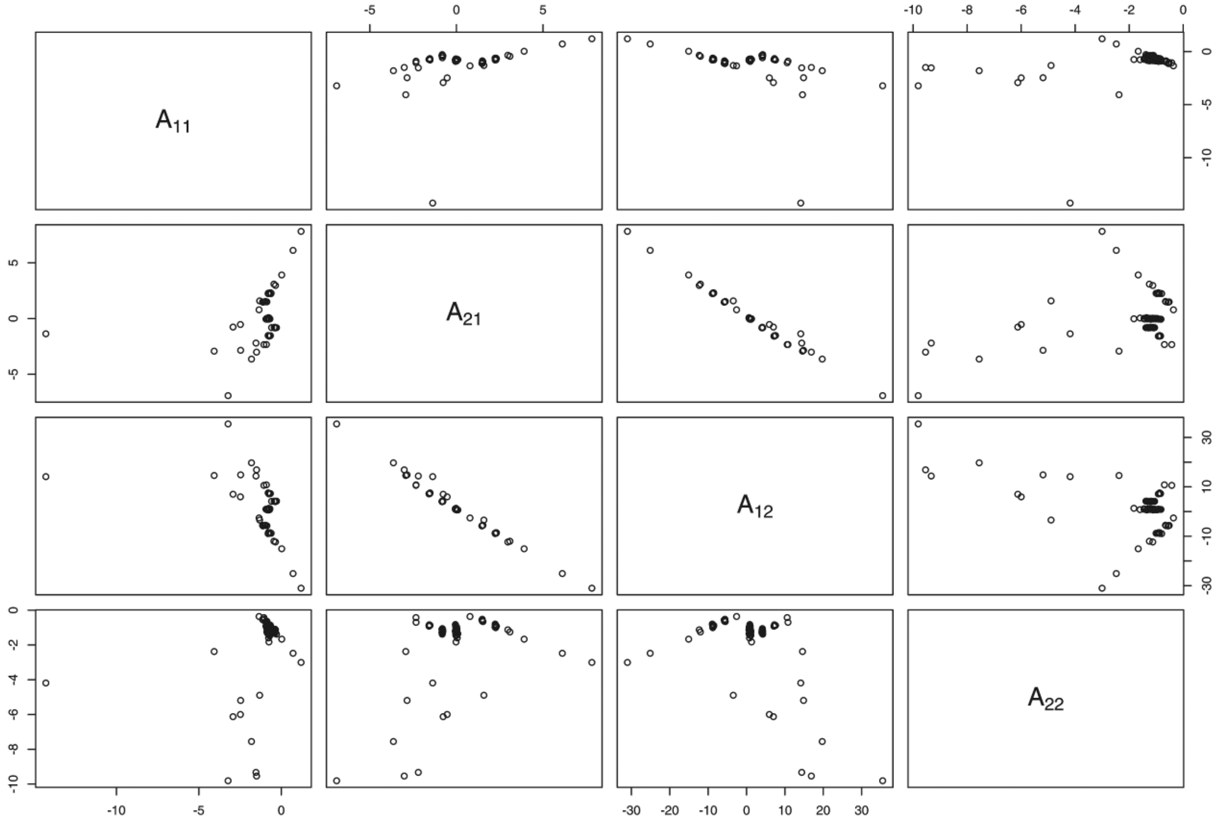

Figure S1. Parameter Estimates with Satisfactory Solver Status in Simulated Data Sets.

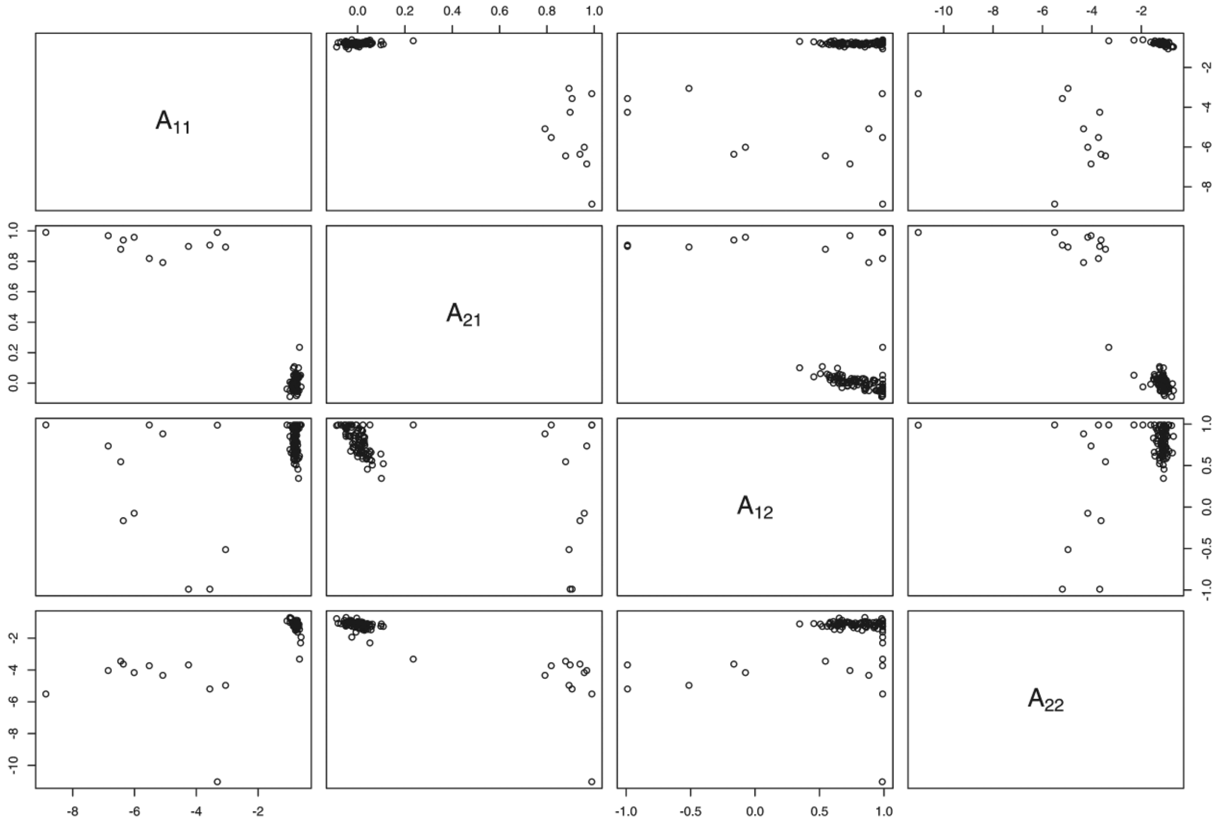

Figure S2. Parameter Estimates with Satisfactory Solver Status in Simulated Data Sets when $|A_{21}| \leq 0.99$ *and* $|A_{12}| \leq 0.99$.

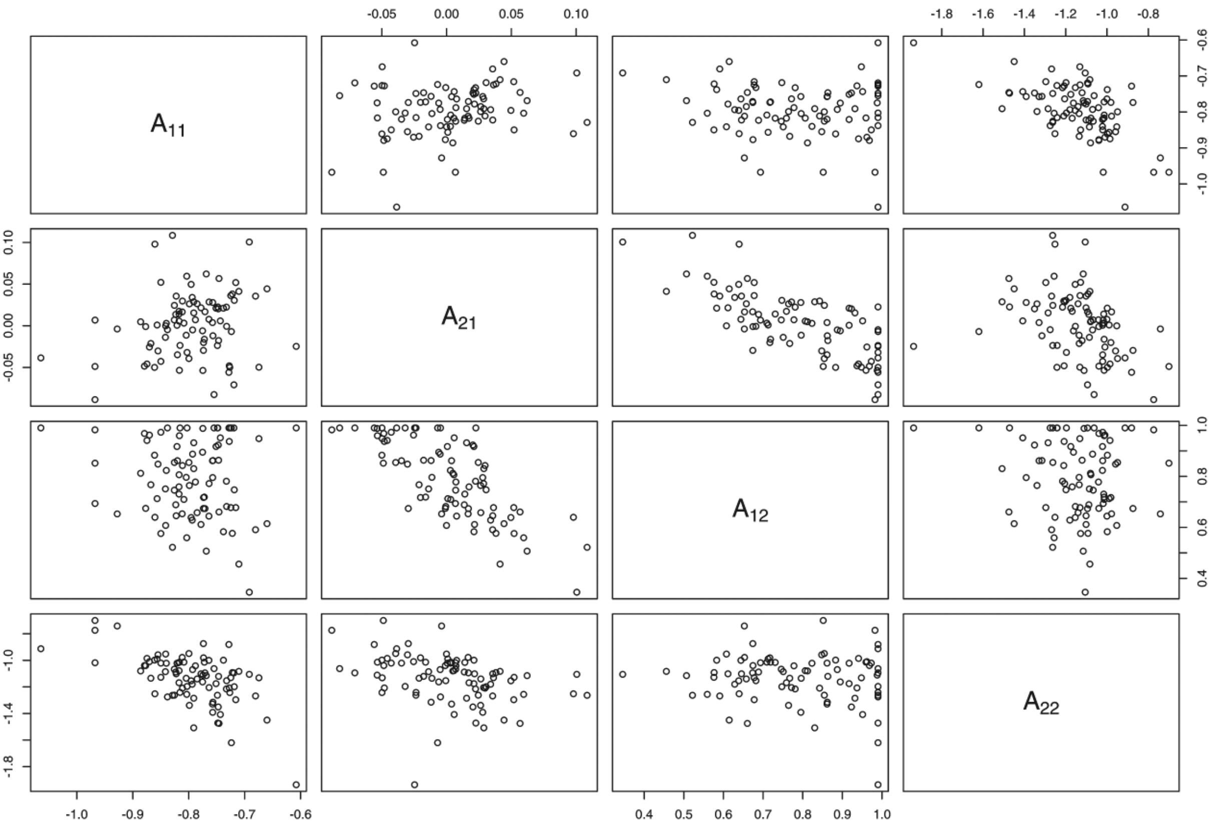

Figure S3. Zoom-in of the Figure S2 into the Region where $|A_{11}| \leq 2$ *and* $|A_{22}| \leq 2$.