Hierarchical Bayesian Continuous Time Dynamic Modeling

Charles C Driver
Max Planck Institute for Human Development

Manuel C Voelkle
Humboldt University Berlin
Max Planck Institute for Human Development

Continuous time dynamic models are similar to popular discrete time models such as autoregressive cross-lagged models, but through use of stochastic differential equations can accurately account for differences in time intervals between measurements, and more parsimoniously specify complex dynamics. As such they offer powerful and flexible approaches to understand ongoing psychological processes and interventions, and allow for measurements to be taken a variable number of times, and at irregular intervals. However, limited developments have taken place regarding the use of continuous time models in a fully hierarchical context, in which all model parameters are allowed to vary over individuals. This has meant that questions regarding individual differences in parameters have had to rely on single-subject time series approaches, which require far more measurement occasions per individual. We present a hierarchical Bayesian approach to estimating continuous time dynamic models, allowing for individual variation in all model parameters. We also describe an extension to the ctsem package for R, which interfaces to the Stan software and allows simple specification and fitting of such models. To demonstrate the approach, we use a subsample from the German socio-economic panel and relate overall life satisfaction and satisfaction with health.

Keywords: Continuous time, dynamic model, state space, non-linear mixed-effects, stochastic differential equation, hierarchical time series

In this work we bring continuous time dynamic modeling together with hierarchical Bayesian modeling, and describe the resultant model and software for estimation. In this approach, the estimation of subject specific continuous time dynamic model parameters is enhanced by using data from other subjects to inform a prior distribution, with no restrictions on the length of time series or number of subjects. This allows for questions regarding individual differences in dynamics, without the requirement of very large numbers of repeated observations that single-subject time series approaches can have, nor the convergence problems with low numbers of subjects that some random-effects approaches have (Eager & Roy, 2017).

Both hierarchical Bayes and continuous time dynamic modeling offer an improved use of data over common alternatives. The hierarchical approach uses information from other subjects to better estimate parameters of each subjects' specific model, while the continuous time approach uses information about the time interval between measurements of each subject. The nature of this improvement is such that it also allows for new questions on individual differences in the continuous time model parameters. That is, in datasets that do not contain sufficient information for single-subject time series approaches – as for instance with the German socioeconomic panel (GSOEP) – we can now estimate subject-specific continuous time dynamic models, by incorporating information from other subjects. As an introduction we first review some of the background and motivation for dynamic models in psychology, consider separately hierarchical modeling and continuous time models, and then consider the combination of both approaches. We then describe the full model more formally, discuss the software implementation we have developed, demonstrate performance of the approach and software with a simulation study, and finish with an example application of a dynamic model of wellbeing based on the GSOEP data. Although the article aims at creating a general understanding of the model, readers who are primarily interested in its application, may easily skip over some of the more detailed description of the subject likelihood and population distribution in the model section.

Charles C Driver, Centre for Lifespan Psychology, Max Planck Institute for Human Development. Manuel C Voelkle, Psychological Research Methods, Humboldt University, Berlin, and Centre for Lifespan Psychology, Max Planck Institute for Human Development. This work also constitutes part of the first authors doctoral thesis, and has been distributed in pre-print form. Correspondence concerning this article should be addressed to Charles C Driver, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. E-mail: driver@mpib-berlin.mpg.de

Dynamic models

Models containing dependencies on earlier states and stochastic inputs, such as the autoregressive and cross lagged model, offer a flexible and powerful approach to examine how psychological constructs function over time within a subject. For the sake of expediency we will refer to such models as *dynamic models*, but note that such a term could in general encompass a broader range of models than dealt with here. Similar to trajectory oriented approaches such as latent growth curves, dynamic models generally involve some smooth long term trajectory. Instead of treating deviations from a smooth overall trajectory as mere uninformative noise, however, such deviations can often be informative as to future states. So, while someone's life satisfaction last year doesn't completely determine their current life satisfaction, it does have some predictive value, over and above their typical level for the last 30 years. If we were to develop a formal model out of such a statement, we would have a simple dynamic model involving some dependence on past states, and a stable baseline state. Since such a model cannot perfectly predict the way satisfaction may change in time, we will also need a variance term to quantify the degree of uncertainty regarding predictions – if satisfaction last year was substantially below baseline, we may expect that satisfaction is still at a level below baseline, but with some uncertainty. The more time that passes before we check satisfaction again, the greater the uncertainty, as more external events and internal processes that we cannot fully predict will occur. As such, the more time that passes, the less informative our previous observation will be, until at some point, the previous observation no longer helps our prediction.

The model described so far assumes we have a perfect measure of life satisfaction, but this is highly unlikely – regardless of exactly how we measure, the same internal satisfaction state could result in a range of measured outcomes due to spurious factors that are unrelated to satisfaction. When such measurement errors are left in the dynamic model, any parameter estimates regarding the dynamics of the true process we are interested in are likely to be biased (Schuurman, Houtveen & Hamaker, 2015). Consequently, *state-space* models (Hamilton, 1986) have been developed, which partition observed deviations from a prediction into innovation variance, representing meaningful but unpredictable fluctuations in the process itself, and measurement error, representing uncorrelated random noise in the observations. The innovation variance may be considered meaningful, in that although it represents unpredictable changes in the process, once such changes in the processes are observed, they are useful for future predictions of the states of the processes. This is in contrast to measurement error, which will not offer any predictive value for the future. By fitting such a dynamic model to the data, we can ask questions like 'how long does a change in this persons' life satisfaction typically persist?'. If we include a second set of observations, this time regarding their health, we can then ask 'to what extent do changes in life satisfaction and health covary?', and also consider temporal dynamics between the two, with 'to what extent does a change in health predict a later change in life satisfaction, after controlling for shared sources of change?'. The major distinction between such an approach and a trajectory only model such as a latent growth curve, is that the relation between fluctuations is addressed, rather than only the shape of overall change. So while it may be that over 30 years, a person exhibits some general decline in health and apparently stable life satisfaction, when the relation between fluctuations is taken into account, one might instead see that life satisfaction tends to fluctuate downwards when health satisfaction does, then slowly recover due to other factors.

Common examples of dynamic models include the autoregressive and cross-lagged panel model, (Finkel, 1995; Hertzog & Nesselroade, 2003) autoregressive moving average models (Box, Jenkins, Reinsel & Ljung, 2015), or the latent change score formulations sometimes used (John J. McArdle, 2009), when they include innovation variance for each time point. They can be applied with observational or experimental data, or a mix of the two, and can be used to tell us to what extent changes in a psychological process are predictive of later changes of either the same process, or some other process of interest. Some examples of the usage of dynamic models within psychology include the analysis of symptom networks in psychopathology (Bringmann et al., 2013), differences in individuals' affective dynamics (Koval, Sütterlin & Kuppens, 2016), and the influence of partners on one another (Hamaker, Zhang & van der Maas, 2009).

Hierarchical Dynamic models

While in some circumstances, a well developed dynamic model for a single subject may be important, in many situations scientists are instead interested in generalising from observed subjects to some population, as well as understanding differences between subjects. For such purposes, one typically needs repeated measures data from multiple subjects. How then, should data from different individuals be combined? Were all subjects exactly alike, combining the information would be very simple and effective. Were the subjects entirely distinct, with no similarities, combining the information would tell us nothing – an estimate of the average weight of a type of leaf, is not improved by including a rock in the sample. The estimation approaches herein endeavour to tread the middle ground, in that while differences between subjects are acknowledged, similarities are leveraged to improve estimation. Such an approach can broadly be called hierarchical dynamic modelling, a term encompassing both frequentist and Bayesian approaches. To understand how hierarchical dynamic models function, and the benefits they offer, it is helpful to first consider two extremes of possible

approaches to dynamic models with multiple subjects. At one end of the continuum are panel models with a single set of fixed-effects parameters governing the dynamics of every subject – 'all leaves weigh exactly the same' – and at the other lies person-specific time series – 'one leaf is to another, as to a rock'.

Panel models containing autoregressive and cross-lagged parameters are regularly estimated with a single set of fixed-effects parameters governing the behavior of the dynamic system for many subjects – one assumes that the system characterizing the psychological processes of one subject is exactly the same as for other subjects[1]. The benefits of this assumption are that the number of parameters in the model is relatively low, and the data from every individual is relevant for estimating every parameter. This assumption usually makes fitting the model to data much simpler, and can increase the precision and accuracy of parameter estimates when the assumption is valid. However, it is a very strong assumption and can result in highly biased parameter estimates. This has long been recognized (i.e., Balestra & Nerlove, 1966) to be the case in regards to the intercept parameter of the dynamic model, in that when subject specific differences in the average level of the processes are not accounted for, the temporal dynamics parameters (auto and cross effects) can be terribly biased – instead of representing some average of the temporal dynamics for all subjects, they instead become a mixture of the within-subject temporal dynamics and the between-subjects differences in average levels. While many poor implementations and claims of large cross effects and causality can still be found, this issue is at least widely recognized and readily resolved. See Halaby (2004) and Hamaker, Kuiper and Grasman (2015) for more details, and Cattell (1963), Molenaar (2004), Voelkle, Brose, Schmiedek and Lindenberger (2014), Wang and Maxwell (2015) for more on between and within subject issues in general.

At the other end of the spectrum is idiographic, or individual specific, time series approaches, where one fits the model separately for each subject (see for example Steele, Ferrer & Nesselroade, 2014). Such approaches ensure that the estimated parameters are directly applicable to a specific individual. With 'sufficiently large' amounts of data for each subject, this is likely to be the simplest and best approach. However, in applications of dynamic models there may be many correlated parameters, coupled with noisy measurements that are correlated in time, ensuring that a 'sufficiently large' amount of data per subject may in fact be 'unattainably large', particularly when one wishes to distinguish measurement error and relate multiple processes. Models estimated with less than ideal amounts of data tend to suffer from finite sample biases, as for example in the autoregressive parameter when the number of time points is low (Marriott & Pope, 1954, 3/4), and also from higher uncertainty and more inaccurate point estimates. Were parameters independent of one another then inaccurate estimates of a parameter of little interest may be tolerable, but in dynamic models there are typically strong dependencies between parameters, such that inaccurate estimation of any single parameter can also reduce accuracy for all other parameters. An example of this dependency is that when, as typically occurs, the parameter governing the subjects' base level is overfit and explains more observed variance than it should, the related auto-effect parameter is typically underestimated, in turn affecting many other parameters in the model. While some may be inclined to view the complete independence between models for each subject as a strength of the single-subject time series approach, there is little value to such independence if it also comes at the cost of making the best model for the subjects empirically unidentifiable, or the estimates substantially biased. Recent work by Liu (2017) demonstrates some of these issues. Liu compared a multilevel and person-specific approach to modelling multiple-subjects autoregressive time series, without measurement error. They examined the effect of time series length, number of subjects, distribution of coefficients, and model heterogeneity (differing order models). So long as the model converges and is not under specified for any subjects (i.e., contains lags of a high enough order), they find that the multilevel approach provides better estimates, even when distributional assumptions regarding the parameters are not met. Of course, not every possibility has been tested, and it is likely possible to find specific cases where this result does not hold.

Hierarchical approaches to dynamic models, such as those from Lodewyckx, Tuerlinckx, Kuppens, Allen and Sheeber (2011), are essentially a generalization which encompasses the two extremes already discussed – a single model for all individuals, or a distinct model for each. Such a hierarchical formulation could take shape either in a frequentist random-effects setting, or a hierarchical Bayesian setting.

Using a hierarchical formulation, instead of estimating either a single set of parameters or multiple sets of independent parameters, one can estimate *population distributions* for model parameters. It is common in hierarchical frequentist approaches to only estimate population distributions, while with Bayesian approaches it is more typical to simultaneously sample subject level parameters *from* the population distribution, with the population distribution essentially serving as a prior distribution for the subject level parameters. The latter approach can help to provide the intuition for the relation among person-specific, hierarchical, and single fixed-effect parameter set approaches. Using a hierarchical model, if one were to fix the variance of the population distribution to different values, one could see that as the

[1]Note that this is distinct from what has become known in econometrics as the 'fixed-effects panel model', which estimates unique intercept terms for every subject, with all other parameters constant across subjects (Halaby, 2004).

variance approaches zero, the subject level parameter estimates get closer and closer to the fixed-effects model with the same set of parameter values for every subject. Conversely, as one fixed the population variance further and further towards infinity, subject level estimates would get closer and closer to those of the person-specific approach, in which the parameter estimates for one subject are not influenced at all by estimates of the others (An R script demonstrating this may be found in the supplementary material). The benefit of a hierarchical approach however, is that rather than fixing the parameters of the population distribution in advance, the population distribution mean and variance can be estimated at the same time as subject level parameters – the extent of similarity across subjects is estimated, rather than fixed in advance to 'not at all similar' or 'completely the same'. An intuitive interpretation of this is that information from all other subjects serves as prior information to help parameter estimation for each specific subject. As is standard with Bayesian methods, one still specifies some prior distributions in advance, but in hierarchical Bayes models these are priors regarding our expectations for the *population distribution*, otherwise known as hyperpriors. So, some initial, possibly very vague, hyperpriors are specified regarding possible population distributions. These hyperpriors, coupled with data and a Bayesian inference procedure, result in an estimated population distribution for the parameters of each subjects' dynamic model. This estimated population distribution, coupled with the likelihood of the dynamic model parameters for a specific subject (calculated just as we would in the single-subject approach), gives the posterior distribution for the subjects' specific parameters.

In the frequentist setting, while it is relatively straightforward to include random-effects for intercept parameters, random-effects for regression and variance parameters of latent factors have typically been more problematic, due to the complex integration required (see for instance Delattre, Genon-Catalot & Samson, 2013; Leander, Almquist, Ahlström, Gabrielsson & Jirstrand, 2015). This *mixed-effects* approach is helpful in that stable differences in level between subjects – likely the largest source of bias due to unobserved heterogeneity – are accounted for, while maintaining a model that can be fit reasonably quickly using the well and commonly understood frequentist inference architecture. Although we show later that when individual differences in latent regression and variance parameters are (incorrectly) not modeled, the magnitude of spurious cross effects is low, many spurious results do nevertheless occur. Further, if not modelled it is of course impossible to ask questions regarding such individual differences, which are a key interest in many cases. Bayesian approaches offer good scope for random-effects over all parameters, and indeed hierarchical Bayesian discrete time dynamic models are implemented and in use, see for instance Schuurman (2016), Schuurman, Ferrer, de Boer-Sonnenschein and Hamaker (2016).

Once a hierarchical dynamic model is estimated, one may be interested in questions regarding the population distribution, one or multiple specific subjects, or one or multiple specific time points of a subject. Questions regarding the population distribution could relate to: means and variances of parameter distributions, such as "what is the average cross effect of health on life satisfaction", and "how much variance is there in the effect across people?" Also possible are questions regarding correlations between population distributions of different parameters, or between parameters and covariates, such as, "do those who typically have worse health show a stronger cross-effect?" Instead of whole population questions, one could instead ask about, for example, the cross effect of health on life satisfaction of a specific subject. Or yet more specific, we might for instance want to predict the health of a subject at some particular time point in the future, given a particular life event and set of circumstances.

Continuous time dynamic models

Within psychology and many fields, classical approaches to dynamic modeling have tended to rely on *discrete time* models, which generally rely on an assumption that the time intervals between measurements are all the same. A discrete time model directly estimates the relation (regression strength) between one measurement occasion and another, without incorporating the time interval information. If all measurement occasions have the same time interval between them, this can be fine. In many cases though the time intervals between occasions may differ, and if the same regression parameter is used to account for different time intervals, it is likely that the parameter is incorrect for some or all occasions – the relation between someone's earlier mood and later mood is likely to be quite different if we compare a ten minute interval to a three day interval, for instance. Obviously, parameter estimates and, thus, scientific conclusions, are biased when observation intervals vary and this is not adequately accounted for. In simple cases, so called phantom variables (Rindskopf, 1984), with missing values for all individuals could be added in order to artificially create equally spaced time intervals. For complex patterns of individually varying time intervals, however, this approach can become untenable (Voelkle & Oud, 2013).

Continuous time models overcome these problems, offering researchers the possibility to estimate parameters free from bias due to unequal intervals, easily compare between studies and datasets with different observation schedules, gather data with variable time intervals between observations, understand changes in observed effects over time, and parsimoniously specify complex dynamics.

Rather than estimate the regression between two measurement occasions directly, the continuous time approach instead estimates a set of *continuous time parameters* that de-

scribe how the process changes at any particular moment, based on the current state of the process. A mathematical function can then be used, in combination with the time interval, to determine the appropriate *discrete time parameters*, such as regression strength, governing the relation between two specific occasions separated by some time interval. This is a non-linear function, that operates in a way such that for first order (dependent only on the previous occasion, and not earlier occasions) discrete-time models with constant time intervals, the discrete and continuous time approaches are equivalent. The same of course holds true for other dynamic model parameters, such as the intercepts and innovation covariance matrix. The ability to naturally and exactly account for different time intervals means continuous time models are particularly suited to the analysis of data from studies with different measurement schemes.

While accounting for different time intervals naturally is one obvious benefit, another reason for interest in continuous time models is that the parameters directly represent elements of differential equations. This is beneficial both in terms of allowing more meaningful interpretation of the parameters, as well as for more readily specifying higher order dynamics such as damped oscillations. More meaningful interpretation of continuous time dynamic models is possible because they are not just models for the specific time points observed as with discrete time models, but rather they describe expected behavior of the processes at *any* time point, whether observed or not. Differential equation models describing more complex (e.g., higher than first order) dynamics have proven of some interest in psychology, with analysis of constructs such as working memory (Gasimova et al., 2014), emotion, (Chow, Ram, Boker, Fujita & Clore, 2005), and addiction (Grasman, Grasman & van der Maas, 2016). For some background on how differential equation models can be applied to psychological constructs, see Deboeck, Nicholson, Bergeman and Preacher (2013) and Boker (2001). Important to note here is that continuous time dynamic models are not the only form of dynamic model using differential equations, as approaches such as generalized local linear approximation (Boker, Deboeck, Edler & Keel, 2010) are also available. Rather, the terminology 'continuous time model' has typically been used to refer to dynamic models that explicitly incorporate the time interval in the equations, leading to an 'exact' solution, that does not require the specification of an embedding dimension in advance. For discussion on the distinction between 'approximate' and 'exact' approaches to estimation of differential equations, see Oud and Folmer (2011) and Steele and Ferrer (2011).

Even for models that explicitly incorporate the time interval, exact, analytic solutions to the stochastic differential equations are usually only available for *linear* stochastic differential equations involving a Gaussian noise term, which are the form of model we are concerned with here. For more complex forms of continuous time model, various approximations are typically necessary, as for instance in Särkkä, Hartikainen, Mbalawata and Haario (2013). For further discussion of the continuous time dynamic model in psychology, Oud (2002) and Voelkle, Oud, Davidov and Schmidt (2012) describe the basic model and applications to panel data, and Voelkle and Oud (2013) detail higher order modelling applications such as the damped linear oscillator. A classic reference for stochastic differential equations more generally is the text from Gardiner (1985).

Hierarchical continuous time dynamic models

While Oud and Jansen (2000) describe a mixed-effects, structural equation modelling approach to continuous time dynamic models, and Driver, Oud and Voelkle (2017) the related software, relatively little work has been done on the generalization to fully random-effects models, in which all parameters may vary over subjects. The continuous time dynamic model has been specified in Bayesian form by Boulton (2014), but this did not extend to a hierarchical approach. The most substantial foray into hierarchical continuous time approaches in psychology has been with the work by Oravecz, Tuerlinckx and Vandekerckhove (2009, 2016) on the hierarchical Ornstein-Uhlenbeck model, which also led to the creation of the BHOUM software for estimation. While the software and modelling approach as it stands is useful, there is a limitation for some applications in that the matrix containing dynamic effects is constrained to be symmetric and positive definite. This means that model structures that require non-symmetric cross effects, such as the autoregressive and cross lagged panel model or damped linear oscillator, cannot be estimated.

Chow, Lu, Sherwood and Zhu (2014) have also looked at hierarchical dynamics in continuous time, fitting nonlinear ordinary differential equation models with random effects on the parameters to ambulatory cardiovascular data from multiple subjects. While the nonlinear aspect offers increased flexibility (at cost of additional complexity), the ordinary differential aspect, rather than stochastic differential, means that the estimated processes are deterministic, and randomness is assumed to occur only at the measurement level. Unless the process is very well understood, with all contributing factors measured, the lack of innovation variance may be quite limiting. Lu, Chow, Sherwood and Zhu (2015) used similar estimation algorithms and data with stochastic differential equations which can account for innovation variance, but avoided a hierarchical approach, and assumed independence between subjects parameters.

This work describes the model and software for a hierarchical Bayesian continuous time dynamic model, without any restriction of symmetric positive-definite dynamics. It builds upon the model and software, ctsem, described in Driver et al. (2017), which offered a mixed-effects modeling

framework for the same linear stochastic differential equation model, in which intercept and mean related parameters could be estimated as random effects across subjects, but subject level variance and regression parameters were restricted to fixed effects estimation. With the change to a hierarchical Bayesian framework, all subject level parameters may now vary across subjects, the covariance between subject level parameters is available, and the estimation of time-invariant relations between covariates and all subject level parameters is now possible (as opposed to relations with only the mean levels of subjects' processes).

The model

There are three main elements to our hierarchical continuous time dynamic model. There is a subject level latent dynamic model, a subject level measurement model, and a population level model that describes the distribution of subject level parameters across subjects. Note that while various elements in the model depend on time, the fundamental parameters of the model are time-invariant. Note also that we ignore subject specific subscripts when discussing the subject level model.

Subject level latent dynamic model

The subject level dynamics are described by the stochastic differential equation:

$$d\boldsymbol{\eta}(t) = \Big(\mathbf{A}\boldsymbol{\eta}(t) + \mathbf{b} + \mathbf{M}\boldsymbol{\chi}(t)\Big)dt + \mathbf{G}d\mathbf{W}(t) \tag{1}$$

Vector $\boldsymbol{\eta}(t) \in \mathbb{R}^v$ represents the state of the v latent processes at time t, so $d\boldsymbol{\eta}(t)$ simply means the direction of change, or gradient, of the latent processes at time t. The matrix $\mathbf{A} \in \mathbb{R}^{v\times v}$ is often referred to as the drift matrix, with auto effects on the diagonal and cross effects on the off-diagonals characterizing the temporal dynamics of the processes. Negative values on the auto effects are typical, and imply that as the latent state becomes more positive, a stronger negative influence on the expected change in the process occurs – the process tends to revert to a baseline (at least in the absence of other influences). Positive auto-effects imply an explosive process, in which deviations from a baseline do not dissipate, but rather accelerate. Cross-effects between processes may be positive or negative, with a positive cross-effect in the first row and second column implying that as the second process becomes more positive, the direction of change in the first process also becomes more positive, while for a negative cross effect the first process is negatively influenced.

The expected auto and cross regression matrix for a given interval of time (i.e., a discrete time effect) can be calculated as per Equation 2. Figure 1 shows this calculation over a range of time intervals, for an example bivariate process of sociability and energy level. The first column of the drift matrix, with values $-0.4, -0.1$, represents the effect of the current state of sociability, on change in sociability and change in energy level respectively. So, when sociability is above baseline – someone is enjoying socialising with friends – both sociability and energy level are likely to go down. The second column of the drift matrix, with values $0.3, -0.2$, represents the effect of the current state of energy level, on change in sociability and change in energy level respectively. So in this case, if energy levels are above baseline then the expected change in sociability will be positively influenced, while the expected change in energy level will be downwards. The converse of course holds true when energy levels are below baseline – sociability will reduce, and energy level will rise.

$$\mathbf{A}^*_{\Delta t_u} = e^{\mathbf{A}(t_u - t_{u-1})} \tag{2}$$

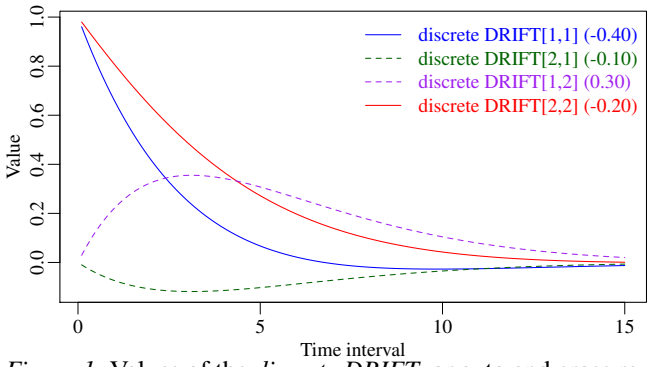

Figure 1. Values of the *discrete DRIFT*, or auto and cross regression matrix, change depending on a function of the continuous time DRIFT matrix and the time interval between observations. The continuous time DRIFT values are shown in brackets in the legend.

The * notation is used in this work to indicate a term that is the discrete time equivalent of the original continuous time parameter, for the time interval Δt_u. The time interval Δt_u is simply the time between measurement occasions u and $u-1$, where $\mathbf{U}$ is the set of all measurement occasions u, from 1 to the total number of measurement occasions. $\mathbf{A}^*_{\Delta t_u}$ then contains the appropriate auto and cross regressions for the effect of latent processes $\boldsymbol{\eta}$ at measurement occasion $u-1$ on $\boldsymbol{\eta}$ at measurement occasion u. So, for a simple univariate process with $\mathbf{A} = -.4$, a time scale of weeks, and a time interval of two weeks between measurement occasions, the discrete time autoregression would be $e^{-.4(2)} = 0.45$. Note that here the exponential of the $\mathbf{A}$ matrix is equivalent to the regular univariate exponential, but once the $\mathbf{A}$ matrix has non-zero off-diagonals, this is no longer the case and a matrix exponential function is necessary.

In Equation 1 the continuous time intercept vector $\mathbf{b} \in \mathbb{R}^v$, provides a constant fixed input to the latent processes $\boldsymbol{\eta}$. In combination with temporal dynamics matrix $\mathbf{A}$, this vector

determines the long-term level at which the processes fluctuate around. Without the continuous time intercept the processes, if mean reverting, would simply fluctuate around zero. The discrete time intercept may be calculated as per Equation 3. For the sociability and energy level example, this discrete time intercept would be a vector of two values, the first representing the value to be added to the sociability state, and the second added to the energy level state. One way to think about this is that when the baseline of the processes is not zero, the autoregressive nature of the system means some portion of the process value is lost for each time interval – the discrete time intercept simply adds a sufficient amount back in to maintain a non-zero baseline (note that this intuition is only strictly valid for stationary systems). For interpretation purposes we are inclined to think the asymptotic level of the processes, $\mathbf{b}_{\Delta t_\infty}$, is more useful. The asymptotic process level is the level the processes tend towards over time, and also the level they start at if stationary.

$$\mathbf{b}^*_{\Delta t_u} = \mathbf{A}^{-1}(\mathbf{A}^*_{\Delta t_u} - \mathbf{I})\mathbf{b} \tag{3}$$

$$\mathbf{b}_{\Delta t_\infty} = -\mathbf{A}^{-1}\mathbf{b} \tag{4}$$

In Equation 1 the time dependent predictors $\boldsymbol{\chi}(t)$ represent exogenous inputs to the system, such as an intervention, that may vary over time and are independent of fluctuations in the system. Equation 1 shows a generalized form for time dependent predictors, that could be treated a variety of ways depending on the predictors assumed shape, or time course (i.e., what values should $\boldsymbol{\chi}$ take on *between* measurement occasions?). We use a simple impulse form shown in Equation 5, in which the predictors are treated as impacting the processes only at the instant of an observation occasion u, and the effects then transmit through the system in accordance with $\mathbf{A}$ as usual. Such a form has the virtue that many alternative shapes are made possible via augmentation of the system state matrices – discussion and examples of this are available in Driver and Voelkle (2017).

$$\boldsymbol{\chi}(t) = \sum_{u \in \mathbf{U}} \mathbf{x}_u \Delta(t - t_u) \tag{5}$$

Here, time dependent predictors $\mathbf{x}_u \in \mathbb{R}^l$ are observed at measurement occasions $u \in \mathbf{U}$. The Dirac delta function $\Delta(t - t_u)$ is a generalized function that is ∞ at 0 and 0 elsewhere, yet has an integral of 1 (when 0 is in the range of integration). It is useful to model an impulse to a system, and here is scaled by the vector of time dependent predictors $\mathbf{x}_u$. The effect of these impulses on processes $\boldsymbol{\eta}(t)$ is then $\mathbf{M} \in \mathbb{R}^{v \times l}$. Put simply, the equation means that at the measurement occasion u a time dependent predictor (e.g., intervention) is observed, the system processes spike upwards or downwards by $\mathbf{M}\mathbf{x}_u$. For a typical intervention that probably only occurs once during the observation window, $\mathbf{x}_u$ would then be zero for every observation u except when the intervention occurred, where it could take on a dummy coding value such as one, or could reflect the strength of the intervention. Because $\mathbf{M}$ is conceptualized as the effect of instantaneous impulses $\mathbf{x}$, which only occur at occasions $\mathbf{U}$ and are not continuously present as for the processes $\boldsymbol{\eta}$, the discrete and continuous time forms are equivalent, at the times when observations are made. This means that the effect of some intervention at measurement occasion $u = 3$ is simply $\mathbf{M}\mathbf{x}_{u=3}$ at the instant of the intervention, and at later measurement occasion $u = 4$, the remaining effect is $\mathbf{A}^*_{\Delta t_{u=4}}\mathbf{M}\mathbf{x}_{u=3}$. If the time interval between occasions 3 and 4 is $\Delta t = 2.30$, using Equation 2 this translates to $e^{\mathbf{A}(2.30)}\mathbf{M}\mathbf{x}_{u=3}$.

In Equation 1, $\mathbf{W}(t) \in \mathbb{R}^v$ represents independent Wiener processes, with a Wiener process being a random-walk in continuous time. $d\mathbf{W}(t)$ is meaningful in the context of stochastic differential equations, and represents the stochastic noise term, an infinitesimally small increment of the Wiener process. Lower triangular matrix $\mathbf{G} \in \mathbb{R}^{v \times v}$ represents the effect of this noise on the change in $\boldsymbol{\eta}(t)$. $\mathbf{Q}$, where $\mathbf{Q} = \mathbf{G}\mathbf{G}^\top$, represents the variance-covariance matrix of this diffusion process in continuous time. Intuitively, one may think of $d\mathbf{W}(t)$ as random fluctuations, and $\mathbf{G}$ as the effect of these fluctuations on the processes. The discrete time innovation covariance matrix, which represents the increase in uncertainty about the process states over a certain time interval, may be calculated as shown in Equations 6 and 7. Figure 2 plots this calculation over a range of time intervals for the example sociability and energy level processes, with diffusion variances of 2 and 3 respectively, and covariance of -1.

$$\mathbf{Q}^*_{\Delta t_u} = \mathbf{Q}_{\Delta t_\infty} - \mathbf{A}^*_{\Delta t_u}\mathbf{Q}_{\Delta t_\infty}(\mathbf{A}^*_{\Delta t_u})^\top \tag{6}$$

$$\mathbf{Q}_{\Delta t_\infty} = \text{irow}(-(\mathbf{A} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{A})^{-1}\,\text{row}(\mathbf{Q})) \tag{7}$$

Equation 6 calculates the discrete time innovation covariance matrix for a given time interval. Intuitively, it shows that the discrete time innovation covariance matrix, which may be thought of as representing the amount and correlation of random noise added to the processes over a specified time interval, equals the asymptotic, or total, innovation covariance matrix $\mathbf{Q}_{\Delta t_\infty}$, minus the amount 'remaining' from the earlier measurement occasion. This remaining amount is determined by the temporal dynamics of $\mathbf{A}$. $\mathbf{Q}_{\Delta t_\infty}$ represents the innovation covariance matrix as Δt approaches infinity, which for a stationary process also represents the total within-subject variance covariance at any point in time. This asymptotic within-person covariance could also be thought of as the uncertainty about process states when no measurements of the processes exist. In the plotted example, we can see that uncertainty regarding the sociability and energy level processes is roughly at this asymptote after a time interval of approximately 10. This asymptotic covariance matrix (as used in Tómasson, 2013) provides a computationally

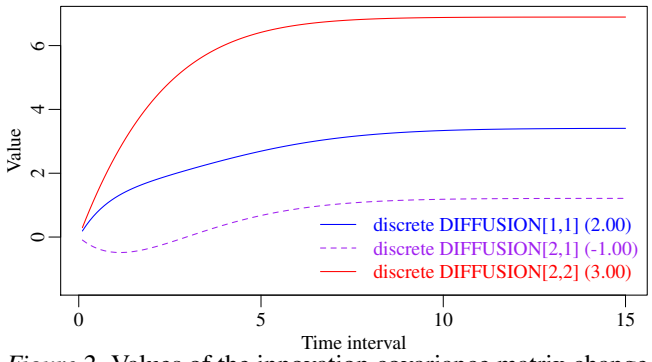

Figure 2. Values of the innovation covariance matrix change depending on a function of the continuous time DRIFT and DIFFUSION matrices and the time interval between observations. The continuous time DIFFUSION values are shown in brackets in the legend (with DRIFT the same as for Figure 1). The values at higher time intervals are no longer changing substantially, thus representing the total within-person covariance matrix (or very close to it).

efficient basis for calculating the additional covariance $\mathbf{Q}^*_{\Delta t_u}$ added to the system over the time interval Δt_u, as the asymptotic matrix only needs to be computed once. $\otimes$ denotes the Kronecker-product, row is an operation that takes elements of a matrix row wise and puts them in a column vector, and irow is the inverse of the row operation.

As the various discrete time calculations formula shown in this section rely on the matrix exponential seen in Equation 2, they can be difficult to intuitively understand. Plotting them with specific parameter values and increasing time interval Δt, can be a valuable tool, as this demonstrates how the implied discrete-time coefficients change as a function of the continuous time parameters and the time interval. An R script to perform such plots for a bivariate latent process model is available in the supplementary material.

Latent dynamic model — discrete time solution

To derive expectations for discretely sampled data, the stochastic differential Equation 1 may be solved and translated to a discrete time representation, for any observation $u \in \mathbf{U}$. Most components for this solution were already shown in Equations 2, 3, 5, 6, and 7, here we simply bring them together.

$$\boldsymbol{\eta}_u = \mathbf{A}^*_{\Delta t_u}\boldsymbol{\eta}_{u-1} + \mathbf{b}^*_{\Delta t_u} + \mathbf{M}\mathbf{x}_u + \boldsymbol{\zeta}_u \quad \boldsymbol{\zeta}_u \sim \mathrm{N}(\mathbf{0}_v, \mathbf{Q}^*_{\Delta t_u}) \tag{8}$$

To reiterate, the * notation is used in this work to indicate a term that is the discrete time equivalent of the original continuous time parameter, for the time interval Δt_u. $\boldsymbol{\zeta}_u$ is the zero mean random error term for the processes at occasion u, which is distributed according to a multivariate normal with covariance $\mathbf{Q}^*_{\Delta t_u}$. The recursive nature of this solution means that at the first measurement occasion $u = 1$, the system must be initialized in some way, with $\mathbf{A}^*_{\Delta t_u}\boldsymbol{\eta}_{u-1}$ replaced by $\boldsymbol{\eta}_{t0}$, and $\mathbf{Q}^*_{\Delta t_u}$ replaced by $\mathbf{Q}^*_{t0}$. These initial states and covariances are later referred to as T0MEANS and T0VAR respectively.

Measurement model

While in principle, non-Gaussian generalizations are possible, for the purposes of this work the latent process vector $\boldsymbol{\eta}(t)$ has the linear measurement model:

$$\mathbf{y}(t) = \boldsymbol{\Lambda}\boldsymbol{\eta}(t) + \boldsymbol{\tau} + \boldsymbol{\epsilon}(t) \quad \text{where } \boldsymbol{\epsilon}(t) \sim \mathrm{N}(\mathbf{0}_c, \boldsymbol{\Theta}) \tag{9}$$

$\mathbf{y}(t) \in \mathbb{R}^c$ is the c length vector of manifest variables, $\boldsymbol{\Lambda} \in \mathbb{R}^{c\times v}$ represents the factor loadings, and $\boldsymbol{\tau} \in \mathbb{R}^c$ the manifest intercepts. The manifest residual vector $\boldsymbol{\epsilon} \in \mathbb{R}^c$ has covariance matrix $\boldsymbol{\Theta} \in \mathbb{R}^{c\times c}$. When such a measurement model is used to adequately account for non-zero equilibrium levels in the data, the continuous time intercept $\mathbf{b}$ may be unnecessary – accounting for such via the measurement model has the virtue that the measurement parameters are not dependent on the temporal dynamics, making optimization or sampling easier.

Subject level likelihood

The subject level likelihood, conditional on time dependent predictors $\mathbf{x}$ and subject level parameters $\boldsymbol{\Phi}$, is as follows:

$$p(\mathbf{y}|\boldsymbol{\Phi}, \mathbf{x}) = \prod_{u\in\mathbf{U}} p(\mathbf{y}_u|\mathbf{y}_{(u-1,u-2,\ldots,1)}, \mathbf{x}_u, \boldsymbol{\Phi}) \tag{10}$$

To avoid the large increase in parameters that comes with sampling or optimizing latent states, we use a continuous-discrete, or hybrid, Kalman filter (Kalman & Bucy, 1961) to analytically compute subject level likelihoods, conditional on subject parameters. For more on filtering see Jazwinski (2007) and Särkkä (2013). The filter operates with a prediction step, in which the expectation $\hat{\boldsymbol{\eta}}_{u|u-1}$ and covariance $\hat{\mathbf{P}}_{u|u-1}$ of the latent states are predicted by:

$$\hat{\boldsymbol{\eta}}_{u|u-1} = \mathbf{A}^*_{\Delta t_u}\hat{\boldsymbol{\eta}}_{u-1|u-1} + \mathbf{b}^*_{\Delta t_u} + \mathbf{M}\mathbf{x}_u \tag{11}$$

$$\hat{\mathbf{P}}_{u|u-1} = \mathbf{A}^*_{\Delta t_u}\hat{\mathbf{P}}_{u-1|u-1}(\mathbf{A}^*_{\Delta t_u})^\top + \mathbf{Q}^*_{\Delta t_u} \tag{12}$$

For the first measurement occasion $u = 1$, the values $\hat{\boldsymbol{\eta}}_{u|u-1}$ and $\hat{\mathbf{P}}_{u|u-1}$ must be provided to the filter. These parameters may in some cases be freely estimated, but in other cases need to be fixed or constrained, either to specific values or by enforcing a dependency to other parameters in the model, such as an assumption of stationarity.

Prediction steps are followed by an update step, wherein rows and columns of matrices are filtered as necessary depending on missingness of the measurements $\mathbf{y}$. The update step involves combining the observed data with the expectation and variance, using the Kalman gain matrix $\mathbf{K} \in \mathbb{R}^{v\times c}$,

which represents the ratio between the process inovation covariance and measurement error.

$$\hat{\mathbf{y}}_{u|u-1} = \mathbf{\Lambda}\hat{\boldsymbol{\eta}}_{u|u-1} + \boldsymbol{\tau} \tag{13}$$

$$\hat{\mathbf{V}}_u = \mathbf{\Lambda}\hat{\mathbf{P}}_{u|u-1}\mathbf{\Lambda}^\top + \mathbf{\Theta} \tag{14}$$

$$\hat{\mathbf{K}}_u = \hat{\mathbf{P}}_{u|u-1}\mathbf{\Lambda}^\top\hat{\mathbf{V}}_u^{-1} \tag{15}$$

$$\hat{\boldsymbol{\eta}}_{u|u} = \hat{\boldsymbol{\eta}}_{u|u-1} + \hat{\mathbf{K}}_u(\mathbf{y}_u - \hat{\mathbf{y}}_{u|u-1}) \tag{16}$$

$$\hat{\mathbf{P}}_{u|u} = (\mathbf{I} - \hat{\mathbf{K}}_u\mathbf{\Lambda})\hat{\mathbf{P}}_{u|u-1} \tag{17}$$

The log likelihood (*ll*) for each subject, conditional on subject level parameters, is typically[2] then (Genz & Bretz, 2009):

$$ll = \sum^{\mathbf{U}}\Big(-1/2(n\ln(2\pi) + \ln\left|\mathbf{V}_u\right| + (\hat{\mathbf{y}}_{(u|u-1)} - \mathbf{y}_u)\mathbf{V}_u^{-1}(\hat{\mathbf{y}}_{u|u-1} - \mathbf{y}_u)^\top)\Big) \tag{19}$$

Where n is the number of non-missing observations at measurement occasion u.

Population distribution

Rather than assume complete independence or dependence across subjects, we assume subject level parameters are drawn from a population distribution, for which we also estimate parameters, conditional on specified hyperpriors. This results in a joint-posterior distribution of:

$$p(\mathbf{\Phi}, \boldsymbol{\mu}, \mathbf{R}, \boldsymbol{\beta}|\mathbf{Y}, \mathbf{z}) \propto p(\mathbf{Y}|\mathbf{\Phi})p(\mathbf{\Phi}|\boldsymbol{\mu}, \mathbf{R}, \boldsymbol{\beta}, \mathbf{z})p(\boldsymbol{\mu}, \mathbf{R}, \boldsymbol{\beta}) \tag{20}$$

Where subject specific parameters $\mathbf{\Phi}_i$ are determined in the following manner:

$$\mathbf{\Phi}_i = \text{tform}\Big(\boldsymbol{\mu} + \mathbf{R}\mathbf{h}_i + \boldsymbol{\beta}\mathbf{z}_i\Big) \tag{21}$$

$$\mathbf{h}_i \sim \text{N}(\mathbf{0}, \mathbf{1}) \tag{22}$$

$$\boldsymbol{\mu} \sim \text{N}(\mathbf{0}, \mathbf{1}) \tag{23}$$

$$\boldsymbol{\beta} \sim \text{N}(\mathbf{0}, \mathbf{1}) \tag{24}$$

$\mathbf{\Phi}_i \in \mathbb{R}^s$ is the s length vector of parameters for the dynamic and measurement models of subject i. $\boldsymbol{\mu} \in \mathbb{R}^s$ parameterizes the means of the raw population distributions of subject level parameters. $\mathbf{R} \in \mathbb{R}^{s\times s}$ is the matrix square root of the raw population distribution covariance matrix, parameterizing the effect of subject specific deviations $\mathbf{h}_i \in \mathbb{R}^s$ on $\mathbf{\Phi}_i$. The matrix square root $\mathbf{R}$ is itself a transformation of parameters sampled and transformed in various ways, as discussed in the following section. $\boldsymbol{\beta} \in \mathbb{R}^{s\times w}$ is the raw effect of time independent predictors $\mathbf{z}_i \in \mathbb{R}^w$ on $\mathbf{\Phi}_i$, where w is the number of time independent predictors. $\mathbf{Y}_i$ contains all the data for subject i used in the subject level model – $\mathbf{y}$ (process related measurements) and $\mathbf{x}$ (time dependent predictors). $\mathbf{z}_i$ contains time independent predictors data for subject i. tform is an operator that applies a transform to each value of the vector it is applied to. The specific transform depends on which subject level parameter matrix the value belongs to, and the position in that matrix — these transforms and rationale are described below, but are in general necessary because many parameters require some bounded distribution, making a purely linear Gaussian approach untenable.

The basic structure of Equation 21 is such that everything inside the brackets – population distribution means, subject specific random deviations, and covariate effects – is on the unconstrained, real number scale. This bracketed portion, which we will later refer to as the *raw* subject level parameters, then undergoes some transformation function (tform) that varies depending on the sort of parameter to be estimated (e.g., standard deviations need to be positive, correlations must be between -1 and 1). This transformation of the raw subject level parameters then leaves us with the subject level parameters we are actually interested in, for example the drift matrix of temporal dynamics. For this reason, we will also refer to the population means and covariate effects inside the brackets as raw population means and raw covariate effects – they are a necessity but we are not interested in them directly. We take this approach to ensure that subject specific parameters do not violate boundary conditions, and that deviations from a mean that is close to a boundary are more likely to be smaller in the direction of the boundary than away from it. For example, for a standard deviation parameter with a population distribution mean of 0.30, a subject specific deviation of −0.40 is not possible because it results in a negative value, while +0.40 would be perfectly reasonable. Figure 3 gives a visual sense to this approach with transformations.

The approach wherein we first sample raw subject spe-

[2]For computational reasons we use an alternate but equivalent form of the log likelihood. We scale the prediction errors across all variables to a standard normal distribution, drop constant terms, calculate the log likelihood of the transformed prediction error vector, and appropriately update the log likelihood for the change in scale, as follows:

$$ll = \sum^{\mathbf{U}}\Big(ln(tr(\mathbf{V}_u^{-1/2})) - \sum 1/2(\mathbf{V}_u^{-1/2}(\hat{\mathbf{y}}_{(u|u-1)} - \mathbf{y}_u))\Big) \tag{18}$$

Where *tr* indicates the trace of a matrix, and $\mathbf{V}^{-1/2}$ is the inverse of the Cholesky decomposition of $\mathbf{V}$. The Stan software manual discusses such a *change of variables* (Stan Development Team, 2016b).

cific deviations $\mathbf{h}_i$ from a standard normal distribution, before multiplying them by the matrix square root of the population distribution $\mathbf{R}$ and add the population mean $\boldsymbol{\mu}$, may be unfamiliar to some. This is a non-centered parameterization, which we implemented to improve sampling efficiency. See Bernardo et al. (2003) and Betancourt and Girolami (2013) for discussion of non-centered parameterizations.

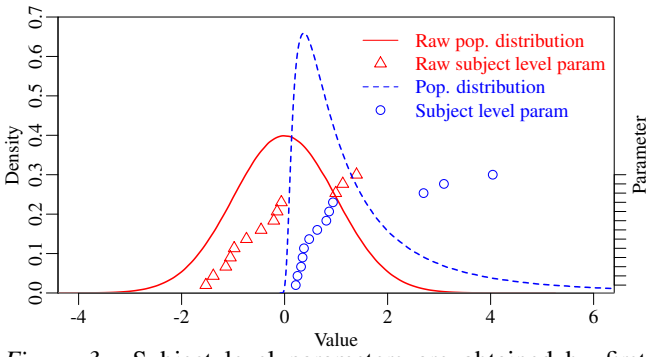

Figure 3. Subject level parameters are obtained by first sampling from a multivariate normal *raw* population distribution, and then transforming by some function to satisfy boundary and other criteria. This example shows a possibility for standard deviation parameters, using an exponential transformation. The height of the individual parameters simply represents the match between raw and transformed.

Population distribution covariance

The matrix $\mathbf{R}$, a square root[3] of the population distribution covariance matrix ($\mathbf{RR}^\intercal$), accounts for parameter correlations such as would be found when, for example, subjects that typically score highly on measurements of one process are also likely to exhibit stronger auto-effects on another. Rather than simply using the conjugate inverse-Wishart prior for the covariance matrix, we opt for an approach that first separates scale and correlation parameters (Barnard, McCulloch & Meng, 2000), wherein:

$$\mathbf{Z} = \mathbf{RR}^\intercal \tag{25}$$

$$\mathbf{R} = \mathbf{SX} \tag{26}$$

$\mathbf{Z}$ is a positive semi-definite covariance matrix, $\mathbf{R}$ is a matrix square root of covariance $\mathbf{Z}$, $\mathbf{X}$ is a matrix square root of a correlation matrix, and $\mathbf{S}$ is a diagonal matrix of standard deviations. These matrices are all of dimension $s \times s$. This separation approach is taken to ensure that the scale of population distributions does not influence the probability of the correlation between them (Tokuda, Goodrich, Van Mechelen, Gelman & Tuerlinckx, 2011), and similarly to ensure that the prior distribution does not become highly informative as variances approach zero (Gelman, 2006). Sampling $\mathbf{X}$ is still somewhat complicated however, as a) $\mathbf{XX}^\intercal$ must result in a correlation matrix, and b) the resulting marginal distribution over $\mathbf{XX}^\intercal$ should not vary across the off-diagonal elements. One approach to handling these concerns is to sample a Cholesky factor correlation matrix using an *LKJ* prior, as implemented in the Stan software (Stan Development Team, 2016a), and is based on the work of Lewandowski, Kurowicka and Joe (2009). Specifics of how the prior is obtained are complex, but the result is that of a uniform distribution over the space of correlation matrices, and prior probabilities of correlations are the same across all elements of the matrix. A drawback of this approach however is that there is no obvious way to maintain these characteristics for the hierarchical case – sampling a population mean correlation matrix, and then subject level correlation matrices based on this mean matrix, again leaves one with the problem of varying marginal correlation probabilities across elements of the matrix. While to obtain $\mathbf{R}$ we only need a single correlation matrix, for subject level covariance matrices we will need the hierarchical form, and for consistency we maintain the same approach in both cases. To deal with these concerns, our approach is as follows: To obtain $\mathbf{X}$, a matrix square root of a correlation matrix, we first sample lower off-diagonal elements from a normal distribution, scale these effects to between -1 and 1 using an inverse logit function (with appropriate scale adjustment), copy these to the upper triangle to form a symmetric matrix, and set the diagonal to 1. This matrix is then scaled by the diagonal matrix containing the inverse of the square roots of the sum of squares of each row to give us $\mathbf{X}$ – similar to scaling a covariance matrix to a correlation, but as we are scaling to a matrix square root we only multiply by the scale matrix once. This ensures that $\mathbf{XX}^\intercal$ always results in a correlation matrix, and the marginal distributions are equal across elements even in the hierarchical case. This algorithm along with a script to plot various outcomes, for any dimension and number of subjects, is included in the supplementary material. As with the LKJ approach, estimates are somewhat regularised towards zero.

To obtain the raw population distribution matrix square root $\mathbf{R}$, the correlation matrix square root $\mathbf{X}$ is pre-multiplied by a diagonal matrix $\mathbf{S}$ containing standard deviations. These standard deviations may be sampled a number of ways depending on how much information one has about their expected scale. By default, we sample from a standard normal distribution and exponentiate, just as for subject level standard deviations. When there are concerns about the informativeness of such a prior at very small values, one may instead sample from a standard normal prior distribution, truncated

[3] When speaking of matrix square roots in this context, we mean a matrix $\mathbf{R}$ that when multiplied by its own transpose as in $\mathbf{RR}^\intercal$, yields the matrix $\mathbf{Z}$. This is not a more regular matrix square root where $\mathbf{Z}$ would be given by $\mathbf{RR}$, but neither is $\mathbf{R}$ a Cholesky or other well known factor, as it is symmetric and not lower or upper triangular.

below at zero, without any transformation. In both cases, the expected scale of subject level deviations for each parameter is set by the tform function, described in the following section. When necessary, the prior for the scale of raw population distributions can be altered, by multiplying the vector of standard deviations by a scaling vector that is fixed in advance.

tform operator - parameter transforms and priors

The tform operator achieves two things. The first, is to convert the raw subject level parameters of each subjects dynamic and measurement models, from the standard normal space we use for sampling, to a range of differently shaped distributions. The second, is to set the prior distribution over our parameter space. Because the raw parameters are on a standard normal scale, applying a simple linear transform multiplying by two would give a prior with a standard deviation of two. In general, the transformations and resulting priors we discuss here are proposed as reasonable starting points for a range of typical situations, not as perfectly robust catch-all solutions. The ctsem software we discuss later allows for them to be easily altered. The transforms we choose, and resulting shape of the prior distributions, depends on the requirements of the specific parameter types. For instance as we have already discussed, standard deviation parameters cannot be negative, so a simple approach for those is for the tform operator to perform an exponential operation. Such parameter boundaries also imply that the subject level parameters are unlikely to be normally, or symmetrically, distributed, particularly as means of the population distributions approach the boundaries – a change in standard deviation from 1.00 to 0.01 is more dramatic than a change from 1.00 to 1.99. Along with the general shape and boundaries of the prior distribution, the scale is of course also important. As a general approach we have aimed for scales that support inference using standardised and centered data, with parameters at relatively normal magnitudes – neither extremely large nor extremely small, as such parameters will tend to generate numeric problems anyway. Further, when subject level parameters are estimated at an upper or lower bound in such a model, it can indicate the need for model respecification (such as including higher order terms) or rescaling the time variable. Another factor to take into account with regard to transformation, is that there is also a need to be able to fix parameters to specific, understandable values, as for instance with elements of the diffusion matrix $\mathbf{Q}$, which for higher order models will generally require a number of elements fixed to 0. This possibility can be lost under certain multivariate transformations. A final important factor for deciding on a transformation is that of sampling efficiency. Sampling efficiency is typically reduced when parameters are correlated (with respect to the sampling procedure), because a random change in one parameter requires a corresponding non-random change in another to compensate, complicating efficient exploration of the parameter space. While the use of modern sampling approaches like Hamiltonian Monte Carlo (Betancourt & Girolami, 2013) and the no U-turn sampler (Homan & Gelman, 2014) mitigate these issues to some extent, minimizing correlations between parameters through transformations still substantially improves performance. A further efficiency consideration is the inclusion of sufficient prior information to guide the sampler away from regions where the likelihood of the data approaches zero and the gradient of the likelihood is relatively flat.

Subject level covariance matrices are created in a similar way to that described for the population distribution covariance matrix, in that we pre-multiply a correlation matrix square root by a diagonal matrix containing the standard deviations to obtain a covariance matrix square root, and then post-multiply this matrix by its transpose to obtain a full covariance matrix. The correlation matrix square root is obtained as per the algorithm already discussed, and standard deviation parameters within the subject level models are obtained by exponentiating a multiple (we have chosen a value of four) of the raw parameter, which results in a prior similar to an independence Jeffreys, or reference scale prior (Bernardo, 1979), but is regularized away from the low or high extremes to ensure a proper posterior.

$$\mathbf{\Phi}_{ij} = \mathrm{e}^{4x} \tag{27}$$

where x is the raw parameter and j denotes the location of any partial correlation parameters in the subject level parameter vector $\mathbf{\Phi}_i$. This approach, wherein mass reduces to 0 at parameter boundaries, is used for all subject level parameters subject to boundaries, because typically at such boundaries other parameters of the model become empirically non-identified and optimization or sampling procedures can run into trouble.

Because intercept and regression type parameters need not be bounded, for these we simply scale the standard normal to the desired range (i.e., level of informativeness) by multiplication. We could of course also add some value if we wanted a non-zero mean.

Diagonals of the drift matrix $\mathbf{A}$ – the temporal auto effects – are transformed to be negative, with probability mass relatively uniformly distributed for discrete time autoregressive effects between 0 and 1, given a time interval of 1, but declining to 0 at the extremes.

$$\mathbf{\Phi}_{ij} = -log(\mathrm{e}^{-1.5x} + 1) \tag{28}$$

where x is the raw parameter and j denotes the location of any partial correlation parameters in the subject level parameter vector $\mathbf{\Phi}_i$. We have opted to use a bounded distribution on the drift auto effects for pragmatic reasons – values greater than 0 represent explosive, non-stationary processes

that are in most cases not theoretically plausible. While allowing for such values may point to misspecification more readily, the constrained form results in what we believe is a computationally simpler and sensible prior distribution for genuine effects – but the model and software allows for this to be easily changed, and we have also successfully tested a simple normal distribution.

We have found that off diagonals of the drift matrix **A** – the temporal cross effects – function best when specified in a problem dependent manner. For basic first order processes, they can simply be left as multiplications of the standard normal distribution. For higher order processes, it may help to parameterize the cross effects between a lower and higher order component, which determine for instance the wavelength of an oscillation, similarly to the auto effects, ensuring negative values.

Figure 4 plots the resulting prior densities when using the described transformations. Note that of course the density for a variance is directly related to the standard deviation, and the density plot for an autoregression assumes that the time interval is 1 with no cross effects involved. For the sake of completeness we include a prior density for all other parameters, such as the drift cross effects, intercepts, and regression type parameters, although these just use a simple multiplication of the standard normal.

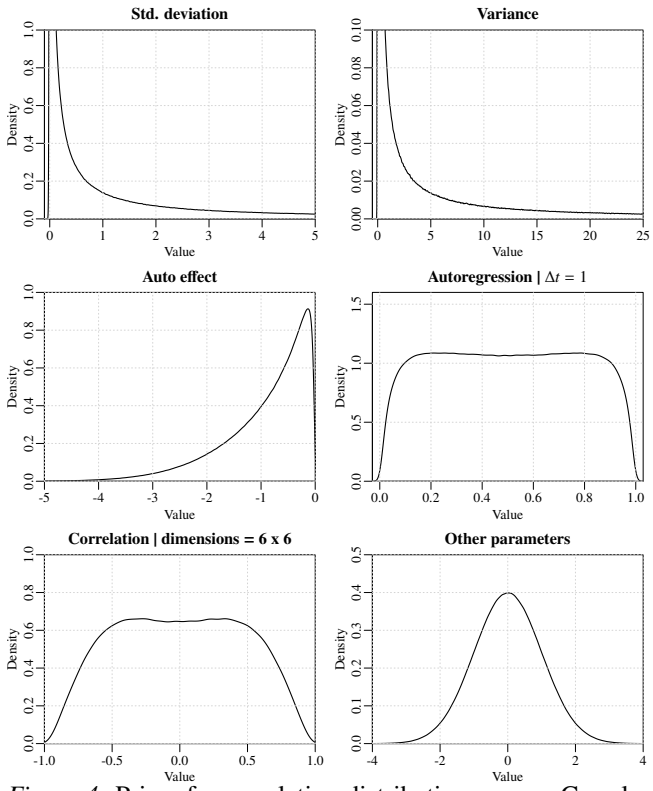

Figure 4. Priors for population distribution means. Correlation plot shown is a marginal distribution for a 6×6 matrix.

Software implementation

The hierarchical continuous time dynamic model has been implemented as an extension to the ctsem software (Driver et al., 2017) for R (R Core Team, 2014). Originally, ctsem was designed to perform maximum likelihood estimation of continuous time structural equation models as they are described in Voelkle et al., 2012, in which the structural equation matrices are set up in the RAM (reticular action model) format (J. Jack McArdle & McDonald, 1984). Individual specific time intervals are accounted for by *definition variables*, and these are coupled with matrix algebra functions to determine the expected means and covariance matrices for each individual. The need for complex functions like the matrix exponential made the OpenMx software (Neale et al., 2016) an obvious choice for fitting the models. In this original form of ctsem however, random-effects are only possible to estimate for intercept parameters. This is a primary motivation for this extension to a hierarchical Bayesian formulation, where all parameters may vary across individuals according to a simultaneously estimated distribution. To fit this new hierarchical form of the model, we use a recursive state-space formulation in which expectations for each time point are modeled conditional on the prior time point, and rely on the Stan software (Carpenter et al., 2017) for model estimation and inference.

Stan is a probabilistic programming language with some similarities to the BUGS language (Bayesian inference using Gibbs sampling) (Spiegelhalter, Thomas, Best, Gilks & Lunn, 1996) language, but greater flexibility. While the switch to a hierarchical Bayesian approach offers a range of benefits, it comes at the price of additional computation time, and the necessity of specifying priors. In cases where computation time or the use of priors is problematic, or one wishes to develop a specific model structure not available with a recursive state-space formulation, the classic form of ctsem for frequentist inference may be used via some different function arguments.

Software usage

The ctsem (Driver et al., 2017) software is available via the R software repository CRAN, using the R code `install.packages('ctsem')`. While full details on usage of the software are provided in the help files of the mentioned functions and the supplementary ctsem package vignette, 'Introduction to Hierarchical Continuous Time Dynamic Models With ctsem', http://cran.r-project.org/package=ctsem/vignettes/hierarchical.pdf, we describe the fundamentals here. The main functions of this extension to ctsem are the ctModel and ctStanFit functions. The ctModel function allows the user to specify the continuous time matrices as any combination of fixed values or freely

estimated parameters. The ctStanFit function translates the specification from ctModel into a model in the Stan language, combines this model with specified data, and estimates the model. Summary and plot functions are available for the output object, and additional details are available by directly applying rstan (Stan Development Team, 2016a) functions to the rstan fit output, available within the ctStanFit output as myfit$stanfit.

Simulation study

To confirm and demonstrate the performance of our specification of the hierarchical Bayesian continuous time dynamic model, we have conducted a small simulation study, using version 2.5.0 of the ctsem software. For the study, we used a model similar to that in the empirical study of wellbeing dynamics we show in the next section, with model structure and true parameter values of the simulation similar to those estimated by the empirical work. The generating model we used specified individual variability over (nearly) all parameters, with the individual parameters distributed according to the default priors we have already discussed – while we do not think this is crucial for performance of the model and software, we will not directly address questions of this form of potential misspecification. The initial latent variance parameters (T0VAR) were not specified as individually varying, as they are problematic to estimate when free across subjects. One of the drift matrix auto effects was also set to have zero variance, to test the performance of the specification when a random effect is specified in cases of no variation (because a standard deviation of exactly zero is not possible to obtain with this approach, for computing simulation quantities such as coverage rates, we specified 0.01 as an arbitrary, approximately zero quantity). Using the ctsem software, we fit data from this generating model using the true model structure, with default priors and start values. We generated data for either 10 or 30 observation time points, 50 or 200 subjects, a cross effect of 0.00 or 0.20, and with or without variation in time intervals between observations. The conditions were fully crossed, and each condition was repeated 200 times. The exact model structure and parameter values used for data generation can be seen in Table 1, and an R script is available in the supplementary material. Three chains with 500 warmup and 500 sampling iterations were used (Hamiltonian Monte Carlo typically generates far more effective samples per iteration than other common sampling approaches). To ensure that only usable simulation runs were included, we rejected runs if any split Rhat scale reduction factor (Gelman & Rubin, 1992) from the samples was greater than 1.10, or if the average effective samples was lower than 100. The split Rhat value refers to the ratio of between chain to within chain variance, while the number of effective samples estimates the number of independent samples out of the total drawn, after accounting for autocorrelation in the chains. In the worst case conditions with only 10 time points, this approach resulted in roughly 12% of runs being dropped. The median number of effective samples across all parameters and conditions was 335, and the median split Rhat was 1.01.

Table 1 shows true values, mean point estimates, RMSE (root mean squared error), 95% credible interval widths, and coverage rates, for all estimated parameters. If the parameter is a subject level parameter, the relevant symbol representing the parameter is also shown. For this case 200 subjects were measured at 30 times, but for other examples (N=200 & T=10, N=50 & T=10, N=50 & T=30) see Appendix C. Due to limitations with the number of simulation runs possible, the simulation measures reported will be subject to some sampling variability, and thus should not be treated as perfectly precise – nevertheless they offer insight into the performance of the model.

From Table 1 we can see that with 200 subjects and 30 time points, inferential properties are for the most part very good – empirical coverage rates of the 95% intervals are approximately 95%, there is minimal to no bias in parameters, and error from point estimates is comparable across approaches. The only deviation from this picture is for correlations between the random effects. While there are far too many (hundreds) of such correlations to table individually, for the most part the generating model had these set to zero. These zero correlations are estimated well, with conservative coverage rates of 1.00 for most parameters, with a few scattered in the 0.90 to 1.00 region. The only poor performer here was a spurious correlation between the correlation parameter in the diffusion matrix and the initial latent mean for the first process – it is not obvious to us why this occurs but it may be due to fixing the initial latent variance across subjects. Stronger correlations, as between the tabled manifestmeans parameters (*corr_manifestmeans_Y1_manifestmeans_Y2*, which sets the correlation between baseline levels of each manifest variable), exhibit a mild bias towards zero.

Table C1 shows that with only 50 subjects and 10 time points, inferential properties are still reasonable, though not optimal. Some biases in the population means are now apparent, similar to those one would expect when fitting time series models to single subject data with too few time points. This pattern could also be predicted by the generally too high estimates of population standard deviations – the population model is providing too little regularization for the subject level parameters. Depending on ones priorities, in cases with less data available such as this one it may be worthwhile to scale the prior for population standard deviations downwards. Tables C2 and C3 show that combinations of either 50 subjects and 30 time points (1500 measurements), or 200 subjects and 10 time points (2000 measurements), are effective for our test model and perform quite similarly. Of course,

Table 1
Simulation results for the full random effects model, with 200 subjects and 30 time points.

Parameter	Symbol	True value	Mean point est.	RMSE	CI width	Coverage
T0mean_eta1	$\boldsymbol{\eta}_{1[1]}$	1.00	0.99	0.10	0.40	0.95
T0mean_eta2	$\boldsymbol{\eta}_{1[2]}$	1.00	1.00	0.10	0.44	0.97
drift_eta1_eta1	$\mathbf{A}_{[1,1]}$	-0.40	-0.40	0.05	0.19	0.96
drift_eta2_eta1	$\mathbf{A}_{[1,2]}$	0.00	0.00	0.02	0.08	0.98
drift_eta1_eta2	$\mathbf{A}_{[2,1]}$	0.10	0.11	0.03	0.15	0.97
drift_eta2_eta2	$\mathbf{A}_{[2,2]}$	-0.20	-0.20	0.03	0.11	0.96
manifestvar_Y1_Y1	$\boldsymbol{\Theta}_{[1,1]}$	1.00	1.00	0.06	0.25	0.95
manifestvar_Y2_Y2	$\boldsymbol{\Theta}_{[2,2]}$	1.00	1.00	0.06	0.23	0.96
diffusion_eta1_eta1	$\mathbf{Q}_{[1,1]}$	1.00	1.00	0.09	0.34	0.94
diffusion_eta2_eta1	$\mathbf{Q}_{[2,1]}$	0.50	0.50	0.05	0.22	0.97
diffusion_eta2_eta2	$\mathbf{Q}_{[2,2]}$	1.00	0.99	0.07	0.27	0.96
T0var_eta1_eta1	$\mathbf{Q}^{*}_{1[1,1]}$	0.50	0.42	0.14	0.57	0.97
T0var_eta2_eta1	$\mathbf{Q}^{*}_{1[2,1]}$	0.10	0.21	0.38	1.73	0.99
T0var_eta2_eta2	$\mathbf{Q}^{*}_{1[2,2]}$	0.51	0.44	0.14	0.60	0.97
manifestmeans_Y1	$\boldsymbol{\tau}_{[1]}$	0.50	0.50	0.09	0.38	0.96
manifestmeans_Y2	$\boldsymbol{\tau}_{[2]}$	0.00	-0.01	0.11	0.42	0.95
hsd_manifestmeans_Y1		1.00	1.01	0.07	0.29	0.96
corr_manifestmeans_Y1_manifestmeans_Y2		0.50	0.46	0.09	0.26	0.86
hsd_manifestmeans_Y2		1.00	1.01	0.08	0.32	0.95
hsd_drift_eta1_eta1		0.15	0.14	0.06	0.22	0.92
hsd_drift_eta1_eta2		0.15	0.15	0.03	0.11	0.95
hsd_drift_eta2_eta1		0.01	0.03	0.02	0.06	0.98
hsd_drift_eta2_eta2		0.08	0.06	0.04	0.14	0.93
hsd_diffusion_eta1_eta1		0.64	0.65	0.07	0.30	0.95
hsd_diffusion_eta2_eta1		0.30	0.24	0.11	0.35	0.89
hsd_diffusion_eta2_eta2		0.64	0.65	0.06	0.25	0.95
hsd_manifestvar_Y1_Y1		0.64	0.65	0.06	0.23	0.95
hsd_manifestvar_Y2_Y2		0.64	0.65	0.05	0.22	0.95

both give less precise estimates than the 200 subjects 30 time points example already discussed.

Table 2 shows an extended set of simulation measures, including empirical power with a 5% alpha level. These measures are shown with respect to the cross effect parameter *drift_eta1_eta2* only, for conditions with and without a true cross effect, over different combinations of N and T. In this case, we also included results from a more restrictive mixed-effects style model, in which only intercept parameters were allowed to vary across subjects. While here this more restrictive model is obviously misspecified, we suspect there are many such cases as the mixed-effects form is much more commonly used, because it is simpler to specify. For cases where a cross effect of 0.2 exists, the results show that even in the worst case condition of N=50 & T=10 power is tolerable, and with N either increased to 200, or T increased to 30, power is very good. In general, it seems that under this model specification, an increase in the number of data points, regardless of whether via more N or more T, seems to improve results similarly. In cases where no cross effect exists, the correctly specified full model is only returning false positives at rates equal to or lower than the 5% alpha, but problems become apparent with the misspecified, mixed effects only model – while we should only wrongly conclude an effect exists 5% of times, as N and T increase, so too does the likelihood of making a spurious inference. However, some comfort for those relying on mixed-effects models can be taken in the mean estimates and RMSE values, indicating that while inference focusing on significance testing is likely to be problematic, actual parameter estimates for the cross effects are unlikely to be too far wrong.

To check performance when misspecification occurs in the opposite direction, with random effects specified in the fitted model while none exist in the generating model, we ran the same simulations with a mixed-effects generating model and a full random effects model fit to the data. As it is not a focus of our investigation we do not provide all the tables, but will mention only that the general trend is that population mean parameters are estimated similarly to when a random-effects generating model is used, though biases related to over-estimation of the population standard deviation parameters are somewhat stronger. Empirical coverage rates for population mean parameters were in general still around 95%.

Dynamics of overall life satisfaction and health satisfaction

To highlight usage of the ctsem software and possibilities of the model, we assessed the dynamics of overall life satisfaction and satisfaction with health, for a selection of subjects from the long running German socioeconomic panel (GSOEP) study, using version 29 of the GSOEP data. Questions regarding the fundamental structure of, and causal relations between, subjective wellbeing constructs are still very much open (Busseri & Sadava, 2011; Schimmack, 2008). Dynamic models have been posed as one way of understanding these constructs better (Headey & Muffels, 2014). Given the long time-span over which such constructs are expected to exhibit substantial change — in the order of months, years, or even decades — gathering sufficient data to reasonably

Table 2
Extended simulation results regarding only the cross-effect parameter. Columns refer to conditions with and without a true cross effect (CE), with N = 50 or 200 subjects, and T = 10 or 30 time points, collapsed over varying intervals conditions. Both full random effects and more restricted mixed effects models were fit. The mixed effects model represents a common approach, in which only intercept parameters vary over subjects.

	CE = 0, N = 50		CE = 0, N = 200		CE = 0.2, N = 50		CE = 0.2, N = 200	
Measure	**T = 10**	**T = 30**	**T = 10**	**T = 30**	**T = 10**	**T = 30**	**T = 10**	**T = 30**
Coverage - full	0.95	0.98	0.95	0.98	0.95	0.96	0.96	0.97
Coverage - mixed	0.89	0.73	0.87	0.60	0.89	0.62	0.74	0.28
Mean Est. - full	0.08	0.02	0.08	0.01	0.37	0.23	0.23	0.21
Mean Est. - mixed	0.04	-0.01	-0.01	-0.02	0.30	0.15	0.16	0.13
RMSE - full	0.15	0.05	0.15	0.03	0.25	0.08	0.09	0.04
RMSE - mixed	0.15	0.06	0.05	0.03	0.23	0.10	0.10	0.08
CI width - full	0.71	0.25	0.71	0.12	0.93	0.33	0.37	0.17
CI width - mixed	0.52	0.14	0.16	0.06	0.79	0.21	0.25	0.09
Power - full	0.05	0.02	0.05	0.02	0.55	0.92	0.91	1.00
Power - mixed	0.11	0.28	0.13	0.40	0.50	0.81	0.77	0.99

fit single-subject models is difficult. Further, although the GSOEP is administered yearly, variability in timing of the questionnaire each year results in some variability of time intervals, which if ignored, may add noise and bias. Thus, a hierarchical continuous time approach, in which we leverage variation in the time intervals between questionnaires as additional information, and inform our estimates of specific subjects dynamics based on many other subjects, seems particularly applicable to such data.

Core questions

While many questions might be asked using this approach, the questions we will address here are the very general ones: What are the temporal dynamics of overall and health satisfaction? How much variation in such dynamics exists? Are there relations between cross-sectional age and dynamics, or between certain aspects of dynamics and other aspects?

Sample details

For this example we randomly sampled 200 subjects from the GSOEP that had been observed at all 29 occasions in our data. Such a sub-sample of course no longer benefits from the population representative nature of the GSOEP. This sample resulted in subject ages (calculated at the midpoint of their participation in the study) from 30 to 77 years (mean = 49.23 and sd=10.69). In our subsample, time intervals between measurements ranged from 0.25 to 1.75 years, with a mean of 1 and standard deviation of 0.11.

Constructs

We are interested in the constructs of satisfaction with health, and overall satisfaction with life. These were measured on an 11 point scale. Translations of the questions from German are as follows: "How satisfied are you today with the following areas of your life?" followed by a range of items including "your health". These scales ranged from 0, totally unhappy, to 10, totally happy. Overall satisfaction was assessed separately, as "How satisfied are you with your life, all things considered?" and ranged from 0, completely dissatisfied, to 10, completely satisfied.

Model

The individual level dynamic model was specified as a first order bivariate model. All parameters of the bivariate latent process and measurement models were left free, except for the process intercept and loading matrices. The process intercepts were set to 0, as the measurement model here accounts for non-zero equilibria, and the factor loading matrix to an identity matrix, for model identification purposes. All free dynamic and measurement model parameters (except initial latent variance) were also free to vary across subjects. Variation in subject level parameters was predicted by age and age squared, with residual variation arising from a multivariate population distribution of parameters. R code to generate this model, plot the population distribution priors, and view the resulting Stan code, is provided in Appendix A. The matrix forms for the subject level model are shown in Figure 5, with underbraced notations indicating the relevant matrix as described in the model section of this paper, and when appropriate, also the name of the matrix in the ctsem software model specification.

Means of population distributions

Shown in Table 3 are the posterior density intervals, point estimates, and diagnostic statistics of the means of the population distributions[4], attained after sampling with four chains of 2000 iterations each, giving potential scale reduction factors (Gelman & Rubin, 1992) below 1.01 and a minimum of 194 effective samples per parameter. Note that the median,

[4]Note that any variance / covariance related parameters are reported as standard deviations and unconstrained correlation square roots – regular covariance matrices are reported in 4.

$$\underbrace{\mathrm{d}\begin{bmatrix}\eta_1\\ \eta_2\end{bmatrix}(t)}_{\mathrm{d}\boldsymbol{\eta}(t)} = \left(\underbrace{\begin{bmatrix}\text{drift\_overallSat\_overallSat} & \text{drift\_overallSat\_healthSat}\\ \text{drift\_healthSat\_overallSat} & \text{drift\_healthSat\_healthSat}\end{bmatrix}}_{\mathbf{A}\ \text{DRIFT}}\underbrace{\begin{bmatrix}\eta_1\\ \eta_2\end{bmatrix}(t)}_{\boldsymbol{\eta}(t)} + \underbrace{\begin{bmatrix}0\\ 0\end{bmatrix}}_{\mathbf{b}\ \text{CINT}}\right)\mathrm{d}t +$$

$$\underbrace{\begin{bmatrix}\text{diffusion\_overallSat\_overallSat} & 0\\ \text{diffusion\_healthSat\_overallSat} & \text{diffusion\_healthSat\_healthSat}\end{bmatrix}}_{\mathbf{G}\ \text{DIFFUSION}}\underbrace{\mathrm{d}\begin{bmatrix}W_1\\ W_2\end{bmatrix}(t)}_{\mathrm{d}\mathbf{W}(t)}$$

$$\underbrace{\begin{bmatrix}Y_1\\ Y_2\end{bmatrix}(t)}_{\mathbf{Y}(t)} = \underbrace{\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}}_{\boldsymbol{\Lambda}\ \text{LAMBDA}}\underbrace{\begin{bmatrix}\eta_1\\ \eta_2\end{bmatrix}(t)}_{\boldsymbol{\eta}(t)} + \underbrace{\begin{bmatrix}\text{manifestmeans\_overallSat}\\ \text{manifestmeans\_healthSat}\end{bmatrix}}_{\boldsymbol{\tau}\ \text{MANIFESTMEANS}} + \underbrace{\begin{bmatrix}\epsilon_1\\ \epsilon_2\end{bmatrix}(t)}_{\boldsymbol{\epsilon}(t)}$$

$$\underbrace{\begin{bmatrix}\epsilon_1\\ \epsilon_2\end{bmatrix}(t)}_{\boldsymbol{\epsilon}(t)} \sim \mathrm{N}\left(\begin{bmatrix}0\\ 0\end{bmatrix}, \underbrace{\begin{bmatrix}\text{manifestvar\_overallSat\_overallSat} & 0\\ 0 & \text{manifestvar\_healthSat\_healthSat}\end{bmatrix}}_{\boldsymbol{\Theta}\ \text{MANIFESTVAR}}\right)$$

Figure 5. Matrix specification of the subject level model of overall life satisfaction and satisfaction with health. Underbraced notations denoting the symbol used to represent the matrix in earlier formulas, and where appropriate also the matrix name in the ctsem specification[a].

[a]Strictly speaking, the diffusion matrix is actually the covariance matrix $\mathbf{G}\mathbf{G}^\top$, but $\mathbf{G}$ is the way it is specified in ctsem.

50%, is reported as the point estimate, as this is typically closest to the true value in simulations reported in Table 1.

Table 3
Posterior intervals and point estimates for means of estimated population distributions

	Symbol	2.5%	50%	97.5%
T0mean_overallSat	$\boldsymbol{\eta}_{1[1]}$	0.53	0.80	1.09
T0mean_healthSat	$\boldsymbol{\eta}_{1[2]}$	1.10	1.40	1.72
drift_overallSat_overallSat	$\mathbf{A}_{[1,1]}$	-0.49	-0.33	-0.22
drift_overallSat_healthSat	$\mathbf{A}_{[1,2]}$	0.02	0.08	0.19
drift_healthSat_overallSat	$\mathbf{A}_{[2,1]}$	-0.07	-0.01	0.04
drift_healthSat_healthSat	$\mathbf{A}_{[2,2]}$	-0.25	-0.17	-0.10
diffusion_overallSat_overallSat	$\mathbf{Q}_{[1,1]}$	0.48	0.58	0.72
diffusion_healthSat_overallSat	$\mathbf{Q}_{[2,1]}$	1.00	1.47	2.30
diffusion_healthSat_healthSat	$\mathbf{Q}_{[2,2]}$	0.51	0.60	0.70
manifestvar_overallSat_overallSat	$\boldsymbol{\Theta}_{[1,1]}$	0.78	0.85	0.91
manifestvar_healthSat_healthSat	$\boldsymbol{\Theta}_{[2,2]}$	0.98	1.05	1.11
manifestmeans_overallSat	$\boldsymbol{\tau}_{[1]}$	6.67	6.87	7.04
manifestmeans_healthSat	$\boldsymbol{\tau}_{[2]}$	6.04	6.30	6.53
T0var_overallSat_overallSat	$\mathbf{Q}^*_{1[1,1]}$	1.28	1.50	1.74
T0var_healthSat_overallSat	$\mathbf{Q}^*_{1[2,1]}$	0.35	0.57	0.91
T0var_healthSat_healthSat	$\mathbf{Q}^*_{1[2,2]}$	1.12	1.48	1.78

Going down the list of parameters shown in Table 3, the T0mean parameters are positive for both overall and health satisfaction. Because the T0means represent initial state estimates for the latent processes, and because we have specified the model such that the latent processes have long run means of 0 (by including non-zero manifest means and leaving the continuous time intercepts fixed to 0), the pop-

Table 4
Posterior intervals and point estimates for means of estimated population distributions of covariance matrices

Matrix	Symbol	2.5%	50%	97.5%
T0VAR	$\mathbf{Q}^*_{1[1,1]}$	1.63	2.24	3.02
T0VAR	$\mathbf{Q}^*_{1[2,1]}$	0.64	1.13	1.73
T0VAR	$\mathbf{Q}^*_{1[2,2]}$	1.25	2.17	3.15
DIFFUSION	$\mathbf{Q}_{[1,1]}$	0.23	0.34	0.52
DIFFUSION	$\mathbf{Q}_{[2,1]}$	0.23	0.31	0.41
DIFFUSION	$\mathbf{Q}_{[2,2]}$	0.26	0.36	0.49
dtDIFFUSION	$\mathbf{Q}_{\Delta t=1,[1,1]}$	0.19	0.27	0.38
dtDIFFUSION	$\mathbf{Q}_{\Delta t=1,[2,1]}$	0.19	0.25	0.33
dtDIFFUSION	$\mathbf{Q}_{\Delta t=1[2,2]}$	0.22	0.30	0.40
asymDIFFUSION	$\mathbf{Q}^*_{\infty[1,1]}$	0.53	0.72	0.96
asymDIFFUSION	$\mathbf{Q}^*_{\infty[2,1]}$	0.59	0.77	0.99
asymDIFFUSION	$\mathbf{Q}^*_{\infty[2,2]}$	0.77	0.99	1.29

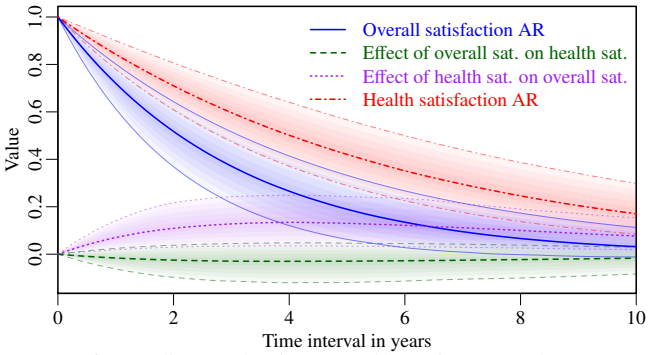

Figure 6. Median and 95% quantiles of auto and cross regressions over time.

ulation means for the T0mean parameters show to what extent subjects‘ initial states tend to be higher or lower than their later states. Combining this structure with the positive values observed, suggests that as time goes by, satisfaction scores decline somewhat, with a larger decline in the health domain. Turning to the auto effect parameters of the drift matrix, *drift_healthSat_healthSat* is higher (closer to zero) than *drift_overallSat_overallSat*, suggesting that changes in health satisfaction typically persist longer than changes in overall satisfaction. Sometimes these negative coefficients are confusing for those used to discrete-time results, but they provide a relatively intuitive interpretation – the higher above baseline a process is, the stronger the downwards pressure due to the auto-effect, and the further below baseline, the stronger the upwards pressure. The cross effect parameter *drift_healthSat_overallSat* is very close to zero, suggesting that changes in overall satisfaction do not predict later changes in health satisfaction. Conversely however, *drift_overallSat_healthSat* is substantially positive, so changes in health satisfaction predict later changes in overall satisfaction, in the same direction. To understand these temporal dynamics due to the drift matrix more intuitively, the expected auto and cross regressions over time are shown in Figure 6, where we see for instance that the expected effect of health satisfaction on overall satisfaction peaks at time intervals of around 2-4 years. Note that this does *not* imply that the relation between the two satisfaction processes is changing with time – yet because the processes are somewhat stable, and a change at one time (in the plot, a change of 1.00 at time zero) persists, this allows for consequences of the initial change to continue building for some time. The diffusion parameters are difficult to interpret on their own because a) they do not reflect total variance in the system, but rather only the rate of incoming variance (unexplained exogenous inputs), and b) also because of the unusual transformation necessary in the off-diagonal parameters. Instead we can consider the diffusion covariance matrix, as well as the discrete time and asymptotic forms, as output by ctsem and shown in Table 4. The discrete time diffusion (dtDIFFUSION) matrix tells us how much change is likely to occur in the latent processes over that interval, and the extent of covariation. The asymptotic diffusion (asymDIFFUSION) covariance matrix gives the total latent process variance and covariance. From Table 4, the dtDIFFUSION variance parameters show that over a time span of 1 year, overall and health satisfaction processe have similar levels of variance. The asymDIFFUSION variance parameters show that in the longer term, there is somewhat more variation in health satisfaction. The off-diagonal, covariance parameters show substantial positive covariation, so when overall satisfaction rises due to unmodeled factors, so too is health satisfaction likely to rise. This is unsurprising, as we might expect that overall and health satisfaction certainly share some common causes. Turning back to Table 3,The two manifest indicators show similar standard deviations for measurement error (manifestvar_overallSat_overallSat and manifestvar_healthSat_healthSat) for each process, which implies that measurement limitations and short term situational influences (e.g., a sunny day) contribute similar levels of variance to each indicator. Further, it seems that such influences contribute a similar amount of variance to the observed scores as the latent process variability – given that we fixed factor loadings to 1.00, they are directly comparable. This is not however suggestive that the measures are unreliable per se, as the measurement error and total latent process variance only reflect within-person variability – to consider reliability one would need to also consider the between-person variance in the manifest means, which in this model account for baseline levels of the processes. The manifest means parameters reflect the intercepts of the manifest indicators, and here both are at roughly similar levels, between 6 and 7. The absolute value here is probably not so interesting, it is rather the individual differences, and relations between individual differences and other parameters or covariates, that are of most interest, and these are discussed later. The T0var parameters reflect the initial variance and covariance of the latent processes, and again, the covariance matrix parameters from Table 4 are likely to be more interpretable.

Covariate effects

Cross-sectional age and age squared were included as time independent predictors for all parameters. Appendix B contains a table of full results, but because we are looking at a combined linear and quadratic effect, examining the multidimensional credible region is much simpler with the plots of Figure 7 (generated using the ctStanTIpredeffects function from ctsem). This figure shows the 50% credible intervals for the effect of age on the model parameters, as well as the implications of the model parameters such as the discrete time effects, the asymptotic diffusion variance. Discrete time (dt) matrices were computed for a time interval of two years, and

when appropriate, correlations (cor) are shown alongside covariances. 50% was chosen for the sake of interpretability of plots in this example, and we do not mean for this interval to be taken as strong support for any interpretations – although the estimated effects are regularised by the standard normal prior, to reduce the chance of large spurious effects.

The top left plot of Figure 7 focuses on initial and baseline levels of the processes, with the strongest effect being that the baseline level of health satisfaction (*manifestmeans_healthSat*) declines with age. Conversely, baseline overall satisfaction seems to rise marginally. Although the t0mean parameters, which reflect the difference between initial and baseline latent values, appear to show some change with age, it is quite modest and we would not make too much of it at this point. For a more complete modelling of within-person trends an additional latent process with zero diffusion could be included (see Driver & Voelkle, 2017, for an example specified using ctsem).

The plots in the centre and top right display the temporal dynamics from the drift matrix, with the centre plot showing the continuous time parameters and the right a discrete time effect matrix. While no effects stand out as highly substantial, there is a rise in the persistence of changes in overall satisfaction (*drift_overallSat_overallSat*) with older ages. There is also something of a decrease in the effect of health satisfaction on overall satisfaction in older ages, and a corresponding increase in the alternate effect – that of overall satisfaction on health. Turning to the lower left and centre plots containing the continuous and discrete time diffusion parameters, younger and older subjects appear to show higher within-subject variability in overall satisfaction, while when it comes to health satisfaction it is only the older subjects showing increased variability. With increasing age, random (in the sense that the model does not predict them) changes to health and overall satisfaction appear to become highly correlated, with the correlation approaching 1.00 – given this it may be plausible to model such subjects using a single latent process. Finally, the lower right plot suggests that measurements of health satisfaction, and to a lesser extent measurements of overall satisfaction, become more error prone with age.

Variance of population distributions

Shown in Table 5 are the posterior density intervals of standard deviation parameters of the population distributions, showing to what extent individual subjects parameter values tended to differ from the population mean values, in ways that could not be predicted by our age covariates – the unexplained between-subjects variance in a parameter. So while every subject has their own particular set of parameters, the estimated mean of the parameter distribution over all subjects is shown in Table 3, and the standard deviation is shown in Table 5. Individual differences in the T0means is unsurprising, as they simply reflect differences in the initial level of the latent process states. Looking at the temporal dynamics parameters, both auto effects (*drift_overallSat_overallSat* and *drift_healthSat_healthSat*) show some variability, reflecting individual differences in the persistence of changes in overall and health satisfaction processes. Regarding cross effects, the effect of health on overall satisfaction *drift_overallSat_healthSat*) shows more variability than the reverse direction *drift_healthSat_overallSat*), which seems consistent with the strength of the effects at the population level. That is, the non-existent or very weak average effect of overall satisfaction on health shows little variability, while the stronger effect of health satisfaction on later overall satisfaction varies more across subjects – the effect is in general more important, but more so for some people than others. The between-subjects variability in both the diffusion diagonals and manifestvar parameters suggests that some subjects exhibit more latent variability than others, and that some subjects measurements are noisier than others. Differences in latent variability may be due to genuine differences, in that some people just experience more change in satisfaction, but could also be due to different scale usage – some people may interpret a change of 1.00 as less meaningful than others, and consequently score themselves with more variability year to year. Differences in the manifestvar parameters may reflect that some people respond to the survey questions with more randomness, or more influence of transient conditions. Between-subjects variation in the manifest means parameters reflects individual differences in baseline levels of the two satisfaction processes, and we see slightly more variation for health satisfaction here.

Table 5
Posterior intervals and point estimates for standard deviations of estimated population distributions

	2.5%	50%	97.5%
T0mean_overallSat	0.03	0.24	0.74
T0mean_healthSat	0.02	0.28	1.06
drift_overallSat_overallSat	0.13	0.25	0.44
drift_overallSat_healthSat	0.01	0.09	0.19
drift_healthSat_overallSat	0.00	0.02	0.06
drift_healthSat_healthSat	0.07	0.16	0.27
diffusion_overallSat_overallSat	0.24	0.34	0.47
diffusion_healthSat_overallSat	0.57	1.02	1.78
diffusion_healthSat_healthSat	0.20	0.30	0.42
manifestvar_overallSat_overallSat	0.25	0.30	0.35
manifestvar_healthSat_healthSat	0.32	0.37	0.43
manifestmeans_overallSat	0.81	0.94	1.09
manifestmeans_healthSat	1.07	1.25	1.46

Correlations in individual differences

Correlations can help to better understand the individual differences in dynamics, and those whose 95% interval do

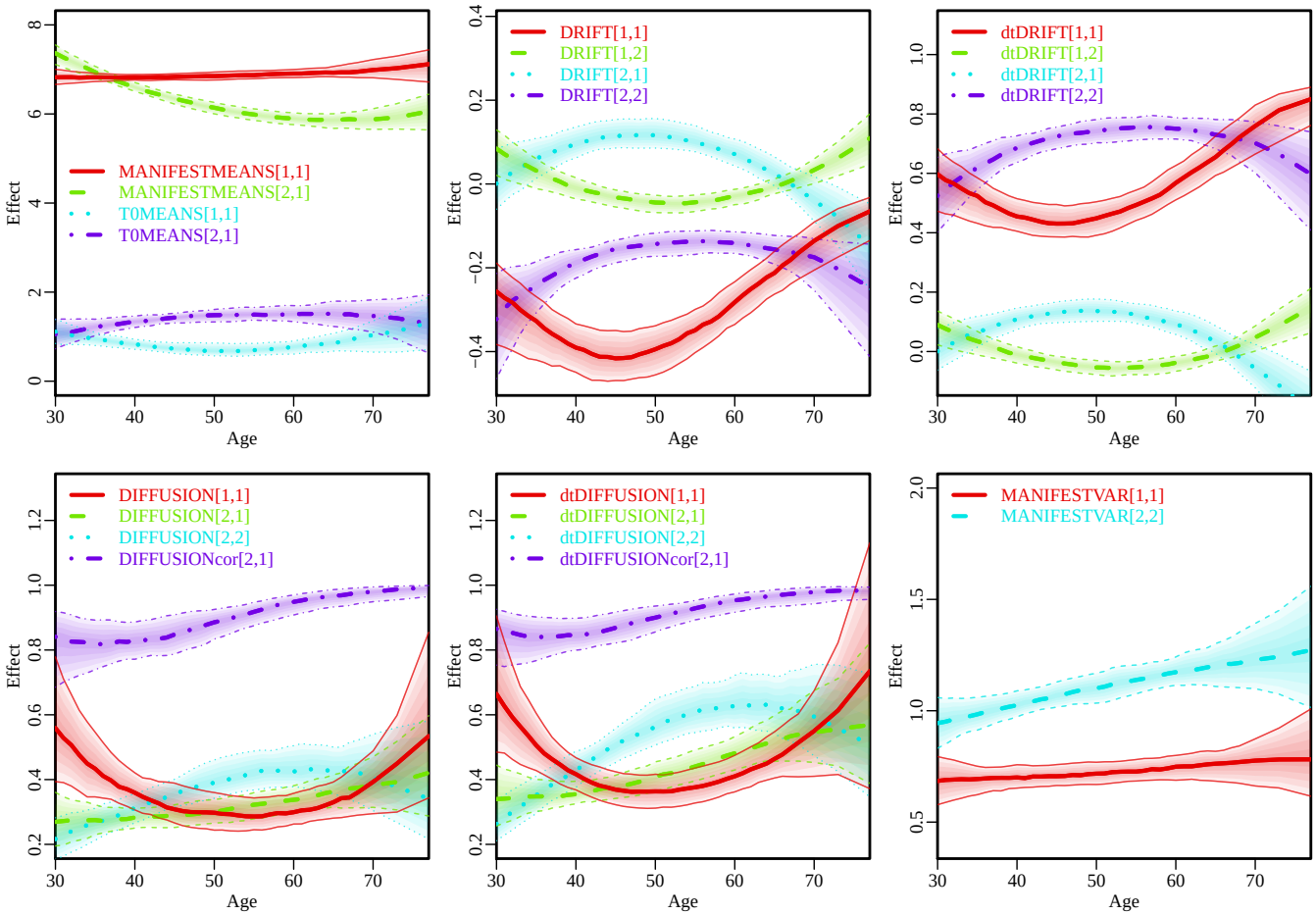

Figure 7. Estimated effect of age on model parameters. Index 1 refers to overall satisfaction, and index 2 to health satisfaction. Diffusion parameters shown are variance and covariance, unless specified as cor (correlation). Discrete time matrices computed with time interval of 2 years.

not contain zero are shown in Table 6, with health and overall satisfaction abbreviated to H and O, respectively. While we note the highly exploratory nature of this approach, we will also remind readers of the regularising prior on such correlations, reducing the likelihood of false positives.

From Table 6, we see that baseline levels of health and overall satisfaction are highly correlated, which is not too surprising. However, what may be somewhat surprising is that the table is largely comprised of correlations involving a baseline level parameter, and the pattern of results is very similar for both health and overall satisfaction parameters. This general pattern is such that increases in baseline levels (manifestmeans) predict reductions in both measurement error (manifestvar) and latent process variance (diffusion diagonals). Specific to health satisfaction, the *manifestmeans_H__drift_H_H* parameter shows that higher baseline levels predicts reduced persistence of changes (drift diagonals). There are also some correlations that do not involve the baseline level parameters, but either regard the cross effect of health on overall satisfaction, or variance terms. The *drift_O_H__drift_O_O* parameter shows a negative relation between persistence of changes in overall satisfaction (the drift auto effect), and the effect of health satisfaction on overall satisfaction (the cross effect). So, subjects for whom changes in overall satisfaction do not persist as long, show a stronger effect of health on overall satisfaction. This is a compensatory correlation, in that it serves to maintain the predictive value of changes in health on later overall satisfaction, even though the predictive value of overall satisfaction for itself at later times is weaker. Regarding correlations in variance parameters, *manifestvar_H_H__manifestvar_O_O* shows that someone who exhibits high measurement error for health satisfaction is likely to also show high measurement error on overall satisfaction, while *manifestvar_H_H__diffusion_H_H* and *manifestvar_O_O__diffusion_H_H* show that subjects who have more measurement error, also have more variance in health satisfaction.

Table 6
Posterior intervals and point estimates for correlations in random-effects, where 95% interval does not include zero.

	2.5%	50%	97.5%
manifestmeans_H__diffusion_O_O	-0.53	-0.29	-0.02
manifestmeans_O__diffusion_H_H	-0.55	-0.29	-0.03
manifestmeans_H__diffusion_H_H	-0.60	-0.31	-0.04
manifestmeans_H__manifestvar_O_O	-0.50	-0.32	-0.13
manifestvar_O_O__diffusion_H_H	0.11	0.38	0.63
manifestmeans_O__manifestvar_H_H	-0.52	-0.39	-0.25
manifestmeans_O__diffusion_O_O	-0.63	-0.40	-0.13
diffusion_H_H__diffusion_O_O	0.01	0.43	0.72
manifestmeans_H__drift_H_H	-0.69	-0.43	-0.04
manifestvar_H_H__diffusion_H_H	0.18	0.49	0.74
manifestmeans_H__manifestvar_H_H	-0.61	-0.49	-0.36
manifestmeans_O__manifestvar_O_O	-0.69	-0.55	-0.38
drift_O_H__drift_O_O	-0.82	-0.56	-0.04
manifestvar_H_H__manifestvar_O_O	0.41	0.56	0.70
manifestmeans_H__manifestmeans_O	0.67	0.76	0.82

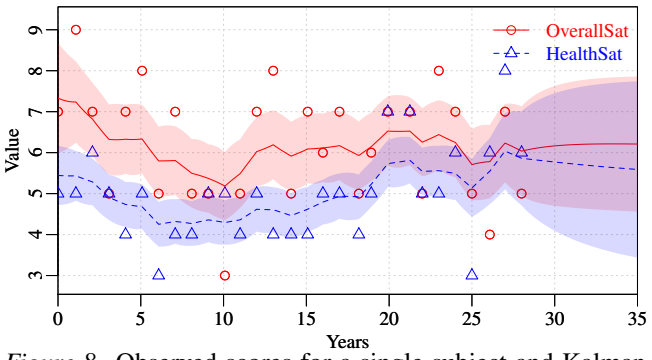

Figure 8. Observed scores for a single subject and Kalman smoothed score estimates from the subjects dynamic model.

Individual level analyses

For individual level analysis, we may compute predicted (based on all prior observations), updated (based on all prior and current observations), and smoothed (based on all observations) expectations and covariances from the Kalman filter, based on specific subjects models. All of this is readily achieved with the ctKalman function from ctsem. This approach allows for: predictions regarding individuals‘ states at any point in time, given any values on the time dependent predictors (external inputs such as interventions or events); residual analysis to check for unmodeled dependencies in the data; or simply as a means of visualization, for comprehension and model sanity checking purposes. An example of such is depicted in Figure 8, where we see observed and smoothed scores with uncertainty (95% intervals), for a randomly selected subject from our sample. Note that we show the uncertainty regarding the latent process – the not displayed measurement uncertainty results in many observations outside of the 95% interval.

Discussion

We have described a hierarchical continuous time dynamic model, for the analysis of repeated measures data of multiple subjects. In addition, we have introduced a Bayesian extension to the free and open-source software ctsem (Driver et al., 2017) that allows fitting this model. The subject level model is flexible enough to allow for many popular longitudinal model structures in psychological research to be specified, including forms of latent trajectories (Delsing & Oud, 2008), autoregressive cross-lagged models (and thus the latent change score formulation, see Voelkle & Oud, 2015), higher order oscillating models, or some mixture of approaches. We can use the model to learn about the measurement structure, variance and covariance of the latent processes stochastic aspects, and temporal dynamics of the processes. Because it is a continuous time model, there are no restrictions on the timing of data collection. Time intervals between observations can vary both within and across individuals, and indeed such variability is likely to improve the estimation of the model (Voelkle & Oud, 2013). The hierarchical aspect ensures that while each individual can have their own distinct model parameters, data collected from other subjects is still informative, leading generally to improved individual and population estimates. The inclusion of subject specific covariates (time independent predictors) may be used to improve estimation and inform about relationships to parameters, but is also not necessary as any sort of control, as heterogeneity at the individual level is accounted for by allowing random subject specific parameters.

Results from a limited simulation study suggest that the Bayesian form offers reliable inferential properties when the correct model is specified, with only marginal reductions when an overly complex population model is specified. In the more limited data conditions, certain point estimates were biased, though coverage rates in general remained good. Under conditions similar or worse than our 50 subject 10 time point example, additional thought regarding model complexity and prior specification may be helpful. Of course, while the generating model was largely based on the empirical results from the GSOEP data, if the generating model had parameter values that differed substantially from what we have posed as a rough norm (see Figure 4), performance of the approach may be different. Thus, care should of course be taken that variables are scaled and centred, and the time scale is such that neither extremely high or extremely low auto effects are expected – often a time scale that gives time intervals around 1.00 serves to achieve this. The only estimates showing noticeable bias in the simulation conditions with more data (200 subjects with 30 time points) were those of strong correlations in random effects, which are slightly pushed towards zero. While alternate approaches to correlation matrices that avoid this are in theory possible, some mild regularisation of the many (hundreds in our example) random

effects correlations has in our experience helped to minimise computational difficulties and improve performance. While more comprehensive power and simulation studies will help to improve the understanding of this and similar models under a wider range of conditions, our results serve to provide some confidence in the procedure and software.

Now, while the model as specified is relatively flexible, there are of course some limitations: Although parameter differences *between* individuals are fully accounted for, for the most part we do not account for parameter variability *within* individuals. Unpredictable changes in the process means can be accounted for through augmentation of the state matrices (Delsing & Oud, 2008; Oud & Jansen, 2000), and known changes can be accounted for using time dependent predictors (Driver & Voelkle, 2017), but changes in dynamic relationships, or randomly fluctuating parameters, are at present not accounted for. Such effects generally imply non-linear stochastic differential equations, and alternative, more complex approaches to dynamic modelling are necessary, as for example in Lu et al. (2015).

In the form described, the model and software is also limited to a linear measurement model with normally distributed errors. However, an option to avoid the Kalman filter and explicitly sample latent states is included in the ctsem software, so although most alternate measurement models are not explicitly specifiable via the ctsem functions at present, non-linearities in the measurement model and link are possible with some modifications to the Stan model that is output by the software.

Additional limitations are those typically found when dealing with either complex hierarchical models, dynamic models, or specifically continuous time dynamic models. These include computation time, parameter dependencies, and parameter interpretation.

Computation time is generally high for hierarchical time series models estimated with sampling approaches typical to Bayesian modeling. In this case the continuous time aspect adds some additional computations for every measurement occasion where the time interval changes. Based on our experience so far, using ctsem and Stan in their present forms on a modern desktop PC requires anywhere from a few minutes for simple univariate process models with limited between-subject random effects, to days for complex multi-process multi-measurement models with many subjects and covariates. The satisfaction example we discussed took approximately 6 hours to complete.

Parameter dependencies in dynamic models pose difficulties both during estimation, and during interpretation. While on the one hand it would be great to specify the model using parameters that were entirely independent of each other, this is not always possible for every parameter, and even when it is, may limit the initial specification and or compound difficulties with interpretation. For instance, rather than parameterizing the innovation covariance matrix using the standard deviations and correlation of the continuous-time diffusion matrix, one alternative we have explored is to estimate directly the asymptotic innovation covariance matrix, $\mathbf{Q}^*_{\Delta t=\infty}$. This is beneficial in that the covariance matrix parameters are made independent of the temporal dynamics parameters, and in this case we think also assists in interpretation. Unfortunately however, it limits the possibility to specify more complex dynamics where many elements of the continuous time diffusion matrix may need to be fixed to zero, but the asymptotic latent variance matrix cannot be determined in advance. While the specific parameterizations we propose may need to be adapted, in this paper we have aimed for a middle ground approach, trying to balance competing factors for typical use cases.

Interpretation of the continuous time dynamic parameters is typically less intuitive for the applied researcher than the related discrete time parameters. While this may be true, we would argue that the two forms can yield different interpretations, and that in general it is helpful to consider both the underlying continuous time parameters as well as the discrete time implications. We hope our explanations of the various parameters encourage people to explore these aspects. We also hope to encourage better intuition for dynamic models in general, by plotting expected effects (as per Figure 6) over a range of time intervals.

While neither hierarchical random-effects models nor continuous time dynamic models are themselves new, there have been few approaches put forward combining the two. We believe it is important to describe such a model and provide the means to estimate it, because accurate estimates for single subject dynamic models may require very large numbers of time points, and because inaccuracies in the estimation of one parameter propagate through the majority of others due to dependencies in the models. We have highlighted a potential application, by showing that we can estimate individual specific models for subjects in long term panel studies such as the GSOEP. Such studies may contain information over a very long amount of time, which is great in terms of investigating developmental questions, but the data are usually not dense enough to estimate individual specific models without the help of information from other individuals. Amongst other things, from this analysis we found some evidence that changes in health satisfaction predict later changes in overall satisfaction, and also that people functioning at different baselines on the satisfaction scales tend to show different dynamics and measurement properties. We hope that this work provides a means for more understanding of processes occurring within individuals, and the factors that relate to differences in such processes between individuals.

Acknowledgments

The data used in this work were made available to us by the German Socio-Economic Panel Study (SOEP) at the German Institute for Economic Research (DIW), Berlin. Software used for the development, analyses and documentation includes knitr (Xie, 2014), texStudio, R (R Core Team, 2014), RStudio (RStudio Team, 2016), Ωnyx (von Oertzen, Brandmaier & Tsang, 2015), OpenMx (Neale et al., 2016), Stan (Carpenter et al., 2017), and RStan (Stan Development Team, 2016a). The formal methods team at the Max Planck Institute for Human Development, and JHL Oud, provided valuable input. Thanks to three reviewers for their diligent and thoughtful reviews.

References

Balestra, P. & Nerlove, M. (1966). Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica*, *34*(3), 585–612. doi:10.2307/1909771. JSTOR: 1909771

Barnard, J., McCulloch, R. & Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 1281–1311. JSTOR: 24306780

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, 113–147. JSTOR: 2985028

Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M. & West, M. (2003). Non-centered parameterisations for hierarchical models and data augmentation. *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting*, 307.

Betancourt, M. J. & Girolami, M. (2013). Hamiltonian Monte Carlo for hierarchical models. arXiv: 1312.0906 [stat]. Retrieved from http://arxiv.org/abs/1312.0906

Boker, S. M. (2001). Differential structural equation modeling of intraindividual variability. *New methods for the analysis of change*, 5–27. doi:10.1037/10409-001

Boker, S. M., Deboeck, P. R., Edler, C. & Keel, P. K. (2010). Generalized local linear approximation of derivatives from time series. In S. -M, E. Ferrer & F. Hsieh (Eds.), *Statistical methods for modeling human dynamics: An interdisciplinary dialogue* (pp. 161–178). The Notre Dame series on quantitative methodology. New York, NY, US: Routledge/Taylor & Francis Group.

Boulton, A. J. (2014). Bayesian estimation of a continuous-time model for discretely-observed panel data. Retrieved from https://kuscholarworks.ku.edu/handle/1808/16843

Box, G. E., Jenkins, G. M., Reinsel, G. C. & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). John Wiley & Sons.

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., … Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLOS ONE*, *8*(4), e60188. doi:10.1371/journal.pone.0060188

Busseri, M. A. & Sadava, S. W. (2011). A review of the tripartite structure of subjective well-being: Implications for conceptualization, operationalization, analysis, and synthesis. *Personality and Social Psychology Review*, *15*(3), 290–314. WOS:000292207700003. doi:10.1177/1088868310391271

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). doi:10.18637/jss.v076.i01

Cattell, R. B. (1963). The structuring of change by P-technique and incremental R-technique. *Problems in measuring change*, 167–198.

Chow, S.-M., Lu, Z., Sherwood, A. & Zhu, H. (2014). Fitting nonlinear ordinary differential equation models with random effects and unknown initial conditions using the stochastic approximation expectation–maximization (SAEM) algorithm. *Psychometrika*, *81*(1), 102–134. doi:10.1007/s11336-014-9431-z

Chow, S.-M., Ram, N., Boker, S. M., Fujita, F. & Clore, G. (2005). Emotion as a thermostat: Representing emotion regulation using a damped oscillator model. *Emotion*, *5*(2), 208–225. doi:10.1037/1528-3542.5.2.208

Deboeck, P. R., Nicholson, J. S., Bergeman, C. S. & Preacher, K. J. (2013). From modeling long-term growth to short-term fluctuations: Differential equation modeling is the language of change. In R. E. Millsap, L. A. van der Ark, D. M. Bolt & C. M. Woods (Eds.), *New Developments in Quantitative Psychology* (66, pp. 427–447). Springer Proceedings in Mathematics & Statistics. Springer New York. doi:10.1007/978-1-4614-9348-8_28

Delattre, M., Genon-Catalot, V. & Samson, A. (2013). Maximum likelihood estimation for stochastic differential equations with random effects. *Scandinavian Journal of Statistics*, *40*(2), 322–343. doi:10.1111/j.1467-9469.2012.00813.x

Delsing, M. J. M. H. & Oud, J. H. L. (2008). Analyzing reciprocal relationships by means of the continuous-time autoregressive latent trajectory model. *Statistica Neerlandica*, *62*(1), 58–82. doi:10.1111/j.1467-9574.2007.00386.x

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). doi:10.18637/jss.v077.i05

Driver, C. C. & Voelkle, M. C. (2017). Understanding the time course of interventions with continuous time dynamic models. *Manuscript submitted for publication.*

Eager, C. & Roy, J. (2017). Mixed effects models are sometimes terrible. *arXiv preprint arXiv:1701.04858*. Retrieved from https://arxiv.org/abs/1701.04858

Finkel, S. E. (1995). *Causal analysis with panel data*. Sage.

Gardiner, C. W. (1985). *Handbook of stochastic methods*. Springer Berlin. Retrieved from http://tocs.ulb.tu-darmstadt.de/12326852.pdf

Gasimova, F., Robitzsch, A., Wilhelm, O., Boker, S. M., Hu, Y. & Hülür, G. (2014). Dynamical systems analysis applied to working memory data. *Quantitative Psychology and Measurement*, *5*, 687. doi:10.3389/fpsyg.2014.00687

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*(3), 515–534. doi:10.1214/06-BA117A

Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. doi:10.1214/ss/1177011136. JSTOR: 2246093

Genz, A. & Bretz, F. (2009). *Computation of multivariate normal and t probabilities*. Lecture Notes in Statistics. Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from http://link.springer.com/10.1007/978-3-642-01689-9

Grasman, J., Grasman, R. P. P. P. & van der Maas, H. L. J. (2016). The dynamics of addiction: Craving versus self-control. *PLOS ONE*, *11*(6). doi:10.1371/journal.pone.0158323

Halaby, C. N. (2004). Panel models in sociological research: Theory into practice. *Annual Review of Sociology*, *30*(1), 507–544. doi:10.1146/annurev.soc.30.012703.110629

Hamaker, E. L., Kuiper, R. M. & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological methods*, *20*(1), 102. doi:10.1037/a0038889

Hamaker, E. L., Zhang, Z. & van der Maas, H. L. J. (2009). Using threshold autoregressive models to study dyadic interactions. *Psychometrika*, *74*(4), 727. doi:10.1007/s11336-009-9113-4

Hamilton, J. (1986). *State-space models*. Elsevier. Retrieved from http://econpapers.repec.org/bookchap/eeeecochp/4-50.htm

Headey, B. & Muffels, R. (2014). Trajectories of life satisfaction: Positive feedback loops may explain why life satisfaction changes in multi-year waves, rather than oscillating around a set-point. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2470527

Hertzog, C. & Nesselroade, J. R. (2003). Assessing psychological change in adulthood: An overview of methodological issues. *Psychology and aging*, *18*(4), 639. doi:10.1037/0882-7974.18.4.639

Homan, M. D. & Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* *15*(1), 1593–1623. Retrieved from http://dl.acm.org/citation.cfm?id=2627435.2638586

Jazwinski, A. H. (2007). *Stochastic processes and filtering theory*. Courier Corporation.

Kalman, R. E. & Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering*, *83*(1), 95–108. doi:10.1115/1.3658902

Koval, P., Sütterlin, S. & Kuppens, P. (2016). Emotional inertia is associated with lower well-being when controlling for differences in emotional context. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.01997. pmid: 26779099

Leander, J., Almquist, J., Ahlström, C., Gabrielsson, J. & Jirstrand, M. (2015). Mixed effects modeling using stochastic differential equations: Illustrated by pharmacokinetic data of nicotinic acid in obese zucker rats. *The AAPS Journal*, *17*(3), 586–596. doi:10.1208/s12248-015-9718-8. pmid: 25693487

Lewandowski, D., Kurowicka, D. & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001. doi:10.1016/j.jmva.2009.04.008

Liu, S. (2017). Person-specific versus multilevel autoregressive models: Accuracy in parameter estimates at the population and individual levels. *British Journal of Mathematical and Statistical Psychology*, n/a–n/a. doi:10.1111/bmsp.12096

Lodewyckx, T., Tuerlinckx, F., Kuppens, P., Allen, N. B. & Sheeber, L. (2011). A hierarchical state space approach to affective dynamics. *Journal of Mathematical Psychology*. Special Issue on Hierarchical Bayesian Models, *55*(1), 68–83. doi:10.1016/j.jmp.2010.08.004

Lu, Z.-H., Chow, S.-M., Sherwood, A. & Zhu, H. (2015). Bayesian analysis of ambulatory blood pressure dynamics with application to irregularly spaced sparse data. *The Annals of Applied Statistics*, *9*(3), 1601–1620. doi:10.1214/15-AOAS846

Marriott, F. H. C. & Pope, J. A. (1954). Bias in the Estimation of Autocorrelations. *Biometrika*, *41*, 390–402. doi:10.2307/2332719. JSTOR: 2332719

McArdle, J. J. [J. Jack] & McDonald, R. P. (1984). Some algebraic properties of the reticular action model for mo-

ment structures. *British Journal of Mathematical and Statistical Psychology*, *37*(2), 234–251. doi:10.1111/j.2044-8317.1984.tb00802.x

McArdle, J. J. [John J.]. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual review of psychology*, *60*, 577–605. 00289. doi:10.1146/annurev.psych.60.110707.163612

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 and Perspectives*, *2*(4), 201–218. doi:10.1207/s15366359mea0204_1

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., ... Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, *81*(2), 535–549. doi:10.1007/s11336-014-9435-8. pmid: 25622929

Oravecz, Z., Tuerlinckx, F. & Vandekerckhove, J. (2009). A hierarchical Ornstein-Uhlenbeck model for continuous repeated measurement data. *Psychometrika*, *74*(3), 395–418. doi:10.1007/s11336-008-9106-8

Oravecz, Z., Tuerlinckx, F. & Vandekerckhove, J. (2016). Bayesian data analysis with the bivariate hierarchial Ornstein-Uhlenbeck process model. *Multivariate Behavioral Research*, (51), 106–119. doi:10 . 1080 / 00273171.2015.1110512

Oud, J. H. L. (2002). Continuous time modeling of the cross-lagged panel design. *Kwantitatieve Methoden*, *69*(1), 1–26. Retrieved from http://members.chello.nl/j.oud7/Oud2002.pdf

Oud, J. H. L. & Folmer, H. (2011). Reply to Steele & Ferrer: Modeling oscillation, approximately or exactly? *Multivariate Behavioral Research*, *46*(6), 985–993. doi:10.1080/00273171.2011.625306. pmid: 26736120

Oud, J. H. L. & Jansen, R. A. R. G. (2000). Continuous time state space modeling of panel data by means of SEM. *Psychometrika*, *65*(2), 199–215. doi:10.1007/BF02294374

R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rindskopf, D. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika*, *49*(1), 37–47. doi:10.1007/BF02294204

RStudio Team. (2016). *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com/

Särkkä, S. (2013). *Bayesian filtering and smoothing*. Cambridge University Press.

Särkkä, S., Hartikainen, J., Mbalawata, I. S. & Haario, H. (2013). Posterior inference on parameters of stochastic differential equations via non-linear Gaussian filtering and adaptive MCMC. *Statistics and Computing*, *25*(2), 427–437. doi:10.1007/s11222-013-9441-1

Schimmack, U. (2008). The structure of subjective well-being. *The science of subjective well-being*, 97–123. 00148.

Schuurman, N. K. (2016). Multilevel autoregressive modeling in psychology: Snags and solutions. *Doctoral dissertation*. Retrieved from http://dspace.library.uu.nl/handle/1874/337475

Schuurman, N. K., Ferrer, E., de Boer-Sonnenschein, M. & Hamaker, E. L. (2016). How to compare cross-lagged associations in a multilevel autoregressive model. *Psychological Methods*, *21*(2), 206–221. doi:10.1037/met0000062

Schuurman, N. K., Houtveen, J. H. & Hamaker, E. L. (2015). Incorporating measurement error in n = 1 psychological autoregressive modeling. *Frontiers in Psychology*, *6*. doi:10 . 3389 / fpsyg . 2015 . 01038. pmid: 26283988

Spiegelhalter, D. J., Thomas, A., Best, N. G., Gilks, W. & Lunn, D. (1996). BUGS: Bayesian inference using Gibbs sampling. *Version 0.5,(version ii) http://www.mrc-bsu. cam. ac. uk/bugs*, *19*.

Stan Development Team. (2016a). RStan: The R interface to Stan (Version 2.11). Retrieved from http://mc-stan.org

Stan Development Team. (2016b). *Stan modeling language users guide and reference manual, version 2.9.0.* Retrieved from http://mc-stan.org

Steele, J. S. & Ferrer, E. (2011). Response to Oud & Folmer: Randomness and residuals. *Multivariate Behavioral Research*, *46*(6), 994–1003. doi:10.1080/00273171.2011.625308. pmid: 26736121

Steele, J. S., Ferrer, E. & Nesselroade, J. R. (2014). An idiographic approach to estimating models of dyadic interactions with differential equations. *Psychometrika*, *79*(4), 675–700. doi:10.1007/s11336-013-9366-9. pmid: 24352513

Tokuda, T., Goodrich, B., Van Mechelen, I., Gelman, A. & Tuerlinckx, F. (2011). Visualizing distributions of covariance matrices. *Unpublished manuscript.* Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.221.680&rep=rep1&type=pdf

Tómasson, H. (2013). Some computational aspects of Gaussian CARMA modelling. *Statistics and Computing*, *25*(2), 375–387. doi:10.1007/s11222-013-9438-9

Voelkle, M. C., Brose, A., Schmiedek, F. & Lindenberger, U. (2014). Toward a unified framework for the study of between-person and within-person structures: Building a bridge between two research paradigms. *Multivariate Behavioral Research*, *49*(3), 193–213. doi:10.1080/00273171.2014.889593

Voelkle, M. C. & Oud, J. H. L. (2013). Continuous time modelling with individually varying time intervals for oscillating and non-oscillating processes. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 103–126. doi:10.1111/j.2044-8317.2012.02043.x

Voelkle, M. C. & Oud, J. H. L. (2015). Relating latent change score and continuous time models. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(3), 366–381. doi:10.1080/10705511.2014.935918

Voelkle, M. C., Oud, J. H. L., Davidov, E. & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, *17*(2), 176–192. doi:10.1037/a0027543. pmid: 22486576

von Oertzen, T., Brandmaier, A. M. & Tsang, S. (2015). Structural Equation Modeling With Onyx. *Structural Equation Modeling*, *22*(1), 148–161. doi:10 . 1080 / 10705511.2014.935842

Wang, L. P. & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods*, *20*(1), 63–83. doi:10 . 1037 / met0000030. pmid: 25822206

Xie, Y. (2014). Knitr: A Comprehensive Tool for Reproducible Research in R. In V. Stodden, F. Leisch & R. D. Peng (Eds.), *Implementing Reproducible Computational Research*. Chapman and Hall/CRC. Retrieved from http://www.crcpress.com/product/isbn/9781466561595

Appendix A
Dynamic model of wellbeing — R script

This script installs and loads ctsem, and specifies the dynamic model of wellbeing used in this paper (although we cannot distribute the GSOEP data). The script also plots the priors used for the population level model, as well as examples of possible priors for the subject level parameters – because the subject level priors depend on the estimated population level parameters.

```
install.packages("ctsem")
require(ctsem)

model<-ctModel(type='stanct', LAMBDA=diag(2),
  n.manifest=2, manifestNames=c('overallSat', 'healthSat'),
  n.latent=2, latentNames=c('overallSat', 'healthSat'),
  n.TIpred=2, TIpredNames=c("meanAge", "meanAgeSq"))

plot(model, wait=TRUE)

fakedata=matrix(1,nrow=5,ncol=6)
colnames(fakedata)=c('id','time','overallSat','healthSat','meanAge','meanAgeSq')
fakedata[,'time']=1:5
fit<-ctStanFit(fakedata,model,fit=FALSE)
cat(fit$stanmodeltext)
```

Appendix B
Time independent predictor effects

Table B1
Posterior distributions for time independent predictor effects

	2.5%	mean	97.5%	z
Age_manifestmeans_healthSat	-0.64	-0.41	-0.17	-3.49
AgeSq_drift_healthSat_overallSat	-0.00	0.03	0.07	1.50
AgeSq_drift_overallSat_overallSat	-0.22	-0.09	0.02	-1.50
Age_diffusion_healthSat_overallSat	-0.07	0.31	0.78	1.45
AgeSq_drift_overallSat_healthSat	-0.10	-0.04	0.02	-1.34
Age_diffusion_healthSat_healthSat	-0.05	0.07	0.18	1.20
AgeSq_manifestmeans_healthSat	-0.10	0.14	0.37	1.11
Age_manifestvar_healthSat_healthSat	-0.03	0.03	0.10	1.03
AgeSq_diffusion_overallSat_overallSat	-0.05	0.05	0.16	1.02
Age_drift_healthSat_healthSat	-0.15	-0.05	0.05	-0.96
Age_drift_overallSat_overallSat	-0.19	-0.06	0.08	-0.87
AgeSq_drift_healthSat_healthSat	-0.05	0.04	0.13	0.87
AgeSq_T0mean_overallSat	-0.16	0.12	0.38	0.84
Age_diffusion_overallSat_overallSat	-0.17	-0.05	0.08	-0.77
Age_T0mean_healthSat	-0.18	0.12	0.42	0.76
Age_drift_healthSat_overallSat	-0.06	-0.02	0.02	-0.76
AgeSq_diffusion_healthSat_healthSat	-0.14	-0.04	0.07	-0.67
Age_manifestmeans_overallSat	-0.12	0.05	0.23	0.57
AgeSq_diffusion_healthSat_overallSat	-0.24	0.11	0.54	0.56
AgeSq_T0mean_healthSat	-0.38	-0.08	0.22	-0.49
Age_T0mean_overallSat	-0.34	-0.06	0.22	-0.42
Age_manifestvar_overallSat_overallSat	-0.05	0.01	0.07	0.33
Age_drift_overallSat_healthSat	-0.07	-0.01	0.06	-0.24
AgeSq_manifestmeans_overallSat	-0.18	0.01	0.19	0.13
AgeSq_manifestvar_overallSat_overallSat	-0.06	-0.00	0.05	-0.11
AgeSq_manifestvar_healthSat_healthSat	-0.06	-0.00	0.06	-0.06

Appendix C
Additional simulation results

Table C1

Simulation results for 50 subjects and 10 time points, with all parameters varying over subjects.

Parameter	Symbol	True value	Mean point est.	RMSE	CI width	Coverage
T0mean_eta1	$\boldsymbol{\eta}_{1[1]}$	1.00	1.03	0.32	1.40	0.95
T0mean_eta2	$\boldsymbol{\eta}_{1[2]}$	1.00	1.00	0.35	1.62	0.96
drift_eta1_eta1	$\mathbf{A}_{[1,1]}$	-0.40	-0.69	0.39	1.38	0.92
drift_eta2_eta1	$\mathbf{A}_{[1,2]}$	0.00	0.12	0.19	0.77	0.95
drift_eta1_eta2	$\mathbf{A}_{[2,1]}$	0.10	0.22	0.20	0.82	0.95
drift_eta2_eta2	$\mathbf{A}_{[2,2]}$	-0.20	-0.46	0.31	1.10	0.91
manifestvar_Y1_Y1	$\boldsymbol{\Theta}_{[1,1]}$	1.00	0.97	0.18	0.77	0.97
manifestvar_Y2_Y2	$\boldsymbol{\Theta}_{[2,2]}$	1.00	0.94	0.18	0.74	0.95
diffusion_eta1_eta1	$\mathbf{Q}_{[1,1]}$	1.00	1.08	0.35	1.42	0.96
diffusion_eta2_eta1	$\mathbf{Q}_{[2,1]}$	0.50	0.50	0.19	1.05	0.99
diffusion_eta2_eta2	$\mathbf{Q}_{[2,2]}$	1.00	1.15	0.32	1.27	0.95
T0var_eta1_eta1	$\mathbf{Q}^*_{1[1,1]}$	0.50	0.56	0.15	1.05	1.00
T0var_eta2_eta1	$\mathbf{Q}^*_{1[2,1]}$	0.10	0.27	0.35	1.87	1.00
T0var_eta2_eta2	$\mathbf{Q}^*_{1[2,2]}$	0.51	0.60	0.17	1.10	1.00
manifestmeans_Y1	$\boldsymbol{\tau}_{[1]}$	0.50	0.48	0.30	1.28	0.94
manifestmeans_Y2	$\boldsymbol{\tau}_{[2]}$	0.00	0.01	0.34	1.50	0.96
hsd_manifestmeans_Y1		1.00	0.96	0.21	0.95	0.96
corr_manifestmeans_Y1_manifestmeans_Y2		0.50	0.27	0.25	0.94	0.94
hsd_manifestmeans_Y2		1.00	0.94	0.24	1.08	0.97
hsd_drift_eta1_eta1		0.15	0.48	0.33	2.25	0.98
hsd_drift_eta1_eta2		0.15	0.22	0.14	0.65	0.99
hsd_drift_eta2_eta1		0.01	0.16	0.14	0.54	0.87
hsd_drift_eta2_eta2		0.08	0.53	0.42	2.76	0.97
hsd_diffusion_eta1_eta1		0.64	0.65	0.28	1.15	0.94
hsd_diffusion_eta2_eta1		0.30	1.26	0.32	8.81	1.00
hsd_diffusion_eta2_eta2		0.64	0.69	0.22	1.06	0.96
hsd_manifestvar_Y1_Y1		0.64	0.65	0.16	0.67	0.96
hsd_manifestvar_Y2_Y2		0.64	0.63	0.16	0.66	0.93

Table C2

Simulation results for 50 subjects and 30 time points, with all parameters varying over subjects.

Parameter	Symbol	True value	Mean point est.	RMSE	CI width	Coverage
T0mean_eta1	$\boldsymbol{\eta}_{1[1]}$	1.00	1.02	0.20	0.84	0.98
T0mean_eta2	$\boldsymbol{\eta}_{1[2]}$	1.00	1.01	0.22	0.92	0.97
drift_eta1_eta1	$\mathbf{A}_{[1,1]}$	-0.40	-0.45	0.11	0.42	0.94
drift_eta2_eta1	$\mathbf{A}_{[1,2]}$	0.00	0.02	0.04	0.19	0.98
drift_eta1_eta2	$\mathbf{A}_{[2,1]}$	0.10	0.13	0.07	0.29	0.97
drift_eta2_eta2	$\mathbf{A}_{[2,2]}$	-0.20	-0.24	0.07	0.26	0.94
manifestvar_Y1_Y1	$\boldsymbol{\Theta}_{[1,1]}$	1.00	0.97	0.13	0.52	0.95
manifestvar_Y2_Y2	$\boldsymbol{\Theta}_{[2,2]}$	1.00	0.98	0.11	0.48	0.96
diffusion_eta1_eta1	$\mathbf{Q}_{[1,1]}$	1.00	1.05	0.18	0.73	0.96
diffusion_eta2_eta1	$\mathbf{Q}_{[2,1]}$	0.50	0.48	0.11	0.42	0.95
diffusion_eta2_eta2	$\mathbf{Q}_{[2,2]}$	1.00	1.04	0.15	0.60	0.95
T0var_eta1_eta1	$\mathbf{Q}^*_{1[1,1]}$	0.50	0.49	0.14	0.87	1.00
T0var_eta2_eta1	$\mathbf{Q}^*_{1[2,1]}$	0.10	0.18	0.34	1.86	1.00
T0var_eta2_eta2	$\mathbf{Q}^*_{1[2,2]}$	0.51	0.51	0.14	0.91	1.00
manifestmeans_Y1	$\boldsymbol{\tau}_{[1]}$	0.50	0.50	0.19	0.76	0.95
manifestmeans_Y2	$\boldsymbol{\tau}_{[2]}$	0.00	-0.00	0.21	0.84	0.94
hsd_manifestmeans_Y1		1.00	1.01	0.16	0.61	0.94
corr_manifestmeans_Y1_manifestmeans_Y2		0.50	0.40	0.17	0.56	0.90
hsd_manifestmeans_Y2		1.00	1.00	0.17	0.70	0.95
hsd_drift_eta1_eta1		0.15	0.18	0.11	0.44	0.96
hsd_drift_eta1_eta2		0.15	0.17	0.06	0.24	0.95
hsd_drift_eta2_eta1		0.01	0.05	0.04	0.14	0.95
hsd_drift_eta2_eta2		0.08	0.11	0.07	0.31	0.98
hsd_diffusion_eta1_eta1		0.64	0.70	0.17	0.69	0.94
hsd_diffusion_eta2_eta1		0.30	0.25	0.15	0.57	0.97
hsd_diffusion_eta2_eta2		0.64	0.70	0.15	0.60	0.93
hsd_manifestvar_Y1_Y1		0.64	0.68	0.12	0.52	0.94
hsd_manifestvar_Y2_Y2		0.64	0.68	0.12	0.51	0.95

Table C3

Simulation results for 200 subjects and 10 time points, with all parameters varying over subjects.

Parameter	Symbol	True value	Mean point est.	RMSE	CI width	Coverage
T0mean_eta1	$\boldsymbol{\eta}_{1[1]}$	1.00	1.02	0.16	0.67	0.96
T0mean_eta2	$\boldsymbol{\eta}_{1[2]}$	1.00	1.01	0.18	0.82	0.97
drift_eta1_eta1	$\mathbf{A}_{[1,1]}$	-0.40	-0.44	0.11	0.47	0.96
drift_eta2_eta1	$\mathbf{A}_{[1,2]}$	0.00	0.02	0.05	0.24	0.97
drift_eta1_eta2	$\mathbf{A}_{[2,1]}$	0.10	0.12	0.07	0.32	0.97
drift_eta2_eta2	$\mathbf{A}_{[2,2]}$	-0.20	-0.24	0.09	0.33	0.95
manifestvar_Y1_Y1	$\boldsymbol{\Theta}_{[1,1]}$	1.00	0.99	0.09	0.37	0.96
manifestvar_Y2_Y2	$\boldsymbol{\Theta}_{[2,2]}$	1.00	0.98	0.08	0.34	0.95
diffusion_eta1_eta1	$\mathbf{Q}_{[1,1]}$	1.00	1.02	0.15	0.65	0.96
diffusion_eta2_eta1	$\mathbf{Q}_{[2,1]}$	0.50	0.51	0.11	0.46	0.97
diffusion_eta2_eta2	$\mathbf{Q}_{[2,2]}$	1.00	1.03	0.14	0.54	0.95
T0var_eta1_eta1	$\mathbf{Q}^*_{1[1,1]}$	0.50	0.46	0.13	0.72	1.00
T0var_eta2_eta1	$\mathbf{Q}^*_{1[2,1]}$	0.10	0.28	0.38	1.81	0.99
T0var_eta2_eta2	$\mathbf{Q}^*_{1[2,2]}$	0.51	0.49	0.14	0.77	1.00
manifestmeans_Y1	$\boldsymbol{\tau}_{[1]}$	0.50	0.48	0.15	0.63	0.95
manifestmeans_Y2	$\boldsymbol{\tau}_{[2]}$	0.00	-0.01	0.18	0.78	0.96
hsd_manifestmeans_Y1		1.00	0.99	0.10	0.46	0.96
corr_manifestmeans_Y1_manifestmeans_Y2		0.50	0.39	0.14	0.53	0.92
hsd_manifestmeans_Y2		1.00	0.98	0.13	0.59	0.96
hsd_drift_eta1_eta1		0.15	0.17	0.10	0.46	0.98
hsd_drift_eta1_eta2		0.15	0.15	0.06	0.25	0.96
hsd_drift_eta2_eta1		0.01	0.07	0.06	0.17	0.94
hsd_drift_eta2_eta2		0.08	0.13	0.09	0.40	0.98
hsd_diffusion_eta1_eta1		0.64	0.65	0.12	0.47	0.95
hsd_diffusion_eta2_eta1		0.30	0.27	0.17	0.74	0.99
hsd_diffusion_eta2_eta2		0.64	0.66	0.09	0.40	0.96
hsd_manifestvar_Y1_Y1		0.64	0.65	0.07	0.28	0.93
hsd_manifestvar_Y2_Y2		0.64	0.65	0.07	0.27	0.94