Bayesian statistical inference for the hierarchical Ornstein-Uhlenbeck model: An online supplement to "A hierarchical latent stochastic differential equation model for affective dynamics"

Zita Oravecz, Francis Tuerlinckx, & Joachim Vandekerckhove
Department of Psychology
Department of Psychology
University of Leuven, Belgium

The hierarchical Ornstein-Uhlenbeck model

Modeling the affect dynamics of a single individual: The OU model with measurement error

The true or latent position in a two-dimensional latent space at time t will be denoted by the vector $\boldsymbol{\Theta}(t)$ defined as: $\boldsymbol{\Theta}(t) = (\Theta_1(t), \Theta_2(t))^{\mathrm{T}}$ and the superscript T indicates the transpose operation. In the core affect application, $\Theta_1(t)$ refers to the position on the first dimension (pleasantness) and $\Theta_2(t)$ to the position on the second dimension (arousal). (We will define the model for two dimensions here and we refer specifically to core affect, but generalizations to more dimensions and other application areas are possible.) In the model formulation, it is assumed that the true core affect changes continuously throughout time, but the measurements are taken at a finite number of time points: $t_1, t_2, \ldots, t_s, \ldots, t_n$, where n stands for the number of measurements. We define the vector $\boldsymbol{Y}(t_s) = (Y_1(t_s), Y_2(t_s))^{\mathrm{T}}$ as the observed pleasantness and arousal scores at time point t_s. The general model can then be written as follows:

$$\begin{cases} d\boldsymbol{\Theta}(t) &= \mathbf{B}\Big(\boldsymbol{\mu} - \boldsymbol{\Theta}(t)\Big)dt + \boldsymbol{\sigma} d\boldsymbol{W}(t) \\ \boldsymbol{Y}(t_s) &= \boldsymbol{\Theta}(t_s) + \boldsymbol{\epsilon}(t_s), \end{cases} \tag{1}$$

where $\boldsymbol{\mu}$ is a vector with two components, $\boldsymbol{\sigma}$ and $\mathbf{B}$ are positive definite 2×2 matrices. The measurement error is represented by $\boldsymbol{\epsilon}(t_s)$, which is a random draw from a bivariate normal distribution

with mean $(0,0)^{\mathrm{T}}$ and covariance matrix $\boldsymbol{\Sigma}_\epsilon$. The component $\boldsymbol{W}(t)$ stands for the standard Wiener process.

Let us first reparametrize the model such that $\boldsymbol{\Sigma}$ is replaced by $\boldsymbol{\Gamma}$, which has the following relation to $\boldsymbol{\Sigma}$ (see Gardiner, 2004):

$$\boldsymbol{\Sigma} = \mathbf{B}\boldsymbol{\Gamma} + \boldsymbol{\Gamma}\mathbf{B}^{\mathrm{T}}.$$

The matrix $\boldsymbol{\Gamma}$ is called the stationary covariance matrix. By integrating over the transition equation (Eq. 1, first line), we arrive at the conditional distribution of the two-dimensional OU process (see Appendix B of the paper for a derivation):

$$\boldsymbol{\Theta}(t)|\boldsymbol{\Theta}(t-d) \sim N_2\left(\boldsymbol{\mu} + e^{-\mathbf{B}d}\Big(\boldsymbol{\Theta}(t-d) - \boldsymbol{\mu}\Big),\ \boldsymbol{\Gamma} - e^{-\mathbf{B}d}\boldsymbol{\Gamma}e^{-\mathbf{B}^{\mathrm{T}}d}\right), \tag{2}$$

where N_2 refers to the bivariate normal distribution. As in the unidimensional case, the two-dimensional process converges to a stationary distribution:

$$\boldsymbol{\Theta}(t) \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Gamma}), \tag{3}$$

provided that all eigenvalues of $\mathbf{B}$ are positive.

The hierarchical OU model

The hierarchical extension requires the specification of population distributions. To proceed with the model description, assumptions about these distributions will be introduced now.

The hierarchical formulation of the OU model is based on the previously presented equations, but some new notation has to be introduced. A specific person p ($p = 1, ..., P$) is measured n_p times at the following sequence of time points: $t_{p1}, t_{p2}, \ldots, t_{ps}, \ldots, t_{p,n_p}$. The index s denotes the s^{th} measurement occasion of that individual. For notational convenience, we will use p and s as the only indicies when denoting parameters or data which are related to the specific observation at t_{ps}.

The model for a single person p for whom the observed data are a function of an underlying OU process and some measurement error can now be written as follows:

$$\boldsymbol{Y}_{ps} = \boldsymbol{\Theta}_{ps} + \boldsymbol{\epsilon}_{ps} \tag{4}$$

where $\boldsymbol{Y}_{ps}$ stands for the observed random vector, $\boldsymbol{\Theta}_{ps}$ for the latent state (or true score) and $\boldsymbol{\epsilon}_{ps}$ for the measurement error. The conditional distribution of $\boldsymbol{\Theta}_{ps}$ given $\boldsymbol{\Theta}_{p,s-1}$ is normally distributed as follows (for $s > 1$):

$$\boldsymbol{\Theta}_{ps}|\boldsymbol{\Theta}_{p,s-1} \sim N_2\left(\boldsymbol{\mu}_{ps} + e^{-\mathbf{B}_p(t_{ps}-t_{p,s-1})}(\boldsymbol{\Theta}_{p,s-1} - \boldsymbol{\mu}_{ps}), \boldsymbol{\Gamma}_p - e^{-\mathbf{B}_p(t_{ps}-t_{p,s-1})}\boldsymbol{\Gamma}_p e^{-\mathbf{B}_p^{\mathrm{T}}(t_{ps}-t_{p,s-1})}\right). \tag{5}$$

For the first observation, $\boldsymbol{\Theta}_{p1}$, it is assumed that $\boldsymbol{\Theta}_{p1} \sim N_2(\boldsymbol{\mu}_{ps}, \boldsymbol{\Gamma}_p)$.

The regression of $\boldsymbol{\mu}_{ps}$ onto the two types of covariates and allowing for a person-specific random deviation is defined as follows:

$$\boldsymbol{\mu}_{ps} = \boldsymbol{\Delta}_\mu \boldsymbol{z}_{ps} + \boldsymbol{A}_\mu \boldsymbol{x}_p + \boldsymbol{E}_{p\mu}, \tag{6}$$

with $\boldsymbol{E}_{p\mu} \sim N_2(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{\mu}})$. The matrices $\boldsymbol{\Delta}_\mu$ and $\boldsymbol{A}_\mu$ are parameter matrices of dimension $2 \times m$ and $2 \times (k+1)$, respectively, containing the regression weights for the covariates. Furthermore, the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\mu}}$ is defined as follows:

$$\boldsymbol{\Sigma}_{\boldsymbol{\mu}} = \begin{pmatrix} \sigma^2_{\mu_1} & \sigma_{\mu_1\mu_2} \\ \sigma_{\mu_1\mu_2} & \sigma^2_{\mu_2} \end{pmatrix} \tag{7}$$

For implementation purposes, it is helpful to use another formulation:

$$\boldsymbol{\mu}_{ps} = \big(\boldsymbol{I}_2 \otimes \boldsymbol{z}_{ps}\big)\boldsymbol{\delta} + \boldsymbol{A}_\mu \boldsymbol{x}_p + \boldsymbol{E}_{p\mu}, \tag{8}$$

with $\otimes$ denoting the Kronecker product. Vector $\boldsymbol{\delta} = \text{vec}(\boldsymbol{\Delta}_\mu)$ is a $(2m) \times 1$ vector of regression coefficients (the coefficient vectors of each dimension are stacked on top of each other).

As explained in the paper, also the other person-specific OU parameters (γ_{1p}, γ_{2p}, ρ_{γ_p}, β_{1p}, β_{2p} and ρ_{β_p}) can be made functions of time-invariant covariates. The population distribution for the other parameters is unidimensional. As an example, the population distribution for γ_{1p} is defined as

$$\gamma_{1p} \sim LN(\boldsymbol{x}_p^{\mathrm{T}}\boldsymbol{\alpha}_{\gamma_1}, \sigma^2_{\gamma_1}),$$

where the density function is:

$$f(\gamma_{1p}) = \frac{1}{\gamma_{1p}\sqrt{2\pi\sigma^2_{\gamma_1}}} e^{-\frac{1}{2}\frac{(\log(\gamma_{1p}) - \boldsymbol{x}_p^{\mathrm{T}}\boldsymbol{\alpha}_{\gamma_1})^2}{\sigma^2_{\gamma_1}}}. \tag{9}$$

The same properties are valid for γ_{2p}, β_{1p} and β_{2p}.

In the case of the standardized off-diagonal elements (ρ_{γ_p} and ρ_{β_p}), it is assumed that their Fisher-z transformed values $F(\rho_{\gamma_p}) = \frac{1}{2}\log\frac{1+\rho_{\gamma_p}}{1-\rho_{\gamma_p}}$ and $F(\rho_{\beta_p}) = \frac{1}{2}\log\frac{1+\rho_{\beta_p}}{1-\rho_{\beta_p}}$ are drawn from a normal population distribution, the mean of which mean depends on covariates. As an example, let us consider ρ_{γ_p} (ρ_{β_p} follows the same reasoning):

$$F(\rho_{\gamma_p}) \sim N(\boldsymbol{x}_p^{\mathrm{T}}\boldsymbol{\alpha}_{\rho_\gamma}, \sigma^2_{\rho_\gamma}).$$

The density of the original ρ_{γ_p} then equals (applying the transformation of variables technique, see e.g. Mood, Graybill, & Boes, 1974):

$$\begin{aligned} f(\rho_{\gamma_p}) =& \left| \frac{dF(\rho_{\gamma_p})}{d\rho_{\gamma_p}} \right| \phi(F(\rho_{\gamma_p}); \boldsymbol{x}_p^{\mathrm{T}} \boldsymbol{\alpha}_{\rho_\gamma}, \sigma^2_{\rho_\gamma}) \\ =& \frac{1}{(1-\rho_{\gamma_p})(1+\rho_{\gamma_p})} \frac{1}{\sqrt{2\pi\sigma^2_{\rho_\gamma}}} \\ & \times \exp\left(-\frac{1}{2} \frac{\left(\frac{1}{2}\log\left(\frac{1+\rho_{\gamma_p}}{1-\rho_{\gamma_p}}\right) - \boldsymbol{x}_p^{\mathrm{T}} \boldsymbol{\alpha}_{\rho_\gamma}\right)^2}{\sigma^2_{\rho_\gamma}} \right), \end{aligned} \tag{10}$$

where $F(\cdot)$ is the Fisher-z transform and $\phi(x; \mu, \sigma^2)$ is the normal density evaluated at x with mean μ and variance σ^2. Again, $\boldsymbol{\alpha}_{\rho_\gamma}$ contains $k+1$ regression coefficients.

Having completed the description of the model, we will summarize in the next section how the statistical inference is carried out.

Bayesian inference in the OU model

As explained in the paper, we opt for the Gibbs sampler whereby, in each iteration, a new value of each parameter is sampled, based on the full conditional distribution of the parameter in question (i.e., the probability distribution of the parameter, given the values of all other parameters as obtained in the previous iteration, as well as the data). The full conditional of all parameters have to be derived. If such a full conditional is a known distribution, then drawing a random sample from it is straightforward. If the full conditional does not correspond to a known distributional form, however, we will implement a Metropolis-Hastings (M-H) step (Gelman, Carlin, Stern, & Rubin, 2004, p. 291) in the Gibbs sampling structure.

As a first general step for describing the inference, we give the contribution of one person to the likelihood:

$$\begin{aligned} L_p &= L(\boldsymbol{\mu}_p, \mathbf{B}_p, \boldsymbol{\Gamma}_p, \boldsymbol{\Sigma}_\epsilon, \{\boldsymbol{\Theta}_{ps}\}_{s=1}^{n_p} \mid \{\boldsymbol{Y}_{ps}\}_{s=1}^{n_p}) \\ &= f(\{\boldsymbol{Y}_{ps}\}_{s=1}^{n_p} \mid \{\boldsymbol{\Theta}_{ps}\}_{s=1}^{n_p}, \boldsymbol{\Sigma}_\epsilon, \boldsymbol{\mu}_p, \boldsymbol{\delta}, \mathbf{B}_p, \boldsymbol{\Gamma}_p), \end{aligned}$$

where $\{\boldsymbol{Y}_{ps}\}_{s=1}^{n_p}$ and $\{\boldsymbol{\Theta}_{ps}\}_{s=1}^{n_p}$ stand for the indexed lists $\{\boldsymbol{Y}_{p1}, \boldsymbol{Y}_{p2}, \ldots, \boldsymbol{Y}_{p,n_p}\}$ and $\{\boldsymbol{\Theta}_{p1}, \boldsymbol{\Theta}_{p2}, \ldots, \boldsymbol{\Theta}_{p,n_p}\}$, respectively. Instead of presenting the posterior distribution hereby, we immediately go on with describing the full conditionals of each parameter. Note that we will represent $\boldsymbol{\mu}_p$ in these derivations as the person-specific home base (i.e., the random intercept) with a mean possibly different from zero.

The full conditional of the latent state $\boldsymbol{\Theta}_{ps}$

We use a single move sampler (as opposed to a multimove sampler) to draw each latent state value separately (Carter & Kohn, 1994). We choose a normal distribution as a prior for $\boldsymbol{\Theta}_{ps}$:

$$\boldsymbol{\Theta}_{ps} \sim N_2(\boldsymbol{\Theta}_0, \boldsymbol{\Phi}_{\boldsymbol{\Theta}}).$$

To achieve a relatively uninformative prior, we set the mean $\boldsymbol{\Theta}_0$ to $\mathbf{0}$ and choose a high variance ($\boldsymbol{\Phi}_{\boldsymbol{\Theta}} = 1000 \times \boldsymbol{I}_2$). If we combine its normal likelihood with this normal prior, the full conditional also has normal distribution.

For convenience, we introduce some new notations, which will stand for time differences. With $d_{p,s+1}$, we denote the time difference between $t_{p,s+1}$ and $t_{p,s}$. Also $d_{p,s}$ stands for the time difference between $t_{p,s}$ and $t_{p,s-1}$. Finally d_{p,n_p} equals $t_{p,n_p} - t_{p,n_p-1}$.

The values of the latent state $\boldsymbol{\Theta}_{ps}$ are drawn sequentially. Since for the first point the likelihood is a bit more simple (it is estimated with the stationary distribution), the full conditional of the first observation ($s = 1$) is different from the rest:

$$(\boldsymbol{\Theta}_{p1} \mid \boldsymbol{\mu}_{p1}, \boldsymbol{\Gamma}_p, \boldsymbol{\Sigma}_\epsilon, \boldsymbol{Y}_{p1}) \sim N_2(\boldsymbol{M}_{\boldsymbol{\Theta}_{p1}}, \boldsymbol{V}_{\boldsymbol{\Theta}_{p1}}),$$

with covariance matrix

$$\boldsymbol{V}_{\boldsymbol{\Theta}_{p1}} = (\boldsymbol{\Phi}_\Theta^{-1} + \boldsymbol{\Sigma}_\epsilon^{-1} + \boldsymbol{\Gamma}_p^{-1})^{-1}$$

and mean

$$\boldsymbol{M}_{\boldsymbol{\Theta}_{p1}} = \boldsymbol{V}_{\boldsymbol{\Theta}_1}(\boldsymbol{\Phi}_\Theta^{-1}\boldsymbol{\Theta}_0 + \boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{Y}_{p1} + \boldsymbol{\Gamma}_p^{-1}\boldsymbol{\mu}_{p1}).$$

As we can see, the full conditional of $\boldsymbol{\Theta}_{p1}$ only depends on $\boldsymbol{\mu}_{p1}$, $\boldsymbol{\Gamma}_p$, $\boldsymbol{\Sigma}_\epsilon$ and $\boldsymbol{Y}_{p1}$. On the other hand, for $s > 1$, $\boldsymbol{\Theta}_{ps}$ is conditional on the previous latent state value ($\boldsymbol{\Theta}_{p,s-1}$), the next latent state value ($\boldsymbol{\Theta}_{p,s+1}$), some of the model parameters ($\boldsymbol{\mu}_p$, $\boldsymbol{\delta}$, $\mathbf{B}_p$, $\boldsymbol{\Gamma}_p$ and $\boldsymbol{\Sigma}_\epsilon$), the corresponding data point ($\boldsymbol{Y}_{ps}$) and the time-varying covariate information ($\boldsymbol{z}_{ps}$) :

$$(\boldsymbol{\Theta}_{ps} \mid \boldsymbol{\mu}_{ps}, \mathbf{B}_p, \boldsymbol{\Gamma}_p, \boldsymbol{\Sigma}_\epsilon, \boldsymbol{Y}_{ps}, \boldsymbol{\Theta}_{p,s-1}, \boldsymbol{\Theta}_{p,s+1}) \sim N_2(\boldsymbol{M}_{\boldsymbol{\Theta}_{ps}}, \boldsymbol{V}_{\boldsymbol{\Theta}_{ps}}),$$

since

$$\boldsymbol{V}_{\boldsymbol{\Theta}_{ps}} = (\boldsymbol{\Phi}_\Theta^{-1} + \boldsymbol{\Sigma}_\epsilon^{-1} + \boldsymbol{V}_{ps}^{-1} + (e^{-\mathbf{B}_p d_{p,s+1}})^{\mathrm{T}} \boldsymbol{V}_{p,s+1}^{-1} (e^{-\mathbf{B}_p d_{p,s+1}}))^{-1}$$

and

$$\begin{aligned}\boldsymbol{M}_{\boldsymbol{\Theta}_{ps}} &= \boldsymbol{V}_{\boldsymbol{\Theta}_{ps}}(\boldsymbol{\Phi}_{\Theta}^{-1}\boldsymbol{\Theta}_0 + \boldsymbol{\Sigma}_{\epsilon_p}^{-1}\boldsymbol{Y}_{ps} + \boldsymbol{V}_{ps}^{-1}\boldsymbol{\mu}_{ps} \\ &\quad + \boldsymbol{V}_{ps}^{-1}e^{-\mathbf{B}_p d_{ps}}\boldsymbol{\Theta}_{p,s-1} - \boldsymbol{V}_{ps}^{-1}e^{-\mathbf{B}_p d_{ps}}\boldsymbol{\mu}_{ps} \\ &\quad + (e^{-\mathbf{B}_p d_{p,s+1}})^{\mathrm{T}}\boldsymbol{V}_{p,s+1}^{-1}\boldsymbol{\Theta}_{p,s+1} \\ &\quad - (e^{-\mathbf{B}_p d_{p,s+1}})^{\mathrm{T}}\boldsymbol{V}_{p,s+1}^{-1}\boldsymbol{\mu}_{p,s+1} \\ &\quad + (e^{-\mathbf{B}_p d_{p,s+1}})^{\mathrm{T}}\boldsymbol{V}_{p,s+1}^{-1}e^{-\mathbf{B}_p d_{p,s+1}}\boldsymbol{\mu}_{p,s+1}),\end{aligned}$$

where $\boldsymbol{V}_{ps}$ is defined as:

$$\boldsymbol{V}_{ps} = \begin{cases} \boldsymbol{\Gamma}_p & \text{if } s = 1 \\ \boldsymbol{\Gamma}_p - e^{-\mathbf{B}_p(t_{ps}-t_{p,s-1})}\boldsymbol{\Gamma}_p e^{-\mathbf{B}_p^{\mathrm{T}}(t_{ps}-t_{p,s-1})} & \text{if } s > 1. \end{cases} \tag{11}$$

. The definition of $\boldsymbol{V}_{p,s+1}$ is analogue (with the appropriate changes of indices).

Naturally, the likelihood of the last point $\boldsymbol{\Theta}_{p,n_p}$ can not depend on the next observation point $\boldsymbol{\Theta}_{p,s+1}$, therefore its posterior has a simpler form:

$$(\boldsymbol{\Theta}_{p,n_p} \mid \boldsymbol{\mu}_{ps}, \mathbf{B}_p, \boldsymbol{\Gamma}_p, \boldsymbol{\Sigma}_\epsilon, \boldsymbol{Y}_{p,n_p}) \sim N_2(\boldsymbol{M}_{\boldsymbol{\Theta}_{p,n_p}}, \boldsymbol{V}_{\boldsymbol{\Theta}_{p,n_p}}),$$

with covariance matrix

$$\boldsymbol{V}_{\boldsymbol{\Theta}_{p,n_p}} = (\boldsymbol{\Phi}_{\Theta}^{-1} + \boldsymbol{\Sigma}_{\epsilon}^{-1} + \boldsymbol{V}_{p,n_p}^{-1})^{-1}$$

and mean

$$\begin{aligned}\boldsymbol{M}_{\boldsymbol{\Theta}_{p,n_p}} &= \boldsymbol{V}_{\boldsymbol{\Theta}_{p,n_p}}(\boldsymbol{\Phi}_{\Theta}^{-1}\boldsymbol{\Theta}_0 + \boldsymbol{\Sigma}_{\epsilon_p}^{-1}\boldsymbol{Y}_{p,n_p} + \boldsymbol{V}_{p,n_p}^{-1}\boldsymbol{\mu}_{p,n_p} \\ &\quad + \boldsymbol{V}_{p,n_p}^{-1}e^{-\mathbf{B}_p d_{p,n_p}}\boldsymbol{\Theta}_{p,n_p-1} - \boldsymbol{V}_{p,n_p}^{-1}e^{-\mathbf{B}_p d_{p,n_p}}\boldsymbol{\mu}_{p,n_p}).\end{aligned}$$

The full conditional of the person-specific parameters

In the case of the two-dimensional parameter $\boldsymbol{\mu}_p$, its bivariate normal prior) combined with its bivariate normal likelihood results in a bivariate normal full conditional density:

$$(\boldsymbol{\mu}_p \mid \{\boldsymbol{\Theta}_{ps}\}_{s=1}^{n_p}, \mathbf{B}_p, \boldsymbol{\Gamma}_p, \boldsymbol{\delta}, \boldsymbol{\alpha}_{\boldsymbol{\mu}}, \boldsymbol{\Sigma}_{\boldsymbol{\mu}}) \sim N_2(\boldsymbol{\Omega}_p, \boldsymbol{\Phi}_p).$$

where

$$\begin{aligned}
\boldsymbol{\Phi}_p &= \Big(\boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{-1} + \boldsymbol{\Gamma}_p^{-1} + \sum_{s=2}^{n_p} \boldsymbol{V}_{ps}^{-1} - \sum_{s=2}^{n_p} \boldsymbol{V}_{ps}^{-1} e^{-\mathbf{B}_p d_{ps}} \\
&\quad - \sum_{s=2}^{n_p} (e^{-\mathbf{B}_p d_{ps}})^{\mathrm{T}} \boldsymbol{V}_{ps}^{-1} + \sum_{s=2}^{n_p} (e^{-\mathbf{B}_p d_{ps}})^{\mathrm{T}} \boldsymbol{V}_{ps}^{-1} e^{-\mathbf{B}_p d_{ps}} \Big)^{-1} \\
\boldsymbol{\Omega}_p &= \boldsymbol{\Phi}_p \Bigg(\boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{-1} \boldsymbol{z}_p^{\mathrm{T}} \boldsymbol{\alpha}_{\boldsymbol{\mu}} - \boldsymbol{\Gamma}_p^{-1} \boldsymbol{F}_{p1} \boldsymbol{\delta} + \boldsymbol{\Gamma}_p^{-1} \boldsymbol{\Theta}_{p1} + \sum_{s=2}^{n_p} \boldsymbol{V}_{ps}^{-1} \boldsymbol{\Theta}_{ps} \\
&\quad - \sum_{s=2}^{n_p} \boldsymbol{V}_{ps}^{-1} e^{-\mathbf{B}_p d_{ps}} \boldsymbol{\Theta}_{p,s-1} - \sum_{s=2}^{n_p} (e^{-\mathbf{B}_p d_{ps}})^{\mathrm{T}} \boldsymbol{V}_{ps}^{-1} \boldsymbol{\Theta}_{ps} \\
&\quad + \sum_{s=2}^{n_p} (e^{-\mathbf{B} d_{ps}})^{\mathrm{T}} \boldsymbol{V}_{ps}^{-1} e^{-\mathbf{B} d_{ps}} \boldsymbol{\Theta}_{p,s-1} - \sum_{s=2}^{n_p} \boldsymbol{V}_{ps}^{-1} \boldsymbol{F}_{ps} \boldsymbol{\delta} \\
&\quad + \sum_{s=2}^{n_p} (e^{-\mathbf{B}_p d_{ps}})^{\mathrm{T}} \boldsymbol{V}_{ps}^{-1} \boldsymbol{F}_{ps} \boldsymbol{\delta} \Bigg),
\end{aligned} \tag{12}$$

and we define $\boldsymbol{F}_{ps} = \boldsymbol{I}_2 \otimes \boldsymbol{z}_{ps} - e^{-\mathbf{B}_p d_{ps}} \boldsymbol{I}_2 \otimes \boldsymbol{z}_{ps}$, and $\boldsymbol{F}_{p1} = \boldsymbol{I}_2 \otimes \boldsymbol{z}_{p1}$.

The conditional distributions for the unidimensional person-specific parameters do not have a known form. Therefore for these parameters we give the product of the likelihood and the prior, which is proportional to the full conditional. The prior distributional assumptions for these parameters are represented by their population densities, which have been specified in Equations 9 and 10. In the derivation of the conditionals, we have to distinguish between two groups of parameters. In the case of γ_{1p}, γ_{2p} and $\rho_{\gamma p}$, the first observation of the chain must be taken into account, while for β_{1p}, β_{2p} and $\rho_{\beta p}$ it does not. As a result, their likelihoods differ slightly. Within these two groups, however, the likelihoods are identical. In the first group, we multiply over all observations:

$$\begin{aligned}
&f(\gamma_{1p} \mid \{\boldsymbol{\Theta}_{ps}\}_{s=1}^{n_p}, \boldsymbol{\mu}_p, \gamma_{2p}, \rho_{\gamma p}, \mathbf{B}_p) \\
&\qquad \propto f(\gamma_{1p}) \prod_{s=1}^{n_p} |\boldsymbol{V}_{ps}|^{-\frac{1}{2}} e^{-\frac{1}{2}\left(\boldsymbol{M}_{ps}^{\mathrm{T}} \boldsymbol{V}_{ps}^{-1} \boldsymbol{M}_{ps}\right)},
\end{aligned} \tag{13}$$

where $\boldsymbol{M}_{ps}$ is defined as:

$$\boldsymbol{M}_{ps} = \begin{cases} \boldsymbol{\mu}_{ps} & \text{if } s = 1 \\ \boldsymbol{\mu}_{ps} + e^{-\mathbf{B}_p (t_{ps} - t_{p,s-1})} (\boldsymbol{\Theta}_{p,s-1} - \boldsymbol{\mu}_{ps}) & \text{if } s > 1 \end{cases} \tag{14}$$

and $\boldsymbol{V}_{ps}$ are defined as in Equation 11. The same derivation can be done easily for γ_{2p} and ρ_γ as well.

For the second group, we start multiplying from the second observation ($s = 2$) onward:

$$f(\beta_{1p} \mid \{\boldsymbol{\Theta}_{ps}\}_{s=1}^{n_p}, \boldsymbol{\mu}_p, \boldsymbol{\Gamma}_p, \beta_{2p}, \rho_{\beta p}) \propto f(\beta_{1p}) \prod_{s=2}^{n_p} |\boldsymbol{V}_{ps}|^{-\frac{1}{2}} e^{-\frac{1}{2}\left(\boldsymbol{M}_{ps}^{\mathrm{T}} \boldsymbol{V}_{ps}^{-1} \boldsymbol{M}_{ps}\right)}. \tag{15}$$

The full conditional of the regression terms

First, we start with the regression coefficients of the time-varying covariates, which have a rather special design: the coefficients of the two dimensions are stacked below each other in a vector $\boldsymbol{\delta}$, as defined in Equation 8. As prior for $\boldsymbol{\delta}$, we choose a multivariate normal distribution:

$$\boldsymbol{\delta} \sim N_{2E}(\boldsymbol{M}_{\boldsymbol{\delta}_0}, \boldsymbol{V}_{\boldsymbol{\delta}_0}).$$

To achieve a relatively uninformative prior, we set $\boldsymbol{M}_{\boldsymbol{\delta}_0}$ to $\mathbf{0}$ and we choose a high variance ($\boldsymbol{V}_{\boldsymbol{\delta}_0} = 1000 \times \boldsymbol{I}_{2E}$). By combining the normal prior with the normal likelihood, we can derive the conditional distribution of $\boldsymbol{\delta}$, which is a multivariate normal distribution:

$$(\boldsymbol{\delta} \mid \{\boldsymbol{\Theta}_{ps}\}_{s=1}^{n_p}, \mathbf{B}_p, \boldsymbol{\Gamma}_p, \boldsymbol{M}_{\boldsymbol{\delta}_0}, \boldsymbol{V}_{\boldsymbol{\delta}_0}) \sim N_{2E}(\boldsymbol{M}_{\boldsymbol{\delta}}, \boldsymbol{V}_{\boldsymbol{\delta}}).$$

with the following parameters:

$$\begin{aligned}
\boldsymbol{V}_{\boldsymbol{\delta}} &= \left(\boldsymbol{V}_{\boldsymbol{\delta}_0}^{-1} + \sum_{p=1}^{P} \boldsymbol{F}_{p1}^{\mathrm{T}} \boldsymbol{\Gamma}_p^{-1} \boldsymbol{F}_{p1} + \sum_{p=1}^{P} \sum_{s=2}^{n_p} \boldsymbol{F}_{ps}^{\mathrm{T}} \boldsymbol{V}_{ps}^{-1} \boldsymbol{F}_{ps} \right)^{-1} \\
\boldsymbol{M}_{\boldsymbol{\delta}} &= \boldsymbol{V}_{\boldsymbol{\delta}} \Bigg(\boldsymbol{V}_{\boldsymbol{\delta}_0}^{-1} \boldsymbol{M}_{\boldsymbol{\delta}_0} + \sum_{p=1}^{P} \boldsymbol{F}_1^{\mathrm{T}} \boldsymbol{\Gamma}_p^{-1} \boldsymbol{\Theta}_{p1} - \sum_{p=1}^{P} \boldsymbol{F}_1^{\mathrm{T}} \boldsymbol{\Gamma}_p^{-1} \boldsymbol{\mu}_p \\
&\quad + \sum_{p=1}^{P} \sum_{s=2}^{n_p} \boldsymbol{F}_{ps}^{\mathrm{T}} \boldsymbol{V}_{ps}^{-1} \boldsymbol{\Theta}_{ps} - \sum_{p=1}^{P} \sum_{s=2}^{n_p} \boldsymbol{F}_{ps}^{\mathrm{T}} \boldsymbol{V}_{ps}^{-1} \boldsymbol{\mu}_p \\
&\quad - \sum_{p=1}^{P} \sum_{s=2}^{n_p} \boldsymbol{F}_{ps}^{\mathrm{T}} \boldsymbol{V}_{ps}^{-1} e^{-\mathbf{B}_p d_{ps}} \boldsymbol{\Theta}_{p,s-1} + \sum_{p=1}^{P} \sum_{s=2}^{n_p} \boldsymbol{F}_{ps}^{\mathrm{T}} \boldsymbol{V}_{ps}^{-1} e^{-\mathbf{B}_p d_{ps}} \boldsymbol{\mu}_p \Bigg),
\end{aligned}$$

where $\boldsymbol{F}_{ps}$ and $\boldsymbol{F}_{p1}$ are defined as for Equation 12.

Second, we deal with regression coefficients of the the time-invariant predictors. They can appear in all population distributions of the person-specific OU parameters, thereby allowing every dynamic model parameter to be turned into a random effect. Let us first specify the priors. A uniform distribution is used as a prior for the regression coefficients

$$f(\boldsymbol{\alpha}_h) \propto 1,$$

where h can be equal to μ_1, μ_2, γ_1, γ_2, ρ_γ, β_1, β_2 or ρ_β. For the residual variance parameters, we choose a uniform prior on $\log \sigma_g$ such that:

$$f(\sigma_h^2) \propto \sigma_h^{-2},$$

where h again equals one of the unidimensional parameters γ_1,γ_2, ρ_γ, β_1, β_2 or ρ_β. For the two-dimensional $\boldsymbol{\mu}$, a prior has to be set on its covariance matrix $\boldsymbol{\Sigma_\mu}$, which can be specified as:

$$f(\boldsymbol{\Sigma_\mu}) \propto \mid \boldsymbol{\Sigma_\mu} \mid^{-3/2},$$

which is the bivariate Jeffreys prior. All these priors can be considered non-informative.

For deriving the full conditional distributions of the regression terms for the person-specific parameters, we again separate $\boldsymbol{\mu}$ from the unidimensional parameters (γ_1, γ_2, ρ_γ, β_1, β_2 and ρ_β). The distributional forms for the regression terms of the latter ones are identical. Here we give the example of the full conditionals of $\boldsymbol{\alpha}_{\gamma_1}$ and $\sigma^2_{\gamma_1}$ (but γ_1 could be substituted by γ_2, ρ_γ, β_1, β_2 or ρ_β as well):

$$f(\boldsymbol{\alpha}_{\gamma_1} \mid \gamma_{11}, \ldots, \gamma_{1P}, \sigma^2_{\gamma_1}) \propto \exp\Big(-\frac{1}{2} (\boldsymbol{\alpha}_{\gamma_1} - \boldsymbol{X}\hat{\boldsymbol{\alpha}}_{\gamma_1})^{\mathrm{T}} \boldsymbol{V}_g^{-1} (\boldsymbol{\alpha}_{\gamma_1} - \boldsymbol{X}\hat{\boldsymbol{\alpha}}_{\gamma_1}) \Big), \tag{16}$$

where $\boldsymbol{X}$ is a $P \times (k+1)$ matrix defined by stacking the person-specific covariate vectors $\boldsymbol{x}_p^{\mathrm{T}}$ underneath each other. If we denote $\boldsymbol{g} = (\log(\gamma_{11}), \ldots, \log(\gamma_{1P}))^{\mathrm{T}}$ such that $\hat{\boldsymbol{\alpha}}_{\gamma_1} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{g}$ and $\boldsymbol{V}_g = \sigma^2_{\gamma_1}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}$, it can be seen that the full conditional of $\boldsymbol{\alpha}_{\gamma_1}$ is a normal density with mean $\boldsymbol{X}\hat{\boldsymbol{\alpha}}_{\gamma_1}$ and covariance matrix $\boldsymbol{V}_g$.

The full conditional for $\sigma^2_{\gamma_1}$, the residual variance of $\boldsymbol{\alpha}_{\gamma_1}$, has the following form:

$$f(\sigma^2_{\gamma_1} \mid \gamma_{11}, \ldots, \gamma_{1P}) \propto (\sigma^2_{\gamma_1})^{-\left(\frac{P-k-1}{2}+1\right)} e^{-\frac{(P-k-1)s^2}{2\sigma^2_{\gamma_1}}},$$

with

$$s^2 = \frac{1}{P-k-1}(\boldsymbol{g} - \boldsymbol{X}\hat{\boldsymbol{\alpha}}_{\gamma_1})^{\mathrm{T}}(\boldsymbol{g} - \boldsymbol{X}\hat{\boldsymbol{\alpha}}_{\gamma_1}),$$

which corresponds to a scaled inverse-χ^2 distribution with scale s^2 and degrees of freedom $P-k-1$ (see e.g. Gelman et al., 2004, Appendix A); and where $\boldsymbol{g}$, $\boldsymbol{X}$, and $\hat{\boldsymbol{\alpha}}_{\gamma_1}$ are defined in the same way as in Equation 16.

The full conditional of $\boldsymbol{\alpha_\mu}$ and its covariance matrix $\boldsymbol{\Sigma_\mu}$ are also known densities. Their derivation involves a bivariate regression problem. Here we just give the solution for the parameters in question (for a general step-by-step treatment of the problem, see Zellner, 1971). First, the data

have to be rearranged. We define the matrix $\boldsymbol{M}$ as the $P \times 2$ matrix of individual home base intercepts, that is $\boldsymbol{M} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_P)^{\mathrm{T}}$. Then the least squares regression coefficient matrix (of the regression of $\boldsymbol{M}$ on $\boldsymbol{X}$, where the latter is defined in Equation 16), equals $\hat{\boldsymbol{A}}_{\boldsymbol{\mu}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{M}$. Stacking columns of $\hat{\boldsymbol{A}}_{\boldsymbol{\mu}}$ below each other results in $\hat{\boldsymbol{\alpha}}_{\boldsymbol{\mu}} = (\hat{\boldsymbol{A}}_{\mu_1}^{\mathrm{T}}, \hat{\boldsymbol{A}}_{\mu_2}^{\mathrm{T}})^{\mathrm{T}}$. In the same vein, let us also define $\vec{\boldsymbol{\alpha}}_{\boldsymbol{\mu}} = (\boldsymbol{\alpha}_{\mu_1}^{\mathrm{T}}, \boldsymbol{\alpha}_{\mu_2}^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{S} = (\boldsymbol{M} - \boldsymbol{X}\hat{\boldsymbol{A}}_{\boldsymbol{\mu}})^{\mathrm{T}}(\boldsymbol{M} - \boldsymbol{X}\hat{\boldsymbol{A}}_{\boldsymbol{\mu}})$. The full conditional of $\boldsymbol{\Sigma}_{\boldsymbol{\mu}}$ then equals

$$f(\boldsymbol{\Sigma}_{\boldsymbol{\mu}} \mid \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_P) \propto |\boldsymbol{\Sigma}_{\boldsymbol{\mu}}|^{-v/2}\, e^{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{-1}\boldsymbol{S})},$$

where $v = P - k + 2$, and $\mathrm{tr}(\cdot)$ denotes the trace operator. This density corresponds to an inverse-Wishart distribution with $P - k + 2$ degrees of freedom and scale $\boldsymbol{S}$. In the estimation algorithm, $\boldsymbol{\Sigma}_{\boldsymbol{\mu}}$ has to be sampled first, and based on its value the matrix $\boldsymbol{\alpha}_{\boldsymbol{\mu}}$ can be sampled from the following conditional density

$$f(\boldsymbol{\alpha}_{\boldsymbol{\mu}} \mid \boldsymbol{\Sigma}_{\boldsymbol{\mu}}, \boldsymbol{\mu}_p, \boldsymbol{x}) \propto e^{-\frac{1}{2}(\vec{\boldsymbol{\alpha}}_{\boldsymbol{\mu}} - \hat{\boldsymbol{\alpha}}_{\boldsymbol{\mu}})^{\mathrm{T}}(\boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{-1} \otimes \boldsymbol{x}_p\boldsymbol{x}_p^{\mathrm{T}})(\vec{\boldsymbol{\alpha}}_{\boldsymbol{\mu}} - \hat{\boldsymbol{\alpha}}_{\boldsymbol{\mu}})},$$

which is a bivariate normal density with mean $\hat{\boldsymbol{\alpha}}_{\boldsymbol{\mu}}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\mu}} \otimes (\boldsymbol{x}_p\boldsymbol{x}_p^{\mathrm{T}})^{-1}$.

The full conditional of $\boldsymbol{\Sigma}_{\epsilon}$

The covariance matrix of the measurement error is constrained to be a diagonal matrix:

$$\boldsymbol{\Sigma}_{\epsilon} = \begin{bmatrix} \sigma_{1\epsilon}^2 & 0 \\ 0 & \sigma_{2\epsilon}^2 \end{bmatrix}.$$

We will sample $\sigma_{1\epsilon}^2$ and $\sigma_{2\epsilon}^2$ in identical ways. First, we assume prior independence for $\sigma_{1\epsilon}^2$ and $\sigma_{2\epsilon}^2$, and we demonstrate the inference with $\sigma_{1\epsilon}^2$. As a prior, we take a scaled inverse-χ^2 distribution:

$$\sigma_{1\epsilon}^2 \sim \text{inv-}\chi^2(\nu_0, s_0^2).$$

If we combine this with the likelihood of $\sigma_{1\epsilon}^2$, which is univariate normal, the result is another scaled inverse-χ^2 distribution:

$$\begin{gathered}(\sigma_{1\epsilon}^2 \mid \{\boldsymbol{Y}(t_{1s})\}_{s=1}^{n_1}, \ldots, \{\boldsymbol{Y}(t_{Ps})\}, \{\boldsymbol{\Theta}(t_{1s})\}_{s=1}^{n_1}, \ldots, \{\boldsymbol{\Theta}(t_{Ps})\}) \\ \sim \text{inv-}\chi^2(\nu_{\sigma_{1\epsilon}^2}, s_{\sigma_{1\epsilon}^2}),\end{gathered}$$

with parameters

$$\nu_{\sigma_{1\epsilon}^2} = \nu_0 + \sum_{p=1}^{P} n_p$$

and

$$s_{\sigma^2_{1\epsilon}} = (\nu_0 s_0^2 + s^2)/\nu_{\sigma^2_{1\epsilon}},$$

where

$$s^2 = \sum_{p=1}^{P} \sum_{s=1}^{n_p} \left(y_{ps1} - \theta_{ps1}\right)^2.$$

Sampling specifications

As has been discussed above, we use the Gibbs sampler for sampling from the full conditionals with a known form, which is the case for most of the parameters. Only for γ_{1p}, γ_{2p}, ρ_{γ_p}, β_{1p}, β_{2p} and ρ_{β_p} (the elements of $\mathbf{\Gamma}_p$ and $\mathbf{B}_p$), we need to implement a Metropolis-Hastings step. For this purpose, reasonable candidate generating distributions have to be assigned. With regard to the two matrices $\mathbf{\Gamma}_p$ and $\mathbf{B}_p$, there is also a constraint which has to be met. According to the theory of the OU process, the matrix product $\mathbf{B}_p\mathbf{\Gamma}_p + \mathbf{\Gamma}_p\mathbf{B}_p^{\mathrm{T}}$ has to be positive definite, since it represents the instantaneous covariance matrix (see e.g., Dunn & Gipson, 1977). Implementing such a restriction is a complex task, so most of the applications on the OU process have opted for an isotropic constraint on $\mathbf{B}_p$ with positive diagonal elements (Blackwell, 1997; Oravecz, Tuerlinckx, & Vandekerckhove, 2009), in which case the aforementioned criterion is always automatically fulfilled. However, our present model incorporates a more general representation of $\mathbf{B}_p$. Accordingly, a technique is needed to preserve the positive definiteness of the matrix product $\mathbf{B}_p\mathbf{\Gamma}_p + \mathbf{\Gamma}_p\mathbf{B}_p^{\mathrm{T}}$. Our method is the following. First of all, $\mathbf{B}_p$ and $\mathbf{\Gamma}_p$ matrices are decomposed into subelements, as has been shown in the paper. The idea is to sample each of the subelements (γ_{1p}, γ_{2p}, ρ_{γ_p}, β_{1p}, β_{2p} and ρ_{β_p}) subsequently in Metropolis-Hastings steps in such a way that their candidate generating distributions are constrained as a function of the previously accepted values of the other five subelements. We infer the form of the function by constraining the determinant of the result of the matrix-product $\mathbf{B}_p\mathbf{\Gamma}_p + \mathbf{\Gamma}_p\mathbf{B}_p^{\mathrm{T}}$ to be always positive (this is a sufficient condition for positive-definiteness in the two-dimensional case with positive diagonal elements, based on Sylvester's criterion). First $\mathbf{B}_p\mathbf{\Gamma}_p + \mathbf{\Gamma}_p\mathbf{B}_p^{\mathrm{T}}$ has to be solved based on the decomposed elements. This way the resulting 2-by-2 matrix consists of four elements which are sums and products of the six subelements in $\mathbf{B}_p$ and $\mathbf{\Gamma}_p$. Second, we calculate the determinant of this form and constrain it to be larger than 0, which results in a fourth order polynomial. Finally, solving this polynomial for the six different subelements gives us its roots. We can sufficiently constrain the candidate generating distributions based on these values. In practice, truncated normal distributions were

implemented with the previously accepted value as a mean and with a variance which ensured a reasonable acceptance ratio (around 0.44, see Gelman et al., 2004, p. 306).

A software program to sample from the joint posterior was written in MATLAB. Since the process is computationally demanding, some subroutines of the code have been written in C++, which then can be called from MATLAB in a straightforward way. Consequently, the computation time is reduced. To demonstrate the program, we present a simulation in the next section.

Simulation

To demonstrate the algorithm, we performed two simulation studies. In each study, we simulated ten datasets with 100 subjects and 100 observations per subject, according to the presented model assumptions. To imitate the properties of a dataset the model is most likely to be used for, we assumed that the data came from consecutive measurement during ten days, ten measurements per day at random intervals. The observation time points are arranged according to this design, and are calculated in minutes. As time-varying covariate information, we use the measurement times and their squared value in hours, this way accounting for linear and quadratic time-effects ($\delta_{L\mu_1}$, $\delta_{Q\mu_1}$, $\delta_{L\mu_2}$, $\delta_{Q\mu_2}$) in the home base.

The presented model was fitted to the simulated datasets by using the aforementioned MATLAB routine. For each parameter, three chains were run with different starting values to explore the full conditional densities. The different starting values are used to test whether the algorithm does not get stuck in a local maximum. The results are based on 30000 draws from each full conditional (10000 iterations per chain). These iterations were preceded by a discarded burn-in period (5000 iterations in each chain), avoiding the incorrect starting values to influence the final estimates. No matter what starting values are set, the chains have to converge around the same value. Convergence can be checked visually and mathematically as well. The mathematical way to check convergence is to calculate the $\hat{R}$ value (Gelman et al., 2004), which expresses the ratio of the between- and within chain variances. In the literature on Bayesian statistics, the chains are considered to have converged if the $\hat{R}$ value is below 1.1. We used the visual and the mathematical criteria as well to asses the convergence of the simulated chains, and we experienced no problems with it in the current analyses.

The computation time per chain was about 4 hours on a computing node with an AMD Opteron250 processor and 2Gb of RAM. Tables 1 and 2 summarize the most important results for the two simulation studies. In the first simulation study, we choose most of the parameters

in correspondence with a real life application, except for the measurement error of the second dimension ($\sigma^2_{2\epsilon}$), which was increased. In the second study, we increased the measurement error even further, and we slightly altered the other parameters as well. The main difference between the two studies is in the magnitude of the supposed measurement error ($\sigma^2_{1\epsilon}$ and $\sigma^2_{2\epsilon}$): the second simulation setting corresponds to a rather noisy dataset. In each table, the first column contains the notation of the selected variables of interest. The second column shows their true values as simulated. The third column is the averaged recovered posterior mean: the average of the posterior means over the simulation studies. The fourth column is the standard deviation of these posterior means.

The recovery in the first simulation study (Table 1) can be considered sufficient. However, there were minor issues in the second study (Table 2), where the parameter settings corresponded to a noisy dataset. As we can see, the algorithm does especially well at estimating the home bases and the related time-varying coefficients. With regard to the more problematic parameters, the estimate of α_{γ_1} seems somewhat biased in the second simulation. The value itself overestimates the true (simulated) value, and its variance $\sigma^2_{\gamma_1}$ is lower than expected. However, we have to remark on the fact that the second simulation study has rather a extreme setting on this parameter, in terms of comparing the level of this so-called stochastic variability to the level of the measurement error variance. The expected stochastic variance in the first dimension (based on the mean (α_{γ_1}) and the variance ($\sigma^2_{\gamma_1}$) of the lognormal distribution for γ_1) is around 3.5, which is lower than the variance of the noise ($\sigma^2_{1\epsilon} = 4.00$). Such a ratio might make it very difficult to correctly estimate these two parameters, as some trade-offs are likely. Clearly, the estimation procedure can still be improved in that respect. However, based on experience, we do not typically expect a very high level of measurement noise in substantive applications. In the second dimension, where the ratio of the measurement error ($\sigma^2_{2\epsilon}$) and the stochastic variability (α_{γ_2}) favors the latter, the estimation algorithm is more accurate.

Table 1:: Summary of the results of simulation study 1.

Model Parameter	Simulated value	Mean posterior estimate	SD of the posterior estimates
α_{μ_1}	6.00	5.99	0.04
α_{μ_2}	5.00	4.97	0.07
σ_{μ_1}	0.40	0.42	0.07
$\sigma_{\mu_1\mu_2}$	0.05	0.04	0.04
σ_{μ_2}	0.30	0.27	0.05
$\delta_{L\mu_1}$	1.00	1.00	0.12
$\delta_{Q\mu_1}$	0.00	−0.01	0.12
$\delta_{L\mu_2}$	4.00	3.96	0.12
$\delta_{Q\mu_2}$	−4.00	−4.00	0.14
α_{γ_1}	0.80	0.77	0.07
$\sigma^2_{\gamma_1}$	0.40	0.38	0.04
α_{γ_2}	1.00	1.01	0.10
$\sigma^2_{\gamma_2}$	0.20	0.20	0.03
α_{γ_ρ}	0.01	0.09	0.03
$\sigma^2_{\gamma_\rho}$	0.10	0.08	0.01
α_{β_1}	−4.20	−4.21	0.04
$\sigma^2_{\beta_1}$	0.40	0.39	0.10
α_{β_2}	−4.00	−3.93	0.09
$\sigma^2_{\beta_2}$	0.50	0.49	0.18
α_{β_ρ}	−0.10	−0.11	0.03
$\sigma^2_{\beta_\rho}$	0.10	0.08	0.02
$\sigma^2_{1\epsilon}$	0.20	0.21	0.04
$\sigma^2_{2\epsilon}$	1.00	0.88	0.13

Table 2:: Summary of the results of simulation study 2, with a higher level of measurement error

Model Parameter	Simulated value	Mean posterior estimate	SD of the posterior estimates
α_{μ_1}	0.00	−0.04	0.12
α_{μ_2}	0.00	−0.01	0.14
σ_{μ_1}	2.00	2.14	0.32
$\sigma_{\mu_1\mu_2}$	0.70	0.71	0.23
σ_{μ_2}	1.00	0.93	0.17
$\delta_{L\mu_1}$	2.00	1.86	0.22
$\delta_{Q\mu_1}$	0.00	0.10	0.22
$\delta_{L\mu_2}$	0.00	−0.01	0.26
$\delta_{Q\mu_2}$	4.00	3.99	0.26
α_{γ_1}	1.00	1.32	0.11
$\sigma^2_{\gamma_1}$	0.50	0.28	0.05
α_{γ_2}	2.00	1.95	0.10
$\sigma^2_{\gamma_2}$	0.10	0.10	0.02
α_{γ_ρ}	0.50	0.41	0.05
$\sigma^2_{\gamma_\rho}$	0.10	0.05	0.01
α_{β_1}	−4.00	−3.75	0.24
$\sigma^2_{\beta_1}$	0.50	0.40	0.26
α_{β_2}	−3.50	−3.44	0.20
$\sigma^2_{\beta_2}$	0.10	0.18	0.13
α_{β_ρ}	0.50	0.38	0.07
$\sigma^2_{\beta_\rho}$	0.10	0.06	0.03
$\sigma^2_{1\epsilon}$	4.00	3.20	0.24
$\sigma^2_{2\epsilon}$	2.00	2.13	0.73

References

Blackwell, P. G. (1997). Random diffusion models for animal movements. *Ecological Modeling*, *100*, 87–102.

Carter, C. K., & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, *81*, 541–553.

Dunn, J. E., & Gipson, P. S. (1977). Analysis of radio telemetry data in studies of home range. *Biometrics*, *33*, 85–101.

Gardiner, C. W. (2004). *Handbook of stochastic methods*. New York: Springer-Verlag.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004). *Bayesian data analysis*. New York: Chapman & Hall.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics*. New York: McGraw-Hill.

Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2009). A hierarchical Ornstein-Uhlenbeck model for continuous repeated measurement data. *Psychometrika*, *74*, 395–418.

Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. New York: Wiley.