Online Supplement: Time-Evolving Psychological Processes Over Repeated Decisions

David Gunawan[1,5], Guy E. Hawkins[2], Robert Kohn[3,5], Minh-Ngoc Tran[4], Scott D. Brown[2]

1: School of Mathematics and Applied Statistics, University of Wollongong
2: School of Psychology, University of Newcastle, Australia
3: Australian School of Business, University of New South Wales, Sydney, Australia
4: Discipline of Business Analytics, University of Sydney Business School
5: Australian Research Council Center of Excellence in Mathematical and Statistical Frontiers

Bayesian Estimation Methods for AR and Trend Models

This section develops Bayesian estimation methods for estimating the dynamic AR and trend LBA models based on recent advances in particle MCMC. We first discuss the notation used in the particle MCMC sampling schemes. Let $\theta \in \Theta \subset \mathbb{R}^{D_\theta}$ be the vector of unknown model parameters and $p(\theta)$ the prior distribution over θ; $\mathbb{R}^d$ means d dimensional Euclidean space. Let $y_{j,t}$ be the vector of observations for the j^{th} subject at the t^{th} block, and define $y_j := (y_{j,1}, ..., y_{j,T})$ as the vector of all observations for subject j and $y := y_{1:S,1:T} = (y_{1,1:T}, ..., y_{S,1:T})$ as the vector of observations for all S subjects. Let $\alpha_{j,t} \in \chi_\alpha \subset \mathbb{R}^{d_\alpha}$ be the vector of parameter values for subject j during time period t. The subject-specific parameters α are called "random effects", and the time periods t are called "blocks". Let $\alpha_j := (\alpha_{j,1}, ..., \alpha_{j,T})$ be all the random effects for subject j and $\alpha_{1:S,1:T} := (\alpha_{1,1:T}, ..., \alpha_{S,1:T})$ the vector of random effects for all S subjects. For the AR model:

$$p(\alpha_{1:S,1:T}|\theta) = \prod_{j=1}^{S} p(\alpha_{j,1}|\theta) \prod_{t=2}^{T} p(\alpha_{j,t}|\alpha_{j,t-1}, \theta)$$

For the trend model:

$$p(\alpha_{1:S,1:T}|\theta) = \prod_{j=1}^{S} \prod_{t=1}^{T} p(\alpha_{j,t}|\theta)$$

Brown and Heathcote (2008) and Terry et al. (2015) give expressions for the density of $p(y_{1:S,1:T}|\alpha_{1:S,1:T}, \theta)$. Our goal is to sample from the posterior density

$$\pi(\theta, \alpha_{1:S,1:T}) := p(y_{1:S,1:T}|\alpha_{1:S,1:T}, \theta)\, p(\alpha_{1:S,1:T}|\theta)\, p(\theta) / p(y_{1:S,1:T}), \tag{1}$$

with marginal likelihood

$$p(y_{1:S,1:T}) = \int \int p(y_{1:S,1:T}|\alpha_{1:S,1:T}, \theta)\, p(\alpha_{1:S,1:T}|\theta)\, p(\theta)\, d\theta d\alpha_{1:S,1:T} \tag{2}$$

In addition to sampling from the posterior density (for parameter inference), estimating the marginal likelihood itself is used for model selection via Bayes factors.

We develop a sampling algorithm using particle Markov chain Monte-Carlo, based on the methods from Andrieu, Doucet, and Holenstein (2010). The core idea is to define a target distribution on an augmented space that includes all the parameters of the model (and random effects) as well as the random variables generated by Monte Carlo sampling, and such that this augmented distribution has as its marginal distribution the joint posterior of the parameters and the random effects. Using this target distributions, we derive the particle Metropolis within Gibbs algorithms for the trend and the AR models. One of the attractions of the proposed methods is that essentially the same code and algorithms, with some modifications, can be used to estimate a variety of time-varying random effects models. All that is needed is to code up a new version of the prior distribution of the random effects, as well as the code to generate the group-level parameters. Any Metropolis within Gibbs steps used to generate the group-level parameters are essentially just the standard MCMC updates conditional on the random effects. The estimation of the AR and trend models is discussed in detail below. We also discuss how to modify some parts of the algorithms to estimate a wide range of time-varying random effects models. In addition, it is straightforward to estimate other dynamic evidence accumulation models (EAMs), for example the diffusion model (Ratcliff, Smith, Brown, & McKoon, 2016). All that is needed is to change the density $p\left(y_{1:S,1:T}|\alpha_{1:S,1:T},\theta\right)$.

The rest of the supplement describes the target distributions for the trend and AR models. The novel particle Metropolis-within-Gibbs sampler is then discussed. We also extend the "Importance Sampling Squared" (IS^2; Tran et al., 2021) algorithm for estimating the marginal likelihoods for dynamic LBA models. We suggest that the non-technical reader skips the sections describing the target distributions for different time-varying LBA models.

Target Distribution for the trend model

This section discusses the augmented target distribution for the polynomial trend model. Let $\{m_{j,t}\left(\alpha_{j,t}|\theta, y_{j,t}\right); j=1,...,S, t=1,...,T\}$ be a family of proposal densities that are used to approximate the conditional posterior densities $\{\pi\left(\alpha_{j,t}|\theta\right); j=1,...,S, t=1,...,T\}$. Let $\alpha_{j,t}^{1:R}=\left(\alpha_{j,t}^{1},...,\alpha_{j,t}^{R}\right)$ refer to all the particles for subject j generated by a standard Monte Carlo algorithm at block t. The joint density of the particles given the parameters for subject j is defined as

$$\psi_j\left(\alpha_{j,1:T}^{1:R}|\theta\right):=\prod_{t=1}^{T}\prod_{r=1}^{R}m_{j,t}\left(\alpha_{j,t}^{r}|\theta\right); \tag{3}$$

hence the joint density of the particles given the parameters for all subjects is

$$\psi\left(\alpha_{1:S,1:T}^{1:R}|\theta\right)=\prod_{j=1}^{S}\psi_j\left(\alpha_{j,1:T}^{1:R}|\theta\right). \tag{4}$$

To define the required augmented target densities, let $k=(k_1,...,k_S)$ and $k_j=(k_{j,1},...,k_{j,T})$, with each of $k_{j,t}\in\{1,...,R\}$, $\alpha_{1:S}^{k_{1:S}}=\left(\alpha_1^{k_1},....,\alpha_S^{k_S}\right)$ be a vector of all selected individual random effects with $\alpha_j^{k_j}=\left(\alpha_{j,1}^{k_{j,1}},...,\alpha_{j,T}^{k_{j,T}}\right)$ and

$\alpha_{1:S}^{(-k_{1:S})} = \left(\alpha_1^{(-k_1)}, ..., \alpha_S^{(-k_S)}\right)$
with $\alpha_j^{(-k_j)} = \left(\alpha_{j,1}^{(-k_{j,1})}, ..., \alpha_{j,T}^{(-k_{j,T})}\right)$ and $\alpha_{j,t}^{(-k_{j,t})} = \left(\alpha_{j,t}^1, ..., \alpha_{j,t}^{k_{j,t-1}}, \alpha_{j,t}^{k_{j,t+1}}, ..., \alpha_{j,t}^R\right)$.

Consider the augmented target density

$$\widetilde{\pi}_R\left(\alpha_{1:S,1:T}^{1:R}, k_{1:S,1:T}, \theta\right) = \frac{\pi\left(\alpha_{1:S,1:T}^{k_{1:S,1:T}}, \theta\right)}{R^{ST}} \prod_{j=1}^{S} \frac{\psi_j\left(\alpha_{j,1:T}^{1:R}|\theta\right)}{\prod_{t=1}^{T} m_{j,t}\left(\alpha_{j,t}^{k_{j,t}}|\theta\right)}. \tag{5}$$

Proposition 1. *The target distribution of Equation 5 has the marginal distribution*

$$\widetilde{\pi}_R\left(\alpha_{1:S,1:T}^{k_{1:S,1:T}}, k_{1:S,1:T}, \theta\right) = \frac{\pi\left(\alpha_{1:S,1:T}^{k_{1:S,1:T}}, \theta\right)}{R^{ST}}; \tag{6}$$

and hence, with some abuse of the notation, we can write $\widetilde{\pi}_R\left(\alpha_{1:S,1:T}, \theta\right) = p\left(\alpha_{1:S,1:T}, \theta|y_{1:S,1:T}\right)$. The proof follows from Andrieu et al. (2010) and Gunawan, Carter, Fiebig, and Kohn (2017).

Target Distributions for the AR model

We first briefly describes the generic sequential Monte Carlo (SMC) methods used to approximate the filtering densities $p\left(\alpha_{j,t}|y_{j,1:t}, \theta\right)$ for $j = 1, ..., S$ and $t = 1, ..., T$. The SMC algorithm recursively produces a set of weighted particles $\left\{\alpha_{j,t}^{(r)}, W_{j,t}^{(r)}\right\}_{r=1}^{R}$ at the t^{th} time period such that

$$\widehat{p}\left(\alpha_{j,t}|y_{j,1:t}, \theta\right) = \sum_{r=1}^{R} W_{j,t}^{(r)} \delta_{\alpha_{j,t}^{(r)}}\left(d\alpha_{j,t}\right) \tag{7}$$

approximates the density

$$p\left(\alpha_{j,t}|y_{j,1:t}, \theta\right) \propto \int p\left(y_{j,t}|\alpha_{j,t}, \theta\right) p\left(\alpha_{j,t}|\alpha_{j,t-1}, \theta\right) p\left(\alpha_{j,t-1}|y_{j,1:t-1}, \theta\right) d\alpha_{j,t-1}; \tag{8}$$

$\delta_a\left(d\alpha\right)$ is the Dirac delta distribution located at a. Equation (8) is used to obtain the particles $\left\{\alpha_{j,t}^{(r)}, W_{j,t}^{(r)}\right\}_{r=1}^{R}$ by first drawing the particles from a proposal distribution $m_{j,t}\left(\alpha_{j,t}|\alpha_{j,t-1}, \theta\right)$ and then computing the weights

$$w_{j,t}^{(r)} = W_{j,t-1}^{(r)} \frac{p\left(y_{j,t}|\alpha_{j,t}^{(r)}, \theta\right) p\left(\alpha_{j,t}^{(r)}|\alpha_{j,t-1}^{(r)}, \theta\right)}{m_{j,t}\left(\alpha_{j,t}^{(r)}|\alpha_{j,t-1}^{(r)}, \theta\right)}$$

to account for the difference between the posterior density and the proposal density. The weights are then normalized as $W_{j,t}^{(r)} = w_{j,t}^{(r)} / \sum_{s=1}^{R} w_{j,t}^{(s)}$. We implement the SMC algorithm for $t = 2, ..., T$ by using a multinomial resampling scheme, denoted as $\mathcal{M}\left(a_{j,t-1}^{1:R}|W_{j,t-1}^{1:R}\right)$. The argument $a_{j,t-1}^r$ means that $\alpha_{j,t-1}^{a_{j,t-1}^r}$ is the ancestor of $\alpha_{j,t}^r$. Algorithm 1 shows the generic SMC algorithm.

Algorithm 1 Generic Sequential Monte Carlo Algorithm

Inputs: $y_{j,1:T}$, R, θ
Outputs: $\alpha_{j,1:T}^{1:R}$, $a_{j,1:T-1}^{1:R}$,$w_{j,1:T}^{1:R}$

1. For $t = 1$

   (a) Sample $\alpha_{j,1}^{r}$ from $m_{j,1}\left(\alpha_{j,1}|y_{j,1},\theta\right)$, for $r = 1,...,R$

   (b) Calculate the importance weights

   $$w_{j,1}^{r} = \frac{p\left(y_{j,1}|\alpha_{j,1}^{r},\theta\right)p\left(\alpha_{j,1}^{r}|\theta\right)}{m_{j,1}\left(\alpha_{j,1}^{r}|y_{j,1},\theta\right)}, r = 1,...,R.$$

   and normalize those to obtain $W_{j,1}^{1:R}$.

2. For $t > 1$

   (a) Sample the ancestral indices $a_{j,t-1}^{1:R} \sim \mathcal{M}\left(a_{j,t-1}^{1:R}|W_{j,t-1}^{1:R}\right)$.

   (b) Sample $\alpha_{j,t}^{r}$ from $m_{j,t}\left(\alpha_{j,t}|\alpha_{j,t-1}^{a_{j,t-1}^{r}},\theta\right)$, $r = 1,...,R$.

   (c) Calculate the importance weights

   $$w_{j,t}^{r} = \frac{p\left(y_{j,t}|\alpha_{j,t}^{r},\theta\right)p\left(\alpha_{j,t}^{r}|\alpha_{j,t-1}^{a_{j,t-1}^{r}},\theta\right)}{m_{j,t}\left(\alpha_{j,t}^{r}|\alpha_{j,t-1}^{a_{j,t-1}^{r}},\theta\right)}, r = 1,...,R.$$

   and normalize those to obtain $W_{j,t}^{1:R}$.

Let $\alpha_{j,t}^{1:R} = \left(\alpha_{j,t}^{1},...,\alpha_{j,t}^{R}\right)$ and $a_{j,t-1}^{1:R} = \left(a_{j,t-1}^{1},...,a_{j,t-1}^{R}\right)$ refer to all the particles and ancestor indices for subject j, respectively, generated by the SMC algorithm at block t. The joint density of the particles given parameters for subject j is

$$\psi_{j}\left(\alpha_{j,1:T}^{1:R},a_{j,1:T-1}^{1:R}|\theta\right) = \prod_{r=1}^{R} m_{j,1}\left(\alpha_{j,1}^{r}|\theta\right)\prod_{t=2}^{T}\prod_{r=1}^{R} W_{j,t-1}^{a_{j,t-1}^{r}} m_{j,t}\left(\alpha_{j,t}^{r}|\alpha_{j,t-1}^{a_{j,t-1}^{r}},\theta\right),$$

and the joint density of the particles given the parameters for all subjects is

$$\psi\left(\alpha_{1:S,1:T}^{1:R},a_{1:S,1:T-1}^{1:R}|\theta\right) = \prod_{j=1}^{S}\psi_{j}\left(\alpha_{j,1:T}^{1:R},a_{j,1:T-1}^{1:R}|\theta\right).$$

Let $\alpha_{j,1:T}^{k_{j,1:T}} = \left(\alpha_{j,1}^{k_{j,1}},...,\alpha_{j,T}^{k_{j,T}}\right)$ be the selected reference trajectory for subject j with associated indices $k_{j,1:T}$. Let $\alpha_{j,1:T}^{-k_{j,1:T}} = \left(\alpha_{j,1}^{-k_{j,1}},...,\alpha_{j,T}^{-k_{j,T}}\right)$ denote the collection of all particles for subject j, except the selected reference trajectory, $\alpha_{j,1:T}^{k_{j,1:T}}$. Let $\alpha_{1:S,1:T}^{k_{1:S,1:T}} =$

$\left(\alpha_{1,1:T}^{k_{1,1:T}}, ..., \alpha_{S,1:T}^{k_{S,1:T}}\right)$ be the collection of the selected reference particles for all S subjects with associated indices $k_{1:S,1:T} = (k_{1,1:T}, ..., k_{S,1:T})$.

Proposition 2. *The target distribution is*

$$\widetilde{\pi}^R\left(\alpha_{1:S,1:T}^{1:R}, a_{1:S,1:T-1}^{1:R}, k_{1:S,1:T}, \theta\right) = \frac{\pi\left(\alpha_{1:S,1:T}^{k_{1:S,1:T}}, \theta\right)}{R^{ST}} \tag{9}$$

$$\prod_{j=1}^{S} \frac{\psi_j\left(\alpha_{j,1:T}^{1:R}, a_{j,1:T-1}^{1:R}|\theta\right)}{m_{j,1}\left(\alpha_{j,1}^{k_{j,1}}|\theta\right)\prod_{t=2}^{T} W_{j,t-1}^{a_{j,t-1}^{k_{j,t}}} m_{j,t}\left(\alpha_{j,t}^{k_{j,t}}|\alpha_{j,t-1}^{a_{j,t-1}^{k_{j,t}}}, \theta\right)}. \tag{10}$$

This target distribution has the marginal distribution

$$\widetilde{\pi}^R\left(\alpha_{1:S,1:T}^{k_{1:S,1:T}}, k_{1:S,1:T}, \theta\right) = \frac{\pi\left(\alpha_{1:S,1:T}^{k_{1:S,1:T}}, \theta\right)}{R^{ST}}, \tag{11}$$

and hence again, with some abuse of notation, we write $\widetilde{\pi}^R\left(\alpha_{1:S,1:T}, \theta\right) = p\left(\alpha_{1:S,1:T}, \theta|y_{1:S,1:T}\right)$. The proof follows from Andrieu et al. (2010). The same target distributions apply to a variety of other time-varying models, such as random walk, random walk with drift, and AR plus trend models. The next section describes the proposed particle Metropolis within Gibbs sampling scheme.

Particle Metropolis within Gibbs Sampling Scheme

Using the target distributions in Equations (5) and (10), we now derive the particle Metropolis within Gibbs algorithms for the trend and AR models. For each component of the sampling scheme, we also explain how to generalise it to other time-varying random effects models. Let $\theta = (\theta_1, ..., \theta_H)$ be a partition of the parameter vector into H components, where each component may be a vector. Algorithm 2 is the general PMwG sampling scheme for dynamic LBA models. The algorithm consists of two main steps: (1) Sampling the group-level parameters conditional on individual level parameters, (2) sampling the individual level parameters conditional on the group-level parameters.

Algorithm 2 Particle Metropolis within Gibbs for Dynamic LBA models

1. For $h = 1, ..., H$
   (a) Sample θ_h^* from the proposal $q_h\left(\cdot|\alpha_{1:S,1:T}^{k_{1:S,1:T}}, k_{1:S,1:T}, \theta^{(-h)}\right)$,
   (b) Accept the proposed values θ_h^* with probability
$$\min\left\{1, \frac{\widetilde{\pi}^R\left(\theta_h^*|\alpha_{1:S,1:T}^{k_{1:S,1:T}}, k_{1:S,1:T}, \theta^{(-h)}\right)}{\widetilde{\pi}^R\left(\theta_h|\alpha_{1:S,1:T}^{k_{1:S,1:T}}, k_{1:S,1:T}, \theta^{(-h)}\right)} \frac{q_h\left(\theta_h|\alpha_{1:S,1:T}^{k_{1:S,1:T}}, k_{1:S,1:T}, \theta^{(-h)}, \theta_h^*\right)}{q_h\left(\theta_h^*|\alpha_{1:S,1:T}^{k_{1:S,1:T}}, k_{1:S,1:T}, \theta^{(-h)}, \theta_h\right)}\right\},$$
2. For $j = 1, ..., S$
   (a) For the trend model
      i. Sample $\left(\alpha_{j,1:T}^{-k_{j,1:T}}\right) \sim \widetilde{\pi}^R\left(\alpha_{j,1:T}^{-k_{j,1:T}}|\theta, \alpha_{j,1:T}^{k_{j,1:T}}, k_{j,1:T}\right)$ using the conditional Monte Carlo algorithm given in Algorithm 3.
      ii. Sample the index $k_{j,1:T}$ with probability given by
$$\widetilde{\pi}_R\left(k_{j,1} = l_1, ..., k_{j,T} = l_T|\theta, \alpha_{j,1:T}^{1:R}\right) = \prod_{t=1}^{T} W_{j,t}^{l_t}.$$
   (b) For the AR model
      i. Sample $\left(\alpha_{j,1:T}^{-k_{j,1:T}}, a_{j,1:T-1}^{-k_{j,2:T}}\right) \sim \widetilde{\pi}^R\left(\alpha_{j,1:T}^{-k_{j,1:T}}, a_{j,1:T-1}^{-k_{j,2:T}}|\theta, \alpha_{j,1:T}^{k_{j,1:T}}, k_{j,1:T}\right)$ using the conditional sequential Monte Carlo given in Algorithm 4.
      ii. Sample $k_{j,T} = m \sim \widetilde{\pi}^R\left(k_{j,T}|\alpha_{j,1:T}^{1:R}, a_{1:T-1}^{1:R}\right)$, where $\widetilde{\pi}^R\left(k_{j,T}|\alpha_{j,1:T}^{1:R}, a_{1:T-1}^{1:R}\right) \propto W_{j,T}^m$.
      iii. Sample $\left(\alpha_{j,t}^{k_{j,t}}, a_{j,t-1}^{k_{j,t}}\right) \sim \widetilde{\pi}^R\left(\alpha_{j,t}^{k_{j,t}}, a_{j,t-1}^{k_{j,t}}|\theta, a_{j,1:t-2}^{1:R}, \alpha_{j,1:t-1}^{1:R}, a_{j,t:T-1}^{k_{j,t+1:T}}, \alpha_{j,t+1:T}^{k_{j,t+1:T}}, k_{j,T}\right)$ for $t = 1, ..., T$ using Bunch, Lindsten, and Singh (2015) Algorithm

Discussion of Algorithm 2. Step 1 of Algorithm 2 samples the group-level parameters θ using standard Gibbs or Metropolis within Gibbs steps. These steps are model dependent and need to be derived for each model. We now show the steps to sample the group-level parameters.

Step 1: Sampling the parameters θ for the AR model

To sample the autoregressive coefficient ϕ from $\widetilde{\pi}^R\left(\cdot|\alpha_{1:S,1:T}^{k_{1:S,1:T}}, k_{1:S,1:T}, \theta^{(-\phi)}\right)$, we draw a proposed value ϕ^* from $\mathcal{N}\left(\mu_\phi, \sigma_\phi^2\right)$ truncated within $(-1, 1)$, where

$$\sigma_\phi^2 = \left(\sum_{j=1}^{S}\sum_{t=2}^{T_j} \text{diag}\left(\alpha_{j,t-1} - \mu\right)' \Sigma^{-1} \text{diag}\left(\alpha_{j,t-1} - \mu\right)\right)^{-1},$$

and

$$\mu_\phi = \sigma_\phi^2 \left(\sum_{j=1}^{S} \sum_{t=2}^{T_j} \text{diag}\left(\alpha_{j,t-1} - \mu\right)' \Sigma^{-1} \left(\alpha_{j,t} - \mu\right) \right).$$

Here, and below, $\text{diag}(x)$ is a diagonal matrix with diagonal elements equal to the vector x. The candidate is accepted with probability

$$\min\left(1, \frac{I\left(-1 < \phi_1^* < 1\right) \times ... \times I\left(-1 < \phi_D^* < 1\right)}{I\left(-1 < \phi_1 < 1\right) \times ... \times I\left(-1 < \phi_D < 1\right)}\right). \tag{12}$$

We sample the μ from $\mathcal{N}\left(\mu_\mu, \sigma_\mu^2\right)$, where

$$\sigma_\mu^2 = \left(S\Sigma^{-1} + \sum_{j=1}^{S} \left(T_j - 1\right)\left(I - \text{diag}\left(\phi\right)\right)' \Sigma^{-1} \left(I - \text{diag}\left(\phi\right)\right) + I_D \right)^{-1}$$

and

$$\mu_\mu = \sigma_\mu^2 \left(\sum_{j=1}^{S} \Sigma^{-1}\alpha_{j,1} + \sum_{j=1}^{S}\sum_{t=2}^{T_j} \left(I - \text{diag}\left(\phi\right)\right)' \Sigma^{-1}\alpha_{j,t} - \left(I - \text{diag}\left(\phi\right)\right)' \Sigma^{-1}\text{diag}\left(\phi\right)\alpha_{j,t-1} \right).$$

We sample the covariance matrix Σ from $IW\left(v_1, S_1\right)$, where $v_1 = v + \sum_{j=1}^{S} T_j$ and

$$S_1 = S_\alpha + \sum_{j=1}^{S} \left(\alpha_{j,1} - \mu\right)\left(\alpha_{j,1} - \mu\right)' +$$

$$\sum_{j=1}^{S}\sum_{t=2}^{T_j} \left(\alpha_{j,t} - \mu - \text{diag}\left(\alpha_{j,t-1} - \mu\right)\phi\right)\left(\alpha_{j,t} - \mu - \text{diag}\left(\alpha_{j,t-1} - \mu\right)\phi\right)'.$$

Step 1: Sampling the parameters θ for the polynomial trend model

We first define $\mu_t = x_t\beta$. The matrix x_t is

$$x_t := \begin{bmatrix} x_{1,t}^\top & 0 & \cdots & \cdots & 0 \\ 0 & x_{2,t}^\top & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & x_{D,t}^\top \end{bmatrix},$$

and the $(k_d \times 1)$ vector $x_{d,t} := \left(1, t, t^2\right)^\top$ for $d = 1, ..., D$, $\beta := \left(\beta_1^\top, ..., \beta_D^\top\right)^\top$ with $\beta_d := \left(\beta_{d,1}, ..., \beta_{d,k_d}\right)^\top$.

We sample the parameter β from $\mathcal{N}\left(\mu_\beta, \sigma_\beta^2\right)$, where

$$\sigma_\beta^2 = \left(\sum_{j=1}^{S}\sum_{t=1}^{T} x_t' \Sigma^{-1} x_t + I_D \right)^{-1}$$

and

$$\mu_\beta = \sigma_\beta^2 \left(\sum_{j=1}^{S} \sum_{t=1}^{T} x_t^{'} \Sigma^{-1} \alpha_{j,t} \right).$$

We sample the covariance matrix Σ from $IW(v_1, S_1)$, where $v_1 = v + \sum_{j=1}^{S} T_j$ and

$$S_1 = S_\alpha + \sum_{j=1}^{S} \sum_{t=1}^{T_j} (\alpha_{j,t} - x_t\beta)(\alpha_{j,t} - x_t\beta)^{'}.$$

Step 2 of the PMwG algorithm

Step 2 of the PMwG in Algorithm 2 is to sample the individual-level parameters for all the subjects for $j = 1, ..., S$; it is discussed below.

Trend Model

Step 2a(i) in Algorithm 2 is the conditional Monte Carlo algorithm given in Algorithm 3 that generates $R-1$ new particles $\alpha_{j,1:T}^{-k_{j,1:T}}$, while keeping the particles $\alpha_{j,1:T}^{k_{j,1:T}}$ fixed. This step produces a collection of particles $\alpha_{j,1:T}^{1:R}$, and (2) the normalised weights $W_{j,1:T}^{1:R}$. Step 2a(ii) in Algorithm 2 samples the new index $k_{j,1:T}$ and updates the selected particles $\alpha_{j,1:T}^{k_{j,1:T}}$. Similar steps can be implemented for higher order polynomial trend and spline models. All that is needed to be change is the density $p(\alpha_{j,t}|\theta)$. For example, $p(\alpha_{j,t}|\theta) \sim \mathcal{N}(\mu_t, \Sigma)$, where the d^{th} component of μ_t is $\beta_{d1} + \beta_{d2}t + \beta_{d3}t^2 + \beta_{d4}t^3$, $d = 1, ..., D$, for the third-degree polynomial trend model.

Algorithm 3 Conditional Monte Carlo Algorithm

1. For $t = 1, ..., T$
   (a) Sample $\alpha_{j,t}^r$ from $m_{j,t}(\alpha_{j,t}|y_{j,t}, \theta)$ for $r = 1, ..., R \setminus \{k_{j,t}\}$.
   (b) Calculate the importance weights
   $$w_{j,t}^r = \frac{p\left(y_{j,t}|\alpha_{j,t}^r, \theta\right) p\left(\alpha_{j,t}^r|\theta\right)}{m_{j,t}\left(\alpha_{j,t}^r|y_{j,t}, \theta\right)}, r = 1, ..., R.$$
   (c) Normalize the weights $W_{j,t}^r = \frac{w_{j,t}^r}{\sum_{k=1}^{R} w_{j,t}^k}$, for $r = 1, ..., R$.

AR Model

Step 2b(i) in Algorithm 2 is the conditional sequential Monte Carlo step that generates $R-1$ particles and ancestor indices while keeping a reference particle $\alpha_{j,1:T}^{k_{j,1:T}}$ and the associated ancestor indices $a_{j,1:T-1}^{k_{j,2:T}}$ unchanged.

Algorithm 4 is the conditional sequential Monte Carlo algorithm. The algorithm produces the set of particles $\alpha_{j,1:T}^{1:R}$, ancestor indices $a_{j,1:T-1}^{1:R}$, and weights $w_{j,1:T}^{1:R}$, given the number of particles R, the group-level parameters θ, the dataset for the jth subject $y_{j,1:T}$, the reference particle $\alpha_{j,1:T}^{k_{j,1:T}}$ and its indices $k_{j,1:T}$.

For $t = 1$, step (1a) samples the set of particles $\alpha_{j,1}^{r}$ for $r \in \{1, ..., R\} \setminus \{k_{j,1}\}$, from the proposal density $m_{j,1}\left(\alpha_{j,1}|y_{j,1}, \theta\right)$. Step (1b) computes the weights $w_{j,1}^{1:R}$ and the normalised weights $W_{j,1}^{1:R}$. For $t > 1$, step (2a) resamples the particles except the reference particle $\alpha_{j,t}^{k_{j,t}}$ using simple multinomial resampling and obtain the ancestor indices $a_{j,t-1}^{1:R}$. Given the resampled particles $\alpha_{j,t-1}^{a_{j,t-1}^{1:R}}$, step (2b) generates the set of particles $\alpha_{j,t}^{r}$ for $r \in \{1, ..., R\} \setminus \{k_{j,t}\}$ from the proposal density $m_{j,t}\left(\alpha_{j,t}|\alpha_{j,t-1}^{a_{j,t-1}^{r}}\right)$. Step (2c) computes the weights $w_{j,t}^{1:R}$ and the normalised weights $W_{j,t}^{1:R}$.

The conditional sequential Monte Carlo can be applied to a wide range of time-varying models. For any new application, all that is required is the prior densities $p\left(\alpha_{1,t}|\theta\right)$ and $p\left(\alpha_{j,t}|\alpha_{j,t-1}, \theta\right)$. For the AR model, the density $p\left(\alpha_{j,t}|\alpha_{j,t-1}, \theta\right) \sim N\left(\mu + \text{diag}\left(\phi\right)\left(\alpha_{j,t-1} - \mu\right), \Sigma\right)$ is given in the main text. If we want to use the random walk model, we replace it with the density $p\left(\alpha_{j,t}|\alpha_{j,t-1}, \theta\right) \sim N\left(\alpha_{j,t-1}, \Sigma\right)$.

Steps 2b(ii) and 2b(iii) in Algorithm 2 sample the indices $k_{j,1:T}$ and update the selected particle $\alpha_{j,1:T}^{k_{j,1:T}} = \left(\alpha_{j,1}^{k_{j,1}}, ..., \alpha_{j,T}^{k_{j,T}}\right)$ using the method developed by Bunch et al. (2015) given in Algorithm 5. Similarly to the conditional sequential Monte Carlo step, this step can be easily modified to estimate a wide range of time-varying random effects models. All that is needed to change is the prior densities $p\left(\alpha_{j,1}|\theta\right)$ and $p\left(\alpha_{j,t}|\alpha_{j,t-1}, \theta\right)$. The next section discusses the proposal densities used in the PMwG sampler.

Tuning Parameters and Proposal Densities for the PMwG Sampler

Sampling efficiency greatly increases when appropriate proposal densities are used for the PMwG sampler. For the PMwG sampler, it is necessary to specify the number of particles R, the proposal densities $m_{j,1}\left(\alpha_{j,1}|\theta\right)$, $m_{j,t}\left(\alpha_{j,t}|\alpha_{j,t-1}, \theta\right)$ and $\omega_{j,t}\left(\alpha_{j,t}|\alpha_{j,t-1}, \theta\right)$ for the AR model and $m_{j,t}\left(\alpha_{j,t}|\theta\right)$ for the trend model for $j = 1, ..., S$ and $t = 1, .., T$. The PMwG sampler has three stages: burn-in, initial adaptation, and sampling. It is initialized at a set of parameters θ and random effects $\alpha_{j,1:T}$ for $j = 1, ..., S$. It then proceeds as in Algorithm 2. Initially, for the AR model, in the burn-in and the initial adaptation stages, the proposal density for subject j is the two component mixture

$$m_{j,1}\left(\alpha_{j,1}|\theta\right) = w_{mix}\mathcal{N}\left(\alpha_{j,1}; \alpha_{j,1}^{(iter-1)}, \epsilon\Sigma_{\alpha}\right) + \left(1 - w_{mix}\right)p\left(\alpha_{j,1}|\theta\right), \tag{13}$$

and

$$m_{j,t}\left(\alpha_{j,t}|\alpha_{j,t-1}, \theta\right) = w_{mix}\mathcal{N}\left(\alpha_{j,t}; \alpha_{j,t}^{(iter-1)}, \epsilon\Sigma_{\alpha}\right) + \left(1 - w_{mix}\right)p\left(\alpha_{j,t}|\alpha_{j,t-1}, \theta\right), \tag{14}$$

for $t > 1$, where $\alpha_{j,t}^{(iter-1)}$ is the previous iterate $\alpha_{j,t}^{k_{j,t}}$ for the individual random effects for subject j at block t. We set $w_{mix} = 0.8$ and ϵ is a scale factor. In this paper, we set $\epsilon = 1$. We apply the proposal density in Equations (13) and (14) for all $t = 1, ..., T$.

Algorithm 4 Conditional Sequential Monte Carlo algorithm

Inputs: N, θ, $y_{j,1:T}$, $\alpha_{j,1:T}^{k_{j,1:T}}$, and $k_{j,1:T}$
Outputs: $\alpha_{j,1:T}^{1:R}$, $a_{j,1:T-1}^{1:R}$,$w_{j,1:T}^{1:R}$

1. For $t = 1$

   (a) Sample $\alpha_{j,1}^{r}$ from $m_{j,1}\left(\alpha_{j,1}|y_{j,1},\theta\right)$, for $r \in \{1,...,R\} \setminus \{k_{j,1}\}$

   (b) Calculate the importance weights

$$w_{j,1}^{r} = \frac{p\left(y_{j,1}|\alpha_{j,1}^{r},\theta\right) p\left(\alpha_{j,1}^{r}|\theta\right)}{m_{j,1}\left(\alpha_{j,1}^{r}|y_{j,1},\theta\right)}, r = 1,...,R.$$

   and normalize them to obtain $W_{j,1}^{1:R}$.

2. For $t > 1$

   (a) Sample the ancestral indices $a_{j,t-1}^{-(k_{j,t})} \sim \mathcal{M}\left(a_{j,t-1}^{-(k_{j,t})}|W_{j,t-1}^{1:R}\right)$.

   (b) Sample $\alpha_{j,t}^{r}$ from $m_{j,t}\left(\alpha_{j,t}|\alpha_{j,t-1}^{a_{j,t-1}^{r}},\theta\right)$, $r = 1,...,R \setminus \{k_{j,t}\}$.

   (c) Calculate the importance weights

$$w_{j,t}^{r} = \frac{p\left(y_{j,t}|\alpha_{j,t}^{r},\theta\right) p\left(\alpha_{j,t}^{r}|\alpha_{j,t-1}^{a_{j,t-1}^{r}},\theta\right)}{m_{j,t}\left(\alpha_{j,t}^{r}|\alpha_{j,t-1}^{a_{j,t-1}^{r}},\theta\right)}, r = 1,...,R.$$

   and normalize to obtain $W_{j,t}^{1:R}$.

In the sampling stage, the posterior MCMC draws $(\alpha_{j,1:T},\theta)$ from the initial adaptation stage are used to build more efficient proposal densities $m_{j,t}\left(\alpha_{j,t}|\alpha_{j,t-1},\theta\right)$ for the AR model and $m_{j,t}\left(\alpha_{j,t}|\theta\right)$ for the trend model for $j = 1,...,S$ and $t = 1,...,T$. The posterior draws of the parameters ϕ and Σ for the AR model, and Σ for the trend model, are transformed so that they all lie on the real line. The covariance matrix Σ is reparameterised in terms of its Cholesky factorisation $\Sigma = LL^{T}$, where L is a lower triangular matrix. We apply a log transformation for the diagonal elements of L, while the subdiagonal elements of L are unrestricted. We also the logit transform each element of the vector ϕ to $\phi_{L} = \log\left(\phi/\left(1-\phi\right)\right)$, which lies on the real line. The proposal densities for $\omega_{j,t}\left(\alpha_{j,t}|\alpha_{j,t-1},\theta\right)$ are the same as $m_{j,t}\left(\alpha_{j,t}|\alpha_{j,t-1},\theta\right)$ for all $t > 1$ for the AR model. For the AR model, a multivariate normal distribution is fitted to the posterior draws of $\alpha_{j,1}$ and θ for $t = 1$ and to the posterior draws of $\alpha_{j,t}$, $\alpha_{j,t-1}$, and θ for $t = 2,...,T$; the conditional distributions $g\left(\alpha_{j,1}|\theta\right) \sim N\left(\alpha_{j,1}|\mu_{j,1}^{prop},\Sigma_{j,1}^{prop}\right)$ and $g\left(\alpha_{j,t}|\alpha_{j,t-1},\theta\right) \sim N\left(\alpha_{j,t}|\mu_{j,t}^{prop},\Sigma_{j,t}^{prop}\right)$ are then obtained. The efficient proposal density for subject j is then the three component

Algorithm 5 Conditional Monte-Carlo for the joint ancestor and random effects conditional distributions

1. Sample $\left(a_{j,t-1}^{(-m)}, \alpha_{j,t}^{(-m)}\right)$ from $\eta\left(a_{j,t-1}^{(-m)}, \alpha_{j,t}^{(-m)} | a_{j,t-1}^{m}, \alpha_{j,t}^{m}, m\right)$,

$$\eta\left(a_{j,t-1}^{-m}, \alpha_{j,t}^{-m} | a_{j,t-1}^{m}, \alpha_{j,t}^{m}, m\right) = \prod_{r \neq m} \frac{v_{j,t-1}^{a_{j,t-1}^{r}}}{\sum_{l} v_{j,t-1}^{l}} \omega_{t}\left(\alpha_{j,t}^{r} | \alpha_{j,t-1}^{a_{j,t-1}^{r}}\right),$$

2. Sample index m from $\eta\left(m = m^{*} | a_{j,t-1}^{1:R}, \alpha_{j,t}^{1:R}\right) \propto \widetilde{W}_{j,t}^{m^{*}}$, where

$$\widetilde{w}_{j,t}^{r} = \frac{w_{j,t-1}^{a_{j,t-1}^{r}} p\left(\alpha_{j,t}^{r} | \alpha_{j,t-1}^{a_{j,t-1}^{r}}, \theta\right) p\left(y_{j,t} | \alpha_{j,t}^{r}, \theta\right) p\left(\alpha_{j,t+1}^{k_{j,t+1}} | \alpha_{j,t}^{r}, \theta\right)}{v_{j,t-1}^{a_{j,t-1}^{r}} \omega_{j,t}\left(\alpha_{j,t}^{r} | \alpha_{j,t-1}^{a_{j,t-1}^{r}}\right)},$$

if $v_{j,t} = w_{j,t}$, then we have that

$$\widetilde{w}_{j,t}^{r} = \frac{p\left(\alpha_{j,t}^{r} | \alpha_{j,t-1}^{a_{j,t-1}^{r}}, \theta\right) p\left(y_{j,t} | \alpha_{j,t}^{r}, \theta\right) p\left(\alpha_{j,t+1}^{k_{j,t+1}} | \alpha_{j,t}^{r}, \theta\right)}{\omega_{j,t}\left(\alpha_{j,t}^{r} | \alpha_{j,t-1}^{a_{j,t-1}^{r}}\right)},$$

and $\widetilde{W}_{j,t}^{r} = \widetilde{w}_{j,t}^{r} / \sum_{l} \widetilde{w}_{j,t}^{l}$.

---

mixture

$$m_{j,1}\left(\alpha_{j,1} | \theta\right) = w_{1,mix} g\left(\alpha_{j,1} | \theta\right) + w_{2,mix} p\left(\alpha_{j,1} | \theta\right) + w_{3,mix} \overline{g}\left(\alpha_{j,1} | \theta\right), \tag{15}$$

and

$$m_{j,t}\left(\alpha_{j,t} | \alpha_{j,t-1}, \theta\right) = w_{1,mix} g\left(\alpha_{j,t} | \alpha_{j,t-1}, \theta\right) + w_{2,mix} p\left(\alpha_{j,t} | \alpha_{j,t-1}, \theta\right) + w_{3,mix} \overline{g}\left(\alpha_{j,t} | \alpha_{j,t-1}, \theta\right), \tag{16}$$

where $\overline{g}\left(\alpha_{j,t} | \alpha_{j,t-1}, \theta\right)$ is also the proposal based on the normal distribution as in $g\left(\alpha_{j,t} | \alpha_{j,t-1}, \theta\right)$ except that we now use $\alpha_{j,t}^{(iter-1)}$ as the means. Similarly, we fit a multivariate normal distribution to the posterior draws of $\alpha_{j,t}$ and θ as in Equation (15), for all $t = 1, ..., T$ for the trend model. Following Hesterberg (1995), including the prior density ensures that the importance weights are bounded if the density $p\left(y_{j,t} | \alpha_{j,t}, \theta\right)$ is bounded, which ensures that the sampler is ergodic. Similar proposals can be used for other time-varying random effects models. The next section discusses the Importance Sampling Squared $\left(\mathrm{IS}^{2}\right)$ method to estimate the marginal likelihood for the time-varying LBA models.

Estimating the Marginal Likelihood for the dynamic LBA models using Importance Sampling Squared $\left(\mathbf{IS}^2\right)$

The marginal likelihood in Equation (2) is frequently used for Bayesian model selection; see, e.g., Chib and Jeliazkov (2001). We now show how the importance sampling squared $\left(\mathrm{IS}^2\right)$ approach can be used to estimate the marginal likelihood of the time-varying LBA models; see Tran et al. (2021) for full details of the IS^2 method. Algorithm 6 outlines the IS^2 algorithm for estimating the marginal likelihood. The unbiased estimate of the likelihood is

$$\widehat{p}\left(y_{1:T}|\theta\right) = \prod_{t=1}^{T}\left\{\frac{1}{R}\sum_{r=1}^{R} w_{j,t}^{(r)}\right\}.$$

Algorithm 6 Importance Sampling Squared $\left(\mathrm{IS}^2\right)$ algorithm for estimating the marginal likelihood

For $i = 1, ..., M$,

1. Generate $\theta_i \overset{iid}{\sim} g_{IS}\left(\theta\right)$ and compute the unbiased estimate of the likelihood $\widehat{p}\left(y_{1:T}|\theta_i\right) = \prod_{t=1}^{T}\left\{\frac{1}{R}\sum_{r=1}^{R} w_{j,t}^{(r)}\right\}$
2. Compute the weights
$$\widetilde{w}\left(\theta_i\right) = \frac{\widehat{p}\left(y_{1:T}|\theta_i\right) p\left(\theta_i\right)}{g_{IS}\left(\theta_i\right)}.$$
3. The IS^2 estimator of the marginal likelihood $p\left(y_{1:T}\right)$
$$\widehat{p}_{IS^2}\left(y\right) = \frac{1}{M}\sum_{i=1}^{M}\widetilde{w}\left(\theta_i\right). \tag{17}$$

Constructing the Proposal Densities

This section outlines how to obtain efficient and reliable proposals: a) $g_{\mathrm{IS}}\left(\theta\right)$ for the LBA group-level parameters; b) $m_{j,1}\left(\alpha_{j,1}|\theta\right)$, $m_{j,t}\left(\alpha_{j,t}|\alpha_{j,t-1}, \theta\right)$ for the random effects $\alpha_{j,t}$, $t = 1, ..., T$ and $j = 1, ..., S$ for the AR model; c) $m_{j,t}\left(\alpha_{j,t}|\theta\right)$, $t = 1, ..., T$ and $j = 1, ..., S$ for the random effects $\alpha_{j,t}$ in the trend model. Our approach to obtain these proposal densities is to run the PMwG sampler for dynamic LBA models to generate efficient posterior samples $\left(\alpha_{1:S,1:T}^{r}, \theta^{r}\right)$, $r = 1, ..., R$, from the posterior density of the parameters and random effects $\pi\left(\theta, \alpha_{1:S,1:T}\right)$. Given the posterior draws of the group-level parameters θ, the proposal density for the parameters $g_{IS}\left(\theta\right)$ is obtained by fitting a mixture of normal densities,

$$g_{\mathrm{IS}}\left(\theta\right) = \sum_{k=1}^{K} w_k^{MIX}\mathcal{N}\left(\theta|\mu_k, \Sigma_k\right), \tag{18}$$

where $\mathcal{N}\left(\mu_k, \Sigma_k\right)$ denotes the normal pdf with mean μ_k, variance-covariance matrix Σ_k, and component weights w_k^{MIX}. In all our examples, the mixture of normals is fitted using the Matlab built-in function $fitgmdist$, and K is set to 2.

Similarly to selecting the proposal densities for $m_{j,1}\left(\alpha_{j,1}|\theta\right)$ and $m_{j,t}\left(\alpha_{j,t}|\alpha_{j,t-1},\theta\right)$ for the PMwG sampler, a normal distribution is first fitted to the posterior draws of $\alpha_{j,1}$ and θ for $t=1$ and to the posterior draws of $\alpha_{j,t}$, $\alpha_{j,t-1}$, and θ for $t=2,...,T$ and $j=1,...,S$; we then obtain the conditional distribution $g\left(\alpha_{j,1}|\theta\right)\sim N\left(\alpha_{j,1}|\mu_{j,1}^{prop},\Sigma_{j,1}^{prop}\right)$ and $g\left(\alpha_{j,t}|\alpha_{j,t-1},\theta\right)\sim N\left(\alpha_{j,t}|\mu_{j,t}^{prop},\Sigma_{j,t}^{prop}\right)$. The group level parameters θ are appropriately transformed so that they all lie on the real line. The efficient proposal density for subject j at block t is the two component mixture

$$m_{j,1}\left(\alpha_{j,1}|\theta\right)=w_1^{mix}g\left(\alpha_{j,1}|\theta\right)+\left(1-w_1^{mix}\right)p\left(\alpha_{j1}|\theta\right), \tag{19}$$

and

$$m_{j,t}\left(\alpha_{j,t}|\alpha_{j,t-1},\theta\right)=w_1^{mix}g\left(\alpha_{j,t}|\alpha_{j,t-1},\theta\right)+\left(1-w_1^{mix}\right)p\left(\alpha_{j,t}|\alpha_{j,t-1},\theta\right), \tag{20}$$

for $t>1$. The mixture weights are set to $w_1^{mix}=0.95$. For the trend model, the proposal density is

$$m_{j,t}\left(\alpha_{j,t}|\theta\right)=w_1^{mix}g\left(\alpha_{j,t}|\theta\right)+\left(1-w_1^{mix}\right)p\left(\alpha_{j,t}|\theta\right), \tag{21}$$

for all $t=1,...,T$. Similar proposals can be used for other time-varying random effects models.

References

Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*, *72*, 1-33.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive psychology*, *57*(3), 153–178.

Bunch, P., Lindsten, F., & Singh, S. S. (2015). Particle Gibbs with refreshed backward simulation. *IEEE International conference on acoustics, speech, and signal processing*.

Chib, S., & Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of American Statistical Association*, *96*(453), 270-281.

Gunawan, D., Carter, C., Fiebig, D. G., & Kohn, R. (2017). Efficient Bayesian estimation for flexible panel models for multivariate outcomes: impact of life events on mental health and excessive alcohol consumption. *arXiv preprint arXiv:1706.03953v1*.

Hesterberg, T. (1995). Weighted average importance sampling and defensive mixture distributions. *Technometrics*, *37*, 185-194.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: current issues and history. *Trends in cognitive sciences*, *20*(4), 260–281.

Terry, A., Marley, A. A. J., Barnwal, A., Wagenmakers, E.-J., Heathcote, A., & Brown, S. D. (2015). Generalising the drift rate distribution for linear ballistic accumulators. *Journal of Mathematical Psychology*, *68*, 49–58.

Tran, M.-N., Scharth, M., Gunawan, D., Kohn, R., Brown, S. D., & Hawkins, G. E. (2021). Robustly estimating the marginal likelihood for cognitive models via importance sampling. *Behavior Research Methods*, *53*, 1148–1165.