SV models

Let P_t be a financial time series (e.g., an asset at time t). Model the returns $r_t = P_t/P_{t-1} - 1$ as

$$\begin{aligned} r_t &\sim N(0, \sigma_t^2), \\ \sigma_t &= \exp(\mu + x_t), \\ x_t &= \phi x_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, v). \end{aligned}$$

This is a non-linear state-space model. However, we can obtain a conditionally Gaussian DLM when modeling $y_t = \log(r_t^2)/2$. Specifically, we obtain

$$\begin{aligned} y_t &= \mu + x_t + \nu_t, \quad \nu_t = \log(\kappa_t)/2, \quad \kappa_t \sim \chi_1^2 \\ x_t &= \phi x_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, v). \end{aligned}$$

SV models

Following, Shephard (1994), Kim, Shephard, and Chib (1998), the distribution of ν_t can be approximated by a mixture of normals, i.e., $\nu_t \sim \sum_{j=1}^{7} q_j N(b_j, v_j)$, with q_j, b_j, v_j known. Furthermore, considering the centered parameterization we have

$$\begin{aligned} y_t &= z_t + \nu_t, \\ z_t &= \mu + \phi(z_{t-1} - \mu) + \epsilon_t. \end{aligned}$$

This is a conditionally Gaussian DLM and standard MCMC methods for state-space models can be used.

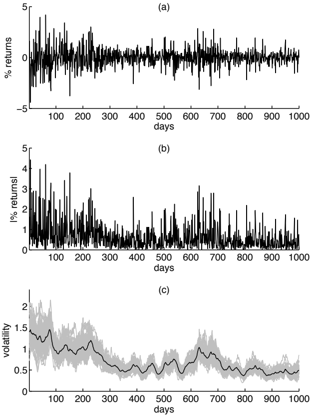
SV models

MCMC algorithm:

- Sample $z_{0:T}$: FFBS algorithm.
- Sample μ : Normal mean.
- v : Inverse-gamma.
- ϕ : Normal linear regression constrained to $0 < \phi < 1$.
- Sample Gaussian mixture indicators: conditionally independent distributions.

SV models

Zero-centered daily returns on $UK:$USA exchange rates, 1000 days from Fall 1992 to Summer 1996 (withdrawal of UK from European Monetary System was in Sep 1992).



Multiprocess Models

Key references: Harrison and Stevens (1971, 1976), West and Harrison (1997) Chapter 12.
Definitions. Let α be the set of uncertain quantities that define a DLM at time t denoted by $\mathcal{M}_t = \mathcal{M}_t(\alpha)$ for $\alpha \in \mathcal{A}$.

- **Class I.** A single model $\mathcal{M}_t(\alpha)$ holds for all t for some $\alpha \in \mathcal{A}$. There is uncertainty about α.
- **Class II.** At time t α takes a value in $\mathcal{A}$ such that $\mathcal{M}_t(\alpha_t)$ holds at time t. Therefore, $\alpha_{1:T}$ defines the set of DLMs for each $t = 1 : T$. No single DLM is appropriate at all times.

Multiprocess Models

Example

- **Mixtures of DLMs with different discount factors.** Let $\alpha = \delta$ and $\mathcal{A} = \{\delta_1, \ldots, \delta_K\}$ with $\delta_k \in (0, 1]$. Then, $\mathcal{M}_t(\delta)$ is

$$\begin{aligned} y_t &= \theta_t + \nu_t, \quad \nu_t \sim N(0, v) \\ \theta_t &= \theta_{t-1} + w_t, \quad w_t \sim N(0, w(\delta)). \end{aligned}$$

Class I: a single $\delta \in \mathcal{A}$ is appropriate for all t. Class II: the model at time t is $\mathcal{M}_t(\delta_t)$ with $\delta_t \in \mathcal{A}$.

Multiprocess Models

Example

- **Mixtures of ARs.** Let $\mathcal{M}_t(\phi^{(k)})$ be

$$y_t = \phi^{(k)} y_{t-1} + \epsilon_t, \;\; \epsilon_t \sim N(0, v),$$

with $k = 1, 2$. Assume $\phi^{(1)} \sim U(0, 1)$ and $\phi^{(2)} \sim U(-1, 0)$. Then, $\alpha_t = \phi^{(k)}$ with $k = 1$ or $k = 2$ and the DLM at time t is $\{\phi^{(k)}, 0, v, 0\}$ (static AR).

Multiprocess Models

Posterior Inference, Class I

- Beginning with $p(\boldsymbol{\alpha}|\mathcal{D}_0)$, we have

$$p(\boldsymbol{\alpha}|\mathcal{D}_t) \propto p(\boldsymbol{\alpha}|\mathcal{D}_{t-1})p(y_t|\boldsymbol{\alpha}, \mathcal{D}_{t-1}).$$

- $p(\boldsymbol{\theta}_t|\mathcal{D}_t)$ is given by

$$p(\boldsymbol{\theta}_t|\mathcal{D}_t) = \int_{\mathcal{A}} p(\boldsymbol{\theta}_t|\boldsymbol{\alpha}, \mathcal{D}_t)p(\boldsymbol{\alpha}|\mathcal{D}_t)d\boldsymbol{\alpha}.$$

- One-step-ahead forecast:

$$p(y_{t+1}|\mathcal{D}_t) = \int_{\mathcal{A}} p(y_{t+1}|\boldsymbol{\alpha}, \mathcal{D}_t)p(\boldsymbol{\alpha}|\mathcal{D}_t)d\boldsymbol{\alpha}.$$

Multiprocess Models

Posterior Inference, Class II. Assume $\mathcal{A} = \{1, \ldots, K\}$, let $\mathcal{M}_t(k)$ refer to the model k at time t. Notation:

- Let $\pi_t(k) = Pr(\mathcal{M}_t(k)|\mathcal{D}_{t-1})$ be the prior probability of selecting model $\mathcal{M}_t(k)$. Let $\pi_t(k|i) = Pr(\mathcal{M}_t(k)|\mathcal{M}_{t-1}(i), \mathcal{D}_{t-1})$. Then,

$$\pi_t(k) = \sum_{i=1}^{K} \pi_t(k|i) p_{t-1}(i),$$

  with $p_{t-1}(i) = Pr(\mathcal{M}_{t-1}(i)|\mathcal{D}_{t-1})$ (posterior at time $t-1$).
- Posterior probability of a path of models:

$$p_t(k_t, \ldots, k_{t-h}) = Pr(\mathcal{M}_t(k_t), \ldots, \mathcal{M}_{t-h}(k_{t-h})|\mathcal{D}_t).$$

- Posterior for model parameters $p(\boldsymbol{\theta}_t|\mathcal{D}_t)$:

$$
\begin{aligned}
p(\boldsymbol{\theta}_t|\mathcal{D}_t) &= \sum_{k_t=1}^{K} p(\boldsymbol{\theta}_t|\mathcal{M}_t(k_t), \mathcal{D}_t) p_t(k_t), \\
&= \sum_{k_t=1}^{K} \cdots \sum_{k_1=1}^{K} p(\boldsymbol{\theta}_t|\mathcal{M}_t(k_t), \ldots, \mathcal{M}_1(k_1), \mathcal{D}_t) \\
&\quad \times p_t(k_t, \ldots, k_1)
\end{aligned}
$$

K^t components $\Rightarrow$ computationally challenging. Assume that dependence on early models is neglibible as time passes and use K^{h+1} components, h small:

$$
\begin{aligned}
p(\boldsymbol{\theta}_t|\mathcal{D}_t) &\approx \sum_{k_t=1}^{K} \cdots \sum_{k_h=1}^{K} p(\boldsymbol{\theta}_t|\mathcal{M}_t(k_t), \ldots, \mathcal{M}_h(k_{t-h}), \mathcal{D}_t) \\
&\quad \times p_t(k_t, \ldots, k_{t-h})
\end{aligned}
$$

Further approximations:

- Ignore components with very small probabilities and combine components that are roughly equal into a single component.
- Replace a collection of components by a single distribution using, for example, the Kullback-Leibler divergence method. E.g., for a mixture of K Gaussians, each with weight $p(k)$, mean $\mathbf{m}(k)$ and variance $\mathbf{C}(k)$, the optimal approx. distribution that minimizes the K-L divergence is a Gaussian with mean $\mathbf{m}$ and variance $\mathbf{C}$:

$$
\begin{aligned}
\mathbf{m} &= \sum_{k=1]}^{K} p(k)\mathbf{m}(k), \\
\mathbf{C} &= \sum_{k=1}^{K} [\mathbf{C}(k) + (\mathbf{m} - \mathbf{m}(k))(\mathbf{m} - \mathbf{m}(k))']p(k).
\end{aligned}
$$

Multiprocess Models

West and Harrison (1997) consider the case of $K = 4$, $k_t = 1:4$, and $\mathcal{M}_t(k_t)$ given by $\{\mathbf{F}, \mathbf{G}, v_t v(k_t), \mathbf{W}_t(k_t)\}$ with $v(k_t)$ and $\mathbf{W}_t(k_t)$ known and $v_t = v = \phi^{-1}$ unknown. Under conjugate Normal-Gamma priors it can be shown that:

- Posteriors at time $t-1$:

$$\begin{aligned}(\boldsymbol{\theta}_{t-1}|\mathcal{M}_{t-1}(k_{t-1}), \mathcal{D}_{t-1}) &\sim T_{n_{t-1}}(\mathbf{m}_{t-1}(k_{t-1}), \mathbf{C}_{t-1}(k_{t-1}))\\ (\phi|\mathcal{M}_{t-1}(k_{t-1}), \mathcal{D}_{t-1}) &\sim G(n_{t-1}/2, d_{t-1}(k_{t-1})/2),\end{aligned}$$

- Priors at time t:

$$(\boldsymbol{\theta}_t|\mathcal{M}_t(k_t), \mathcal{M}_{t-1}(k_{t-1}), \mathcal{D}_{t-1}) \sim T_{n_{t-1}}(\mathbf{a}_t(k_{t-1}), \mathbf{R}_t(k_t, k_{t-1})),$$

with $\mathbf{a}_t(k_{t-1}) = \mathbf{G}\mathbf{m}_t(k_{t-1})$, and $\mathbf{R}_t(k_t, k_{t-1}) = \mathbf{G}\mathbf{C}_{t-1}(k_{t-1})\mathbf{G}' + \mathbf{W}_t(k_t)$.

Multiprocess Models

- One-step-ahead forecast:

$$(y_t|\mathcal{M}_t(k_t), \mathcal{M}_{t-1}(k_{t-1}), \mathcal{D}_{t-1}) \quad \sim \quad T_{n_{t-1}}(f_t(k_{t-1}), q_t(k_t, k_{t-1})),$$

with $f_t(k_{t-1}) = \mathbf{F}'\mathbf{a}_t(k_{t-1})$, and
$q_t(k_t, k_{t-1}) = \mathbf{F}'\mathbf{R}_t(k_t, k_{t-1})\mathbf{F} + s_{t-1}(k_{t-1})v_t(k_t)$, and

$$p(y_t|\mathcal{D}_{t-1}) \quad = \quad \sum_{k_t=1}^{K} \sum_{k_{t-1}=1}^{K} [p(y_t|\mathcal{M}_t(k_t), \mathcal{M}_{t-1}(k_{t-1}), \mathcal{D}_{t-1}) \times \pi(k_t) p_{t-1}(k_{t-1})] \,.$$

- Posteriors at time t:

$$(\boldsymbol{\theta}_t|\mathcal{M}_t(k_t), \mathcal{M}_{t-1}(k_{t-1}), \mathcal{D}_t) \quad \sim \quad T_{n_t}(\mathbf{m}_t(k_t, k_{t-1}), \mathbf{C}_t(k_t, k_{t-1})),$$
$$(\phi|\mathcal{M}_t(k_t), \mathcal{M}_{t-1}(k_{t-1}), \mathcal{D}_t) \quad \sim \quad G(n_t/2, d_t(k_t, k_{t-1})/2),$$

Multiprocess Models

$$
\begin{aligned}
\mathbf{m}_t(k_t, k_{t-1}) &= \mathbf{a}_t(k_{t-1}) + \mathbf{A}_t(k_t, k_{t-1}) e_t(k_{t-1}), \\
\mathbf{C}_t(k_t, k_{t-1}) &= [s_t(k_t, k_{t-1})/s_{t-1}(k_{t-1})] \times \\
&\quad [\mathbf{R}_t(k_t, k_{t-1}) - \mathbf{A}_t(k_t, k_{t-1})\mathbf{A}_t'(k_t, k_{t-1}) q_t(k_t, k_{t-1})], \\
e_t(k_{t-1}) &= y_t - f_t(k_{t-1}), \\
\mathbf{A}_t(k_t, k_{t-1}) &= \mathbf{R}_t(k_t, k_{t-1})\mathbf{F}/q_t(k_t, k_{t-1}), \\
d_t(k_t, k_{t-1}) &= d_{t-1}(k_{t-1}) + s_{t-1}(k_{t-1}) e_t(k_{t-1})^2 / q_t(k_t, k_{t-1}),
\end{aligned}
$$

$s_t(k_t, k_{t-1}) = d_t(k_t, k_{t-1})/n_t$ and $n_t = n_{t-1} + 1$. In addition,

- Posterior model probabilities:

$$
p_t(k_t, k_{t-1}) \propto \frac{\pi(k_t) p_{t-1}(k_{t-1})}{q_t(k_t, k_{t-1})^{1/2} [n_{t-1} + e_t(k_{t-1})^2 / q_t(k_t, k_{t-1})]^{n_t/2}}.
$$

Multiprocess Models

Using $h = 1$ we obtain the following:

$$p(\boldsymbol{\theta}_t|\mathcal{D}_t) = \sum_{k_t=1}^{K} \sum_{k_{t-1}=1}^{K} p(\boldsymbol{\theta}_t|\mathcal{M}_t(k_t), \mathcal{M}_{t-1}(k_{t-1}), \mathcal{D}_t) p_t(k_t, k_{t-1}).$$

These can be collapsed to obtain:

$$\begin{aligned} (\boldsymbol{\theta}_t|\mathcal{M}_t(k_t), \mathcal{D}_t) &\approx T_{n_t}(\mathbf{m}_t(k_t), \mathbf{C}_t(k_t)), \\ (\phi|\mathcal{M}_t(k_t), \mathcal{D}_t) &\approx G(n_t/2, d_t(k_t)/2), \end{aligned}$$

where $d_t(k_t)$, $\mathbf{m}_t(k_t)$, $\mathbf{C}_t(k_t)$ are computed using the K-L method.

Multiprocess Models

In addition, we have:

- $p_t(k_t) = Pr(\mathcal{M}_t(k_t)|\mathcal{D}_t) = \sum_{k_{t-1}=1}^{K} p_t(k_t, k_{t-1})$,
- $Pr(\mathcal{M}_{t-1}(k_{t-1})|\mathcal{D}_t) = \sum_{k_t=1}^{K} p_t(k_t, k_{t-1})$,
- $Pr(\mathcal{M}_{t-1}(k_{t-1})|\mathcal{M}_t(k_t), \mathcal{D}_t) = p_t(k_t, k_{t-1})/p_t(k_t)$.

Mixtures of structured autoregressions

- We have a collection of $q = 1 : Q$ consecutive time series of length T.
- Each of the Q time series is described by an AR process.
- Different time series may be generated from different AR processes. Assume we have K possible AR processes with $K << Q$.
- Structured prior distributions on the AR parameters.
- **Motivation: Cognitive fatigue data.** Multiple electroencephalograms.

Cognitive fatigue data

- EEG data collected at NASA Ames by L. Trejo and collaborators.
- **Experiment:** Multiple subjects were asked to solve simple arithmetic equations continuously for 3 hours.
- 64 EEG channels were recorded per subject.
- For each channel we have a collection of “consecutive” **epochs.** An **epoch** is a time series of 1,664 observations. It corresponds to 13 seconds of recording: 5s prior to the stimulus and 8s post-stimulus $\Rightarrow$ sampling rate is 128 Hz.

Goal: <u>Online</u> detection of cognitive fatigue. (1) Can we detect fatigue from EEGs? If so, what characterizes fatigue? (3) Are there several states of mental alertness?

Cognitive fatigue data

Epochs

Problem solving 3+5-7 >=< 2? Time: 13s (-5s,8s)	⇒	Response	⇒	Inter-trial interval Time: 1s	⇒	Problem solving 4+2-1 >=< 3? Time: 13s (-5s,8s)	⇒

Let $q = 1 : Q$ be the number of epochs. We consider K models $\{\mathcal{M}_q(1), \ldots, \mathcal{M}_q(K)\}$. Each model represents one of K brain states.

AR mixture component

$$\mathcal{M}_q(k) : \qquad y_{q,t} = \sum_{i=1}^{p} \phi_i^{(k)} y_{q,t-i} + \epsilon_{q,t}^{(k)}, \quad \epsilon_{q,t}^{(k)} \sim N(0, v).$$

AR characteristic polynomial

$$\phi^{(k)}(u) = 1 - \phi_1^{(k)} u - \ldots \phi_p^{(k)} u^p.$$

Let $\boldsymbol{\alpha}^{(k)} = (\alpha_1^{(k)}, \ldots, \alpha_p^{(k)})$ denote the reciprocal characteristic roots.

Structured priors: Prior structure on the reciprocal roots. Suppose we have n_c pairs of complex reciprocal roots and n_r real reciprocal roots.

- **Complex roots.** Appear in conjugate pairs:

$$\alpha_{2j-1}^{(k)} = r_j^{(k)} \exp(-2\pi i/\lambda_j^{(k)}), \quad \alpha_{2j}^{(k)} = r_j^{(k)} \exp(2\pi i/\lambda_j^{(k)}),$$

$r_j^{(k)}$: modulus and $\lambda_j^{(k)}$: period.

- **Real roots.**

$$\alpha_j^{(k)} = r_j^{(k)}$$

for $j = 1 : n_r$.

- Assume that all the roots are different.
- Stationarity: $|r_j^{(k)}| < 1$.

We follow an approach similar to Huerta and West (1999):

- **Complex roots:**

$$(r_j|\mathcal{D}_0) \sim f(r_j), \quad (\lambda_j|\mathcal{D}_0) \sim g(\lambda_j),$$

  $f(\cdot)$ continuous on $(a, b) \subseteq (0, 1)$ and $g(\cdot)$ continuous on $(c, d) \subseteq (2, \lambda_{\max})$.

- **Real roots:**

$$(r_j|\mathcal{D}_0) \sim h(r_j)$$

  $h(\cdot)$ continuous on $(e, f) \subseteq (-1, 1)$.

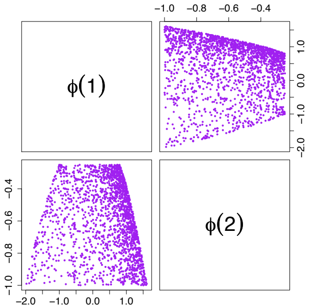
- **Variance**

$$(v|\mathcal{D}_0) \sim \mathrm{IG}(n_0/2, d_0/2).$$

Structured priors: Examples

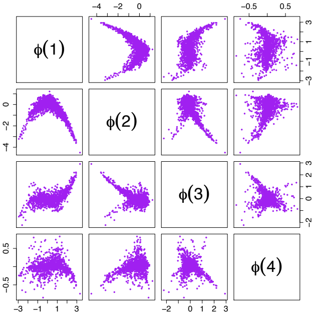
One pair of complex roots

$r \sim U(0.5, 1),\ \lambda \sim U(2, 10)$



One pair of complex roots and two real roots

$r_1 \sim U(0, 1),\ \lambda_1 \sim U(2, 10),$
$r_2, r_3 \sim U(-1, 1)$



Structured priors: Examples

- **AR**(1) **with** $K = 2$**:**

$$r^{(1)} \sim U(0.8, 1), \quad r^{(2)} \sim U(0, 0.3).$$

- **AR**(2) **with** $K = 2$:

$$\begin{aligned}
(r^{(1)}|\mathcal{D}_0) &\sim TN(0.95, 0.001, \mathcal{R}_1^{(1)}), \\
(\lambda^{(1)}|\mathcal{D}_0) &\sim TN(10, 4, \mathcal{R}_2^{(1)}), \\
(r^{(2)}|\mathcal{D}_0) &\sim TN(0.95, 0.001, \mathcal{R}_1^{(2)}), \\
\lambda^{(2)}|\mathcal{D}_0) &\sim TN(17, 4, \mathcal{R}_2^{(2)}),
\end{aligned}$$

with $\mathcal{R}_1^{(k)} = (0.8, 1)$, $\mathcal{R}_2^{(1)} = (8, 12)$ and $\mathcal{R}_2^{(2)} = (14, 20)$.

Further model structure. Let $\mathcal{D}_{q-1} = \{\mathcal{D}_0, \mathbf{y}_{1:(q-1)}\}$.

Prior and posterior for q

$$\pi_q(k) \equiv Pr[\mathcal{M}_q(k)|\mathcal{D}_{q-1}] \quad \overset{\overset{\mathbf{y}_q}{\downarrow}}{\Longrightarrow} \quad p_q(k) \equiv Pr[\mathcal{M}_q(k)|\mathcal{D}_q]$$

Transition probabilities

$$Pr[\mathcal{M}_q(k)|\mathcal{M}_{q-1}(i), \mathcal{D}_{q-1}] = Pr[\mathcal{M}_q(k)|\mathcal{M}_{q-1}(i), \mathcal{D}_0] \equiv \pi(k|i)$$

Posterior inference

$$
\begin{aligned}
p(\phi^{(1:K)}, v|\mathcal{D}_q) &= \sum_{k_q=1}^{K} p(\phi^{(1:K)}, v|\mathcal{M}_q(k_q), \mathcal{D}_q) p_q(k_q) \\
&= \sum_{k_q=1}^{K} \cdots \sum_{k_1=1}^{K} p(\phi^{(1:K)}, v|\mathcal{M}_1(k_1), \ldots, \mathcal{M}_q(k_q), \mathcal{D}_q) \\
&\quad \times p_q(k_q, \ldots, k_1),
\end{aligned}
$$

where $p_q(k_q, \ldots, k_{(q-h)}) = Pr[M_q(k_q), \ldots, \mathcal{M}_{q-h}(k_{q-h})|\mathcal{D}_q]$

Problem 1: *Number of components in the mixture is K^q.*
Problem 2: $p_q(k) \propto \pi_q(k) p(\mathbf{y}_q|\mathcal{M}_q(k), \mathcal{D}_{q-1})$, but $p(\mathbf{y}_q|\mathcal{M}_q(k), \mathcal{D}_{q-1})$ is not available in closed form.

Posterior inference

- Mixtures of AR(1) and AR(2) components: Approximations (Prado, 2010)
  - Approximate the K^q components by K^h for $h = 1, 2, 3$ for all q.
  - Filtering step: Combine several components into one via Kullback-Leibler divergence.
  - Under truncated normal-inverse gamma priors we obtain approximate truncated normal-inverse gamma posteriors.
- General case: Sequential Monte Carlo (Prado, 2012). Combines approches of Liu and West (2001) + Djurić (2001) + Structured priors on AR components.

Approximations

Under truncated normal priors on the AR coefficients we have the following:

- For $k = 1 : K$,

$$(\phi^{(k)}|v, \mathcal{M}_q(k_q), \mathcal{D}_q) \approx TN(\phi^{(k)}|m_q^{(k)}(k_q), vC_q^{(k)}(k_q), \mathcal{R}^{(k)})$$

- Posterior for v :

$$(v|\mathcal{M}_q(k_q), \mathcal{D}_q) \approx IG(n_q/2, d_q(k_q)/2)$$

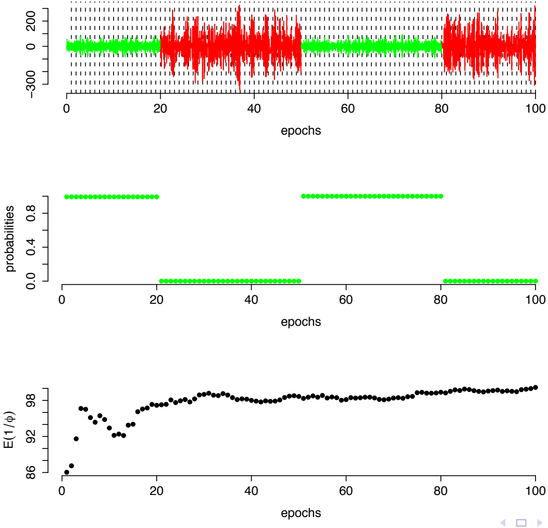
- Approximate expressions for $p_q(k_q, k_{q-1})$ and $p_q(k_q)$ can also be obtained.

Approximations

Example: Simulated AR(2) data. $T = 10,000$ data points were simulated from two AR(2) models in batches of 100 observations, to obtain 100 epochs, with $r_1 = 0.95$, $\lambda_1 = 6$, $r_2 = 0.99$ and $\lambda_2 = 16$ and $v = 100$.

- Priors: Truncated normal with approx. regions $(0.7, 1) \times (3, 8)$ for (r_1, λ_1) and $(0.7, 1) \times (11, 17)$ for (r_2, λ_2).
- Posteriors: $E(r_1|\mathcal{D}_{100}) \approx 0.956$, $E(\lambda_1|\mathcal{D}_{100}) \approx 6.015$, $E(r_2|\mathcal{D}_{100}) \approx 0.988$ and $E(\lambda_2|\mathcal{D}_{100}) \approx 16.104$.

Approximations



Analysis of latent EEG components from subject **skh** using multi-AR(1) approximations, model with $K = 2$.