1 Description of the problem

Consider the following stochastic volatility model:

$$y_t = \exp\left(\frac{x_t}{2}\right)\varepsilon_t \quad \text{such that } y_t \sim g(.|x_t) \tag{1}$$
$$x_t = \alpha x_{t-1} + \sigma\eta_t \quad \text{where } \eta_t \sim N(0,1). \tag{2}$$

It can be noticed that

$$x_t|\alpha\sigma \sim N\left(0, \frac{\sigma^2}{1-\alpha^2}\right) \tag{3}$$
$$x_t|x_{t-1}\alpha\sigma \sim N\left(\alpha x_{t-1}, \sigma^2\right). \tag{4}$$

Therefore it seems reasonable to assume

$$q_1(x_1) = N\left(0, \frac{\sigma^2}{1-\alpha^2}\right) \tag{5}$$
$$q_k(x_k|x_{k-1}) = N\left(\alpha x_{k-1}, \sigma^2\right), \tag{6}$$

where $q_{[.]}()$ is the generic sampling distribution. Suppose we are interested in the posterior distribution

$$\Pi(x_{1:n}, y_{1:n}) \propto \Pi(x_1)\prod_{k=2}^{n} N(\alpha x_{k-1})\prod_{k=1}^{n} g(y_k|x_k) := \gamma(x_{1:n}) \tag{7}$$

and suppose that $q_1(x_1) = \Pi(x_1)$ then a simple sequential importance sampling scheme would be

1.

$$x_1^i \sim q_1(x_1) \tag{8}$$
$$\gamma(x_1^i) = \Pi(x_1^i)g(y_1|x_1^i) \tag{9}$$
$$w(x_1^i) = \frac{q_1(x_1^i)g(y_1|x_1^i)}{q_1(x_1^i)} = g(y_1|x_1^i) \tag{10}$$

2.

$$x_{1:2}^i \sim q_1(x_1)q_2(x_2|x_1) \tag{11}$$
$$\gamma(x_{1:2}^i) = \Pi(x_1^i)N(x_2^i|\alpha x_1^i, \sigma^2)g(y_1|x_1^i)g(y_2|x_2^i) \tag{12}$$
$$w(x_{1:2}^i) = \frac{q_1(x_1^i)q_2(x_2^i|x_1^i)g(y_1|x_1^i)g(y_2|x_2^i)}{q_1(x_1^i)q_2(x_2^i|x_1^i)} = g(y_1|x_1^i)g(y_2|x_2^i) \tag{13}$$

3. ...

N.

$$
\begin{aligned}
x^i_{1:n} &\sim q_1(x_1)\prod_{k=2}^{n} q_2(x_k|x_{k-1}) && (14)\\
\gamma(x^i_{1:n}) &= \Pi(x^i_1)\prod_{k=2}^{n} N(x^i_k|\alpha x^i_{k-1},\sigma^2)\prod_{k=1}^{n} g(y_k|x^i_k) && (15)\\
w(x^i_{1:n}) &= \prod_{k=1}^{n} g(y_k|x^i_k) && (16)
\end{aligned}
$$

Can we apply the same scheme to a stochastic volatility model with long memory? Consider the following model

$$
\begin{aligned}
y_t &= \exp\left(\frac{x_t}{2}\right)\varepsilon_t \qquad \text{such that } y_t \sim g(.|x_t) && (17)\\
(1-\mathrm{L})^d x_t &= \sigma\eta_t \qquad \text{where } \eta_t \sim N(0,1). && (18)
\end{aligned}
$$

where

$$
\begin{aligned}
(1-\mathrm{L})^d &= 1+\sum_{j=1}^{\infty}\phi_j \mathrm{L}^j && (19)\\
\phi_j &= \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} = \prod_{0<k\leq j}\frac{k-1-d}{k} && (20)\\
\phi_j &\approx \frac{j^{-d-1}}{\Gamma(-d)} \qquad \text{as } j\to\infty && (21)\\
Cov(h) &= \frac{\Gamma(h+d)\Gamma(1-2d)}{\Gamma(h-d+1)\Gamma(1-d)\Gamma(d)}\sigma^2. && (22)
\end{aligned}
$$

It can be noticed that

$$
\begin{aligned}
x_t|d\sigma &\sim N\left[0, Cov(0)\right] && (23)\\
x_t|x_{(t-1)\dots x_{-\infty}}d\sigma &\sim N\left(-\sum_{j=1}^{\infty}\phi_j x_{t-j},\sigma^2\right). && (24)
\end{aligned}
$$

Does it make sense to consider the following proposals?

$$
\begin{aligned}
q_1(x_1) &= N\left[0, Cov(0)\right] && (25)\\
q_k(x_k|x_{k-1}\dots x_1) &= N\left(-\sum_{j=1}^{k+1}\phi_j x_{k-j},\sigma^2\right) && (26)
\end{aligned}
$$

NOTICE That the structure of long memory is such that the initial prior distribution $q_1()$ will affect the path for a long time (MIXING CONDITION PROBLEM??).

Hsu and Breidt (01), apparently propose a similar scheme. However, they use a Kullback-Leibler discrepancy criterion in order to choose the best truncated model to sample from. Then, by an importance sampling scheme they adjust for the truncation error. Assuming that the proposals are fine, the following scheme can be developed:

1.

$$\begin{aligned} x_1^i &\sim q_1(x_1) && (27)\\ \gamma(x_1^i) &= \Pi(x_1^i)g(y_1|x_1^i) && (28)\\ w(x_1^i) &= \frac{q_1(x_1^i)g(y_1|x_1^i)}{q_1(x_1^i)} = g(y_1|x_1^i) && (29) \end{aligned}$$

2.

$$\begin{aligned} x_{1:2}^i &\sim q_1(x_1)q_2(x_2|x_1)\\ \gamma(x_{1:2}^i) &= \Pi(x_1^i)p(x_2^i|x_1^i d\sigma^2)g(y_1|x_1^i)g(y_2|x_2^i)\\ w(x_{1:2}^i) &= \frac{p(x_2^i|x_1^i d\sigma^2)g(y_1|x_1^i)g(y_2|x_2^i)}{q_2(x_2^i|x_1^i)} && (30) \end{aligned}$$

3. ...

n.

$$\begin{aligned} x_{1:n}^i &\sim q_1(x_1)\prod_{k=2}^{n} q_2(x_k|x_{k-1})\\ \gamma(x_{1:n}^i) &= \Pi(x_1^i)\prod_{k=2}^{n} p(x_k^i|x_{k-1}^i d\sigma^2)\prod_{k=1}^{n} g(y_k|x_k^i)\\ w(x_{1:n}^i) &= \frac{\prod_{k=2}^{n} p(x_k^i|x_{k-1}^i d\sigma^2)}{\prod_{k=2}^{n} q_2(x_k^i|x_{k-1}^i)}\prod_{k=1}^{n} g(y_k|x_k^i) && (31) \end{aligned}$$

By using the Whittle approximation, it turns out that

$$\left[p(x_{2:n}^i|d\sigma^2)\right]^2 \approx (2\pi)^{-2n}\prod_{j=0}^{n-2}\frac{1}{f(\omega_j)}e^{-I(\omega_j)/f(\omega_j)} \quad (32)$$

Where $I(.)$ is the periodogram evaluated on $x_{2:n}^i$ and $f(.)$ is the theoretical spectral density. What about

$$\prod_{k=2}^{n} p(x_k^i|x_{k-1}^i d\sigma^2) = p(x_{2:n}^i|x_1 d\sigma^2) = ? \quad (33)$$

It is important to notice that

$$\lim_{\omega \to 0} f(\omega) = \infty, \tag{34}$$

therefore by using the Whittle approximation (32), there is going to be a problem. One possible solution could the expected value of the periodogram:

$$f(0) \approx E\left[I(0)\right] = \frac{1}{2\pi n}\left[n\gamma(0) + 2(n-1)\gamma(1) + 2(n-2)\gamma(2) + \ldots + 2\gamma(n-1)\right]. \tag{35}$$

By doing so we have a finite approximation of the spectral density which depends on the theoretical autocovariances of the ARFIMA model. Notice however, that the theoretical autocovariance structure is known only for an ARFIMA(0,d,0). Generalization for ARFIMA(p,d,q) can be obtain by specific algorithm based on circular convolution. I am not sure this is the right way to go because, eventually the algorithm for a general ARFIMA(p,d,q) will be extremely expensive.

2 Naive algorithm

Suppose that the following approximation holds (I am not sure, because long memory is strongly affected by the initial conditions)

$$p(x^i_{2:n}|x_1 d\sigma^2) \approx p(x^i_{2:n}|d\sigma^2) \tag{36}$$

then we could you implement the whittle approximation (32) to evaluate the weights in (31)

Chan and Petris propose to reduce the Infinite order Markovian representation as follows

$$\begin{aligned} x_t(1-L)^d &= \varepsilon_t && (37)\\ \Delta x_t(1-\mathrm{L})^{d-1} &= \varepsilon_t && (38)\\ \Delta x_t &= \sum_{j=0}^{\infty} \phi_j \varepsilon_{t-j} && (39)\\ \Delta x_t &\approx \sum_{j=0}^{M} \phi_j \varepsilon_{t-j} && (40) \end{aligned}$$

where ϕ_j derive from the expansion of $(1-\mathrm{L})^{1-d}$. Then, they use approximation (37) to write the likelihood. Can this be useful?

In the work of Comets, Fernandez and Ferrari (2002) they propose a regenerative construction and perfect simulation algorithm where they propose the following idea (this is Ferrari's answer to an email of mine)

to construct a random variable k_i that tells you how far in the past you have to go to compute the probability of x_i (the value of the chain at time i). You may be lucky and have a data big enough such that it is not necessary to go *further, but if not, you have to truncate and will commit some (hopefully small) error.*

From what I understand this may be an alternative way to choose the proposal distribution. Hsu and Breidt (01) use a Kullback-Leibler discrepancy criterion. On the other hand (26) implements all the info available (probably inefficient when we have a long series).

3 Better procedure?

Consider the following LMSV model

$$\begin{aligned} y_t &= \exp\left(\frac{x_t}{2}\right)\varepsilon_t \qquad \text{such that } y_t \sim g(.|x_t) \\ (1-\mathrm{L})^d x_t &= \sigma\eta_t \qquad \text{where } \eta_t \sim N(0,1), \end{aligned} \tag{41}$$

and consider the following reparametrization:

$$\begin{aligned} x_t &= -\sum_{j=1}^{\infty}\phi_j x_{t-j} + \sigma\mu_t && (42) \\ &= -\sum_{j=1}^{L-1}\phi_j x_{t-j} - \sum_{j=L}^{\infty}\phi_j x_{t-j} + \sigma\mu_t && (43) \\ &= -\sum_{j=1}^{L-1}\phi_j x_{t-j} - R_{t-L} + \sigma\mu_t. && (44) \end{aligned}$$

$$\Pi(x_t|I_{t-1}, d, \sigma_\mu, \sigma_\varepsilon) = N\left[-\sum_{j=1}^{L-1}\phi_j x_{t-j} - R_{t-L}, \sigma_\mu^2\right]. \tag{45}$$

It can be noticed that

$$\Pi(x_{t:-\infty}) = \underbrace{\Pi(R_0)\Pi(x_1|R_0)\prod_{L\geq 2|i=2}^{L}\Pi(x_i|x_{1:i-1}, R_0)}_{\text{block or move sample}}\prod_{i=L+1}^{t}\Pi(x_i|x_{i-L:i-1}, R_{L-i}) \tag{46}$$

and

$$\begin{aligned}
\Pi(R_0) &= N[0, \gamma(0)] \quad ????? && (47)\\
\Pi(x_1|R_0, \sigma^2) &= N(-R_0, \sigma^2) && (48)\\
\Pi(x_i|x_{1:i-1}, R_0, \sigma^2) &= N\left(-\sum_{j=1}^{i-1} \phi_j x_{i-j} - R_0, \sigma^2\right) && (49)\\
\Pi(R_1|R_0, \sigma_R^2) &= N[R_0, \gamma(0)] \qquad ????? && (50)\\
\Pi(x_i|x_{i-L+1:i-1}, R_{L-i}, \sigma^2) &= N\left(-\sum_{j=1}^{L-1} \phi_j x_{i-j} - R_{L-j}, \sigma^2\right) && (51)
\end{aligned}$$

Therefore

L

$$\begin{aligned}
R_0^i &\sim \Pi(R_0, d)\\
x_{1:L}^i &\sim \Pi(x_{1:L}|R_0^i, \sigma^2)\\
\gamma(x_{1:L}^i, R_0^i) &= \Pi(R_0^i)\Pi(x_{1:L}^i|R_0^i)\prod_{j=1}^{L} g(y_j|x_j^i)\\
w(x_{1:L}^i, R_0^i) &= \prod_{j=1}^{L} g(y_j|x_j^i) && (52)
\end{aligned}$$

L+1

$$\begin{aligned}
R_1^i &\sim \Pi(R_1|R_0^i, d)\\
x_{L+1}^i &\sim \Pi(x_{L+1}|x_{2:L}^i, R_1^i, \sigma^2)\\
\gamma(x_{1:L+1}^i, R_{0:1}^i) &= \Pi(R_0^i)\Pi(R_1^i|R_0^i)\Pi(x_{1:L}^i|R_0^i)\Pi(x_{L+1}|x_{2:L}^i, R_1^i, \sigma^2)\prod_{j=1}^{L+1} g(y_j|x_j^i)\\
w(x_{1:L+1}^i, R_{0:1}^i) &= \prod_{j=1}^{L+1} g(y_j|x_j^i) && (53)
\end{aligned}$$

$\vdots$

L+(t-L)

$$\begin{aligned}
R_{t-L}^i &\sim \Pi(R_{t-L}|R_{t-L-1}^i, d)\\
x_t^i &\sim \Pi(x_t|x_{t-L+1:t-1}^i, R_{t-L}^i, \sigma^2)\\
\gamma(x_{1:t}^i, R_{0:t-L}^i) &= \Pi(R_0^i)\prod_{j=1}^{t-L}\Pi(R_j^i|R_{j-1}^i)\Pi(x_{1:L}^i|R_0^i)\prod_{j=L+1}^{t}\Pi(x_j|x_{j-L+1:j-1}^i, R_{j-L}^i, \sigma^2)\prod_{j=1}^{t} g(y_j|x_j^i)\\
w(x_{1:t}^i, R_{0:1}^i) &= \prod_{j=1}^{t} g(y_j|x_j^i) && (54)
\end{aligned}$$

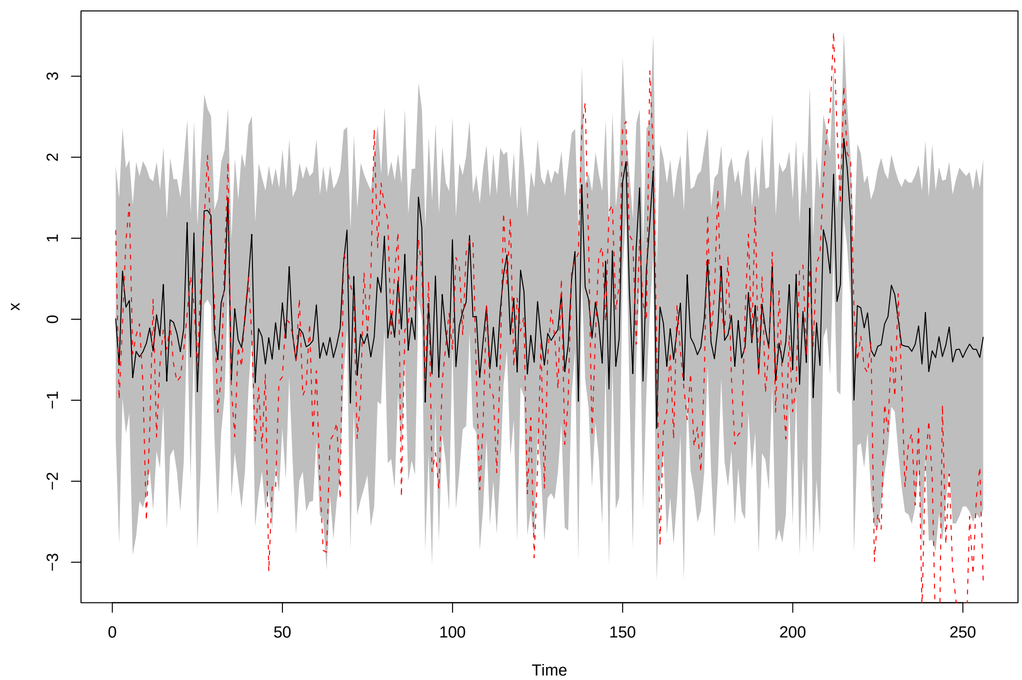

Figure 1: Sampling from Prior. No R_t augmentation. $L = 1$, $n = 2^8$, $N = 2^{12}$, $\sigma = 1$, $d = 0.49$, $\eta = 0.001$

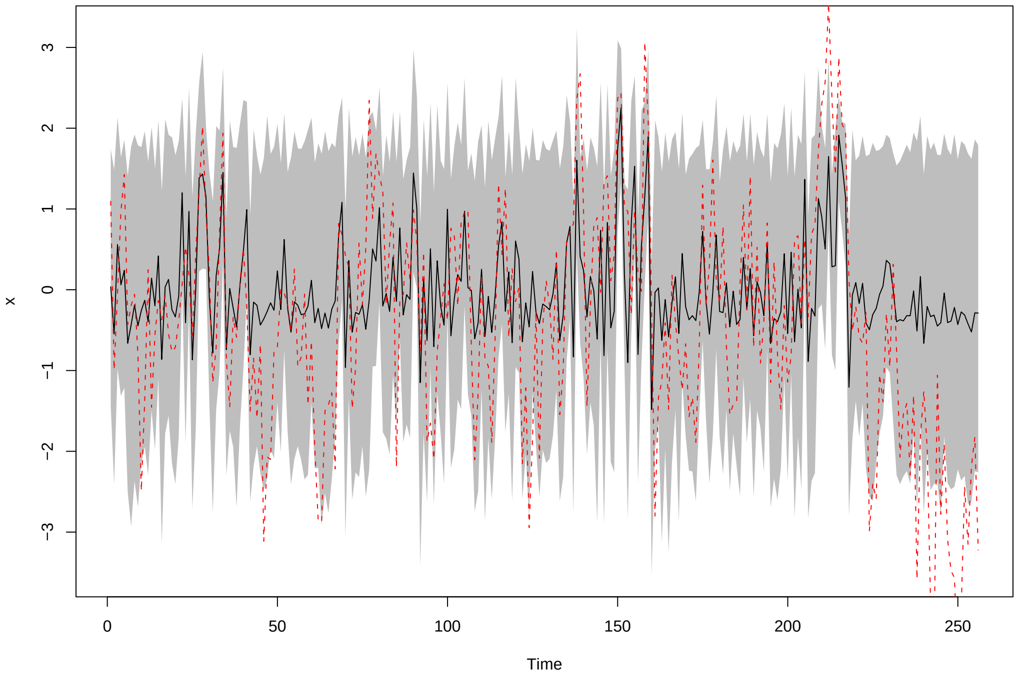

Figure 2: Sampling from Prior. No R_t augmentation. $L = 19$, $n = 2^8$, $N = 2^{12}$, $\sigma = 1$, $d = 0.49$, $\eta = 0.001$

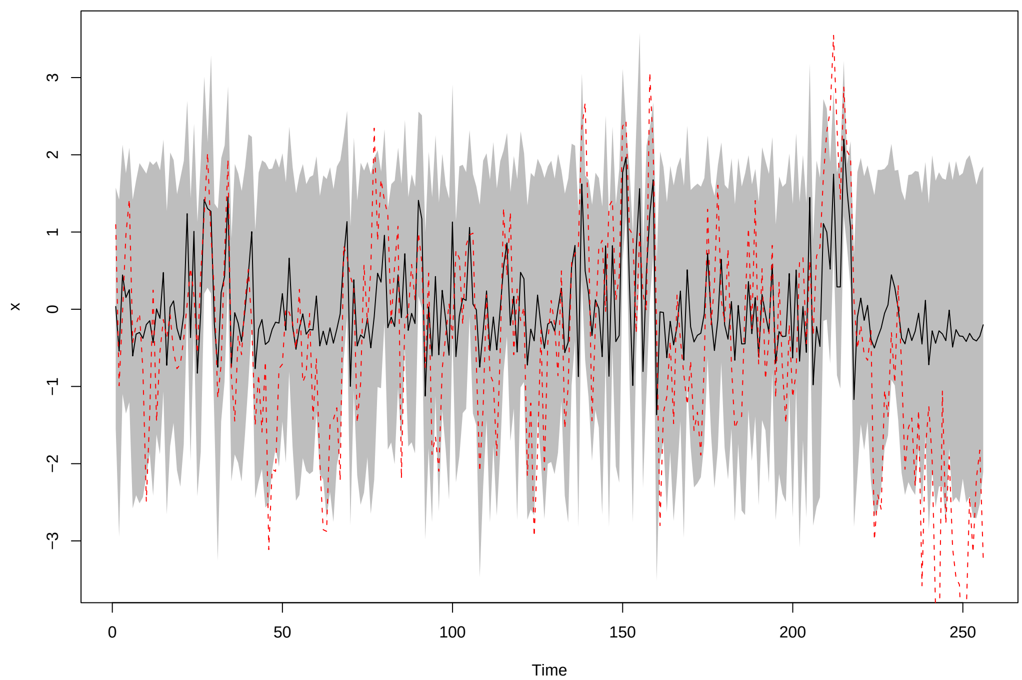

Figure 3: Sampling from Prior. No R_t augmentation. $L = 39$, $n = 2^8$, $N = 2^{12}$, $\sigma = 1$, $d = 0.49$, $\eta = 0.001$

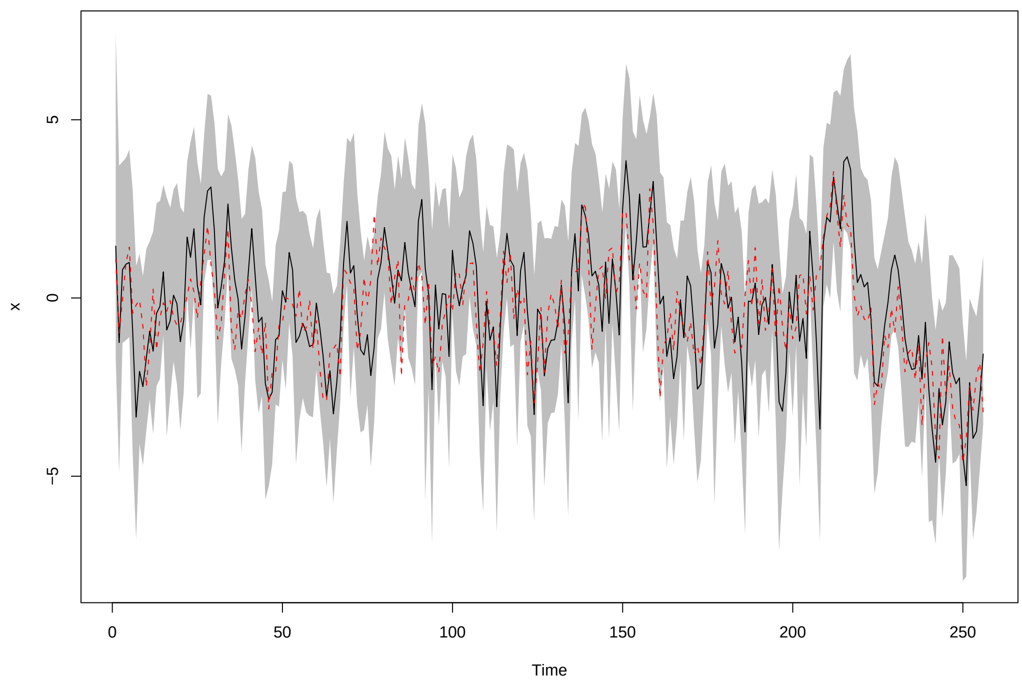

Figure 4: Sampling from Prior. R_t augmentation. $L = 1$, $n = 2^8$, $N = 2^{12}$, $\sigma = 1$, $d = 0.49$, $\eta = 0.001$

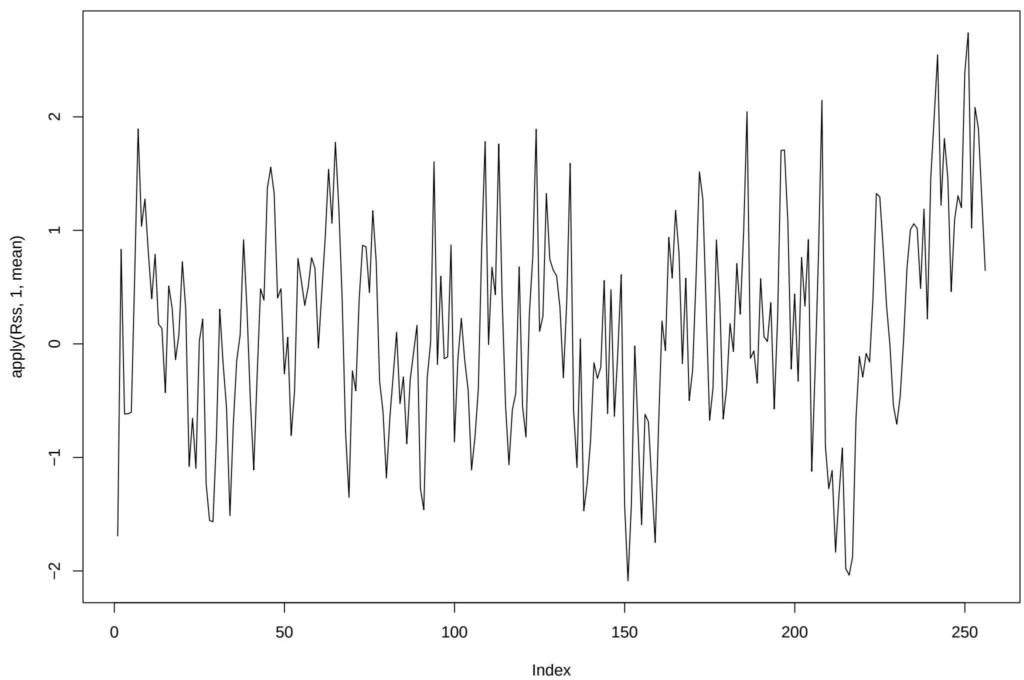

Figure 5: Sampling from Prior. R_t augmentation. $L = 1$, $n = 2^8$, $N = 2^{12}$, $\sigma = 1$, $d = 0.49$, $\eta = 0.001$

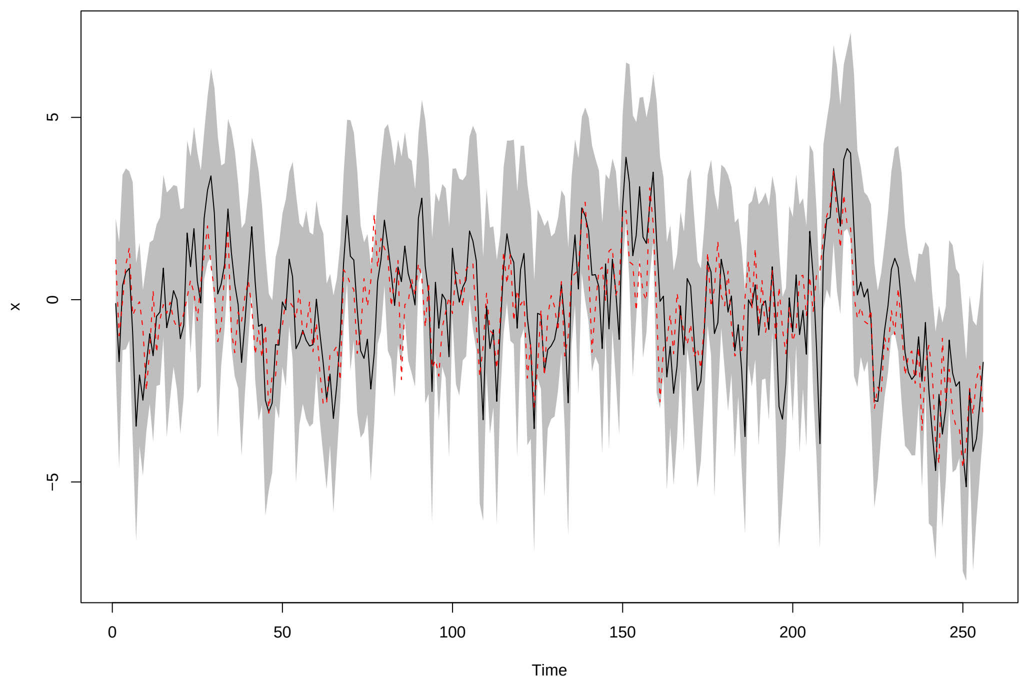

Figure 6: Sampling from Prior. R_t augmentation. $L = 19$, $n = 2^8$, $N = 2^{12}$, $\sigma = 1$, $d = 0.49$, $\eta = 0.001$

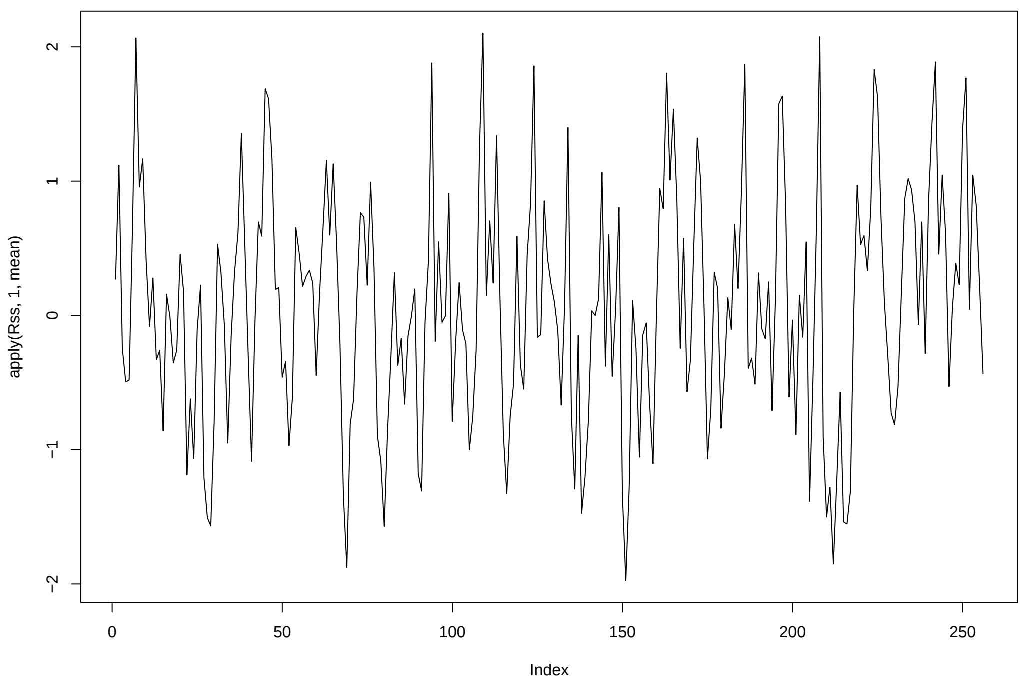

Figure 7: Sampling from Prior. R_t augmentation. $L = 19$, $n = 2^8$, $N = 2^{12}$, $\sigma = 1$, $d = 0.49$, $\eta = 0.001$

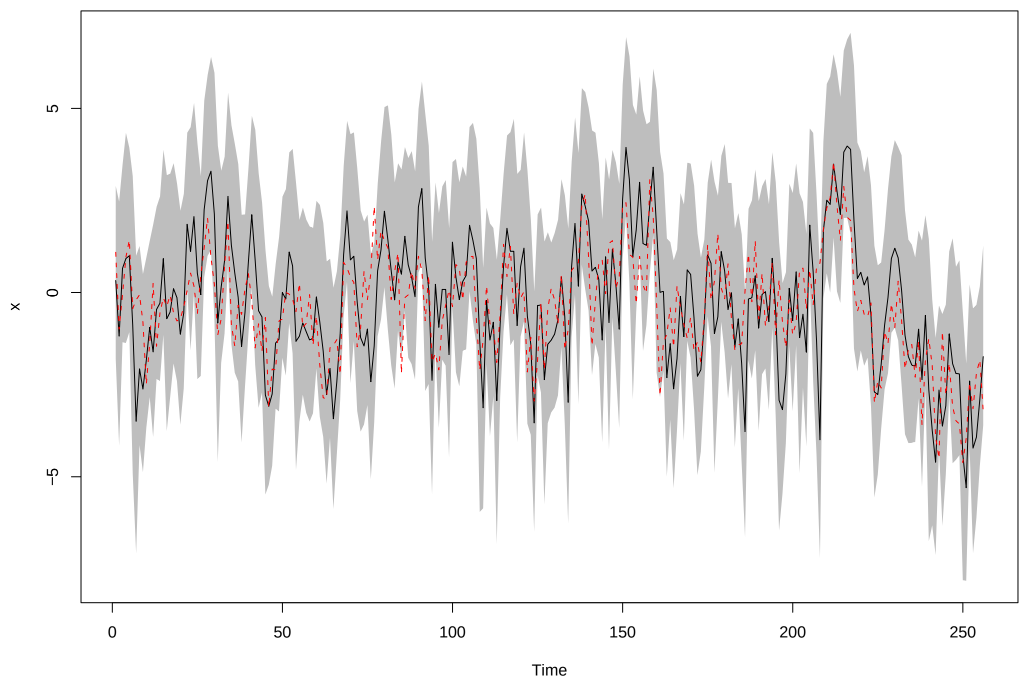

Figure 8: Sampling from Prior. R_t augmentation. $L = 39$, $n = 2^8$, $N = 2^{12}$, $\sigma = 1$, $d = 0.49$, $\eta = 0.001$

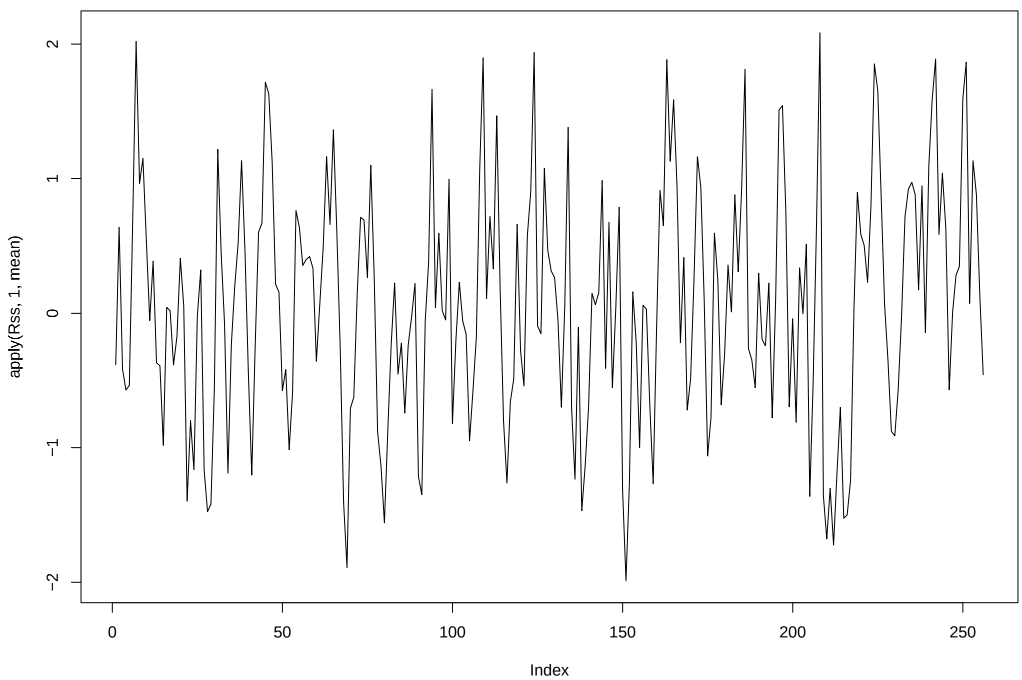

Figure 9: Sampling from Prior. R_t augmentation. $L = 39$, $n = 2^8$, $N = 2^{12}$, $\sigma = 1$, $d = 0.49$, $\eta = 0.001$

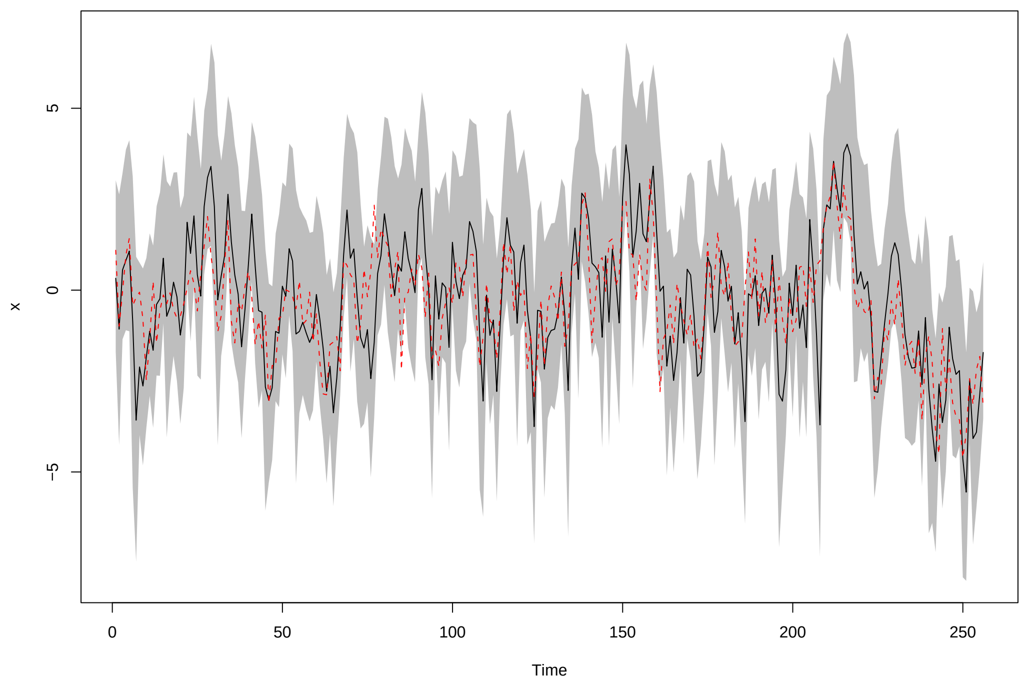

Figure 10: Sampling from Prior. R_t augmentation. $L = 99$, $n = 2^8$, $N = 2^{12}$, $\sigma = 1$, $d = 0.49$, $\eta = 0.001$

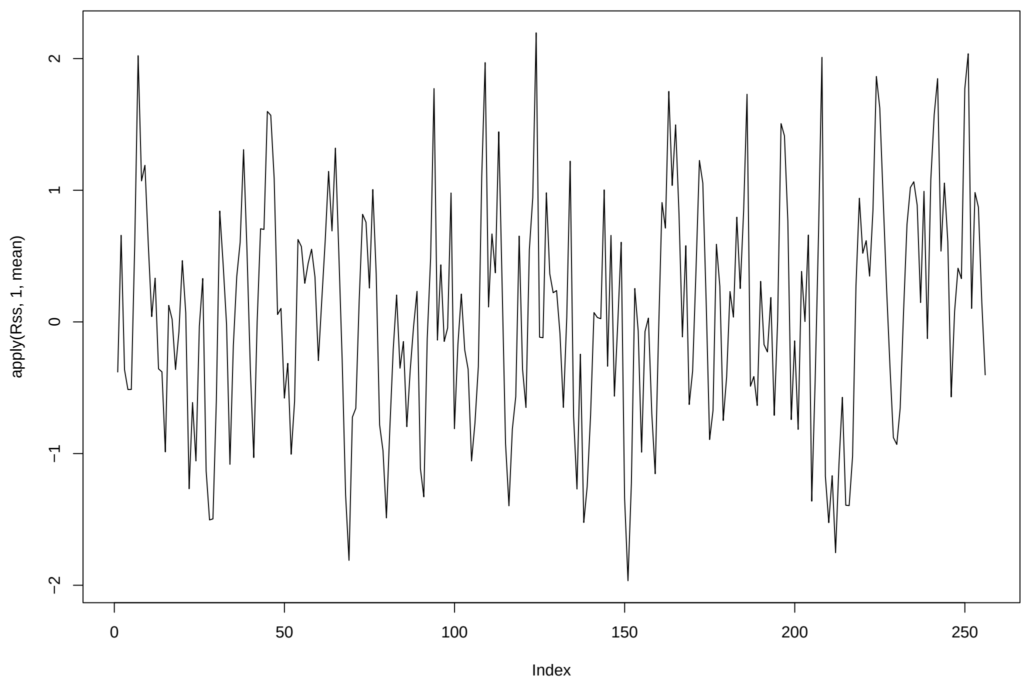

Figure 11: Sampling from Prior. R_t augmentation. $L = 99$, $n = 2^8$, $N = 2^{12}$, $\sigma = 1$, $d = 0.49$, $\eta = 0.001$