MCMC Documentation for Newcode

Georgi Dinolov

September 12, 2016

1 Simplified Model

We will consider the simplified stochastic volatility model with noise:

$$
\begin{aligned}
y_t &= \log(S_t) + \xi\varepsilon_t, & (1)\\
\log(S_t) &= \log(S_{t-\Delta}) + \exp(h_t)\varepsilon_{t,1}, & (2)\\
h_{t+\Delta} &= \alpha(\Delta) + \theta(\Delta)(h_t - \alpha(\Delta)) + \tau(\Delta)\varepsilon_{t,2}, & (3)\\
\varepsilon_t, \varepsilon_{t,1}, \varepsilon_{t,2} &\sim N(0,1).
\end{aligned}
$$

As before,

$$\alpha(\Delta) := \hat{\alpha} + \frac{1}{2}\log(\Delta), \quad \theta(\Delta) := e^{-\hat{\theta}\Delta}, \quad \tau^2(\Delta) := \frac{\hat{\tau}^2}{2\hat{\theta}}(1 - e^{-2\hat{\theta}\Delta}), \quad \hat{\theta} := 1/\text{Timescale}, \quad \frac{\hat{\tau}^2}{2\hat{\theta}} := \text{Long-run spot variance.}$$

Further, if the $\log(\hat{\sigma}_t)$ is the continuous-time OU process governing the volatility,

$$\exp(h_t) = \sqrt{\Delta}\hat{\sigma}_t, \qquad h_t = \log(\sqrt{\Delta}\hat{\sigma}_t).$$

I set the timescale for the volatility to 10 minutes (10*60*1000 milliseconds). Further, I set

$$\frac{\hat{\tau}^2}{2\hat{\theta}} = 0.116,$$

which is obtained by transforming the average VVX to the h_t (millisecond) scale. With $\hat{\theta}$ determined, this also determines $\hat{\tau}^2$. Finally,

$$\hat{\alpha} = -13,$$

which I got from the VIX. Finally,

$$\xi^2 = 6.5 \cdot 10^{-7},$$

which corresponds to a rather large (nickle to a dime) bid-ask spread. To summarize

$$\hat{\alpha} = -13, \qquad \hat{\theta} = 1.67 \cdot 10^{-6}, \qquad \hat{\tau}^2 = 1.93 \cdot 10^{-7}, \qquad \xi^2 = 6.50 \cdot 10^{-7}. \tag{4}$$

Using these parameters, I simulate data (implemented in the `generate.simulated.data(...)` function), with $\Delta = 10$ milliseconds. I will denote this data-generation increment as Δ_{gen}. Keeping Δ_{gen} this small ensures that the data record is a very close approximation to sampling from the continuous-time model. The thus-generated data vector of either true or noise-contaminated prices is then sub-sampled over index increments of length Δ/Δ_{gen} to obtain data samples from the (approximately) continuous-time model, observed every Δ milliseconds. Figure (1) below show a typical data set over a single trading day (6.5*60*60*1000 milliseconds). The difference between the true (red dashed) and noisy (solid black) prices is noticeable.

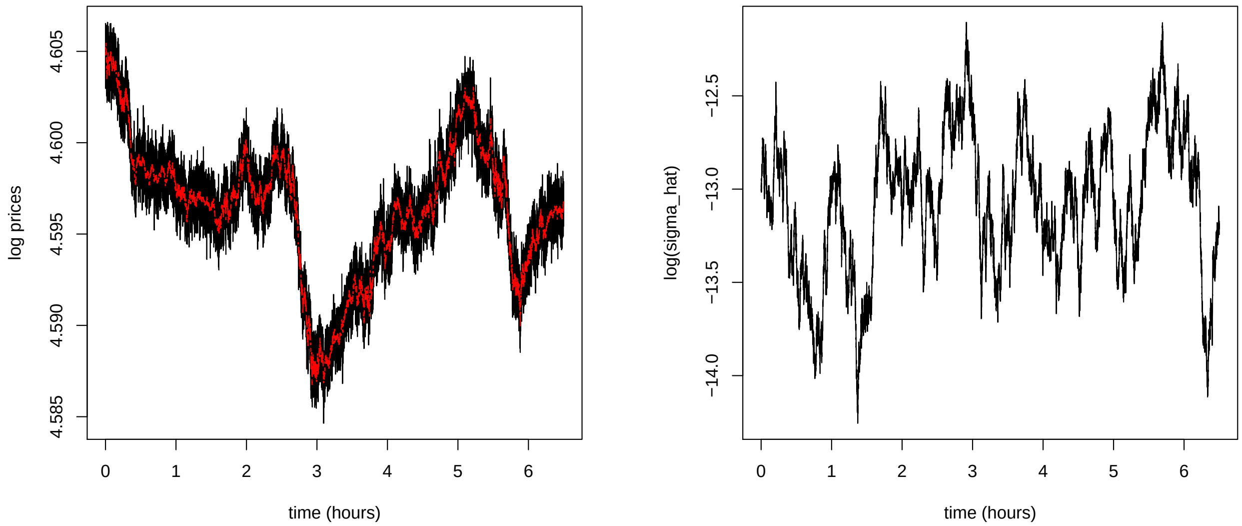

Figure 1: In the left panel are the true (red dashed) and noisy log-prices. The right panel shows the continuous-time volatility path. Each timeseries is sampled on the second from the (approximately) continuous-time model.

1.1 Linearizing the Volatility Term

This subsection is relevant when sampling h_t. We want to make (2) linear in h_t, with Gaussian innovations, to sample the posterior of $(h_\Delta, \ldots, h_T, h_{T+\Delta})$ using the Kalman Filter. To do this, we isolate the $\exp(h_t)\varepsilon_{t,1}$ and take the half-log of the square of each side of the equation:

$$\log(S_t) = \log(S_{t-\Delta}) + \exp(h_t)\varepsilon_{t,1} \qquad\qquad \underbrace{\frac{1}{2}\log\left((\log S_t/S_{t-\Delta})^2\right)}_{y_t^*} = h_t + \underbrace{\frac{1}{2}\log(\varepsilon_{t,1}^2)}_{\varepsilon_{t,*}}$$

The innovation term $\varepsilon_{t,*}$ is distributed as a scaled log-χ^2 random variable. The pdf of a log-χ^2 RV can be represented a weighted sum of normal densities, as in Omori et al. [2007]. In particular,

$$p(\log(\chi^2)) = \sum_{k=1}^{9} p_k\, N\left(\log(\chi^2)|m_k, v_k^2\right),$$

$$\Rightarrow p(\varepsilon_{t,*}) = p\left(\frac{1}{2}\log(\chi^2)\right) = \sum_{k=1}^{9} p_k\, N\left(\varepsilon_{t,*}\,\middle|\,\frac{m_k}{2}, \frac{v_k^2}{4}\right).$$

Introducing the latent variables $(\gamma_\Delta, \ldots, \gamma_T)$, with $\gamma_t \in \{1, \ldots, 9\}$,

$$p(\varepsilon_{t,*}|\gamma_t) = N(\varepsilon_{t,*}|m_{\gamma_t}/2, v_{\gamma_t}^2/4).$$

Subject to the above transformations, the model for the latent prices and volatilities becomes

$$y_t^* := \frac{1}{2}\log\left((\log(S_t/S_{t-\Delta}))^2\right) = h_t + \varepsilon_{t,*}, \quad (5)$$

$$h_{t+\Delta} = \alpha(\Delta) + \theta(\Delta)(h_t - \alpha(\Delta)) + \tau(\Delta)\varepsilon_{t,2}, \quad (6)$$

$$\varepsilon_{t,*}|\gamma_t \sim N(m_{\gamma_t}/2, v_{\gamma_t}^2/4), \quad (7)$$

$$p(\gamma_t) = p_{\gamma_t} \quad (8)$$

2 Posterior Sampling

For this version of the document (and code), I will only sample $\log(S_t), \gamma_t$, h_t, (α, θ, τ^2), and fixing ξ^2 to the true, data-generating parameter described above. The observational duration T is equal to a single trading day. Given a sampling period of Δ, the observational times are

$$t = 0, \Delta, \ldots, n\Delta,$$

where $n\Delta = T$, and there is a total of $n+1$ elements in the data vector. Let $\phi = (\alpha, \theta, \tau^2)$. The posterior sampling steps are:

1. Sample the joint posterior $p\left(\log(S_0), \ldots \log(S_{n\Delta}), \gamma_0, \ldots, \gamma_{n\Delta} | y_0, \ldots, y_{n\Delta}, h_0, \ldots, h_{n\Delta}, \phi, \xi^2\right)$
   a) Sample $p\left(\log(S_0), \ldots \log(S_{n\Delta}) | y_0, \ldots, y_{n\Delta}, h_0, \ldots, h_{n\Delta}, \phi, \xi^2\right)$
   b) Sample $p\left(\gamma_0, \ldots, \gamma_{n\Delta} | \log(S_0), \ldots \log(S_{n\Delta}), y_0, \ldots, y_{n\Delta}, h_0, \ldots, h_{n\Delta}, \phi, \xi^2\right)$
2. Sample the joint posterior $p\left(\phi, h_0, \ldots, h_{n\Delta} | \gamma_0, \ldots, \gamma_{n\Delta}, \log(S_0), \ldots \log(S_{n\Delta}), \xi^2\right)$ by
   a) Sampling $p\left(\phi | \gamma_0, \ldots, \gamma_{n\Delta}, \log(S_0), \ldots \log(S_{n\Delta}), \phi, \xi^2\right)$
   b) Sampling $p\left(h_0, \ldots, h_{n\Delta} | \phi, \gamma_0, \ldots, \gamma_{n\Delta}, \log(S_0), \ldots \log(S_{n\Delta}), \phi, \xi^2\right)$

2.1 Sampling $\log(S_t)|y, h_t, \phi, \xi^2$

To sample $\log(S_t)$, we can consider equations (1) and (2). This is equivalent to sampling from

$$p\left(\log(S_0), \ldots \log(S_{n\Delta}) | y_0, \ldots, y_{n\Delta}, h_0, \ldots, h_{n\Delta}, \phi, \xi^2\right).$$

Note that using equations (1) and (2) is equivalent to *integrating out* γ_t from the posterior. **Sampling** $\log(S_t)$ **has be done before sampling** γ_t.

For the distribution of the latent term $\log(S_{-\Delta})$, I will assume

$$\log(S_{-\Delta}) \sim N(y_0, \xi^2),$$
$$\mu_{-\Delta} := y_{-\Delta}, \quad \sigma_{-\Delta}^2 := \xi^2.$$

The forward filtering equations are

$$\begin{aligned}
p(\log(S_t)|y_{0:t-\Delta}, h_{\Delta:T+\Delta}, \phi, \xi^2) &= N(\log(S_t)|m_t, s_t^2), \\
m_t &:= \mu_{t-\Delta}, \\
s_t^2 &:= \sigma_{t-\Delta}^2 + \exp(2h_t),
\end{aligned}$$

$$\begin{aligned}
p(\log(S_t)|y_{0:t}, h_{\Delta:T+\Delta}, \phi, \xi^2) &= N(\log(S_t)|\mu_t, \sigma_t^2), \\
\mu_t &:= \frac{y_t}{\frac{\xi^2}{s_t^2} + 1} + \frac{m_t}{\frac{s_t^2}{\xi^2} + 1}, \\
\sigma_t^2 &:= \left(\frac{1}{\xi^2} + \frac{1}{s_t^2}\right)^{-1}.
\end{aligned}$$

Iterating over all t we obtain the posterior distribution of $\log(S_T)$

$$\log(S_T)|y_{0:T} \sim N(\mu_T, \sigma_T^2).$$

The Markov structure of the model permits us to write

$$p(\log(S_t)|y_{0:T}, \log(S_T), \ldots, \log(S_{t+\Delta})) = p(\log(S_t)|y_{0:t}, \log(S_{t+\Delta}))$$

Thus, conditional on $\log(S_{t+\Delta})$,

$$\begin{aligned} p(\log(S_t)|y_{0:t}, \log(S_{t+\Delta})) &\propto p(\log(S_{t+\Delta})|\log(S_t))p(\log(S_t)|y_{0:t}) \\ &\propto N(\log(S_{t+\Delta})|\log(S_t), \exp(2h_{t+\Delta}))N(\log(S_t)|\mu_t, \sigma_t^2) \\ &\propto N\left(\log(S_t)\middle|\frac{\log(S_{t+\Delta})}{\exp(2h_{t+\Delta})/\sigma_t^2+1} + \frac{\mu_t}{\sigma_t^2/\exp(2h_{t+\Delta})+1}\right). \end{aligned}$$

In this way we obtain the joint posterior sample for $(\log(S_0), \ldots, \log(S_T))$.

2.2 Sampling $\gamma_t|\log(S_t), y_t, h_t, \phi, \xi^2$

The posterior samples for γ_t are particularly easy since there is no leverage in this model. Note here the for γ_t, $t = \Delta, \ldots, n\Delta$, for a total of n latent variables.

$$p(\gamma_t|y_t^*, h_t) \propto N(y_t^*|h_t + m_{\gamma_t}/2, v_{\gamma_t}^2/4)p(\gamma_t)$$

2.3 Sampling $\phi|y_t^*, \gamma_t, \xi^2$

To sample ϕ, we first integrate out h_t using the forward filtering equations of the Kalman Filter. I'll use the same notation as for sampling $\log(S_t)$. The prior for h_0 is

$$\begin{gathered} h_0 \sim N(\alpha, \tau^2/(1-\theta^2)), \\ \mu_0 := \alpha, \quad \sigma_0^2 := \frac{\tau^2}{1-\theta^2}. \end{gathered}$$

The forward filtering equations for h_t are

$$\begin{aligned} p(h_t|y_{0:t-\Delta}^*, \gamma_{0:T}, \phi, \gamma_t) &= N(h_t|M_t, s_t^2), \\ M_t &:= \alpha(\Delta)(1-\theta(\Delta)) + \theta(\Delta)\mu_{t-\Delta}, \\ s_t^2 &:= \tau^2(\Delta) + \theta^2(\Delta)\sigma_{t-\Delta}^2, \end{aligned} \tag{9}$$

$$\begin{aligned} p(h_t|y_{0:t}^*, \gamma_{0:T}, \phi, \gamma_t) &= N(h_t|\mu_t, \sigma_t^2), \\ \mu_t &:= \frac{y_t^* - m_{\gamma_t}/2}{\frac{v_{\gamma_t}^2/4}{s_t^2}+1} + \frac{M_t}{\frac{s_t^2}{v_{\gamma_t}^2/4}+1}, \\ \sigma_t^2 &:= \left(\frac{1}{v_{\gamma_t}^2/4} + \frac{1}{s_t^2}\right)^{-1}. \end{aligned} \tag{10}$$

Then

$$p(y_t^*|y_\Delta^*, \ldots, y_{t-\Delta}^*, \phi, \gamma_t) = \int p(y_t^*|h_t, y_\Delta^*, \ldots, y_{t-\Delta}^*, \phi, \gamma_t)p(h_t|y_\Delta^*, \ldots, y_{t-\Delta}^*, \phi, \gamma_t)dh_t = N(y_t^*|M_t + m_{\gamma_t}/2, v_{\gamma_t}^2/4 + s_t^2),$$

so

$$p(y^*_\Delta, \ldots, y^*_T | \phi, \gamma_{\Delta:T}) = \prod_{t=\Delta}^{T} p(y^*_t | y^*_\Delta, \ldots, y^*_{t-\Delta}, \phi, \gamma_t).$$

The posterior

$$p(\phi | y^*_{\Delta:T}, \gamma_{\Delta:T}) \propto p(y^*_{\Delta:T} | \phi, \gamma_{\Delta:T}) p(\phi)$$

is sampled through a Metropolis-Hastings step, where the proposal distribution is subject to tuning. The priors for the parameters in ϕ are described in the paper.

2.4 Sampling $h_t | \phi, y^*_t, \gamma_t, \xi^2$

We sample h_t for all t using the Forward Filtering equations (9) and (10) and the Backward Sampling step

$$\begin{aligned}
p(h_t | y^*_{0:t}, h_{t+\Delta}) &\propto p(h_{t+\Delta} | h_t) p(h_t | y^*_{0:t}) \\
&\propto N(h_{t+\Delta} | \alpha(\Delta)(1-\theta(\Delta)) + \theta(\Delta) h_t, \tau^2(\Delta)) N(h_t | \mu_t, \sigma^2_t) \\
&\propto N\left(h_t \,\middle|\, \frac{[h_{t+\Delta} - \alpha(\Delta)(1-\theta(\Delta))]/\theta(\Delta)}{(\tau^2(\Delta)/\theta(\Delta)^2)/\sigma^2_t + 1} + \frac{\mu_t}{\sigma^2_t/(\tau^2(\Delta)/\theta(\Delta)^2) + 1}\right).
\end{aligned}$$

References

Yasuhiro Omori, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima. Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140(2):425–449, 2007.