Power-law SV model

by Hoai Nam Nguyen

April 4, 2018

In this document, I refer to Chan and Grant (2014) unless stated otherwise. The algorithms described in this paper are implemented by the code we are working on. I want to adapt these algorithms to estimating the parameters in our new PL-SV model. To be consistent with their notations, I write the observation equation of our model in terms of log volatility h_t as follow:

$$y_t = \mu + \lambda(e^{h_t})^{\kappa/2} + \epsilon_t$$

where $\epsilon_t \sim N(0, e^{h_t})$

On page 7, they propose a 5-step algorithm to estimate the standard SV model by sequentially sampling from the posterior distributions as follow:

1. $p(\mathbf{h}|\mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2)$
2. $p(\mu|\mathbf{y}, \mathbf{h}, \mu_h, \phi_h, \omega_h^2) = p(\mu|\mathbf{y}, \mathbf{h})$
3. $p(\mu_h|\mathbf{y}, \mu, \mathbf{h}, \phi_h, \omega_h^2) = p(\mu_h|\mathbf{h}, \phi_h, \omega_h^2)$
4. $p(\omega_h^2|\mathbf{y}, \mu, \mathbf{h}, \mu_h, \phi_h) = p(\omega_h^2|\mathbf{h}, \mu_h, \phi_h)$
5. $p(\phi_h|\mathbf{y}, \mu, \mathbf{h}, \mu_h, \omega_h^2) = p(\phi_h|\mathbf{h}, \mu_h, \omega_h^2)$

We don't really care about the last 3 steps since they only involve the measurement equation, which stays unchanged. But we have to adjust step 1 and step 2 to make it work for our model.

STEP 1

In a nutshell, they come up with an approximation of the log density function:

$$log\ p(\mathbf{h}|\mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2) \approx -\tfrac{1}{2}(\mathbf{h'K_hh} - 2\mathbf{h'k_h}) + const$$

Please refer to page 17 for the definitions of the matrix $\mathbf{K_h}$ and the vector $\mathbf{k_h}$. Then, they sample from it numerically using Newton-Raphson method until the difference between two consecutive iterations are close enough within a specified threshold. That is, the iteration is given by:

$$\mathbf{h^{(t+1)}} = \mathbf{h^{(t)}} + \mathbf{K_h^{-1}}(-\mathbf{K_hh(t)} + \mathbf{k_h}) = \mathbf{K_h^{-1}k_h}$$

The iterations stop when $||\mathbf{h^{(t+1)}} - \mathbf{h^{(t)}}|| < c$ for some specified tolerance level c. Hence, we can see that they have a while loop in the code, which accounts for most of the computational time of the algorithm.

Note that the matrix $\mathbf{G}$ and the vector $\mathbf{f}$, which are among of the components of the matrix $\mathbf{K_h}$ and the vector $\mathbf{k_h}$ above, are defined in terms of the following derivatives:

$$\frac{\partial}{\partial h_t} log\ p(y_t|\mu, h_t)$$

$$\frac{\partial^2}{\partial h_t^2} log\ p(y_t|\mu, h_t)$$

This is easy to figure out since we know that $y_t|\mu, h_t \sim N(\mu, e^{h_t})$ for the standard SV model.

On page 19, they modify the method slightly to adapt it to the SV-M model. In particular, the log volatility $log\ p(y_t|\mu, \lambda, h_t)$ is now different because $y_t|\mu, \lambda, h_t \sim N(\mu + \lambda e^{h_t/2}, e^{h_t})$. Therefore, the derivatives above need to be recalculated, leading to different $\mathbf{G}$ and $\mathbf{f}$. Once the components have been modified, we can sample h as before by iterating until convergence.

Therefore, I suggest we adopt the same approach here. It's easy to see that $y_t|\mu, \lambda, \kappa, h_t \sim N(\mu + \lambda(e^{h_t})^{\kappa/2}, e^{h_t})$. I have recalculated the derivatives as follow:

$$\frac{\partial(log\ p)}{\partial h_t} = -\frac{1}{2} + \frac{1}{2}(y_t - \mu)^2 e^{-h_t} - \frac{1}{2}\lambda^2(\kappa - 1)e^{(\kappa-1)h_t} + \lambda(y_t - \mu)(\frac{\kappa}{2} - 1)(e^{h_t})^{\kappa/2-1}$$

$$\frac{\partial^2(log\ p)}{\partial h_t^2} = -\frac{1}{2}(y_t - \mu)^2 e^{-h_t} - \frac{1}{2}\lambda^2(\kappa - 1)^2 e^{(\kappa-1)h_t} + \lambda(y_t - \mu)(\frac{\kappa}{2} - 1)^2(e^{h_t})^{\kappa/2-1}$$

The SV-M model is a special case of our model with $\kappa = 2$. If we substitute $\kappa = 2$ in the equations above, we will recover the derivatives on page 19 for the SV-M model. Then, we can redefine $\mathbf{G}$ and $\mathbf{f}$ and conduct the iterations as before.

STEP 2

We know that the SV-M model is written as follow:

$$y_t = \mu + \lambda e^{h_t} + \epsilon_t$$

where $\epsilon_t \sim N(0, e^{h_t})$

On page 20, they rewrite the model in matrix form:

$$\boldsymbol{y} = \boldsymbol{X_\beta \beta} + \boldsymbol{\epsilon}$$

where $\boldsymbol{y} = (y_1, \ldots, y_T)'$, $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_T)'$, $\boldsymbol{\beta} = (\mu, \lambda)'$ and the matrix $\boldsymbol{X_\beta}$ is given by:

$$\boldsymbol{X_\beta} = \begin{pmatrix} 1 & e^{h_1} \\ \vdots & \vdots \\ 1 & e^{h_T} \end{pmatrix}$$

Given that $\mu \sim N(\mu_0, V_\mu)$ and $\lambda \sim N(\lambda_0, V_\lambda)$ are the prior densities, we can deduce that $\boldsymbol{\beta} \sim N(\boldsymbol{\beta_0}, \boldsymbol{V_\beta})$ is the prior for $\boldsymbol{\beta}$, where $\boldsymbol{V_\beta} = diag(V_\mu, V_\lambda)$ and $\boldsymbol{\beta_0} = (\mu_0, \lambda_0)'$. By using standard results, we have the following posterior density for the joint distribution of μ and λ:

$$(\mu, \lambda | \boldsymbol{y}, \boldsymbol{h}) \sim N(\hat{\boldsymbol{\beta}}, \boldsymbol{D_\beta})$$

This is a bivariate normal distribution with $\boldsymbol{D_\beta^{-1}} = \boldsymbol{V_\beta^{-1}} + \boldsymbol{X'_\beta \Sigma_y^{-1} X_\beta}$ and $\hat{\boldsymbol{\beta}} = \boldsymbol{D_\beta}(\boldsymbol{V_\beta^{-1}\beta_0} + \boldsymbol{X'_\beta \Sigma_y^{-1} y})$ with $\boldsymbol{\Sigma_y} = diag(e^{h_1}, \dots, e^{h_T})$

For our PL-SV model, I suggest adopting this approach to sample from $(\mu, \lambda | \boldsymbol{y}, \boldsymbol{h}, \kappa)$. The only difference will be the definition of the matrix $\boldsymbol{X_\beta}$:

$$\boldsymbol{X_\beta} = \begin{pmatrix} 1 & (e^{h_1})^{\kappa/2} \\ \vdots & \vdots \\ 1 & (e^{h_T})^{\kappa/2} \end{pmatrix}$$

Sampling $\boldsymbol{\kappa}$

Between step 2 and step 3 in the algorithm above, we need to add an intermediate step to sample κ. That is, we do it after sampling μ and λ jointly but before moving to the measurement equation. Specifically, we need to sample from the following density:

$$p(\kappa | \boldsymbol{y}, \boldsymbol{h}, \mu, \lambda)$$

Then, we have the following implications:

$$\begin{aligned} p(\kappa | \boldsymbol{y}, \boldsymbol{h}, \mu, \lambda) &\propto p(\boldsymbol{y}, \boldsymbol{h}, \mu, \lambda | \kappa) p(\kappa) \\ &= p(\boldsymbol{y} | \boldsymbol{h}, \mu, \lambda, \kappa) p(\boldsymbol{h}, \mu, \lambda | \kappa) p(\kappa) \\ &= p(\boldsymbol{y} | \boldsymbol{h}, \mu, \lambda, \kappa) p(\boldsymbol{h}, \mu, \lambda) p(\kappa) \quad \text{by independence} \\ &\propto p(\boldsymbol{y} | \boldsymbol{h}, \mu, \lambda, \kappa) p(\kappa) \end{aligned}$$

Thus, we can sample from $p(\boldsymbol{y} | \boldsymbol{h}, \mu, \lambda, \kappa) p(\kappa)$, where $p(\kappa)$ is the prior for κ. The conditional density $p(\boldsymbol{y} | \boldsymbol{h}, \mu, \lambda, \kappa)$ will just be a multivariate Normal distribution because the errors are normally distributed.

Although we can find the posterior density for κ as above, I imagine that it's impossible to describe this density by any well-known distribution. Therefore, we may need to resort to numerical methods to do it. The simplest method to sample from a continuous distribution with CDF $F(x)$ is to sample U from $Uniform(0,1)$ first and calculate $F^{-1}(U)$. Of course, getting $F(x)$ in analytic form can be difficult in this case. However, we can divide the interval [-1,1] (the domain of κ) into small sub-intervals. For each end point x^*, we calculate the following integral numerically:

$$\textstyle\int_{-1}^{x^*} p(\kappa|\boldsymbol{y}, \boldsymbol{h}, \mu, \lambda)$$

Then, we choose the x^* such that the value of the integral above is closest to U.

Please let me know if you know any numerical method to sample from an arbitrary density.

Side note

Now I'm aware that λ must be strictly positive, we cannot sample μ and λ simultaneously. Therefore, I suggest sampling from the 3 parameters separately:

1. Sample from $p(\mu|\mathbf{y}, \mathbf{h}, \lambda, \kappa)$

   Just like before, we can prove that this is proportional to $p(\boldsymbol{y}|\boldsymbol{h}, \mu, \lambda, \kappa)p(\mu)$. We can assume a Normal prior for μ as usual.

2. Sample from $p(\lambda|\mathbf{y}, \mathbf{h}, \lambda, \kappa)$

   This is proportional to $p(\boldsymbol{y}|\boldsymbol{h}, \mu, \lambda, \kappa)p(\lambda)$. We can use a Gamma prior for λ

3. Sample from $p(\kappa|\mathbf{y}, \mathbf{h}, \lambda, \kappa)$

   This is proportional to $p(\boldsymbol{y}|\boldsymbol{h}, \mu, \lambda, \kappa)p(\kappa)$. We can try a Beta or a Uniform prior for κ