Estimation of Stochastic Volatility Models

A simple univariate SV model

$$\begin{aligned} r_t &= \beta_0 + \beta_1 r_{t-1} + a_t \\ a_t &= \sqrt{h_t}\epsilon_t, \quad \epsilon_t \sim \text{iid} \quad N(0,1) \\ \ln h_t &= \alpha_0 + \alpha_1 \ln(h_{t-1}) + v_t, \quad v_t \sim \text{iid} \quad N(0, \sigma_v^2) \end{aligned}$$

For simplicity, the shocks $\{\epsilon_t\}$ and $\{v_t\}$ are assumed to be independent. See Section 10.7 of the text for a general SV model.

Parameters of the model:

- Traditional: $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ and $\boldsymbol{\omega} = (\alpha_0, \alpha_1, \sigma_v^2)'$
- Auxiliary: unobservable volatility $\boldsymbol{H} = (h_1, h_2, \cdots, h_n)'$, where n is the sample size.

Likelihood function:

$$f(r_1, \cdots, r_n | \boldsymbol{\beta}, \boldsymbol{\omega}) = \int f(r_1, \cdots, r_n | \boldsymbol{H}, \boldsymbol{\beta}) f(\boldsymbol{H} | \boldsymbol{\omega}) d\boldsymbol{H},$$

which is complicated involving a n-dimensional integration over $\boldsymbol{H}$. What can be done?

Answer: Markov Chain Monte Carlo (MCMC) methods

Gibbs sampling: An iterative procedure based on the properties of Markov chain to obtain a joint distribution from conditional distributions.

Procedure: Consider the case of three parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$. Suppose that the joint distribution is complicated, but the three conditional distributions

$$f(\theta_1|\theta_2, \theta_3), \quad f(\theta_2|\theta_3, \theta_1), \quad f(\theta_3|\theta_1, \theta_2)$$

are simple, i.e. it is easy to draw random samples from the three conditional distributions. Then, we can employ the Gibbs sampler as follows:

1. Specify some arbitrary initial values $(\theta_{1,0}, \theta_{2,0}, \theta_{3,0})$.
2. Draw a random variate $\theta_{1,1}$ from $f(\theta_1|\theta_{2,0}, \theta_{3,0})$.
3. Draw a random variate $\theta_{2,1}$ from $f(\theta_2|\theta_{3,0}, \theta_{1,1})$.
4. Draw a random variate $\theta_{3,1}$ from $f(\theta_3|\theta_{1,1}, \theta_{2,1})$.

   Steps 2 to 4 give a new set of parameters $(\theta_{1,1}, \theta_{2,1}, \theta_{3,1})$
5. Use the new parameters as given and repeat Steps 2 to 4.

The procedure is iterated for many times, say m. Under some regularity conditions, the parameter $(\theta_{1,m}, \theta_{2,m}, \theta_{3,m})$ constitutes a random draw from the joint distribution $f(\theta_1, \theta_2, \theta_3)$.

In practice, one can further use properties of Markov chain and repeat the above procedure for many many times, say N, then discard the first m iterations as *burn-ins*. This results in a random sample

$$(\theta_{1,m+1}, \theta_{2,m+1}, \theta_{3,m+1}), \cdots, (\theta_{1,N}, \theta_{2,N}, \theta_{3,N}).$$

The sample size is $N - m$.

We can use this random sample to make inference about the parameters. For instance, how to estimate θ_1?

Answer: sample mean of θ_1, i.e.

$$\hat{\theta}_1 = \frac{1}{N-m} \sum_{i=1}^{N-m} \theta_{1,m+i}.$$

The sample variance of θ_1 is

$$\hat{\sigma}_1^2 = \frac{1}{N-m-1} \sum_{i=1}^{N-m} (\theta_{1,m+i} - \hat{\theta}_1)^2.$$

As another example, what is the estimate of $\theta_1 - \theta_2$?

Answer: use the Gibbs sample to obtain a random sample

$$\{\theta_{1,i} - \theta_{2,i}\}_{i=m+1}^{N}$$

for the quantity of interest $\theta_1 - \theta_2$.

The key question then is: How to obtain the conditional distributions? Check convergence?

Bayesian inference becomes useful.

Concept: *Posterior dist* $\propto$ *Likelihood* $\times$ *Prior dist.*

For the SV model considered, we have

$$f(\boldsymbol{\beta}, \boldsymbol{\omega} | r_1, \cdots, r_n) \propto f(r_1, \cdots, r_n | \boldsymbol{\beta}, \boldsymbol{\omega}) f(\boldsymbol{\beta}, \boldsymbol{\omega}).$$

In Gibbs sampling, you may think of the prior distributions as the distributions used to draw the initial estimates for the iteration.

In practice, independent prior distributions are often used, i.e.

$$f(\boldsymbol{\beta}, \boldsymbol{\omega}) = f(\boldsymbol{\beta}) f(\boldsymbol{\omega}) = f(\boldsymbol{\beta}) f(\alpha_0, \alpha_1) f(\sigma_v^2).$$

Furthermore, simplicity often plays a role in prior selection.

Return to SV model

Let $\boldsymbol{R} = (r_1, \cdots, r_n)$ be the return series. The cond. dists needed are

$$f(\boldsymbol{\beta} | \boldsymbol{R}, \boldsymbol{H}, \boldsymbol{\omega}), \quad f(\boldsymbol{\omega} | \boldsymbol{R}, \boldsymbol{H}, \boldsymbol{\beta}), \quad f(\boldsymbol{H} | \boldsymbol{R}, \boldsymbol{\beta}, \boldsymbol{\omega}).$$

Details are given below: (Sec. 10.7.1 of the text)

- Given $\boldsymbol{H}$, one can transform the mean equ into a simple linear regression model

$$\frac{r_t}{\sqrt{h_t}} = \beta_0 \frac{1}{\sqrt{h_t}} + \beta_1 \frac{r_{t-1}}{\sqrt{h_t}} + \epsilon_t, \quad t = 2, \cdots, n$$

where $\epsilon_t \sim$ iid $N(0,1)$. Consequently, β_0 and β_1 can be estimated by the linear regression method. In particular, if one uses a normal prior distribution for $\boldsymbol{\beta}$, then the conditional dist of $\boldsymbol{\beta}$ is normal. The random draw can easily be done.

- For $\boldsymbol{\omega}$, we further divide it into $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)'$ and σ_v^2.
  - Given $\boldsymbol{H}$ and σ_v^2, the volatility equ is

    $$\ln(h_t) = \alpha_0 + \alpha_1 \ln(h_{t-1}) + v_t, \quad t = 2, \cdots, n.$$

    Again, this is a simple linear regression and $\boldsymbol{\alpha}$ can be estimated easily. It is also normal if a normal prior is used.
  - Given $\boldsymbol{H}$ and $\boldsymbol{\alpha}$, we can compute

    $$v_t = \ln(h_t) - \alpha_0 - \alpha_1 \ln(h_{t-1}), \quad t = 2, \cdots, n.$$

    This is a random sample from a normal distribution with mean zero and variance σ_v^2. Thus,

    $$\sum_{t=2}^{n} \frac{v_t^2}{\sigma_v^2} \sim \chi^2_{n-1}.$$

If one uses the prior dist $(m\lambda)/\sigma_v^2 \sim \chi^2_m$, then

$$\frac{m\lambda + \Sigma_{t=2}^n v_t^2}{\sigma_v^2} \sim \chi^2_{n+m-1}.$$

Consequently, random draws of σ_v^2 can be done easily.

- Finally, for $\boldsymbol{H}$, we draw h_t one by one because there is no close-from dist available.

  Consider h_t and assume that all other parameters are known. Then h_t depends on h_{t-1}, h_{t+1} and a_t, i.e.

$$f(h_t|\boldsymbol{R}, \boldsymbol{H}_{(-t)}, \boldsymbol{\omega}, \boldsymbol{\beta})$$

$$\begin{aligned} &\propto \; f(a_t|h_t, r_t, r_{t-1}, \boldsymbol{\beta}) f(h_t + h_{t-1}, \boldsymbol{\omega}) f(h_{t+1}|h_t, \boldsymbol{\omega}) \\ &\propto \; h_t^{-1.5} \exp[-(r_t - \beta_0 - \beta_1 r_{t-1})^2/(2h_t) - (\ln(h_t) - \mu_t)^2/(2\sigma^2)] \end{aligned}$$

  where $\mu_t = [\alpha_0(1-\alpha_0) + \alpha_1(\ln h_{t+1} + \ln h_{t-1})]/(1+\alpha_1^2)$ and $\sigma^2 = \sigma_v^2/(1+\alpha_1^2)$. Why?

  This result is not belonged to a known dist. **What next**?

Clever solutions have been developed, e.g.

- Metropolis-Hasting (MH) algorithm

- Griddy Gibbs

Metropolis-Hasting algorithm: Use an approximate dist to draw the new parameter and a clever decision rule to decide accept or reject the new parameter.

Giddy Gibbs: Specify a grid for possible values of h_t, say m points. Evaluate the posterior density function of h_t for all m points. These m values from an empirical density for h_t. Draw h_t from the empirical density.

MH applies to high dimensional parameters. Griddy Gibbs is basically for scalar parameter only.

For SV models, both methods have been used.

Example: Monthly log returns of S&P 500 index from 1962 to 1999. Figure 10.3.

The program used is "svm" on GSBMBA. Typically, intensive computation is needed.

Example: Monthly IBM stock returns from 1926 to 1997 (864 obs). Comparison SV model with GARCH(1,1) model.

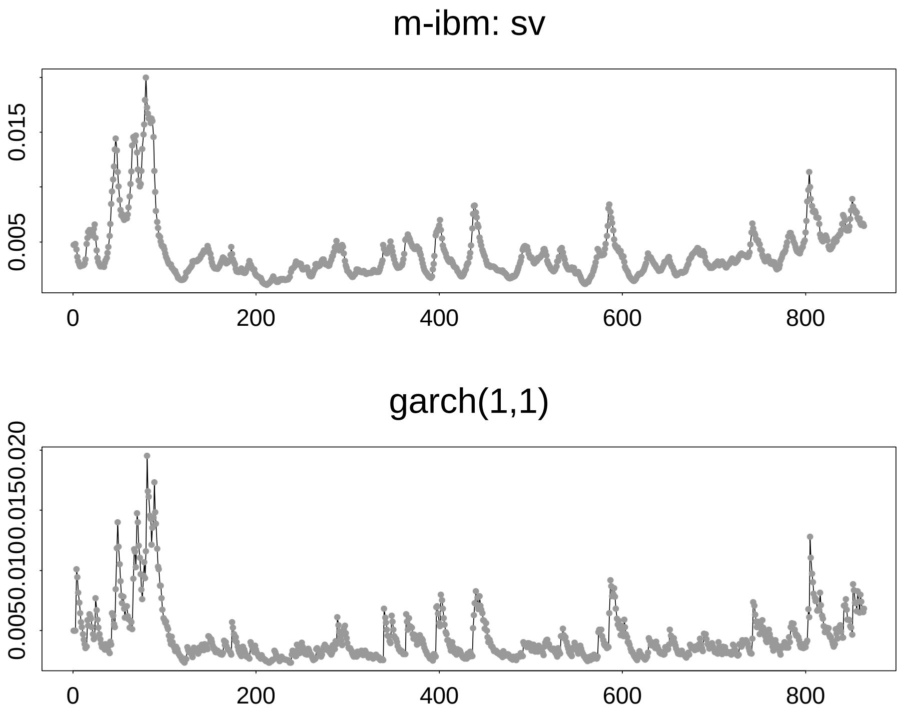

Figure 1: Estimated volatilities of IBM monthly returns

Illustration of estimating SV models

```
gsbmba% svm
 mu, k, iseed, ini, iter: 1 1 327 300 300
 order of the svm: 1
 File name for return series: m-ibm.dat
 Number of columns in the file: 1
 number of returns:   864
 Do regressors contain lagged returns only? (1=y): 1
 lags of the explanatory variables: 1
 prior mean for beta (const. at last): 0.1 .01
 Prior Covariance mtx of beta (row by row):
.4 0.
```

```
0. 1.
 prior mean for alpha: 0.01 .7
 prior covariance mtx for alpha (row by row)
4. 0.
0. .3
 prior for sigma_v^2: (m & lambda): 5. .2
 input initial volatility values? (1=y): 0
 Initial values for alpha & sigma_v^2: -1. .7 1.
 number of grid points: 400

 summary of alpha & sigm_v:
      Pos.Mean           Pos.Std.Error
      -0.59956            0.15169
       0.89554            0.02628
       0.08356            0.01387
 summary of beta (constant appears last):
       0.06314            0.03126
       0.01298            0.00199
 Pos.Mean of Volatility: fort.25
```