Modelling Exchange Rate Volatility

Yu Bai

Department of Finance, Bocconi University

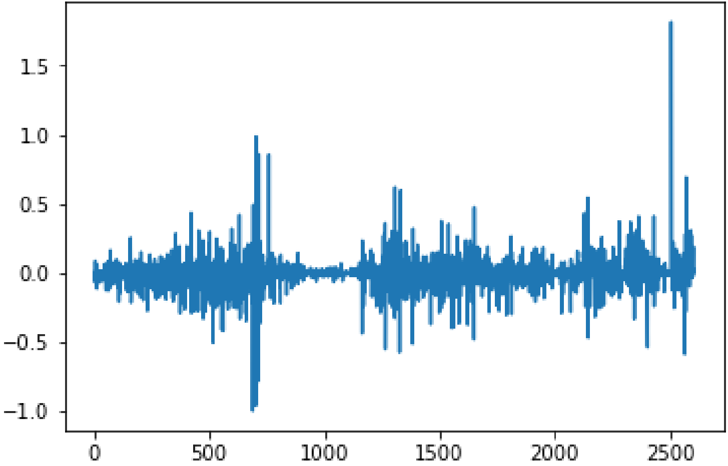
October 23, 2017

Outline

Motivation

- **Volatility clustering** is often observed in macroeconomic and financial time series.
- For example, during financial crises movements in financial asset returns tend be large (of either sign). In normal periods, the same asset returns might exhibit little time variation.
- There are mainly two types of models dealing with time varying variance.
  - GARCH type: The variance process is a deterministic function of past observations with associated parameters.
  - Stochastic Volatility: The variance process is just ”stochastic”, derived from theoretical option pricing formula.

Data



GARCH-type Models

- Bollerslev(1986) proposed the generalized *ARCH*, or *GARCH*(p,q) model

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 = \omega + \alpha(L)\varepsilon_{t-i}^2 + \beta(L)\sigma_{t-j}^2$$

- Rearranging the *GARCH*(p,q) model, we get

$$\varepsilon_t^2 = \omega + [\alpha(L) + \beta(L)]\varepsilon_{t-1}^2 - \beta(L)v_{t-1} + v_t$$

  where $v_t = \varepsilon_t^2 - \sigma_t^2$
- Stationary condition requires that all the roots of $\alpha(x) + \beta(x) = 1$ lie outside the unit circle.

- Glosten, Jagannathan and Runkle(1993) introduces a modified version of *GARCH* model to deal with asymmetric effects.

$$r_t = \mu + \varepsilon_t, \quad r_t|\mathscr{F}_{t-1} \sim N(\mu, \sigma_t^2)$$
$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{o} \gamma_j \varepsilon_{t-j}^2 \mathcal{I}_{\{\varepsilon_{t-j}<0\}} + \sum_{k=1}^{q} \beta_k \sigma_{t-k}^2$$

- Parameter restrictions for $GJR - GARCH(1,1,1)$

$$\omega > 0, \alpha \geqslant 0, \alpha + \gamma \geqslant 0, \beta \geqslant 0$$

- Stationary requires

$$\alpha + \frac{1}{2}\gamma + \beta < 1$$

Estimation

In practice, to implement *MLE*, we need to propose

- Starting value
- Optimization method
- The way to compute Hessian

In general, *GARCH* type model has the following likelihood function formula

$$\ln f(r_t|\mu, \sigma_t^2) = -\frac{1}{2}(\ln 2\pi + \ln \sigma_t^2 + \frac{(r_t - \mu)^2}{\sigma_t^2})$$

It can be shown that under certain regular conditions, quasi-maximum likelihood estimator (When the data follows a non-normal distribution) is consistent and asymptotically normal

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \mathcal{I}^{-1}\mathcal{J}\mathcal{I}^{-1})$$

where

$$\mathcal{I} = -E[\frac{\partial^2 \ln(\theta, r_t)}{\partial\theta\partial\theta'}]$$
$$\mathcal{J} = E[\frac{\partial \ln(\theta, r_t)}{\partial\theta}\frac{\partial \ln(\theta, r_t)}{\partial\theta'}]$$

- Since in many finance applications, data are generally non-normal. It remains to know how to estimate the variance of the QMLE.
- Note that, if $\mathcal{I} = \mathcal{J}$, the variance of *QMLE* simplifies to the *MLE* variance.
- However, Bollerslev and Wooldridge(1992) have shown that $\mathcal{I} = \mathcal{J}$ does not generally hold. They propose a sandwich formula for estimating the variance

$$\hat{\mathcal{I}}^{-1}\hat{\mathcal{J}}\hat{\mathcal{I}}^{-1}$$

I will use the following formula to approximate the Hessian and the score

$$\mathcal{I}_{ij}^{-1} \approx \frac{f(\theta + e_i h_i + e_j h_j) - f(\theta + e_i h_i) - f(\theta + e_j h_j) + f(\theta)}{h_i h_j}$$

$$\frac{\partial \ln f(\theta)}{\partial \theta} \approx \frac{f(\theta + e_i h_i) - f(\theta)}{h_i}$$

Estimation Results-GARCH

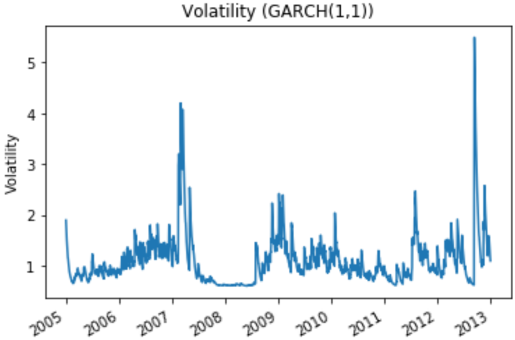


Table: GARCH model estimation

parameter	estimation	std. error	t statistics
μ	-0.008364	0.00939	-0.89081
ω	0.000143	0.000102	1.40525
α	0.03	0.004563	6.57486
β	0.9	0.030462	29.54542

Estimation Results-GJR-GARCH

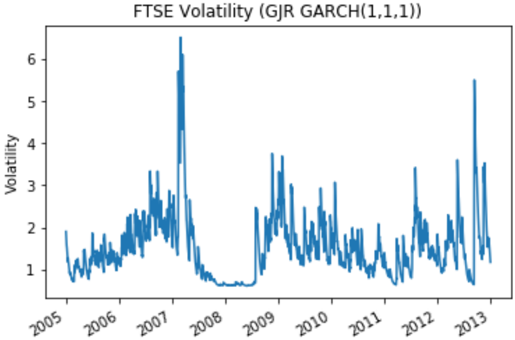


Table: GJR-GARCH model estimation

parameter	estimation	std. error	t statistics
μ	-0.008364	0.002826	-2.96001
ω	0.000143	0.000145	0.98981
α	0.03	0.013401	2.23867
γ	0.09	0.018150	4.95860
β	0.9	0.053656	16.77354

Basic Stochastic Volatility (SV) Model

State space representation of basic SV model:

$$y_t = e^{\frac{1}{2}h_t}\epsilon_t, \quad \epsilon \sim N(0,1)$$
$$h_t = h_{t-1} + u_t, \quad u_t \sim N(0, \sigma_h^2)$$

where $Var(y_t|h_t) = e^{h_t}$, h_0 is treated as an unknown parameter.

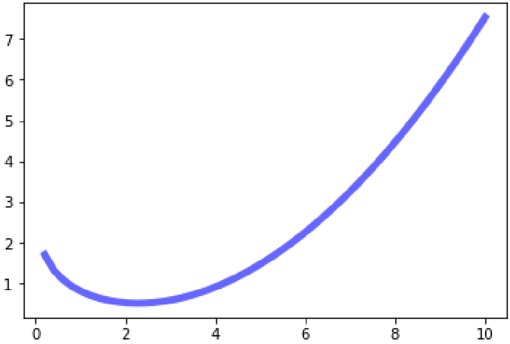
- h_t is often called the log-volatility.
- This basic SV model can be estimated by QMLE, using Kalman Filter algorithm. However, there are some problems.
    - The joint density of $h = (h_1, ..., h_T)'$ given the model parameters and the data is high-dimensional.
    - Normal assumption may not be a good approximation.

Let $y_t^* = \log y_t^2$, $\epsilon_t^* = \log \epsilon_t^2$, the model can be transformed in a linear state space form:

$$\begin{aligned} y_t^* &= h_t + \epsilon_t^* \\ h_t &= h_{t-1} + u_t \end{aligned}$$

Note that, ϵ_t^* follows a log-χ_1^2 distribution.

- Question: Is it good to approximate log-χ_1^2 distribution by normal distribution?


7
6
5
4
3
2
1
0
2
4
6
8
10

A Step Back

Reference: Jacquier, E., Polson, N.G., Rossi, P., 1994. Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 371-418.

- First paper which introduces Bayesian approach to estimate SV model, citation: 1790.
- Gibbs sampler:
  1. $p(\sigma_h|h, y)$
  2. $p(h|\sigma_h, y)$
- Step 1 is easy, but step 2 is not. One can show that:

$$p(h_t|h_{t+1}, h_{t-1}, \sigma_h, y) \propto \frac{1}{\sqrt{h_t}} \exp \frac{-y_t^2}{2h_t} \times \frac{1}{h_t} \exp \frac{-\log^2 h_t}{2\sigma_h^2}$$

  which is a product of normal and log-normal.

- We can impose an independence Metropolis-Hastings algorithm to draw from this density.
- How to choose the proposal? If we choose $q(\phi^*) = \frac{1}{h_t} \exp \frac{-\log^2 h_t}{2\sigma_h^2}$, then the acceptance probability simplifies to the likelihood ratio

$$\alpha = \min\left(\frac{(h_t^*)^{-0.5} \exp(\frac{-y_t^2}{2h_t^*})}{(h_t^m)^{-0.5} \exp(\frac{-y_t^2}{2h_t^m})}, \quad 1 \right)$$

- Single-move, exactly distribution, low convergence (maybe).
- Is there a way to jointly draw for $\boldsymbol{h}$?

Auxiliary Mixture Sampler

- Kim, Shepherd and Chib (1998) suggest an alternative way to deal with the log-χ^2_1 distribution to let us draw $\boldsymbol{h}$ jointly.
- The idea is to approximate the log-χ^2_1 distribution with a 7-components Gaussian mixture:

$$f(\epsilon_t^*) = \Sigma_{i=1}^{7} p_i \varphi(\epsilon_t^*; \mu_i - 1.2704, \sigma_i^2)$$

- The mixing parameters are fixed and also, the means are adjusted.
- The mixing parameters are obtained by matching the moments of the log-χ^2_1 distribution.

Table: A seven-component Gaussian mixture for approximating the log-χ_1^2 distribution.

component	p_i	μ_i	σ_i^2
1	0.00730	-10.12999	5.79596
2	0.10556	-3.97281	2.61369
3	0.00002	-8.56686	5.17950
4	0.04395	2.77786	0.16735
5	0.34001	0.61942	0.64009
6	0.24566	1.79518	0.34023
7	0.25750	-1.08819	1.26261

Gibbs Sampler

- Note that, conditioning on the component, ϵ_t^* becomes normal:

$$(\epsilon_t^* | s_t = i) \sim N(\mu_i, \sigma_i^2), \quad p(s_t = i) = p_i,$$

- Assume independent prior and complete the model specification:

$$\sigma_h^2 \sim IG(v_h, S_h), \quad h_0 \sim N(a_0, b_0).$$

- The posterior draw from $p(\boldsymbol{h}, \boldsymbol{s}, \sigma_h^2, h_0 | \boldsymbol{y})$ can be obtained via Gibbs sampler in 4 steps
  1. $p(\boldsymbol{s} | \boldsymbol{y}, \boldsymbol{h}, \sigma_h^2, h_0)$
  2. $p(\boldsymbol{h} | \boldsymbol{y}, \boldsymbol{s}, \sigma_h^2, h_0)$
  3. $p(\sigma_h^2 | \boldsymbol{y}, \boldsymbol{h}, \boldsymbol{s}, h_0)$
  4. $p(h_0 | \boldsymbol{y}, \boldsymbol{h}, \sigma_h^2, \boldsymbol{s})$

Step 1

- Note that, conditioning on $\boldsymbol{y}, \boldsymbol{h}$, $\boldsymbol{s}$ is known.

$$p(\boldsymbol{s}|\boldsymbol{y}, \boldsymbol{h}, \sigma_h^2, h_0) = \amalg_{t=1}^{T} p(s_t|y_t, h_t, \sigma_h^2, h_0)$$

- By Bayes' theorem,

$$p(s_t = i|y_t, h_t, \sigma_h^2, h_0) = \frac{p_i \varphi(y_t^*; h_t + \mu_i - 1.2704, \sigma_i^2)}{\Sigma_{j=1}^{7} p_j \varphi(y_t^*; h_t + \mu_j - 1.2704, \sigma_j^2)}$$

- This is a discrete distribution. We can easily sample it by a inverse-transformation method.

Inverse Transformation Method

- $X \sim F$, non-decreasing
- $F^{-1}(y) = \inf\{x : F(x) \geqslant y\}, \quad 0 \leqslant y \leqslant 1$
- Let $U \sim \mathcal{U}(0,1)$,

$$P(F^{-1}(u) \leqslant x) = P(u \leqslant F(x)) = F(x)$$

- Then, we can draw from F by
  1. $U \sim \mathcal{U}(0,1)$
  2. $X = F^{-1}(u)$

Step 2

- First, note that
$$(\boldsymbol{\epsilon}^*|\boldsymbol{s}) \sim N(\boldsymbol{d}_s, \boldsymbol{\Sigma}_s)$$
where $\boldsymbol{d}_s = (\mu_{s_1} - 1.2704, ..., \mu_{s_7} - 1.2704)'$, $\boldsymbol{\Sigma}_s = diag(\sigma^2_{s_1}, ..., \sigma^2_{s_7})$.
- By simple transformation of variables, we get the likelihood
$$(\boldsymbol{y}^*|\boldsymbol{s}, \boldsymbol{h}) \sim N(\boldsymbol{h} + \boldsymbol{d}_s, \boldsymbol{\Sigma}_s)$$
- Let $\boldsymbol{H}$ be the usual first-differencing matrix, rewrite the state equation as
$$\boldsymbol{Hh} = \tilde{\boldsymbol{\alpha}}_h + \boldsymbol{u},$$
where $\tilde{\boldsymbol{\alpha}}_h = (h_0, 0, ..., 0)'$, $\boldsymbol{u} \sim N(0, \sigma^2_h I_T)$.

- Then, it is trivial to show that,

$$(\boldsymbol{h}|\sigma_h^2, h_0) \sim N(h_0 I_T, \sigma_h^2 (\boldsymbol{H}'\boldsymbol{H})^{-1}).$$

- Now we have obtained the prior and likelihood, posterior can be done easily by applying directly the regression results:

$$(\boldsymbol{h}|\boldsymbol{y}, \boldsymbol{s}, \sigma_h^2, h_0) \sim N(\hat{h}, K_h^{-1})$$

where

$$K_h = \frac{1}{\sigma_h^2}\boldsymbol{H}'\boldsymbol{H} + \boldsymbol{\Sigma}_s^{-1}, \quad \hat{h} = K_h^{-1}\left(\frac{1}{\sigma_h^2}\boldsymbol{H}'\boldsymbol{H} h_0 I_T + \boldsymbol{\Sigma}_s^{-1}(\boldsymbol{y}^* - \boldsymbol{d}_s)\right)$$

- This step can be done by precision sampler, instead of Kalman-Filter algorithm by KSC(1998).

Precision Sampler

We can generate R independent draw from $N(\mu, K^{-1})$ of dimension n by the following 4 steps

1. Compute the lower Cholesky factorization $K = BB'$.
2. Generate $Z = (Z_1, ..., Z_n)'$ by drawing $Z_1, ..., Z_n \sim N(0, 1)$.
3. Return $U = \mu + B'^{-1}Z$.
4. Repeat step 2 and step 3independently R times.

Step 3 and 4

- These steps are easy, since both can be done by standard regression results.

$$(\sigma_h^2|\boldsymbol{y}, \boldsymbol{h}, \boldsymbol{s}, h_0) \sim IG\Big(v_h + \frac{T}{2}, S_h + \frac{1}{2}\Sigma_{t=1}^{T}(h_t - h_{t-1})^2\Big)$$

$$(h_0|\boldsymbol{y}, \boldsymbol{h}, \sigma_h^2) \sim N(\hat{h}_0, K_{h_0}^{-1}),$$

where

$$K_{h_0} = \frac{1}{b_0} + \frac{1}{\sigma_h^2}, \quad \hat{h}_0 = K_{h_0}^{-1}\Big(\frac{a_0}{b_0} + \frac{h_1}{\sigma_h^2}\Big)$$

Estimation Results