Bayesian Estimation of GARCH(1,1) Models with Normal and Student-t Innovations

Valentin Meylan

Bayesian Computation

June 19^{th}, 2019

Sections

Data features

Daily Adj. close from June 2009 to April 2019

FIGURE – Intel Corp. stock price from June 2009 to April 2019

Data features

Daily returns from June 2009 to April 2019

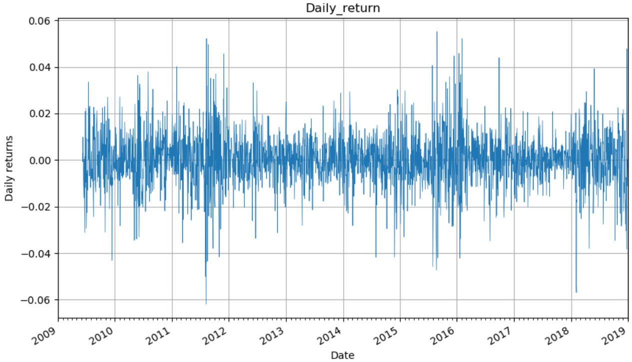


FIGURE – Intel Corp. daily log-return from June 2009 to April 2019

Model specifications
GARCH(1,1) with normal innovations

The process $\{y_t,\, t \in \mathcal{Z}\}$ is said to be a **GARCH(1,1)** process with normal innovations if, for $t = 1, \ldots, T$,

$$y_t = \epsilon_t \cdot h_t^{1/2}$$

where ϵ_t is a sequence of $i.i.d.$ variables $\sim \mathcal{N}(0,1)$, and

$$h_t = \omega + \alpha \cdot y_{t-1}^2 + \beta \cdot h_{t-1}$$

where $\omega > 0\,;\, \alpha, \beta \geq 0\,;\, \mathcal{N}(0,1)$ denotes the standard normal distribution. The restriction on the GARCH parameters ω, α, β guarantee the conditional variance's positivity.

Model specifications
GARCH(1,1) with normal innovations

We choose normal priors over ω, α, β :

$$p(\omega) \propto \phi_{\mathcal{N}}(\omega|\mu_\omega, \Sigma_\omega) \cdot \mathbf{1}_{\omega>0}$$

$$p(\alpha) \propto \phi_{\mathcal{N}}(\alpha|\mu_\alpha, \Sigma_\alpha) \cdot \mathbf{1}_{\alpha\geq 0}$$

$$p(\beta) \propto \phi_{\mathcal{N}}(\beta|\mu_\beta, \Sigma_\beta) \cdot \mathbf{1}_{\beta\geq 0}$$

We will enforce the constraints on the parameters in the MH sampling algorithm.
We assume prior independence between the parameters, i.e. :

$$p(\theta) = p(\omega)p(\alpha)p(\beta)$$

Model specifications
GARCH(1,1) with normal innovations

Likelihood function for our model :

$$L(\theta|y) \propto \prod_{t=1}^{T}(h_t)^{-1/2}exp(-\frac{1}{2}\frac{y_t^2}{h_t})$$

$$log(L(\theta|y)) \propto -\frac{1}{2}\left(\sum_{t=1}^{T}log(h_t) + \frac{y_t^2}{h_t}\right)$$

We define the log posterior as :

$$log(p(\theta|y)) \propto log(L(\theta|y)) \cdot log(p(\theta))$$

Metropolis-Hastings

Random-walk Metropolis-Hastings algorithm

Idea : leverage MCMC sampling to generate $\theta_0, \ldots, \theta_i, \ldots, \theta_n$ samples of θ, our vector of variables.

We would like to re-organize the chain into delivering sampling from the distribution of interest $p(\theta|d)$ as the number of draws n goes to ∞.

Such fixing requires the implementation of an accept/reject step as detailed in the following slide.

Metropolis-Hastings

Random-walk Metropolis-Hastings algorithm

Given :

- A target $\widetilde{f}(\theta)$ (possibly unnormalized)
- A symmetric proposal density : $h(\eta)$

Then :

- Generate a sequence of samples (θ_n)
    - Sample a proposal from h :

$$\theta_{proposal} = \theta_n + \eta_n$$

$$U_n \sim U([0,1])$$

    - Compute the MH acceptance ratio :

$$R = min\left[\frac{f(\theta_{proposal}|d)}{f(\theta_n|d)}, 1\right]$$

    - If $U_n \leq R$, accept : $\theta_{n+1} = \theta_{proposal}$
      Else, reject : $\theta_{n+1} = \theta_n$

Metropolis-Hastings

MH applied to GARCH(1,1) with normal innovations

	# batches	sims / batch	time_per_MH	array_stepsize	Skip_rate	U>R_rate	A_rate
0	1	2000	0min 5s	[1.0e-04 1.12591e-01 8.67884e-02]	0.4925	0.5040	0.0035
1	1	2000	0min 11s	[1.0e-09 4.8e-03 2.0e-05]	0.0000	0.5435	0.4565
2	1	10000	0min 55s	[1.0e-09 4.8e-03 2.0e-05]	0.0000	0.5219	0.4781

	# batches	sims / batch	avg_omega / MH_sampling	avg_alpha / MH_sampling	avg_beta / MH_sampling
0	1	2000	0.000044	0.099667	0.728261
1	1	2000	0.000005	0.052443	0.930147
2	1	10000	0.000005	0.053129	0.929335

TABLE – MH sampling results

Metropolis-Hastings

MH applied to GARCH(1,1) with normal innovations

	omega	alpha	beta
Asset	0.000005	0.05	0.93

TABLE – MLE values

	# batches	sims / batch	avg_omega / MH_sampling	avg_alpha / MH_sampling	avg_beta / MH_sampling
0	1	2000	0.000044	0.099667	0.728261
1	1	2000	0.000005	0.052443	0.930147
2	1	10000	0.000005	0.053129	0.929335

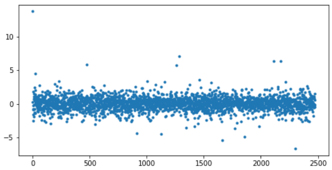
TABLE – MH sampling results

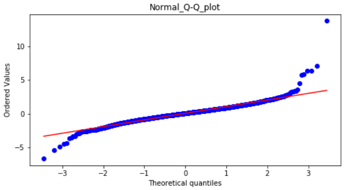
Metropolis-Hastings

MH applied to GARCH(1,1) with normal innovations

We compute $\widehat{h}_t$ from the average posterior sample and define the standardized residuals as :

$$\widehat{Z}_t = \frac{y_t}{\widehat{h}_t^{1/2}}$$





GARCH(1,1) with student-t innovations

Model specification

We will consider the GARCH(1,1) with Student-t innovations model of the log-returns process $\{y_t,\, t \in \mathcal{Z}\}$, for $t = 1, \ldots, T$:

$$y_t = \epsilon_t \cdot h_t^{1/2}$$

where ϵ_t is a sequence of $i.i.d.$ variables $\sim \mathbf{t}(\nu)$,

$$h_t = \omega + \alpha \cdot y_{t-1}^2 + \beta \cdot h_{t-1}$$

for $\omega > 0\,;\, \alpha, \beta \geq 0\,;\, \mathbf{t}(\nu)$ denotes the Students-t distribution with ν degrees of freedom.

GARCH(1,1) with student-t innovations

Model specification

The term $(\alpha + \beta)$ is the degree of persistence in the autocorrelation of the squares which controls the intensity of the clustering in the variance process. With a value close to one, past shocks and past variances will have a longer impact on the future conditional variance.

	omega	alpha	beta	nu
values	0.0	0.985204	0.014677	49.515688

TABLE – MLE values from a GARCH(1,1) with student-t innovations fit, on Intel Corp. stock

GARCH(1,1) with student-t innovations

Model specification

We chose normal priors over ω, α, β :

$$p(\omega) \propto \phi_{\mathcal{N}}(\omega|\mu_\omega, \Sigma_\omega) \cdot \mathbf{1}_{\omega>0}$$

$$p(\alpha) \propto \phi_{\mathcal{N}}(\alpha|\mu_\alpha, \Sigma_\alpha) \cdot \mathbf{1}_{\alpha\geq 0}$$

$$p(\beta) \propto \phi_{\mathcal{N}}(\beta|\mu_\beta, \Sigma_\beta) \cdot \mathbf{1}_{\beta\geq 0}$$

We assume prior independence between the parameters, i.e. :

$$p(\theta) = p(\omega)p(\alpha)p(\beta)$$

GARCH(1,1) with student-t innovations

Model specification

The likelihood function applied to our GARCH(1,1) with student innovations yields :

$$L(\theta|y) \propto \prod_{t=1}^{T} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)\nu^{1/2}\pi^{1/2}h_t^{1/2}} \cdot \left[1 + \frac{1}{\nu}\frac{y_t^2}{h_t}\right]^{-\frac{\nu+1}{2}}$$

$$log(L(\theta|y)) \propto T\left[log(\Gamma(\frac{\nu+1}{2}) - log(\Gamma(\frac{\nu}{2})) - \frac{1}{2}log(\nu) - \frac{1}{2}log(\pi)\right]$$

$$-\sum_{t=1}^{T}\frac{1}{2}log(h_t) + (\frac{\nu+1}{2})\left[1 + \frac{1}{\nu}\frac{y_t^2}{h_t}\right]$$

Metropolis-Hastings

MH applied to GARCH(1,1) with student-t innovations

	omega	alpha	beta	nu
Stock	0.0	0.985204	0.014677	49.515688

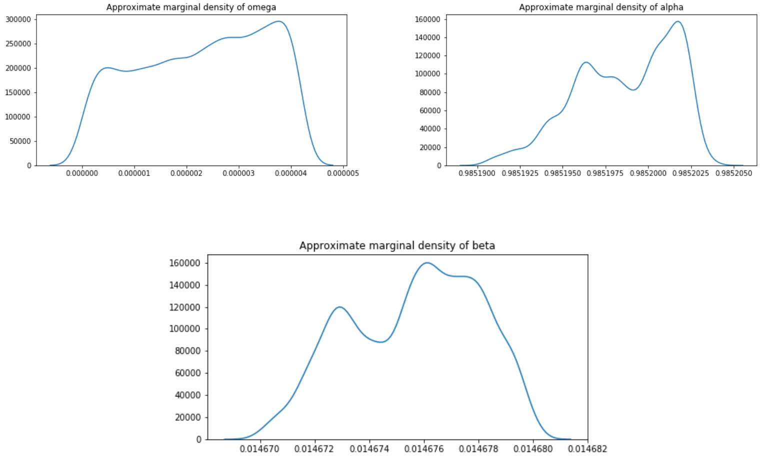
TABLE – MLE values

Time to compute 10,000 simulations : $\sim 55s$
Acceptance rate : $\sim 66\%$
Results obtained with student-5 proposal density $h(\eta)$

	# batches	sims/batch	avg_omega/MH_sampling	avg_alpha/MH_sampling	avg_beta/MH_sampling
23	1	10000	0.000002	0.985201	0.014781
24	1	10000	0.000002	0.985204	0.014688
25	1	10000	0.000002	0.985204	0.014688
26	1	10000	0.000002	0.985207	0.014733
27	1	10000	0.000002	0.985199	0.014676

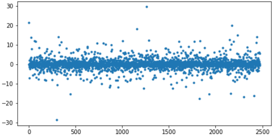
Metropolis-Hastings

MH applied to GARCH(1,1) with student-t innovations

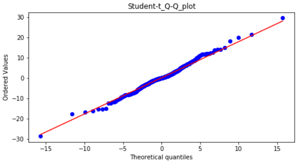


Metropolis-Hastings

MH applied to GARCH(1,1) with student-t innovations



(a) Scatterplot of normalized residuals



(b) Q-Q plot for $\nu = 3$

Augmented GARCH(1,1) with Student-t innovations

Model specification

We will consider the augmented model from (D. Ardia, L. Hoogerheide, 2009) of the log-returns process $\{y_t,\, t \in \mathcal{Z}\}$ of a GARCH(1,1) with Student-t innovations, for $t = 1, \ldots, T$:

$$y_t = \epsilon_t (\frac{\nu - 2}{\nu} \eta_t h_t)^{1/2}$$

The parameter $\frac{\nu - 2}{\nu}$ is a scaling factor which ensures the conditional variance of y_t to be h_t.

Augmented GARCH(1,1) with Student-t innovations

Model specification

$$y_t = \epsilon_t \left(\frac{\nu - 2}{\nu} \eta_t h_t\right)^{1/2}$$

where ϵ_t is a sequence of $i.i.d.$ variables $\sim \mathcal{N}(0, 1)$,
where η_t is a sequence of $i.i.d.$ variables $\sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu}{2})$,

$$h_t = \omega + \alpha \cdot y_{t-1}^2 + \beta \cdot h_{t-1}$$

for $\omega > 0\,;\ \alpha, \beta \geq 0\,;\ \mathcal{N}(0, 1)$ denotes the standard normal distribution, $\mathcal{IG}$ the inverted gamma distribution.

Augmented GARCH(1,1) with Student-t innovations

Model specification

Define : $\psi \doteq (\omega, \alpha, \beta, \nu)$. We can define the T x T diagonal matrix :

$$\Sigma \doteq \Sigma(\psi, \eta) = diag\left(\eta_t \frac{\nu - 2}{\nu} h_t(\omega, \alpha, \beta)\right)$$

for $t = 1, \ldots, T$
We can express the likelihood of (ψ, η) as :

$$\mathcal{L}(\psi, \eta | y) \propto (det\Sigma)^{-1/2} exp[-\frac{1}{2} y' \Sigma^{-1} y]$$

Augmented GARCH(1,1) with Student-t innovations

Model specification

$$p(\omega) \propto \phi_{\mathcal{N}}(\omega|\mu_\omega, \Sigma_\omega) \cdot \mathbf{1}_{\omega>0}$$

$$p(\alpha) \propto \phi_{\mathcal{N}}(\alpha|\mu_\alpha, \Sigma_\alpha) \cdot \mathbf{1}_{\alpha\geq 0}$$

$$p(\beta) \propto \phi_{\mathcal{N}}(\beta|\mu_\beta, \Sigma_\beta) \cdot \mathbf{1}_{\beta\geq 0}$$

The components η_t are $i.i.d.$ distributed from the inverted gamma density, hence the prior distribution on η conditional on ν is :

$$log(p(\eta|\nu)) = \frac{T\nu}{2}log(\frac{\nu}{2}) - T \cdot log(\Gamma(\frac{\nu}{2})) - (\frac{\nu}{2}+1)\sum_{t=1}^{T} log(\eta_t) - \frac{1}{2}\sum_{t=1}^{T}\frac{\nu}{\eta_t}$$

Augmented GARCH(1,1) with Student-t innovations

Model specification

Following (Deschamps, 2006) in the choice of the prior distribution on the degrees of freedom parameter ν, our prior is a translated exponential with parameters $\lambda > 0$ and $\delta \geq 2$:

$$p(\nu) = \lambda exp[-\lambda(\nu - \delta)] \cdot \mathbf{1}_{\nu > \delta}$$

For large values of λ, the mass of the prior is concentrated in the neighborhood of δ and a constraint on the degrees of freedom can be imposed in this manner. Normality of the errors is assumed when δ is chosen large.

Rejection sampling

Idea

Idea : we would like to sample from the posterior $f(\theta|d)$. The algorithm consists in sampling from the hypograph of $\widetilde{f}(\theta|d)$ i.e. the set f points of $\mathbb{R}^{d+1}$ which is between the graph $\widetilde{f}(\theta|d)$ and that of the nul function : $\theta \mapsto 0$:

$$Hypograph(\widetilde{f}) = \{(\theta, h); 0 \leq h \leq f(\theta|d)\}$$

Such that if we can sample from the hypograph, then we can drop the h coordinate to get samples from the posterior.

Rejection sampling
Algorithm

- Generate a sample θ_i with distribution $h(\theta)$
- Sample $u_i \sim U([0,1])$
- If :
$$u_i \leq \frac{\widetilde{f}(\theta_i|d)}{Kh(\theta_i)}$$
then we accept the sample, and reject otherwise.
This corresponds to sampling the height and checking whether it falls in the epigraph of $\widetilde{f}(\theta|d)$.

Accepted samples are i.i.d. from the normalized target $f(\theta|d)$

Major weakness : $h(\theta)$ sets the value of K : $K \geq max \left[\frac{\widetilde{f}(\theta|d)}{h(\theta)} \right]$

Metropolis-Hastings Rejection Sampling

MH & RS applied to augmented GARCH(1,1) with student-t innovations

Asset	omega	alpha	beta	nu
Stock	0.0	0.985204	0.014677	49.515688

TABLE – MLE values

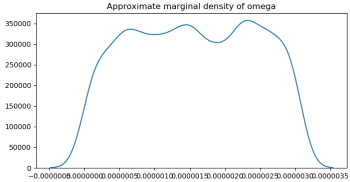
Time to compute 10,000 simulations : $\sim 2min$ (vs. $55s$ previously)
Acceptance rate : $\sim 30\%$

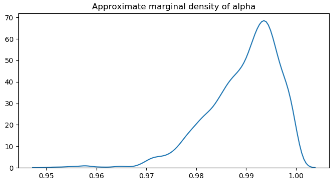
	# batches	sims / batch	avg_omega / MH_sampling	avg_alpha / MH_sampling	avg_beta / MH_sampling
85	1.0	10000.0	0.000002	0.972059	0.014832
86	1.0	10000.0	0.000002	0.980215	0.014839
87	1.0	10000.0	0.000002	0.989472	0.014841

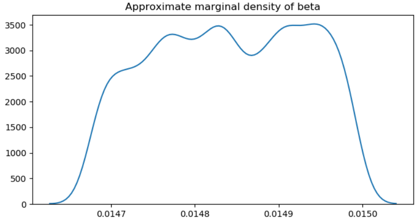
TABLE – MH sampling

Metropolis-Hastings

MH & RS applied to augmented GARCH(1,1) with student-t innovations

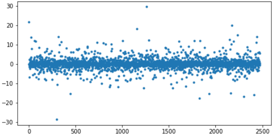




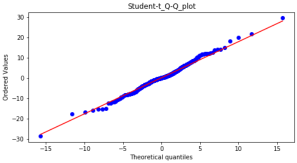


Metropolis-Hastings

MH & RS applied to augmented GARCH(1,1) with student-t innovations



(a) Scatterplot of normalized residuals



(b) Q-Q plot for $\nu = 3$

Importance Sampling
Idea

Idea : Having an accept-reject step, itself function of a unique draw might not be flexible.

Furthermore, the outcome of the accept-reject step is subject to the choice of the proposal density $h(.)$.

In fact, this proposal directly sets the minimum K value which scales the factor. We aim to compute weights, function of the sampled values. The next slide will detail the algorithm.

Importance Sampling

Algorithm

Two ingredients :

- A target $\widetilde{f}(\theta|d)$ (can be unnormalized)
- A proposal $h(\theta)$

The algorithm :

- Generate i.i.d. samples from $h : \theta_1, \ldots, \theta_l$
- Compute their weights :

$$w_i = \frac{\widetilde{f}(\theta_i|d)}{h(\theta_i)}$$

- Approximation of $\mathbb{E}_{\theta \sim f(\theta|d)}[S(\theta)]$:

$$\mathbb{E}_{\theta \sim f(\theta|d)}[S(\theta)] \approx \frac{\sum w_i S(\theta_i)}{\sum w_i}$$

Metropolis-Hastings Importance Sampling

MH & IS applied to augmented GARCH(1,1) with student-t innovations

Asset	omega	alpha	beta	nu
values	0.0	0.985204	0.014677	49.515688

TABLE – MLE values

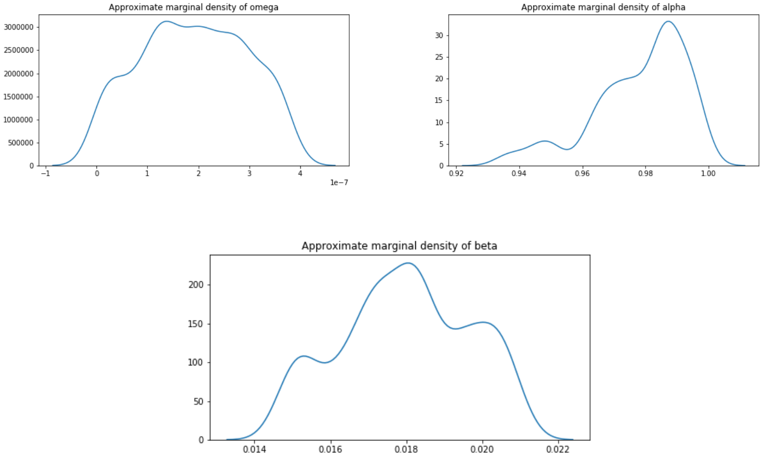
Time to compute 10,000 simulations : $\sim 10min$ (vs. $2min$ for IS and $55s$ for MH-only sampling)
Acceptance rate : $\sim 30\%$

	# batches	sims / batch	avg_omega / MH_sampling	avg_alpha / MH_sampling	avg_beta / MH_sampling
12	1	1000	0.000000	0.971696	0.017663
13	1	1000	0.000000	0.935570	0.017709
14	1	1000	0.000000	0.984308	0.017837

TABLE – MH sampling

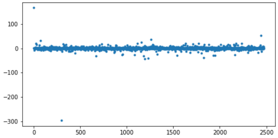
Metropolis-Hastings Importance Sampling

MH & IS applied to augmented GARCH(1,1) with student-t innovations

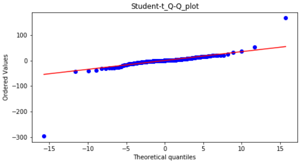


Metropolis-Hastings Importance Sampling

MH & IS applied to augmented GARCH(1,1) with student-t innovations



(a) Scatterplot of normalized residuals



(b) Q-Q plot for $\nu = 3$

Bibliography

1. D. Ardia and L. Hoogerheide F., "Bayesian Estimation of the GARCH(1,1) Model with Student-t Innovations", The R Journal, Available : https ://journal.r-project.org/archive/2010/RJ-2010-014/index.html