

6. Volatility Modelling in Financial Time Series

It has long been recognized that financial time series exhibit changes in volatility over time that are serially correlated.

In particular, financial returns demonstrate *volatility clustering*; large changes tend to be followed by large changes and vice versa. Such models fall into two classes:

- ▶ *Observation-driven* models allow the variance of the observed series to depend on its lagged values
- ▶ *Parameter-driven* models specify that the variance of the observations is a function of some unobserved or latent process.

Both types of model have the general form of state-space models.

6.1 Types of Volatility Models

In a *parameter-driven* model


$$p(y_t | x_t, \boxed{y}_{1:(t-1)}, y_{1:(t-1)}) \equiv p(y_t | x_t)$$

$$p(x_{t+1} | x_t, \boxed{y}_{1:(t-1)}, y_{1:t}) \equiv p(x_{t+1} | x_t)$$

so that

$$p(x_{2:n}, y_{1:n} | x_1) = \left\{ \prod_{t=1}^n p(y_t | x_t) \right\} \left\{ \prod_{t=2}^n p(x_t | x_{t-1}) \right\}$$

6.1 Types of Volatility Models

x not y

In an *observation-driven* model

$$p(y_t|x_t, \boxed{y}_{1:(t-1)}, y_{1:(t-1)}) \equiv p(y_t|x_t)$$

$$p(x_{t+1}|x_t, \boxed{y}_{1:(t-1)}, y_{1:t}) \equiv p(x_{t+1}|y_{1:t})$$

so that

$$p(x_{1:n}, y_{1:n}|x_1) = \prod_{t=1}^n p(y_t|x_t)p(x_t|y_{1:(t-1)})$$

where $p(x_1|y_0) \equiv p(x_1)$, say.

6.1 Types of Volatility Models

Example (Model for Count Data)

Suppose that

$$p(y_t|x_t) \equiv \text{Poisson}(\pi x_t)$$

$$p(x_{t+1}|x_t) = 1 \text{ if } x_{t+1} = ax_t, 0 \text{ otherwise}$$

$$p(x_1) \equiv \text{Gamma}(\alpha, \lambda)$$

for $t = 1, \dots, n$ for some constants π, a, α, λ .

6.1 Types of Volatility Models

Example (Model for Count Data)

We need to compute the filtering and prediction distributions:
Suppose that

$$p(x_t|y_{1:t}) \equiv \text{Gamma}(\alpha_t, \lambda_t)$$

$$p(x_{t+1}|y_{1:t}) \equiv \text{Gamma}(\alpha_t, \lambda_t/a)$$

$$p(y_{t+1}|y_{1:t}) \equiv \text{NegBinomial}(\alpha_t, 1 - \pi/\lambda_t)$$

where

$$\alpha_t = \alpha + \sum_{j=1}^t y_j \qquad \lambda_t = \lambda a^{1-t} + \pi \left(\frac{1 - a^{-t}}{1 - a^{-1}} \right)$$

6.1 Types of Volatility Models

Example (Model for Count Data)

Then

$$\mathbb{E}[Y_{t+1}|Y_{1:t} = y_{1:t}] = \frac{\pi\alpha_t}{\lambda_{t+1} - \pi}$$

and

$$\mathbb{V}ar[Y_{t+1}|Y_{1:t} = y_{1:t}] = \frac{\pi\alpha_t\lambda_t}{(\lambda_{t+1} - \pi)^2}$$

6.1 Types of Volatility Models

Example (AR(1) as an observation-driven model)

Suppose that

$$p(x_{t+1}|y_{1:t}) = 1 \text{ if } x_{t+1} = y_t, 0 \text{ otherwise}$$

$$p(y_t|x_t) \equiv f_Z(y_t - \phi x_t)$$

for some pdf f_Z . Then

$$Y_t - \phi X_t \stackrel{d}{=} Z_t \quad \text{or} \quad Y_t = \phi X_t + Z_t$$

so with $X_t = Y_{t-1}$ we have an AR(1) model.

6.1 Types of Volatility Models

Example (Exponential family observation-driven model)

Suppose that

$$p(y_t|x_t) = \exp\{y_t x_t - b(x_t) + c(y_t)\}$$

that is, a natural exponential family model with canonical parameter x_t . We have that

$$\exp\{b(x_t)\} = \int \exp\{y_t x_t - b(x_t) + c(y_t)\} dy_t$$

and

$$\dot{b}(x_t) = \mathbb{E}[Y_t|X_t = x_t] \qquad \ddot{b}(x_t) = \mathbb{V}ar[Y_t|X_t = x_t]$$

6.1 Types of Volatility Models

Example (Exponential family observation-driven model)

Suppose that

$$p(x_t|y_{1:(t-1)}) = f(x_t|\alpha_{t|t-1}, \lambda_{t|t-1})$$

where

$$f(x|\alpha, \lambda) = \exp\{\alpha x - \lambda b(x) + A(\alpha, \lambda)\}$$

6.1 Types of Volatility Models

Example (Exponential family observation-driven model)

Then

$$\begin{aligned} p(x_t|y_{1:(t-1)}, y_t) &\propto p(y_t|x_t)f(x_t|y_{1:(t-1)}) \\ &\propto \exp\{y_t x_t - b(x_t) + c(y_t)\} \exp\{\alpha_{t|t-1} x_t - \lambda_{t|t-1} b(x_t)\} \\ &\propto \exp\{\alpha_{t|t} x_t - \lambda_{t|t} b(x_t)\} \end{aligned}$$

say. That is

$$f(x_t|y_{1:(t-1)}) \quad \text{and} \quad f(x_t|y_{1:t})$$

have the same functional form, with parameters updated

$$(\alpha_{t|t-1}, \lambda_{t|t-1}) \longrightarrow (\alpha_{t|t}, \lambda_{t|t})$$

6.1 Types of Volatility Models

A standard model for financial times series is based on the *Stochastic Differential Equation* (SDE)

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

where $\{W_t\}$ is a continuous time process with Gaussian increments dW_t .

The equivalent equation for $Y_t = \log S_t$ is

$$dY_t = \mu dt + \sigma dW_t$$

6.1 Types of Volatility Models

The usual interpretation of this equation is in terms of a constant time difference discrete time process where we replace dY_t by $Y_{t+1} - Y_t$, and dt by Δ , and imagine the process observed at times

$$\Delta, 2\Delta, 3\Delta, \dots$$

to yield data $y_1, y_2, y_3, \dots$. Note that

$$Y_{t+1} - Y_t = \log S_{t+1} - \log S_t = \log(S_{t+1}/S_t)$$

6.1 Types of Volatility Models

Thus, in the discrete approximation

$$Y_{t+1} - Y_t = \mu\Delta + \sigma Z_t$$

where $\{Z_t\} \sim \mathcal{N}(0, 1)$, so that

$$Y_{t+1}|Y_t = y_t \sim \mathcal{N}(y_t + \Delta\mu, \sigma^2)$$

However, this model does not reflect the behaviour of real financial time series.

6.1 Types of Volatility Models

A better model is based on the SDE

$$dY_t = \mu dt + \sigma(t)dW_t$$

Thus, in the discrete approximation

$$Y_{t+1} - Y_t = \mu\Delta + \sigma_t Z_t$$

where again $\{Z_t\} \sim \mathcal{N}(0, 1)$, so that

$$Y_{t+1}|Y_t = y_t \sim \mathcal{N}(y_t + \Delta\mu, \sigma_t^2)$$

This is referred to as a *conditional heteroscedastic* model with time-varying variance.

6.1 Types of Volatility Models

The most popular examples of observation-driven models in finance are

- ▶ the *Autoregressive Conditional Heteroscedasticity* (ARCH)
- ▶ *Generalized ARCH* (GARCH) models.

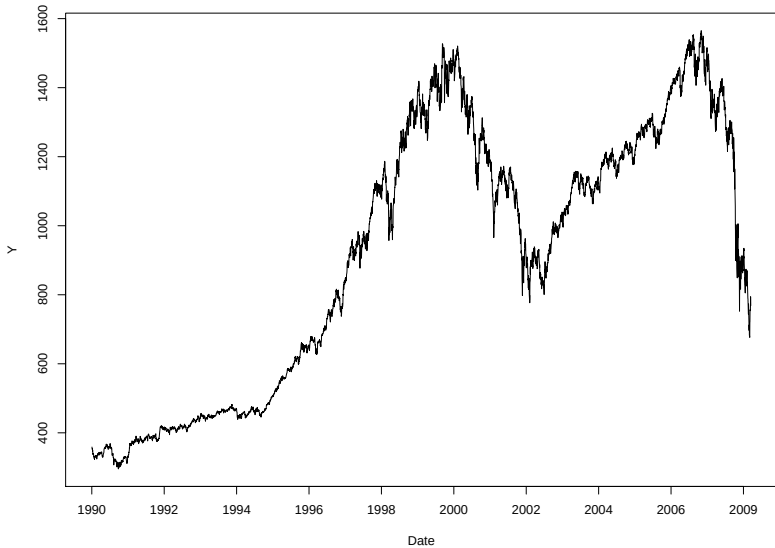
In particular, let y_t be a realization, at time t , of the time series of interest. Typically, y_t is taken to be the compounded return of the underlying asset, so that

$$y_t = 100 \log(s_t/s_{t-1})$$

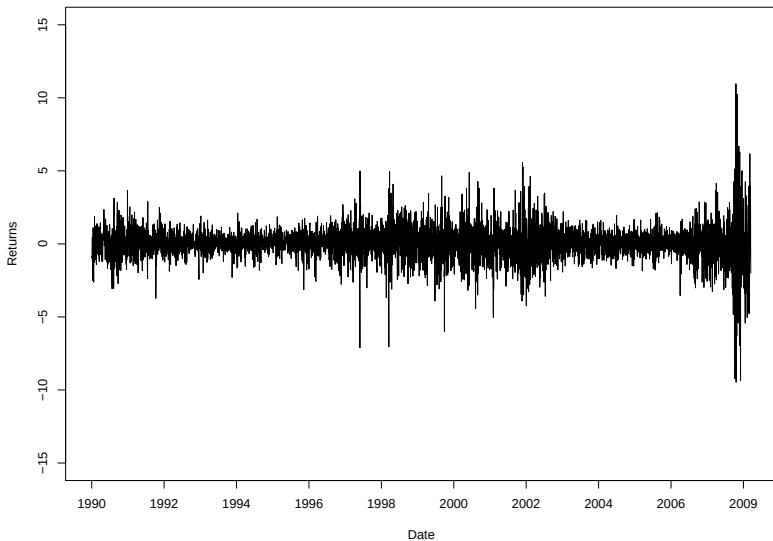
where s_t denotes the observed price of the asset.

ARCH type models specify the distribution of the current observation as a one-step-ahead prediction density.

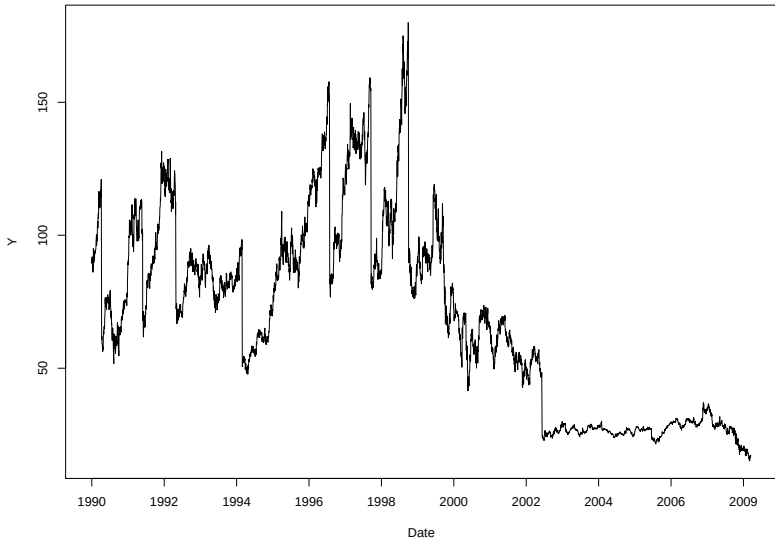
S & P 500 : 1990-2009



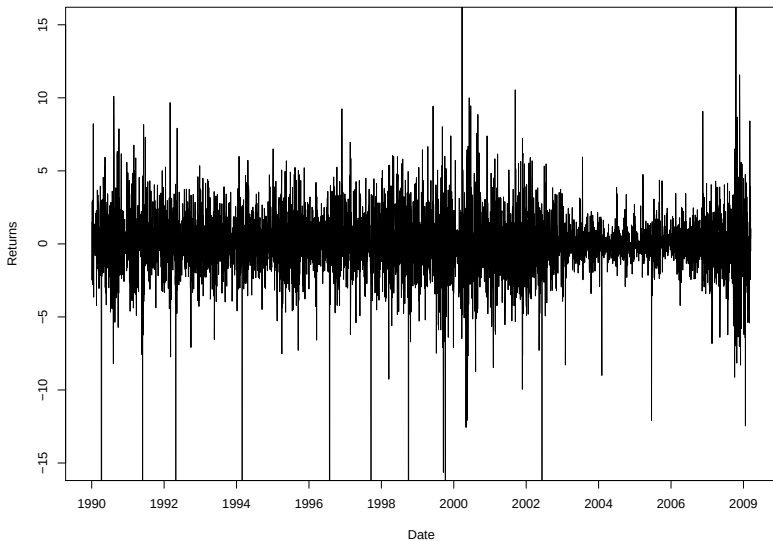
S & P 500 : 1990-2009 log Returns



Microsoft : 1990-2009



Microsoft : 1990-2009 log Returns



6.2 Observation Driven Models

For the observation-driven models,

$$Y_t \mid \mathcal{Y}_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$$

where $\mathcal{Y}_{t-1}$ contains all the information up to time $t - 1$, so that $\mathcal{Y}_t = y_{1:(t-1)}$.

$$Y_t = \sigma_t Z_t,$$

where $\{Z_t\}$ is a sequence of independent $\mathcal{N}(0, 1)$ random variables.

The ARCH(p) model allows the conditional variance σ_t^2 of Y_t to be a linear combination of past squared observations, so that

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i}^2.$$

Properties of the ARCH(1) model

The parameters α_0 and α_1 have to be non-negative, and the process is stationary if and only if $\alpha_1 < 1$, with

$$\mathbb{V}ar[Y_t] = \mathbb{E}[Y_t^2] = \alpha_0 / (1 - \alpha_1).$$

All the odd moments of Y_t are zero by symmetry, while the fourth moment exists if and only if $3\alpha_1^2 < 1$ and is

$$\mathbb{E}[Y_t^4] = \frac{3\alpha_0^2 (1 - \alpha_1^2)}{(1 - \alpha_1)^2 (1 - 3\alpha_1^2)}$$

and thus the kurtosis is

$$-3 + \frac{\mathbb{E}[Y_t^4]}{\mathbb{E}[Y_t^2]^2}$$

and is greater than zero and hence the marginal distribution of Y_t is *leptokurtic* (that is, has fat tails).

GARCH(p, q) model

The GARCH(p, q) process is an extension to the ARCH(p) model which models σ_t^2 as dependent on its lagged values;

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i}^2 + \sum_{i=1}^q \beta_i \sigma_{t-i}^2.$$

The most widely used GARCH model is that of order (1, 1). Sufficient conditions for $\sigma_t^2 \geq 0$ are $\alpha_i \geq 0, i = 0, 1$ and $\beta_1 \geq 0$.

The GARCH(1, 1) process has zero mean, and is second order stationary if and only if $\alpha_1 + \beta_1 < 1$, with

$$\mathbb{V}ar[Y_t] = \frac{\alpha_0}{(1 - \alpha_1 - \beta_1)}$$

and all the odd moments equal to zero.

GARCH(p, q) model

If in addition,

$$3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$$

the fourth moment exists and is equal to

$$\mathbb{E} \left[Y_t^4 \right] = \frac{3\alpha_0^2 (1 + \alpha_1 + \beta_1)}{(1 - \alpha_1 - \beta_1) (1 - 3\alpha_1^2 - 2\alpha_1\beta_1 - \beta_1^2)}$$

and again y_t exhibits leptokurtosis.

A special case of the GARCH(1, 1) model has $\alpha_1 + \beta_1 = 1$, which is called the Integrated GARCH(IGARCH) model.

There exist many other versions of the ARCH type models,

- Exponential GARCH(EGARCH)
- ARCH-in-Mean (ARCH-M)
- TGARCH
- MGARCH

6.2 Observation Driven Models

Observation-driven models are built out of one-step-ahead prediction densities. These densities allow the likelihood function to be constructed via the usual prediction error decomposition.

Therefore, the maximum likelihood estimation of the unknown parameters in the model is in principle straightforward. However, there are also a number of drawbacks to ARCH type models.

- the parameter constraints, imposed so that the conditional variance σ_t^2 remains non-negative, are often violated when estimating these coefficients.
- GARCH models rule out a random oscillatory behavior of the conditional variance process.

6.3 Inference For ARCH/GARCH Models

The ARCH(1) process is defined as

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2,$$

where $\alpha_0 \geq 0, \alpha_1 \geq 0$ are the two parameters about which inference is required.

The ARCH(p) process is defined as

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i}^2,$$

where $\alpha_0 \geq 0, \alpha_i \geq 0$ are the parameters of the ARCH(p) model.

6.3 Inference For ARCH/GARCH Models

Recall that the moments of the ARCH(1) model are given as follows

- (i) $\mathbb{E}[Y_t] = \mathbb{E}[Y_t^3] = 0$
- (ii) The second moment of Y_t is

$$\mathbb{E}[Y_t^2] = \frac{\alpha_0}{(1 - \alpha_1)}, \quad \text{for } 0 \leq \alpha_1 < 1.$$

- (iii) The fourth moment of Y_t is

$$\mathbb{E}[Y_t^4] = 3\mathbb{E}[\sigma_t^4] = \frac{3\alpha_0^2 (1 - \alpha_1^2)}{(1 - \alpha_1)^2 (1 - 3\alpha_1^2)}, \quad \text{for } 0 \leq \alpha_1^2 < \frac{1}{3}.$$

6.3 Inference For ARCH/GARCH Models

(iv) The kurtosis of Y_t is given by

$$\kappa = 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2}, \quad \text{for } \alpha_1^2 < \frac{1}{3}.$$

- If $\alpha_1 = 0$ then $\kappa = 3$, and the distribution is Normal.
- If $\alpha_1 > 0$ then $\kappa > 3$, and the distribution is heavy-tailed.

(v) The autocorrelation function (ACF) of Y_t^2 is given by

$$\rho_{Y_t^2}(h) = \alpha_1^h,$$

where $h = 0, 1, \dots, n$ for all $n \geq 0$.

The variance characteristics are solely dependent on the nature of the parameter α_1 .

Stationarity in ARCH(1) Variance

The condition for stationarity in variance is that $\alpha_1 < 1$; in this case, sudden changes to the error variance have an impact that decrease at an exponential rate and will eventually diminish in subsequent periods.

The conditional variance, σ_t^2 , varies over time and is dependent on past squared error terms. Note that

- ▶ The sequence Y_t is white noise and Y_t^2 is an autoregressive process, hence the existence of *volatility clustering* is partly controlled by α_1 .
- ▶ Note that Y_t^2 is not necessarily covariance stationary; its variance will be finite only if $3\alpha_1^2 < 1$.

Persistence

When $\alpha_1 > 1$ shocks to the variance in period one will have a more than proportionate impact in subsequent periods,

- ▶ Effects in the previous period causing greater shocks in the next, leading to instability in the system.
- ▶ The marginal variance is not finite, and the conditional variance grows at a more than proportionate rate (dependent on α_1) in every subsequent period.

The conditional time-varying error variance should always be positive; thus parameter constraints or reparameterization is needed.

Likelihood Function For ARCH(1) Model

Let y_0 be a fixed constant used to seed the recursion for σ_t . The likelihood for ARCH(1) takes the form

$$f(y_{1:n}|y_0, \alpha_0, \alpha_1) = \prod_{t=1}^n \left(\frac{1}{2\sigma_t^2} \right)^{\frac{1}{2}} \exp \left(-\frac{y_t^2}{2\sigma_t^2} \right),$$

so the log-likelihood is

$$\begin{aligned} \log f(y_{1:n}|y_0, \alpha_0, \alpha_1) &= \sum_{t=1}^n \log f(y_t|y_{t-1}, \alpha_0, \alpha_1) \\ &= \text{const.} - \frac{1}{2} \sum_{t=1}^n \left[\log \sigma_t^2 + \frac{y_t^2}{\sigma_t^2} \right]. \end{aligned}$$

Likelihood Function For ARCH(1) Model

To obtain the ML estimates we differentiate with respect to α_0 , α_1 respectively to obtain the *score equations*:

$$\begin{aligned}\frac{\partial \log f}{\partial \alpha_0} &= \frac{1}{2} \sum_{t=1}^n \left(\frac{\partial \sigma_t^2}{\partial \alpha_0} \right) \frac{1}{\sigma_t^2} \left(\frac{y_t^2}{\sigma_t^2} - 1 \right), & \frac{\partial \sigma_t^2}{\partial \alpha_0} &= 1, \\ \frac{\partial \log f}{\partial \alpha_1} &= \frac{1}{2} \sum_{t=1}^n \left(\frac{\partial \sigma_t^2}{\partial \alpha_1} \right) \frac{1}{\sigma_t^2} \left(\frac{y_t^2}{\sigma_t^2} - 1 \right), & \frac{\partial \sigma_t^2}{\partial \alpha_1} &= y_{t-1}^2.\end{aligned}$$

and set these derivatives to zero.

For ARCH(p), $(\alpha_0, \alpha_1)^n$ becomes $(\alpha_0, \alpha_1, \dots, \alpha_p)^n$ so

$$\left(\frac{\partial \sigma_t^2}{\partial \alpha_0}, \dots, \frac{\partial \sigma_t^2}{\partial \alpha_p} \right)^n = \left(1, y_{t-1}^2, \dots, y_{t-p}^2 \right)^n.$$

The GARCH(1, 1) model.

The GARCH(1, 1) process is defined by

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

with parameters $\alpha_0 \geq 0, \alpha_1 \geq 0, \beta_1 \geq 0$.

The GARCH(p, q) process is defined by

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2,$$

where $\alpha_0 \geq 0, \alpha_i \geq 0, \beta_j \geq 0$ are the parameters of the GARCH(p, q) model.

GARCH models

The moments of the GARCH(1,1) model take the following values:

(i) $\mathbb{E}[Y_t] = \mathbb{E}[Y_t^3] = 0$

(ii) If $0 \leq \alpha_1 + \beta_1 < 1$,

$$\mathbb{E}[Y_t^2] = \frac{\alpha_0}{(1 - \alpha_1 - \beta_1)},$$

(iii) If $0 \leq \alpha_1 + \beta_1 < 1$ and $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$

$$\mathbb{E}[Y_t^4] = \frac{3\alpha_0^2 (1 + \alpha_1 + \beta_1)}{(1 - \alpha_1 - \beta_1) (1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2)},$$

The fourth moment does not exist when the sum of $\alpha_1 + \beta_1$ is close to one, and the value of α_1 is not close to zero.

(iv) If $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$, the kurtosis is

$$\kappa = \frac{3(1 + \alpha_1 + \beta_1)(1 - \alpha_1 - \beta_1)}{(1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2)},$$

When $\beta_1 = 0$, this condition is the same as the ARCH(1) model, but when $\beta_1 > 0$, α_1 has to be less than $\sqrt{\frac{1}{3}}$. For example, in the typical case where α_1 is not close to zero and β_1 is near to one, κ does not exist.

(v) The ACF of Y_t^2 is given by

$$\begin{aligned}\rho(1) &= \frac{\alpha_1(1 - \beta_1^2 - \alpha_1\beta_1)}{(1 - \beta_1^2 - 2\alpha_1\beta_1)} \\ \rho(h) &= (\alpha_1 + \beta_1)\rho_{h-1}, \quad \text{for } h \geq 2.\end{aligned}$$

The ACF declines geometrically at the rate of $\alpha_1 + \beta_1$. If α_1 is sufficiently small and the sum of $\alpha_1 + \beta_1$ is close to one, then there exists a slowly decreasing autocorrelation function with finite kurtosis.

The stationarity of the GARCH(p, q) model is ensured if the coefficients in the conditional variance equation sum to less than one (i.e. $\alpha_1 + \dots + \alpha_p + \beta_1 + \dots + \beta_q < 1$), in which case the unconditional variance of Y_t ,

$$\frac{\alpha_0}{1 - (\alpha_1 + \dots + \alpha_p + \beta_1 + \dots + \beta_q)},$$

is a finite constant.

In this case, shocks to the variance term do not have a permanent effect, but fade over time.

For the GARCH(1, 1) model, the following is known:

- (i) If $\alpha_1 + \beta_1 < 1$, under normality of the residual errors,

$$\mathbb{V}ar[Y_t] = \text{const.} \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)}$$

and $\mathbb{C}ov[Y_t, Y_s] \neq 0$.

- (ii) If $\alpha_1 + \beta_1 \rightarrow 1$, the ACF will decay quite slowly, indicating a relatively slow change in conditional variance. This has often been observed to occur in practice especially with high frequency data. This indicates that a shock at time t will persist for many future periods.

- (iii) If $\alpha_1 + \beta_1 = 1$, then a shock at time t will lead to a permanent change in all future periods; this also refers to the Integrated-GARCH(I-GARCH) model, where the conditional variance is non-stationary and the unconditional variance does not exist.
- (iv) If $\alpha_1 + \beta_1 > 1$, then a shock at time t will have a destabilizing effect, not only leading to a permanent change in future periods, but reinforcing itself over time.

It is widely thought that the GARCH(1, 1) is an adequate model in a wide range of volatility modelling situations; it is a simple model, and thus avoids the problems of overfitting, and yet has been found to have the main features present in more complex models.

Likelihood Function For GARCH(1,1) Model

ML estimation can be constructed for all GARCH-type models; it is identical to that for the ARCH model, with the addition of score equations for β

$$\frac{\partial \log f}{\partial \beta_1} = \frac{1}{2} \sum_{t=1}^n \left(\frac{\partial \sigma_t^2}{\partial \beta_1} \right) \frac{1}{\sigma_t^2} \left(\frac{y_t^2}{\sigma_t^2} - 1 \right),$$

where

$$\frac{\partial \sigma_t^2}{\partial \beta_1} = \sigma_{t-1}^2 + \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \beta_1}.$$

Likelihood Estimation For GARCH(1,1) Model

To obtain the ML estimates, we need to implement a numerical calculation for the partial derivatives recursively for $t = 1, \dots, n$.

This is more complicated than the ARCH case due to the recursive term in the score equation for β_1 . The resulting estimators have properties of asymptotic normality and consistency.

Quasi Maximum Likelihood (QML) estimation may also asymptotic normal distribution for the QML estimates and are in practice close to the ML estimates.

Bayesian inference For GARCH(1,1) Model

Assuming a log-normal prior for α_0 , and a Dirichlet prior for (α_1, β_1) , the Bayesian posterior distribution is

$$\begin{aligned} p(\alpha_0, \alpha_1, \beta_1 | y_{1:n}) &\propto \prod_{t=1}^n \left(\frac{1}{2\sigma_t^2} \right)^{\frac{1}{2}} \exp \left(-\frac{y_t^2}{2\sigma_t^2} \right) \alpha_0^{-1} \exp \left(-\frac{(\log \alpha_0)^2}{2\sigma_{\alpha_0}^2} \right) \\ &\quad \times \alpha_1^{\gamma_1-1} \beta_1^{\gamma_2-1} (1 - \alpha_1 - \beta_1)^{\gamma_3-1} \\ &\propto \exp \left\{ -\frac{1}{2} \sum_{t=1}^n \log \left(\sigma_t^2 + \frac{y_t^2}{\sigma_t^2} \right) - \frac{(\log \alpha_0)^2}{2\sigma_{\alpha_0}^2} \right\} \\ &\quad \times \alpha_0^{-1} \times \alpha_1^{\gamma_1-1} \beta_1^{\gamma_2-1} (1 - \alpha_1 - \beta_1)^{\gamma_3-1}. \end{aligned}$$

Bayesian inference For GARCH(1,1) Model

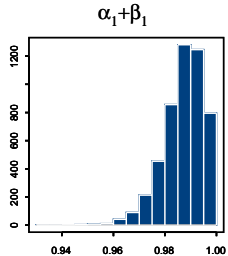
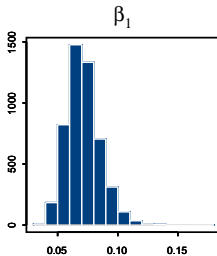
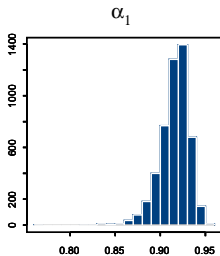
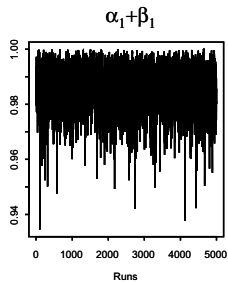
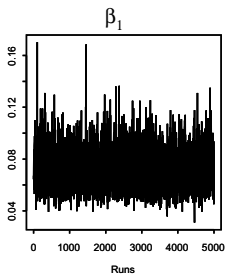
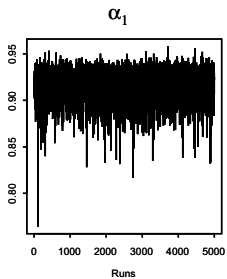
Example : FX Returns in a number of twelve Far Eastern and other currencies

- Daily data
- Hourly data

taken around the time of the market crash in the late nineties.

Following results from a Bayesian analysis via Markov chain Monte Carlo (MCMC). In the MCMC algorithm, used 2,500,000, and recorded parameters at every 500th iteration.

Daily JPY GARCH(1,1)



Daily Results GARCH(1,1)

	α_1 Mean, Median, Std	β_1 Mean, Median, Std	$(\alpha_1 + \beta_1)$ Mean, Median, Std
D-THB	0.9036, 0.9039, 0.0091	0.0961, 0.0957, 0.0091	0.9997, 0.9998, 0.0004
D-SGD	0.0347, 0.0199, 0.0409	0.6248, 0.6403, 0.2144	0.6594, 0.6743, 0.2097
D-JPY	0.9159, 0.9181, 0.0160	0.0714, 0.0701, 0.0138	0.9872, 0.9883, 0.0079
D-HKD	0.0355, 0.0204, 0.0421	0.5823, 0.5804, 0.2208	0.6177, 0.6147, 0.2166
D-GBP	0.0441, 0.0289, 0.0469	0.2045, 0.1925, 0.0740	0.2486, 0.2381, 0.0810
D-CHF	0.2244, 0.1365, 0.2402	0.1127, 0.1130, 0.0419	0.3371, 0.2622, 0.2155
D-CAD	0.9223, 0.9233, 0.0096	0.0707, 0.0702, 0.0095	0.9930, 0.9940, 0.0052
D-AUD	0.0394, 0.0239, 0.0441	0.4568, 0.4224, 0.2047	0.4963, 0.4641, 0.2003
H-THB	0.0269, 0.0174, 0.0284	0.6468, 0.6416, 0.0981	0.7295, 0.6684, 0.0973
H-SGD	0.2945, 0.2932, 0.0363	0.6845, 0.6866, 0.0414	0.9790, 0.9849, 0.0192
H-JPY	0.9193, 0.9200, 0.0103	0.0797, 0.0789, 0.0102	0.9990, 0.9993, 0.0010
H-HKD	0.6618, 0.6621, 0.0350	0.3182, 0.3183, 0.0368	0.9800, 0.9838, 0.0162

Posterior statistics of GARCH(1,1) model for 12 FX series.

Daily Results GARCH(1,1)

The table above contains posterior summaries for the three parameters in the GARCH(1, 1) model for all FX series.

To explore the stability and persistence of GARCH(1, 1) model, the sum of the $\alpha_1 + \beta_1$ should be examined. From the table, five data series (D-THB, H-JPY, D-CAD, H-JPY, H-SGD) yield values of $(\alpha_1 + \beta_1)$ to be very close to one

In addition, the estimated values of α_1 are close to one and β_1 are close to zero. Thus there exists considerable persistence in volatility, moving towards non-stationarity.

A Constrained GARCH(1,1) Model

We introduced a constrained model to ensure the existence of higher order moments; the kurtosis exists for the observable GARCH(1, 1) process only when the inequality

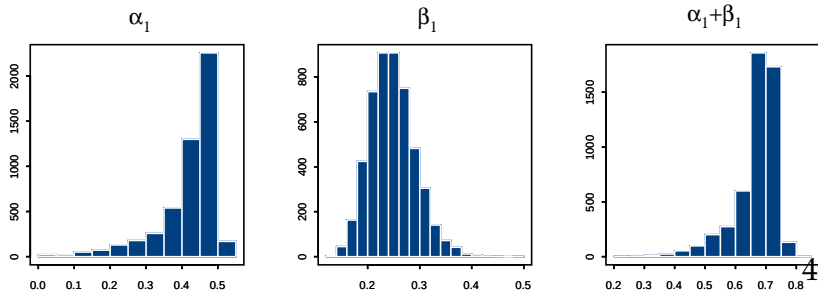
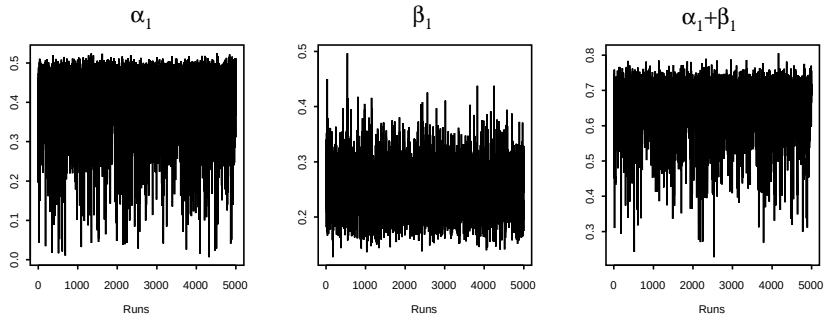
$$3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$$

holds; further, the fourth moment only exists for a certain range of values of α_1, β_1 .

The additional constraint can be explicitly incorporated into the MCMC simulation scheme; we reject points generated by the proposal mechanism that violate the constraint

Note that such constraints are typically problematic in conventional (non-simulation based) classical and Bayesian inference.

A Constrained GARCH(1,1) Model



Daily Results GARCH(1,1)

GARCH (1,1)	α_1 Mean, Median, Std	β_1 Mean, Median, Std	$(\alpha_1 + \beta_1)$ Mean, Median, Std
D-THB	0.3838, 0.3837, 0.0086	0.4551, 0.4558, 0.0164	0.8389, 0.8394, 0.0079
D-SGD	0.0358, 0.0214, 0.0214	0.6170, 0.622788, 0.6228	0.6528, 0.6582, 0.2078
D-JPY	0.4200, 0.4481, 0.0839	0.2479, 0.2450, 0.0435	0.6679, 0.6878, 0.0715
D-HKD	0.0356, 0.0210, 0.0413	0.5841, 0.5859, 0.2207	0.6197, 0.6182, 0.2140
D-GBP	0.0418, 0.0251, 0.0472	0.2025, 0.1924, 0.0732	0.2443, 0.2344, 0.0820
D-CHF	0.1452, 0.1120, 0.1216	0.1213, 0.1182, 0.0383	0.2665, 0.2431, 0.1111
D-CAD	0.4803, 0.4974, 0.0548	0.1517, 0.1503, 0.0288	0.6319, 0.6444, 0.0498
D-AUD	0.0405, 0.0254, 0.0449	0.4561, 0.4277, 0.1990	0.4966, 0.4721, 0.1950
H-THB	0.2566, 0.2561, 0.0280	0.6586, 0.6603, 0.0450	0.9152, 0.9173, 0.0205
H-SGD	0.0285, 0.0193, 0.0293	0.6480, 0.644049, 0.0994	0.6765, 0.6737, 0.0987
H-JPY	0.4227, 0.4235, 0.0186	0.3708, 0.369196, 0.0378	0.7934, 0.7930, 0.0199
H-HKD	0.4315, 0.4326, 0.0192	0.3484, 0.3475, 0.0380	0.7800, 0.7797, 0.0205

Posterior statistics for the constrained model of GARCH(1,1)
for 12 FX series

Daily Results GARCH(1,1)

Results: The posterior statistics values for this GARCH(1, 1) model are displayed above.

No currency estimates the values of $(\alpha_1 + \beta_1)$ to be very close to 1, although the H-THB obtains the highest estimated posterior mean value of 0.9152.

We conclude that this constrained model, where the existence of kurtosis is required in the model, produces very different parameter estimates; this may have serious consequences for prediction.

The Student- t GARCH(1, 1) Model

The leptokurtosis of the observed returns series can be modelled explicitly. The Student- t GARCH(1, 1) model can be formulated as

$$\begin{aligned}Y_t &= Z_t \sigma_t^2 \\Z_t &\sim \mathcal{N}(0, k\lambda_t) \\ \lambda_t &\sim \text{IGamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)\end{aligned}$$

and for stationarity, $0 < \alpha_1 + \beta_1 < 1$.

The Student- t GARCH(1, 1) Model

- ▶ The parameters λ_t , $t = (1, \dots, n)$ modify the model so that

$$Y_t | \sigma_t^2 \sim St(0, k\sigma_t^2, \nu),$$

where ν takes some positive value, and k is a constant term.

- ▶ For the conditional variance of Y_t to be finite, we require $\nu > 2$. Again, choosing a constant term

$$k = \frac{(\nu - 2)}{\nu}$$

ensures that the conditional variance of y_t remains as σ_t^2 , and setting each $\lambda_t = 1$ recovers the original GARCH(1, 1)

The Student- t GARCH(1,1) Model with ν unknown

- ▶ For $4 < \nu < \infty$, the conditional kurtosis for the t -GARCH(1,1) model is

$$3(\nu - 2)/(\nu - 4)$$

which is greater than that of a normal.

- ▶ The kurtosis for the Student- t GARCH(1,1) only exists if $\nu > 4$.
- ▶ As $\nu \rightarrow \infty$, the Student density tends to a normal.
- ▶ All odd moments are zero.

Results for the t -GARCH(1,1) Model.

t -GARCH(1,1) $\nu = 5$	Median, Std α_1	Median, Std β_1	Median, Std ($\alpha_1 + \beta_1$)
D-THB	0.8672, 0.8677, 0.0148	0.1252, 0.1247, 0.0157	0.9924, 0.9924, 0.0048
D-SGD	0.5928, 0.5948, 0.0361	0.2628, 0.2612, 0.0310	0.8556, 0.8561, 0.0289
D-JPY	0.9507, 0.9521, 0.0107	0.0250, 0.0244, 0.0057	0.9757, 0.9769, 0.0072
D-HKD	0.3955, 0.3947, 0.0275	0.5754, 0.5776, 0.0347	0.9709, 0.9760, 0.0224
D-GBP	0.0118, 0.0057, 0.0165	0.1150, 0.1124, 0.0326	0.1268, 0.1246, 0.0348
D-CHF	0.6865, 0.9027, 0.3607	0.0382, 0.0282, 0.0258	0.7247, 0.9291, 0.3399
D-CAD	0.9337, 0.9352, 0.0140	0.0380, 0.0373, 0.0080	0.9716, 0.9729, 0.0099
D-AUD	0.1207, 0.0097, 0.0347	0.0264, 0.1189, 0.0320	0.1471, 0.1432, 0.0433
H-THB	0.2597, 0.2589, 0.0578	0.4477, 0.4441, 0.0673	0.7074, 0.7062, 0.0670
H-SGD	0.2352, 0.2340, 0.0499	0.4140, 0.4097, 0.0625	0.6493, 0.6488, 0.0579
H-JPY	0.9086, 0.9103, 0.0181	0.0683, 0.0668, 0.0154	0.9769, 0.9775, 0.0084
H-HKD	0.2308, 0.2259, 0.0767	0.4339, 0.4308, 0.0629	0.6646, 0.6644, 0.0732

Posterior statistics of t -GARCH(1,1) model with $\nu = 5$ for 12 FX series.

Student- t Model with ν unknown

For the Bayesian t -GARCH(1,1) model, if ν is also to be included as an unknown parameter, inference can also be made about it.

t -GARCH(1,1) ν unknown	ν Mean, Median, Std
D-THB	6.8834, 6.8834, 0.3640
D-SGD	6.9735, 6.9735, 0.3619
D-JPY	8.3251, 8.3251, 0.5570
D-HKD	5.0460, 5.0460, 0.1667
D-GBP	7.5149, 7.5149, 0.4321
D-CHF	8.7967, 8.7967, 0.6852
D-CAD	9.7992, 9.7992, 0.8881
D-AUD	7.2669, 7.2669, 0.3898
H-THB	6.5218, 6.5218, 0.3756
H-SGD	6.3907, 6.3907, 0.3614
H-JPY	8.1712, 8.1712, 0.6470
H-HKD	6.7072, 6.7071, 0.4109

Posterior statistics for ν in t -GARCH(1,1) model for 12 FX series

6.4 Stochastic Volatility Models

The main alternative to ARCH type models is the stochastic volatility (SV), a class of *parameter-driven* models and allows the variance of the observations to be an unobserved *random* process.

SV models overcome the drawbacks encountered with GARCH models and fit more naturally into the theoretical framework within which much of modern finance theory has been developed. In particular, SV models can easily be seen to have simple continuous-time analogues used for option pricing.

6.4 Stochastic Volatility Models

The simplest SV model is

$$\begin{aligned}Y_t &= \exp(h_t/2) Z_t \\h_t &= \gamma + \phi h_{t-1} + U_t\end{aligned}$$

where y_t is, as usual, the observation at time t , the Z_t s are independent identically distributed (i.i.d.) $\mathcal{N}(0, 1)$ random variables, the U_t s are also i.i.d. $\mathcal{N}(0, \sigma_\eta^2)$ random variables.

The latent process $h_t \sim AR(1)$ is interpreted as the random and uneven flow of new information into the market, and ϕ is the persistence in the volatility.

6.4 Stochastic Volatility Models

Leverage: The advantage of using SV models lies in the fact that they provide greater flexibility in describing stylized facts such as *leverage*, which causes the conditional variance to respond asymmetrically to rises and falls in y_t .

Falling stock prices cause the debt to equity ratio of firms to increase and this entails more uncertainty and in turn increased volatility, whereas rising stock prices decrease a firm's debt to equity ratio, while increasing investor's confidence causing lower levels of volatility.

The leverage effect cannot be described by the ARCH or GARCHmodel, because the conditional variance depends only on the size of lagged y_t 's and not on their sign; however, it can be captured by the EGARCHmodel.

6.4 Stochastic Volatility Models

The SV model with leverage again has

$$\begin{aligned}Y_t &= \exp(h_t/2) Z_t \\h_t &= \gamma + \phi h_{t-1} + U_t\end{aligned}$$

but now the two processes $\{Z_t\}$ and $\{U_t\}$ are correlated, that is one of

$$\mathbb{C}orr[Z_t, U_t] = \rho \quad \mathbb{C}orr[Z_{t+1}, U_t] = \rho \quad \mathbb{C}orr[Z_t, U_{t+1}] = \rho$$

holds.

6.4 Stochastic Volatility Models

Persistence in Volatility: Asset returns have been found to have quite high autocorrelations for long lags. The SV model can capture this phenomenon very easily. The parameter ϕ in the AR process is interpreted as the persistence in the volatility and the restriction

$$|\phi| < 1$$

is typically imposed to ensure that the series h_t of the log-volatilities is stationary.

Most studies in the SV literature have found evidence of near unit root behavior of the process h_t with values of ϕ ranging from 0.8 to 0.995 demonstrating that the volatility of asset returns is indeed highly persistent. However, h_t can also be allowed to follow a random walk by setting $\phi = 1$.

Properties of the Stochastic Volatility Model

Suppose $\{Z_t\}$ and $\{U_t\}$ are independent. If $|\phi| < 1$, the process $\{h_t\}$ is strictly stationary with unconditional mean and variance given respectively by

$$\begin{aligned}\mu_h &= \mathbb{E}[h_t] = \frac{\gamma}{1 - \phi} \\ \sigma_h^2 &= \mathbb{V}ar[h_t] = \frac{\sigma_\eta^2}{1 - \phi^2}.\end{aligned}$$

Since Y_t is the product of two processes, Z_t and $\exp(h_t/2)$, and Z_t is always stationary, y_t will also be stationary if and only if h_t is stationary.

Properties of the Stochastic Volatility Model

Then

$$\mathbb{E}[Y_t] = \mathbb{E}[Y_t | \mathcal{Y}_{t-1}] = 0$$

so that Y_t is zero mean and the autocorrelation function (ACF) of Y_t is

$$\rho_{Y_t}(\tau) = \mathbb{E}[Y_t Y_{t-\tau}] = \mathbb{E}\left[\exp\left(\frac{h_t}{2} + \frac{h_{t-\tau}}{2}\right)\right] \mathbb{E}[Z_t Z_{t-\tau}] = 0.$$

Thus the series Y_t is a *martingale difference*. Furthermore, if the distribution of Z_t is symmetric, it follows that all the odd moments of Y_t are zero.

Properties of the Stochastic Volatility Model

By assumption, $\exp(h_t)$ is log-normal distributed, so from standard properties of the log-normal distribution, we have

$$\mathbb{E} \left[\exp(h_t)^j \right] = \exp \left\{ j\mu_h + \frac{1}{2}j^2\sigma_h^2 \right\},$$

so that, if r is even and h_t is stationary, all the even moments of y_t exist and are given by the formula

$$\mathbb{E} [Y_t^r] = \mathbb{E} \left[(\exp h_t)^{r/2} \right] \mathbb{E} [Z_t^r] = \exp \left\{ \frac{r\mu_h}{2} + \frac{r^2}{8}\sigma_h^2 \right\} \frac{r!}{2^{r/2} \left(\frac{r}{2}\right)!}.$$

Properties of the Stochastic Volatility Model

In particular

$$\mathbb{V}ar[Y_t] = \mathbb{E}[Y_t^2] = \exp\left\{\mu_h + \frac{1}{2}\sigma_h^2\right\}$$

Hence if h_t is stationary, Y_t is a white-noise process.

The fourth moment is

$$\mathbb{E}[Y_t^4] = 3 \exp\{2\mu_h + 2\sigma_h^2\}$$

and so the kurtosis for Y_t is

$$-3 + \mathbb{E}[Y_t^4] / \mathbb{E}[Y_t^2]^2 = 3(\exp(\sigma_h^2) - 1),$$

which is greater than 0 if σ_h^2 is positive. Thus, Y_t has a leptokurtic, symmetric distribution.

Linearization

The dynamic properties of the SV model appear most clearly if we square Y_t and take logarithms, so that

$$\log Y_t^2 = h_t + \log Z_t^2.$$

If Z_t has a standard normal distribution, then $\log Z_t^2$ has a log-chi-square distribution with mean $\psi(1) - \log 2 \simeq -1.2704$ and variance $\pi^2/2 \simeq 4.9348$, where $\psi(\cdot)$ is the digamma function.

Thus, if we define $\xi_t = \log Z_t^2 + 1.2704$, then clearly ξ_t is i.i.d. with mean zero and variance $\pi^2/2$ and we may rewrite the model as

$$\log Y_t^2 = -1.2704 + h_t + \xi_t.$$

Therefore, it follows that $\log Y_t^2$ is a linear process which is the sum of the AR(1) process h_t and white-noise. Hence, $\log Y_t^2$ behaves approximately as an ARMA(1, 1) process, with its ACF being equivalent to that of an ARMA(1, 1) process and given by

$$\rho_{\log Y_t^2}(\tau) = \frac{\phi^\tau}{1 + (\pi^2/2\sigma_h^2)}, \quad \tau = 1, 2, \dots$$

The ACF of the powers of the absolute values of y_t are also available; the ACF of Y_t^2 is approximately proportional to that of the AR(1) process h_t .

Linearization

When the errors Z_t have a Student t -distribution with ν degrees of freedom, Y_t is also white noise if and only if the process h_t is stationary, and if $Z_t \sim t_\nu$, then

$$\mathbb{V}ar [Z_t] = \mathbb{E} [Z_t^2] = \nu / (\nu - 2), \nu > 2$$

and

$$\mathbb{E} [Z_t^4] = 3\nu^2 / (\nu - 2)(\nu - 4), \nu > 4.$$

Hence, it follows immediately that the unconditional variance of Y_t generalizes in this case to

$$\mathbb{V}ar [Y_t] = \mathbb{E} [Y_t^2] = \frac{\nu}{\nu - 2} \exp \left\{ \mu_h + \frac{1}{2} \sigma_h^2 \right\}, \nu > 2$$

The kurtosis for $\nu > 4$ is $3 [-1 + (\nu - 2) \exp (\sigma_h^2) / (\nu - 4)]$.

Linearization

The SV model with $Z_t \sim t_\nu$ can also be transformed into a linear form; let

$$Z_t = \zeta_t \kappa_t^{-1/2},$$

where $\zeta_t \sim \mathcal{N}(0, 1)$ and $\nu \kappa_t$ is independent of ζ_t and has a chi-square distribution with ν degrees of freedom.

Therefore, $\log Z_t^2 = \log \zeta_t^2 - \log \kappa_t$ and it follows that

$$\mathbb{E}[\log \kappa_t] = \psi(\nu/2) - \log(\nu/2)$$

and

$$\mathbb{V}ar[\log \kappa_t] = \psi'(\nu/2)$$

with $\psi(\cdot)$ and $\psi'(\cdot)$ the digamma and trigamma functions, respectively.

Linearization

Therefore, if we now define

$$\xi_t = (\log \zeta_t^2 + 1.2704) + (\log \kappa_t - \psi(\nu/2) + \log(\nu/2)),$$

then clearly ξ_t is i.i.d. with mean zero and variance $\pi^2/2 + \psi'(\nu/2)$. Squaring y_t and taking logarithms gives

$$\log Y_t^2 = -1.2704 - \psi(\nu/2) + \log(\nu/2) + h_t + \xi_t,$$

which is again a linear process which adds the i.i.d. ξ_t to the AR(1) h_t . The ACF is

$$\rho_{\log Y_t^2}(\tau) = \frac{\phi^\tau}{1 + [\psi'(\nu/2) + \pi^2/2] / \sigma_h^2}.$$

6.5 Inference for the Stochastic Volatility Model

No analytic expression exists for the densities $p(y_t | \mathcal{Y}_{t-1})$, and this makes the likelihood function hard to evaluate; the distribution of Y_t conditional on past information $\mathcal{Y}_{t-1}$ does not possess an analytic expression.

The likelihood can be computed by integrating the latent log-volatilities out of the joint probability distribution. If $h_{1:n} = (h_1, \dots, h_n)^\top$ and $\boldsymbol{\theta} = (\gamma, \phi, \sigma_\eta^2)$, then the likelihood is

$$L(y_{1:n}; \boldsymbol{\theta}) = \int p(y_{1:n}, \mathbf{h} | \boldsymbol{\theta}) d\mathbf{h}_{1:n} = \int p(y_{1:n} | \mathbf{h}_{1:n}, \boldsymbol{\theta}) p(\mathbf{h} | \boldsymbol{\theta}) d\mathbf{h}_{1:n}.$$

This last integral requires the use of numerical procedures and this makes the estimation $\boldsymbol{\theta}$ quite involved.

Generalized Method-of-Moments (GMM)

The simplest estimation procedure of SV models is the Method-of-Moments. The key advantage of GMM is that it does not require the specification of the likelihood function, but only certain moment conditions are needed.

Given a sample of size n , $y_{1:n}$, the GMM procedure requires the construction of a vector $\mathbf{g}$, whose elements will be the differences between the unconditional expectations and the sample moments.

For the SV model there are three parameters we need to estimate, namely $\boldsymbol{\theta} = (\gamma, \phi, \sigma_\eta^2)$, and a large number of moments to use.

Generalized Method-of-Moments (GMM)

For example, we might define estimating function g with components

$$\begin{aligned} & \frac{1}{n} \sum y_t^2 - \mathbb{E}[Y_t^2] \\ & \frac{1}{n} \sum y_t^4 - \mathbb{E}[Y_t^4] \\ & \frac{1}{n} \sum y_t^2 y_{t-1}^2 - \mathbb{E}[Y_t^2 Y_{t-1}^2] \\ & \vdots \\ & \frac{1}{n} \sum y_t^2 y_{t-\tau}^2 - \mathbb{E}[Y_t^2 Y_{t-\tau}^2] \end{aligned}$$

where the theoretical values of $\mathbb{E}[Y_t^2 Y_{t-\tau}^2]$, for $\tau \geq 1$, can be found analytically.

Generalized Method-of-Moments (GMM)

The objective function to be minimized is then

$$Q = \mathbf{g}^\top \mathbf{W} \mathbf{g}$$

where $\mathbf{W}$ is a $(\tau + 2) \times (\tau + 2)$ positive definite, symmetric weighting matrix.

An advantage of the GMM method is simplicity; the main disadvantage is that it can be inefficient in small samples, although GMM estimators are consistent and asymptotically normal even when the residual errors are non-Gaussian.

GMM estimators are consistent if the observations y_t are stationary. When the persistence in the latent process, h_t , is high, i.e. ϕ is close to unity, as is usually the case in practice, the GMM estimator works poorly.

Generalized Method-of-Moments (GMM)

There are disadvantages:

- ▶ Estimates can be substantially biased especially for σ_{η}^2 , have large mean squared errors (MSE) when there is high persistence and low *coefficient of variation*

$$C.V. = \text{Var} [\exp (h_t)] / \mathbb{E} [\exp (h_t)]^2 = \exp (\sigma_h^2) - 1.$$

- ▶ GMM parameter estimates are not invariant to reparameterization

$$\psi = f(\theta) \quad \text{then} \quad \hat{\psi} \neq f(\hat{\theta})$$

- ▶ GMM estimation does not deliver filtered or smoothed estimates of h_t

6.6 Quasi-Maximum Likelihood estimation

The *Quasi-Maximum Likelihood* (QML) method is based on the linearization approach described earlier. Assuming that the errors $Z_t \sim \mathcal{N}(0, 1)$, and denoting $W_t = \log Y_t^2$, as has already been seen, the SV model can be written as

$$\begin{aligned}W_t &= -1.2704 + h_t + \xi_t, \\h_t &= \gamma + \phi h_{t-1} + U_t,\end{aligned}$$

where $\xi_t = \log Z_t^2 - \mathbb{E}[\log Z_t^2]$, with $\sigma_\xi^2 = \mathbb{V}ar[\xi_t] = \pi^2/2$.

This is a linear but non-Gaussian state-space model. The QML approach treats the observation errors, ξ_t , as though they were i.i.d $\mathcal{N}(0, \pi^2/2)$ and apply the standard Kalman filter.

6.6 Quasi-Maximum Likelihood estimation

The Kalman filter produces one-step-ahead forecasts of the observations, w_t , and the log-volatilities, h_t , as well as filtered estimates of the latter. Given a set of observations $\{y_1, \dots, y_n\}$, or equivalently $\{\log y_1^2, \dots, \log y_n^2\}$, the recursions can also be used to construct the Gaussian likelihood of the data via the prediction error decomposition.

If this Gaussian form of the likelihood is then maximized with respect to the hyperparameters of the model, typically using numerical procedures, it will yield QML estimates of the unknown parameters.

6.6 Quasi-Maximum Likelihood estimation

Assume that $|\phi| < 1$, so that h_t is stationary. Taking expectations on both sides of the observation equation, we obtain

$$\mathbb{E}[W_t] = \gamma^* = -1.2704 + \mu_h = -1.2704 + \gamma / (1 - \phi) .$$

Moreover, if we denote $W_t^* = W_t - \gamma^*$ to be the new observations centered around their unconditional mean and $\alpha_t = h_t - \mu_h$ be the mean-centered states, then the model can be rewritten as follows

$$\begin{aligned} W_t^* &= \alpha_t + \xi_t, \\ \alpha_t &= \phi \alpha_{t-1} + U_t. \end{aligned}$$

6.6 Quasi-Maximum Likelihood estimation

The latter state-space model does not explicitly contain the constant term γ of the state-transition equation, and a consistent estimator of γ^* is given by the sample mean of w_t , or equivalently $\log y_t^2$, and is also the QML estimator of γ^* .

The parameter γ has been “concentrated” out of the likelihood; we can apply the Kalman filter to the model with the centered observations, and obtain the QML estimates of $\theta = (\phi, \sigma_\eta^2)^\top$.

Once the estimates $\hat{\phi}$ and $\hat{\sigma}_\eta^2$ are available, the QML estimator of γ will be given by

$$\hat{\gamma} = (1 - \hat{\phi}) \left(1.2704 + \frac{1}{n} \sum \log y_t^2 \right).$$

6.6 Quasi-Maximum Likelihood estimation

The Kalman filter recursions then compute the *one-step-ahead prediction*, $a_{t|t-1}$, and the *filtered estimates*, $a_{t|t}$, of the unobserved states α_t assuming that the observations sequentially become available in the usual way. Initializing with

$$a_{0|0} = \mathbb{E}[\alpha_t] = 0 \quad P_{0|0} = \text{Var}[\alpha_t] = \sigma_\eta^2 / (1 - \phi^2),$$

the one-step-ahead prediction estimates of α_t and their mean square errors (MSEs) are respectively given by

$$a_{t|t-1} = \phi a_{t-1|t-1}, \quad P_{t|t-1} = \phi^2 P_{t-1|t-1} + \sigma_\eta^2, \quad t = 1, \dots, n,$$

while the filtered estimates, $a_{t|t}$.

6.6 Quasi-Maximum Likelihood estimation

The MSEs, $P_{t|t}$, are respectively given by

$$\begin{aligned}a_{t|t} &= a_{t|t-1} + P_{t|t-1}f_t^{-1} (w_t^* - a_{t|t-1}) \\P_{t|t} &= P_{t|t-1} - P_{t|t-1}^2f_t^{-1}, \quad t = 1, \dots, n,\end{aligned}$$

where the terms $w_t^* - a_{t|t-1}$ are the innovations in predicting w_t^* given past observations $\{w_{t-1}^*, \dots, w_1^*\}$ and $f_t = P_{t|t-1} + \sigma_\xi^2$ are the MSEs of the one-step-ahead prediction estimates of w_t^* .

Due to non-Gaussianity, the filtered and smoothed estimators $a_{t|t-1}$ and $a_{t|t}$ are only minimum mean square linear estimators (MMSLEs) of the unobserved variable α_t , given the observations up to time $t - 1$ and t ; they are optimal in the class of linear estimators.

6.6 Quasi-Maximum Likelihood estimation

A Gaussian (quasi) log-likelihood can be constructed

$$l_q(\boldsymbol{\theta}; w^*) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n \log f_t - \frac{1}{2} \sum_{t=1}^n \frac{(w_t^* - a_{t|t-1})^2}{f_t}.$$

The resulting QML estimators of $\boldsymbol{\theta}$ are consistent with asymptotically normal distribution.

The backward recursions produce the smoothed estimates $a_{t|n}$ of α_t along with their MSE $P_{t|n}$

$$\begin{aligned} a_{t|n} &= a_{t|t} + \phi P_{t|t} P_{t+1|t}^{-1} (a_{t+1|n} - \phi a_{t|t}) \\ P_{t|n} &= P_{t|t} + \phi^2 P_{t|t}^2 P_{t+1|t}^{-2} (P_{t+1|n} - P_{t+1|t}), \quad t = n-1, \dots, 1. \end{aligned}$$

6.6 Quasi-Maximum Likelihood estimation

The QML procedure can be applied to the SV model when ϕ is set equal to one and the log-volatilities are allowed to follow a random walk. When $\phi = 1$, the state-transition equation becomes

$$\alpha_t = \alpha_{t-1} + U_t$$

and the linearized SV model becomes a random walk plus noise model for W_t^* with the only unknown parameter being σ_η^2 .

The Kalman filter prediction and update equations and the recursions need to be initialized with a diffuse prior for α_1 , by setting $P_{1|0} = \kappa$, where κ is some large positive constant and $a_{1|0}$ an arbitrary constant.

6.6 Quasi-Maximum Likelihood estimation

QML estimation is not restricted only to the case when $Z_t \sim \mathcal{N}(0, 1)$, but with minor modifications can also be used to estimate a SV model with $Z_t \sim t_\nu$.

As before, if $|\phi| < 1$ and $Z_t \sim t_\nu$, let $Z_t = \zeta_t \kappa_t^{-1/2}$, with $\nu \kappa_t \sim \chi_\nu^2$ independent of $\zeta_t \sim \mathcal{N}(0, 1)$, which results in $w_t = \log y_t^2$, with $\mathbb{E}[\xi_t] = 0$ and

$$\sigma_\xi^2 = \text{Var}[\xi_t] = \pi^2/2 + \psi'(\nu/2).$$

In addition, w_t^* is obtained from w_t by subtracting the unconditional mean γ^* , in which case it is given by

$$\gamma^* = -1.2704 - \psi(\nu/2) + \log(\nu/2) + \gamma/(1 - \phi)$$

and thus the state-space form of the model has the same form as above.

6.6 Quasi-Maximum Likelihood estimation

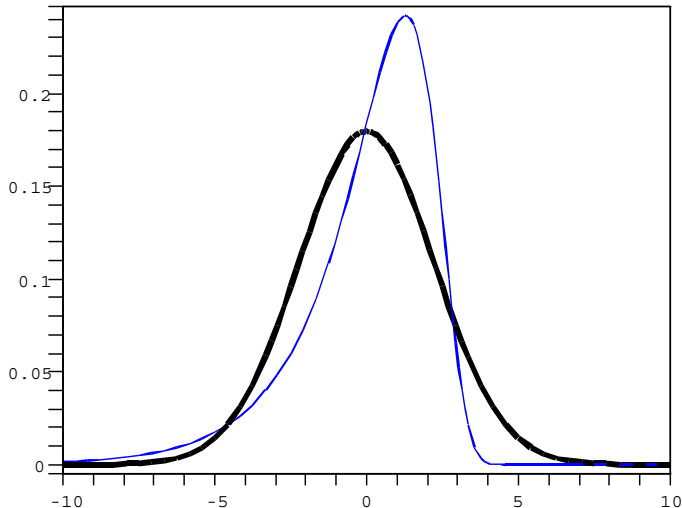
The QML procedure is inefficient compared to ML, as it approximates the density of a $\log(\chi_1^2)$ variable by a normal density.

A comparison of these densities (below) illustrates that this approximation is rather inappropriate; the adequacy of the approximation depends critically on the true parameter values

- ▶ For large values of σ_η^2 , the AR(1) process, h_t , dominates ξ_t , the non-Gaussian error term in the observation equation, and the normal approximation may be adequate and the QML approach is close to optimal.
- ▶ However, as σ_η^2 decreases, the approximation worsens and for small values of σ_η^2 , usually found in practice, the QML estimates can be extremely biased and have high root mean square error.

6.6 Quasi-Maximum Likelihood estimation

Comparison of the $\log(\chi_1^2)$ density (thin solid line) with the $\mathcal{N}(0, \pi^2/2)$ density (thick solid line).



6.6 Quasi-Maximum Likelihood estimation

Example: Simulated a sample of 1,000 values from a SV model with parameters $\gamma = 0$, so that $\mu_h = \gamma / (1 - \phi) = 0$, $\phi = 0.9$ and $\sigma_\eta^2 = 0.1$. The size of the sample is typical for financial data, as are the chosen parameter values.

A plot of the likelihood function over a range of values of ϕ and σ_η^2 shows that it is rather flat. For this reason and to avoid convergence difficulties usually encountered with some of the numerical optimization procedures, we use stochastic optimization (*simulated annealing*) algorithm to find an approximate maximum of the quasi-likelihood function.

6.6 Quasi-Maximum Likelihood estimation

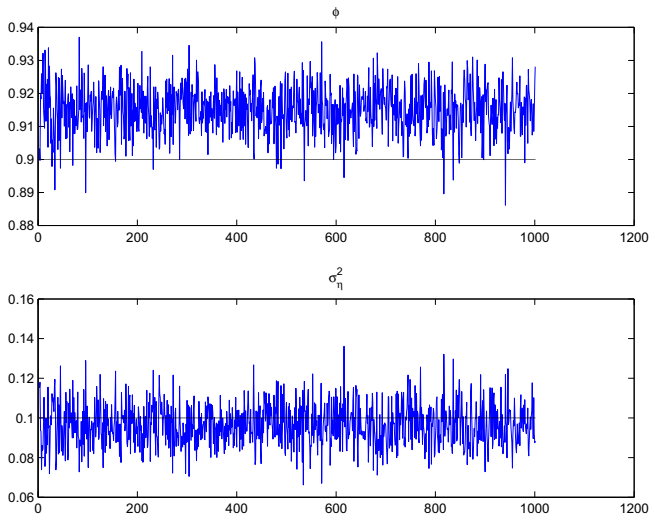
Simulated annealing works with a powered version of the (quasi-)likelihood

$$\mathcal{L}_M(\mu_h, \phi, \sigma_\eta^2) = \mathcal{L}(\mu_h, \phi, \sigma_\eta^2)^{1/M}.$$

Starting with M large (so that the function is $\mathcal{L}_M$ is relatively flat), values of the parameters are proposed, and accepted if they lead to an increase in the function. Slowly, M is decreased, and the samples become concentrated near the mode of $\mathcal{L}$.

The following plots display the QML estimates found using simulated annealing of ϕ and σ_η^2 once the algorithm has converged, while the straight horizontal lines indicate the true value of the parameters.

6.6 Quasi-Maximum Likelihood estimation



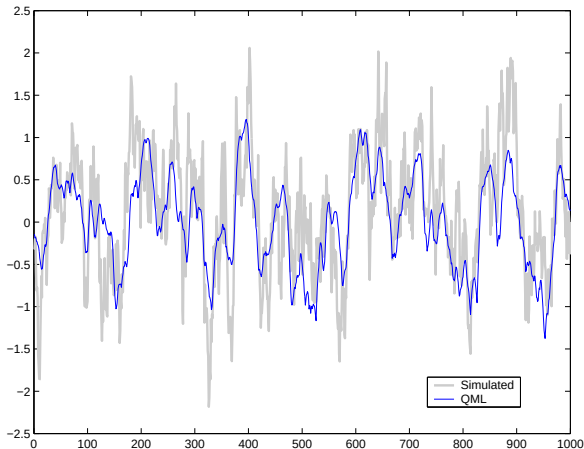
6.6 Quasi-Maximum Likelihood estimation

The QML estimate of ϕ exhibits a positive bias while the corresponding estimate of σ_η^2 is quite accurate. In fact, the mean of the sample obtained for ϕ is equal to 0.9162 while the corresponding sample mean for σ_η^2 is equal to 0.0967, an estimate which does not exhibit any significant bias.

The estimate for μ_h is obtained, as noted above, by adding 1.2704 to the sample mean of the log-squared observations, which results in a value equal to 0.1133. The estimate for γ is obtained by the transformation $\gamma = (1 - \phi)\mu_h$ and turns out to be equal to 0.0095, which is quite accurate.

Volatility Estimates

Simulated underlying volatility process (thick grey line) and estimated smoothed volatilities via the QML method (thin black line).



A drawback to the QML procedure is the so-called *inlier* problems encountered by taking logarithms of very small numbers. In particular, when the asset returns, y_t , are close to zero $\log y_t^2$ is a large negative number and in the extreme case where $y_t = 0$, $\log y_t^2$ is not defined.

Instead of transforming to $w_t = \log y_t^2$, it is possible to work with the series

$$\omega_t = \log (y_t^2 + \delta s^2) - \frac{\delta s^2}{y_t^2 + \delta s^2},$$

where s^2 is the sample variance of y_t and δ is a small user-specified constant.

6.7 Bayesian Inference

After we have collected a set of data, $\mathbf{Y}$, which are assumed to have come from a density $p(\cdot|\boldsymbol{\theta})$, we can investigate the distribution of the parameters $\boldsymbol{\theta}$ given $\mathbf{Y}$ using Bayes Theorem. In essence, given the data, we update our degree of belief about $\boldsymbol{\theta}$ and obtain a posterior distribution of the parameters, which is denoted $p(\boldsymbol{\theta}|\mathbf{Y})$ and is given by

$$p(\boldsymbol{\theta}|\mathbf{Y}) = \frac{p(\mathbf{Y}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta})}{\int_{\Theta} p(\mathbf{Y}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}} \propto p(\mathbf{Y}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}),$$

In the SV case, we will explore the posterior distribution using Markov chain Monte Carlo (MCMC).

6.7 Bayesian Inference

The estimation of the SV model via MCMC considers the hierarchical structure of conditional distributions.

Let

- ▶ $\boldsymbol{\theta} = (\gamma, \phi, \sigma_\eta^2)^\top$ denote the vector of hyperparameters,
- ▶ $\mathbf{h}_{1:n} = (h_1, \dots, h_n)^\top$ denote the vector of log-volatilities
- ▶ $\mathbf{y}_{1:n} = (y_1, \dots, y_n)^\top$ the vector of observations,

then the hierarchy is specified by the sequence of three conditional distributions.

6.7 Bayesian Inference

- ▶ the distribution of the observations conditional on the log-volatilities, $p(y_{1:n}|h_{1:n})$,
- ▶ the distribution of the log-volatilities conditional on the hyperparameters, $p(h_{1:n}|\theta)$
- ▶ the prior distribution of the hyperparameters, $p(\theta)$.

The joint posterior distribution of the log-volatilities and hyperparameters is

$$p(h_{1:n}, \theta | y_{1:n}) \propto p(y_{1:n} | h_{1:n}) p(h_{1:n} | \theta) p(\theta).$$

Gibbs sampler for the SV model

1. Choose arbitrary starting values $h_{1:n}^{(0)}$, $\theta^{(0)}$ and let $i = 0$.
2. Sample $h_{1:n}^{(i+1)} \sim p(h_{1:n} | y_{1:n}, \theta^{(i)})$.
3. Sample $\theta^{(i+1)} \sim p(\theta | y_{1:n}, h_{1:n}^{(i+1)})$.
4. Set $i = i + 1$ and goto 1.

Gibbs sampler for the SV model

Step (2) of the Gibbs algorithm is relatively simple to implement, but sampling from

$$p\left(h_{1:n}|y_{1:n}, \boldsymbol{\theta}^{(i)}\right)$$

is not that straightforward.

Single-move algorithms circumvent this difficult part of the procedure by decomposing further the density $p\left(h_{1:n}|y_{1:n}, \boldsymbol{\theta}^{(i)}\right)$ into the conditionals

$$p\left(h_t|h_{1:n\setminus t}^{(i)}, y_{1:n}, \boldsymbol{\theta}^{(i)}\right)$$

where

$$h_{1:n\setminus t}^{(i)} = \left(h_1^{(i+1)}, \dots, h_{t-1}^{(i+1)}, h_{t+1}^{(i)}, \dots, h_n^{(i)}\right).$$

Gibbs sampler for the SV model

Step 2 of the Gibbs Sampler algorithm becomes:

2a. For $t = 1, \dots, n$, sample

$$h_t^{(i+1)} \sim p \left(h_t | h_{1:n \setminus t}^{(i)}, y_{1:n}, \boldsymbol{\theta}^{(i)} \right)$$

The common feature of all single-move algorithms is that they exploit the Markovian structure of the log-volatilities process;

$$\begin{aligned} p(h_t | h_{1:n \setminus t}, y_{1:n}, \boldsymbol{\theta}) &= p(h_t | h_{t-1}, h_{t+1}, y_t, \boldsymbol{\theta}) \\ &\propto p(y_t | h_t) p(h_{t+1} | h_t, \boldsymbol{\theta}) p(h_t | h_{t-1}, \boldsymbol{\theta}), \end{aligned}$$

where the second line is deduced from Bayes theorem.

Rejection Metropolis-Hastings

Consider the parameterization of the SV model:

$$\begin{aligned}y_t &= \sqrt{h_t}Z_t, \\ \log h_t &= \gamma + \phi \log h_{t-1} + U_t, \quad t = 1, \dots, n,\end{aligned}$$

where Z_t and U_t are contemporaneously and serially independent random variables with distributions $\mathcal{N}(0, 1)$ and $\mathcal{N}(0, \sigma_\eta^2)$, respectively.

Gibbs sampler for the SV model

In the standard model, with $|\phi| < 1$, the logarithm of the latent volatilities follows a stationary, Gaussian AR(1) process, so that

$$\log h_t | h_{t-1}, \boldsymbol{\theta} \sim \mathcal{N}(\gamma + \phi \log h_{t-1}, \sigma_\eta^2),$$

which implies that $h_t | h_{t-1}, \boldsymbol{\theta}$ has a log-normal distribution $\mathcal{LN}(\gamma + \phi \log h_{t-1}, \sigma_\eta^2)$ and in particular,

$$p(h_t | h_{t-1}, \boldsymbol{\theta}) \propto \frac{1}{h_t} \exp \left\{ -\frac{(\log h_t - \gamma - \phi \log h_{t-1})^2}{2\sigma_\eta^2} \right\}.$$

Gibbs sampler for the SV model

In addition, noting that $y_t|h_t \sim \mathcal{N}(0, h_t)$, it follows that

$$p(h_t|h_{1:n \setminus t}, y_{1:n}, \boldsymbol{\theta}) \propto \frac{1}{h_t^{1/2}} \exp \left\{ -\frac{y_t^2}{2h_t} \right\} \\ \times \frac{1}{h_t} \exp \left\{ -\frac{(l_{t+1} - \gamma - \phi l_t)^2 + (l_t - \gamma - \phi l_{t-1})^2}{2\sigma_\eta^2} \right\}$$

where $l_t = \log h_t$.

Gibbs sampler for the SV model

After some algebra it follows that

$$p(h_t | h_{1:n \setminus t}, y_{1:n}, \boldsymbol{\theta}) \propto \frac{1}{h_t^{1/2}} \exp \left\{ -\frac{y_t^2}{2h_t} \right\} \times \frac{1}{h_t} \exp \left\{ -\frac{(l_t - m_t)^2}{2\sigma_*^2} \right\} = f(h_t),$$

where

$$m_t = \frac{\gamma(1 - \phi) + \phi(l_{t+1} + l_{t-1})}{(1 + \phi^2)} \quad \text{and} \quad \sigma_*^2 = \frac{\sigma_\eta^2}{1 + \phi^2}.$$

This conditional cannot be sampled directly, but can be sampled using *rejection sampling*.

Gibbs sampler for the SV model

The idea is to use rejection sampling within an independence Metropolis-Hastings algorithm; this approach is based on a density g and a constant c such that

$$p(x|h_{1:n\setminus t}, y_{1:n}, \boldsymbol{\theta}) \leq cg(x)$$

for most, but not necessarily all, x , so that g is a pseudo-dominating density.

For each time t , proposals h'_t are generated from the density g , until one of this proposals is accepted with probability

$$\min \left\{ 1, \frac{p(h'_t|h_{1:n\setminus t}^{(i)}, y_{1:n}, \boldsymbol{\theta})}{cg(h'_t)} \right\}.$$

Gibbs sampler for the SV model

The accepted h'_t then enters as a proposed value in an independence Metropolis-Hastings accept-reject step and we set $h_t^{(i+1)} = h'_t$ with probability

$$\min \left[1, \frac{p \left(h'_t | h_{1:n \setminus t}^{(i)}, y_{1:n}, \boldsymbol{\theta} \right) \min \left\{ p \left(h_t^{(i)} | h_{1:n \setminus t}^{(i)}, y_{1:n}, \boldsymbol{\theta} \right), c g \left(h_t^{(i)} \right) \right\}}{p \left(h_t^{(i)} | h_{1:n \setminus t}^{(i)}, y_{1:n}, \boldsymbol{\theta} \right) \min \left\{ p \left(h'_t | h_{1:n \setminus t}^{(i)}, y_{1:n}, \boldsymbol{\theta} \right), c g \left(h'_t \right) \right\}} \right].$$

If h'_t is not accepted at this second stage, then we set $h_t^{(i+1)} = h_t^{(i)}$.

The problem is reduced to finding the density g and the constant c .

Gibbs sampler for the SV model

The posterior density of h_t in can be seen as the product of two densities,

- an improper inverse-gamma density, $IGamma(-0.5, 0.5y_t^2)$,
- a log-normal density.

The log-normal part can be approximated by an inverse-gamma density, $IG(\alpha, \beta_t)$, with the same mean and variance; this approximation amounts to setting

$$\alpha = \frac{1 - 2 \exp(\sigma_*^2)}{1 - \exp(\sigma_*^2)} \quad \text{and} \quad \beta_t = (\alpha - 1) \exp(m_t + 0.5\sigma_*^2).$$

Gibbs sampler for the SV model

The product of the two inverse-gamma densities, the improper one and the approximating one, is also the density of an $IG(\nu, \lambda_t)$ random variable, with $\nu = \alpha + 0.5$ and $\lambda_t = \beta_t + 0.5y_t^2$.

The pseudo-dominating density g is given by

$$g(x) \propto x^{-(\alpha+0.5+1)} \exp \left\{ -\frac{\beta_t + 0.5y_t^2}{x} \right\}.$$

Gibbs sampler for the SV model

This method has been regarded as inefficient, as the posterior density $p(h_t|h_{1:n\setminus t}, y_{1:n}, \theta)$ can be written as

$$p(h_t|h_{1:n\setminus t}, y_{1:n}, \theta) \propto p(y_t|h_t) p(h_t|h_{t-1}, h_{t+1}, \theta).$$

In essence, the initial method attempts to approximate the density

$$p(h_t|h_{t-1}, h_{t+1}, \theta) \quad \text{by} \quad p(y_t|h_t).$$

However, in practice, the prior density

$$p(h_t|h_{t-1}, h_{t+1}, \theta)$$

dominates the likelihood, and therefore one should focus in approximating $p(h_t|h_{1:n\setminus t}, y_{1:n}, \theta)$ by a density of the same form as $p(h_t|h_{t-1}, h_{t+1}, \theta)$.

Gibbs sampler for the SV model

Consider the following parameterization of the SV model:

$$\begin{aligned}y_t &= \exp(h_t/2)Z_t, \\h_t &= \mu_h + \phi(h_{t-1} - \mu_h) + U_t, \quad t = 1, \dots, n,\end{aligned}$$

where Z_t and U_t are as before. The log-volatilities have been centered around their unconditional mean and

$$h_t|h_{t-1}, \boldsymbol{\theta} \sim \mathcal{N}(\mu_h + \phi(h_{t-1} - \mu_h), \sigma_\eta^2)$$

while $y_t|h_t \sim \mathcal{N}(0, \exp(h_t))$. Then $h_t|h_{t-1}, h_{t+1}, \boldsymbol{\theta} \sim \mathcal{N}(m_t, \sigma_*^2)$, where

$$m_t = \mu_h + \frac{\phi[(h_{t+1} - \mu_h) + (h_{t-1} - \mu_h)]}{1 + \phi^2} \quad \text{and} \quad \sigma_*^2 = \frac{\sigma_\eta^2}{1 + \phi^2}.$$

Gibbs sampler for the SV model

Thus $\log p(h_t | h_{1:n \setminus t}, y_{1:n}, \theta)$ is given by

$$\begin{aligned}\log p(h_t | -) &= \text{const} + \log p(h_t | h_{t-1}, h_{t+1}, \theta) + \log p(y_t | h_t) \\ &= \text{const} - \frac{(h_t - m_t)^2}{2\sigma_*^2} - \frac{h_t}{2} - \frac{y_t^2}{2} \exp(-h_t) \\ &\simeq \text{const} - \frac{(h_t - m_t)^2}{2\sigma_*^2} - \frac{h_t}{2} \\ &\quad - \frac{y_t^2}{2} \exp(-m_t) \left[1 - (h_t - m_t) + \frac{(h_t - m_t)^2}{2} \right] \\ &= \log g(h_t),\end{aligned}$$

say, where the third line comes from a second order Taylor expansion of $\exp(-h_t)$ about m_t .

Gibbs sampler for the SV model

In addition, the terms in $\log g$ can be combined and after some algebra, it is easily seen that

$$g(h_t) \propto f_N(h_t | \mu_t, \sigma_t^2)$$

where $f_N(x | \alpha, \beta^2)$ denotes the density of a Gaussian random variable with mean α and variance β^2 , and where

$$\begin{aligned}\mu_t &= \frac{\sigma_t^2}{\sigma_*^2} m_t + \frac{\sigma_t^2}{2} [y_t^2 \exp(-m_t) (1 + m_t) - 1] \\ \sigma_t^2 &= \left[\frac{1}{\sigma_*^2} + \frac{y_t^2}{2} \exp(-m_t) \right]^{-1}.\end{aligned}$$

Gibbs sampler for the SV model

Rejection Sampling : The function $\exp(-h_t)$ is convex and hence, by considering the tangent line of $\exp(-h_t)$ at the point m_t , where m_t is as before, it follows that

$$\exp(-h_t) \geq \exp(-m_t)(1 + m_t - h_t).$$

Using the latter result

$$\begin{aligned} \log p(h_t | h_{1:n \setminus t}, y_{1:n}, \theta) &\leq \text{const} - \frac{(h_t - m_t)^2}{2\sigma_*^2} - \frac{h_t}{2} \\ &\quad - \frac{y_t^2}{2} \exp(-m_t)(1 + m_t - h_t) \\ &= \log g(h_t). \end{aligned}$$

Gibbs sampler for the SV model

The terms in $\log g$ can be combined and thus

$$g(h_t) \propto f_N(h_t | \mu_t, \sigma_*^2)$$

where σ_*^2 is as previously, and μ_t is

$$\mu_t = m_t + \frac{\sigma_*^2}{2} [y_t^2 \exp(-m_t) - 1].$$

In this case the density $f_N(\cdot | \mu_t, \sigma_*^2)$ does bound $p(\cdot | h_{1:n \setminus t}, y_{1:n}, \boldsymbol{\theta})$ and so rejection sampling can be used.

Gibbs sampler for the SV model

In order to update the log-volatilities, for each time t , proposals h'_t are generated from a $\mathcal{N}(\mu_t, \sigma_*^2)$ density until one of these is accepted with probability

$$\min \left\{ 1, \frac{p(h'_t | h_{1:n \setminus t}^{(i)}, y_{1:n}, \boldsymbol{\theta})}{f_N(h'_t | \mu_t, \sigma_*^2)} \right\}$$

we then set $h_t^{(i+1)} = h'_t$ and proceed to update $h_{t+1}^{(i)}$ in the same manner.

Gibbs sampler for the SV model

Sampling θ : The Gibbs sampler algorithm is split into three parts:

3a Sample $\sigma_{\eta}^{2(i+1)} \sim p\left(\sigma_{\eta}^2 | h_{1:n}^{(i+1)}, \mu_h^{(i)}, \phi^{(i)}\right).$

3b Sample $\phi^{(i+1)} \sim p\left(\phi | h_{1:n}^{(i+1)}, \mu_h^{(i)}, \sigma_{\eta}^{2(i+1)}\right),$ under the restriction $|\phi| < 1$

3c Sample $\mu_h^{(i+1)} \sim p\left(\mu_h | h_{1:n}^{(i+1)}, \phi^{(i+1)}, \sigma_{\eta}^{2(i+1)}\right).$

Gibbs sampler for the SV model

The likelihood function of the log-volatilities can be expressed as

$$\begin{aligned} p(h_{1:n}|\boldsymbol{\theta}) &= p(h_1|\boldsymbol{\theta}) \prod_{t=1}^{n-1} p(h_{t+1}|h_t, \boldsymbol{\theta}) \\ &= \left(\frac{1}{2\pi\sigma_\eta^2} \right)^{\frac{n}{2}} (1 - \phi^2)^{\frac{1}{2}} \exp \left\{ -\frac{SSQ_h}{2\sigma_\eta^2} \right\}. \end{aligned}$$

where

$$SSQ_h = (h_1 - \mu_h)^2 (1 - \phi^2) + \sum_{t=1}^{n-1} [(h_{t+1} - \phi h_t) - (1 - \phi) \mu_h]^2$$

Gibbs sampler for the SV model

Sampling σ_η^2 .

Under an inverse-gamma conjugate prior, $\sigma_\eta^2 \sim IG(\alpha_\sigma/2, \beta_\sigma/2)$ the posterior for the parameter σ_η^2 takes the form

$$p(\sigma_\eta^2 | h_{1:n}, \mu_h, \phi) \propto (\sigma_\eta^2)^{-[1+(\alpha_\sigma+n)/2]} \exp\left\{-\frac{\beta_\sigma + SSQ_h}{2\sigma_\eta^2}\right\}.$$

Therefore, in the Gibbs sampler for the SV model, the variance σ_η^2 can be directly updated by drawing from an $IG(\nu_\sigma, \lambda_\sigma)$, where

$$\nu_\sigma = (\alpha_\sigma + n)/2 \quad \lambda_\sigma = (\beta_\sigma + SSQ_h)/2.$$

Setting $\alpha_\sigma = 5$ and $\beta_\sigma = 0.05$ yields prior mean of approximately 0.0167 .

Gibbs sampler for the SV model

Sampling ϕ .

To respect the restriction of ϕ to the stationary region, a transformed Beta prior is used

$$(\phi + 1) / 2 \sim \text{Beta}(\alpha_\phi, \beta_\phi)$$

The full conditional posterior for ϕ is

$$p(\phi | h_{1:n}, \mu_h, \sigma_\eta^2) \propto p(\phi) (1 - \phi^2)^{1/2} \exp \left\{ -\frac{(h_1 - \mu_h)^2 (1 - \phi^2)}{2\sigma_\eta^2} \right\} \\ \times \exp \left\{ -\frac{\sum_{t=1}^{n-1} [(h_{t+1} - \mu_h) - \phi (h_t - \mu)]^2}{2\sigma_\eta^2} \right\}.$$

Gibbs sampler for the SV model

This density is log-concave in ϕ and hence, the *adaptive rejection sampling procedure* can be used. A simpler approach is to use the independence M-H algorithm.

Expanding terms in the second exponent yields that the full conditional is proportional to $f_N(\cdot | \mu_\phi, \sigma_\phi^2)$, the density of a Normal random variable with mean and variance respectively given by

$$\mu_\phi = \frac{\sum_{t=1}^{n-1} (h_{t+1} - \mu_h) (h_t - \mu_h)}{\sum_{t=1}^{n-1} (h_t - \mu_h)^2} \quad \text{and} \quad \sigma_\phi^2 = \frac{\sigma_\eta^2}{\sum_{t=1}^{n-1} (h_t - \mu_h)^2}.$$

Gibbs sampler for the SV model

The posterior density of ϕ is:

$$p(\phi|h_{1:n}, \mu_h, \sigma_\eta^2) \propto p(\phi) (1 - \phi^2)^{1/2} \exp \left\{ -\frac{(h_1 - \mu_h)^2 (1 - \phi^2)}{2\sigma_\eta^2} \right\} \\ \times f_N(\phi|\mu_\phi, \sigma_\phi^2).$$

Within a M-H setting, we choose an initial value $|\phi^{(0)}| < 1$. We sample draws $\phi_* \sim f_N$ and then provided $|\phi_*| < 1$, the proposed ϕ_* enters a MH accept-reject step.

Typical choices for the parameters α_ϕ and β_ϕ of the prior density are $\alpha_\phi = 20$ and $\beta_\phi = 3/2$ so that the prior mean for ϕ is $-1 + 2\alpha_\phi / (\alpha_\phi + \beta_\phi) \simeq 0.8605$.

Gibbs sampler for the SV model

Sampling μ_h .

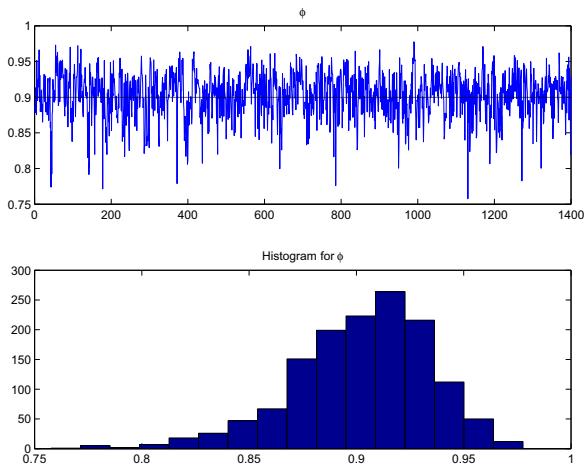
Under a non-informative prior distribution for μ_h , $p(\mu_h) \propto 1$, it is easily seen that the posterior density satisfies

$$p(\mu_h | h_{1:n}, \phi, \sigma_\eta^2) \propto \exp \left\{ -\frac{SSQ_h}{2\sigma_\eta^2} \right\}.$$

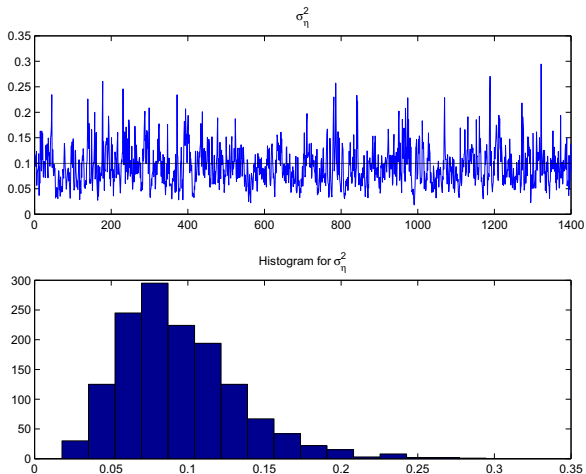
After some algebra, it follows that the parameter μ_h can also be directly updated in the Gibbs sampler, $\mu_h \sim \mathcal{N}(m_h, s_h^2)$, where

$$m_h = \frac{(1 + \phi) h_1 + \sum_{t=1}^{n-1} (h_{t+1} - \phi h_t)}{(n-1)(1-\phi) + (1+\phi)}$$
$$s_h^2 = \frac{\sigma_\eta^2}{\left[(n-1)(1-\phi)^2 + (1-\phi^2) \right]}.$$

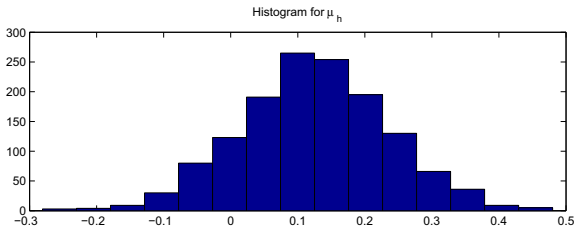
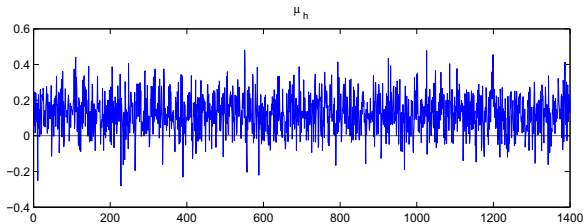
Gibbs sampler for the SV model



Gibbs sampler for the SV model



Gibbs sampler for the SV model



Gibbs sampler for the SV model

Notes:

- ▶ it is also possible to obtain multi-step-ahead forecasts of future log-volatilities. Finally, with some further (non-trivial) modifications the procedure can also be adapted to handle fat-tailed distributions and correlated errors, as well as multivariate SV models.
- ▶ the procedure is not well suited to the problem of obtaining smoothed estimates of the log-volatilities as more data become available.

Gibbs sampler for the SV model

- if the persistence in the volatility is high, i.e. $\phi \simeq 1$, and σ_η^2 is small, the log-volatility process is highly correlated. As a result, the draws $h_t^{(i)}$ sampled from the full conditional densities exhibit high levels of correlation for long lags and consequently, there is little movement in the chain.

Thus, the single-move MCMC procedure converges slowly in this case. In order to overcome this problem, another approach has been proposed in the literature, which attempts to sample the whole vector of the log-volatilities with a single draw from the density $p(h_{1:n}|y_{1:n}, \theta)$.

Multi-state move MCMC Samplers

The *multi-state move* approach begins by write the SV model as

$$\begin{aligned}w_t &= \log y_t^2 = h_t + z_t, \\h_t &= \mu_h + \phi(h_{t-1} - \mu_h) + U_t, \quad t = 1, \dots, n.\end{aligned}$$

Instead of approximating the $\log \chi^2$ distribution of $z_t = \log Z_t^2$ with a $\mathcal{N}(-1.2704, \pi^2/2)$, which is suboptimal and results in poor sample properties, a better choice is to approximate the $\log \chi^2$ distribution by a mixture of J normals,

Multi-state move MCMC Samplers

$$f_{z_t}(z_t) \simeq \sum_{j=1}^7 q_j f_{z_t|s_t=j}(z_t|s_t=j),$$

where

$$z_t|s_t=j \sim \mathcal{N}(m_j - 1.2704, v_j^2)$$

s_t is the indicator variable at time t , and the $q_j, j = 1, \dots, J$ are the weights attached to each component, so that $p(s_t = j) = q_j$ and $\sum_{j=1}^J q_j = 1$.

The elements $\{q_j, m_j, v_j^2, j = 1, \dots, J\}$ are selected by matching the first four moments of the $\log \chi^2$ distribution to the approximating mixture of normals.

Multi-state move MCMC Samplers

In addition, given the indicator variable at each time t , the state-space model is Gaussian, that is

$$w_t | s_t = j, h_t \sim \mathcal{N} \left(h_t + m_j - 1.2704, v_j^2 \right).$$

Given the indicator variables, the model can be thought of as a time-inhomogeneous, Gaussian state-space model, a standard algorithm can be used to sample $h_1, \dots, h_n$ at once from $p(h_{1:n} | \mathbf{w}, s_{1:n})$, where $\mathbf{w} = (\log y_1^2, \dots, \log y_n^2)$ and $s_{1:n} = (s_1, \dots, s_n)$.

This approach augments the vector of parameters and log-volatilities, $(\boldsymbol{\theta}, \mathbf{h})$, with the vector $s_{1:n}$ of indicator variables and construct a Gibbs sampler which will produce samples from the density $p(\boldsymbol{\theta}, h_{1:n}, s_{1:n} | \mathbf{w})$.

Multi-state move MCMC Samplers

The general form of the multi-state move Gibbs sampler for the SV model is as follows:

1. Choose arbitrary starting values $\boldsymbol{\theta}^{(0)}, h_{1:n}^{(0)}, s_{1:n}^{(0)}$ and let $i = 0$.
2. Sample $h_{1:n}^{(i+1)} \sim p\left(\mathbf{h}|\mathbf{w}, s_{1:n}^{(i)}, \boldsymbol{\theta}^{(i)}\right)$.
3. Sample $s_{1:n}^{(i+1)} \sim p\left(s_{1:n}|\mathbf{w}, h_{1:n}^{(i+1)}\right)$.
4. Sample $\boldsymbol{\theta}^{(i+1)} \sim p\left(\boldsymbol{\theta}|h_{1:n}^{(i+1)}\right)$.
5. Set $i = i + 1$ and go to 1.

Multi-state move MCMC Samplers

Step 1. is best conducted via the *simulation smoother*. To sample the vector $s_{1:n}$ from its full conditional density, we sample each s_t independently using rejection sampling;

$$p(s_{1:n}|\mathbf{w}, h_{1:n}) = \prod_{t=1}^n p(s_t|w_t, h_t) \propto \prod_{t=1}^n p(w_t|s_t, h_t) p(s_t)$$
$$p(s_t = j|w_t, h_t) \propto q_j f_N(w_t|h_t + m_j - 1.2704, v_j^2), \quad j = 1, \dots, K,$$

where $f_N(\cdot|a, b^2)$ denotes the density of a normal random variable with mean a and variance b^2 .

Multi-state move MCMC Samplers

Therefore, for each time t in the sample, we can draw a proposed value $s_t^* = j$, with probability q_j , which is then accepted with probability

$$f_N \left(w_t | h_t^{(i+1)} + m_j - 1.2704, v_j^2 \right).$$

Finally, to update the parameters θ , exactly the same approach as with the single-move Gibbs sampler can be used.

This procedure greatly reduces the correlation between successive draws. The method has also been extended to handle fat-tailed observation errors and leverage effect. However, this method suffers from the inlier problem discussed above.

Multi-state move MCMC Samplers

Another multi-state move sampler for general non-Gaussian time series models has also been proposed, where the main idea is, on any given iteration, to **fix** a given number of randomly chosen log-volatilities fixed and update the rest conditional on the observations and the fixed volatilities.

- ▶ if h_{t-1} and h_{t+k+1} are two elements of the log-volatilities vector that have been chosen to remain fixed, then the block

$$(h_t, \dots, h_{t+k})$$

is updated by sampling from

$$p(h_t, \dots, h_{t+k} | y_t, \dots, y_{t+k}, h_{t-1}, h_{t+k+1}, \boldsymbol{\theta})$$