

Chapter 7

Financial Time Series Models

7.1 Volatility Modelling

One form of non-stationarity not yet examined is the relaxation of the constant variance assumption.

We use the general term **volatility** to capture the concept of time-varying variance of a process. This could be considered to arise in

- the **unconditional** or marginal model of the observation process
- the **conditional** model for the process (eg in a state space setting).

7.1 Volatility Modelling

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

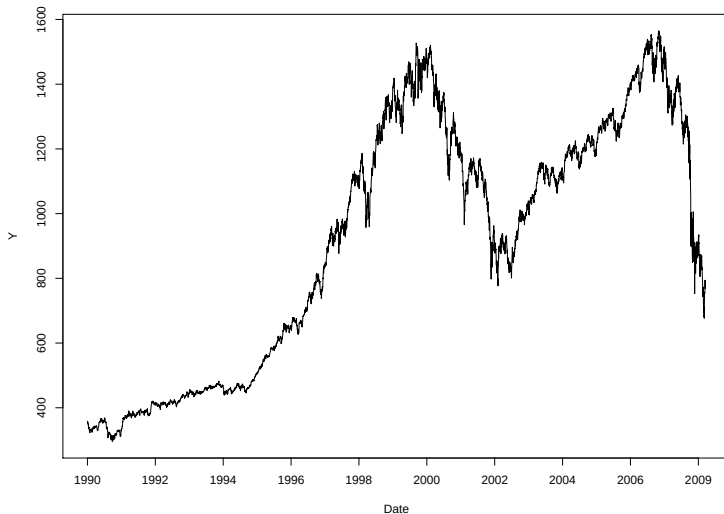
In particular, financial returns demonstrate

volatility clustering

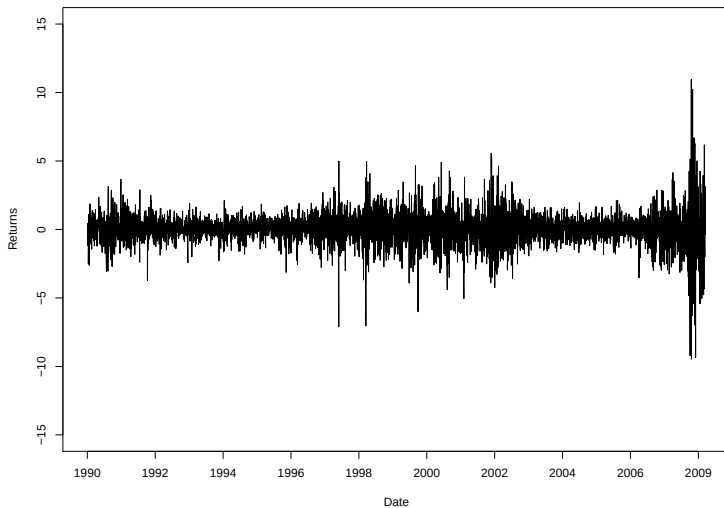
that is, large changes tend to be followed by large changes and vice versa.

Models need to be developed to capture this behaviour.

S & P 500 : 1990-2009



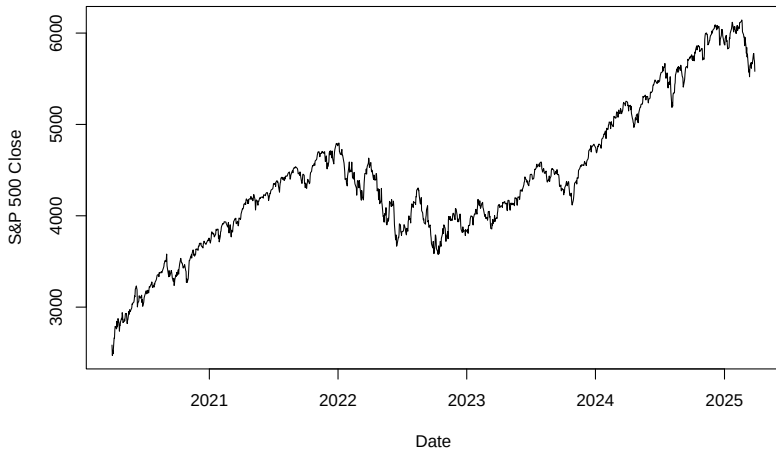
S & P 500 : 1990-2009 log Returns



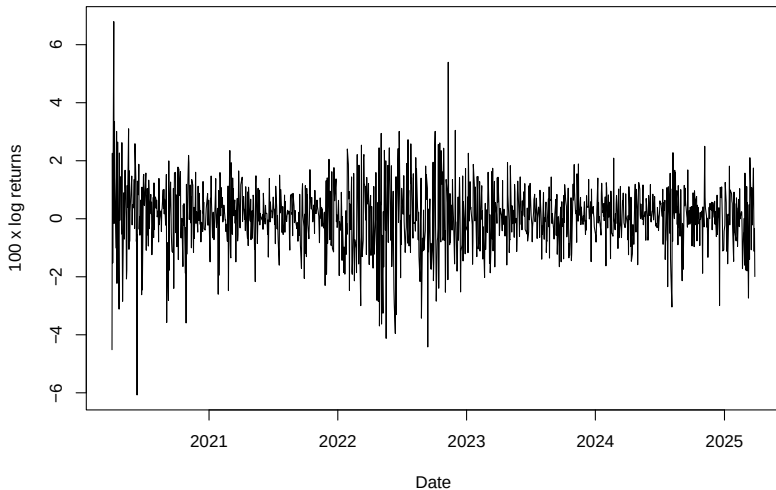
S & P 500 : 2020-2025

```
my.url<- 'https://www.math.mcgill.ca/dstephens/545/Data/HistoricalData_1743471926857.csv'
sp25<-read.csv(my.url,header=T)
dvec<-rev(as.character(sp25[,1]))
date.vec<-as.Date(dvec,format="%m/%d/%Y")
Yt<-rev(sp25[,2])
Yt.ret<-diff(log(Yt))
par(mar=c(4,4,2,2))
plot(date.vec,Yt,type='l',xlab='Date',ylab='S&P 500 Close')
```

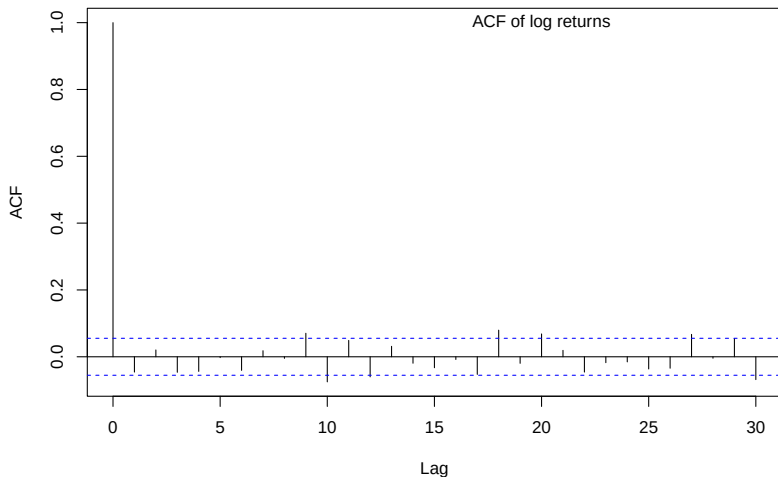
S & P 500 : 2020-2025



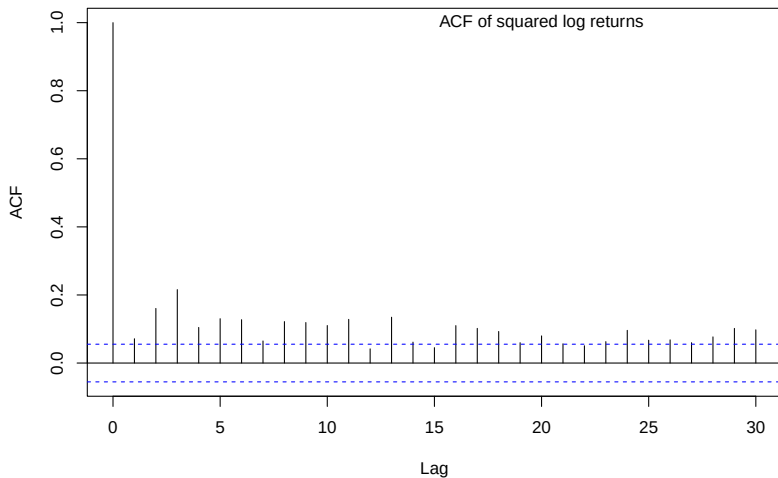
S & P 500 : 2020-2025



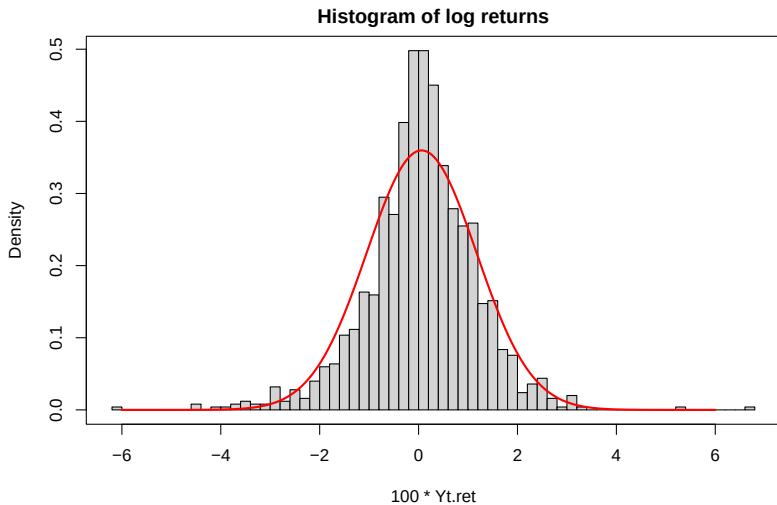
S & P 500 : 2020-2025



S & P 500 : 2020-2025



S & P 500 : 2020-2025



7.2 Types of Volatility Models

Volatility 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.

These models have the general form of state-space models.

7.2 Types of Volatility Models

In a **parameter-driven** model

$$p(y_t | \mathbf{x}_t, \mathbf{x}_{1:(t-1)}, y_{1:(t-1)}) \equiv p(y_t | \mathbf{x}_t)$$

$$p(\mathbf{x}_{t+1} | \mathbf{x}_t, \mathbf{x}_{1:(t-1)}, y_{1:t}) \equiv p(\mathbf{x}_{t+1} | \mathbf{x}_t)$$

so that

$$p(\mathbf{x}_{2:n}, y_{1:n} | \mathbf{x}_1) = \left\{ \prod_{t=1}^n p(y_t | \mathbf{x}_t) \right\} \left\{ \prod_{t=2}^n p(\mathbf{x}_t | \mathbf{x}_{t-1}) \right\}$$

7.2 Types of Volatility Models

In an **observation-driven** model

$$p(y_t | \mathbf{x}_t, \mathbf{x}_{1:(t-1)}, y_{1:(t-1)}) \equiv p(y_t | \mathbf{x}_t)$$

$$p(\mathbf{x}_{t+1} | \mathbf{x}_t, \mathbf{x}_{1:(t-1)}, y_{1:t}) \equiv p(\mathbf{x}_{t+1} | y_{1:t})$$

so that

$$p(\mathbf{x}_{1:n}, y_{1:n} | \mathbf{x}_1) = \prod_{t=1}^n p(y_t | \mathbf{x}_t) p(\mathbf{x}_t | y_{1:(t-1)})$$

where $p(\mathbf{x}_1 | y_0) \equiv p(\mathbf{x}_1)$, say.

7.2 Types of Volatility Models

Example: Model for Count Data

Suppose that for $t = 2, \dots, n$

$$p(y_t | \mathbf{x}_t) \equiv \text{Poisson}(\xi \mathbf{x}_t)$$

$$p(\mathbf{x}_{t+1} | \mathbf{x}_t) = \begin{cases} 1 & \text{if } \mathbf{x}_{t+1} = a\mathbf{x}_t \\ 0 & \text{otherwise} \end{cases}$$

and that where $p(\mathbf{x}_1) \equiv \text{Gamma}(\alpha, \lambda)$, for some constants ξ, a, α, λ .

7.2 Types of Volatility Models

Example: Model for Count Data

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

$$p(\mathbf{x}_t | \mathbf{y}_{1:t}) \equiv \text{Gamma}(\alpha_t, \lambda_t)$$

$$p(\mathbf{x}_{t+1} | \mathbf{y}_{1:t}) \equiv \text{Gamma}(\alpha_t, \lambda_t/a)$$

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

where

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

7.2 Types of Volatility Models

Example: Model for Count Data

Then

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

and

$$\text{Var}[Y_{t+1} | Y_{1:t} = y_{1:t}] = \frac{\xi \alpha_t \lambda_t}{(\lambda_{t+1} - \xi)^2}$$

7.2 Types of Volatility Models

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

Suppose that

$$p(\mathbf{x}_{t+1} | \mathbf{y}_{1:t}) = \begin{cases} 1 & \text{if } \mathbf{x}_{t+1} = \mathbf{y}_t \\ 0 & \text{otherwise} \end{cases}$$

$$p(\mathbf{y}_t | \mathbf{x}_t) \equiv f_Z(\mathbf{y}_t - \phi \mathbf{x}_t)$$

for some pdf f_Z . Then

$$\mathbf{Y}_t - \phi \mathbf{X}_t \stackrel{d}{=} \mathbf{Z}_t \quad \text{or} \quad \mathbf{Y}_t = \phi \mathbf{X}_t + \mathbf{Z}_t$$

so with $\mathbf{X}_t = \mathbf{Y}_{t-1}$ we have an AR(1) model.

7.2 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

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

7.2 Types of Volatility Models

Consider the discrete time process where we replace dY_t by $Y_{t+1} - Y_t$, and dt by Δ , with observation times

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

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

$$\begin{aligned} Y_{t+1} - Y_t &= \log S_{t+1} - \log S_t = \log(S_{t+1}/S_t) \\ &= \mu\Delta + \sigma Z_t \end{aligned}$$

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

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

7.2 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.

7.2 Types of Volatility Models

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

- Autoregressive Conditional Heteroscedasticity (ARCH)
- Generalized ARCH (GARCH).

models.

7.2 Types of Volatility Models

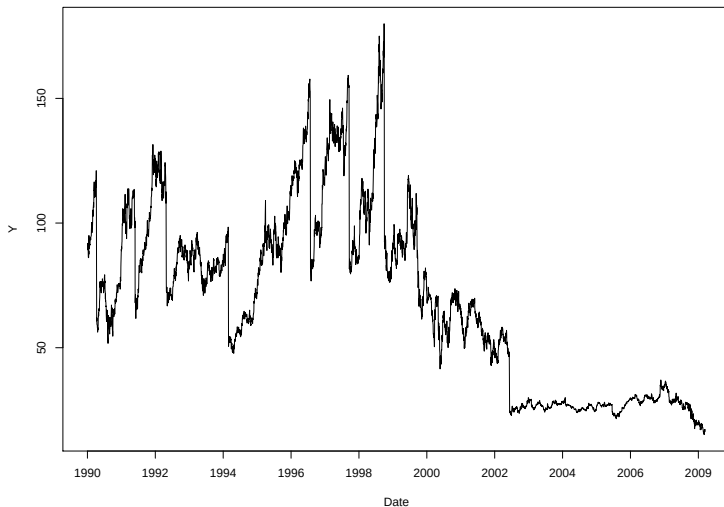
In particular, let y_t be a realization, at time t , of the time series of interest. Typically, y_t is taken to be the compounded return of the underlying asset, so that

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

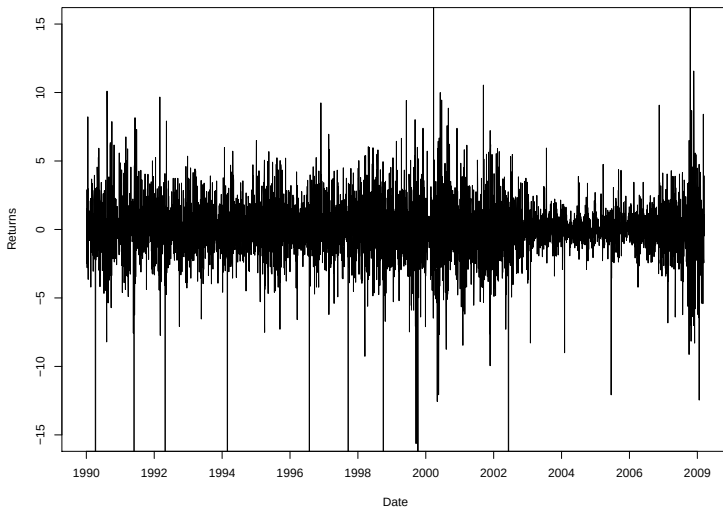
where s_t denotes the observed price of the asset.

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

Microsoft : 1990-2009



Microsoft : 1990-2009 log Returns



7.3 Observation Driven Models

For the observation-driven models,

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

so that

$$Y_t = \sigma_t Z_t,$$

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

7.3 Observation Driven Models

The **ARCH**(p) model allows the **conditional variance** σ_t^2 of Y_t to be a linear combination of previous 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 simplest ARCH model is the ARCH(1)

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

Parameters α_0 and α_1 are non-negative, and by iterated expectation, the process is stationary if and only if $\alpha_1 < 1$, with

$$\text{Var}[Y_t] = \mathbb{E}[Y_t^2] = \frac{\alpha_0}{1 - \alpha_1}.$$

Properties of the ARCH(1) model

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 the **kurtosis** (a measure of distributional heavy tailedness)

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

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

Stationarity in ARCH(1) Variance

The ACF of $\{Y_t^2\}$ is given by

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

where $h = 0, 1, 2, \dots$. Note that

- $\{Y_t\}$ is a white noise process,
- $\{Y_t^2\}$ is an autoregressive process that exhibits volatility clustering controlled by α_1 .
- $\{Y_t^2\}$ is not necessarily covariance stationary; its variance will be finite only if $3\alpha_1^2 < 1$.

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

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

so the log-likelihood is

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

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

$$\text{Var}[Y_t] = \frac{\alpha_0}{(1 - \alpha_1 - \beta_1)}$$

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

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

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.

GARCH models

The moments of the GARCH(1,1) model are:

(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 $\sqrt{1/3}$. 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

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

The ACF declines geometrically at the rate of $\alpha_1 + \beta_1$.

GARCH models

GARCH(p, q) is stationary if the coefficients in the conditional variance equation sum to less than one

$$\alpha_1 + \dots + \alpha_p + \beta_1 + \dots + \beta_q < 1,$$

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.

GARCH models

For the GARCH(1, 1) model: if $\alpha_1 + \beta_1 < 1$, under normality of the residual errors,

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

and $\text{Cov}[Y_t, Y_s] = 0$, but $\text{Cov}[Y_t^2, Y_s^2] \neq 0$.

ARCH/GARCH model extensions

There exist many other versions of the ARCH type models,

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

7.4 Stochastic Volatility Models

The **stochastic volatility** (SV) model is a form of non-linear state space and **parameter-driven** model where the variance of the observations to be an unobserved **random** process.

7.4 Stochastic Volatility Models

The simplest SV model takes the form

$$\begin{aligned}Y_t &= \exp(X_t/2) Z_t \\X_t &= \phi_0 + \phi X_{t-1} + W_t\end{aligned}$$

where Y_t is the observation at time t , $Z_t \sim \mathcal{N}(0, 1)$, and $W_t \sim \mathcal{N}(0, \sigma_W^2)$ are IID series.

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

7.4 Stochastic Volatility Models

The SV model with **leverage** has

$$\begin{aligned}Y_t &= \exp(X_t/2) Z_t \\X_t &= \phi_0 + \phi X_{t-1} + W_t\end{aligned}$$

but now the two processes $\{Z_t\}$ and $\{W_t\}$ are correlated.

We could have that one of

$$\text{Corr}[Z_t, W_t] = \rho \quad \text{Corr}[Z_{t+1}, W_t] = \rho \quad \text{Corr}[Z_t, W_{t+1}] = \rho$$

holds.

7.4 Stochastic Volatility Models

The parameter ϕ captures the **persistence** in the volatility, and the restriction

$$|\phi| < 1$$

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

However, X_t can also be allowed to follow a random walk by setting $\phi = 1$.

Properties of the Stochastic Volatility Model

Suppose $\{Z_t\}$ and $\{W_t\}$ are independent. Under stationarity, the mean and variance of $\{X_t\}$ given respectively by

$$\mu_X = \mathbb{E}[X_t] = \frac{\phi_0}{1 - \phi} \quad \sigma_X^2 = \text{Var}[X_t] = \frac{\sigma_W^2}{1 - \phi^2}.$$

Since

$$Y_t = Z_t \exp(X_t/2)$$

and $\{Z_t\}$ is always stationary, $\{Y_t\}$ will also be stationary if and only if $\{X_t\}$ is stationary. If the distribution of Z_t is symmetric, then all the odd moments of Y_t are zero.

Properties of the Stochastic Volatility Model

By iterated expectation

$$\mathbb{E}[Y_t] = \mathbb{E}[\mathbb{E}[Y_t \mid Y_{1:(t-1)}]] = 0$$

so that Y_t is zero mean and the ACF of Y_t is

$$\gamma_Y(h) = \mathbb{E}[Y_t Y_{t-h}] = \mathbb{E}\left[\exp\left(\frac{X_t}{2} + \frac{X_{t-h}}{2}\right)\right] \mathbb{E}[Z_t Z_{t-h}] = 0$$

for $h \geq 1$, as $\{Z_t\}$ is an uncorrelated process.

Properties of the Stochastic Volatility Model

By assumption, $\exp(X_t)$ has a **log-normal** distributed, and

$$\mathbb{E} \left[\exp(X_t)^j \right] = \exp \left\{ j\mu_X + \frac{1}{2}j^2\sigma_X^2 \right\},$$

so that, if $r = 2s$ is even and $\{X_t\}$ is stationary, all the even moments of Y_t exist and are given by the formula

$$\mathbb{E}[Y_t^r] = \mathbb{E}[(\exp(X_t))^s] \mathbb{E}[Z_t^r] = \exp \left\{ s\mu_X + \frac{s^2}{2}\sigma_X^2 \right\} \frac{(2s)!}{2^s s!}.$$

Then

$$\text{Var}[Y_t] = \mathbb{E}[Y_t^2] - (\mathbb{E}[Y_t])^2 = \exp \left\{ \mu_X + \frac{1}{2}\sigma_X^2 \right\} \left(\frac{2!}{2^1 1!} - \left(\frac{1!}{2^0 0!} \right)^2 \right)$$

Hence if $\{X_t\}$ is stationary, $\{Y_t\}$ is an uncorrelated process.

Properties of the Stochastic Volatility Model

Also

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

and so the kurtosis for Y_t is

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

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

Linearization

If we square Y_t and take logarithms, then

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

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

$$\psi(z) = \frac{d}{dz} \log \Gamma(z) = \frac{\dot{\Gamma}(z)}{\Gamma(z)}.$$

Linearization

Thus, if $V_t = \log Z_t^2 + 1.2704$, then $\{V_t\}$ is IID with mean zero and variance $\pi^2/2$, we may rewrite the model as

$$\log Y_t^2 = -1.2704 + X_t + V_t.$$

Linearization

Hence $\{\log Y_t^2\}$ is a **linear** process which is the sum of

- the AR(1) process $\{X_t\}$
- a white noise process.

and $\log Y_t^2$ behaves (approximately) as an ARMA(1, 1) process, with ACF given by

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

7.5 Inference for the SV Model

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

The likelihood can be computed by integrating the latent log-volatilities out of the joint probability distribution.

7.5 Inference for the SV Model

The likelihood is

$$\begin{aligned} p(y_{1:n}; \boldsymbol{\theta}) &= \int p(y_{1:n}, \mathbf{x}_{1:n}; \boldsymbol{\theta}) d\mathbf{x}_{1:n} \\ &= \int p(y_{1:n} \mid \mathbf{x}_{1:n}; \boldsymbol{\theta}) p(\mathbf{x}_{1:n}; \boldsymbol{\theta}) d\mathbf{x}_{1:n}. \end{aligned}$$

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

Generalized Method-of-Moments (GMM)

The **generalized method of moments** (GMM) 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 $\boldsymbol{\theta} = (\phi_0, \phi, \sigma_W^2)$, and a large number of moments to use.

Generalized Method-of-Moments (GMM)

For example, we could define $\mathbf{g}$ via

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

where the theoretical values of $\mathbb{E}[Y_t^2 Y_{t-h}^2]$, for $h \geq 1$, 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 positive definite, symmetric weighting matrix.

GMM estimators are consistent if the observed process $\{Y_t\}$ is stationary. However,

- when $|\phi|$ is near 1, the GMM estimator can work poorly.
- the method does not deliver filtered or smoothed estimates of X_t .

7.6 Quasi-Maximum Likelihood estimation

The **Quasi-Maximum Likelihood** (QML) method is based on the linearization approach described earlier.

Let $Y_t^* = \log Y_t^2$. Then the SV model can be written as

$$Y_t^* = -1.2704 + X_t + V_t$$

$$X_t = \phi_0 + \phi X_{t-1} + W_t$$

where

$$V_t = \log Z_t^2 - \mathbb{E}[\log Z_t^2]$$

with $\sigma_V^2 = \text{Var}[V_t] = \pi^2/2$.

7.6 Quasi-Maximum Likelihood estimation

This is a linear but non-Gaussian state-space model. The QML approach treats the observation errors as if

$$V_t \sim \mathcal{N}(0, \pi^2/2)$$

and applies the standard Kalman filter

Given transformed data $\{y_1^*, \dots, y_n^*\}$, the Kalman recursions can be used to construct the approximate Gaussian likelihood.

7.7 Bayesian Inference

In a Bayesian procedure, we update our **prior** belief about θ and obtain a **posterior** distribution of the parameters, which is denoted $\pi_n(\theta | y_{1:n})$ and is given by

$$\pi_n(y_{1:n} | \theta) = \frac{p(y_{1:n} | \theta) \pi_0(\theta)}{\int p(y_{1:n} | \mathbf{t}) \pi_0(\mathbf{t}) d\mathbf{t}} \propto p(y_{1:n} | \theta) \pi_0(\theta),$$

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

7.7 Bayesian Inference

Let

- $\boldsymbol{\theta} = (\phi_0, \phi, \sigma_W^2)^\top$ denote the vector of parameters,
- $X_{1:n} = (X_1, \dots, X_n)^\top$ denote the vector of log-volatilities
- $y_{1:n} = (y_1, \dots, y_n)^\top$ denote the vector of observations,

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

7.7 Bayesian Inference

Key elements are

- the distribution of the observations conditional on the log-volatilities, $p(y_{1:n}|X_{1:n})$,
- the distribution of the log-volatilities conditional on the hyperparameters, $p(X_{1:n}|\theta)$
- the prior distribution of the parameters, $\pi_0(\theta)$.

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

$$\pi_n(X_{1:n}, \theta | y_{1:n}) \propto p(y_{1:n} | X_{1:n}) p(X_{1:n} | \theta) \pi_0(\theta).$$

7.7 Bayesian Inference

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

MCMC samples iteratively from the conditional posterior distributions: we initialize the values

$$\mathbf{x}_{1:n}^{(1)}, \boldsymbol{\theta}^{(1)}$$

and then for iteration $s = 2, 3, \dots$, sample

- $\boldsymbol{\theta}^{(s)}$ from $\pi_n(\boldsymbol{\theta} | X_{1:n}, y_{1:n})$, assuming $X_{1:n} = \mathbf{x}_{1:n}^{(s-1)}$
- $\mathbf{x}_{1:n}^{(s)}$ from $\pi_n(X_{1:n} | \boldsymbol{\theta}, y_{1:n})$, assuming $\boldsymbol{\theta} = \boldsymbol{\theta}^{(s)}$

and store the samples $\{\boldsymbol{\theta}^{(s)}, \mathbf{x}_{1:n}^{(s)}\}$ over many iterations.

7.7 Bayesian Inference

Particle filtering methods may also be used to perform filtering based inference;

- see previous knitr sheet.

It is also possible to integrate particle methods into MCMC:

- particle methods are used to estimate the likelihood
- the estimated likelihood used in the MCMC calculations
- can be effective if the number of particles is large.