Statistical Inference and Methods

David A. Stephens

Department of Mathematics
Imperial College London

d.stephens@imperial.ac.uk
http://stats.ma.ic.ac.uk/∼das01/

14th February 2006

Part VII

Session 7: Volatility Modelling

Session 7: Volatility Modelling

Volatility Modelling

- ARCH
- GARCH
- Stochastic Volatility
- Multivariate Volatility
- Methods of Inference

Session 7: Volatility Modelling

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

In particular, financial returns demonstrate volatility clustering, meaning that large changes tend to be followed by large changes and vice versa.

A conceptually useful division of these models into *observation-driven* and *parameter-driven* models.

- 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.

Session 7: Volatility Modelling

The most popular examples of observation-driven models are the *Autoregressive Conditional Heteroscedasticity* (*ARCH*) and *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 $y_t = 100 \log(x_t/x_{t-1})$, where x_t denotes the price of the asset. ARCH type models specify the distribution of the current observation as a one-step-ahead prediction density.

Session 7: Volatility Modelling

More precisely, for the observation-driven models, we assume $y_t \mid \Psi_{t-1} \sim N\left(0, \sigma_t^2\right)$, where Ψ_{t-1} contains all the information up to time $t-1$, so that $\Psi_t = \{y_t, y_{t-1}, \ldots\}$.

$$y_t = \sigma_t \varepsilon_t,$$

where $\{\varepsilon_t\}$ is a sequence of independent $N(0,1)$ 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

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

Session 7: Volatility Modelling

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

$$Var\left(y_t\right) = E\left(y_t^2\right) = \alpha_0 / \left(1 - \alpha_1\right).$$

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

$$E\left(y_t^4\right) = \frac{3\alpha_0^2\left(1 - \alpha_1^2\right)}{\left(1 - \alpha_1\right)^2\left(1 - 3\alpha_1^2\right)}.$$

The implied kurtosis is

$$-3 + E\left(y_t^4\right) / E\left(y_t^2\right)^2$$

and is greater than zero and hence y_t is *leptokurtotic* (fat tails).

Session 7: Volatility Modelling

The 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 y_t is zero mean, second order stationary if and only if $\alpha_1 + \beta_1 < 1$, with

$$Var(y_t) = \alpha_0 / (1 - \alpha_1 - \beta_1)$$

and all the odd moments equal to zero.

Session 7: Volatility Modelling

If in addition,

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

the fourth moment exists and is equal to

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

and 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.

Session 7: Volatility Modelling

There exist many other versions of the ARCH type models,

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

Session 7: Volatility Modelling

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

Inference For ARCH/GARCH Models

In this section, we will study the likelihood function for the ARCH and GARCH models to illustrate the Bayesian approach for the two univariate GARCH models.

The first model is an ordinary GARCH(1,1) and the second model is a Student-t GARCH(1,1). For both models, parameters are α_1, β_1, and $(\alpha_1 + \beta_1)$, which is recognized as a measure of *persistence*.

Session 7: Volatility Modelling

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.

Session 7: Volatility Modelling

Summary: The moments of the ARCH(1) model are given as follows

(i) $E(Y_t) = E(Y_t^3) = 0$

(ii) The second moment of Y_t is

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

(iii) The fourth moment of Y_t is

$$E(Y_t^4) = 3E(\sigma_t^4) = \frac{3\alpha_0^2\left(1-\alpha_1^2\right)}{(1-\alpha_1)^2\left(1-3\alpha_1^2\right)}, \qquad \text{for } 0 \leq \alpha_1^2 < \frac{1}{3}.$$

Session 7: Volatility Modelling

(iv) The kurtosis of Y_t is given by

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

- 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

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

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

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

Stationarity and Persistence in ARCH(1) Variance

- Condition for stationarity: $\alpha_1 < 1$;
- 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.
- 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 $3\alpha_1^2 < 1$.

Session 7: Volatility Modelling

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 unconditional 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; we may ensure this in the ARCH(1) case by using α_0^2 instead of α_0, and α_1^2 instead of $\alpha_1$as starting values in the Maximum Likelihood (ML) calculations if these parameters should be negative. Doing so imposes positive parameter values from new ML results.

Session 7: Volatility Modelling

The likelihood for ARCH(1) can be written as

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

where $y=(y_1,y_2,...,y_n)$. Thus the log likelihood is

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

Session 7: Volatility Modelling

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

$$\frac{\partial \log f}{\partial \alpha_0} = \frac{1}{2}\sum_{t=1}^{n}\left(\frac{\partial \sigma_t^2}{\partial \alpha_0}\right)\frac{1}{\sigma_t^2}\left(\frac{y_t^2}{\sigma_t^2}-1\right), \qquad \frac{\partial \sigma_t^2}{\partial \alpha_0} = 1,$$

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

For ARCH(p), $(\alpha_0, \alpha_1)^T$ becomes $(\alpha_0, \alpha_1, .., \alpha_p)^T$ so

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

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.

Session 7: Volatility Modelling

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

(i) $E(Y_t) = E(Y_t^3) = 0$

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

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

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

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

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

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

When $\beta_1 = 0$, this condition is the same as the ARCH(1) model, but when $\beta_1 > 0$, α_1 has to be lower 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.

Session 7: Volatility Modelling

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

$$\rho_1 = \frac{\alpha_1(1-\beta_1^2-\alpha_1\beta_1)}{(1-\beta_1^2-2\alpha_1\beta_1)}, \qquad \rho_s = (\alpha_1+\beta_1)\rho_{s-1}, \qquad \text{for } s \geq 2$$

Clearly ρ_s depends on the values of α_1 and β_1.

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.

Stationarity and Persistence in GARCH(1,1) Volatility

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. $\alpha_1 + ... + \alpha_p + \beta_1 + ... + \beta_q < 1$), in which case the unconditional variance of Y_t,

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

is a finite constant.

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

Session 7: Volatility Modelling

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

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

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

and $Cov(Y_t, Y_s) \neq 0$.

(ii) If $\alpha_1 + \beta_1 \longrightarrow 1$, 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.

Session 7: Volatility Modelling

(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 broadly an adequate model that has been successfully used 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

An ML estimation structure 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

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

Session 7: Volatility Modelling

To obtain the ML estimates. we need to implement a numerical calculation for the partial derivatives recursively for $t = 1, .., n$. Unlike ARCH(p), the ML for GARCH(1,1) is more complicated than just implementing the previous procedure 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

The Bayesian posterior distribution is

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

Session 7: Volatility Modelling

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

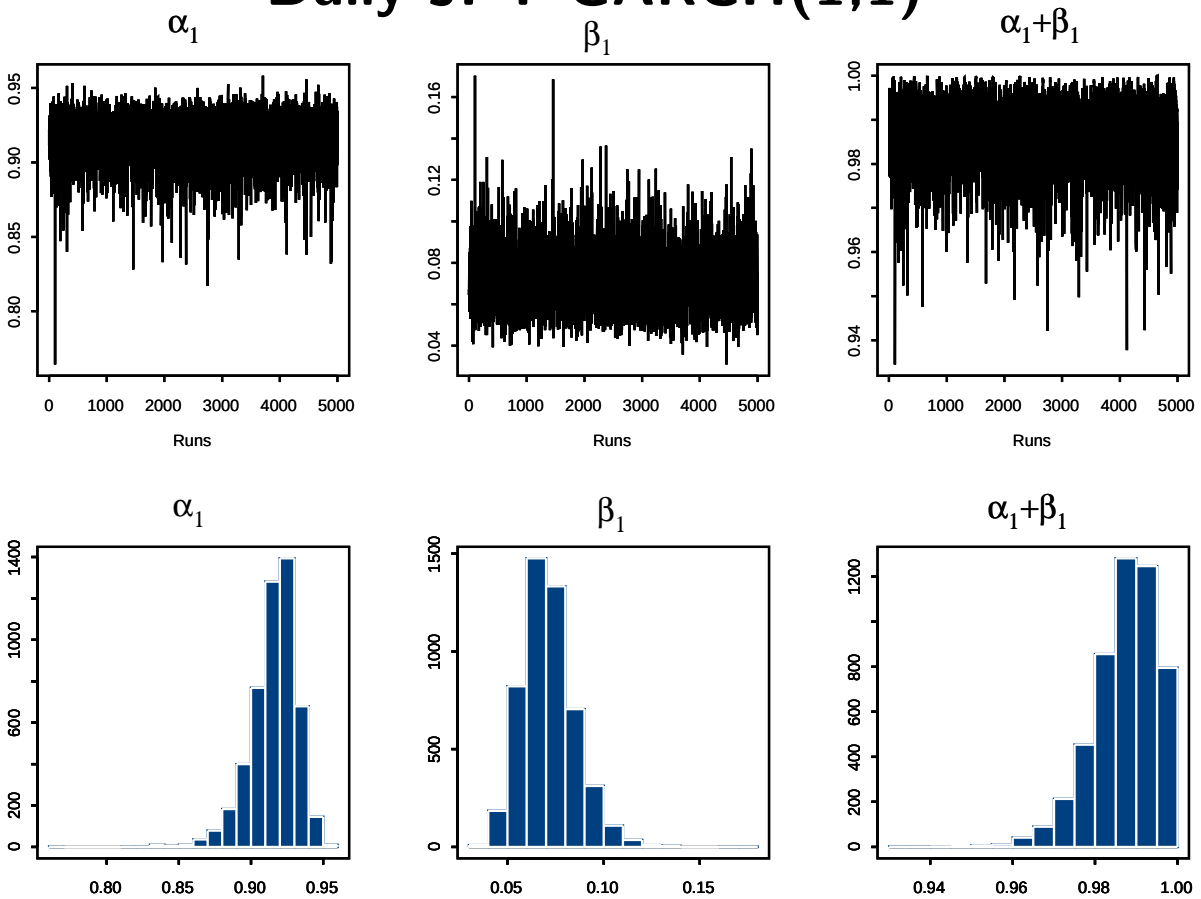
- Daily data
- Hourly data

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 500^{th} iteration.

Session 7: Volatility Modelling

Daily JPY GARCH(1,1)



Session 7: Volatility Modelling

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.

Session 7: Volatility Modelling

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 $GARCH(p,q)$ 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 significantly 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.

Session 7: Volatility Modelling

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

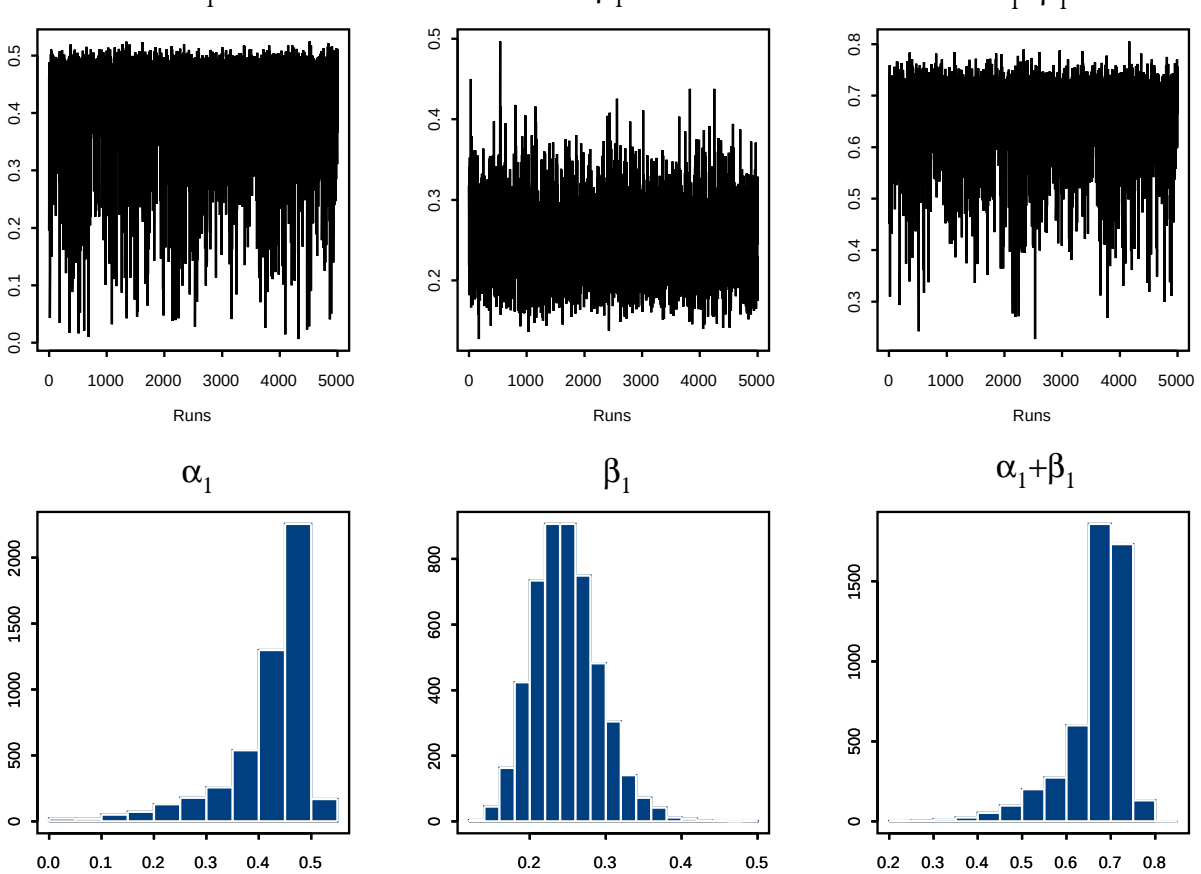
$$3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$$

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.

Daily JPY GARCH(1,1): Constrained Model



Session 7: Volatility Modelling

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

Session 7: Volatility Modelling

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

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

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

Session 7: Volatility Modelling

- The parameters λ_t, $t = (1, .., n)$ 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$(1, 1)$

Session 7: Volatility Modelling

- 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 $\nu \to \infty$, 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 ν =5 for 12 FX series.

The Student-t GARCH(1,1) 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

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.

Session 7: Volatility Modelling

The most popular SV model is

$$\begin{aligned} y_t &= \exp\left(h_t/2\right)\varepsilon_t \\ h_t &= \gamma + \phi h_{t-1} + \eta_t \end{aligned}$$

where y_t is, as usual, the observation at time t, the ε_ts are independent identically distributed (i.i.d.) $N(0,1)$ random variables, the η_ts are also i.i.d. $N\left(0,\sigma_\eta^2\right)$ random variables.

The latent process h_t can be interpreted as the random and uneven flow of new information into the market, and ϕ is the persistence in the volatility.

Session 7: Volatility Modelling

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.

More precisely, 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 GARCH model, 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 EGARCH model.

Session 7: Volatility Modelling

Persistence in Volatility: A sset returns has been found to have quite high autocorrelations for long lags. The SV model can capture this phenomenon very easily. As has already been mentioned, 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

For simplicity, the error processes, ε_t and η_t, in the SV model are initially presumed independent. If $|\phi| < 1$, the process $\{h_t\}$ is strictly stationary with unconditional mean and variance given respectively by

$$\begin{aligned} \mu_h &= E(h_t) = \frac{\gamma}{1-\phi} \\ \sigma_h^2 &= Var(h_t) = \frac{\sigma_\eta^2}{1-\phi^2}. \end{aligned}$$

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

Then

$$E(y_t) = E(y_t \mid \Psi_{t-1}) = 0$$

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

$$\rho_{y_t}(\tau) = E(y_t y_{t-\tau}) = E\left(\exp\left(\frac{h_t}{2} + \frac{h_{t-\tau}}{2}\right)\right) E(\varepsilon_t \varepsilon_{t-\tau}) = 0.$$

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

Session 7: Volatility Modelling

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

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

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

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

Session 7: Volatility Modelling

In particular

$$Var\left(y_t\right) = E\left(\left(y_t\right)^2\right) = \exp\left\{\mu_h + \frac{1}{2}\sigma_h^2\right\}$$

and hence if h_t is stationary, y_t is a *white-noise* process. The fourth moment is

$$E\left[\left(y_t\right)^4\right] = 3\exp\left\{2\mu_h + 2\sigma_h^2\right\}$$

and so the kurtosis for y_t is

$$-3 + E\left(y_t^4\right)/E\left(y_t^2\right)^2 = 3\left(\exp\left(\sigma_h^2\right) - 1\right),$$

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

Session 7: Volatility Modelling

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

$$\log y_t^2 = h_t + \log \varepsilon_t^2.$$

If ε_t has a standard normal distribution, then $\log \varepsilon_t^2$ 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 $\xi_t = \log \varepsilon_t^2 + 1.2704$, then clearly ξ_t is i.i.d. with mean zero and variance $\pi^2/2$ and we may rewrite the model as

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

Session 7: Volatility Modelling

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

$$\rho_{\log y_t^2}(\tau) = \frac{\phi^\tau}{1 + \left(\pi^2/2\sigma_h^2\right)}, \quad \tau = 1, 2, \ldots.$$

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.

Session 7: Volatility Modelling

When the errors ε_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 $\varepsilon_t \sim t_\nu$, then

$$Var\left(\varepsilon_t\right) = E\left(\varepsilon_t^2\right) = \nu/\left(\nu-2\right),\ \nu > 2$$

and

$$E\left(\varepsilon_t^4\right) = 3\nu^2/\left[\left(\nu-2\right)\left(\nu-4\right)\right],\ \nu > 4.$$

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

$$Var\left(y_t\right) = E\left(y_t^2\right) = \frac{\nu}{\nu-2}\exp\left\{\mu_h + \frac{1}{2}\sigma_h^2\right\},\ \nu > 2$$

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

Session 7: Volatility Modelling

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

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

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

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

$$E(\log \kappa_t) = \psi(\nu/2) - \log(\nu/2)$$

and

$$Var(\log \kappa_t) = \psi'(\nu/2)$$

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

Session 7: Volatility Modelling

Therefore, if we now define

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

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

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

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

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

Inference for the Stochastic Volatility Model

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

One way of deriving the likelihood is by integrating the latent log-volatilities out of the joint probability distribution. In particular, denote by $\mathbf{y} = (y_1, \ldots, y_T)^{\mathsf{T}}$ the vector of observations for T consecutive periods, $\mathbf{h} = (h_1, \ldots, h_T)^{\mathsf{T}}$ the vector of the corresponding log-volatilities and $\boldsymbol{\theta} = (\gamma, \phi, \sigma_\eta^2)$.

Session 7: Volatility Modelling

Then, the likelihood is given by

$$L(\mathbf{y};\boldsymbol{\theta}) = \int p(\mathbf{y},\mathbf{h}|\boldsymbol{\theta})\, d\mathbf{h} = \int p(\mathbf{y} \mid \mathbf{h},\boldsymbol{\theta})\, p(\mathbf{h} \mid \boldsymbol{\theta})\, d\mathbf{h}.$$

This last integral is of dimension equal to the sample size, T, its evaluation requires the use of numerical procedures and this makes the estimation of the hyperparameters, $\boldsymbol{\theta}$, via the Maximum Likelihood method quite involved.

Session 7: Volatility Modelling

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 T, $\mathbf{y}$, 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 $\boldsymbol{\theta} = \left(\gamma, \phi, \sigma_\eta^2\right)$, and a large number of moments to use.

Session 7: Volatility Modelling

For example, we might define estimating function $\mathbf{g}$ with components

$$\frac{1}{T}\sum y_t^2 - E\left(y_t^2\right)$$
$$\frac{1}{T}\sum y_t^4 - E\left(y_t^4\right)$$
$$\frac{1}{T}\sum y_t^2 y_{t-1}^2 - E\left(y_t^2 y_{t-1}^2\right)$$
$$\vdots$$
$$\frac{1}{T}\sum y_t^2 y_{t-\tau}^2 - E\left(y_t^2 y_{t-\tau}^2\right)$$

where the theoretical values of $E(y_t^2 y_{t-\tau}^2)$, for $\tau \geq 1$, can be found analytically.

Session 7: Volatility Modelling

The objective function to be minimized is then

$$Q = \mathbf{g}^{\mathsf{T}}\mathbf{W}\mathbf{g}$$

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

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

Furthermore, GMM is asymptotically 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.

Session 7: Volatility Modelling

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*

$$C.V. = Var\left(\exp\left(h_t\right)\right) / \left\{E\left(\exp\left(h_t\right)\right)\right\}^2 = \exp\left(\sigma_h^2\right) - 1.$$

- GMM parameter estimates are not invariant to reparameterization

$$\boldsymbol{\psi} = f\left(\boldsymbol{\theta}\right) \quad \text{then} \quad \widehat{\boldsymbol{\psi}} \neq f\left(\widehat{\boldsymbol{\theta}}\right)$$

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

Session 7: Volatility Modelling

Quasi-Maximum Likelihood (QML): The QML method is based on the linearization of the SV model by squaring y_t and taking logarithms. Assuming that the errors $\varepsilon_t \sim N(0,1)$ and denoting $w_t = \log y_t^2$, as has already been seen, the SV model can be written as

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

where $\xi_t = \log \varepsilon_t^2 - E\left(\log \varepsilon_t^2\right)$, with $\sigma_\xi^2 = Var\left(\xi_t\right) = \pi^2/2$.

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 $N\left(0, \pi^2/2\right)$ and apply the standard Kalman filter.

Session 7: Volatility Modelling

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 $\{y_1, \ldots, y_T\}$, or equivalently $\left\{\log y_1^2, \ldots, \log y_T^2\right\}$, 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. Before we proceed with describing the method in more detail, we make one more simplifying transformation of the model,

Session 7: Volatility Modelling

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

$$E(w_t) = \gamma^* = -1.2704 + \mu_h = -1.2704 + \gamma/(1-\phi).$$

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

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

Session 7: Volatility Modelling

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^*.

Therefore, by applying this last transformation on the SV model, we have managed to "concentrate" the parameter γ out of the likelihood; we can apply the Kalman filter to the model with the mean centered observations, and obtain the QML estimates of $\boldsymbol{\theta} = \left(\phi, \sigma_\eta^2\right)^{\mathsf{T}}$.

Once the estimates ϕ and σ_η^2 are available, the QML estimator of γ will be given by $\gamma = (1-\phi)\left(1.2704 + \frac{1}{T}\sum \log y_t^2\right)$.

Session 7: Volatility Modelling

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

$$a_{0|0} = E(\alpha_t) = 0 \qquad P_{0|0} = Var(\alpha_t) = \sigma_\eta^2/(1-\phi^2),$$

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

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

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

Session 7: Volatility Modelling

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

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

where the terms $w_t^* - a_{t|t-1}$ are the innovations in predicting w_t^* given past observations $\left\{ w_{t-1}^*, \ldots, w_1^* \right\}$ and $f_t = P_{t|t-1} + \sigma_\xi^2$ are the MSE's 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.

Session 7: Volatility Modelling

A Gaussian (quasi) log-likelihood can be constructed

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

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

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

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

Session 7: Volatility Modelling

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

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

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.

Session 7: Volatility Modelling

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

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

$$\sigma_\xi^2 = Var(\xi_t) = \pi^2/2 + \psi'(\nu/2).$$

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.

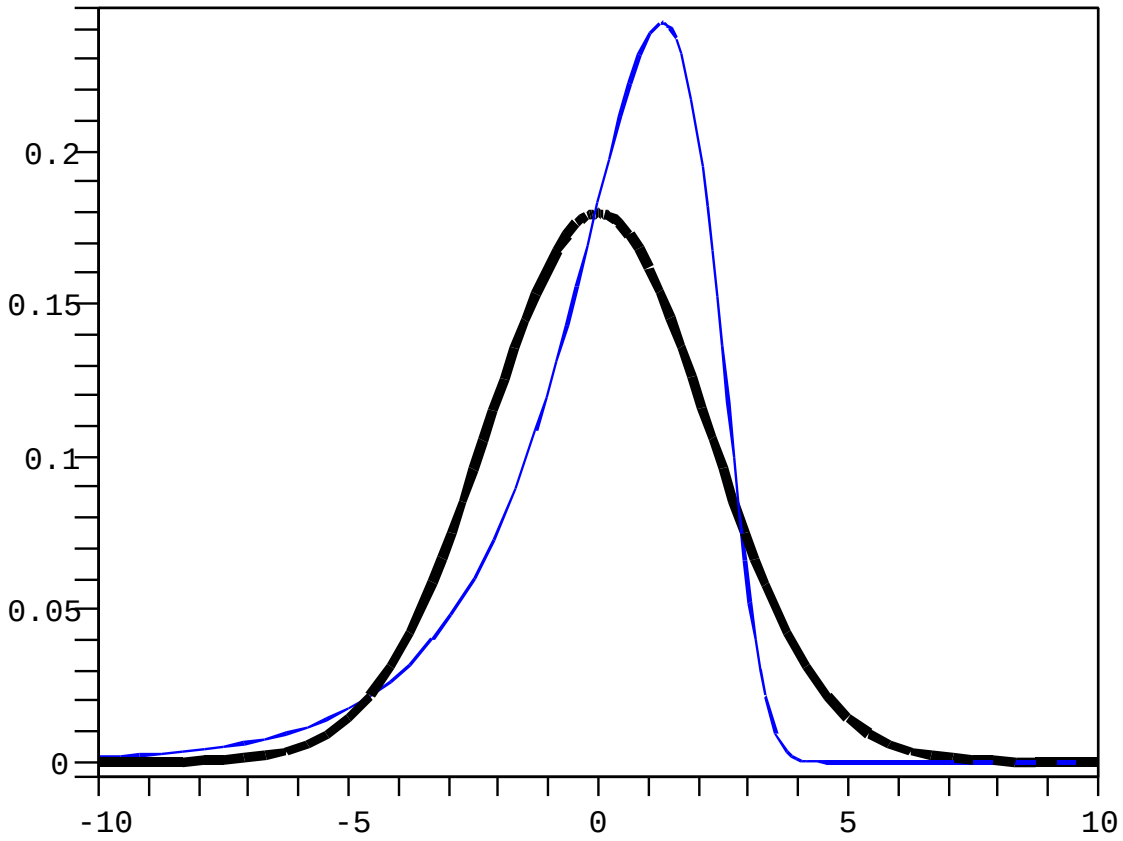
Session 7: Volatility Modelling

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

Session 7: Volatility Modelling

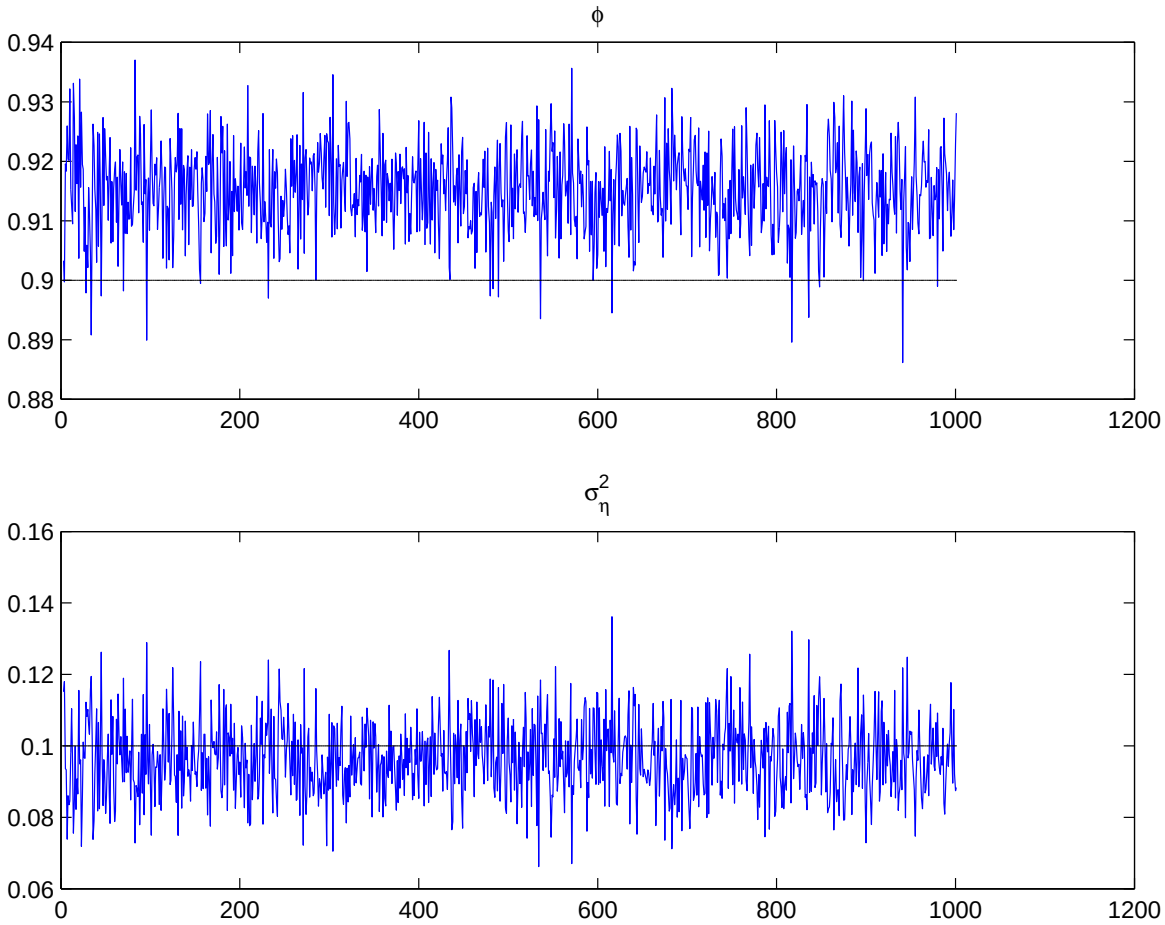


Session 7: Volatility Modelling

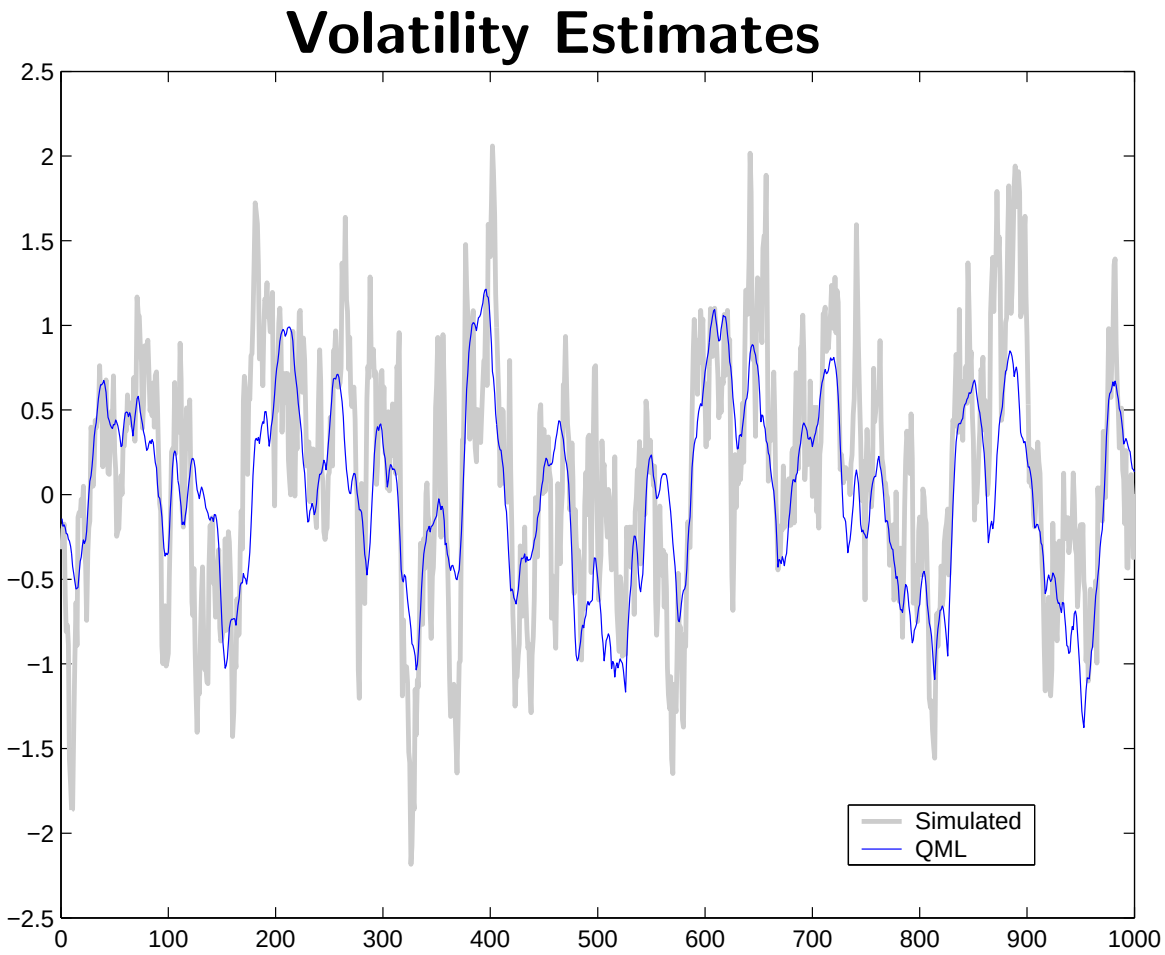
Example: Simulated a sample of 1,000 values from a SV model with parameters $\gamma = 0$, so that $\mu_h = \gamma/(1-\phi) = 0$, $\phi = 0.9$ and $\sigma_\eta^2 = 0.1$. The size of the sample is typical for financial data, as are the chosen parameter values.

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.

Session 7: Volatility Modelling



Session 7: Volatility Modelling



Session 7: Volatility Modelling

Inliers: A drawback to the QML procedure worth noting 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

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

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

Bayesian Approaches to 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).

Session 7: Volatility Modelling

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

Let

- $\boldsymbol{\theta} = \left(\gamma, \phi, \sigma_\eta^2\right)^{\mathsf{T}}$ denote the vector of hyperparameters,
- $\mathbf{h} = (h_1, \ldots, h_T)^{\mathsf{T}}$ denote the vector of log-volatilities
- $\mathbf{y} = (y_1, \ldots, y_T)^{\mathsf{T}}$ the vector of observations,

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

Session 7: Volatility Modelling

- the distribution of the observations conditional on the log-volatilities, $p(\mathbf{y}|\mathbf{h})$,
- the distribution of the log-volatilities conditional on the hyperparameters, $p(\mathbf{h}|\theta)$
- the prior distribution of the hyperparameters, $p(\theta)$.

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

$$p(\mathbf{h}, \theta|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{h})\, p(\mathbf{h}|\theta)\, p(\theta).$$

Gibbs sampler for the SV model

1. Choose arbitrary starting values $\mathbf{h}^{(0)}$, $\boldsymbol{\theta}^{(0)}$ and let $i = 0$.
2. Sample $\mathbf{h}^{(i+1)} \sim p\left(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}^{(i)}\right)$.
3. Sample $\boldsymbol{\theta}^{(i+1)} \sim p\left(\boldsymbol{\theta}|\mathbf{y}, \mathbf{h}^{(i+1)}\right)$.
4. Set $i = i + 1$ and goto 1.

Session 7: Volatility Modelling

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

$$p\left(\mathbf{h}|\mathbf{y},\boldsymbol{\theta}^{(i)}\right)$$

is not that straightforward.

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

$$p\left(h_t|\mathbf{h}_{\setminus t}^{(i)},\mathbf{y},\boldsymbol{\theta}^{(i)}\right)$$

where

$$\mathbf{h}_{\setminus t}^{(i)}=\left(h_1^{(i+1)},\ldots,h_{t-1}^{(i+1)},h_{t+1}^{(i)},\ldots,h_T^{(i)}\right).$$

Session 7: Volatility Modelling

Step 2 of the Gibbs Sampler algorithm becomes:

2a. For $t = 1, \ldots, T$, sample

$$h_t^{(i+1)} \sim p\left(h_t | \mathbf{h}_{\setminus t}^{(i)}, \mathbf{y}, \boldsymbol{\theta}^{(i)}\right)$$

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

$$\begin{aligned} p\left(h_t | \mathbf{h}_{\setminus t}, \mathbf{y}, \boldsymbol{\theta}\right) &= p\left(h_t | h_{t-1}, h_{t+1}, y_t, \boldsymbol{\theta}\right) \\ &\propto p\left(y_t | h_t\right) p\left(h_{t+1} | h_t, \boldsymbol{\theta}\right) p\left(h_t | h_{t-1}, \boldsymbol{\theta}\right), \end{aligned}$$

where the second line is deduced from Bayes theorem.

Session 7: Volatility Modelling

Rejection Metropolis-Hastings: An first approach to the estimation of the SV model via MCMC was offered in the literature using ideas from non-Gaussian and non-linear state-space modeling. Consider the parameterization of the SV model:

$$\begin{aligned} y_t &= \sqrt{h_t}\varepsilon_t, \\ \log h_t &= \gamma + \phi \log h_{t-1} + \eta_t, \quad t = 1, \ldots, T, \end{aligned}$$

where ε_t and η_t are contemporaneously and serially independent random variables with distributions $N(0,1)$ and $N\left(0, \sigma_\eta^2\right)$, respectively.

Session 7: Volatility Modelling

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

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

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

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

Session 7: Volatility Modelling

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

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

where $l_t = \log h_t$.

Session 7: Volatility Modelling

After some algebra it follows that

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

where

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

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

Session 7: Volatility Modelling

The idea is to place the rejection sampling method within an independence M-H algorithm; this approach is based on a density g and a constant c such that

$$p\left(x|\mathbf{h}_{\backslash t}, \mathbf{y}, \boldsymbol{\theta}\right) \leq cg\left(x\right)$$

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

For each time t, proposals x_t are generated from the density g, until one of this proposals is accepted with probability $\min\left\{1, p\left(x_t|\mathbf{h}_{\backslash t}^{(i)}, \mathbf{y}, \boldsymbol{\theta}\right)/cg\left(x_t\right)\right\}$.

Session 7: Volatility Modelling

The accepted x_t then enters an independence M-H accept-reject step and we set $h_t^{(i+1)} = x_t$ with probability

$$\min\left[1, \frac{p\left(x_t|\mathbf{h}_{\backslash t}^{(i)}, \mathbf{y}, \boldsymbol{\theta}\right) \min\left\{p\left(h_t^{(i)}|\mathbf{h}_{\backslash t}^{(i)}, \mathbf{y}, \boldsymbol{\theta}\right), cg\left(h_t^{(i)}\right)\right\}}{p\left(h_t^{(i)}|\mathbf{h}_{\backslash t}^{(i)}, \mathbf{y}, \boldsymbol{\theta}\right) \min\left\{p\left(x_t|\mathbf{h}_{\backslash t}^{(i)}, \mathbf{y}, \boldsymbol{\theta}\right), cg\left(x_t\right)\right\}}\right].$$

If x_t is not accepted, then we set $h_t^{(i+1)} = h_t^{(i)}$.

Session 7: Volatility Modelling

Thus, the problem is reduced to finding the pseudo-dominating density g and choosing accordingly the constant c.

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

- an improper inverse-gamma density, $IG\left(-0.5, 0.5y_t^2\right)$,
- a log-normal density.

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

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

Session 7: Volatility Modelling

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

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

Session 7: Volatility Modelling

This method has been regarded as inefficient, as the posterior density $p\left(h_t|\mathbf{h}_{\setminus t}, \mathbf{y}, \boldsymbol{\theta}\right)$ can clearly be written as

$$p\left(h_t|\mathbf{h}_{\setminus t}, \mathbf{y}, \boldsymbol{\theta}\right) \propto p\left(y_t|h_t\right) p\left(h_t|h_{t-1}, h_{t+1}, \boldsymbol{\theta}\right).$$

In essence, the initial method attempts to approximate the density

$$p\left(h_t|h_{t-1}, h_{t+1}, \boldsymbol{\theta}\right) \qquad by \qquad p\left(y_t|h_t\right).$$

However, in practice, the prior density

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

dominates the likelihood, and therefore one should focus in approximating $p\left(h_t|\mathbf{h}_{\setminus t}, \mathbf{y}, \boldsymbol{\theta}\right)$ by a density of the same form as $p\left(h_t|h_{t-1}, h_{t+1}, \boldsymbol{\theta}\right)$.

Session 7: Volatility Modelling

Consider the following parameterization of the SV model:

$$\begin{aligned} y_t &= \exp(h_t/2)\varepsilon_t, \\ h_t &= \mu_h + \phi(h_{t-1} - \mu_h) + \eta_t, \quad t = 1, \ldots, T, \end{aligned}$$

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

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

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

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

Session 7: Volatility Modelling

Thus $\log p\left(h_t|\mathbf{h}_{\setminus t}, \mathbf{y}, \boldsymbol{\theta}\right)$ is given by

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

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

Session 7: Volatility Modelling

In addition, the terms in $\log g$ can be combined and after some algebra, it is easily seen that $g(h_t) \propto f_N\left(h_t|\mu_t, \sigma_t^2\right)$, where $f_N\left(x|\alpha, \beta^2\right)$ denotes the density of a Gaussian random variable with mean α and variance β^2, and where

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

Session 7: Volatility Modelling

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 given as before by (15), it follows that

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

Using the latter result

$$\begin{aligned} \log p\left(h_t | \mathbf{h}_{\setminus t}, \mathbf{y}, \boldsymbol{\theta}\right) &\leq \text{const} - \frac{(h_t - m_t)^2}{2\sigma_*^2} - \frac{h_t}{2} \\ &\quad - \frac{y_t^2}{2}\exp(-m_t)(1 + m_t - h_t) \\ &= \log g(h_t). \end{aligned}$$

Session 7: Volatility Modelling

The terms in $\log g$ can be combined and thus $g(h_t) \propto f_N\left(h_t|\mu_t, \sigma_*^2\right)$, where σ_*^2 is as previously, and μ_t is

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

In this case the density $f_N\left(\cdot|\mu_t, \sigma_*^2\right)$ does bound $p\left(\cdot|\mathbf{h}_{\setminus t}, \mathbf{y}, \boldsymbol{\theta}\right)$ and so rejection sampling can be used.

In order to update the log-volatilities, for each time t, proposals x_t are generated from a $N\left(\mu_t, \sigma_*^2\right)$ density until one of these is accepted with probability

$$\min\left\{1, \frac{p\left(x_t|\mathbf{h}_{\setminus t}^{(i)}, \mathbf{y}, \boldsymbol{\theta}\right)}{f_N\left(x_t|\mu_t, \sigma_*^2\right)}\right\}$$

we then set $h_t^{(i+1)} = x_t$ and proceed to update $h_{t+1}^{(i)}$ in the same

Session 7: Volatility Modelling

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

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

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

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

Session 7: Volatility Modelling

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

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

where

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

Session 7: Volatility Modelling

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

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

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

$$\nu_\sigma = (\alpha_\sigma + T)/2 \qquad \lambda_\sigma = (\beta_\sigma + SSQ_h)/2.$$

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

Session 7: Volatility Modelling

Sampling ϕ.

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

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

The full conditional posterior for ϕ is

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

Session 7: Volatility Modelling

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 $f_N\left(\cdot|\mu_\phi, \sigma^2_\phi\right)$, the density of a Normal random variable with mean and variance respectively given by

$$\mu_\phi = \frac{\sum_{t=1}^{T-1}\left(h_{t+1}-\mu_h\right)\left(h_t-\mu_h\right)}{\sum_{t=1}^{T-1}\left(h_t-\mu_h\right)^2} \quad \text{and} \quad \sigma^2_\phi = \frac{\sigma^2_\eta}{\sum_{t=1}^{T-1}\left(h_t-\mu_h\right)^2}.$$

Session 7: Volatility Modelling

The posterior density of ϕ is:

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

Within a M-H setting, we choose an initial value $\left|\phi^{(0)}\right|<1$. 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 $-1+2\alpha_\phi/\left(\alpha_\phi+\beta_\phi\right) \simeq 0.8605$.

Session 7: Volatility Modelling

Sampling μ_h.

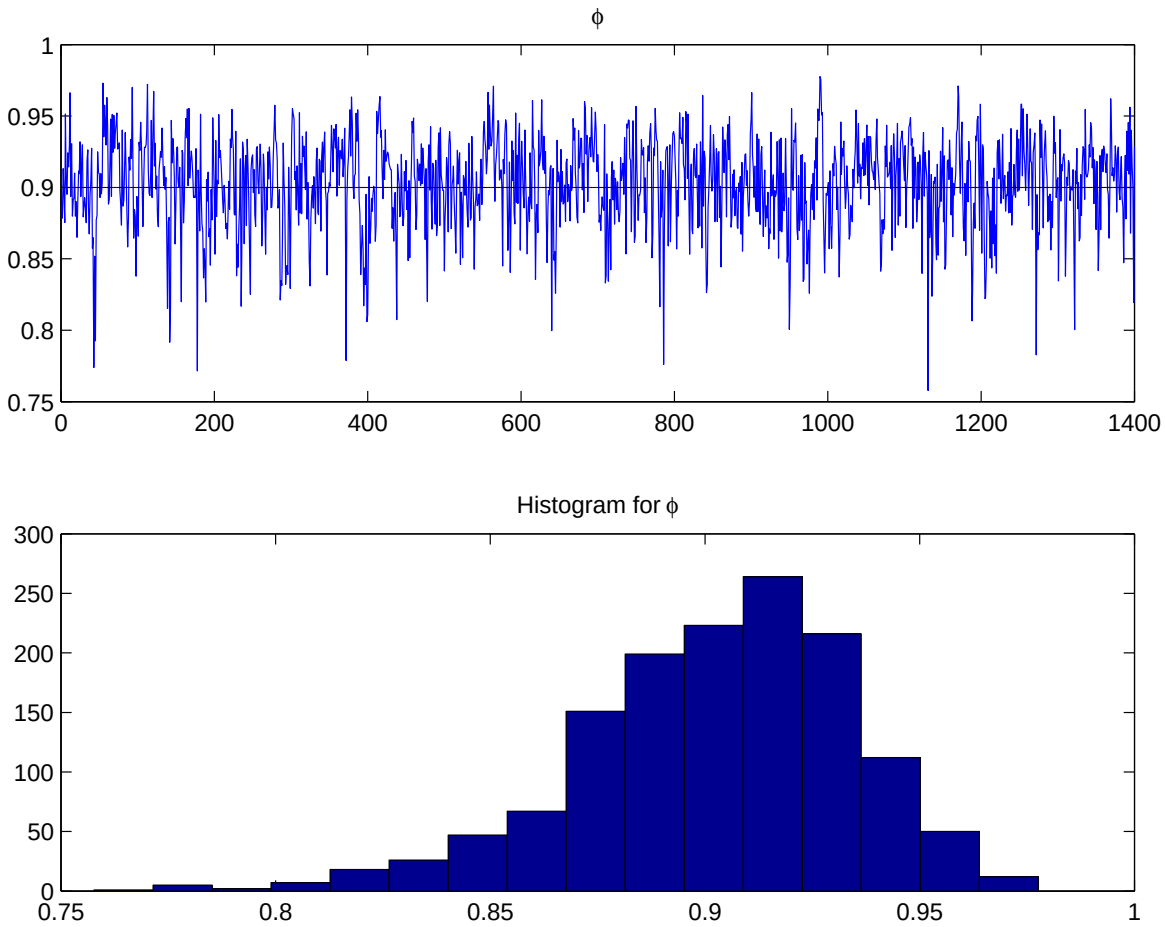
Under a non-informative prior distribution for μ_h, $p(\mu_h) \propto 1$, it is easily seen that the posterior density satisfies

$$p\left(\mu_h|\mathbf{h}, \phi, \sigma_\eta^2\right) \propto \exp\left\{-\frac{SSQ_h}{2\sigma_\eta^2}\right\}.$$

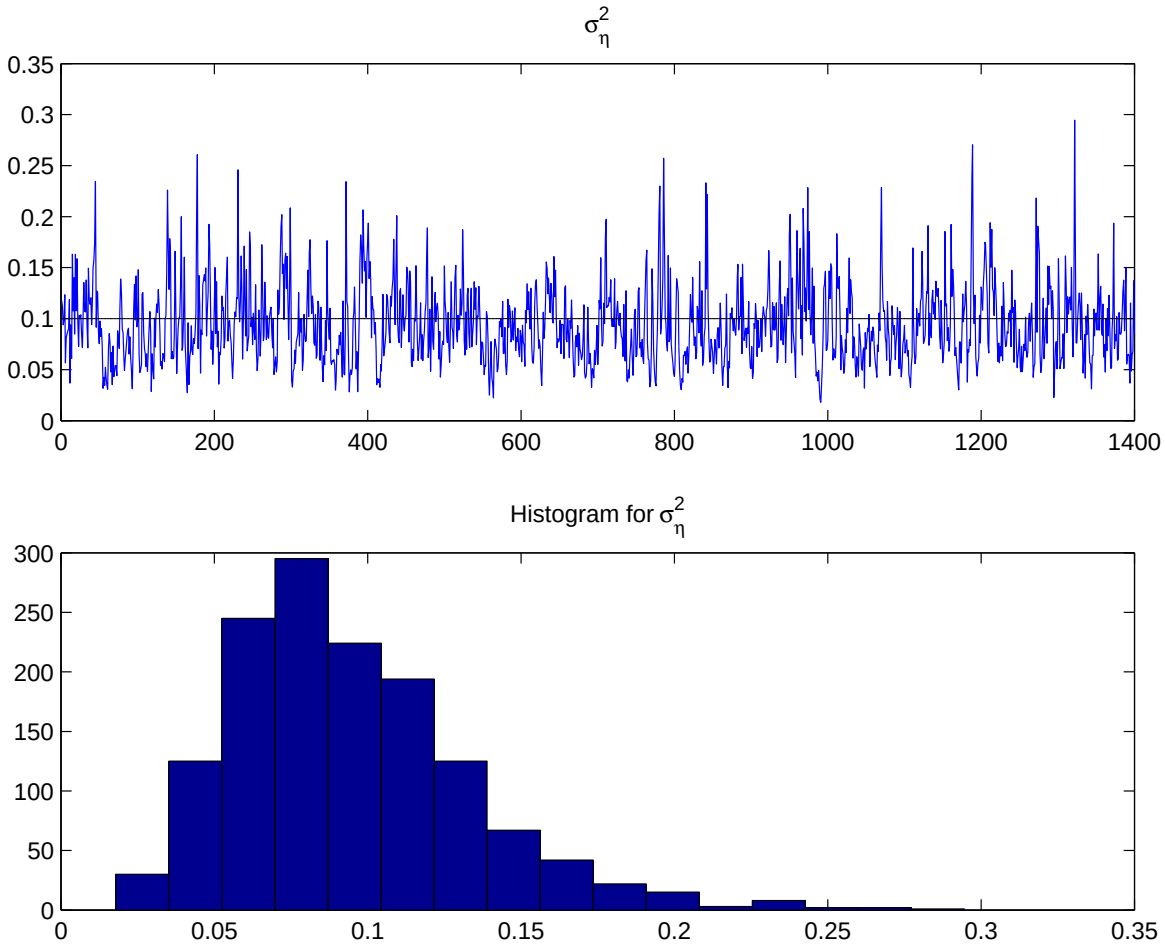
After some algebra, it follows that the parameter μ_h can also be directly updated in the Gibbs sampler, $\mu_h \sim N\left(m_h, s_h^2\right)$, where

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

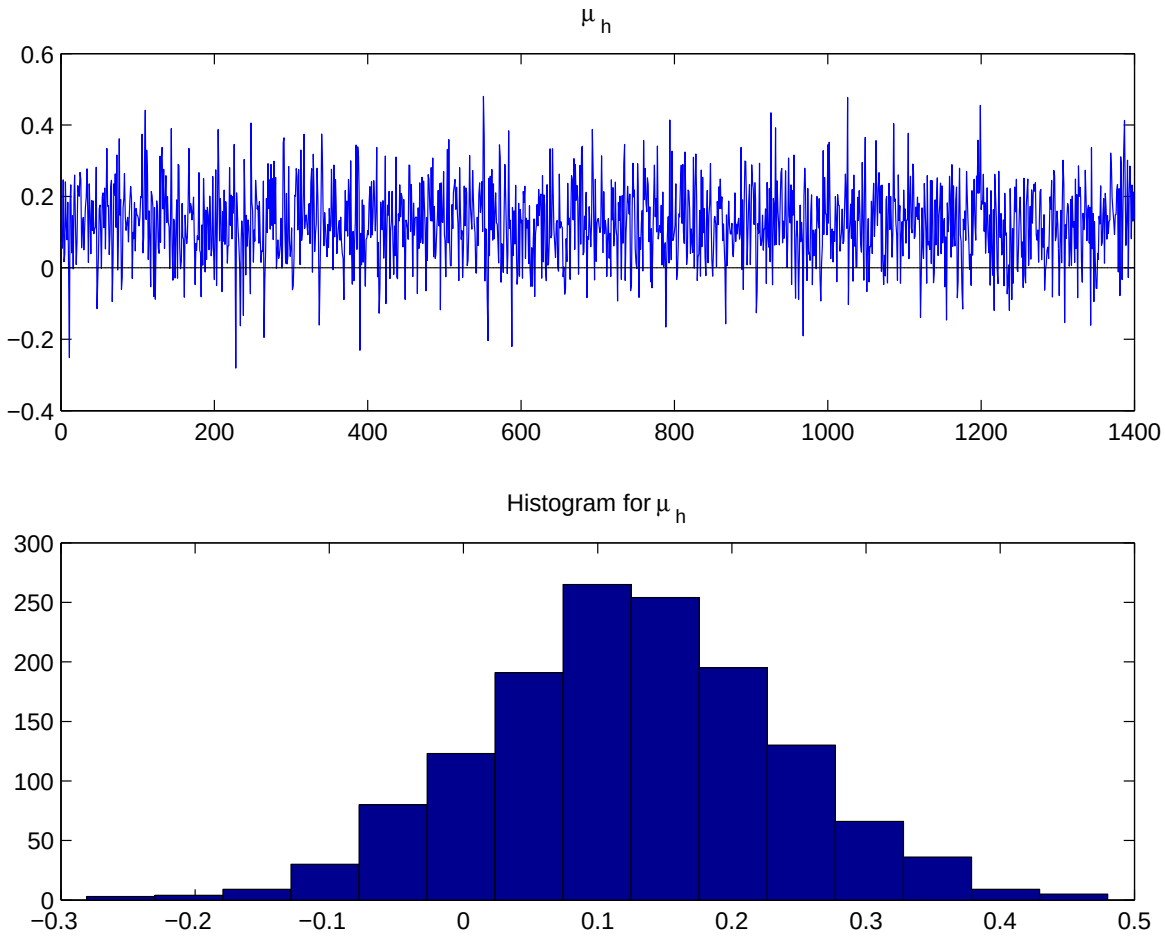
Session 7: Volatility Modelling



Session 7: Volatility Modelling

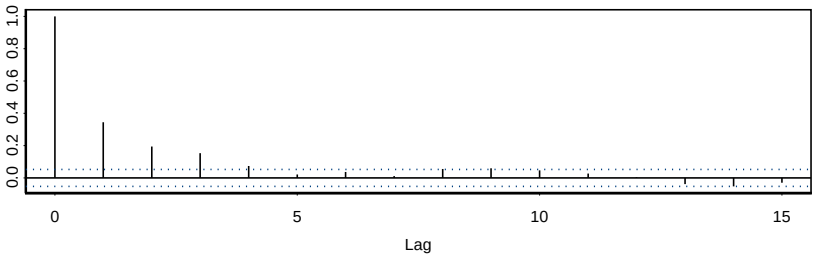


Session 7: Volatility Modelling

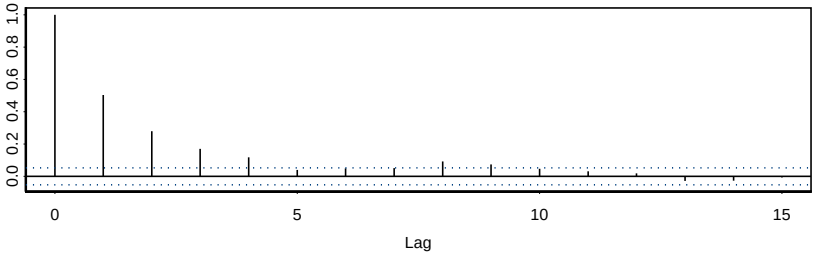


Session 7: Volatility Modelling

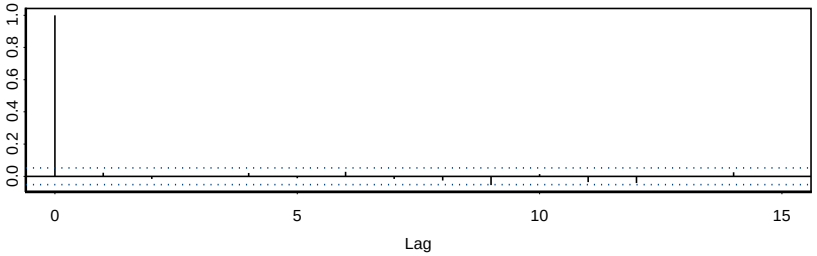
ϕ



σ_η^2



μ_h



Session 7: Volatility Modelling

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.

Session 7: Volatility Modelling

- 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 $p(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$.

Multimove MCMC Samplers

Multimove MCMC samplers, based on the linear state-space representation, have had a great impact on other approaches.

The multimove approach begins by write the SV model as

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

Session 7: Volatility Modelling

Instead of approximating the $\log \chi^2$ distribution of $z_t = \log \varepsilon_t^2$ with a $N\left(-1.2704, \pi^2/2\right)$, which is suboptimal and results in poor sample properties, a better choice is to approximate the $\log \chi^2$ distribution by a mixture of seven normals, so that

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

where $z_t|s_t = j \sim N\left(m_j - 1.2704, v_j^2\right)$, s_t is the indicator variable at time t, and the q_j, $j = 1, \ldots, 7$ are the weights attached to each component, so that $p(s_t = j) = q_j$ and $\sum_{j=1}^{7} q_j = 1$. The elements $\{q_j, m_j, v_j^2, j = 1, \ldots, 7\}$ are selected by matching the first four moments of the $\log \chi^2$ distribution to the approximating mixture of normals.

Session 7: Volatility Modelling

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

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

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 $h_1, \ldots, h_T$ at once from $p(\mathbf{h}|\mathbf{w}, \mathbf{s})$, where $\mathbf{w} = \left(\log y_1^2, \ldots, \log y_T^2\right)$ and $\mathbf{s} = (s_1, \ldots, s_T)$.

Therefore, the main idea of this approach is to augment further the vector of parameters and log-volatilities, $(\boldsymbol{\theta}, \mathbf{h})$, with the vector $\mathbf{s}$ of indicator variables and construct a Gibbs sampler which will produce samples from the density $p(\boldsymbol{\theta}, \mathbf{h}, \mathbf{s}|\mathbf{w})$.

Session 7: Volatility Modelling

The general form of the multimove Gibbs sampler for the SV model is as follows:

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

Session 7: Volatility Modelling

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

$$p(\mathbf{s}|\mathbf{w},\mathbf{h}) = \prod_{t=1}^{T} p(s_t|w_t,h_t) \propto \prod_{t=1}^{T} p(w_t|s_t,h_t)\, p(s_t)$$

$$p(s_t = j|w_t,h_t) \propto q_j\, f_N\left(w_t|h_t + m_j - 1.2704, v_j^2\right), \quad j = 1,\ldots,K,$$

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

Session 7: Volatility Modelling

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 $\boldsymbol{\theta}$, 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.

Session 7: Volatility Modelling

Another multimove 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

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

is updated by sampling from

$$p\left(h_t, \ldots, h_{t+k} | y_t, \ldots, y_{t+k}, h_{t-1}, h_{t+k}, \theta\right)$$

Multivariate Stochastic Volatility Models

Multivariate modelling allows representation of the volatility co-movements amongst financial time series.

- Interactive movement between markets, or sectors, or stocks in a sector, or exchange rates is often observed.
- Volatilities of different series move together.
- The flow of new information (trading volume, quote arrivals, dividend announcements, policy announcements) will affect the volatility of all the assets in the market/sector.

Session 7: Volatility Modelling

- Arbitrage pricing theory suggests indicates a relationship between the expected return of a stock and the covariance of the returns.
- Portfolio optimization theory and 'value at risk' theory, which studies the extreme downside of a portfolio of assets, both require knowledge of the joint distribution of the asset returns.

Session 7: Volatility Modelling

Examples of Multivariate ARCH models:

- the *diagonal vech* model,
- the *constant conditional correlation* model,
- the *factor-ARCH model*,
- the *BEKK model*
- the *latent factor ARCH model*

Most of these models highly parameterized, with constraints on the parameters needed to ensure symmetry and positivity are complicated and hard to interpret. The *dynamic conditional correlation GARCH model*, the *flexible multivariate GARCH model* and the *orthogonal GARCH* also have potential.

Session 7: Volatility Modelling

Far fewer multivariate models have been proposed in the stochastic volatility literature. The main reason for this is that stochastic volatility models are expressed in terms of normal log-volatilities, which in turn are hard to be extended to a multivariate counterpart.

- The most well known multivariate SV model overcomes this difficulty by defining vectors of log-volatilities which interact through a constant correlation structure.
- Another approach is to use *factor structures* which are more flexible and allow for a large number of assets to be modelled simultaneously.
- The *multivariate random walk* model has received significant attention
- The Wishart distribution is also a commonly used model basis.

Constant Correlation Model

This model does not allow the covariances between different time series to evolve independently of the variances; it can be considered as the stochastic volatility analogue to the constant conditional correlation ARCH model.

For all $t = 1, \ldots, T$ and for each $i = 1, \ldots, N$

$$\begin{aligned} y_{it} &= \exp(h_{it}/2)\varepsilon_{it}, \\ h_{it} &= \gamma_i + \phi_i h_{i\,t-1} + \eta_{it}, \end{aligned}$$

$\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})^{\mathsf{T}}$ and $\boldsymbol{\eta}_t = (\eta_{1t}, \ldots, \eta_{Nt})^{\mathsf{T}}$ are serially and mutually independent, both normal distributed with zero means, $Var(\boldsymbol{\eta}_t) = \boldsymbol{\Sigma}_\eta$.

Session 7: Volatility Modelling

$$Var\left(\varepsilon_t\right) = \mathbf{\Sigma}_\varepsilon = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1N} \\ \rho_{12} & 1 & & \rho_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1N} & \rho_{2N} & & 1 \end{bmatrix}$$

so that $\mathbf{\Sigma}_\varepsilon$ is a correlation matrix, i.e. $\left|\rho_{ij}\right| < 1$.

Therefore, the model can be written as:

$$\mathbf{y}_t|\mathbf{h}_t \sim N\left(\mathbf{0}, \mathbf{H}_t^{1/2}\mathbf{\Sigma}_\varepsilon\mathbf{H}_t^{1/2}\right),$$

$\mathbf{H}_t$ =diag$(\exp\left(h_{1t}\right), \ldots, \exp\left(h_{Nt}\right))$ and $\mathbf{\Sigma}_\varepsilon$ is given above.

Session 7: Volatility Modelling

The vector of log-volatilities $\mathbf{h}_t = (h_{1t}, \ldots, h_{Nt})^{\mathsf{T}}$ follows a vector-AR(1) (VAR(1)) process so that

$$\mathbf{h}_t = \boldsymbol{\gamma} + \boldsymbol{\Phi}\mathbf{h}_{t-1} + \boldsymbol{\eta}_t,$$

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_N)^{\mathsf{T}}$, $\boldsymbol{\Phi}$ =diag$(\phi_1, \ldots, \phi_N)$ and $\boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$.

The matrix $\boldsymbol{\Phi}$ is taken to be the identity matrix, and $\boldsymbol{\gamma}$ is equal to the N-dimensional zero vector, so that the log-volatilities follow a random walk and hence, the fact that volatility movements are persistent is stressed.

Session 7: Volatility Modelling

For $i \neq j$, the model has constant correlations as

$$Cov\left(y_{it}, y_{jt}|\mathbf{h}_t\right) = E\left(y_{it}^2 y_{jt}^2|\mathbf{h}_t\right) = \rho_{ij} \exp((h_{it} + h_{jt})/2)$$

- It is thought that the assumption of constant correlations may be empirically reasonable.
- The model can be written in a linear, non-Gaussian state space form, and thus can be easily estimated by QML

Denoting $\mathbf{w}_t = \left(\log y_{1t}^2, \ldots, \log y_{Nt}^2\right)^{\mathsf{T}}$ and by $\mathbf{1}$ the $N \times 1$ vector of ones,

$$\mathbf{w}_t = -1.2704 \times \mathbf{1} + \mathbf{h}_t + \boldsymbol{\xi}_t,$$

where $\boldsymbol{\xi}_t = \left(\log \varepsilon_{1t}^2, \ldots, \log \varepsilon_{Nt}^2\right)^{\mathsf{T}} + 1.2704 \times \mathbf{1}$ is a zero-mean vector, serially independent vector process whose covariance matrix is denoted $\boldsymbol{\Sigma}_\xi$.

Session 7: Volatility Modelling

Since Φ is a diagonal matrix, the state vector, $\mathbf{h}_t$, and the observation vector, $\mathbf{w}_t$, are linked together only through the off-diagonal elements of the covariance matrices Σ_η and Σ_ξ.

The linearized model resembles a *Seemingly Unrelated Time Series* (SUTSE) model, where the N time series, $w_{1t}, \ldots, w_{Nt}$, are modelled together not because they interact with each other, but because they are all subject to the same economic environment and thus are contemporaneously correlated.

Furthermore, it can be shown that the covariance matrix, Σ_ξ, of $\boldsymbol{\xi}_t$ can be expressed in terms of Σ_ε.

Session 7: Volatility Modelling

Let c_{ij} denote the ij-th element of $\mathbf{\Sigma}_\xi$. Since for $i = 1, \ldots, N$, $\xi_{it} = \log \varepsilon_{it}^2 + 1.2704$, the diagonal elements are all equal to $Var(\xi_{it})$, so that $c_{ii} = \pi^2/2$, while the off-diagonal elements are given by

$$c_{ij} = \sum_{n=1}^{\infty} \frac{(n-1)!}{n\,(1/2)_n} \rho_{ij}^{2n}, \quad i \neq j$$

where $(\alpha)_n = \alpha(\alpha+1)\cdots(\alpha+n-1)$.

Session 7: Volatility Modelling

Note: This transformation of the model incurs loss of information, since the correlation coefficients ρ_{ij} have been squared, all the relevant information on their sign has been lost. Thus, if the transformed observations are used to estimate the model, then we can recover the information on the magnitude of the original correlations ρ_{ij}, **but not their sign**.

A final feature of the model is that it can incorporate common factors and hence allow common trends and cycles in the volatility. In a common factor model each of the N elements of a multivariate time series are related to $K \leq N$ *common factors* via a linear function.

Session 7: Volatility Modelling

This model can be considered as a generalization of a SUTSE model where the N elements have certain common components. For the case where the log-volatilities follow a random walk, the model can be written as a common factor model, as follows

$$\begin{aligned} \mathbf{w}_t &= -1.2704 \times \mathbf{1} + \mathbf{B}\mathbf{h}_t^* + \bar{\mathbf{h}} + \boldsymbol{\xi}_t \\ \mathbf{h}_t^* &= \mathbf{h}_{t-1}^* + \boldsymbol{\eta}_t^*. \end{aligned}$$

$\mathbf{w}_t$ and $\boldsymbol{\xi}_t$ are defined as before, $\mathbf{B}$ is an $(N \times K)$ matrix of factor loadings, $\bar{\mathbf{h}}$ is an N-dimensional vector, while the second term is a K-dimensional random walk model with $\boldsymbol{\eta}_t^* \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}_\eta^*\right)$.

Session 7: Volatility Modelling

Unless restrictions are placed on the elements β_{ij} of $\mathbf{B}$ and on the vector $\bar{\mathbf{h}}$, the model above is not identifiable. The usual restrictions are to set $\beta_{ij} = 0$ for $j > i$ and $\mathbf{\Sigma}_\eta^*$ equal to the identity matrix, while the first K elements of the vector $\bar{\mathbf{h}}$ are restricted to be zero and the rest are unconstrained.

An appealing feature implied by these restrictions is that the common factors $\mathbf{h}_t^*$ are uncorrelated with each other and this simplifies estimation and the theoretical properties of the model. Conversely, the restrictions placed on the matrix $\mathbf{B}$ are displeasing, in the sense that they imply that the first observation w_{1t} depends only on the first factor, w_{2t} depends only on the first two factors and only the last K observations, $w_{k\,t}, \ldots, w_{N\,t}$ are related to all of the common factors.

Session 7: Volatility Modelling

- Depending on the ordering for the N series in order to model them, we will get different relations between the N observations and the K factors; this is an unappealing feature.
- In the latter case once estimation of the model has been completed, using an orthogonal matrix $\mathbf{R}$ to give a factor rotation and a more useful interpretation to the factors.
- Finally, the idea of common factors can also be extended to stationary models of the log-volatilities. For a stationary model, we allow $\mathbf{h}_t^*$ in to follow a stationary VAR(1) process and this also removes the need to include $\bar{\mathbf{h}}$ in the observation equation).

Factor SV models

The basic advantage of *factor models* is that they reduce the dimensionality of the parameter space; a factor model, with unobserved latent ARCH factors has also been proposed.

Multivariate factor SV models resemble this model, in the sense that the returns $\mathbf{y}_t$ from N series are linear functions of K unobserved latent factors, where each one is driven by a SV model.

The difference between the common factor model is that in this case the common factors are incorporated for the untransformed observations.

Session 7: Volatility Modelling

The simplest factor model uses factors $\mathbf{f}_t = (f_{1t}, \ldots, f_{Kt})^{\mathsf{T}}$ the vector of K common factors. Then

$$\begin{aligned} \mathbf{y}_t &= \mathbf{B}\mathbf{f}_t + \boldsymbol{\omega}_t, \\ f_{it} &= \exp(h_{it}/2)\,\varepsilon_{it}, \\ h_{it} &= \mu_i + \phi_i(h_{it-1} - \mu_i) + \eta_{it}\ , \quad i = 1, \ldots, K. \end{aligned}$$

In the factor SV model, the matrix $\mathbf{B}$ is a constant $(N \times K)$ matrix of factor loadings with $K < N$, the errors $\boldsymbol{\omega}_t \sim N(\mathbf{0}, \boldsymbol{\Omega})$ and are serially and mutually independent of all the other terms.

In the model K factors f_{it} are introduced with each one modelled as an independent SV process.

Session 7: Volatility Modelling

Clearly, since $\mathbf{f}_t$ and $\boldsymbol{\omega}_t$ are independent, zero mean processes, we have $E(\mathbf{y}_t) = 0$ and

$$Var\left(\mathbf{y}_t|\mathbf{h}_t\right) = \mathbf{B}\mathbf{H}_t\mathbf{B}^{\mathsf{T}} + \boldsymbol{\Omega},$$

where

$$\mathbf{H}_t = \text{diag}\left(\exp\left(h_{1t}\right), \ldots, \exp\left(h_{K\,t}\right)\right).$$

MCMC methods can be used to estimate the model, as the QML method is not effective in this case. However, the method has not been implemented and another more complicated multivariate factor SV model is usually preferred.

Session 7: Volatility Modelling

The second factor SV model allows the errors $\boldsymbol{\omega}_t = (\omega_{1t}, \ldots, \omega_{Nt})$ to be driven by independent SV processes also. That is,

$$\begin{aligned} \omega_{jt} &= \exp(\alpha_{jt}/2)\, e_{jt}, \\ \alpha_{jt} &= m_j + \delta_j(\alpha_{jt-1} - m_j) + u_{jt}, \quad j = 1, \ldots, N. \end{aligned}$$

As before, the e_{jt} are i.i.d $N(0,1)$ random variables with no serial dependence and the u_{jt} are i.i.d $N(0, \sigma^2_{uj})$ serially and mutually independent of the e_{jt}. As usual we place the constraint $|\delta_j| < 1$ to ensure the stationarity of the log-volatility processes.

Session 7: Volatility Modelling

The intuition behind this extension of the model is that the K factors can account for the off-diagonal elements of the covariance matrix of the returns but cannot account for all the marginal persistence in the volatility.

Loosely speaking the volatility resulting from the fact that the N series are subject to the same economic environment is explained by the volatility of the common factors $\mathbf{f}_t$, while each of the idiosyncratic errors, or series-specific shocks $\boldsymbol{\omega}_t$ account for the volatility associated with a particular series.

The inclusion of SV effects in the errors $\boldsymbol{\omega}_t$ makes the model more robust, in the sense that if an unusual return is observed for a specific series, then this can be attributed to the idiosyncratic error associated with that particular series.

Session 7: Volatility Modelling

Identifiability: As with the common factors model the parameters of the model are not identifiable unless constraints are placed on the elements β_{ij} of the matrix $\mathbf{B}$. There are two ways that this can be done.

- set $\beta_{ii} = 1$ for $i = 1, \ldots, K$ and $\beta_{ij} = 0$ for $j > i$.
- set $\beta_{ii} = 1$ for $i = 1, \ldots, K$, while

$$\beta_{i(i+1)} = 0, \ i = 1, \ldots, K-1$$

In this case the first series of returns depends in all but the second common factor, the second series is influenced by all the factors apart from the third and so on, while only the last $N - K + 1$ series depend on all the factors.

In both cases, an optimal factor rotation can be used to give a more meaningful interpretation to the factors.

Session 7: Volatility Modelling

Denote by $\mathbf{h}_t$ the vector $(h_{1t}, \ldots, h_{Kt})^{\mathsf{T}}$ of log-volatilities of the factors, while $\boldsymbol{\alpha}_t = (\alpha_{1t}, \ldots, \alpha_{Nt})^{\mathsf{T}}$.

Since the errors ε_{it} $(i = 1, \ldots, K)$and e_{jt} $(j = 1, \ldots, N)$ are zero mean, we have

$$E(\mathbf{f}_t) = E(\boldsymbol{\omega}_t) = E(\mathbf{f}_t|\mathbf{h}_t) = E(\boldsymbol{\omega}_t|\boldsymbol{\alpha}_t) = \mathbf{0}.$$

Since the model for the returns is

$$\mathbf{y}_t = \mathbf{B}\mathbf{f}_t + \boldsymbol{\omega}_t,$$

both the unconditional and conditional mean of $\mathbf{y}_t$ are equal to the zero vector.

Session 7: Volatility Modelling

Since each of the elements of $\mathbf{f}_t$ and $\boldsymbol{\omega}_t$ follow mutually independent SV processes,

$$\left(\begin{array}{c} \mathbf{f}_t \\ \boldsymbol{\omega}_t \end{array} \right) \Big| \mathbf{h}_t, \boldsymbol{\alpha}_t \sim NID \left(\mathbf{0}, \left[\begin{array}{cc} \mathbf{H}_t & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_t \end{array} \right] \right),$$

where

$$\begin{aligned} \mathbf{H}_t &= \text{diag}\left(\exp\left(h_{1t}\right), \ldots, \exp\left(h_{Kt}\right)\right) \\ \boldsymbol{\Omega}_t &= \text{diag}\left(\exp\left(\alpha_{1t}\right), \ldots, \exp\left(\alpha_{Nt}\right)\right) \end{aligned}$$

where *NID* denotes normally and serially independently distributed. Since $\mathbf{f}_t$ is independent of $\boldsymbol{\omega}_t$ and both are normally distributed,

$$\mathbf{y}_t | \mathbf{h}_t, \boldsymbol{\omega}_t \sim NID\left(\mathbf{0}, \mathbf{B}\mathbf{H}_t\mathbf{B}^{\mathsf{T}} + \boldsymbol{\Omega}_t\right).$$

Session 7: Volatility Modelling

Expanding the matrix multiplications it follows that

$$\begin{aligned} Var\left(y_{it}|\mathbf{h}_t,\boldsymbol{\omega}_t\right) &= \beta_{i1}^2\exp\left(h_{1t}\right)+\cdots+\beta_{iK}^2\exp\left(h_{K\,t}\right)+\exp\left(\alpha_{it}\right) \\ Cov\left(y_{it},y_{jt}|\mathbf{h}_t,\boldsymbol{\omega}_t\right) &= \beta_{i1}\beta_{j1}\exp\left(h_{1t}\right)+\cdots+\beta_{iK}\beta_{jK}\exp\left(h_{K\,t}\right), \end{aligned}$$

The evolution of the covariances in the factor SV model depends on the variances but in a very complicated way.

The unconditional covariance matrix of $\mathbf{y}_t$ is deduced using the properties of log-normal distributions.

From the properties of the AR processes, we know that

$$h_{it}\sim N\left(\mu_i,\sigma_{\eta\,i}^2/\left(1-\phi_i^2\right)\right) \qquad \alpha_{jt}\sim N\left(m_j,\sigma_{u\,j}^2/\left(1-\delta_i^2\right)\right).$$

Session 7: Volatility Modelling

$E(f_{it}) = E(\omega_{jt}) = 0$ while

$$\begin{aligned} \sigma^2_{f_i} &= Var(f_{it}) = \exp\left(\mu_i + \frac{\sigma^2_{\eta i}}{2(1-\phi_i^2)}\right) \\ \sigma^2_{\omega_j} &= Var\left(m_j + \frac{\sigma^2_{uj}}{2(1-\delta_j^2)}\right). \end{aligned}$$

Therefore,

$$Var(\mathbf{y}_t) = \mathbf{BHB}^{\mathsf{T}} + \mathbf{\Omega}$$

where $\mathbf{H} = \text{diag}\left(\sigma^2_{f_1}, \ldots, \sigma^2_{f_K}\right)$ and $\mathbf{\Omega} = \text{diag}\left(\sigma^2_{\omega_1}, \ldots, \sigma^2_{\omega_N}\right)$.

Session 7: Volatility Modelling

More complicated dynamics could be introduced in the latent SV processes. For example, the errors $\mathbf{e}_t$ in the idiosyncratic shocks can be allowed to be correlated, so that $\mathbf{e}_t \sim NID(\mathbf{0}, \boldsymbol{\Sigma}_e)$.

Alternatively, consider a factor SV model where the errors in the latent AR processes of the log-volatilities $\mathbf{h}_t$ are contemporaneously correlated, so that

$$\mathbf{h}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\eta}_t,$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)^{\mathsf{T}}$, $\boldsymbol{\Phi}$ =diag$(\phi_1, \ldots, \phi_K)$ and $\boldsymbol{\eta}_t \sim NID(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$.

Session 7: Volatility Modelling

- The model can be extended in two ways: by incorporating Student-t distributed errors ω_t with stochastic volatility and allowing the series of returns to have jumps.
- The estimation of this model is performed using MCMC methods, but the number of factors is user-controlled and is not estimated from the data. If the number of factors is too large, then the MCMC behaves poorly and does not converge; if the factors are too few, there are co-movements in the estimated variances of the idiosyncratic errors and the variances of one or more factors.

Multivariate SV models based on Wishart processes

One of the first multivariate SV models exploits the conjugacy between the Wishart distribution and the *matrix variate beta distribution* and allows the variance of the observations to evolve according to a *matrix random walk*.

An attractive feature of the model is that it leads to closed form prediction and update equations that are generalizations of the standard Kalman filter recursions.

Session 7: Volatility Modelling

Stationarity: Related results have become popular in the literature of matrix normal dynamic linear models, but the model has never been applied in a SV setting for financial time series; the conjugacy between the Wishart and the matrix variate beta distribution cannot be sustained if we try to move away from the matrix random walk formulation and impose stationarity to the process generating the volatility matrix.

From a financial point of view imposing non-stationarity on the volatility matrix is an unattractive assumption.

Session 7: Volatility Modelling

Model Description: If $U \sim B_n(\nu/2, m/2)$ with $\nu, m > n-1$ and $V = R^{\mathsf{T}}R \sim W_n(\nu+m, \Sigma)$ for some positive definite symmetric $(n \times n)$ matrix Σ and for an upper-triangular matrix R with positive diagonal elements, then

$$S = R^{\mathsf{T}}UR \sim W_n(\nu, \Sigma).$$

Note also that the matrix beta distribution and its density in are only defined for $\nu, m > n-1$ and is singular otherwise.

An extension to this definition of the matrix variate beta distribution for positive integers $m > 0$, and the conjugacy result still holds for $\nu > n-1$ and integer-valued $m > 0$. In particular, the density of a singular $B_n(\nu/2, 1/2)$ -distributed random matrix can be found.

Session 7: Volatility Modelling

Denote by $\mathbf{y}_t = (y_{1t}, \ldots, y_{Nt})$ the vector of one-period returns from N assets at time t and by $\mathbf{\Sigma}_t$ the $(N \times N)$ covariance matrix of $\mathbf{y}_t$.

In addition $\mathbf{T}_t$ denotes the Cholesky factor of $\mathbf{\Sigma}_t$, i.e. the upper-triangular matrix with positive diagonal elements such that $\mathbf{T}_t^{\mathsf{T}}\mathbf{T}_t = \mathbf{\Sigma}_t$. Then

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{T}_t^{-1}\varepsilon_t, \quad \varepsilon_t \sim NID\left(\mathbf{0}, \mathbf{I}_N\right) \\
\mathbf{\Sigma}_t &= \frac{\nu+1}{\nu}\mathbf{T}_{t-1}^{\mathsf{T}}\mathbf{Q}_t\mathbf{T}_{t-1} \\
\mathbf{Q}_t &\sim B_N\left(\frac{\nu}{2}, \frac{1}{2}\right), \quad t = 1, \ldots, T.
\end{aligned}$$

Session 7: Volatility Modelling

The random variables $\mathbf{Q}_t$ are serially and mutually independent of all other variables in the model and the initial condition

$$\boldsymbol{\Sigma}_0 \sim W_N\left(\nu+1, (\nu+1)^{-1}\mathbf{S}_0^{-1}\right)$$

for some $(N \times N)$ matrix $\mathbf{S}_0$ (positive definite and symmetric).

Note that

$$\mathbf{y}_t|\boldsymbol{\Sigma}_t \sim NID\left(\mathbf{0}, \boldsymbol{\Sigma}_t^{-1}\right).$$

Using the conjugacy between the Wishart and matrix beta distributions it follows that

$$\mathbf{T}_0^{\mathsf{T}}\mathbf{Q}_1\mathbf{T}_0 \sim W_N\left(\nu, (\nu+1)^{-1}\mathbf{S}_0^{-1}\right),$$

so that prior to observing $\mathbf{y}_1$, $\boldsymbol{\Sigma}_1 \sim W_N\left(\nu, \nu^{-1}\mathbf{S}_0^{-1}\right)$ with $E\left(\boldsymbol{\Sigma}_1\right) = \mathbf{S}_0^{-1}$.

Session 7: Volatility Modelling

Once the first observation $\mathbf{y}_1$ becomes available, the posterior for $\boldsymbol{\Sigma}_1$ is a Wishart $W_N\left(\nu+1,(\nu+1)^{-1}\mathbf{S}_1^{-1}\right)$, with

$$\mathbf{S}_1 = \frac{1}{\nu+1}\left(\mathbf{y}_1\mathbf{y}_1^{\mathsf{T}} + \nu\mathbf{S}_0\right)$$

with

$$E\left(\boldsymbol{\Sigma}_1|\mathbf{y}_1\right) = \mathbf{S}_1^{-1}.$$

This posterior distribution of $\boldsymbol{\Sigma}_1$ suggests a prior $W_N\left(\nu,\nu^{-1}\mathbf{S}_1^{-1}\right)$ distribution for $\boldsymbol{\Sigma}_2$, so that

$$E\left(\boldsymbol{\Sigma}_2|\mathbf{y}_1\right) = \mathbf{S}_1^{-1} = E\left(\boldsymbol{\Sigma}_1|\mathbf{y}_1\right)$$

.

Session 7: Volatility Modelling

The filtered distribution for Σ_2 has the same form with that of Σ_1; an induction argument shows that at each time t, the prior for Σ_t is a $W_N\left(\nu, \nu^{-1}\mathbf{S}_{t-1}^{-1}\right)$ distribution, while the posterior is $W_N\left(\nu+1, (\nu+1)^{-1}\mathbf{S}_t^{-1}\right)$, where

$$\mathbf{S}_t = \left(\mathbf{y}_t\mathbf{y}_t^{\mathsf{T}} + \nu\mathbf{S}_{t-1}\right)/(\nu+1).$$

This filtering algorithm delivers the one-step-ahead prediction distributions for $\mathbf{y}_t$ and hence the estimation of the model by maximum likelihood is is principle straightforward.

Session 7: Volatility Modelling

For a more general model denote by $\mathbf{\Sigma}_t$ the $(N \times N)$ covariance matrix of the returns at time t. We also write $\mathbf{\Sigma}_t = \mathbf{T}_t^{\mathsf{T}} \mathbf{T}_t$ for the Cholesky decomposition of $\mathbf{\Sigma}_t$.

Then, the new Wishart model sets for $t = 1, \ldots, T$

$$\mathbf{y}_t = \mathbf{T}_t^{\mathsf{T}} \varepsilon_t, \quad \varepsilon_t \sim NID\left(\mathbf{0}, \mathbf{I}_N\right)$$

so that $\mathbf{y}_t | \mathbf{\Sigma}_t \sim N\left(\mathbf{0}, \mathbf{\Sigma}_t\right)$ and hence the density of the returns at time t given the covariance structure is

$$p\left(\mathbf{y}_t | \mathbf{\Sigma}_t\right) = \frac{1}{(2\pi)^{\frac{N}{2}} \det\left(\mathbf{\Sigma}_t\right)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} \mathbf{y}_t^{\mathsf{T}} \mathbf{\Sigma}_t^{-1} \mathbf{y}_t\right\}.$$

Session 7: Volatility Modelling

The evolution of the covariance matrices $\mathbf{\Sigma}_t$ is specified through one-step-ahead prediction densities, specifically, the density of the inverse of the current covariance structure $\mathbf{\Sigma}_t$ conditional on $\mathbf{\Sigma}_{t-1}$.

In particular,

$$\mathbf{\Sigma}_t^{-1} | \mathbf{\Sigma}_{t-1} \sim W_N\left(\nu, \mathbf{S}_{t-1}\right),$$

where $\nu \geq N$. To define the scale matrix $\mathbf{S}_{t-1}$ of the Wishart distribution, we introduce a positive definite symmetric $(N \times N)$ parameter matrix $\mathbf{A}$, with Cholesky factor $\mathbf{R}$, i.e. $\mathbf{A} = \mathbf{R}^{\mathsf{T}}\mathbf{R}$ and a scalar parameter d.

Then, for $t = 0, \ldots, T-1$, $\mathbf{S}_t$ is defined as follows

$$\mathbf{S}_t = \frac{1}{\nu}\mathbf{R}^{\mathsf{T}} \left(\boldsymbol{\Sigma}_t^{-1}\right)^d \mathbf{R},$$

where we assume a known initial condition $\boldsymbol{\Sigma}_0 > 0$ for the covariance matrix. Thus,

$$\boldsymbol{\Sigma}_t^{-1} | \boldsymbol{\Sigma}_{t-1} \sim W_N\left(\nu, \frac{1}{\nu}\mathbf{R}^{\mathsf{T}} \left(\boldsymbol{\Sigma}_{t-1}^{-1}\right)^d \mathbf{R}\right)$$

and hence the conditional density of $\boldsymbol{\Sigma}_t^{-1}$ is available in straightforward form. The latter distribution defines the evolution of the covariance matrices.

Session 7: Volatility Modelling

The time-variation of the covariances can be expressed in terms of $\boldsymbol{\Sigma}_t$. Since $\boldsymbol{\Sigma}_t^{-1}|\boldsymbol{\Sigma}_{t-1}$ is $W_N(\nu, \mathbf{S}_{t-1})$-distributed, it follows that

$$\begin{aligned}\boldsymbol{\Sigma}_t|\boldsymbol{\Sigma}_{t-1} &\sim IW_N\left(\nu+N+1, \mathbf{S}_{t-1}^{-1}\right)\\ &\equiv IW_N\left(\nu+N+1, \nu\left(\mathbf{R}\right)^{-1}\left(\boldsymbol{\Sigma}_{t-1}\right)^d\left(\mathbf{R}^{\mathsf{T}}\right)^{-1}\right).\end{aligned}$$

Using the properties of the Wishart/inverse Wishart distributions

$$\begin{aligned}E\left(\boldsymbol{\Sigma}_t^{-1}|\boldsymbol{\Sigma}_{t-1}\right) &= \nu\mathbf{S}_{t-1} = \mathbf{R}^{\mathsf{T}}\left(\boldsymbol{\Sigma}_{t-1}^{-1}\right)^d\mathbf{R} \quad \text{and}\\ E\left(\boldsymbol{\Sigma}_t|\boldsymbol{\Sigma}_{t-1}\right) &= \frac{1}{\nu-N-1}\mathbf{S}_{t-1}^{-1}.\end{aligned}$$

Hyperparameters:

- the elements of $\mathbf{A}^{-1}$ determine the dependence of the current variances and covariances of the returns depend on the previous periods corresponding values.
- the scalar d defines the strength of the relationship between the current values of the volatilities and covariances of the assets and those in the previous period, like a persistence parameter.

Session 7: Volatility Modelling

A value of d close to zero implies small persistence in the covariance structure; if $d = 0$,

$$\begin{aligned} E\left(\boldsymbol{\Sigma}_t|\boldsymbol{\Sigma}_{t-1}\right) &= \frac{\nu}{\nu - N - 1}\left(\mathbf{R}^{\mathsf{T}}\mathbf{R}\right)^{-1} = \frac{\nu}{\nu - N - 1}\mathbf{A}^{-1} \\ E\left(\boldsymbol{\Sigma}_t^{-1}|\boldsymbol{\Sigma}_{t-1}\right) &= \mathbf{A}. \end{aligned}$$

that is, the conditional expectation of the current covariance structure does not depend on the previous period.

The only values of d that result in the process generating the conditional covariance matrix being stationary indeed lie in the interval $(-1, 1)$.

The Wishart Autoregressive process

The idea underlying the WAR(1) processes relies on the fact that if a $(p \times n)$ matrix $\mathbf{X}$ is formed by setting its columns to be $n \geq p$ random draws from a multivariate $N(\mathbf{0}, \mathbf{\Sigma})$ distribution, where the $(p \times p)$ matrix $\mathbf{\Sigma} > 0$, then the random variable

$$\mathbf{X}\mathbf{X}^{\mathrm{T}} \sim W_p(n, \mathbf{\Sigma})$$

To introduce intertemporal dependence between the covariance matrices formed in this way, the columns of the matrix $\mathbf{X}$ are presumed to be observations from a latent Gaussian VAR(1) process.

As before let $\mathbf{\Sigma}_t$ be the covariance structure of the returns-vector $\mathbf{y}_t$ from N assets. Consider now, $\nu > N-1$ identical, N-dimensional, Gaussian VAR(1) processes $\mathbf{x}_{1t}, \ldots, \mathbf{x}_{\nu t}$, where each $\mathbf{x}_{it}$ $(i = 1, \ldots, \nu)$ is generated for $t = 1, \ldots, T$ from

$$\mathbf{x}_{it} = \mathbf{\Phi}\mathbf{x}_{it-1} + \boldsymbol{\eta}_{it} \ , \quad \boldsymbol{\eta}_{it} \sim NID\left(\mathbf{0}, \mathbf{\Sigma}_{\eta}\right),$$

where $\mathbf{\Phi}$ is a $(N \times N)$ matrix.

The VAR(1) process is stationary if and only if all the eigenvalues λ_i $(i = 1, \ldots, N)$ of the matrix $\mathbf{\Phi}$ lie inside the unit circle; the ν VAR(1) processes $\mathbf{x}_{1t}, \ldots, \mathbf{x}_{\nu t}$ are stationary, and the unconditional distribution of all of the $\mathbf{x}_{it}$ $(i = 1, \ldots, \nu)$ is multivariate normal.

The unconditional mean of each $\mathbf{x}_{it}$ is the zero vector while the unconditional variance $Var(\mathbf{x}_{it}) = \mathbf{S}$ is given by

$$\mathbf{S} = \mathbf{\Phi}\mathbf{S}\mathbf{\Phi}^{\mathsf{T}} + \mathbf{\Sigma}_{\eta}.$$

Applying the $vec(\cdot)$ operator on both sides, it is seen that $vec(\mathbf{S}) = (\mathbf{I}_{N^2} - \mathbf{\Phi} \otimes \mathbf{\Phi})\, vec(\mathbf{\Sigma}_{\eta})$. In short, the VAR(1) processes $\mathbf{x}_{1t}, \ldots, \mathbf{x}_{\nu t}$ are i.i.d. $N(\mathbf{0}, \mathbf{S})$ -distributed variables.

The ν processes can be represented as a matrix normal linear model

$$\mathbf{X}_t = \mathbf{\Phi}\mathbf{X}_{t-1} + \mathbf{H}_t,$$

where the $(N \times \nu)$ matrices $\mathbf{X}_t$ and $\mathbf{H}_t$ are defined as

$$\begin{aligned} \mathbf{X}_t &= [\mathbf{x}_{1t}, \ldots, \mathbf{x}_{\nu t}] \\ \mathbf{H}_t &= [\boldsymbol{\eta}_{1t}, \ldots, \boldsymbol{\eta}_{\nu t}] \end{aligned}$$

Since the vector $\boldsymbol{\eta}_{it}$ $(i = 1, \ldots, \nu)$ are i.i.d $N(\mathbf{0}, \mathbf{\Sigma}_\eta)$ we have

$$\mathbf{H}_t \sim N_{N,\nu}(\mathbf{0}, \mathbf{\Sigma}_\eta \otimes \mathbf{I}_\nu).$$

Most importantly, using the unconditional distributions of the ν VAR(1) processes, we also have that

$$\mathbf{X}_t \sim N_{N,\nu}(\mathbf{0}, \mathbf{S} \otimes \mathbf{I}_\nu).$$

Therefore if

$$\boldsymbol{\Sigma}_t = \mathbf{X}_t \mathbf{X}_t^{\mathsf{T}} \qquad t = 1, \ldots, T$$

then $\boldsymbol{\Sigma}_t > 0$ with probability one and the unconditional distribution of the covariance matrix at each time t is

$$\boldsymbol{\Sigma}_t \sim W_N(\nu, \mathbf{S}) \qquad t = 1, \ldots, T$$

and so the unconditional density is

$$p(\boldsymbol{\Sigma}_t) = \frac{1}{2^{\frac{\nu N}{2}} \Gamma_N\left(\frac{\nu}{2}\right)} \frac{\det(\boldsymbol{\Sigma}_t)^{\frac{(\nu - N - 1)}{2}}}{\det(\mathbf{S})^{\frac{\nu}{2}}} \exp\left\{-\frac{1}{2} tr\left(\mathbf{S}^{-1} \boldsymbol{\Sigma}_t\right)\right\}.$$

Moreover, using the properties of the Wishart distribution, $E(\boldsymbol{\Sigma}_t) = \nu \mathbf{S}$.

Session 7: Volatility Modelling

The distribution of $\mathbf{\Sigma}_t$ conditional on $\mathbf{\Sigma}_{t-1}$ is computed by considering the conditional distribution of the latent VAR(1) processes contained in $\mathbf{X}_t$ conditional on $\mathbf{X}_{t-1}$.

Since $\mathbf{H}_t \sim N_{N,\nu}(\mathbf{0}, \mathbf{\Sigma}_\eta \otimes \mathbf{I}_\nu)$ using the dynamic linear model representation,

$$\mathbf{X}_t|\mathbf{X}_{t-1} \sim N_{N,\nu}(\mathbf{\Phi}\mathbf{X}_{t-1}, \mathbf{\Sigma}_\eta \otimes \mathbf{I}_\nu).$$

We have that the distribution of $\mathbf{\Sigma}_t$ conditional on $\mathbf{X}_{t-1}$ is a non-central Wishart distribution

$$\mathbf{\Sigma}_t|\mathbf{X}_{t-1} \sim W_N\left(\nu, \mathbf{\Sigma}_\eta, \mathbf{\Sigma}_\eta^{-1}\mathbf{\Phi}\mathbf{X}_{t-1}\mathbf{X}_{t-1}^{\mathsf{T}}\mathbf{\Phi}^{\mathsf{T}}\right)$$

The conditional distribution of $\mathbf{\Sigma}_t$ depends on $\mathbf{X}_{t-1}$ only through the product

$$\mathbf{X}_{t-1}\mathbf{X}_{t-1}^{\mathsf{T}} = \mathbf{\Sigma}_{t-1},$$

which implies that the process is Markov. That is, the current covariance matrix $\mathbf{\Sigma}_t$ depends only on the previous period covariance matrix $\mathbf{\Sigma}_{t-1}$ and not on the latent VAR(1) processes contained in $\mathbf{X}_{t-1}$.

In particular, the conditional density of $\boldsymbol{\Sigma}_t$ is given by

$$\begin{aligned} p\left(\boldsymbol{\Sigma}_t|\boldsymbol{\Sigma}_{t-1}\right) &= \frac{1}{2^{\frac{\nu N}{2}}\Gamma_N\left(\frac{\nu}{2}\right)}\frac{\det\left(\boldsymbol{\Sigma}_t\right)^{\frac{(\nu-N-1)}{2}}}{\det\left(\boldsymbol{\Sigma}_\eta\right)^{\frac{\nu}{2}}} \times \\ &\exp\left\{-\frac{1}{2}tr\left[\boldsymbol{\Sigma}_\eta^{-1}\left(\boldsymbol{\Sigma}_t+\boldsymbol{\Phi}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\Phi}^{\mathsf{T}}\right)\right]\right\} \times \\ &{}_0F_1\left(\frac{\nu}{2};\frac{1}{4}\boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Phi}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\Phi}^{\mathsf{T}}\boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Sigma}_t\right). \end{aligned}$$

where ${}_0F_1$ is the hypergeometric function of matrix argument. It follows that

$$E\left(\boldsymbol{\Sigma}_t|\boldsymbol{\Sigma}_{t-1}\right) = \nu\boldsymbol{\Sigma}_\eta + \boldsymbol{\Phi}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\Phi}^{\mathsf{T}}.$$

Session 7: Volatility Modelling

The fact that the WAR(1) process is Markov is important because it implies that the underlying VAR(1) processes can be ignored, and define the process generating the covariance structures through the unconditional and conditional distributions of $\boldsymbol{\Sigma}_t$, as in the previous model. That is, we may specify

$$\begin{aligned} \boldsymbol{\Sigma}_t|\boldsymbol{\Sigma}_{t-1} &\sim W_N\left(\nu, \boldsymbol{\Sigma}_\eta, \boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Phi}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\Phi}^{\mathsf{T}}\right) \quad \text{with} \\ \boldsymbol{\Sigma}_t &\sim W_N\left(\nu, \mathbf{S}\right) \ , \quad \text{where} \quad \mathbf{S} = \boldsymbol{\Sigma}_\eta + \boldsymbol{\Phi}\mathbf{S}\boldsymbol{\Phi}^{\mathsf{T}}. \end{aligned}$$

This allows us also to consider non-integer degrees of freedom $\nu > N-1$, since both the Wishart and the non-central Wishart distributions can be defined for any $\nu > N-1$.

In this case the WAR(1) process which generates the covariance structures loses its interpretation in terms of the VAR(1) processes.

The returns $\mathbf{y}_t$ can be made to depend on the covariance $\mathbf{\Sigma}_t$ generated by the WAR(1) process in the following ways

- set
$$\mathbf{y}_t|\mathbf{\Sigma}_t \sim N(\mathbf{0}, \mathbf{\Sigma}_t).$$
- allow the returns to depend on the inverse $\mathbf{\Sigma}_t^{-1}$, that is
$$\mathbf{y}_t|\mathbf{\Sigma}_t \sim N\left(\mathbf{0}, \mathbf{\Sigma}_t^{-1}\right).$$

Since the unconditional distribution of $\mathbf{\Sigma}_t$ is $W_N(\nu, \mathbf{S})$ it follows, using the relation between the Wishart and the inverse Wishart distributions that

$$\mathbf{\Sigma}_t^{-1} \sim IW_N\left(\nu + N + 1, \mathbf{S}^{-1}\right)$$

Therefore, since

$$\mathbf{y}_t | \mathbf{\Sigma}_t \sim N\left(\mathbf{0}, \mathbf{\Sigma}_t^{-1}\right)$$

and, marginally

$$\mathbf{\Sigma}_t^{-1} \sim IW_N\left(\nu + N + 1, \mathbf{S}^{-1}\right)$$

it follows that the marginal distribution of $\mathbf{y}_t$ is a multivariate t-distribution, denoted

$$t_p\left(\nu - N + 1, 1, \mathbf{0}, \mathbf{S}^{-1}\right)$$

Session 7: Volatility Modelling

This latter result is interesting from a modelling point of view; it is known that the distribution of the returns $\mathbf{y}_t$ has fatter tails than the normal distribution and that the kurtosis arising by accounting for heteroscedasticity only cannot account for all of the excess kurtosis.

This is the reason that most researchers consider extending the standard SV model by allowing the observation errors to have fat-tailed distributions. Recently, there have been attempts to construct stationary time series which account for heteroscedasticity and have specified marginal distributions with heavy tails. Therefore, it is possible that using the above distribution for the returns vector, we might be able to get a better fit to actual financial returns data.