

6.8 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.

6.8 Multivariate Stochastic Volatility Models

- ▶ 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.

6.8 Multivariate Stochastic Volatility Models

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.

6.8 Multivariate Stochastic Volatility Models

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 $t = 1, \dots, n$ and for each $i = 1, \dots, N$

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

$\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})^\top$ and $\eta_t = (\eta_{1t}, \dots, \eta_{Nt})^\top$ are serially and mutually independent, both normal distributed with zero means, $\mathbb{V}ar[\eta_t] = \Sigma_\eta$.

Constant Correlation Model

$$\text{Var}[\boldsymbol{\varepsilon}_t] = \boldsymbol{\Sigma}_\varepsilon = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1N} \\ \rho_{12} & 1 & & \rho_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1N} & \rho_{2N} & & 1 \end{bmatrix}$$

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

Therefore, the model can be written as:

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

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

Constant Correlation Model

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

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

where $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_N)^\top$, $\boldsymbol{\Phi} = \text{diag}(\phi_1, \dots, \phi_N)$ and $\boldsymbol{\eta}_t \sim \mathcal{N}_N(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$.

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.

Constant Correlation Model

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

$$\text{Cov}[Y_{it}, Y_{jt} | \mathbf{h}_t] = \mathbb{E}[Y_{it}^2 Y_{jt}^2 | \mathbf{h}_t] = \rho_{ij} \exp((h_{it} + h_{jt})/2)$$

- ▶ 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 $\mathbf{w}_t = (\log y_{1t}^2, \dots, \log y_{Nt}^2)^\top$ and by $\mathbf{1}$ the $N \times 1$ vector of ones,

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

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

Constant Correlation Model

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, $w_{1t}, \dots, w_{Nt}$, 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 ξ_t can be expressed in terms of Σ_ε .

Constant Correlation Model

Let c_{ij} denote the ij -th element of Σ_ξ . Since for $i = 1, \dots, N$, $\xi_{it} = \log \varepsilon_{it}^2 + 1.2704$, the diagonal elements are all equal to $\text{Var} [\xi_{it}]$, 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 $(\alpha)_n = \alpha (\alpha + 1) \cdots (\alpha + n - 1)$.

Constant Correlation Model

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.

Constant Correlation Model

This model is 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

$$\begin{aligned}w_t &= -1.2704\mathbf{1} + \mathbf{B}h_t^* + \bar{h} + \boldsymbol{\xi}_t \\h_t^* &= h_{t-1}^* + \boldsymbol{\eta}_t^*\end{aligned}$$

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

Constant Correlation Model

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 Σ_{η}^* 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 which simplifies estimation

However, the restrictions placed on $\mathbf{B}$ are displeasing, as 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, $w_{kt}, \dots, w_{Nt}$ are related to all of the common factors.

Constant Correlation Model

- ▶ 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.

Factor SV models

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

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

In the factor SV model, the matrix $\mathbf{B}$ is a constant ($N \times K$) matrix of factor loadings with $K < N$, the errors $\boldsymbol{\omega}_t \sim \mathcal{N}_N(\mathbf{0}, \boldsymbol{\Omega})$ 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.

Clearly, since f_t and ω_t are independent, zero mean processes, we have $\mathbb{E}[Y_t] = \mathbf{0}$ and

$$\text{Var}[Y_t|h_t] = BH_tB^\top + \Omega,$$

where

$$H_t = \text{diag}(\exp(h_{1t}), \dots, \exp(h_{Kt})).$$

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.

Factor SV models

The second factor SV model allows the errors

$$\boldsymbol{\omega}_t = (\omega_{1t}, \dots, \omega_{Nt})$$

to be driven by independent SV processes also. That is,

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

As before, the e_{jt} are i.i.d $\mathcal{N}(0, 1)$ random variables with no serial dependence and the u_{jt} are i.i.d $\mathcal{N}(0, \sigma_{uj}^2)$ 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.

Factor SV models

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.

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 f_t , while each of the idiosyncratic errors, or series-specific shocks ω_t account for the volatility associated with a particular series.

The inclusion of SV effects in the errors ω_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.

Factor SV models

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

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

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.

Factor SV models

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

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

$$\mathbb{E}[\mathbf{f}_t] = \mathbb{E}[\boldsymbol{\omega}_t] = \mathbb{E}[\mathbf{f}_t | \mathbf{h}_t] = \mathbb{E}[\boldsymbol{\omega}_t | \boldsymbol{\alpha}_t] = \mathbf{0}.$$

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.

Factor SV models

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

$$\begin{pmatrix} \mathbf{f}_t \\ \boldsymbol{\omega}_t \end{pmatrix} \Big| \mathbf{h}_t, \boldsymbol{\alpha}_t \sim \mathcal{N}_{N+K} \left(\mathbf{0}, \begin{bmatrix} \mathbf{H}_t & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_t \end{bmatrix} \right),$$

where

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

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

$$\mathbf{Y}_t | \mathbf{h}_t, \boldsymbol{\omega}_t \sim \mathcal{N}_N(\mathbf{0}, \mathbf{B}\mathbf{H}_t\mathbf{B}^\top + \boldsymbol{\Omega}_t).$$

Factor SV models

Expanding the matrix multiplications it follows that

$$\begin{aligned}\mathbb{V}ar[Y_{it}|\mathbf{h}_t, \boldsymbol{\omega}_t] &= \beta_{i1}^2 \exp(h_{1t}) + \cdots + \beta_{iK}^2 \exp(h_{Kt}) + \exp(\alpha_{it}) \\ \mathbb{C}ov[Y_{it}, Y_{jt}|\mathbf{h}_t, \boldsymbol{\omega}_t] &= \beta_{i1}\beta_{j1} \exp(h_{1t}) + \cdots + \beta_{iK}\beta_{jK} \exp(h_{Kt}),\end{aligned}$$

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

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

Factor SV models

We have $\mathbb{E}[f_{it}] = \mathbb{E}[\omega_{jt}] = 0$ while

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

Therefore,

$$\text{Var}[\mathbf{Y}_t] = \mathbf{B}\mathbf{H}\mathbf{B}^\top + \mathbf{\Omega}$$

where $\mathbf{H} = \text{diag}(\sigma_{f_1}^2, \dots, \sigma_{f_K}^2)$ and $\mathbf{\Omega} = \text{diag}(\sigma_{\omega_1}^2, \dots, \sigma_{\omega_N}^2)$.

More complicated dynamics could be introduced in the latent SV processes. For example, the errors e_t in the idiosyncratic shocks can be allowed to be correlated, so that $e_t \sim \mathcal{N}_K(\mathbf{0}, \Sigma_e)$.

Alternatively, consider a factor SV model where the errors in the latent AR processes of the log-volatilities h_t are contemporaneously correlated, so that

$$h_t = \mu + \Phi (h_{t-1} - \mu) + \eta_t,$$

where $\mu = (\mu_1, \dots, \mu_K)^\top$, $\Phi = \text{diag}(\phi_1, \dots, \phi_K)$ and $\eta_t \sim \mathcal{N}_K(\mathbf{0}, \Sigma_\eta)$.

- ▶ 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.

Models based on Wishart processes

The *Wishart* distribution is a distribution for positive definite matrix-valued random quantities. It has two parameters, a degrees of freedom parameter, ν , and a scale matrix Σ , which is positive definite.

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

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

Models based on Wishart processes

Let $\mathbf{X}$ be a $p \times p$ positive definite matrix. For $\mathbf{X} \sim \text{Wishart}_p(\nu, \Sigma)$, we require $\nu > p - 1$, and the Wishart pdf is

$$\frac{|\mathbf{X}|^{(\nu-p-1)/2}}{2^{\nu p/2} |\Sigma|^{\nu/2} \Gamma_p(\nu/2)} \exp \left\{ -\frac{1}{2} \text{Tr}(\Sigma^{-1} \mathbf{X}) \right\}$$

where

$$\Gamma_p \left(\frac{\nu}{2} \right) = \pi^{p(p-1)/4} \prod_{j=1}^p \Gamma(\nu/2 + (1-j)/2)$$

is the *multivariate Gamma function*. The expected value of $\mathbf{X}$ is

$$\mathbb{E}[\mathbf{X}] = \nu \Sigma$$

Models based on Wishart processes

The Wishart distribution arises from the multivariate Normal distribution as follows; if $\mathbf{X}_1, \dots, \mathbf{X}_n \sim \mathcal{N}_p(0, \Sigma)$, then

$$\mathbf{S} = \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^{\top} \quad (p \times p)$$

yields that

$$\mathbf{S} \sim \text{Wishart}_p(n, \Sigma)$$

Models based on Wishart processes

The Inverse-Wishart Distribution

Let $\mathbf{X}$ is a $p \times p$ positive definite matrix. For

$$\mathbf{X} \sim IWishart_p(\nu, \Psi),$$

we require $\nu > p - 1$, and the Inverse Wishart pdf is

$$\frac{|\Psi|^{\nu/2}}{2^{\nu p/2} |\mathbf{X}|^{-(\nu+p+1)/2} \Gamma_p(\nu/2)} \exp \left\{ -\frac{1}{2} \text{Tr}(\Psi \mathbf{X}^{-1}) \right\}$$

The expected value of $\mathbf{X}$ is $\Psi/(\nu - p - 1)$.

If $\mathbf{X} \sim Wishart_p(\nu, \Sigma)$, then

$$\mathbf{X}^{-1} \sim IWishart_p(\nu, \Sigma^{-1}).$$

The Matrix Beta Distribution

A $p \times p$ random symmetric positive definite matrix $\mathbf{X}$ is said to have a *matrix beta* distribution with parameters (a, b) if its pdf takes the form

$$\frac{1}{\mathcal{B}_p(a, b)} |\mathbf{X}|^{a-(p+1)/2} |I_p - \mathbf{X}|^{b-(p+1)/2} \quad \mathbf{0} < \mathbf{X} < I_p$$

where

$$a > (p-1)/2 \quad b > (p-1)/2$$

and

$$\mathcal{B}_p(a, b) = \frac{\Gamma_p(a)\Gamma_p(b)}{\Gamma_p(a+b)}$$

is the multivariate beta function.

Models based on Wishart processes

The inequality defining the support of this pdf

$$\mathbf{0} < \mathbf{X} < \mathbf{I}_p$$

is to be interpreted as the collection of $p \times p$ matrices such that both

$$\mathbf{X} \quad \text{and} \quad \mathbf{I}_p - \mathbf{X}$$

are positive definite.

By considering the transform

$$\mathbf{X} = (\mathbf{I}_p + \mathbf{Z})^{-1}\mathbf{Z} \quad \Longleftrightarrow \quad \mathbf{Z} = (\mathbf{I}_p - \mathbf{X})^{-1}\mathbf{X}$$

we have the alternative form

$$\frac{1}{\mathcal{B}_p(a, b)} |\mathbf{Z}|^{a-(p+1)/2} |\mathbf{I}_p + \mathbf{Z}|^{b-(p+1)/2} \quad \mathbf{Z} > \mathbf{0}.$$

Models based on Wishart processes

If $\mathbf{U} \sim \text{Beta}_p(a/2, b/2)$ with $a, b > p - 1$ and

$$\mathbf{V} = \mathbf{R}^T \mathbf{R} \sim \text{Wishart}_p(a + b, \Sigma)$$

for some positive definite symmetric $(p \times p)$ matrix Σ and for an upper-triangular matrix $\mathbf{R}$ with positive diagonal elements, then

$$\mathbf{S} = \mathbf{R}^T \mathbf{U} \mathbf{R} \sim \text{Wishart}_p(a, \Sigma).$$

Note that the matrix beta distribution and its density in are only defined for $a, b > p - 1$, and is singular otherwise.

Note: An extension to the definition of the matrix variate beta distribution for positive integers $b > 0$, and the conjugacy result still holds for $a > p - 1$ and integer-valued $b > 0$. In particular, the density of a singular $\text{Beta}_p(a/2, 1/2)$ -distributed random matrix can be found.

Models based on Wishart processes

To construct a multivariate stochastic volatility model: if $\mathbf{y}_t = (y_{1t}, \dots, y_{Nt})^\top$ the vector of one-period returns from N assets at time t and by Σ_t the $(N \times N)$ covariance matrix of $\mathbf{y}_t$.

Let $\mathbf{n}_t$ denote the Cholesky factor of Σ_t , defines by

$$\mathbf{n}_t^\top \mathbf{n}_t = \Sigma_t$$

that is, an upper-triangular matrix with positive diagonal elements. Then

$$\boldsymbol{\varepsilon}_t \sim \mathcal{N}_N(\mathbf{0}, \mathbf{I}_N)$$

$$\mathbf{y}_t = \mathbf{n}_t^{-1} \boldsymbol{\varepsilon}_t,$$

$$\Sigma_t = \frac{\nu + 1}{\nu} \mathbf{n}_{t-1}^\top \mathbf{Q}_t \mathbf{n}_{t-1}$$

$$\mathbf{Q}_t \sim \text{Beta}_N(\nu/2, 1/2), \quad t = 1, \dots, n.$$

Models based on Wishart processes

The random variables Q_t are serially and mutually independent of all other variables in the model.

Suppose also the initial condition that

$$\Sigma_0 \sim \text{Wishart}_N \left(\nu + 1, (\nu + 1)^{-1} S_0^{-1} \right)$$

for some $(N \times N)$ matrix S_0 (positive definite and symmetric).

Note that

$$\mathbb{E}[\Sigma_0] = S_0^{-1}$$

Models based on Wishart processes

Note that

$$\mathbf{Y}_t | \boldsymbol{\Sigma}_t \sim \mathcal{N}_N \left(\mathbf{0}, \boldsymbol{\Sigma}_t^{-1} \right).$$

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

$$\mathbf{n}_0^\top \mathbf{Q}_1 \mathbf{n}_0 \sim \text{Wishart}_N \left(\nu, (\nu + 1)^{-1} \mathbf{S}_0^{-1} \right),$$

so that prior to observing $\mathbf{Y}_1$, $\boldsymbol{\Sigma}_1 \sim \text{Wishart}_N \left(\nu, \nu^{-1} \mathbf{S}_0^{-1} \right)$ with

$$\mathbb{E} [\boldsymbol{\Sigma}_1] = \mathbf{S}_0^{-1}.$$

Models based on Wishart processes

Once the first observation $\mathbf{y}_1$ becomes available, the posterior for Σ_1 is a $Wishart_N(\nu + 1, (\nu + 1)^{-1}\mathbf{S}_1^{-1})$, with

$$\mathbf{S}_1 = \frac{1}{\nu + 1} (\mathbf{y}_1 \mathbf{y}_1^\top + \nu \mathbf{S}_0)$$

with

$$\mathbb{E} [\Sigma_1 | \mathbf{y}_1] = \mathbf{S}_1^{-1}.$$

This posterior distribution of Σ_1 suggests a prior $Wishart_N(\nu, \nu^{-1}\mathbf{S}_1^{-1})$ distribution for Σ_2 , so that

$$\mathbb{E} [\Sigma_2 | \mathbf{y}_1] = \mathbf{S}_1^{-1} = \mathbb{E} [\Sigma_1 | \mathbf{y}_1]$$

.

Models based on Wishart processes

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 $Wishart_N \left(\nu, \nu^{-1} S_{t-1}^{-1} \right)$ distribution,
- the posterior is $Wishart_N \left(\nu + 1, (\nu + 1)^{-1} S_t^{-1} \right)$, where

$$S_t = \frac{(\mathbf{y}_t \mathbf{y}_t^\top + \nu S_{t-1})}{(\nu + 1)}.$$

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 in principle straightforward.

Models based on Wishart processes

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

Then, the new Wishart model has $\varepsilon_t \sim \mathcal{N}_N(\mathbf{0}, \mathbf{I}_N)$ for $t = 1, \dots, n$

$$\mathbf{y}_t = \mathbf{n}_t^\top \varepsilon_t$$

so that $\mathbf{Y}_t | \Sigma_t \sim \mathcal{N}_N(\mathbf{0}, \Sigma_t)$ and hence the density of the returns at time t given the covariance structure is

$$p(\mathbf{y}_t | \Sigma_t) = \frac{1}{(2\pi)^{N/2} |\Sigma_t|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \mathbf{y}_t^\top \Sigma_t^{-1} \mathbf{y}_t \right\}.$$

Models based on Wishart processes

The evolution of the covariance matrices Σ_t is specified through one-step-ahead prediction densities, specifically, the density of the inverse of the current covariance structure Σ_t conditional on Σ_{t-1} .

In particular,

$$\Sigma_t^{-1} | \Sigma_{t-1} \sim \text{Wishart}_N(\nu, S_{t-1}),$$

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

Models based on Wishart processes

Then, for $t = 0, \dots, n - 1$, S_t is defined as follows

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

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

$$\boldsymbol{\Sigma}_t^{-1} | \boldsymbol{\Sigma}_{t-1} \sim \text{Wishart}_N \left(\nu, \frac{1}{\nu} \mathbf{R}^\top \left(\boldsymbol{\Sigma}_{t-1}^{-1} \right)^d \mathbf{R} \right)$$

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.

Models based on Wishart processes

The time-variation of the covariances can be expressed in terms of Σ_t . Since $\Sigma_t^{-1}|\Sigma_{t-1}$ is $Wishart_N(\nu, S_{t-1})$ -distributed, it follows that

$$\begin{aligned}\Sigma_t|\Sigma_{t-1} &\sim IWishart_N\left(\nu + N + 1, S_{t-1}^{-1}\right) \\ &\equiv IWishart_N\left(\nu + N + 1, \nu (\mathbf{R})^{-1} (\Sigma_{t-1})^d (\mathbf{R}^\top)^{-1}\right).\end{aligned}$$

Using the properties of the Wishart/inverse Wishart distributions

$$\begin{aligned}\mathbb{E}\left[\Sigma_t^{-1}|\Sigma_{t-1}\right] &= \nu S_{t-1} = \mathbf{R}^\top \left(\Sigma_{t-1}^{-1}\right)^d \mathbf{R} \\ \mathbb{E}\left[\Sigma_t|\Sigma_{t-1}\right] &= \frac{1}{\nu - N - 1} S_{t-1}^{-1}.\end{aligned}$$

Models based on Wishart processes

Hyperparameters:

- the elements of 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.

Models based on Wishart processes

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

$$\begin{aligned}\mathbb{E}[\boldsymbol{\Sigma}_t | \boldsymbol{\Sigma}_{t-1}] &= \frac{\nu}{\nu - N - 1} (\mathbf{R}^\top \mathbf{R})^{-1} = \frac{\nu}{\nu - N - 1} \mathbf{A}^{-1} \\ \mathbb{E}[\boldsymbol{\Sigma}_t^{-1} | \boldsymbol{\Sigma}_{t-1}] &= \mathbf{A}.\end{aligned}$$

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 X is formed by setting its columns to be $n \geq p$ random draws from a multivariate $\mathcal{N}_p(\mathbf{0}, \Sigma)$ distribution, where the $(p \times p)$ matrix $\Sigma > 0$, then the random variable

$$XX^T \sim \text{Wishart}_p(n, \Sigma)$$

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

The Wishart Autoregressive process

Let Σ_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 $\mathbf{x}_{1t}, \dots, \mathbf{x}_{\nu t}$, where each $\mathbf{x}_{it}$ ($i = 1, \dots, \nu$) is generated for $t = 1, \dots, n$ from

$$\mathbf{x}_{it} = \Phi \mathbf{x}_{it-1} + \boldsymbol{\eta}_{it},$$

where $\boldsymbol{\eta}_{it} \sim \mathcal{N}_N(\mathbf{0}, \Sigma_\eta)$, and Φ is a $(N \times N)$ matrix.

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

The Wishart Autoregressive process

The unconditional mean of each $\mathbf{x}_{it}$ is the zero vector while the unconditional variance $\mathbb{V}ar[\mathbf{x}_{it}] = \mathbf{S}$ is given by

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

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 $\mathbf{x}_{1t}, \dots, \mathbf{x}_{\nu t}$ are i.i.d. $\mathcal{N}_N(\mathbf{0}, \mathbf{S})$ -distributed variables.

The Wishart Autoregressive process

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

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

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

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

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

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

The Wishart Autoregressive process

Therefore if

$$\Sigma_t = \mathbf{X}_t \mathbf{X}_t^\top \quad t = 1, \dots, n$$

then $\Sigma_t > 0$ with probability one and the unconditional distribution of the covariance matrix at each time t is

$$\Sigma_t \sim \text{Wishart}_N(\nu, \mathbf{S}) \quad t = 1, \dots, n$$

and so the unconditional density is

$$p(\Sigma_t) = \frac{1}{2^{\frac{\nu N}{2}} \Gamma_N(\frac{\nu}{2})} \frac{|\Sigma_t|^{\frac{(\nu-N-1)}{2}}}{|\mathbf{S}|^{\frac{\nu}{2}}} \exp \left\{ -\frac{1}{2} \text{tr} \left(\mathbf{S}^{-1} \Sigma_t \right) \right\}.$$

Moreover, using the properties of the Wishart distribution, $\mathbb{E}[\Sigma_t] = \nu \mathbf{S}$.

The Wishart Autoregressive process

The distribution of Σ_t conditional on Σ_{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 $\mathbf{H}_t \sim \mathcal{N}_{N,\nu}(\mathbf{0}, \Sigma_\eta \otimes \mathbf{I}_\nu)$ using the dynamic linear model representation,

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

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

$$\Sigma_t | \mathbf{X}_{t-1} \sim \text{Wishart}_N \left(\nu, \Sigma_\eta, \Sigma_\eta^{-1} \Phi \mathbf{X}_{t-1} \mathbf{X}_{t-1}^\top \Phi^\top \right)$$

The Wishart Autoregressive process

The conditional distribution of Σ_t depends on X_{t-1} only through the product

$$X_{t-1}X_{t-1}^\top = \Sigma_{t-1},$$

which implies that the process is Markov. That is, the current covariance matrix Σ_t depends only on the previous period covariance matrix Σ_{t-1} and not on the latent VAR(1) processes contained in X_{t-1} .

The Wishart Autoregressive process

In particular, the conditional density of Σ_t is given by

$$\begin{aligned} p(\Sigma_t | \Sigma_{t-1}) &= \frac{1}{2^{\frac{\nu N}{2}} \Gamma_N(\frac{\nu}{2})} \frac{|\Sigma_t|^{\frac{(\nu-N-1)}{2}}}{|\Sigma_\eta|^{\frac{\nu}{2}}} \times \\ &\exp \left\{ -\frac{1}{2} \text{Tr} \left[\Sigma_\eta^{-1} (\Sigma_t + \Phi \Sigma_{t-1} \Phi^\top) \right] \right\} \times \\ &{}_0F_1 \left(\frac{\nu}{2}; \frac{1}{4} \Sigma_\eta^{-1} \Phi \Sigma_{t-1} \Phi^\top \Sigma_\eta^{-1} \Sigma_t \right). \end{aligned}$$

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

$$\mathbb{E}[\Sigma_t | \Sigma_{t-1}] = \nu \Sigma_\eta + \Phi \Sigma_{t-1} \Phi^\top.$$

The Wishart Autoregressive process

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 Σ_t , as in the previous model. That is

$$\begin{aligned}\Sigma_t | \Sigma_{t-1} &\sim \text{Wishart}_N \left(\nu, \Sigma_\eta, \Sigma_\eta^{-1} \Phi \Sigma_{t-1} \Phi^\top \right) \quad \text{with} \\ \Sigma_t &\sim \text{Wishart}_N (\nu, S) \quad , \quad \text{where} \quad S = \Sigma_\eta + \Phi S \Phi^\top.\end{aligned}$$

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 Wishart Autoregressive process

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

- set

$$\mathbf{Y}_t | \Sigma_t \sim \mathcal{N}_N(\mathbf{0}, \Sigma_t).$$

- allow the returns to depend on the inverse Σ_t^{-1} , that is

$$\mathbf{Y}_t | \Sigma_t \sim \mathcal{N}_N(\mathbf{0}, \Sigma_t^{-1}).$$

The Wishart Autoregressive process

Since the unconditional distribution of Σ_t is $Wishart_N(\nu, S)$ it follows, using the relation between the Wishart and the inverse Wishart distributions that

$$\Sigma_t^{-1} \sim IWishart_N(\nu + N + 1, S^{-1})$$

Therefore, since

$$Y_t | \Sigma_t \sim \mathcal{N}_N(\mathbf{0}, \Sigma_t^{-1})$$

and, marginally

$$\Sigma_t^{-1} \sim IWishart_N(\nu + N + 1, S^{-1})$$

it follows that the marginal distribution of Y_t is a multivariate t -distribution,

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

The Wishart Autoregressive process

This latter result is interesting from a modelling point of view; it is known that the distribution of the returns 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.