Financial Data Analysis

Stochastic Volatility Models

June 21, 2010
Stochastic Volatility Models

• The basic first–order autoregressive stochastic volatility (ARSV) model is given by

\[
\epsilon_t = u_t \exp\{\lambda_t/2\}, \quad u_t \overset{iid}{\sim} N(0, 1),
\]

\[
\lambda_t = \omega + \phi \lambda_{t-1} + \eta_t, \quad |\phi| < 1, \quad \eta_t \overset{iid}{\sim} N(0, \sigma_\eta^2),
\]

and the processes \( \{u_t\} \) and \( \{\eta_t\} \) are independent, and therefore also the processes \( \{u_t\} \) and \( \{\lambda_t\} \).

• Stochastic volatility (SV) processes with the property of \( \{u_t\} \) and \( \{\lambda_t\} \) being independent are referred to as independent SV processes.

• \( \lambda_t \) is the log–variance of \( r_t \), i.e., \( \lambda_t = \log \sigma_t^2 \), so \( \exp\{\lambda_t/2\} = \sigma_t \).
• Equation (2) describes a standard first-order Gaussian autoregression for $\log \sigma^2_t$, and it follows that the unconditional distribution

$$\lambda_t = \log \sigma^2_t \sim \text{Normal} \left( \frac{\omega}{1 - \phi}, \frac{\sigma^2_\eta}{1 - \phi^2} \right),$$

and volatility has a lognormal distribution.

• Motivation for this specification:

• In the stochastic volatility framework, volatility is partially determined by unpredictable events on the same day.

• Volatility can be expected to be linked to the flow of information arrival in the market, and there is some unpredictable component as invariably some news is not scheduled.

• Autocorrelation in volatility is then generated by autocorrelation in the news process.
• In contrast to GARCH, volatility in the SV framework is driven by an (unobserved) process separately from the return process (e.g., news arrival).

• In the GARCH framework, $\sigma_i^2$ is an (at least in principle) observable deterministic function of past shocks $\{\epsilon_s : s < t\}$. 
Basic Properties of Stochastic Volatility Models

- The moments of the SV model can be derived by using the fact that
  \[ x \sim N(\mu, \sigma^2) \Rightarrow \mathbb{E}(e^{tx}) = \exp \left\{ t\mu + \frac{t^2\sigma^2}{2} \right\}, \quad t \in \mathbb{R}. \]

- This gives, for example,
  \[
  \begin{align*}
  \mathbb{E}(\epsilon_t^2) &= \mathbb{E}(u_t^2) \exp \left\{ \mu_\lambda + \frac{\sigma_\lambda^2}{2} \right\} \\
  \mathbb{E}(\epsilon_t^4) &= \mathbb{E}(u_t^4) \exp \left\{ 2\mu_\lambda + 2\sigma_\lambda^2 \right\},
  \end{align*}
  \]

so that the kurtosis, \( \kappa \), is

\[
\kappa(\epsilon_t) = \frac{\mathbb{E}(\epsilon_t^4)}{\mathbb{E}^2(\epsilon_t^2)} = \frac{\mathbb{E}(u_t^4) \exp \left\{ 2\mu_\lambda + 2\sigma_\lambda^2 \right\}}{\mathbb{E}^2(u_t^2) \exp \left\{ 2\mu_\lambda + \sigma_\lambda^2 \right\}} = \kappa(u_t) \exp\{\sigma_\lambda^2\}, \quad (4)
\]

which becomes

\[
\kappa(\epsilon_t) = 3 \exp\{\sigma_\lambda^2\}
\]

in case of normality of \( u_t \).
• Equation (4) shows that the implied kurtosis of the unconditional return distribution consists of two components:
  – The kurtosis of the errors $u_t$, and
  – the kurtosis that is due to the variation in the volatility process $\lambda_t$.

• Note that excess kurtosis is positive, i.e., $\kappa(\epsilon_t) > \kappa(u_t)$, even if $\phi = 0$ (i.e., volatility is not persistent) due to stochastic $\eta_t$.

• The autocorrelations of the squared process are given by
  \[
  \text{Corr}(\epsilon_t^2, \epsilon_{t-\tau}^2) = \frac{\exp\{\sigma^2_\lambda \phi^\tau\} - 1}{\kappa(u_t) \exp\{\sigma^2_\lambda\} - 1}, \quad \tau > 0,
  \]
  which for small $\sigma^2_\lambda$ and/or $\phi$ close to one (which often occurs in financial time series) turns out to be approximately
  \[
  \text{Corr}(\epsilon_t^2, \epsilon_{t-\tau}^2) \approx \phi^\tau \frac{\exp\{\sigma^2_\lambda\} - 1}{\kappa(u_t) \exp\{\sigma^2_\lambda\} - 1},
  \]
  so we observe that the shape of the autocorrelation function of $\lambda_t$ (which is a standard AR(1)) approximately carries over to the squared process.
The role of $\phi$ in the stochastic volatility model is thus comparable with that of the sum $\alpha + \beta$ in the GARCH(1,1) process, i.e., it can serve as a measure of volatility persistence.

If fat–tailed distributions are called for, then $u_t$ may be taken to follow a GED or Student’s $t$ distribution, or any other fat–tailed alternative.

Asymmetric volatility dynamics (leverage effects) can also be incorporated into the model by abandoning the assumption of independence of $\{u_t\}$ and $\{\eta_t\}$ made in the definition of (1) and (2).

In particular, if we assume that

$$
\begin{bmatrix}
    u_t \\
    \eta_{t+1}
\end{bmatrix} \overset{iid}{\sim} \mathcal{N}
\left( 
\begin{bmatrix}
    0 \\
    0
\end{bmatrix},
\begin{bmatrix}
    1 & \delta \sigma \eta \\
    \delta \sigma \eta & \sigma^2 \eta
\end{bmatrix}
\right),
$$

where $\delta$ is the correlation between $u_t$ and $\eta_{t+1}$, then a negative $\delta$ introduces a negative correlation between $\epsilon_t$ and $\sigma^2_{t+1}$, i.e., a leverage effect.
Estimation of Stochastic Volatility Models

- Since the volatility process is unobservable, estimation of SV models is considerably more involved than estimation of GARCH processes.

- See Chapter 11.6 of Taylor (2005) for an overview of approaches to estimating SV models.\(^1\)

Regime–switching Models

1. Independent Mixture Models

- We first consider independent normal mixture distributions.

- A $k$–component (discrete) normal mixture distribution is described by the density

$$f(x) = \sum_{j=1}^{k} \lambda_j \phi(x; \mu_j, \sigma_j^2), \quad \phi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\},$$

where $\lambda_j > 0$, $j = 1, \ldots, k$, are the mixing weights, satisfying $\sum_j \lambda_j = 1$, and the $\mu_j$s and $\sigma_j^2$s are the component means and variances respectively.

- The mean and the variance of a normal mixture random variable follows from the moments of the normal distribution and are given by

$$\mu := \mathbb{E}(x) = \sum_{j=1}^{k} \lambda_j \mu_j,$$
\[
\sigma^2 := \text{Var}(x) = \mathbb{E}(x^2) - \mathbb{E}^2(x) = \sum_{j=1}^{k} \lambda_j (\sigma_j^2 + \mu_j^2) - \left( \sum_{j=1}^{k} \lambda_j \mu_j \right)^2 \\
= \sum_{j=1}^{k} \lambda_j \sigma_j^2 + \sum_{j=1}^{k} \lambda_j (\mu_j - \mu)^2,
\]
respectively.

- A possible interpretation of the normal mixture is that returns are normally distributed, but that return expectation and variance depend on the market regime, e.g., bull vs. bear markets.

- The regimes are not directly observable, however.

- Flexible with respect to skewness and kurtosis, bimodality.

- For example, consider the \textit{scale normal mixture}, where only the variances
are component-specific, whereas the component means are all equal to \( \mu \). This gives rise to a leptokurtic density.

- Consider example

\[
\lambda = 0.8, \quad \mu_1 = \mu_2 = 0, \quad \sigma_1^2 = 1, \quad \sigma_2^2 = 10,
\]

and the normal distribution with the same mean \( \mu = 0 \) and variance \( \sigma^2 = 0.8 \cdot 1 + 0.2 \cdot 10 = 2.8 \).
• A skewed density can be generated by allowing the component means to differ.

• The typical situation in financial applications where left–skewness emerges is where the component with the smaller probability (mixing weight) has the greater variance and the smaller mean.

• Consider example

\[ \lambda = 0.8, \quad \mu_1 = 1, \quad \mu_2 = -1, \quad \sigma_1^2 = 1, \quad \sigma_2^2 = 10. \]
• Bimodality is the result of the component means being sufficiently far apart, relative to the magnitude of the variances.
\( \lambda_1 = 0.5, \mu_1 = -1.5, \mu_2 = 1.5, \sigma^2_1 = \sigma^2_2 = 1 \)
Maximum Likelihood Estimation of Mixture Distributions via the EM Algorithm

• The parameter vector to be estimated is \( \theta = (\lambda_1, \ldots, \lambda_{k-1}, \mu_1, \ldots, \mu_k, \sigma_1^2, \ldots, \sigma_k^2) \).

• The **Expectation Maximization** (EM) algorithm provides a convenient tool for estimating mixture models even in higher dimensions (i.e., multivariate models).

• The EM algorithm is a general tool for handling estimation problems that may be *interpreted* as missing data problems.

• For mixture models, it gives rise to a very convenient recursive estimation scheme, where each iteration is given in closed form.

• Each iteration of the EM algorithm consists of two steps: The E–step and the M–step (this its name).

• Suppose we have a sample of return data \( \{ r_t \}_{t=1}^T \).
• Start with initial parameter values \( \hat{\lambda}_1, \ldots, \hat{\lambda}_k, \hat{\mu}_1, \ldots, \hat{\mu}_k, \hat{\sigma}_1^2, \ldots, \hat{\sigma}_k^2 \).

• **E–step:** Set

\[
\pi_{jt} = \frac{\lambda_j \phi(r_t; \mu_j, \sigma_j^2)}{\sum_{i=1}^k \lambda_i \phi(r_t; \mu_i, \sigma_i^2)}, \quad j = 1, \ldots, k, \quad t = 1, \ldots, T. \tag{7}
\]

• How can we interpret this quantity?

• Let the regime that the process is in at date \( t \) be indexed by an unobserved random variable \( s_t \), so \( s_t \in \{1, \ldots, k\} \), and

\[
P(s_t = j; \theta) = \lambda_j,
\]

and we have the conditional densities

\[
f(r_t|s_t = j; \theta) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp \left\{ -\frac{(r_t - \mu_j)^2}{2\sigma_j^2} \right\}, \quad j = 1, \ldots, k.
\]
Now recall
\[ P(A \cap B) = P(B) \cdot P(A|B), \]
and if we are interested in the probability of the joint event that \( s_t = j \) and \( r_t \) falls in some interval, we would integrate over the joint density

\[
p(r_t, s_t = j; \theta) = P(s_t = j; \theta) f(r_t|s_t = j; \theta)
\]

\[
= \frac{\lambda_j}{\sqrt{2\pi}\sigma_j} \exp \left\{ -\frac{(r_t - \mu_j)^2}{2\sigma_j^2} \right\}
\]

\[
= \lambda_j \phi(r_t; \mu_j, \sigma_j^2).
\]

Recall that the unconditional distribution of \( r_t \) is

\[
f(r_t; \theta) = \sum_j \lambda_j \phi(r_t; \mu_j, \sigma_j^2) = \sum_j p(r_t, s_t = j; \theta).
\]
• Then, using Bayes’ Theorem, $f(x|y) = \frac{f(x, y)}{f(y)} = f(x)f(y|x)/f(y)$,

$$P(s_t = j|r_t; \theta) = \frac{p(r_t, s_t = j; \theta)}{f(r_t; \theta)} = \frac{\lambda_j \phi(r_t; \mu_j, \sigma_j^2)}{\sum_{i=1}^{k} \lambda_i \phi(r_t; \mu_i, \sigma_i^2)} ,$$

that is, (7) is the posterior probability that $r_t$ has been generated by component $j$.

• I.e., (7) represents the probability, given the observed data, that the unobserved regime responsible for the observation at time $t$ was regime $j$.

• How are these probabilities used in calculating the parameter updates?
• **M–step**: Obtain new parameter values as follows:

$$
\hat{\lambda}_j = \frac{1}{T} \sum_{t=1}^{T} z_{jt}, \quad j = 1, \ldots, k,
$$

$$
\hat{\mu}_j = \left( \sum_{t} z_{jt} \right)^{-1} \sum_{t=1}^{T} z_{jt} r_t = \frac{1}{T\hat{\lambda}_j} \sum_{t=1}^{T} z_{jt} r_t, \quad j = 1, \ldots, k,
$$

$$
\hat{\sigma}_j^2 = \frac{1}{T\hat{\lambda}_j} \sum_{t=1}^{T} z_{jt} (r_t - \hat{\mu}_j)^2, \quad j = 1, \ldots, k.
$$

• That is, in updating the parameters for each component, each observation is weighted with this component’s posterior probability.

• If $z_{jt}$ is rather small (probability that $r_t$ was generated by regime $j$ is small), then this observation will only have a very small weight in calculating the parameter updates for regime $j$, and vice versa if $z_{jt}$ is very large.

• Then return to the E–step.
• It can be shown that each iteration of the EM algorithm increases the log–likelihood function,

\[
\log L = \sum_{t=1}^{T} \log \left\{ \sum_{j=1}^{k} \lambda_j \phi(r_t; \mu_j, \sigma_j^2) \right\}
\]

(8)

until a stationary point has been reached.

• The E–step and the M–step are then repeated until a prespecified convergence criterion is met.
• For example, we may choose to quit as soon as the improvement in the value of the log–likelihood function from one step to the next is less than some small number \( \epsilon > 0 \).

• Alternatively, we may stop when the maximum absolute difference between all elements of the parameter vector in two successive iterations is less than \( \epsilon \).

• Clearly the algorithm may converge to local maxima, so that different starting values should be used.
Example: European Stock Market Returns
- Consider two–component mixture models for the major European stock markets.

Table 1: Two–component normal mixture parameter estimates for European stock markets

<table>
<thead>
<tr>
<th></th>
<th>( \hat{\lambda}_1 )</th>
<th>( \hat{\mu}_1 )</th>
<th>( \hat{\sigma}^2_1 )</th>
<th>( \hat{\lambda}_2 )</th>
<th>( \hat{\mu}_2 )</th>
<th>( \hat{\sigma}^2_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAC 40</td>
<td>0.8396 (0.0199)</td>
<td>0.0679 (0.0190)</td>
<td>1.0310 (0.0462)</td>
<td>0.1604 (0.0199)</td>
<td>-0.2023 (0.1102)</td>
<td>7.0192 (0.6372)</td>
</tr>
<tr>
<td>DAX 30</td>
<td>0.8221 (0.0205)</td>
<td>0.0898 (0.0195)</td>
<td>1.0253 (0.0500)</td>
<td>0.1779 (0.0205)</td>
<td>-0.2851 (0.1085)</td>
<td>7.5843 (0.6527)</td>
</tr>
<tr>
<td>FTSE 100</td>
<td>0.8357 (0.0174)</td>
<td>0.0542 (0.0145)</td>
<td>0.6120 (0.0268)</td>
<td>0.1643 (0.0174)</td>
<td>-0.0944 (0.0880)</td>
<td>4.9196 (0.4118)</td>
</tr>
</tbody>
</table>

Table 2: Three–component normal mixture parameter estimates for European stock markets

<table>
<thead>
<tr>
<th></th>
<th>( \hat{\lambda}_1 )</th>
<th>( \hat{\mu}_1 )</th>
<th>( \hat{\sigma}^2_1 )</th>
<th>( \hat{\lambda}_2 )</th>
<th>( \hat{\mu}_2 )</th>
<th>( \hat{\sigma}^2_2 )</th>
<th>( \hat{\lambda}_3 )</th>
<th>( \hat{\mu}_3 )</th>
<th>( \hat{\sigma}^2_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAC 40</td>
<td>0.672 (0.0842)</td>
<td>0.098 (0.0257)</td>
<td>0.829 (0.0897)</td>
<td>0.290 (0.0726)</td>
<td>-0.144 (0.1035)</td>
<td>3.133 (0.7571)</td>
<td>0.037 (0.4205)</td>
<td>13.931 (3.8774)</td>
<td></td>
</tr>
<tr>
<td>DAX 30</td>
<td>0.703 (0.0495)</td>
<td>0.046 (0.0333)</td>
<td>1.509 (0.1383)</td>
<td>0.187 (0.0563)</td>
<td>0.154 (0.0446)</td>
<td>0.239 (0.0848)</td>
<td>0.109 (0.1608)</td>
<td>9.913 (1.1378)</td>
<td></td>
</tr>
<tr>
<td>FTSE 100</td>
<td>0.662 (0.0596)</td>
<td>0.060 (0.0182)</td>
<td>0.480 (0.0407)</td>
<td>0.300 (0.0528)</td>
<td>-0.008 (0.0579)</td>
<td>2.034 (0.3675)</td>
<td>0.036 (0.3065)</td>
<td>10.580 (2.4531)</td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Three–component normal mixture parameter estimates for European stock markets

<table>
<thead>
<tr>
<th></th>
<th>$k = 2$</th>
<th></th>
<th>$k = 3$</th>
<th></th>
<th>$k = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>log $L$</td>
<td>BIC</td>
<td>log $L$</td>
<td>BIC</td>
<td>log $L$</td>
</tr>
<tr>
<td>CAC 40</td>
<td>−8451.7</td>
<td>16946</td>
<td>−8439.9</td>
<td>16948</td>
<td>−8438.5</td>
</tr>
<tr>
<td>DAX 30</td>
<td>−8628.4</td>
<td>17299</td>
<td>−8602.4</td>
<td>17273</td>
<td>−8599.8</td>
</tr>
<tr>
<td>FTSE 100</td>
<td>−7313.7</td>
<td>14670</td>
<td>−7295.3</td>
<td>14659</td>
<td>−7294.7</td>
</tr>
</tbody>
</table>

- BIC = $-2 \times \log L + (3k - 1) \times \log T$, where $3k - 1$ is the number of parameters of a mixture model with $k$ components. Bold entries indicate the best model according to BIC.

- The cdf of the normal mixture is calculated as

$$F(r_t; \theta) = \sum_{j=1}^{k} \lambda_j \Phi \left( \frac{r_t - \mu_j}{\sigma_j} \right),$$

where $\Phi$ is the standard normal cdf.
• The sequence (9) should be uniformly distributed over the unit interval, and then

\[ z = \Phi^{-1} \left( \sum_{j=1}^{k} \lambda_j \Phi \left( \frac{r_t - \mu_j}{\sigma_j} \right) \right) \]  

(10)

should have a standard normal distribution.