Leture 2: Univariate Volatility Modelling - Part I

FM321: Risk Management and Modelling

Linyan Zhu

4 October 2022

LSE Finance

Main topics

- Introduction to financial returns and risks
- Volatility as the simplest risk measure

  Volatility is useful for various contexts:

  - Investment decisions
  - Portfolio construction
  - Risk management
  - Derivatives pricing
- Other risk measures beyond volatility
- Implementing risk forecasts
- Backtesting and stress-testing

- Univariate volatility modelling (one financial asset)
  - Moving average models
  - ARCH
  - GARCH
  - Model estimation
  - Diagnostics
  - Alternative approaches
- Multivariate volatility modelling

Unconditional vs. conditional variance

Recall two different concepts of variance for a series $\{r_t\}$

- Unconditional variance

  variance of r_t given no information

$$\sigma^2 = Var(r_t)$$

- Conditional variance

  variance of r_t given past information

$$\sigma_t^2 = Var(r_t | r_{t-1}, r_{t-2}, \ldots)$$

- Unconditional variance can be estimated by the sample variance of returns

$$\hat{\sigma}^2 = \frac{1}{T-1}\sum_{t=1}^{T}(r_t - \bar{r})^2$$

when r_t is stationary (implying that the unconditional variance is a constant).

- For high-frequency returns (such as daily returns), $\bar{r} \approx 0$, so

$$\hat{\sigma}^2 \approx \frac{1}{T-1}\sum_{t=1}^{T} r_t^2$$

$$\sigma_t^2 = Var(r_t | r_{t-1}, r_{t-2}, \ldots) \quad \leftarrow \text{ need a model to estimate}$$

- What features of data should the model capture?

$$\sigma_t^2 = Var(r_t | r_{t-1}, r_{t-2}, \dots) \quad \leftarrow \text{ need a model to estimate}$$

- What features of data should the model capture?

- Volatility clusters

$$\sigma_t^2 = Var(r_t | r_{t-1}, r_{t-2}, \ldots) \quad \leftarrow \text{ need a model to estimate}$$

- What features of data should the model capture?

- Volatility clusters
  - An unexpected shock to returns is usually followed by a period of high conditional volatility

Models of conditional variance

$$\sigma_t^2 = Var(r_t | r_{t-1}, r_{t-2}, \ldots) \quad \leftarrow \text{ need a model to estimate}$$

- What features of data should the model capture?

- Volatility clusters
  - An unexpected shock to returns is usually followed by a period of high conditional volatility

  - High conditional volatility in a period is followed by high conditional volatility in the next few periods, but shocks eventually die out

- Let W_E denote the estimation window.

- The conditional variance is the average sum of squared returns over the estimation window:

$$\hat{\sigma}_t^2 = \frac{1}{W_E} \sum_{i=1}^{W_E} r_{t-i}^2$$

- What's good about the model?

Moving average model: a starting point

- Let W_E denote the estimation window.
- The conditional variance is the average sum of squared returns over the estimation window:

$$\hat{\sigma}_t^2 = \frac{1}{W_E} \sum_{i=1}^{W_E} r_{t-i}^2$$

- What's good about the model?

  - Adaptive model that captures some of the desired properties (shocks to return lead to higher conditional variance, but eventually die out).

- Let W_E denote the estimation window.
- The conditional variance is the average sum of squared returns over the estimation window:

$$\hat{\sigma}_t^2 = \frac{1}{W_E} \sum_{i=1}^{W_E} r_{t-i}^2$$

- What's good about the model?

  - Adaptive model that captures some of the desired properties (shocks to return lead to higher conditional variance, but eventually die out).
  - Easy to implement (no parameter to be estimated).

Moving average model: a starting point

- What's wrong with the model?

- What's wrong with the model?
- Sensitive to the estimation window W_E.

- What's wrong with the model?
- Sensitive to the estimation window W_E.
- When W_E is small, model specification leads to abrupt changes in conditional volatility estimates when there is a return shock without a true volatility shock.

Moving average model: a starting point

- What's wrong with the model?
- Sensitive to the estimation window W_E.

- When W_E is small, model specification leads to abrupt changes in conditional volatility estimates when there is a return shock without a true volatility shock.

- When W_E is large, conditional volatility estimate react too slowly to a true volatility shock and also die out too slowly.

- What's wrong with the model?
- Sensitive to the estimation window W_E.
- When W_E is small, model specification leads to abrupt changes in conditional volatility estimates when there is a return shock without a true volatility shock.
- When W_E is large, conditional volatility estimate react too slowly to a true volatility shock and also die out too slowly.
- Equally weighted scheme cannot capture volatility clustering.

Moving average model: a starting point

- What's wrong with the model?
- Sensitive to the estimation window W_E.
- When W_E is small, model specification leads to abrupt changes in conditional volatility estimates when there is a return shock without a true volatility shock.
- When W_E is large, conditional volatility estimate react too slowly to a true volatility shock and also die out too slowly.
- Equally weighted scheme cannot capture volatility clustering.
- Exponential decay was introduced to remove some of the drawbacks of moving average models

- Model specification:

$$\sigma_t^2 = (1-\lambda)r_{t-1}^2 + \lambda\sigma_{t-1}^2$$

- λ: decay factor satisfying $0 < \lambda < 1$

- Higher λ implies greater persistence of the impact of shocks.

- Allows for greater influence of more recent observations on volatility estimates than more distant ones.

- Conditional variance is a weighted sum of past squared returns,

$$\sigma_t^2 = w_1 r_{t-1}^2 + w_2 r_{t-2}^2 + \cdots + w_{W_E} r_{t-W_E}^2$$

- and the weights are exponentially declining:

$$w_2 = \lambda w_1$$

$$w_3 = \lambda^2 w_1$$

$$w_4 = \lambda^3 w_1$$

$$\ldots$$

$$w_k = \lambda^{k-1} w_1$$

- w_1 is an appropriate normalizing constant chosen so that the weights sum up to 1.

$$1 = \sum_{i=1}^{W_E} w_i = w_1 + \lambda w_1 + \lambda^2 w_1 + \dots \lambda^{W_E - 1} w_1$$
$$= (1 + \lambda + \lambda^2 + \dots + \lambda^{W_E - 1}) w_1$$
$$= \frac{1 - \lambda^{W_E}}{1 - \lambda} w_1$$

- Hence,

$$w_1 = \frac{1 - \lambda}{1 - \lambda^{W_E}}$$

- We have

$$\sigma_t^2 = \frac{1 - \lambda}{1 - \lambda^{W_E}} (r_{t-1}^2 + \lambda r_{t-2}^2 + \dots + \lambda^{W_E - 1} r_{t-W_E}^2)$$
$$= \frac{1 - \lambda}{\lambda(1 - \lambda^{W_E})} \sum_{i=1}^{W_E} \lambda^i r_{t-i}^2$$

- How to get the EWMA equation:

$$\sigma_t^2 = (1-\lambda) r_{t-1}^2 + \lambda \sigma_{t-1}^2$$

EWMA: derivation

- How to get the EWMA equation:

$$\sigma_t^2 = (1-\lambda) r_{t-1}^2 + \lambda \sigma_{t-1}^2$$

- When W_E is large, λ^i becomes negligible for all $i \geq W_E$. Hence approximate the model by setting $W_E \to \infty$

$$\begin{aligned}
\sigma_t^2 &= (1-\lambda) \sum_{i=1}^{\infty} \lambda^{i-1} r_{t-i}^2 \\
&= (1-\lambda) \left(r_{t-1}^2 + \sum_{i=2}^{\infty} \lambda^{i-1} r_{t-i}^2 \right) \\
&= (1-\lambda) r_{t-1}^2 + (1-\lambda) \sum_{i=2}^{\infty} \lambda^{i-1} r_{t-i}^2 \\
&= (1-\lambda) r_{t-1}^2 + \lambda \underbrace{(1-\lambda) \sum_{i=1}^{\infty} \lambda^{i-1} r_{t-i-1}^2}_{\sigma_{t-1}^2} \\
&= (1-\lambda) r_{t-1}^2 + \lambda \sigma_{t-1}^2
\end{aligned}$$

EWMA: what value does λ take?

- EWMA was implemented in large scale by JP Morgan in the late 1980s and early 1990s, and made available broadly under the brand name RiskMetrics (around the same time as the concept of Value-at-Risk)

- JP Morgan set for daily data: $\lambda = 0.94$

- When we use the term EWMA we are assuming that value.

Unconditional EWMA variance is not defined!

- Suppose a return series $\{r_t\}$ satisfies:

$$\sigma_t^2 = (1-\lambda) r_{t-1}^2 + \lambda \sigma_{t-1}^2$$

Unconditional EWMA variance is not defined!

- Suppose a return series $\{r_t\}$ satisfies:

$$\sigma_t^2 = (1-\lambda)r_{t-1}^2 + \lambda\sigma_{t-1}^2$$

- For a scaled version of the return, $r_t' = cr_t$ $(c > 0)$:

$$\begin{aligned}
(\sigma_t')^2 &= Var[(r')_t^2 | r_{t-1}', \dots] \\
&= Var[cr_t^2 | cr_{t-1}, \dots] \\
&= c^2 Var(r_t^2 | r_{t-1}, \dots) \\
&= c^2\sigma_t^2
\end{aligned}$$

- Suppose a return series $\{r_t\}$ satisfies:

$$\sigma_t^2 = (1-\lambda) r_{t-1}^2 + \lambda \sigma_{t-1}^2$$

- For a scaled version of the return, $r_t' = c r_t$ $(c > 0)$:

$$\begin{aligned}
(\sigma_t')^2 &= Var[(r')_t^2 | r_{t-1}', \dots] \\
&= Var[c r_t^2 | c r_{t-1}, \dots] \\
&= c^2 Var(r_t^2 | r_{t-1}, \dots) \\
&= c^2 \sigma_t^2
\end{aligned}$$

- Hence, r_t' also satisfies the same EWMA equation above.

$$(\sigma_t')^2 = c^2 \sigma_t^2 = c^2[(1-\lambda) r_{t-1}^2 + \lambda \sigma_{t-1}^2] = (1-\lambda)(r_{t-1}')^2 + \lambda(\sigma_{t-1}')^2$$

Unconditional EWMA variance is not defined!

- Suppose a return series $\{r_t\}$ satisfies:

$$\sigma_t^2 = (1-\lambda)r_{t-1}^2 + \lambda\sigma_{t-1}^2$$

- For a scaled version of the return, $r'_t = cr_t$ $(c > 0)$:

$$\begin{aligned}(\sigma'_t)^2 &= Var[(r')_t^2 | r'_{t-1}, \dots] \\ &= Var[cr_t^2 | cr_{t-1}, \dots] \\ &= c^2 Var(r_t^2 | r_{t-1}, \dots) \\ &= c^2\sigma_t^2\end{aligned}$$

- Hence, r'_t also satisfies the same EWMA equation above.

$$(\sigma'_t)^2 = c^2\sigma_t^2 = c^2[(1-\lambda)r_{t-1}^2 + \lambda\sigma_{t-1}^2] = (1-\lambda)(r'_{t-1})^2 + \lambda(\sigma'_{t-1})^2$$

- Knowledge of model parameters is insufficient to pin down the unconditional variance of the model.

- Why do we care whether we can or cannot characterize the unconditional variance of EWMA, or any volatility model?

- What exactly is in the specification of EWMA that makes the unconditional variance non-existent?

- Unconditional variance resembles the level that we would expect the conditional variance to stay at if there were no shocks to the system.

- Once a shock hits, conditional variance will rise, and eventually mean-revert to the original level.

- Non-existence of the unconditional variance suggests that future conditional variances will instead drift away like a random walk.

- Key problem: the sum of coefficients is 1, and there is no constant term.

- What happens to $E_t[\sigma^2_{t+k}]$ when there is a shock to σ^2_t?

Standardized residual

- Write log return r_t as

$$r_t = \sigma_t z_t$$

where z_t is called the **standardized residual** satisfying $E_{t-1}[z_t] = 0$ and $Var_{t-1}[z_t] = 1$.

- z_t represents the component of return that is independent of all information up to time $t-1$.

- Write log return r_t as

$$r_t = \sigma_t z_t$$

where z_t is called the **standardized residual** satisfying $E_{t-1}[z_t] = 0$ and $Var_{t-1}[z_t] = 1$.

- z_t represents the component of return that is independent of all information up to time $t-1$.

- Verify that σ_t^2 is indeed the conditional variance of r_t:

$$Var_{t-1}(r_t) = E_{t-1}(r_t^2) = E_{t-1}(\sigma_t^2 z_t^2) = \sigma_t^2 E_{t-1}(z_t^2) = \sigma_t^2$$

Properties of the standardized residual

- For any future date $t + k$ ($k > 0$), we have

$$
\begin{aligned}
E_t[r_{t+k}^2] &= E_t[\sigma_{t+k}^2 z_{t+k}^2] \\
&= E_t[E_{t+k-1}[\sigma_{t+k}^2 z_{t+k}^2]] \\
&= E_t[\sigma_{t+k}^2 E_{t+k-1}[z_{t+k}^2]] \\
&= E_t[\sigma_{t+k}^2]
\end{aligned}
$$

- Allow us to focus on modeling σ_t^2, leaving z_t as a standardized white noise that does not require modeling (other than assigning a distribution, usually the standard normal distribution).

EWMA - effect of a shock

- Shocks to volatility in EWMA are not mean-reverting, so a jump in conditional volatility is not expected to come down. Reminds us of the random walk model (that you might have seen elsewhere), which does not have a stationary mean and has an infinite variance.

$$E_t[\sigma^2_{t+k}] = E_t[\sigma^2_{t+k-1}] = \cdots = E_t[\sigma^2_{t+1}] = \sigma^2_t$$

- Shocks to volatility in EWMA are not mean-reverting, so a jump in conditional volatility is not expected to come down. Reminds us of the random walk model (that you might have seen elsewhere), which does not have a stationary mean and has an infinite variance.

$$E_t[\sigma^2_{t+k}] = E_t[\sigma^2_{t+k-1}] = \cdots = E_t[\sigma^2_{t+1}] = \sigma^2_t$$

- To see this, start from the EWMA equation:

$$\sigma^2_{t+k} = (1-\lambda) r^2_{t+k-1} + \lambda \sigma^2_{t+k-1},$$

we have

$$\begin{aligned} E_t[\sigma^2_{t+k}] &= (1-\lambda) E_t[r^2_{t+k-1}] + \lambda E_t[\sigma^2_{t+k-1}] \\ &= (1-\lambda) E_t[\sigma^2_{t+k-1}] + \lambda E_t[\sigma^2_{t+k-1}] \\ &= E_t[\sigma^2_{t+k-1}] \end{aligned}$$

- Shocks to volatility in EWMA are not mean-reverting, so a jump in conditional volatility is not expected to come down. Reminds us of the random walk model (that you might have seen elsewhere), which does not have a stationary mean and has an infinite variance.

$$E_t[\sigma^2_{t+k}] = E_t[\sigma^2_{t+k-1}] = \cdots = E_t[\sigma^2_{t+1}] = \sigma^2_t$$

- To see this, start from the EWMA equation:

$$\sigma^2_{t+k} = (1-\lambda) r^2_{t+k-1} + \lambda \sigma^2_{t+k-1},$$

we have

$$\begin{aligned} E_t[\sigma^2_{t+k}] &= (1-\lambda) E_t[r^2_{t+k-1}] + \lambda E_t[\sigma^2_{t+k-1}] \\ &= (1-\lambda) E_t[\sigma^2_{t+k-1}] + \lambda E_t[\sigma^2_{t+k-1}] \\ &= E_t[\sigma^2_{t+k-1}] \end{aligned}$$

- Key problem: the sum of coefficients is 1, and there is no constant term.

Autoregressive conditional heteroskedasticity (ARCH)

- ARCH models were proposed by Engle (1982)
- Bollerslev (1986) generalized ARCH to GARCH (Generalized ARCH) models.
- Serve as the basis for most univariate volatility models used in practice
- Model conditional volatility as a function of:
  - Past squared returns (in ARCH);
  - Past conditional volatility (in GARCH);
  - Possibly additional factors (in more complex specifications)

ARCH - model specification

- The defining equation of an ARCH(1) process is

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2$$

- More generally, ARCH(N) allows for an arbitrary number N of lags:

$$\sigma_t^2 = \omega + \sum_{n=1}^{N} \alpha_n r_{t-n}^2$$

- It is necessary to assume that $\omega > 0$, $\alpha_n \geq 0$, and at least one of the α_n is strictly positive.

ARCH - comments

- ARCH models effectively incorporate the effects of past shocks on current volatility estimates.
- They improve on moving average models in that, under appropriate conditions, they do not suffer the drawbacks related to unconditional variance.

ARCH(1) - unconditional variance

- Assume that the ARCH process has a finite unconditional variance, and denote it by σ^2.

- Recall that $E[r_t^2] = E[\sigma_t^2] = \sigma^2$ for every t.

- Therefore, for an ARCH(1) process:

$$
\begin{aligned}
E[\sigma_{t+1}^2] &= \omega + \alpha E[r_t^2] \\
\Rightarrow \sigma^2 &= \omega + \alpha\sigma^2 \\
\Rightarrow \sigma^2 &= \frac{\omega}{1-\alpha}
\end{aligned}
$$

- In order for the unconditional variance of the process to exist, we need $\alpha < 1$.

ARCH(N) - unconditional variance

- A similar argument can be used to establish that, for an ARCH(N) process we have: $\sigma^2 = \frac{\omega}{1-\sum_{n=1}^{N} \alpha_n}$.

- In order for the unconditional variance of the process to exist, we need $\sum_{n=1}^{N} \alpha_n < 1$

- Unlike with EWMA, knowledge of the model parameters is sufficient to pin down the unconditional variance of the model

- Therefore, for an ARCH(1) process:

$$
\begin{aligned}
\sigma^2_{t+k} &= \omega + \alpha r^2_{t+k-1} \\
E_t[\sigma^2_{t+k}] &= \omega + \alpha E_t[r^2_{t+k-1}] \\
E_t[\sigma^2_{t+k}] &= \omega + \alpha E_t[\sigma^2_{t+k-1}] \\
E_t[\sigma^2_{t+k}] &= \sigma^2(1-\alpha) + \alpha E_t[\sigma^2_{t+k-1}] \\
E_t[\sigma^2_{t+k}] - \sigma^2 &= \alpha \left[E_t[\sigma^2_{t+k-1}] - \sigma^2 \right]
\end{aligned}
$$

- Therefore, for an ARCH(1) process:

$$\sigma^2_{t+k} = \omega + \alpha r^2_{t+k-1}$$
$$E_t[\sigma^2_{t+k}] = \omega + \alpha E_t[r^2_{t+k-1}]$$
$$E_t[\sigma^2_{t+k}] = \omega + \alpha E_t[\sigma^2_{t+k-1}]$$
$$E_t[\sigma^2_{t+k}] = \sigma^2(1-\alpha) + \alpha E_t[\sigma^2_{t+k-1}]$$
$$E_t[\sigma^2_{t+k}] - \sigma^2 = \alpha\left[E_t[\sigma^2_{t+k-1}] - \sigma^2\right]$$

- Hence:

$$E_t[\sigma^2_{t+k}] - \sigma^2 = \alpha^{k-1}\left[E_t[\sigma^2_{t+1}] - \sigma^2\right] = \alpha^{k-1}(\sigma^2_{t+1} - \sigma^2)$$

- Therefore, for an ARCH(1) process:

$$
\begin{aligned}
\sigma_{t+k}^2 &= \omega + \alpha r_{t+k-1}^2 \\
E_t[\sigma_{t+k}^2] &= \omega + \alpha E_t[r_{t+k-1}^2] \\
E_t[\sigma_{t+k}^2] &= \omega + \alpha E_t[\sigma_{t+k-1}^2] \\
E_t[\sigma_{t+k}^2] &= \sigma^2(1-\alpha) + \alpha E_t[\sigma_{t+k-1}^2] \\
E_t[\sigma_{t+k}^2] - \sigma^2 &= \alpha\left[E_t[\sigma_{t+k-1}^2] - \sigma^2\right]
\end{aligned}
$$

- Hence:

$$
E_t[\sigma_{t+k}^2] - \sigma^2 = \alpha^{k-1}\left[E_t[\sigma_{t+1}^2] - \sigma^2\right] = \alpha^{k-1}(\sigma_{t+1}^2 - \sigma^2)
$$

- Conditional variance forecasts are mean reverting.

ARCH(1) - tail behavior

- What can we say about the tail behavior of an ARCH(1) process?

- Assume that z_t is normally distributed conditional on information up to time $t-1$, so that $E_{t-1}[z_t^4] = 3$

- Note that

$$\begin{aligned} E[r_t^4] &= E[E_{t-1}[r_t^4]] \\ &= E[E_{t-1}[\sigma_t^4 z_t^4]] \\ &= E[\sigma_t^4 E_{t-1}[z_t^4]] \\ &= 3E[\sigma_t^4] \end{aligned}$$

- We have

$$\sigma_t^4 = \omega^2 + 2\alpha\omega r_{t-1}^2 + \alpha^2 r_{t-1}^4$$

$$E[\sigma_t^4] = \omega^2 + 2\alpha\omega \frac{\omega}{1-\alpha} + 3\alpha^2 E[\sigma_{t-1}^4]$$

$$(1-3\alpha^2)E[\sigma_t^4] = \omega^2 \left[1 + \tfrac{2\alpha}{1-\alpha}\right]$$

$$E[\sigma_t^4] = \omega^2 \left[\tfrac{1+\alpha}{(1-\alpha)(1-3\alpha^2)}\right]$$

- We can compute the unconditional kurtosis as

$$\frac{E[r_t^4]}{E[r_t^2]^2} = 3\omega^2 \left[\tfrac{1+\alpha}{(1-\alpha)(1-3\alpha^2)}\right] \times \frac{(1-\alpha)^2}{\omega^2}$$

$$= 3\frac{(1-\alpha^2)}{(1-3\alpha^2)}$$

ARCH(1) - tail behavior

- In order for the kurtosis to exist, we need $3\alpha^2 < 1$, or $\alpha < 0.577$

- So long as this condition is satisfied, the unconditional kurtosis is greater than 3

- The ARCH(1) process has unconditionally heavy tails even though we assumed that the standardized residuals are conditionally normally distributed

ARCH vs. EWMA

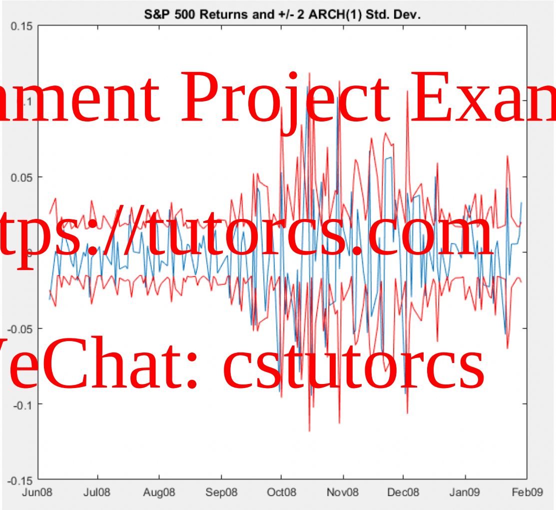
- Both processes allow for temporary shocks to returns to have an effect on conditional volatility.

- Both processes allow for temporary shocks to returns to have an effect on conditional volatility.
- However, ARCH can replicate some usual features of financial time series which EWMA cannot:

  - Unconditional variance is finite.
  - Shocks to volatility eventually die out.

  - Tails are heavy even if standardized residuals are normal (if parameter values are such that fourth moment exists).

ARCH(1): very unstable estimates



- In spite of all of the improvements over EWMA, ARCH models are not quite satisfactory (almost nobody uses it).

- Volatility estimates tend to be very unstable, which creates practical problems.

- This problem can be solved if we include a large number of lags, but that leads to estimation difficulties.

- In addition, conditions on parameters for existence of fourth moments tend to be too restrictive.