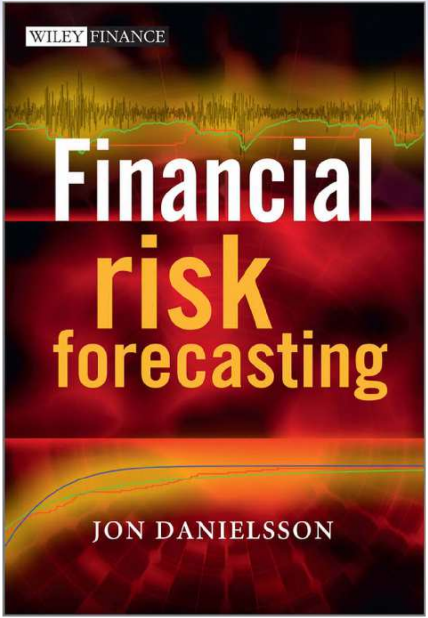
Financial Risk Forecasting
Chapter 2-a
Univariate Volatility Modelling — Part A

Jon Danielsson ©2025
London School of Economics

To accompany
Financial Risk Forecasting
FinancialRiskForecasting.com

Published by Wiley 2011

Version 10.0, August 2025


WILEY FINANCE
Financial risk forecasting
JON DANIELSSON

Univariate Volatility Modelling

Volatility

- Volatility is the main measure of risk (see Chapter 4)
- Investment decisions
- Portfolio construction
- Derivative pricing

Estimation and Forecasting

- In this Chapter, we focus on univariate volatility estimation and forecasting for a single asset
- The next Chapter focusses on implementation
- The Chapter after next does multivariate volatility
- The focus in this Chapter is on estimation, both of a model and in-sample volatilities
- Chapter 5 introduces (out-of-sample) forecasting of volatilities

Structure

- The theoretical specification of common volatility models
- The theory of estimating the models
- Practical implementation in estimation
- Diagnostics of estimated volatility models

Univariate Volatility Models

- Moving average (MA)
- Exponentially weighted moving average (EWMA)
- GARCH and its extension models

Notation new to this Chapter

W_E	Estimation window
λ	Decay factor in EWMA
ϵ_t	Residuals
ω, α, β	Main model parameters
ζ, δ	Other model parameters
θ	Set of parameters
L_1, L_2	Lags in volatility models

Risk (and hence volatility) are latent variables

- It is not possible to measure volatility
- Instead, it has to be inferred from observed data
- Using some sort of model
- And that means there are multiple alternative ways to measure volatility
- And it can be very hard to discriminate between them

Learning outcomes

1. Know the difference between conditional and unconditional volatility
2. Understand why a return series can be conditionally normal and unconditionally fat-tailed
3. Recognise the various types of models used for volatility forecasting
4. Understand why the MA model is generally not recommended
5. Understand and derive the EWMA model
6. Identify the main characteristics of the ARCH and GARCH models, including conditional and unconditional volatility and tail thickness, memory and news, and parameter restrictions
7. Recognise the most important extensions to these models

Main challenges

- Understand the difference between conditional and unconditional
- For example, how can returns both be fat tailed and normal?
- What is a nested and un-nested model

Other reading

- Alexios Galanos, 2024, GARCH Models
  cran.r-project.org/web/packages/tsgarch/vignettes/garch_models.pdf
- Christian Francq and Jean-Michel Zakoian (2019), "GARCH Models: Structure, Statistical Inference and Financial Applications", 2nd Edition.

Simple Models

The Mean of Returns

- Recall from the last Chapter that the mean of daily returns is very low
- In one case mean is 0.0001007711
- While the standard error is more than 100 times larger, at 0.01245144
- With daily returns it is usually quite safe to assume the mean is zero
- It is convenient because it makes the mathematics simpler
- We will return to this issue several times later in the book

Volatility

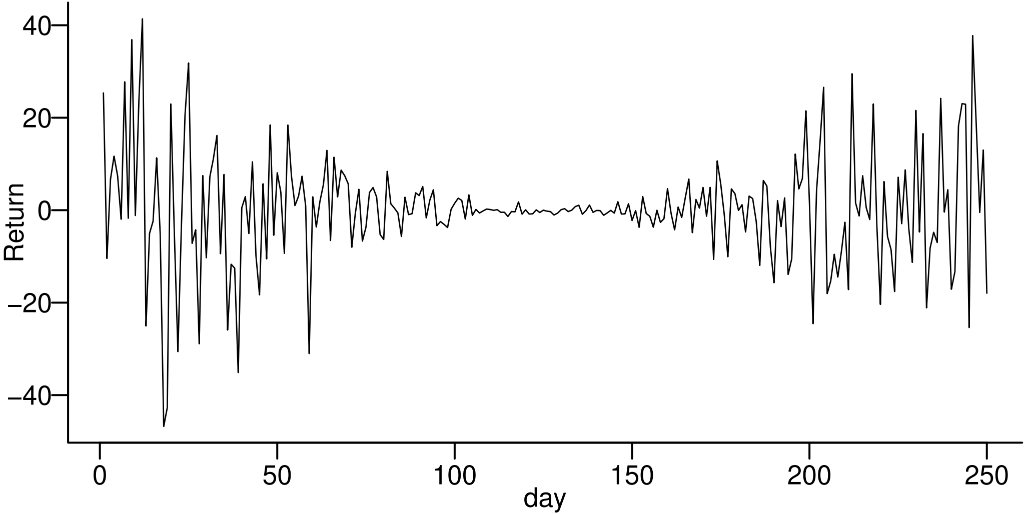
- Volatility is the standard error (square root of variance) of returns
- The definition of variance is

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

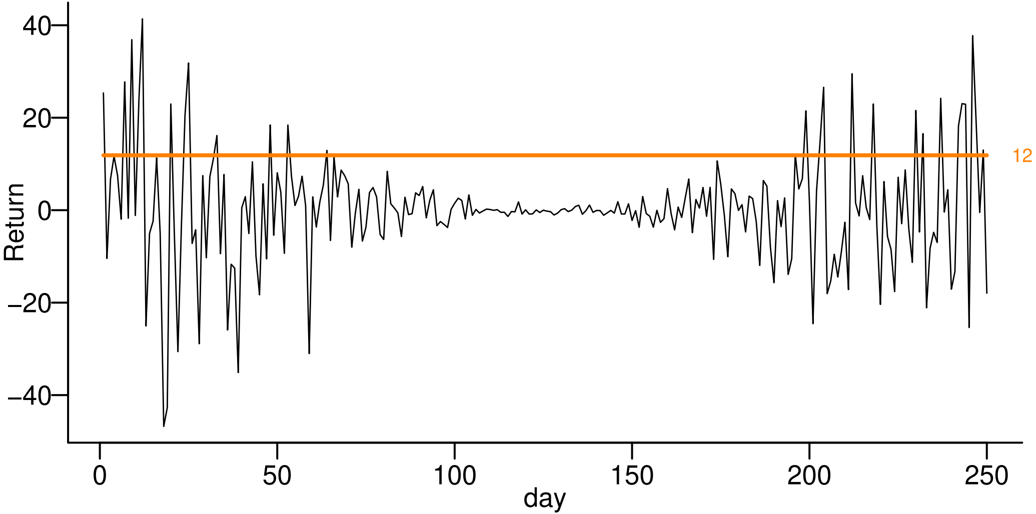
- And since we assume the mean is zero, this simplifies to

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2$$

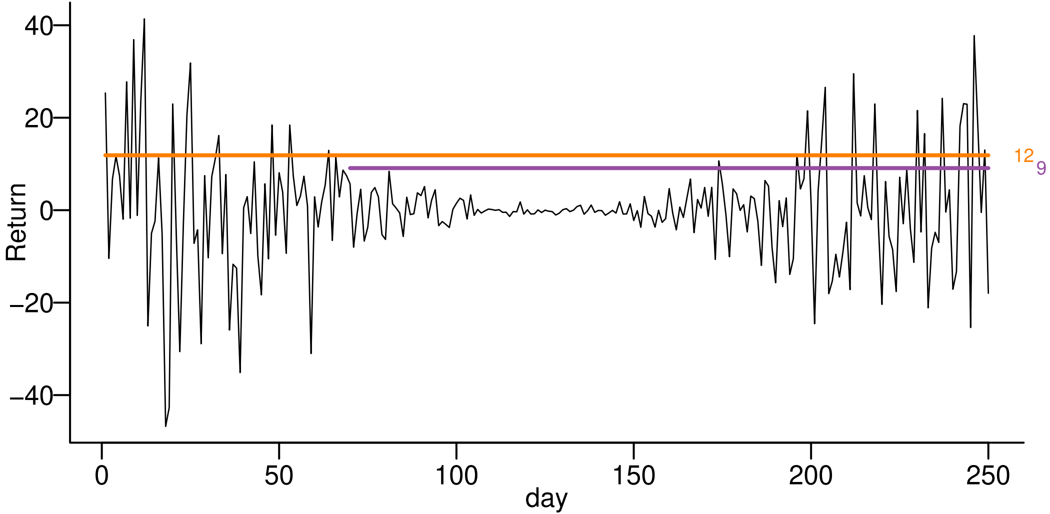
Hypothetical Returns



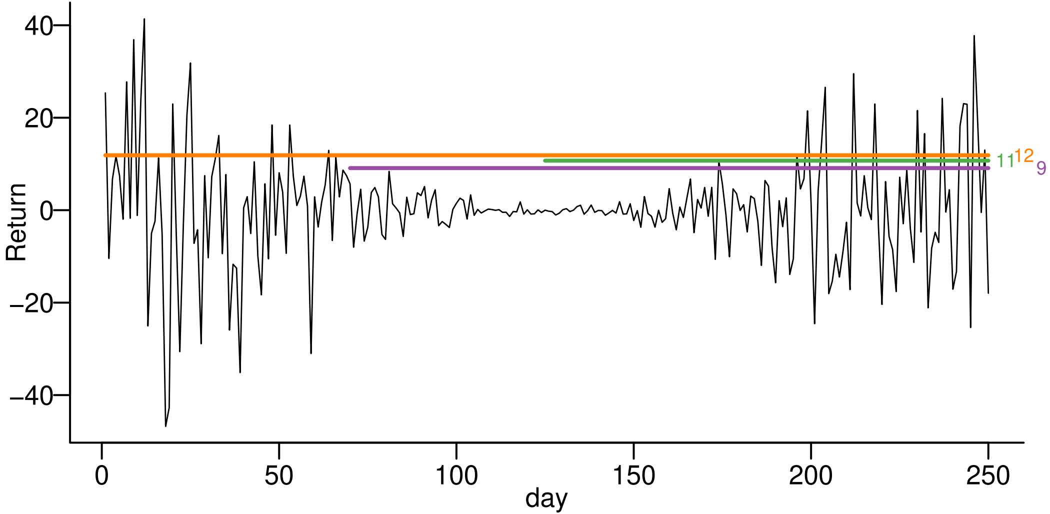
Hypothetical Returns (cont.)



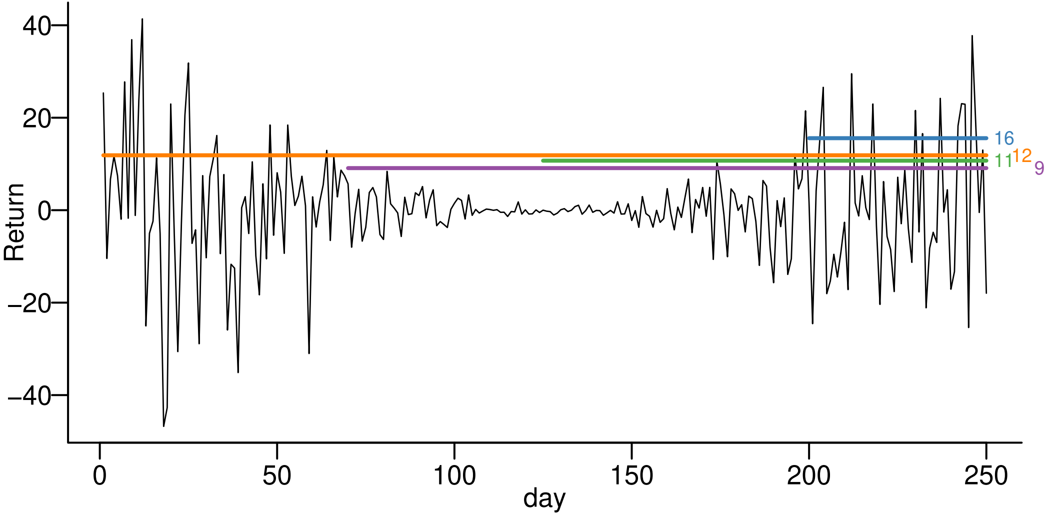
Hypothetical Returns (cont.)



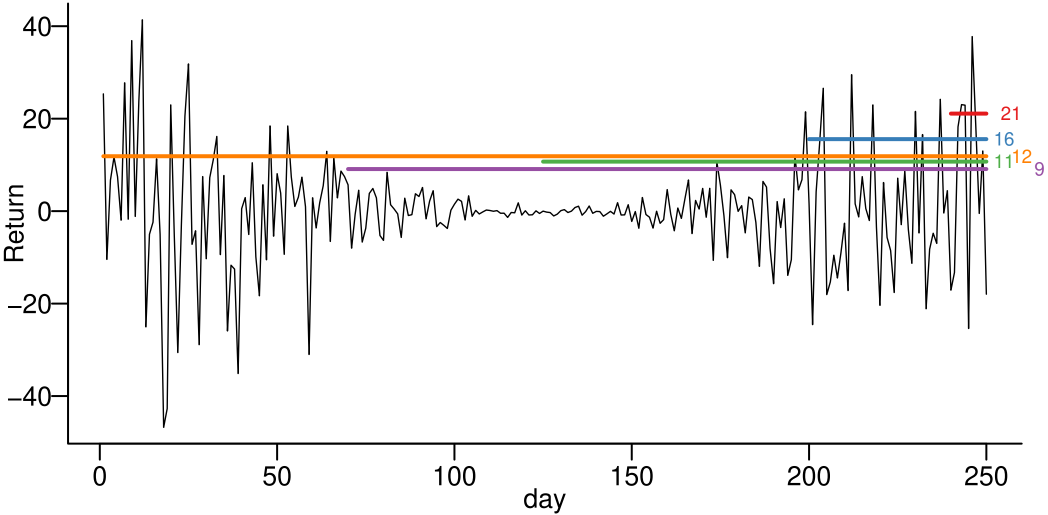
Hypothetical Returns (cont.)



Hypothetical Returns (cont.)



Hypothetical Returns (cont.)



Two Concepts of Volatility

- Unconditional volatility (σ)
  Volatility over an entire time period

- Conditional volatility (σ_t)
  Volatility conditional on a given time period, the past history, model and model parameters

Moving Average Model

Moving Average

- The simplest volatility model is *moving average*
- Where the *conditional* volatility is the square root of the average of squared returns over the *estimation window*, W_E

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

- Note that this is a *one-day-ahead* forecast
- Because it uses information until time $t-1$ to do the calculation for time t

Moving Average Pros and Cons

- The moving average model will generally perform quite badly
- Because it is quite sensitive to the size of the estimation window (see the next plots)
- And because it does not weigh history

Exponentially Weighted Moving Average (EWMA)

Exponentially Weighted Moving Average (EWMA)

- JP Morgan proposed a model called *RiskMetrics* in 1993
- Since they later used that name for a consulting company they spun off, we use the more pedestrian
- *Exponentially weighted moving average* (EWMA)

EWMA

- Volatility a *exponentially weighted* sum of past returns,

$$\hat{\sigma}_t^2 = \lambda y_{t-1}^2 + \lambda^2 y_{t-2}^2 \ldots + \lambda^{W_E} y_{t-W_E}^2$$

$$1 > \lambda > 0$$

$$\sum_{i=1}^{W_E} \lambda^i = 1$$

- If W_E is large enough, the terms λ^i are negligible for all $i \geq W_E$
- So set $W_E = \infty$
- We show in the Appendix that the EWMA equation is then

$$\hat{\sigma}_t^2 = (1 - \lambda) y_{t-1}^2 + \lambda \hat{\sigma}_{t-1}^2$$

What Is λ ?

$$\hat{\sigma}_t^2 = (1-\lambda)y_{t-1}^2 + \lambda\hat{\sigma}_{t-1}^2$$

- It is possible to estimate EWMA with the methods discussed later in this Chapter
  - Since the EWMA is a reduced GARCH model, we could simply estimate λ with maximum likelihood (see the GARCH discussion below)
- JP Morgan set for daily data

$$\lambda = 0.94$$

- And when we use the term EWMA we are assuming that value

Unconditional EWMA Volatility

- The EWMA is a reduced GARCH model
- So, using an equation shown later, we get that the unconditional volatility is

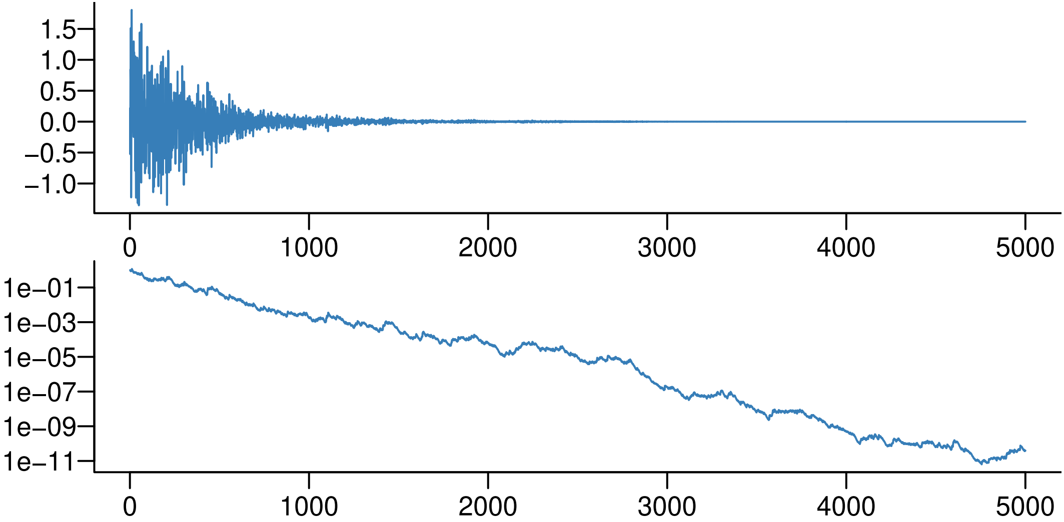
$$\sigma^2 = \frac{0}{0}$$

- In other words, undefined
- We can explore that more with simulations

Simulation Analysis

```
N=1000
y=rnorm(N)
lambda=0.94
sigma2=vector(length=N)
sigma2[1]=1
for(i in 2:N){
  sigma2[i]=lambda*sigma2[i-1]+(1-lambda)*y[i-1]^2
  y[i]=y[i]*sqrt(sigma2[i])
}
plot(y,type='l')
plot(sigma2,type='l',log='y')
```

Simulate EWMA



Conclusion From Simulation

- As we simulated longer and longer paths
- The σ become smaller and smaller
- Eventually zero
- So the unconditional volatility is undefined

EWMA Pros and Cons

- Pros
  - It is really simple to implement
  - And not that inaccurate compared to the more sophisticated GARCH
  - And multivatiate (see Chapter 3) versions are really easy
- Cons
  - By definition it is less accurate than GARCH
  - Which can become important in some cases
  - The unconditional volatility is not defined, which can be a problem

The ARCH Family

ARCH

- Robert Engle proposed a model in 1982 called autoregressive conditionally heteroscadastic (ARCH)
- Most volatility models derive from this
- Returns have a *conditional* distribution (here assumed to be normal)

$$y_t \sim \mathcal{N}(0, \sigma_t^2)$$

- Or we can write:

$$y_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1)$$

- Where ϵ_t is called *residual*

ARCH

- The volatility is weighted average of past returns
- ARCH(L_1)

$$\sigma_t^2 = \omega + \sum_{i=1}^{L_1} \alpha_i y_{t-i}^2$$

- The number of lags is L_1
- The most common form is *ARCH(1)*

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

- ω, α are parameters to be estimated with maximum likelihood

ARCH(1) Fat Tails

- The most common distributional assumption for residuals ϵ is standard normality; that is:

$$\epsilon_t \sim N(0,1)$$

- Because ϵ_t is normal, so must conditional returns, $\sigma_t \epsilon_t$ because σ_t is a constant
- What about the *unconditional* distribution of the returns?
- We show in the Appendix that

$$\text{Kurtosis} = \frac{3(1-\alpha^2)}{1-3\alpha^2} > 3 \text{ if } 3\alpha < 1$$

A common misunderstanding

- Returns can be both fat and normal
- That is, unconditionally fat
- And conditionally normal
- Be very careful not to confuse these two
- A very common reason for losing points in exams and assignments
- Returns can also be conditionally normal and unconditionally normal
- And conditionally fat and unconditionally fat

Parameter Restrictions for ARCH(1)

- The ARCH(1) model has two parameters, ω and α
- Can we allow those two to take any value on the real line?
- No; there are two restrictions on the values the parameters can take
- One we always impose, and the other sometimes
- To ensure positive volatility, ensure both parameters are positive

$$\alpha > 0, \quad \omega > 0$$

Preventing Explosions – Stationarity

- Suppose $\alpha > 1$ then we expect σ_t to become bigger and bigger over time
- Which would mean that the unconditional variance is undefined

$$\sigma^2 = \frac{\omega}{1-\alpha}$$

- Because, something to the power 2 cannot be negative (these are not complex numbers)
- We might think therefore to restrict α to be less than one

$$0 < \alpha < 1$$

- This is not needed except in special circumstances (see discussion on GARCH below)

Generalised ARCH (GARCH)

Generalised ARCH (GARCH)

- ARCH models have significant limitations and are rarely used in practice
- The reason is that it needs to use information from many days before t to calculate volatility on day t
- That is, it needs a lot of lags
- The solution is to write it as an ARMA *type* model
- That is, add one component to the equation, $\beta\sigma_{t-1}^2$

GARCH Equation

GARCH(L_1,L_2)

$$\sigma_t^2 = \omega + \sum_{i=1}^{L_1} \alpha_i y_{t-i}^2 + \sum_{j=1}^{L_2} \beta_j \sigma_{t-j}^2$$

GARCH(1,1)

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$$

- GARCH(1,1) is the most common specification

GARCH(1,1) Unconditional Volatility

- The unconditional volatility is the unconditional expectation of volatility on a given day

$$\sigma^2 = \mathrm{E}[\sigma_t^2]$$

- So plug in the parameters

$$\sigma^2 = \mathrm{E}(\omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2) = \omega + \alpha \sigma^2 + \beta \sigma^2$$

- And solve

$$\sigma^2 = \frac{\omega}{1 - \alpha - \beta}$$

Parameter Restrictions

- To ensure positive volatility forecasts

$$\omega, \alpha, \beta \geq 0$$

- Because if any parameter is negative, σ_{t+1} may be negative

Stationarity

- Should we impose

$$\alpha + \beta < 1$$

- Because

$$\sigma^2 = \frac{\omega}{1 - \alpha - \beta}$$

- Not advisable except in special circumstances for two reasons
  1. Can lead to multiple parameter combinations satisfying the constraint so volatility forecasts can be non-unique
  2. Model is misspecified, and the non-restricted model could give more accurate forecasts

EWMA Unconditional Variance

- EWMA is

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

- The GARCH model is

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2$$

- And GARCH becomes EWMA when $\omega = 0, \beta = \lambda, \alpha = 1 - \lambda$
- So for EWMA

$$\sigma^2 = \frac{0}{0}$$

- In other words, EWMA has no unconditional variance

tGARCH

Conditional Distributions

- The conditional distribution in the GARCH model in the previous section is the normal

$$\epsilon_t \sim \mathcal{N}(0,1)$$

- We showed in Chapter 1 that *unconditional returns* are fat and we also showed in a previous section that the *unconditional returns* in GARCH are fat even if the conditional distribution is normal
- That leaves the question of whether they are fat enough, which becomes important in the risk measure chapters later in the book
- We can make the GARCH model *fatter* by using a different conditional distribution

Student-t GARCH (tGARCH)

- As we discussed in the previous chapter, the Student-t distribution is fat, where the degrees of freedom parameter – ν – controls the *conditional fatness*
- If $\nu = \infty$ the Student becomes the normal
- The Student-t GARCH then replaces the normal innovation distribution with the Student

$$\epsilon_t \sim t_{(\nu)}$$

- When it comes to estimation, ν becomes yet another parameter to be estimated along with the three GARCH parameters

Standardised or Non-Standardised t

- An important complication is that by convention the Student-t in tGARCH uses the *standardised* Student-t density, that is, one where it is standardised to have *variance=1*
- While the Student-t density density has *variance*$=\nu/\nu-2$
- That becomes important later
- The second moment (variance) is not defined for $\nu \leq 2$, so the above needs $\nu > 2$

Skew t

- One variant of the Student-t is skewed Student-t
- That is, one side of the distribution is fatter than the other
- For example, there might be a bigger chance of large losses than large gains

Downsides

- There are two downsides to tGARCH
  1. The tGARCH needs more observations in estimation than the normal GARCH, usually at least several thousand
  2. The ν parameter is often estimated with high standard error so it is imprecise and can move around when we use estimation windows in risk forecasting

Asymmetric Power GARCH — apARCH

Leverage Effect

- If the price of equity falls, the company's *debt to equity* ratio increases and the company becomes riskier as a consequence
- We might, therefore, expect the volatility to increase
- That is known as the *leverage effect*
- Stock returns are negatively correlated with changes in volatility
- However, the standard GARCH model assumes symmetry

- So separate out the impact of the positive and negative returns with an extra parameter, ζ

$$\sigma_t^2 = \omega + \alpha \left(|y_{t-1}| - \zeta y_{t-1} \right)^2 + \beta \sigma_{t-1}^2$$

- If $\zeta = 0$ this model reduces to the standard GARCH model

Power Effect

- In the standard GARCH model, the power on lagged returns and volatility is 2

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$$

- However, there is no reason to believe that it should be 2, and it is sometimes found in estimation that the GARCH model is improved if the power is different
- In the *power GARCH* we also estimate the parameter for the power, δ

$$\sigma_t^2 = \omega + \alpha |y_{t-1}|^\delta + \beta \sigma_{t-1}^\delta$$

- If

$$\delta \neq 2$$

- The model has power effects

Asymmetric Power GARCH – apARCH

- These two effects come together in the *asymmetric power* GARCH, or apARCH, model

$$\sigma_t^2 = \omega + \alpha \left(|y_{t-1}| - \zeta y_{t-1} \right)^{\delta} + \beta \sigma_{t-1}^{\delta}$$

- The model allows for leverage effects when $\zeta \neq 0$ and power effects when $\delta \neq 2$
- This model can be difficult to estimate and typically requires thousands of observations

Mean

What About the Mean?

- In the standard GARCH model the mean is assumed to be zero

$$y_t \sim (0, \sigma_t^2)$$

- What is often done in practice is to subtract the mean (de-mean) from the returns prior to estimation
- Usually best for risk applications
- However, in other applications, like price forecasting, it can be beneficial to forecast the mean
- While there are many complicated ways to do so, two are relatively simple and common

(G)ARCH in Mean

- Incorporating a mean

$$y_t = \mu_t + \sigma_t \epsilon_t$$

- Return can be correlated with volatility

$$\mu_t = a\sigma_t$$

- Where a captures the impact of volatility on the mean
- And is estimated as an extra parameter

(G)ARCH in Mean With ARMA

- We can also make the mean follow some autoregressive process like ARMA (autoregressive–moving-average)

$$\mu_t = a + by_{t-1} + c\sigma^2_{t-1}$$

- This is built into the rugarch package

Regressors

GARCH with External Regressors — GARCH-X

- Standard GARCH models use only past shocks and past volatility
- Financial volatility can also be influenced by external variables, like
  - Trading volume
  - News sentiment
  - Geopolitical risk
  - Policy announcements, like tariffs
  - Environmental events
- These can be added to the variance equation as known exogenous inputs

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma x_t$$

Why Use External Regressors?

- External signals can:
  - Improve volatility forecasts
  - Capture non-market sources of uncertainty
  - Help explain volatility during abnormal events
- Questions we can ask:
  - Do macro or geopolitical shocks increase volatility?
  - Are market reactions asymmetric across event types?
  - Did an oil spill affect markets?
  - How do the markets react to the Trump tariff announcements?
- Regressors must be known at time t — no future data allowed
- GARCH-X models give context to volatility — not just memory

Sentiment as a Volatility Driver

- Market sentiment reflects the overall mood or tone in the market
  - News headlines
  - Analyst commentary
  - Social media signals (e.g., Twitter, Reddit)
- Sentiment scores can be constructed using perhaps
  - Textual analysis
  - Machine learning / NLP models
  - Proprietary indices (e.g., Bloomberg, S&P Global, ...)
- In GARCH-X:

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma \, \text{Sentiment}_t$$

Sentiment helps explain volatility driven by perception, not just fundamentals.

Geopolitical Risk, Tariffs and Volatility

- GPR_t geopolitical uncertainty
- Trump's 2025 tariffs — D_t is 1 or 0 depending on whether an announcement was made

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma_1 \text{GPR}_t + \gamma_2 D_t$$

Geopolitical Risk Indices

Several indices measure geopolitical risk

- Caldara & Iacoviello GPR Index — based on news-based economic narratives
- Bloomberg Geopolitical Risk Index — uses proprietary media signals
- BlackRock Geopolitical Risk Dashboard
- GeoQuant Political Risk Scores — AI-driven quantitative assessments

EWMA

EWMA

- Volatility a *weighted* sum of past returns, with weights w_i

$$\hat{\sigma}_t^2 = w_1 y_{t-1}^2 + w_2 y_{t-2}^2 \ldots + w_{W_E} y_{t-W_E}^2$$

- Let the weights be *exponentially declining*, and denote them by w^i, start by:

$$\hat{\sigma}_t^2 = w y_{t-1}^2 + w^2 y_{t-2}^2 \ldots + w^{W_E} y_{t-W_E}^2$$

- $1 > w > 0$
- If W_E is large enough, the terms w^n are negligible for all $n \geq W_E$
- So set $W_E = \infty$

Deriving the EWMA Model

- Let weight be an arbitrary $1 > \lambda > 0$

$$\hat{\sigma}_t^2 \propto \lambda y_{t-1}^2 + \lambda^2 y_{t-2}^2 \ldots + \lambda^{W_E} y_{t-W_E}^2$$

- Sum of an infinite power series

$$\text{Sum} = \frac{\lambda}{1-\lambda} = \lambda + \lambda^2 + \lambda^3 + \ldots + \lambda^\infty$$

- So

$$\hat{\sigma}_t^2 = \frac{1}{\text{Sum}} \left(\lambda y_{t-1}^2 + \lambda^2 y_{t-2}^2 \ldots + \lambda^\infty y_{t-\infty}^2 \right)$$

- We get

$$\hat{\sigma}_t^2 = \frac{1}{\text{Sum}} \sum_{i=1}^{\infty} \lambda^i y_{t-i}^2 = \frac{1-\lambda}{\lambda} \sum_{i=1}^{\infty} \lambda^i y_{t-i}^2$$

Deriving the EWMA Model (cont.)

- Rewriting

$$\hat{\sigma}_t^2 = (1-\lambda)y_{t-1}^2 + \frac{1-\lambda}{\lambda}\sum_{i=2}^{\infty}\lambda^i y_{t-i}^2$$

- Since

$$(1-\lambda)\sum_{i=1}^{\infty}\lambda^i = (1-\lambda)(\lambda^1 + \cdots + \lambda^{\infty}) = \lambda$$

- We get the EWMA equation

$$\hat{\sigma}_t^2 = (1-\lambda)y_{t-1}^2 + \lambda\hat{\sigma}_{t-1}^2$$

Back to main EWMA slides

Fat Tails

Some Mathematics of Moments

- The expected value (unconditionally) of a time dependent variable y_t is:

$$\mathrm{E}(y^m) = \mathrm{E}\left(\mathrm{E}_t(y^m)\right) = \mathrm{E}(y_t^m)$$

for all t. Therefore when $\mu = 0$:

$$E(y^2) = \sigma^2 = \mathrm{E}\left(y_t^2\right)$$

Some Mathematics of Moments (cont.)

- If we write that in terms of residuals

$$\mathrm{E}(y^2) = \mathrm{E}\left(\sigma_t^2 \epsilon_t^2\right) = \mathrm{E}\left(\sigma_t^2\right)$$

- Then:

$$\sigma^2 = \mathrm{E}(\omega + \alpha y_{t-1}^2) = \omega + \alpha\sigma^2$$

- Because the parameters are constant
- So, the *unconditional* volatility of the ARCH(1) model is:

$$\sigma^2 = \frac{\omega}{1-\alpha}$$

ARCH(1) Fat Tails (cont.)

- The 4th moment is

$$\mathrm{E}(y^4) = \mathrm{E}(y_t^4) = \mathrm{E}(\sigma_t^4 \epsilon_t^4)$$

- Because of the normality of ϵ_t and because it has variance one

$$\mathrm{E}[\epsilon_t^4] = 3$$

- Then the 4th moment of y_t is

$$3\,\mathrm{E}[\sigma_t^4]$$

ARCH(1) Fat Tails (cont.)

- Recall the definition of kurtosis

$$\text{Kurtosis} = \frac{\mathrm{E}\left(y^4\right)}{\left(\mathrm{E}\left(y^2\right)\right)^2} = \frac{\mathrm{E}\left(y^4\right)}{\sigma^4}$$

- Plug in the ARCH parameters

$$\begin{aligned}\mathrm{E}(y^4) &= 3\,\mathrm{E}\left(\left(\omega + \alpha y_{t-1}^2\right)^2\right) \\ &= 3\omega^2 + 6\alpha\omega\,\mathrm{E}(y^2) + 3\alpha^2\,\mathrm{E}(y^4) \\ &= 3\omega^2 + 6\alpha\omega\frac{\omega}{1-\alpha} + 3\alpha^2\,\mathrm{E}(y^4)\end{aligned}$$

ARCH(1) Fat Tails (cont.)

- Then

$$\begin{aligned} \mathsf{E}(y^4)(1-3\alpha^2) &= 3\omega^2 + 6\alpha\omega\frac{\omega}{1-\alpha} \\ &= 3\frac{\omega^2(1+\alpha)}{1-\alpha} \end{aligned}$$

ARCH(1) Fat Tails (cont.)

- Solve for

$$\begin{aligned} E(y^4) &= \frac{3\omega^2(1+\alpha)}{(1-\alpha)(1-3\alpha^2)} \\ &= \frac{3\sigma^4(1-\alpha^2)}{1-3\alpha^2} \end{aligned}$$

- If that exceeds three, the unconditional fatness of the returns, y, must be fat

$$\text{Kurtosis} = \frac{3\left(1-\alpha^2\right)}{1-3\alpha^2} > 3 \quad \text{if } 3\alpha^2 < 1.$$

Back to main ARCH slides