Financial Risk Forecasting
Chapter 2
Univariate Volatility Modeling

Jon Danielsson ©2020
London School of Economics

To accompany
Financial Risk Forecasting
www.financialriskforecasting.com
Published by Wiley 2011
Version 5.0, August 2020
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 the estimation and forecasting of volatility for a single asset (univariate)
- The next Chapter 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
- Implied volatilities (like VIX)
- Realized volatilities
Notation

- $W_E$ Estimation window
- $\lambda$ Decay factor in EWMA
- $Z_t$ Residuals
- $\omega, \alpha, \beta$ Main model parameters
- $\zeta, \delta$ Other model parameters
- $L_1, L_2$ Lags in volatility models
Simple models
The mean of returns

• Recall from the last Chapter that the mean of daily returns is very low
• For www.financialriskforecasting.com/data/sp-500.csv
• It 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
• 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

\[
\hat{\sigma}^2 = \frac{1}{T-1} \sum_{i=1}^{T} (x_i - \bar{x})^2
\]

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

\[
\hat{\sigma}^2 = \frac{1}{T} \sum_{i=1}^{T} x_i^2
\]
Moving average

- The simplest volatility model is moving average
- Where the conditional volatility is the average sum of squared returns over the estimation window, $W_E$

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

- 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

- 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
Hypothetical returns
Hypothetical returns

![Graph showing hypothetical returns over time. The graph displays a series of returns ranging from -40 to 40, with a horizontal line at 12 indicating a specific threshold. The x-axis represents days from 0 to 250, and the y-axis represents returns from -40 to 40.](image)
Hypothetical returns
Hypothetical returns

Day

Return

Simple

ARCH

GARCH

Extensions

ML

Code

Sim

Diagnosis

Alt

Sum

Covid

Financial Risk Forecasting © 2011-2020 Jon Danielsson, page 17 of 143
Hypothetical returns

The graph shows hypothetical returns over a period of 250 days. The returns are represented on the y-axis, ranging from -40 to 40, and the days on the x-axis, starting from 0 to 250. The returns exhibit volatility and are marked with specific values at certain days, indicated by numbers on the graph.
Two concepts of volatility

• Unconditional volatility ($\sigma$)
  volatility over an entire time period

• Conditional volatility ($\sigma_t$)
  volatility conditional on a given time period, the past history, model and model parameters
• 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 \textit{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 \textit{exponentially declining}, and denote them by \( \lambda^i \):

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

• \( 1 > \lambda > 0 \)

• If \( W_E \) is large enough, the terms \( \lambda^n \) are negligible for all \( n \geq W_E \)

• So set \( W_E = \infty \)

Financial Risk Forecasting © 2011-2020 Jon Danielsson, page 21 of 143
Deriving the EWMA model

- Sum of an infinite power series
  \[
  \text{Sum} = \frac{\lambda}{1 - \lambda} = \lambda + \lambda^2 + \lambda^3 + \ldots + \lambda^\infty
  \]

- Then
  \[
  \hat{\sigma}_t^2 = \lambda y_{t-1}^2 + \lambda^2 y_{t-2}^2 \ldots + \lambda^\infty y_{t-\infty}^2
  \]

- Then
  \[
  \hat{\sigma}_t^2 = \frac{1 - \lambda}{\lambda} \sum_{i=1}^{\infty} \lambda^i y_{t-i}^2
  \]
• 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

\[ \hat{\sigma}_t^2 = (1 - \lambda)y_{t-1}^2 + \lambda\hat{\sigma}_{t-1}^2 \]
What is $\lambda$?

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

- 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 $\lambda$ with maximum likelihood (See 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
sigma=vector(length=N)
sigma[1]=1
for (i in 2:N){
  sigma[i]=lambda* sigma[i-1] +
           (1-lambda)*y[i-1]^2
  y[i]=y[i]*sqrt(sigma[i])
}
plot(y,type='l')
plot(sigma,type='l',log='y')
Simulate EWMA
Conclusion from simulation

- As we simulated longer and longer paths
- The $\sigma$ 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 multivariate (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 GARCH Family
ARCH and GARCH

- Robert Engle proposed a model in 1982 called autoregressive conditionally heteroscedastic (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
  \[ Y_t = \sigma_t Z_t, \quad Z_t \sim \mathcal{N}(0, 1) \]
- Where \( Z_t \) 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)
    \[
    \sigma_t^2 = \omega + \alpha y_{t-1}^2
    \]
  - $\omega, \alpha$ are parameters to be estimated with maximum likelihood
Moments

- Moments are the expected value of a random variable to some power
- So the $m$th moment is

$$E[X^m] = \int_{-\infty}^{\infty} x^m \, dx$$

- The mean is the first moment

$$\mu := E[X] = \int_{-\infty}^{\infty} x \, dx$$

- The variance is a function of the first and the second moment

$$E[(X - \mu)^2] = E[X^2] + \mu^2 - 2\mu E[X]$$
Some mathematics of moments

• The expected value (unconditionally) of a time dependent variable $Y_t$ is:

$$E(Y^m) = E(E_t(Y^m)) = E(Y_t^m)$$

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

$$E(Y^2) = \sigma^2 = E(Y_t^2)$$
Some mathematics of moments

• If we write that in terms of residuals

\[ E(Y^2) = E(\sigma_t^2 Z_t^2) = E(\sigma_t^2) \]

• Then:

\[ \sigma^2 = E(\omega + \alpha Y_{t-1}^2) = \omega + \alpha \sigma^2 \]

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

\[ \sigma^2 = \frac{\omega}{1 - \alpha} \]
Fat tails

- Recall the discussion in the last chapter which said that a distribution is fat tailed if it has more extreme outcomes than a normal with the same mean and variance.
- Kurtosis is normalized fourth moment:
  \[
  \text{Kurtosis} = \frac{E(Y^4)}{(E(Y^2))^2}
  \]
- Kurtosis is 3 for the normal.
- Regardless of what the mean and variance are.
ARCH(1) fat tails

- The most common distributional assumption for residuals $Z$ is standard normality; that is:
  
  \[ Z_t \sim N(0, 1) \]

- Because $Z_t$ is normal, so must conditional returns, $\sigma_t Z_t$ because $\sigma_t$ is a constant

- What about the *unconditional* distribution of the returns?
The 4th moment is

$$E(Y^4) = E(Y_t^4) = E(\sigma_t^4 Z_1^4)$$

Because of the normality of $Z_t$ and because it has variance one

$$E[Z_t^4] = 3$$

Then the 4th moment of $Y_t$ is

$$3 E[\sigma_t^4]$$
• Recall the definition of kurtosis

\[ \text{Kurtosis} = \frac{E(Y^4)}{(E(Y^2))^2} = \frac{E(Y^4)}{\sigma^4} \]

• Plug in the ARCH parameters

\[ E(Y^4) = 3E \left( (\omega + \alpha Y_{t-1}^2)^2 \right) \]
\[ = 3\omega^2 + 6\alpha\omega E(Y^2) + 3\alpha^2 E(Y^4) \]
\[ = 3\omega^2 + 6\alpha\omega \frac{\omega}{1 - \alpha} + 3\alpha^2 E(Y^4) \]

• Then

\[ E(Y^4)(1 - 3\alpha^2) = 3\omega^2 + 6\alpha\omega \frac{\omega}{1 - \alpha} \]
\[ = 3 \frac{\omega^2(1 + \alpha)}{1 - \alpha} \]
• Solve for

\[
E(Y^4) = \frac{3\omega^2(1 + \alpha)}{(1 - \alpha)(1 - 3\alpha^2)} = \frac{\sigma^2(1 - \alpha^2)}{1 - 3\alpha^2}
\]

• If that exceeds three then \( Y_t \) must be fat tailed

\[
\text{Kurtosis} = \frac{3(1 - \alpha^2)}{1 - 3\alpha^2} > 3 \quad \text{if} \quad 3\alpha^2 < 1.
\]
Extreme value analysis (see Chapter 9)

- The table below shows that the higher the $\alpha$ the fatter the tails (lower tail index $\iota$ implies fatter tails)

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>0.10</th>
<th>0.50</th>
<th>0.90</th>
<th>0.99</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\iota$</td>
<td>26.48</td>
<td>4.73</td>
<td>2.30</td>
<td>2.02</td>
</tr>
</tbody>
</table>
Parameter restrictions for ARCH(1)

- The ARCH(1) model has two parameters, $\omega$ and $\alpha$
- 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 $\sigma_t$ to become bigger and bigger over time

• Which would mean that the unconditional variance undefined

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

• Because, something to the power to cannot be negative

• We might think therefore to restrict $\alpha$ to be less than one

$$0 < \alpha < 1$$

• This is not needed except in special circumstances (See discussion on GARCH below)
Generalized ARCH (GARCH)

• As it turns out, ARCH is not a very good model and almost nobody uses it
• 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 model
• That is, add one component to the equation, \( \beta \sigma^2_{t-1} \)
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

\[ \sigma^2 = \mathbb{E}[\sigma_t^2] \]

- So plug in the parameters

\[ \sigma^2 = \mathbb{E}(\omega + \alpha Y_{t-1}^2 + \beta \sigma_{t-1}^2) = \omega + \alpha \sigma^2 + \beta \sigma^2 \]

- 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 \( \sigma_{t+1} \) 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 2 reasons

1. Can lead to multiple parameter combinations satisfying the constraint so volatility forecasts can be non-unique

2. Model is misspecified anyways and the non-restricted could give more accurate forecasts
**EWMA unconditional variance**

- EWMA is
  
  \[ \sigma_t^2 = (1 - \lambda)y_{t-i}^2 + \lambda \sigma_{t-1}^2 \]

- 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
  
  \[ \sigma^2 = \frac{0}{0} \]
Meaning of parameters

- $\alpha$ is news — how volatility reacts to new information
- $\beta$ is memory — how much volatility remembers from the past
- The size of $(\alpha + \beta)$ determines how quickly the predictability (memory) of the process dies out:
  - If $(\alpha + \beta)$ is close to zero, predictability will die out very quickly
  - If $(\alpha + \beta)$ is close to one, predictability will die out slowly
GARCH news and memory

- \( \text{alpha}=0.1, \ \text{beta}=0.8 \)
  - Initial peak
  - Slow decay

- \( \text{alpha}=0.15, \ \text{beta}=0.8 \)
  - Larger initial peak
  - Steep decay

- \( \text{alpha}=0.1, \ \text{beta}=0.5 \)
  - Initial peak
  - Slow decay

- \( \text{alpha}=0.15, \ \text{beta}=0.5 \)
  - Larger initial peak
  - Steep decay
Half-life

- In physics, half-life indicates how quickly a radioactive material decays to have the radiation.
- The half-life of GARCH is how quickly a shock to volatility dies out to have the impact of the shock.
- You can see that labelled on the next figure.
GARCH news and memory

\(\alpha = 0.1, \beta = 0.8\)

\(\alpha = 0.1, \beta = 0.5\)

\(\alpha = 0.15, \beta = 0.8\)

\(\alpha = 0.15, \beta = 0.5\)


GARCH Half-life

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

As

\[
E_{t-1}[y_t^2] = \sigma_t^2
\]

Then

\[
y_t^2 = \omega + (\alpha + \beta)y_{t-1}^2
\]

Subtract \( \sigma^2 = \omega/(1 - \alpha - \beta) \) from both sides

\[
y_t^2 - \sigma^2 = \omega + (\alpha + \beta)y_{t-1}^2 - \frac{\omega}{1 - \alpha - \beta}
\]

\[
=(\alpha + \beta)(y_{t-1}^2 - \sigma^2)
\]
• Iterate $n$ times then

$$y_{t+n}^2 - \sigma^2 = (\alpha + \beta)^n(y_{t-1}^2 - \sigma^2)$$

• The number of periods, $n^*$, it takes for conditional variance to revert back halfway towards unconditional variance

$$\sigma_{t+n^*}^2 - \sigma^2 = \frac{1}{2}(\sigma_{t+1}^2 - \sigma^2)$$

• So we want to solve for $n^*$ as a function of the parameters
• So

\[(\alpha + \beta)^{n^*-1}(\sigma^2_{t+1} - \sigma^2) = \frac{1}{2}(\sigma^2_{t+1} - \sigma^2)\]

• So

\[n^* = 1 + \frac{\log \left( \frac{1}{2} \right)}{\log(\alpha + \beta)}\]

• And as \((\alpha + \beta) \to 1, n^* \to \infty\), memory is infinite
Other GARCH type models
Conditional distributions

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

\[ Z_t \sim N(0, 1) \]

- We showed in Chapter 1 that unconditional, returns are fat and we also showed in the previous section that the unconditional returns in GARCH are fat even if the condition 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 — $\nu$ — controls the fatness
• If $\nu = \infty$ the Student becomes the normal
• The Student–t GARCH then replaces the normal innovation distribution with the Student

$$Z_t \sim t(\nu)$$

• When it comes to estimation, $\nu$ becomes yet another parameter to be estimated along with the three GARCH parameters
There are two downsides to this

1. The tGARCH needs more observation in estimation, several thousand
2. The $\nu$ parameter is often estimated with high standard error so it is imprecise and can move around when we use estimation windows in risk forecasting
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 from the returns prior to estimation
- However, in many 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

\[ Y_t = \mu_t + \sigma_t Z_t \]

- Return can be positively related to volatility

\[ \mu_t = \delta \sigma_t^2 \]

- Where \( \delta \) captures the impact of volatility on the mean
- We can also make the mean follow some autoregressive process like ARMA

\[ \mu_t = a + b y_{t-1} + c \epsilon_{t-1} \]

- This is build into the rugarch package
- We will evaluate the quality of the mean in the chapter on backtesting
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 therefore might expect the volatility to increase.
- That is known as the leverage effects.
- 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 $\zeta$

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

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

• In the standard GARCH model the power on lagged 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, $\delta$

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

• If

$$\delta \neq 2$$

• The model has power effects
Asymmetric power GARCH – APARCH

- These two affects come together in the APARCH model
  \[ \sigma_t^2 = \omega + \alpha (|Y_{t-1}| - \zeta Y_{t-1})^\delta + \beta \sigma_{t-1}^\delta \]

- The model allows for leverage effects when \( \zeta \neq 0 \) and
  power effects when \( \delta \neq 2 \)

- This model can difficult to estimate and typically requires
  thousands of observations
The news impact curve captures the relationship between shocks at time $t - 1$ ($Z_{t-1}$) to the variance at time $t$, $\sigma_t^2$. 
SP-500

- 10,000 observations
- 1980-11-12 - 2020-07-09
- We discuss the significance a bit later

<table>
<thead>
<tr>
<th></th>
<th>( \omega )</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \zeta )</th>
<th>( \delta )</th>
<th>( \nu )</th>
<th>half life</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00000018</td>
<td>0.09</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
<td>48</td>
</tr>
<tr>
<td>2</td>
<td>0.00000010</td>
<td>0.08</td>
<td>0.91</td>
<td></td>
<td></td>
<td>5.9</td>
<td>110</td>
</tr>
<tr>
<td>3</td>
<td>0.00000001</td>
<td>0.06</td>
<td>0.89</td>
<td>0.35</td>
<td>2.6</td>
<td></td>
<td>47</td>
</tr>
</tbody>
</table>
GARCH news and memory (same SP-500)
Maximum Likelihood
Maximum likelihood estimation

- A linear model is

\[ y = a + bx + \epsilon \]

- i.e. \( \epsilon \) relate linearly to the dependent variable \( y \)
- Why we can do OLS
- But GARCH is non-linear
- Do maximum likelihood
- Consistent parameter estimates even if true density is non-normal
What is maximum likelihood?

- Ask the question which parameters most likely generated the data we have.

- Suppose we have a sample of 

\((-0.2, 3, 4, -1, 0.5)\)

- Of the three possibilities, which is most likely?

<table>
<thead>
<tr>
<th>(\mu)</th>
<th>(\sigma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>-2</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>
Normal density

- If 
  \[ x \sim \mathcal{N}(\mu, \sigma^2) \]
- Then its density is 
  \[ f(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left( -\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2} \right) \]
ARCH(1)

- $Z_t$ in ARCH(1) is standard normally distributed

\[
Y_t = \sigma_t Z_t \\
\sigma_t^2 = \omega + \alpha Y_{t-1}^2 \\
Z_t \sim \mathcal{N}(0, 1)
\]
ARCH(1) density

• Presence of lagged returns implies the density function for $t = 1$ is unknown since $y_0$ is unknown.

• The $t = 2$ density is:

$$f(y_2|y_1) = \frac{1}{\sqrt{2\pi(\omega + \alpha y_1^2)}} \exp \left( -\frac{1}{2} \frac{y_2^2}{\omega + \alpha y_1^2} \right)$$

• Joint density

$$\prod_{t=2}^{T} f(y_t|y_{t-1}) = \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi(\omega + \alpha y_{t-1}^2)}} \exp \left( -\frac{1}{2} \frac{y_t^2}{\omega + \alpha y_{t-1}^2} \right)$$
ARCH(1) likelihood function

- The likelihood function is then the density, except we condition the parameters $\theta = (\alpha, \beta)$ on the data

$$\log \mathcal{L}(\theta | \text{data})$$

- The log likelihood function is then

$$\log \mathcal{L} = -\left(\frac{T-1}{2} \log(2\pi) \right)$$

$$- \frac{1}{2} \sum_{t=2}^{T} \left( \log(\omega + \alpha y_{t-1}^2) + \frac{y_t^2}{\omega + \alpha y_{t-1}^2} \right)$$
GARCH(1,1)

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

$$f(y_2|y_1) = \frac{1}{\sqrt{2\pi(\omega + \alpha y_1^2 + \beta \hat{\sigma}_1^2)}} \exp \left( -\frac{1}{2} \frac{y_2^2}{\omega + \alpha y_1^2 + \beta \hat{\sigma}_1^2} \right)$$

$$\log L = -\frac{T-1}{2} \log(2\pi)$$

$$- \frac{1}{2} \sum_{t=2}^{T} \left( \log(\omega + \alpha y_{t-1}^2 + \beta \hat{\sigma}_{t-1}^2) + \frac{y_t^2}{\omega + \alpha y_{t-1}^2 + \beta \hat{\sigma}_{t-1}^2} \right)$$
Importance of $\sigma_1$

- Value of $\sigma_1$ can make a large difference
- Especially when the sample size is small
- Typically set $\sigma_1 = \hat{\sigma}$
Volatility targeting

• Since we know that
  \[ \sigma^2 = \frac{\omega}{1 - \alpha - \beta} \]

• We can set
  \[ \omega = \hat{\sigma}^2(1 - \alpha - \beta) \]

• where \( \hat{\sigma}^2 \) is the sample variance

• And save one parameter in the estimation
Issues in estimation

- Parameters are obtained by maximizing the likelihood function

$$\hat{\theta}_{\text{ML}} = \max_{\theta \in \Theta} \log L(\theta | y)$$

- Done with an algorithm called optimizer
- There often are numerical problems
Optimizing

- The computer uses an algorithm called optimizer to maximize the likelihood function
- There are many optimizers available
- A classical, and not the best, optimizer is *Newton–Raphson*
Newton–Raphson
Newton–Raphson

\[
\frac{df(x)}{dx} > 0 \quad \frac{d^2f(x)}{dx^2} > 0
\]
Newton–Raphson

\[
\begin{align*}
\frac{df(x)}{dx} &< 0 \\
\frac{d^2f(x)}{dx^2} &< 0
\end{align*}
\]
Newton–Raphson

\[
\frac{df(x)}{dx} > 0 \quad \frac{d^2f(x)}{dx^2} < 0
\]
\[
\begin{align*}
\frac{df(x)}{dx} &= 0 \\
\frac{d^2f(x)}{dx^2} &< 0
\end{align*}
\]
Pathologies

- It is straightforward to optimize the ARCH model
- And the GARCH
- For other models they can be a number of problems
- A common one is multiple local maxima
Pathologies
Pathologies
Pathologies

Parameter value vs. log likelihood

- The graph shows the log likelihood values for various parameter values.
- There is a peak in the likelihood around a parameter value of approximately 100.
- The likelihood decreases as the parameter value moves away from this peak.

Financial Risk Forecasting © 2011-2020 Jon Danielsson, page 90 of 143
Pathologies

![Graph showing log likelihood vs parameter value](image)
Issues

- Problems are rare for smaller models such as GARCH(1,1)
- The more parameters there are, the more likely problems are
- And the more data we need for estimation
- GARCH might need at least 500 observations, Student–t GARCH 3,000
- For the multivariate models discussed in the next chapter estimation problems are common
- For example multiple local minima
Estimation options

• At least two R packages support estimating GARCH style models
• The best is rugarch by Alexios Ghalanos

  cran.r-project.org/web/packages/rugarch/vignettes/Introduction_to_the_rugarch_package

• cran.r-project.org/web/packages/rugarch/rugarch.pdf
• Provides useful plots (try plotting the result)
• And useful information when printing result
Specifying models with `rugarch`

- There are a large number of GARCH type models and specification
- `rugarch` separates out model specification and estimation
- `ugarchspec` creates model, some common options are
  - `variance.model`
  - `mean.model`
  - `distribution.model`
- See next slide
- Then model is estimated with `ugarchfit`
- It returns a structure with estimated parameters, fitted volatility, log likelihood and other useful results
variance.model

- List containing the variance model specification
- Standard GARCH is ’sGARCH’ (default). Some others ”apARCH”
- garchOrder The ARCH and GARCH orders (lags)
- external.regressors
- variance.targeting
- variance.model = list( garchOrder = c(1, 1))
**mean.model**

- List containing the mean model specification
- It is possible to make it a constant or use a more rich model
- For our purposes, we just tell the code not to do it
- `mean.model = list( armaOrder = c(0,0),include.mean = FALSE)`)
distribution.model

- The conditional density to use for the innovations
- The default is "norm" for the normal distribution
- We also use "std" for the student-t
- But there are many others that could be used
- distribution.model = "std"
### R estimation with rugarch

Same SP-500 10,000 observations. 1980-11-12 - 2020-07-09

```r
library(rugarch)

spec1 = ugarchspec(variance.model = list(garchOrder = c(1, 1)),
                    mean.model = list(armaOrder = c(0, 0), include.mean = FALSE))
res1 = ugarchfit(spec = spec1, data = y)

spec2 = ugarchspec(variance.model = list(garchOrder = c(1, 0)),
                    mean.model = list(armaOrder = c(0, 0), include.mean = FALSE))
res2 = ugarchfit(spec = spec2, data = y)

spec3 = ugarchspec(variance.model = list(garchOrder = c(1, 1)),
                    mean.model = list(armaOrder = c(0, 0), include.mean = FALSE),
                    distribution.model = "std")
res3 = ugarchfit(spec = spec3, data = y)

plot(res3)
```
Choosing optimizers

- rugarch allows us to choose optimizers. Try (see the 1,0 in garchOrder)

```r
spec4 = ugarchspec(variance.model = list(garchOrder = c(1, 0)),
                   mean.model = list(armaOrder = c(0, 0), include.mean = FALSE),
                   distribution.model = "std")
res4 = ugarchfit(spec = spec4, data = y)
```

Warning message: In .sgarchfit(spec = spec, data = data, out.sample = out.sample, :
ugarchfit-->warning: solver failure to converge.

- But

```r
res4 = ugarchfit(spec = spec4, data = y, solver="hybrid")
```

- Works

- The “hybrid” strategy solver first tries the “solnp” solver, in failing to converge tries “nlminb”, then “gosolnp” and finally “nloptr” solvers.
Simulation analysis of GARCH models

- We use thousand observations of the S&P 500
- First, estimate a normal GARCH model and include the mean effects
SP-500

Graph showing historical data of the SP-500 index from 2017 to 2020, with a plot of the index values on the left and a plot of the residuals on the right.
SP-500 with 1 simulation
Summary

• The simulated data looks quite different from the true data
• Of course, it should be different but what we’re looking for is the dynamic patterns in it
• Booms and busts
• Volatility clusters
• The following plot shows 10 simulations
SP-500 with 10 simulations
10 parameter estimates from true and simulated data, Normal

<table>
<thead>
<tr>
<th>Sim</th>
<th>ω</th>
<th>α</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>3.8e-06</td>
<td>0.218</td>
<td>0.752</td>
</tr>
<tr>
<td>2</td>
<td>3.9e-06</td>
<td>0.272</td>
<td>0.727</td>
</tr>
<tr>
<td>3</td>
<td>4e-06</td>
<td>0.273</td>
<td>0.701</td>
</tr>
<tr>
<td>4</td>
<td>4.1e-06</td>
<td>0.281</td>
<td>0.718</td>
</tr>
<tr>
<td>5</td>
<td>2.8e-06</td>
<td>0.224</td>
<td>0.753</td>
</tr>
<tr>
<td>6</td>
<td>4.2e-06</td>
<td>0.246</td>
<td>0.716</td>
</tr>
<tr>
<td>7</td>
<td>3.1e-06</td>
<td>0.26</td>
<td>0.725</td>
</tr>
<tr>
<td>8</td>
<td>4.6e-06</td>
<td>0.271</td>
<td>0.703</td>
</tr>
<tr>
<td>9</td>
<td>2.4e-06</td>
<td>0.18</td>
<td>0.798</td>
</tr>
<tr>
<td>10</td>
<td>4.9e-06</td>
<td>0.257</td>
<td>0.712</td>
</tr>
<tr>
<td>11</td>
<td>6.9e-06</td>
<td>0.242</td>
<td>0.687</td>
</tr>
</tbody>
</table>
10 parameter estimates from true and simulated data, t-garch

<table>
<thead>
<tr>
<th></th>
<th>ω</th>
<th>α</th>
<th>β</th>
<th>ν</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>2.2e-06</td>
<td>0.2</td>
<td>0.799</td>
<td>4.118</td>
</tr>
<tr>
<td>2</td>
<td>1.5e-06</td>
<td>0.19</td>
<td>0.809</td>
<td>3.901</td>
</tr>
<tr>
<td>3</td>
<td>2.2e-06</td>
<td>0.231</td>
<td>0.768</td>
<td>4.773</td>
</tr>
<tr>
<td>4</td>
<td>3.2e-06</td>
<td>0.252</td>
<td>0.735</td>
<td>4.317</td>
</tr>
<tr>
<td>5</td>
<td>2e-06</td>
<td>0.204</td>
<td>0.777</td>
<td>5.128</td>
</tr>
<tr>
<td>6</td>
<td>2.8e-06</td>
<td>0.192</td>
<td>0.807</td>
<td>4.139</td>
</tr>
<tr>
<td>7</td>
<td>2.1e-06</td>
<td>0.18</td>
<td>0.796</td>
<td>5.185</td>
</tr>
<tr>
<td>8</td>
<td>9e-07</td>
<td>0.117</td>
<td>0.881</td>
<td>3.928</td>
</tr>
<tr>
<td>9</td>
<td>2e-06</td>
<td>0.179</td>
<td>0.81</td>
<td>4.693</td>
</tr>
<tr>
<td>10</td>
<td>2.1e-06</td>
<td>0.237</td>
<td>0.762</td>
<td>4.32</td>
</tr>
<tr>
<td>11</td>
<td>2.1e-06</td>
<td>0.261</td>
<td>0.738</td>
<td>4.743</td>
</tr>
</tbody>
</table>
Diagnosing Volatility Models
Diagnosing volatility models

- Ideally we would want to know how a model performs operationally.
- But usually we have to make do with an in–sample comparison.
- In out–of–sample forecast comparisons, more parsimonious models perform better even if more flexible model is significantly better in–sample.
- If more flexible model performs better in sample, it is unlikely to perform better out of sample.
Alternatives

1. Likelihood ratio tests
2. Residual analysis

- Do not confuse the discussion here with backtesting as presented later in the book
Likelihood ratio test

- Consider two models, where one is *nested* inside the other
- That is, one is a strict subset of the other
- Like ARCH(1) is a subset of GARCH(1,1)
- Is the same as a standard t-test for one parameter
- But also allows testing of multiple parameters
- Call the nested model *restricted* ($R$) and the other *unrestricted* ($U$)
- By definition

\[ \log \mathcal{L}_R \leq \log \mathcal{L}_U \]
Likelihood ratios

- Parameter value
- Log likelihood
Likelihood ratios

- **Restricted**
- **Unrestricted**

Parameter value vs. log likelihood graph.
Likelihood ratios
Likelihood ratios

\[ \log L_U - \log L_R \]

![Graph showing likelihood ratios for restricted and unrestricted models.](image)
Likelihood ratio test

- 2 times the difference between the unrestricted and restricted likelihoods is $\chi^2$ distributed.
- The degrees of freedom is equal to the number of restrictions.

\[
LR = 2(L_U - L_R) \sim \chi^2_{\text{(number of restrictions)}}
\]

\[
q_{\text{chisq}}(p=1-0.05, df=1) = 3.841459
\]
\( \chi^2 \) in R

\[
\begin{align*}
x &= \text{seq}(0.0, 9, \text{by}=0.1) \\
\text{plot}(x, \text{dchisq}(x, 3), \text{type}='l') \\
\text{plot}(x, \text{pchisq}(x, 3), \text{type}='l') \\
\text{plot}(x, \text{dchisq}(x, 1), \text{type}='l') \\
\text{plot}(x, \text{pchisq}(x, 1), \text{type}='l')
\end{align*}
\]
**Compare GARCH(1,1) to ARCH(1)**

SP-500 10,000 observations 1980-11-12 - 2020-07-09

```r
spec0 = ugarchspec(variance.model = list(garchOrder = c(1, 0)),
                    mean.model = list(armaOrder = c(0, 0), include.mean = FALSE))
res0 = ugarchfit(spec = spec0, data = y, solver="hybrid")
likelihood(res0)

spec1 = ugarchspec(variance.model = list(garchOrder = c(1, 1)),
                    mean.model = list(armaOrder = c(0, 0), include.mean = FALSE))
res1 = ugarchfit(spec = spec1, data = y, solver="hybrid")
likelihood(res1)

LR=(likelihood(res1)-likelihood(res0))
LR
740.9976
pvalue=1-pchisq(LR, 1)
pvalue
0
```
Sample tests

<table>
<thead>
<tr>
<th>Unrestricted</th>
<th>Restricted</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARCH(4)</td>
<td>ARCH(1)</td>
<td>$H_0 : \alpha_2 = \alpha_3 = \alpha_4 = 0$</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>ARCH(1)</td>
<td>$H_0 : \beta = 0$</td>
</tr>
<tr>
<td>GARCH(2,2)</td>
<td>GARCH(1,1)</td>
<td>$H_0 : \beta_2 = \alpha_2 = 0$</td>
</tr>
<tr>
<td>APARCH</td>
<td>GARCH</td>
<td>$H_0 : \delta = 2, \zeta = 0$</td>
</tr>
</tbody>
</table>
SP-500 10,000 observations 1980-11-12 - 2020-07-09

<table>
<thead>
<tr>
<th>Model</th>
<th>$\omega$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\zeta$</th>
<th>$\delta$</th>
<th>$\nu$</th>
<th>lik</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARCH(1)</td>
<td>0.0000861</td>
<td>0.325</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31347</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>0.0000018</td>
<td>0.095</td>
<td>0.891</td>
<td></td>
<td></td>
<td></td>
<td>32693</td>
</tr>
<tr>
<td>tGARCH(1,1)</td>
<td>0.0000010</td>
<td>0.079</td>
<td>0.915</td>
<td></td>
<td></td>
<td>5.926</td>
<td>33012</td>
</tr>
<tr>
<td>APGARCH(1,1)</td>
<td>0.0000001</td>
<td>0.06</td>
<td>0.889</td>
<td>0.352</td>
<td>2.573</td>
<td></td>
<td>32794</td>
</tr>
</tbody>
</table>

The tGARCH it is significantly the best model, followed by the APGARCH. However, it might be worthwhile to estimate a tAPGARCH, which might be significantly better than the others.
Residual analysis

- Test if a single model is correctly specified
- Is based on the residuals (innovations) of the estimated model being what was assumed in the model specification
- For example, for a normal GARCH are the residuals iid normal
Residual analysis of GARCH(1,1)

- Returns have a distribution (here assumed to be normal)

\[ Y_t \sim \mathcal{N}(0, \sigma^2_t) \]
\[ Y_t = \sigma_t Z_t, \quad Z_t \sim \mathcal{N}(0, 1) \]
\[ \sigma^2_t = \omega + \alpha Y^2_{t-1} + \beta \sigma^2_{t-1} \]
\[ \hat{z}_t = \frac{y_t}{\hat{\sigma}_t} \]

- Can test the residuals \( z \) with a Jarque-Bera test for normality and Ljung-Box test for autocorrelation of \( z \) and \( z^2 \). Can also do a QQ plot for \( z \)
Residual tests in R

```r
residuals(res1)

Box.test(residuals(res1), lag = 20,
        type = c("Ljung-Box"))

library(tseries)
jarque.bera.test(residuals(res1))

library(car)
qqPlot(residuals(res1))
qqPlot(residuals(res1), distribution="t", df=4)
```
S&P500 return
Conditional volatility
Conditional volatility and returns

Returns and $\pm 2 \times \sigma_t$
QQ plot of residuals
ACF of squared residuals

![ACF of squared residuals graph]
Residual analysis

\[ \hat{z}_t = \frac{y_t}{\hat{\sigma}_t} \]

- Jarque-Bera test for normality and Ljung-Box test for autocorrelation

<table>
<thead>
<tr>
<th>Model</th>
<th>Jarque-Bera test</th>
<th>Ljung-Box test (20 squared lags)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARCH(1)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>ARCH(4)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>GARCH(4,1)</td>
<td>0.00</td>
<td>0.99</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>0.00</td>
<td>0.53</td>
</tr>
</tbody>
</table>
Alternative volatility models
Realized Volatility

• Based on taking intraday data, sampled at regular intervals and using them to obtain covariance matrix

• Pros
  • Purely data driven and no reliance on parametric methods

• Cons
  • Intraday data must be available, data is difficult to obtain, hard to use and not very clean and expensive
  • Intraday data has complicated patterns that have to be addressed
Implied volatility

- Black Scholes (BS) equation

\[ \text{price} = BS(T, r, S, X, \sigma) \]

- Volatility implied by the price

\[ \sigma = BS^{-1}(T, r, S, X, \text{price}) \]

- Based on current market prices and not historical data
- Depends critically on accuracy of BS model (constant volatility and normal innovations)
VIX

- From the Chicago Board Options Exchange (CBOE)
- Volatility Index — VIX
- It captures the annualized one month volatility of the S&P-500
- Similar to implied volatilities
- It is known as the *Fear Index* of the financial markets
- And even has inspired a thriller by Robert Harris with the same title
- The following figure shows the VIX along with key market events
VIX
GARCH and VIX on SP-500

- Red: GARCH
- Blue: VIX
Summary
Summary

- 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
SP-500 ACF with and without 2020

2016-03-07 to 2020-02-25

2016-07-28 to 2020-07-17
GARCH analysis

- Consider the normal and student to GARCH
- When we include and exclude the crisis period
- The outcome is exactly what we would expect
### SP-500

#### 2016-02-22 to 2020-02-10

<table>
<thead>
<tr>
<th>$\omega$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\nu$</th>
<th>$\alpha + \beta$</th>
<th>Half</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6e-06</td>
<td>0.19</td>
<td>0.73</td>
<td></td>
<td>0.93</td>
<td>9</td>
</tr>
<tr>
<td>2.7e-06</td>
<td>0.19</td>
<td>0.79</td>
<td>4.24</td>
<td>0.97</td>
<td>27</td>
</tr>
</tbody>
</table>

#### 2016-07-14 to 2020-07-02

<table>
<thead>
<tr>
<th>$\omega$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\nu$</th>
<th>$\alpha + \beta$</th>
<th>Half</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.7e-06</td>
<td>0.22</td>
<td>0.75</td>
<td></td>
<td>0.97</td>
<td>23</td>
</tr>
<tr>
<td>2.1e-06</td>
<td>0.22</td>
<td>0.80</td>
<td>3.86</td>
<td>1.02</td>
<td>NA</td>
</tr>
</tbody>
</table>

Why do you think the half-life is NA for the student-t?
News impact (normal GARCH)

2016–02–18 to 2020–02–06
2016–07–12 to 2020–06–30
The graphs show the quantile-quantile plots for the SP-500 index returns from different periods:

1. **2016–02–18 to 2020–02–06**
   - **Sample quantiles**
   - **Theoretical quantiles normal**

   - **Sample quantiles**
   - **Theoretical quantiles t(3)**

The plots compare the sample quantiles (blue dots) against the theoretical quantiles under a normal distribution and a Student's t-distribution with 3 degrees of freedom. The data points are labeled with the number of observations.