Quantitative Finance

Lectures in Quantitative Finance

Spring Term 2022

4. Fundamentals of Financial Econometrics

Prof. Dr. Erich Walter Farkas

walter.farkas@bf.uzh.ch

Universität Zürich[UZH]

ETH zürich

s:fi

Introduction

What is *volatility*?

- The volatility of a financial asset alludes to the fact that asset price and hence the associated return, is *random*.
- The volatility of a stock, σ, is a measure of our uncertainty about the stock returns.
- Technically, the volatility of the asset is the *standard deviation* of the *return distribution*. Stocks typically have a volatility between 20% and 50%.
- A return and its volatility are always expressed relative to a period of time.

Why does *volatility* matter?

- Financial markets are quite often driven by *unanticipated shocks*.
- Investors, traders and asset managers revise their expectations in response to those shocks and *rebalance* positions.
- As a result, we observe *fluctuations* (simply ups and downs) in asset prices. This mechanism holds for various markets including stocks, bonds, FX, commodities, and many others.

Main goal: explain how to use historical data to produce estimates of the current and future levels of volatilities (and correlations).

- Modeling volatility allows us to consider various financial implications:
  - the choice of optimal portfolios,
  - hedging portfolios,
  - Value-at-Risk (VaR) evaluation and forecasting,
  - option pricing,
  - market news-reaction analysis,
  - fundamental (stock) trading and technical trading.
- We consider the following three models:
  1. the *Exponentially Weighted Moving Average (EWMA)* model;
  2. the *AutoRegressive Conditional Heteroscedascity (ARCH)* model;
  3. the *Generalized ARCH (GARCH)* model.

Definition. A sequence (or a vector) of random variables is *homoscedastic* if all its random variables have the *same finite variance*. This is also known as homogeneity of variance.

Definition. A sequence (or a vector) of random variables is *heteroscedastic* if the *variances* (i.e., random disturbances) are *different* across the random variables. Thus, heteroscedasticity is the absence of homoscedasticity.

Definition. *Volatility clustering*: "Large returns tend to be followed by large returns of either sign, and small returns tend to be followed by small returns of either sign", as first noted by Mandelbrot (1963). Quantitatively, while returns (u_t) themselves are uncorrelated, *squared* (or *absolute*) returns display a *positive*, significant, and slowly decaying *autocorrelation function*: $\mathrm{corr}(u_t^2, u_{t+\tau}^2) > 0$ for τ ranging from a few minutes to several weeks.

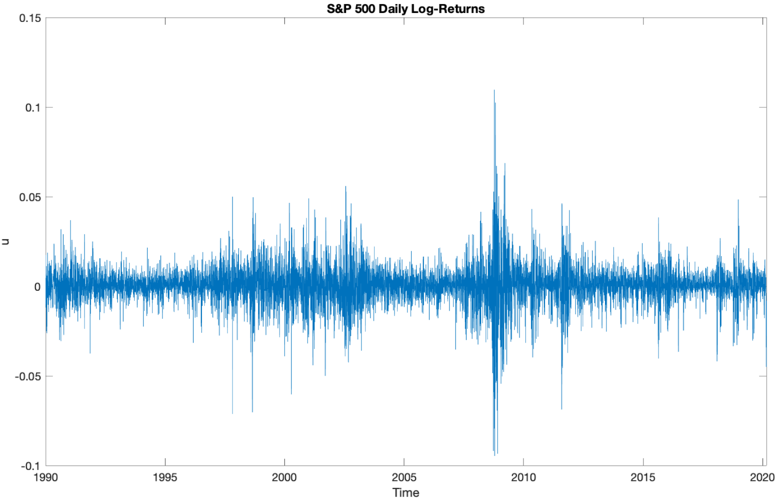

Figure 1: Changing volatility exemplified by the S&P 500 daily log-returns. Observe the heteroscedasticity and clustering of volatility.

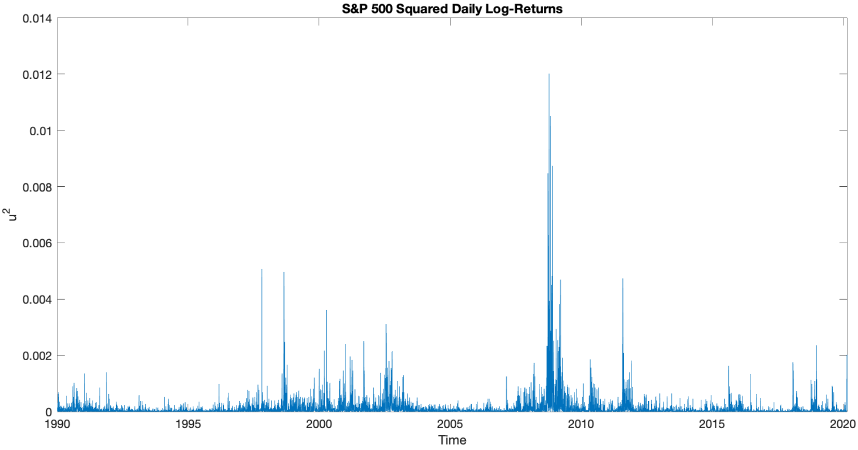

Figure 2: Daily squared log-returns are plotted here. They exhibit heteroscedasticity and positive-autocorrelation (which are caused by the periods with mostly large or small values).

Volatility clustering in the returns

- Simple chart shows the periods with mostly large values and periods with mostly small values - this holds for both returns and squared (or absolute) returns. This is actually a form of *heteroscedasticity*.
- *Volatility clustering*: The volatility *changes over time* and its degree shows a *tendency* to *persist*, i.e., there are periods of low volatility and periods where volatility is high.
- *Large swings*, positive or negative, tend to be *followed* by *large swings*; thus, the *positive autocorrelations* found in the *squared* (or absolute) *returns*.
- The autocorrelations are not large (usually < 0.3), and they decrease more or less slowly with the lag order (depending on the frequency and the series).
- (G)ARCH models were developed to model such stylized facts; *heteroscedasticity* and *volatility clustering*.

Volatility clustering in the returns

- *Financial theory*: the price of an asset is the expected present value of its future income flows.
- An asset price changes because the *expectations* of investors about these future incomes change over time. As time passes, new information (*news*) about these future incomes is released, which modifies the expectations.
- As a result, returns are random and therefore *volatile*.
- Volatility fluctuates over time because the *arrival rate* of *news* fluctuates. For example, *crisis periods* correspond to more *news releases*: in particular bad news tend to spread in clusters.
- *Volatility clustering* is thus due to clusters of arrivals of different types of news.

To estimate the volatility of a stock from (empirical) data, the price is observed at fixed intervals of time (e.g. every day, week, or month).

Consider

$n+1$: number of observations.

S_i : stock price at the end of the i^{th} interval, with $i = 0, 1, ...n$.

τ : length of the time intervals in years (1 month: $\tau = 1/12$ etc.).

and define the *daily log-returns* as

$$u_i = \log\left(\frac{S_i}{S_{i-1}}\right) \quad i = 1, 2, \ldots, n.$$

- An *unbiased* estimator of the variance v of the u_i's is given by:

$$\widehat{v} = \frac{1}{n-1}\sum_{i=1}^{n}(u_i - \overline{u})^2 \,,$$

where $\overline{u}$ is the sample mean of u_i.

- The *annualized volatility* $\widehat{\sigma}$, under the assumption of Gaussianity, can be estimated as

$$\widehat{\sigma} = \frac{\sqrt{v}}{\sqrt{\tau}}.$$

- The *standard error* of this estimate can be shown to be approximately $\frac{\widehat{\sigma}}{\sqrt{2n}}$.

Choosing an appropriate value for n is not easy:

- More data generally leads to more accuracy, but σ does change over time and *old data may not be relevant* for predicting the future volatility.
- A reasonably good *compromise*: use closing prices from daily data over the most recent 90 to 180 days.
- A popular *rule of thumb*: set n equal to the number of days to which the volatility is applied.
- For example, if the volatility estimate is used to value a 2-year option, daily data for the last 2 years are used.

Example: Stock prices over one month

Day	Closing price	S_i/S_{i-1}	$\log(S_i/S_{i-1})$
0	20.00		
1	20.10	1.00500	0.00499
2	19.90	0.99005	0.01000
3	20.00	1.00503	0.00501
4	20.50	1.02500	0.02469
5	20.25	0.98780	-0.01227
6	20.90	1.03210	0.03159
7	20.90	1.00000	0.00000
8	20.90	1.00000	0.00000
9	20.75	0.99282	-0.00720
10	20.75	1.00000	0.00000
11	21.00	1.01205	0.01198
12	21.10	1.00476	0.00475
13	20.90	0.99052	-0.00952
14	20.90	1.00000	0.00000
15	21.25	1.01675	0.01661
16	21.40	1.00706	0.00703
17	21.40	1.00000	0.00000
18	21.25	0.99299	-0.00703
19	21.75	1.02353	0.02326
20	22.00	1.01149	0.01143

Universität Zürich[UZH] ETH zürich s:fi

In this case

$$\sum_{i=1}^{20} u_i = 0.09531 \quad \text{and} \quad \sum_{i=1}^{20} u_i^2 = 0.00326\,.$$

- ▷ The *estimate* of the *standard deviation* of daily returns is

$$\sqrt{\frac{0.00326}{19} - \frac{0.09531^2}{20 \cdot 19}} = 0.01216 \quad \text{(or } 1.216\%\text{)}.$$

- ▷ Assuming that there are 252 trading days per year, i.e., $\tau = 1/252$, an estimate for the *volatility per annum* is

$$0.01216 \times \sqrt{252} = 0.193 \quad \text{(or } 19.3\%\text{)}.$$

▷ The *standard error* of this estimate is

$$\frac{0.193}{\sqrt{2 \times 20}} = 0.031 \quad \text{(or 3.1\% per annum).}$$

▷ With *dividend paying stocks*: the return u_i during a time interval that includes an ex-dividend day is given by

$$u_i = \log \frac{S_i + D_i}{S_{i-1}} ,$$

where D_i is the amount of the dividend paid out at time i.

An important issue is whether time should be measured in *calendar days* or *trading days* when volatility parameters are being estimated and used.

- Practitioners tend to *ignore days on which the exchange is closed* when estimating volatility from historical data (and when calculating the life of an option).
- This may be justified by the empirical evidence suggesting that *volatility* is to a large extent *caused* by *trading* itself (and not only by new information reaching the market).
- The volatility *per annum* (p.a.) is calculated from the volatility per trading day using the formula:

$$\text{Vola p.a.} = \text{Vola per trading day} \times \sqrt{\#\text{ trading days p.a.}}\,.$$

Estimating volatility

Estimating volatility

- Recall, volatility is inherently *not observable* (latent).

- Like returns, volatility also evolves randomly through time. Of course, this does not imply that it is not partially predictable (why?).

- Volatility can be "indirectly" measured through econometric models using *observed* returns. That is, the models we utilize link the "observed" returns to the "unobserved" volatility.

- Inferring volatility from returns is one of the primary research focuses in *financial econometrics*.

- Measures based on the *empirical standard deviation* of most recent returns (i.e., rolling window approach).
- Model-based measures:
  - **E**xponential **W**eighted **M**oving **A**verage model (EWMA);
  - **A**uto**R**egressive **C**onditional **H**eteroskedastic models (ARCH);
  - **G**eneralized **A**uto**R**egressive **C**onditional **H**eteroskedastic models (GARCH);
  - stochastic volatility;
  - stochastic volatility + jumps, only jumps.
- Measures based on *high frequency* returns (so-called model-free), such as *realized volatility*.

- Denote with σ_n the volatility of a market variable on day n, as estimated at the end of day $n-1$.
- The square of the volatility σ_n^2 on day n is the variance rate.
- Recall that the variable u_i is defined as the continuously compounded return between the end of day $i-1$ and the end of day i:

$$u_i = \log \frac{S_i}{S_{i-1}}.$$

- An *unbiased estimate* of the variance rate per day, σ_n^2, using the most recent m observations on the u_i is

$$\sigma_n^2 = \frac{1}{m-1} \sum_{i=1}^{m} (u_{n-i} - \overline{u})^2 \,, \tag{1}$$

where the mean $\overline{u}$ is given by

$$\overline{u} = \frac{1}{m} \sum_{i=1}^{m} u_{n-i}.$$

- For the purpose of monitoring daily volatility the last formula can be changed in a number of ways:

  1. u_i can be defined as the percentage change in the market variable between the end of day $i-1$ and the end of day i, so that
  $$u_i = \frac{S_i - S_{i-1}}{S_{i-1}};$$
  2. One can assume $\overline{u} = 0$;
  3. $m-1$ can be replaced by m.

- These three changes (i.e., assumptions) make very little difference to the calculated estimates, whereas they allow us to *simplify* the formula for the variance rate from Equation (1) that now becomes:

$$\sigma_n^2 = \frac{1}{m} \sum_{i=1}^{m} u_{n-i}^2.$$

- The last expression gives *equal weight* to $u_{n-1}^2, u_{n-2}^2, \ldots, u_{n-m}^2$.
- Note: All expressions that follow could also be derived without assuming $\overline{u} = 0$. We assume this for simplicity and improved intuition.
- Idea: Our objective is to estimate the *current level* of volatility σ_n, therefore, it would make sense to give *more weight* to *recent data*.

- We can accomplish this with the following model,

$$\sigma_n^2 = \sum_{i=1}^{m} \alpha_i u_{n-i}^2 \,. \tag{2}$$

- The coefficient $\alpha_i > 0$ is the *weight* given to the observation i days ago.
- If we choose them so that $\alpha_i < \alpha_j$ when $i > j$, *less weight* is given to *older observations*.
- The weights must *sum up* to *unity*,

$$\sum_{i=1}^{m} \alpha_i = 1.$$

An extension of the idea in Equation (2) is to assume that there is a *long-run average variance rate* and that this should be given some weight.

- This leads to a model that takes the form

$$\sigma_n^2 = \gamma V_L + \sum_{i=1}^{m} \alpha_i u_{n-i}^2 \,, \tag{3}$$

  where V_L is the long-run variance rate and γ is the weight assigned to V_L.

- Because the weights must sum to unity, we have

$$\gamma + \sum_{i=1}^{m} \alpha_i = 1.$$

ARCH(m) model

The representation from Equation (3) is known as the *ARCH(m) model* and it was first suggested by Robert Engle in 1982 in the journal *Econometrica*.

- The ARCH model class for asset returns was designed to *capture* the *dependence* (in the form of positive autocorrelations) in the *squared* (or absolute) *returns* $\Longrightarrow$ *volatility clustering* (a form of heteroscedasticity).

- From a *statistical* viewpoint, taking account of heteroscedasticity provides more *efficient estimates* of the *conditional mean* parameters and more realistic *confidence bands* for the *forecasts*.

- In the ARCH(m) model, the *estimate* of the *variance* is based on a *long-run average variance* and m observations: the *older* an *observation*, the *less weight* it is given.

- Defining $\omega = \gamma V_L$, the ARCH(m) model from Equation (3) can be written as

$$\sigma_n^2 = \omega + \sum_{i=1}^{m} \alpha_i u_{n-i}^2.$$

- This is the version of the model used when the parameters are being estimated.

EWMA model

The *Exponentially Weighted Moving Average* (EWMA) model is a particular case of the model in Equation (2), and consequently also a particular case of an ARCH(1) model with $\gamma = 0$.

- Here, the weights α_i *decrease exponentially* as we move back through time,

$$\alpha_{i+1} = \lambda \alpha_i \, , \text{ where } \lambda \in [0, 1] \, .$$

- It turns out that this weighting scheme leads to a particularly simple formula for *updating volatility estimates*:

$$\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1 - \lambda) u_{n-1}^2 . \quad (4)$$

- The estimate σ_n is the volatility for day n (made at the end of day $n-1$) and is calculated from σ_{n-1} (the estimate that was made at the end of day $n-2$ of the volatility for day $n-1$) and u_{n-1} (the most recent percentage change).

- To understand why Equation (4) corresponds to *weights* that *decrease exponentially*, we substitute for σ^2_{n-1} to get

$$\sigma^2_n = \lambda\left[\lambda\sigma^2_{n-2} + (1-\lambda)u^2_{n-2}\right] + (1-\lambda)u^2_{n-1},$$

or, when rearranged,

$$\sigma^2_n = (1-\lambda)(u^2_{n-1} + \lambda u^2_{n-2}) + \lambda^2\sigma^2_{n-2}.$$

- Substituting in a similar way for σ^2_{n-2} yields

$$\sigma^2_n = (1-\lambda)(u^2_{n-1} + \lambda u^2_{n-2} + \lambda^2 u^2_{n-3}) + \lambda^3\sigma^2_{n-3}.$$

- Continuing in this way, we see that

$$\sigma^2_n = (1-\lambda)\sum_{i=1}^{m}\lambda^{i-1}u^2_{n-i} + \lambda^m\sigma^2_{n-m}.$$

- To repeat, we have

$$\sigma_n^2 = (1-\lambda)\sum_{i=1}^{m}\lambda^{i-1}u_{n-i}^2 + \lambda^m \sigma_{n-m}^2. \tag{5}$$

- Note that for large m the term $\lambda^m \sigma_{n-m}^2$ is *sufficiently small* to be *ignored* so that Equation (5) is the same as Equation (2) with $\alpha_i = (1-\lambda)\lambda^{i-1}$.

- The weights for the u_i *decline* at *rate* λ as we move back through time; each weight is λ times the previous weight.

EWMA model: Example

- Suppose that $\lambda = 0.90$, the volatility estimated for a market variable for day $n-1$ is 1% per day, and during day $n-1$ the market variable increased by 2%.
- This means: $\sigma_{n-1}^2 = 0.01^2 = 0.0001$ and $u_{n-1}^2 = 0.02^2 = 0.0004$.
- Equation (4) yields

$$\sigma_n^2 = 0.9 \times 0.0001 + 0.1 \times 0.0004 = 0.00013.$$

- The *estimate* of the *volatility* σ_n for the day n is therefore $\sqrt{0.00013}$, or 1.14% per day.

EWMA model: Example

- Note that the *expected value* of u_{n-1}^2 is σ_{n-1}^2 or 0.0001.
- In this example, the realized value of u_{n-1}^2 is *greater* than the expected value and as a result our volatility estimate *increases*.
- If the realized value of u_{n-1}^2 had been *less* than its expected value, our estimate of the volatility would have *decreased*.

The EWMA approach has the attractive feature that *relatively little data needs to be stored*.

- At any given time we need to remember only the *current estimate* of the *variance rate* and the *most recent observation* of the value of the *market variable*.
- When we get a new observation of the value of the market variable, we calculate a new daily percentage change and use Equation (4) to *update* our estimate of the variance rate.
- The old estimate of the variance rate and the old value of the market variable can then be *discarded*.

- The EWMA approach is designed to *track changes* in the *volatility*.
- The Risk Metrics database, which was originally created by J. P. Morgan and made publicly available in 1994, uses the EWMA model with $\lambda = 0.94$ for *updating daily volatility estimates*.
- The company found that, across a range of different market variables, this value of λ yields *forecasts* of the *variance* rate that come *closest* to the *realized variance* rate.

GARCH(1,1) model

The *GARCH(1,1) model* was first proposed by Tim Bollerslev in 1986.

Definition. In the most simple GARCH(1,1) model, returns are conditionally *normally distributed*:

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

and the conditional variance σ_t^2 is calculated from a long-run average variance rate V_L, the past variance σ_{t-1}^2, and the past return u_{t-1}:

$$\sigma_t^2 = \underbrace{\gamma \cdot V_L}_{\omega} + \beta \cdot \sigma_{t-1}^2 + \alpha \cdot u_{t-1}^2, \qquad \text{with} \quad 1 = \gamma + \beta + \alpha.$$

- ▷ We have *heteroscedasticity* because σ_t^2 is the conditional (on past returns) variance of u_t:

$$\mathrm{Var}\left(u_t | u_0, \ldots, u_{t-1}\right) = \sigma_t^2.$$

- ▷ The conditional mean $\mathbb{E}(u_t | u_0, \ldots, u_{t-1})$ is assumed equal to zero. Hence, u_t is not autocorrelated and $\mathbb{E}(u_t) = 0$.

- The *EWMA* model is a *particular case* of GARCH(1,1) where $\gamma = 0$, $\alpha = 1 - \lambda$, $\beta = \lambda$.

- The *ARCH(1)* model is a *particular case* of GARCH(1,1) where $\beta = 0$.

- The (1,1) in GARCH(1,1) indicates that σ_t^2 is based on the most recent observation of the (squared) return, namely u_{t-1}^2, and the most recent estimate of the variance rate, namely $\sigma_{t-1}^2 \implies$ generalization to GARCH(p,q).

- The more general GARCH(p,q) model calculates σ_t^2 from the most recent *p observations of u^2* and the most recent *q estimates of the variance rate*. Exercise: Write it down.

- Note: GARCH(1,1) is by far the most popular GARCH model.

- Setting $\omega = \gamma V_L$, the GARCH(1,1) model can also be written as

$$\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2 \,, \tag{6}$$

and this is the form that is usually used for the purposes of estimating the parameters.

- There are *two effects* of a large value of σ_{t-1}^2 in the GARCH(1,1) representation from Equation (6):
  - a *direct* effect: σ_{t-1}^2 is large $\implies$ σ_t^2 is large;
  - an *indirect* effect: σ_{t-1}^2 is large $\implies$ u_{t-1}^2 tends to be large because $u_{t-1} \sim \mathcal{N}(0, \sigma_{t-1}^2)$ $\implies$ σ_t^2 tends to be large.

- Hence, the GARCH(1,1) model captures *volatility clustering.*

- Note that for a *stable GARCH(1,1) process* we require $\alpha + \beta < 1$, otherwise the weight term applied to the long-term variance is negative.

Universität Zürich[UZH] ETH zürich s:fi

- *Volatility clustering* often generates more *extreme values* and less central values compared to case when returns were independent (i.e. if $\alpha = \beta = 0$).
- *Large* (positive or negative) *returns* (u_t) tend to *follow large returns*, *small returns* tend to *cluster* as well. Return observations mingle with each other.
- This gives a *higher proportion* of *extreme returns* (and of returns close to 0) than if returns are independent through time.
- Even if the "conditional" distribution is assumed to be normal, the *"unconditional" distribution* is *not* necessarily *normal* (it has a *larger kurtosis* than the normal).

GARCH(1,1) model: Example

Suppose that a GARCH(1,1) model is estimated from daily data as

$$\sigma_n^2 = 0.000002 + 0.13u_{n-1}^2 + 0.86\sigma_{n-1}^2 \,.$$

- This corresponds to $\alpha = 0.13$, $\beta = 0.86$, and $\omega = 0.000002$.
- Since $\gamma = 1 - \alpha - \beta$, it follows that $\gamma = 0.01$.
- Since $\omega = \gamma V_L$, it follows that $V_L = 0.0002$.
- In other words, the long-run average variance per day implied by the model is 0.0002.
- This corresponds to a volatility of $\sqrt{0.0002} = 0.014$ or 1.4% per day.

Suppose that the estimate of the volatility on the day $n-1$ is 1.6% per day, so that $\sigma_{n-1}^2 = 0.016^2 = 0.000256$, and that on the day $n-1$ the market variable decreased by 1% so that $u_{n-1}^2 = 0.01^2 = 0.0001$.

- Then,

$$\sigma_n^2 = 0.000002 + 0.13 \times 0.0001 + 0.86 \times 0.000256 = 0.00023516.$$

- The new estimate of the volatility is therefore $\sqrt{0.00023516} = 0.0153$ or 1.53%.

- Substituting for σ^2_{n-1} and, afterwards, for σ^2_{n-2} in Equation (6), we obtain

$$\sigma_n^2 = \omega + \beta\omega + \beta^2\omega + \alpha u_{n-1}^2 + \alpha\beta u_{n-2}^2 + \alpha\beta^2 u_{n-3}^2 + \beta^3\sigma_{n-3}^2 .$$

- Continuing this, we see that the weight applied to u^2_{n-i} is $\alpha\beta^{i-1}$.
- The *weights decline exponentially* at rate β.
- The parameter β can be interpreted as a *decay rate*. It is similar to the parameter λ in the EWMA model.
- The GARCH(1,1) model is similar to the EWMA model except that, in addition to assigning weights that decline exponentially to past u_i^2, it also *assigns* some *weight* to the *long-run average volatility*.

The GARCH(1,1) model recognizes that *over time* the variance tends to get pulled back to a *long-run average level* of V_L.

- The amount of weight assigned to V_L is $\gamma = 1 - \alpha - \beta$.
- The *unconditional variance* is given by:

$$\mathrm{Var}[u_t] = V_L = \frac{\omega}{1 - \alpha - \beta} \quad \text{if} \quad \alpha + \beta < 1.$$

- The *conditional variances* (σ_t^2) fluctuate around the unconditional one:

$$V_L = \mathrm{E}(\sigma_t^2).$$

GARCH(1,1) model: Mean reversion

- The GARCH(1,1) is equivalent to a model where the variance V follows the *stochastic process*

$$\mathrm{d}V = a(V_L - V)\mathrm{d}t + \xi V \mathrm{d}z \,,$$

where time is measured in days, $a = 1 - \alpha - \beta$, and $\xi = \alpha\sqrt{2}$. This is the *mean reverting model*.

- The variance has a *drift* that pulls it back to V_L at *rate* a.
- When $V > V_L$, the variance has a *negative drift*, when $V < V_L$ it has a *positive drift*.
- In *practice*, variance rates tend to be *mean reverting*.
- The GARCH(1,1) model *incorporates mean reversion*, whereas the EWMA model *does not*.

GARCH(1,1) model: Parameter estimation

Universität Zürich[UZH] ETH zürich s:fi

- A important question that needs to be discussed is how the best-fit parameters ω, α, β in GARCH(1,1) can be estimated.
- *Maximum likelihood estimation* (MLE) method can be used to obtain GARCH parameter values, see Appendix. Link to Appendix
- GARCH(1,1) is *more general* and hence more appealing than the ARCH(1) and EWMA models.
- However, in circumstances where the best-fit value of ω turns out to be negative, the GARCH(1,1) model is *not stable* and it makes sense to switch to the EWMA model.
- *Trade-off* between model generality and statistical estimation (i.e., parameter uncertainty).

Using GARCH(1,1) to forecast volatility

- The variance rate estimated at the end of day $n-1$ for day n, when GARCH(1,1) is used, is

$$\sigma_n^2 = (1-\alpha-\beta)V_L + \alpha u_{n-1}^2 + \beta\sigma_{n-1}^2$$

so that

$$\sigma_n^2 - V_L = \alpha(u_{n-1}^2 - V_L) + \beta(\sigma_{n-1}^2 - V_L).$$

- On day $n+t$ in the future, we have

$$\sigma_{n+t}^2 - V_L = \alpha(u_{n+t-1}^2 - V_L) + \beta(\sigma_{n+t-1}^2 - V_L),$$

which implies a *recursive scheme* for forecasting volatility.

- The expected value of u_{n+t-1}^2 is σ_{n+t-1}^2, hence

$$\mathbb{E}[\sigma_{n+t}^2 - V_L] = (\alpha + \beta)\mathbb{E}[\sigma_{n+t-1}^2 - V_L].$$

- Using this equation *repeatedly* yields

$$\mathbb{E}[\sigma_{n+t}^2 - V_L] = (\alpha + \beta)^t(\sigma_n^2 - V_L)$$

or

$$\mathbb{E}[\sigma_{n+t}^2] = V_L + (\alpha + \beta)^t(\sigma_n^2 - V_L)\,. \tag{7}$$

- This equation *forecasts* the *volatility* on day $n+t$ using the information available at the end of the day $n-1$.

- In the EWMA model we have $\alpha + \beta = 1$ and the last equation shows that the expected future variance rate *equals* the *current* variance rate.
- When $\alpha + \beta < 1$, the final term in the equation becomes *progressively smaller* as t increases.
- As mentioned earlier, the variance rate exhibits *mean reversion* with a reversion level of V_L and a reversion rate of $1 - \alpha - \beta$.
- Our *forecast* of the future variance rate tends towards V_L as we look further and further ahead.
- This analysis emphasizes the point that we must have $\alpha + \beta < 1$ for a *stable* GARCH(1,1) process.
- When $\alpha + \beta > 1$ the weight given to the long-term average variance is negative and the process is *mean fleeing* rather than mean reverting.

Using GARCH(1,1) to forecast volatility

- In the yen-dollar exchange rate example considered earlier we calculated $\alpha + \beta = 0.9602$ and $V_L = 0.00004422$.
- *Suppose* that our estimate of the current variance rate per day is 0.00006 (this corresponds to a volatility of 0.77% per day).
- In 10 days the *expected variance rate* is

$$0.00004422 + 0.9602^{10}(0.00006 - 0.00004422) = 0.00005473.$$

- The expected volatility per day is 0.0074, still well *above* the long-term volatility of 0.00665 per day.
- However, the expected variance rate in 100 days is

$$0.00004422 + 0.9602^{100}(0.00006 - 0.00004422) = 0.00004449$$

and the expected volatility per day is 0.00667 very *close* to the long-term volatility.

Extensions of GARCH

Many *extensions* of the GARCH(p,q) model have been presented in the existing literature. Some of the most prominent cases are:

- Exponential GARCH (EGARCH), and
- Threshold GARCH (TGARCH).

Universität Zürich[UZH] ETH zürich s:fi

- The *EGARCH model* is a GARCH variant that models the logarithm of the conditional variance.
- It includes a *leverage* term to capture the *asymmetric effects* between positive and negative asset returns.
- The EGARCH(1,1) model takes the following form:

$$\log \sigma_n^2 = \omega + \alpha g\left(\epsilon_{n-1}\right) + \beta \log \sigma_{n-1}^2 \,,$$

where $\epsilon_n = u_n/\sigma_n$ and $g(\epsilon_n) = \theta\epsilon_n + \gamma(|\epsilon_n| - E[|\epsilon_n|])$.

- Since *negative returns* have a *more pronounced effect* on *volatility* than positive returns of the same magnitude, the parameter θ usually takes negative values.

- The *TGARCH model* is a specification of conditional variance.
- Like the EGARCH model it allows positive returns to have a larger/smaller impact on volatility than negative returns.
- The TGARCH(1,1) model has the following form:

$$\sigma_n^2 = \omega + (\alpha + \gamma N_{n-1}) u_{n-1}^2 + \beta \sigma_{n-1}^2 \,,$$

where N_{n-1} is an indicator for negative u_{n-1}, that is

$$N_{n-1} = \begin{cases} 1 & \text{if } u_{n-1} < 0 \,, \\ 0 & \text{if } u_{n-1} \geq 0 \,. \end{cases}$$

Conclusion

Conclusion

- Empirical *stylized facts* of financial returns: Among others, financial time series exhibit *time-varying volatility* and *volatility clustering*, i.e., periods of swings interspersed with periods of relative calm.
- Financial econometrics: (G)ARCH models were developed to model such stylized facts; *heteroscedasticity* and *volatility clustering*.
- The GARCH class of models are the most *general* $\Longrightarrow$ ARCH and EWMA models can be seen as particular cases of the GARCH class.
- However, EWMA can be preffered when *statistical estimation* of the GARCH parameters turns out to be unstable (e.g., the best-fit w is negative). Moreover, the EWMA approach has the attractive feature that relatively *little data* needs to be stored.
- Inferring volatility from returns is one of the primary research focuses in *financial econometrics*.

References

Books:

- Hull, John C., *Options, futures and other derivatives*, Sixth Edition, Prentice Hall, 2006, Chapter 19: Estimating Volatilities, pages 461 - 480.
- Tsay, Ruey S., *Analysis of Financial Time Series*, Third Edition, Wiley, 2010, Chapter 3: Conditional Heteroscedastic Models, pages 143 - 150.

Appendix: Estimation

How can the *maximum likelihood estimation* (MLE) method be used to estimate the parameters when the GARCH(1,1) model (or some other volatility updating scheme)?

- MLE attempts to find the parameter values that maximize the *likelihood function* $\mathcal{L}$, given the observations.
- The resulting estimate is called a *maximum likelihood estimate*.
- Denote $v_i = \sigma_i^2$ as the variance estimated for day i.
- In order to perform the estimation described here, we assume that the *probability distribution* of u_i *conditional* on the *variance* is *normal*.

Use the *maximum likelihood method* to estimate the constant variance v of a $\mathcal{N}(0, v)$ random variable from n observations:

$$u_1, u_2, \cdots, u_n.$$

- The *likelihood* of u_i being observed is given by the normal pdf:

$$\frac{1}{\sqrt{2\pi v}} \exp\left(\frac{-u_i^2}{2v}\right).$$

- The likelihood of n (independent) observations is the *product*:

$$\begin{aligned}\mathcal{L}(v) &= \frac{1}{\sqrt{2\pi v}} \exp\left(\frac{-u_1^2}{2v}\right) \times \cdots \times \frac{1}{\sqrt{2\pi v}} \exp\left(\frac{-u_n^2}{2v}\right) = \\ &= \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi v}} \exp\left(\frac{-u_i^2}{2v}\right)\right].\end{aligned}$$

- The *best estimate* of v is the value that *maximizes* this expression.

- Taking the *logarithm*, we wish to *maximize* the *log-likelihood* function:

$$\log \mathcal{L}(v) = -\frac{1}{2}\sum_{i=1}^{n}\left(\log(v) + \frac{u_i^2}{v}\right).$$

- Under the *assumption* of Gaussianity, the normal log-likelihood function for a sample of n observations is:

$$\log \mathcal{L}(\omega,\, \beta,\, \alpha) = -\frac{1}{2}\sum_{i=1}^{n}\left\{\log(\sigma_i^2) + \frac{u_i^2}{\sigma_i^2}\right\}, \tag{8}$$

  where σ_i^2 is replaced by its chosen specification, for example the GARCH(1,1) model:

$$\sigma_i^2 = \omega + \alpha u_{i-1}^2 + \beta\sigma_{i-1}^2.$$

- We *search iteratively* to find the parameters of the model that maximize the expression in Equation (8). Note that more *efficient* methods exist as well.

Example. The data below shows the exchange rate between the Japanese yen and the US dollar for the time period between January 6th 1988 and August 15th 1997.

Date	Day i	S_i	u_i	$v_i = \sigma_i^2$	$-\log(v_i) - u_i^2/v_i$
06-Jan-88	1	0.007728			
07-Jan-88	2	0.007779	0.006599		
08-Jan-88	3	0.007746	-0.004242	0.00004355	9.6283
11-Jan-88	4	0.007816	0.009037	0.00004198	8.1329
12-Jan-88	5	0.007837	0.002687	0.00004455	9.8568
13-Jan-88	6	0.007924	0.011101	0.00004220	7.1529
...	...	...	...	...	...
13-Aug-97	2421	0.008643	0.003374	0.00007626	9.3321
14-Aug-97	2422	0.008493	-0.017309	0.00007092	5.3294
15-Aug-97	2423	0.008495	0.000144	0.00008417	9.3824
					$\sum = 22063.5763$

Estimating GARCH(1,1) parameters

- The fifth column shows the *estimate* of the variance rate $v_i = \sigma_i^2$ for day i made at the end of day $i-1$.
- On day 3 we start things off by setting the variance equal to u_2^2.
- On subsequent days, we use equation

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2.$$

- The sixth column tabulates the *likelihood* measure $-\log(v_i) - u_i^2/v_i$.
- The values in the fifth and sixth columns are based on the current trial estimates of ω, α and β: we are interested in *maximizing* the *sum* of the members in the sixth column.
- This involves an *iterative search* procedure.

- On subsequent days, we use

$$\omega = 0.00000176, \quad \alpha = 0.0626, \quad \beta = 0.8976.$$

- The numbers shown in the above table were calculated on the *final iteration* of the search for the optimal ω, α, and β.

- The *long-term* variance rate V_L in our example is

$$V_L = \frac{\omega}{1 - \alpha - \beta} = \frac{0.00000176}{0.0398} = 0.00004422.$$

Estimating GARCH(1,1) parameters

- When the EWMA model is used, the *estimation procedure* is relatively simple: we set $\omega = 0$, $\alpha = 1 - \lambda$, and $\beta = \lambda$.
- In the table above, the value of λ that *maximizes* the *objective function* is 0.9686 and the value of the objective function is 21995.8377.
- Both GARCH(1,1) and the EWMA method can be implemented by using the solver routine in Excel to search for the values of the parameters that *maximize* the *likelihood function*.
- However, we suggest you switch from Excel to Python, Matlab, or R.

Appendix: Correlations

The discussion so far has centered on the estimation and forecasting of volatility. The goal of this section is to show how *correlation estimates* can be *updated* in a similar way to volatility estimates.

- Recall that the *covariance* between two random variables X and Y is defined as

$$\mathrm{cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)],$$

where μ_X and μ_Y are the *means* of X and Y, respectively.

- The *correlation* (Pearson's correlation coefficient) between two random variables X and Y is

$$\rho_{X,Y} = \mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y},$$

where σ_X and σ_Y are the *standard deviations* of X and Y, respectively.

- Define x_i and y_i as the percentage changes (simple returns) in X and Y between the end of the day $i-1$ and the end of day i:

$$x_i = \frac{X_i - X_{i-1}}{X_{i-1}} \quad \text{and} \quad y_i = \frac{Y_i - Y_{i-1}}{Y_{i-1}},$$

where X_i and Y_i are the values of X and Y at the end of the day i.

- We also define

$\sigma_{x,n}$: daily volatility of variable X estimated for day n;

$\sigma_{y,n}$: daily volatility of variable Y estimated for day n;

cov_n : estimate of covariance between daily changes in X and Y, calculated on day n.

- Our *estimate* of the *correlation* between X and Y on day n is

$$\frac{\mathrm{cov}_n}{\sigma_{x,n}\sigma_{y,n}}.$$

- Using an *equal-weighting scheme* and assuming that the means of x_i and y_i are zero, we can estimate the *variance* of X and Y from the most recent m observations as

$$\sigma_{x,n}^2 = \frac{1}{m}\sum_{i=1}^{m} x_{n-i}^2 \quad \text{and} \quad \sigma_{y,n}^2 = \frac{1}{m}\sum_{i=1}^{m} y_{n-i}^2 .$$

- A similar estimate for the *covariance* between X and Y is

$$\mathrm{cov}_n = \frac{1}{m}\sum_{i=1}^{m} x_{n-i} y_{n-i} .$$

One *alternative* for *updating covariances* is the EWMA model, as previously discussed.

- The formula for *updating* the *covariance estimate* under EWMA becomes

$$\text{cov}_n = \lambda cov_{n-1} + (1 - \lambda) x_{n-1} y_{n-1}.$$

- A similar analysis to that presented for the EWMA volatility model shows that the weights given to observations on the x_i and y_i *decline* as we move back through time.

- The *lower* the value of λ, the *greater* the weight that is given to *recent observations*.

- Assume $\lambda = 0.95$ and that the estimate of the correlation between two random variables X and Y on the day $n-1$ is 0.6.
- Assume that the estimate of the volatilities for X and Y on the day $n-1$ are 1% and 2%, respectively.
- From the relationship between the correlation and the covariance, the estimate for the covariance between X and Y on the day $n-1$ is

$$0.6 \times 0.01 \times 0.02 = 0.00012.$$

- Suppose that the percentage changes in X and Y on the day $n-1$ are 0.5% and 2.5%, respectively.
- The variance and covariance for the day n would be updated as follows:

$$\sigma_{x,n}^2 = 0.95 \times 0.01^2 + 0.05 \times 0.005^2 = 0.00009625;$$
$$\sigma_{y,n}^2 = 0.95 \times 0.02^2 + 0.05 \times 0.025^2 = 0.00041125;$$
$$\mathrm{cov}_n = 0.95 \times 0.00012 + 0.05 \times 0.005 \times 0.025 = 0.00012025.$$

- The new volatility of X is $\sqrt{0.00009625} = 0.981\%$.
- The new volatility of Y is $\sqrt{0.00041125} = 2.028\%$.
- The new *coefficient of correlation* between X and Y is

$$\frac{0.00012025}{0.00981 \times 0.02028} = 0.6044.$$

- GARCH models can also be used for *updating covariance estimates* and *forecasting* the future level of covariances.

- For example the GARCH(1,1) model for updating a covariance is

$$\mathrm{cov}_n = \omega + \alpha x_{n-1} y_{n-1} + \beta \mathrm{cov}_{n-1},$$

  and the long-term average covariance is $\omega/(1-\alpha-\beta)$.

- Similar formulas to those discussed above can be developed for forecasting future covariances and calculating the average covariance during the life time of an option.

Once all variances and pairwise covariances have been calculated, a *(variance-)covariance matrix* can be constructed. (Also recall Chapter 1).

- ▷ When $i \neq j$, the (i, j) element of this matrix represents the covariance between variables i and j. When $j = i$ it represents the variance of the variable i.

- ▷ Not all covariance matrices are *internally consistent*. The condition for an $N \times N$ covariance matrix Σ to be internally consistent is

$$w^{\intercal} \Sigma w \geq 0 \,,$$

for all $N \times 1$ vectors w, where $w^{\intercal}$ is the transpose of w. In general, such matrices are called *positive-semidefinite*.

- To understand why the last condition must hold, suppose that $w = (w_1, \ldots, w_n)^\intercal$. The expression $w^\intercal \Sigma w$ is the *variance* of $w_1 x_1 + \ldots + w_n x_n$ where x_i is the value of the variable i. As such it cannot be negative.

- To ensure that a positive-semidefinite matrix is produced, variances and covariances should be calculated *consistently*.

- For example, if variances are calculated by giving equal weight to the last m data items, the *same* should be done for the covariances.

- If variances are updated using an EWMA model with $\lambda = 0.94$, the *same* should be done for the covariances.

- An example of a covariance matrix that it is *not internally consistent* is

$$\begin{bmatrix} 1 & 0 & 0.9 \\ 0 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \end{bmatrix}$$

- The variance of each variable in this example is 1.0 and so the covariances are also coefficients of *correlation*.

- The first variable is highly correlated with the third variable and the second variable is highly correlated with the third variable.

- However, there is no correlation at all between the first and the second variables. This seems strange. When we set w equal to $(1, 1, -1)$ we find that the *positive semi-definiteness condition* above is *not satisfied* proving that the matrix is not positive-semidefinite.

Acknowledgements

Thanks goes to Alexander Smirnow and Urban Ulrych for their support in preparing these slides and for challenging an earlier version of them!