

# Financial Econometrics

Sol Yates

April 25, 2024

## Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Financial Time Series and their Characteristics</b> | <b>1</b>  |
| 1.1      | Asset Returns . . . . .                                | 1         |
| 1.2      | Descriptive Statistics . . . . .                       | 5         |
| 1.3      | Distribution of Returns . . . . .                      | 6         |
| <b>2</b> | <b>Time Series Basics</b>                              | <b>8</b>  |
| <b>3</b> | <b>ARMA Processes</b>                                  | <b>14</b> |
| 3.1      | ARMA . . . . .   | 16        |
| 3.2      | ARMA estimation . . . . .                              | 17        |
| 3.3      | Conditional Heteroskedacity . . . . .                  | 19        |
| 3.4      | GARCH . . . . .  | 23        |
| 3.5      | Recap . . . . .  | 26        |
| 3.6      | Estimation, Model choice and forecasting . . . . .     | 27        |
| 3.7      | Model Choice and Diagnostics . . . . .                 | 28        |
| 3.8      | Variance Forecast Evaluation . . . . .                 | 30        |
| 3.9      | Tutorial . . . . .                                     | 44        |

## Lecture 1: Introduction Lecture

Wed 31 Jan 11:19

### 1 Financial Time Series and their Characteristics

#### 1.1 Asset Returns

[Lecture 1] [Reading]

Financial studies involve returns, instead of prices of assets.

Returns :

- Is a complete and scale free summary of the investment opportunity
- Are easier to handle than price series

$p_t$  is the price of an asset at time index  $t$ . And assuming an asset pays no dividends.

#### Continuous Compounding

### One period Simple Returns

Holding the asset for one period from date  $t - 1$  to date  $t$  would result in a simple gross return :

$$1 + R_t = \frac{P_t}{P_{t-1}} \text{ or } P_t = P_{t-1}(1 + R_t).$$

The corresponding one period simple net return or simple return is :

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}}.$$

### Multi period Simple Returns

Holding the asset for  $k$  periods between dates  $t-k$  and  $t$  gives a  $k$ -period simple gross return :

$$\begin{aligned} 1 + R_t[k] &= \frac{P_t}{P_{t-k}} = \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \times \dots \times \frac{P_{t-k+1}}{P_{t-k}}. \\ &= (1 + R_t)(1 + R_{t-1}) \dots (1 + R_{t-k+1}). \\ &= \prod_{j=0}^{k-1} (1 + R_{t-j}). \end{aligned}$$

That is, the  $k$ -period simple gross return is just the product of the  $k$  one period simple gross returns involved. A compound return.

The actual time interval is important in discussing and comparing returns, if not given, it is implicitly assumed to be one year.

If an asset is held for  $k$  years, then the annualized average return is defined as

$$\text{Annualized } R_t[k] = \left( \prod_{j=0}^{(k-1)} (1 + R_{t-j}) \right)^{\left(\frac{1}{k}\right)} - 1.$$

Which is a geometric mean of the  $k$  one period simple gross returns involved and can be computed by

$$= \exp\left(\frac{1}{k} \sum_{j=0}^{(k-1)} \ln(1 + r_{t-j})\right) - 1$$

Where it is easier to compute the arithmetic average than the geometric mean and the one-period returns tend to be small, one can use a first order Taylor expansion to approximate the annualized return and obtain

$$\approx \frac{1}{k} \sum_{j=0}^{(k-1)} R_{t-j}.$$

### Continuous Compounding

Assume the interest rate of a bank deposit is 10% per annul, and the initial deposit is \$1

If the bank pays interest once a year, then the net value of the deposit becomes 1.1\$. If the bank pays interest semi-annually, the 6-month interest rate is 5% and the net value is  $1(1 + \frac{0.1}{2})^2 = \$1.1025$  after the first year.

In general if the bank pays interest  $m$  times a year, then the interest rate for each payment is 10% and the net value of the deposit becomes  $1(1 + \frac{0.1}{m})^m$  one year later.

### Continuously Compounded Returns

The natural logarithm of the simple gross return of an asset is called the continuously compounded return or log return :

$$R_t = \ln(1 + R_t) = \ln(P_t/P_{t-1}) = p_t - p_{t-1} \quad (1)$$

Where  $p_t = \ln(P_t)$ . Continuously compounded returns are advantageous since they are the sum of continuously compounded multi period return.

### Portfolio Return

Simple net return of a portfolio consisting of  $N$  assets is a weighted average of the simple net returns of the assets involved, where the weight on each asset is the percentage of the portfolio's value invested in that asset. Where  $p$  is a portfolio that places weight  $w_i$  on asset  $i$ . Then the simple return of  $p$  at time  $t$  is

$$R_{p,t} = \sum_{i=1}^N w_i R_{it}.$$

Where  $R_{it}$  is the simple return of asset  $i$ .

The continuously compounded returns of a portfolio, do not have this property. Instead,

$$R_{p,t} \text{sim} \sum_{i=1}^N w_i r_{it}.$$

Where  $r_{p,t}$  is the continuously compounded return of the portfolio at time  $t$

### Dividend Payment

If an asset pays periodically. Let  $D_t$  be the dividend payment of an asset between dates  $t-1$  and  $P_t$  be the price of the asset at the end of period  $t$ . The dividend is this not included in  $P_t$  The simple net return and continuously compounded return at time  $t$  become

$$R_t = \frac{P_t + D_t}{P_{t-1}} - 1 \quad , \quad r_t = \ln(P_t + D_t) - \ln(P_{t-1}).$$

### Excess Return

The difference between the asset's return and return on some reference asset, often taken to be riskless such as short term US treasury bill. Simple excess return and log excess return of an asset are then defined as

$$Z_t = R_t - R_{0t} \quad , \quad z_t = r_t - r_{0t}.$$

Where  $R_{0t}$  and  $r_{0t}$  are the simple and log returns of the reference asset (resp)

## Distributional Properties of Returns

### Review of statistical distributions and their moments

#### Joint Distribution

$$F_{X,Y}(x, y : \theta) = P(X \leq x, Y \leq y : \theta).$$

Where  $x \in R^{(p)}$ ,  $y \in R^{(q)}$  and the inequality  $\leq$  is a joint distribution function of X and Y with parameter  $\theta$ . The behavior of X and Y is characterized by  $F_{X,Y}(x, y : \theta)$

If the joint probability density function  $f_{x,y}(x, y : \theta)$  exists then

$$F_{X,Y}(x, y : \theta) = \int_{-\infty}^x \int_{-\infty}^y f_{x,y}(w, z; \theta) dz dw.$$

Where X and Y are continuous random vectors

#### Marginal Distribution

Given by

$$F_X(X; \theta) = F_{X,Y}(x, \infty, \dots, \infty, \theta).$$

Thus, the marginal distribution of X is obtained by integrating out Y. A similar definition applies to the marginal distribution of Y If  $k = 1$  X is a scalar random variable and the distribution function becomes

$$F_X(x) = P(X \leq x; \theta).$$

Which is the CDF of X. The CDF of a random variable is nondecreasing and satisfies  $F_X(-\infty) = 0$  and  $F_X(\infty) = 1$  For a given probability p, the smallest real number  $x_p$  such that  $p \leq F_X(x_p)$  is called the 100 p th quantile of the random variable X

#### Conditional Distribution

The conditional distribution of X given  $y \leq y$  is given by

$$F_{X|Y \leq y}(x; \theta) = \frac{P(X \leq X, Y \leq Y : \theta)}{P(Y \leq Y : \theta)}.$$

#### Moments of a Random Variable

The l-th moment of a continuous random variable X is defined as

$$M'_l = E[X^l] = \int_{-\infty}^{\infty} x^l f(x) dx$$

Where E stands for expectation and  $f(x)$  is the probability density function of x. The first moment is called the mean or expectation, measuring the central location of the distribution.

The l-th central moment of X is defined as

$$M_l = E[(X - \mu_x)^l] = \int_{-\infty}^{\infty} (x - \mu_x)^l f(x) dx$$

The second central moment, denoted  $\sigma_x^2$  measures the variability of X and is called the variance of X. The positive square root  $\sigma_x$  of variance is the *standard deviation* of X.

The first two moments of a random variable uniquely determine a normal distribution.

The *Third Central* moment measures the symmetry of  $X$  with respect to its mean, whereas the *fourth central moment* measures the tail behaviour of  $X$ .

*Skewness* and *kurtosis* are normalised third and fourth central moments of  $X$ , are often used to summarise the extent of asymmetry and tail thickness

## 1.2 Descriptive Statistics

Let  $Y_t$  be a time-series of random variables with a history of realisations  $y_t$  with  $t = 1, \dots, T$

Mean

$$E[Y_t] = \mu \quad , \quad \hat{m}u = \frac{1}{T} \sum_{t=1}^T y_t$$

Variance

$$V[Y_t] = E[(Y_t - \mu)^2] \quad , \quad \hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{m}u)^2$$

Skewness

$$S[Y_t] = E\left[\frac{(Y_t - \mu)^3}{\sigma^3}\right] \quad , \quad \hat{S} = \frac{1}{T} \sum_{t=1}^T \left[\frac{(y_t - \hat{m}u)^3}{\hat{\sigma}^3}\right]$$

Kurtosis

$$K[Y_t] = E\left[\frac{(Y_t - \mu)^4}{\sigma^4}\right] \quad , \quad \hat{K} = \frac{1}{T} \sum_{t=1}^T \left[\frac{(y_t - \hat{m}u)^4}{\hat{\sigma}^4}\right]$$

Jarque-Bera test, tests  $H_0$  of normality of the series :

$$JB = \frac{T}{6} \left( \hat{S}^2 + \frac{(\hat{K} - 3)^2}{4} \right)$$

Where  $k$  is the number of estimated parameters. This test statistic has a  $\chi^2$  distribution with 2 degrees of freedom (always). Tests 2 parameters jointly. Rejection when skewness is not 0 or kurtosis is not 3. Skewed or heavy tailed. Then use individual tests against 0 or 3 using WLLN and CLT. T test, standardizing appropriately.

Quantile-Quantile plots : plot theoretical quantiles against the empirical ones

### Stylized Facts

- Return series do not follow a normal distribution
- The normal distribution does not explain the occurrence probability of extreme events
- Better assumptions are student-t or stable distributions
- On higher frequencies (intraday) the deviation from normality is more pronounced than on lower frequencies
- Aggregated return series, do however, tend to normality

- Return series posses fat tails
- Return series are leptokurtic or posses an overkurtosis (kurtosis  $> 3$ )
- Large returns occur more often than expected
- Large returns are more often negative than positive which yields left skewed returns (skewness  $< 0$ )
- Intraday returns are subject to typical trading session effects (seasonality, opening and closing issues)
- Returns are subject to volatility clustering, which is again more pronounced on higher frequencies
- Volatility is time varying
- Financial time series are correlated
- Correlations are also time varying

### Standardized Return

$$\left(\frac{r_t - \mu}{\hat{\sigma}}\right).$$

Kurtosis is probably the most important, telling you about the number of extreme events. Say coca-cola vs tesla (kurtosis of 50). Can be seen as number of outliers around mean

Plotting histogram, kurtosis is heavy tails, extreme distribution lands exactly to the tails.

## 1.3 Distribution of Returns

The most general model for the log returns is its joint distribution function  $F_r(r_{11}, \dots, r_N : r_{12}, \dots, r_{N2} : \dots r_{IT} \dots r_{NT} : Y; \theta)$

Where  $Y$  is a state vector consisting of variables that summarise the environment in which asset returns are determined and  $\theta$  is a vector of parameters that uniquely determines the distribution function  $F_r(\cdot)$ , which governs the stochastic behaviour the returns  $r_{it}$  and  $Y$ .

Often the state vector  $Y$  is treated as given and the main concern is the conditional distribution of  $\{r\}_{it}$  given  $Y$ .

Some financial theories (CAPM) focus on the joint distirbution of  $N$  returns at a single tome index  $t$ . Whilst others look at the dynamic structure of individual asset returns

Since our main concern is the joint distribution of  $\{r_{it}\}_{t=1}^T$  for asset  $i$ , it is useful to partition the joint distribution as :

$$\begin{aligned} F(r_{i1}, \dots, r_{iT} : \theta) &= F(r_{i1})F(r_{i2}|r_{i1}) \dots F(r_{iT}|r_{iT-1}, \dots, r_{i1}) \\ &= F(r_{i1}) \prod_{t=2}^T F(r_{iT}|r_{it-1}, \dots, r_{i1}) \end{aligned}$$

Where the parameter  $\theta$  is omitted for brevity.

This partitions the temporal dependencies of the log return  $r_{it}$ . With the main issue the specification of the conditional distribution  $F(r_{it}|r_{i,t-1})$  since different distributional specifications lead to different theories in finance.

For instance the *random walk hypothesis* in which one version entails the conditional distribution  $F(r_{it}|r_{i,t-1}, \dots, r_{i1})$  is equal to the marginal distribution  $F(r_{it})$  meaning returns are temporally independent and thus not predictable.

### Normal Distribution

A traditional assumption is that the simple returns  $\{R_{it}|t = 1, \dots, T\}$  are **independently and identically distributed** as normal with fixed mean and variance.

However, this assumption encounters difficulties empirically,

- The lower bound of a simple return is -1, but the normal distribution may assume any value in the real line and hence has no lower bound
- If  $R_{it}$  is normally distributed then the multi period simple return  $R_{it}[k]$  is not normally distributed because it is a product of one period returns
- The normality assumption is not supported by many empirical asset returns

### Log normal Distribution

Another commonly used assumption is that the log returns  $r_t$  of an asset are independent and identically distributed (iid) as normal with mean  $\mu$  and variance  $\sigma^2$ . The simple returns are then iid lognormal random variables with mean and variance given by

$$E[R_t] = \exp\left(\mu + \frac{\sigma^2}{2}\right) - 1$$

And

$$Var[R_t] = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1]$$

### Stable Distribution

The stable distribution are a natural generalisation of normal in that they are stable under addition, meeting the need of continuously compounded returns  $r_t$ . Furthermore, stable distributions are capable of capturing excess kurtosis, shown by historical stock returns

### Hypothesis Test

Null  $H_0 : s = 0$  vs  $H_1 : s \neq 0$

$$\hat{t} + CLT \rightarrow^{(d)} N(0, 1).$$

Tells you distribution under the null, then 95% of probability mass is between critical values, then outside of this, either suff evidence against the null or a type I error (5%) (at tails). Fundamentally, we cannot trust the null hypothesis.

Whatever we want to test, we put into the alternative. NO conclusion can be made if we fail to reject the null. If we collect evidence against the null then this is fundamentally different.

Na

## Lecture 2: Second Lecture - Review of Time Series

Fri 02 Feb 15:45

[Lecture 2] [Reading]

## 2 Time Series Basics

### Stationarity

A time series  $\{r_t\}$  is *strict stationary* if the joint distributing of  $(r_{t_1}, \dots, r_{t_k})$  is identical to that of  $(r_{t_1+t}, \dots, r_{t_k+t})$  for all  $t$  where  $k$  is an arbitrary positive integer and  $t_1, \dots, t_k$  is a collection of  $k$  positive integers.

That is, *strict stationarity* requires that the joint distributing is *time invariant* under a time shift, but of course this is hard to verify empirically and a very strong condition.

A weaker condition is that a time series  $\{r_t\}$  is *weakly stationary* if both the mean of  $r_t$  and the covariance between  $r_t$  and  $r_{t-l}$  are time invariant.

That is,

$$E[r_t] = \mu \text{ a constant } Cov(r_t, r_{t-l}) = \gamma_l \text{ which only depends on } l \quad (2)$$

Where weak stationarity implies the time plot of the data would show that the  $T$  values fluctuate with *constant* variation around a fixed level. Enabling one to make inference conceding future observations. But, implicitly we have assumed that the first 2 moments of  $r_t$  are finite

Where the covariance  $\gamma_l = Cov(r_t, r_{t-l})$  is called the lag- $\downarrow$  auto covariance of  $r_t$ . With 2 important properties :

1.  $\gamma_0 = Var(r_t)$
2.  $\gamma_{-l} = \gamma_l$

### Correlation and autocorrelation functions

The correlation coefficient between 2 random variables  $X$  and  $Y$  is defined as

$$\rho_{x,y} = \frac{Cov(X,Y)}{\sqrt{Var(x)Var(Y)}} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sqrt{E(X - \mu_x)^2 E(Y - \mu_y)^2}} \quad (3)$$

Where  $\mu_x$  and  $\mu_y$  are the mean of  $X$  and  $Y$  resp, assuming the variances exist also. This measures the *strength* of linear dependence between  $X$  and  $Y$ , and it can be shown that  $-1 \leq \rho_{x,y} \leq 1$  and  $\rho_{x,y} = \rho_{y,x}$ . Where the two RV are uncorrelated if  $\rho_{x,y} = 0$ , which occurs iff  $X$  and  $Y$  are independent.

**Autocorrelation Function (ACF)** For a weakly stationary return series  $r_t$  when the linear dependence between  $r_t$  and its past values  $r_{t-i}$ , we can generalise the correlation concept to autocorrelation.

The correlation coefficient between  $r_t$  and  $r_{t-i}$  is called the lag- $\ell$  autocorrelation of  $r_t$  and is commonly denoted by  $\rho_\ell$ , which under the weak stationarity assumption is a function of  $\ell$  only

We define

$$\rho_\ell = \frac{Cov(r_t, r_{t-\ell})}{\sqrt{Var(r_t)Var(r_{t-\ell})}} = \frac{Cov(r_t, r_{t-\ell})}{Var(r_t)} = \frac{\gamma_\ell}{\gamma_0}$$

Where  $Var(r_t) = Var(r_{t-\ell})$  for a weakly stationary series



For a given sample of returns  $\{r_t\}_{t=1}^T$  let  $\bar{r}$  be the sample mean ( $\bar{r} = \sum_{t=1}^T r_t / T$ ). Then the lag-1 sample autocorrelation of  $r_t$  is

$$\hat{p}_1 = \frac{\sum_{t=2}^T (r_t - \bar{r})(r_{t-1} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}$$

Which, under some general conditions is a consistent estimator of  $p_1$  IE, if  $\{r_t\}$  is an iid sequence and  $E[r_t^2] < \infty$ , then  $\hat{p}_1$  is asymptotically normal with mean zero and variance  $\frac{1}{T}$ .

In practice we can use this to test the null hypothesis  $H_0 : p_1 = 0$  vs  $H_a : p_1 \neq 0$  With the test stat the t ratio which is  $\sqrt{T}\hat{p}_1$  and follows asymptotically the standard normal distribution.

The null hypothesis is *rejected* if the t ratio is large in magnitude or (equivalently) the p value of the t ratio is small ( $< 0.05$ )

lag -  $\ell$  sample autocorrelation of  $r_t$  is defined as

$$\hat{p}_\ell = \frac{(r_t - \bar{r})(r_{t-\ell} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2} \quad (4)$$

where  $0 \leq \ell < T - 1$

**Testing Individual ACF** For the previous result, we can test  $H_0 : p_\ell = 0$  vs  $H_a : p_\ell \neq 0$  using test stat

$$\text{t ratio} = \frac{\hat{p}_\ell}{\sqrt{\left(1 + 2 \sum_{i=1}^{\ell-1} \hat{p}_i^2\right) / T}}$$

Then, if  $\{r_t\}$  is a stationary Gaussian series satisfying  $\rho_j = 0$  for  $j > l$  the t ratio is *asymptotically* distributed as a standard normal RV. And hence the decision rule of the test

Reject  $H_0$  if  $|\text{t ratio}| > Z_{\frac{\alpha}{2}}$  where  $Z_{\frac{\alpha}{2}}$  is the  $100(1 - \alpha/2)$ th percentile of the standard normal distribution.

In finite samples,  $\hat{p}_\ell$  is a biased estimator of  $p_\ell$ , of the order  $\frac{1}{T}$  which is substantial with a small sample size (although large samples in financial so OK).

**Portmanteau Test** Portmanteau statistic

$$Q^*(m) = T \sum_{i=1}^m \hat{p}_i^2$$

to test jointly that several autocorrelations of  $r_t$  are zero. Null hypothesis  $H_0 : p_i \neq 0$  for some  $i \in \{1, \dots, m\}$ . Under the assumption that  $\{r_t\}$  is an iid sequence,  $Q^*(m)$  is asymptotically topically a chi-squared RV with m degrees of freedom.

**Ljung and Box (1978)** modify this to increase the power of the test in finite samples

$$Q(m) = T(T+2) \sum_{l=1}^m \frac{\hat{p}_l^2}{T-l}$$

where the decision rule is to redirect  $H_0$  if  $Q(m) > \chi_\alpha^2$  where  $\chi_\alpha^2$  denotes the  $100(1 - \alpha)$ th percentile of a chi squared distribution (m dof). The decision rule to reject  $H_0$  if the p value is less than or equal to  $\alpha$

The choice of  $m$  may affect performance of the statistic, several values of  $m$  are often used.

The statistics  $\hat{p}_1, \hat{p}_2, \dots$  in eq. (4) are the *sample autocorrelation functions* (ACF) of  $r_t$ .

A linear time series model can be characterised by its ACF, and linear time series modelling makes use of the sample ACF to capture the linear dynamic of the data.

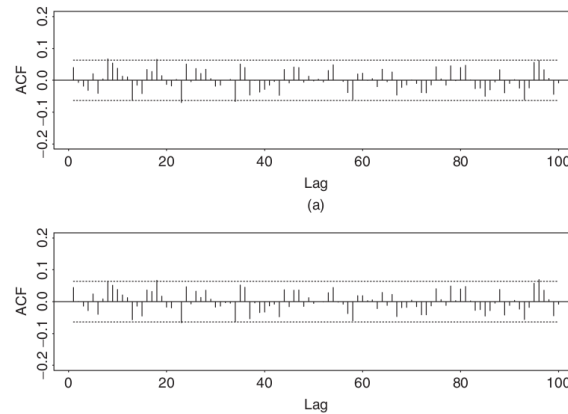


Figure 1: Sample Autocorrelation Functions of Monthly Simple And Log Returns

Indeed, the two sample ACFs are very close to each other, and suggest the serial correlations of monthly IBM stock returns are *very small*. The sample ACFs are all within their two SE limit, indicating they are not significantly different from zero at the 5% level. Additionally, the Ljung-Box stats give  $Q(5) = 3.37$  and  $Q(10) = 13.99$ , corresponding p values of 0.64 and 0.17 based on chi-squared distributions.

Often, a version of the CAPM theory is that the return  $\{r_t\}$  of an asset is not predictable and should have no autocorrelations. Thus, testing for zero autocorrelations has been used as a tool to check the efficient market assumption

Jjk

### Stochastic Processes

- Chronologically ordered equidistant observations
- Generated by stochastic process
- Stochastic process - collection of RV (each  $Y_i$  is generated by different member of stochastic processes)
- **assumption** time series data has been generated by *stochastic process*

**Definition 1. stochastic process** is a family of random variables defined on a probability space

**Definition 2. time series** is a realisation of a stochastic process

**Definition 3. time series analysis** - only one history  $Y_t(w)$ , one state of the world  $w \in \omega$  is available, but the goal is to derive the properties of  $Y_t(\cdot)$  for a given  $t$  for different states of the world

Idea - how can we understand what is driving omegas? Different states of the world, since we observe  $y_t$ . So place some structure on  $y_t$

Should be able to recognise :

- Non-stationary time series
- Autoregressive time series
- Kurtosis time series

**Definition 4. Auto Covariance**

Time series often show correlation between successive observations, this feature is called serial correlating or **autocorrelation**

Dependencies over time are described by auto covariance and autocorrelation functions

The  $j$ -th autocovariance of  $Y_t$  is given by

$$\text{Cov}[Y_t, y_{t-j}] = \gamma_{t,t-j} = E[Y_t - E[Y_t]][Y_{t-j} - E[Y_{t-j}]]$$

Correspondingly the variance of  $Y_t$  is defined as :

$$V[Y_T] = \gamma_{t,t} = E[Y_t - E[Y_t]]^2$$

**Definition 5. Autocorrelation**

The  $j$ -th autocorrelation of  $Y_t$  is given by :

$$\rho_{t,t-j} = \frac{\text{Cov}[Y_t, Y_{t-j}]}{V[Y_t]^{(\frac{1}{2})} V[Y_{t-j}]^{(\frac{1}{2})}}$$

**Definition 6. Covariance Stationary** A time series  $\{Y_t\}_{t=-\infty}^{(\infty)}$  is called covariance stationary, or weakly stationary, if :

$$\begin{aligned} E[Y_t] &= \mu_Y \\ V[Y_t] &= \gamma_{t,t} = \gamma_0 = \sigma_Y^2 < \infty \\ \text{Cov}[Y_t, Y_{t-j}] &= \gamma_{t,t-j} = \gamma_j < \infty \end{aligned}$$

For a covariance stationary process the  $j$ -th autocorrelation is given by :

**White Noise**

**Definition 7.** white noise A time series is called white noise if it satisfies:

$$E[Y_t] = 0, V[Y_t] = \sigma_Y^2, Cov[Y_t, Y_s] = E[Y_t, Y_s] = 0$$

White noise is a weakly stationary process - all the ACFs are 0.

Particularly, if  $r_t$  is normally distributed with mean 0 and variance  $= \sigma^2$  the series is a Gaussian white noise

**Definition 8.** Autocorrelation Function Autocorrelation function of a covariance stationary process  $\{Y_t\}_{t=-\infty}^{(\infty)}$  is the sequence of autocorrelations  $\rho_j$  for all  $j = 0, 1, 2, \dots$

**Definition 9.** Empirical Autocorrelation Function

The empirical (or sample) autocorrelation function of a time series  $Y_t$  is the sequence of sample autocorrelation coefficients  $\hat{\rho}_j$  for all  $j = 0, 1, 2, \dots$ :

$$\hat{\rho}_j = \frac{\hat{\gamma}_j}{\hat{\gamma}_0} = \frac{\sum_{t=j+1}^T (Y_t - \bar{Y})(Y_{t-j} - \bar{Y})}{\sum_{t=1}^T (Y_t - \bar{Y})^2}$$

And

$$\hat{\gamma}_j = \frac{1}{T} \sum_{t=j+1}^T (Y_t - \bar{Y})(Y_{t-j} - \bar{Y}) \quad \bar{Y} = \frac{1}{T} \sum_{t=1}^T Y_t$$

The graphical depictions of the empirical autocorrelation function is called an autocorrelogram

**Definition 10.** Partial Autocorrelation Function

Partial autocorrelation between  $Y_t$  and  $Y_{t-j}$  is the conditional correlation between  $Y_t$  and  $Y_{t-j}$  given (holding fixed)  $Y_{t-1}, \dots, Y_{t-j+1}$

$$A_j = Cor[Y_t, Y_{t-j} | Y_{t-1}, \dots, Y_{t-j+1}]$$

Effects of in-between values are controlled for

Corresponding sample quantity  $\hat{a}_j$  is called sample partial autocorrelation and is obtained as the OLS estimator of the coefficient  $a_j$  in model

$$Y_t = a_0 + a_1 Y_{t-1} + \dots + a_j Y_{t-j} + \mu_t$$

**Definition 11.** Sample Autocorrelation Function

If data generating process is a white noise process, then for large T:

$$\hat{\rho}_j \approx N(0, \frac{1}{T}), j = 1, 2, \dots$$

Means :  $H_0 : \hat{\rho}_j = 0$  is rejected, if zero does not fall within the approximate 95% confidence interval

$$[\hat{rho}_j - \frac{2}{\sqrt{T}}, \hat{rho}_j + \frac{2}{\sqrt{T}}]$$

Equivalently, autocorrelations are not significant when  $\hat{\rho}_j$  is within the approximate two standard error bound  $\pm 2/\sqrt{T}$

**Linear Time Series** A time series  $r_t$  is said to be linear if it can be written as

$$r_t = \mu + \sum_{i=0}^{\infty} \psi_i \alpha_{t-i} \quad (5)$$

Where  $\mu$  is the mean of  $r_t$ ,  $\psi_0 = 1$  and  $\{\alpha_t\}$  is a sequence of iid RV with mean zero and well defined distributions (ii a white noise)

For this equation, the dynamic structure of  $r_t$  is governed by the coefficients  $\psi_i$  which are called the  $\psi$  weights of  $r_t$

If  $r_t$  is weakly stationary, we can obtain its mean and variance easily by using the independence of  $\{\alpha_t\}$  as

$$E[r_t] = \mu \quad \text{Var}(r_t) = \sigma_\alpha^2 \sum_{i=0}^{\infty} \psi_i^2 \quad (6)$$

Where  $\sigma_u^2$  is the variance of  $a_t$

### Simple AR Models

If a monthly return of a value weighted index has a statistically significant lag-1 autocorrelation indicates that the lagged return  $r_{t-1}$  may be useful in predicting  $r_t$ , we can implement this in a model such as

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t \quad (7)$$

Where  $\{a_t\}$  is assumed to be a white noise series with mean zero and variance  $\sigma_a^2$

This is analogous to the simple linear regression model in which  $r_t$  is the dependent variable and  $r_{t-1}$  is the explanatory variable.

This is actually an autoregressive (AR) model of order 1 (AR(1))

**Note.** Conditional on the past return, the current return is centred around  $\phi_0 + \phi_1 r_{t-1}$  with CID  $\sigma_a$  AR(1) model implies that conditional on past return  $r_{t-1}$ , we have

$$E[r_t | r_{t-1}] = \phi_0 + \phi_1 r_{t-1} \quad \text{Var}(r_t | r_{t-1}) = \text{Var}(a_t) = \sigma_a^2$$

This is a Markov property such that conditional on  $r_{t-1}$ , the return  $r_t$  is not correlated with  $r_{t-i}$  for  $i > 1$

A straightforward generalisation of the AR(1) model is the AR(p) model

### 3 ARMA Processes

**Definition 12.** AR(p)-Process A time series is called an autoregressive process of order p if it satisfies a relationship of the type :

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

Where  $\varepsilon_t$  is a white noise error term

**AR(1) process** : the simplest form of an AR(p) process is obtained for  $p = 1$  as

$$Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t \quad (8)$$

#### AR(1) Model

Starting with the sufficient and necessary condition for weak stationarity of the AR(1) model, assuming that the series is *weakly stationary*, we have  $E[r_t] = \mu$ ,  $\text{Var}(r_t) = \gamma_0$  and  $\text{Cov}(r_t, r_{t-j}) = \gamma_j$  where  $\mu$  and  $\gamma_0$  are constant and  $\gamma_j$  is a function of j, not t. To obtain the mean variance and autocorrelations of the series,

Taking the expectation of eq. (8) (with slightly different notation), and since  $E[a_t] = 0$ , we obtain

$$E[r_t] = \phi_0 + \phi_1 E[r_{t-1}]$$

Then, under the stationarity condition  $E[r_t] = E[r_{t-1}] = \mu$  and hence

$$\mu = \phi_0 + \phi_1 \mu \quad \text{or} \quad E[r_t] = \mu = \frac{\phi_0}{1 - \phi_1}$$

Which has two implications for  $r_t$

1. The mean of  $r_t$  exists if  $\phi_1 \neq 1$
2. The mean of  $r_t$  is zero iff  $\phi_0 = 0$

Thus, for a stationary AR(1) process, the constant term  $\phi_0$  is related to the mean of  $r_t$  via  $\phi_0 = (1 - \phi_1)\mu$  and  $\phi_0 = 0$  implies that  $E[r_t] = 0$

Then, using  $\phi_0 = (1 - \phi_1)\mu$ , the AR(1) model can be rewritten as

$$r_t - \mu = \phi_1 (r_{t-1} - \mu) + \alpha_t$$

Which, using repeated substitutions implies

$$\begin{aligned} r_t - \mu &= a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \dots \\ &= \sum_{i=0}^{\infty} \phi_1^i a_{t-i} \end{aligned} \quad (9)$$

Which expresses a AR(1) model in the form of a linear time series

$$r_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + a_t$$

Where p is a non-negative integer and  $\{a_t\}$  is defined previously.

Essentially, this says that the past p variables  $r_{t-i}$   $i = 1, \dots, p$  jointly determined the conditional expectation of  $r_t$  given the past data.

**AR(1) Properties**

Considering an  $AR(1)$  process:

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \quad -1 < \rho < 1$$

where  $u_t$  is a white noise process. This  $AR(1)$  process has the following properties

1.  $\varepsilon_t = \sum_{i=0}^{\infty} \rho^i u_{t-i} = \mathbf{MA}(\infty)$
2.  $E[\varepsilon_t] = 0$
3.  $V[\varepsilon_t] = \gamma_0 = \frac{\sigma_u^2}{1-\rho^2}$  with  $V[u_t] = \sigma_u^2$
- 4.

$$\begin{aligned} \text{Cov}(\varepsilon_t, \varepsilon_{t-1}) &= \gamma_1 = \rho_1 \frac{\sigma_u^2}{1-\rho^2} \\ \text{Cov}(\varepsilon_t, \varepsilon_{t-s}) &= \gamma_s = \rho_s \frac{\sigma_u^2}{1-\rho^2} \\ \text{Cov}(\varepsilon_t, \varepsilon_{t-1}) &= \gamma_1 = \rho_1 \frac{\sigma_u^2}{1-\rho^2} \end{aligned}$$

**Example. AR(1) Process**

$$\sum_{i=0}^{\infty} \rho^{(i)} u_{t-i} = (\text{wald}) \mathbf{MA}(\infty)$$

**Autocorrelation function of an AR(1) model** Then, multiplying eq. (9) by  $\alpha_t$ , using the independence between  $a_t$  and  $r_{t-1}$ , and taking expectations we obtain

$$E[a_t(r_t - \mu)] = \phi_1 E[a_t(r_{t-1} - \mu)] + E[a_t^2] = E[a_t^2] = \sigma_a^2$$

**identifying AR models in practice**

In reality, the order  $p$  of an AR time series is unknown, it must be specified empirically, referred to as the *order determination*

There are two approaches, either the PACF or information criteria.

**Partial Autocorrelation Function (PACF)** The PACF of a stationary time series is a function of its ACF and is a useful tool for determining the order  $p$  of an AR model.

Considering AR models in consecutive orders

$$r_t = \phi_{0,1} + \phi_{1,1}r_{t-1} + e_{1t} \quad (10)$$

$$r_t = \phi_{0,2} + \phi_{1,2}r_{t-1} + \phi_{2,2}r_{t-2} + e_{2t} \quad (11)$$

Where  $\phi_{0,j}$ ,  $\phi_{i,j}$  and  $\{e_{jt}\}$  are respectively, the constant term, coefficient of  $r_{t-i}$  and the error term of an  $AR(j)$  model. For a stationary Gaussian  $AR(p)$  model, it can be shown that the sample PACF has the following properties:

- $\hat{\phi}_{p,p}$  converges to  $\phi_p$  as the sample size  $T$  tends to infinity

- $\hat{\phi}_{l,l}$  converges to zero for all  $\ell > p$
- The asymptotic variance of  $\hat{\phi}_{l,l}$  is  $1/T$  for all  $l > p$

Thus, for an AR(p) series, the sample PACF cuts off at the lag p.

### 3.1 ARMA

**Definition 13.** MA(q)-Process A time series is called a **moving average process of order q** if it satisfies a relationship of the type

$$Y_t = \mu = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

Where  $\varepsilon_t$  is a white noise error term

MA(1) Process: the simplest form of an MA(q) process is obtained for  $q = 1$  as

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Often times the AR or MA models become cumbersome, so one may need a higher order model with many parameters to describe the dynamic structure of the data.

ARMA models are introduced essentially combining both AR and MA models, such that the number of parameters is kept small. Importantly, the GARCH model can be regarded as an ARMA model.

A time series  $r_t$  follows an ARMA(1, 1) model if it satisfies

$$r_t - \phi_1 r_{t-1} = \phi_0 + a_t - \theta_1 a_{t-1}$$

where  $\{a_t\}$  is a white noise series. With the left and right giving the AR and MA part resp.

**Lag operator** let  $\{Y_t\}_{t=-\infty}^{(\infty)}$  be a time series, then the lag operator  $\mathcal{L}$  is defined by the relation

$$L^{(j)} \equiv Y_{t-j}$$

If  $\{Y_t = c\}_{t=-\infty}^{(\infty)}$  where  $c \in \mathbb{R}$ , then  $\mathcal{L}^{(j)} Y_t = L^{(j)} c = c$

**ARMA(p,q)** is a time series  $\{Y_t\}_{t=-\infty}^{(\infty)}$  of the following form

$$\begin{aligned} \phi_p(L) Y_t &= c + \Theta(L) \varepsilon_t \text{ where} \\ \phi_p(L) &= 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p \\ \Theta(L) &= 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q \end{aligned}$$

With  $\varepsilon_t$  being a white noise and  $\phi_p$  and  $\Theta_q$  are called lag polynomials

#### Properties of ARMA(1, 1) Models

[2.6] These are generalisation of those of AR(1) models with some modifications to handle the MA(1) component.

Starting with the stationarity condition and taking expectation of the ARMA(1, 1) model :

$$E[r_t] - \phi_1 E[r_{t-1}] = \phi_0 + E[a_t] - \theta_1 E[a_{t-1}]$$



Because  $E[\alpha_i] = 0$  for all  $i$ , the mean of  $r_t$  is

$$E[r_t] = \mu = \frac{\phi_0}{1 - \phi_0}$$

provided the series is weakly stationary.

Then, assuming for simplicity that  $\phi_0 = 0$  we consider the autocovariance function of  $r_t$

Multiplying the model by  $a_t$  and taking expectations we have

$$E[r_t a_t] = E[\alpha_t^2] - \theta_1 E[a_t a_{t-1}] = E[\alpha_t^2] = \sigma_a^2 \quad (12)$$

Then rewriting the model as

$$r_t = \phi_1 r_{t-1} + \alpha_t - \theta_1 a_{t-1}$$

and taking the variance of the prior equation, we have

$$\text{Var}(r_t) = \phi_1^2 \text{Var}(r_{t-1}) + \sigma_a^2 + \theta_1^2 \sigma_a^2 - 2\phi_1 \theta_1 E[r_{t-1} a_{t-1}]$$

Where we make use of the fact that  $r_{t-1}$  and  $a_t$  are uncorrelated, then using 7 we obtain

$$\text{Var}(r_t) - \phi_1^2 \text{Var}(r_{t-1}) = (1 - 2\phi_1 \theta_1 + \theta_1^2) \sigma_a^2$$

Therefore if the series is weakly stationary, then  $\text{Var}(r_t) = \text{Var}(r_{t-1})$  and we have

$$\text{Var}(r_t) = \frac{(1 - 2\phi_1 \theta_1 + \theta_1^2) \sigma_a^2}{1 - \phi_1^2}$$

### 3.2 ARMA estimation

ARMA(p, q) process:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

- Estimation via conditional Max likelihood
- **conditional** : derive the likelihood function under the assumption that the initial values of  $Y_t$  and  $\varepsilon_t$  are available
- **assume** :  $\varepsilon_t \sim^{(iid)} N(0, \sigma^2)$
- ML parameter estimators are derived under the assumption of normality are quasi ML estimators
- Our goal is to estimate the vector  $\theta = (c, \phi_1, \phi_2, \dots, \theta_1, \theta_2, \dots, \theta_q, \sigma^2)'$

ARMA estimation

Conditional log likelihood

Estimation is done under assumption that error term is normal.

LBJ test

Whether  $p$  is sufficiently long, if model specified correctly, then residuals shouldn't be correlated with each other.

Tells whether white noise property is plausible assumption

Critical values is from chi-squared dist, we test for absence of autocorrelation upto chosen lag order, leading into next weeks lecture of conditional heteroskedacity.

ARCH-ML test

Tests for conditional heteroskedacity in regression residuals

Pick ARMA based on this, if modelled successfully then null of LBJ test shouldn't be rejected and there shouldn't be any conditional heteroskedacity

## Lecture 3: ARCH Models

Mon 12 Feb 09:01

[Lecture 3] [Reading]

### Review

Week 1

Leptokurtic Property - How to measure a lot of outliers? Kurtosis. The kurtosis of our distribution is larger than 3 (4th moment of distribution). Since  $K[r_t] > 3$  where  $e \sim \mathcal{N}(\mu, \sigma^2)$

Left-Skewness - more negative returns than positive ones.  $S[r_t] < 0$

Volatility clustering - periods of high volatility are followed by periods of high volatility. The volatile periods on the markets (across S of return distribution) they *cluster*. Market volatility is persistent.

Shape of daily returns - Compared to say a normal bell curve, is this a good distribution? Weekly more normal then daily, monthly more normal than weekly. Thus *aggregate returns tend to normality*

- Should know these by heart
- And be able to apply them and tell graphically

Week 2

Time series analysis

ARMA Models ( $ARMA(1, 1)$   $y_t = c + \rho_1 + y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$

Stationarity ADF test

Model selection ACF /DACF

Information criteria - Bayesian information criteria helps to choose whether ARMA(1, 2) or MA(1) is better for data.

At the end we estimate by quasi-likelihood since  $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ , which is key to today's material.

It is important to realise this assumption is quite strong, the shortcut for this type of estimator is quasi-likelihood.

$\Theta = (c, \rho_1, \theta_1, \sigma^2)$ , in empirical terms the maximum likelihood estimators minimise the negative log-likelihood, we can only find the minimum using gradient descent, hence minimising the negative.

$$\hat{\theta}_{QML} = \underset{\Theta}{argmin}$$

Autoregressive process order 1

Log likelihood - maximises function to find combination of parameters of model such that our  $\varepsilon_t$ 's are normal

For Financial Econometrics, once plot ACF and PCF, once looking at squared residuals, we have seen a lot of significant lags in the squared residuals. LBQ test and ARCH-LM test whether there is remaining autocorrelation within the squares residuals.

These tests tell us that  $\hat{\sigma}^2$  tell us there is autocorrelation across time within the residuals, only problem of model misspecification comes from squared residuals, variance of error term.

### 3.3 Conditional Heteroskedacity

In any ARMA model there is some expectation

$$Y_t = E[y_t|F_{t-1}] + \varepsilon_t$$

$c + \rho y_{t-1} + \theta \varepsilon_{t-1}$ .  $F$  is filtration, past information and  $\varepsilon_t$  is new information/shock today.

White noise (DSA) :

$$\begin{aligned}\varepsilon_t &\sim WN \\ E[\varepsilon_t] &= 0 \\ V[\varepsilon_t] &= \sigma^2 \\ Cov(\varepsilon_t, \varepsilon_t) &= 0\end{aligned}$$

#### Conditional vs Unconditional Variance

$$V\varepsilon_t|\mathcal{F}_{t-1} = \sigma_t^2 V[z_t|\mathcal{F}_{t-1}] = \sigma_t^2 V[z_t] = \sigma_t^2 = \sigma_t^2(\mathcal{F}_{t-1})$$

Conditional :  $V[\varepsilon_t]$  and Unconditional  $V[\varepsilon_t|F_{t-1}]$ .  $\sigma_t^2$  is hence the conditional variance of  $\varepsilon_t$ , ie  $\varepsilon_t$  is conditional heteroskedatic.  $\sigma_t = \sqrt{\sigma_t^2}$  gives the conditional variance of  $\varepsilon_t$

Where  $V[\varepsilon_t|\mathcal{F}_{t-1}]$  depends on past filtration.

The unconditional variance of  $\varepsilon_t$  is

$$\sigma_\varepsilon^2 = V[\varepsilon_t]$$

$$\begin{aligned}Y_t &= c + \rho y_{t-1} + \varepsilon_t \\ E[y_t] &= E[c + \rho y_{t-1} + \varepsilon_t] \\ &= c + \rho E[y_{t-1}] \\ E[y_t] &= E[y_{t-1}] \\ E[y_t] &= \frac{c}{1-\rho}\end{aligned}$$

That is,

$$\frac{c}{1-\rho} \text{ vs } c + \phi y_{t-1} \quad (*)$$

$$\begin{aligned}
& E[y_t|F_{t-1}] \\
& E[c + \rho y_{t-1} + \varepsilon_t|F_{t-1}] \\
& C + \rho E[y_{t-1}|F_{t-1}] + E[\varepsilon_t|F_{t-1}] \\
& C + \rho y_{t-1} + 0
\end{aligned}$$

### White Noise

- $E[\varepsilon_t] = 0$
- $V[\varepsilon_t] = \sigma^2$
- $cov[\varepsilon_t, \varepsilon_t] = 0$

The unconditional moment in (\*) is more important.

White noise assumption, assumes both conditional and unconditional are constant over time, that is

$$V[\varepsilon_t] = V[\varepsilon_t, |F_{t-1}] = \sigma^2$$

$V[\varepsilon_t] = \sigma^2$  but  $V[\varepsilon_t, F_{t-1}]$  is time varying (conditional second moment).

We start with  $\varepsilon_t = \mathcal{L}_t \cdot \sigma_t$  where  $\mathcal{L}_t \sim \mathcal{N}(0, 1)$  and ARCH (1) :  $\sigma_t^2 = w + \alpha \varepsilon_{t-1}^2$

As we have just done with AR1, now look at conditional and unconditional second moment of ARCH(1).

$V[\varepsilon_t]$  and

$$\begin{aligned}
E[\varepsilon_t] &= E[\mathcal{L}_t \sigma_t] = \\
& E[\mathcal{L}_t] E[\sigma_t] \\
& 0 \cdot E[\sigma_t] = 0 \\
V[\varepsilon_t] &= E[\varepsilon_t]^2 E[\mathcal{L}_t^2 \cdot \sigma_t^2] = \\
& E[\mathcal{L}_t^2] \cdot E[\sigma_t^2] = E[\sigma_t^2]
\end{aligned}$$

$V[\varepsilon_t|F_{t-1}]$  (not right yet)

$$\begin{aligned}
E[\varepsilon_t|F_{t-1}] &= \\
& E[\mathcal{L}_t] E[\sigma_t] \\
& 0 \cdot E[\sigma_t] = 0 \\
V[\varepsilon_t] &= E[\varepsilon_t]^2 E[\mathcal{L}_t^2 \cdot \sigma_t^2] = \\
& E[\mathcal{L}_t^2] \cdot E[\sigma_t^2] = E[\sigma_t^2]
\end{aligned}$$

### General Settings

So far we have focused on the estimation of the conditional mean function  $E[Y_t|F_{t-1}]$  :

$$Y_t = E[Y_t|F_{t-1}] + \varepsilon_t$$

Where  $\varepsilon_t$  is a weak white noise, that is,  $\varepsilon_t$  is serially uncorrelated :  $Cov[\varepsilon_t, \varepsilon_{t-j}] = 0 \quad \forall j \neq 1$

Although, the weak white noise assumption does not say anything about *serial dependency* in the higher moments of  $\varepsilon_t$ , eg  $Cov(\varepsilon_t^2, \varepsilon_{t-j}^2) \neq 0$  for some  $j \neq 0$  But so far we have assume that  $\varepsilon_t$  is conditionally and unconditionally *homoskedastic*:

$$V[\varepsilon_t] = V[\varepsilon_t|F_{t-1}] \equiv \sigma_\varepsilon^2 \forall t$$

## GARCH-Type Models

If we relax the conditionally homoskedastic assumption and assume the following decomposition of the error term :

$$\varepsilon_t = z_t \sqrt{\sigma_t^2}$$

where  $z_t$  is an iid error term with zero expectation and unit variance. The function  $\sigma_t^2$  is assumed to be a function *only of* past information  $\sigma_t^2 \equiv \sigma_t^2(\mathcal{F}_{t-1})$

## ARCH(1) Processes

A process  $\sigma_t^2$  is called an ARCH(1) process if

$$\sigma_t^2 = w + \alpha \varepsilon_{t-1}^2$$

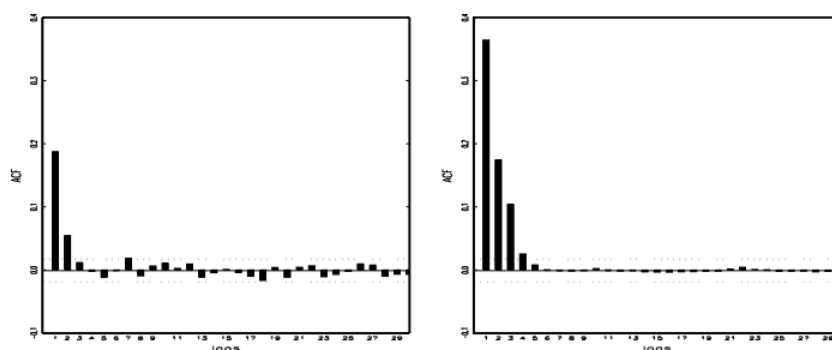
With  $w > 0$  and  $\alpha \geq 0$

## Properties of Arch(1)

- Arch (!) Conditional variance  $\sigma_t^2$  is strictly positive if  $w > 0$  and  $\alpha \geq 0$
- Opposite to the historical volatility estimator, the arch 1 volatility is a weighted average of past information that gives more weight to the recent information than to the distant one
- Arch 1 process can be written as an A(1) process in  $\varepsilon_t^2$
- Consequently  $\varepsilon_t^2$  is stationary if  $|\alpha| < 1$
- Given that both process  $\varepsilon_t$  and  $\varepsilon_t^2$  and  $E[\varepsilon_t] = 0$  then the unconditional variance of  $\varepsilon_t$ ,  $E[\varepsilon_t]$  is given by

$$\sigma_\varepsilon^2 = V[\varepsilon_t] = E[\varepsilon_t^2] = \frac{w}{1 - \alpha}$$

- ARCH(1) captures the clustering effect : when volatility is high, it more probably stays high
- The kurtosis is always large



Autocorrelation functions of squared time series with ARCH(1) conditional variance with  $\alpha = 0.2$  (left panel) and  $\alpha = 0.7$  (right panel)

Figure 2

Conditional variance moment, we observe a high persistence in daily log returns in order to cauterises this lag persistence, this lag has to be large too. But the estimation of this A(50) model becomes very cumbersome, likelihoods optimise numerically, once you start imposing Stationarity conditions this it rot ensure generating something with a stationary second moments, these are some solutions to polynomial equations so we run into large p issues.

In tutorial we look at arch's in simulation study

## Lecture 4: GARCH

Mon 19 Feb 09:00

[FE-L4] [3.5, 3.6, 3.8,3.9]

### Recap

ARMA

1.  $E[\varepsilon_t] = 0$
2.  $V\varepsilon_t = \sigma^2$
3.  $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$  that is no serial correlation

Tutorial 2 : S&P 500 Daily log returns  $\rightarrow$  ARMA(p,q)  $\rightarrow$  BIC then use residual diagnostics

$$@_t = y_t - \hat{E}[y_t|F_{t-1}] \rightarrow MA(\mathcal{L})$$

Week 3

NP / Rob Engel 2003

$$\begin{aligned}\varepsilon_t &= \sigma_t \mathcal{L}_t \\ \mathcal{L}_t &\sim \mathcal{N}(0, 1) \\ \sigma_t^2 &= w + \alpha \varepsilon_{t-1}^2 < -\end{aligned}$$

1.  $a \geq 0$  and  $\omega > 0$  - to ensure positivity of conditional variance
2.  $|\alpha| < 1$  Stationarity of conditional variance

ARCH(1)

$$\left\{ \begin{array}{l} \sigma = w + \alpha \varepsilon_{t-1} < - \\ \varepsilon_t + \mathcal{L}_t \sigma_t \\ \text{rewrite } \sigma_t^2 = w + \alpha \varepsilon_{t-1}^2 + \varepsilon_t^2 - \varepsilon_t^2 \\ AR(1) \text{ in } \varepsilon_t^2 \rightarrow \varepsilon_t^2 = w + \alpha \varepsilon_{t-1}^2 + (\varepsilon_t^2 - \sigma_t^2) \end{array} \right. \quad Video \left\{ \begin{array}{l} E[V_t] = 0 \\ V[v_t] < \infty v_t = \sigma^2 \\ Cov(v_t, v_{t-s}) = 0 \end{array} \right.$$

Pros

- Volatility clustering (video)
- Rise persistence at the cost of ARCH (p)
- Leptokurtic property  $\alpha^2 \in (0, \frac{1}{3})$

Cons

- Leverage effect :  $E[\mathcal{L}_t^3] = 0$
- Long memory (ACF)

What can we do with our Garch models to capture all remaining things in ACF?

### 3.4 GARCH

A process  $\sigma_t^2$  is called an GARCH(1, 1) process if

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

With  $\omega > 0$ ,  $\alpha \geq 0$  and  $\beta \geq 0$

#### Properties

- $\varepsilon_t^2$  is stationary if  $\alpha + \beta < 1$
- Both processes  $\varepsilon_t$  and  $\varepsilon_t^2$  are stationary and  $E[\varepsilon_t] = 0$  then the unconditional variance of  $\varepsilon_t$   $V[\varepsilon_t]$  which is equal to the unconditional mean of  $\varepsilon_t^2$
- No leverage effects as in the ARCH
- 

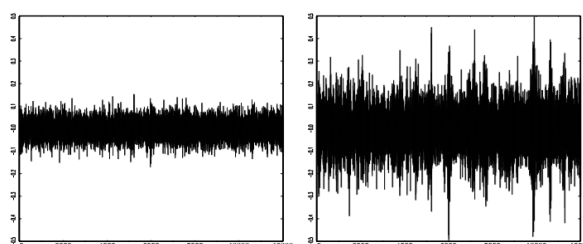


Figure 3: Simulated GARCH Models

Left  $\alpha = 0.01$  and  $\beta = 0.8$ . Right  $\alpha = 0.08$  and  $\beta = 0.9$  If allow close to 1 then can generate longer persistence, usually the memory of the daily log returns is us more persistent. Most have very low memory, thus people came up with GARCH(p, q)

M1 GARCH(1, 1)

- It takes into account / able to model more persistent conditional volatility processes
- Mitigating the tradeoff between generating a leptokurtic distribution of  $\varepsilon_t$  and the persistence of the ACF if  $\varepsilon_t^2$  as compared to ARCH(1)

M2 ARCH(1)

GARCH captures over kurtosis, even if we could like sum of  $\alpha + \beta$  to 1, we still have an opportunity to generate a over kurtosis ( $>3$ )

We can also show GARCH reveals larger excess kurtosis than the arch model, we can compare which is larger than the other,  $\frac{6\alpha^2}{1-2\alpha^2-(\alpha+\beta)^2}$

Can show  $A(1)$  is equal to  $MA(\infty)$ , same applies for GARCH for  $ARCH(\infty)$

$\alpha + \beta$  provides the necessary information on the degree of volatility clustering

### GARCH(p, q)

Just extension of GARCH(1, 1), key notation is polynomial for lag operator, lags shift an observation 1 period ahead (power 2 = 2 period ahead). But except for notation, nothing fundamental changes.

To lie outside of the root circle, in practice to estimate such a model, ensure positivity constraints, then also have to ensure process modelling is stationary - the constraints on stationary on highly non linear. This very quickly becomes a complicated non linear constraint, thus a numerical issue driven by Stationarity constraint (non linear) imposed by IRMA (p, q), but if allow for more p and q lags, then model is able to generate over kurtosis then the persistence of the series, the properties become better but at the cost of optimising over something with highly non-linear constraint.

### Further Types of GARCH models

ARCH provides an exponential decay, have to know GARCHS for risk modelling.

### Integrated GARCH(1, 1)

- Specific to high frequency time series
- Describes a very large persistence in the conditional variance
- Is strictly stationary
- Propose  $\alpha$  and  $\beta$  sum upto 1, GARCH STRUCTURE there to ensure non stationary process
- Risk metrics assumes that daily log returns follows process with infinite variance, that is we are not dealing with well defined statistical processes in real life, as seen by lack of first 2 moments

**RiskMetrics<sup>TM</sup>** A special case of the IGARCH(1, 1) process

- From estimating the
- Gives forecast
- $\lambda$  calibrates on loads of different stocks in the 90s
- Fix the  $\beta$  with  $\lambda$

**Exponential GARCH** Aimed at capturing asymmetric shocks, now modelling  $h_{t-1}$  log transformation of  $\sigma_t^2$ , assuming it follows GARCH looking process, and modify the ARCH part

- Modelling logs of variance because we want to get rid of parameter constraints, if modelling logs can be positive, negative, get rid of these issues by modelling logs
- 

**Threshold GARCH** TGARCH (1, 1) with indicator function, if shock was negative, bit easier to look at, if  $\gamma$  is positive, then ...

Tgarch, E garch if model left skewed

Tgarch(1,1) GJR-Garch

Usual garch(1, 1) :  $\sigma_t^2 = w + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$



$$\text{Tgarch}(1, 1) : \sigma_t^2 = w + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

$$\text{News impact curve} : \text{NIC}(\varepsilon_t | \sigma_{t-1}^2 = \sigma_{t-2}^2, \dots, \sigma_t^2)$$

$$\text{GARCH}(1, 1) : w + \beta \sigma_t^2 + \alpha \varepsilon_{t-1}^2 \quad \text{TGARCH}(1, 1) =$$

$$\begin{cases} W + \beta \sigma_t^2 + \alpha \varepsilon_{t-1}^2 \varepsilon_{t-1} < 0 \\ W + \beta \sigma_t^2 + \alpha + \delta \varepsilon_{t-1}^2, \varepsilon_{t-1} < 0 \end{cases}$$

$$\text{NIC} : \text{Egarch}(1,1)$$

$$H_t = \ln(\sigma_t^2) = w + \alpha \mathcal{L}_{t-1} + \gamma(|z_{t-1}| - \sqrt{\frac{2}{a}}) \exp(h_t) = \sigma_t^2 = \exp^w \cdot \exp^{\alpha z_{t-1}} \cdot \exp^{\gamma(|z_{t-1}| - \sqrt{\frac{2}{a}})} \sigma_t^2 = \exp^w \cdot \sigma^2$$

$$\varepsilon_t > 0$$

$$\varepsilon_t < 0$$

If shock positive then  $\exp^{\alpha+\gamma} \cdot \varepsilon_t / \sigma_t$

NIC: once you write down NIC, then it becomes more evident what model parameters give you which response, EGarch  $\alpha < 0$ ,  $z_t$  between 0 and 1

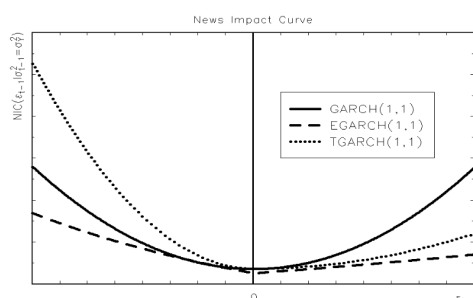


Figure 4: News Impact Curve

Model quality based on one picture, isn't exactly true, in order to plot NIC. Plug in  $\sigma$ ,  $\gamma$ ,  $\beta$  ( $\omega$ ), plot is based on one set of parameters, can easily be reversed.

So has something to do with data rather than overall quality of model,

**Recap** ARCH, GARCH, IGARCH, EGARCH, TGARCH. Financial econometrics model conditional second moment, but what about first moment?

- Conditional mean? (1st moment), why are we interested in the second moment?
- We are risk averse etc, but
- In week 2 we have talked about how to model, ARMA - expected value of  $y_t$  then T2 we estimated conditional mean models, but the returns are on average 0, there is very slight autoregressive coefficients, but overall there is **no time series structure** in the conditional mean :

$$E[r_t | F_{t-1}] = 0$$

- WE have compared the ACF for daily log returns  $r_t$ , but in the actual return series, the history of returns is completely uninformative of the future
- In autocorrelation function few squared return we see a lot going on, and it doesn't die out, squared return is a proxy of conditional variance

Why do we model conditional second moment?

There is no time series structure to first moment, but there is in conditional second moment. Then we think how can we model our conditional variance of return process?

Nobel prize given for ARMA framework where  $\varepsilon_t$  can be white noise process. Then, even GARCH is not enough.

Then RiskMetrics comes and assumes infinite variance of daily returns, albeit a popular way of thinking. How much does turbulence persevere in market, how long after do we have to be conservative in our risk approaches

EGARCH, TGARCH. TGARCH more intuitive, EGARCH model the log variances and so can relax the positivity constraints, we don't care whether shocks are negative. Essentially a philosophical introduction to risk-modelling

## Lecture 5: Model Estimation and Forecasting

Mon 26 Feb 08:58

[FE-L5] [Reading]

### 3.5 Recap

Week 1 : Leverage effects (skewness + testing whether neg) , volatility clustering (time series) , long memory (ACF of squared returns series) , leptokurtic property (sample skewness testing against 3). Properties (plots/ test)

Week 2 : limitations of ARMA modelling, which assume innovations are white noise - nothing about conditional heteroskedasticity). Unconditional - variance of innovations is constant over time, but evidence empirically that conditional 2nd moment seems to be time variant.

Week 3 : Rob engles ARCH ARCH(1) model  $\sigma_t^2 = f(\{t-1\})$ . Pro - volatility clustering, con - leverage of  $\{t$ , but long memory for very large p, kurtosis  $\alpha^2 \in (0, \frac{1}{s})$

Week 4 : GARCH(1,1) - pro - volatility clustering and long memory and overkurtosis, con - leverage TGARCH, EGARCH  $\rightarrow$  leverage. M (IGARCH) .

### Maximum Likelihood

Quasi Maximum Likelihood

Maximum likelihood - have data  $x_1, \dots, x_t$  then **assume** this data follows *some* distribution.

Which is function of the parameters, say  $x_t \sim N(\mu, \sigma^2)$  and  $\Theta = (\mu, \sigma^2)$

Then have PDF of data  $f(x_t, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2} \exp(-\frac{(x-\mu)^2}{2\sigma^2})}$ . If assume normal dist, then each and every value of  $x_t$  you know probability this data came from this distributing, then voter the entire sample you can take the likelihood function

$$\begin{aligned}\mathcal{L}|\mu, \sigma^2 &= \prod_{t=1}^T f(x_t|\mu, \sigma^2) \\ &= f(x_1|\mu, \sigma^2) \cdot f(x_2|\mu, \sigma^2) \dots\end{aligned}$$

So take log likelihood that is a function of data for given value of parameters  $\mu, \sigma^2$

$$\log \mathcal{L}(x_1, \dots, x_T|\mu, \sigma^2) = \ln\left(\prod_{t=1}^T f(x_t|\mu, \sigma^2)\right)$$

In any time series we work with quasi likelihood, in classical ML you must be able to evaluate likelihood function at each and every point. At an autoregressive process of order 1 (AA(1)).

Have  $\varepsilon_t \sim N(0, \sigma^2)$  so  $\varepsilon_t = y_t - c - \phi y_{t-1}$  which is  $N(0, 1)$

Then we have

Why quasi-likelihood?

Likelihood for first population :  $f(e_1|c, \phi, \sigma^2)$ , we assume  $y_0$  is ...

Now likelihood function becomes function of data and parameters, but *also initial values* depending on how many autoregressive lags are there.  $\rightarrow$  it is not really a likelihood. The conditioning makes it a quasi-likelihood

### 3.6 Estimation, Model choice and forecasting

Use knowledge of Max likelihood to ascertain which model fits the data best

Assume

$$\begin{aligned}R_t &= c + \varepsilon_t \\ E_t &= \mathcal{L}_t \sigma_t^2 = w + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2\end{aligned}$$

Estimate with Max likelihood.

$$\varepsilon_t = r_t - c E_t | \mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$$

$$E[c_t | \mathcal{F}_{t-1}] \text{ and } V[\varepsilon_t | \mathcal{F}_{t-1}]$$

Where  $\sigma_t = f(\mathcal{F}_{t-1})$  and  $\varepsilon_t | \mathcal{F}_{t-1} = \mathcal{L} | F_{t-1} \sigma_t | \mathcal{F}_{t-1}$

Where  $f(\varepsilon | F_{t-1}, \theta) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{\sigma_t^2}\right)$

And

$$\theta = (c, w, \alpha, \rho)$$

$$= \frac{1}{\sqrt{2\pi(w + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2)}} \cdot \exp\left(-\frac{(v_t - c)^2}{2(w + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2)}\right)$$

$$\sigma_0^2 = \frac{w}{1 - \alpha - \beta} = \frac{1}{T} 2(r_t - \hat{\mu})^2$$

Normally distributed innovations. From likelihood theory, the best is the one with the largest likelihood.

Estimation of GARCH Models

Model :  $Y_t = X_t' \gamma + \varepsilon_t$

The conditional variance of  $\varepsilon_t$  follows a GARCH(p, q) model

- $M = \max(p, q)$  Numbers of initial observations  $t = -m + 1, -m + 2, \dots, 0$

### Conditional maximum likelihood

Normal  $Z_t$

Student t  $Z_t$

Assume  $z_t \sim T(\nu)$  (std student t dist) , then :

$$\begin{aligned} E[Z_t] &= 0 \\ V[Z_t] &= \frac{\nu}{\nu - 2} \\ \text{Density Function} &= \frac{\Gamma[(\nu + 1)/2]}{(\pi\nu)^{1/2} \Gamma[\nu/2]} \left[ 1 + \frac{z_t^2}{\nu} \right]^{-(\nu+1)/2} \end{aligned}$$

Often estimated using standardised student t distribution, which is symmetric so expected value is 0, and

In PS1, there was ex on student t distribution with different degrees of freedom - the larger the dof, the closer to normal RV, smaller the heavier the tails (more outliers). 1 dof - Cauchy distribution

## 3.7 Model Choice and Diagnostics

Verify if there are ARCH effects in

- The original series of interest  $Y_t$
- The residuals from a mean regression  $\hat{\varepsilon}_t$  The residuals standardised by the estimated GARCHS  $\hat{z}_t = \frac{\hat{\varepsilon}_t}{\sqrt{\hat{\sigma}_t^2}}$

**Test for ARCH effects**

**ARCH-M test**

Auxiliary regression on the series of interest  $\bar{x}_t$  (original series, residuals, standardised residuals):

$$\bar{x}_t^2 = \psi + \alpha_1 \bar{x}_{t-1}^2 + \alpha_2 \bar{x}_{t-2}^2 + \dots + \alpha_m \bar{x}_{t-m}^2 + \varepsilon_t$$

With  $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_m = 0$  and  $H_A : H_0$  is not true

**Standardised Residual Diagnostics**

Assuming you already estimate a GARCH model for series

Verify if there are still ARCH effects left in the series (if the estimated GARCH model is correctly specified) by performing standardised residual diagnostic tests on the residuals standardised by the estimated GARCH conditional volatility ( $\hat{z}_t = \frac{\varepsilon_t}{\sqrt{\hat{\sigma}_t^2}}$ )

| Model             | parameters           |                     |                     |                    |                    |                   | AIC    | ARCH-LM test       | JB test              |
|-------------------|----------------------|---------------------|---------------------|--------------------|--------------------|-------------------|--------|--------------------|----------------------|
|                   | c                    | $\omega$            | $\alpha$            | $\beta$            | $\gamma$           | d                 |        |                    |                      |
| ARCH(1)           | 7.94E-05<br>(0.5838) | 0.0002<br>(0.0000)  | 0.2502<br>(0.0000)  |                    |                    |                   | -5.388 | 71.985<br>(0.0000) | 29533.63<br>(0.0000) |
| GARCH(1,1)        | 0.0004<br>(0.0019)   | 2.41E-6<br>(0.0000) | 0.0603<br>(0.0000)  | 0.9339<br>(0.0000) |                    |                   | -5.543 | 3.948<br>(0.4130)  | 15747.28<br>(0.0000) |
| Risk Metrics (RM) | 0.0004<br>(0.0109)   |                     |                     | 0.9615<br>(0.0000) |                    |                   | -5.529 | 10.395<br>(0.0349) | 22530.63<br>(0.0000) |
| EGARCH(1,1)       | 0.0001<br>(0.3297)   | -0.1474<br>(0.0000) | -0.0475<br>(0.0000) | 0.9920<br>(0.0000) | 0.1099<br>(0.0000) |                   | -5.565 | 5.998<br>(0.1117)  | 8276.66<br>(0.0000)  |
| TGARCH(1,1)       | 0.0001<br>(0.2409)   | 2.57E-6<br>(0.0000) | 0.0274<br>(0.0000)  | 0.9343<br>(0.0000) | 0.0653<br>(0.0000) |                   | -5.557 | 2.6153<br>(0.624)  | 8865.977<br>(0.0000) |
| FIGARCH(0.4,1)    | 0.0003<br>(0.0081)   | 1.87E-6<br>(0.0000) |                     | 0.2898<br>(0.0000) |                    | 0.370<br>(0.0000) | -5.502 | 3.248<br>(0.4251)  | 18833.38<br>(0.0000) |

Note: The round parentheses give the p-values; The JB test is done on the standardized residuals.

Figure 5: Estimation of different GARCH Models

Arch(1) is capturing overkurtosis, since it is able to generate outliers ( $\alpha$  is sig diff from 0). But intercept is not sig different from 0.

Arch-LM test and JB test are tested on  $\dots$ , both tests are redirected, there is remaining heteroskedacity,  $\alpha$  relatively mild.

*GARCH* - passing arch lm test, decay in ACF is very slow,  $\alpha, \beta$  close to 1, very persistent, but able to measure conditional heteroskedacity

RM - re estimated on data, p val for ARCH lm is 0.04, depends on confidence interval determines rejection. But none are looking like norm RV

E(T) GARCH - neagtive shocks (response to future volatlity)  $\gamma$  positive. Egarch model log variances,

EGARCH -  $\alpha$  - if shock negative then log of variance should be multiplied with negative variance (asymmetric response, how much is shock differnt from abs value of expected shock)

$\alpha$  and  $\gamma$ ? Negative and positive for egarch - at 5% sig level, all garchs seem to model sufficiently long memory using model parameters, out of these (ignoring fact dont past JB test of normality)

When we talked about ARMA we talked about AIC, BIC allowing us to compare different models estimated using ML, but different models have different parameters, so to control for this have different penalty functions (k denotes parameters).

Even asymmetric GARCH are unable to account for negative ( $\beta$ ), we see in the data. Thus we require advanced financial econometrics

Garch loved since it is easy to forecast risk with them, central banks require risk forecasting on daily basis - using GARCH(1,1) is very easy for this.

**Exercise 1.** TGARCH(1,1) Estimated  $\hat{\sigma}_t^2 = \hat{w} + \hat{\alpha}\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \dots$

$$\begin{aligned} E[\sigma_{t-1}^2 | \mathcal{F}_t] &= \hat{w} + \alpha r_t^2 + \hat{\beta}\hat{\sigma}_t^2 + \dots \\ E[\sigma_{t-1}^2 | \mathcal{F}_t] &= w + \alpha E[\varepsilon_t^2 | \mathcal{F}_t] + \beta E[\sigma_{t+1}^2 | \mathcal{F}_t] + \dots \end{aligned}$$

Expected value

$$W + \alpha E[\sigma_{t+1}^2 | \mathcal{F}_t] + \beta E[\sigma_t^2 | \mathcal{F}_t] + \gamma E[\pi(z_t)]$$

### Forecasting with Risk Metrics

Let  $\sigma_t^2$  follow a riskmetrics model:

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

Where  $\lambda = 0.94$

### 3.8 Variance Forecast Evaluation

$\sigma^2$  is not observed, it may be replaced by proxies such as

- $\sigma_{t+h}^2 = r_{t+h}^2$  (squared daily returns)
- $\sigma_{t+h}^2 = RV_{t+h}$  daily realised variance

Or alternatively, we evaluate the variance forecasts within economic applications :

- Value at risk, expected shortcuts,
- Asset pricing etc

Good forecasting performance does not translate to good in sample fit (tradeoff?)

**Tutorial 1.** 5 Last week simulated GARCH, this week estimating GARCH and forecasting based on the estimates. In PS4 we have simulated  $y_t = c + \psi y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t$  Chose some parameters, then simulated based upon those parameters, and once we had innovations, we simulated for values of ARMA parameters, simulated the ARMA recursions This week have the daily log returns of SNP500, estimate ARMA and GARCH parameters which are coming from data, why cant we just plug these parameters in and use them in the simulation? Taking our  $\hat{\sigma}^2$  and simulate returns, why cant we do this and why instead do we forecast where  $\hat{\sigma}_t^2 = \hat{w} + \gamma r_{t-1}^2 + \hat{\beta}\sigma_{t-1}^2$  Simulated series which resemble data properties is defined as  $E[\sigma_{t+1}^2 | \mathcal{F}_t]$

## Lecture 6: Kalman Filter

Mon 04 Mar 09:04

[Lecture]

$$E[\sigma_{t+1}^2 | \mathcal{F}_t] = w + \alpha \varepsilon_t^2 + \beta \sigma_t^2$$

If we think about the classical ARMA-GARCH framework, we have

1. returns with some conditional mean  $+ \varepsilon_t$  where

$$\begin{aligned}
r_t &= E[r_t | \mathcal{F}_{t-1}] + \varepsilon_t \\
\varepsilon_t &= \mathcal{L}_t \cdot \sigma_t(\mathcal{F}_{t-1}) \\
\mathcal{L}_t &\sim \mathcal{N}(0, 1)
\end{aligned}$$

If we would like to assume that returns is driven by

$$\begin{aligned}
r_t &= \varepsilon_t = \mathcal{L}_t \sigma_t \\
\sigma_t^2 &= f(\mathcal{F}_{t-1}) + q_t \\
\varepsilon_t | \mathcal{F}_{t-1} &\sim \mathcal{N}(0, 1) \\
f(\varepsilon_t) &= \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right)
\end{aligned}$$

Multivariate normal distribution

$$\begin{bmatrix} r \\ y \end{bmatrix} \approx N\left(\begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_{yy} \end{pmatrix}\right)$$

We need the  $x|y$  distribution to derive the Kalman filter, the transformation can get us certain properties

Expected value (these are population parameters, fixed values, numbers measuring fixed variance)

$$\begin{aligned}
E[z] &= E[x] - \Sigma_{xy} \Sigma_{yy}^{-1} (E[y] - \mu_y) \\
&= E[Z'Z] \quad \text{if scalar, then } E[z^2] \\
&= E[(x - \Sigma_{xy} \Sigma_{yy}^{-1} (y - \mu_y)) (x - \Sigma_{xy} \Sigma_{yy}^{-1} (y - \mu_y))] \\
&\quad \text{algebra} \dots \\
\text{cov}(x, y) &= E[xy] - E[x]E[y] \\
&\rightarrow u_x u_y - E[xy'] = -\text{cov}(x, y) = -\Sigma_{xy} \\
&\quad \text{and so the whole term} \\
&= E[X'X] - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma'_{xy} \\
\text{cov}(z, z) &= E[ZZ'] - E[Z]E[Z]' \\
&= E[xx']
\end{aligned}$$

We have 2 RV with joint distribution, we want to understand the conditional distribution

$$X|y \approx N(\dots)$$

Then if we take the transformation  $z$

$$z = x - \Sigma_{xy}$$

1.  $E[z] = E[x]$
2.  $V[z] = \Sigma_{xx} - \Sigma_{xy}^{-1} \Sigma_{xy}$

3.  $\text{Cov}(y, z) = 0$

$$x = z + \Sigma_{xy} + \Sigma_{yy}^{-1}(y - \mu_y)$$

$$E[x|y] = E[z|y] + \Sigma_{xy}\Sigma_{yy}^{-1}(E[y|y] - \mu_y)$$

First term =  $E[z]$ , second term is same third is  $y$  so

$$E[x|y] = \mu_x = \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)$$

$$V[x|y]$$

$$x|y \approx N(\mu_x)$$

### Local Trend Model

In order to understand the SV model, consider a simple local trend model first

$$y_t = \mu_t + e_t, \quad e_t \sim N(0, \sigma_e^2), \quad (13)$$

$$\mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2), \quad (14)$$

Where  $\{e_t\}$  and  $\{\eta_t\}$  are 2 independent Gaussian white noise series and  $t = 1, \dots, T$

The initial value  $\mu_1$  is either *given or follows a known distribution* and is independent of  $\{e_t\}$  and  $\{\eta_t\}$  for  $t > 0$ .

Here  $\mu_t$  is a pure *random walk* with initial value  $\mu_1$  and  $y_t$  is an observed version of  $\mu_t$  with added noise  $e_t$ .  $\mu_t$  is referred to as the *trend* of the series which is not directly observable, and  $y_t$  is the observed data with observational noise  $e_t$ .

The models above can be used to analyse the *realised* volatility of an asset price, where  $\mu_t$  represents the underlying log volatility of the asset price and  $y_t$  is the log of realised volatility.

The model is a special linear *gaussian state space model*, with the variable  $\mu_t$  called the *state* of the system at time  $t$  (not directly observed).

The  $y$  model provides the link between the data  $y_t$  and the state  $\mu_t$  and is called the *observation equation* with measurement error  $e_t$

The next  $\mu_{t+1}$  governs the time evolution of the state variable and is the state equation with innovate  $\eta_t$ .

If  $\sigma_e = 0$  then  $y_t = \mu_t$  and there is no measurement error, which is an ARMA(0, 1, 0) model

If  $\sigma_e > 0$  then there exist measurement error and  $y_t$  is an ARMA(0, 1, 1) model satisfying

$$(1 - B)y_t = (1 - \theta B)a_t \quad (15)$$

where  $\{a_t\}$  is a gaussian white noise with mean zero and variance  $\sigma_a^2$

Then, the values of  $\theta$  and  $\sigma_{eta}$  are determined by  $\sigma_e$  and  $\sigma_\eta$

From the initial model we have

$$(1 - B)\mu_{t+1} = \eta_t \quad \text{or} \quad \mu_{t+1} = \frac{1}{1 - B}\eta_t \quad (16)$$

$$\text{then we can rewrite } y_t = \mu_t + e_t = y_t = \frac{1}{1 - B}\eta_{t-1} + e_t$$



And multiplying by B we have

$$(1 - B)y_t = \eta_{t-1} + e_t - e_{t-1}$$

Then letting  $(1 - B)y_t = w_t$  we have  $w_t = \eta_{t-1} + e_t - e_{t-1}$ . And under the model assumptions it is easy to see that  $w_t$  is gaussian,  $\text{Var}(w_t) = 2\sigma_e^2 + \sigma_\eta^2$  and  $\text{Cov}(w_t, w_{t-1}) = -\sigma_e^2$  and  $\text{Cov}(w_t, w_{t-j}) = 0$  for  $j > 1$ .

Then consequently  $w_t$  follows an MA(1) model and can be written as  $w_t = (1 - \theta B)a_t$ .

And by equating the variance and lag-1 autocovariance of  $w_t = (1 - \theta B)a_t = \eta_{t-1} + e_t - e_{t-1}$  and we have

$$\begin{aligned}(1 + \theta^2)\sigma_a^2 &= 2\sigma_e^2 + \sigma_\eta^2 \\ \theta\sigma_a^2 &= \sigma_e^2\end{aligned}$$

Then for a given  $\sigma_e^2$  and  $\sigma_\eta^2$  we consider the ratio of these to form a quadratic function of  $\theta$ , having 2 solutions which we select the one that satisfies  $|\theta| < 1$ .

Idea is that if you have state equation with large variance, you won't be able to recover much.

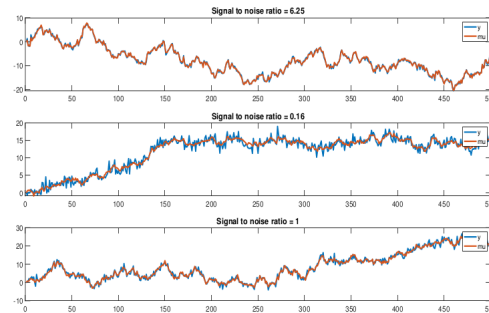


Figure 6: State Space Model

If signal to noise ratio = 0.16, observe blue try to recover red. It is not very informative, if ratio is 6 then signal is very informative.

### Kalman Filter

The aim of the analysis is to infer properties of the state  $\mu_t$  alone from the data and the model. Let  $F_t = \{y_1, \dots, y_t\}$  be the information available at time  $t$  (inclusive) and assume that the model is known, including all parameters.

Three estimates of interest

1. Filtering : recover  $\mu_t$  (remove measurement error)
2. Smoothing : estimate  $\mu_t$  given all available information up to time  $T$
3. Prediction : forecast  $\mu_{t+k}$

Analogy - filtering is figuring out the word you are reading based on knowledge accumulated from the beginning of the note, predicting is to guess the next word and smoothing is to decipher a particular word once you have read through the note.

**Properties of Multivariate Normal Distribution** Considering a multivariate normal distribution

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_{yy} \end{pmatrix}\right)$$

Kalman filter is a tool which characterises the conditional distribution of  $\mu_t$  given the data. Given the data we observe what is the distribution of  $\mu_t$

### Notation

Let  $\mu_{t|j} = E[\mu_t|F_j]$  and  $\Sigma_{t|j} = \text{Var}(\mu_t|F_j)$  be the conditional mean and variance of  $\mu_t$  given  $F_j$ .  $y_{t|j}$  is the conditional mean of  $y_t$  given  $F_j$

And  $v_t = y_t - y_{t|t-1}$  and  $V_t = \text{Var}(v_t|F_{t-1})$  be the 1 step ahead forecast error and its variance of  $y_t$  given  $F_{t-1}$

The forecast error  $v_t$  is independent of  $F_{t-1}$  so that the conditional variance is the same as the unconditional variance, that is  $\text{Var}(v_t|F_{t-1}) = \text{Var}(v_t)$

Then

$$Y_{t|t-1} = E[y_t|F_{t-1}] = E[\mu_t + e_t|F_{t-1}] = E[\mu_t|F_{t-1}] = \mu_{t|t-1}$$

And consequently,

$$v_t = y_t - y_{t|t-1} = y_t - \mu_{t|t-1}$$

and

$$V_t = \text{Var}(y_t - \mu_{t|t-1}|F_{t-1}) = \text{Var}(\mu_t + e_t - \mu_{t|t-1}|F_{t-1})$$

$$= \text{Var}((\mu_t - \mu_{t|t-1}|F_{t-1})) + \text{Var}(e_t|F_{t-1}) = \Sigma_{t|t-1} + \sigma_e^2$$

And then it is easy to see that

$$E[v_t] = E[E[y_t - y_{t|t-1}]|F_{t-1}] = E[y_{t|t-1} - y_{t|t-1}] = 0$$

$$\text{Cov}(v_t, y_j) = E[v_t, y_j] = E[E[v_t y_j|F_{t-1}]] = E[y_j|E[v_t|F_{t-1}]] = 0, \quad j < t$$

Then as expected the 1 step ahead forecast error is uncorrelated with  $y_j$  for  $j < t$ . And furthermore for the linear model in eq. (13) and eq. (14)  $\mu_{t|t} = E[\mu_t|F_t] = E[\mu_t|F_{t-1}, v_t]$  and  $\Sigma_{tZt} = \text{Var}(\mu_t|F_t) = \text{Var}(\mu_t|F_{t-1}, v_t)$

That is, the information set  $f_t$  can be written as  $F_t = \{F_{t-1}, y_t\}$

### Theorem 1 :

Properties of MV normal distribution useful to the Kalman filter under normality Suppose that  $x$ ,  $y$  and  $z$  are 3 RV such that their joint distribution is MV normal, additionally assume that the diagonal block covariance  $\Sigma_{ww}$  is non singular for  $w = x, y, z$  and  $\Sigma_{yx} = 0$ , then

1.  $E[x|y] = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)$
2.  $\text{Var}(x|y) = \Sigma_{xx} - \Sigma_{xx}\Sigma_{yy}^{-1}\Sigma_{yx}$
3.  $E[x|y, z] = E[x|y] + \Sigma_{xz}\Sigma_{zz}^{-1}(z - \mu_z)$
4.  $\text{Var}(x|y, z) = \text{Var}(x|y) - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}$

Then the conditional distribution of  $x$  given  $y$  is

$$x|y \sim \mathcal{N}(\mu_x + \Sigma_x y \Sigma_y^{-1} (y - \mu_y), \Sigma -)$$

$$\begin{pmatrix} \mu_t \\ \nu_t \end{pmatrix} | \mathcal{F}_{t-1}$$

Goal is the conditional distribution  $\mu_t | F_t$  based on new data  $y_t$  and the conditional distribution  $\mu_t | F_{t-1}$

$$\begin{pmatrix} \mu_t \\ \nu_t \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} \mu_{t|t-1} \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{t|t-1} & \Sigma_{t|t-1} \\ \Sigma_{t|t-1} & V_t \end{pmatrix} \right)$$

### Prediction

Initialise idea on conditional mean and variance of unobserved latent state variable (signals), take data minus initial value/expectation.

Look at forecast error variance

How does it work?

First remove measurement error then estimate  $\mu_t$  given all available information, then forecast. Distribution of  $\mu_t$  given information set  $\mathcal{F}_t$  today. In order to today recover the value of  $\mu_t$  need to update conditional expectation so that take into account signal to noise ratio. How much new noise contributes to the conditional variance expectation.

Filter latent process based on information  $t-1$  then we update forecast once new information has arrived. Kalman gain measures how much information does the new shock at time  $t$  add to uncertainty (?). Don't take information as given  $y_t$  has noise itself  $e$  so we only update conditional expectations proportionally to the signal to noise ratio.

Recover, then smoothing re estimating  $\mu_t$  (trying to mitigate effect of starting values), then based on this forecast latent process. All based on one property of MVR norm.

After we know this we can write it down given this formula

Major idea of KF is to write down some expectations of latent process, then update these according to Kalman gain which measures model uncertainty plus new variance originating from noisy data. Which are inherently small in financial data. New data is not very informative (nothing in autocorrelation structure), so strongly depends on starting values. These values are not eaten up by new data as they may in physics.

[11]

### Kalman Filter

The goal of the Kalman filter is to update knowledge of the state variable recursively when a new data point becomes available. That is, knowing the conditional distribution of  $\mu_t$  given  $F_{t-1}$  and the new data  $y_t$ , we would like to obtain the conditional distribution of  $\mu_t$  given  $F_t$  where as before  $F_j = \{y_1 \dots, y_j\}$  since  $F_t = \{F_{t-1}, v_t\}$  giving  $y_t$  and  $F_{t-1}$  is equivalent to giving  $v_t$  and  $F_{t-1}$ .

To derive the KF, it suffices to consider the joint conditional distribution of  $(\mu_t, v_t)'$  given  $F_{t-1}$  before applying the above theorem

The conditional distribution of  $v_t$  given  $F_{t-1}$  is normal with mean zero and variance given by

$$V_t = \text{Var}(\mu_t - \mu_{t|t-1} | F_{t-1}) + \text{Var}(e_t | F_{t-1}) = \sum_{t|t=1} + \sigma_e^2$$

And that of  $\mu_t$  given  $F_{t-1}$  is also normal with mean  $\mu_{t|t-1}$  and variance  $\Sigma_{t|t-1}$

Then, what remains to be solved is the conditional covariance between  $\mu_t$  and  $v_t$  given  $F_{t-1}$

From the definition

We can obtain

$$\begin{pmatrix} \mu_t \\ v_t \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} \mu_{t|t-1} \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{t|t-1} & \Sigma_{t|t-1} \\ \Sigma_{t|t-1} & V_t \end{pmatrix} \right)$$

Then, using theorem 11.1 [11.1.2], the conditional distribution of  $\mu_t$  given  $F_t$  is normal with mean and variance

$$\begin{aligned} \mu_{t|t} &= \mu_{t|t-1} + \frac{\Sigma_{t|t-1} v_t}{V_t} = \mu_{t|t-1} + K_t v_t \\ \Sigma_{t|t} &= \Sigma_{t|t-1} - \frac{\Sigma_{t|t-1}^2}{V_t} = \Sigma_{t|t-1} (1 - K_t) \end{aligned} \quad (17)$$

Where  $K_t = \Sigma_{t|t-1}/V_t$ , is the *Kalman Gain*, the regression coefficient of  $\mu_t$  on  $v_t$  then from eq. (17) the Kalman gain is the factor that governs the contribution of the new shock  $v_t$  to the state variable  $\mu_t$

Then, one can use the knowledge of  $\mu_t$  given  $F_t$  to predict  $\mu_{t+1}$  via eq. (18)

$$\mu_{t+1|t} = E[\mu_t + \eta_t | F_t] = E[\mu_t | F_t] = \mu_{t|t} \quad (18)$$

$$\Sigma_{t+1|t} = \text{Var}(\mu_{t+1} | F_t) = \text{Var}(\mu_t | F_t) + \text{Var}(\eta_t) = \Sigma_{t|t} + \sigma_\eta^2 \quad (19)$$

So, once the new data  $y_{t+1}$  is observed, one can repeat the above procedure to update knowledge of  $\mu_{t+1}$ . Which is the famous *Kalman filter* algorithm.

### Kalman Filter For Local Trend Model

$$\begin{aligned} v_t &= y_t - \mu_{t|t-1} \\ V_t &= \Sigma_{t|t-1} + \sigma_e^2 \\ K_t &= \Sigma_{t|t-1}/V_t \\ \mu_{t|t+1} &= \mu_{t|t-1} + K_t v_t \\ \Sigma_{t+1|t} &= \Sigma_{t|t-1} (1 - K_t) + \sigma_\eta^2, t = 1, \dots, T \end{aligned}$$

### Tutorial 6

#### [PS6]

Is the signal to noise ratio the same no matter the number of simulations?

SNR is defined as  $\frac{\sigma_e^2}{\sigma_\eta^2}$ , the variances determine this, here they do not depend on anything, it is always defined by the variance of the error term in state equation and in the observation model.

B) mu and filter s

Our forecast error is defined as in [slide 11]

To understand the code, write the recursions using this slide

---

```

1 for (t in 1:T){
2   predict_mu[t] = filter_mu[t]
3   predict_S[t] = filter_S[t]
4   v[t] = y[t] - predict_mu[t]
5   V[t] = predict_S[t] + s_e^2
6   K[t] = predict_S[t] / V[t]
7   filter_mu[t+1] = predict_mu[t] + K[t] * v[t]
8   filter_S[t+1] = predict_S[t] * (1 - K[t]) + s_eta^2
9   print(V[t])
10 }

```

---

Old conditional expectation  $\mu_{t|t+1} = \mu_{t|t} + K_t \cdot \nu_t$

In the for loop we take our conditional predictions and feed them 1 step ahead in the next iteration, so we have filtered out in iteration t becomes a prediction in t + 1 so  $\nu_{t+1} = y_{t+1} - \mu_{t+1|t}$

Kalman filter updating - the filtered at t becomes a prediction for t+1

Also possible to start at 2 and do filtering at t-1

What are the filter initialisations?

Can also draw from MV norm, this is local linear trend model with strong SNR, starting values matter less here, though this is the usual way to initialise.

Filter gets updated proportional to Kalman gain.

$$\begin{aligned}
 \nu &= y_1 - u_{1|0} = y_1 - \text{predict\_mu}[1] \\
 V_1 &= \text{predict\_S}[1] + \sigma_e^2 \\
 K_1 &= \dots \\
 \text{filter}[1] &= \text{predict\_mu}[1] + V[1] \cdot K[1] \\
 &\text{then for } t = 2
 \end{aligned}$$

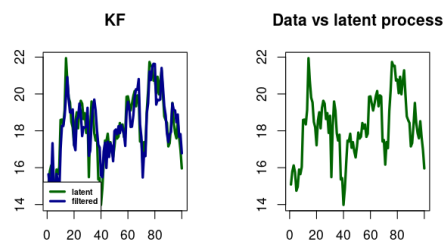


Figure 7

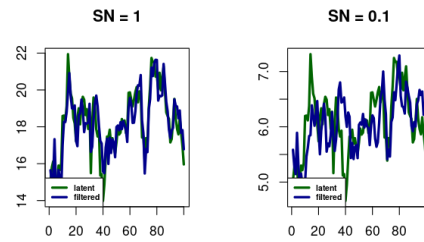


Figure 8

CI

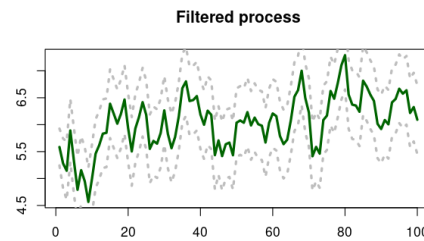


Figure 9

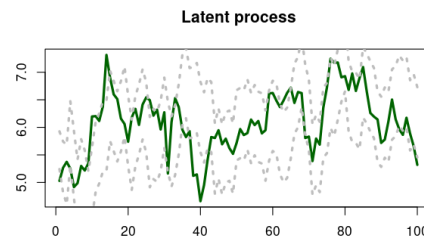


Figure 10

Fundamental property / assumption on  $\mu_t$ , we assumed distribution given is **normal** and thus once we obtain filtration we can use this to write down CI. Conditional variance is given by filtered  $\Sigma$ .

**Hypothetical Exam Question** Why would CI be much wider where  $\sigma_e^2 = 0.9$  and  $\sigma_q^2 = 0.3$  than  $\sigma_e^2 = 0.9$  and  $\sigma_q^2 = 0.9$

The second has a SNR of 1, so higher ratio means the closer the observation and state equations are. Thinking about how CI are calculated,  $\pm 2 \cdot \Sigma_{t|t}$  or  $\Sigma_{t|t} = \Sigma_{t|t} + \sigma_q^2$

[recursive slide] - what about Kalman gain, need to think about how  $\sigma_\eta^2$  influences variance, what does increase in  $\sigma_e^2$ , this is in denominator of Kalman gain :

This is all a recursive process

Why do we opt to calculate negative log likelihood?

---

```

1  kf_loglik = function(y,s_e,s_eta){
2  fit = kf_recurions(y,s_e,s_eta)
3  # compute the negative log likelihood
4  l = 0.5*log(2*pi)+0.5*(log(fit$V))+ 0.5*((fit$v^2)/fit$V)
5  ll =sum(l)
6  return(ll)
7  }

```

---

Estimates are RV, they can lie anywhere so have to do negative.

## Lecture 7

Fri 15 Mar 09:13

## Lecture 8: Realised Volatility

Mon 15 Apr 09:02

[ECON60332 Financial Econometrics]

GARCH etc developed a long time ago, these days we mostly work with realised variances and advances in this domain

### Idea

- A more consistent estimator of daily variance
- Computed from the intra day (high-frequency) information
- Use *current* intraday data to extract information about current level of volatilities
- Based on a continuous time model for the price process
- It is not based on a deterministic parametric specification (a non parametric measure)
- Measure of the variance that has been realised over a particular period

### Mathematical framework

- Assume a logarithmic asset price process  $p(t)$  that evolves continually in time for each  $t \in [0, \infty)$
- Then the continuously compounded return over a n-th period (usually a day) is defined by

$$r_n = p(nh) - p((n-1)h), \quad n = 1, 2, \dots, N$$

where h is the length of the period (h = 1 is a day) and N denotes the number of periods (days) in which we observe the process

- Assume we have  $M + 1$  intra-daily observations during each day
- The intra-daily returns, for the n-th day, are defined as :

$$r_{j,n} = p((n-1)h + \frac{hj}{M}) - p((n-1)h + \frac{h(j-1)}{M}), \quad j = 1, \dots, M$$

M is the number of intraday observations, j is the iterator that goes over those continuous increments

This formula gives you the log price difference over a time increment j (small time period, almost 0) The actual time  $\frac{h}{m}$  is the length of the time increment ( $10^{-6}$  for NYSE) Time increment is very small, the increment over a millisecond is very small, plotting the continuous time evolution will be very smooth line M - intraday periods

### realised variance (RV)

The realized variance for day n based on M intra-daily returns is defined as the sum of the squared intra-daily returns over that day:

$$RV_n^M = \sum_{j=1}^M r_{j,n}^2$$

- If intra-daily returns are 5-minute returns, then  $M = 78$  for six and a half hours of trading during a day
- The realised volatility does not estimate the variance of 5-min returns, but the daily variance

Parametric estimator is where you assume some parameters, once you know them you can describe the DGP Non-parametric - DGP is some function, index by time, there is no parameter that covers the intraday prices for instance

Point of confusion - RV is a estimate of the variance of 5 min returns - **it is not** Realised over a day, not over a time period that is sitting in a 5 minute period, also not divided by M

We are not measuring a risk at the beginning at the beginning and end of interval, we are measuring the risk that has been realised over this time period.

In continuous time, calculating this over many time increments (in fact infinitely small), if we would like to measure the risk over this period, we would need to integrate the risk which can of course also be written as infinite sum.

Integrate the risk, risk ( $\sigma^2$ ) is continuous process as well as price

### Quadratic Variation

- Assume further that the log-price  $p(t)$  follows a continuous-time marginal process :
  - $E[|p(t)|] < \infty$  for all  $t \geq 0$
  - $E[(p(t)|p(s)), 0 \leq s < t] = p(s)$
  - Examples :  $W(t)$ , the standard Brownian motion
- The quadratic variation (QV) process associated with the martingale  $p(t)$  process is given by :

$$QV(t) = \underbrace{\lim}_{M \rightarrow \infty} \sum_{j=1}^{M-1} (p(t_{j+1}) - p(t_j))^2$$

for any sequence of partitions  $t_0 = 0 < t_1 < \dots < t_M = t$  with  $\sup t_{j+1} - t_j \rightarrow 0$  as  $M \rightarrow \infty$

For any continuous marginal process we can write down the RV, that accumulates the risk from period 0 to t



RV estimator approximates the increments of the QV (risk that accumulates from period 0 to t), which is the quantity we are interested in.

Assume continuous martingale on prices

### Relation between QV and RV

Thus, when  $M \rightarrow \infty$  :

$$RV_n^M \xrightarrow{p} QV(hn) - qv(H(N-1))$$

- QV(t) accumulates the variance of the martingale process from 0 to t
- RV measures the increment of "quadratic variation" (over a day) associated with the continuous martingale log price process
- The asymptotics of RV are infill (the time period, n, is held fixed; the number of observations within the period, M, tends to infinity)
- Assume further that the log price  $p(t)$  follows a continuous time stochastic volatility process without jumps (a very broad subclass of the martingale class processes)

$$p(t) = \int_0^t \sigma(s) dW(s), \quad t \geq 0$$

- $\sigma$  is the spot volatility,  $\sigma^2$  the spot variance
- $\sigma$  is independent of W
- Model can capture volatility clustering and fat tails

### Integrated Variance

Really learn the difference

Because of the Brownian motion assumption on the measure, the returns will be normal and will have 0 mean.

- The increments of the log-price process (daily returns) have the following distribution, conditional on the process  $\sigma(t)$  :

$$r_n | \sigma_n^{[2]} \sim N(0, \sigma_n^{[2]})$$

where

- $\sigma_n^{[2]} = \sigma^2 * (nh) - \sigma^2 * ((n-1)h)$
- $M \sigma^2 * (t) = \int_0^t \sigma^2(t)$
- The process  $\sigma^2 * (t)$  is called the integrated variance

Assuming price is continuous time martingale, realised variance don't have to assume anything

### Facts about RV

martingale process best price of tomorrows price is today?

Consistency of RV holds under the assumption that the price process is a *continuous martingale* process  
Problems:

- The price discreteness leads to discrimination error in RV measurement
- The martingale property is distorted by spurious autocorrelations in the high-frequency data stemming from market microstructure effects:
  1. Price discreteness
  2. Rounding effects
  3. Bid-ask bounces
  4. Gradual response of price to a block-trade
  5. Strategic order flows
  6. Data recording mistakes

Market microstructure (particularly strategic order flows) can be taken advantage of

But often, data is cleaned rather than taking advantage of microstructures in raw data

- In order to reach efficiency (smallest variance), one should choose the highest possible sampling
- The ultra-high sampling induces bias in RV
- Bias-variance trade-off: assessment through volatility signature plot
- Volatility signature plot depicts the average of a very long sample of RV's against the sampling frequency
- Empirically, the plot stabilises at frequencies ranging from 5 to 40 minute

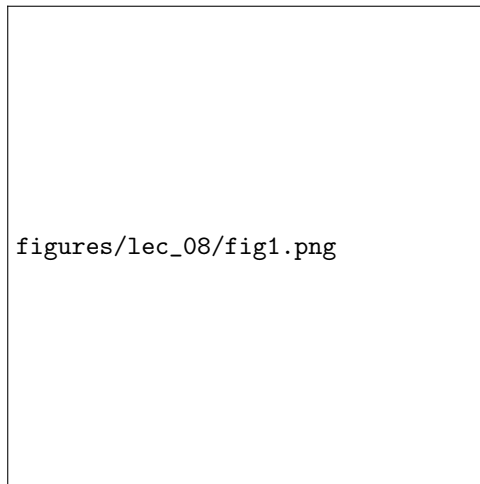


Figure 11: Volatility signature plot for daily RV of IBM from 01.01.2001 until 30.06.2006

Ideally, we would like to pick sample frequency as high as possible, however there is a tradeoff. The efficiency tells us to go high frequency, however this results in biased RV due to microstructure.

So typically, 5-minutes so see less and less of these patterns

Problem is that it is realised, still need some model to forecast it going into the future, these are long memory models.

**HAR model for  $\sqrt{RV}$** 

Heterogeneous autoregressive model

$$X_t = x + \beta^{(d)} X_{t-1}^{(d)} + \beta^{(w)} X_{t-1}^{(w)} + \beta^{(m)} X_{t-1}^{(m)} + \varepsilon_t$$

where

- $M$   $X_t = \sqrt{RV_t^M}$
- $d$  stands for the daily
- $w$  for weekly (5 days)
- $m$  for monthly (22 days) frequency.  $X_t^{(\cdot)}$  are averages of past values of  $RV_t$ , EG  $X_t^{(w)} = \sqrt{\frac{1}{5} \sum_{i=0}^4 RV_{t-i}}$
- The future volatility is predicted by daily, weekly and monthly volatilities

**sample properties of RV**

- The logarithm of daily realised standard deviations (squared root of daily RV) are approximately normally distributed
- The daily returns standardised by realised standard deviations are approximately normal distributed
- The autocorrelation function of daily RV series is slowly decaying: long memory property of financial volatility
- In order to obtain forecast of future volatility from a RV time-series apply eg, long-memory models: ARFIMA or HAR of Corsi (2009)
- Further approaches of modelling series of RV: mixed data sampling etc

Historical volatility is calculated based on CID of daily returns, in contrast we have better

- Prices are unit root non-stationary (unpredictable ) processes
- Have better measure of daily risk (no longer no need for GARCH)
- GARCH still popular in practice due to RiskMetrics
- Not a lot of people in industry use realised variance since intra day data is expensive, thus less popular
- GARCH for latent processes

**Summary**

- More consistent estimator of daily variance
- Computed from high frequency (expensive ) intraday data
- Almost error-free - sometimes market microstructure meaning that martingale assumption is sometimes violated (not by much)
- No parametric assumptions, just that log price is martingale
- Understanding more about model, parameters etc need model then

- Realised over a day, we are looking at continuous time, some time interval, what is the risk over this interval

Market microstructure

### 3.9 Tutorial

Long memory is the persistence of financial risk over time, meaning that for several months (IBM 80 / 90 days) We cannot model this by any means of standard models since we would need an AR(90) model, instead, we start looking for a model which is capable of modelling  $\sigma_t^2$  enough such that A GARCH model is pretty bad, either some kurtosis or not all the memory is captured well enough To check whether realised variance is empirically a good measure of daily risk, we take day returns and divide by  $\sqrt{RV_t}$

Ideally the standardised returns should look like an iid variable with standardised distribution. If there is some systematic Then that We learn about stochastic modelling Realised variance is a non parametric estimator, so far we have not estimated anything, we have took information, summed and squared it, and with this we are able to model the realised volatility. LBQ test, all autocorrelations are jointly 0 Pictures are suggestive evidence rather than concrete evidence

Writing Economic interpretation of QV(t)

This is the financial risk accumulated upto period t (+1) , assuming  $p(t)$  is continuous, liquid assets. QV(t) measures the price variation of the continuous martingale process for over infinitely small periods of time

Key is to demonstrate knowledge, think about people and their money and trading, subject is mitigating peoples risks

## Lecture 9: Value at Risk

Mon 22 Apr 09:04

### Idea

- It is defined to be the worst expected loss with probability  $p$
- It is a quantile-based risk measure
- It is mainly concerned with the market risk
- It is used by financial institutions to assess their risks
- It is used by regulatory committee to set margin requirements to financial institutions

Probability  $p$  is any level you can choose, but in Basel committee it is 1% Market risk is mostly driver of the risk but risky mortgages can increase this

**Quantile** For any univariate cumulate distribution function  $F(x)$  and probability  $p$ , such that  $0 < p < 1$ , the quantity

$$\inf\{x|F(x) \geq p\}$$

is called the  **$p$ -th quantile** of  $F(x)$ . We use the term  $F^{-1}(p)$  to denote the  $p$ -th quantile of  $F(x)$

Smallest  $x$  at the support of  $F(x)$  such that Inverse of the CDF

### Value at risk for a long position

- $\Delta V_t(l)$  - change in the value of the financial asset  $V_t$  from  $t$  to  $t + l$  (in a given currency)
- $F_t(x)$  - cumulative distribution function (CDF) of  $\Delta V_t(l)$

We implicitly define at time  $t$  to the VaR of a long position in this financial asset over time horizon  $l$  with probability  $p$  s.t.

$$p = \Pr[\Delta V_t(l) \leq \text{VaR}_t(p, l)] = F_t(\text{VaR}_t(p, l))$$

Then

$$\text{VaR}_t(p, l) = F_t^{-1}(p)$$

- VaR is the  $p$ -th quantile of the distribution of the asset of profits and losses
- For a long position, the var is defined as a negative value (loss)

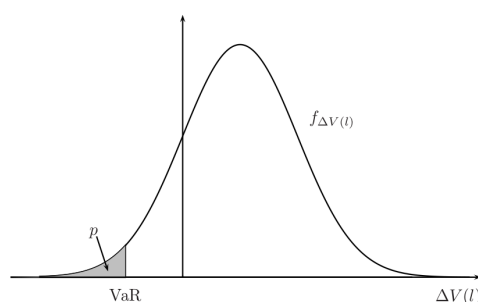


Figure 12

### Interpretation 1

- The Var of a portfolio at level  $p$  is the minimal amount that can be lost with probability less than  $p$
- The var is a loss bound that at most will be exceeded with probability  $p$
- Example 1 :  $\text{Var}_t(1\%, 1)$  consider the investment over the horizon of one trading day. Assume  $\text{Var}_t(1\%, 1) = 100,000$ . given 100 (independent) trading days then we can expect that in 1 out of 100 trading days the realised loss will exceed 100,000 (the VAR)

### Basel Accords

- In 1996 the Basel Committee imposes banks to use the var as a risk measure of their investment and sets minimum capital requirements
- Banks are allowed to use their own internal models to compute VAR
- Impose a method for testing the accuracy of the internal models used by banks to measure their market risk : **back testing**
- In 2001, the Basel committee sets capital charges to market risk from the Var measure
- In 2012, Basel 3 sets further regulation on the minimum capital requirement and recommends the stress-testing

### Type of VaR measurements

How to get predicted density from profit and loss?

- (Un)conditional extreme value theory based approaches - EVT
- (Un)conditional non-parametric based approaches - historical simulations
- Unconditional parametric approaches under independence
- Conditional parametric approaches - conditional quantile regressions, GARCH with non-parametric estimates of the tails

**VAR for net returns** Let  $r_t(l)$  denote the l-period net return of  $V_t$ ,  $r_t(l) = \frac{V_{t+l} - V_t}{V_t}$ , then

$$p = \Pr[V_{t+l} - V_t \leq VAR_t(p, l)] = \Pr\left[\frac{V_{t+l} - V_t}{V_t} \leq \frac{VAR_t(p, l)}{V_t}\right]$$

It follows that  $Var_t(p, l) = F^{-1}(p)V_t$  where  $F_t$  is the cdf of the net returns  $r_t(l)$

**VAR for log returns** Let  $r_t(l)$  denote the l-period log return of  $V_t$ ,  $r_t(l) = \ln(\frac{V_{t+l}}{V_t})$ , then

It follows that  $Var_t(p, l) = F^{-1}(p)V_t$  where  $F_t$  is the cdf of the net returns  $r_t(l)$

If sample is a bit more volatile, and you need a 30 day ahead forecast, then you multiply mean by 30 days (cumulative return over 30 days)

Var computation with RiskMetrics model Subtract mean of k-period ahead mean standardise with variance, then same for LHS, then rearrange with terms

### Basel Committee

- Don't have legal power, but local central bank is heavily influenced by the reports
- Developed to try to avoid the collapse of financial systems, introducing the first notion of reserves

### Back testing

- Punishment of banks which face more violations (losses below the VAR) that can reasonably be expected given the confidence level : are required to hold higher levels of capital
- Banks with more than 9 violations out of 250 trading days are required to follow the standardized approach (8%)
- Imposition of penalty leads to reputation loss, higher capital charges and possibly to the introduction of a more stringent external model to forecast the VAR thresholds

**Procedure** Unconditional coverage test :

$$E[H_t] = p$$

where  $H_t = 1(r_t < \hat{Var}_t(p))$   $t = T + 1, T + 2, \dots, T + S$

A green light is max 4 violations - internal models A yellow light - between 5 and 9 violations : progressive penalties A red light - more than 9 violations : "standardised approach"

Independence test

$$E[H_t|F_{t-1}] = p \text{ and } H_t|F_{t-1} \text{ iid } Ber(p)$$

Conditional coverage test

| Distribution | Stock Type | Start date | ND         |        |        |              | TD         |        |        |              | EVT        |
|--------------|------------|------------|------------|--------|--------|--------------|------------|--------|--------|--------------|------------|
|              |            |            | ARMA-GARCH | RM-est | RM-fix | ARMA-FIGARCH | ARMA-GARCH | RM-est | RM-fix | ARMA-FIGARCH | ARMA-GARCH |
| small        | 1987       |            | 2.79       | 2.79   | 2.79   | 2.09         | 2.09       | 2.09   | 2.09   | 2.09         | 0.69       |
|              | 1996       |            | 2.79       | 2.79   | 2.79   | 2.09         | 2.09       | 2.09   | 2.09   | 2.09         | 0.69       |
|              | 2001       |            | 2.09       | 2.09   | 2.79   | 2.09         | 2.09       | 2.09   | 2.09   | 2.09         | 0.69       |
|              | 2005       |            | 2.09       | 2.09   | 2.79   | 2.09         | 2.09       | 2.09   | 2.09   | 2.09         | 2.09       |
| middle       | 1987       |            | 2.09       | 2.79   | 2.79   | 3.49         | 1.39       | 1.39   | 2.09   | 1.39         | 0.69       |
|              | 1996       |            | 1.39       | 2.79   | 2.79   | 1.39         | 1.39       | 2.09   | 2.09   | 1.39         | 0.69       |
|              | 2001       |            | 2.09       | 2.79   | 2.79   | 2.79         | 1.39       | 2.79   | 2.09   | 2.09         | 1.39       |
|              | 2005       |            | 2.79       | 3.49   | 2.79   | 2.09         | 1.39       | 2.79   | 2.09   | 1.39         | 1.39       |
| large        | 1987       |            | 2.09       | 3.49   | 3.49   | 3.49         | 2.09       | 2.09   | 2.09   | 2.79         | 0.69       |
|              | 1996       |            | 2.79       | 3.49   | 3.49   | 2.79         | 2.09       | 2.79   | 2.09   | 2.09         | 0.69       |
|              | 2001       |            | 2.79       | 3.49   | 3.49   | 2.09         | 2.09       | 2.79   | 2.09   | 2.09         | 1.39       |
|              | 2005       |            | 3.49       | 4.19   | 3.49   | 2.79         | 2.79       | 2.09   | 2.09   | 2.79         | 2.09       |

Figure 13: Halbleib Pohlmeier (2011): VaR through calm period

Percentage rate of violations of VaR(1%) for the period January 1, 2007 – July 18, 2007 (total of 143 days). Bold type entries refer to p-values of conditional coverage test smaller than 0.05, italic to p-values between and 0.05 and 0.10 and no mark refers to p-values larger than

VaR had bad performance in the financial crisis in 2008, inherently due to the fact it is a time series, it has not seen this data before.

Tradeoff - do you go with conservative model to be in green zone or use ARMA garch quantile

| Distribution | Stock Type | Start date | ND         |        |        |              | TD         |        |        |              | EVT        |
|--------------|------------|------------|------------|--------|--------|--------------|------------|--------|--------|--------------|------------|
|              |            |            | ARMA-GARCH | RM-est | RM-fix | ARMA-FIGARCH | ARMA-GARCH | RM-est | RM-fix | ARMA-FIGARCH | ARMA-GARCH |
| small        | 1987       |            | 2.94       | 2.54   | 2.54   | 3.13         | 2.35       | 1.96   | 1.76   | 1.96         | 0.39       |
|              | 1996       |            | 3.13       | 2.54   | 2.54   | 3.13         | 2.54       | 1.96   | 1.76   | 2.15         | 1.17       |
|              | 2001       |            | 3.33       | 2.54   | 2.54   | 2.15         | 2.94       | 2.15   | 1.76   | 1.56         | 1.56       |
|              | 2005       |            | 2.94       | 2.35   | 2.54   | 3.13         | 2.74       | 2.35   | 1.76   | 1.56         | 2.74       |
| middle       | 1987       |            | 3.52       | 2.74   | 2.74   | 3.52         | 2.54       | 2.15   | 2.15   | 2.15         | 1.17       |
|              | 1996       |            | 3.52       | 2.74   | 2.74   | 2.94         | 2.54       | 2.15   | 2.15   | 1.96         | 1.37       |
|              | 2001       |            | 3.52       | 2.74   | 2.74   | 2.54         | 2.74       | 2.35   | 2.15   | 2.15         | 1.96       |
|              | 2005       |            | 4.11       | 2.54   | 2.74   | 3.13         | 2.74       | 2.35   | 2.15   | 1.96         | 1.96       |
| large        | 1987       |            | 3.72       | 3.13   | 2.74   | 3.72         | 1.96       | 1.96   | 1.76   | 2.15         | 0.98       |
|              | 1996       |            | 3.72       | 3.13   | 2.74   | 2.74         | 1.96       | 1.96   | 1.76   | 1.76         | 0.98       |
|              | 2001       |            | 4.11       | 3.13   | 2.74   | 2.74         | 2.35       | 2.15   | 1.76   | 1.76         | 1.17       |
|              | 2005       |            | 4.51       | 9.21   | 2.74   | 5.68         | 2.35       | 2.15   | 1.76   | 3.13         | 1.17       |

Figure 14: Halbleib Pohlmeier (2011): Var through crisis period

Major critique if VAR is that it is very sensitive to model risk, and doesn't necessarily prevent banks from large trading losses, since notion of being conservative in risk modelling isn't a feature and it doesn't capture systematic risk

### Coherent risk measures

1. Translation invariance  $\rho(+c) = \rho(X) - c$  for all  $c$  - hedge risk by investing into something constant to reduce risk

2. Sub additivity  $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$  - risk of investing into 2 things is not larger than sum of 2 risks
3. Positive homogeneity :  $\rho(\lambda X) = \lambda \rho(X)$  for all  $\lambda \geq 0$
4. Monotonicity  $X \leq Y \rightarrow \rho(Y) \leq \rho(X)$

1,3,4 is from value at risk is quantile

Why no subadditivity in VAR?

Value at risk approach has the disadvantage that it doesn't take into account the distribution of the log, given that it has occurred, e the shape of the tail of the profit-loss distribution. Therefore we consider here another risk measure called "expected shortfall"

### Expected Shortfall

The expected shortfall (ESF) is defined as

$$\begin{aligned} ESF_t(p, l) &= E[X_t | X_t < VAR_t(p, l)] \\ &= \int_{-\infty}^{VAR_t(p, l)} x dF_t(x | x < VAR_t(p, l)) \end{aligned}$$

- ESF is the expected loss under the conditional that the VAR threshold has already been crossed. as such, it can discriminate between positions with the same VAR'S but with different distributions of the extreme losses
- ESF is a coherent risk measure

$$\sum_{i=1}^n$$