Financial Econometrics

Sol Yates

February 5, 2024

Contents

Lecture 1: Introduction Lecture

Wed 31 Jan 11:19

1 Financial Time Series and their Characteristics

1.1 Asset Returns

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 :

$$1 + R_t[k] = \frac{P_t}{P_{t-k}} = \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \times \ldots \times \frac{P_{t-k+1}}{P_{t-k}}.$$

$$= (1 + R_t)(1 + R_{t-1}) \ldots (1 + R_{t-k+1}).$$

$$= \prod_{j=0}^{k-1} (1 + R_{t-j}).$$

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 gold for k years, then the annualized average return is defined as

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

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

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

Where it is easier to compute the arithmetic average than the geometric mean and the one-period returns tend so 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 $10\%/m$ 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 :

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

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,

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

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 rissoles 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 θ. 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

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

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

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

Variance

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

Skewness

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

Kurtosis

$$K[Y_t] = E[\frac{(Y_t - \mu)^4}{\sigma^4}] \quad , \quad \hat{S} = \frac{1}{T}\sum_{t=1}^{T}[\frac{(Y_t - \mu)^4}{\sigma^4}][].$$

Jargue-Bera test, tests H_0 of normality of the series :

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

where k is the number of estimated parameters. This test statistic has a χ^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

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

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.

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.

Lecture 2: Second Lecture - Review TS

Fri 02 Feb 15:45

2 Time Series Basics

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 processs

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
Definition

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

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

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 :

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

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

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

for a covaraince stationary process the j-th autocorrelation is given by :

Definition 7. white noise a TS is called this if it satisfies the following

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

white noise is a weakly stationary process

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

Definition 9. the empirical (or sample) autocorrelatino function of a time series Y_t is the sequence of sample autocorrelation coefficients $\hat{\rho}_j$ for all $j = 0, 1, 2, \ldots$:

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

and

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

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) $Y_{t-1}, \ldots, Y_{t-j+1}$

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

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

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

Definition 11. sample autocorrelation function

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

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

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

$$[r\hat{h}o_j - \frac{2}{\sqrt{T}}, r\hat{h}o_j + \frac{2}{\sqrt{T}}]$$

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

3 Arma Processes

Definition 12. a time series is called an autoregressive process of order p if it satisfies a relationship of the type :

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

where ε_t is a white noise error term

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

$$Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t$$

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

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

where ε_t is a white noise error term

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

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

Example. AR(1) Process

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

3.1 ARMA

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

$$\phi_p(L) Y_t = c + \Theta(L)\varepsilon_t where$$

$$\phi_p(L) = 1 - \phi_1 L - \phi_2 L^{(2)} - \ldots . - \phi_p L^{(p)}$$

$$\Theta(L) = 1 + \theta_1 L + \theta_2 L^{(2)} + \ldots + \theta_q L^{(q)}$$

with ε_t being a white noise and ϕ_p and Θ_q are called lag polynomials

3.2 ARMA estimation

ARMA(p, q) process:

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

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

ARMA estimation
Conditional log likelihood
estimation is done under asssumption that error term is normal.
LBQ 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-LM test

tests for conditional heteroskecadity in regression residuals

pick ARMA based on this, if modelled succesfully then null of LBQ test shouldnt be rejected and there shouldnt be any conditional heterskedacity