Capital University of Economics and Business

ISEM

Financial Econometrics

Summary

Instructor:
CHEUNG Ying Lun

Term:
Fall Semester, 2019

November 11, 2019

1 Asset returns

1.1 Simple returns

The return of holding an asset for one period of time is

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

If one holds an asset for k periods of time, the cumulative k-period return is just the product of the previous k one-period returns, namely

$$1 + R_t[k] = \frac{P_t}{P_{t-k}} = \frac{P_t}{P_{t-1}} \frac{P_{t-1}}{P_{t-2}} \cdots \frac{P_{t-k+1}}{P_{t-k}} = (1 + R_t) \cdots (1 + R_{t-k+1}).$$

The average return $\bar{R}$ is defined as the geometric mean of the past one-period returns.

$$1 + R_t[k] = (1 + \bar{R})^k \iff 1 + \bar{R} = \left(\prod_{j=0}^{k-1} (1 + R_{t-j}) \right)^{1/k}.$$

1.2 Compounding

The effective interest rate (or return) can be affected by frequency of compounding. Suppose *per annum* interest rate is r, and the frequency of compounding is m, then for every $1/m$ year, there is a r/m interest. Therefore, the effective interest rate is

$$1 + R = \left(1 + \frac{r}{m}\right)^m.$$

For example, suppose the *per annum* interest rate is 10%, then the effective interest rate for different frequency of compounding is given by the following table. It can be obser-

Type	**No. of payments**	**Interest rate per period**	**Total value**
Annual	1	10%	\$11000.00
Semiannual	2	5%	\$11025.00
Quarterly	4	2.5%	\$11038.13
Monthly	12	0.833%	\$11047.13
Weekly	52	0.192%	\$11050.65
Daily	365	0.027%	\$11051.56

Table 1: Values of a loan with 10% per annum interest rate

ved that the effective interest rate is higher, if the frequency of compounding increases. As $m \to \infty$, the effective interest rate becomes

$$1 + R = \lim_{m \to \infty} \left(1 + \frac{r}{m}\right)^m = e^r.$$

Equivalently, $r = \ln(1+R)$. Therefore, r is also called the *log return*. One advantage of using the log return is that, the multi-period log return is just the sum of the single-period returns, and so the average log return is the simple average of the one-period log returns.

$$r_t[k] = \sum_{j=0}^{k-1} r_{t-j}, \qquad \bar{r} = \frac{1}{k}\sum_{j=0}^{k-1} r_{t-j}.$$

1.3 Statistical distribution of asset returns

In this course, we would like to study the statistical properties of the asset returns. Here we first review some basics of statistical distributions.

1.3.1 Review of moments

Suppose that the marginal density function of X exists and is given by $f_X(x)$, then we can define the *moments* and *centered moments* of X as

$$m'_j = \mathbb{E}\left[X^j\right] = \int_{-\infty}^{\infty} x^j f_X(x)dx,$$
$$m_j = \mathbb{E}\left[(X-\mu_X)^j\right] = \int_{-\infty}^{\infty} (x-\mu_X)^j f_X(x)dx,$$

where $\mu_X = m'_1 = \mathbb{E}[X]$. We are mostly interested in the first four moments of X:

Mean The first moment measures the average location of X.

Variance The second centered moment measures the dispersion of X around its mean.

Skewness The third moment of the standardized variable measures the degree of asymmetry in the distribution of X.

Kurtosis The fourth moment of the standardized variable measures the fatness of the tails of the distribution of X, i.e., the probability of extreme values in X.

1.3.2 Review of statistical distributions

Normal distribution If $R \sim \mathcal{N}(\mu, \sigma^2)$, then it has the following properties:

- $S(R) = 0$ and $K(R) = 3$.
- $\hat{S}(R) \xrightarrow{d} \mathcal{N}\left(0, \frac{6}{T}\right)$ and $\hat{K}(R) \xrightarrow{d} \mathcal{N}\left(3, \frac{24}{T}\right)$.
- Therefore, one can construct a t-test to test for example $H_0 : S(R) = 0$ against $H_1 : S(R) \neq 0$. If the null hypothesis is rejected, then one can conclude that R is not normally distributed.

Log normal distribution Since the normal distribution does not have a lower bound, it is not suitable to be used to describe the behavior of asset returns. Instead, we can assume the log-return is normally distributed, i.e., $r = \ln(1+R) \sim \mathcal{N}(\mu, \sigma^2)$. Then we can show that

$$\lim_{r \to -\infty} R = \lim_{r \to -\infty} \mathrm{e}^r - 1 = -1.$$

Scale mixture of normal distributions In reality, the sample kurtosis of asset (log) returns is usually larger than 3, and therefore the normality test is rejected. The scale mixture of normal distributions can provide a higher kurtosis by increasing the probability of extreme events.

1.3.3 Risk management

If we know the statistical distribution of the asset return, we can then assess its risks by the following methods:

Value-at-risk Defined as $P(\Delta V \leq VaR_\alpha) = \alpha$, it is the maximum loss of an asset given a probability α. For example, if $\alpha = 5\%$, then there is 95% chance that the performance of the asset would be better than VaR_α.

Expected shortfall The VaR cannot fully describe the risk of an asset since it ignores the tail risk. The expected shortfall $ES_\alpha(\Delta V) = \mathbb{E}[\Delta V | \Delta V \leq VaR_\alpha]$ can complement the use of VaR by providing an assessment of the tail risk.

2 Stationary stochastic process

In Chapter 1, we are treating the asset returns as a random variable, say $r \sim (\mu, \sigma^2)$, and the return in each time period, r_t, as a realization of the random variable r. However, this ignores the dynamic relations among each r_t. Therefore, in this chapter, we will treat it as a *stochastic process*, i.e., a sequence of random variables sorted by time.

2.1 Linear process

If X_t has finite mean and variance, and has the representation

$$X_t = \mu_X + \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \qquad \varepsilon_t \overset{\text{iid}}{\sim} \left(0, \sigma^2\right),$$

then we say X_t is a linear process. Here we study four types of linear processes:

White noise $X_t = \mu_X + \varepsilon_t$.

- In this case, $c_0 = 1$ and $c_j = 0$ for all $j > 0$.

AR(p) model $X_t = a_0 + a_1 X_{t-1} + \cdots + a_p X_{t-p} + \varepsilon_t$.

- Suppose $p = 1$ and $|a_1| < 1$, then we can write

$$X_t = \mu_X + \sum_{j=0}^{\infty} a_1^j \varepsilon_{t-j}.$$

In this case, $c_j = a_1^j$.

MA(q) model $X_t = \mu_X + \varepsilon_t - b_1\varepsilon_{t-1} - \cdots - b_q\varepsilon_{t-q}$.

- In this case, $c_0 = 1$, $c_j = -b_j$ for $j = 1, \ldots, q$ and $c_j = 0$ for $j > q$.

ARMA(p, q) model $X_t = a_0 + a_1 X_{t-1} + \cdots + a_p X_{t-p} + \varepsilon_t - b_1\varepsilon_{t-1} - \cdots - b_q\varepsilon_{t-q}$.

- Suppose $p = q = 1$ and $|a_1| < 1$, then we can write

$$X_t = \mu_X + \left(1 + \sum_{j-1}^{\infty} a_1^{j-1}(a_1 - b_1)L^j\right)\varepsilon_t.$$

In this case, $c_0 = 1$ and $c_j = a_1^{j-1}(a_1 - b_1)$ for $j > 0$.

2.2 Properties of linear processes

2.2.1 White noise

If $X_t = \mu_X + \varepsilon_t$, and $\varepsilon \overset{\text{iid}}{\sim} (0, \sigma^2)$, then it is obvious that $X_t \overset{\text{iid}}{\sim} (\mu_X, \sigma^2)$. By the iid property, we know also that

$$\gamma_X(h) = 0, \qquad \rho_X(h) = 0, \qquad \text{for all } h > 0.$$

Moreover, if X_t is iid, then its sample autocorrelation follows

$$\hat{\rho}_X(h) \xrightarrow{d} \mathcal{N}(0, T^{-1}), \qquad h = 1, 2, \ldots.$$

Therefore, we can construct a t-test to test the following hypothesis

$$H_0^h : \hat{\rho}_X(h) = 0 \qquad \text{vs} \qquad H_1^h : \hat{\rho}_X(h) \neq 0$$

If we reject the null hypothesis, then we conclude that X_t is not a white noise. Alternatively, we can also test the joint hypothesis

$$\begin{cases} H_0 : \rho_X(1) = \cdots = \rho_X(m) = 0 \\ H_1 : \rho_X(i) \neq 0 \text{ for some } i \in \{1, \ldots, m\} \end{cases}$$

with the statistic

$$Q(m) = T(T+2)\sum_{j=1}^{m} \frac{\hat{\rho}_X(j)^2}{T-j} \xrightarrow{d} \chi_m^2.$$

Again, if we reject the null hypothesis, then we conclude that X_t is not a white noise.

2.2.2 AR(p) model

Stationarity Writing the AR process as

$$(1 - a_1 L - \cdots - a_p L^p)(X_t - \mu_X) = \varepsilon_t$$

and let $A(L) = 1 - a_1 L - \ldots a_p L^p$ be the lag polynomial. Then X_t is weakly stationary if and only if the root of the lag polynomial $A(z)$ is *larger than one* in absolute value. That is, the solutions of

$$A(z) = 1 - a_1 z - \cdots - a_p z^p = 0$$

are larger than one in absolute value. Note that if $z = r_j^{-1}$ are the solutions of the polynomial $A(z) = 0$, then we can factor the polynomial as

$$(z - r_1^{-1}) \cdots (z - r_p^{-1}) = 0 \iff (1 - r_1 z) \ldots (1 - r_p z) = 0.$$

Therefore, we can factor the lag polynomial as

$$(1 - r_1 L) \cdots (1 - r_p L)(X_t - \mu_X) = \varepsilon_t.$$

Equivalently, X_t is weakly stationary if $|r_j|$ are *smaller than one* for all $j = 1, \ldots, p$.

Moments If X_t is weakly stationary, then its mean can be obtained simply by taking expectation of the model and solve for μ_X. We have

$$\mu_X = \frac{a_0}{1 - a_1 - \cdots - a_p}.$$

The variance and autocovariance function of X_t can be obtained by the *Yule-Walker equations*, which is obtained by multiplying $X_{t-h} - \mu_X$ to the model and taking expectation,

$$\gamma_X(h) - a_1 \gamma_X(h-1) - \cdots - a_p \gamma_X(h-p) = \begin{cases} 0 & \text{if } h > 0 \\ \sigma^2 & \text{if } h = 0 \end{cases}$$

The Yule-Walker equations are useful in the following two ways:

1. We can use the relation to find higher order autocovariance.
2. We can estimate the coefficients a_j by using sample autocovariance and those equations.

For example, when $p = 1$, the autocovariance is given by

$$\sigma_X^2 = \gamma_X(0) = \frac{\sigma^2}{1 - a_1^2}, \qquad \gamma_X(h) = a_1^h \gamma_X(0).$$

Therefore, the autocovariance converges to zero as $h \to \infty$.

2.2.3 MA(q) model

Moments To find the mean of an MA process, we can simply take the expectation of the model and obtain $\mathbb{E}[X_t] = \mu_X$. By the iid assumption, the variance of X_t can be given by

$$\text{var}(X_t) = (1 + b_1^2 + \cdots + b_q^2)\sigma^2.$$

Similarly, the autocovariance can be simply obtained by

$$\gamma_X(h) = \begin{cases} (-b_h + b_{h+1}b_1 + \cdots + b_q b_{q-h})\sigma^2 & h \leq q \\ 0, & h > q \end{cases}$$

Here we can make two observations:

1. The autocovariance has a different pattern from that of an AR process. It becomes zero when $h > q$, while for an AR process, the autocovariance only becomes zero as $h \to \infty$.

2. The mean, variance and autocovariance do not depend on t, and are finite for any values of b_j. Therefore, an MA(q) process is always weakly stationary.

2.2.4 ARMA(p, q) model

Stationarity The stationary of an ARMA(p, q) model is the same as an AR(p) model.

Moments Similar to an AR process, the mean of a weakly stationary ARMA process can be obtained by taking the expectation of model, and obtain

$$\mu_X = \frac{a_0}{1 - a_1 - \cdots - a_p}.$$

Similarly, the autocovariance function can be obtained by using the Yule-Walker equations. For example, if $p = q = 1$, then

$$\gamma_X(1) = a_1\gamma_X(0) - b_1\sigma^2$$
$$\gamma_X(h) = a_1\gamma_X(h-1)$$

Similar to that of an AR(1) process, the autocovariance only goes to zero when $h \to \infty$.

2.3 Estimation

In reality, we only observe the sequence of data X_t. Therefore, we need to estimate the model. Here we introduce two methods.

2.3.1 Least square method

The AR(p) model has the same form as a linear regression model. Therefore, we can estimate the model by regressing X_t on its lags $\{X_{t-j}\}$.

2.3.2 Maximum likelihood estimation

For an ARMA model, since the lags X_{t-j} are correlated with the error terms e_{t-j}, the least square method leads to a bias. Therefore, we use the maximum likelihood estimation (MLE) instead. The idea of MLE is to choose a set of parameters, such that the probability of getting the set of observations $\{X_t\}$ is maximized. Therefore, the MLE is given by

$$(\hat{a}_0, \ldots, \hat{a}_p, \hat{b}_1, \ldots, \hat{b}_q, \hat{\sigma}^2) = \hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}; X_1, \ldots, X_T)$$

where the likelihood function $\mathcal{L}(\boldsymbol{\theta}; X_1, \ldots, X_T)$ is the joint density function of ε_t

$$\mathcal{L}(\boldsymbol{\theta}; X_1, \ldots, X_T) = f_{\varepsilon_{p+1},\ldots,\varepsilon_T}(\varepsilon_{p+1}, \ldots \varepsilon_T; \boldsymbol{\theta}).$$

Assuming $\varepsilon_{p+1-q} = \cdots = \varepsilon_p = 0$, we can express ε_t as a function of X_t and its lags, as well as the parameters $\boldsymbol{\theta}$.

2.4 Order selection

2.4.1 Autocorrelation function

If X_t is an MA(q) process, then we know that its autocorrelation becomes zero as $h > q$. Therefore, by plotting the autocorrelation function, we can choose q such that $\gamma_X(q+1)$ is close to 0.

2.4.2 Partial autocorrelation function

If X_t is an AR(p) process, then we can also express it as an AR($p+1$) process, with $a_{p+1} = 0$. Therefore, we can perform a series of estimation

$$\begin{aligned} X_t &= a_{0,1} + a_{1,1}X_{t-1} + e_{1t} \\ X_t &= a_{0,2} + a_{1,2}X_{t-1} + a_{2,2}X_{t-2} + e_{2t} \\ X_t &= a_{0,3} + a_{1,3}X_{t-1} + a_{2,3}X_{t-2} + a_{3,3}X_{t-3} + e_{3t} \\ &\vdots \end{aligned}$$

We can then choose the AR order such that $a_{p+1,p+1}$ is close to zero.

2.4.3 Information criteria

The order of an ARMA(p, q) model can be chosen by minimizing the information criteria functions

$$\begin{aligned} AIC(p,q) &= \ln(\tilde{\sigma}^2(p,q)) + \frac{2(p+q)}{T} \\ BIC(p,q) &= \ln(\tilde{\sigma}^2(p,q)) + \frac{(p+q)\ln T}{T} \end{aligned}$$

2.4.4 Model adequacy

By assumption, ε_t is a white noise sequence. Therefore, if X_t is an ARMA(p,q) process, and we are estimating an ARMA(i,j) model, where $i \geq p$ and $j \geq q$, then the estimated residuals should also behave like a white noise. Therefore, we can apply the white noise test on the estimated residuals. If the null hypothesis is not rejected, then we say the model is adequate. If the null hypothesis is rejected, then we need to increase either i or j to capture the remaining serial correlation in the residuals.

2.5 Forecasting

The optimal forecast of a stochastic process is given by the conditional expectation

$$\hat{X}_t(h) = \mathbb{E}\left[X_{t+h} | \mathcal{I}_t\right] = \mathbb{E}\left[X_{t+h} | X_t, \ldots, X_1\right].$$

In an MA or ARMA model, we need to further assume that the initial values of ε_t are zero. Then, we can express ε_t as a function of X_t. For example, in an ARMA(p,q) model, we have to assume that $\varepsilon_{p+1-q} = \cdots = \varepsilon_p = 0$, then we have

$$\begin{aligned}
\varepsilon_{p+1} &= X_{p+1} - a_1 X_p - \cdots - a_p X_1 - a_0 \\
\varepsilon_{p+2} &= X_{p+2} - a_1 X_{p+1} - \cdots - a_p X_2 - a_0 + b_1 \varepsilon_{p+1} \\
&\vdots
\end{aligned}$$

In this case, $\mathcal{I}_t = \{X_1, \ldots, X_t\} = \{\varepsilon_{p+1}, \ldots, \varepsilon_t\}$.

3 Nonstationary stochastic process

3.1 Unit root process

Consider the ARMA(p,q) process

$$(1 - a_1 L - \cdots - a_p L^p)(X_t - \mu_X) = (1 - b_1 L - \cdots - b_q L^q)\varepsilon_t.$$

If d roots of the lag polynomial $A(L) = 1 - a_1 L - \cdots - a_p L^p$ are equal to one, then

$$(1-L)^d (X_t - \mu_X) = (1 - r_{d+1} L)^{-1} \cdots (1 - r_p L)^{-1} (1 - b_1 L - \cdots - b_q L^q)\varepsilon_t.$$

Since the right side of the equation is an ARMA$(p-d,q)$ process, we say X_t is an ARIMA$(p-d,d,q)$ process. More generally, we say X_t is an $I(d)$ process if $(1-L)^d X_t = \Delta^d X_t$ is weakly stationary.

Random walk process The simplest $I(1)$ model is the random walk process, defined as

$$X_t = X_{t-1} + \varepsilon_t.$$

Given an initial observation X_0, a random walk process can be written as $X_t = X_0 + \varepsilon_1 + \cdots + \varepsilon_t$. Obviously, the variance of X_t is $t\sigma^2$, which increases with t. Therefore, X_t is not stationary. We say X_t has a stochastic trend.

Random walk with drift If there is an intercept term in the random walk process, then

$$X_t = a_0 + X_{t-1} + \varepsilon_t = a_0 \cdot t + \sum_{s=1}^{t} \varepsilon_s.$$

The constant term a_0 here acts as a linear trend term.

Trend stationary process If Y_t is weakly stationary and $X_t = \gamma \cdot t + Y_t$, then we say X_t is trend stationary.

3.2 Unit root test

The most popular unit root test is the Dickey-Fuller test, which estimate the model

$$X_t = a_1 X_{t-1} + \varepsilon_t$$

and test the hypothesis

$$H_0 : a_1 = 1 \qquad \text{vs} \qquad H_1 : a_1 < 1$$

If X_t is an AR(p) process, then we estimate

$$\Delta X_t = \beta X_{t-1} + \sum_{j=1}^{p-1} \phi_j \Delta X_{t-j} + \varepsilon_t$$

and test the hypothesis

$$H_0 : \beta = 0 \qquad \text{vs} \qquad H_1 : \beta < 0$$

We call it the augmented Dickey-Fuller test. Note that the asymptotic distribution of $\hat{\beta}$ and $\hat{a}_1$ are not standard, and so the critical values are obtained by simulation.