Notes for QF603: Quantitative Analysis of Financial Market

Zhou Shen
Singapore Management University

Version: 1.0.0

Update: October 6, 2021

1 Session1 Introduction

1.1 Basic of Probability Theory

1.1.1 Variables

- (Discrete Random Variable) $\mathbb{P}(X = x_i) = p_i, i = 1, 2, \cdots, n$
- (Continuous Random Variable) $\mathbb{P}(r_1 < X < r_2) = p$

1.1.2 Probability Density Functions(PDF)

For the probability p of X lying between r_1 and r_2, we define the probability density function $f(x)$ as follows:

$$\int_{r_1}^{r_2} f(x)dx = p$$

1.1.3 Cumulative Distribution Functions(CDF)

Let f(x) be the CDF. A cumulative distribution function $F(a)$ tells us probability of a random variable X being less than a certain value a:

$$F(a) = \int_{-\infty}^{a} f(x)dx = \mathbb{P}(X \leq a)$$

For CDF, we have following properties:

- $f(x) = \frac{dF(x)}{dx}$
- $\mathbb{P}(a < X \leq b) = \int_a^b f(x)dx = F(b) - F(a)$
- $\mathbb{P}(X > a) = 1 - F(a)$

1.1.4 Inverse Cumulative Distribution Functions

Let $F(a)$ be the cumulative distribution function. We define the inverse function $F^{-1}(p)$, the inverse cumulative distribution, as follows:

$$F^{-1}(p) = a$$

The inverse distribution function is also called the **quantile function**.

The inverse distribution function has following properties:

- $F^{-1}(p)$ is non-decreasing.
- $F^{-1}(y) \leq x$ if and only if $y \leq F(x)$.
- If Y has a uniform distribution in the interval $[0, 1]$, then $F^{-1}(Y)$ is a random variable with distribution F.

Remark *If Y has a uniform distribution in the interval $[0, 1]$, we get $F(Y) = Y$, then $F^{-1}(Y) = Y$.*

1.1.5 Mutually Exclusive Events

For a given random variable, the probability of any of two mutually exclusive events A and B occurring is just the sum of their individual probabilities.

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$

1.1.6 Independent Events and Joint Probability

If the outcome of one random variable is not influenced by the outcome of the other random variable, then we say those variables are independent. The joint probability of A and B is such that

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

1.1.7 Probability Matrices

When dealing with the joint probabilities of two variables, it is often convenient to summarize the various probabilities in a probability matrix or probability table.

Example 1.1

		Stock		
		Outperform	Underperform	Total %
Bonds	Upgrade	15%	5%	20%
	No Change	30%	25%	55%
	Downgrade	5%	20%	25%
	Total %	50%	50%	100%

Figure 1: Stock Grading by Equity Analyst and Credit Rating Agency

1.1.8 Conditional Probability

Probability of A given that B is:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}$$

If $\mathbb{P}(A|B) = \mathbb{P}(A)$, the two random variables A and B, are independent.

1.2 Bayesian Analysis

1.2.1 Definition

For two random variables, A and B, Bayes' theorem states that

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

Remark *To see sample problems, go for QF603session1 pdf from page45 to page68.*

2 Statistics and inference

2.1 Basics of Statistics

2.1.1 Statistical Population

Statistical population is the set of all possible elements that are of interest for a statistical analysis.

Example 2.1 *The time series of split-adjusted daily stock prices of Dell Inc. since IPO on June 22, 1988 till taken private on October 29, 2013.*

2.1.2 Moment Generating Function

Let X be a random variable with pdf $f(x)$. The moment generating function(mgf) of X is:

$$
\begin{aligned}
M_x(\theta) &= \mathbb{E}[e^{\theta x}] \\
&= 1 + \mathbb{E}[x]\theta + \mathbb{E}[x^2]\frac{\theta^2}{2!} + \mathbb{E}[x^3]\frac{\theta^3}{3!} + \cdots
\end{aligned}
$$

$M_x^n(0)$ gives us the nth moment of the distribution of the random variable x:

- $\mu = M_x^1(0)$
- $\sigma^2 = \mathbb{E}[x^2] - (\mathbb{E}[x])^2 = M_x^2(0) - (M_x^1(0))^2$
- Skewness $= \mathbb{E}(x^3) - 3\mathbb{E}(x)\mathbb{E}(x^2) + 2(\mathbb{E}(x))^3$
  $= M_x^3(0) - 3(M_x^1(0))^2 M_x^2(0) + 2(M_x^1(0))^3$
- Kurtosis $= \mathbb{E}(x^4) - 4\mathbb{E}(x)\mathbb{E}(x^3) + 6(\mathbb{E}(x))^2\mathbb{E}(x^2) - 3(\mathbb{E}(x))^4$

Remark *Because typing seperated equations under environment "itemize" will cause some problems, so I just typed skewness as an example of subsitute moment of x with mgf.*

2.1.3 Linear Combination of Variables: Mean and Variance

Let a,b and c be constant. Let X and Y be two random variables, with means μ_X and μ_Y respectively. Also, the corresponding variances are σ_X^2, σ_Y^2, Then

- $\mathbb{E}(aX + bY + c) = a\mathbb{E}(x) + b\mathbb{E}(Y) + c$
- $\mathbb{V}(aX + bY + c) = a^2\mathbb{V}(X) + b^2\mathbb{V}(Y) + 2ab\mathbb{C}(X, Y)$

Where $\mathbb{C}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(X)$.

2.1.4 Population versus Sample

Population parameters (e.g. mean, variance, etc) are usually unobserved. But if we assume that the dataset follows a certain parametric distribution (say "normal"), then we can try to estimate the parameters of that distribution from the data.

Properties of estimators:

- **Bias**: Bias is the difference between the expected value of the estimator and the true (unobserved) parameter value being estimated. A zero bias estimator is an unbiased estimator.
- **Consistency**: Estimators are typically computed over a finite sample size n. Consistency refers to how, as n increases, the estimated value gets closer and closer to the true parameter value.
- **Efficiency**: For an unbiased estimator, efficiency refers to how much its precision is lower than the theoretically highest possible precision. A more efficient estimator needs fewer observations to achieve a given level of variance (precision).

Unbiasedness

A statistic $\Psi(X)$is an unbiased estimator of θ if

$$\mathbb{E}[\Psi(X)] = \theta$$

For population's observations $R_1, R_2, \cdots, R_n$, we have **unbiased estimate of mean and variance** for population as follows:

$$\mu = \frac{1}{n}\sum_{t=1}^{n} R_t$$
$$\sigma^2 = \frac{1}{n-1}\sum_{t=1}^{n}(R_t - \mu)^2$$

Consistency

A sequence of estimators $\theta_n(X)$ of θ from sample X of size n is said to be a consistent estimator if

$$\lim_{x \to +\infty} \mathbb{P}(\theta_n - \theta| < \epsilon) = 1$$

2.2 Statistical Distribution

The statistical Distribution of a variable's population is a description of the relative numbers of times each possible outcome will occur in a number of trials.

2.2.1 Normal Distribution

If $X \overset{d}{\sim} \mathcal{N}(\mu, \sigma^2)$, the PDF $f(x)$ is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2)$$

Mean and Variance:

$$\mathbb{E}(x) = \int_{-\infty}^{\infty} x f(x) dx = \mu$$
$$\mathbb{V}(x) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \sigma^2$$

Let $\mu = 0, \sigma^2 = 1$, we have pdf:

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$$

2.3 Statistical Inference

2.3.1 Law of Large Number(LLN)

Theorem 2.1 *Law of Large Number(LLN) For random sample* $X_1, X_2, \cdots, X_n$, *if n is large enough, the expectation of variance X equals to mean of samples:*

$$\mathbb{E}[X] = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} X_t$$

2.3.2 Central Limit Theorem(CLT)

Theorem 2.2 *Central Limit Theorem(CLT) Suppose $X_1, X_2, \cdots, X_n$ is a sequence of **independent and identically distributed**(i.i.d.) random variables with $\mathbb{E}[X_t] = \mu$ and variance $\mathrm{Var} = \sigma^2 < \infty$. As n approches infinity, the distribution of sum X_t converges to a normal distribution:*

$$\sum_{t=1}^{n} X_t \sim \mathcal{N}(n\mu, n\sigma^2)$$

2.3.3 Confidence Interval

Suppose data $X_1, X_2, \cdots, X_n$ are randomly sampled from $X \sim \mathcal{N}(\mu, \sigma^2)$ such that for an $a > 0$:

$$\mathbb{P}(-a < t < +a) = 95\%$$

Where $\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} = t \sim \mathcal{N}(0, 1)$(CLT), so we know the value of a.

Thus the probability of μ falling within the confidence interval $[\overline{X} - a\frac{\sigma}{\sqrt{n}}, \overline{X} + a\frac{\sigma}{\sqrt{n}}]$ is 95%:

$$\mathbb{P}(\overline{X} - a\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + a\frac{\sigma}{\sqrt{n}}) = 95\%$$

This probability of 5% is known as the **significance level**. The **critical regions** correspond to the significance level's areas.

2.3.4 Type I and Type II Errors

Result of the Test	Reality: H_0 is true	Reality: H_0 is false
Reject H_0	Type I error	Correct inference
Do not reject H_0	Correct inference	Type II error

Figure 2: Type I and Type II Errors

3 Distributions

3.1 Discrete Distributions

3.1.1 Bernoulli Distribution

A Bernoulli random variable X is either zero or one.

$$\mathbb{P}(X = 0) = p, \mathbb{P}(X = 1) = 1 - p$$

PMF and CDF:

$$\mathbb{P}(X = x) = p^x (1 - p)^{1-x}$$

$$F(x) = \begin{cases} 0 & x < 0 \\ p & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

Mean and variance of Bernoulli distribution:

$$\mu = p$$
$$\sigma^2 = p(1 - p)$$

3.1.2 Binomial Distribution

PMF and CDF:

$$\mathbb{P}(n; N, p) = \binom{N}{n} p^n (1-p)^{N-n}$$

$$F(k; N, p) = \mathbb{P}(X \leq k) = \sum_{i=0}^{|k|} \binom{N}{i} p^i (1-p)^{N-i}$$

Mean and variance:

$$\mu = Np$$

$$Var(X) = Np(1-p)$$

3.1.3 Poisson Distribution

$$\mathbb{P}(X = n) = \frac{\lambda^n}{n!} e^{-\lambda}$$

Mean and variance:

$$\mu = \lambda$$

$$Var(X) = \lambda$$

3.2 Continuous Distributions

3.2.1 Normal Distribution

If $X \overset{d}{\sim} \mathcal{N}(\mu, \sigma^2)$, the PDF and CDF:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2)$$

$$F(x) = \frac{1}{2}(1 + \mathrm{erf}(\frac{x-\mu}{\sigma\sqrt{2}}))$$

Where erf denotes error function:

$$\mathrm{erf} z = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$$

Mean and Variance:

$$\mathbb{E}(x) = \int_{-\infty}^{\infty} x f(x) dx = \mu$$
$$\mathrm{Var}(x) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \sigma^2$$

3.2.2 Lognormal Distribution

If $Y \sim \mathcal{N}(\mu, \sigma)$, we define $X = e^Y$, then we say X follows Lognormal Distribution.

PDF:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp(-\frac{(\ln x - \mu)^2}{2\sigma^2})$$

CDF:

$$F(x) = \frac{1}{2}(1 + \mathrm{erf}(\frac{\ln x - \mu}{\sigma\sqrt{2}}))$$

Mean and Variance:

$$\mathbb{E}(x) = \exp(\mu + \frac{\sigma^2}{2})$$
$$\mathrm{Var}(x) = (\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)$$

3.2.3 Chi-Squared Distribution

Suppose we have k independent normal variables $z_1, \cdots, z_k$, let $S = \sum_{i=1}^{k} {z_i}^2$, then we say S follows Chi-Squared Distribution:

$$S \sim \chi_k^2$$

PDF:

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}$$

Where Γ function is defined:

$$\Gamma(n) = \int_0^{\infty} x^{n-1} e^{-x} dx$$

Mean and Variance:

$$\mathbb{E}(x) = k$$
$$\mathrm{Var}(x) = 2k$$

3.2.4 Student's t Distribution

Let $Z \sim \mathcal{N}(0, 1)$, $U \sim \chi^2_k$, we define:

$$X = \frac{Z}{\sqrt{U/k}}$$

Then X follows Student's t distribution:

$$X \sim t(K)$$

PDF:

$$f(x) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\Gamma(\frac{k}{2})}(1 + \frac{x^2}{k})^{-\frac{k+1}{2}}$$

Mean and Variance:

$$\mathbb{E}(x) = 0$$
$$\mathrm{Var}(x) = \frac{k}{k-2}$$

Note that t-distributed random variable will be the test statistic of sample mean esitimates.

3.2.5 F-Distribution

If $U_1 \sim \chi^2_{k_1}$, $U_2 \sim \chi^2_{k_2}$, U_1 and U_2 are independent, we define

$$X = \frac{U_1/k_1}{U_2/k_2}$$

Then we say X follows F-Distribution:

$$X \sim \mathrm{F}(k_1, k_2)$$

PDF:

$$f(x) = \frac{\sqrt{\frac{(k_1x)^{k_1}k_2^{k_2}}{(k_1x+k_2)^{k_1+k_2}}}}{x\mathrm{B}(\frac{k_1}{2}, \frac{k_2}{2})}$$

Where

$$\mathrm{B}(x, y) = \int_0^1 z^{x-1}(1-z)^{y-1}dz$$

Mean and variance:

$$\mu = \frac{k_2}{k_2 - 2}$$
$$\sigma^2 = \frac{2k_2^2(k_1 + k_2 - 2)}{k_1(k_2 - 2)^2(k_2 - 4)}$$

F-distribution has properties:

- When k_1, k_2 aprroximate infinity, F-distribution equals to standard normal distribution.
- If $X \sim t(k)$, then $X^2 \sim \mathrm{F}(1, k)$.

4 Cross Sectional Estimation Frameworks

4.1 Ordinary Least Squares(OLS)

4.1.1 Simple OLS

Model 0 is $Y_i = a + e_i$.

But given n pairs of observations on explanatory variable X_i and dependent variable Y_i, we can have Model 1 by postulating that

$$Y_i = a + bX_i + e_i, \ i = 1, 2, \cdots, n,$$

where e_i is the noise.

Assumptions:

1. $\mathbb{E}(e_i) = 0$ for every i
2. $\mathbb{E}(e_i^2) = \sigma_e^2$
3. $\mathbb{E}(e_i, e_j) = 0$ for every i, j
4. X_i, e_j are independednt for each i, j
5. $e_i \sim \mathcal{N}(0, \sigma_e^2)$

Least Sqaures: Minimizing the sum of squared errors:

$$\min_{\hat{a},\hat{b}} \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{a} - \hat{b}X_i)^2$$

Set the first derivative of a and b as 0, we get:

$$\begin{aligned}
\hat{a} &= \bar{Y} - \hat{b}\bar{X} \\
\hat{b} &= \frac{\sum_{i=1}^{n} X_i(Y_i - \bar{Y})}{\sum_{i=1}^{n} X_i(X_i - \bar{X})} \\
&= \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})} \\
&= \frac{\mathbb{C}(Y, X)}{\mathbb{V}(X)}
\end{aligned}$$

Properties of $\hat{a}, \hat{b}$:

$$\begin{aligned}
\hat{a} &\sim \mathcal{N}\left(a, \sigma_e^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\right)\right) \\
\hat{b} &\sim \mathcal{N}\left(b, \sigma_e^2\left(\frac{1}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\right)\right)
\end{aligned}$$

Note *Because $\mathbb{E}(\hat{a}) = a$,$\mathbb{E}(\hat{b}) = b$, the $\hat{a}, \hat{b}$ are unbiased. Then Gauss-Markov Theorem states $\hat{a}$ and $\hat{b}$ has smallest variances, so $\hat{a}$ and $\hat{b}$ are efficient.*

4.1.2 Gauss-Markov Theorem

Theorem 4.1 *Gauss-Markov Theorem states that among all linear and unbiased estimators, the OLS estimators $\hat{a}$ andb $\hat{b}$ have the minimum variances.*

4.1.3 OLS in Matrix

Write model as $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$, by calculation, we get:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

4.1.4 Hypothesis Testing

Series of residuals:

$$\hat{e}_i = Y_i - \hat{a} - \hat{b}X_i,\ i = 1, 2, \cdots, n$$

Testing null hypothesis $H_0 : b = \beta(e.g.\ \beta = 0)$:

$$t_{n-2} = \frac{\hat{b} - \beta}{\hat{\sigma^e}\sqrt{\frac{1}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}}$$

Testing null hypothesis $H_0 : a = \alpha(e.g.\ \alpha = 0)$:

$$t_{n-2} = \frac{\hat{a} - \alpha}{\hat{\sigma^e}\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}}$$

Note *The denominators in testings are* ***standard errors*** *of $\hat{b}$ and $\hat{a}$.*

4.1.5 Consistent Properties of OLS

1. OLS $\hat{b}$ estimator is consistent:

$$\lim_{n\to\infty} \hat{b} = b$$

2. OLS $\hat{a}$ estimator is consistent:

$$\lim_{n\to\infty} \hat{a} = a$$

4.1.6 Decomposition

- TSS: Total Sum of Squares

- ESS: Explained Sum of Squares
- RSS: Residual Sum of Squares

$$\text{TSS} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$
$$\text{ESS} = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$
$$\text{RSS} = \sum_{i=1}^{n} \hat{e}_i{}^2 = \sum_{i=1}^{n} (Y_i - \hat{Y})^2$$
$$\text{TSS} = \text{ESS} + \text{RSS}$$
$$R^2 := \frac{ESS}{TSS}$$

4.1.7 Point Forecast and Confidence Interval

The point forecast is

$$\widehat{Y_{n+1}} = \hat{a} + \hat{b} X_{n+1}$$

With 95% probability, the forecast value falls within the confidence interval bounded by

$$\hat{Y}_{n+1} \pm t_{n-2,97.5\%} \times \hat{\sigma_e} \sqrt{1 + \frac{1}{n} + \frac{(X_{n+1} - \bar{X})^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}}$$

4.2 Discrete dependent variables

- Probit/logit model: used for binary dependent variables
- Multinomial probit/logit: Categorical dependent variables
- Ordered probit/logit: Discrete dependent variables

Probit:

$$\phi(X) = \mathbb{P}(Z \leq X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{X} \exp(-\frac{u^2}{2}) du$$

Logit:

$$\phi(X) = \mathbb{P}(Z \leq X) = \frac{1}{1 + \exp(-k(X - X_0))}$$

Nature of Dependent Variable	Methodology
Continuous	OLS
Discrete 1/0	Probit/Logit
Multivalued Discrete, with a value order [i.e. lower numbers are 'better' or worse']	Ordered Probit/Logit
Categorical labels	Multinomial Probit/Logit

Figure 3: Methodology roadmap

5 Cross Sectional Estimation Techniques

5.1 Some notes

```
result = sm.OLS(data_w_dummies['lnepratio'], sm.add_constant(data_w_dummies[['lnoperatingmargin', 'Agriculture Forestry And Fishing', 'Construction','Finance Insurance And Real Estate',
'Manufacturing', 'Mining','Retail Trade', 'Services','Transportation Communications Electric Gas And Sanitary Service']]), missing='drop').fit(cov_type='cluster', cov_kwds={'groups':
data_w_dummies['siccode']})
result.summary()
```

To avoid cluster effect on estimation, we set cov_type as 'cluster'

```
industrydummies = pd.get_dummies(df_left['sicsector'])
```

```
data_w_dummies['lnoperatingmargin'] = np.log(data_w_dummies['operatingmargin'])
result = sm.OLS(data_w_dummies['epratio'], sm.add_constant(data_w_dummies[['lnoperatingmargin']]), missing='drop').fit()
result.summary()
When r is very low, do log
```

```
#use logit instead
result = smdiscrete.Logit(data_w_dummies['paydividend'], sm.add_constant(data_w_dummies[['lnoperatingmargin', 'Agriculture Forestry And Fishing', 'Construction','Finance Insurance And Real Estate', 'Manufacturing', 'Mining','Retail Trade', 'Services','Transportation Communications Electric Gas And Sanitary Service']]), missing='drop').fit(cov_type='cluster', cov_kwds={'groups': data_w_dummies['siccode']})
result.summary()          Use logit to do estimation
```

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
features = scaler.fit_transform(features) # Standardization
from sklearn.decomposition import PCA
pca = PCA(n_components = 4)
principal_components = pca.fit_transform(features)
principal_components
pca.explained_variance_ratio_
```

Remark *Some key words: Fixed effects: dummy variables; t-statistic low: log; intra-cluster correlations: clustering; Collinearity: PCA; adding squared term; Hazard model; survival analysis*

6 CAPM (Cross Sectional Application), and Tests

6.1 Capital Market Line(CML)

$$\text{Sharpe Ratio} = \frac{\mathbb{E}(\mathrm{r_i} - \mathrm{r_f})}{\sigma_\mathrm{i}}$$

6.2 Security Market Line(SML)

$$r_i - r_f = \beta_i(r_m - r_f)$$

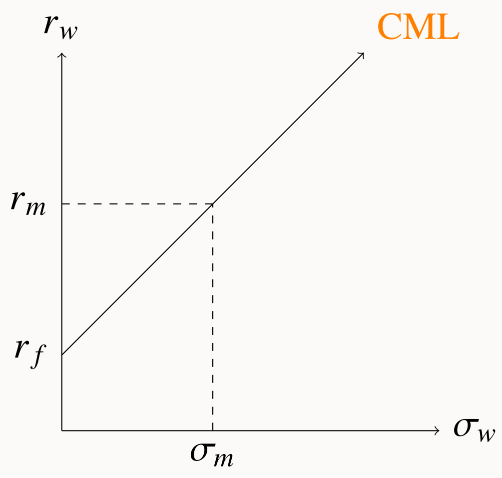

Figure 4: CML

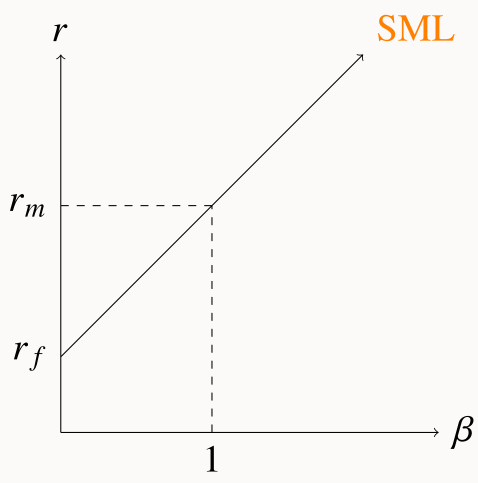

Figure 5: SML

$$\text{Treynor Ratio} = \frac{\mathrm{r_i} - \mathrm{r_f}}{\beta_\mathrm{i}} = \frac{r_m - r_f}{1}$$

6.3 Market Model

Market model assumes that any stock's log return r_{it} is bivariate normally distriuted with r_{mt}.

A linear regression model of r_{it} on r_{mt} is

$$r_{it} = a + br_{mt} + e_{it}$$

where

$$a = \mathbb{E}(r_{it}) - \frac{\sigma_{im}}{\sigma_m^2}\mathbb{E}(r_{mt})$$
$$b = \frac{\sigma_{im}}{\sigma_m^2}$$

Remark *In QF600, the we do regression on simple returns, not log returns. I do not know if both methods work fine or only exists one right version.*

6.4 Capital Asset Pricing Model(CAPM)

CAPM:

$$\mathbb{E}(r_{it} - r_{ft}) = b_i\mathbb{E}(r_{mt} - r_{ft})$$

Do regression:

$$r_{it} - r_{ft} = a_i + b_i(r_{mt} - r_{ft}) + e_{it}$$

OLS estimation of beta:

$$\hat{b}_i = \frac{\sum_{t=1}^{T}(r_{mt} - \bar{r_m})(r_{it} - \bar{r_i})}{\sum_{t=1}^{T}(r_{mt} - \bar{r_m})^2}$$

6.4.1 Some measurements

- standard error of a:

$$\hat{\sigma_e}\sqrt{\frac{1}{T} + \frac{\bar{X}^2}{\sum_{t=1}^{T}(X_t - \bar{X})^2}}$$

- standard error of b:

$$\hat{\sigma_e}\sqrt{\frac{1}{\sum_{t=1}^{T}(X_t - \bar{X})^2}}$$

- sum squared resid/residual Sum of Squares(SSR/RSS):

$$\text{SSR/RSS} = \sum_{t=1}^{T}\hat{e_t}^2$$

- standard error of regression:

$$\sigma_e = \sqrt{\frac{1}{T-2}\text{SSR}}$$

- Treynor ratio:

$$\text{Treynor ratio} := r_{mt} - r_{ft} = \frac{r_{it} - r_{ft}}{\beta_i}$$

- Jensen's measure:

$$\text{Jensen's measure} := \mathbb{E}(r_{it} - r_{ft}) - b_i\mathbb{E}(r_{mt} - r_{ft})$$

- Sharpe ratio:

$$\text{Shapre ratio} := \frac{\mathbb{E}(r_{it} - r_{ft})}{\sigma_i}$$

- $M^2 measurement$:

$$M^2 := \mathbb{E}(r_{it} - r_{ft})\frac{\sigma_m}{\sigma_i} - \mathbb{E}(r_{mt} - r_{ft})$$

- Information ratio:

$$\text{Information ratio} := \frac{\mathbb{E}(r_{it} - r_{ft})}{\sigma_{i-f}}$$

- active return:

$$\text{active return} := r_{it} - r_{ft}$$

- tracking error/active risk:

$$\text{tracking error} := \sigma_{i-f}$$