Introduction to Financial Econometrics
Chapter 3 The Constant Expected Return Model

Eric Zivot
Department of Economics
University of Washington

January 6, 2000
This version: February 19, 2000

1 The Constant Expected Return Model of Asset Returns

Let R_{it} denote the continuously compounded return on an asset i at time t. We make the following assumptions regarding the probability distribution of R_{it} for $i = 1, \ldots, N$ assets over the time horizon $t = 1, \ldots, T$.

1. Normality of returns: $R_{it} \sim N(\mu_i, \sigma_i^2)$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.
2. Constant variances and covariances: $cov(R_{it}, R_{jt}) = \sigma_{ij}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.
3. Independence over time: R_{it} is independent of R_{js} for $t \neq s$ and $i, j = 1, \ldots, N$.

Assumption 1 states that in every time period asset returns are normally distributed and that the mean and the variance of each asset return is constant over time. In particular, we have for each asset i

$$\begin{aligned} E[R_{it}] &= \mu_i \text{ for all values of } t \\ var(R_{it}) &= \sigma_i^2 \text{ for all values of } t \end{aligned}$$

The second assumption states that the contemporaneous covariances between assets are constant over time. Given assumption 1, assumption 2 implies that the contemporaneous correlations between assets are constant over time as well. That is, for all assets

$$corr(R_{it}, R_{jt}) = \rho_{ij} \text{ for all values of } t.$$

The third assumption stipulates that all of the asset returns are independent over time. In particular, for a given asset i the returns on the asset are *serially independent* which implies that

$$cov(R_{it}, R_{is}) = 0 \text{ for all } t \neq s.$$

Additionally, the returns on all possible pairs of assets i and j are serially independent which implies that

$$cov(R_{it}, R_{js}) = 0 \text{ for all } i \neq j \text{ and } t \neq s.$$

Assumptions 1-3 indicate that all asset returns at a given point in time are jointly (multivariate) normally distributed and that this joint distribution stays constant over time. Clearly these are very strong assumptions. However, they allow us to development a straightforward probabilistic model for asset returns as well as statistical tools for estimating the parameters of the model and testing hypotheses about the parameter values and assumptions.

1.0.1 Regression Model Representation

A convenient mathematical representation or model of asset returns can be given based on assumptions 1-3. This is the constant expected return (CER) regression model. For assets $i = 1, \ldots, N$ and time periods $t = 1, \ldots, T$ we have

$$\begin{aligned} R_{it} &= \mu_i + \varepsilon_{it} & (1) \\ \varepsilon_{it} &\sim i.i.d.\ N(0, \sigma_i^2) \\ cov(\varepsilon_{it}, \varepsilon_{jt}) &= \sigma_{ij} & (2) \end{aligned}$$

where μ_i is a constant and we assume that ε_{it} is independent of ε_{js} for all time periods $t \neq s$. The notation $\varepsilon_{it} \sim i.i.d.\ N(0, \sigma_i^2)$ stipulates that the random variable ε_{it} is serially independent and identically distributed as a normal random variable with mean zero and variance σ_i^2. In particular, note that, $E[\varepsilon_{it}] = 0$, $var(\varepsilon_{it}) = \sigma_i^2$ and $cov(\varepsilon_{it}, \varepsilon_{js}) = 0$ for $t \neq s$.

Using the basic properties of expectation, variance and covariance discussed in chapter 2, we can derive the following properties of returns. For expected returns we have that,

$$E[R_{it}] = E[\mu_i + \varepsilon_{it}] = \mu_i + E[\varepsilon_{it}] = \mu_i$$

since μ_i is constant and $E[\varepsilon_{it}] = 0$. Regarding the variance of returns, we have that

$$var(R_{it}) = var(\mu_i + \varepsilon_{it}) = var(\varepsilon_{it}) = \sigma_i^2$$

which uses the fact that the variance of a constant is zero. For covariances of returns, we have

$$cov(R_{it}, R_{jt}) = cov(\mu_i + \varepsilon_{it}, \mu_j + \varepsilon_{jt}) = cov(\varepsilon_{it}, \varepsilon_{jt}) = \sigma_{ij}$$

and

$$cov(R_{it}, R_{js}) = cov(\mu_i + \varepsilon_{it}, \mu_j + \varepsilon_{js}) = cov(\varepsilon_{it}, \varepsilon_{js}) = 0,\ t \neq s,$$

which use the fact that adding constants to two random variables does not affect the covariance between them. Given that covariances and variances of returns are constant over time gives the result that correlations between returns over time are also constant:

$$\begin{aligned} corr(R_{it}, R_{jt}) &= \frac{cov(R_{it}, R_{jt})}{\sqrt{var(R_{it})var(R_{jt})}} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} = \rho_{ij}, \\ corr(R_{it}, R_{js}) &= \frac{cov(R_{it}, R_{js})}{\sqrt{var(R_{it})var(R_{js})}} = \frac{0}{\sigma_i \sigma_j} = 0,\ t \neq s. \end{aligned}$$

Finally, since the random variable ε_{it} is independent and identically distributed $(i.i.d.)$ normal the asset return R_{it} will also be $i.i.d.$ normal:

$$R_{it} \sim i.i.d.\ N(\mu_i, \sigma_i^2).$$

The CER model has a very simple form and is identical to the measurement error model in the statistics literature. In words, the model states that each asset return is equal to a constant μ_i (the expected return) plus a normally distributed random variable ε_{it} with mean zero and constant variance. The random variable ε_{it} can be interpreted as representing the unexpected news concerning the value of the asset that arrives at time t. To see this, note that using (1) we can write ε_{it} as

$$\begin{aligned} \varepsilon_{it} &= R_{it} - \mu_i \\ &= R_{it} - E[R_{it}] \end{aligned}$$

so that ε_{it} is defined to be the deviation of the random return from its expected value. If the news is good, then the realized value of ε_{it} is positive and the observed return is above its expected value μ_i. If the news is bad, then ε_{jt} is negative and the observed return is less than expected. The assumption that $E[\varepsilon_{it}] = 0$ means that news, on average, is neutral; neither good nor bad. The assumption that $var(\varepsilon_{it}) = \sigma_i^2$ can be interpreted as saying that volatility of news arrival is constant over time. The random news variable affecting asset i, e_{it}, is allowed to be contemporaneously correlated with the random news variable affecting asset j, ε_{jt}, to capture the idea that news about one asset may spill over and affect another asset. For example, let asset i be Microsoft and asset j be Apple computer. Then one interpretation of news in this context is general news about the computer industry and technology. Good news should lead to positive values of ε_{it} and ε_{jt}. Hence these variables will be positively correlated.

The CER model with continuously compounded returns has the following nice property with respect to the interpretation of ε_{it} as news. Consider the default case where R_{it} is interpreted as the continuously compounded monthly return. Since multiperiod continuously compounded returns are additive we can interpret, for example, R_{it} as the sum of 30 daily continuously compounded returns[1]:

$$R_{it} = \sum_{k=0}^{29} R^d_{it-k}$$

where R^d_{it} denotes the continuously compounded daily return on asset i. If we assume that daily returns are described by the CER model then

$$\begin{aligned} R^d_{it} &= \mu^d_i + \varepsilon^d_{it}, \\ \varepsilon^d_{it} &\sim i.i.d\ N(0, (\sigma^d_i)^2), \\ cov(\varepsilon^d_{it}, \varepsilon^d_{jt}) &= \sigma^d_{ij}, \end{aligned}$$

and the monthly return may then be expressed as

$$\begin{aligned} R_{it} &= \sum_{k=0}^{29} (\mu^d_i + \varepsilon^d_{it-k}) \\ &= 30 \cdot \mu^d_i + \sum_{k=0}^{29} \varepsilon^d_{it-k} \\ &= \mu_i + \varepsilon_{it}, \end{aligned}$$

where

$$\begin{aligned} \mu_i &= 30 \cdot \mu^d_i, \\ \varepsilon_{it} &= \sum_{k=0}^{29} \varepsilon_{it-k}. \end{aligned}$$

Hence, the monthly expected return, μ_i, is simply 30 times the daily expected return. The interpretation of ε_{it} in the CER model when returns are continuously compounded is the accumulation of news over the month.

2 Estimating the CER Model

2.1 The Random Sampling Environment

The CER model of asset returns gives us a rigorous way of interpreting the time series behavior of asset returns. At the beginning of every month t, R_{it} is a random

[1]For simplicity of exposition, we will ignore the fact that some assets do not trade over the weekend.

variable representing the return to be realized at the end of the month. The CER model states that $R_{it} \sim i.i.d.\ N(\mu_i, \sigma_i^2)$. Our best guess for the return at the end of the month is $E[R_{it}] = \mu_i$, our measure of uncertainty about our best guess is captured by $\sigma_i = \sqrt{var(R_{it})}$ and our measure of the direction of linear association between R_{it} and R_{jt} is $\sigma_{ij} = cov(R_{it}, R_{jt})$. The CER model assumes that the economic environment is constant over time so that the normal distribution characterizing monthly returns is the same every month.

Our life would be very easy if we knew the exact values of μ_i, σ_i^2 and σ_{ij}, the parameters of the CER model. In actuality, however, we do not know these values with certainty. A key task in financial econometrics is estimating the values of μ_i, σ_i^2 and σ_{ij} from a history of observed data.

Suppose we observe monthly returns on N different assets over the horizon $t = 1, \ldots, T$. Let $r_{i1}, \ldots, r_{iT}$ denote the observed history of T monthly returns on asset i for $i = 1, \ldots, N$. It is assumed that the observed returns are realizations of the random variables $R_{i1}, \ldots, R_{iT}$, where R_{it} is described by the CER model (1). We call $R_{i1}, \ldots, R_{iT}$ a random sample from the CER model (1). In this case, we can use the observed returns to estimate the unknown parameters of the CER model

2.2 Estimation Theory

Before we describe the estimation of the CER model, it is useful to summarize some concepts in estimation theory. Let θ denote some characteristic of the CER model (1) we are interested in estimating. For example, if we are interested in the expected return then $\theta = \mu_i$; if we are interested in the variance of returns then $\theta = \sigma_i^2$. The goal is to estimate θ based on the observed data $r_{i1}, \ldots, r_{iT}$.

Definition 1 *An estimator of θ is a rule or algorithm for forming an estimate for θ.*

Definition 2 *An estimate of θ is simply the value of an estimator based on the observed data.*

To establish some notation, let $\hat{\theta}(R_{i1}, \ldots, R_{iT})$ denote an estimator of θ treated as a function of the random variables $R_{i1}, \ldots, R_{iT}$. Clearly, $\hat{\theta}(R_{i1}, \ldots, R_{iT})$ is a random variable. Let $\hat{\theta}(r_{i1}, \ldots, r_{iT})$ denote an estimate of θ based on the realized values $r_{i1}, \ldots, r_{iT}$. $\hat{\theta}(r_{i1}, \ldots, r_{iT})$ is simply an number. We will often use $\hat{\theta}$ as shorthand notation to represent either an estimator of θ or an estimate of θ. The context will determine how to interpret $\hat{\theta}$.

2.2.1 Properties of Estimators

Consider $\hat{\theta} = \hat{\theta}(R_{i1}, \ldots, R_{iT})$ as a random variable. In general, the p.d.f. of $\hat{\theta}$, $p(\hat{\theta})$, depends on the p.d.fs of the random variables $R_{i1}, \ldots, R_{iT}$. The exact form of $p(\hat{\theta})$

may be very complicated. For analysis purposes, we often focus on certain characteristics of $p(\hat{\theta})$. Four important properties of estimators are (1) *bias*; (2) *precision*; (3) *efficiency*; and (4) *consistency*.

Bias concerns the location or center of $p(\hat{\theta})$. If $p(\hat{\theta})$ is centered away from θ then we say $\hat{\theta}$ is *biased*. If $p(\hat{\theta})$ is centered at θ then we say that $\hat{\theta}$ is *unbiased*. Formally we have the following definitions:

Definition 3 *The bias of an estimator $\hat{\theta}$ of θ is given by*

$$bias(\hat{\theta}, \theta) = E[\hat{\theta}] - \theta.$$

Definition 4 *An estimator $\hat{\theta}$ of θ is unbiased if $bias(\hat{\theta}, \theta) = 0$; i.e., $E[\hat{\theta}] = \theta$.*

Precision - to be completed
Efficiency - to be completed
Consistency - to be completed

2.3 Some Special Probability Distributions Used in Statistical Inference

2.3.1 The Chi-Square distribution with T degrees of freedom

Let $Z_1, Z_2, \ldots, Z_T$ be independent standard normal random variables. That is,

$$Z_i \sim i.i.d.\ N(0,1),\ i = 1, \ldots, T.$$

Define a new random variable X such that

$$X = Z_1^2 + Z_2^2 + \cdots Z_T^2 = \sum_{i=1}^{T} Z_i^2.$$

Then X is a chi-square random variable with T degrees of freedom. Such a random variable is often denoted χ_T^2 and we use the notation $X \sim \chi_T^2$. The p.d.f. of X is illustrated in Figure xxx for various values of T. Notice that X is only allowed to take non-negative values. The p.d.f. is highly right skewed for small values of T and becomes symmetric as T gets large. Furthermore, it can be shown that

$$E[X] = T.$$

The chi-square distribution is used often in statistical inference and probabilities associated with chi-square random variables are needed. Critical values, which are just quantiles of the chi-square distribution, are used in typical calculations. To illustrate, suppose we wish to find the critical value of the chi-square distribution with T degrees of freedom such that the probability to the right of the critical value is α. Let $\chi^2_T(\alpha)$ denote this critical value[2]. Then

$$\Pr(X > \chi^2_T(\alpha)) = \alpha.$$

For example, if $T = 5$ and $\alpha = 0.05$ then $\chi^2_5(0.05) = 11.07$; if $T = 100$ then $\chi^2_{100}(0.05) = 124.34$.

2.3.2 Student's t distribution with T degrees of freedom

Let Z be a standard normal random variable, $Z \sim N(0,1)$, and let X be a chi-square random variable with T degrees of freedom, $X \sim \chi^2_T$. Assume that Z and X are independent. Define a new random variable t such that

$$t = \frac{Z}{\sqrt{X/T}}.$$

Then t is a Student's t random variable with T degrees of freedom and we use the notation $t \sim t_T$ to indicate that t is distributed Student-t. Figure xxx shows the p.d.f of t for various values of the degrees of freedom T. Notice that the p.d.f. is symmetric about zero and has a bell shape like the normal. The tail thickness of the p.d.f. is determined by the degrees of freedom. For small values of T, the tails are quite spread out and are thicker than the tails of the normal. As T gets large the tails shrink and become close to the normal. In fact, as $T \to \infty$ the p.d.f. of the Student t converges to the p.d.f. of the normal.

The Student-t distribution is used heavily in statistical inference and critical values from the distribution are often needed. Let $t_T(\alpha)$ denote the critical value such that

$$\Pr(t > t_T(\alpha)) = \alpha.$$

For example, if $T = 10$ and $\alpha = 0.025$ then $t_{10}(0.025) = 2.228$; if $T = 100$ then $t_{60}(0.025) = 2.00$. Since the Student-t distribution is symmetric about zero, we have that

$$\Pr(-t_T(\alpha) \leq t \leq t_T(\alpha)) = 1 - 2\alpha.$$

For example, if $T = 60$ and $\alpha = 2$ then $t_{60}(0.025) = 2$ and

$$\Pr(-t_{60}(0.025) \leq t \leq t_{60}(0.025)) = \Pr(-2 \leq t \leq 2) = 1 - 2(0.025) = 0.95.$$

2.4 Least Squares Estimation of the CER Model

Let $R_{i1}, \ldots, R_{iT}$ denote a random sample from the CER model and let $r_{i1}, \ldots, r_{iT}$ denote the realized values from the random sample. Consider the problem of estimating the parameter μ_i in the CER model (1). As an example, consider the observed

[2]Excel has functions for computing probabilities from the chi-square distribution.

monthly continuously compounded returns, $r_{i1}, \ldots, r_{iT}$, for Microsoft stock over the period January 1987 through May 1997. These data are illustrated in Figure 1.

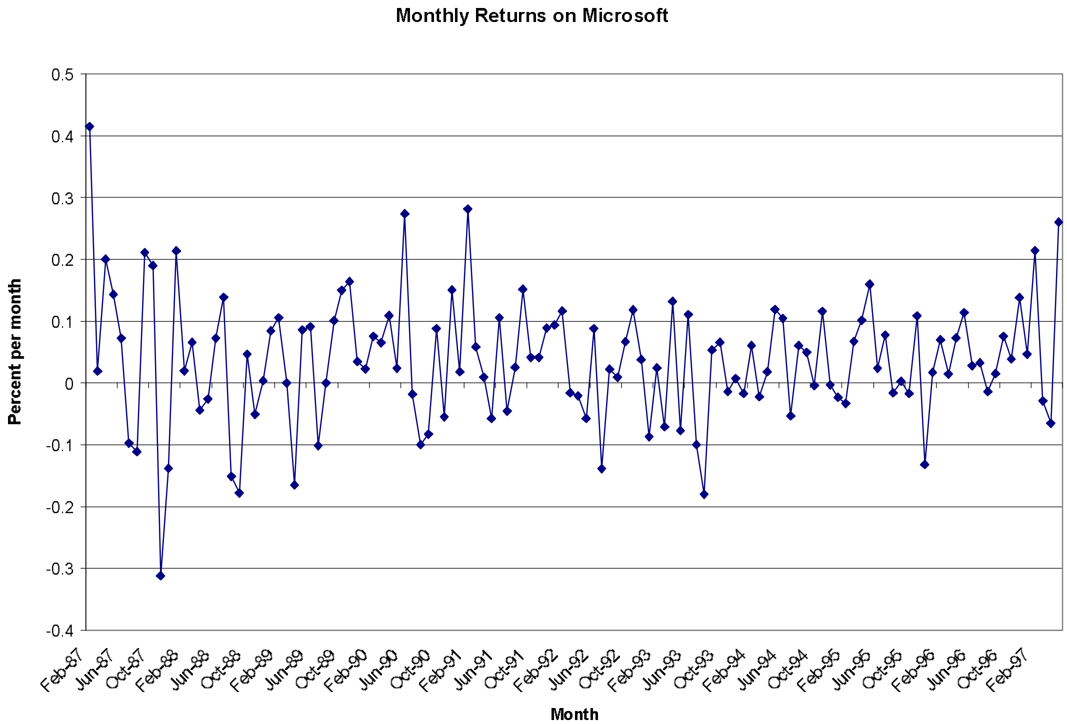

Figure 1

Notice that the data seem to fluctuate up and down about some central value near 0.05. Intuitively, the parameter μ_i represents this central value. Let $\hat{\mu}_i$ denote a prospective estimate of μ_i. The *prediction error* or *residual* at time t associated with this estimate is defined as

$$\hat{\varepsilon}_{it} = r_{it} - \hat{\mu}_i, \;\; t = 1, \ldots, T,$$

and the squared prediction error (squared residual) is

$$\hat{\varepsilon}_{it}^2 = (r_{it} - \hat{\mu}_i)^2, \;\; t = 1, \ldots, T.$$

The method of *least squares* determines the estimate $\hat{\mu}_i$ that minimizes the sum of squared prediction errors (residuals). Formally, to determine the least squares estimate of μ_i we solve

$$\min_{\hat{\mu}_i} RSS(\hat{\mu}_i) = \sum_{t=1}^{T} \hat{\varepsilon}_{it}^2 = \sum_{t=1}^{T} (r_{it} - \hat{\mu}_i)^2,$$

where $RSS(\hat{\mu}_i)$ denotes the residual sum of squares associated with the estimate $\hat{\mu}_i$. Using simple calculus, the first order conditions for a minimum satisfy

$$0 = \frac{dRSS(\hat{\mu}_i)}{d\widehat{\mu}_i} = \frac{d}{d\widehat{\mu}_i}\sum_{t=1}^{T}(r_{it} - \hat{\mu}_i)^2 = -2\sum_{t=1}^{T}(r_{it} - \hat{\mu}_i) = -2\sum_{t=1}^{T} r_t + 2T\widehat{\mu}_i.$$

Solving for $\hat{\mu}_i$ gives the least squares estimate

$$\hat{\mu}_i = \frac{1}{T}\sum_{t=1}^{T} r_{it}. \quad (3)$$

Hence, the least squares estimate of μ_i $(i = 1, \ldots, N)$ in the CER model is simply the *sample average* of the observed returns for asset i.

2.4.1 The least squares estimates of σ_i and σ_{ij}

Given the least squares estimate of μ_i $(i = 1, \ldots, N)$, $\hat{\mu}_i$, the least square estimates of σ_i^2 and σ_{ij} are defined to be

$$\hat{\sigma}_i^2 = \frac{1}{T-1}\sum_{t=1}^{T}(r_{it} - \hat{\mu}_i)^2 \quad (4)$$

and

$$\hat{\sigma}_{ij} = \frac{1}{T-1}\sum_{t=1}^{T}(r_{it} - \hat{\mu}_i)(r_{jt} - \hat{\mu}_j). \quad (5)$$

Notice that (4) is simply the *sample variance* of the observed returns for asset i and (5) is the *sample covariance* of the observed returns on assets i and j.

2.4.2 Statistical properties of the least squares estimator of μ_i

To determine the statistical properties of $\hat{\mu}_i$ in the CER model we treat it as a function of the random sample $R_{i1}, \ldots, R_{iT}$:

$$\hat{\mu}_i = \hat{\mu}_i(R_{i1}, \ldots, R_{iT}) = \frac{1}{T}\sum_{t=1}^{T} R_{it} \quad (6)$$

where R_{it} is assumed to be generated by the CER model (1).

In the CER model, the random variables R_{it} $(t = 1, \ldots, T)$ are normally distributed. Since the least squares estimator (6) is an average of these normal random variables it is also normally distributed. That is, $p(\hat{\mu}_i)$ is a normal density. To determine the mean of this distribution we compute

$$E[\hat{\mu}_i] = E\left[\frac{1}{T}\sum_{t=1}^{T} R_{it}\right]$$

$$
\begin{aligned}
&= E\left[\frac{1}{T}\sum_{t=1}^{T}(\mu_i + \varepsilon_{it})\right] \quad \text{(in the CER model } R_{it} = \mu_i + \varepsilon_{it}\text{)} \\
&= \frac{1}{T}\sum_{t=1}^{T}\mu_i + \frac{1}{T}\sum_{t=1}^{T}E[\varepsilon_{it}] \quad \text{(by the linearity of } E[\cdot]\text{)} \\
&= \frac{1}{T}\sum_{t=1}^{T}\mu_i \text{ (since } E[\varepsilon_{it}] = 0,\ t = 1, \ldots, T\text{)} \\
&= \frac{1}{T}T\mu_i \text{ (since } \mu_i \text{ is a constant)} \\
&= \mu_i.
\end{aligned}
$$

Hence $\hat{\mu}_i$ is an unbiased estimator for μ_i.

To determine the variance of $\hat{\mu}_i$ we compute

$$
\begin{aligned}
var(\hat{\mu}_i) &= var\left(\frac{1}{T}\sum_{t=1}^{T}R_{it}\right) \\
&= var\left(\frac{1}{T}\sum_{t=1}^{T}(\mu_i + \varepsilon_{it})\right) \quad \text{(in the CER model } R_{it} = \mu_i + \varepsilon_{it}\text{)} \\
&= var\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\right) \quad \text{(since } \mu_i \text{ is a constant)} \\
&= \frac{1}{T^2}\sum_{t=1}^{T}var(\varepsilon_{it}) \quad \text{(since } \varepsilon_{it} \text{ are independent over time)} \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sigma_i^2 \quad \text{(since } var(\varepsilon_{it}) = \sigma_i^2,\ t = 1, \ldots, T\text{)} \\
&= \frac{1}{T^2}T\sigma_i^2 \\
&= \frac{\sigma_i^2}{T}.
\end{aligned}
$$

Notice that the variance of $\hat{\mu}_i$ is equal to the variance of R_{it} divided by the sample size and is therefore much smaller than the variance of R_{it}.

Using the above results we deduce that

$$
\hat{\mu}_i \sim N\left(\mu_i, \frac{\sigma_i^2}{T}\right). \tag{7}
$$

This distribution is illustrated in Figure xxx. Notice that the variance of $\hat{\mu}_i$ is inversely related to the sample size T. Given σ_i^2, $var(\hat{\mu}_i)$ is smaller for larger sample sizes than for smaller sample sizes. This makes sense since we expect to have a more precise estimator when we have more data. If the sample size is very large (as $T \to \infty$) then $var(\hat{\mu}_i)$ will be approximately zero and the normal distribution of $\hat{\mu}_i$ given by (7) will be essentially a spike at μ_i. In other words, if the sample size is very large then we essentially know the true value of μ_i. Hence, $\hat{\mu}_i$ is a *consistent* estimator of μ_i.

The standard deviation of $\hat{\mu}_i$ is referred to as the *standard error* of μ_i and is given by

$$SE(\hat{\mu}_i) = \sqrt{var(\hat{\mu}_i)} = \frac{\sigma_i}{\sqrt{T}}. \tag{8}$$

$SE(\hat{\mu}_i)$ is in the same units as $\hat{\mu}_i$ and measures the precision of $\hat{\mu}_i$ as an estimate. If $SE(\hat{\mu}_i)$ is small relative to $\hat{\mu}_i$ then $\hat{\mu}_i$ is a relatively precise of μ_i; if $SE(\hat{\mu}_i)$ is large relative to μ_i then $\hat{\mu}_i$ is a relatively imprecise estimate of μ_i.

Unfortunately, $SE(\hat{\mu}_i)$ is not a practically useful measure of the precision of $\hat{\mu}_i$ because it depends on the unknown value of σ_i. To get a practically useful measure of precision for $\hat{\mu}_i$ we compute the *estimated standard error*

$$\widehat{SE}(\hat{\mu}_i) = \sqrt{\widehat{var}(\hat{\mu}_i)} = \frac{\widehat{\sigma}_i}{\sqrt{T}}$$

which is just (8) with the unknown value of σ_i replaced by the least squares estimate $\widehat{\sigma}_i = \sqrt{\widehat{\sigma}_i^2}$.

The precision of $\hat{\mu}_i$ is best communicated by computing a confidence for the unknown value of μ_i. A confidence interval is an interval estimate of μ_i such that we can put an explicit probability statement about the likelihood that the confidence interval covers μ_i. The construction of a confidence interval for μ_i is based on the following statistical result :

$$\frac{\hat{\mu}_i - \mu_i}{\widehat{SE}(\hat{\mu}_i)} \sim t_{T-1}. \tag{9}$$

That is, the standardized value of $\hat{\mu}_i$ has a Student-t distribution with $T-1$ degrees of freedom[3]. To compute a $(1-\alpha)\cdot 100\%$ confidence interval for μ_i we use (9) and the critical value $t_{T-1}(\alpha/2)$ to give

$$\Pr\left(-t_{T-1}(\alpha/2) \leq \frac{\hat{\mu}_i - \mu_i}{\widehat{SE}(\hat{\mu}_i)} \leq t_{T-1}(\alpha/2)\right) = 1-\alpha,$$

which can be rearranged as

$$\Pr\left(\hat{\mu}_i - t_{T-1}(\alpha/2)\cdot\widehat{SE}(\hat{\mu}_i) \leq \mu_i \leq \hat{\mu}_i + t_{T-1}\cdot\widehat{SE}(\hat{\mu}_i)\right) = 0.95.$$

Hence, the interval

$$[\hat{\mu}_i - t_{T-1}(\alpha/2)\cdot\widehat{SE}(\hat{\mu}_i),\ \hat{\mu}_i + t_{T-1}\cdot\widehat{SE}(\hat{\mu}_i)] = \hat{\mu}_i \pm t_{T-1}(\alpha/2)\cdot\widehat{SE}(\hat{\mu}_i)$$

covers the true unknown value of μ_i with probability $1-\alpha$.

[3]This resut follows from the fact that $\hat{\mu}_i$ is normally distributed and $\widehat{SE}(\hat{\mu}_i)$ is equal to the square root of a chi-square random variable divided by its degrees of freedom.

For example, suppose we want to compute a 95% confidence interval for μ_i. In this case $\alpha = 0.05$ and $1 - \alpha = 0.95$. Suppose further that $T - 1 = 60$ so that $t_{T-1}(\alpha/2) = t_{60}(0.025) = 2$. Then the 95% confidence for μ_i is given by

$$\hat{\mu}_i \pm 2 \cdot \widehat{SE}(\hat{\mu}_i).$$

2.4.3 The Gauss-Markov Theorem

To be completed

2.4.4 Statistical properties of the least squares estimators of σ_i^2 and σ_{ij}.

To determine the statistical properties of $\hat{\sigma}_i^2$ we need to treat it as a function of the random sample $R_{i1}, \ldots, R_{iT}$:

$$\hat{\sigma}_i^2 = \hat{\sigma}_i^2(R_{i1}, \ldots R_{iT}) = \frac{1}{T-1} \sum (R_{it} - \hat{\mu}_i)^2.$$

Note also that $\hat{\mu}_i$ is to be treated as a random variable. Similarly, to determine the statistical properties of $\hat{\sigma}_{ij}$ we need to treat it as a function of $R_{i1}, \ldots, R_{iT}$ and $R_{i1}, \ldots, R_{jT}$:

$$\hat{\sigma}_{ij} = \hat{\sigma}_{ij}(R_{i1}, \ldots, R_{iT}; R_{j1}, \ldots, R_{jT}) = \frac{1}{T-1} \sum_{t=1}^{T} (R_{it} - \hat{\mu}_i)(R_{jt} - \hat{\mu}_j).$$

Assuming that returns are generated by the CER model (1), we have the following results:

$$E[\hat{\sigma}_i^2] = \sigma_i^2$$

and

$$E[\hat{\sigma}_{ij}] = \sigma_{ij}.$$

That is, both $\hat{\sigma}_i^2$ and $\hat{\sigma}_{ij}$ are unbiased estimators.

Unfortunately, the derivations of the variances of $\hat{\sigma}_i^2$ and $\hat{\sigma}_{ij}$ are complicated and the results are extremely messy ard hard to work with. As a result, estimated standard errors for $\hat{\sigma}_i^2$ and $\hat{\sigma}_{ij}$ are not usually reported in statistical analysis. Since $\hat{\sigma}_i^2$ and $\hat{\sigma}_{ij}$ are key inputs to portfolio analysis not knowing their precision casts some doubts on the precision of the portfolio analysis[4].

[4]The precision of $\hat{\sigma}_i^2$ and $\hat{\sigma}_{ij}$ can be evaluated numerically using a simulation technique called bootstrapping.

2.4.5 Application: Estimation of the CER model for Microsoft and Apple

Consider estimating the CER model for Microsoft and Apple using monthly continuously compounded returns over the period January 1987 through May 1997. This gives a sample size of $T = 124$. These returns are illustrated in Figure xxx. The least squares estimates of μ_i and the associated estimated standard errors are given by

$$\begin{aligned} \hat{\mu}_{msft} &= 0.036, \ \widehat{SE}(\hat{\mu}_{msft}) = 0.009, \\ \hat{\mu}_{apple} &= -0.001, \ \widehat{SE}(\hat{\mu}_{apple}) = 0.012. \end{aligned}$$

The estimated monthly mean return for Microsoft is 3.6%, which is quite remarkable, whereas the estimated monthly mean return for Apple is -0.1%. The implied estimates for the expected annual continuously compounded returns on Microsoft and Apple are $12 \cdot (0.036) = 43.2\%$ and $12 \cdot (-0.001) = -1.2\%$. The expected monthly return for Microsoft is estimated fairly precisely since the estimated standard error, $\widehat{SE}(\hat{\mu}_{msft}) = 0.009$, is small relative to the estimated mean. The mean for Apple, however, is not estimated very precisely since the estimated standard error, $\widehat{SE}(\hat{\mu}_{apple}) = 0.012$, is quite large relative to the estimated mean.

To better communicate the precision of the estimated means, interval estimates in the form of confidence intervals can be computed. 95% confidence intervals μ_i are

$$\begin{aligned} \hat{\mu}_{msft} \pm 1.98 \cdot t_{123}(0.25) \cdot \widehat{SE}(\hat{\mu}_{msft}) &= 0.036 \pm 1.98 \cdot 0.009 \\ &= [0.018, 0.054], \\ \hat{\mu}_{apple} \pm 1.98 \cdot t_{123}(0.25) \cdot \widehat{SE}(\hat{\mu}_{apple}) &= -0.001 \pm 1.98 \cdot 0.012 \\ &= [-0.025, 0.023]. \end{aligned}$$

The 95% confidence interval for μ_{msft} indicates that the expected monthly return Microsoft can be as small as 1.8% per month or as high as 5.4% per month. The interval for μ_{apple} indicates that the expected monthly return can be as low as -2.5% per month or as high as 2.3%. Clearly, there is more uncertainty about the expected return for Apple than there is for Microsoft.

The estimates of σ_i^2 and σ_{ij} are given by

$$\begin{aligned} \hat{\sigma}^2_{msft} &= 0.011, \ \hat{\sigma}_{msft} = 0.104, \\ \hat{\sigma}^2_{apple} &= 0.018, \ \hat{\sigma}_{apple} = 0.134, \\ \hat{\sigma}_{msft,apple} &= 0.007. \end{aligned}$$

The estimated standard deviations of the monthly returns on Microsoft and Apple are both quite large - around 10% per month. The estimated covariance between the returns on Microsoft and Apple is positive which indicates that the returns tend to move in the same direction. To measure the strength of the linear dependence between the returns we estimate the correlation between the returns, $\rho_{msft,apple}$. The estimated correlation coefficient is given by

$$\hat{\rho}_{msft,apple} = \frac{\hat{\sigma}_{msft,apple}}{\hat{\sigma}_{msft}\hat{\sigma}_{apple}} = \frac{0.007}{(0.104)(0.134)} = 0.497,$$

and shows that there is a fairly strong positive relationship between the returns on Microsoft and the returns on Apple.

3 Problems

4 References