Chapter 1

The Constant Expected Return Model

Date: October 25, 2009

The first model of asset returns we consider is the very simple *constant expected return* (CER) model. This model assumes that an asset's return over time is normally distributed with a constant (time invariant) mean and variance. The model allows for the returns on different assets to be contemporaneously correlated but that the correlations are constant over time. Although this model is very simple, it allows us to discuss and develop several important econometric topics such as Monte Carlo simulation, estimation, bootstrapping, hypothesis testing, forecasting and model evaluation.

1.1 CER Model Assumptions

Let $r_{it} = \ln(P_{it}/P_{t-1})$ denote the continuously compounded return on 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 $t = 1, \ldots, T$.

Assumption 1

(i) *Covariance stationarity*: $\{r_{i1}, \ldots, r_{iT}\} = \{r_{it}\}_{t=1}^{T}$ is a covariance stationary and ergodic stochastic process with $E[r_{it}] = \mu_i$, $\mathrm{var}(r_{it}) = \sigma_i^2$, $\mathrm{cov}(r_{it}, r_{jt}) = \sigma_{ij}$ and $\mathrm{cor}(r_{it}, r_{jt}) = \rho_{ij}$.

(ii) Normality: $r_{it} \sim N(\mu_i, \sigma_i^2)$

(iii) No serial correlation: $\text{cov}(r_{it}, r_{js}) = \text{cor}(r_{it}, r_{is}) = 0$ for $t \neq s$ and $i, j = 1, \ldots, N$.

Assumption 1 states that in every time period asset returns are jointly (multivariate) normally distributed, that the means and the variances of all asset returns, and all of the pairwise contemporaneous covariances and correlations between assets are constant over time. In addition, all of the asset returns are *serially uncorrelated*

$$\text{cor}(r_{it}, r_{is}) = \text{cov}(r_{it}, r_{is}) = 0 \text{ for all } i \text{ and } t \neq s,$$

and the returns on all possible pairs of assets i and j are serially uncorrelated

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

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, testing hypotheses about the parameter values and assumptions.

1.1.1 Regression Model Representation

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

$$\begin{aligned} r_{it} &= \mu_i + \varepsilon_{it}, \\ \{\varepsilon_{it}\}_{t=1}^{T} &\sim \text{GWN}(0, \sigma_i^2), \\ \text{cov}(\varepsilon_{it}, \varepsilon_{js}) &= \begin{cases} \sigma_{ij} & t = s \\ 0 & t \neq s \end{cases}. \end{aligned} \tag{1.1}$$

The notation $\varepsilon_{it} \sim \text{GWN}(0, \sigma_i^2)$ stipulates that the stochastic process $\{\varepsilon_{it}\}_{t=1}^{T}$ is a Gaussian white noise process with $E[\varepsilon_{it}] = 0$ and $\text{var}(\varepsilon_{it}) = \sigma_i^2$. In addition, the random error term ε_{it} is independent of ε_{js} for all assets $i \neq j$ and all time periods $t \neq s$.

Using the basic properties of expectation, variance and covariance, we can derive the following properties of returns in the CER model:

$$\begin{aligned}
E[r_{it}] &= E[\mu_i + \varepsilon_{it}] = \mu_i + E[\varepsilon_{it}] = \mu_i, \\
\mathrm{var}(r_{it}) &= \mathrm{var}(\mu_i + \varepsilon_{it}) = \mathrm{var}(\varepsilon_{it}) = \sigma_i^2, \\
\mathrm{cov}(r_{it}, r_{jt}) &= \mathrm{cov}(\mu_i + \varepsilon_{it}, \mu_j + \varepsilon_{jt}) = \mathrm{cov}(\varepsilon_{it}, \varepsilon_{jt}) = \sigma_{ij}, \\
\mathrm{cov}(r_{it}, R_{js}) &= \mathrm{cov}(\mu_i + \varepsilon_{it}, \mu_j + \varepsilon_{js}) = \mathrm{cov}(\varepsilon_{it}, \varepsilon_{js}) = 0,\ t \neq s.
\end{aligned}$$

Given that covariances and variances of returns are constant over time implies that the correlations between returns over time are also constant:

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

Finally, since $\{\varepsilon_{it}\}_{t=1}^T \sim \mathrm{GWN}(0, \sigma_i^2)$ it follows that $\{r_{it}\}_{t=1}^T \sim iid\ N(\mu_i, \sigma_i^2)$. Hence, the CER regression model (1.1) for r_{it} is equivalent to the model implied by Assumption 1.

1.1.2 Interpretation of the CER Regression Model

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 between time $t-1$ and time t. To see this, (1.1) implies that

$$\varepsilon_{it} = r_{it} - \mu_i = r_{it} - E[r_{it}],$$

so that ε_{it} is defined as the deviation of the random return from its expected value. If the news between times $t-1$ and t 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_{it} is negative and the observed return is less than expected. The assumption $E[\varepsilon_{it}] = 0$ means that news, on average, is neutral; neither good nor bad. The assumption that $\mathrm{var}(\varepsilon_{it}) = \sigma_i^2$ can be interpreted as saying that volatility, or typical magnitude, of news arrival is constant over time. The random news variable affecting asset i, ε_{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, if asset i is Microsoft stock and asset j is Apple Computer stock, 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 both ε_{it} and ε_{jt}. Hence these variables will be positively correlated due to a positive reaction to a common news component. Finally, the news on asset j at time s is unrelated to the news on asset i at time t for all times $t \neq s$. For example, this means that the news for Apple in January is not related to the news for Microsoft in February.

Sometimes it is convenient to re-express the CER model (1.1) as

$$\begin{aligned} r_{it} &= \mu_i + \varepsilon_{it} = \mu_i + \sigma_i \cdot z_{it} \\ \{z_{it}\}_{t=1}^{T} &\sim \text{GWN}(0,1), \end{aligned} \tag{1.2}$$

In this form, the random news shock is the *iid* standard normal random variable z_{it} scaled by the volatility σ_i. This form is particularly convenient for value-at-risk calculations.

1.1.3 Time Aggregation and the CER Model

The CER model for continuously compounded returns has the following nice aggregation property with respect to the interpretation of ε_{it} as news. Suppose that t represents months so that r_{it} is the continuously compounded monthly return on asset i. Now, instead of the monthly return, suppose we are interested in the annual continuously compounded return $r_{it} = r_{it}(12)$. Since multi-period continuously compounded returns are additive, $r_{it}(12)$ is the sum of 12 monthly continuously compounded returns:

$$r_{it} = r_{it}(12) = \sum_{t=0}^{11} R_{it-k} = r_{it} + r_{it-1} + \cdots + r_{it-11}.$$

Using the CER regression model (1.1) for the monthly return r_{it}, we may express the annual return $r_{it}(12)$ as

$$r_{it}(12) = \sum_{t=0}^{11} (\mu_i + \varepsilon_{it}) = 12 \cdot \mu_i + \sum_{t=0}^{11} \varepsilon_{it} = \mu_i^A + \varepsilon_{it}(12),$$

where $\mu_i^A = 12 \cdot \mu_i$ is the annual expected return on asset i, and $\varepsilon_{it}(12) = \sum_{k=0}^{11} \varepsilon_{it-k}$ is the annual random news component. The annual expected return, μ_i^A, is simply 12 times the monthly expected return, μ_i. The annual random news component, $\varepsilon_{it}(12)$, is the accumulation of news over the year. As a result, the variance of the annual news component is 12 time the variance of the monthly news component:

$$
\begin{aligned}
\operatorname{var}(\varepsilon_{it}(12)) &= \operatorname{var}\left(\sum_{k=0}^{11} \varepsilon_{it-k}\right) \\
&= \sum_{k=0}^{11} \operatorname{var}(\varepsilon_{it-k}) \text{ since } \varepsilon_{it} \text{ is uncorrelated over time} \\
&= \sum_{k=0}^{11} \sigma_i^2 \text{ since } \operatorname{var}(\varepsilon_{it}) \text{ is constant over time} \\
&= 12 \cdot \sigma_i^2 = \left(\sigma_i^A\right)^2.
\end{aligned}
$$

It follows that the standard deviation of the annual news is equal to $\sqrt{12}$ times the standard deviation of monthly news:

$$
\mathrm{SD}(\varepsilon_{it}(12)) = \sqrt{12}\mathrm{SD}(\varepsilon_{it}) = \sqrt{12}\sigma_i.
$$

This result is know as the *square root of time rule*. Similarly, due to the additivity of covariances, the covariance between $\varepsilon_{it}(12)$ and $\varepsilon_{jt}(12)$ is 12 times the monthly covariance:

$$
\begin{aligned}
\operatorname{cov}(\varepsilon_{it}(12), \varepsilon_{jt}(12)) &= \operatorname{cov}\left(\sum_{k=0}^{11} \varepsilon_{it-k}, \sum_{k=0}^{11} \varepsilon_{jt-k}\right) \\
&= \sum_{k=0}^{11} \operatorname{cov}(\varepsilon_{it-k}, \varepsilon_{jt-k}) \text{ since } \varepsilon_{it} \text{ and } \varepsilon_{jt} \text{ are uncorrelated over time} \\
&= \sum_{k=0}^{11} \sigma_{ij} \text{ since } \operatorname{cov}(\varepsilon_{it}, \varepsilon_{jt}) \text{ is constant over time} \\
&= 12 \cdot \sigma_{ij} = \sigma_{ij}^A.
\end{aligned}
$$

The above results imply that the correlation between the annual errors $\varepsilon_{it}(12)$ and $\varepsilon_{jt}(12)$ is the same as the correlation between the monthly errors ε_{it} and

ε_{jt} :

$$\begin{aligned}
\mathrm{cor}(\varepsilon_{it}(12), \varepsilon_{jt}(12)) &= \frac{\mathrm{cov}(\varepsilon_{it}(12), \varepsilon_{jt}(12))}{\sqrt{\mathrm{var}(\varepsilon_{it}(12)) \cdot \mathrm{var}(\varepsilon_{jt}(12))}} \\
&= \frac{12 \cdot \sigma_{ij}}{\sqrt{12\sigma_i^2 \cdot 12\sigma_j^2}} \\
&= \frac{\sigma_{ij}}{\sigma_i \sigma_j} = \rho_{ij} = \mathrm{cor}(\varepsilon_{it}, \varepsilon_{jt}).
\end{aligned}$$

1.1.4 The Random Walk Model of Asset Prices

The CER model of asset returns (1.1) gives rise to the so-called *random walk* (RW) model of the *logarithm* of asset prices. To see this, recall that the continuously compounded return, r_{it}, is defined from asset prices via $r_{it} = \ln\left(\frac{P_{it}}{P_{it-1}}\right) = \ln(P_{it}) - \ln(P_{it-1})$. Letting $p_{it} = \ln(P_{it})$ and using the representation of r_{it} in the CER model (1.1), we can express the log-price as:

$$p_{it} = p_{it-1} + \mu_i + \varepsilon_{it}. \tag{1.3}$$

The representation in (1.3) is known as the RW model for log-prices. It is a representation of the CER model in terms of log-prices.

In the RW model (1.3), μ_i represents the expected change in the log-price (continuously compounded return) between months $t-1$ and t, and ε_{it} represents the unexpected change in the log-price. That is,

$$\begin{aligned}
E[\Delta p_{it}] &= E[r_{it}] = \mu_i, \\
\varepsilon_{it} &= \Delta p_{it} - E[\Delta p_{it}].
\end{aligned}$$

where $\Delta p_{it} = p_{it} - p_{it-1}$. Further, in the RW model, the unexpected changes in log-price, ε_{it}, are uncorrelated over time ($\mathrm{cov}(\varepsilon_{it}, \varepsilon_{is}) = 0$ for $t \neq s$) so that future changes in log-price cannot be predicted from past changes in the log-price[1].

The RW model gives the following interpretation for the evolution of log prices. Let p_{i0} denote the initial log price of asset i. The RW model says that the log-price at time $t = 1$ is

$$p_{i1} = p_{i0} + \mu_i + \varepsilon_{i1},$$

[1] The notion that future changes in asset prices cannot be predicted from past changes in asset prices is often referred to as the weak form of the efficient markets hypothesis.

where ε_{i1} is the value of random news that arrives between times 0 and 1. Notice that at time $t = 0$ the expected log-price at time $t = 1$ is

$$E[p_{i1}] = p_{i0} + \mu_i + E[\varepsilon_{i1}] = p_{i0} + \mu_i,$$

which is the initial price plus the expected return between times 0 and 1. Similarly, by recursive substitution the log-price at time $t = 2$ is

$$\begin{aligned} p_{i2} &= p_{i1} + \mu_i + \varepsilon_{i2} \\ &= p_{i0} + \mu_i + \mu_i + \varepsilon_{i1} + \varepsilon_{i2} \\ &= p_{i0} + 2 \cdot \mu_i + \sum_{t=1}^{2} \varepsilon_{it}, \end{aligned}$$

which is equal to the initial log-price, p_{i0}, plus the two period expected return, $2 \cdot \mu_i$, plus the accumulated random news over the two periods, $\sum_{t=1}^{2} \varepsilon_{it}$. By recursive substitution, the log price at time $t = T$ is

$$p_{iT} = p_{i0} + T \cdot \mu_i + \sum_{t=1}^{T} \varepsilon_{it}.$$

At time $t = 0$, the expected log-price at time $t = T$ is

$$E[p_{iT}] = p_{i0} + T \cdot \mu_i,$$

which is the initial price plus the expected growth in prices over T periods. The actual price, p_{iT}, deviates from the expected price by the accumulated random news:

$$p_{iT} - E[p_{iT}] = \sum_{t=1}^{T} \varepsilon_{it}.$$

At time $t = 0$, the variance of the log-price at time T is

$$\mathrm{var}(p_{iT}) = \mathrm{var}\left(\sum_{t=1}^{T} \varepsilon_{it}\right) = T \cdot \sigma^2$$

Hence, the RW model implies that the stochastic process of log-prices $\{p_{it}\}$ is non-stationary because the variance of p_{it} increases with t. Finally, because $\varepsilon_{it} \sim iid\ N(0, \sigma^2)$ it follows that $p_{iT} \sim N(T\mu, T\sigma^2)$.

The term *random walk* was originally used to describe the unpredictable movements of a drunken sailor staggering down the street. The sailor starts at an initial position, p_0, outside the bar. The sailor generally moves in the direction described by μ but randomly deviates from this direction after each step t by an amount equal to ε_t. After T steps the sailor ends up at position $p_T = p_0 + \mu \cdot T + \sum_{t=1}^{T} \varepsilon_t$. The sailor is expected to be at location μT, but where he actually ends up depends on the accumulation of the random changes in direction $\sum_{t=1}^{T} \varepsilon_t$. Because $\mathrm{var}(p_T) = \sigma^2 T$, the uncertainty about where the sailor will be increases with each step.

The RW model for log-price implies the following model for price:

$$P_t = P_0 e^{\mu \cdot t + \sum_{s=1}^{t} \varepsilon_s} = P_0 e^{\mu t} e^{\sum_{s=1}^{t} \varepsilon_s}.$$

The term $e^{\mu t}$ represents the expected exponential growth rate in prices between times 0 and time t, and the term $e^{\sum_{s=1}^{t} \varepsilon_s}$ represents the unexpected exponential growth in prices. Here, P_t is log-normally distributed because $p_t = \ln P_t$ is normally distributed.

1.1.5 The CER Model in Matrix Notation

Define the $N \times 1$ vectors $\mathbf{R}_t = (R_{1t}, \ldots, R_{Nt})'$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)'$, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$ and the $N \times N$ symmetric covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1N} & \sigma_{2N} & \cdots & \sigma_N^2 \end{pmatrix}.$$

Then the CER model matrix notation is

$$\begin{aligned} \mathbf{R}_t &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t, \\ \boldsymbol{\varepsilon}_t &\sim GWN(\mathbf{0}, \boldsymbol{\Sigma}), \end{aligned} \tag{1.4}$$

which implies that $\mathbf{R}_t \sim iid\ N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

1.2 Monte Carlo Simulation of the CER Model

A simple technique that can be used to understand the probabilistic behavior of a model involves using computer simulation methods to create pseudo data

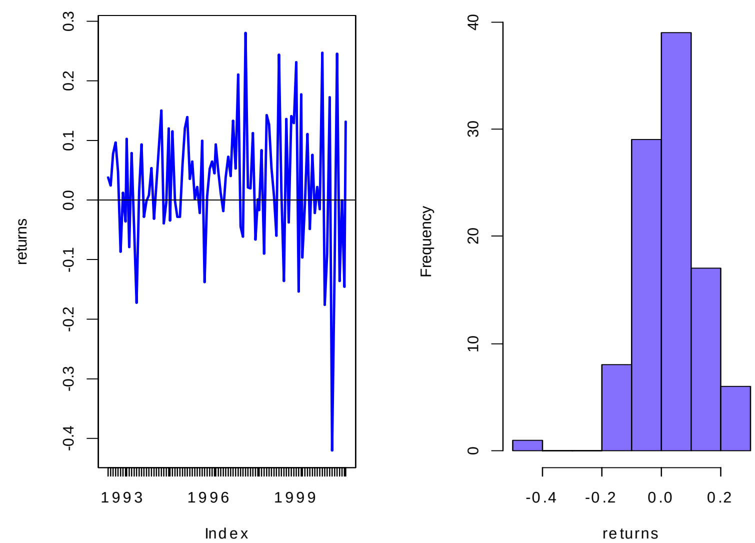

Figure 1.1: Monthly continuously compounded returns on Microsoft. Source: finance.yahoo.com.

from the model. The process of creating such pseudo data is called *Monte Carlo simulation*[2]. To illustrate the use of Monte Carlo simulation, consider creating pseudo return data from the CER model (1.1) for a single asset. The steps to create a Monte Carlo simulation from the CER model are:

1. Fix values for the CER model parameters μ and σ.

2. Determine the number of simulated values, T, to create.

3. Use a computer random number generator to simulate T *iid* values of ε_t from a $N(0, \sigma^2)$ distribution. Denote these simulated values are $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_T$.

4. Create the simulated return data $\tilde{R}_t = \mu + \tilde{\varepsilon}_t$ for $t = 1, \ldots, T$.

To motivate plausible values for μ and σ in the simulation, Figure 1.1 shows the actual monthly continuously compounded return data on Microsoft

[2] Monte Carlo referrs to the famous city in Monaco where gambling is legal.

stock over the ten-year period July, 1992 through October, 2000. The parameter $\mu_i = E[r_{it}]$ in the CER model represents this central value, and σ_i represents the typical size of a deviation about μ_i. In Figure 1.1, the returns seem to fluctuate up and down about a central value near 0.03, or 3%, and the typical size of a return deviation about 0.03 is roughly 0.10, or 10%. In fact, the sample average return is 2.8% and the sample standard deviation is 10.7%.

To mimic the monthly return data on Microsoft in the Monte Carlo simulation, the values $\mu = 0.03$ and $\sigma = 0.10$ are used as the model's parameters and $T = 100$ is the number of simulated values (sample size). Let $\{\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_{100}\}$ denote the 100 simulated values of the news variable $\varepsilon_t \sim \mathrm{GWN}(0, (0.10)^2)$.The simulated returns are then computed using[3]

$$\tilde{R}_t = 0.03 + \tilde{\varepsilon}_t,\ t = 1, \ldots, 100 \tag{1.5}$$

Example 1 *Simulating observation from the CER model using R*

To create and plot the simulated returns from (1.5) use

```
> mu = 0.03
> sd.e = 0.10
> nobs = 100
> set.seed(111)
> sim.e = rnorm(nobs, mean=0, sd=sd.e)
> sim.ret = mu + sim.e
> par(mfrow=c(1,2))
> ts.plot(sim.ret, main="",
+        xlab="months",ylab="return", lwd=2, col="blue")
> abline(h=mu)
> hist(sim.ret, main="", xlab="returns", col="slateblue1")
> par(mfrow=c(1,1))
```

The simulated returns $\{\tilde{R}_t\}_{t=1}^{100}$ are shown in Figure 1.2. The simulated return data fluctuate randomly about $\mu = 0.03$, and the typical size of the fluctuation is approximately equal to $\sigma = 0.10$. The simulated return data look somewhat like the actual monthly return data for Microsoft. The main difference is that the return volatility for Microsoft appears to have increased in the latter part of the sample whereas the simulated data has constant volatility over the entire sample. ■

[3] Alternatively, the returns can be simulated directly by simulating observation from a normal distribution with mean 0.05 and standard deviation 0.10.

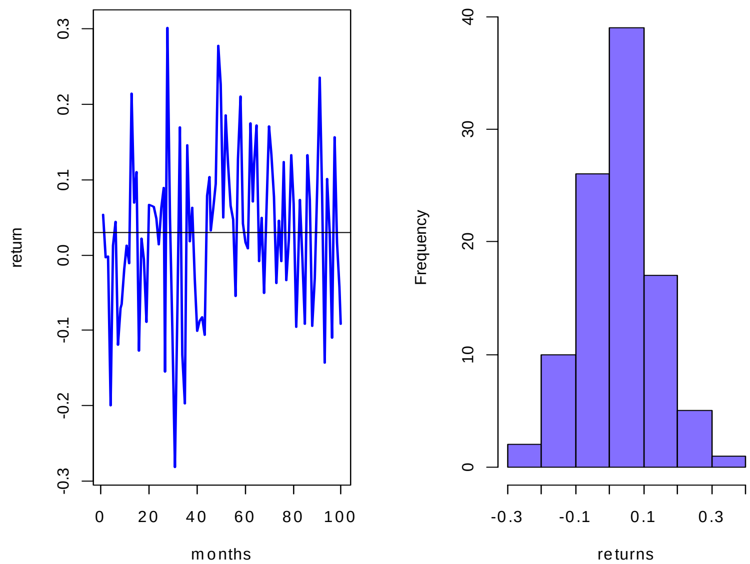

Figure 1.2: Simulated returns from the CER model $r_t = 0.03 + \varepsilon_t$, $\varepsilon_t \sim \text{GWN}(0, (0.10)^2)$.

Monte Carlo simulation of a model can be used as a first pass reality check of the model. If simulated data from the model do not look like the data that the model is supposed to describe, then serious doubt is cast on the model. However, if simulated data look reasonably close to the actual data then the first step reality check is passed.

Example 2 *Simulating log-prices from the RW model*

The RW model for log-price based on the CER model (1.5) is

$$p_t = p_0 + 0.03 \cdot t + \sum_{j=1}^{t} \varepsilon_j, \varepsilon_t \sim \text{GWN}(0, (0.10)^2)$$

A Monte Carlo simulation of this RW model with $p_0 = 1$ can be create in R using

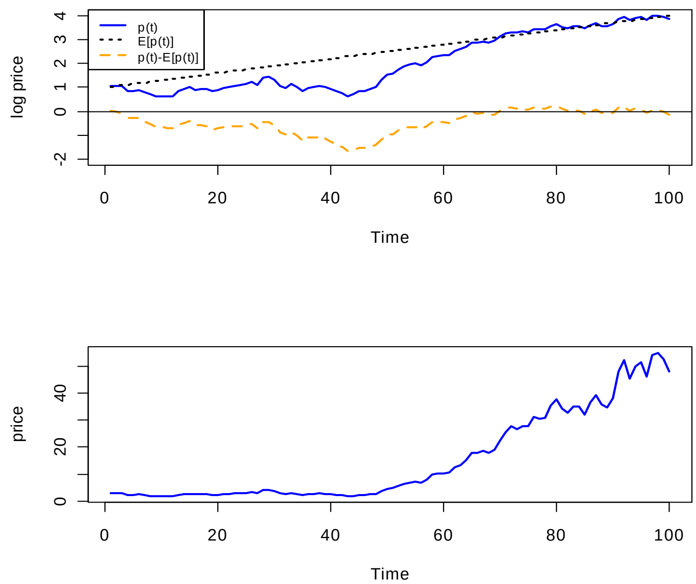

Figure 1.3: Simulated values from the RW model $p_t = p_0 + 0.03 \cdot t + \sum_{j=1}^{t} \varepsilon_j$, $\varepsilon_t \sim \text{GWN}(0, (0.10)^2)$.

```
> mu = 0.03
> sd.e = 0.10
> nobs = 100
> set.seed(111)
> sim.e = rnorm(nobs, mean=0, sd=sd.e)
> sim.p = 1 + mu*seq(nobs) + cumsum(sim.e)
> sim.P = exp(sim.p)
```

Figure 1.3 shows the simulated values. The top panel shows the log price, p_t, the expected price $E[p_t] = p_0 + 0.03 \cdot t$ and the accumulated random news $p_t - E[p_t] = \sum_{s=1}^{t} \tilde{\varepsilon}_s$. The bottom panel shows the simulated price levels $P_t = e^{p_t}$. Figure 1.4 shows the actual log prices and price levels for Microsoft stock. Notice the similarity between the simulated random walk data and the actual data.

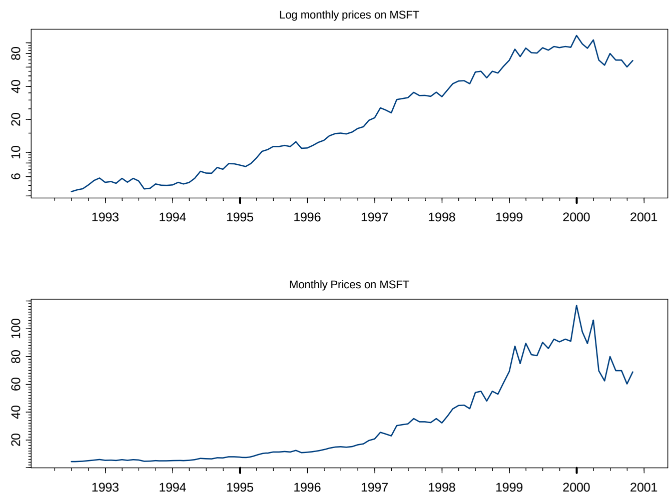

Figure 1.4:

1.2.1 Simulating Returns on More than One Asset

Creating a Monte Carlo simulation of more than one return from the CER model requires simulating observations from a multivariate normal distribution. This follows from the matrix representation of the CER model given in (1.4). The steps required to create a Monte Carlo simulation are:

1. Fix values for the CER model parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

2. Determine the number of simulated values, T, to create.

3. Use a computer random number generator to simulate T *iid* values of the random vector $\boldsymbol{\varepsilon}_t$ from the multivariate normal distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$. Denote these simulated values are $\tilde{\boldsymbol{\varepsilon}}_1, \ldots, \tilde{\boldsymbol{\varepsilon}}_T$.

4. Create the simulated return vector $\tilde{\mathbf{R}}_t = \boldsymbol{\mu} + \tilde{\boldsymbol{\varepsilon}}_t$ for $t = 1, \ldots, T$.

To motivate the parameters for a multivariate simulation of the CER model, consider the monthly continuously compounded returns for Microsoft, Starbucks and the S & P 500 index over the ten-year period July, 1992 through October, 2000 illustrated in Figures 1.5 and 1.6. All returns seem

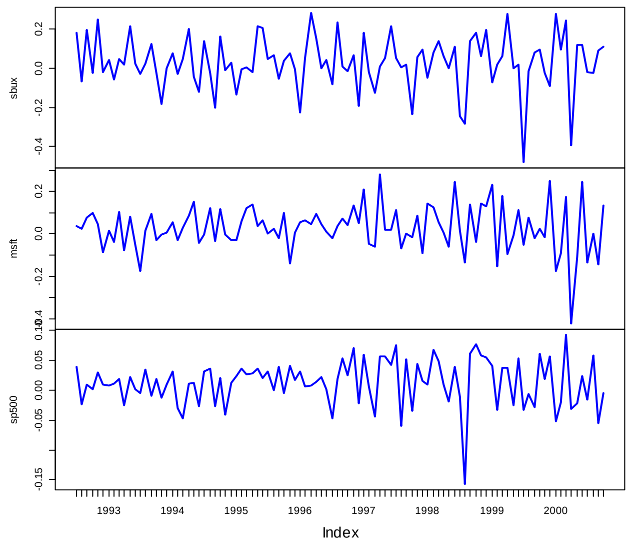

Figure 1.5: Monthly continuously compounded returns on Microsoft, Starbucks and the S&P 500 index. Source: finance.yahoo.com.

to fluctuate around a mean value near zero. The volatilities of Microsoft and Starbucks are similar with typical magnitudes around 0.10, or 10%. The volatility of the S&P 500 index is considerably smaller at about 0.05, or 5%. The pairwise scatterplots show that all returns appear to be positively related. The pairs (MSFT,SP500) and (SBUX,SP500) appear to be the most correlated, and the shape of the scatter indicates correlation values around 0.5. The pair (MSFT, SBUX) shows a weak positive correlation around 0.2. Let $\mathbf{R}_t = (R_{msft}, R_{sbux}, R_{sp500})'$. Then, a plausible value for $\boldsymbol{\mu}$ is $\boldsymbol{\mu} = (0, 0, 0)'$,

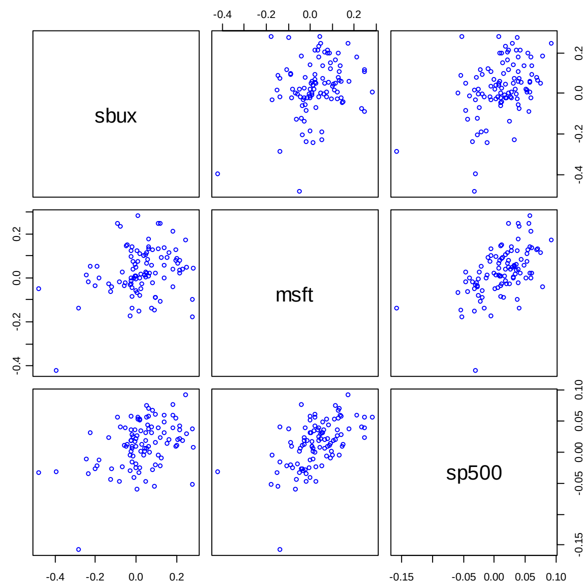

Figure 1.6: Scatterplot matrix of the monthly continuously compounded returns on Microsoft, Starbucks and the S&P 500 index.

and a plausible value for $\mathbf{\Sigma}$ is

$$\mathbf{\Sigma} = \begin{pmatrix} (0.10)^2 & (0.10)(0.10)(0.2) & (0.10)(0.05)(0.5) \\ (0.10)(0.05)(0.5) & (0.10)^2 & (0.10)(0.05)(0.5) \\ (0.10)(0.05)(0.5) & (0.10)(0.05)(0.5) & (0.05)^2 \end{pmatrix}$$

$$= \begin{pmatrix} 0.0100 & 0.0020 & 0.0025 \\ 0.0020 & 0.0100 & 0.0025 \\ 0.0025 & 0.0025 & 0.0025 \end{pmatrix}$$

Example 3 *Monte Carlo simulation of CER model for three assets*

Simulating values from the multivariate CER model (1.4) requires simulating multivariate normal random variables. In R, this can be done using the function `rmvnorm()` from the package mvtnorm. To create a Monte Carlo simulation from the CER model calibrated to the month continuously returns on Microsoft, Starbucks and the S&P 500 index use

```
> mu = rep(0,3)
> sig.msft = 0.10
> sig.sbux = 0.10
> sig.sp500 = 0.05
> rho.msft.sbux = 0.2
> rho.msft.sp500 = 0.5
> rho.sbux.sp500 = 0.5
> sig.msft.sbux = rho.msft.sbux*sig.msft*sig.sbux
> sig.msft.sp500 = rho.msft.sp500*sig.msft*sig.sp500
> sig.sbux.sp500 = rho.sbux.sp500*sig.sbux*sig.sp500
> Sigma = matrix(c(sig.msft^2, sig.msft.sbux, sig.msft.sp500,
+                 sig.msft.sbux, sig.sbux^2, sig.sbux.sp500,
+                 sig.msft.sp500, sig.sbux.sp500, sig.sp500^2),
+                 nrow=3, ncol=3, byrow=TRUE)
> nobs = 100
> set.seed(123)
> returns.sim = rmvnorm(nobs, mean=mu, sigma=Sigma)
```

The simulated returns are shown in Figure 1.7. They look fairly similar to the actual returns given in Figure 1.5.

1.3 Estimating the Parameters of the CER Model

The CER model of asset returns gives us a simple framework for interpreting the time series behavior of asset returns and prices. At the beginning of every month t, r_{it} is a random variable representing the continuously compounded return to be realized at the end of the month. The CER model states that $r_{it} \sim iid\ N(\mu_i, \sigma_i^2)$. 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 $\mathrm{SD}(r_{it}) = \sigma_i$, and our measures of the direction and strength of linear association between r_{it} and r_{jt} are $\sigma_{ij} = \mathrm{cov}(r_{it}, r_{jt})$ and $\rho_{ij} = \mathrm{cor}(r_{it}, r_{jt})$, respectively. 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.

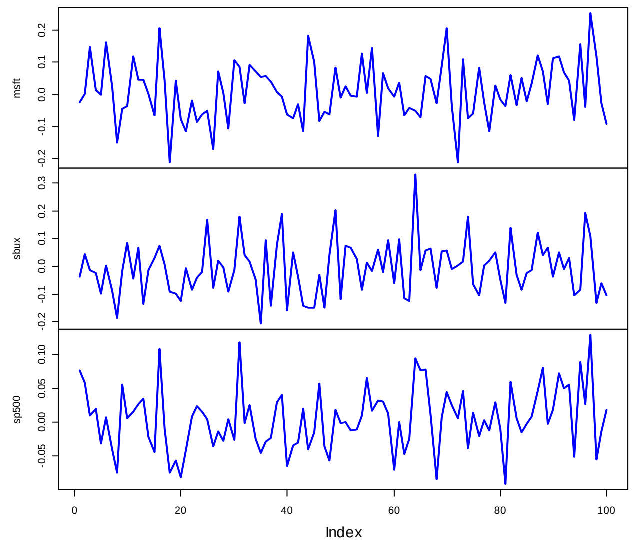

Figure 1.7: Monte Carlo simulation of monthly continuously compounded returns on Microsoft, Starbuck and the S&P 500 index from the multivariate CER model.

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

Suppose we observe monthly continuously compounded returns on N different assets over the horizon $t = 1, \ldots, T$. It is assumed that the observed returns are realizations of the stochastic process $\{r_{i1}\}_{t=1}^{T}$, where r_{it} is described by the CER model (1.1). Under these assumptions, we can use the observed returns to estimate the unknown parameters of the CER model. However, before we describe the estimation of the CER model, it is necessary to review some fundamental concepts in the statistical theory of estimation.

1.3.1 Estimators and Estimates

Let θ denote some characteristic of the CER model (1.1) 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 a sample of size T of the observed data.

Definition 4 *An estimator of θ is a rule or algorithm for estimating θ.*

Definition 5 *An estimate of θ is simply the value of an estimator based on the observed data.*

Example 6 *The sample average as an estimator and an estimate*

The sample average $\frac{1}{T}\sum_{t=1}^{T} r_t$ is an algorithm for computing an estimate of the expected return μ. Before the sample is observed, the sample average is a simple linear function of the random variables $\{r_1, \ldots, r_T\}$ and so is itself a random variable. After the sample is observed, the sample average can be evaluated using the observed data which produces the estimate. For example, suppose $T = 5$ and the realized values of the returns are $r_1 = 0.1, r_2 = 0.05, r_3 = 0.025, r_4 = -0.1, r_5 = -0.05$. Then the estimate of μ_i using the sample average is

$$\hat{\mu} = \frac{1}{5}(0.1 + 0.05 + 0.025 + -0.1 + -0.05) = 0.005$$

■

To discuss the properties of estimators it is necessary to establish some notation. Let $\hat{\theta}$ denote an estimator of θ when it is treated as a function of the random variables $\{r_{i1}, \ldots, r_{iT}\}$. In this case, $\hat{\theta}$ is a random variable. Let $\hat{\theta}$ denote an estimate of θ when it is based on the observed data. Then $\hat{\theta}$ represents the realized value of the estimator and is simply a number. We use $\hat{\theta}$ to represent either an estimator of θ or an estimate of θ. The context will determine how to interpret $\hat{\theta}$.

1.3.2 Properties of Estimators

Consider $\hat{\theta}$ as a random variable. In general, the pdf of $\hat{\theta}$, $f(\hat{\theta})$, depends on the pdf's of the random variables $\{r_{i1}, \ldots, r_{iT}\}$. The exact form of $f(\hat{\theta})$ may be very complicated. For analysis purposes, we often focus on certain characteristics of $f(\hat{\theta})$, like its expected value (center), variance and standard deviation (spread about expected value). The expected value of an estimator is related to the concept of estimator *bias*, and the variance/standard deviation of an estimator is related to the concept of estimator *precision*. Different realizations of the random variables $\{r_{i1}, \ldots, r_{iT}\}$ will produce different values of $\hat{\theta}$. Some values of $\hat{\theta}$ will be bigger than θ and some will be smaller. Intuitively, a *good* estimator of θ is one that is on average correct (unbiased) and never gets to far away from θ (small variance). That is, a good estimator will have small bias and high precision.

Bias

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

Definition 7 *The estimation error is the difference between the estimator and the parameter being estimated:*

$$\text{error}(\hat{\theta}, \theta) = \hat{\theta} - \theta.$$

Definition 8 *The bias of an estimator $\hat{\theta}$ of θ is the expected estimation error:*

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

Definition 9 *An estimator $\hat{\theta}$ of θ is unbiased if* $\text{bias}(\hat{\theta}, \theta) = 0$; *i.e., if* $E[\hat{\theta}] = \theta$ *or* $E[\text{error}(\hat{\theta}, \theta)] = 0$.

Unbiasedness is a desirable property of an estimator. It means that the estimator produces the correct answer "on average", where "on average" means over many hypothetical realizations of the random variables $\{r_{i1}, \ldots, r_{iT}\}$. It is important to keep in mind that an unbiased estimator for θ may not be very close to θ for a particular sample, and that a biased estimator may be actually be quite close to θ. For example, consider the pdf of $\hat{\theta}_1$ in Figure 1.8. The center of the distribution is at the true value $\theta = 0$, $E[\hat{\theta}_1] = 0$, but the distribution is very widely spread out about $\theta = 0$. That is, $\text{var}(\hat{\theta}_1)$ is large. On average (over many hypothetical samples) the value of $\hat{\theta}_1$ will be close to θ, but in any given sample the value of $\hat{\theta}_1$ can be quite a bit above or below θ. Hence, unbiasedness by itself does not guarantee a good estimator of θ. Now consider the pdf for $\hat{\theta}_2$. The center of the pdf is slightly higher than $\theta = 0$, i.e., $\text{bias}(\hat{\theta}_2, \theta) > 0$, but the spread of the distribution is small. Although the value of $\hat{\theta}_2$ is not equal to 0 *on average* we might prefer the estimator $\hat{\theta}_2$ over $\hat{\theta}_1$ because it is generally closer to $\theta = 0$ *on average* than $\hat{\theta}_1$.

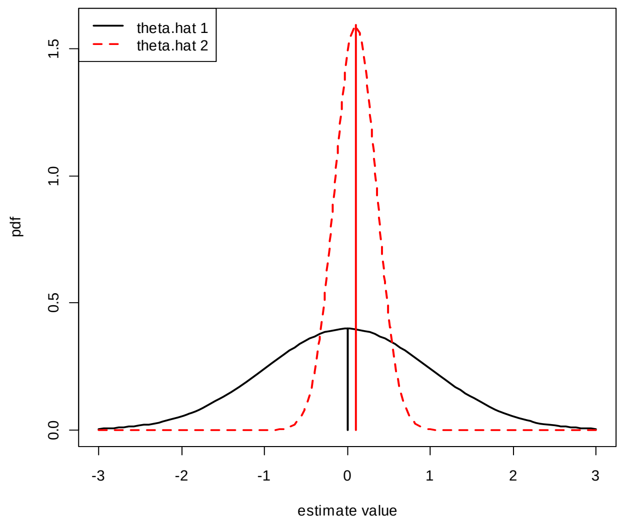

Figure 1.8: Distributions of competiting estimators for $\theta = 0$. $\hat{\theta}_1$ is unbiased but has high variance, and $\hat{\theta}_2$ is biased but has low variance.

Precision

An estimate is, hopefully, our best guess of the true (but unknown) value of θ. Our guess most certainly will be wrong, but we hope it will not be too far off. A precise estimate is one in which the variability of the estimation error is small. The variability of the estimation error is captured by the *mean squared error.*

Definition 10 *The mean squared error of an estimator $\hat{\theta}$ of θ is given by*

$$\text{mse}(\hat{\theta}, \theta) = E[(\hat{\theta} - \theta)^2] = E[\text{error}(\hat{\theta}, \theta)^2]$$

The mean squared error measures the expected squared deviation of $\hat{\theta}$ from θ. If this expected deviation is small, then we know that $\hat{\theta}$ will almost always be close to θ. Alternatively, if the mean squared is large then it is possible to see samples for which $\hat{\theta}$ is quite far from θ. A useful decomposition of $\text{mse}(\hat{\theta}, \theta)$ is

$$\text{mse}(\hat{\theta}, \theta) = E[(\hat{\theta} - E[\hat{\theta}])^2] + \left(E[\hat{\theta}] - \theta\right)^2 = \text{var}(\hat{\theta}) + \text{bias}(\hat{\theta}, \theta)^2$$

The derivation of this result is straightforward and is given in the appendix. The result states that for any estimator $\hat{\theta}$ of θ, $\mathrm{mse}(\hat{\theta},\theta)$ can be split into a variance component, $\mathrm{var}(\hat{\theta})$, and a bias component, $\mathrm{bias}(\hat{\theta},\theta)^2$. Clearly, $\mathrm{mse}(\hat{\theta},\theta)$ will be small only if both components are small. If an estimator is unbiased then $\mathrm{mse}(\hat{\theta},\theta) = \mathrm{var}(\hat{\theta}) = E[(\hat{\theta}-\theta)^2]$ is just the squared deviation of $\hat{\theta}$ about θ. Hence, an unbiased estimator $\hat{\theta}$ of θ is *good,* if it has a small variance.

1.3.3 Estimators for the Parameters of the CER Model

To estimate the CER model parameters μ_i, σ_i^2, σ_{ij} and ρ_{ij} we can use the plug-in principle from statistics:

Plug-in-Principle: Estimate model parameters using corresponding sample statistics

For the CER model, the plug-in principle estimates for the CER model parameters μ_i, σ_i^2, σ_{ij} and ρ_{ij} are the following sample descriptive statistics:

$$\hat{\mu}_i = \frac{1}{T}\sum_{t=1}^{T} r_{it} = \bar{r}_i, \tag{1.6}$$

$$\hat{\sigma}_i^2 = \frac{1}{T-1}\sum_{t=1}^{T}(r_{it}-\hat{\mu}_i)^2, \tag{1.7}$$

$$\hat{\sigma}_i = \sqrt{\hat{\sigma}_i^2}, \tag{1.8}$$

$$\hat{\sigma}_{ij} = \frac{1}{T-1}\sum_{t=1}^{T}(r_{it}-\hat{\mu}_i)(r_{jt}-\hat{\mu}_j), \tag{1.9}$$

$$\hat{\rho}_{ij} = \frac{\hat{\sigma}_{ij}}{\hat{\sigma}_i\hat{\sigma}_j}. \tag{1.10}$$

Example 11 *Estimating the CER model parameters for Microsoft, Starbucks and the S&P 500 index.*

To illustrate typical estimates of the CER model parameters, we use data on $T = 100$ monthly continuously compounded returns for Microsoft, Starbucks and the S & P 500 index over the period July 1992 through October 2000. These data are illustrated in Figures 1.5 and 1.6. The estimates of μ_i using (1.6) can be computed using the R functions `apply()` and `mean()`

```
> muhat.vals = apply(returns.mat,2,mean)
> muhat.vals
   sbux    msft   sp500
0.02777 0.02756 0.01253
```

giving $\hat{\mu}_{sbux} = 0.02777$, $\hat{\mu}_{msft} = 0.02756$ and $\hat{\mu}_{sp500} = 0.01253$. The expected return estimates for Microsoft and Starbucks are very similar at about 2.8% per month, whereas the S&P 500 expected return estimate is smaller at only 1.25% per month. The estimates of the parameters σ_i^2, σ_i, using (**??**) and (**??**) can be computed using `apply()`, `var()` and `sd()`

```
> sigma2hat.vals = apply(returns.mat,2,var)
> sigma2hat.vals
    sbux     msft    sp500
0.018459 0.011411 0.001432
> sigmahat.vals = apply(returns.mat,2,sd)
> sigmahat.vals
   sbux    msft   sp500
0.13586 0.10682 0.03785
```

giving

$$
\begin{aligned}
\hat{\sigma}^2_{sbux} &= 0.0185,\ \hat{\sigma}_{sbux} = 0.1359, \\
\hat{\sigma}^2_{msft} &= 0.0114,\ \hat{\sigma}_{msft} = 0.1068, \\
\hat{\sigma}^2_{sp500} &= 0.0014,\ \hat{\sigma}_{sp500} = 0.0379.
\end{aligned}
$$

Starbucks has the most variable monthly returns, and the S&P 500 index has the smallest. The scatterplots of the returns are illustrated in figure **??**. All returns appear to be positively related. The pairs (MSFT, SP500) and (SBUX, SP500) appear to be the most correlated. The estimates of σ_{ij} and ρ_{ij} using (1.9) and (1.10) can be computed using the functions `var()` and `cor()`

```
> cov.mat
          sbux     msft    sp500
sbux  0.018459 0.004031 0.002158
msft  0.004031 0.011411 0.002244
sp500 0.002158 0.002244 0.001432
> cor.mat = cor(returns.z)
```

```
> cor.mat
        sbux   msft  sp500
sbux  1.0000 0.2777 0.4198
msft  0.2777 1.0000 0.5551
sp500 0.4198 0.5551 1.0000
```

Notice when `var()` and `cor()` are given a matrix of returns they return the estimated variance matrix and the estimated correlation matrix, respectively. To extract the unique pairwise values use

```
> covhat.vals = cov.mat[lower.tri(cov.mat)]
> rhohat.vals = cor.mat[lower.tri(cor.mat)]
> names(covhat.vals) <- names(rhohat.vals) <-
+ c("sbux,msft","sbux,sp500","msft,sp500")
> covhat.vals
 sbux,msft sbux,sp500 msft,sp500
  0.004031   0.002158   0.002244
> rhohat.vals
 sbux,msft sbux,sp500 msft,sp500
    0.2777     0.4198     0.5551
```

The pairwise covariances and correlations are

$$\begin{aligned}
\hat{\sigma}_{msft,sbux} &= 0.0040, \hat{\sigma}_{msft,sp500} = 0.0022, \hat{\sigma}_{sbux,sp500} = 0.0022, \\
\hat{\rho}_{msft,sbux} &= 0.2777, \hat{\rho}_{msft,sp500} = 0.5551, \hat{\rho}_{sbux,sp500} = 0.4198.
\end{aligned}$$

These estimates confirm the visual results from the scatterplot matrix in Figure 1.6.

1.4 Statistical Properties of the CER Model Estimates

To determine the statistical properties of plug-in principle estimators $\hat{\mu}_i, \hat{\sigma}_i^2, \hat{\sigma}_i, \hat{\sigma}_{ij}$ and $\hat{\rho}_{ij}$ in the CER model, we treat them as functions of the stochastic process $\{r_{i1}\}_{t=1}^{T}$ where r_{it} is assumed to be generated by the CER model (1.1) for $i = 1, \ldots, N$. We first consider the statistical properties of $\hat{\mu}_i$ because the derivations are the most straightforward. We then summarize the properties of the remaining estimators.

1.4.1 Bias of $\hat{\mu}_i$

Consider first the estimator $\hat{\mu}_i$ given by (1.6). In the CER model, $r_{it} \sim$ *iid* $N(\mu_i, \sigma_i^2)$ and since $\hat{\mu}_i$ is an average of these normal random variables, it is also normally distributed. That is, the pdf of $\hat{\mu}_i$, $f(\hat{\mu}_i)$, is a normal density. To determine the mean of this distribution we must compute $E[\hat{\mu}_i] = E[T^{-1}\sum_{t=1}^{T} r_{it}]$. Using results about the expectation of a linear combination of random variables, it follows that

$$
\begin{aligned}
E[\hat{\mu}_i] &= E\left[\frac{1}{T}\sum_{t=1}^{T} r_{it}\right] \\
&= E\left[\frac{1}{T}\sum_{t=1}^{T}(\mu_i + \varepsilon_{it})\right] \quad \text{(since } r_{it} = \mu_i + \varepsilon_{it}\text{)} \\
&= \frac{1}{T}\sum_{t=1}^{T}\mu_i + \frac{1}{T}\sum_{t=1}^{T}E[\varepsilon_{it}] \quad \text{(by the linearity of } E[\cdot]\text{)} \\
&= \frac{1}{T}\sum_{t=1}^{T}\mu_i \ \text{(since } E[\varepsilon_{it}] = 0,\ t = 1, \ldots, T\text{)} \\
&= \frac{1}{T}T \cdot \mu_i \ = \mu_i.
\end{aligned}
$$

Hence, the mean of $f(\hat{\mu}_i)$ is equal to μ_i. We have just proved that $\hat{\mu}_i$ an *unbiased estimator* for μ_i in the CER model.

1.4.2 Precision of $\hat{\mu}_i$

To determine the variance of $\hat{\mu}_i$ we must compute $\mathrm{var}(\hat{\mu}_i) = \mathrm{var}(T^{-1}\sum_{t=1}^{T} r_{it})$. Using the results about the variance of a linear combination of uncorrelated random variables, we have

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

Hence,

$$\mathrm{var}(\hat{\mu}_i) = \frac{\sigma_i^2}{T}. \tag{1.11}$$

Notice $\mathrm{var}(\hat{\mu}_i) = \mathrm{var}(r_{it})/T$, and so is much smaller than $\mathrm{var}(r_{it})$. Also, as the sample size gets larger and larger, $\mathrm{var}(\hat{\mu}_i)$ gets smaller and smaller.

Using (1.11), the standard deviation of $\hat{\mu}_i$ is

$$\mathrm{SD}(\hat{\mu}_i) = \sqrt{\mathrm{var}(\hat{\mu}_i)} = \frac{\sigma_i}{\sqrt{T}}. \tag{1.12}$$

The standard deviation of $\hat{\mu}_i$ is often called the *standard error* of $\hat{\mu}_i$ and is denoted $\mathrm{SE}(\hat{\mu}_i)$:

$$\mathrm{SE}(\hat{\mu}_i) = \mathrm{SD}(\hat{\mu}_i) = \frac{\sigma_i}{\sqrt{T}}. \tag{1.13}$$

The value of $\mathrm{SE}(\hat{\mu}_i)$ is in the same units as $\hat{\mu}_i$ and measures the precision of $\hat{\mu}_i$ as an estimate. If $\mathrm{SE}(\hat{\mu}_i)$ is small relative to $\hat{\mu}_i$ then $\hat{\mu}_i$ is a relatively precise of μ_i because $f(\hat{\mu}_i)$ will be tightly concentrated around μ_i; if $\mathrm{SE}(\hat{\mu}_i)$ is large relative to μ_i then $\hat{\mu}_i$ is a relatively imprecise estimate of μ_i because $f(\hat{\mu}_i)$ will be spread out about μ_i. Figure 1.9 illustrates these relationships

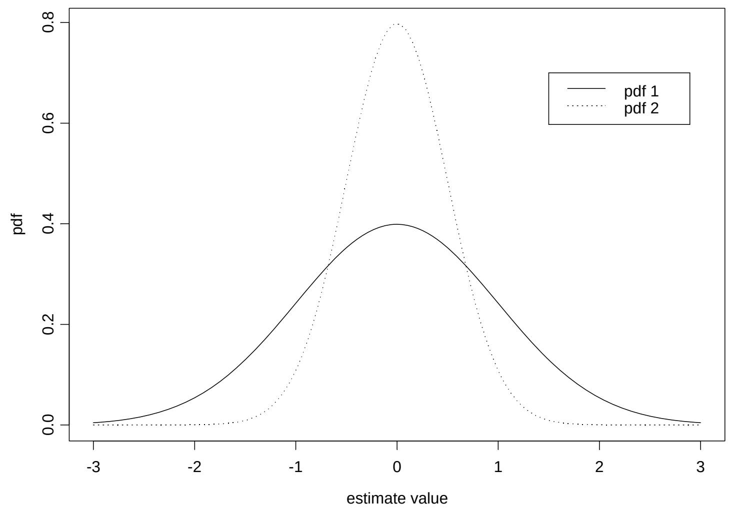

Figure 1.9: Pdfs for $\hat{\mu}_i$ with small and large values of $SE(\hat{\mu}_i)$. True value of $\mu_i = 0$.

Unfortunately, $\mathrm{SE}(\hat{\mu}_i)$ 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*

$$\widehat{\mathrm{SE}}(\hat{\mu}_i) = \sqrt{\widehat{\mathrm{var}}(\hat{\mu}_i)} = \frac{\widehat{\sigma}_i}{\sqrt{T}} \tag{1.14}$$

which is just (1.13) with the unknown value of σ_i replaced by the estimate $\widehat{\sigma}_i$ given by (1.8) .

Example 12 $\widehat{\mathrm{SE}}(\hat{\mu}_i)$ *values for Microsoft, Starbucks and the S&P 500*

For the Microsoft, Starbucks and S&P 500 estimates of $\hat{\mu}$, the values of $\widehat{\mathrm{SE}}(\hat{\mu}_i)$

using (1.14) are

$$\begin{aligned}
\widehat{\mathrm{SE}}(\hat{\mu}_{msft}) &= \frac{0.1068}{\sqrt{100}} = 0.01068 \\
\widehat{\mathrm{SE}}(\hat{\mu}_{sbux}) &= \frac{0.1359}{\sqrt{100}} = 0.01359 \\
\widehat{\mathrm{SE}}(\hat{\mu}_{sp500}) &= \frac{0.0379}{\sqrt{100}} = 0.003785
\end{aligned}$$

Clearly, the mean return μ_i is estimated more precisely for the S&P 500 index than it is for Microsoft and Starbucks. This occurs because $\hat{\sigma}_{sp500}$ is much smaller than $\hat{\sigma}_{msft}$ and $\hat{\sigma}_{sbux}$. It is useful to compare the magnitude of $\widehat{\mathrm{SE}}(\hat{\mu}_i)$ to the value of $\hat{\mu}_i$ to evaluate if $\hat{\mu}_i$ is a precise estimate:

$$\begin{aligned}
\frac{\hat{\mu}_{msft}}{\widehat{\mathrm{SE}}(\hat{\mu}_{msft})} &= \frac{0.02756}{0.01068} = 2.580, \\
\frac{\hat{\mu}_{sbux}}{\widehat{\mathrm{SE}}(\hat{\mu}_{sbux})} &= \frac{0.02777}{0.01359} = 2.044, \\
\frac{\hat{\mu}_{sp500}}{\widehat{\mathrm{SE}}(\hat{\mu}_{sp500})} &= \frac{0.01253}{0.00378} = 3.312.
\end{aligned}$$

Here we see that $\hat{\mu}_{msft}$ and $\hat{\mu}_{sbux}$ are over two times their estimated standard error values, and $\hat{\mu}_{sp500}$ is more than three times its estimated standard error value.

1.4.3 The Sampling Distribution of $\hat{\mu}_i$

Using the results that pdf of $\hat{\mu}_i$ is normal with $E[\hat{\mu}_i] = \mu_i$ and $\mathrm{var}(\hat{\mu}_i) = \frac{\sigma_i^2}{T}$ we can write

$$\hat{\mu}_i \sim N\left(\mu_i, \frac{\sigma_i^2}{T}\right). \tag{1.15}$$

The distribution for $\hat{\mu}_i$ is centered at the true value μ_i, and the spread about the average depends on the magnitude of σ_i^2, the variability of r_{it}, and the sample size, T. For a fixed sample size, T, the uncertainty in $\hat{\mu}_i$ is larger for larger values of σ_i^2. Notice that the variance of $\hat{\mu}_i$ is inversely related to the sample size T. Given σ_i^2, $\mathrm{var}(\hat{\mu}_i)$ is smaller for larger sample sizes than for smaller sample sizes. This makes sense since we expect to have a more

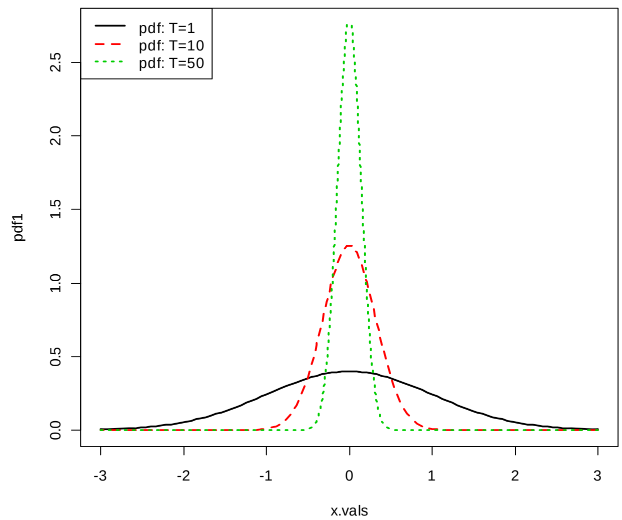

Figure 1.10: $N(0, 1/\sqrt{T})$ sampling distributions for $\hat{\mu}$ for $T = 1, 10$ and 50.

precise estimator when we have more data. If the sample size is very large (as $T \to \infty$) then $\text{var}(\hat{\mu}_i)$ will be approximately zero and the normal distribution of $\hat{\mu}_i$ given by (1.15) 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. In the language of statistics we say that $\hat{\mu}_i$ is a *consistent* estimator of μ_i as the sample size goes to infinity.

The distribution of $\hat{\mu}_i$, with $\mu_i = 0$ and $\sigma_i^2 = 1$ for various sample sizes is illustrated in figure **??**. Notice how fast the distribution collapses at $\mu_i = 0$ as T increases.

1.4.4 Confidence intervals for μ_i

The precision of $\hat{\mu}_i$ is best communicated by computing a *confidence interval* 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 interval covers μ_i. The construction of a confidence interval for μ_i is based on the following statistical result (see the appendix for details).

Result: Let $\{r_{it}\}_{t=1}^{T}$ denote a stochastic process generated from the CER model (1.1). Then the t-ratio

$$t_i = \frac{\hat{\mu}_i - \mu_i}{\widehat{\mathrm{SE}}(\hat{\mu}_i)} \sim t_{T-1},$$

where t_{T-1} denotes a Student's-t random variable with $T-1$ degrees of freedom.

The Student's-t distribution with $v > 0$ degrees of freedom is a symmetric distribution centered at zero, like the standard normal. The tail-thickness (kurtosis) of the distribution is determined by the degrees of freedom parameter v. For values of v close to zero, the tails of the Student's-t distribution are much fatter than the tails of the standard normal distribution. As v gets large, the Student's t distribution approaches the standard normal distribution.

For $\alpha \in (0,1)$, we compute a $(1-\alpha)\cdot 100\%$ confidence interval for μ_i using (**??**) and the $1-\alpha/2$ quantile (critical value) $t_{T-1}(1-\alpha/2)$ to give

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

which can be rearranged as

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

Hence, the interval

$$[\hat{\mu}_i - t_{T-1}(1-\alpha/2)\cdot\widehat{\mathrm{SE}}(\hat{\mu}_i),\ \hat{\mu}_i + t_{T-1}(1-\alpha/2)\cdot\widehat{\mathrm{SE}}(\hat{\mu}_i)] \quad (1.16)$$

$$= \hat{\mu}_i \pm t_{T-1}(1-\alpha/2)\cdot\widehat{\mathrm{SE}}(\hat{\mu}_i) \quad (1.17)$$

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

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 $T-1 = 60$ (five years of monthly return data) so that $t_{T-1}(1-\alpha/2) = t_{60}(0.975) = 2$. Then the 95% confidence for μ_i is given by

$$\hat{\mu}_i \pm 2\cdot\widehat{\mathrm{SE}}(\hat{\mu}_i). \quad (1.18)$$

The above formula for a 95% confidence interval is often used as a rule of thumb for computing an approximate 95% confidence interval for moderate sample sizes. It is easy to remember and does not require the computation of the quantile $t_{T-1}(1-\alpha/2)$ from the Student's-t distribution.

Example 13 *95% confidence intervals for μ_i for Microsoft, Starbucks and the S & P 500.*

Consider computing 95% confidence intervals for μ_i using (1.16) based on the estimated results for the Microsoft, Starbucks and S&P 500 data. The degrees of freedom for the Student's t distribution is $T-1=99$. The 97.5% quantile, $t_{99}(0.975)$, can be computed using the R function `qt()`

```
> qt(0.975, df=99)
[1] 1.984
```

Then the 95% confidence intervals are given by

$$\begin{aligned}
MSFT &: 0.02756 \pm 1.984 \cdot 0.01068 = [0.0064, 0.0488] \\
SBUX &: 0.02777 \pm 1.984 \cdot 0.01359 = [0.0008, 0.0547] \\
SP500 &: 0.01253 \pm 1.984 \cdot 0.003785 = [0.0050, 0.0200]
\end{aligned}$$

With probability 0.95, the above intervals will contain the true mean values assuming the CER model is valid. The 95% confidence intervals for MSFT and SBUX are fairly wide. The widths are almost 5%, with lower limits near 0 and upper limits near 5%. In contrast, the 95% confidence interval for SP500 is about half the width of the MSFT or SBUX confidence interval. The lower limit is near 0.5% and the upper limit is near 2%. This clearly shows that the mean return for SP500 is estimated much more precisely than the mean return for MSFT or SBUX.

1.4.5 Interpreting $E[\hat{\mu}_i]$, $\mathrm{SE}(\hat{\mu}_i)$ and Confidence Intervals Using Monte Carlo Simulation

The exact meaning of unbiased, $E[\mu_i] = \mu_i$, the interpretation of $\mathrm{SE}(\hat{\mu}_i)$ as a measure of precision, and the interpretation of the coverage probability of a confidence interval can be a bit hard to grasp at first. Strictly speaking, $E[\hat{\mu}_i] = \mu_i$ means that over an infinite number of repeated samples of $\{r_{it}\}_{t=1}^{T}$ the average of the $\hat{\mu}_i$ values computed over the infinite samples is equal to the true value μ_i. The value of $\mathrm{SE}(\hat{\mu}_i)$ represents the standard deviation of these $\hat{\mu}_i$ values, and the 95% confidence interval for μ_i will actually contain μ_i in 95% of the samples. We can think of these hypothetical samples as different Monte Carlo simulations of the CER model. In this way we can approximate the computations involved in evaluating $E[\hat{\mu}_i]$, $SE(\hat{\mu}_i)$ and the

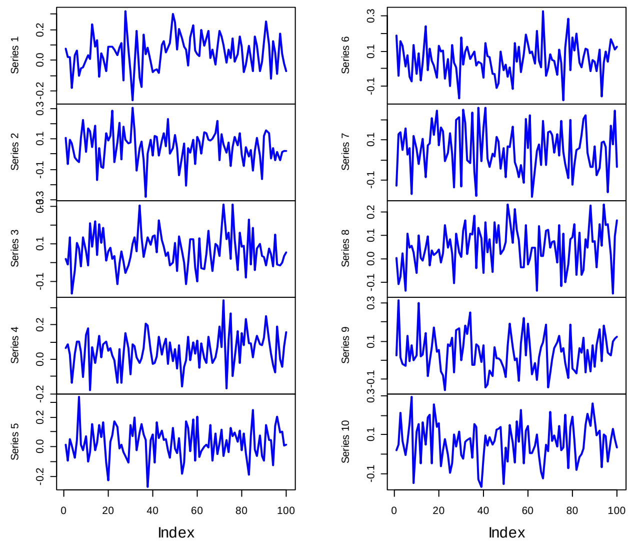

Figure 1.11: Ten simulated samples of size $T = 50$ from the CER model $r_t = 0.05 + \varepsilon_t$, $\varepsilon_t \sim GWN(0, (0.10)^2)$.

coverage probability of a confidence interval using a large but finite number of Monte Carlo simulations.

To illustrate, consider the CER model

$$\begin{aligned} r_t &= 0.05 + \varepsilon_t, t = 1, \ldots, 100 \\ \varepsilon_t &\sim GWN(0, (0.10)^2) \end{aligned} \tag{1.19}$$

Using Monte Carlo simulation, we can simulate $N = 1000$ samples of size $T = 100$ from (1.19) giving the sample realizations $\{\tilde{r}_t^j\}_{t=1}^{50}$ for $j = 1, \ldots, 1000$. The first 10 of these sample realizations are illustrated in Figure 1.11. Notice that there is considerable variation in the simulated samples but that all of the simulated samples fluctuate about the true mean value of $\mu = 0.05$. For each of the 1000 simulated samples the estimate $\hat{\mu}$ is formed giving 1000 mean estimates $\{\hat{\mu}^1, \ldots, \hat{\mu}^{1000}\}$. A histogram of these 1000 mean values is illustrated in figure **??**. The histogram of the estimated means, $\hat{\mu}^j$, can be thought of as an estimate of the underlying pdf, $f(\hat{\mu})$, of the estimator $\hat{\mu}$ which we know is a normal pdf centered at $E[\hat{\mu}] = \mu = 0.05$ with

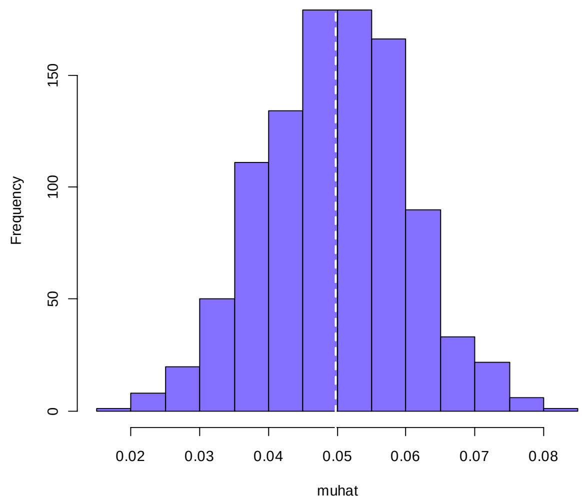

Figure 1.12: Distribution of μ computed from 1000 Monte Carlo simulations from the CER model (1.19).

$\mathrm{SE}(\hat{\mu}_i) = \frac{0.10}{\sqrt{100}} = 0.01$. Notice that the center of the histogram is very close to the true mean value $\mu = 0.05$. That is, on average over the 1000 Monte Carlo samples the value of $\hat{\mu}$ is about 0.05. In some samples, the estimate is too big and in some samples the estimate is too small but *on average* the estimate is correct. In fact, the average value of $\{\hat{\mu}^1, \ldots, \hat{\mu}^{1000}\}$ from the 1000 simulated samples is

$$\overline{\hat{\mu}} = \frac{1}{1000}\sum_{j=1}^{1000} \hat{\mu}^j = 0.04969,$$

which is very close to the true value 0.05. If the number of simulated samples is allowed to go to infinity then the sample average $\overline{\hat{\mu}}$ will be exactly equal to $\mu = 0.05$:

$$\lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} \hat{\mu}^j = \mu = 0.05.$$

The typical size of the spread about the center of the histogram represents $\mathrm{SE}(\hat{\mu}_i)$ and gives an indication of the precision of $\hat{\mu}_i$.The value of $\mathrm{SE}(\hat{\mu}_i)$ may be approximated by computing the sample standard deviation of the 1000 $\hat{\mu}^j$ values

$$\sqrt{\frac{1}{999}\sum_{j=1}^{1000}(\hat{\mu}^j - 0.04969)^2} = 0.01041$$

Notice that this value is very close to $\mathrm{SE}(\hat{\mu}_i) = \frac{0.10}{\sqrt{100}} = 0.01$. If the number of simulated sample goes to infinity then

$$\lim_{N\to\infty}\sqrt{\frac{1}{N-1}\sum_{j=1}^{N}(\hat{\mu}^j - \frac{1}{N}\sum_{j=1}^{N}\hat{\mu}^j)^2} = \mathrm{SE}(\hat{\mu}_i) = 0.10.$$

Example 14 *Monte Carlo simulation using R*

The R code to perform the Monte Carlo simulation presented in this section is

```
> mu = 0.05
> sd = 0.10
> n.obs = 100
> n.sim = 1000
> set.seed(111)
> sim.means = rep(0,n.sim)# initialize vectors
> for (sim in 1:n.sim) {
+ sim.ret = rnorm(n.obs,mean=mu,sd=sd)
+ sim.means[sim] = mean(sim.ret)
+ }
> mean(sim.means)
[1] 0.04969
> sd(sim.means)
[1] 0.01041
```

■

1.4.6 Statistical properties of the estimators of $\sigma_i^2, \sigma_i, \sigma_{ij}$ and ρ_{ij}.

To determine the statistical properties of $\hat{\sigma}_i^2$, $\hat{\sigma}_i$, $\hat{\sigma}_{ij}$ and $\hat{\rho}_{ij}$ we need to treat them as a functions of the random variables $\{r_{it}\}_{t=1}^{T}$.

Bias

Assuming that returns are generated by the CER model (1.1), the sample variances and covariances are unbiased estimators,

$$\begin{aligned} E[\hat{\sigma}_i^2] &= \sigma_i^2, \\ E[\hat{\sigma}_{ij}] &= \sigma_{ij}, \end{aligned}$$

but the sample standard deviations and correlations are biased estimators,

$$\begin{aligned} E[\hat{\sigma}_i] &\neq \sigma_i, \\ E[\hat{\rho}_{ij}] &\neq \rho_{ij}. \end{aligned}$$

The proofs of these results are beyond the scope of this book. However, they may be easily be evaluated using Monte Carlo methods.

Precision

The derivations of the variances of $\hat{\sigma}_i^2, \hat{\sigma}_i$, $\hat{\sigma}_{ij}$ and $\hat{\rho}_{ij}$ are complicated and the exact results are extremely messy and hard to work with. However, there are simple approximate formulas for the variances of $\hat{\sigma}_i^2, \hat{\sigma}_i$ and $\hat{\rho}_{ij}$ that are valid if the sample size, T, is reasonably large [4]. These large sample approximate formulas are given by

$$\begin{aligned} \mathrm{SE}(\hat{\sigma}_i^2) &\approx \frac{\sigma_i^2}{\sqrt{T/2}}, \\ \mathrm{SE}(\hat{\sigma}_i) &\approx \frac{\sigma_i}{\sqrt{2T}}, \\ \mathrm{SE}(\rho_{ij}) &\approx \frac{(1-\rho_{ij}^2)}{\sqrt{T}}, \end{aligned}$$

where "≈" denotes approximately equal. The approximations are such that the approximation error goes to zero as the sample size T gets very large. As with the formula for the standard error of the sample mean, the formulas for the standard errors above are inversely related to the square root of the sample size. Interestingly, $\mathrm{SE}(\hat{\sigma}_i)$ goes to zero the fastest and $\mathrm{SE}(\hat{\sigma}_i^2)$ goes to zero the slowest. Hence, for a fixed sample size, it appears that σ_i is generally

[4] The large sample approximate formula for the variance of $\hat{\sigma}_{ij}$ is too messy to work with so we omit it here.

estimated more precisely than σ_i^2 and ρ_{ij}, and ρ_{ij} is estimated generally more precisely than σ_i^2.

The above formulas are not practically useful, however, because they depend on the unknown quantities σ_i^2, σ_i and ρ_{ij}. Practically useful formulas replace σ_i^2, σ_i and ρ_{ij} by the estimates $\hat{\sigma}_i^2, \hat{\sigma}_i$ and $\hat{\rho}_{ij}$ and give rise to the *estimated standard errors:*

$$\begin{aligned}
\widehat{\mathrm{SE}}(\hat{\sigma}_i^2) &\approx \frac{\hat{\sigma}_i^2}{\sqrt{T/2}}, \\
\widehat{\mathrm{SE}}(\hat{\sigma}_i) &\approx \frac{\hat{\sigma}_i}{\sqrt{2T}}, \\
\widehat{\mathrm{SE}}(\rho_{ij}) &\approx \frac{(1-\hat{\rho}_{ij}^2)}{\sqrt{T}}.
\end{aligned}$$

Example 15 *Computing $\widehat{\mathrm{SE}}(\hat{\sigma}_i^2)$, $\widehat{\mathrm{SE}}(\hat{\sigma}_i)$, and $\widehat{\mathrm{SE}}(\hat{\rho}_i)$ for Microsoft, Starbucks and the S & P 500.*

For the Microsoft, Starbucks and S&P 500 return data, the values of $\widehat{\mathrm{SE}}(\hat{\sigma}_i^2)$, $\widehat{\mathrm{SE}}(\hat{\sigma}_i)$ are

$$\begin{aligned}
\widehat{\mathrm{SE}}(\hat{\sigma}_{msft}^2) &= \frac{(0.1068)^2}{\sqrt{100/2}} = 0.00161,\ \widehat{\mathrm{SE}}(\hat{\sigma}_{msft}) = \frac{0.1068}{\sqrt{2\cdot 100}} = 0.00755 \\
\widehat{\mathrm{SE}}(\hat{\sigma}_{sbux}^2) &= \frac{(0.1359)^2}{\sqrt{100/2}} = 0.00261,\ \widehat{\mathrm{SE}}(\hat{\sigma}_{sbux}) = \frac{0.1359}{\sqrt{2\cdot 100}} = 0.00961 \\
\widehat{\mathrm{SE}}(\hat{\sigma}_{sp500}^2) &= \frac{(0.0379)^2}{\sqrt{100/2}} = 0.00020,\ \widehat{\mathrm{SE}}(\hat{\sigma}_{sp500}) = \frac{0.0379}{\sqrt{2\cdot 100}} = 0.00268
\end{aligned}$$

The values of $\widehat{\mathrm{SE}}(\hat{\rho}_i)$ are

$$\begin{aligned}
\widehat{\mathrm{SE}}(\hat{\rho}_{msft,sbux}) &= \frac{1-(0.2777)^2}{\sqrt{100}} = 0.09229 \\
\widehat{\mathrm{SE}}(\hat{\rho}_{msft,sp500}) &= \frac{1-(0.5551)^2}{\sqrt{100}} = 0.06919 \\
\widehat{\mathrm{SE}}(\hat{\rho}_{sbux,sp500}) &= \frac{1-(0.4198)^2}{\sqrt{100}} = 0.08238
\end{aligned}$$

Sampling distribution

To be completed

Confidence Intervals for σ_i^2, σ_i and ρ_{ij}

Approximate 95% confidence intervals for σ_i^2, σ_i and ρ_{ij} are give by

$$\begin{aligned}
\hat{\sigma}_i^2 \pm 2 \cdot \widehat{\mathrm{SE}}(\hat{\sigma}_i^2) &= \hat{\sigma}_i^2 \pm 2 \cdot \frac{\hat{\sigma}_i^2}{\sqrt{T/2}}, \\
\hat{\sigma}_i \pm 2 \cdot \widehat{\mathrm{SE}}(\hat{\sigma}_i) &= \hat{\sigma}_i \pm 2 \cdot \frac{\hat{\sigma}_i}{\sqrt{2T}} \\
\hat{\rho}_{ij} \pm 2 \cdot \widehat{\mathrm{SE}}(\hat{\rho}_{ij}) &= \hat{\rho}_{ij} \pm 2 \cdot \frac{(1-\hat{\rho}_{ij}^2)}{\sqrt{T}}
\end{aligned}$$

Example 16 *95% confidence intervals for σ_i^2, σ_i and ρ_{ij} for Microsoft, Starbucks and the S&P 500.*

For the Microsoft, Starbucks and S&P 500 return data the approximate 95% confidence intervals for σ_i^2,are

$$\begin{aligned}
MSFT &: 0.01141 \pm 2 \cdot (0.00161) = [0.00818, 0.01464] \\
SBUX &: 0.01846 \pm 2 \cdot (0.00261) = [0.01324, 0.02368] \\
SP500 &: 0.00143 \pm 2 \cdot (0.00020) = [0.00103, 0.00184]
\end{aligned}$$

The approximate 95% confidence intervals for σ_i are

$$\begin{aligned}
MSFT &: 0.1068 \pm 2 \cdot (0.007554) = [0.09172, 0.1219] \\
SBUX &: 0.1359 \pm 2 \cdot (0.009607) = [0.11665, 0.1551] \\
SP500 &: 0.03785 \pm 2 \cdot (0.002676) = [0.03249, 0.0432]
\end{aligned}$$

The approximate 95% confidence intervals for ρ_{ij} are

$$\begin{aligned}
MSFT, SBUX &: 0.2777 \pm 2 \cdot (0.09229) = [0.09314, 0.4623] \\
MSFT, SP500 &: 0.5551 \pm 2 \cdot (0.06919) = [0.41674, 0.6935] \\
SBUX, P500 &: 0.4198 \pm 2 \cdot (0.08238) = [0.25499, 0.5845]
\end{aligned}$$

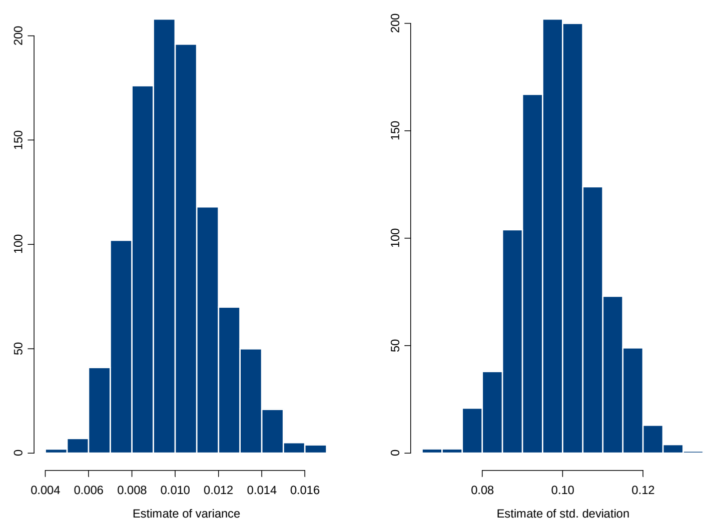

Figure 1.13: Histograms of $\hat{\sigma}^2$ and $\hat{\sigma}$ computed from $N = 1000$ Monte Carlo samples from CER model.

Evaluating the Statistical Properties of $\hat{\sigma}_i^2$ and $\hat{\sigma}_i$ by Monte Carlo simulation

We can evaluate the statistical properties of $\hat{\sigma}_i^2$ and $\hat{\sigma}_i$by Monte Carlo simulation in the same way that we evaluated the statistical properties of $\hat{\mu}_i$. Consider first the variability estimates $\hat{\sigma}_i^2$ and $\hat{\sigma}_i$. We use the simulation model (1.19) and $N = 1000$ simulated samples of size $T = 50$ to compute the estimates $\{\left(\hat{\sigma}^2\right)^1, \ldots, \left(\hat{\sigma}^2\right)^{1000}\}$ and $\{\hat{\sigma}^1, \ldots, \hat{\sigma}^{1000}\}$. The histograms of these values are displayed in figure 1.13.The histogram for the $\hat{\sigma}^2$ values is bell-shaped and slightly right skewed but is centered very close to $0.010 = \sigma^2$. The histogram for the $\hat{\sigma}$ values is more symmetric and is centered near $0.10 = \sigma$.

The average values of σ^2 and σ from the 1000 simulations are

$$\begin{aligned}\frac{1}{1000}\sum_{j=1}^{1000}\hat{\sigma}^2 &= 0.009952 \\ \frac{1}{1000}\sum_{j=1}^{1000}\hat{\sigma} &= 0.09928\end{aligned}$$

The sample standard deviation values of the Monte Carlo estimates of σ^2 and σ give approximations to $SD(\hat{\sigma}^2)$ and $SD(\hat{\sigma})$.

Evaluating the Statistical Properties of $\hat{\sigma}_{ij}$ and $\hat{\rho}_{ij}$ by Monte Carlo simulation

1.5 Further Reading

To be completed

1.6 Appendix

1.6.1 Proofs of Some Technical Results

Result: $\text{mse}(\hat{\theta},\theta) = E[(\hat{\theta}-E[\hat{\theta}])^2] + \left(E[\hat{\theta}]-\theta\right)^2 = \text{var}(\hat{\theta}) + \text{bias}(\hat{\theta},\theta)^2$

Proof. Recall, $\text{mse}(\hat{\theta},\theta) = E[(\hat{\theta}-\theta)^2]$. Write

$$\hat{\theta}-\theta = \hat{\theta}-E[\hat{\theta}]+E[\hat{\theta}]-\theta.$$

Then

$$(\hat{\theta}-\theta)^2 = \left(\hat{\theta}-E[\hat{\theta}]\right)^2 + 2\left(\hat{\theta}-E[\hat{\theta}]\right)\left(E[\hat{\theta}]-\theta\right) + \left(E[\hat{\theta}]-\theta\right)^2.$$

Taking expectations of both sides gives

$$\begin{aligned}\text{mse}(\hat{\theta},\theta) &= E\left[\left(\hat{\theta}-E[\hat{\theta}]\right)\right]^2 + E\left[\left(E[\hat{\theta}]-\theta\right)^2\right] \\ &= \text{var}(\hat{\theta}) + \text{bias}(\hat{\theta},\theta)^2.\end{aligned}$$

■

1.6.2 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,

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

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 pdf of X is illustrated in Figure xxx for various values of T. Notice that X is only allowed to take non-negative values. The pdf is highly right skewed for small values of T and becomes symmetric as T gets large. The mean of the distribution is

$$E[X] = E[Z_1^2] + E[Z_2^2] + \cdots E[Z_T^2] = T,$$

since $E[Z_i^2] = \mathrm{var}(Z_i) = 1$.

1.6.3 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, $X \sim \chi_T^2$. 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's-t. Figure xxx shows the pdf of t for various values of the degrees of freedom T. Notice that the pdf is symmetric about zero and has a bell shape like the normal. The tail thickness of the pdf 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 pdf of the Student's t converges to the pdf of the normal.

The Student's-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's-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

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

1.7 Problems

To be completed

Bibliography

[1] Campbell, Lo and MacKinley (1998). *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ.