Chapter 1

Estimation of The CER Model

Date: February 17, 2015

The CER model of asset returns presented in the previous chapter gives us a simple framework for interpreting the time series behavior of asset returns and prices. At the beginning of time $t-1$, $\mathbf{R}_t$ is an $N \times 1$ random vector representing the returns (simple or continuously compounded) on assets $i = 1, \ldots, N$ to be realized at time t. The CER model states that $\mathbf{R}_t \sim iid$ $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Our best guess for the return at t on asset i 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 multivariate normal distribution characterizing returns is the same for all time periods t.

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 returns on N different assets over the sample $t = 1, \ldots, T$. Denote this sample $\{\mathbf{r}_1, \ldots, \mathbf{r}_T\} = \{\mathbf{r}_t\}_{t=1}^T$, where $\mathbf{r}_t = (\mathbf{r}_{1t}, \ldots, \mathbf{r}_{Nt})'$ is the $N \times 1$ vector of returns on N assets observed at time t. It is assumed that the observed returns are realizations of the random variables $\{\mathbf{R}_t\}_{t=1}^T$, where $\mathbf{R}_t = (R_{1t}, \ldots, R_{Nt})'$ is a vector of N asset returns described by the CER model

$$\mathbf{R}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t, \;\; \boldsymbol{\varepsilon}_t \sim iid\ N(\mathbf{0}, \boldsymbol{\Sigma}). \tag{1.1}$$

Under these assumptions, we can use the observed returns $\{\mathbf{r}_t\}_{t=1}^T$ to estimate the unknown parameters in $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ of the CER model. However, before we describe the estimation of the CER model in detail, it is necessary to review some fundamental concepts in the statistical theory of estimation.

In this chapter we use the R packages **PerformanceAnalytics**, **tseries** and **zoo**. Make sure these packages are installed and loaded prior to running the examples in the chapter.

1.1 Estimators and Estimates

Let R_t be the return on a single asset described by the CER model and let θ denote some characteristic (parameter) of the CER model we are interested in estimating. For simplicity, assume that $\theta \in \mathbb{R}$ is a single parameter. For example, if we are interested in the expected return on the asset, then $\theta = \mu$; if we are interested in the variance of asset i returns, then $\theta = \sigma^2$; if we are interested in the first lag autocorrelation then $\theta = \rho_1$. The goal is to estimate θ based on a sample of size T of the observed data.

Definition 1 *Let $\{R_1, \ldots, R_T\}$ denote a collection of T random returns from the CER model, and let θ denote some characteristic of the model. An estimator of θ, denoted $\hat{\theta}$, is a rule or algorithm for estimating θ as a function of the random variables $\{R_1, \ldots, R_T\}$. Here, $\hat{\theta}$ is a random variable.*

Definition 2 *Let $\{r_1, \ldots, r_T\}$ denote an observed sample of size T from the CER model, and let θ denote some characteristic of the model. An estimate of θ, denoted $\hat{\theta}$, is simply the value of the estimator for θ based on the observed sample $\{r_1, \ldots, r_T\}$. Here, $\hat{\theta}$ is a number.*

Example 3 *The sample average as an estimator and an estimate*

Let R_t be the return on a single asset described by the CER model, and suppose we are interested in estimating $\theta = \mu = E[R_t]$ from the sample of observed returns $\{r_t\}_{t=1}^T$. The sample average $\hat{\mu} = \frac{1}{T}\sum_{t=1}^T r_t$ is an algorithm for computing an estimate of the expected return μ. Before the sample is observed, we can think of $\hat{\mu} = \frac{1}{T}\sum_{t=1}^T R_t$ as a simple linear function of the random variables $\{R_t\}_{t=1}^T$ and so is itself a random variable. After the sample is observed, the sample average can be evaluated using the observed data $\{r_t\}_{t=1}^T$ which produces the estimate of μ. 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 using the sample average is

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

■

The example above illustrates the distinction between an estimator and an estimate of a parameter θ. However, typically in the statistics literature we use the same symbol, $\hat{\theta}$, to denote both an estimator and an estimate. When $\hat{\theta}$ is treated as a function of the random returns it denotes an estimator and is a random variable. When $\hat{\theta}$ is evaluated using the observed data it denotes an estimate and is simply a number. The context in which we discuss $\hat{\theta}$ will determine how to interpret it.

1.1.1 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 $\{R_t\}_{t=1}^T$. The exact form of $f(\hat{\theta})$ may be very complicated. Sometimes we can use analytical calculations to determine the exact form of $f(\hat{\theta})$. In general, the exact form of $f(\hat{\theta})$ is often too difficult to derive exactly. When $f(\hat{\theta})$ is too difficult to compute we can often approximate $f(\hat{\theta})$ using either Monte Carlo simulation techniques or the *Central Limit Theorem* (CLT). In Monte Carlo simulation, we use the computer the simulate many different realizations of the random returns $\{R_t\}_{t=1}^T$ and on each simulated sample we evaluate the estimator $\hat{\theta}$. The Monte Carlo approximation of $f(\hat{\theta})$ is the empirical distribution $\hat{\theta}$ over the different simulated samples. For a given sample size T, Monte Carlo simulation gives a very accurate approximation to $f(\hat{\theta})$ if the number of simulated samples is very large. The CLT approximation of $f(\hat{\theta})$ is a normal distribution approximation that becomes more accurate as the sample size T gets very large. An advantage of the CLT approximation is that it is often easy to compute. The disadvantage is that the accuracy of the approximation depends on the estimator $\hat{\theta}$ and sample size T.

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 $\{R_t\}_{t=1}^T$ 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 too 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 a *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 4 *The estimation error is the difference between the estimator and the parameter being estimated:*

$$\text{error}(\hat{\theta}, \theta) = \hat{\theta} - \theta. \tag{1.2}$$

Definition 5 *The bias of an estimator $\hat{\theta}$ of θ is the expected estimation error:*

$$\text{bias}(\hat{\theta}, \theta) = E[\text{error}(\hat{\theta}, \theta)] = E[\hat{\theta}] - \theta. \tag{1.3}$$

Definition 6 *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 $\{R_t\}_{t=1}^T$. 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 two estimators of θ, $\hat{\theta}_1$ and $\hat{\theta}_2$. The pdfs of $\hat{\theta}_1$ and $\hat{\theta}_2$ are illustrated in Figure 1.1. $\hat{\theta}_1$ is an unbiased estimator of θ with a large variance, and $\hat{\theta}_2$ is a biased estimator of θ with a small variance. Consider first, the pdf of $\hat{\theta}_1$. 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

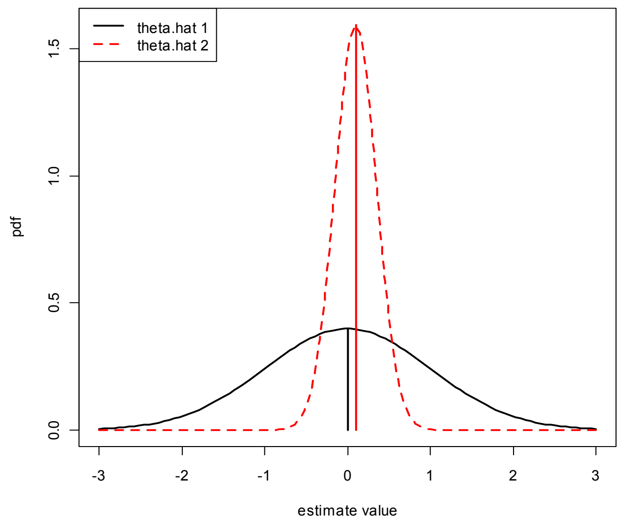

Figure 1.1: Distributions of competiting estimators for $\theta = 0$. $\hat{\theta}_1$ is unbiased but has high variance, and $\hat{\theta}_2$ is biased but has low variance.

consider the pdf for $\hat{\theta}_2$. The center of the pdf is slightly higher than $\theta = 0$, i.e., 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 $\hat{\theta}_1$.

While unbiasedness is a desirable property of an estimator $\hat{\theta}$ of θ, it by itself, is not enough determine if $\hat{\theta}$ is a good estimator. Being correct on average means that $\hat{\theta}$ is seldom exactly correct for any given sample. In some samples $\hat{\theta}$ is less than θ, and some samples $\hat{\theta}$ is greater than θ. In addition we need to know how far $\hat{\theta}$ typically is from θ. That is, we need to know about the magnitude of the spread of the distribution of $\hat{\theta}$ about its average value. This will tell us the precision of $\hat{\theta}$.

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

$$\mathrm{mse}(\hat{\theta},\theta) = E[(\hat{\theta}-\theta)^2] = E[\mathrm{error}(\hat{\theta},\theta)^2] \tag{1.4}$$

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 $\mathrm{mse}(\hat{\theta},\theta)$ is

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

The derivation of this result is straightforward. 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

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

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.

The $\text{mse}(\hat{\theta}, \theta)$ and $\text{var}(\hat{\theta})$ are based on squared deviations and so are not in the same units of measurement as θ. Measures of precision that are in the same units as θ are the *root mean square error*

$$\text{rmse}(\hat{\theta}, \theta) = \sqrt{\text{mse}(\hat{\theta}, \theta)}, \tag{1.5}$$

and the *standard error*

$$\text{se}(\hat{\theta}) = \sqrt{\text{var}(\hat{\theta})}. \tag{1.6}$$

If $\text{bias}(\hat{\theta}, \theta) \approx 0$ then the precision of $\hat{\theta}$ is typically measured by $\text{se}(\hat{\theta})$.

Good Estimators

With the concepts of bias and precision in hand, we can state what defines a good estimator.

Definition 8 *A good estimator $\hat{\theta}$ of θ has a small bias (1.3) and a small standard error (1.6).*

1.1.2 Asymptotic Properties of Estimators

Estimator bias and precision are finite sample properties. That is, they are properties that hold for a fixed sample size T. Very often we are also interested in properties of estimators when the sample size T gets very large. For example, analytic calculations may show that the bias and mse of an estimator $\hat{\theta}$ depend on T in a decreasing way. That is, as T gets very large the bias and mse approach zero. So for a very large sample, $\hat{\theta}$ is effectively unbiased with high precision. In this case we say that $\hat{\theta}$ is a *consistent* estimator of θ. In addition, for large samples the CLT says that $f(\hat{\theta})$ can often be well approximated by a normal distribution. In this case, we say that $\hat{\theta}$ is asymptotically normally distributed. The word "asymptotic" means "in an infinitely large sample" or "as the sample size T goes to infinity". Of course, in the real world we don't have an infinitely large sample and so the asymptic results are only approximations. How good these approximation are for a given sample size T depends on the context. Monte Carlo simulations can often be used to evaluate asymptotic approximations in a given context.

Consistency

Let $\hat{\theta}$ be an estimator of θ based on the random returns $\{R_t\}_{t=1}^T$.

Definition 9 *$\hat{\theta}$ is consistent for θ (converges in probability to θ) if for any $\varepsilon > 0$*

$$\lim_{T\to\infty} \Pr(|\hat{\theta} - \theta| > \varepsilon) = 0$$

Intuitively, consistency says that as we get enough data then $\hat{\theta}$ will eventually equal θ. In other words, if we have enough data then we know the truth.

Theorems in probability theory known as *Laws of Large Numbers* are used to determine if an estimator is consistent or not. In general, we have the following result: an estimator $\hat{\theta}$ is consistent for θ if

- $\text{bias}(\hat{\theta}, \theta) = 0$ as $T \to \infty$
- $\text{se}\left(\hat{\theta}\right) = 0$ as $T \to \infty$

Equivalently, $\hat{\theta}$ is consistent for θ if $\text{mse}(\hat{\theta}, \theta) \to 0$ as $T \to \infty$. Intuitively, if $f(\hat{\theta})$ collapses to θ as $T \to \infty$ then $\hat{\theta}$ is consistent for θ.

Asymptotic Normality

Let $\hat{\theta}$ be an estimator of θ based on the random returns $\{R_t\}_{t=1}^T$.

Definition 10 *An estimator $\hat{\theta}$ is asymptotically normally distributed if*

$$\hat{\theta} \sim N(\theta, \text{se}(\hat{\theta})^2) \tag{1.7}$$

for large enough T.

Asymptotic normality means that $f(\hat{\theta})$ is well approximated by a normal distribution with mean θ and variance $\text{se}(\hat{\theta})^2$. The justification for asymptotic normality comes from the famous *Central Limit Theorem*.

Central Limit Theorem There are actually many versions of the CLT with different assumptions.[1] In its simplist form, the CLT says that the sample average of a collection of iid random variables $X_1, \ldots, X_T$ with $E[X_i] = \mu$ and $var(X_i) = \sigma^2$ is asymptotically normal with mean μ and variance σ^2/T. In particular, the CDF of the standardized sample mean

$$\frac{\bar{X} - \mu}{\text{se}(\bar{X})} = \frac{\bar{X} - \mu}{\sigma/\sqrt{T}} = \sqrt{T}\left(\frac{\bar{X} - \mu}{\sigma}\right),$$

converges to the CDF of a standard normal random variable Z as $T \to \infty$. This result can be expressed as

$$\sqrt{T}\left(\frac{\bar{X} - \mu}{\sigma}\right) \sim Z \sim N(0, 1),$$

for large enough T. Equivalently,

$$\bar{X} \sim \mu + \frac{\sigma}{\sqrt{T}} \times Z \sim N\left(\mu, \frac{\sigma^2}{T}\right) = N\left(\mu, \text{se}(\bar{X})^2\right),$$

for large enough T. This form shows that $\bar{X}$ is asymptotically normal with mean μ and variance σ^2/T.

Asymptotic Confidence Intervals

For an asymptotically normal estimator $\hat{\theta}$ of θ, the precision of $\hat{\theta}$ is measure by $\widehat{\text{se}}(\hat{\theta})$ but is best communicated by computing a (asymptotic) *confidence interval* for the unknown value of θ. A confidence interval is an *interval estimate* of θ such that we can put an explicit probability statement about the likelihood that the interval covers θ.

The construction of an asymptotic confidence interval for θ uses the asymptotic normality result

$$\frac{\hat{\theta} - \theta}{\text{se}(\hat{\theta})} = Z \sim N(0, 1). \tag{1.8}$$

Then, for $\alpha \in (0, 1)$, we compute a $(1 - \alpha) \cdot 100\%$ confidence interval for θ using (1.8) and the $1 - \alpha/2$ standard normal quantile (critical value) $q^Z_{(1-\alpha/2)}$ to give

$$\Pr\left(-q^Z_{(1-\alpha/2)} \leq \frac{\hat{\theta} - \theta}{\text{se}(\hat{\theta})} \leq q^Z_{(1-\alpha/2)}\right) = 1 - \alpha,$$

[1] White (1984) gives a comprehensive discussion of CLTs useful in econometrics.

which can be rearranged as

$$\Pr\left(\hat{\theta} - q^Z_{(1-\alpha/2)} \cdot \mathrm{se}(\hat{\theta}) \leq \mu_i \leq \hat{\theta} + q^Z_{(1-\alpha/2)} \cdot \mathrm{se}(\hat{\theta})\right) = 1 - \alpha.$$

Hence, the interval

$$[\hat{\theta} - q^Z_{(1-\alpha/2)} \cdot \mathrm{se}(\hat{\theta}),\ \hat{\theta} + q^Z_{(1-\alpha/2)} \cdot \mathrm{se}(\hat{\theta})] = \hat{\theta} \pm q^Z_{(1-\alpha/2)} \cdot \mathrm{se}(\hat{\theta}) \tag{1.9}$$

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

In practice, typical values for α are 0.05 and 0.01 for which $q^Z_{(0.975)} = 1.96$ and $q^Z_{(0.995)} = 2.58$. Then, approximate 95% and 99% asymptotic confidence intervals for θ have the form $\hat{\theta} \pm 2 \cdot \mathrm{se}(\hat{\theta})$ and $\hat{\theta} \pm 2.5 \cdot \mathrm{se}(\hat{\theta})$, respectively.

1.2 Estimators for the Parameters of the CER Model

Let $\{\mathbf{r}_t\}_{t=1}^T$denote a sample of size T of observed returns on N assets from the CER model (1.1). To estimate the unknown CER model parameters μ_i, σ_i^2, σ_{ij} and ρ_{ij} from $\{\mathbf{r}_t\}_{t=1}^T$ we can use the *plug-in principle* from statistics:

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

For the CER model parameters, the plug-in principle estimates are the following sample descriptive statistics discussed in Chapter xxx (Descriptive Statistics):

$$\hat{\mu}_i = \frac{1}{T}\sum_{t=1}^T r_{it}, \tag{1.10}$$

$$\hat{\sigma}_i^2 = \frac{1}{T-1}\sum_{t=1}^T (r_{it} - \hat{\mu}_i)^2, \tag{1.11}$$

$$\hat{\sigma}_i = \sqrt{\hat{\sigma}_i^2}, \tag{1.12}$$

$$\hat{\sigma}_{ij} = \frac{1}{T-1}\sum_{t=1}^T (r_{it} - \hat{\mu}_i)(r_{jt} - \hat{\mu}_j), \tag{1.13}$$

$$\hat{\rho}_{ij} = \frac{\hat{\sigma}_{ij}}{\hat{\sigma}_i \hat{\sigma}_j}. \tag{1.14}$$

The plug-in principle is appropriate because the the CER model parameters $\mu_i, \sigma_i^2, \sigma_{ij}$ and ρ_{ij} are characteristics of the underlying distribution of returns that are naturally estimated using sample statistics.

The plug-in principle sample statistics (1.10) - (1.14) are given for a single asset and the statistics (1.13) - (1.14) are given for one pair of assets. However, these statistics can be computed for a collection of N assets using the matrix sample statistics

$$\underset{(N\times 1)}{\hat{\boldsymbol{\mu}}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{r}_t = \begin{pmatrix} \hat{\mu}_1 \\ \vdots \\ \hat{\mu}_N \end{pmatrix}, \tag{1.15}$$

$$\underset{(N\times N)}{\hat{\boldsymbol{\Sigma}}} = \frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{r}_t - \hat{\boldsymbol{\mu}})(\mathbf{r}_t - \hat{\boldsymbol{\mu}})' = \begin{pmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1N} \\ \hat{\sigma}_{12} & \hat{\sigma}_2^2 & \cdots & \hat{\sigma}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{1N} & \hat{\sigma}_{2N} & \cdots & \hat{\sigma}_N^2 \end{pmatrix}. \tag{1.16}$$

Here $\hat{\boldsymbol{\mu}}$ is called the sample mean vector and $\hat{\boldsymbol{\Sigma}}$ is called the sample covariance matrix. The sample variances are the diagonal elements of $\hat{\boldsymbol{\Sigma}}$ and the sample covariances are the off diagonal elements of $\hat{\boldsymbol{\Sigma}}$. To get the sample correlations, define the $N \times N$ diagonal matrix

$$\hat{\mathbf{D}} = \begin{pmatrix} \hat{\sigma}_1 & 0 & \cdots & 0 \\ 0 & \hat{\sigma}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\sigma}_N \end{pmatrix}.$$

Then the sample correlation matrix $\hat{\mathbf{R}}$ is computed as

$$\hat{\mathbf{R}} = \hat{\mathbf{D}}^{-1}\hat{\boldsymbol{\Sigma}}\hat{\mathbf{D}}^{-1} = \begin{pmatrix} 1 & \hat{\rho}_{12} & \cdots & \hat{\rho}_{1N} \\ \hat{\rho}_{12} & 1 & \cdots & \hat{\rho}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{1N} & \hat{\rho}_{2N} & \cdots & 1 \end{pmatrix}. \tag{1.17}$$

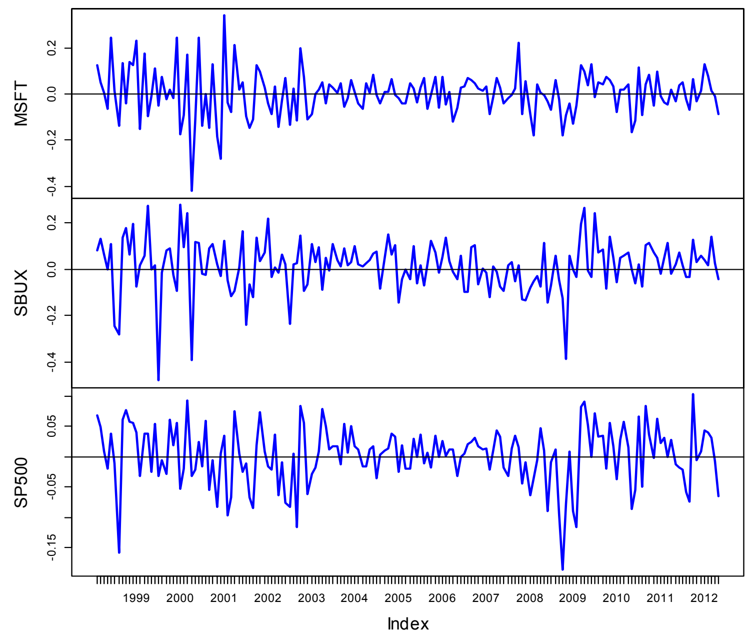

Figure 1.2: Monthly cc returns on Microsoft stock, Starbucks stock, and the S&P 500 index, over the period January 1998 through May 2012.

Here, the sample correlations are the off diagonal elements of $\hat{\mathbf{R}}$.

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 monthly continuously compounded returns for Microsoft, Starbucks and the S & P 500 index over the period January 1998 through May 2012. The data is the same as that used in Chapters xxx (Descriptive Statistics & CER Model) and is retrieved from Yahoo! using the **tseries** function `get.hist.quote()` as follows

```
> msftPrices = get.hist.quote(instrument="msft", start="1998-01-01",
+                             end="2012-05-31", quote="AdjClose",
+                             provider="yahoo", origin="1970-01-01",
```

```
+                              compression="m", retclass="zoo")
> sbuxPrices = get.hist.quote(instrument="sbux", start="1998-01-01",
+                              end="2012-05-31", quote="AdjClose",
+                              provider="yahoo", origin="1970-01-01",
+                              compression="m", retclass="zoo")
> sp500Prices = get.hist.quote(instrument="^gspc", start="1998-01-01",
+                               end="2012-05-31", quote="AdjClose",
+                               provider="yahoo", origin="1970-01-01",
+                               compression="m", retclass="zoo")
> colnames(msftPrices) = "MSFT"
> colnames(sbuxPrices) = "SBUX"
> colnames(sp500Prices) = "SP500"
> index(msftPrices) = as.yearmon(index(msftPrices))
> index(sbuxPrices) = as.yearmon(index(sbuxPrices))
> index(sp500Prices) = as.yearmon(index(sp500Prices))
> cerPrices = merge(msftPrices, sbuxPrices, sp500Prices)
> msftRetS = Return.calculate(msftPrices, method="simple")
> sbuxRetS = Return.calculate(sbuxPrices, method="simple")
> sp500RetS = Return.calculate(sp500Prices, method="simple")
> cerRetS = Return.calculate(cerPrices, method="simple")
> msftRetS = msftRetS[-1]
> sbuxRetS = sbuxRetS[-1]
> sp500RetS = sp500RetS[-1]
> cerRetS = cerRetS[-1]
> msftRetC = log(1 + msftRetS)
> sbuxRetC = log(1 + sbuxRetS)
> sp500RetC = log(1 + sp500RetS)
> cerRetC = merge(msftRetC, sbuxRetC, sp500RetC)
> colnames(cerRetC) = c("MSFT", "SBUX", "SP500")
```

These data are illustrated in Figures 1.2 and 1.3.

The estimates of μ_i $(i = msft, sbux, sp500)$ using (1.10) or (1.15) can be computed using the R functions `apply()` and `mean()`

```
> muhat = apply(cerRetC,2,mean)
> muhat
    MSFT     SBUX    SP500
0.004127 0.014657 0.001687
```

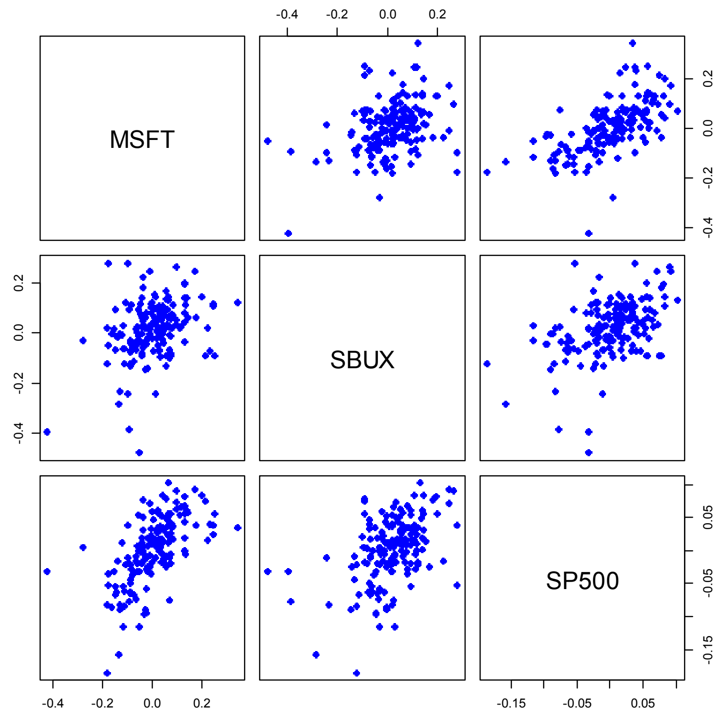

Figure 1.3: Scatterplot matrix of the monthly cc returns on Microsoft stock, Starbucks stock, and the S&P 500 index.

Starbucks has the highest average monthly return at 1.5% and the S&P 500 index has the lowest at 0.2%.

The estimates of the parameters σ_i^2, σ_i, using (1.11) and (1.12) can be computed using apply(), var() and sd()

```
> sigma2hat = apply(cerRetC,2,var)
> sigma2hat
    MSFT     SBUX    SP500
0.010051 0.012465 0.002349
> sigmahat = apply(cerRetC,2,sd)
> sigmahat
   MSFT    SBUX   SP500
0.10026 0.11164 0.04847
```

Starbucks has the most variable monthly returns at 11%, and the S&P 500 index has the smallest at 5%.

The scatterplots of the returns are illustrated in Figure 1.3. All returns appear to be positively related. The covariance and correlation matrix estimates using (1.16) and (1.17) can be computed using the functions `var()` (or `cov()`) and `cor()`

```
> covmat = var(cerRetC)
> covmat
          MSFT     SBUX    SP500
MSFT  0.010051 0.003819 0.003000
SBUX  0.003819 0.012465 0.002476
SP500 0.003000 0.002476 0.002349
> cormat = cor(cerRetC)
> cormat
        MSFT   SBUX  SP500
MSFT  1.0000 0.3412 0.6173
SBUX  0.3412 1.0000 0.4575
SP500 0.6173 0.4575 1.0000
```

To extract the unique pairwise values of σ_{ij} and ρ_{ij} from the matrix objects `covmat` and `cormat` use

```
> covhat = covmat[lower.tri(covmat)]
> rhohat = cormat[lower.tri(cormat)]
> names(covhat) <- names(rhohat) <-
+ c("msft,sbux","msft,sp500","sbux,sp500")
> covhat
 msft,sbux msft,sp500 sbux,sp500
  0.003819   0.003000   0.002476
> rhohat
 msft,sbux msft,sp500 sbux,sp500
    0.3412     0.6173     0.4575
```

The pairs (MSFT, SP500) and (SBUX, SP500) are the most correlated. These estimates confirm the visual results from the scatterplot matrix in Figure 1.3. ■

1.3 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 random variables $\{\mathbf{R}_t\}_{t=1}^T$ where $\mathbf{R}_t$ is assumed to be generated by the CER model (1.1).

1.3.1 Bias

Assuming that returns are generated by the CER model (1.1), $\hat{\mu}_i$, $\hat{\sigma}_i^2$ and $\hat{\sigma}_{ij}$ are unbiased estimators,

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

but $\hat{\sigma}_i$ and $\hat{\rho}_{ij}$ are biased estimators,

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

It can be shown that the biases in $\hat{\sigma}_i$ and $\hat{\rho}_{ij}$, are very small and decreasing in T such that $\text{bias}(\hat{\sigma}_i, \sigma_i) = \text{bias}(\hat{\rho}_{ij}, \rho_{ij}) = 0$ as $T \to \infty$. The proofs of these results are beyond the scope of this book and may be found, for example, in Goldberger (1991). As we shall see, these results about bias can be easily verified using Monte Carlo methods.

It is instructive to illustrate how to derive the result $E[\hat{\mu}_i] = \mu_i$. Using results about the expectation of a linear combination of random variables, it follows that

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

The derivation of the results $E[\hat{\sigma}_i^2] = \sigma_i^2$ and $E[\hat{\sigma}_{ij}] = \sigma_{ij}$ are similar but are considerably more involved and so are omitted.

1.3.2 Precision

Because the CER model estimators are either unbiased or the bias is very small, the precision of these estimators is measured by their standard errors. The standard error for $\hat{\mu}_i$, $\mathrm{se}(\hat{\mu}_i)$, can be calculated exactly and is given by

$$
\mathrm{se}(\hat{\mu}_i) = \frac{\sigma_i}{\sqrt{T}}. \tag{1.18}
$$

The derivation of this result is straightforward. Using the results about the variance of a linear combination of uncorrelated random variables, we have

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

Then $\mathrm{se}(\hat{\mu}_i) = \mathrm{SD}(\hat{\mu}_i) = \frac{\sigma_i}{\sqrt{T}}$. We make the following remarks:

1. 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.

2. The magnitude of $\mathrm{se}(\hat{\mu}_i)$ depends positively on the volatility of returns, $\sigma_i = \mathrm{SD}(R_{it})$. For a given sample size T, assets with higher return volatility have larger values of $\mathrm{se}(\hat{\mu}_i)$ than assets with lower return volatility. In other words, estimates of expected return for high volatility assets are less precise than estimates of expected returns for low volatility assets.

3. For a given return volatility σ_i, $\mathrm{se}(\hat{\mu}_i)$ is smaller for larger sample sizes T. In other words, $\hat{\mu}_i$ is more precisely estimated for larger samples. Moreover, $\mathrm{se}(\hat{\mu}_i) \to 0$ as $T \to \infty$ at rate $\sqrt{T}$

The derivations of the standard errors for $\hat{\sigma}_i^2, \hat{\sigma}_i$, $\hat{\sigma}_{ij}$ 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 standard errors of $\hat{\sigma}_i^2, \hat{\sigma}_i$ and $\hat{\rho}_{ij}$ based on the CLT that are valid if the sample size, T, is reasonably large.[2] These large sample approximate formulas are given by

$$\mathrm{se}(\hat{\sigma}_i^2) \approx \frac{\sqrt{2}\sigma_i^2}{\sqrt{T}} = \frac{\sigma_i^2}{\sqrt{T/2}}, \tag{1.19}$$

$$\mathrm{se}(\hat{\sigma}_i) \approx \frac{\sigma_i}{\sqrt{2T}}, \tag{1.20}$$

$$\mathrm{se}(\rho_{ij}) \approx \frac{(1-\rho_{ij}^2)}{\sqrt{T}}, \tag{1.21}$$

where "$\approx$" denotes approximately equal. The approximations are such that the approximation error goes to zero as the sample size T gets very large. We make the following remarks:

1. As with the formula for the standard error of the sample mean, the formulas for $\mathrm{se}(\hat{\sigma}_i^2)$ and $\mathrm{se}(\hat{\sigma}_i)$ depend on σ_i^2. Larger values of σ_i^2 imply less precise estimates of $\hat{\sigma}_i^2$ and $\hat{\sigma}_i$.

2. The formula for $\mathrm{se}(\rho_{ij})$, does not depend on σ_i^2 but rather depends on ρ_{ij}^2 and is smaller the closer ρ_{ij}^2 is to unity. Intuitively, this makes sense because as ρ_{ij}^2 approaches one the linear dependence between R_{it} and R_{jt} becomes almost perfect and this will be easily recognizable in the data (scatterplot will almost follow a straight line).

3. The formulas for the standard errors above are inversely related to the square root of the sample size, $\sqrt{T}$, which means that larger sample sizes imply smaller values of the standard errors.

4. 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, these formulas suggest that σ_i is generally estimated more precisely than σ_i^2 and ρ_{ij}, and ρ_{ij} is estimated generally more precisely than σ_i^2.

[2] The large sample approximate formula for the variance of $\hat{\sigma}_{ij}$ is too messy to work with so we omit it here. In practice, we can use the bootstrap to provide an estimated standard error for $\hat{\sigma}_{ij}$.

The above formulas (1.19) - (1.21) 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:*

$$\widehat{\mathrm{se}}(\hat{\mu}_i) = \frac{\widehat{\sigma}_i}{\sqrt{T}} \tag{1.22}$$

$$\widehat{\mathrm{se}}(\hat{\sigma}_i^2) \approx \frac{\hat{\sigma}_i^2}{\sqrt{T/2}}, \tag{1.23}$$

$$\widehat{\mathrm{se}}(\hat{\sigma}_i) \approx \frac{\hat{\sigma}_i}{\sqrt{2T}}, \tag{1.24}$$

$$\widehat{\mathrm{se}}(\rho_{ij}) \approx \frac{(1-\hat{\rho}_{ij}^2)}{\sqrt{T}}. \tag{1.25}$$

It is good practice to report estimates together with their estimated standard errors. In this way the precision of the estimates is transparent to the user. Typically, estimates are reported in a table with the estimates in one column and the estimated standard errors in an adjacent column.

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

For Microsoft, Starbucks and S&P 500, the values of $\widehat{\mathrm{se}}(\hat{\mu}_i)$ are easily computed in R using

```
> n.obs = nrow(cerRetC)
> seMuhat = sigmahat/sqrt(n.obs)
```

The values of $\hat{\mu}_i$ and $\widehat{\mathrm{se}}(\hat{\mu}_i)$ shown together are

```
> cbind(muhat, seMuhat)
        muhat seMuhat
MSFT  0.00413 0.00764
SBUX  0.01466 0.00851
SP500 0.00169 0.00370
```

For Microsoft and Starbucks, the values of $\widehat{\mathrm{se}}(\hat{\mu}_i)$ are similar because the values of $\hat{\sigma}_i$ are similar, and $\widehat{\mathrm{se}}(\hat{\mu}_i)$ is smallest for the S&P 500 index. This occurs because $\hat{\sigma}_{sp500}$ is much smaller than $\hat{\sigma}_{msft}$ and $\hat{\sigma}_{sbux}$. Hence, $\hat{\mu}_i$ is estimated more precisely for the S&P 500 index (a highly diversified portfolio) than it is for Microsoft and Starbucks stock (individual assets).

It is tempting 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

```
> seMuhat/muhat
 MSFT  SBUX SP500
1.852 0.581 2.191
```

Here we see that $\widehat{\text{se}}(\hat{\mu}_{msft})$ and $\widehat{\text{se}}(\hat{\mu}_{sp500})$ are about twice as large as $\hat{\mu}_{msft}$ and $\hat{\mu}_{sp500}$, respectively, whereas $\widehat{\text{se}}(\hat{\mu}_{sbux})$ is about half the size of $\hat{\mu}_{sbux}$. This seems to indicate that $\hat{\mu}_{sbux}$ is most precisely estimated. However, this comparison is misleading. To see why, consider the range of values determined by $\hat{\mu}_i \pm 2 \times \widehat{\text{se}}(\hat{\mu}_i)$.

```
> upper = muhat + 2*seMuhat
> lower = muhat - 2*seMuhat
> width = upper - lower
> cbind(lower, upper, width)
         lower   upper  width
MSFT  -0.01116 0.01942 0.0306
SBUX  -0.00237 0.03168 0.0341
SP500 -0.00570 0.00908 0.0148
```

For normally distributed $\hat{\mu}_i$, this range contains the true value of μ_i with probability around 0.95. For all assets, this range contains both positive and negative values but the range is smallest for the S&P 500 index. ■

Example 13 *Computing* $\widehat{\text{se}}(\hat{\sigma}_i^2)$, $\widehat{\text{se}}(\hat{\sigma}_i)$, *and* $\widehat{\text{se}}(\hat{\rho}_{ij})$ *for Microsoft, Starbucks and the S&P 500.*

For Microsoft, Starbucks and S&P 500, the values of $\widehat{\text{se}}(\hat{\sigma}_i^2)$, $\widehat{\text{se}}(\hat{\sigma}_i)$ and $\widehat{\text{se}}(\hat{\rho}_{ij})$ (together with the estimates $\hat{\sigma}_i^2$, $\hat{\sigma}_i$ and $\hat{\rho}_{ij}$) are

```
> seSigma2hat = sigma2hat/sqrt(n.obs/2)
> seSigmahat = sigmahat/sqrt(2*n.obs)
> cbind(sigma2hat, seSigma2hat, sigmahat, seSigmahat)
      sigma2hat seSigma2hat sigmahat seSigmahat
MSFT    0.01005    0.001084   0.1003    0.00541
SBUX    0.01246    0.001344   0.1116    0.00602
SP500   0.00235    0.000253   0.0485    0.00261
```

Notice that σ^2 and σ for the S&P 500 index are estimated much more precisely than the values for Microsoft and Starbucks. Also notice that σ_i is estimated more precisely than μ_i for all assets: the values of $\widehat{\text{se}}(\hat{\sigma}_i)$ relative to $\hat{\sigma}_i$ are much smaller than the values of $\widehat{\text{se}}(\hat{\mu}_i)$ to $\hat{\mu}_i$.

The values of $\widehat{\text{se}}(\hat{\rho}_{ij})$ (together with $\hat{\rho}_{ij}$) are

```
> seRhohat = (1-rhohat^2)/sqrt(n.obs)
> cbind(rhohat, seRhohat)
           rhohat seRhohat
msft,sbux   0.341   0.0674
msft,sp500  0.617   0.0472
sbux,sp500  0.457   0.0603
```

The values of $\widehat{\mathrm{se}}(\hat{\rho}_{ij})$ are moderate in size (relative to $\hat{\rho}_{ij}$). Notice that $\hat{\rho}_{sbux,sp500}$ has the smallest estimated standard error because $\hat{\rho}^2_{sbux,sp500}$ is closest to one. ■

1.3.3 Sampling Distributions and Confidence Intervals

Sampling Distribution for $\hat{\mu}_i$

In the CER model, $R_{it} \sim iid\ N(\mu_i, \sigma_i^2)$ and since $\hat{\mu}_i = \frac{1}{T}\sum_{t=1}^{T} R_{it}$ is an average of these normal random variables, it is also normally distributed. The mean of $\hat{\mu}_i$ is μ_i and its variance is $\frac{\sigma_i^2}{T}$. Therefore, the exact probability distribution of $\hat{\mu}_i$, $f(\hat{\mu}_i)$, for a fixed sample size T is the normal distribution $\hat{\mu}_i \sim N\left(\mu_i, \frac{\sigma_i^2}{T}\right)$ where

$$f(\hat{\mu}_i) = \left(\frac{2\pi\sigma_i^2}{T}\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma_i^2/T}(\hat{\mu}_i - \mu_i)^2\right\}. \tag{1.26}$$

The probability curve $f(\hat{\mu}_i)$ is centered at the true value μ_i, and the spread about μ_i 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 precise estimator when we have more data. If the sample size is very large (as $T \to \infty$) then $\mathrm{var}(\hat{\mu}_i)$ will be approximately zero and the normal distribution of $\hat{\mu}_i$ given by (1.26) 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, we have established that $\hat{\mu}_i$ is a *consistent* estimator of μ_i as the sample size goes to infinity.

Example 14 *Sampling distribution of $\hat{\mu}$ with different sample sizes.*

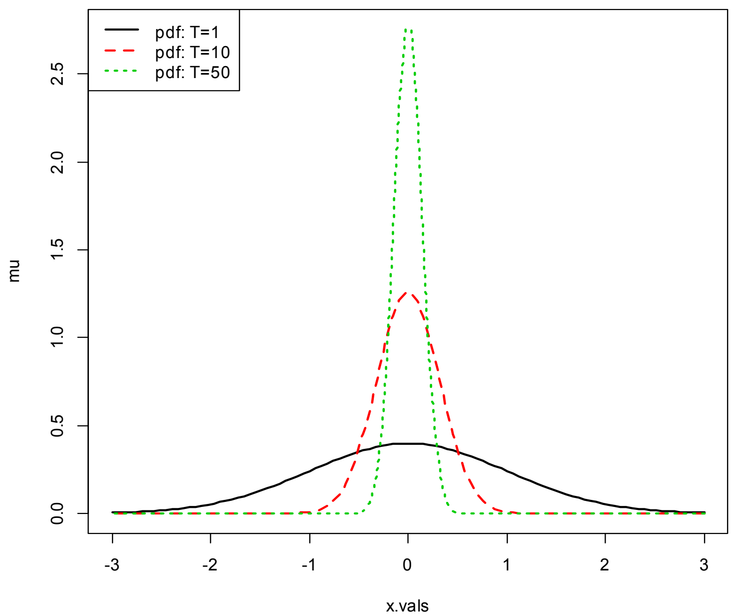

Figure 1.4: $N(0, 1/T)$ sampling distributions for $\hat{\mu}$ for $T = 1, 10$ and 50.

The distribution of $\hat{\mu}_i$, with $\mu_i = 0$ and $\sigma_i^2 = 1$ for various sample sizes is illustrated in figure 1.4. Notice how fast the distribution collapses at $\mu_i = 0$ as T increases.

Confidence intervals for μ_i

The precision of $\hat{\mu}_i$ is measure by $\widehat{\mathrm{se}}(\hat{\mu}_i)$ but 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 an exact confidence interval for μ_i is based on the following statistical result (see the appendix for details).

Result: Let $\{R_{it}\}_{t=1}^T$ be generated from the CER model (1.1). Define the t-ratio as

$$t_i = \frac{\hat{\mu}_i - \mu_i}{\widehat{\mathrm{se}}(\hat{\mu}_i)} = \frac{\hat{\mu}_i - \mu_i}{\hat{\sigma}/\sqrt{T}}, \tag{1.27}$$

Then $t_i \sim t_{T-1}$ 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 (1.27) 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

$$\begin{aligned} & [\hat{\mu}_i - t_{T-1}(1-\alpha/2)\cdot\widehat{\mathrm{se}}(\hat{\mu}_i),\ \hat{\mu}_i + t_{T-1}(1-\alpha/2)\cdot\widehat{\mathrm{se}}(\hat{\mu}_i)] \\ = {} & \hat{\mu}_i \pm t_{T-1}(1-\alpha/2)\cdot\widehat{\mathrm{se}}(\hat{\mu}_i) \end{aligned} \tag{1.28}$$

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

Example 15 *Computing 95% confidence intervals for μ_i*

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 $T-1 = 60$ (e.g., five years of monthly return data) so that $t_{T-1}(1-\alpha/2) = t_{60}(0.975) = 2$. This can be verified in R using the function `qt()`

Then the 95% confidence for μ_i is given by

$$\hat{\mu}_i \pm 2 \cdot \widehat{\mathrm{se}}(\hat{\mu}_i). \tag{1.29}$$

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. It is also an approximate 95% confidence interval that is based the asymptotic normality of $\hat{\mu}_i$. Recall, for a normal distribution with mean μ and variance σ^2 approximately 95% of the probability lies between $\mu \pm 2\sigma$. ■

The coverage probability associated with the confidence interval for μ_i is based on the fact that the estimator $\hat{\mu}_i$ is a random variable. Since confidence interval is constructed as $\hat{\mu}_i \pm t_{T-1}(1-\alpha/2) \cdot \widehat{\mathrm{se}}(\hat{\mu}_i)$ it is also a random variable. An intuitive way to think about the coverage probability associated with the confidence interval is to think about the game of "horseshoes".[3] The horse shoe is the confidence interval and the parameter μ_i is the post at which the horse shoe is tossed. Think of playing game 100 times (i.e, simulate 100 samples of the CER model). If the thrower is 95% accurate (if the coverage probability is 0.95) then 95 of the 100 tosses should ring the post (95 of the constructed confidence intervals should contain the true value μ_i).

Example 16 *95% confidence intervals for μ_i for Microsoft, Starbucks and the S & P 500 index.*

Consider computing 95% confidence intervals for μ_i using (1.28) 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 $T-1=171$. The 97.5% quantile, $t_{99}(0.975)$, can be computed using the R function `qt()`

```
> t.975 = qt(0.975, df=(n.obs-1))
> t.975
[1] 1.97
```

Notice that this quantile is very close to 2. Then the exact 95% confidence intervals are given by

[3] Horse shoes is a game commonly played at county fairs. See http://en.wikipedia.org/wiki/Horseshoes for a complete description of the game.

```
> lower = muhat - t.975*seMuhat
> upper = muhat + t.975*seMuhat
> width = upper - lower
> cbind(lower, upper, width)
         lower   upper  width
MSFT  -0.01096 0.01922 0.0302
SBUX  -0.00215 0.03146 0.0336
SP500 -0.00561 0.00898 0.0146
```

With probability 0.95, the above intervals will contain the true mean values assuming the CER model is valid. The 95% confidence intervals for Microsoft and Starbucks are fairly wide (about 3%) and contain both negative and positive values. The confidence interval for the S&P 500 index is tighter but also contains negative and positive values. For Microsoft, the confidence interval is $[-1.1\%, 1.9\%]$. This means that with probability 0.95, the true monthly expected return is somewhere between -1.1% and 1.9%. The economic implications of a -1.1% expected monthly return and a 1.9% expected return are vastly different. In contrast, the 95% confidence interval for the SP500 is about half the width of the intervals for Microsoft or Starbucks. The lower limit is near -0.5% and the upper limit is near 1%. This result clearly shows that the monthly mean return for the S&P 500 index is estimated much more precisely than the monthly mean returns for Microsoft or Starbucks.

Sampling distributions for $\hat{\sigma}_i^2$, $\hat{\sigma}_i$ and $\hat{\rho}_{ij}$

The exact distributions of $\hat{\sigma}_i^2$, $\hat{\sigma}_i$ and $\hat{\rho}_{ij}$ based on a fixed sample size T are difficult to derive.[4] However, approximate normal distributions of the form (1.7) based on the CLT are readily available:

$$\hat{\sigma}_i^2 \sim N\left(\sigma_i^2, \mathrm{se}(\hat{\sigma}_i^2)^2\right) = N\left(\sigma_i^2, \frac{4\sigma_i^4}{T}\right), \tag{1.30}$$

$$\hat{\sigma}_i \sim N\left(\sigma_i, \mathrm{se}(\hat{\sigma}_i)^2\right) = N\left(\sigma_i, \frac{\sigma_i^2}{2T}\right), \tag{1.31}$$

$$\hat{\rho}_{ij} \sim N\left(\rho_{ij}, \mathrm{se}(\hat{\rho}_{ij})^2\right) = N\left(\rho_{ij}, \frac{(1-\hat{\rho}_{ij}^2)^2}{T}\right). \tag{1.32}$$

[4] For example, the exact sampling distribution of $(T-1)\hat{\sigma}_i^2/\sigma_i^2$ is chi-square with $T-1$ degrees of freedom.

These appoximate normal distributions can be used to compute approximate confidence intervals for σ_i^2, σ_i and ρ_{ij}.

Approximate 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 given by

$$\hat{\sigma}_i^2 \pm 2 \cdot \widehat{\mathrm{se}}(\hat{\sigma}_i^2) = \hat{\sigma}_i^2 \pm 2 \cdot \frac{\hat{\sigma}_i^2}{\sqrt{T/2}}, \tag{1.33}$$

$$\hat{\sigma}_i \pm 2 \cdot \widehat{\mathrm{se}}(\hat{\sigma}_i) = \hat{\sigma}_i \pm 2 \cdot \frac{\hat{\sigma}_i}{\sqrt{2T}}, \tag{1.34}$$

$$\hat{\rho}_{ij} \pm 2 \cdot \widehat{\mathrm{se}}(\hat{\rho}_{ij}) = \hat{\rho}_{ij} \pm 2 \cdot \frac{(1 - \hat{\rho}_{ij}^2)}{\sqrt{T}}. \tag{1.35}$$

Example 17 *Approximate 95% confidence intervals for σ_i^2, σ_i and ρ_{ij} for Microsoft, Starbucks and the S&P 500.*

Using (1.33) - (1.34), the approximate 95% confidence intervals for σ_i^2 and σ_i ($i =$ Microsoft, Starbucks, S&P 500) are

```
> lowerSigma2 = sigma2hat - 2*seSigma2hat
> upperSigma2 = sigma2hat + 2*seSigma2hat
> widthSigma2 = upperSigma2 - lowerSigma2
> cbind(lowerSigma2, upperSigma2, widthSigma2)
      lowerSigma2 upperSigma2 widthSigma2
MSFT      0.00788     0.01222     0.00434
SBUX      0.00978     0.01515     0.00538
SP500     0.00184     0.00286     0.00101
> upperSigma = sigmahat + 2*seSigmahat
> widthSigma = upperSigma - lowerSigma
> cbind(lowerSigma, upperSigma, widthSigma)
      lowerSigma upperSigma widthSigma
MSFT      0.0894     0.1111     0.0216
SBUX      0.0996     0.1237     0.0241
SP500     0.0432     0.0537     0.0105
```

The 95% confidence intervals for σ and σ^2 are larger for Microsoft and Starbucks than for the S&P 500 index. For all assets, the intervals for σ are fairly narrow (2% for Microsoft and Starbucks and 1% for S&P 500 index) indicating that σ is precisely estimated.

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

```
> cbind(lowerRho, upperRho, widthRho)
           lowerRho upperRho widthRho
msft,sbux     0.206    0.476    0.269
msft,sp500    0.523    0.712    0.189
sbux,sp500    0.337    0.578    0.241
```

The 95% confidence intervals for ρ_{ij} are not too wide and all contain just positive values away from zero. The smallest interval is for $\rho_{msft,sp500}$ because $\hat{\rho}_{msft,sp500}$ is closest to 1.

1.4 Using Monte Carlo Simulation to Understand the Statistical Properties of Estimators

Let R_t be the return on a single asset described by the CER model, let θ denote some characteristic (parameter) of the CER model we are interested in estimating, and let $\hat{\theta}$ denote an estimator for θ based on a sample of size T. The exact meaning of estimator bias, $\text{bias}(\hat{\theta}, \theta)$, the interpretation of $\text{se}(\hat{\theta})$ as a measure of precision, the sampling distribution $f(\hat{\theta})$, and the interpretation of the coverage probability of a confidence interval for θ, can all be a bit hard to grasp at first. If $\text{bias}(\hat{\theta}, \theta) = 0$ so that $E[\hat{\theta}] = \theta$ then over an infinite number of repeated samples of $\{R_{it}\}_{t=1}^{T}$ the average of the $\hat{\theta}$ values computed over the infinite samples is equal to the true value θ. The value of $\text{se}(\hat{\theta})$ represents the standard deviation of these $\hat{\theta}$ values. The sampling distribution $f(\hat{\theta})$ is the smoothed histogram of these $\hat{\theta}$ values. And the 95% confidence intervals for θ will actually contain θ 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 $E[\hat{\theta}]$, $\text{se}(\hat{\theta})$, $f(\hat{\theta})$, and the coverage probability of a confidence interval using a large, but finite, number of Monte Carlo simulations.

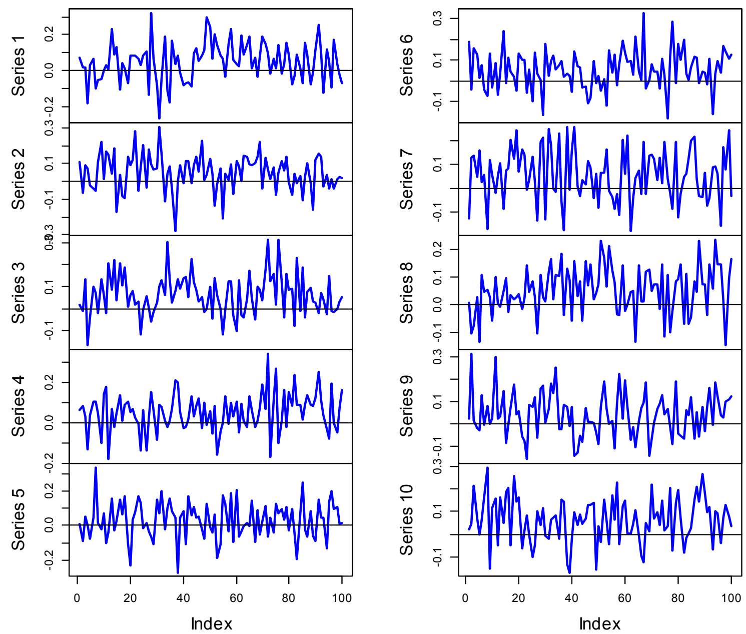

Figure 1.5: Ten simulated samples of size $T = 100$ from the CER model $R_t = 0.05 + \varepsilon_t$, $\varepsilon_t \sim iid\ N(0, (0.10)^2)$.

1.4.1 Evaluating the Statistical Properties of $\hat{\mu}$ Using Monte Carlo Simulation

Consider the CER model

$$
\begin{aligned}
R_t &= 0.05 + \varepsilon_t, t = 1, \ldots, 100 \qquad (1.36)\\
&\varepsilon_t \sim \text{GWN}(0, (0.10)^2).
\end{aligned}
$$

Here, the true parameter values are $\mu = 0.05$ and $\sigma = 0.10$. Using Monte Carlo simulation, we can simulate $N = 1000$ different samples of size $T = 100$ from (1.36) giving the sample realizations $\{r_t^j\}_{t=1}^{100}$ for $j = 1, \ldots, 1000$. The first 10 of these simulated samples are illustrated in Figure 1.5. Notice that there is considerable variation in the appearance of the simulated samples, but that all of the simulated samples fluctuate about the true mean value of $\mu = 0.05$ and have a typical deviation from the mean of about 0.10. For each of the 1000 simulated samples we can estimate $\hat{\mu}$ 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 from (1.26) is a normal pdf centered at $E[\hat{\mu}] = \mu = 0.05$ with $\mathrm{se}(\hat{\mu}_i) = \frac{0.10}{\sqrt{100}} = 0.01$. This normal curve (solid orange line) is superimposed on the histogram in Figure 1.6. Notice that the center of the histogram (white dashed vertical line) 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

$$\overline{\hat{\mu}} = \frac{1}{1000} \sum_{j=1}^{1000} \hat{\mu}^j = 0.0497,$$

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 $\mu = 0.05$:

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

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:

$$\hat{\sigma}_{\hat{\mu}} = \sqrt{\frac{1}{999} \sum_{j=1}^{1000} (\hat{\mu}^j - 0.04969)^2} = 0.0104.$$

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

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

The coverage probability of the 95% confidence interval for μ can also be illustrated using Monte Carlo simulation. For each simulation j, the interval

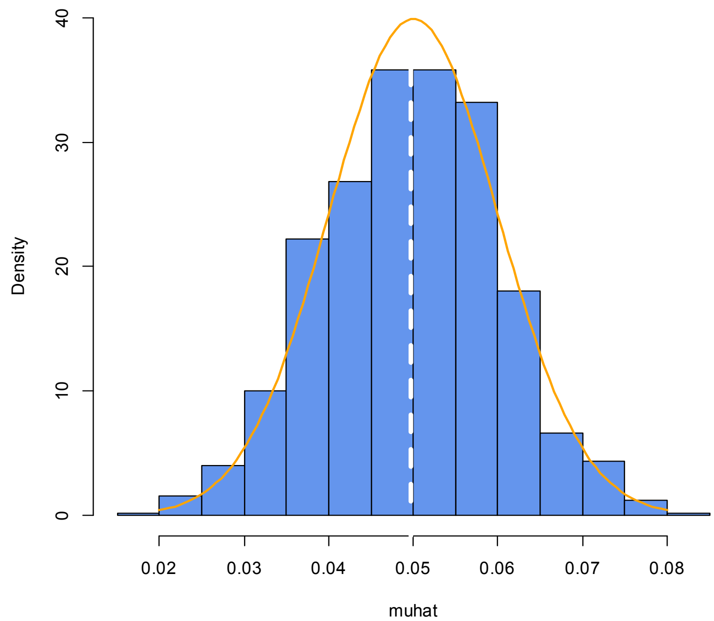

Figure 1.6: Distribution of $\hat{\mu}$ computed from 1000 Monte Carlo simulations from the CER model (1.36). White dashed line is the average of the μ values, and orange curve is the true $f(\hat{\mu})$.

$\hat{\mu} \pm t_{100}(0.975) \times \widehat{\mathrm{se}}(\hat{\mu}^j)$ is computed. The coverage probability is approximated by the fraction of intervals that contain (cover) the true $\mu = 0.05$. For the 1000 simulated samples, this fraction turns out to be 0.931. As the number of simulations goes to infinity, the Monte Carlo coverage probability will be equal to 0.95.

Example 18 *Monte Carlo simulation to evaluate $E[\hat{\mu}]$, $\mathrm{se}(\hat{\mu})$ and 95% confidence intervals for μ*

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

```
> mu = 0.05
```

```
> sigma = 0.10
> n.obs = 100
> n.sim = 1000
> set.seed(111)
> sim.means = rep(0,n.sim)
> mu.lower = rep(0,n.sim)
> mu.upper = rep(0,n.sim)
> qt.975 = qt(0.975, nobs-1)
> for (sim in 1:n.sim) {
+    sim.ret = rnorm(n.obs,mean=mu,sd=sigma)
+    sim.means[sim] = mean(sim.ret)
+    se.muhat = sd(sim.ret)/sqrt(n.obs)
+    mu.lower[sim] = sim.means[sim]-qt.975*se.muhat
+    mu.upper[sim] = sim.means[sim]+qt.975*se.muhat
+ }
```

The 1000×1 vectors `sim.means`, `mu.lower` and `mu.upper` contain the values of $\hat{\mu}^j$, $\hat{\mu}^j - t_{100}(0.975) \times \widehat{\mathrm{se}}(\hat{\mu}^j)$ and $\hat{\mu}^j + t_{100}(0.975) \times \widehat{\mathrm{se}}(\hat{\mu}^j)$ computed from each of the simulated samples $j = 1, \ldots, 1000$. The mean and standard deviation of $\{\hat{\mu}^1, \ldots, \hat{\mu}^{1000}\}$ are

```
> mean(sim.means)
[1] 0.0497
> sd(sim.means)
[1] 0.0104
```

To evaluate the converage probability of the 95% confidence intervals, we count the number of times each interval actually contains the true value of μ

```
> in.interval = mu >= mu.lower & mu <= mu.upper
> sum(in.interval)/n.sim
[1] 0.931
```

■

1.4.2 Evaluating the Statistical Properties of $\hat{\sigma}_i^2$ and $\hat{\sigma}_i$ Using 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 $\hat{\mu}_i$. We use the simulation model (1.36) and $N = 1000$ simulated samples of size $T = 100$ to compute the estimates $\{(\hat{\sigma}^2)^1, \ldots, (\hat{\sigma}^2)^{1000}\}$ and $\{\hat{\sigma}^1, \ldots, \hat{\sigma}^{1000}\}$. The histograms of these values, with the asymptotic normal distributions overlayed, are displayed in Figure 1.7. The histogram for the $\hat{\sigma}^2$ values is bell-shaped and slightly right skewed but is centered very close to $\sigma^2 = 0.010$. The histogram for the $\hat{\sigma}$ values is more symmetric and is centered near $\sigma = 0.10$. The average values of $\hat{\sigma}^2$ and $\hat{\sigma}$ from the 1000 simulations are

$$\begin{aligned}
\overline{\hat{\sigma}^2} &= \frac{1}{1000}\sum_{j=1}^{1000}(\hat{\sigma}^2)^j = 0.00999, \\
\overline{\hat{\sigma}} &= \frac{1}{1000}\sum_{j=1}^{1000}\hat{\sigma}^j = 0.0997.
\end{aligned}$$

The Monte Carlo estimate of the bias for $\hat{\sigma}^2$ is $0.00999 - 0.01 = -0.0000$, and the estimate of bias for $\hat{\sigma}$ is $0.0997 - 0.010 = -0.0003$. This confirms that $\hat{\sigma}^2$ is unbiased and that the bias in $\hat{\sigma}$ is extremely small. If the number of simulated samples, N, goes to infinity then $\overline{\hat{\sigma}^2} \to E[\hat{\sigma}^2] = \sigma^2 = 0.01$, and $\overline{\hat{\sigma}} \to E[\hat{\sigma}] = \sigma + \text{bias}(\hat{\sigma}, \sigma)$.

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

$$\begin{aligned}
\hat{\sigma}_{\hat{\sigma}^2} &= \sqrt{\frac{1}{999}\sum_{j=1}^{1000}((\hat{\sigma}^2)^j - 0.00999)^2} = 0.00135, \\
\hat{\sigma}_{\hat{\sigma}} &= \sqrt{\frac{1}{999}\sum_{j=1}^{1000}(\hat{\sigma}^j - 0.0997)^2} = 0.00676.
\end{aligned}$$

The approximate values for $\text{se}(\hat{\sigma}^2)$ and $\text{se}(\hat{\sigma})$ based on the CLT are

$$\begin{aligned}
\text{se}(\hat{\sigma}^2) &= \frac{(0.10)^2}{\sqrt{100/2}} = 0.00141, \\
\text{se}(\hat{\sigma}^2) &= \frac{0.10}{\sqrt{2 \times 100}} = 0.00707.
\end{aligned}$$

Notice that the Monte Carlo estimates of $\text{se}(\hat{\sigma}^2)$ and $\text{se}(\hat{\sigma})$ are a bit different from the CLT based estimates. The reason is that the CLT based estimates

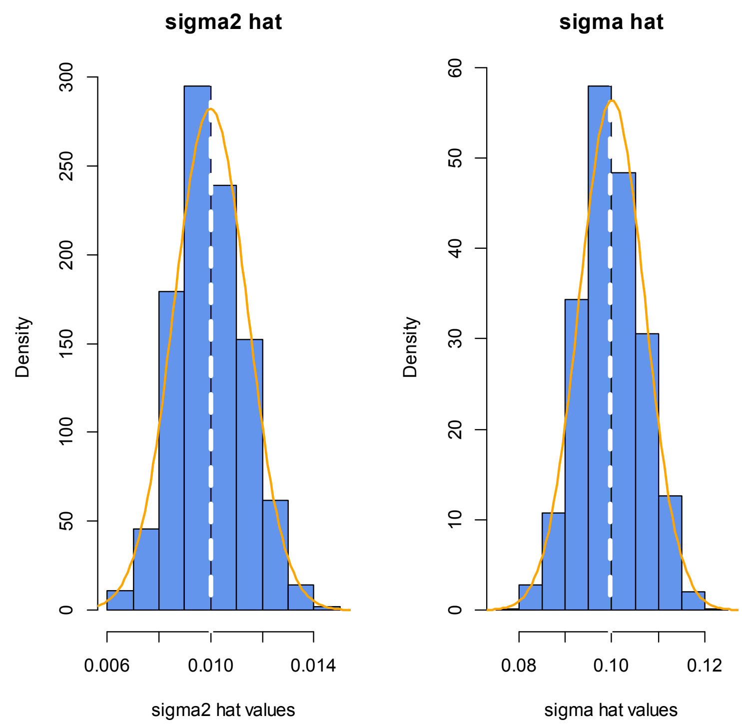

Figure 1.7: Histograms of $\hat{\sigma}^2$ and $\hat{\sigma}$ computed from $N = 1000$ Monte Carlo samples from CER model.

are approximations that hold when the sample size T is large. Because $T = 100$ is not too large, the Monte Carlo estimates of $\mathrm{se}(\hat{\sigma}^2)$ and $\mathrm{se}(\hat{\sigma})$ are likely more accurate (and will be more accurate if the number of simulations is larger).

For each simulation j, the approximate 95% confidence intervals $\left(\hat{\sigma}^2\right)^j \pm 2 \times \widehat{\mathrm{se}}(\left(\hat{\sigma}^2\right)^j)$ and $\hat{\sigma}^j \pm 2 \times \widehat{\mathrm{se}}(\hat{\sigma}^j)$ are computed. The coverage probabilities of these intervals is approximated by the fractions of intervals that contain (cover) the true values $\sigma^2 = 0.01$ and $\sigma = 0.10$, respectively. For the 1000 simulated samples, these fractions turn out to be 0.951 and 0.963, respectively. As the number of simulations and the sample size goes to infinity, the Monte Carlo coverage probability will be equal to 0.95.

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

To evaluate the statistical properties of $\hat{\rho}_{ij} = \text{cor}(R_{it}, R_{jt})$, we must simulate from the CER model in matrix form (1.1). For example, consider the bivariate CER model

$$\begin{pmatrix} R_{1t} \\ R_{2t} \end{pmatrix} = \begin{pmatrix} 0.05 \\ 0.03 \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},\ t = 1, \ldots, 100, \tag{1.37}$$

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim iid\ N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} (0.10)^2 & (0.75)(0.10)(0.05) \\ (0.75)(0.10)(0.05) & (0.05)^2 \end{pmatrix} \right), \tag{1.38}$$

where $\mu_1 = 0.05$, $\mu_2 = 0.03$, $\sigma_1 = 0.10$, $\sigma_2 = 0.05$, and $\rho_{12} = 0.75$. We use the simulation model (1.37)-(1.38) with $N = 1000$ simulated samples of size $T = 100$ to compute the estimates $\{\hat{\rho}_{12}^1, \ldots, \hat{\rho}_{12}^{1000}\}$. The histogram of these values, with the asymptotic normal distribution overlaid, is displayed in Figure 1.8. The histogram for the $\hat{\rho}_{12}$ values is bell-shaped with a mild left skewness and centered close to $\rho_{12} = 0.75$. The sample mean and standard deviation values of $\hat{\rho}_{12}$ across the 1000 simulations are, respectively,

$$\begin{aligned} \bar{\hat{\rho}}_{12} &= \frac{1}{1000} \sum_{j=1}^{1000} \hat{\rho}_{12}^j = 0.747, \\ \hat{\sigma}_{\hat{\rho}_{12}} &= \sqrt{\frac{1}{999} \sum_{j=1}^{1000} (\hat{\rho}_{12}^j - 0.747)^2} = 0.045. \end{aligned}$$

There is very slight downward bias in $\hat{\rho}_{12}$. The Monte Carlo standard deviation, $\hat{\sigma}_{\hat{\rho}_{12}}$, is very close to the approximate standard error $\text{se}(\hat{\rho}_{12}) = (1 - 0.75^2)/\sqrt{100} = 0.044$. For each simulation j, the approximate 95% confidence interval $\hat{\rho}_{12}^j \pm 2 \times \widehat{\text{se}}(\hat{\rho}_{12}^j)$ is computed. The coverage probability of this interval is approximated by the fractions of intervals that contain (cover) the true values $\rho_{12} = 0.75$. For the 1000 simulated samples, this fractions turns out to be 0.952.

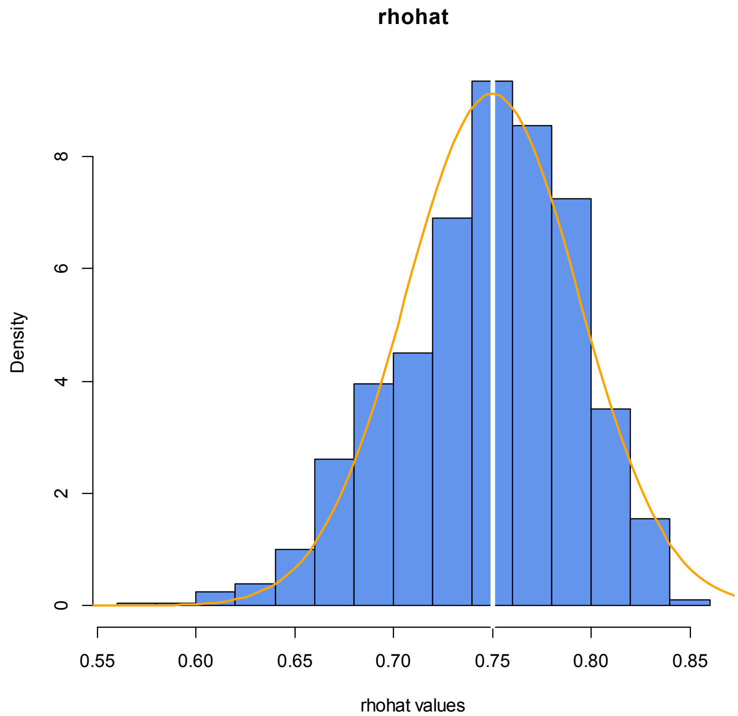

Figure 1.8: Histograms of $\hat{\rho}_{12}$ computed from $N = 1000$ Monte Carlo samples from the bivariate CER model (1.37) - (1.38).

1.4.4 Estimating Value-at-Risk in the CER Model

Consider the CER model for the simple return R_t. From the location-scale representation

$$R_t = \mu + \sigma \times Z_t,$$

the α-quantile of R_t is

$$q_\alpha^R = \mu + \sigma \times q_\alpha^Z,$$

where q_α^Z is the α-quantile of $Z_t \sim iid\ N(0,1)$. Then, for an initial investment W_0 and $\alpha \in (0,1)$ the $\alpha \cdot 100\%$ value-at-risk (VaR_α) is given by

$$\mathrm{VaR}_\alpha = W_0 \times q_\alpha^R = W_0(\mu + \sigma \times q_\alpha^Z).$$

Given the CER model estimators $\hat{\mu}$ and $\hat{\sigma}$, the plug-in principle estimator for VaR_α is

$$\widehat{\text{VaR}}_\alpha = W_0 \times \hat{q}_\alpha^R = W_0(\hat{\mu} + \hat{\sigma} \times q_\alpha^Z). \tag{1.39}$$

Here $\widehat{\text{VaR}}_\alpha$ is a linear function of $\hat{q}_\alpha^R$, which itself is a linear function of $\hat{\mu}$ and $\hat{\sigma}$. The statistical properties of $\widehat{\text{VaR}}_\alpha$ can then be easily derived from the statistical properties of $\hat{\mu}$ and $\hat{\sigma}$.

If the CER model is applied to the cc return R_t then

$$\text{VaR}_\alpha = W_0 \times \left(e^{q_\alpha^R} - 1\right) = W_0 \left(e^{\mu + \sigma \times q_\alpha^Z} - 1\right),$$

and the estimate of VaR_α is

$$\widehat{\text{VaR}}_\alpha = W_0 \times \left(e^{\hat{q}_\alpha^R} - 1\right) = W_0 \left(e^{\hat{\mu} + \hat{\sigma} \times q_\alpha^Z} - 1\right). \tag{1.40}$$

Here, $\widehat{\text{VaR}}_\alpha$ is a nonlinear function of $\hat{q}_\alpha^R$ (and $\hat{\mu}$ and $\hat{\sigma}$). In this case, the statistical properties of $\widehat{\text{VaR}}_\alpha$ cannot be easily derived from the statistical properties of $\hat{\mu}$ and $\hat{\sigma}$.

Statistical Properties of $\hat{q}_\alpha^R$ for Simple Returns

Regarding bias we have

$$E[\hat{q}_\alpha^R] = E[\mu] + q_\alpha^Z \times E[\hat{\sigma}] \approx \mu + \sigma \times q_\alpha^Z. \tag{1.41}$$

Hence, $\hat{q}_\alpha^R$ is approximately unbiased for q_α^R because $\hat{\mu}$ is unbiased for μ and $\hat{\sigma}$ is approximately unbiased for σ.

To derive results regarding precision and the sampling distribution we require the following result.

Result. In the CER model

$$\begin{pmatrix} \hat{\mu} \\ \hat{\sigma} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \\ \sigma \end{pmatrix}, \begin{pmatrix} \text{se}(\hat{\mu})^2 & 0 \\ 0 & \text{se}(\hat{\sigma})^2 \end{pmatrix}\right) \tag{1.42}$$

for large enough T. Hence, $\hat{\mu}$ and $\hat{\sigma}$ are (asymptotically) jointly normally distributed and $\text{cov}(\hat{\mu}, \hat{\sigma}) = 0$ which implies that they are also independent.

Using (1.42) we have

$$\begin{aligned}
\mathrm{var}(\hat{q}_\alpha^R) &= \mathrm{var}(\hat{\mu} + \hat{\sigma} \times q_\alpha^Z) \\
&= \mathrm{var}(\hat{\mu}) + \left(q_\alpha^Z\right)^2 \mathrm{var}(\hat{\sigma}) + 2q_\alpha^Z \mathrm{cov}(\hat{\mu}, \hat{\sigma}) \\
&= \mathrm{var}(\hat{\mu}) + \left(q_\alpha^Z\right)^2 \mathrm{var}(\hat{\sigma}) \text{ (since } \mathrm{cov}(\hat{\mu}, \hat{\sigma}) = 0) \\
&= \frac{\sigma^2}{T} + \frac{\left(q_\alpha^Z\right)^2 \sigma^2}{2T} \\
&= \frac{\sigma^2}{T}\left[1 + \frac{1}{2}\left(q_\alpha^Z\right)^2\right].
\end{aligned}$$

Then

$$\mathrm{se}(\hat{q}_\alpha^R) = \sqrt{\mathrm{var}(\hat{q}_\alpha^R)} = \frac{\sigma}{\sqrt{T}}\left[1 + \frac{1}{2}\left(q_\alpha^Z\right)^2\right]^{1/2}. \tag{1.43}$$

Using the above results, the sampling distribution of $\hat{q}_\alpha^R$ can be approximated by the normal distribution

$$\hat{q}_\alpha^R \sim N(q_\alpha^R,\ \mathrm{se}(\hat{q}_\alpha^R)^2), \tag{1.44}$$

for large enough T.

Remarks

1. $\mathrm{se}(\hat{q}_\alpha^R)$ increases with σ and q_α^Z, and decreases with T. In particular, $\mathrm{se}(\hat{q}_\alpha^R)$ increases as α goes to zero (show plot)
2. The formula for $\mathrm{se}(\hat{q}_\alpha^R)$ is not practically useful because it depends on the unknown value σ. The practically useful estimated standard error replaces the unknown value of σ with the estimate $\hat{\sigma}$ and is given by

$$\widehat{\mathrm{se}}(\hat{q}_\alpha^R) = \frac{\hat{\sigma}}{\sqrt{T}}\left[1 + \frac{1}{2}\left(q_\alpha^Z\right)^2\right]^{1/2}. \tag{1.45}$$

[Insert figure of $\widehat{\mathrm{se}}(\hat{q}_\alpha^R)$ vs. α here]

Example 19 *Estimating $\hat{q}_\alpha^R$ for Microsoft, Starbucks and the S&P 500 index.*

The estimates of q_α^R and $\mathrm{se}(\hat{q}_\alpha^R)$, for $\alpha = 0.05$ and $\alpha = 0.01$, from the simple monthly returns for Microsoft, Starbucks and the S&P 500 index are

```
> n.obs = length(msftRetC)
> muhatS = colMeans(cerRetS)
> sigmahatS = apply(cerRetS, 2, sd)
> qhat.05 = muhatS + sigmahatS*qnorm(0.05)
> qhat.01 = muhatS + sigmahatS*qnorm(0.01)
> seQhat.05 = (sigmahatS/sqrt(n.obs))*sqrt(1 + 0.5*qnorm(0.05)^2)
> seQhat.01 = (sigmahatS/sqrt(n.obs))*sqrt(1 + 0.5*qnorm(0.01)^2)
> cbind(qhat.05, seQhat.05, qhat.01, seQhat.01)
      qhat.05 seQhat.05 qhat.01 seQhat.01
MSFT  -0.1578   0.01187  -0.227   0.01490
SBUX  -0.1587   0.01276  -0.233   0.01602
SP500 -0.0758   0.00559  -0.108   0.00702
```

For Microsoft and Starbucks, the values of $\widehat{\mathrm{se}}(\hat{q}_{\alpha}^{R})$ are about 1.2%, 1.5% and for $\alpha = 0.05$ and $\alpha = 0.01$, respectively. For the S&P 500 index, the corresponding values of $\widehat{\mathrm{se}}(\hat{q}_{\alpha}^{R})$ are only 0.5% and 0.7%, respectively. To see that these standard errors values are actually fairly large consider the 95% confidence intervals for $q_{0.05}^{R}$ and $q_{0.01}^{R}$

```
> lowerQhat.05 = qhat.05 - 2*seqhat.05
> upperQhat.05 = qhat.05 + 2*seqhat.05
> widthQhat.05 = upperQhat.05 - lowerQhat.05
> cbind(lowerQhat.05, upperQhat.05, widthQhat.05)
      lowerQhat.05 upperQhat.05 widthQhat.05
MSFT        -0.197      -0.1181       0.0794
SBUX        -0.203      -0.1144       0.0884
SP500       -0.095      -0.0566       0.0384
> lowerQhat.01 = qhat.01 - 2*seqhat.01
> upperQhat.01 = qhat.01 + 2*seqhat.01
> widthQhat.01 = upperQhat.01 - lowerQhat.01
> cbind(lowerQhat.01, upperQhat.01, widthQhat.01)
      lowerQhat.01 upperQhat.01 widthQhat.01
MSFT        -0.277      -0.1772       0.0996
SBUX        -0.289      -0.1775       0.1110
SP500       -0.132      -0.0842       0.0482
```

For example, the 95% confidences for $q_{0.05}^{R}$ and $q_{0.01}^{R}$ for Microsoft are $[-19.7\%, -11.8\%]$ and $[-27.7\%, -17.7\%]$, respectively, which are quite large. Hence, the 5% and 1% simple monthly return quantiles are not estimated very precisely. ■

Statistical Properties of $\hat{q}_\alpha^R$ for CC Returns

To be completed

The Delta Method To be completed

Statistical Properties of VaR_α

First, consider estimated VaR for simple returns (1.39). Then, using (1.41) - (1.44)

$$
\begin{aligned}
E[\widehat{\mathrm{VaR}}_\alpha] &= W_0 \times E[\hat{q}_\alpha^R] \approx W_0 \times q_\alpha^R = \mathrm{VaR}_\alpha, \\
\mathrm{se}(\widehat{\mathrm{VaR}}_\alpha) &= W_0 \times \mathrm{se}(\hat{q}_\alpha^R), \\
& \widehat{\mathrm{VaR}}_\alpha \sim N(\mathrm{VaR}_\alpha, \mathrm{se}(\widehat{\mathrm{VaR}}_\alpha)^2).
\end{aligned}
$$

Hence, $\widehat{\mathrm{VaR}}_\alpha$ is approximately unbiased and normally distributed with

$$
\mathrm{se}(\widehat{\mathrm{VaR}}_\alpha) = W_0 \times \frac{\sigma}{\sqrt{T}} \left[1 + \frac{1}{2}\left(q_\alpha^Z\right)^2\right]^{1/2}.
$$

The practically useful estimated standard error replaces the unknown value of σ with the estimate $\hat{\sigma}$ and is given by

$$
\mathrm{se}(\widehat{\mathrm{VaR}}_\alpha) = W_0 \times \frac{\hat{\sigma}}{\sqrt{T}} \left[1 + \frac{1}{2}\left(q_\alpha^Z\right)^2\right]^{1/2}.
$$

Example 20 *Estimating* VaR_α *for Microsoft, Starbucks and the S&P 500 index.*

Consider a \$100,000 investment for one month in Microsoft, Starbucks and the S&P 500 index. The estimates of VaR_α and $\mathrm{se}(\widehat{\mathrm{VaR}}_\alpha)$, for $\alpha = 0.05$ and $\alpha = 0.01$, from the simple monthly returns for Microsoft, Starbucks and the S&P 500 index are

```
> w0=100000
> VaR.05 = qhat.05*w0
> seVaR.05 = w0*seQhat.05
> VaR.01 = qhat.01*w0
> seVaR.01 = w0*seQhat.01
```

```
> cbind(VaR.05, seVaR.05, VaR.01, seVaR.01)
      VaR.05 seVaR.05 VaR.01 seVaR.01
MSFT  -15780     1187 -22696     1490
SBUX  -15865     1276 -23303     1602
SP500  -7577      559 -10834      702
```

The 95% confidence intervals for $\mathrm{VaR}_{0.05}$ and $\mathrm{VaR}_{0.01}$ are

```
> lowerVaR.05 = VaR.05 - 2*seVaR.05
> upperVaR.05 = VaR.05 + 2*seVaR.05
> widthVaR.05 = upperVaR.05 - lowerVaR.05
> cbind(lowerVaR.05, upperVaR.05, widthVaR.05)
      lowerVaR.05 upperVaR.05 widthVaR.05
MSFT       -18154      -13405        4748
SBUX       -18418      -13312        5106
SP500       -8695       -6459        2236
> lowerVaR.01 = VaR.01 - 2*seVaR.01
> upperVaR.01 = VaR.01 + 2*seVaR.01
> widthVaR.01 = upperVaR.01 - lowerVaR.01
> cbind(lowerVaR.01, upperVaR.01, widthVaR.01)
      lowerVaR.01 upperVaR.01 widthVaR.01
MSFT       -25676      -19717        5959
SBUX       -26507      -20099        6408
SP500      -12237       -9431        2806
```

1.5 Further Reading

To be completed

Bibliography

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