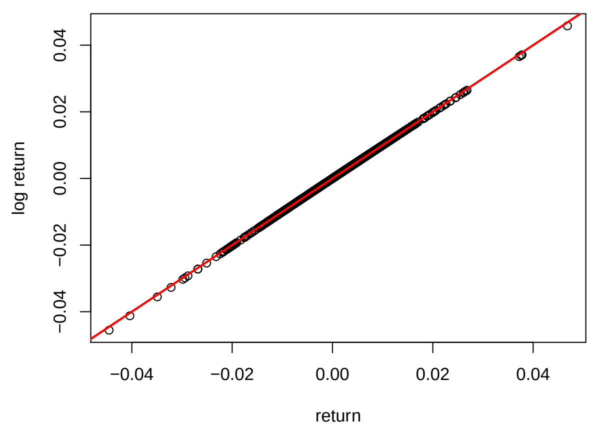
Solutions to Selected Computer Lab Problems and Exercises in Chapter 2 of *Statistics and Data Analysis for Financial Engineering, 2nd ed.* by David Ruppert and David S. Matteson

Problem 12. Code to produce the plot is below.

```
data = read.csv("MCD_PriceDaily.csv")
head(data)
adjPrice = data[,7]
logReturn = diff(log(adjPrice))
n = length(adjPrice)
return = adjPrice[-1]/adjPrice[-n] - 1
plot(return,logReturn,ylab = "log return")
abline(a = 0,b = 1,col = "red",lwd = 2)
```

We see that the return and log return for any day are almost equal. This is reasonable in light of the discussion in Section 2.1.3, especially Figure 2.1.



Problem 13. The R output is below. We see that the means of the returns and log returns are somewhat similar and the standard deviations of the returns and log returns agree to three significant digits. (`sd(return) = 0.00890` but R did not print the final zero.)

```
> options(digits = 4)
> mean(return)
[1] 0.0005027
> mean(logReturn)
[1] 0.0004631
```

```
> sd(return)
[1] 0.0089
> sd(logReturn)
[1] 0.008901
```

Problem 14. It is crucial to use a paired t-test because the return and log return on any given day are highly correlated and the independent samples t-test assumes that they are independent. The paired t-test has a p-value of 2.2×10^{-16}. Therefore, using any reasonable significance level including, of course, 5%, we reject the null hypothesis that the mean of the returns and mean of the log returns are equal.

```
> t.test(return,logReturn,paired = TRUE)

Paired t-test

data:  return and logReturn
t = 16, df = 1200, p-value <2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 3.478e-05 4.460e-05
sample estimates:
mean of the differences
              3.969e-05
```

Note: The p-value would be non-significant if we incorrectly used an unpaired t-test. See below.

```
> t.test(return,logReturn,paired = FALSE)

Welch Two Sample t-test

data:  return and logReturn
t = 0.11, df = 2300, p-value = 0.9
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.0006801  0.0007595
sample estimates:
mean of x mean of y
0.0005027 0.0004631
```

Let d_t be the difference between the return and log return on day t. The assumptions of the paired t-test are that $d_1, \ldots, d_n$ are independent, that they are normally distributed, and that they have a constant variance. Here n is the sample size.

The normality assumption is unlikely to hold, but it is not crucial since, by the central limit theorem, the mean of $d_1, \ldots, d_n$ will be approximately normal and that is enough for the t-test to be valid.

The assumption of a constant variance is likely to be met, although it is not crucial since the t-test is robust to this assumption.

The independence assumption is unlikely to be met, since we might expect some serial correlation and volatility clustering is likely. However, the serial correlation should be small and the t-test is not much affected by volatility clustering.

In summary, the assumptions of the t-test are unlikely to hold exactly, but the conclusion of the t-test should be valid, nonetheless.

Problem 16. The `R` output is below. The expect value of the bet is slightly positive, with a 95% confidence equal to (0.05583, 0.06517), so you should be willing to make this bet.

```
> niter = 10000
> value = matrix(niter)
> set.seed(2015)
> for (i in 1:niter)
+ {
+   logR = rnorm(20,mean = mean(logReturn),sd = sd(logReturn))
+   prices = 93.07 * exp(cumsum(logR))
+   ind = (min(prices) < 85)
+   value[i] = 100*ind - (1 - ind)
+ }
> mu = mean(value)
> mu
[1] 0.0605
> mu + c(-1,1)* sqrt(mu*(1-mu)/niter)*qnorm(0.975)
[1] 0.05583 0.06517
```

Exercise 2. (a) $R_2 = (P_2 + D_2)/P_1 - 1 = 54.2/52 - 1$

(b) $\frac{P_4+D_4}{P_3}\frac{P_3+D_3}{P_2}\frac{P_2+D_2}{P1} - 1 = \frac{59.25}{53}\frac{53.2}{54}\frac{54.2}{52} - 1$

(c) $\log\{(P_3 + D_3)/P_2\} = \log(53.2/54)$

Exercise 7. a) $r_t(4)$ is $N\Big((4)(0.06), (4)(0.47)\Big)$.

(b) Using `R`, $P(r_1(4) < 2) =$ `pnorm(2,4*0.06,sd = sqrt(4*0.47))` $= 0.9004$.

```
> pnorm(2,4*0.06,sd = sqrt(4*0.47))
[1] 0.9004
```

(c) $r_2(1) = r_2$ and $r_2(2) = r_2 + r_1$. Therefore, $\text{cov}(r_2(1), r_2(2)) = \text{var}(r_2) = 0.47$.

(d) $r_t(3) = r_t + r_{t-1} + r_{t-2}$. Therefore, given $r_{t-2} = 0.6$, $r_t(3)$ is $N\Big(0.6 + (2)(0.06), (2)(0.47)\Big)$.

Exercise 8. (a) $P(X_2 > 1.3X_0) = P(r_1 + r_2 > \log(1.3))$ and $r_1 + r_2$ is $N(2\mu, 2\sigma^2)$. This probability can be found using `R`:

```
pnorm(log(1.3),mean = 2*mu,sd = sqrt(2)*sigma,lower.tail = FALSE)
```

Of course, numerical values of `mu` and `sigma` must be supplied to use `R`.

(b) In (A.4), use $g(x) = \exp(x)$ and $h(y) = \log(y)$. The density is at x

$$\frac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\frac{1}{2\sigma^2}\{\log(x) - (\log(X_0) + \mu)\}^2\right].$$

(c) $\exp\Big[\{\log(X_0) + k\mu\} + \sigma\sqrt{k}\Phi^{-1}(0.9)\Big]$ where Φ is the standard normal cumulative distribution function. See page 676 for a discussion of normal quantiles.

(d) Use the fact that X_k^2 is lognormal$(2k\mu, 4k\sigma^2)$. Then

$$E(X_k^2) = X_0^2 \exp(2k\mu + 2k\sigma^2).$$

(e) Using $\text{Var}(X_k) = E(X_k^2) - (E(X_k))^2$ we have

$$X_0^2 \exp(2k\mu + k\sigma^2)\{\exp(k\sigma^2) - 1\}.$$

Exercise 9. (a) $P(X_3 > 1.2X_0) = P(\log(X_3/X_0) > \log(1.2)) =$ `pnorm(log(1.2),mean = 0.1 * 3,sd = sqrt(3) * 0.2,lower.tail = FALSE)` $= 0.633$.

```
> pnorm(log(1.2),mean = 0.1 * 3,sd = sqrt(3) * 0.2,lower.tail = FALSE)
[1] 0.633
```

(c) $P(X_t/X_0 \geq 2) = P(\log(X_t/X_0) \geq \log(2)) =$ `pnorm(log(2),mean = t * 0.1, sd = sqrt(t) * 0.2, lower.tail = FALSE)`

From the output below, we see that this probability exceeds 0.9 if t is 18 or more. I found that the probability exceeds 0.9 if $t = 100$ so I tried all t between 1 and 1000.

```
> t = 1:100
> prob = pnorm(log(2),mean = t * 0.1, sd = sqrt(t) * 0.2,
    lower.tail = FALSE)
> min(which(prob > 0.9))
[1] 18
```