Solutions to Selected Computer Lab Problems and Exercises in Chapter 11 of *Statistics and Data Analysis for Financial Engineering, 2nd ed.* by David Ruppert and David S. Matteson

Problem 6. The MLE of α is 0.1414. `boxcox()` returns the values of α and their log-likelihoods, which are called `x` and `y`, respectively. The following finds the MLE:

```
alpha = fit2$x[fit2$y==max(fit2$y)]
> alpha
[1] 0.1414141
```

Problem 7. Yes, there is a huge decrease in AIC when α changes from 1 (no transformation) to the MLE.

```
> AIC(fit1)
[1] 12094.19
> AIC(fit3)
[1] 1583.145
```

Problem 8. No, checking for residual correlation is pointless, since the data are cross-sectional and are in no particular order.

Problem 9. We see in the output below that `price`, `bathrooms`, `stories`, and `gasheat` are all in the final model. The probability of having air conditioning increases with `price` and `stories` and is lower for houses with gas heat compared to those with other types of heating. The variable `bathrooms` is only marginally significant with a p-value of 0.067 and it has a negative coefficient which seems counterintuitive.

```
> summary(fit2)

Call:
glm(formula = aircon ~ price + bathrooms + stories + gasheat,
    family = "binomial", data = HousePrices)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8433  -0.7278  -0.5121   0.6876   3.0753


Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.045e+00  4.050e-01  -9.987  < 2e-16 ***
price        4.782e-05  6.008e-06   7.959 1.73e-15 ***
bathrooms   -4.723e-01  2.576e-01  -1.833 0.066786 .
stories      3.224e-01  1.317e-01   2.449 0.014334 *
gasheatyes  -3.657e+00  1.082e+00  -3.378 0.000729 ***
---
(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 681.92  on 545  degrees of freedom
Residual deviance: 529.36  on 541  degrees of freedom
AIC: 539.36


Number of Fisher Scoring iterations: 6
```

To investigate whether the variable `bathrooms` might be dropped, the model with this variable deleted was fit.

```
fit3 = glm(aircon~price+stories+gasheat,family="binomial")
summary(fit3)
AIC(fit2)
AIC(fit3)


> AIC(fit2)
[1] 539.3585
> AIC(fit3)
[1] 540.8747


> dim(HousePrices)[1]
[1] 546
> n=dim(HousePrices)[1]
> log(n)
[1] 6.302619
```

We see that the models with and without `bathroom` have similar AIC values. Moreover, the sample size is 546, so $\log(n) = 6.3$. Therefore, the model without `bathrooms` will have a smaller BIC. This suggests that `bathrooms` should be deleted.

Problem 10. If the model without `bathrooms` is used, then the coefficients are:

```
> coef(fit3)
  (Intercept)          price       stories     gasheatyes
-4.2756900399   0.0000431954  0.2944715944  -3.6895704723
```

The coefficients should be multiplied by $(1, 42000, 2, 0)$. The probability of this house having air conditioning is 0.133:

```
> prob = plogis( c(1, 42000, 2, 0) %*% coef(fit3))
> prob
          [,1]
[1,] 0.1332481
```

Exercise 3a. $\theta_1 = 0.0128$, $\theta_2 = 0.0256$, $\theta_3 = 0.0170$, and $\theta_4 = 0.0718$.

The Nelson–Siegel yield curve is

$$y_t(\theta_1, \theta_2, \theta_3, \theta_4) = \theta_1 + \left(\theta_2 + \frac{\theta_3}{\theta_4}\right) \frac{1 - \exp(-\theta_4 t)}{\theta_4 t} - \frac{\theta_3}{\theta_4} \exp(-\theta_4 t).$$

The following `R` code uses the yield curve to fit the Nelson–Siegel model.

```
NSYield = function(theta)
  {theta[1] + (theta[2] + theta[3]/theta[4]) *
   (1 - exp(-theta[4]*T))/(theta[4]*T)-
   (theta[3]/theta[4])*exp(-theta[4]*T)}
dat = read.table("ZeroPrices.txt",header=T)
attach(dat)
start = c(0.01,0.025,0.01,0.075)
T = maturity
SS = function(theta) sum((price-100*exp(-T*NSYield(theta)))^2)
fitnls = optim(start,SS)
thetahat = fitnls$par
round(thetahat,4)
```

The estimates are:

```
> round(thetahat,4)
[1] 0.0176 0.0416 0.0023 0.0696
```

Exercise 3b. The following continuation of the R code above defines a function to compute the forward rate and plots the forward rate and yield curve. It also adds a legend.

```
NSForward = function(theta){
   theta[1] + (theta[2] + theta[3]*T) * exp(-theta[4]*T)
}
plot(T,NSForward(thetahat),type="l",ylab="rate",xlab="T",lwd=2)
lines(T,NSYield(thetahat),lty=2,lwd=2,col="red")
legend("topright",c("forward rate","yield"),lty=c(1,2),
    lwd=2,col=c("black","red"))
```
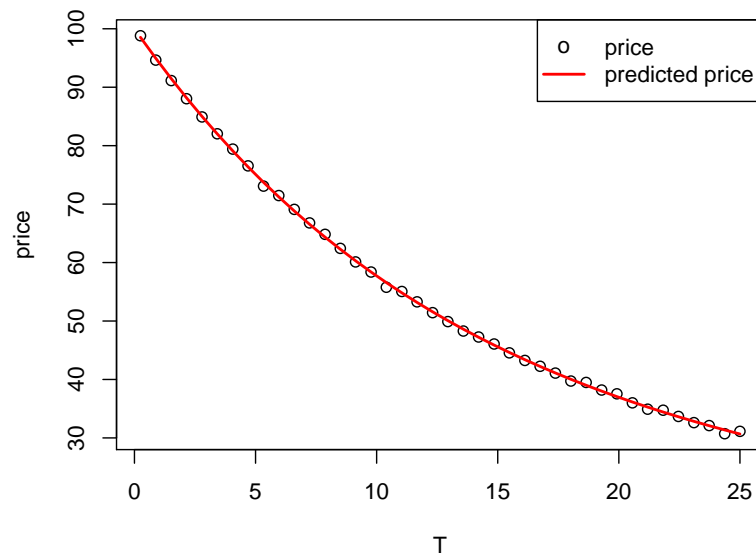
The plot is below.
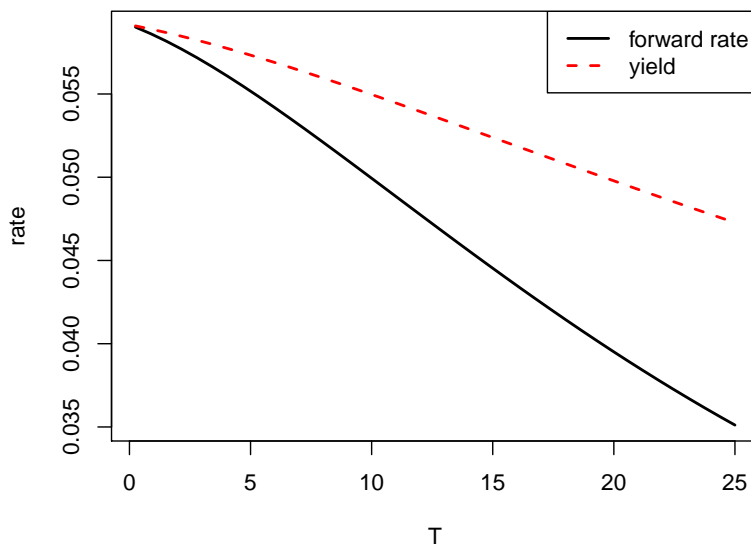
As a check that the computations are correct, the following code plots the prices and the predicted prices:

```
plot(T,price)
lines(T,100*exp(-T*NSYield(fitnls$par)),col="red",lwd=2 )
legend("topright",c("price","predicted price"),lty=c(NA,1),
    pch=c("o",NA),col=c("black","red"),lwd=2)
```

The plot is below. We see that the prices are predicted very accurately by the Nelson-Siegel model.

Exercise 4. There is no sign of bias. The deviation between $r = 0.06$ and $\hat{r} = 0.0585$ is 0.0015 and approximately the same as the standard of $\hat{r}$ which is 0.00149. A deviation approximately one standard error in size is small and could be due entirely to random variation. This does not prove that there is no bias, but only that there is no evidence of bias.