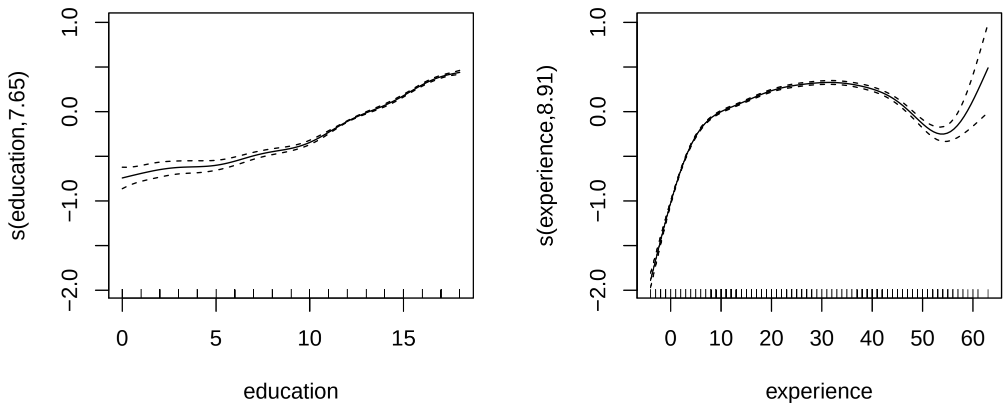
Solutions to Selected Computer Lab Problems and Exercises in Chapter 21 of *Statistics and Data Analysis for Financial Engineering, 2nd ed.* by David Ruppert and David S. Matteson

Problem 1. The summary below shows that $\widehat{\beta}_0 = 6.19$ and $\widehat{\beta}_1 = -0.241280$.

```
> summary(fitGam)

Family: gaussian
Link function: identity

Formula:
log(wage) ~ s(education) + s(experience) + ethnicity

Parametric coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    6.189742   0.003558 1739.89   <2e-16 ***
ethnicityafam -0.241280   0.012697  -19.00   <2e-16 ***
---

Approximate significance of smooth terms:
                edf Ref.df      F p-value
s(education)  7.653  7.653  672.2  <2e-16 ***
s(experience) 8.906  8.906 1207.6  <2e-16 ***
---

R-sq.(adj) =   0.36   Deviance explained = 36.1%
GCV score = 0.32802  Scale est. = 0.3278    n = 28155
```

Problem 2. s_1 is slightly wiggly but it is increasing and, in general, its slope is greater for people with more than 10 years of education compared to those with 9 or less years.



s_2 rises quickly up to about 10 years of experience, rises less quickly to a peak at about 30 years of experience, and then drops until about 50 years of experience. After 50 years of experience, it rises ago. This does NOT mean that an individual's wages rise after 50 years of experience. These data are cross-sectional, not longitudinal, so individuals are not followed during their careers. The workers with 50 or more years of experience might be quite different than the workers with less experience because, for example, workers with lower wages may retire earlier. Less than 1% of worker have 50 or more years of experience as can be determined by:

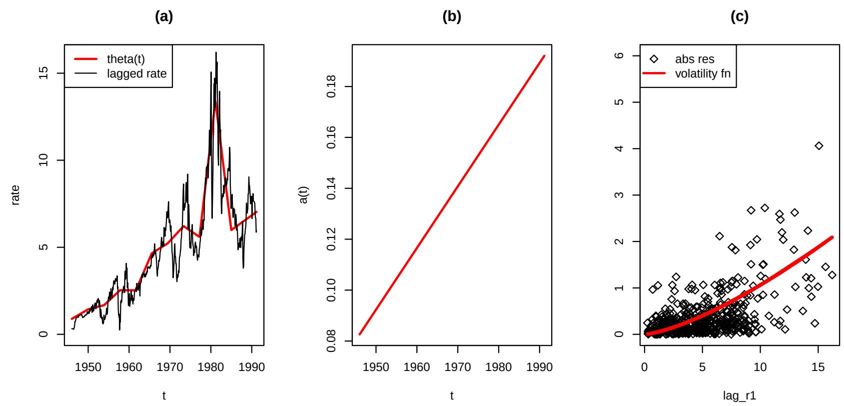
```
> 100* mean((experience > 50))
[1] 0.8808382
```

Problem 3. Ten knots are are being used. Here the `outer` function creates at `n` $\times$ 10 matrix whose i, jth element is $t_i - \kappa_j$ where t_i is the ith element of `t` and κ_j is the jth knot.

The statement `X2 = X1 * (X1>0)` replaces all negative elements of `X1` by 0. Stated differently, `X2` contains the elements $(t_i - \kappa_j)_+$.

The variable `X3` is the matrix $\boldsymbol{X}$ in (21.16). It contains all of the spline basis functions evaluated at the elements of `t`. Here, $p = 1$ and the spline basis functions are $1, t, (t - \kappa_1)_+, \ldots, (t - \kappa_{10})_+$.

Problem 4. We saw in the previous problem that `X3` contains the spline basis functions, so that `X3%*%theta` is a linear combination of these basis function (with weights in `theta`) and so is a spline. Also, the first two basis functions are 1 and t so `X3[,1:2]` contains the basis of linear functions.



Problem 5.

A time-varying θ says that the interest rate reverts to a mean that is changing with time. Therefore, it makes sense that the estimate of θ appears tracks the interest rate, since under the model the interest rate is tracking θ.

Problem 6. The following code prints the AIC of the model just fit and of a new model where $a(t)$ is constant. Note that AIC is smaller for the model with constant $a(t)$. This suggests that we accept the null hypothesis that $a(t)$ is constant.

One might prefer a formal test instead of AIC. Since AIC is -2 times the log-likelihood plus twice number of parameters, we can use AIC to compute the log-likelihood of both models and compute a likelihood ratio test. The likelihood ratio statistic on the left hand side of (5.27) is 957.1095 - 955.9084 $-2 = -0.8$. The "2" is twice the number of parameters that are being tested (1). A likelihood ratio statistic should always be nonnegative, and the negative value must be due to numerical error. In any case, the likelihood ratio test is very small which means we should accept that $a(t)$ is constant.

```
> AIC(nlmod_CKLS_ext)
[1] 957.1095
> nlmod_CKLS_ext2 = nls(delta_r1 ~ X3[,1]*a[1] * (X3%*%theta-lag_r1),
+     start=list(theta = c(10,rep(0,m)),
+     a=.01),control=list(maxiter=200))
> AIC(nlmod_CKLS_ext2)
[1] 955.9084
> AIC(nlmod_CKLS_ext)-AIC(nlmod_CKLS_ext2)-2
[1] -0.7988272
```

Exercise 1. (a) Because $s(0) = 1$ and $s(1) = 1.3$, $s(t) = 1 + 0.3t$ for $0 \leq t \leq 1$. Therefore, $s(0.5) = 1.15$.

(b) Since $s(t)$ is linear for $t > 3$ and $s(4) = s(5) = 6$, $s(t) \equiv 6$ for $t \geq 3$. Therefore, $s(3) = 6$.

(c)
$$\int_2^4 s(t)dt = \int_2^3 5.5 + 0.5(t-2)dt + \int_3^4 6dt = 5.75 + 6 = 11.75.$$

Exercise 2. (a) $E(r_t|R_{t-1} = 0.04) = 0.1(0.035 - 0.04) = 0.0005$

(b) $\text{Var}(r_t|r_{t-1} = 0.02) = \{(2.3)(0.02)\}^2 = 0.002116$