Statistics 435/535/711 16 February 2022
P. Shaman

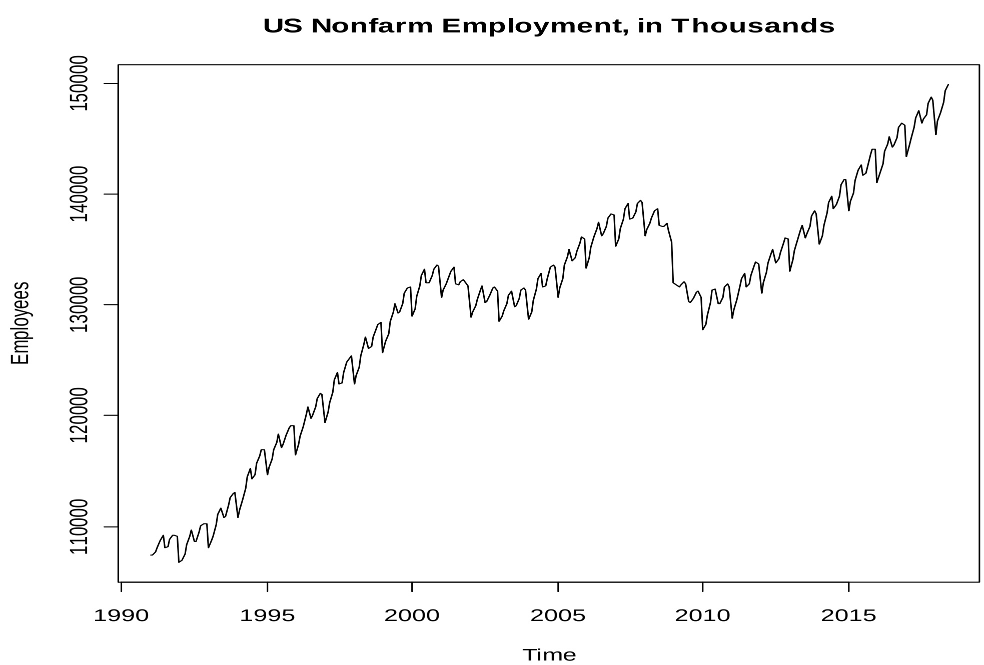
U.S. Nonfarm Employment

The data in USnonfarmemp.txt give monthly U.S. nonfarm employment, in thousands, for the period January 1991 to June 2018. The data are provided by the U.S. Bureau of Labor Statistics (BLS) and are not seasonally adjusted. The figures exclude proprietors, private household employees, unpaid volunteers, farm employees, and unincorporated self-employed persons. Those included account for approximately 80 per cent of the workers who contribute to GDP.

```
> emp<-read.csv("F:/Stat71122Spring/USnonfarmemp.txt")
> attach(emp)
> head(emp)
  Year Month employees   logemp      dlogemp
1 1991     1    107426 11.58456 -0.023459922
2 1991     2    107394 11.58426 -0.000297924
3 1991     3    107683 11.58695  0.002687411
4 1991     4    108072 11.59055  0.003605946
5 1991     5    108754 11.59684  0.006290779
6 1991     6    109235 11.60126  0.004413074
```

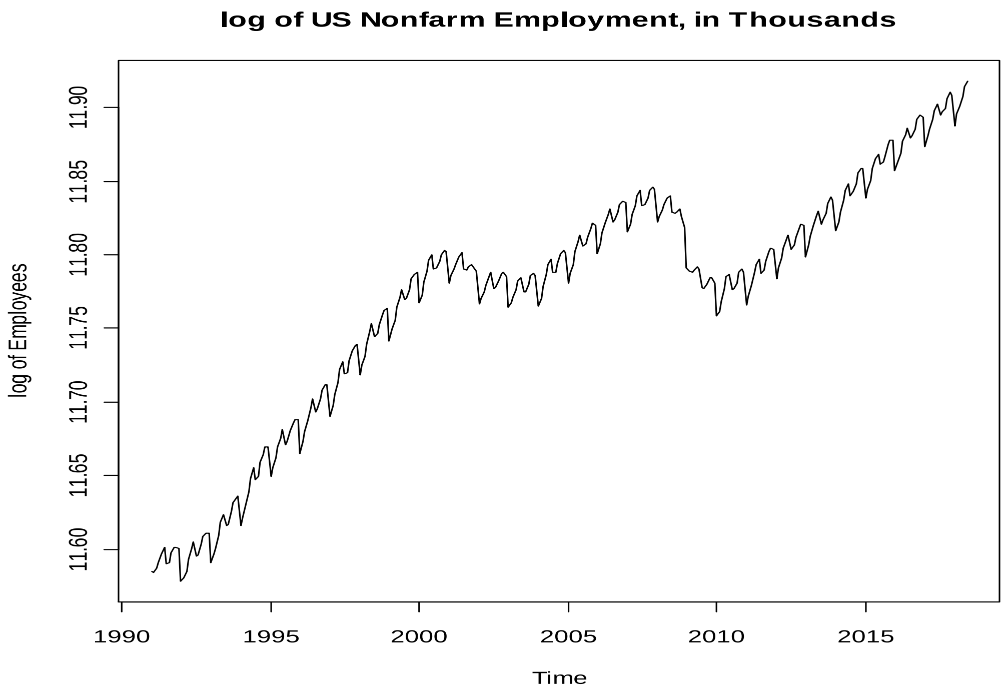
Let's begin with a plot of the employment data vs. time.

```
>
plot(ts(employees,start=c(1991,1),freq=12),xlab="Time",ylab="Employees",main=
"US Nonfarm Employment, in Thousands")
```



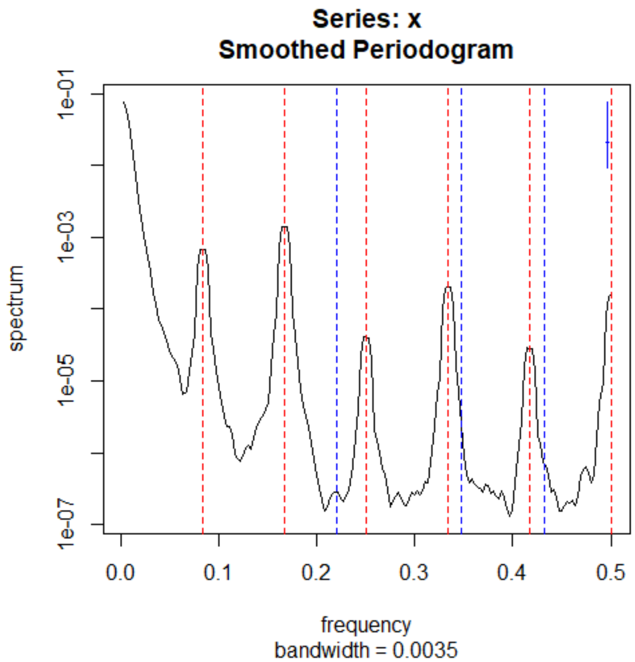
The U.S. economy experienced a strong rise from 1991 to the start of 2001. In the aftermath of the Asian financial crisis and the bursting of the dot com bubble (the NASDAQ began to crash in March 2000), a recession occurred in 2001. Subsequently, another strong increase took place, until the middle of 2007. Then, as the recession took effect, employment declined, until the summer of 2009. This decline included a precipitous drop from December 2007 to January 2008. (Every year, one should note, there is a noticeable decline in U.S. employment from December to January, a seasonal effect.) The Business Cycle Dating Committee of the National Bureau of Economic Research, which is acknowledged to be the arbiter of U.S. business cycles, ruled that the U.S. recession started in January 2008 and ended in June 2009. Finally, the graph shows that the current rise in employment started in the summer of 2009 and increased steadily throughout the remaining range of the data.

A plot of the log of employment follows.



Before turning to estimation, let's plot the spectral density of the log employment series. This will give us information about the structure of the series and provide guidance on what variables to include in a model fit to the data. In plotting the spectral density, we need to choose a span. It is common to use half of the square root of the length of the series, or a nearby value. For this series, this calculation gives 9. Sometimes a smaller span value is chosen to show more clearly some of the features of the spectrum, especially when there are likely to be two structural components which are close to each other in frequency. The plot which follows uses 5 for the span.

```
> spectrum(logemp,span=5)
> abline(v=c(1/12,2/12,3/12,4/12,5/12,6/12),col="red",lty=2)
> abline(v=c(0.220,0.348,0.432),col="blue",lty=2)
```



The peak at low frequency indicates the presence of trend. The dashed red lines are at frequencies 1/12, 2/12, 3/12, 4/12, 5/12, and 6/12, which are the seasonal frequencies for monthly data. The dashed blue lines are at frequencies 0.220, 0.348, and 0.432, the calendar frequencies. The plot shows the presence of a strong seasonal component and no calendar structure. To begin, let's fit a multiplicative decomposition model with trend and seasonal variables and, to confirm their lack of significance, calendar trigonometric pairs for the frequencies 0.348 and 0.432. The calendar pairs showed no significance (all *p*-values for the *t*-tests were greater than 0.93).

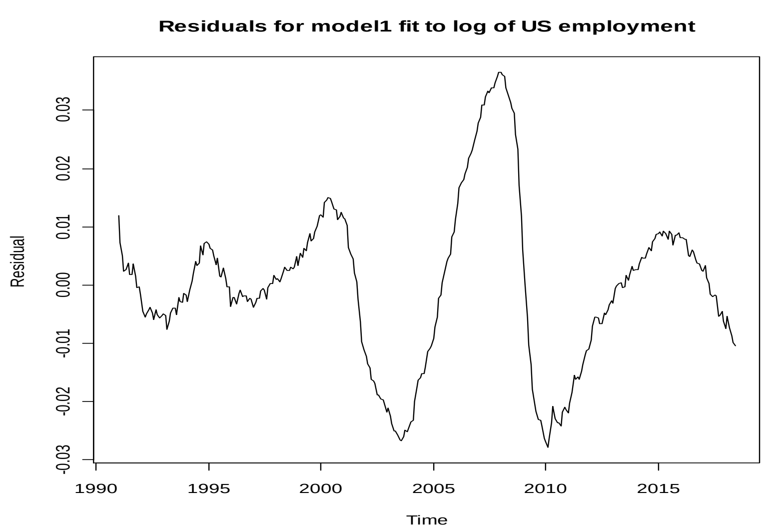
So let's fit a multiplicative decomposition model with just trend and seasonal components. The trend is difficult to track with a polynomial in time. We started by fitting a model with a sixth-degree polynomial trend (there are five turns in the graph) and a seasonal component. We found the sixth power of time coefficient is not significant, and so we use a fifth-degree polynomial trend.

```
> Time<-as.numeric(1:length(employees))
> fMonth<-as.factor(Month)

> model1<-
lm(logemp~Time+I(Time^2)+I(Time^3)+I(Time^4)+I(Time^5)+fMonth);summary(model1
)

Call:
lm(formula = logemp ~ Time + I(Time^2) + I(Time^3) + I(Time^4) +
    I(Time^5) + fMonth)
```

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.027878 -0.006551  0.000301  0.007037  0.036591

Coefficients:
              Estimate Std. Error  t value Pr(>|t|)
(Intercept)  1.157e+01  5.455e-03 2121.550  < 2e-16 ***
Time         2.801e-05  2.962e-04    0.095 0.924732
I(Time^2)    4.514e-05  5.523e-06    8.173 7.56e-15 ***
I(Time^3)   -3.929e-07  4.223e-08   -9.304  < 2e-16 ***
I(Time^4)    1.217e-09  1.405e-10    8.660 2.55e-16 ***
I(Time^5)   -1.254e-12  1.690e-13   -7.418 1.12e-12 ***
fMonth2      4.278e-03  3.838e-03    1.114 0.265951
fMonth3      8.972e-03  3.839e-03    2.337 0.020061 *
fMonth4      1.494e-02  3.839e-03    3.891 0.000122 ***
fMonth5      2.047e-02  3.840e-03    5.330 1.88e-07 ***
fMonth6      2.327e-02  3.840e-03    6.060 3.90e-09 ***
fMonth7      1.360e-02  3.875e-03    3.510 0.000513 ***
fMonth8      1.381e-02  3.875e-03    3.563 0.000424 ***
fMonth9      1.796e-02  3.876e-03    4.633 5.29e-06 ***
fMonth10     2.266e-02  3.876e-03    5.846 1.26e-08 ***
fMonth11     2.389e-02  3.876e-03    6.163 2.20e-09 ***
fMonth12     2.174e-02  3.877e-03    5.607 4.52e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01436 on 313 degrees of freedom
Multiple R-squared: 0.9717,     Adjusted R-squared: 0.9702
F-statistic: 671.6 on 16 and 313 DF,  p-value: < 2.2e-16
```

Here is the residual plot:

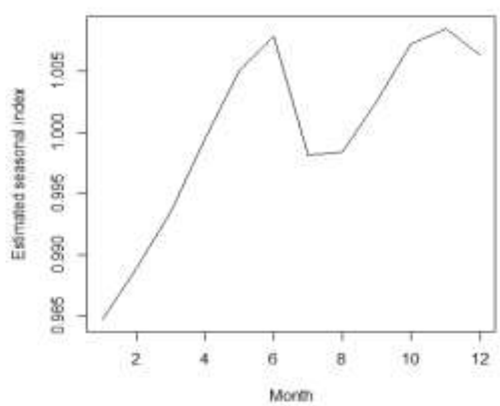
```
> plot(ts(resid(model1),start=c(1991,1),freq=12),xlab="Time",ylab="Residual
",main="Residuals for model1 fit to log of US employment")
```



As the plot shows, the fifth-degree polynomial cannot capture the shape of the trend adequately, primarily during the economic downturns. To assess it properly, though, pay attention to the span of the vertical scale here, in comparison to the scale for the plot of the logged data.

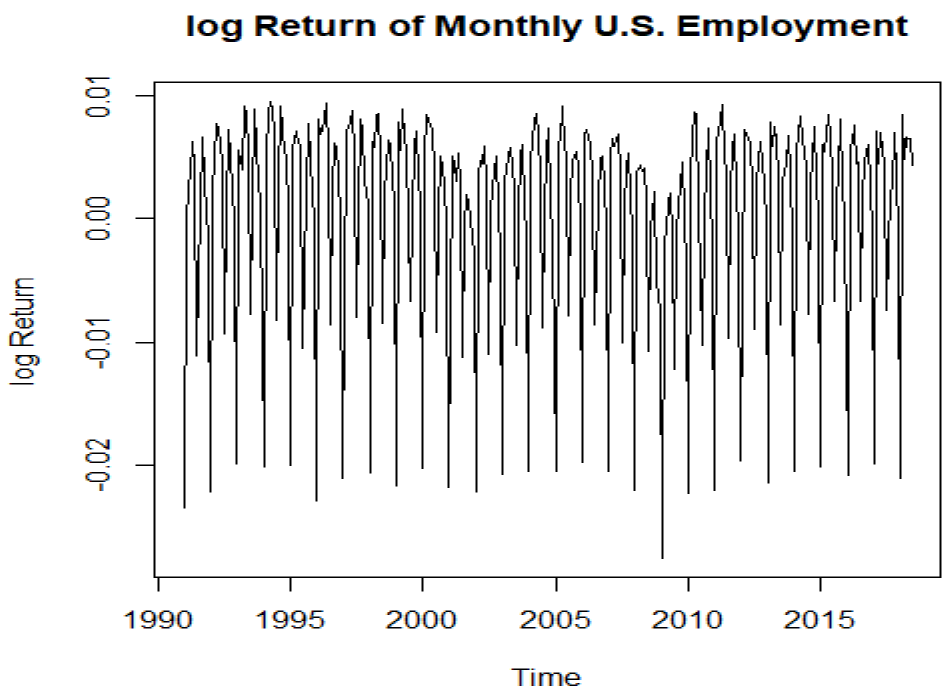
Perhaps a better strategy for trend estimation would be to focus on a smoothing procedure which follows the trend contours. But let's first look at the seasonal index estimates from this model.

```
> b1<-coef(model1)[1]
> b2<-coef(model1)[7:17]+bl
> b3<-c(b1,b2)
> seas<-exp(b3-mean(b3))
> cbind(seas)
             seas
Month1      0.9847
Month2      0.9889
Month3      0.9935
Month4      0.9995
Month5      1.0050
Month6      1.0078
Month7      0.9981
Month8      0.9983
Month9      1.0025
Month10     1.0072
Month11     1.0085
Month12     1.0063

> seas.ts<-ts(seas)
> plot(seas.ts,xlab="Month",ylab="Estimated seasonal index")
```



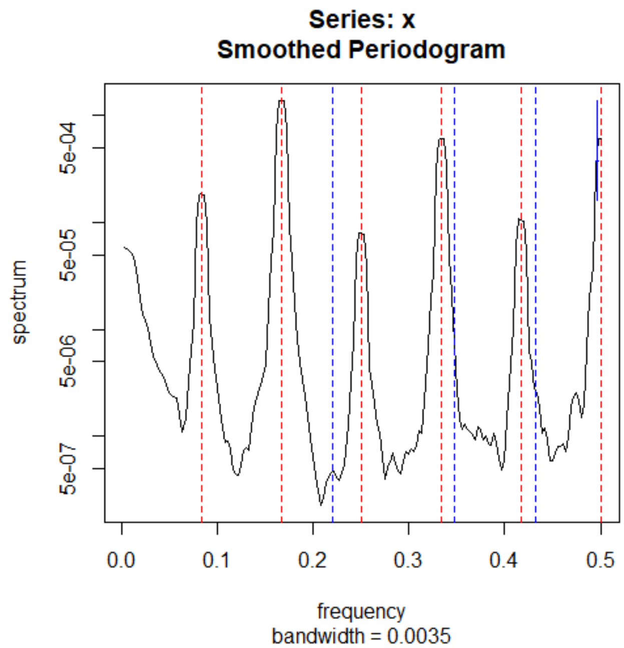
The estimation shows two seasonal peaks, one in May and June, and a second in October, November, and December. The low point is in January, where employment is estimated to be 1.53 per cent below the level of the trend. There is a steady increase in the index from January to June. The pattern also contains a dip in July and August. Overall, the estimated seasonal adjustments to employment range monthly from 1.53 per cent below the trend level (January) to 0.85 per cent above the trend level (November).

Next, noting that the trend estimation is inadequate, let's analyze the monthly changes (the differences) of the logged values, the log return data. Here is the plot of this series:



For this data set, the differencing operation has not killed the trend structure. And the large negative value is for the change from December 2008 to January 2009.

The plot of the spectrum of this log return series is on the next page. Compare it to the spectral plot on page 3. Here the trend component is seen to be much less prominent than that for the log series. Moreover, the seasonal peaks at the higher frequencies have become relatively more prominent. Also, the plot suggests that the calendar pairs at frequencies 0.348 and 0.432 might now require attention (admittedly, the visual evidence is not convincing). The changes have occurred because the differencing operation giving the log return data dampens low frequency components and enhances high frequency components. The calendar frequencies are in the high frequency range.



The following fit includes both trend (a sixth-degree polynomial in time) and seasonal structure, two calendar trigonometric pairs, and a dummy for the January 2009 difference.

```
> obs217<-c(rep(0,216),1,rep(0,113))
> c348<-cos(pi*0.696*Time);s348<-sin(pi*0.696*Time)
> c432<-cos(pi*0.864*Time);s432<-sin(pi*0.864*Time)

> model2<-
lm(dlogemp~Time+I(Time^2)+I(Time^3)+I(Time^4)+I(Time^5)+I(Time^6)+fMonth+obs2
17+c348+s348+c432+s432);summary(model2)

Call:
lm(formula = dlogemp ~ Time + I(Time^2) + I(Time^3) + I(Time^4) +
    I(Time^5) + I(Time^6) + fMonth + obs217 + c348 + s348 + c432 +
    s432)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0061835 -0.0007854  0.0001046  0.0008955  0.0033661

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.429e-02  6.728e-04 -36.100  < 2e-16 ***
Time         2.745e-04  5.155e-05   5.325 1.95e-07 ***
I(Time^2)   -5.677e-06  1.349e-06  -4.209 3.38e-05 ***
I(Time^3)    5.358e-08  1.526e-08   3.511 0.000513 ***
I(Time^4)   -2.663e-10  8.375e-11  -3.180 0.001626 **
I(Time^5)    6.704e-13  2.196e-13   3.052 0.002468 **
I(Time^6)   -6.667e-16  2.204e-16  -3.024 0.002704 **
fMonth2      2.632e-02  4.177e-04  63.020  < 2e-16 ***
fMonth3      2.672e-02  4.176e-04  63.985  < 2e-16 ***
fMonth4      2.800e-02  4.178e-04  67.015  < 2e-16 ***
fMonth5      2.756e-02  4.178e-04  65.980  < 2e-16 ***
```

```
fMonth6       2.484e-02  4.179e-04  59.430  < 2e-16 ***
fMonth7       1.190e-02  4.217e-04  28.210  < 2e-16 ***
fMonth8       2.218e-02  4.217e-04  52.596  < 2e-16 ***
fMonth9       2.611e-02  4.218e-04  61.885  < 2e-16 ***
fMonth10      2.665e-02  4.218e-04  63.188  < 2e-16 ***
fMonth11      2.320e-02  4.218e-04  55.007  < 2e-16 ***
fMonth12      1.980e-02  4.220e-04  46.929  < 2e-16 ***
obs217       -5.628e-03  1.596e-03  -3.527 0.000484 ***
c348         -1.792e-04  1.211e-04  -1.481 0.139761
s348         -2.261e-04  1.205e-04  -1.876 0.061629 .
c432         -1.248e-04  1.207e-04  -1.034 0.301896
s432          1.197e-04  1.208e-04   0.991 0.322340
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.001548 on 307 degrees of freedom
Multiple R-squared:  0.9667,    Adjusted R-squared:  0.9643
F-statistic:   405 on 22 and 307 DF,  p-value: < 2.2e-16
```

The calendar pair with frequency 0.432 is not significant. Let's drop it and refit.

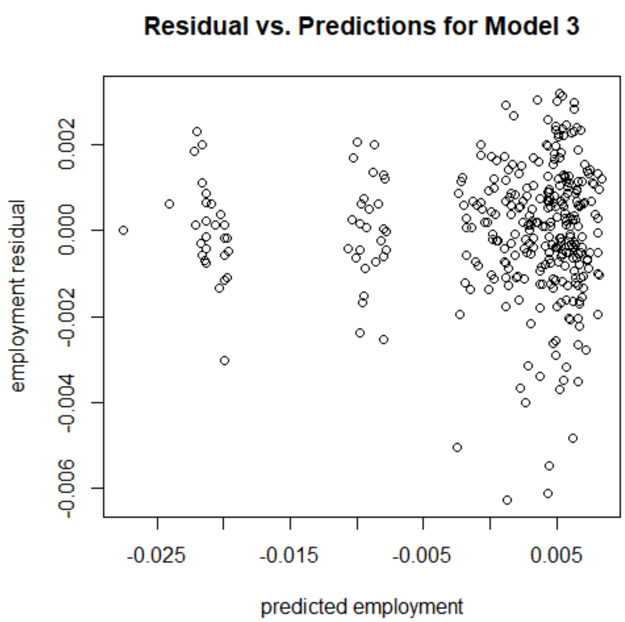
```
 > model3<-
lm(dlogemp~Time+I(Time^2)+I(Time^3)+I(Time^4)+I(Time^5)+I(Time^6)+fMonth+obs2
17+c348+s348);summary(model3)

Call:
lm(formula = dlogemp ~ Time + I(Time^2) + I(Time^3) + I(Time^4) +
    I(Time^5) + I(Time^6) + fMonth + obs217 + c348 + s348)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0062743 -0.0007293  0.0001163  0.0009428  0.0032004

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.427e-02  6.727e-04 -36.075  < 2e-16 ***
Time         2.736e-04  5.155e-05   5.307 2.13e-07 ***
I(Time^2)   -5.653e-06  1.349e-06  -4.191 3.63e-05 ***
I(Time^3)    5.331e-08  1.526e-08   3.493 0.000547 ***
I(Time^4)   -2.648e-10  8.375e-11  -3.162 0.001725 **
I(Time^5)    6.664e-13  2.196e-13   3.034 0.002616 **
I(Time^6)   -6.626e-16  2.204e-16  -3.006 0.002866 **
fMonth2      2.631e-02  4.176e-04  63.001  < 2e-16 ***
fMonth3      2.672e-02  4.176e-04  63.969  < 2e-16 ***
fMonth4      2.798e-02  4.177e-04  67.000  < 2e-16 ***
fMonth5      2.755e-02  4.177e-04  65.961  < 2e-16 ***
fMonth6      2.483e-02  4.179e-04  59.417  < 2e-16 ***
fMonth7      1.189e-02  4.217e-04  28.192  < 2e-16 ***
fMonth8      2.217e-02  4.217e-04  52.575  < 2e-16 ***
fMonth9      2.610e-02  4.218e-04  61.867  < 2e-16 ***
fMonth10     2.664e-02  4.218e-04  63.170  < 2e-16 ***
fMonth11     2.319e-02  4.218e-04  54.988  < 2e-16 ***
fMonth12     1.979e-02  4.220e-04  46.911  < 2e-16 ***
obs217      -5.753e-03  1.591e-03  -3.617 0.000348 ***
c348        -1.782e-04  1.211e-04  -1.472 0.142082
s348        -2.256e-04  1.205e-04  -1.872 0.062200 .
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.001548 on 309 degrees of freedom
Multiple R-squared:  0.9665,    Adjusted R-squared:  0.9643
F-statistic: 445.3 on 20 and 309 DF,  p-value: < 2.2e-16
```

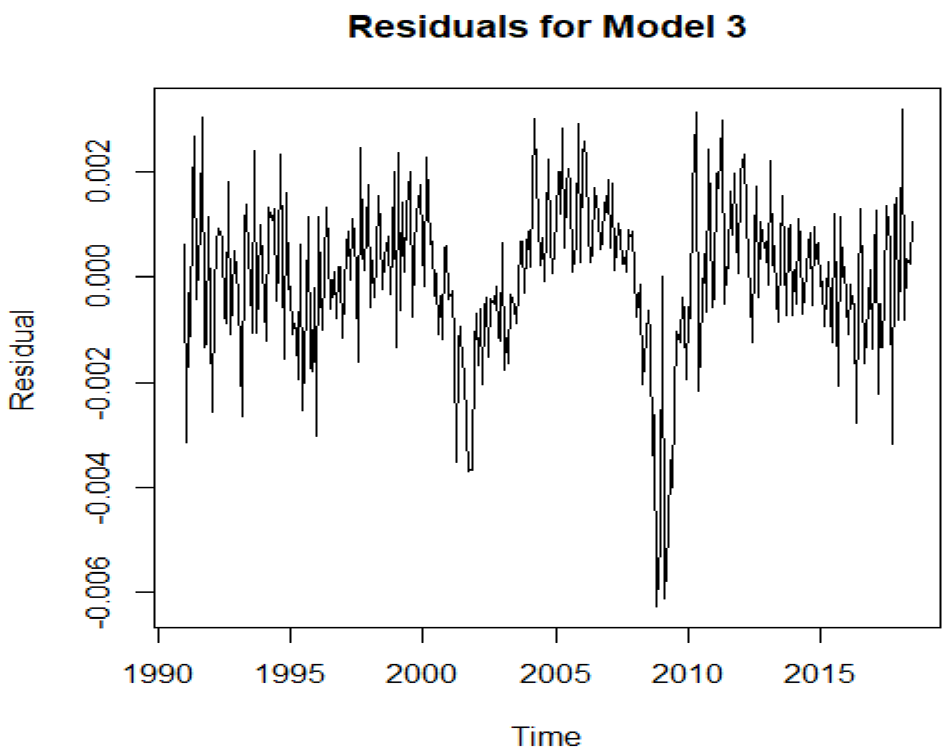
Let's test the significance of the calendar pair.

```
> model4<-
lm(dlogemp~Time+I(Time^2)+I(Time^3)+I(Time^4)+I(Time^5)+I(Time^6)+fMonth+obs2
17)

> anova(model4,model3)
Analysis of Variance Table

Model 1: dlogemp ~ Time + I(Time^2) + I(Time^3) + I(Time^4) + I(Time^5) +
    I(Time^6) + fMonth + obs217
Model 2: dlogemp ~ Time + I(Time^2) + I(Time^3) + I(Time^4) + I(Time^5) +
    I(Time^6) + fMonth + obs217 + c348 + s348
  Res.Df        RSS Df  Sum of Sq      F  Pr(>F)
1    311 0.00075384
2    309 0.00074030  2 1.3538e-05 2.8253 0.06082 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The calendar trigonometric pair with frequency 0.348 is marginally significant via the partial *F*-test, and we choose to retain it in the model. This calendar effect is marginally statistically significant for the fit to the log returns (the differences of the log of employment), but not if the dependent variable is taken to be the log of employment—this makes sense because differencing enhances fast movements of the time series, and frequency 0.348 is in the high frequency band.



The plot above reveals some mild heteroscedasticity, but not enough to cause concern. A plot of the residuals vs. time follows.



The trend estimation is still unsatisfactory. There are two large dips where the model consistently overestimates the level of employment. The first is from mid-2000 to the end of 2004, surrounding the 2001 recession, and the second is from mid-2007 to mid-2010, the period surrounding the 2008–2009 recession. In each case, the model cannot capture the extent of the drop in employment.

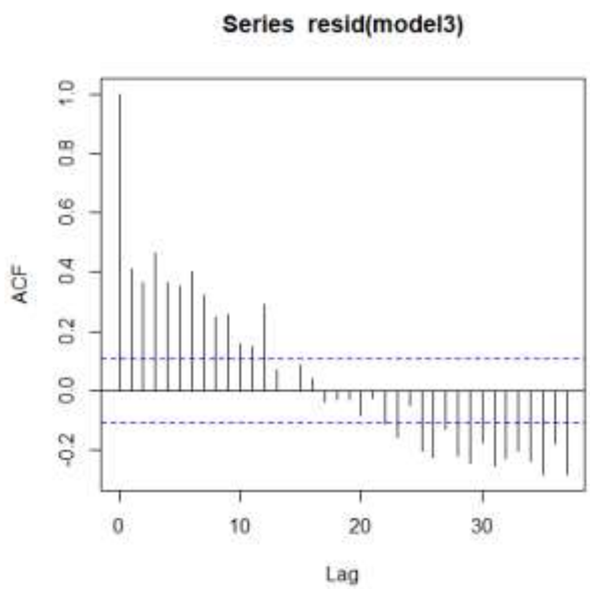
Let's compare the seasonal index estimation from model 3 with that from the first model. The table below shows the estimations from the two models. We use the methodology described in the 24 January notes to obtain seasonal index estimates for the original data in model 3.

```
> b1<-coef(model3)[1]
> b2<-coef(model3)[8:18]+b1
> b3<-c(b1,b2)
> x<-b3-mean(b3)
> s12<-0
> for(j in 2:12){
+ xsub<-x[j:12]
+ s12<-s12+sum(xsub)
+ }
> s12<-s12/12
> s<-c(rep(0,times=12))
> s[12]<-s12
> for(j in 1:11){
+ xsub<-x[1:j]
+ s[j]<-s[12]+sum(xsub)
+ }
```

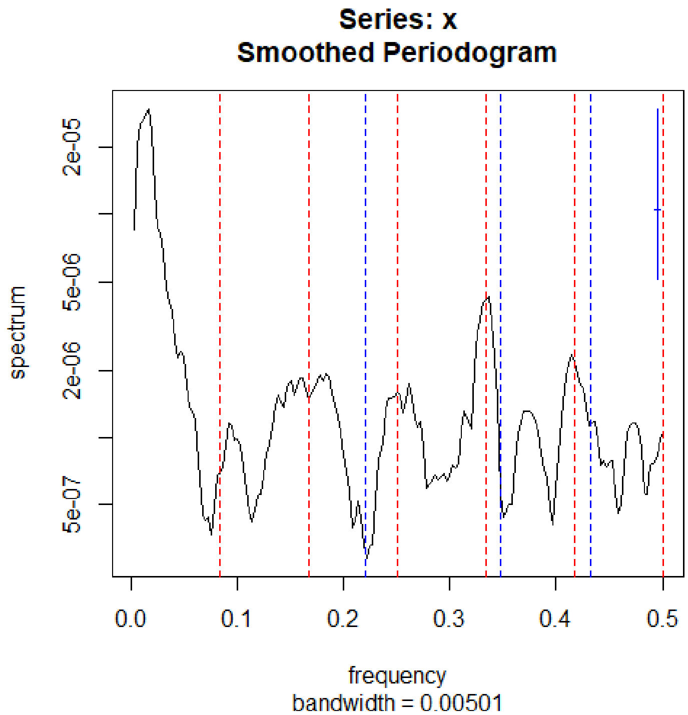
```
> seas3<-exp(s)
> cbind(seas,seas3)
             seas   seas3
Month1       0.9847 0.9845
Month2       0.9889 0.9888
Month3       0.9935 0.9935
Month4       0.9995 0.9996
Month5       1.0050 1.0052
Month6       1.0078 1.0081
Month7       0.9981 0.9980
Month8       0.9983 0.9983
Month9       1.0025 1.0024
Month10      1.0072 1.0072
Month11      1.0085 1.0085
Month12      1.0063 1.0063
```

The two estimations are virtually the same.

Let's consider some additional residual diagnostics for model 3.



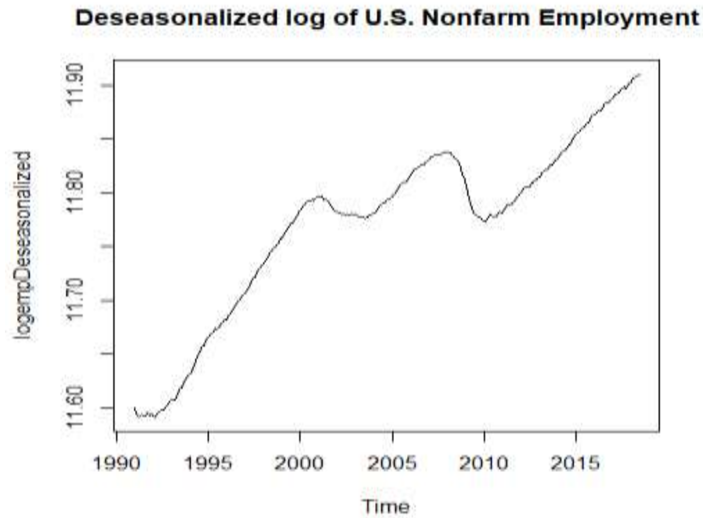
This residual acf plot indicates there are uncaptured trend and some dynamic seasonal structure; for the latter note the significant lag 12 correlation. The residual spectral plot follows.



The peak at low frequency reveals remaining trend structure, and the peaks at frequencies 4/12 and 5/12 are indicative of remaining dynamic seasonal structure. There is no remaining calendar structure.

Next, we attempt to use a smoothing technique to construct representations of the trend and noise components of the log of employment. First, we deseasonalize the data, using the estimated seasonal indices from the first model fit to the log data. To do so, we subtract the estimated log seasonal indices from the log data. The following plot shows the result. Note we are operating on the log scale.

```
> lseas<-c(rep(log(seas),27),log(seas)[1:6])
> plot(ts(logemp-
lseas,start=c(1991,1),freq=12),xlab="Time",ylab="logempDeseasonalized",main="
Deseasonalized log of U.S. Nonfarm Employment")
```



The result shown represents an estimate of the sum of the trend and noise components of the log of employment.

In the next step we apply a simple data smoother to dampen the noise, leaving an estimate of just the trend component. There are many options available for smoothing data. The aim is to replace a somewhat rough time trace by one that is smoother. For illustration, let's apply a running average of length five. That is, if the observation at time t is y_t, we replace it by the centered moving average

$$\frac{1}{5}\left(y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}\right).$$

This is an operation similar to that given by the `decompose` function in R. It is, of course, a shorter moving average than the one employed by `decompose`.

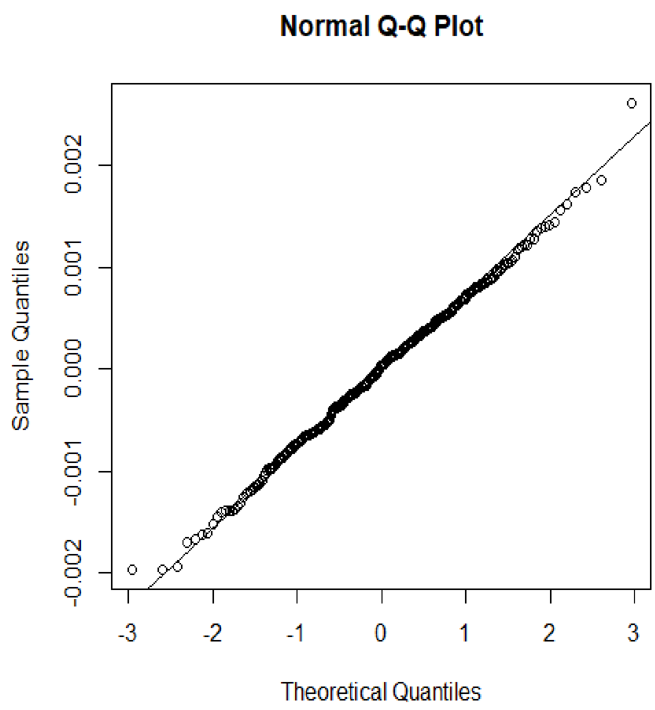
We apply this smoother to the deseasonalized log employment data. The result of the smoothing should be an estimate of the trend, because the smoothing operation diminishes the noise component. Then we subtract the result of the smooth from the deseasonalized log employment series. Hopefully this will provide reduction to just the noise component, which we hope will have white noise structure. A normal quantile plot and time series plot for the outcome of this procedure follow.

```
> #deseasonalize the log data and define the result as a time series (time
> #series class is required to use the lag operator)
> dslemp<-ts(logemp-lseas)

> #calculate a smooth of the deseasonalized log data, giving a trend estimate
> smdslemp<-(lag(dslemp,2)+lag(dslemp,1)+dslemp+lag(dslemp,-1)+lag(dslemp,-
2))/5

> #note that the class of smdslemp is time series
> class(smdslemp)
```

```
[1] "ts"

> #subtract smooth from the deseasonalized log data, giving a noise estimate
> rndm<-dslemp-smdslemp

> #note that the class of rndm is time series
> class(rndm)
[1] "ts"

> length(rndm)
[1] 326
```

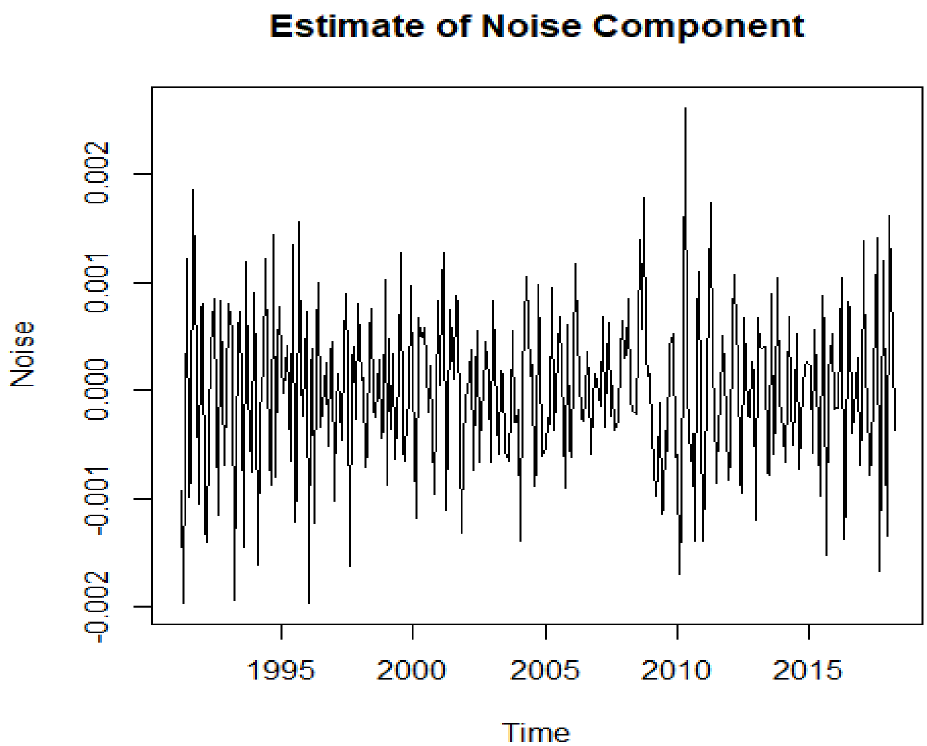
The use of the lag operator has resulted in the loss of the first two and the last two months of the time span.

```
> qqnorm(rndm)
> qqline(rndm)
```



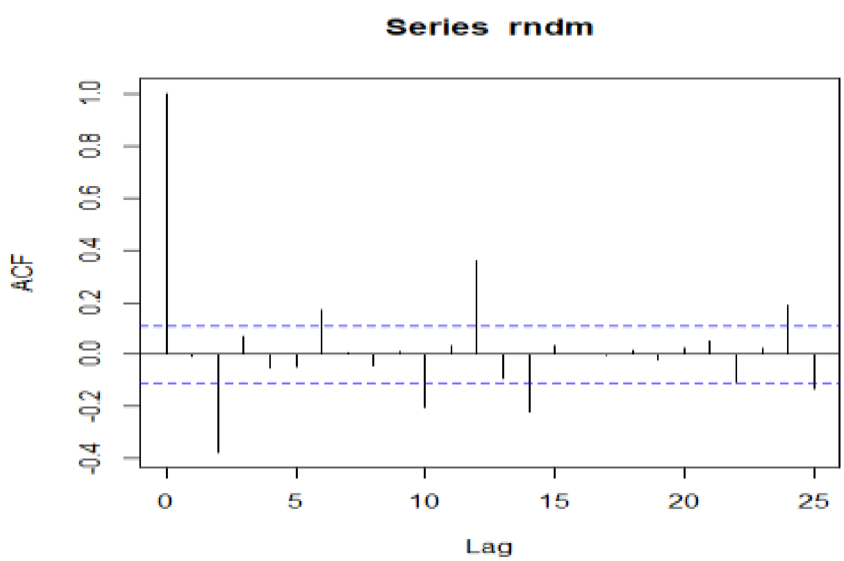
The residuals from the smooth operation estimate the noise component and are very well described by a normal distribution.

```
>
plot(ts(rndm,start=c(1991,3),freq=12),xlab="Time",ylab="Noise",main="Estimate
of Noise Component")
```



The result is not completely satisfactory. Trend estimation is off base during the 2008–2009 recession, and there is some amount of changing volatility. Let's look at the autocorrelation plot.

```
> acf(rndm)
```



Reduction to white noise has not been achieved. There are significant autocorrelations at lags 2, 6, 10, 12, 14, 24, and 25. The correlations at lags 6, 12, and 24 suggest some remaining seasonal structure has not been captured by the analysis. That is, the seasonal is dynamic rather than static.

Now let's employ the decomposition procedure supported by R. Recall that it uses a moving average of length 13 to dampen the trend. The model we use is multiplicative, as we want to work with the log of employment.

```
> emp.ts<-ts(employees,freq=12)
> emp.decmps<-decompose(emp.ts,type="mult")
> seasr<-emp.decmps$seasonal[1:12]/prod(emp.decmps$seasonal[1:12])^(1/12)
> prod(seasr)
[1] 1

> seasr
 [1] 0.9843595 0.9886928 0.9935261 0.9996480 1.0052323 1.0080483 0.9979983
 [8] 0.9982659 1.0024642 1.0072302 1.0085046 1.0063734
```

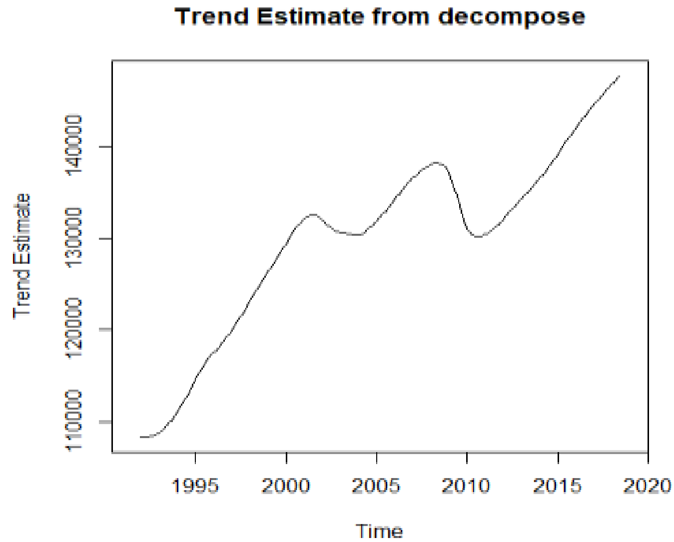
These seasonal index estimates are essentially identical to those obtained from model 1 and model 3.

Next, let's look at the trend.

```
> trnd<-emp.decmps$trend
> length(trnd)
[1] 330
> trnd[1:12]
 [1]       NA       NA       NA       NA       NA       NA 108397.3 108350.9
 [9] 108325.9 108329.0 108356.0 108388.5
> trnd[319:330]
 [1] 146709.2 146888.1 147078.0 147268.7 147463.2 147663.2       NA       NA
 [9]       NA       NA       NA       NA
```

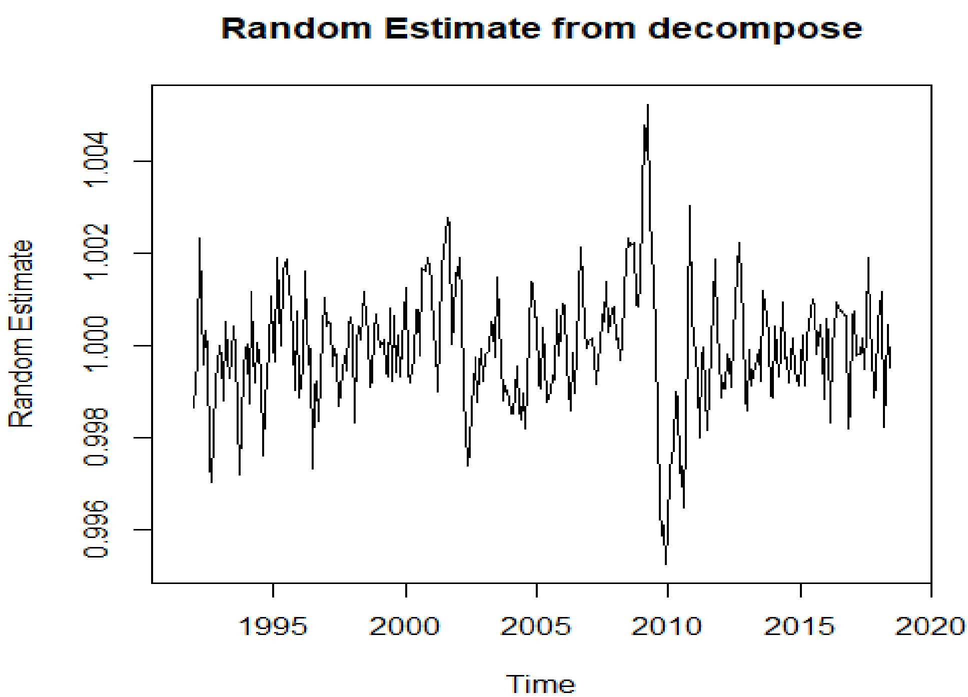
The trend estimation does not apply to the first six and the last six observations. So let's be careful in plotting the estimate.

```
> plot(ts(trnd,start=c(1991,7),freq=12),xlab="Time",ylab="Trend
Estimate",main="Trend Estimate from decompose")
```

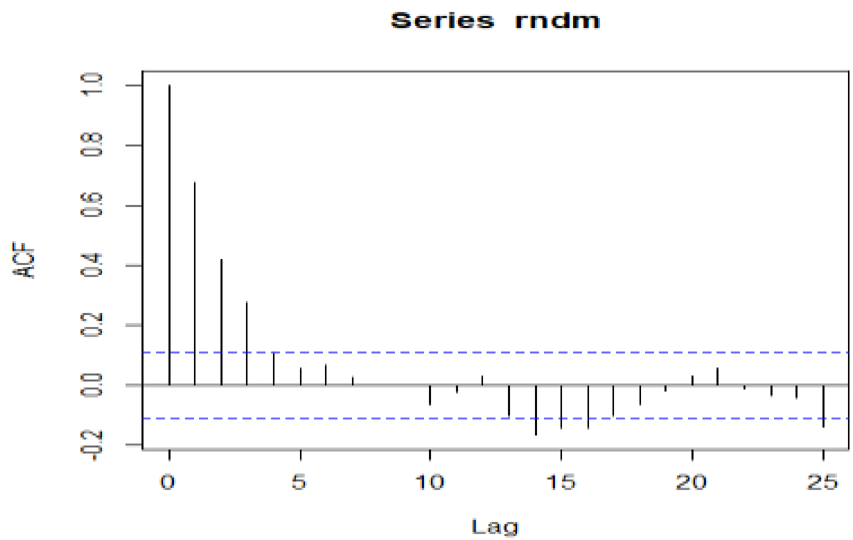


Finally, diagnostics for the random estimate are given.

```
> rndm<-emp.decmps$random
> plot(ts(rndm,start=c(1991,7),freq=12),xlab="Time",ylab="Random
Estimate",main="Random Estimate from decompose")
```



```
> rndm<-rndm[7:324]
> rndm.ts<-ts(rndm)
> acf(rndm)
```



This analysis does not adequately address the onset of the 2008–2009 recession.

The BLS monthly data are available starting in 1939.

Recall that the 24 January notes describe the procedures that R employs in using `decompose` to calculate the components of an additive decomposition model and of a multiplicative decomposition model. Here, in considering the employment data, a multiplicative decomposition model has been fit. Once the trend and seasonal components have been estimated, the random component (giving the residuals) is obtained by division, not by subtraction. That is, one divides the employment data by the trend and seasonal estimates.

Before ending this discussion, let's analyze just the first 16 years of data, 1991 through 2006. This will avoid structure stemming from the 2008–2009 recession. We use the log return series.

```
> emp<-data.frame(emp,Time,fMonth,c348,s348)

> model5<-
lm(dlogemp~Time+I(Time^2)+I(Time^3)+I(Time^4)+I(Time^5)+I(Time^6)+fMonth+c348
+s348,data=emp[1:192,]);summary(model5)

Call:
lm(formula = dlogemp ~ Time + I(Time^2) + I(Time^3) + I(Time^4) +
    I(Time^5) + I(Time^6) + fMonth + c348 + s348, data = emp[1:192,
    ])

Residuals:
       Min         1Q     Median         3Q        Max
-2.772e-03 -6.499e-04 -2.656e-05  6.711e-04  2.796e-03
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -2.543e-02  6.641e-04 -38.289  < 2e-16 ***
Time         5.092e-04  8.804e-05   5.784 3.36e-08 ***
I(Time^2)   -1.952e-05  3.930e-06  -4.966 1.64e-06 ***
I(Time^3)    3.720e-07  7.605e-08   4.892 2.28e-06 ***
I(Time^4)   -3.622e-09  7.148e-10  -5.066 1.04e-06 ***
I(Time^5)    1.694e-11  3.213e-12   5.272 4.01e-07 ***
I(Time^6)   -3.006e-14  5.530e-15  -5.436 1.84e-07 ***
fMonth2      2.624e-02  4.067e-04  64.529  < 2e-16 ***
fMonth3      2.699e-02  4.068e-04  66.354  < 2e-16 ***
fMonth4      2.811e-02  4.067e-04  69.111  < 2e-16 ***
fMonth5      2.811e-02  4.071e-04  69.048  < 2e-16 ***
fMonth6      2.550e-02  4.072e-04  62.614  < 2e-16 ***
fMonth7      1.181e-02  4.073e-04  28.996  < 2e-16 ***
fMonth8      2.199e-02  4.078e-04  53.918  < 2e-16 ***
fMonth9      2.699e-02  4.079e-04  66.181  < 2e-16 ***
fMonth10     2.629e-02  4.081e-04  64.431  < 2e-16 ***
fMonth11     2.333e-02  4.087e-04  57.094  < 2e-16 ***
fMonth12     2.041e-02  4.089e-04  49.906  < 2e-16 ***
c348        -1.414e-04  1.178e-04  -1.200   0.2316    
s348        -2.766e-04  1.174e-04  -2.355   0.0196 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.00115 on 172 degrees of freedom
Multiple R-squared:  0.9821,    Adjusted R-squared:  0.9802 
F-statistic: 498.1 on 19 and 172 DF,  p-value: < 2.2e-16
```

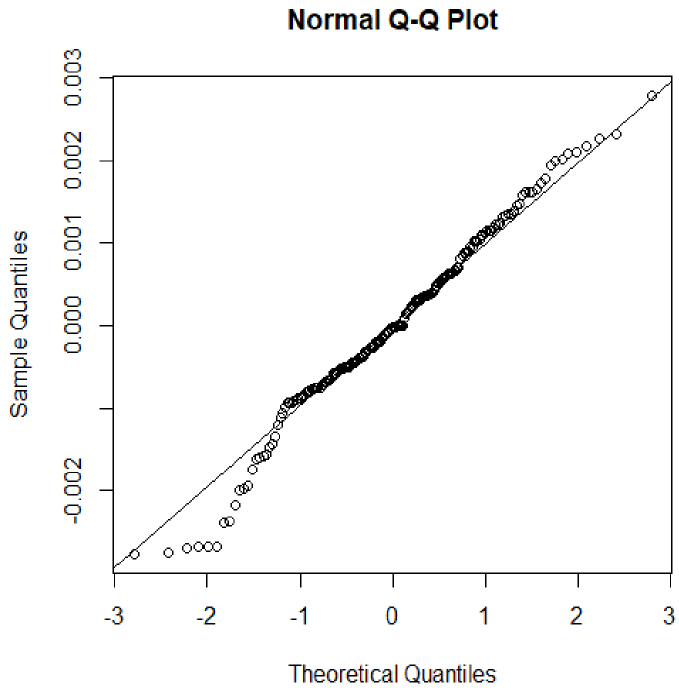
Calculation of seasonal index estimates for the employment data follows.

```
> b1<-coef(model5)[1]
> b2<-coef(model5)[8:18]+b1
> b3<-c(b1,b2)
> x<-b3-mean(b3)
> s12<-0
> for(j in 2:12){
+ xsub<-x[j:12]
+ s12<-s12+sum(xsub)
+ }
> s12<-s12/12
> s<-c(rep(0,times=12))
> s[12]<-s12
> for(j in 1:11){
+ xsub<-x[1:j]
+ s[j]<-s[12]+sum(xsub)
+ }
> seas5<-exp(s)
> seas5
 [1] 0.9844646 0.9885045 0.9933026 0.9992402 1.0052171 1.0085910 0.9982176
 [8] 0.9980568 1.0029057 1.0070712 1.0082662 1.0065118
```

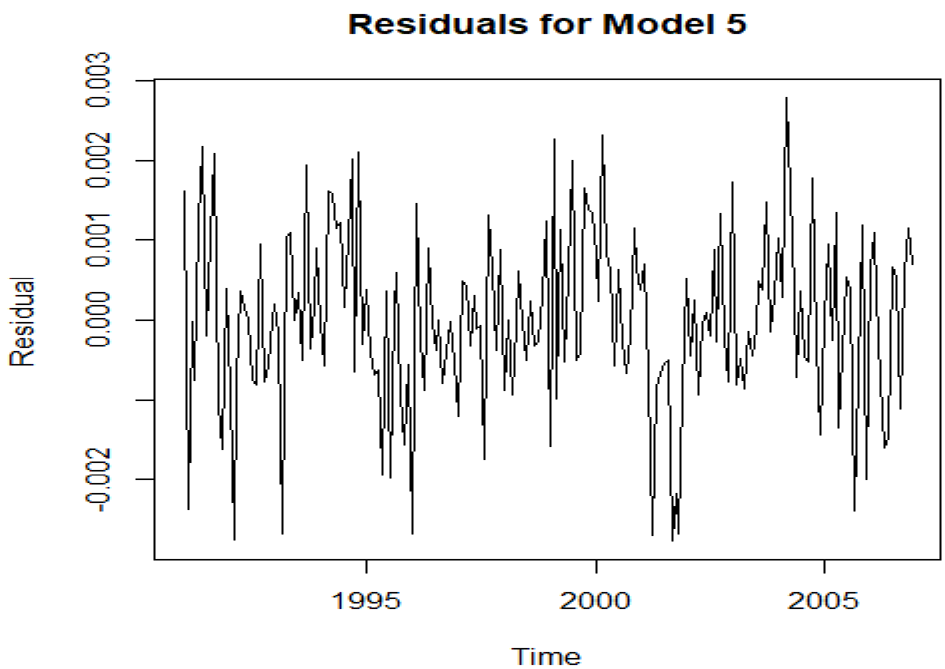
Again, the seasonal index estimates are essentially the same as those obtained before, with the first and third models, and the `decompose` function in R.

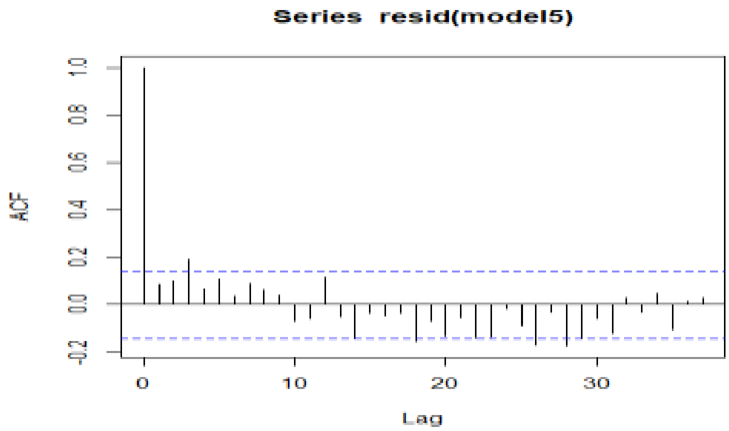
Diagnostics for the residuals from the model 5 fit to the 1991–2006 log return series follow.

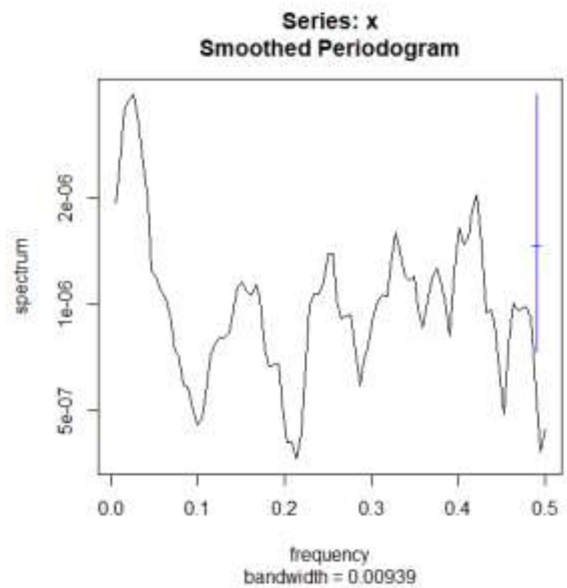
```
> qqnorm(resid(model5))
> qqline(resid(model5))
```




```
> shapiro.test(resid(model5))

        Shapiro-Wilk normality test

data:  resid(model5)
W = 0.98538, p-value = 0.04372
```


Residuals for Model 5
Residual
0.003
0.002
0.001
0.000
-0.002
1995
2000
2005
Time





The lower tail of the residual distribution is nonnormal. The plot of the residuals vs. time shows remaining trend structure, and this is also revealed in the spectral plot. However, the acf and spectral plots indicate that, despite the remaining trend, model 5 almost achieves reduction to white noise. In the spectral plot double the length of the segment of the blue line above the notch, and compare this doubled length to the vertical range of the estimated spectrum. If the doubled segment length exceeds the vertical range, we can say there is reduction to white noise by the model. By this approximate measure it appears that model 5 almost gives reduction to white noise, but not quite.

It is evident that the difficulties arising with models 1 and 3 stem from their inability to address properly features of the 2008–2009 recession. All of the three models also do not adequately address the downturn in employment from mid-2000 to the end of 2001. This downturn is less severe than that of the 2008–2009 recession.

Summary and additional remarks

1. The U.S. nonfarm employment monthly time series spans the time range 1991 to 2018(6). It shows two noticeable downturns. One of these originated in 2000, arising from several factors, including the Asian financial crisis and the bursting of the dot-com bubble. The second downturn, in 2008–2009, was more severe.

2. A multiplicative decomposition model is fit to the monthly employment data. Estimation of the trend structure with a polynomial in time is very poor during the two downturn periods. Despite poor estimation of trend structure via the multiplicative model, estimation of the static seasonal pattern appears to be reasonable.

3. The log return series contains prominent trend and seasonal components. That is, the differencing operation applied to the log data does not kill the trend structure, leaving trends around the two downturns. A model fit to the log return series is also not able to capture the trend structure during the two downturn periods. However, the estimated seasonal indices obtained from this model are in extremely close agreement with the seasonal index estimates from the multiplicative decomposition model. The estimation shows two seasonal peaks, one in May and June, and a second in October, November, and December. The low point is in January, where employment is estimated to be 1.5 per cent below the level of the trend.

4. A data smoother, using five adjacent months, is used to produce an estimate of the trend from the multiplicative decomposition model.

5. Two other models are fit to the log data, one using the decomposition procedure in R, and a second restricting analysis to the years 1991–2006. The seasonal index estimates obtained from these fits are consistent with the estimates obtained from models 1 and 3.