Solutions to Time Series Analysis: with Applications in R

Johan Larsson

2017-04-05

Contents

Preface

This book contains solutions to the problems in the book *Time Series Analysis: with Applications in R*, third edition, by Cryer and Chan. It is provided as a github repository so that anybody may contribute to its development.

Dependencies

You will need these packages to reproduce the examples in this book:

```r
install.packages(c(
  "latticeExtra",
  "lattice",
  "TSA",
  "pander"
))
```

```r
# Load the packages
library(latticeExtra)
library(TSA)
library(pander)
library(dplyr)
```

In order for some of the content in this book to be reproducible, the random seed is set as follows.

```r
set.seed(1234)
```

Chapter 1

Introduction

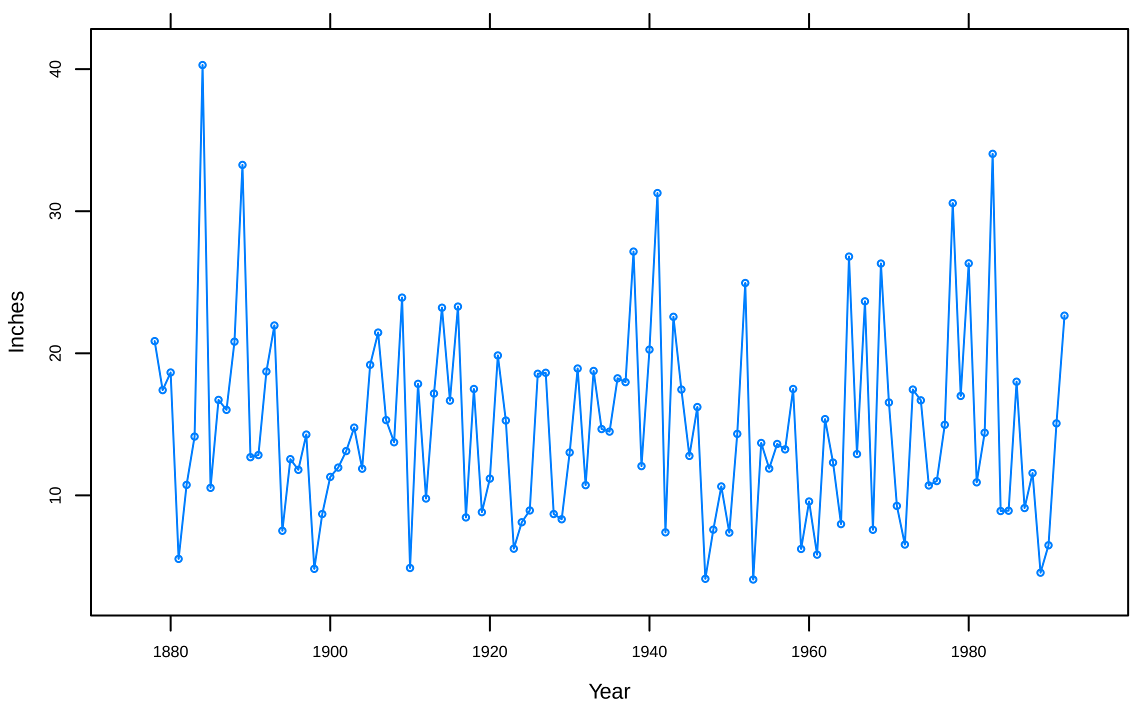
1.1 Larain

Use software to produce the time series plot shown in Exhibit 1.2, on page 2. The data are in the file named larain.

```r
library(TSA)
library(latticeExtra)

data(larain, package = "TSA")
```

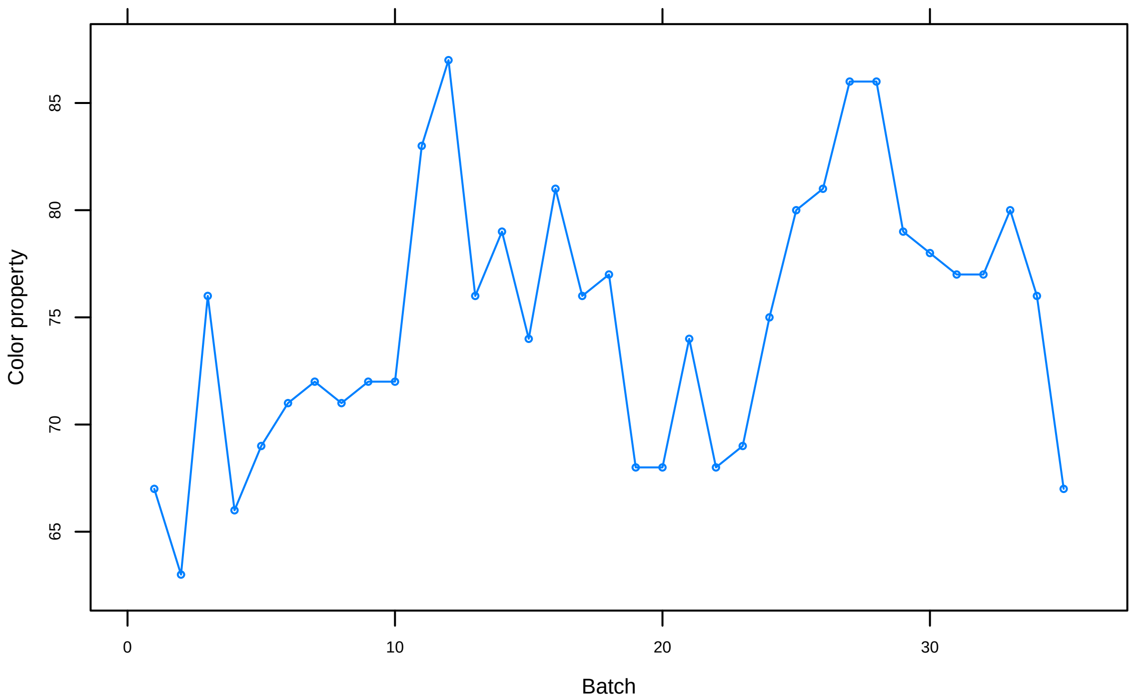
```r
xyplot(larain, ylab = "Inches", xlab = "Year", type = "o")
```



1.2 Colors

Produce the time series plot displayed in Exhibit 1.3, on page 3. The data file is named color.

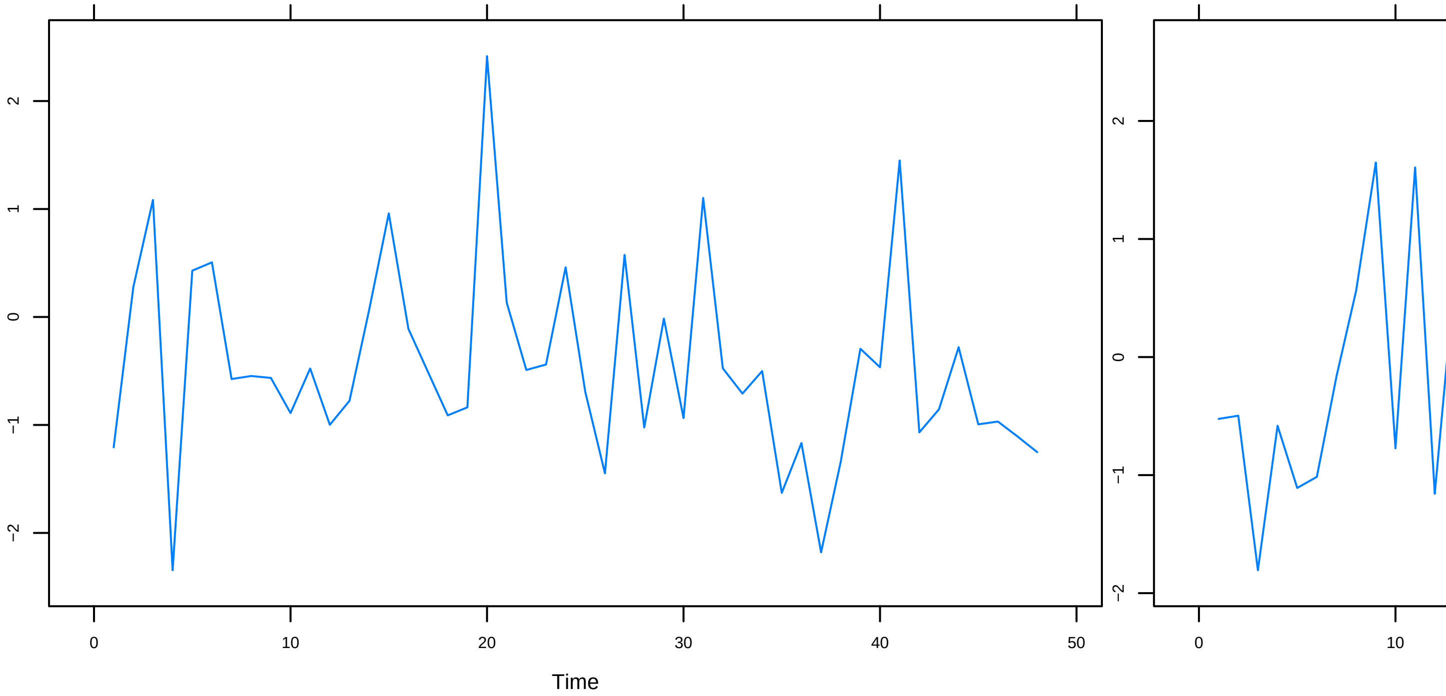
```
data(color)
xyplot(color, ylab = "Color property", xlab = "Batch", type = "o")
```



1.3 Random, normal time series

Simulate a completely random process of length 48 with independent, normal values. Plot the time series plot. Does it look "random"? Repeat this exercise several times with a new simulation each time.

```
xyplot(as.ts(rnorm(48)))
xyplot(as.ts(rnorm(48)))
```

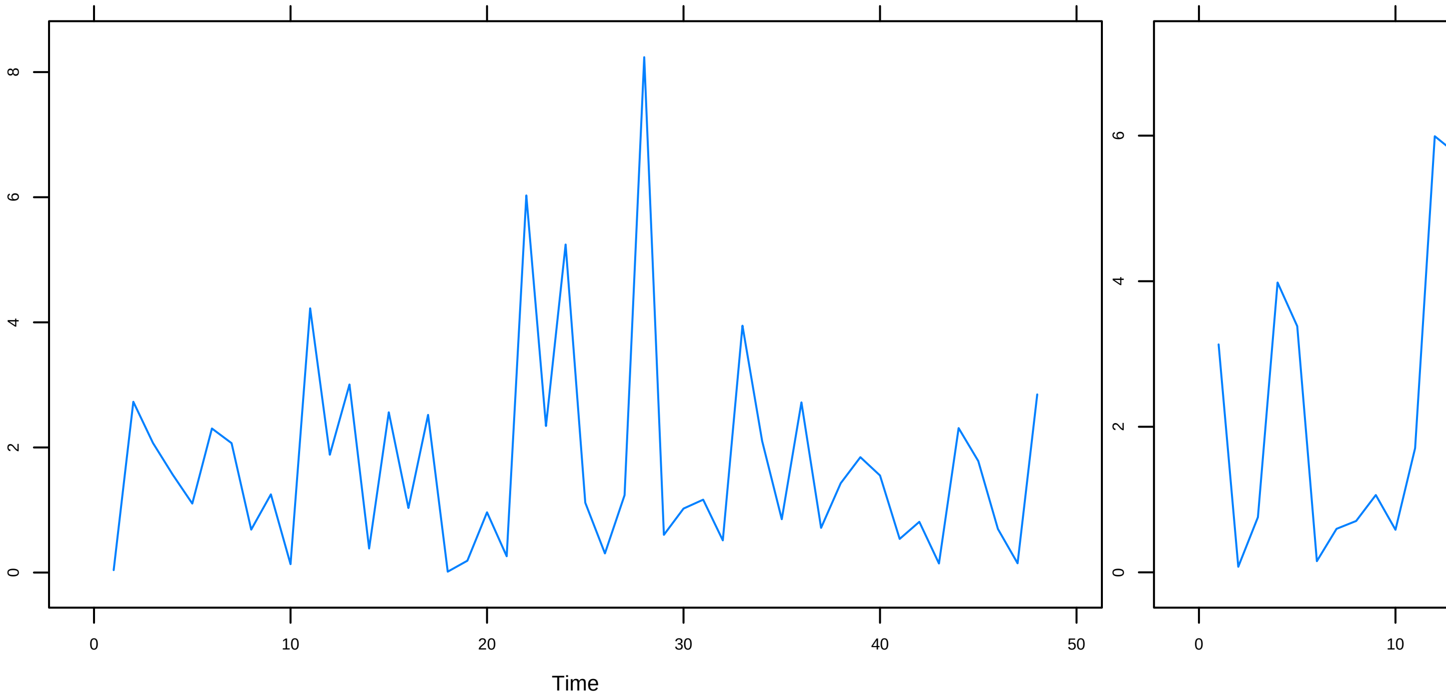


As far as we can tell there is no discernable pattern here.

1.4 Random, χ^2-distributed time series

Simulate a completely random process of length 48 with independent, chi-square distributed values, each with 2 degrees of freedom. Display the time series plot. Does it look “random” and nonnormal? Repeat this exercise several times with a new simulation each time.

```
xyplot(as.ts(rchisq(48, 2)))
xyplot(as.ts(rchisq(48, 2)))
```

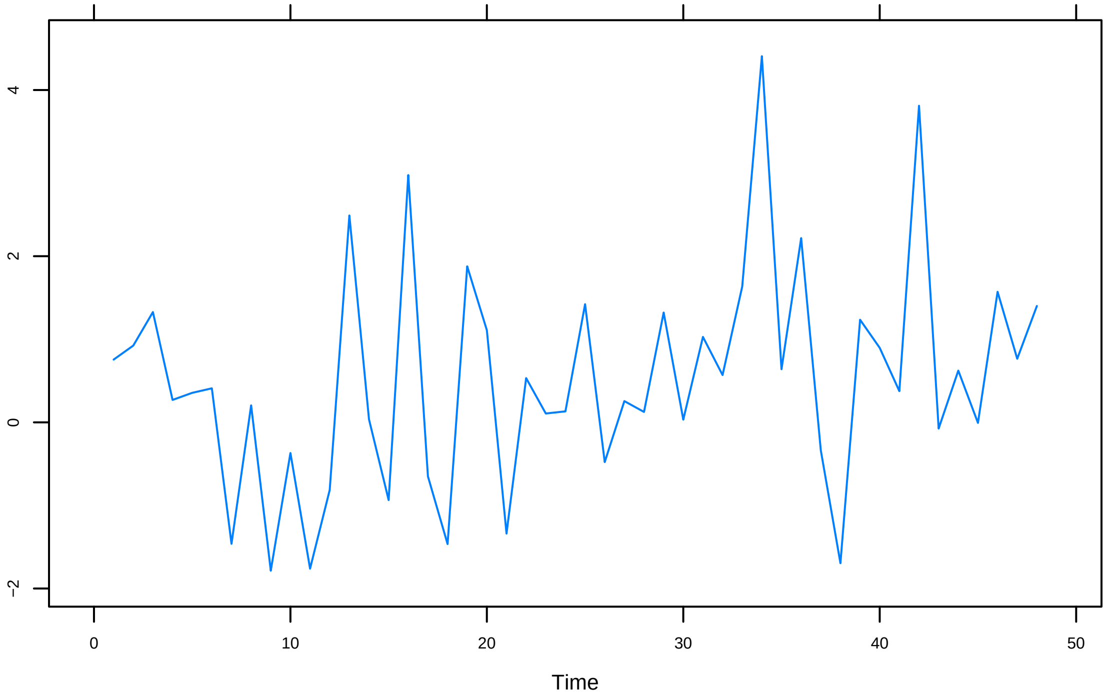


The process appears random, though non-normal.

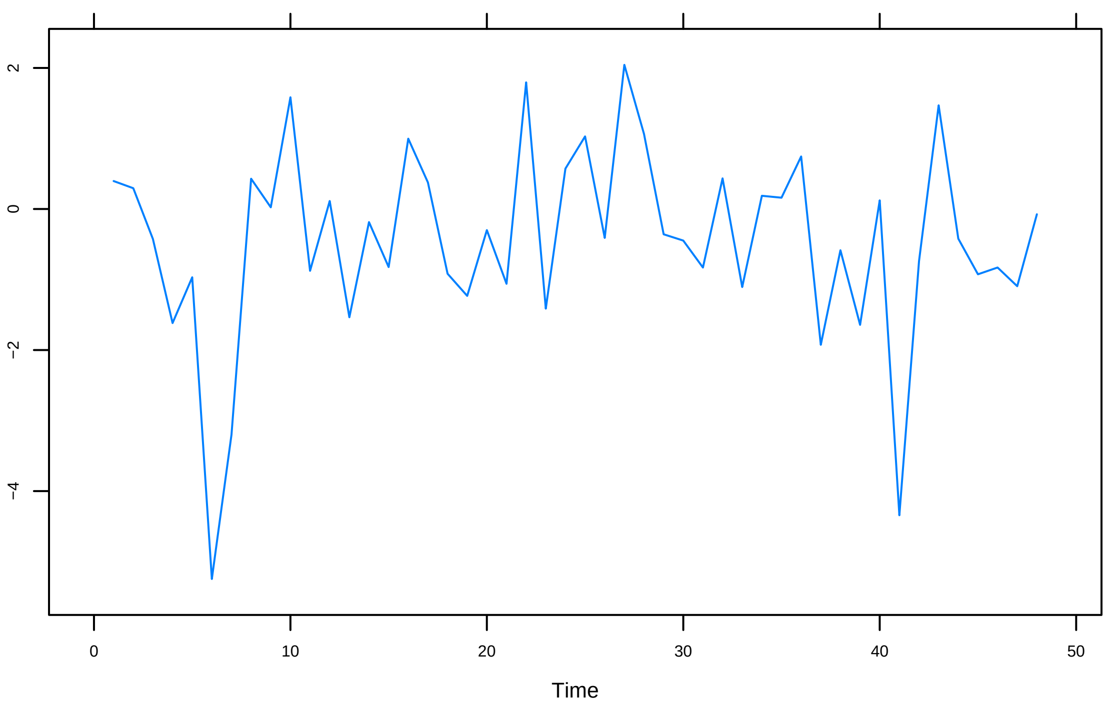
1.5 $t(5)$-distributed, random values

Simulate a completely random process of length 48 with independent, t-distributed values each with 5 degrees of freedom. Construct the time series plot. Does it look “random” and nonnormal? Repeat this exercise several times with a new simulation each time.

```
xyplot(as.ts(rt(48, 5)))
```




```r
xyplot(as.ts(rt(48, 5)))
```

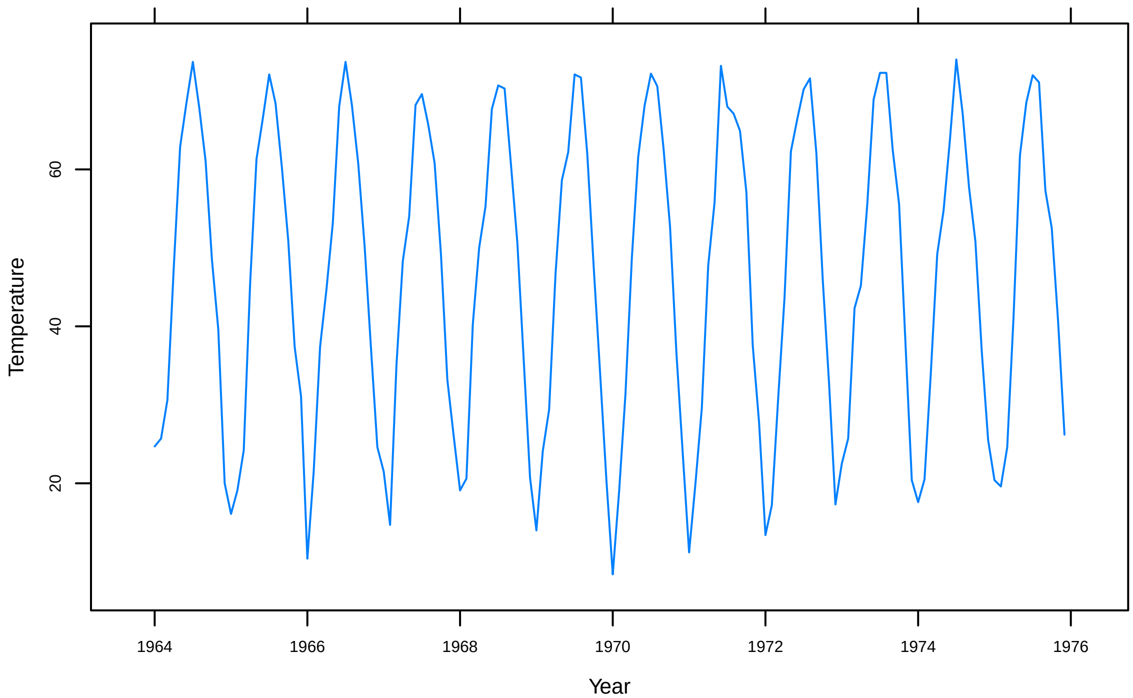


It looks random but not normal, though it should be approximately so, considering the distribution that we have sampled from.

1.6 Dubuque temperature series

Construct a time series plot with monthly plotting symbols for the Dubuque temperature series as in Exhibit 1.7, on page 6. The data are in the file named tempdub.

```
data(tempdub)
xyplot(tempdub, ylab = "Temperature", xlab = "Year")
```



Chapter 2

Fundamental concepts

2.1 Basic properties of expected value and covariance

a

$$\begin{aligned}
\text{Cov}[X,Y] &= \text{Corr}[X,Y]\sqrt{Var[X]Var[Y]} && (2.1)\\
&= 0.25\sqrt{9\times 4} = 1.5 && (2.2)\\
\text{Var}[X,Y] &= Var[X] + Var[Y] + 2Cov[X,Y] && (2.3)\\
&= 9 + 4 + 2\times 3 = 16 && (2.4)\\
& && (2.5)
\end{aligned}$$

b

$$\text{Cov}[X, X+Y] = \text{Cov}[X,X] + \text{Cov}[X,Y] = \text{Var}[X] + \text{Cov}[X,Y] = 9 + 1.5 = 10.5$$

c

$$\begin{aligned}
\text{Corr}[X+Y, X-Y] =& \text{Corr}[X,X] + \text{Corr}[X,-Y] + \text{Corr}[Y,X] + \text{Corr}[Y,-Y] && (2.6)\\
=& \text{Corr}[Y,X] + \text{Corr}[Y,-Y] && (2.7)\\
=& 1 - 0.25 + 0.25 - 1 && (2.8)\\
=& 0 && (2.9)\\
& && (2.10)
\end{aligned}$$

2.2 Dependence and covariance

$$\text{Cov}[X+Y, X-Y] = \text{Cov}[X,X] + \text{Cov}[X,-Y] + \text{Cov}[Y,X] + \text{Cov}[Y,-Y] = \\ Var[X] - Cov[X,Y] + Cov[X,Y] - Var[Y] = 0$$

since $Var[X] = Var[Y]$.

2.3 Strict and weak stationarity

a

We have that

$$
\begin{gathered}
P(Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}) = \\
P(X_1, X_2, \dots, X_n) = \\
P(Y_{t_1-k}, Y_{t_2-k}, \dots, Y_{t_n-k}),
\end{gathered}
$$

which satisfies our requirement for strict stationarity.

b

The autocovariance is given by

$$\gamma_{t,s} = \text{Cov}[Y_t, Y_s] = \text{Cov}[X, X] = \text{Var}[X] = \sigma^2.$$

c

```
library(lattice)
tstest <- ts(runif(100))

lattice::xyplot(tstest,
                panel = function(x, y, ...) {
                  panel.abline(h = mean(y), lty = 2)
                  panel.xyplot(x, y, ...)
                })
```

2.4 Zero-mean white noise

a

$$
\begin{gathered}
E[Y_t] = E[e_t + \theta e_{t-1}] = E[e_t] + \theta E[e_{t-1}] = 0 + 0 = 0 \\
V[Y_t] = V[e_t + \theta e_{t-1}] = V[e_t] + \theta^2 V[e_{t-1}] = \sigma_e^2 + \theta^2 \sigma_e^2 = \sigma_2^2(1 + \theta^2)
\end{gathered}
$$

For $k = 1$ we have

$$
\begin{gathered}
C[e_t + \theta e_{t-1}, e_{t-1} + \theta e_{t-2}] = \\
C[e_t, e_{t-1}] + C[e_t, \theta e_{t-2}] + C[\theta e_{t-1}, e_{t-1}] + C[\theta e_{t-1}, \theta e_{t-2}] = \\
0 + 0 + \theta V[e_{t-1}] + 0 = \theta \sigma_e^2, \\
\text{Corr}[Y_t, Y_{t-k}] = \frac{\theta \sigma_e^2}{\sqrt{(\sigma_e^2(1+\theta^2))^2}} = \frac{\theta}{1+\theta^2}
\end{gathered}
$$

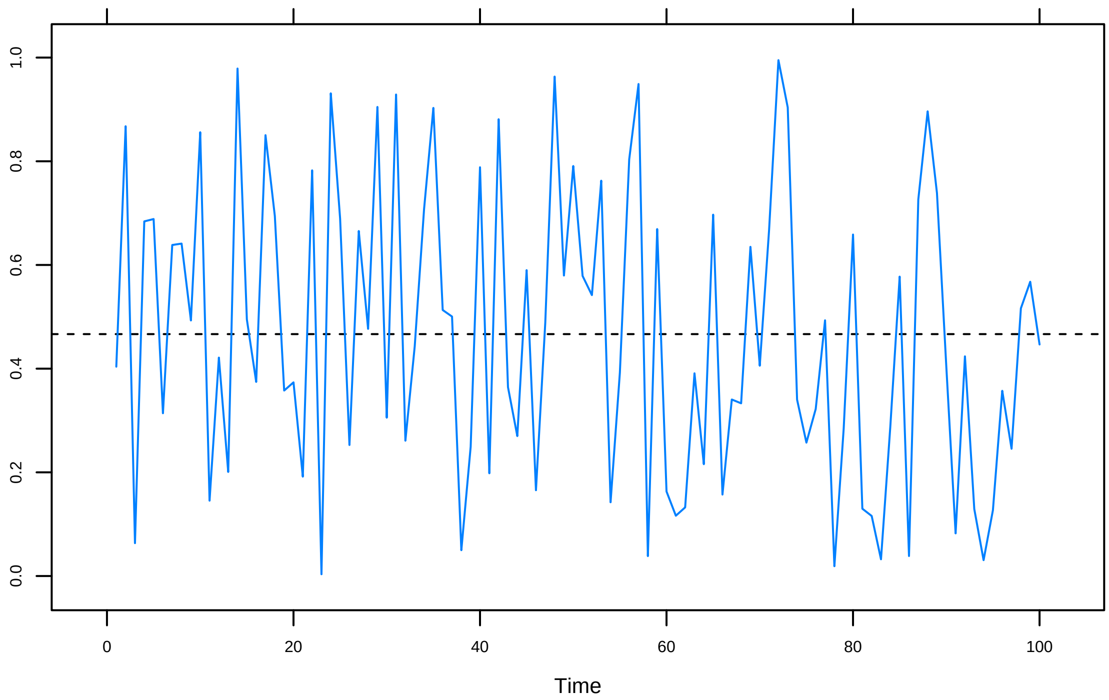


Figure 2.1: A white noise time series: no drift, independence between observations.

and for $k = 0$ we get

$$\text{Corr}[Y_t, Y_{t-k}] = \text{Corr}[Y_t, Y_t] = 1$$

and, finally, for $k > 0$:

$$C[e_t + \theta e_{t-1}, e_{t-k} + \theta e_{t-k-1}] =$$
$$C[e_t, e_{t-k}] + C[e_t, e_{t-1-k}] + C[\theta e_{t-1}, e_{t-k}] + C[\theta e_{t-1}, \theta e_{t-1-k}] = 0$$

given that all terms are independent. Taken together, we have that

$$\text{Corr}[Y_t, Y_{t-k}] = \begin{cases} 1 & \text{for } k = 0 \\ \frac{\theta}{1+\theta^2} & \text{for } k = 1 \\ 0 & \text{for } k > 1 \end{cases}.$$

And, as required,

$$\text{Corr}[Y_t, Y_{t-k}] = \begin{cases} \frac{3}{1+3^2} = \frac{3}{10} & \text{if } \theta = 3 \\ \frac{1/3}{1+(1/3)^2} = \frac{1}{10/3} = \frac{3}{10} & \text{if } \theta = 1/3 \end{cases}.$$

b

No, probably not. Given that ρ is standardized, we will not be able to detect any difference in the variance regardless of the values of k.

2.5 Zero-mean stationary series

a

$$\mu_t = E[Y_t] = E[5 + 2t + X_t] = 5 + 2E[t] + E[X_t] = 5 + 2t + 0 = 2t + 5$$

b

$$\gamma_k = \text{Corr}[5 + 2t + X_t, 5 + 2(t - k) + X_{t-k}] = \text{Corr}[X_t, X_{t-k}]$$

c

No, the mean function (μ_t) is constant and the aurocovariance ($\gamma_{t,t-k}$) free from t.

2.6 Stationary time series

a

$$\text{Cov}[a + X_t, b + X_{t-k}] = \text{Cov}[X_t, X_{t-k}],$$

which is free from t for all k because X_t is stationary.

b

$$\mu_t = E[Y_t] = \begin{cases} E[X_t] & \text{for odd } t \\ 3 + E[X_t] & \text{for even } t \end{cases}.$$

Since μ_t varies depending on t, Y_t is not stationary.

2.7 First and second-order difference series

a

$$\mu_t = E[W_t] = E[Y_t - Y_{t-1}] = E[Y_t] - E[Y_{t-1}] = 0$$

because Y_t is stationary.

$$\text{Cov}[W_t] = \text{Cov}[Y_t - Y_{t-1}, Y_{t-k} - Y_{t-1-k}] =$$
$$\text{Cov}[Y_t, Y_{t-k}] + \text{Cov}[Y_t, Y_{t-1-k}] + \text{Cov}[-Y_{t-k}, Y_{t-k}] + \text{Cov}[-Y_{t-k}, -Y_{t-1-k}] =$$
$$\gamma_k - \gamma_{k+1} - \gamma_{k-1} + \gamma_k = 2\gamma_k - \gamma_{k+1} - \gamma_{k-1}. \quad \square$$

b

In (a), we discovered that the difference between two stationary processes, ∇Y_t itself was stationary. It follows that the difference between two of these differences, $\nabla^2 Y_t$ is also stationary.

2.8 Generalized difference series

$$
\begin{aligned}
E[W_t] &= c_1 E[Y_t] + c_2 E[Y_t] + \cdots + c_n E[Y_t] && (2.11) \\
&= E[Y_t](c_1 + c_2 + \cdots + c_n), && (2.12)
\end{aligned}
$$

and thus the expected value is constant. Moreover,

$$
\begin{aligned}
\text{Cov}[W_t] &= \text{Cov}[c_1 Y_t + c_2 Y_{t-1} + \cdots + c_n Y_{t-k}, c_1 Y_{t-k} + c_2 Y_{t-k-1} + \cdots + c_n Y_{t-k-n}] && (2.13) \\
&= \sum_{i=0}^{n} \sum_{j=0}^{n} c_i c_j \text{Cov}[Y_{t-j} Y_{t-i-k}] && (2.14) \\
&= \sum_{i=0}^{n} \sum_{j=0}^{n} c_i c_j \gamma_{j-k-i}, && (2.15)
\end{aligned}
$$

which is free of t; consequently, W_t is stationary.

2.9 Zero-mean stationary difference series

a

$$E[Y_t] = \beta_0 + \beta_1 t + E[X_t] = \beta_0 + \beta_1 t + \mu_{t_x},$$

which is not free of t and hence *not* stationary.

$$\text{Cov}[Y_t] = \text{Cov}[X_t, X_t - 1] = \gamma_{t-1}$$

$$
\begin{gathered}
E[W_t] = E[Y_t - Y_{t-1}] = E[\beta_0 + \beta_1 t + X_t - (\beta_0 + \beta_1(t-1) + X_{t-1})] = \\
\beta_0 + \beta_1 t - \beta_0 - \beta_1 t + \beta_1 = \beta_1,
\end{gathered}
$$

is free of t and, furthermore, we have

$$
\begin{gathered}
\text{Cov}[W_t] = \text{Cov}[\beta_0 + \beta_1 t + X_t, \beta_0 + \beta_1(t-1) + X_{t-1}] = \\
\text{Cov}[X_t, X_{t-1}] = \gamma_k,
\end{gathered}
$$

which is also free of t, thereby proving that W_t is stationary.

b

$$E[Y_t] = E[\mu_t + X_t] = \mu_t + \mu_t = 0 + 0 = 0, \quad \text{and}$$
$$\text{Cov}[Y_t] = \text{Cov}[\mu_t + X_t, \mu_{t-k} + X_{t-k}] = \text{Cov}[X_t, X_{t-k}] = \gamma_k$$

$$\triangledown^m Y_t = \triangledown(\triangledown^{m1} Y_t)$$

Currently unsolved.

2.10 Zero-mean, unit-variance process

a

$$\mu_t = E[Y_t] = E[\mu_t + \sigma_t X_t] = \mu_t + \sigma_t E[X_t] = \mu_t + \sigma_t \times 0 = \mu_t$$
$$\gamma_{t,t-k} = \text{Cov}[Y_t] = \text{Cov}[\mu_t + \sigma_t X_t, \mu_{t-k} + \sigma_{t-k} X_{t-k}] = \sigma_t \sigma_{t-k} \text{Cov}[X_t, X_{t-k}] = \sigma_t \sigma_{t-k} \rho_k$$

b

First, we have

$$\text{Var}[Y_t] = \text{Var}[\mu_t + \sigma_t X_t] = 0 + \sigma_t^2 \text{Var}[X_t] = \sigma_t^2 \times 1 = \sigma_t^2$$

since $\{X_t\}$ has unit-variance. Futhermore,

$$\text{Corr}[Y_t, Y_{t-k}] = \frac{\sigma_t \sigma_{t-k} \rho_k}{\sqrt{\text{Var}[Y_t] \text{Var}[Y_{t-k}]}} = \frac{\sigma_t \sigma_{t-k} \rho_k}{\sigma_t \sigma_{t-k}} = \rho_k,$$

which depends only on the time lag, k. However, $\{Y_t\}$ is not necessarily stationary since μ_t may depend on t.

c

Yes, ρ_k might be free from t but if σ_t is not, we will have a non-stationary time series with autocorrelation free from t and constant mean.

2.11 Drift

a

$$\text{Cov}[X_t, X_{t-k}] = \gamma_k$$
$$E[X_t] = 3t$$

$\{X_t\}$ is not stationary because μ_t varies with t.

b

$$E[Y_t] = 3 - 3t + E[X_t] = 7 - 3t - 3t = 7$$
$$\mathrm{Cov}[Y_t, Y_{t-k}] = \mathrm{Cov}[7 - 3t + X_t, 7 - 3(t-k) + X_{t-k}] = \mathrm{Cov}[X_t, X_{t-k}] = \gamma_k$$

Since the mean function of $\{Y_t\}$ is constant (7) and its autocovariance free of t, $\{Y_t\}$ is stionary.

2.12 Periods

$$E[Y_t] = E[e_t - e_{t-12}] = E[e_t] - E[e_{t-12}] = 0$$
$$\mathrm{Cov}[Y_t, Y_{t-k}] = \mathrm{Cov}[e_t - e_{t-12}, e_{t-k} - e_{t-12-k}] =$$
$$\mathrm{Cov}[e_t, e_{t-k}] - \mathrm{Cov}[e_t, e_{t-12-k}] - \mathrm{Cov}[e_{t-12}, e_{t-k}] + \mathrm{Cov}[e_{t-12}, e_{t-12-k}]$$

Then, as required, we have

$$\mathrm{Cov}[Y_t, Y_{t-k}] = \begin{cases} \mathrm{Cov}[e_t, e_{t-12}] - \mathrm{Cov}[e_t, e_t] - \\ \mathrm{Cov}[e_{t-12}, e_{t-12}] + \mathrm{Cov}[e_{t-12}, e_t] = \\ \mathrm{Var}[e_t] - \mathrm{Var}[e_{t-12}] \neq 0 & \text{for } k = 12 \\ \\ \mathrm{Cov}[e_t, e_{t-k}] - \mathrm{Cov}[e_t, e_{t-12-k}] - \\ \mathrm{Cov}[e_{t-12}, e_{t-k}] + \mathrm{Cov}[e_{t-12}, e_{t-12-k}] = \\ 0 + 0 + 0 + 0 = 0 & \text{for } k \neq 12 \end{cases}$$

2.13 Drift, part 2

a

$$E[Y_t] = E[e_t - \theta e_{t-1}^2] = E[e_t] - \theta E[e_{t-1}^2] = 0 - \theta \mathrm{Var}[e_{t-1}] = -\theta\sigma_e^2$$

And thus the requirement of constant variance is fulfilled. Moreover,

$$\mathrm{Var}[Y_t] = \mathrm{Var}[e_t - \theta e_{t-1}^2] = \mathrm{Var}[e_t] + \theta^2 \mathrm{Var}[e_{t-1}^2] = \sigma_e^2 + \theta^2 (E[e_{t-1}^4] - E[e_{t-1}^2]^2),$$

where

$$E[e_{t-1}^4] = 3\sigma_e^4 \quad \text{and} \quad E[e_{t-1}^2]^2 = \sigma_e^4,$$

gives us

$$\mathrm{Var}[Y_t] = \sigma_e^2 + \theta(3\sigma_e^4 - \sigma_e^2) = \sigma_e^2 + 2\theta^2\sigma_e^4$$

and

$$
\begin{gathered}
\text{Cov}[Y_t, Y_{t-1}] = \text{Cov}[e_t - \theta e_{t-1}^2, e_{t-1} - \theta e_{t-2}^2] = \\
\text{Cov}[e_t, e_{t-1}] + \text{Cov}[e_t, -\theta e_{t-2}^2] + \text{Cov}[-\theta e_{t-1}^2, e_{t-1}]\text{Cov}[-\theta e_{t-1}^2, -\theta e_{t-2}^2] = \\
\text{Cov}[e_t, e_{t-1}] - \theta \text{Cov}[e_t, e_{t-2}^2] - \theta \text{Cov}[e_{t-1}^2, e_{t-1}] + \theta^2 \text{Cov}[e_{t-1}^2, e_{t-2}^2] = \\
-\theta \text{Cov}[e_{t-1}^2, e_{t-1}] = -\theta(E[e_{t-1}^3] + \mu_{t-1} + \mu_t) = 0
\end{gathered}
$$

which means that the autocorrelation function $\gamma_{t,s}$ also has to be zero.

b

The autocorrelation of $\{Y_t\}$ is zeor and its mean function is constant, thus $\{Y_t\}$ must be stationary.

2.14 Stationarity, again

a

$$
\begin{gathered}
E[Y_t] = E[\theta_0 + te_t] = \theta_0 + E[e_t] = \theta_0 + t \times 0 = \theta_0 \\
\text{Var}[Y_t] = \text{Var}[\theta_0] + \text{Var}[te_t] = 0 + t^2\sigma_e^2 = t^2\sigma_e^2
\end{gathered}
$$

So $\{Y_t\}$ is not stationary.

b

$$
\begin{gathered}
E[W_t] = E[\triangledown Y_t] = E[\theta_0 + te_t - \theta_0 - (t-1)e_{t-1}] = tE[e_t] - tE[e_{t-1} + E[e_{t-1}] = 0 \\
\text{Var}[\triangledown Y_t] = \text{Var}[te_t] = -\text{Var}[(t-1)e_{t-1}] = t^2\sigma_e^2 - (t-1)^2\sigma_e^2 = \sigma_e^2(t^2 - t^2 + 2t - 1) = (2t-1)\sigma_e^2,
\end{gathered}
$$

which varies with t and means that $\{W_t\}$ is not stationary.

c

$$
\begin{gathered}
E[Y_t] = E[e_t e_{t-1}] = E[e_t]E[e_{t-1}] = 0 \\
\text{Cov}[Y_t, Y_{t-1}] = \text{Cov}[e_t e_{t-1}, e_{t-1}e_{t-2}] = E[(e_t e_{t-1} - \mu_t^2)(e_{t-1}e_{t-2} - \mu_t^2)] = \\
E[e_t]E[e_{t-1}]E[e_{t-1}]E[e_{t-2}] = 0
\end{gathered}
$$

Both the covariance and the mean function are zero, hence the process is stationary.

2.15 Random variable, zero mean

a

$E[Y_t] = (-1)^t E[X] = 0$

b

$$\text{Cov}[Y_t, Y_{t-k}] = \text{Cov}[(-1)^t X, (-1)^{t-k} X] = (-1)^{2t-k}\text{Cov}[X, X] = (-1)^k \text{Var}[X] = (-1)^k \sigma_t^2$$

c

Yes, the covariance is free of t and the mean is constant.

2.16 Mean and variance

$$\begin{gathered} E[Y_t] = E[A + X_t] = E[A] + E[X_t] = \mu_A + \mu_X \\ \text{Cov}[Y_t, Y_{t-k}] = \text{Cov}[A + X_t, A + X_{t-k}] = \\ \text{Cov}[A, A] + \text{Cov}[A, X_{t-k}] + \text{Cov}[X_t, A] + \text{Cov}[X_t, X_{t-k}] = \sigma_A^2 + \gamma_{k_k} \end{gathered}$$

2.17 Variance of sample mean

$$\begin{gathered} \text{Var}[\bar{Y}] = \text{Var}\left[\frac{1}{n}\sum_{t=1}^{n} Y_t\right] = \frac{1}{n^2}\text{Var}\left[\sum_{t=1}^{n} Y_t\right] = \\ \frac{1}{n^2}\text{Cov}\left[\sum_{t=1}^{n} Y_t, \sum_{s=1}^{n} Y_s\right] = \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\gamma_{t-s} \end{gathered}$$

Setting $k = t - s, j = t$ gives us

$$\begin{gathered} \text{Var}[\bar{Y}] = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{j-k=1}^{n}\gamma_k = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{j=k+1}^{n+k}\gamma_k = \\ \frac{1}{n^2}\left(\sum_{k=1}^{n-1}\sum_{j=k+1}^{n}\gamma_k + \sum_{k=-n+1}^{0}\sum_{j=1}^{n+k}\gamma_k\right) = \\ \frac{1}{n^2}\left(\sum_{k=1}^{n-1}(n-k)\gamma_k + \sum_{k=-n+1}^{0}(n+k)\gamma_k\right) = \\ \frac{1}{n^2}\sum_{k=-n+1}^{n-1}((n-k)\gamma_k + (n+k)\gamma_k) = \\ \frac{1}{n^2}\sum_{k=-n+1}^{n-1}(n-|k|)\gamma_k = \frac{1}{n}\sum_{k=-n+1}^{n-1}\left(1 - \frac{|k|}{n}\right)\gamma_k \quad \square \end{gathered}$$

2.18 Sample variance

a

$$\sum_{t=1}^{n}(Y_t-\mu)^2=\sum_{t=1}^{n}((Y_t-\bar{Y})+(\bar{Y}-\mu))^2=$$
$$\sum_{t=1}^{n}((Y_t-\bar{Y})^2-2(Y_t-\bar{Y})(\bar{Y}-\mu)+(\bar{Y}-\mu)^2)=$$
$$n(\bar{Y}-\mu)^2+2(\bar{Y}-\mu)\sum_{t=1}^{n}(Y_t-\bar{Y})+\sum_{t=1}^{n}(Y_t-\bar{Y})^2=$$
$$n(\bar{Y}-\mu)^2+\sum_{t=1}^{n}(Y_t-\bar{Y})^2 \quad \square$$

b

$$E[s^2]=E\left[\frac{n}{n-1}\sum_{t=1}^{n}(Y_t-\bar{Y})^2\right]=\frac{n}{n-1}E\left[\sum_{t=1}^{n}\left((Y_t-\mu)^2+n(\bar{Y}-\mu)^2\right)\right]=$$
$$\frac{n}{n-1}\sum_{t=1}^{n}\left(E[(Y_t-\mu)^2]+nE[(\bar{Y}-\mu)^2]\right)=\frac{1}{n-1}\left(n\text{Var}[Y_t]-n\text{Var}[\bar{Y}]\right)=$$
$$\frac{n}{n-1}\gamma_0-\frac{n}{n-1}\text{Var}[\bar{Y}]=\frac{1}{n-1}\left(n\gamma_0-n\left(\frac{\gamma_0}{n}+\frac{2}{n}\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\gamma_k\right)\right)=$$
$$\frac{1}{n-1}\left(n\gamma_0-\gamma_0+2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\gamma_k\right)=\frac{1}{n-1}\left(\gamma_0(n-1)+2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\gamma_k\right)=$$
$$\gamma_0+\frac{2}{n-1}\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\gamma_k \quad \square$$

c

Since $\gamma_k=0$ for $k\neq 0$, in our case for all k, we have

$$E[s^2]=\gamma_0-\frac{2}{n-1}\sum_{t=1}^{n}\left(1-\frac{k}{n}\right)\times 0=\gamma_0$$

2.19 Random walk with drift

a

$$Y_1=\theta_0+e_1$$
$$Y_2=\theta_0+\theta_0+e_2+e_1$$
$$Y_t=\theta_0+\theta_0+\cdots+\theta_0+e_t+e_{t-1}+\cdots+e_1=$$
$$Y_t=t\theta_0+e_t+e_{t-1}+\cdots+e_1 \quad \square$$

b

$$\mu_t = E[Y_t] = E[t\theta_0 + e_t + e_{t-1} + \cdots + e_1] = t\theta_0 + E[e_t] + E[e_{t-1}] + \cdots + E[e_1] =$$
$$t\theta_0 + 0 + 0 + \cdots + 0 = t\theta_0$$

c

$$\gamma_{t,t-k} = \mathrm{Cov}[Y_t, Y_{t-k}] = \mathrm{Cov}[t\theta_0 + e_t, +e_{t-1} + \cdots + e_1, (t-k)\theta_0 + e_{t-k}, +e_{t-1-k} + \cdots + e_1] =$$
$$\mathrm{Cov}[e_{t-k}, +e_{t-1-k} + \cdots + e_1, e_{t-k}, +e_{t-1-k} + \cdots + e_1] \quad \text{(since all other terms are 0)} =$$
$$\mathrm{Var}[e_{t-k}, +e_{t-1-k} + \cdots + e_1, e_{t-k}, +e_{t-1-k} + \cdots + e_1] = (t-k)\sigma_e^2$$

2.20 Random walk

a

$$\mu_1 = E[Y_1] = E[e_1] = 0$$
$$\mu_2 = E[Y_2] = E[Y_1 - e_2] = E[Y_1] - E[e_2] = 0 - 0 = 0$$
$$\cdots$$
$$\mu_{t-1} = E[Y_{t-1}] = E[Y_{t-2} - e_{t-1}] = E[Y_{t-2}] - E[e_{t-1}] = 0$$
$$\mu_t = E[Y_t] = E[Y_{t-1} - e_t] = E[Y_t] - E[e_t] = 0,$$

which implies $\mu_t = \mu_{t-1}$ Q.E.D.

b

$$\mathrm{Var}[Y_1] = \sigma_e^2$$
$$\mathrm{Var}[Y_2] = \mathrm{Var}[Y_1 - e_2] = \mathrm{Var}[Y_1] + \mathrm{Var}[e_1] = \sigma_e^2 + \sigma_e^2 = 2\sigma_e^2$$
$$\cdots$$
$$\mathrm{Var}[Y_{t-1}] = \mathrm{Var}[Y_{t-2} - e_{t-1}] = \mathrm{Var}[Y_{t-2}] + \mathrm{Var}[e_{t-1}] = (t-1)\sigma_e^2$$
$$\mathrm{Var}[Y_t] = \mathrm{Var}[Y_{t-1} - e_t] = \mathrm{Var}[Y_{t-1}] + \mathrm{Var}[e_t] = (t-1)\sigma_e^2 + \sigma_e^2 = t\sigma_e^2 \quad \square$$

c

$$\mathrm{Cov}[Y_t, Y_s] = \mathrm{Cov}[Y_t, Y_t + e_{t+1} + e_{t+2} + \cdots + e_s] = \mathrm{Cov}[Y_t, Y_t] = \mathrm{Var}[Y_t] = t\sigma_e^2$$

2.21 Random walk with random starting value

a

$$E[Y_t] = E[Y_0 + e_t + e_{t-1} + \cdots + e_1] =$$
$$E[Y_0] + E[e_t] + E[e_{t-1}] + E[e_{t-2}] + \cdots + E[e_1] =$$
$$\mu_0 + 0 + \cdots + 0 = \mu_0 \quad \square$$

b

$$\begin{gathered}\mathrm{Var}[Y_t] = \mathrm{Var}[Y_0 + e_t + e_{t-1} + \cdots + e_1] = \\ \mathrm{Var}[Y_0] + \mathrm{Var}[e_t] + \mathrm{Var}[e_{t-1}] + \cdots + \mathrm{Var}[e_1] = \\ \sigma_0^2 + t\sigma_e^2 \quad \square\end{gathered}$$

c

$$\begin{gathered}\mathrm{Cov}[Y_t, Y_s] = \mathrm{Cov}[Y_t, Y_t + e_{t+1} + e_{t+2} + \cdots + e_s] = \\ \mathrm{Cov}[Y_t, Y_t] = \mathrm{Var}[Y_t] = \sigma_0^2 + t\sigma_e^2 \quad \square\end{gathered}$$

d

$$\mathrm{Corr}[Y_t, Y_s] = \frac{\sigma_0^2 + t\sigma_e^2}{\sqrt{(\sigma_0^2 + t\sigma_e^2)(\sigma_0^2 + s\sigma_e^2)}} = \sqrt{\frac{\sigma_0^2 + t\sigma_e^2}{\sigma_0^2 + s\sigma_e^2}} \quad \square$$

2.22 Asymptotic stationarity

a

$$\begin{gathered}E[Y_1] = E[e_1] = 0 \\ E[Y_2] = E[cY_1 + e_2] = cE[Y_1] + E[e_2] = 0 \\ \cdots \\ E[Y_t] = E[cY_{t-1} + e_t] = cE[Y_{t-1}] + E[e_t] = 0 \quad \square\end{gathered}$$

b

$$\begin{gathered}\mathrm{Var}[Y_1] = \mathrm{Var}[e_1] = \sigma_e^2 \\ \mathrm{Var}[Y_2] = \mathrm{Var}[cY_1 + e_2] = c^2\mathrm{Var}[Y_{t-1}] + \mathrm{Var}[e_2] = c^2\sigma_e^2 + \sigma_e^2 = \sigma_e^2(1 + c^2) \\ \cdots \\ \mathrm{Var}[Y_t] = \sigma_e^2(1 + c^2 + c^4 + \cdots + c^{2t-2}) \quad \square\end{gathered}$$

$\{Y_t\}$ is not stationary, given that its variance varies with t.

c

$$\begin{gathered}\mathrm{Cov}[Y_t, Y_{t-1}] = \mathrm{Cov}[cY_{t-1} + e_t, Y_{t-1}] = c\mathrm{Cov}[Y_{t-1}, Y_{t-1}] = c\mathrm{Var}[Y_{t-1}] \quad \text{giving} \\ \mathrm{Corr}[Y_t, Y_{t-1}] = \frac{c\mathrm{Var}[Y_{t-1}]}{\sqrt{\mathrm{Var}[Y_t]\mathrm{Var}[Y_{t-1}]}} = c\sqrt{\frac{\mathrm{Var}[Y_{t-1}]}{\mathrm{Var}[Y_t]}} \quad \square\end{gathered}$$

And, in the general case,

$$\begin{gathered}\operatorname{Cov}[Y_t, Y_{t-k}] = \operatorname{Cov}[cY_{t-1} + e_t, Y_{t-k}] = \\ c\operatorname{Cov}[cY_{t-2} + e_{t-1}, Y_{t-k}] = \\ c^3\operatorname{Cov}[Y_{t-2} + e_{t-1}, Y_{t-k}] = \dots \\ = c^k\operatorname{Var}[Y_{t-k}]\end{gathered}$$

giving

$$\operatorname{Corr}[Y_t, Y_{t-k}] = \frac{c^k\operatorname{Var}[Y_{t-k}]}{\sqrt{\operatorname{Var}[Y_t]\operatorname{Var}[Y_{t-k}]}} = c^k\sqrt{\frac{\operatorname{Var}[Y_{t-k}]}{\operatorname{Var}[Y_t]}} \quad \square$$

d

$$\operatorname{Var}[Y_t] = \sigma_e^2(1 + c^2 + c^4 + \dots + c^{2t-2}) = \sigma_e^2\sum_{t=1}^{n} c^{2(t-1)} = \sigma_e^2\sum_{t=0}^{n-1} c^{2t} = \sigma_e^2\frac{1-c^{2t}}{1-c^2}$$

And because

$$\lim_{t\to\infty}\sigma_e^2\frac{1-c^{2t}}{1-c^2} = \sigma_e^2\frac{1}{1-c^2} \quad \text{since } |c| < 1,$$

which is free of t, $\{Y_t\}$ can be considered *asymptotically* stationary.

e

$$\begin{gathered}Y_t = c(cY_{t-2} + e_{t-1}) + e_t = \dots = e_t + ce_{t-1} + c^2e_{t-2} + \dots + c^{t-2}e_2 + \frac{c^{t-1}}{\sqrt{1-c^2}}e_1 \\ \operatorname{Var}[Y_t] = \operatorname{Var}[e_t + ce_{t-1} + c^2e_{t-2} + \dots + c^{t-2}e_2 + \frac{c^{t-1}}{\sqrt{1-c^2}}e_1] = \\ \operatorname{Var}[e_t] + c^2\operatorname{Var}[e_{t-1}] + c^4\operatorname{Var}[e_{t-2}] + \dots + c^{2(t-2)}\operatorname{Var}[e_2] + \frac{c^{2(t-1)}}{1-c^2}\operatorname{Var}[e_1] = \\ \sigma_e^2(1 + c^2 + c^4 + \dots + c^{2(t-2)} + \frac{c^{2(t-1)}}{1-c^2}) = \sigma_e^2\left(\sum_{t=1}^{n} c^{2(t-1)} - c^{2(t-1)} + \frac{c^{2(t-1)}}{1-c^2}\right) = \\ \sigma_e^2\frac{1-c^{2t}+c^{2t-2+2}}{1-c^2} = \sigma_e^2\frac{1}{1-c^2} \quad \square\end{gathered}$$

Futhermore,

$$\begin{gathered}E[Y_1] = E\left[\frac{e_1}{\sqrt{1-c^2}}\right] = \frac{E[e_1]}{\sqrt{1-c^2}} = 0 \\ E[Y_2] = E[cY_1 + e_2] = cE[Y_1] = 0 \\ \dots \\ E[Y_t] = E[cY_{t-1} + e_2] = cE[Y_{t-1}] = 0,\end{gathered}$$

which satisfies our first requirement for weak stationarity. Also,

$$\mathrm{Cov}[Y_t, Y_{t-k}] = \mathrm{Cov}[cY_{t-1} + e_t, Y_{t-1}] = c^k \mathrm{Var}[Y_{t-1}] =$$
$$c^k \frac{\sigma_e^2}{1-c^2},$$

which is free of t and hence $\{Y_t\}$ is now stationary.

2.23 Stationarity in sums of stochastic processes

$$E[W_t] = E[Z_t + Y_t] = E[Z_t] + Y[Z_t] = \mu_{Z_t} + \mu_{Y_s}$$

Since both processes are stationary – and hence their sums are constant – the sum of both processes must also be constant.

$$\mathrm{Cov}[W_t, W_{t-k}] = \mathrm{Cov}[Z_t + Y_t, Z_{t-k} + Y_{t-k}] =$$
$$\mathrm{Cov}[Z_t, Z_{t-k}] + \mathrm{Cov}[Z_t, Y_{t-k}] + \mathrm{Cov}[Y_t, Z_{t-k}] + \mathrm{Cov}[Y_t, Y_{t-k}] =$$
$$\mathrm{Cov}[Z_t, Z_{t-k}] + \mathrm{Cov}[Z_t, Y_{t-k}] + \mathrm{Cov}[Y_t, Z_{t-k}] + \mathrm{Cov}[Y_t, Y_{t-k}] = \mathrm{Cov}[Z_t, Z_{t-k}] + \mathrm{Cov}[Y_t, Y_{t-k}] = \gamma_{Z_k} + \gamma_{Y_k},$$

both free of t.

2.24 Measurement noise

$$E[Y_t] = E[Y_t + e_t] = E[X_t] + E[e_t] - \mu_t$$
$$\mathrm{Var}[Y_t] = \mathrm{Var}[X_t + e_t] = \mathrm{Var}[X_t] + \mathrm{Var}[e_t] = \sigma_X^2 + \sigma_e^2$$
$$\mathrm{Cov}[Y_t, Y_{t-k}] = \mathrm{Cov}[X_t + e_t, X_{t-k} + e_{t-k}] = \mathrm{Cov}[X_t, X_{t-k}] = \rho_k$$
$$\mathrm{Corr}[Y_t, Y_{t-k}] = \frac{\rho_k}{\sqrt{(\sigma_X^2 + \sigma_e^2)(\sigma_X^2 + \sigma_e^2)}} = \frac{\rho_k}{\sigma_X^2 + \sigma_e^2} = \frac{\rho_k}{1 + \frac{\sigma_e^2}{\sigma_X^2}} \quad \square$$

2.25 Random cosine wave

$$E[Y_t] = E\left[\beta_0 + \sum_{i=1}^{k}(A_i\cos(2\pi f_i t) + B_i\sin(2\pi f_i t))\right] =$$
$$\beta_0 + \sum_{i=1}^{k}(E[A_i]\cos(2\pi f_i t) + E[B_i]\sin(2\pi f_i t) = \beta_0$$
$$\mathrm{Cov}[Y_t, Y_s] = \mathrm{Cov}\left[\sum_{i=1}^{k} A_i\cos(2\pi f_i t) + B_i\sin(2\pi f_i t), \sum_{j=1}^{k} A_j\cos(2\pi f_j s) + B_j\sin(2\pi f_j s)\right] =$$
$$\sum_{i=1}^{k}\mathrm{Cov}[A_i\cos(2\pi f_i t) + A_i\sin(2\pi f_i s)] + \sum_{i=1}^{k}\mathrm{Cov}[B_i\cos(2\pi f_j t) + B_i\sin(2\pi f_j s)] =$$
$$\sum_{i=1}^{k}\mathrm{Var}[A_i](\cos(2\pi f_i t) + \sin(2\pi f_i s)) + \sum_{i=1}^{k}\mathrm{Var}[B_i](\cos(2\pi f_j t) + \sin(2\pi f_j s)) =$$
$$\frac{\sigma_i^2}{2}\sum_{i=1}^{k}(\cos(2\pi f_i(t-s)) + \sin(2\pi f_i(t+s))) + \frac{\sigma_i^2}{2}\sum_{i=1}^{k}(\cos(2\pi f_j(t-s)) + \sin(2\pi f_j(t+s))) =$$
$$\sigma_i^2\sum_{i=1}^{k}\cos(2\pi f_i(t-s)) = \sigma_i^2\sum_{i=1}^{k}\cos(2\pi f_i k),$$

and is thus free of t and s.

2.26 Semivariogram

a

$$\Gamma_{t,s} = \frac{1}{2}E[(Y_t - Y_s)^2] = \frac{1}{2}E[Y_t^2 - 2Y_tY_s + Y_s^2] =$$
$$\frac{1}{2}\left(E[Y_t^2] - 2E[Y_tY_s] + E[Y_s^2]\right) = \frac{1}{2}\gamma_0 + \frac{1}{2}\gamma_0 - 2\times\frac{1}{2}\gamma_{|t-s|} = \gamma_0 - \gamma_{|t-s|}$$
$$\mathrm{Cov}[Y_t, Y_s] = E[Y_tY_s] - \mu_t\mu_s = E[Y_tY_s] = \gamma_{|t-s|} \quad \square$$

b

$$Y_t - Y_s = e_t + e_{t-1} + \cdots + e_1 - e_s - e_{s-1} - \cdots - e_1 =$$
$$e_t + e_{t-1} + \cdots + e_{s+1}, \quad \text{for } t > s$$
$$\Gamma_{t,s} = \frac{1}{2}E[(Y_t - Y_s)^2] = \frac{1}{2}\mathrm{Var}[e_t + e_{t-1} + \cdots + e_{s-1}] =$$
$$\frac{1}{2}\sigma_e^2(t-s) \quad \square$$

2.27 Polynomials

a

$$\begin{gathered}
E[Y_t] = E[e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \cdots + \phi^r e_{t-r}] = 0 \\
\mathrm{Cov}[Y_t, Y_{t-k}] = \mathrm{Cov}[e_t + \phi e_{t-1} + \cdots + \phi^r e_{t-r}, e_{t-k} + \phi e_{t-1-k} + \cdots + \phi^r e_{t-r-k}] = \\
\mathrm{Cov}[e_1 + \cdots + \phi^k e_{t-k} + \phi^{k+1} e_{t-k-1} + \cdots + \phi^r e_{t-r}, e_{t-r}, e_{t-k} + \cdots + \phi^k e_{t-k-1} + \cdots + \phi^r e_{t-k-r}] = \\
\sigma_e^2(\phi^k + \phi^{k+2} + \phi^{k+4} + \cdots + \phi^{k+2(r-k)}) = \sigma_e^2 \phi^k (1 + \phi^2 + \phi^4 + \cdots + \phi^{2(r-k)})
\end{gathered}$$

Hence, because of the zero mean and covariance free of t, it is a stationary process.

b

$$\begin{gathered}
\mathrm{Var}[Y_t] = \mathrm{Var}[e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \cdots + \phi^r e_{t-r}] = \sigma_e^2(1 + \phi + \phi^2 + \cdots + \phi^{2r}) \\
\mathrm{Corr}[Y_t, Y_{t-k}] = \frac{\sigma_e^2 \phi^k (1 + \phi^2 + \phi^4 + \cdots + \phi^{2(r-k)})}{\sqrt{(\sigma_e^2(1 + \phi + \phi^2 + \cdots + \phi^{2r}))^2}} = \frac{\phi^k (1 + \phi^2 + \phi^4 + \cdots + \phi^{2(r-k)})}{(1 + \phi + \phi^2 + \cdots + \phi^{2r})} \quad \square
\end{gathered}$$

2.28 Random cosine wave extended

a

$$\begin{gathered}
E[Y_t] = E[R\cos(2\pi(ft+\phi))] = E[R]\cos(2\pi(ft+\phi)) = \\
E[R]\int_0^1 \cos(E[R\cos(2\pi(ft+\phi))])d\phi = E[R]\left[\frac{1}{2\pi}\sin(2\pi(ft+\phi))\right]_0^1 = \\
E[R]\left(\frac{1}{2\pi}(\sin(2\pi(ft+1)) - \sin(2\pi(ft)))\right) = \\
E[R]\left(\frac{1}{2\pi}(\sin(2\pi ft + 2\pi) - \sin(2\pi ft + 1))\right) = \\
E[R]\,(0) = 0
\end{gathered}$$

b

$$\begin{gathered}
\gamma_{t,s} = E[R\cos(2\pi(ft+\phi))R\cos(2\pi(fs+\phi))] = \\
\frac{1}{2}E[R^2]\int_0^1 \left(\cos(2\pi(f(t-s)) + \frac{1}{4\pi}\sin(2\pi(f(t+s)+2\phi))\right) = \\
\frac{1}{2}E[R^2]\left[\cos(2\pi f(t-s)) + \frac{1}{4\pi}\sin(2\pi(f(t+s)+2\phi))\right]_0^1 = \\
\frac{1}{2}E[R^2]\left(\cos(2\pi(f|t-s|))\right),
\end{gathered}$$

which is free of t.

2.29 Random cosine wave further

a

$$E[Y_t] = \sum_{j=1}^{m} E[R_j]E[\cos(2\pi(f_j t + \phi))] = \text{via } 2.28 = \sum_{j=1}^{m} E[R_j] \times 0 = 0$$

b

$$\gamma_k = \sum_{j=1}^{m} E[R_j]\cos(2\pi f_j k), \text{ also from 2.28.}$$

2.30 Rayleigh distribution

$$Y = R\cos(2\pi(ft+\phi)), \quad X = R\sin(2\pi(ft+\phi))$$
$$\begin{bmatrix} \frac{\partial X}{\partial R} & \frac{\partial X}{\partial \Phi} \\ \frac{\partial Y}{\partial R} & \frac{\partial X}{\partial \Phi} \end{bmatrix} = \begin{bmatrix} \cos(2\pi(ft+\Phi)) & 2\pi R\sin(2\pi(ft+\Phi)) \\ \sin(2\pi(ft+\Phi)) & 2\pi R\cos(2\pi(ft+\Phi)) \end{bmatrix},$$

with jacobian

$$-2\pi R = -2\pi\sqrt{X^2+Y^2}$$

and inverse Jacobian

$$\frac{1}{-2\pi\sqrt{X^2+Y^2}}.$$

Furthermore,

$$f(r,\Phi) = re^{-r^2/2}$$

and

$$f(x,y) = \frac{e^{-(x^2+y^2)/2}\sqrt{x^2+y^2}}{2\pi\sqrt{x^2+y^2}} = \frac{e^{-x^2/2}}{\sqrt{2\pi}}\frac{e^{-y^2/2}}{\sqrt{2\pi}} \quad \square$$

Chapter 3

Trends

3.1 Least squares estimation for linear regression trend

We begin by taking the partial derivatives with respect to β_0.

$$\frac{\partial}{\partial \beta_0}\mathcal{Q}(\beta_0, \beta_1) = -2\sum_{t=1}^{n}(Y_t - \beta_0 - \beta_1 t)$$

We set it to 0 and from this retrieve

$$\begin{aligned}
-2\sum_{t=1}^{n}(Y_t - \beta_0 - \beta_1 t) =& 0 \implies \\
\sum_{t=1}^{n} Y_t - n\beta_0 - \beta_1 \sum_{t=1}^{n} t =& 0 \implies \\
\beta_0 = \frac{\sum_{t=1}^{n} Y_t - \beta_1 \sum_{t=1}^{n} t}{n} =& \bar{Y} - \beta_1 \bar{t}
\end{aligned}$$

Next, we take the partial derivative with respect to β_1;

$$\frac{\partial}{\partial \beta_1}\mathcal{Q}(\beta_0, \beta_1) = -2\sum_{t=1}^{n} t(Y_t - \beta_0 - \beta_1 t)$$

Setting this to 0 as well, multiplying both sides with $-1/2$ and rearranging results in

$$\begin{aligned}
-2\sum_{t=1}^{n} t(Y_t - \beta_0 - \beta_1 t) =& 0 \implies \\
\beta_1 \sum_{t=1}^{n} t^2 =& \sum_{t=1}^{n} Y_t t - \beta_0 \sum_{t=1}^{n} t
\end{aligned}$$

Then, substituting with the result gained previously for β_0, we get

$$\beta_1 \sum_{t=1}^n t^2 = \sum_{t=1}^n Y_t t - \left(\frac{\sum_{t=1}^n Y_t}{n} - \beta_1 \frac{\sum_{t=1}^n}{n} \right) \sum_{t=1}^n t \iff$$

$$\beta_1 \left(\sum_{t=1}^n t^2 - \frac{(\sum_{t=1}^n t)^2}{n} \right) = \sum_{t=1}^n Y_t t - \frac{\sum_{t=1}^n Y_t \sum_{t=1}^n t}{n} \iff$$

$$\beta_1 = \frac{n \sum_{t=1}^n Y_t t - \sum_{t=1}^n Y_t \sum_{t=1}^n t}{n \sum_{t=1}^n t^2 - \left(\sum_{t=1}^n t\right)^2} = \frac{\sum_{t=1}^n (Y_t - \bar{Y})(t - \bar{t})}{\sum_{t=1}^n (t - \bar{t})^2} \quad \square$$

3.2 Variance of mean estimator

$$\bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t = \frac{1}{n} \sum_{t=1}^n (\mu + e_t - e_{t-1}) = \mu + \frac{1}{n} \sum_{t=1}^n (e_t - e_{t-1}) = \mu + \frac{1}{n}(e_n - e_0)$$

$$\mathrm{Var}[\bar{Y}] = \mathrm{Var}[\mu + \frac{1}{n}(e_n - e_0)] = \frac{1}{n^2}(\sigma_e^2 + \sigma_e^2) = \frac{2\sigma_e^2}{n^2}$$

It is uncommon for the sample size to have such a large impact on the variance estimator for the sample mean.

Setting $Y_t = \mu + e_t$ instead gives

$$\bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t = \frac{1}{n} \sum_{t=1}^n (\mu + e_t) = \mu + \frac{1}{n} \sum_{t=1}^n e_t$$

$$\mathrm{Var}[\bar{Y}] = \mathrm{Var}\left[\mu + \frac{1}{n} \sum_{t=1}^n e_t\right] = 0 + \frac{1}{n^2} \times n\sigma_e^2 = \frac{\sigma_e^2}{n}.$$

3.3 Variance of mean estimator #2

$$\bar{Y} = \frac{1}{n} \sum_{t=1}^n (\mu + e_t + e_{t-1}) = \mu + \frac{1}{n} \sum_{t=1}^n (e_t + e_{t-1}) = \mu + \frac{1}{n} \left(e_n + e_0 + 2 \sum_{t=1}^{n-1} t \right)$$

$$\mathrm{Var}[\bar{Y}] = \frac{1}{n^2}(\sigma_e^2 + \sigma_e^2 + 4(n-1)\sigma_e^2) = \frac{1}{n^2} 2(2n-1)\sigma_e^2$$

Setting $Y_t = \mu + e_t$ instead gives the result from 3.2. We note that for large n the variance if approximately four times larger with $Y_t = \mu + e_t + e_{t-1}$.

3.4 Hours

a

```
library(TSA)
data("hours")
xyplot(hours)
```

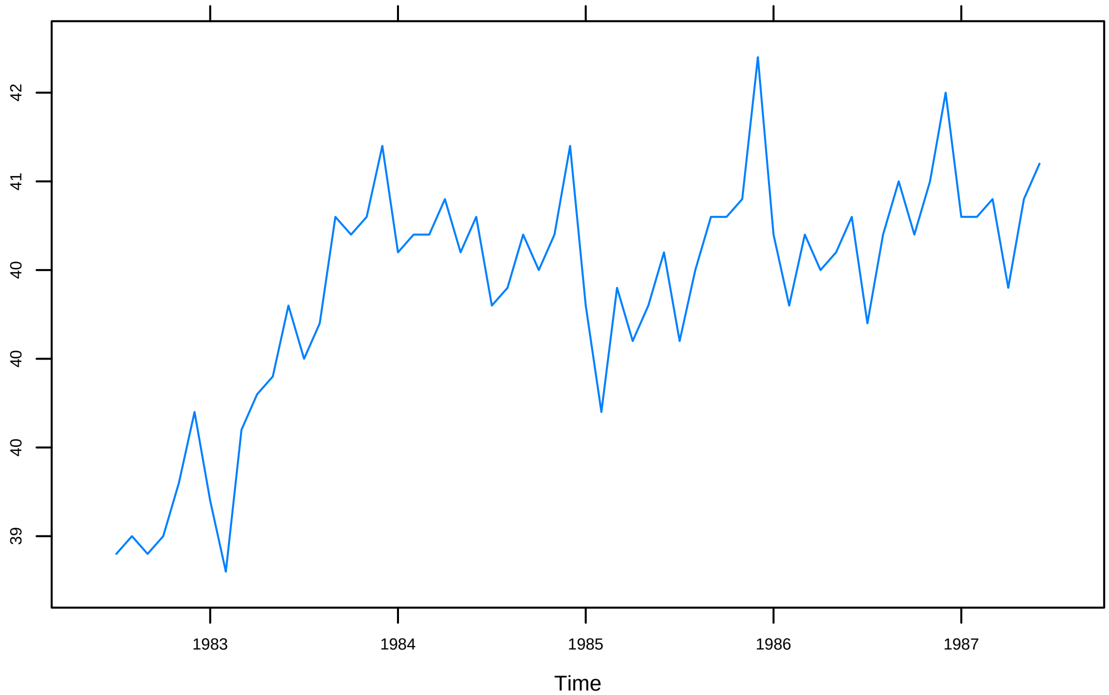


Figure 3.1: Monthly values of the average hours worked per week in the U.S. manufacturing sector.

In Figure 1 we see a steep incline between 83 and 84. There also appears to be a seasonal trend with generally longer work hours later in the year apart from the summer; 1984, however, does not exhibit as clear a pattern.

b

```
months <- c("J", "A", "S", "O", "N", "D", "J", "F", "M", "A", "M", "J")

xyplot(hours, panel = function(x, y, ...) {
  panel.xyplot(x, y, ...)
  panel.text(x = x, y = y, labels = months)
})
```

Here, in Figure 2, our interpretation is largely the same. It is clear that December stands out as the month with the longest weekly work hours whilst February and January are low-points, demonstrating a clear trend.

3.5 Wages

a

```
data("wages")
xyplot(wages, panel = function(x, y, ...) {
  panel.xyplot(x, y, ...)
```

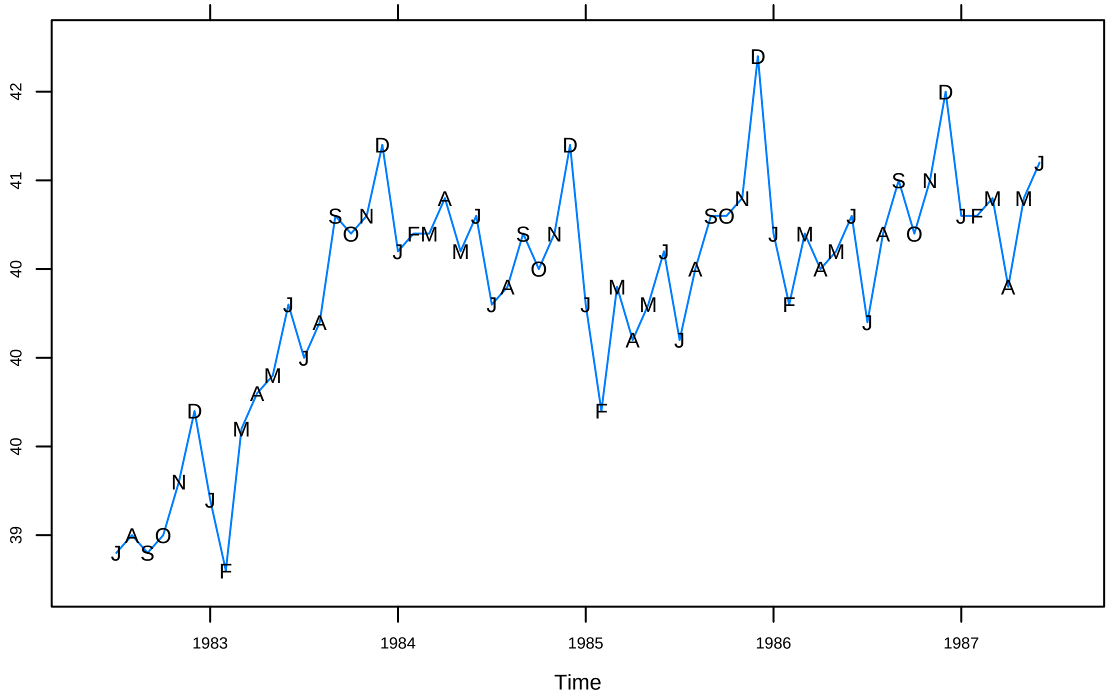


Figure 3.2: Monthly values of average hours worked per week with superposed initials of months.

```
  panel.text(x, y, labels = months)
})
```

There is a positive trend with seasonality: August is a low-point for wages. Generally, there seems to be larger increases in the fall.

b

```
wages_fit1 <- lm(wages ~ time(wages))
summary(wages_fit1)
```

```
##
## Call:
## lm(formula = wages ~ time(wages))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.2383 -0.0498  0.0194  0.0585  0.1314
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.49e+02   1.11e+01   -49.2   <2e-16 ***
## time(wages)  2.81e-01   5.62e-03    50.0   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

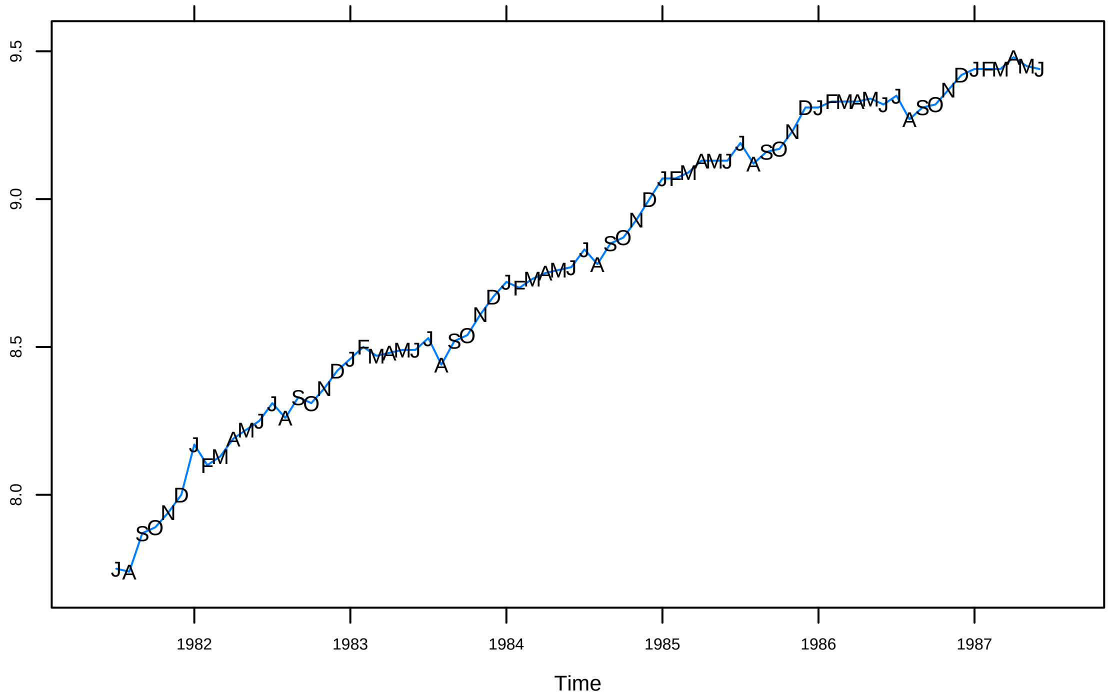


Figure 3.3: Monthly average hourly wages for workers in the U.S. apparel and textile industry.

```
## Residual standard error: 0.083 on 70 degrees of freedom
## Multiple R-squared:  0.973,  Adjusted R-squared:  0.972
## F-statistic: 2.5e+03 on 1 and 70 DF,  p-value: <2e-16
```

```
wages_rst <- rstudent(wages_fit1)
```

c

```
xyplot(wages_rst ~ time(wages_rst), type = "l",
       xlab = "Time", ylab = "Studentized residuals")
```

We still seem to have autocorrelation related to the time and not white noise.

d

```
wages_fit2 <- lm(wages ~ time(wages) + I(time(wages)^2))
summary(wages_fit2)
```

```
##
## Call:
## lm(formula = wages ~ time(wages) + I(time(wages)^2))
##
## Residuals:
```

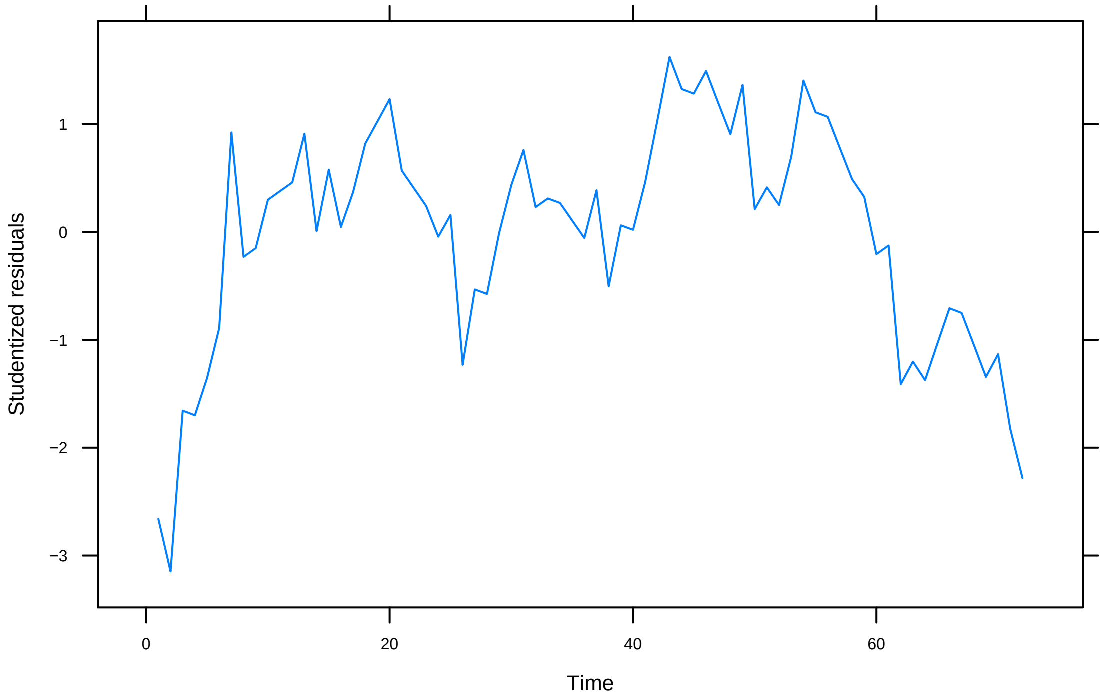


Figure 3.4: (#fig:wages_resid)Residual plot

```
##      Min       1Q   Median       3Q      Max
## -0.14832 -0.04144  0.00156  0.05009  0.13984
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -8.49e+04   1.02e+04   -8.34  4.9e-12 ***
## time(wages)       8.53e+01   1.03e+01    8.31  5.4e-12 ***
## I(time(wages)^2) -2.14e-02   2.59e-03   -8.28  6.1e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.059 on 69 degrees of freedom
## Multiple R-squared:  0.986,  Adjusted R-squared:  0.986
## F-statistic: 2.49e+03 on 2 and 69 DF,  p-value: <2e-16
```

```
wages_rst2 <- rstudent(wages_fit2)
```

e

```
xyplot(wages_rst2 ~ time(wages_rst), type = "l",
       xlab = "Time", ylab = "Studentized residuals")
```

This looks more like random noise but there is still clear autocorrelation between the fitted residuals that we have yet to capture in our model.

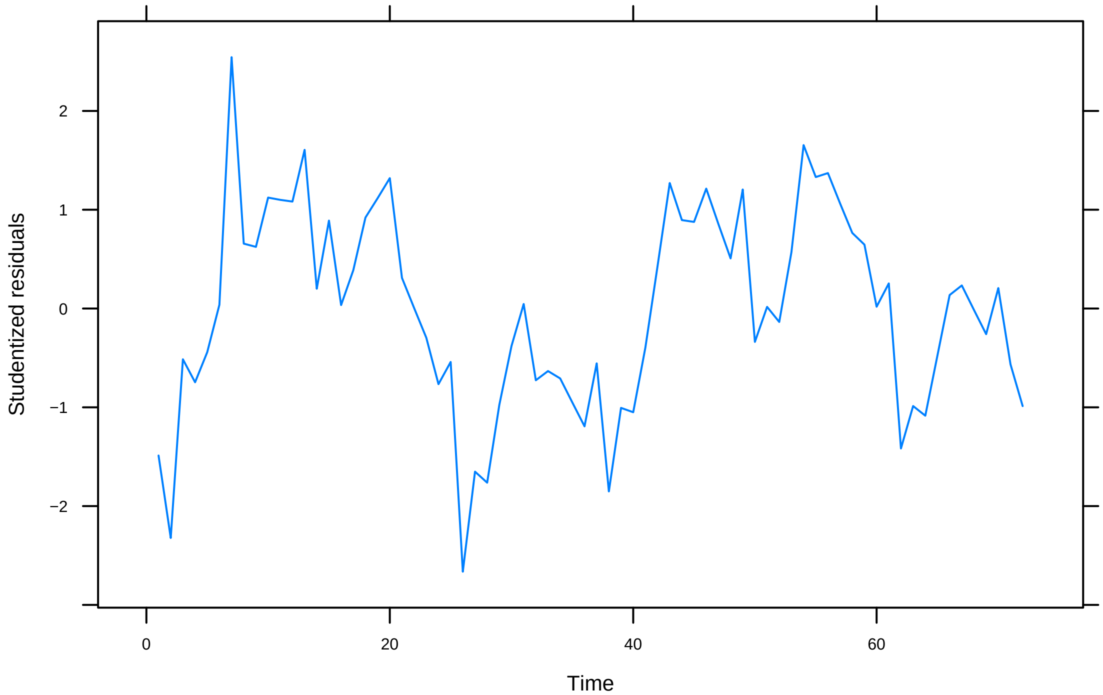


Figure 3.5: (#fig:wages_quad_resid)Residual plot for our quadratic model.

3.6 Beer sales

a

```
data(beersales)
xyplot(beersales)
```

Clear seasonal trends. There is an initial positive trend from 1975 to around 1981 that then levels out.

b

```
months <- c("J", "F", "M", "A", "M", "J", "J", "A", "S", "O", "N", "D")

xyplot(beersales,
       panel = function(x, y, ...) {
         panel.xyplot(x, y, ...)
         panel.text(x, y, labels = months)
       })
```

It is now evident that the peaks are in the warm months and the slump in the winter and fall months. December is a particular low point, while May, June, and July seem to be the high points.

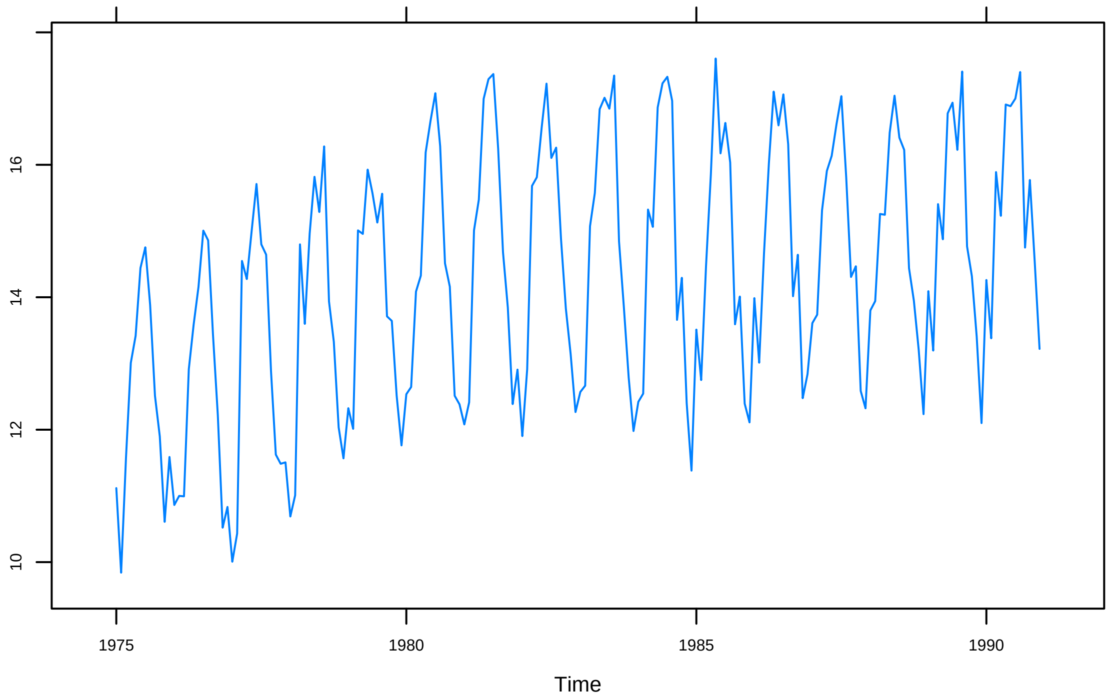


Figure 3.6: Monthly U.S. beer sales.

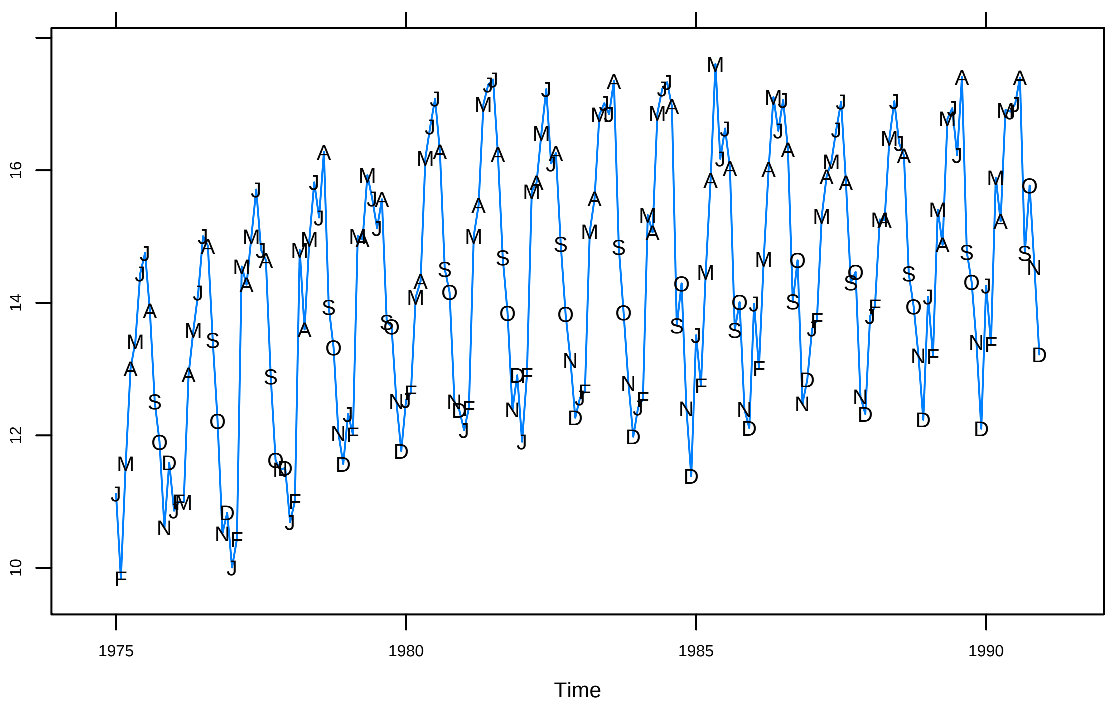


Figure 3.7: Monthly U.S. beer sales annotated with the months' initials.

c

```r
beer_fit1 <- lm(beersales ~ season(beersales))
pander(summary(beer_fit1))
```

	Estimate	Std. Error	t value	Pr(>|t|)
season(beersales)February	-0.1426	0.3732	-0.382	0.7029
season(beersales)March	2.082	0.3732	5.579	8.771e-08
season(beersales)April	2.398	0.3732	6.424	1.151e-09
season(beersales)May	3.599	0.3732	9.643	5.322e-18
season(beersales)June	3.85	0.3732	10.31	6.813e-20
season(beersales)July	3.769	0.3732	10.1	2.812e-19
season(beersales)August	3.609	0.3732	9.669	4.494e-18
season(beersales)September	1.573	0.3732	4.214	3.964e-05
season(beersales)October	1.254	0.3732	3.361	0.0009484
season(beersales)November	-0.04797	0.3732	-0.1285	0.8979
season(beersales)December	-0.4231	0.3732	-1.134	0.2585
(Intercept)	12.49	0.2639	47.31	1.786e-103

Table 3.2: Fitting linear model: beersales ~ season(beersales)

Observations	Residual Std. Error	R^2	Adjusted R^2
192	1.056	0.7103	0.6926

All comparisons are made against january. The model helpfully explains approximately 0.71 of the variance and is statistically significant. Most of the factors are significant (mostly the winter months as expected).

d

```r
xyplot(rstudent(beer_fit1) ~ time(beersales), type = "l",
       xlab = "Time", ylab = "Studentized residuals",
       panel = function(x, y, ...) {
         panel.xyplot(x, y, ...)
         panel.xyplot(x, y, pch = as.vector(season(beersales)), col = 1)
       })
```

Looking at the residuals in 3.8 We don't have a good fit to our data; in particular, wee're not capturing the long-term trend.

e

```r
beer_fit2 <- lm(beersales ~ season(beersales) + time(beersales) +
                  I(time(beersales) ^ 2))
pander(summary(beer_fit2))
```

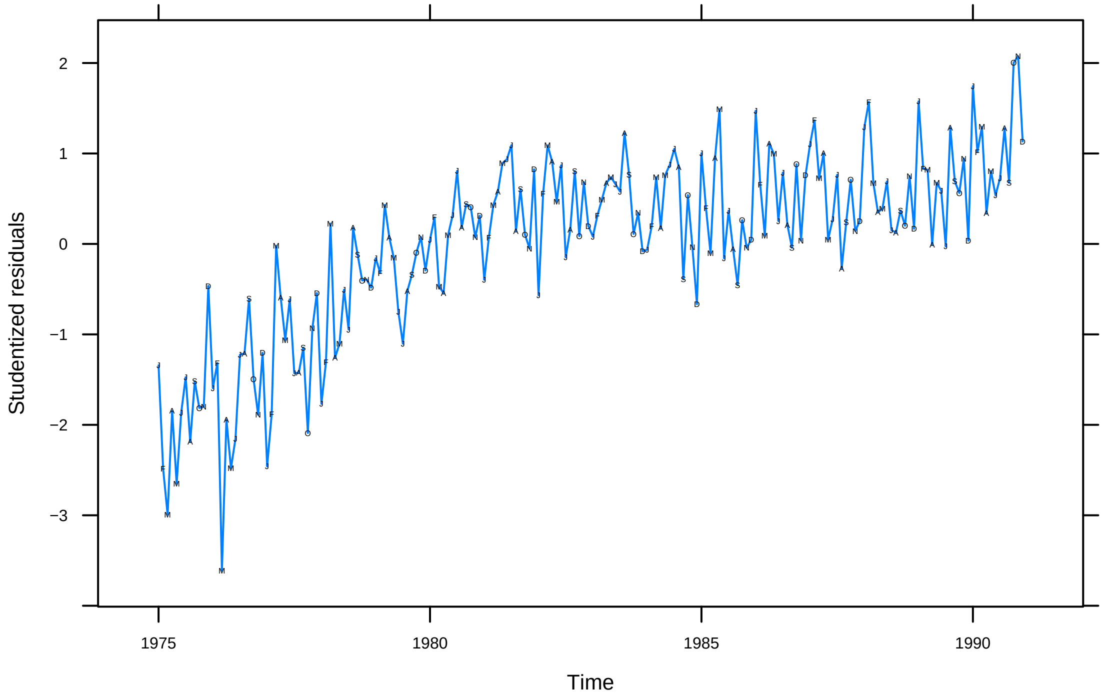


Figure 3.8: Beer sales residual plot.

	Estimate	Std. Error	t value	Pr(>|t|)
season(beersales)February	-0.1579	0.209	-0.7554	0.451
season(beersales)March	2.052	0.209	9.818	1.864e-18
season(beersales)April	2.353	0.209	11.26	1.533e-22
season(beersales)May	3.539	0.209	16.93	6.063e-39
season(beersales)June	3.776	0.209	18.06	4.117e-42
season(beersales)July	3.681	0.209	17.61	7.706e-41
season(beersales)August	3.507	0.2091	16.78	1.698e-38
season(beersales)September	1.458	0.2091	6.972	5.89e-11
season(beersales)October	1.126	0.2091	5.385	2.268e-07
season(beersales)November	-0.1894	0.2091	-0.9059	0.3662
season(beersales)December	-0.5773	0.2092	-2.76	0.00638
time(beersales)	71.96	8.867	8.115	7.703e-14
I(time(beersales)^2)	-0.0181	0.002236	-8.096	8.633e-14
(Intercept)	-71498	8791	-8.133	6.932e-14

Table 3.4: Fitting linear model: beersales ~ season(beersales) + time(beersales) + I(time(beersales)^2)

Observations	Residual Std. Error	R^2	Adjusted R^2
192	0.5911	0.9102	0.9036

This model fits the data better, explaining roughly 0.91 of the variance.

f

```
xyplot(rstudent(beer_fit2) ~ time(beersales), type = "l",
       xlab = "Time", yla = "Studentized residuals",
       panel = function(x, y, ...) {
         panel.xyplot(x, y, ...)
         panel.xyplot(x, y, pch = as.vector(season(beersales)), col = 1)
```

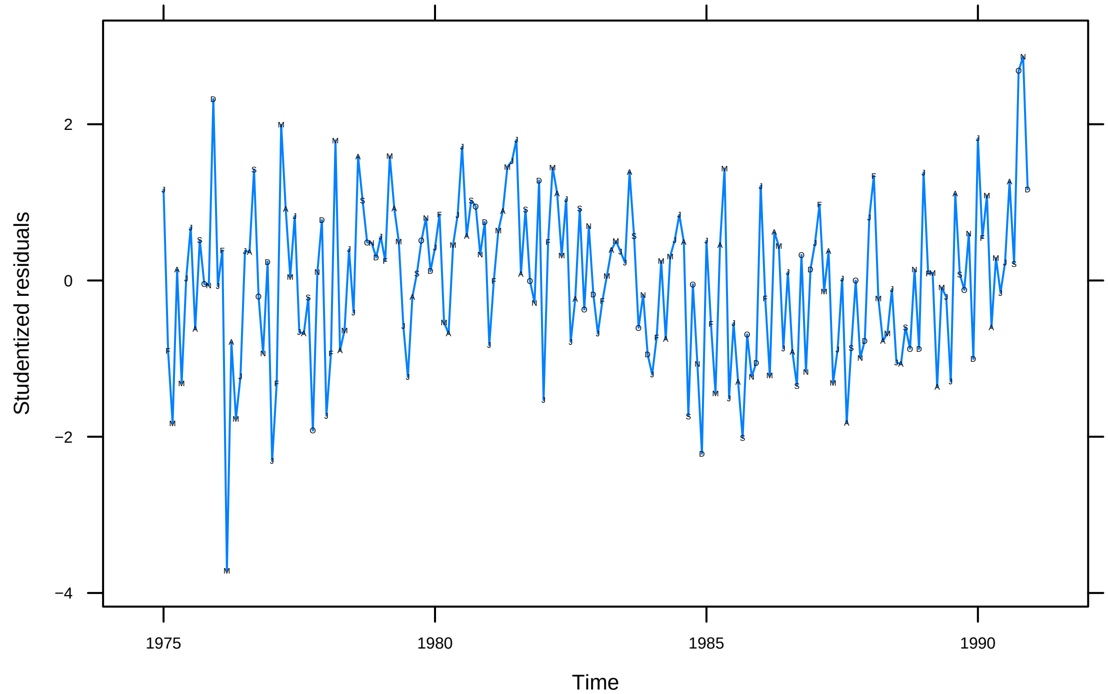


Figure 3.9: Beer sales residual plot from the quadratic fit.

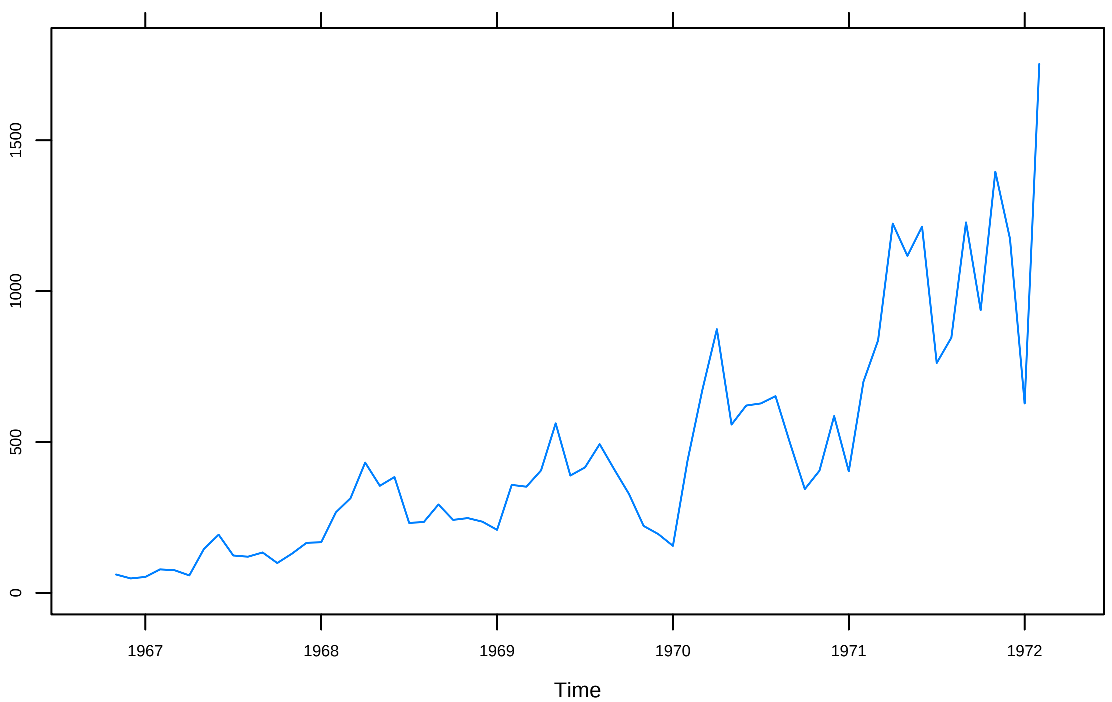


Figure 3.10: Monthly unit sales of recreational vehicles from Winnebago.

b

```
winn_fit1 <- lm(winnebago ~ time(winnebago))
summary(winn_fit1) %>%
  pander()
```

	Estimate	Std. Error	t value	Pr(>|t|)
time(winnebago)	200.7	17.03	11.79	1.777e-17
(Intercept)	-394886	33540	-11.77	1.87e-17

Table 3.6: Fitting linear model: winnebago ~ time(winnebago)

Observations	Residual Std. Error	R^2	Adjusted R^2
64	209.7	0.6915	0.6865

The model is significant and explains 0.69 of the variance.

```
xyplot(rstudent(winn_fit1) ~ time(winnebago), type = "l",
       xlab = "Time", ylab = "Studentized residuals")
```

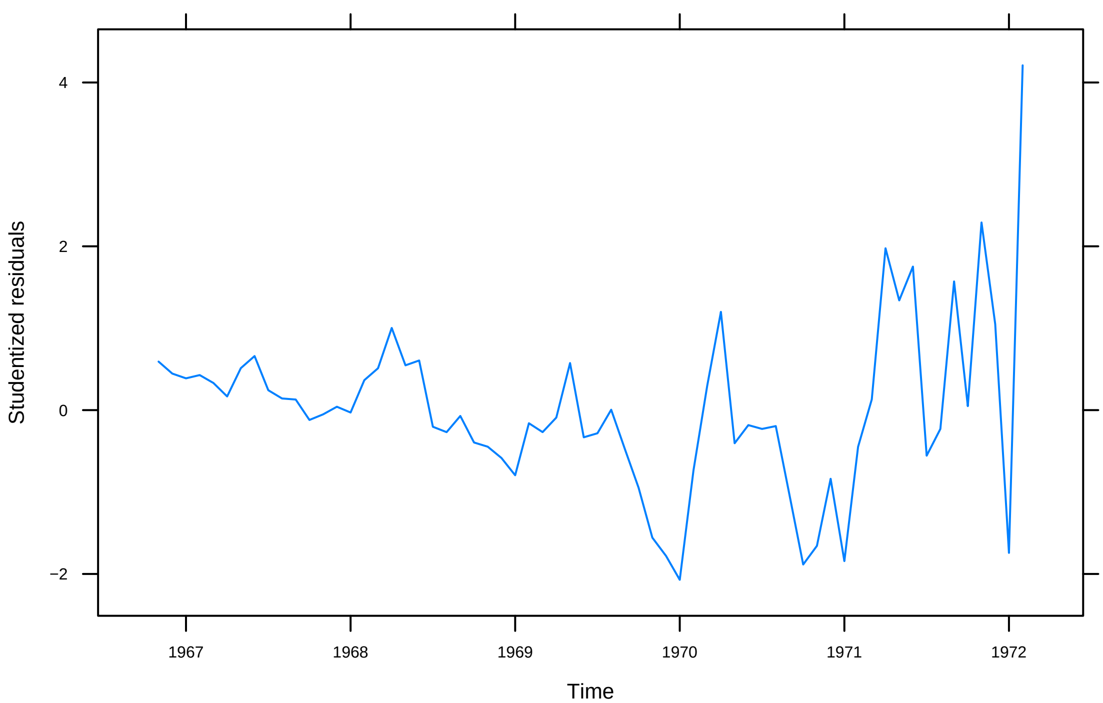


Figure 3.11: Residuals for the linear fit for the winnebago data.

The fit is poor (Figure **??**. It is not random and it is clear that we're making worse predictions for later yers.

c

To produce a better fit, we transform the outcome with the natural logarithm.

```r
winn_fit_log <- lm(log(winnebago) ~ time(winnebago))
pander(summary(winn_fit_log))
```

	Estimate	Std. Error	t value	Pr(>|t|)
time(winnebago)	0.5031	0.03199	15.73	2.575e-23
(Intercept)	-984.9	62.99	-15.64	3.45e-23

Table 3.8: Fitting linear model: log(winnebago) ~ time(winnebago)

Observations	Residual Std. Error	R^2	Adjusted R^2
64	0.3939	0.7996	0.7964

The model is better, explaining almost 0.8 of the variance.

d

```r
xyplot(rstudent(winn_fit_log) ~ time(winnebago), type = "l",
       xlab = "Time", ylab = "Studentized residuals",
       panel = function(x, y, ...) {
         panel.xyplot(x, y, ...)
         panel.xyplot(x, y, pch = as.vector(season(winnebago)), col = 1)
       })
```

This looks more like random noise (Figure **??**. Values still cling together somewhat but it is certainly better than the linear model. We're still systematically overpredictinig the values for some months, however.

e

```r
winn_fit_seasonal <- lm(log(winnebago) ~ season(winnebago) + time(winnebago))
pander(summary(winn_fit_seasonal))
```

	Estimate	Std. Error	t value	Pr(>|t|)
season(winnebago)February	0.6244	0.1818	3.434	0.001188
season(winnebago)March	0.6822	0.1909	3.574	0.0007793
season(winnebago)April	0.8096	0.1908	4.243	9.301e-05
season(winnebago)May	0.8695	0.1907	4.559	3.246e-05
season(winnebago)June	0.8631	0.1907	4.526	3.627e-05
season(winnebago)July	0.5539	0.1907	2.905	0.00542
season(winnebago)August	0.5699	0.1907	2.988	0.004305
season(winnebago)September	0.5757	0.1907	3.018	0.00396

	Estimate	Std. Error	t value	Pr(>|t|)
season(winnebago)October	0.2635	0.1908	1.381	0.1733
season(winnebago)November	0.2868	0.1819	1.577	0.1209
season(winnebago)December	0.248	0.1818	1.364	0.1785
time(winnebago)	0.5091	0.02571	19.8	1.351e-25
(Intercept)	-997.3	50.64	-19.69	1.718e-25

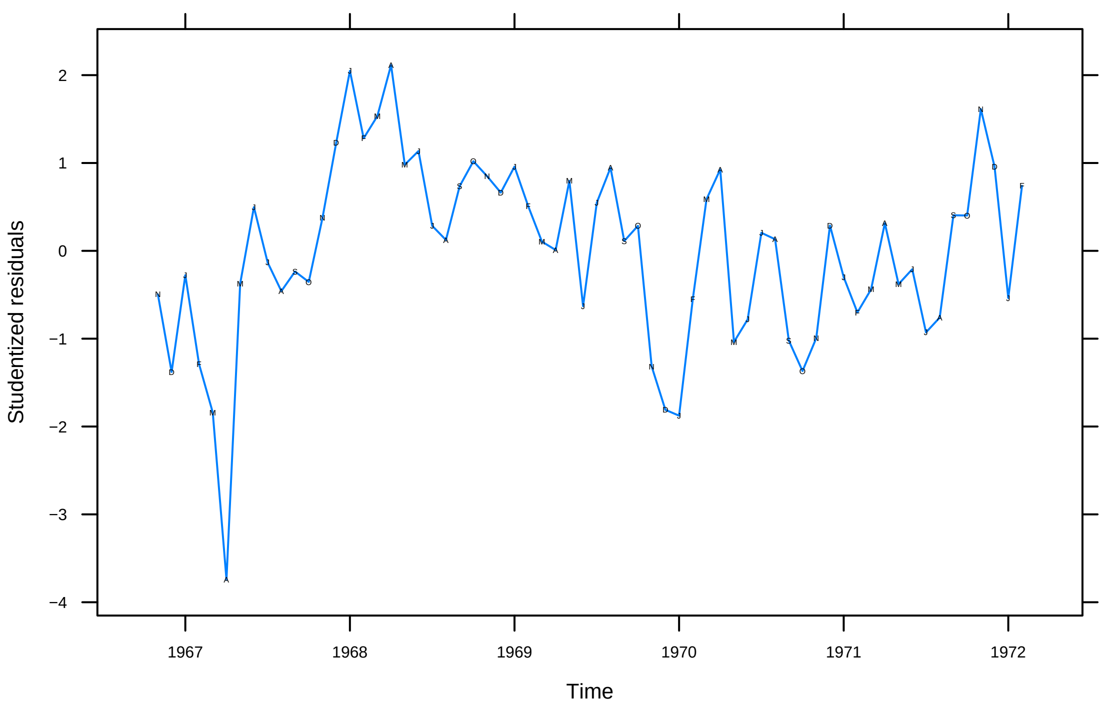
Table 3.10: Fitting linear model: log(winnebago) ~ season(winnebago) + time(winnebago)

Observations	Residual Std. Error	R^2	Adjusted R^2
64	0.3149	0.8946	0.8699

The fit is improved further. We have a R^2 of 0.89 and significance for most of our seasonal means as well as the time trend.

f

```r
xyplot(rstudent(winn_fit_seasonal) ~ time(winnebago), type = "l",
       xlab = "Time", ylab = "Studentized residuals",
       panel = function(x, y, ...) {
         panel.xyplot(x, y, ...)
         panel.xyplot(x, y, col = 1, pch = as.vector(season(winnebago)))
       })
```



This is acceptable even if our residuals are quite large for some of the values, notably at the start of the series.

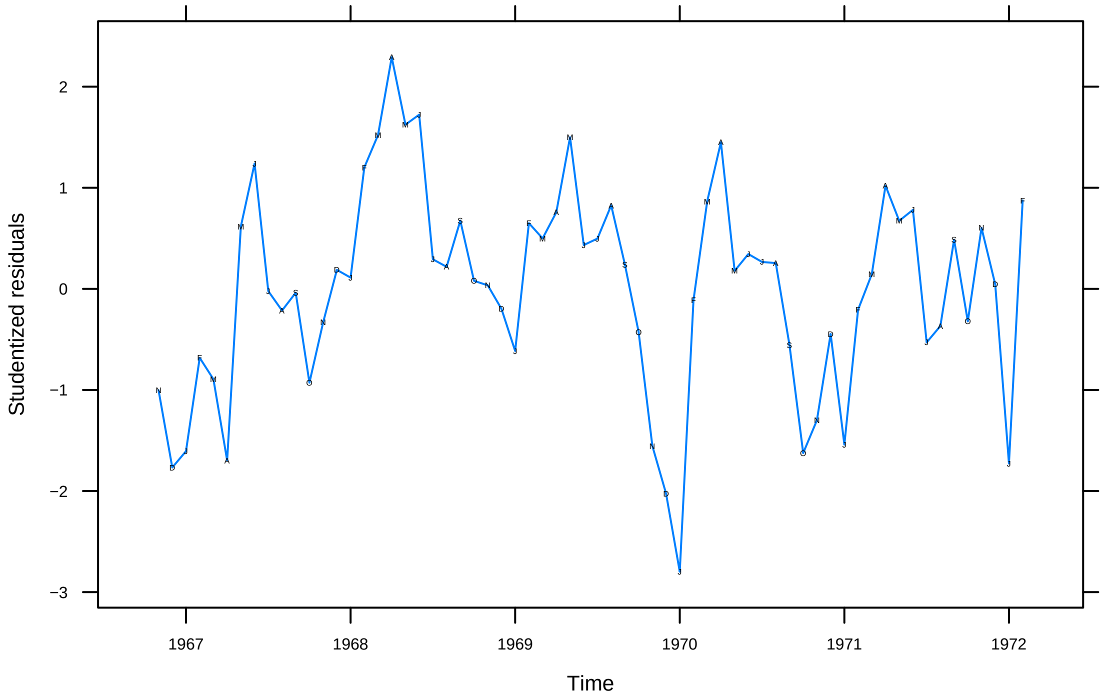


Figure 3.12: Residual plot after natural log transformation.

3.8 Retail

a

```
data(retail)
xyplot(retail, panel = function(x, y, ...) {
  panel.xyplot(x, y, ...)
  panel.xyplot(x, y, pch = as.vector(season(retail)), col = 1)
})
```

Plotting the retail sales trend there seems to be a long-term linear trend as well as heavy seasonality in tht December – and to slighter extent also November and October – exhibit regular surges in retail sales.

b

```
retail_lm <- lm(retail ~ season(retail) + time(retail))
pander(summary(retail_lm))
```

	Estimate	Std. Error	t value	Pr(>|t|)
season(retail)February	-3.015	1.29	-2.337	0.02024
season(retail)March	0.07469	1.29	0.05791	0.9539
season(retail)April	3.447	1.305	2.641	0.008801
season(retail)May	3.108	1.305	2.381	0.01803
season(retail)June	3.074	1.305	2.355	0.01932

	Estimate	Std. Error	t value	Pr(>|t|)
season(retail)July	6.053	1.305	4.638	5.757e-06
season(retail)August	3.138	1.305	2.404	0.01695
season(retail)September	3.428	1.305	2.626	0.009187
season(retail)October	8.555	1.305	6.555	3.336e-10
season(retail)November	20.82	1.305	15.95	1.274e-39
season(retail)December	52.54	1.305	40.25	3.169e-109
time(retail)	3.67	0.04369	84	5.206e-181
(Intercept)	-7249	87.24	-83.1	6.41e-180

Table 3.12: Fitting linear model: retail ~ season(retail) + time(retail)

Observations	Residual Std. Error	R^2	Adjusted R^2
255	4.278	0.9767	0.9755

This *seems* like an effective model, explaining 0.98 of the variance in retail sales.

c

```
xyplot(rstudent(retail_lm) ~ time(retail), type = "l",
       xlab = "Time", ylab = "Studentized residuals",
       panel = function(x, y, ...) {
         panel.xyplot(x, y, ...)
         panel.xyplot(x, y, pch = as.vector(season(retail)), col = 1)
       })
```

The residual plot (Figure 3.14) tells a different story: we're underpredicting values for early period and overpredicting values for the later years – however, this should be an easy fix.

3.9 Prescriptions

a

```
data(prescrip)
xyplot(prescrip, ylab = "Prescription costs",
       panel = function(x, y, ...) {
         panel.xyplot(x, y, ...)
         panel.xyplot(x, y, pch = as.vector(season(prescrip)), col = 1)
       })
```

Figure 3.15 shows a clear, smooth, and cyclical seasonal trend. Values are genereally higher for the summer months and there seems to be an exponential increase long-term.

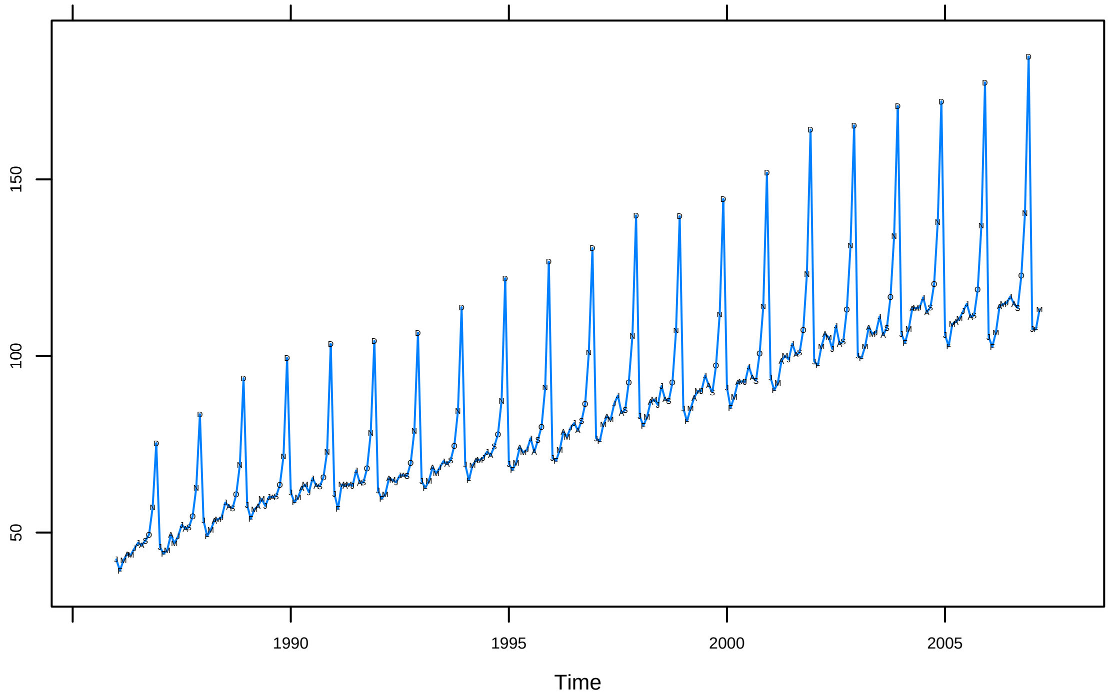


Figure 3.13: Total retail sales in the U.K. in billions pounds.

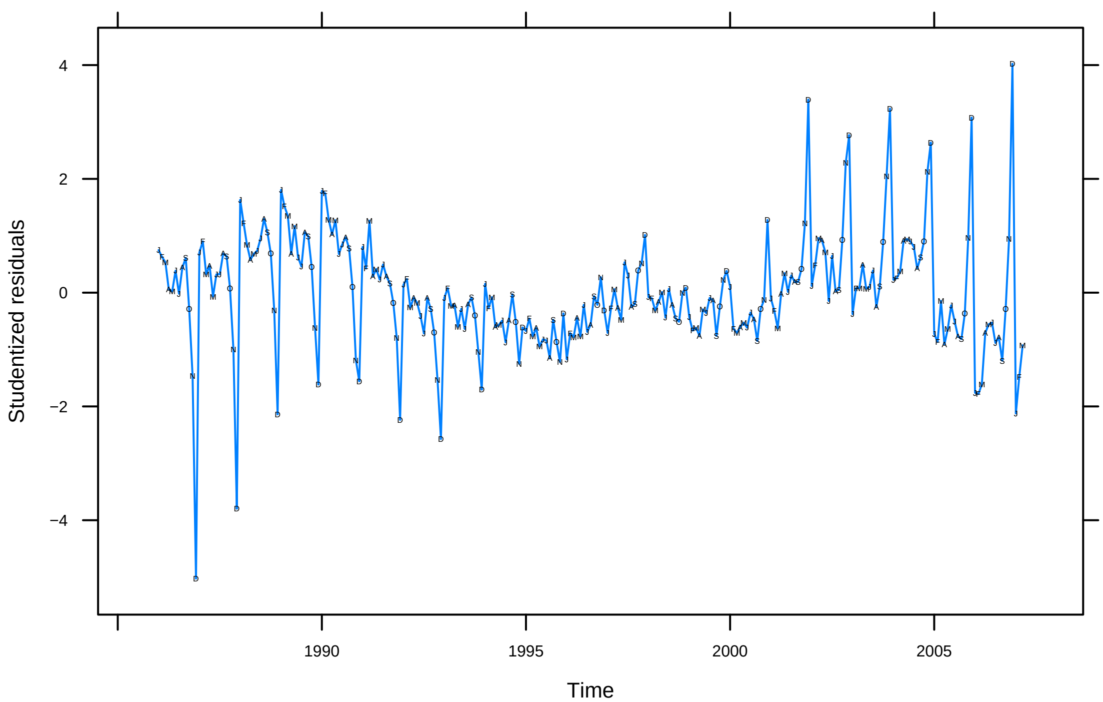


Figure 3.14: Studentized residuals for our seasonality + linear model of retail sales.

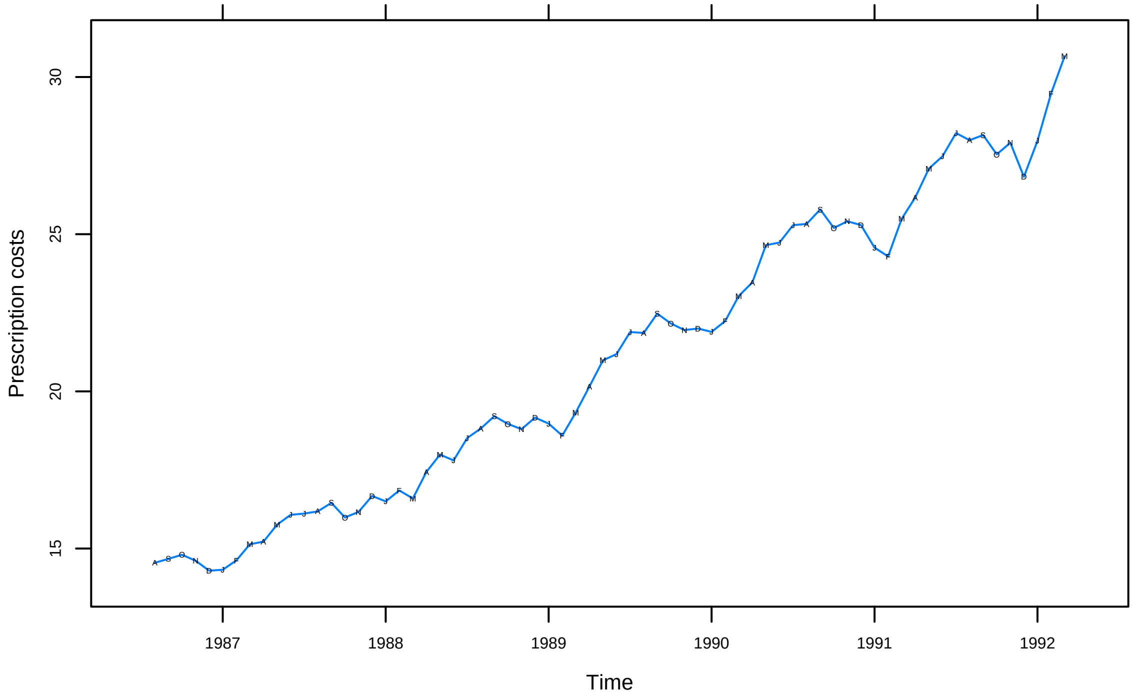


Figure 3.15: Monthly U.S. prescription costs.

b

```
pchange <- diff(prescrip) / prescrip
xyplot(pchange ~ time(prescrip), type = "l",
       panel = function(x, y, ...) {
         panel.xyplot(x, y, ...)
         panel.xyplot(x, y, pch = as.vector(season(pchange)), col = 1)
       })
```

The monthly percentage difference series looks rather stationary.

c

```
pres_cos <- lm(pchange ~ harmonic(pchange))
pander(summary(pres_cos))
```

	Estimate	Std. Error	t value	Pr(>|t|)
harmonic(pchange)cos(2*pit*)	-0.006605	0.003237	-2.041	0.04542
harmonic(pchange)sin(2*pit*)	0.01612	0.003208	5.026	4.291e-06
(Intercept)	0.01159	0.002282	5.08	3.508e-06

Table 3.14: Fitting linear model: pchange ~ harmonic(pchange)

Observations	Residual Std. Error	R^2	Adjusted R^2
67	0.01862	0.3126	0.2912

We explain 0.31 of the variance. The model is significant though.

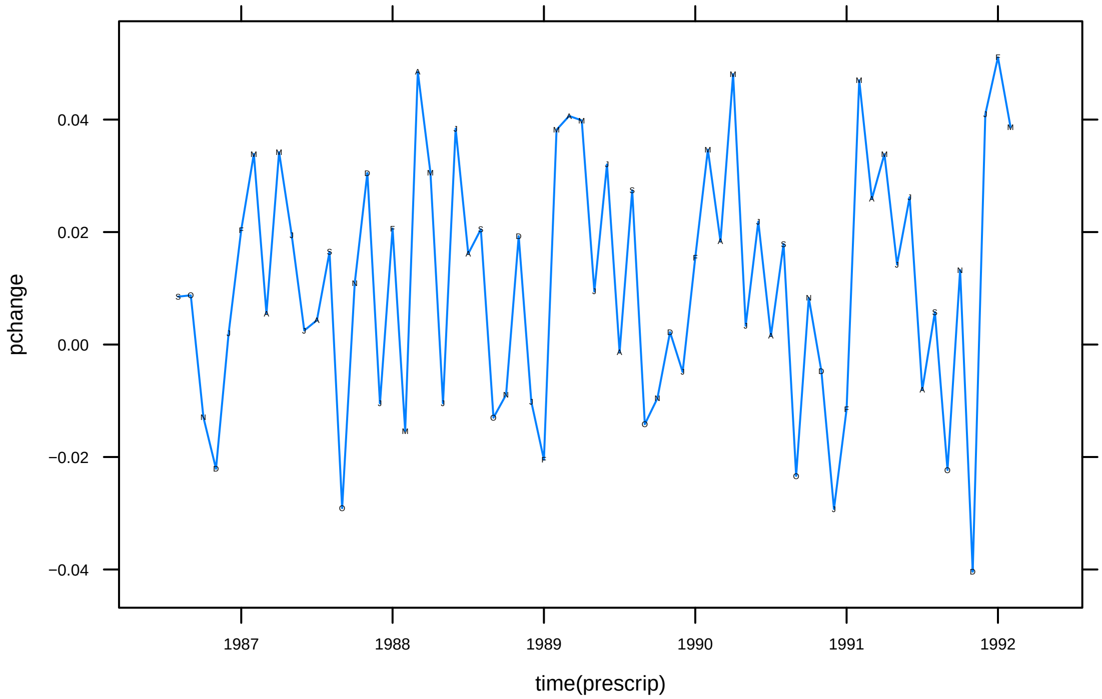


Figure 3.16: Percentage changes from month-to-month in prescription costs.

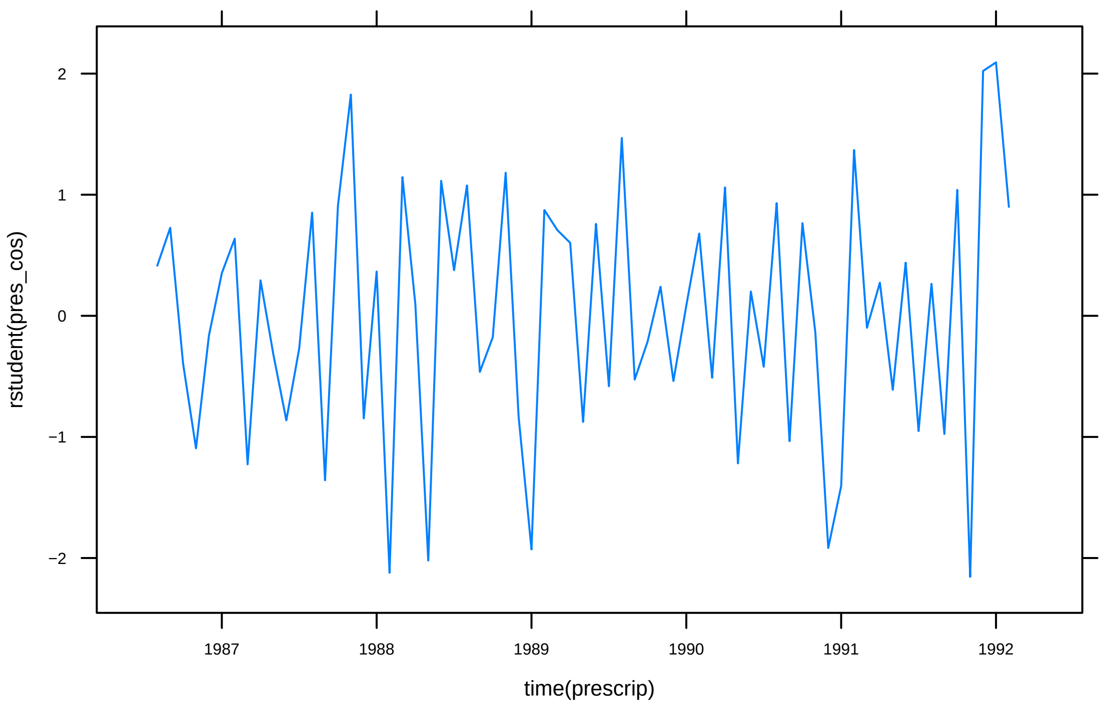


Figure 3.17: Residuals for our cosine model.