Solutions to Time Series Analysis: with Applications in R

Johan Larsson

2017-04-04

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. Unlike the book, the solutions here use lattice graphics when possible instead of base graphics.

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.

```
library(TSA)
library(latticeExtra)

data(larain, package = "TSA")

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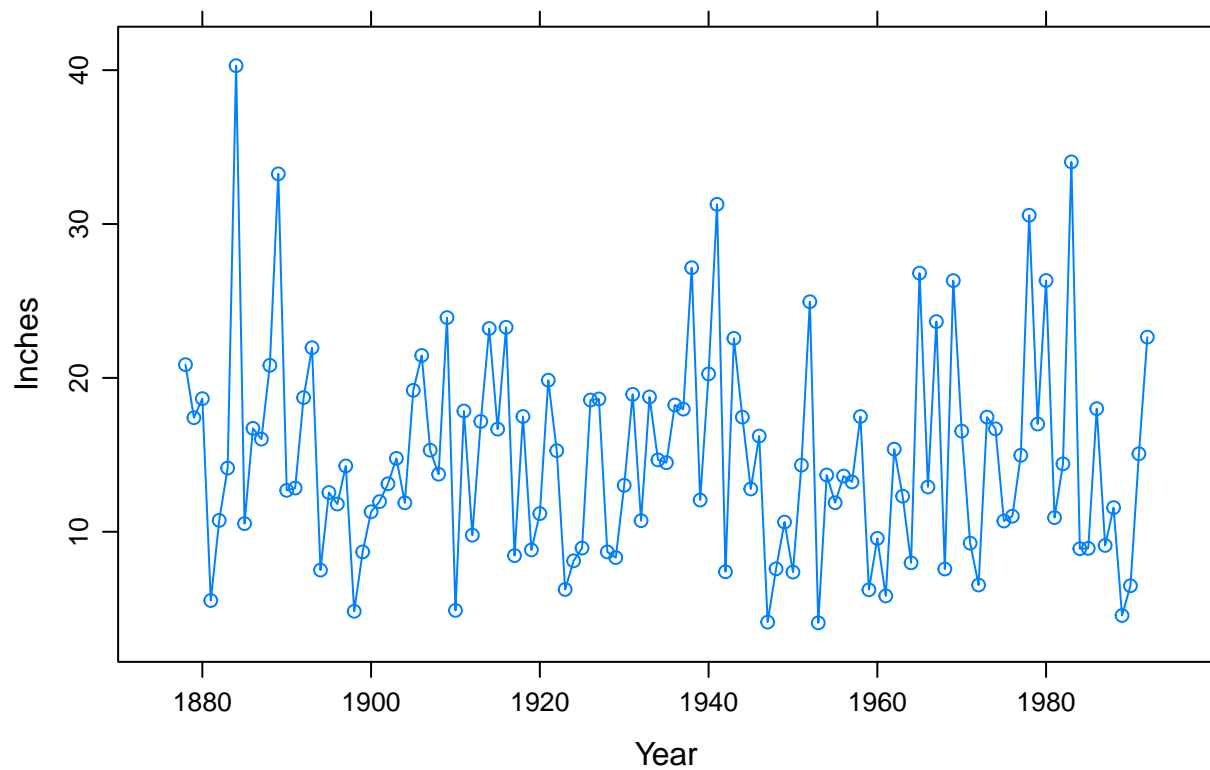
1.2   Colors

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")
```
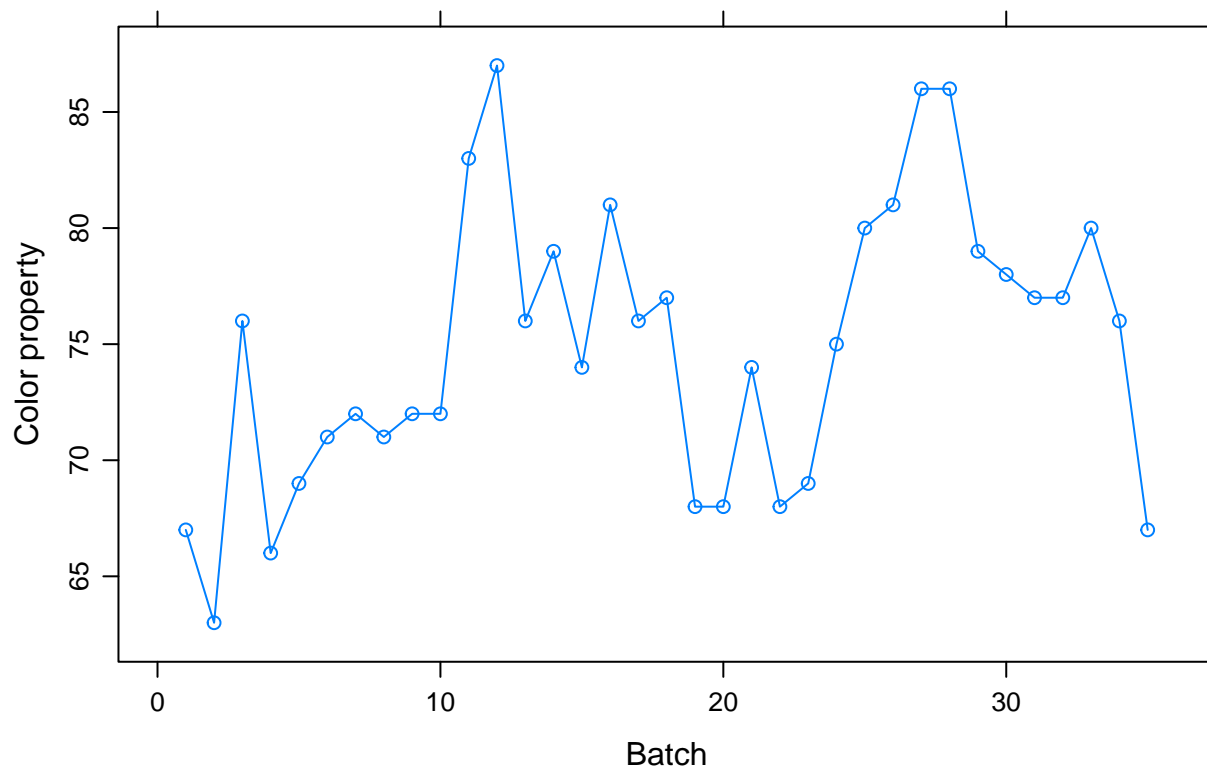

1.3   Random, normal time series

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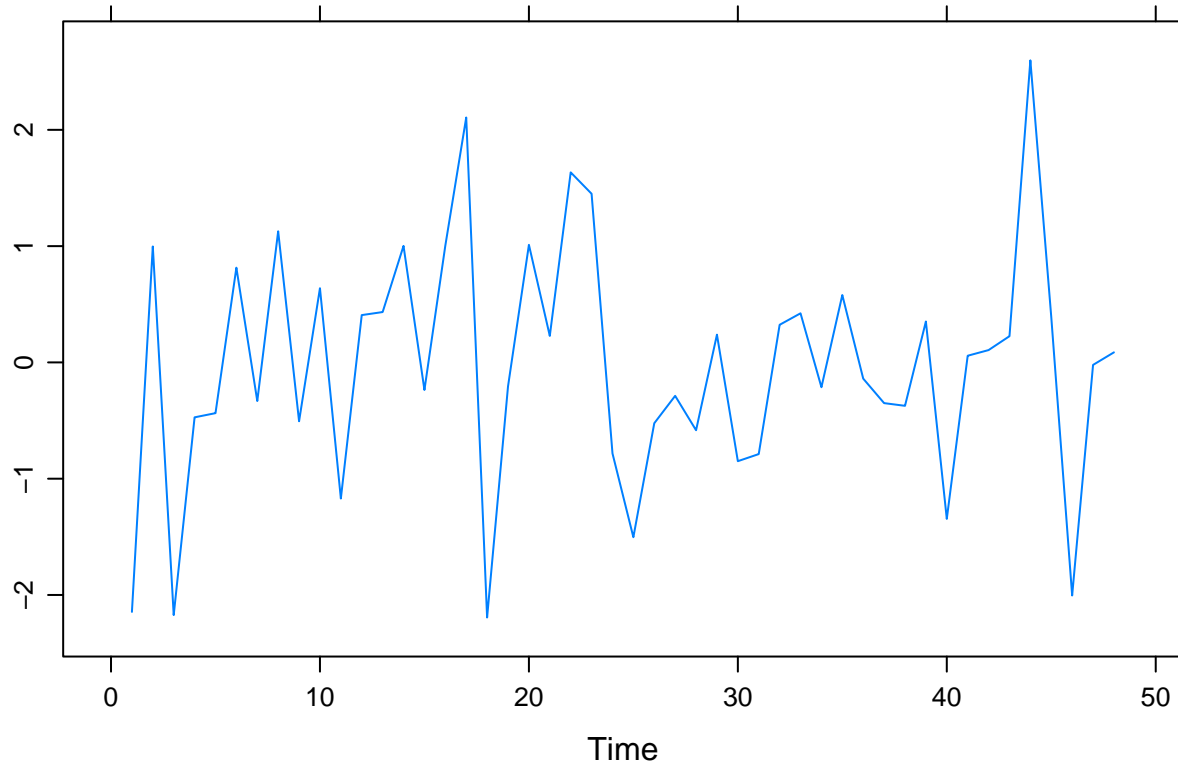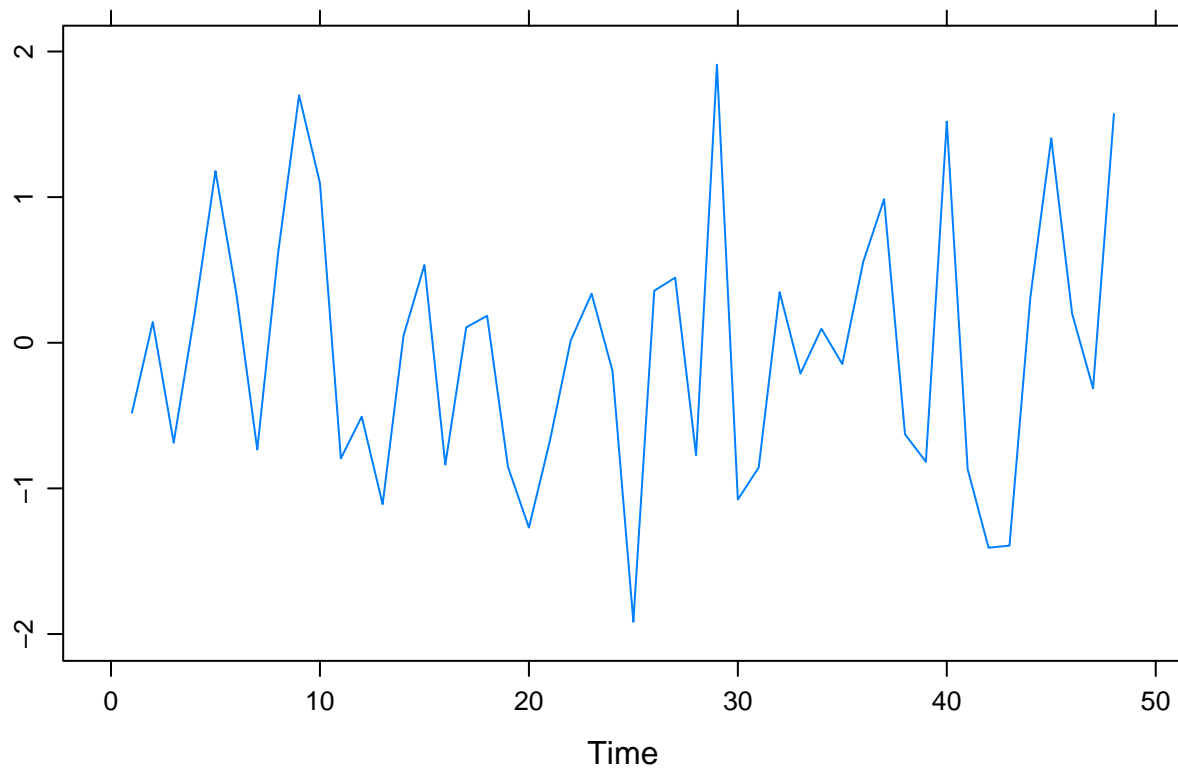
As far as we can tell there is no discernable pattern here.

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)))
```
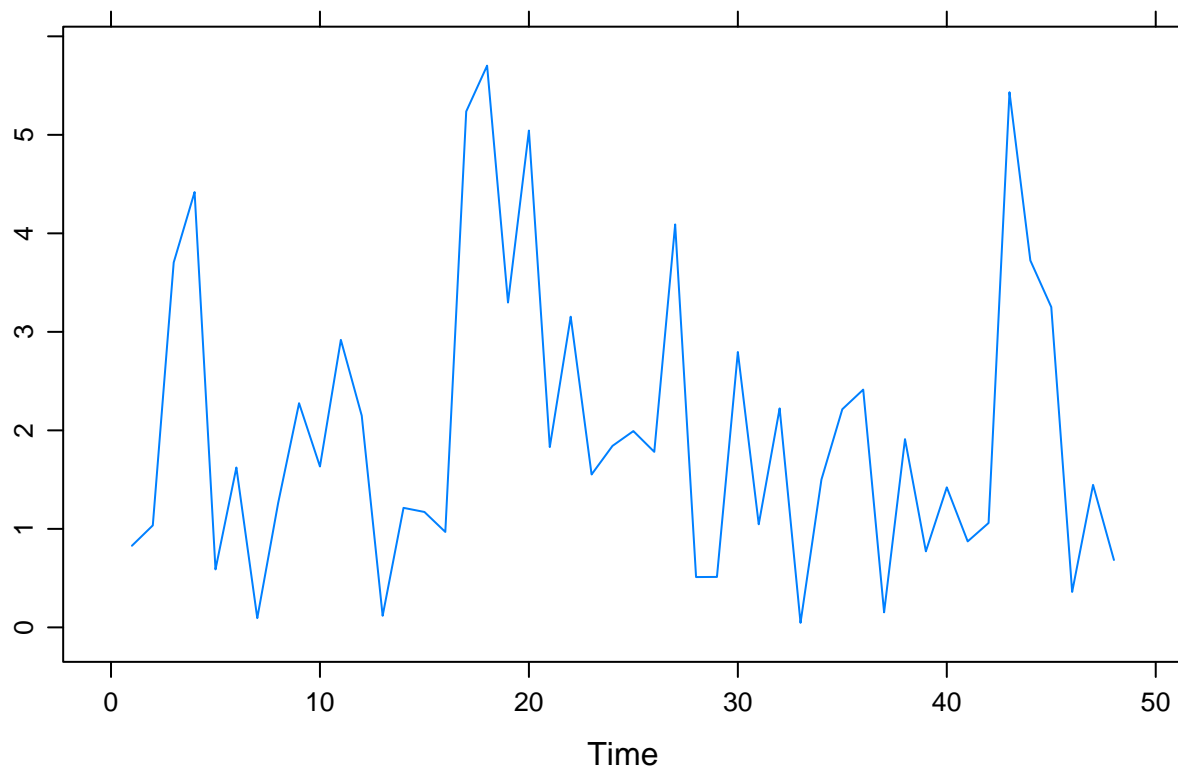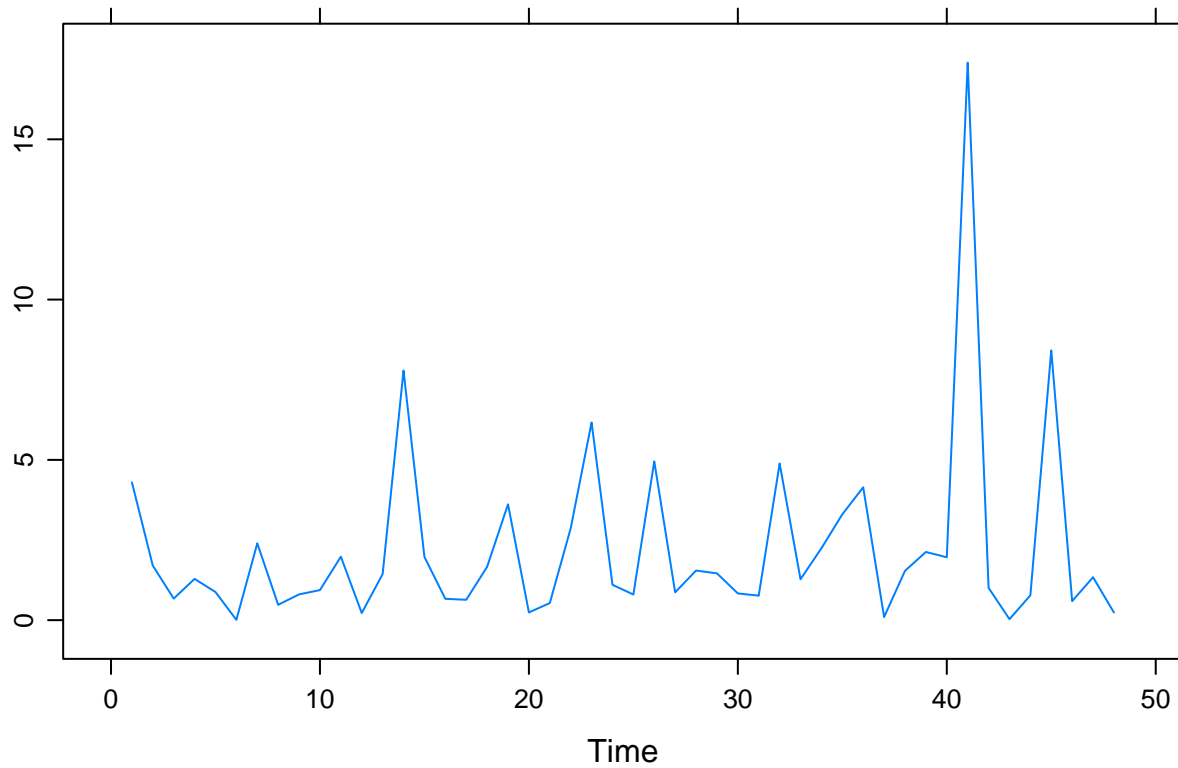
The process appears random, though non-normal.

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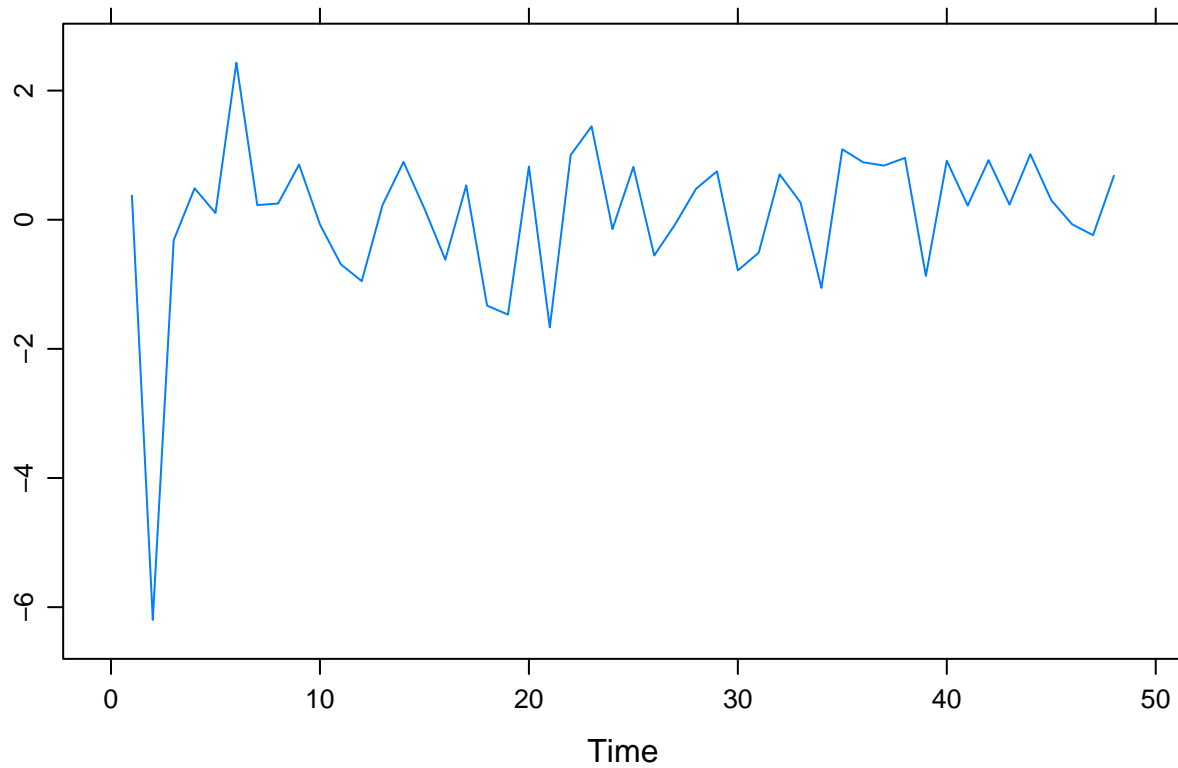
```
xyplot(as.ts(rt(48, 5)))
```
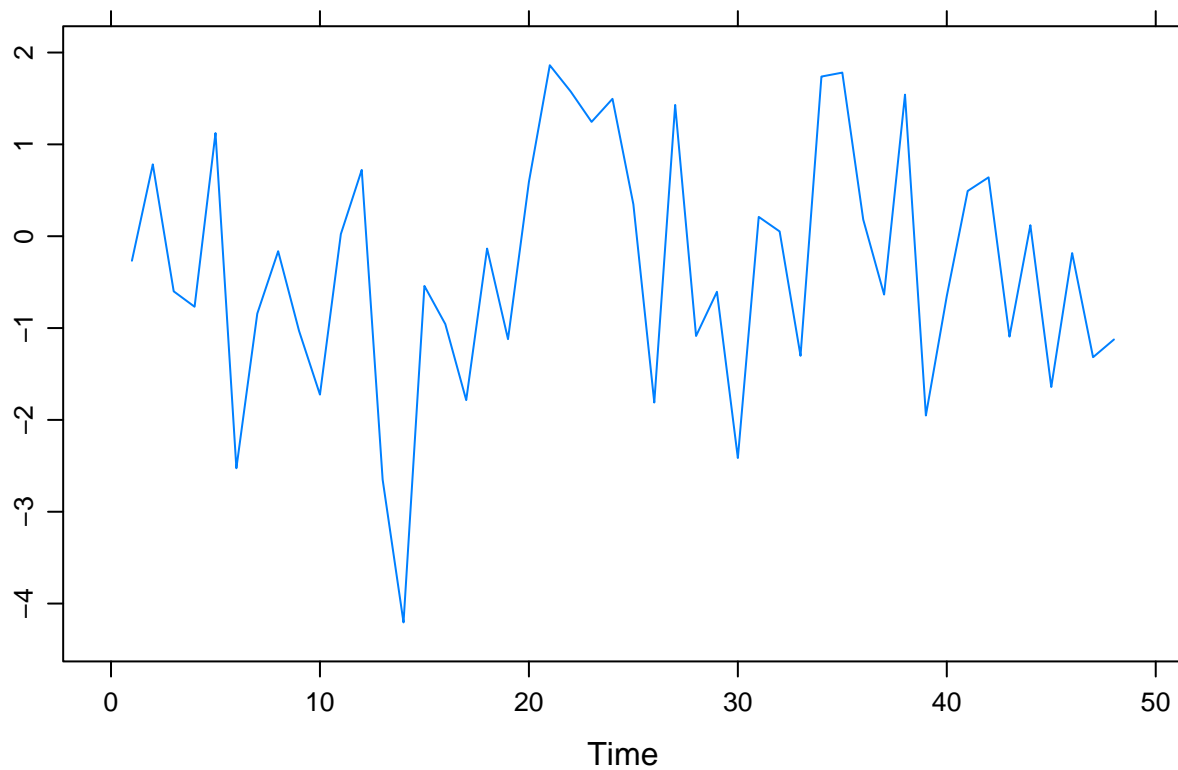

It looks random but not normal, though it should be approximately so, considering the distribution that we have sampled from.

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")
```
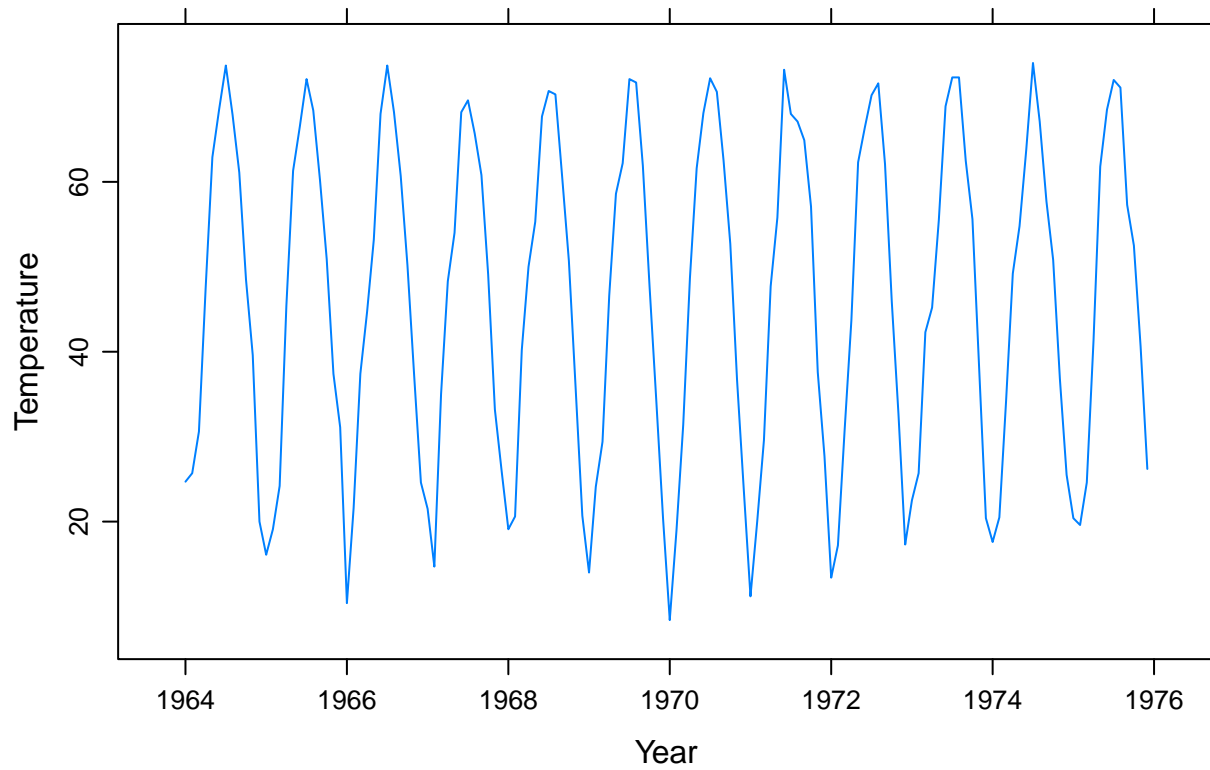
Chapter 2

Fundamental concepts

2.1 Basic properties of expected value and covariance

a

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

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

$$\text{Corr}[X+Y, X-Y] = \text{Corr}[X,X] + \text{Corr}[X,-Y] + \text{Corr}[Y,X] + \text{Corr}[Y,-Y] \tag{2.6}$$
$$= \text{Corr}[Y,X] + \text{Corr}[Y,-Y] \tag{2.7}$$
$$= 1 - 0.25 + 0.25 - 1 \tag{2.8}$$
$$= 0 \tag{2.9}$$
$$\tag{2.10}$$

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

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

which satisfies our requirement for strict stationarity.

b

The autocovariance is given by

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

c

```r
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

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

For $k = 1$ we have

$$
\begin{aligned}
C[e_t + \theta e_{t-1}, e_{t-1} + \theta e_{t-2}] &= \\
C[e_t, e_{t-1}] + C[e_t, \theta e_{t-2}] + C[\theta e_{t-1}, e_{t-1}] + C[\theta e_{t-1}, \theta e_{t-2}] &= \\
0 + 0 + \theta V[e_{t-1}] + 0 = \theta \sigma_e^2,&
\end{aligned}
$$
$$
\mathrm{Corr}[Y_t, Y_{t-k}] = \frac{\theta \sigma_e^2}{\sqrt{(\sigma_e^2(1 + \theta^2))^2}} = \frac{\theta}{1 + \theta^2}
$$
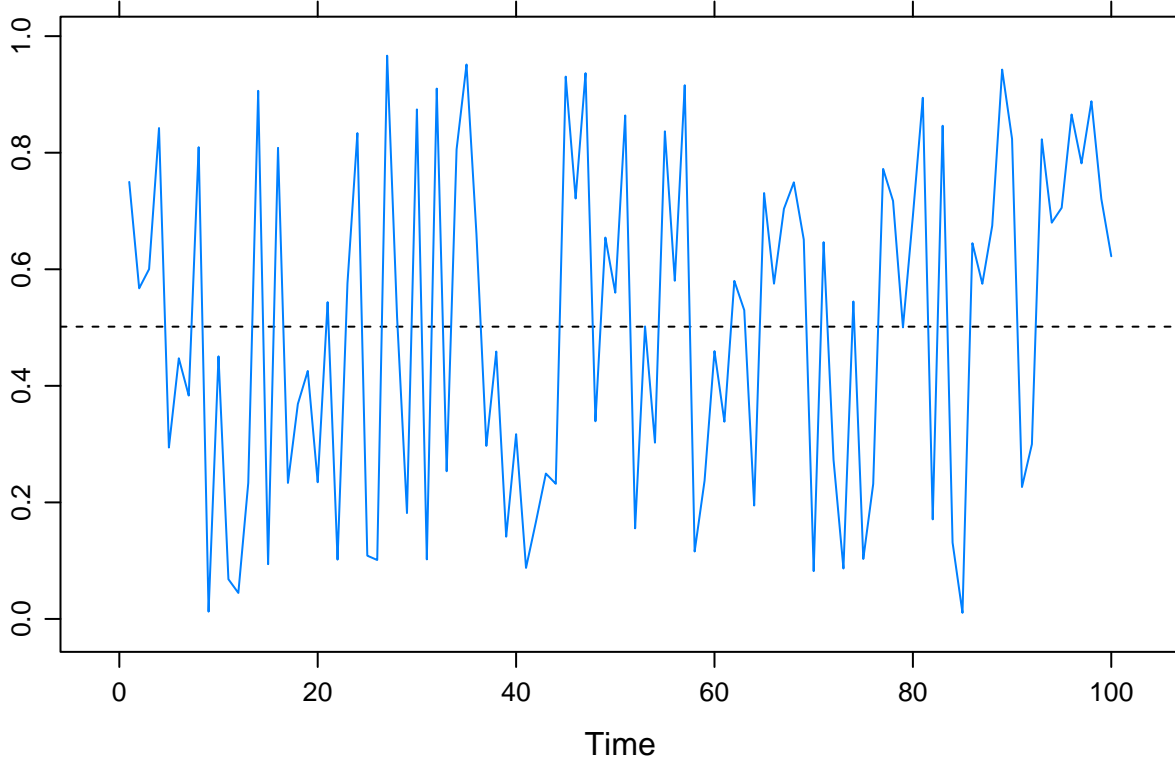
Figure 2.1: A white noise time series: no drift, independence between observations.

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

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

and thus the expected value is constant. Moreover,

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

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.

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

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

is free of t and, furthermore, we have

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

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

b

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

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

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

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

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

2.12 Periods

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

Then, as required, we have

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

2.13 Drift, part 2

a

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

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

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

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

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

and

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

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

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

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

b

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

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

c

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

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

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

2.17 Variance of sample mean

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

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

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

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

$$E[s^2] = E\left[\frac{n}{n-1}\sum_{t=1}^{n}(Y_t - \bar{Y})^2\right] = \frac{n}{n-1}E\left[\sum_{t=1}^{n}\left((Y_t - \mu)^2 + n(\bar{Y} - \mu)^2\right)\right] =$$

$$\frac{n}{n-1}\sum_{t=1}^{n}\left(E[(Y_t - \mu)^2] + nE[(\bar{Y} - \mu)^2]\right) = \frac{1}{n-1}\left(n\mathrm{Var}[Y_t] - n\mathrm{Var}[\bar{Y}]\right) =$$

$$\frac{n}{n-1}\gamma_0 - \frac{n}{n-1}\mathrm{Var}[\bar{Y}] = \frac{1}{n-1}\left(n\gamma_0 - n\left(\frac{\gamma_0}{n} + \frac{2}{n}\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\gamma_k\right)\right) =$$

$$\frac{1}{n-1}\left(n\gamma_0 - \gamma_0 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\gamma_k\right) = \frac{1}{n-1}\left(\gamma_0(n-1) + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\gamma_k\right) =$$

$$\gamma_0 + \frac{2}{n-1}\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\gamma_k \quad \square$$

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

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

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

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

c

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

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

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

c

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

d

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

2.22   Asymptotic stationarity

a

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

b

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

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

c

$$\text{Cov}[Y_t, Y_{t-1}] = \text{Cov}[cY_{t-1} + e_t, Y_{t-1}] = c\text{Cov}[Y_{t-1}, Y_{t-1}] = c\text{Var}[Y_{t-1}] \quad \text{giving}$$

$$\text{Corr}[Y_t, Y_{t-1}] = \frac{c\text{Var}[Y_{t-1}]}{\sqrt{\text{Var}[Y_t]\text{Var}[Y_{t-1}]}} = c\sqrt{\frac{\text{Var}[Y_{t-1}]}{\text{Var}[Y_t]}} \quad \square$$

And, in the general case,

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

giving

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

d

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

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

$$Y_t = c(cY_{t-2} + e_{t-1}) + e_t = \cdots = e_t + ce_{t-1} + c^2e_{t-2} + \cdots + c^{t-2}e_2 + \frac{c^{t-1}}{\sqrt{1 - c^2}}e_1$$

$$\text{Var}[Y_t] = \text{Var}[e_t + ce_{t-1} + c^2e_{t-2} + \cdots + c^{t-2}e_2 + \frac{c^{t-1}}{\sqrt{1 - c^2}}e_1] =$$

$$\text{Var}[e_t] + c^2\text{Var}[e_{t-1}] + c^4\text{Var}[e_{t-2}] + \cdots + c^{2(t-2)}\text{Var}[e_2] + \frac{c^{2(t-1)}}{1 - c^2}\text{Var}[e_1] =$$

$$\sigma_e^2(1 + c^2 + c^4 + \cdots + c^{2(t-2)} + \frac{c^{2(t-1)}}{1 - c^2}) = \sigma_e^2\left(\sum_{t=1}^{n}c^{2(t-1)} - c^{2(t-1)} + \frac{c^{2(t-1)}}{1 - c^2}\right) =$$

$$\sigma_e^2\frac{1 - c^{2t} + c^{2t-2+2}}{1 - c^2} = \sigma_e^2\frac{1}{1 - c^2} \quad \square$$

Futhermore,

$$E[Y_1] = E\left[\frac{e_1}{\sqrt{1 - c^2}}\right] = \frac{E[e_1]}{\sqrt{1 - c^2}} = 0$$

$$E[Y_2] = E[cY_1 + e_2] = cE[Y_1] = 0$$

$$\cdots$$

$$E[Y_t] = E[cY_{t-1} + e_2] = cE[Y_{t-1}] = 0,$$

which satisfies our first requirement for weak stationarity. Also,

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

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.

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

both free of t.

2.24   Measurement noise

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

2.25 Random cosine wave

$$E[Y_t] = E\left[\beta_0 + \sum_{i=1}^{k}(A_i\cos(2\pi f_i t) + B_i\sin(2\pi f_i t))\right] =$$

$$\beta_0 + \sum_{i=1}^{k}(E[A_i]\cos(2\pi f_i t) + E[B_i]\sin(2\pi f_i t) = \beta_0$$

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

$$\sum_{i=1}^{k}\text{Cov}[A_i\cos(2\pi f_i t) + A_i\sin(2\pi f_i s)] + \sum_{i=1}^{k}\text{Cov}[B_i\cos(2\pi f_j t) + B_i\sin(2\pi f_j s)] =$$

$$\sum_{i=1}^{k}\text{Var}[A_i](\cos(2\pi f_i t) + \sin(2\pi f_i s)) + \sum_{i=1}^{k}\text{Var}[B_i](\cos(2\pi f_j t) + \sin(2\pi f_j s)) =$$

$$\frac{\sigma_i^2}{2}\sum_{i=1}^{k}(\cos(2\pi f_i(t-s)) + \sin(2\pi f_i(t+s))) + \frac{\sigma_i^2}{2}\sum_{i=1}^{k}(\cos(2\pi f_j(t-s)) + \sin(2\pi f_j(t+s))) =$$

$$\sigma_i^2\sum_{i=1}^{k}\cos(2\pi f_i(t-s)) = \sigma_i^2\sum_{i=1}^{k}\cos(2\pi f_i k),$$

and is thus free of t and s.

2.26 Semivariogram

a

$$\Gamma_{t,s} = \frac{1}{2}E[(Y_t - Y_s)^2] = \frac{1}{2}E[Y_t^2 - 2Y_tY_s + Y_s^2] =$$

$$\frac{1}{2}\left(E[Y_t^2] - 2E[Y_tY_s] + E[Y_s^2]\right) = \frac{1}{2}\gamma_0 + \frac{1}{2}\gamma_0 - 2\times\frac{1}{2}\gamma_{|t-s|} = \gamma_0 - \gamma_{|t-s|}$$

$$\text{Cov}[Y_t, Y_s] = E[Y_tY_s] - \mu_t\mu_s = E[Y_tY_s] = \gamma_{|t-s|} \quad \square$$

b

$$Y_t - Y_s = e_t + e_{t-1} + \cdots + e_1 - e_s - e_{s-1} - \cdots - e_1 =$$

$$e_t + e_{t-1} + \cdots + e_{s+1}, \quad \text{for } t > s$$

$$\Gamma_{t,s} = \frac{1}{2}E[(Y_t - Y_s)^2] = \frac{1}{2}\text{Var}[e_t + e_{t-1} + \cdots + e_{s-1}] =$$

$$\frac{1}{2}\sigma_e^2(t-s) \quad \square$$

2.27   Polynomials

a

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

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

b

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

2.28   Random cosine wave extended

a

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

b

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

which is free of t.

2.29 Random cosine wave further

a

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

b

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

2.30 Rayleigh distribution

$$Y = R\cos\left(2\pi(ft + \phi)\right), \quad X = R\sin\left(2\pi(ft + \phi)\right)$$

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

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

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

$$\sum_{t=1}^{n} Y_t - n\beta_0 - \beta_1 \sum_{t=1}^{n} t = 0 \implies$$

$$\beta_0 = \frac{\sum_{t=1}^{n} Y_t - \beta_1 \sum_{t=1}^{n} t}{n} = \bar{Y} - \beta_1 \bar{t}$$

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

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

$$\beta_1 \sum_{t=1}^{n} t^2 = \sum_{t=1}^{n} Y_t t - \beta_0 \sum_{t=1}^{n} t$$

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

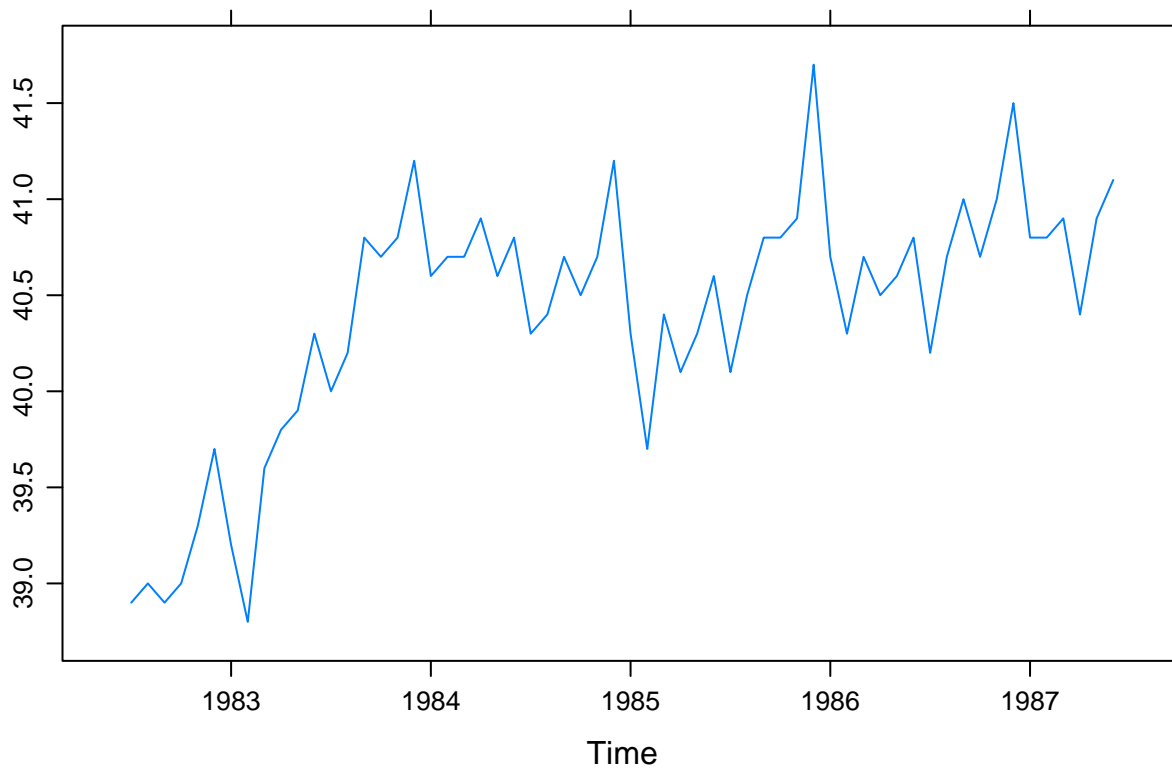```
library(TSA)
data("hours")
xyplot(hours)
```

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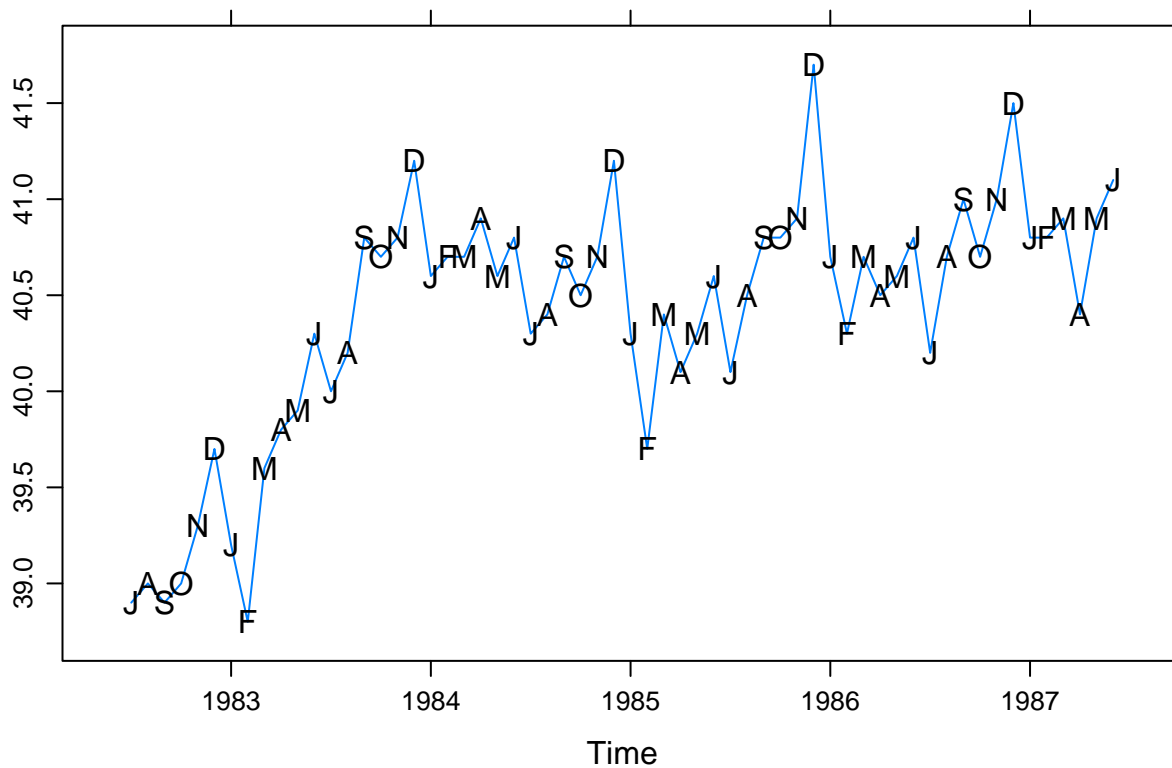
Figure 3.2: Monthly values of average hours worked per week with superposed initials of months.

```
  panel.xyplot(x, y, ...)
  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
## -0.23828 -0.04981  0.01942  0.05845  0.13136
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.490e+02  1.115e+01  -49.24   <2e-16 ***
```
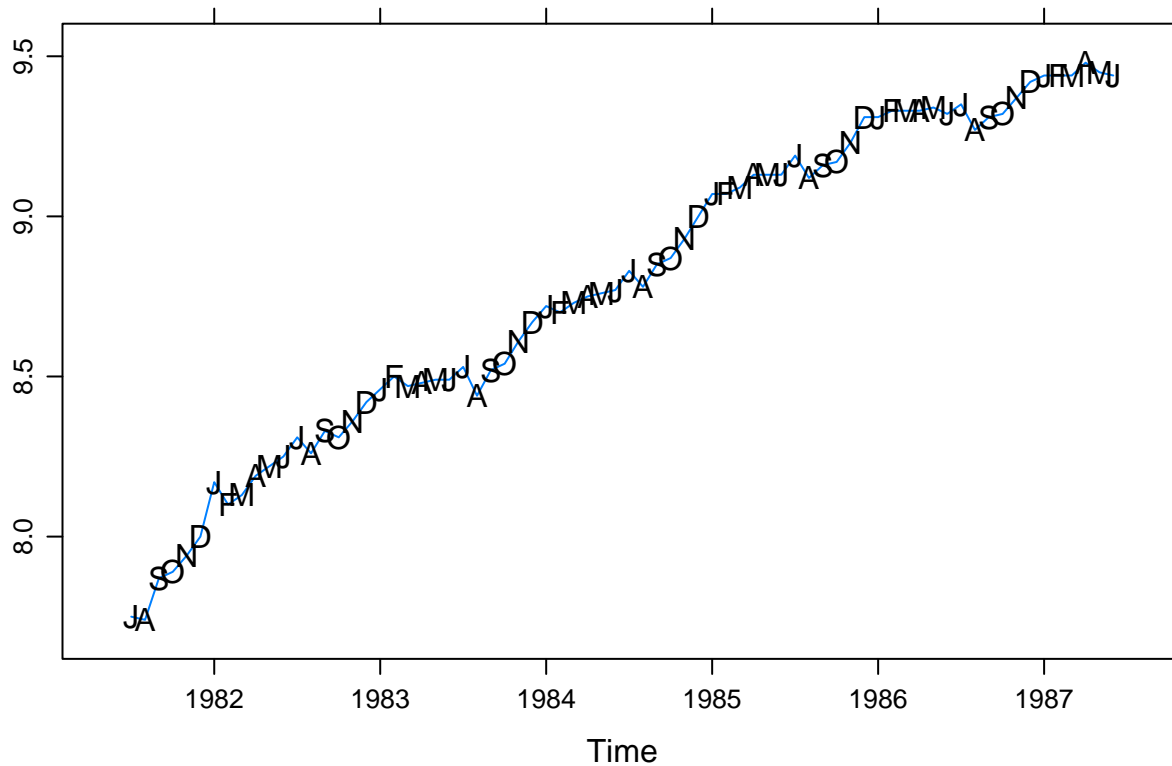
Figure 3.3: Monthly average hourly wages for workers in the U.S. apparel and textile industry.

```
## time(wages)  2.811e-01  5.618e-03    50.03   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08257 on 70 degrees of freedom
## Multiple R-squared:  0.9728, Adjusted R-squared:  0.9724
## F-statistic:  2503 on 1 and 70 DF,  p-value: < 2.2e-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)
```
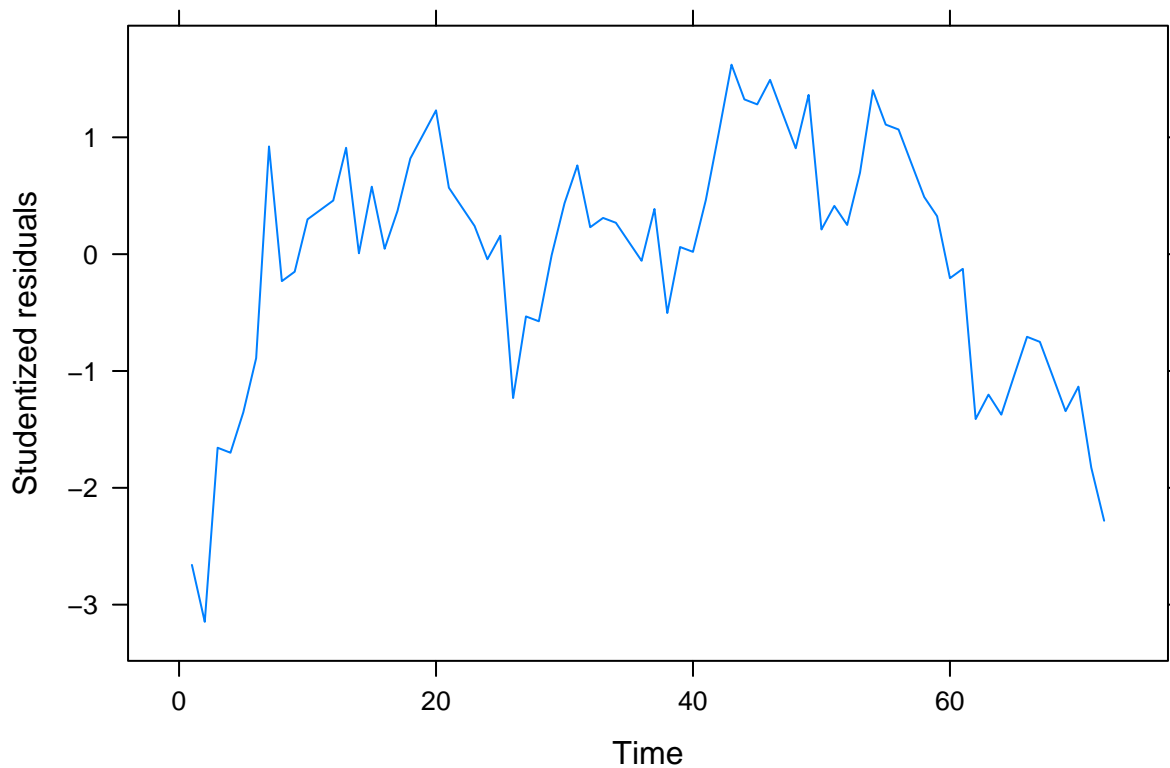
Figure 3.4: (#fig:wages_resid)Residual plot

```
## 
## Call:
## lm(formula = wages ~ time(wages) + I(time(wages)^2))
## 
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.148318 -0.041440  0.001563  0.050089  0.139839
## 
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -8.495e+04  1.019e+04  -8.336 4.87e-12 ***
## time(wages)      8.534e+01  1.027e+01   8.309 5.44e-12 ***
## I(time(wages)^2) -2.143e-02  2.588e-03  -8.282 6.10e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.05889 on 69 degrees of freedom
## Multiple R-squared:  0.9864, Adjusted R-squared:  0.986
## F-statistic:  2494 on 2 and 69 DF,  p-value: < 2.2e-16
```

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

e

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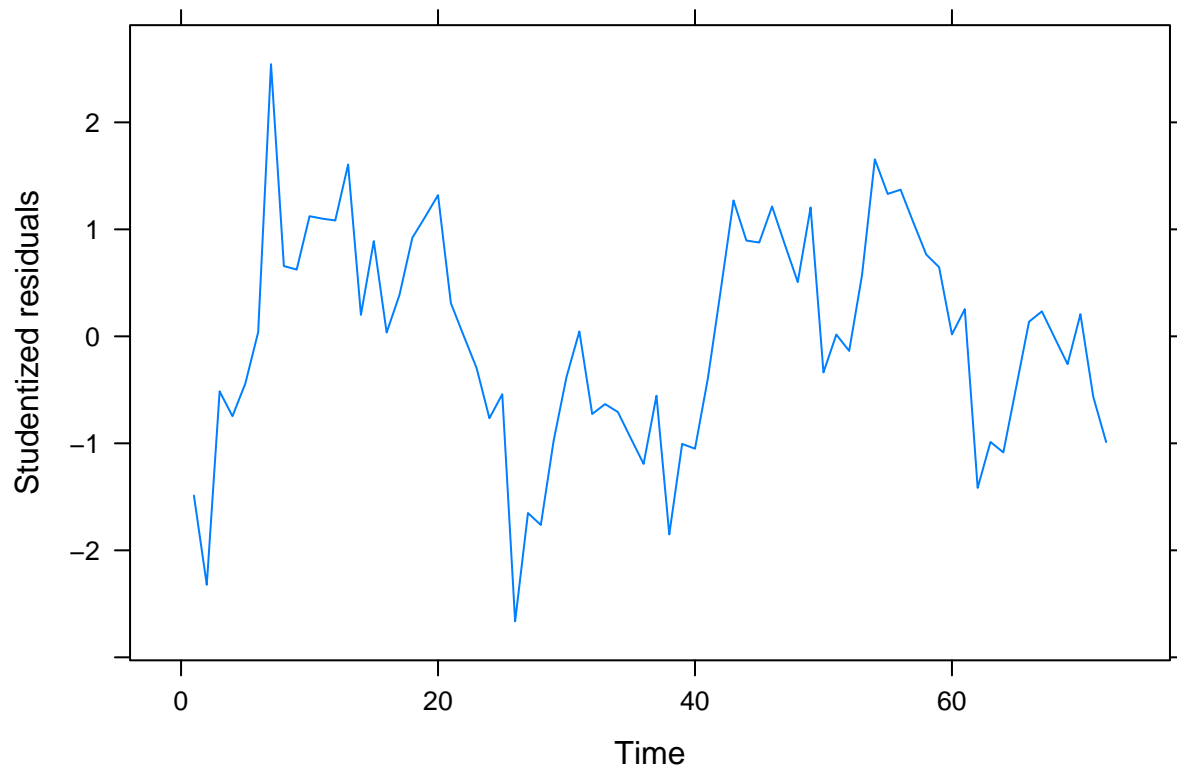

Figure 3.5: (#fig:wages_quad_resid)Residual plot for our quadratic model.

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.

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.
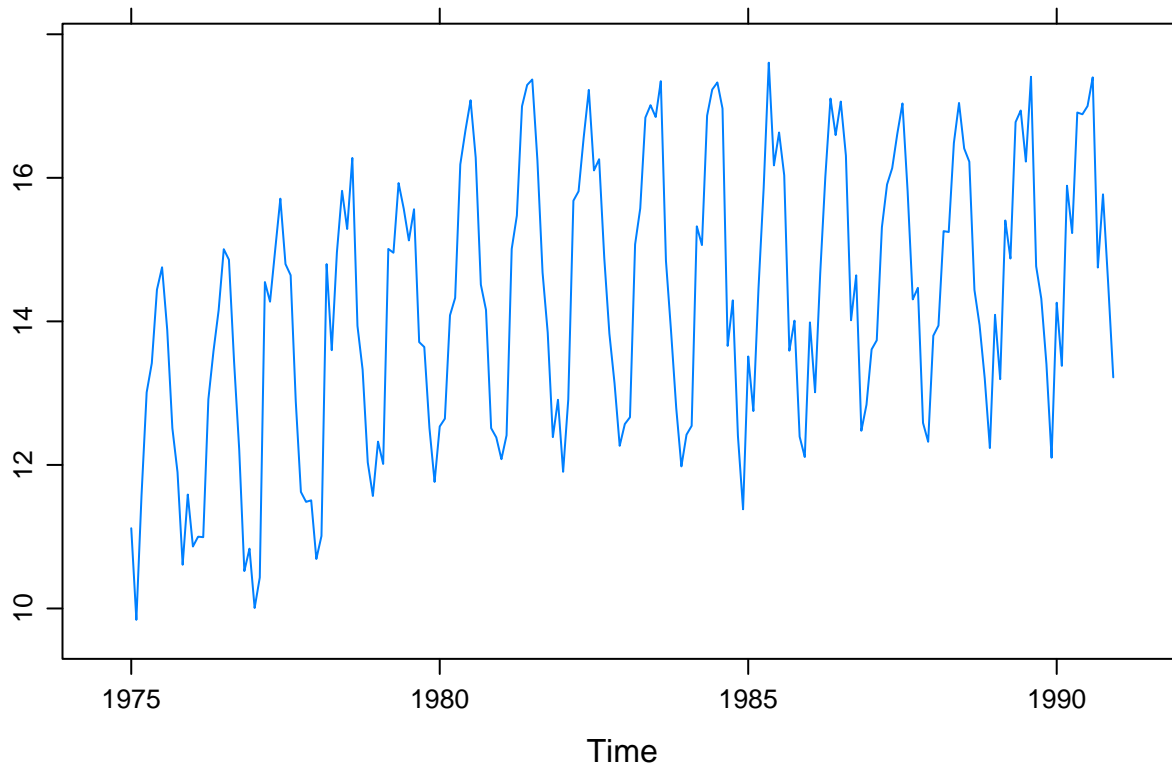
b

Figure 3.6: Monthly U.S. beer sales.

```r
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.

c

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

```
##
## Call:
## lm(formula = beersales ~ season(beersales))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5745 -0.4772  0.1759  0.7312  2.1023
```
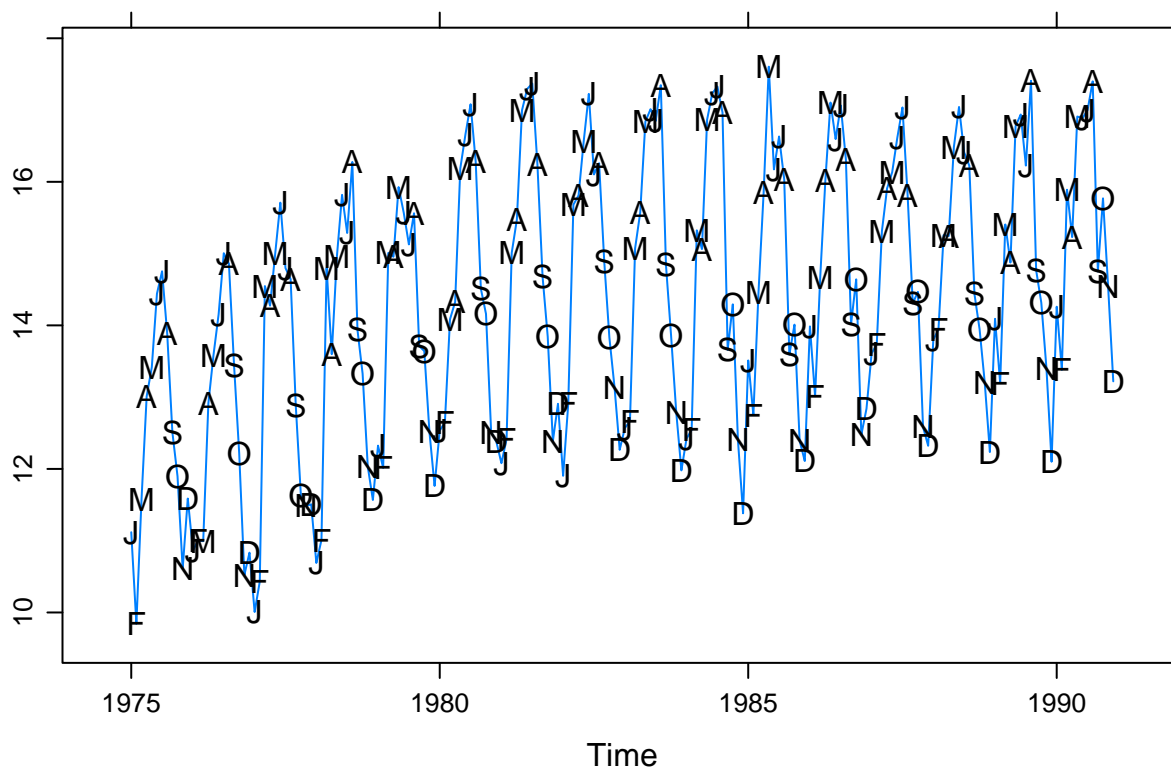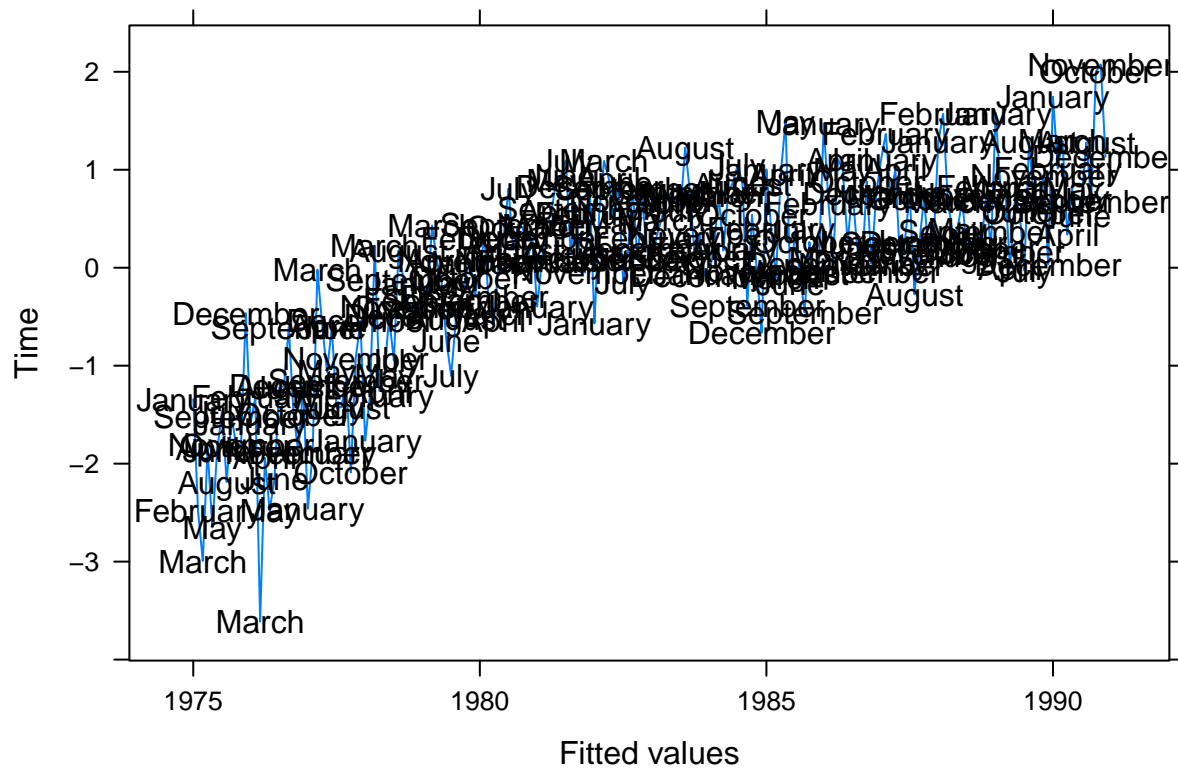
Figure 3.7: Monthly U.S. beer sales annotated with the months' initials.

```
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               12.48568    0.26392  47.309  < 2e-16 ***
## season(beersales)February -0.14259    0.37324  -0.382 0.702879
## season(beersales)March     2.08219    0.37324   5.579 8.77e-08 ***
## season(beersales)April     2.39760    0.37324   6.424 1.15e-09 ***
## season(beersales)May       3.59896    0.37324   9.643  < 2e-16 ***
## season(beersales)June      3.84976    0.37324  10.314  < 2e-16 ***
## season(beersales)July      3.76866    0.37324  10.097  < 2e-16 ***
## season(beersales)August    3.60877    0.37324   9.669  < 2e-16 ***
## season(beersales)September 1.57282    0.37324   4.214 3.96e-05 ***
## season(beersales)October   1.25444    0.37324   3.361 0.000948 ***
## season(beersales)November -0.04797    0.37324  -0.129 0.897881
## season(beersales)December -0.42309    0.37324  -1.134 0.258487
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.056 on 180 degrees of freedom
## Multiple R-squared:  0.7103, Adjusted R-squared:  0.6926
## F-statistic: 40.12 on 11 and 180 DF,  p-value: < 2.2e-16
```

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

```
xyplot(rstudent(beer_fit1) ~ time(beersales), type = "l",
       xlab = "Fitted values", ylab = "Time",
       panel = function(x, y, ...) {
         panel.xyplot(x, y, ...)
         panel.text(x, y, labels = season(beersales))
       })
```


e