Time Series Analysis and Its Applications
✂ 4th Edition ✂
Instructor's Manual

We assume that the package for the text, `astsa`, has been loaded. See the *Package Notes* at the web page for the text: `http://www.stat.pitt.edu/stoffer/tsa4/`

Chapter 1

Solutions

1.1 The major differences are how quickly the signal dies out in the explosion versus the earthquake and the larger amplitude of the signals in the explosion.

```
plot(EQ5, ylab="Earthquate/Explosion")
lines(EXP6, col=2)
legend("topright", c("EQ","EXP"), col=1:2, lty=1)
```

1.2 Figure 1.1 shows contrived data simulated according to this model. The modulating functions are also plotted. The code is given for part (b); part (a) is given in the text. For part (c), basically remove the `cos()` part.

```
s = c(rep(0,100), 10*exp(-(1:100)/200)*cos(2*pi*1:100/4))  # part (b)
x = ts(s + rnorm(200, 0, 1))
plot(x)
lines(c(rep(0,100), 10*exp(-(1:100)/200)))  # modulator on same plot as the series
```

The first signal bears a striking resemblance to the two arrival phases in the explosion. The second signal decays more slowly and looks more like the earthquake. The periodic behavior is emulated by the cosine function which will make one cycle every four points. If we assume that the data are sampled at 4 points per second, the data will make 1 cycle in a second, which is about the same rate as the seismic series.

1.3 Below is R code for parts (a)-(c). In all cases the moving average nearly annihilates (completely in the 2nd case) the signal. The signals in part (a) and (c) are similar.

```
w = rnorm(150,0,1)          # 50 extra to avoid startup problems
x = filter(w, filter=c(0,-.9), method="recursive")[-(1:50)]  # AR
x2 = 2*cos(2*pi*(1:100)/4)    # sinusoid
x3 = x2 + rnorm(100,0,1)      # sinusoid + noise
v = filter(x, rep(1,4)/4)     # moving average
v2 = filter(x2, rep(1,4)/4)   # moving average
v3 = filter(x3, rep(1,4)/4)   # moving average
par(mfrow=c(3,1))
plot.ts(x, main="autoregression")
lines(v,lty="dashed")
plot.ts(x2, main="sinusoid")
lines(v2,lty="dashed")
plot.ts(x3, main="sinusoid + noise")
lines(v3,lty="dashed")
```

1.4 Simply expand the binomial product inside the expectation and use the fact that μ_t is a nonrandom constant, i.e.,

$$
\begin{aligned}
\gamma(s,t) &= \mathrm{E}[(x_s x_t - \mu_s x_t - x_s \mu_t + \mu_s \mu_t] \\
&= \mathrm{E}(x_s x_t) - \mu_s \mathrm{E}(x_t) - \mathrm{E}(x_s)\mu_t + \mu_s \mu_t
\end{aligned}
$$

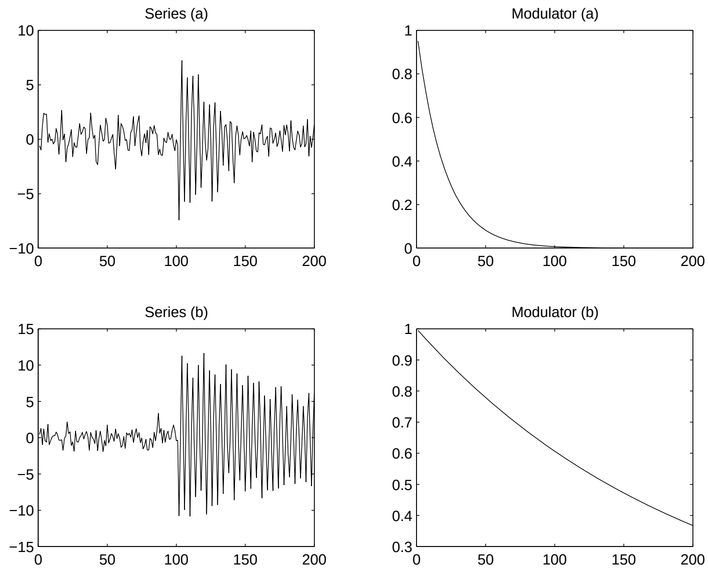

Fig. 1.1. Simulated series with exponential modulations

$$= \mathrm{E}(x_s x_t) - \mu_s\mu_t - \mu_s\mu_t + \mu_s\mu_t$$

1.5 (a) In each case the signals are fixed, so $\mathrm{E}(x_t) = \mathrm{E}(s_t + w_t) = s_t + \mathrm{E}(w_t) = s_t$. To plot the means, repeat Problem 1.2 and just plot the signals (s) without the noise.
(b) The autocovariance function $\gamma(t, u) = \mathrm{E}[(x_t - s_t)(x_u - s_u) = \mathrm{E}(w_t w_u)$, which is one (1) when $t = u$ and zero (0) otherwise.

1.6 (a) Since $\mathrm{E}x_t = \beta_1 + \beta_2 t$, the mean is not constant, so the process is not stationary. Note that

$$\begin{aligned} x_t - x_{t-1} &= \beta_1 + \beta_2 t + w_t - \beta_1 - \beta_2(t-1) - w_{t-1} \\ &= \beta_2 + w_t - w_{t-1}, \end{aligned}$$

which is clearly stationary. Verify that the mean is β_2 and the autocovariance is 2 for $s = 2$ and -1 for $|s - t| = 1$ and is zero for $|s - t| > 1$.
(b) First, write

$$\begin{aligned} \mathrm{E}(y_t) &= \frac{1}{2q+1} \sum_{j=-q}^{q} [(\beta_1 + \beta_2(t - j))] \\ &= \frac{1}{2q+1} \left[(2q+1)(\beta_1 + \beta_2 t) - \beta_2 \sum_{j=-q}^{q} j \right] \\ &= \beta_1 + \beta_2 t \end{aligned}$$

because the positive and negative terms in the last sum cancel out. To get the covariance write the process as

$$y_t = \sum_{j=-\infty}^{\infty} a_j w_{t-j},$$

where $a_j = 1, j = -q, \ldots, 0, \ldots, q$ and is zero otherwise. To get the covariance, note that we need

$$
\begin{aligned}
\gamma_y(h) &= \mathrm{E}[(y_{t+h} - \mathrm{E}y_{t+h})(y_t - \mathrm{E}y_t)] \\
&= (2q+1)^{-2} \sum_j \sum_k a_j a_k \mathrm{E} w_{t+h-j} w_{t-k} \\
&= \frac{\sigma^2}{(2q+1)^2} \sum_{j,k} a_j a_k \delta_{h+k-j}, \\
&= \sum_{j=-\infty}^{\infty} a_{j+h} a_j,
\end{aligned}
$$

where $\delta_{h+k-j} = 1, j = k+h$ and is zero otherwise. Writing out the terms in $\gamma_y(h)$, for $h = 0, \pm 1, \pm 2, \ldots$, we obtain

$$
\gamma_y(h) = \frac{\sigma^2(2q+1-|h|)}{(2q+1)^2}
$$

for $h = 0, \pm 1, \pm 2, \ldots, \pm 2q$ and zero for $|h| > q$.

1.7 By a computation analogous to that appearing in Example 1.17, we may obtain

$$
\gamma(h) = \begin{cases} 6\sigma_w^2 & h = 0 \\ 4\sigma_w^2 & \text{h=}\pm 1 \\ \sigma_w^2 & \text{h=}\pm 2 \\ 0 & |h|\text{>2.} \end{cases}
$$

The autocorrelation is obtained by dividing the autocovariances by $\gamma(0) = 6\sigma_w^2$.

1.8 (a) Simply substitute $\delta s + \sum_{k=1}^{s} w_k$ for x_s to see that

$$
\underbrace{\delta t + \sum_{k=1}^{t} w_k}_{x_t} = \delta + \underbrace{\left(\delta(t-1) + \sum_{k=1}^{t-1} w_k\right)}_{x_{t-1}} + w_t.
$$

Alternately, the result can be shown by induction.

(b) For the mean,

$$
\mathrm{E}x_t = \mathrm{E}\left(\delta t + \sum_{k=1}^{t} w_k\right) = \delta t + \sum_{k=1}^{t} \mathrm{E}w_k = \delta t.
$$

For the covariance, without loss of generality, consider the case $s \le t$.

$$
\begin{aligned}
\gamma(s,t) &= \mathrm{cov}(x_s, x_t) = \mathrm{E}\{(x_s - \delta s)(x_t - \delta t)\} \\
&= \mathrm{E}\left\{\sum_{j=1}^{s} w_j \sum_{k=1}^{t} w_k\right\} \\
&= \mathrm{E}\left\{(w_1 + \cdots + w_s)(w_1 + \cdots + w_s + w_{s+1} + \ldots + w_t)\right\} \\
&= \sum_{j=1}^{s} \mathrm{E}(w_j^2) = s\,\sigma_w^2. \qquad [\text{or } \min(s,t)\,\sigma_w^2]
\end{aligned}
$$

(c) The series is nonstationary because both the mean function and the autocovariance function depend on time, t.

(d) From (b), $\rho_x(t-1,t) = (t-1)\sigma_w^2 / \sqrt{(t-1)\sigma_w^2}\sqrt{t\sigma_w^2}$, which yields the result. The implication is that the series tends to change slowly.

(e) One possibility is to note that $\nabla x_t = x_t - x_{t-1} = \delta + w_t$, which is stationary because $\mu_{x,t} = \delta$ and $\gamma_x(t+h,t) = \sigma_w^2 \delta_0(h)$ are both independent of time t, where $\delta_0(h)$ is the delta measure.

1.9 Because $\mathrm{E}(U_1) = \mathrm{E}(U_2) = 0$, we have $\mathrm{E}(x_t) = 0$. Then,

$$
\begin{aligned}
\gamma(h) &= \mathrm{E}(x_{t+h}x_t) \\
&= \mathrm{E}\Big\{ \big(U_1 \sin[2\pi\omega_0(t+h)] + U_2 \cos[2\pi\omega_0(t+h)]\big)\big(U_1 \sin[2\pi\omega_0 t] + U_2 \cos[2\pi\omega_0 t]\big) \Big\} \\
&= \sigma_w^2 \Big(\sin[2\pi\omega_0(t+h)] \sin[2\pi\omega_0 t] + \cos[2\pi\omega_0(t+h)] \cos[2\pi\omega_0 t] \Big) \\
&= \sigma_w^2 \cos[2\pi\omega_0(t+h) - 2\pi\omega_0 t] \\
&= \sigma_w^2 \cos[2\pi\omega_0 h]
\end{aligned}
$$

by the standard trigonometric identity, $\cos(A-B) = \sin A \sin B + \cos A \cos B$.

1.10 (a)

$$
MSE(A) = \mathrm{E}\Big\{ x_{t+\ell}^2 - 2A\mathrm{E}(x_{t+\ell}x_t) + A^2\mathrm{E}(x_t^2) \Big\} = \gamma(0) - 2A\gamma(\ell) + A^2\gamma(0).
$$

Setting the derivative with respect to A to zero yields

$$
-2\gamma(\ell) + 2A\gamma(0) = 0
$$

and solving gives the required value.

(b)

$$
MSE(A) = \gamma(0)\Big[1 - 2\frac{\rho(\ell)\gamma(\ell)}{\gamma(0)} + \rho^2(\ell)\Big] = \gamma(0)\Big[1 - 2\rho^2(\ell) + \rho^2(\ell)\Big] = \gamma(0)\Big[1 - \rho^2(\ell)\Big].
$$

(c) If $x_{t+\ell} = Ax_t$ with probability one, then

$$
\mathrm{E}(x_{t+\ell} - Ax_t)^2 = \gamma(0)\Big[1 - \rho^2(\ell)\Big] = 0
$$

implying that $\rho(\ell) = \pm 1$. Since $A = \rho(\ell)$, the conclusion follows.

1.11 (a) Since $x_t = \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}$,

$$
\gamma(h) = \mathrm{E} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j w_{t+h-j} w_{t-k} \psi_k = \sigma_w^2 \sum_{j,k} \psi_j \psi_k \delta_{h-j+k} = \sigma_w^2 \sum_{k=-\infty}^{\infty} \psi_{k+h}\psi_k,
$$

where $\delta_t = 1$ for $t = 0$ and is zero otherwise.

(b) The proof is identical to the one given in Appendix A, Example A.2.

1.12

$$
\gamma_{xy}(h) = \mathrm{E}[(x_{t+h} - \mu_x)(y_t - \mu_y)] = \mathrm{E}[(y_t - \mu_y)(x_{t+h} - \mu_x)] = \gamma_{yx}(-h)
$$

1.13 (a)

$$
\gamma_y(h) = \begin{cases} \sigma_w^2(1+\theta^2) + \sigma_u^2 & h = 0 \\ -\theta\sigma_w^2 & h = \pm 1 \\ 0 & |h| > 1. \end{cases}
$$

(b)

$$\gamma_{xy}(h) = \begin{cases} \sigma_w^2 & h = 0 \\ -\theta\sigma_w^2 & h = -1 \\ 0 & \text{otherwise.} \end{cases}$$

$$\gamma_x(h) = \begin{cases} \sigma_w^2 & h = 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$\rho_{xy}(h) = \frac{\gamma_{xy}(h)}{\sqrt{\gamma_x(0)\gamma_y(0)}}.$$

(c) The processes are jointly stationary because the autocovariance and cross-covariance functions depend only on lag h.

1.14 (a) For the mean, write

$$\mathrm{E}(y_t) = \mathrm{E}(\exp\{x_t\}) = \exp\left\{\mu_x + \frac{1}{2}\gamma_x(0)\right\},$$

using the given equation at $\lambda = 1$.

(b) For the autocovariance function, note that

$$\begin{aligned} \mathrm{E}(y_{t+h}y_t) &= \mathrm{E}\big(\exp\{x_{t+h}\}\exp\{x_t\}\big) \\ &= \mathrm{E}\big(\exp\{x_{t+h} + x_t\}\big) \\ &= \exp\{2\mu_x + \gamma_x(0) + \gamma_x(h)\}, \end{aligned}$$

since $x_t + x_{t+h}$ is the sum of two correlated normal random variables and will be normally distributed with mean $2\mu_x$ and variance

$$\gamma_x(0) + \gamma_x(0) + 2\gamma_x(h) = 2\big(\gamma_x(0) + \gamma(h)\big).$$

For the autocovariance of y_t,

$$\begin{aligned} \gamma_y(h) &= \mathrm{E}(y_{t+h}y_t) - \mathrm{E}(y_{t+h})\mathrm{E}(y_t) \\ &= \exp\{2\mu_x + \gamma_x(0) + \gamma_x(h)\} - \Big(\exp\big\{\mu_x + \frac{1}{2}\gamma_x(0)\big\}\Big)^2 \\ &= \exp\{2\mu_x + \gamma_x(0)\}\big(\exp\{\gamma_x(h)\} - 1\big). \end{aligned}$$

1.15 The process is stationary because

$$\mu_{x,t} = \mathrm{E}(x_t) = \mathrm{E}(w_t w_{t-1}) = \mathrm{E}(w_t)\mathrm{E}(w_{t-1}) = 0;$$

$$\gamma_x(0) = \mathrm{E}(w_t w_{t-1} w_t w_{t-1}) = \mathrm{E}(w_t^2)\mathrm{E}(w_{t-1}^2) = \sigma_w^2\sigma_w^2 = \sigma_w^4,$$

$$\gamma_x(1) = \mathrm{E}(w_{t+1} w_t w_t w_{t-1}) = \mathrm{E}(w_{t+1})\mathrm{E}(w_t^2)\mathrm{E}(w_{t-1}) = 0 = \gamma(-1),$$

and similar computations establish that $\gamma_x(h) = 0$, for $|h| \geq 1$. The series is white noise.

1.16 (a) For $t = 1, 2, \ldots$,

$$\mathrm{E}(x_t) = \int_0^1 \sin(2\pi ut)du = -\frac{1}{2\pi t}\cos(2\pi ut)\Big|_0^1 = -\frac{1}{2\pi t}\big[\cos(2\pi t) - 1\big] = 0;$$

$$\gamma(h) = \int_0^1 \sin[2\pi u(t+h)]\sin[2\pi ut]\,du.$$

Using the identity $2\sin(\alpha)\sin(\beta) = \cos(\alpha - \beta) - \cos(\alpha + \beta)$ gives $\gamma(0) = 1/2$ and $\gamma(h) = 0$, for $h \neq 0$.

(b) **This part of the problem is harder than it seems at first and it might be a good idea to omit it in more elementary presentations.** The easiest way to tackle the problem is to calculate some probabilities (which can be given as a hint), e.g., because U is uniform on $[0, 1]$,

$$\begin{aligned}&\Pr\{x_1 \le 0,\, x_3 \le 0\} = \Pr\{\sin(2\pi U) \le 0,\, \sin(2\pi 3U) \le 0\}\\ &= \Pr\Big\{U \in [1/2, 1] \cap \Big([1/6, 1/3] \cup [1/2, 2/3] \cup [5/6, 1]\Big)\Big\} = 0 + 1/6 + 1/6 = \frac{1}{3}\end{aligned}$$

but, similarly,

$$\Pr[x_2 \le 0, x_4 \le 0] = \frac{1}{4}.$$

1.17 (a) The essential part of the exponent of the characteristic [or moment generating] function is

$$\begin{aligned}\sum_{j=1}^{n} \lambda_j x_j &= \sum_{j=1}^{n} \lambda_j (w_j - \theta w_{j-1})\\ &= -\lambda_1 \theta w_0 + \sum_{j=1}^{n-1} (\lambda_j - \theta \lambda_{j+1}) w_j + \lambda_n w_n.\end{aligned}$$

Because the w_t are independent and identically distributed, the characteristic function can be written as

$$\phi(\lambda_1, \dots, \lambda_n) = \phi_w(-\lambda_1\theta) \prod_{j=1}^{n-1} \phi_w(\lambda_j - \theta\lambda_{j+1}) \phi_w(\lambda_n)$$

(b) Because the joint distribution of the w_j will not change simply by shifting $x_1, \dots, x_n$ to $x_{1+h}, \dots, x_{n+h}$, the characteristic function [or MGF] remains the same.

1.18 Letting $k = j + h$, holding j fixed after substituting from (1.29) yields

$$\begin{aligned}\sum_{h=-\infty}^{\infty} |\gamma(h)| &= \sigma_w^2 \sum_{h=-\infty}^{\infty} \Big| \sum_{j=-\infty}^{\infty} \psi_{j+h}\psi_j \Big| \le \sigma_w^2 \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |\psi_{j+h}||\psi_j|\\ &= \sigma_w^2 \sum_{k=-\infty}^{\infty} |\psi_k| \sum_{j=-\infty}^{\infty} |\psi_j| < \infty.\end{aligned}$$

1.19 (a) $\mathrm{E}(x_t) = \mathrm{E}(\mu + w_t + \theta w_{t-1}) = \mu$.

(b) $\gamma(h) = \mathrm{cov}(w_{t+h} + \theta w_{t+h-1}, w_t + \theta w_{t-1}$, so $\gamma_x(0) = (1 + \theta^2)\sigma_w^2$, $\gamma_x(\pm 1) = \theta\sigma_w^2$, and 0 otherwise.

(c) From (a) and (b) we see that, for any θ, both the mean function and autocovariance function are independent of time.

(d) From the given formula, and because $\gamma_x(h) = 0$ for $|h| > 1$, we have

$$\mathrm{var}(\bar{x}) = \frac{1}{n}\left[\gamma_x(0) + \frac{2(n-1)}{n}\gamma_x(1)\right].$$

- When $\theta = 0$, $\gamma_x(\pm 1) = 0$ and it's the classical case, $\mathrm{var}(\bar{x}) = \sigma_w^2/n$.
- When $\theta = 1$, $\gamma_x(0) = 2\sigma_w^2$ and $\gamma_x(\pm 1) = \sigma_w^2$, so $\mathrm{var}(\bar{x}) = \frac{\sigma_w^2}{n}[2 + \frac{2(n-1)}{n}] = \frac{\sigma_w^2}{n}[4 - \frac{2}{n}]$. In this case, the variance is about 4 times as large as the uncorrelated case.
- When $\theta = -1$, $\gamma_x(0) = 2\sigma_w^2$ and $\gamma_x(\pm 1) = -\sigma_w^2$, so $\mathrm{var}(\bar{x}) = \frac{\sigma_w^2}{n}[2 - \frac{2(n-1)}{n}] = \frac{\sigma_w^2}{n}[\frac{2}{n}]$. In this case, the variance is smaller ($n > 2$) than the uncorrelated case and the variance is nearly zero for moderate n.

(e) It's easier to estimate the mean when θ is negative. Negatively correlated data vary around the mean more tightly than positively correlated data. In essence, you need fewer data to identify the mean if the data are negatively correlated.

1.20 Code for parts (a) and (b) is below. Students should have about 1 in 20 ACF values within the bounds, but the values for part (b) will be larger in general than for part (a).

```
wa = rnorm(500,0,1)
wb = rnorm(50,0,1)
par(mfrow=c(2,1))
(acf(wa, 20))    # plot and print results
(acf(wb, 20))    # plot and print results
```

1.21 This is similar to the previous problem. Generate 2 extra observations due to loss of the end points in making the MA.

```
wa = rnorm(502,0,1)
wb = rnorm(52,0,1)
va = filter(wa, sides=2, rep(1,3)/3)
vb = filter(wb, sides=2, rep(1,3)/3)
par(mfrow=c(2,1))
(acf(va, 20, na.action = na.pass))   # plot and print results
(acf(vb, 20, na.action = na.pass))   # plot and print results
```

1.22 Generate the data as in Problem 1.2 and then type `acf(x)`. The sample ACF will exhibit significant correlations at one cycle every four lags, which is the same frequency as the signal. (The process is not stationary because the mean function is the signal, which depends on time t.)

1.23 The sample ACF should look sinusoidal, making one cycle every 50 lags.

```
x = 2*cos(2*pi*(1:500)/50 + .6*pi)+ rnorm(500,0,1)
acf(x,100)
```

1.24 $\gamma_y(h) = \text{cov}(y_{t+h}, y_t) = \text{cov}(x_{t+h} - .7x_{t+h-1}, x_t - .7x_{t-1}) = 0$ if $|h| > 1$ because the x_ts are independent. When $h = 0$, $\gamma_y(0) = \sigma_x^2(1+.7^2)$, where σ_x^2 is the variance of x_t. When $h = 1$, $\gamma_y(1) = -.7\sigma_x^2$. Thus, $\rho_y(1) = -.7/(1 + .7^2) = -.47$

1.25 (a) The variance is always non-negative, so for x_t a stationary series

$$\text{var}\left\{\sum_{s=1}^{n} a_s x_s\right\} = \text{cov}\left\{\sum_s a_s x_s, \sum_t a_t x_t\right\} = \sum_{s=1}^{n}\sum_{t=1}^{n} a_s \gamma(s-t) a_t = \boldsymbol{a}'\mathit{\Gamma}\boldsymbol{a} \geq 0,$$

thus $\mathit{\Gamma} = \{\gamma(s-t), s, t = 1, \dots, n\}$ is a non-negative definite matrix.

(b) Let $Y_t = x_t - \bar{x}$ for $t = 1 \dots n$ and construct the $n \times 2n$ matrix

$$D = \begin{pmatrix} 0 & 0 & \cdots & 0 & Y_1 & Y_2 & \cdots & Y_n \\ 0 & \cdots & 0 & Y_1 & Y_2 & \cdots & Y_n & 0 \\ \vdots & & & & \vdots & & & \vdots \\ 0 & Y_1 & Y_2 & \cdots & Y_n & 0 & \cdots & 0 \end{pmatrix}$$

With $\hat{\mathit{\Gamma}}_n = \{\hat{\gamma}(s-t)\}_{s,t=1}^{n}$, it is easy to show that

$$\hat{\mathit{\Gamma}}_n = \frac{1}{n} DD'.$$

Then, for any vector $\boldsymbol{a} \in \mathbb{R}^n$,

$$\boldsymbol{a}'\hat{\mathit{\Gamma}}_n\boldsymbol{a} = \frac{1}{n}\boldsymbol{a}'DD'\boldsymbol{a} = \frac{1}{n}\boldsymbol{c}'\boldsymbol{c} = \sum_{i=1}^{n} c_i^2 \geq 0$$

for $\boldsymbol{c} = D'\boldsymbol{a}$. Noting that $\hat{\Gamma}_n$ is symmetric, it will be positive definite (p.d.) if its eigenvalues are positive. Since the diagonal elements of $\hat{\Gamma}_n$ are $\hat{\gamma}(0)$, the sum of the eigenvalues of $\hat{\Gamma}_n$ is $n\hat{\gamma}(0)$. Consequently, $\hat{\Gamma}_n$ is p.d. as long as $\hat{\gamma}(0) > 0$. In other terms, if the sample variance of the data is not zero, $\hat{\Gamma}_n$ is p.d.

1.26 (a)

$$\mathrm{E}\bar{x}_t = \frac{1}{N}\sum_{j=1}^{N} x_{jt} = \frac{1}{N}\sum_{j=1}^{n} \mu_t = \frac{N\mu_t}{N} = \mu_t$$

(b)

$$\mathrm{E}[(\bar{x}_t - \mu_t)^2] = \frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N}\mathrm{E}(x_{jt}-\mu_t)(x_{kt}-\mu_t) = \frac{1}{N^2}\sum_{j=1}^{N} e_{jt}^2 = \frac{1}{N}\gamma_e(t,t)$$

(c) As long as the separate series are observing the same signal, we may assume that the variance goes down proportionally to the number series as in the iid case. If normality is reasonable, pointwise $100(1-\alpha)$ % intervals can be computed as

$$\bar{x}_t \pm z_{\alpha/2}\gamma_e(t,t)/\sqrt{N}$$

1.27

$$V_x(h) = \frac{1}{2}\mathrm{E}[(x_{s+h}-\mu)-(x_s-\mu)]^2 = \frac{1}{2}[\gamma(0)-\gamma(h)-\gamma(-h)+\gamma(0)] = \gamma(0)-\gamma(h).$$

1.28 The numerator and denominator of $\hat{\rho}(h)$ are

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}[\beta_1(t-\bar{t})+\beta_1 h][\beta_1(t-\bar{t})] = \frac{\beta_1^2}{n}\left[\sum_{t=1}^{n-h}(t-\bar{t})^2 + h\sum_{t=1}^{n-h}(t-\bar{t})\right]$$

and

$$\hat{\gamma}(0) = \frac{\beta_1^2}{n}\sum_{t=1}^{n}(t-\bar{t})^2.$$

Now, write the numerator as

$$\hat{\gamma}(h) = \hat{\gamma}(0) + \frac{\beta_1^2}{n}\left[-\sum_{t=n-h+1}^{n}(t-\bar{t})^2 - h\sum_{t=n-h+1}^{n}(t-\bar{t})\right]$$

Hence, we can write

$$\hat{\rho}(h) = 1 + R$$

where

$$R = \frac{\beta_1^2}{n\hat{\gamma}(0)}\left[-\sum_{t=n-h+1}^{n}(t-\bar{t})^2 - h\sum_{t=n-h+1}^{n}(t-\bar{t})\right]$$

is a remainder term that needs to converge to zero. We can evaluate the terms in the remainder using

$$\sum_{t=1}^{m} t = \frac{m(m+1)}{2}$$

and

$$\sum_{t=1}^{m} t^2 = \frac{m(2m+1)(m+1)}{6}$$

The denominator reduces to

$$
\begin{aligned}
n\hat{\gamma}(0) &= \beta_1^2\Bigg[\sum_{t=1}^{n} t^2 - n\bar{t}^2\Bigg] \\
&= \beta_1^2\Bigg[\frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4}\Bigg] \\
&= \beta_1^2\frac{n(n+1)(n-1)}{12},
\end{aligned}
$$

whereas the numerator can be simplified by letting $s = t - n + h$ so that

$$
R = \frac{\beta_1^2}{n\hat{\gamma}(0)}\Bigg[-\sum_{s=1}^{h}(s+n-h-\bar{t})^2 - h\sum_{s=1}^{h}(s+n-h-\bar{t})\Bigg]
$$

The terms in the numerator of R are $O(n^2)$, whereas the denominator is $O(n^3)$ so that the remainder term converges to zero.

1.29 (a)

$$
\Pr\{\sqrt{n}|\bar{x}| > \epsilon\} \le n\frac{\mathrm{E}[\bar{x}^2]}{\epsilon^2}
$$

Note that,

$$
n\mathrm{E}[\bar{x}^2] \to \sum_{u=-\infty}^{\infty} \gamma(h) = 0,
$$

where the last step employs the summability condition. The variance of $\bar{x}$ is derived in (1.33).

(b) An example of such a process is $x_t = \nabla w_t = w_t - w_{t-1}$, where w_t is white noise. This situation arises when a stationary process is over-differenced (i.e., w_t is already stationary, so ∇w_t would be considered over-differencing).

1.30 Let $y_t = x_t - \mu_x$ and write the difference as

$$
\begin{aligned}
n^{1/2}\big(\tilde{\gamma}(h) - \hat{\gamma}(h)\big) &= n^{-1/2}\sum_{t=1}^{n} y_{t+h}y_t - n^{-1/2}\sum_{t=1}^{n-h}(y_{t+h}-\bar{y})(y_t-\bar{y}) \\
&= n^{-1/2}\Bigg[\sum_{t=n-h+1}^{n} y_{t+h}y_t + \bar{y}\sum_{t=1}^{n-h} y_t + \bar{y}\sum_{t=1}^{n-h} y_{t+h} - (n-h)\bar{y}^2\Bigg]
\end{aligned}
$$

For the first term

$$
\begin{aligned}
\mathrm{E}\Bigg[n^{-1/2}\Big|\sum_{t=n-h+1}^{n} y_{t+h}y_t\Big|\Bigg] &\le n^{-1/2}\mathrm{E}\sum_{t=n-h+1}^{n} |y_{t+h}y_t| \\
&\le n^{-1/2}\sum_{t=n-h+1}^{n} \mathrm{E}^{1/2}[y_{t+h}^2]\mathrm{E}^{1/2}[y_t^2] \\
&= n^{-1/2}h\gamma_x(0) \\
&\to 0,
\end{aligned}
$$

as $n \to \infty$. Applying the Markov inequality in the hint then shows that the first term is $o_p(1)$. In order to handle the other terms, which differ trivially from $n^{-1/2}n\bar{y}^2$, note that, from Theorem A.5, $n^{1/2}\bar{y}$ converging in distribution to a standard normal implies that $n\bar{y}^2$ converges in distribution to a chi-square random variable with 1 degree of freedom and hence $n\bar{y}^2 = O_p(1)$. Hence, $n^{-1/2}n\bar{y}^2 = n^{-1/2}O_p(1) = o_p(1)$ and the result is proved.

1.31 To apply Theorem A.7, we need the ACF of x_t. Note that

$$\begin{aligned}\gamma_x(h) &= \sum_{j,k} \phi^j \phi^k \mathrm{E}[w_{t+h-j} w_{t-k}] \\ &= \sigma_w^2 \sum_k \phi^{h+k} \phi^k \\ &= \sigma_w^2 \phi^h \sum_{k=0}^{\infty} \phi^{2k} \\ &= \frac{\sigma_w^2 \phi^h}{1-\phi^2},\end{aligned}$$

and we have $\rho_x(h) = \phi^h$ for the ACF. Now, from (A.55), we have

$$\begin{aligned}w_{11} &= \sum_{u=1}^{\infty} \big[\rho_x(u+1) + \rho_x(u-1) - 2\rho_x(1)\rho_x(u)\big]^2 \\ &= \sum_{u=1}^{\infty} \big[\phi^{u+1} + \phi^{u-1} - 2\phi^{u+1}\big] \\ &= \frac{(1-\phi^2)^2}{\phi^2} \sum_{u=1}^{\infty} \phi^{2u} \\ &= 1-\phi^2.\end{aligned}$$

The limiting result implies that

$$\hat{\rho}(1) \sim AN\left\{\phi, \frac{1-\phi^2}{n}\right\}.$$

In order to derive a $100(1-\alpha)$ % confidence interval, note that

$$\frac{n(\hat{\rho}(1)-\phi)^2}{1-\phi^2} \le z_{\alpha/2}^2$$

with probability $1-\alpha$. Looking at the roots of

$$A\phi^2 + B\phi + C = 0,$$

where

$$\begin{aligned}A &= (1 + \frac{z_{\alpha/2}^2}{n}), \\ B &= -2\hat{\rho}(1),\end{aligned}$$

and

$$C = \hat{\rho}^2(1) - \frac{z_{\alpha/2}^2}{n}$$

gives the interval

$$-\frac{B}{2A} \pm \frac{\sqrt{B^2-4AC}}{2A}$$

Taking $\hat{\rho}(1) = .64, n = 100, z_{.025} = 1.96$ gives the approximate 95% confidence interval $(.47, .77)$.

1.32 (a) $\mathrm{E}(x_t x_{t+h}) = 0$ and $\mathrm{E}(x_t x_{t+h} x_s x_{s+k}) = 0$ unless all subscripts match. But $t \neq s$, and $h, k \geq 1$, so all subscripts can't match and hence $\mathrm{cov}(x_t x_{t+h}, x_s x_{s+k}) = 0$.

(b) Define $y_t = x_t \sum_{j=1}^h \lambda_j x_{t+j}$, for $\lambda_1, \ldots, \lambda_h \in \mathbf{R}$ arbitrary. Then y_t is strictly stationary, h-dependent, and $\text{var}(y_t) = \sigma^4 \sum_{j=1}^h \lambda_j^2$. Hence, with $\bar{y}_n = \sum_1^n y_t/n$, the M-dependent CLT of Appendix A implies that

$$\sqrt{n}\bar{y}_n \to_d N\Big(0, \sum_{\ell=-\infty}^{\infty} \gamma_y(\ell)\Big) \equiv N\Big(0, \gamma_y(0)\Big) \equiv N\Big(0, \sigma^4 \sum_{j=1}^h \lambda_j^2\Big).$$

Thus

$$\sigma^{-2} n^{-1/2} \sum_{t=1}^n \sum_{j=1}^h \lambda_j x_t x_{t+j} \to_d N\Big(0, \sum_{j=1}^h \lambda_j^2\Big)$$

which, by the Cramer-Wold device, implies

$$\sigma^{-2} n^{-1/2} \sum_{t=1}^n (x_t x_{t+1}, \ldots, x_t x_{t+h})' \to_d (z_1, \ldots, z_h)'.$$

(c) This part follows from the proof of Problem 1.30, noting that $\mu_x = 0$.

(d) Using part (c), for large n,

$$\sqrt{n}\hat{\rho}(h) \sim \frac{\sqrt{n}\sum_{t=1}^n x_t x_{t+h}/n}{\sum_{t=1}^n x_t^2/n}.$$

Since the denominator $\to_p \sigma^2$, using Slutsky's Theorem,

$$\left(\frac{n^{-1/2}\sum_{t=1}^n x_t x_{t+h}/n}{\sigma^2} \frac{\sigma^2}{\sum_{t=1}^n x_t^2/n};\ j = 1, \ldots, h\right)' \to_d (z_1, \ldots, z_h)'.$$

Chapter 2

Solutions

2.1 (a)–(e) Detailed code is given in Appendix R. We reproduce it here with additional code that will produce all the necessary results. The model is overparameterized if an intercept is included (the terms for each Q are intercepts); most packages will kick out Q1 or Q4 (R kicks out Q1). The estimated annual increase is $\hat{\beta} = .167$. In general, $\alpha_i - \alpha_j$ is the average increase (decrease) from quarter i to quarter j. The question asks for $\hat{\alpha}_4 - \hat{\alpha}_3 = .88 - 1.15 = -.27$, indicating a decrease of about .27. There is substantial correlation left in the residuals, even at the yearly cycle.

```
trend = time(jj) - 1970   # helps to 'center' time
Q = factor(rep(1:4,21))   # make (Q)uarter factors
reg = lm(log(jj)~0 + trend + Q, na.action=NULL)  # without intercept
   Coefficients:
         Estimate Std. Error t value
   trend 0.167172   0.002259   74.00
   Q1    1.052793   0.027359   38.48
   Q2    1.080916   0.027365   39.50
   Q3    1.151024   0.027383   42.03
   Q4    0.882266   0.027412   32.19
   ---
   Residual standard error: 0.1254 on 79 degrees of freedom
   Multiple R-squared:  0.9935,    Adjusted R-squared:  0.9931
   F-statistic:  2407 on 5 and 79 DF,  p-value: < 2.2e-16
reg2 = lm(log(jj)~ trend + Q, na.action=NULL)    # with intercept
plot(log(jj), type="o")   # plot data and fit
lines(fitted(reg), col=2)
dev.new()
plot.ts(resid(reg))       # plot residuals
dev.new()
acf(resid(reg), 20)       # ACF of the resids
```

2.2 (a)–(b) The following code will produce the output. Note that P_{t-4} is significant in the regression and highly correlated (zero-order correlation is .52) with mortality. Note: P_t and P_{t-4} are also highly correlated.

The AIC and BIC are smaller than the values in Table 2.2, so the addition of the lagged particulate is an improvement.

```
temp = tempr-mean(tempr)
ded = ts.intersect(cmort, trend=time(cmort), temp, temp2=temp^2,  part,
    partL4=lag(part,-4))
summary(fit <- lm(cmort~trend + temp + temp2 + part + partL4, data=ded))
pairs(ded)      # easiest way is to do all of them
cor(ded)
AIC(fit)/nrow(ded) - log(2*pi)
 [1] 4.641492
```

```
BIC(fit)/nrow(ded) - log(2*pi)
 [1] 4.699677
```

2.3 The code is given below. The trend stationary data (y) are very regular and the true and estimated mean functions are typically close to each other. On the other hand, the random walk data (x) are very irregular, they're all over the place and the fitted line and true mean functions are rarely close.

```
par(mfrow=c(4,2), mar=c(2.5,2.5,0,0)+.5, mgp=c(1.6,.6,0))  # set up
 for (i in 1:4){
  x = ts(cumsum(rnorm(100,.01,1)))        # data
  y = ts(.01*1:100 + rnorm(100))
  regx = lm(x~0+time(x), na.action=NULL)  # regressions
  regy = lm(y~0+time(y), na.action=NULL)
  plot(x)                                 # plots
   abline(a=0, b=.01, col=2, lty=2)       # true mean
   abline(regx, col=4)                    # fitted line
  plot(y)
   abline(a=0, b=.01, col=2, lty=2)
   abline(regy, col=4)
 }
```

2.4 For the normal regression models we have $x_t \sim N(\boldsymbol{\beta}_j' \boldsymbol{z}_t, \sigma_j^2)$, for $j = 1, 2$. Then

$$\ln \frac{f_1(\boldsymbol{x}; \boldsymbol{\beta}_1, \sigma_1^2)}{f_2(\boldsymbol{x}; \boldsymbol{\beta}_2, \sigma_2^2)} = -\frac{n}{2} \ln \sigma_1^2 + \frac{n}{2} \ln \sigma_2^2 - \frac{1}{2\sigma_1^2} \sum_{t=1}^{n} (x_t - \boldsymbol{\beta}_1' \boldsymbol{z}_t)^2 + \frac{1}{2\sigma_2^2} \sum_{t=1}^{n} (x_t - \boldsymbol{\beta}_2' \boldsymbol{z}_t)^2$$

Taking expectations, the fourth term in the above becomes by adding and subtracting $\boldsymbol{\beta}_1' \boldsymbol{z}_t$ inside the parentheses

$$\mathrm{E}_1[(x_t - \boldsymbol{\beta}_2' \boldsymbol{z}_t)^2] = n\sigma_1^2 + (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)' Z' Z (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)$$

and, dividing through by n and collecting terms, we obtain the quoted result.

2.5 Using the quoted results and the independence of $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$, we have

$$\begin{aligned}
\mathrm{E}_1[I(\boldsymbol{\beta}, \sigma^2; \hat{\boldsymbol{\beta}}, \hat{\sigma}^2)] &= \frac{1}{2}\left[\mathrm{E}_1\left(\frac{n}{\chi^2_{n-k}}\right) - \ln \sigma_1^2 + \mathrm{E}_1 \ln \hat{\sigma}^2 + \mathrm{E}_1\left(\frac{\chi^2_k}{\chi^2_{n-k}}\right)\right] \\
&= \frac{1}{2}\left[\mathrm{E}_1\left(\frac{n}{\chi^2_{n-k}}\right) - \ln \sigma_1^2 + \mathrm{E}_1 \ln \hat{\sigma}^2 + \mathrm{E}_1(\chi^2_k)\mathrm{E}_1\left(\frac{1}{\chi^2_{n-k}}\right)\right] \\
&= -\frac{1}{2}\left[\ln \sigma_1^2 + \mathrm{E}_1 \ln \hat{\sigma}^2 + \frac{n}{n-k-2} + \frac{k}{n-k-2}\right],
\end{aligned}$$

which simplifies to the desired result.

2.6 (a) Since $\mathrm{E}x_t = \beta_0 + \beta_1 t$, the mean depends on t and hence the process in not stationary. (Note that the points will be randomly distributed around a straight line.)

(b) Note that $\nabla x_t = \beta_1 + w_t - w_{t-1}$ so that $\mathrm{E}(\nabla x_t) = \beta_1$ and

$$\mathrm{cov}(\nabla x_{t+h}, \nabla x_t) = \begin{cases} 2\sigma_w^2 & h = 0 \\ -\sigma_w^2 & h = \pm 1 \\ 0 & |h| > 1. \end{cases}$$

(c) Here $\nabla x_t = \beta_1 + y_t - y_{t-1}$, so $\mathrm{E}(\nabla x_t) = \beta_1 + \mu_y - \mu_y = \beta_1$. Also,

$$\mathrm{cov}(\nabla x_{t+h}, \nabla x_t) = \mathrm{cov}(y_{t+h} - y_{t+h-1}, y_t - y_{t-1}) = 2\gamma_y(h) - \gamma_y(h+1) - \gamma_y(h-1),$$

which is independent of t.

2.7 This is similar to part (c) of the previous problem except that now we have $\mathrm{E}(x_t - x_{t-1}) = \delta$, with autocovariance function

$$\text{cov}(w_{t+h} + y_{t+h} - y_{t+h-1}, w_t + y_t - y_{t-1}) = \gamma_w(h) + 2\gamma_y(h) - \gamma_y(h+1) - \gamma_y(h-1).$$

2.8 (a) The variance in the second half of the varve series is obviously larger than that in the first half. Dividing the data in half gives $\hat{\gamma}_x(0) = 133, 594$ for the first and second parts respectively and the variance is about 4.5 times as large in the second half. The transformed series $y_t = \ln x_t$ has $\hat{\gamma}_y(0) = .27, .45$ for the two halves, respectively and the variance of the second half is only about 1.7 times as large. Histograms, computed for the two series in Figure 2.1 indicate that the transformation improves the normal approximation.

```
varv1 = varve[1:317]
varv2 = varve[318:634]
var(varv1)       # = 133.4574
var(varv2)       # = 594.4904
var(log(varv1))  # = 0.2707217
var(log(varv2))  # = 0.451371
par(mfrow=c(1,2))
hist(varve)
hist(log(varve))
plot(log(varve))     # for part (b)
acf(log(varve))      # for part (c)
plot(diff(log(varve))) # for part (d)
acf(diff(log(varve)))  # for part (d)
```

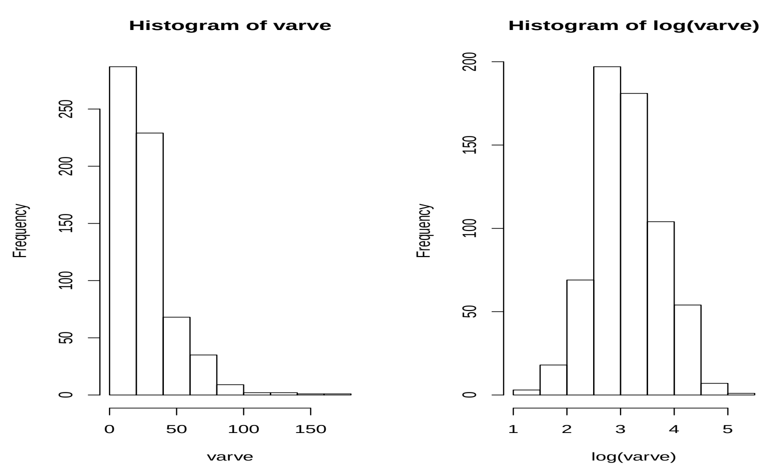

Fig. 2.1. Histograms for varve series x_t and $y_t = \ln x_t$.

(b) The data between 300 and 450 show a positive trend that is similar to the global temperature data. (Presumably, this is due to a difference in the Earth's rotation.)

(c) The ACF of the y_t is positive for a large number of lags and decreases in a linear fashion.

(d) The plot of u_t and its ACF seem to indicate stationarity. The ACF has one significant value at lag 1 (with a value of $-.3974$).

Because u_t can be written in the form

$$u_t = \log\left(\frac{x_t}{x_{t-1}}\right) = \log\left(1 + \frac{x_t - x_{t-1}}{x_{t-1}}\right) \approx \frac{x_t - x_{t-1}}{x_{t-1}},$$

it can be interpreted as the proportion of annual change.

(e)-(f) Note that

$$\gamma_u(0) = \mathrm{E}[u_t - \mu_u)^2] = \mathrm{E}[w_t^2] + \theta^2\mathrm{E}[w_{t-1}]^2 = \sigma_w^2(1+\theta^2)$$

and

$$\gamma_u(1) = \mathrm{E}[(w_{t+1} - \theta w_t)(w_t - \theta_{t-1})] = -\theta\mathrm{E}[w_t^2] = -\theta\sigma_w^2,$$

with $\gamma_u(h) = 0$ for $|h| > 1$. The ACF is

$$\rho(1) = \frac{\theta}{1+\theta^2}$$

or

$$\rho(1)\theta^2 - \theta + \rho(1) = 0$$

and we may solve for

$$\theta = \frac{1 \pm \sqrt{1 - 4\rho^2(1)}}{2\rho(1)}$$

using the quadratic formula. Hence, for $\hat{\rho}(1) = -.3974$

$$\hat{\theta} = \frac{1 \pm \sqrt{1 - 4(-.3974)^2}}{2(-.3974)},$$

yielding the roots $\hat{\theta} = -.4946, \ -2.0217$. We take the root $\theta = -.4946$ (this is the invertible root, see Chapter 3). Then,

$$\sigma_w^2 = \frac{\hat{\gamma}_u(0)}{1+\theta^2} = \frac{.3317}{1+(-.4946)^2} = .2665$$

2.9 The R code for this problem is below.

```
summary(fit <- lm(soi~time(soi), na.action=NULL)) # part (a)
                Estimate  Std. Error  t value  Pr(>t)
   (Intercept)  13.70367     3.18873    4.298  2.12e-05
   time(soi)    -0.00692     0.00162   -4.272  2.36e-05  <- significant slope
soi.d = resid(fit)   # part (b), detrended data in soi.d
plot(soi.d)
(length(soi.d))  # = 453
per = as.vector(abs(fft(soi.d))^2/453)  # as.vector removes the ts attributes
ord = 1:227
freq = (ord-1)/453
plot(freq, per[ord], type="l")  # graph
cbind(freq , per[ord])          # list
```

The El Niño peak is around .024 or approx 1 cycle/42 months (`freq = 0.02428` and local max `per = 1.22`). The annual peak is at `freq = 0.0838` with `per = 9.33`.

2.10 (a) `ts.plot(oil, gas, col=1:2)` will plot the data on the same graph. The data resemble the random walks shown in Figure 1.10. We showed in Chapter 1, Example 1.18, that random walks are not stationary [this is also covered in Problem 1.8]. See Figure 2.2. Use the following URL for more details on the data set or more concurrent series: `http://tonto.eia.doe.gov/dnav/pet/pet_pri_spt_s1_w.htm`.

(b) See the answer to 2.8(d).

(c) The code is below and follows the hints. The transformed series look stationary and there is very little autocorrelation left after transformation, so a random walk seems plausible for each series. There are some very extreme values [students are pointed to note the outliers in part (e)]. The two series seem to be moving at the same time [this is stressed in parts (d) and (e)].

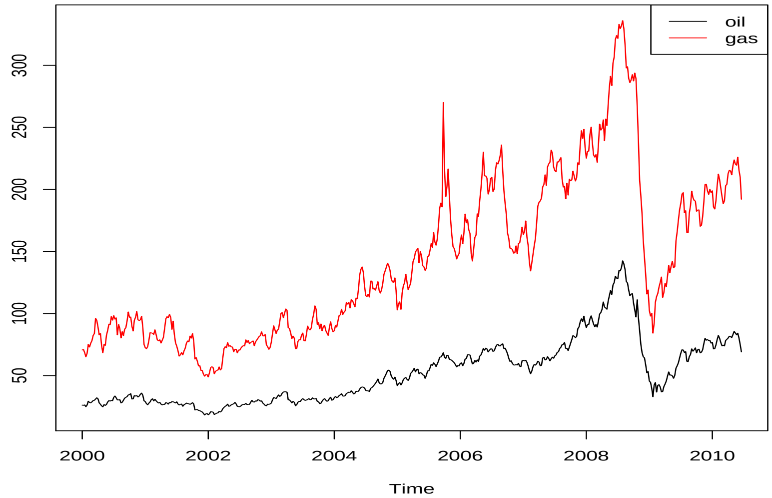

Fig. 2.2. Oil in dollars per barrel; gas in cents per gallon

```
poil = diff(log(oil))
pgas = diff(log(gas))
ts.plot(poil, pgas, col=1:2)
acf(poil)
acf(pgas)
```

(d) There is strong cross-correlation at the zero lag [.66] and `poil` one week ahead [.18]. There is also feedback for `pgas` three weeks ahead [.17]. As noted in the hint, `ccf(poil, pgas)` has oil leading for negative values of `Lag`; i.e., R computes `corr[poil(t+Lag), pgas(t)]` for `Lag = 0, ±1, ±2, ...`.

(e) Using either `astsa` or `tsa3.rda`, issuing `lag2.plot(poil, pgas, 3)` will give the plots. Aside from the aforementioned extreme outliers (10-30% changes in oil/gas prices in one week) the data seem fairly linear. Also, there may be some asymmetry noted in the lag 1 plot, but this appearance could be caused by outliers.

(f) (i) The hint shows how to set up the regression. Note that the interactions are not significant if you include them. The results are

```
             Estimate Std. Error t value Pr(>t)
(Intercept) -0.006445   0.003464  -1.860  0.06338 .
poil         0.683127   0.058369  11.704  < 2e-16
poilL        0.111927   0.038554   2.903  0.00385
indi         0.012368   0.005516   2.242  0.02534

Residual standard error: 0.04169 on 539 degrees of freedom
Multiple R-squared: 0.4563,     Adjusted R-squared: 0.4532
F-statistic: 150.8 on 3 and 539 DF,  p-value: < 2.2e-16
```

(ii) Note $-0.006445 + 0.012368 = .006$. Hence, the two models are

$$\hat{G}_t = -.006 + .68 O_t + .11 O_{t-1} \quad \text{if } O_t < 0$$

$$\hat{G}_t = \ \ .006 + .68 O_t + .11 O_{t-1} \quad \text{if } O_t \geq 0.$$

which suggests that there is no asymmetry for these data. The prices are FOB (*Free On Board: A sales transaction in which the seller makes the product available for pick up at a specified port or terminal at a specified price and the buyer pays for the subsequent transportation and insurance.*) so the asymmetry might be found in the gas pump prices.

(iii) `plot.ts(resid(fit))` and `acf(resid(fit))` seem to suggest the fit is not bad except for the very extreme outliers.

2.11 These are interesting (see Figure 2.3). Students should realize that we're all going to die.

```
plot(globtemp)
lines(lowess(globtemp),col=2)
lines(smooth.spline(globtemp, spar=1),col=4)
```

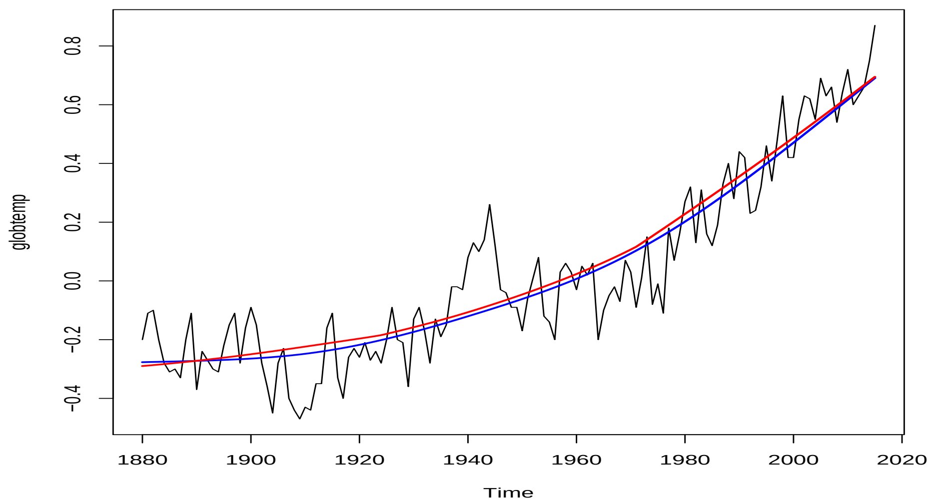

Fig. 2.3. lowess is red; smoothing spline is blue

Chapter 3

Solutions

3.1 Note $\rho_x(1) = \frac{\theta}{1+\theta^2}$ and is bounded. Thus $\frac{\partial \rho_x(1)}{\partial \theta} = \frac{1-\theta^2}{(1+\theta^2)^2} = 0$ when $\theta = \pm 1$. We conclude $\rho_x(1)$ has a maximum at $\theta = 1$ wherein $\rho_x(1) = 1/2$ and a minimum at $\theta = -1$ wherein $\rho_x(1) = -1/2$.

3.2 (a) Use induction or insert the solution into the equation, i.e.,

$$\sum_{j=0}^{t} \phi^j w_{t-j} = \phi\Big(\sum_{j=0}^{t-1} \phi^j w_{t-1-j}\Big) + w_t \,.$$

(b) $\mathrm{E}(x_t) = \sum_{j=0}^{t} \phi^j \mathrm{E}(w_{t-j}) = 0$

(c) $\mathrm{var}(x_t) = \sum_{j=0}^{t} \phi^{2j}\mathrm{var}(w_{t-j}) = \sigma_w^2 \sum_{j=0}^{t} \phi^{2j} = \frac{\sigma_w^2}{1-\phi^2}(1-\phi^{2(t+1)})$, using the fact that w_t is noise.

(d) Notice that

$$x_{t+h} = \sum_{j=0}^{t+h} \phi^j w_{t+h-j} = \sum_{j=0}^{h-1} \phi^j w_{t+h-j} + \sum_{j=h}^{t+h} \phi^j w_{t+h-j} = \sum_{j=0}^{h-1} \phi^j w_{t+h-j} + \phi^h \sum_{k=0}^{t} \phi^k w_{t-k} = \sum_{j=0}^{h-1} \phi^j w_{t+h-j} + \phi^h x_t.$$

Alternately, just iterate x_{t+h} back h time units. For example, below (3.5) we iterated back k units of time; so in that result put $t \mapsto t+h$ and $k \mapsto h$ to get

$$x_{t+h} = \phi^h x_t + \sum_{j=0}^{h-1} \phi^j w_{t+h-j} \,,$$

immediately.
Since x_t involves the w_s for $s \le t$,

$$\mathrm{cov}(x_{t+h}, x_t) = \mathrm{cov}\Big(\sum_{j=0}^{h-1} \phi^j w_{t+h-j} + \phi^h x_t,\ x_t\Big) = \phi^h \mathrm{var}(x_t).$$

(e) x_t is NOT stationary because the (co)variance depends on time t.

(f) As $t \to \infty$, $\mathrm{var}(x_t) \to \sigma_w^2/(1-\phi^2)$ by (c), and hence by (d), the autocovariance function is independent of t.

(g) Generate more than n observations, for example, generate $n + n_0$ observations, where n_0 is fairly large (like 50), and discard the first n_0.

(h) Write $x_t = \phi^t w_0 + \sum_{j=0}^{t-1} \phi^j w_{t-j}$. Now, $\mathrm{var}(x_t) = \frac{\phi^{2t}}{1-\phi^2}\sigma_w^2 + \sigma_w^2 \sum_{j=0}^{t-1} \phi^{2j} = \sigma_w^2 \Big[\frac{\phi^{2t}}{1-\phi^2} + \frac{1-\phi^{2t}}{1-\phi^2}\Big] = \frac{\sigma_w^2}{1-\phi^2}$ independent of t.

3.3 (a) Most of the work is done in the example. For ease, write $\theta = \phi^{-1}$. Then

$$\begin{aligned}\gamma_x(h) &= \text{cov}\left\{(\theta w_{t+1} + \cdots + \theta^{h+1} w_{t+h+1} + \theta^{h+2} w_{t+h+2} + \cdots), (\theta w_{t+h+1} + \theta^2 w_{t+h+2} + \cdots)\right\} \\ &= \sigma_w^2(\theta^{h+2} + \theta^{h+4} + \cdots) = \sigma_w^2 \theta^h [\theta^2/(1-\theta^2)].\end{aligned}$$

(b) This part follows from the discussion of the causal model. Here we have $y_t = \theta y_{t-1} + v_t$, so $y_t = \sum_{j=0}^{\infty} \theta^j v_{t-j}$. Noting (3.6), (3.7) and that $\sigma_v^2 = \sigma_w^2 \theta^2$ finishes the verification.

3.4 (a) Write this as $(1 - .3B)(1 - .5B)x_t = (1 - .3B)w_t$ and reduce to $(1 - .5B)x_t = w_t$. Hence the process is a causal and invertible AR(1): $x_t = .5x_{t-1} + w_t$.

(b) The AR polynomial is $1 - 1z + .5z^2$ which has complex roots $1 \pm i$ outside the unit circle (note $|1 \pm i|^2 = 2$). The MA polynomial is $1 - z$ which has root unity. Thus the process is a causal but not invertible ARMA$(2, 1)$.

3.5 Let ξ_1 and ξ_2 be the roots of $\phi(z)$, that is, $\phi(z) = (1 - \xi_1^{-1} z)(1 - \xi_2^{-1} z)$. The causal condition is $|\xi_1| > 1,\ |\xi_2| > 1$. Let $u_1 = \xi_1^{-1}$ and $u_2 = \xi_2^{-1}$ so that $\phi(z) = (1 - u_1 z)(1 - u_2 z)$ with causal condition $|u_1| < 1,\ |u_2| < 1$. To show $|u_1| < 1,\ |u_2| < 1$ if and only if the three given inequalities hold. In terms of u_1 and u_2, the inequalities are:

(i) $\phi_2 + \phi_1 - 1 = -(1 - u_1)(1 - u_2) < 0$ (note $\phi_1 = u_1 + u_2$ and $\phi_2 = -u_1 u_2$)
(ii) $\phi_2 - \phi_1 - 1 = -(1 + u_1)(1 + u_2) < 0$
(iii) $|\phi_2| = |u_1 u_2| < 1$

[IF] If (i)–(iii) hold, then (iii), which is $|u_1 u_2| < 1$, implies at least one of u_1, u_2 must be less than 1 in absolute value (both if they are complex). Thus, (iii) is enough to imply $|u_1| < 1,\ |u_2| < 1$ in the case of complex roots.

Now suppose the roots are real. Suppose WOLOG, $|u_1| < 1$. But, if $|u_1| < 1$, then $(1 \pm u_1) > 0$ so for (i) and (ii) to hold, we must have $(1 \pm u_2) > 0$ or $|u_2| < 1$ as desired.

[ONLY IF] If $|u_1| < 1,\ |u_2| < 1$ and they are real, then (i) and (ii) hold because $(1 \pm u_j) > 0$ for $j = 1, 2$; (iii) is obvious.

If $|u_1| < 1,\ |u_2| < 1$ and they are complex, $u_2 = \overline{u}_1$ and (i) $-|1 - u_1|^2 < 0$, (ii) $-|1 + u_1|^2 < 0$, (iii) $|u_1|^2 < 1$. $\square$

3.6 The roots of $\phi(z) = 1 + .9z^2$ are $\pm i/\sqrt{.9}$. Because the roots are purely imaginary, $\theta = \arg(i/\sqrt{.9}) = \pi/2$ and consequently, $\rho(h) = a\sqrt{.9}^h \cos(\frac{\pi}{2}h + b)$, or $\rho(h)$ makes one cycle every 4 values of h. Because $\rho(0) = 1$ and $\rho(1) = \phi_1/(1 - \phi_2) = 0$, it follows that $a = 1$ and $b = 0$ in which case $\rho(h) = \sqrt{.9}^h \cos(\frac{\pi}{2}h)$. Thus $\rho(h) = \{1, 0, -\sqrt{.9}, 0, \sqrt{.9}^5, \ldots\}$ for $h = 0, 1, 2, 3, 4, \ldots$.

```
u = ARMAacf(ar=c(0,-.9), lag.max=25)
plot(0:25, u, type="h", xlab="Lag", ylab="ACF")
points(0:25, u)
abline(h=0)
```

3.7 These problems can be tedious, so we suggest not assigning all three parts. Also, students can check their answers in R using `ARMAacf(ar=c(_,_), lag.max=10)`; fill in the blanks. For the ACFs we have $\rho(0) = 1$, $\rho(1) = \phi_1/(1 - \phi_2)$ and

distinct roots: $\rho(h) = c_1 z_1^{-h} + c_2 z_2^{-h}$ equal roots: $\rho(h) = z_0^{-h}(c_1 + c_2 h)$

(a) $\phi(z) = 1 + 1.6z + .64z^2 = (1 + .8z)^2$. This is equal roots case with $z_0 = -1/.8$. Thus $\rho(h) = -.8^h(a + bh)$. To solve for a and b, note for $h = 0$ we have $\rho(0) = 1 = a$ and for $h = 1$ we have $\rho(1) = -1.6/(1 + .64) = -.8(1 + b)$ or $b = .22$. Finally, $\rho(h) = -.8^h(1 + .22h)$ for $h = 0, 1, 2, \ldots$.

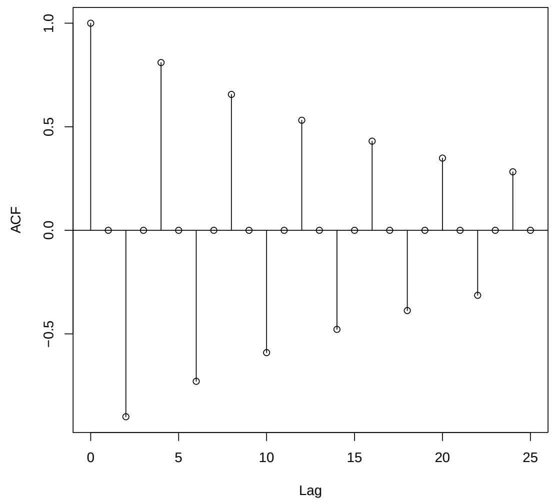

Fig. 3.1. ACF for Problem 3.6

(b) $\phi(z) = 1 - .4z - .45z^2 = (1 - .9z)(1 + .5z)$. This is the unequal roots case with $z_1 = 1/.9$ and $z_2 = -1/.5$. For the ACF, $\rho(h) = a.9^h - b.5^h$ where a and b are found by solving $1 = a + b$ and $.4/(1 - .45) = .9a - .5b$ or $a = .88$ and $b = .12$.

(c) $\phi(z) = 1 - 1.2z + .85z^2$. This is the complex roots case, with inverse roots $.6 \pm .7i$. `Arg(.6+.7i)` and `Mod(.6+.7i)` give $\theta = \arg(.6+.7i) = .86$ radians and $|.6+.7i| = .92$. Thus, $\rho(h) = a\,.92^h \cos(.86h+b)$ where a and b are found by solving $1 = a\cos(b)$ $[h = 0]$ and $1.2/(1+.85) = a\,.92\cos(.86+b)$ $[h = 1]$. Solving, $b = \pi$ and $a = -1$.

3.8 There's not much to do in verifying the calculations... almost everything is done in the example. The ACF distinguishes the MA(1) case but not the ARMA(1,1) or AR(1) cases, which look similar to each other (see Figure 3.2).

```
u1 = ARMAacf(ar=.6, ma=.9, lag.max=10)
u2 = ARMAacf(ar=.6, ma=0, lag.max=10)
u3 = ARMAacf(ar=0,ma=.9, lag.max=10)
plot(u1[-1], type="o", col=1, ylab="ACF", xlab="Lag")
lines(u2[-1], type="o", col=2)
lines(u3[-1], type="o", col=4)
legend("topright", c("ARMA","AR","MA"), col=c(1,2,4), lty=1)
```

3.9 R code:

```
ar = arima.sim(list(order=c(1,0,0), ar=.6), n=100)
ma = arima.sim(list(order=c(0,0,1), ma=.9), n=100)
arma = arima.sim(list(order=c(1,0,1), ar=.6, ma=.9), n=100)
acf2(ar)
acf2(ma)
acf2(arma)
```

The results should be close to Table 3.1 with randomness taken into consideration.

3.10 R code and output:

```
(reg = ar.ols(cmort, order=2, demean=FALSE, intercept=TRUE))
  Coefficients
       1       2
```

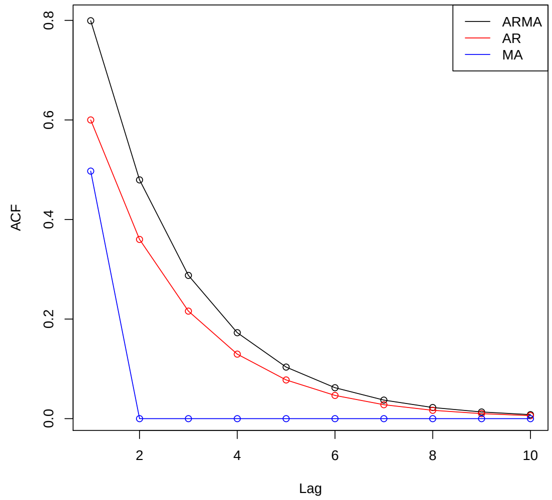

Fig. 3.2. ACFs for Problem 3.8

```
  0.4286  0.4418
  Intercept: 11.45 (2.394)
  Order selected 2  sigma^2 estimated as  32.32
predict(reg, n.ahead=4)
 $pred
  Time Series:
  Start = c(1979, 41)
  End = c(1979, 44)
  Frequency = 52
  [1] 87.59986 86.76349 87.33714 87.21350
 $se
  Time Series:
  Start = c(1979, 41)
  End = c(1979, 44)
  Frequency = 52
  [1] 5.684848 6.184973 7.134227 7.593357
```

3.11 (a) The model can be written as $x_{n+1} = \sum_{j=1}^{\infty}(-\theta)^j x_{n+1-j} + w_{n+1}$. From this we conclude that $\tilde{x}_{n+1} = \sum_{j=1}^{\infty}(-\theta)^j x_{n+1-j}$ and MSE $= \mathrm{E}(x_{n+1} - \tilde{x}_{n+1})^2 = \mathrm{E}w_{n+1}^2 = \sigma_w^2$.

(b) Truncating, we have $\widetilde{x}_{n+1}^n = \sum_{j=1}^{n}(-\theta)^j x_{n+1-j}$. Thus

$$\begin{aligned}
\text{MSE} = \mathrm{E}(x_{n+1} - \widetilde{x}_{n+1}^n)^2 &= \mathrm{E}\left[\sum_{j=n+1}^{\infty}(-\theta)^j x_{n+1-j} + w_{n+1}\right]^2 \\
&= \mathrm{E}\left[(-\theta)^{(n+1)}\sum_{j=n+1}^{\infty}(-\theta)^{j-(n+1)} x_{n+1-j} + w_{n+1}\right]^2 \\
&= \mathrm{E}\left[(-\theta)^{(n+1)} w_0 + w_{n+1}\right]^2 = \sigma_w^2\left(1 + \theta^{2(n+1)}\right).
\end{aligned}$$

There can be a substantial difference between the two MSEs for small values of n, but for large n the difference is negligible.

3.12 The proof is by contradiction. Assume there is a Γ_n that is singular. Because $\gamma(0) > 0$, $\Gamma_1 = \{\gamma(0)\}$ is non-singular. Thus, there is an $r \ge 1$ such that Γ_r is non-singular. Consider the ordered sequence $\Gamma_1, \Gamma_2, \ldots$ and suppose Γ_{r+1} is the first singular Γ_n in the sequence. Then x_{r+1} is a linear combination of $x = (x_1, \ldots, x_r)'$, say, $x_{r+1} = b'x$ where $b = (b_1, ..., b_r)'$. Because of stationarity, it must also be true that $x_{r+h+1} = b'x_h$, where $x_h = (x_h, \ldots, x_{r+h-1})'$ for all $h \ge 1$. This means that for any $n \ge r+1$, x_n is a linear combination of $x_1, \ldots, x_r$, i.e., $x_n = b_n'x$ where $b_n = (b_{n1}, ..., b_{nr})'$. Thus, $\gamma(0) = \text{var}(x_n) = b_n'\Gamma_r b_n = b_n' Q \Lambda Q' b_n$ where QQ' is the identity matrix and $\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_r\}$ is the diagonal matrix of the positive eigenvalues $(0 < \lambda_1 \le \cdots \le \lambda_r)$ of Γ_r. From this result we conclude

$$\gamma(0) \ge \lambda_1 b_n' QQ' b_n = \lambda_1 \sum_{j=1}^{r} b_{nj}^2;$$

this shows that for each j, b_{nj} is bounded in n. In addition, $\gamma(0) = \text{cov}(x_n, x_n) = \text{cov}(x_n, b_n'x)$ from which it follows that

$$0 < \gamma(0) \le \sum_{j=1}^{r} |b_{nj}|\, |\gamma(n-j)|.$$

From this inequality it is seen that because the b_{nj} are bounded, it is not possible to have $\gamma(0) > 0$ and $\gamma(h) \to 0$ as $h \to \infty$.

3.13 First take the prediction equations (3.63) with $n = h$ and divide both sides by $\gamma(0)$ to obtain $R_h \phi_h = \rho_h$. Partition the equation as in the hint with $\phi_h = (\phi_{h-1}', \phi_{hh})'$ [note $\rho(0) = 1$]:

$$\begin{pmatrix} R_{h-1} & \tilde\rho_{h-1} \\ \tilde\rho_{h-1}' & 1 \end{pmatrix} \begin{pmatrix} \phi_{h-1} \\ \phi_{hh} \end{pmatrix} = \begin{pmatrix} \rho_{h-1} \\ \rho(h) \end{pmatrix},$$

and solve. We get

$$R_{h-1}\phi_{h-1} + \tilde\rho_{h-1}\phi_{hh} = \rho_{h-1} \tag{1}$$

$$\tilde\rho_{h-1}'\phi_{h-1} + \phi_{hh} = \rho(h). \tag{2}$$

Solve equation (1) for ϕ_{h-1} to obtain

$$\phi_{h-1} = R_{h-1}^{-1}\left(\rho_{h-1} - \tilde\rho_{h-1}\phi_{hh}\right).$$

Substitute this into equation (2) and solve for ϕ_{hh}:

$$\phi_{hh} = \frac{\rho(h) - \tilde\rho_{h-1}' R_{h-1}^{-1}\rho_{h-1}}{1 - \tilde\rho_{h-1}' R_{h-1}^{-1}\tilde\rho_{h-1}}. \tag{3}$$

Next, we must show that the PACF,

$$\frac{\text{E}(\epsilon_t \delta_{t-h})}{\sqrt{\text{E}(\epsilon_t^2)\text{E}(\delta_{t-h}^2)}},$$

can be written in the form of equation (3). To this end, let $x = (x_{t-1}, ..., x_{t-h+1})'$. The regression of x_t on x is $(\Gamma_{h-1}^{-1}\gamma_{h-1})'x$. The regression of x_{t-h} on x is $(\Gamma_{h-1}^{-1}\tilde\gamma_{h-1})'x$. Thus

$$\epsilon_t = x_t - \gamma_{h-1}'\Gamma_{h-1}^{-1}x$$

$$\delta_{t-h} = x_{t-h} - \tilde\gamma_{h-1}'\Gamma_{h-1}^{-1}x.$$

From this we calculate [the calculations below are all similar to the verification of equation (3.66); also, note for vectors a and b, $a'b = b'a$]

$$\text{E}(\epsilon_t\delta_{t-h}) = \text{cov}(\epsilon_t, \delta_{t-h}) = \gamma(h) - \tilde\gamma_{h-1}'\Gamma_{h-1}^{-1}\gamma_{h-1}.$$

Similar calculations show that

$$\mathrm{E}(\delta_{t-h}^2) = \mathrm{var}(\delta_{t-h}) = \gamma(0) - \tilde{\gamma}_{h-1}' \Gamma_{h-1}^{-1} \tilde{\gamma}_{h-1}.$$

Also note that the error of the regression of x_t on x is the same as the error of the regression of x_t on $\tilde{x}$, where $\tilde{x} = (x_{t-h+1}, ..., x_{t-1})'$; that is, $\epsilon_t = x_t - (\Gamma_{h-1}^{-1}\tilde{\gamma}_{h-1})'\tilde{x}$. From this we conclude that

$$\mathrm{E}(\epsilon_t^2) = \mathrm{var}(\epsilon_t) = \gamma(0) - \gamma_{h-1}' \Gamma_{h-1}^{-1} \gamma_{h-1} = \gamma(0) - \tilde{\gamma}_{h-1}' \Gamma_{h-1}^{-1} \tilde{\gamma}_{h-1}.$$

This proves the result upon factoring out $\gamma(0)$ in the numerator and denominator.

3.14 (a) We want to find $g(x)$ to minimize $\mathrm{E}[y - g(x)]^2$. Write this as $\mathrm{E}[\mathrm{E}\{(y - g(x))^2 \mid x\}]$. Minimize the inner expectation: $\partial \mathrm{E}\{(y - g(x))^2 \mid x\}/\partial g(x) = 2[\mathrm{E}(y|x) - g(x)] = 0$ from which we conclude $g(x) = \mathrm{E}(y|x)$ is the required minimum.

(b) $g(x) = \mathrm{E}(y|x) = \mathrm{E}(x^2 + z|x) = x^2 + \mathrm{E}(z) = x^2$. MSE $= \mathrm{E}(y - g(x))^2 = \mathrm{E}(y - x^2)^2 = \mathrm{E}(z^2) = \mathrm{var}(z) = 1$.

(c) Let $g(x) = a + bx$. Using the prediction equations, $g(x)$ satisfies

$$\text{(i) } \mathrm{E}[y - g(x)] = 0 \quad \text{(ii) } \mathrm{E}[(y - g(x))x] = 0$$

or

$$\text{(i) } \mathrm{E}[y] = \mathrm{E}[a + bx] \quad \text{(ii) } \mathrm{E}(xy) = \mathrm{E}[(a + bx)x]$$

From (i) we have $a + b\mathrm{E}(x) = \mathrm{E}(y)$, but $\mathrm{E}(x) = 0$ and $\mathrm{E}(y) = 1$ so $a = 1$. From (ii) we have $a\mathrm{E}(x) + b\mathrm{E}(x^2) = \mathrm{E}(xy)$, or $b = \mathrm{E}[x(x^2 + z)] = \mathrm{E}(x^3) + \mathrm{E}(xz) = 0 + 0$. Finally $g(x) = a + bx = 1$ and MSE $= \mathrm{E}(y - 1)^2 = \mathrm{E}(y^2) - 1 = \mathrm{E}(x^4) + \mathrm{E}(z^2) - 1 = 3 + 1 - 1 = 3$.
Conclusion: In this case, the best linear predictor has three times the error of the optimal predictor (conditional expectation).

3.15 For an AR(1), equation (3.86) is exact; that is, $\mathrm{E}(x_{t+m} - x_{t+m}^t)^2 = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2$. For an AR(1), $\psi_j = \phi^j$ and thus $\sigma_w^2 \sum_{j=0}^{m-1} \phi^{2j} = \sigma_w^2(1 - \phi^{2m})/(1 - \phi^2)$, the desired expression.

3.16 From Example 3.7, $x_t = 1.4 \sum_{j=1}^{\infty} (-.5)^{j-1} x_{t-j} + w_t$, so the truncated one-step-ahead prediction using (3.81) is $\tilde{x}_{n+1}^n = 1.4 \sum_{j=1}^{n} (-.5)^{j-1} x_{n+1-j}$.
From Property 3.7,

$$\begin{aligned}
\tilde{x}_{n+1}^n &= .9x_n + .5\tilde{w}_n^n = .9x_n + .5(x_n - .9x_{n-1} - .5\tilde{w}_{n-1}^n) \\
&= 1.4x_n - .9(.5)x_{n-1} - .5\tilde{w}_{n-1}^n = 1.4x_n - .9(.5)x_{n-1} - .5^2(x_{n-1} - .9x_{n-2} - .5\tilde{w}_{n-2}^n) \\
&= 1.4x_n - 1.4(.5)x_{n-1} + .9(.5^2)x_{n-2} + .5^3\tilde{w}_{n-2}^n \\
&= 1.4x_n - 1.4(.5)x_{n-1} + 1.4(.5^2)x_{n-2} - .9(.5^3)x_{n-3} - .5^4\tilde{w}_{n-3}^n \\
&\;\;\vdots \\
&= 1.4 \sum_{j=1}^{n} (-.5)^{j-1} x_{n+1-j}
\end{aligned}$$

3.17 $\mathrm{E}(x_{n+m} - \tilde{x}_{n+m})(x_{n+m+k} - \tilde{x}_{n+m+k}) = \mathrm{E}\Big(\sum_{j=0}^{m-1} \psi_j w_{n+m-j}\Big)\Big(\sum_{\ell=0}^{m+k-1} \psi_\ell w_{n+m+k-\ell}\Big)$

$$= \mathrm{E}\Big(\sum_{j=0}^{m-1} \psi_j w_{n+m-j}\Big)\Big(\sum_{\ell=k}^{m+k-1} \psi_\ell w_{n+m+k-\ell}\Big) = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j \psi_{j+k}$$

3.18 (a)–(b) Below `reg1` is least squares and `reg2` is Yule-Walker. The standard errors for each case are also evaluated; the Yule-Walker run uses Property 3.8. The two methods produce similar results.

```
(reg1 = ar.ols(cmort, order=2))  #  coefs: [1] 0.4286 [2] 0.4418;  sigma^2 estimated as
    32.32
(reg2 = ar.yw(cmort, order=2))   #  coefs: [1] 0.4339 [2] 0.4376 ; sigma^2 estimated as
    32.84
(reg1$asy.se.coef)               #     se: [1] 0.0397 [2] 0.0397
(sqrt(diag(reg2$asy.var.coef)))  #     se: [1] 0.0400 [2] 0.0400
```

3.19 (a) For an AR(1) we have, $x_1^n = x_1$, $x_0^n = \phi x_1$, $x_{-1}^n = \phi x_0^n = \phi^2 x_1$, and in general, $x_t^n = \phi^{1-t} x_1$ for $t = 1, 0, -1, -2, \ldots$.

(b) $\tilde{w}_t(\phi) \stackrel{\text{def}}{=} x_t^n - \phi x_{t-1}^n = \phi^{1-t} x_1 - \phi \phi^{2-t} x_1 = \phi^{1-t}(1-\phi^2) x_1$.

(c) $\sum_{t=-\infty}^{1} \tilde{w}_t^2(\phi) = (1-\phi^2)^2 x_1^2 \sum_{t=-\infty}^{1} \phi^{2(1-t)} = (1-\phi^2)^2 x_1^2 \sum_{t=1}^{\infty} \phi^{2t-2} = (1-\phi^2)^2 x_1^2 \frac{1}{1-\phi^2} = (1-\phi^2) x_1^2$.

(d) From (3.107), $S(\phi) = (1-\phi^2)x_1^2 + \sum_{t=2}^{n}(x_t - \phi x_{t-1})^2 = \sum_{t=-\infty}^{1} \tilde{w}_t^2(\phi) + \sum_{t=2}^{n}(x_t - \phi x_{t-1})^2 = \sum_{t=-\infty}^{n} \tilde{w}_t^2(\phi)$ using (c) and the fact that $\tilde{w}_t(\phi) = x_t - \phi x_{t-1}$ for $1 \le t \le n$.

(e) For $t = 2, \ldots, n$, $x_t^{t-1} = \phi x_{t-1}$ and $x_t - x_t^{t-1} = x_t - \phi x_{t-1}$. For $t = 1$, $x_1^0 = \mathrm{E}(x_1) = 0$ so $x_1 - x_1^0 = x_1$. Also, for $t = 2, \ldots, n$, $P_t^{t-1} = \mathrm{E}(x_t - \phi x_{t-1})^2 = \mathrm{E}(w_t^2) = \sigma_w^2$ so $r_t^{t-1} = 1$. For $t = 1$, $P_1^0 = \mathrm{E}(x_1)^2 = \sigma_w^2/(1-\phi^2)$ so $r_1^0 = 1/(1-\phi^2)$ we may write $S(\phi)$ in the desired form.

3.20 Although the results will vary, the data should behave like observations from a white noise process and each run should yield different parameter estimates where one is approximately the negative of the other. Students may also get warnings such as non-convergence or the SEs being `NaN`; it may also happen that the `ar` estimate is negative one.

```
x = arima.sim(list(order=c(1,0,1), ar=.9, ma=-.9), n=500)
plot(x)
acf2(x)
sarima(x, 1, 0 ,1)
```

3.21 The following R program can be used. The estimates are close to the actual values.

```
rep(NA, 10) -> phi -> theta -> sigma2
for (i in 1:10){
 x = arima.sim(n = 200, list(ar = .9, ma = .5))
 fit = arima(x, order=c(1,0,1))
 phi[i] = fit$coef[1]; theta[i] = fit$coef[2]; sigma2[i] = fit$sigma2
 }
cbind("phi"=phi, "theta"=theta, "sigma2"=sigma2)
```

3.22 Below is R code for this example using Yule-Walker. The asymptotic distribution is normal with mean .99 and standard error $\sqrt{(1-.99^2)/50} \approx .02$. The bootstrap distribution should be very different than the asymptotic distribution. (If students use MLE, there might be problems because ϕ is very near the boundary. You might alert students to this fact or let them find out on their own.)

```
x = arima.sim(list(ar=.99), n=50)
fit = ar.yw(x, order=1)
phi = fit$ar              # estimate of phi
nboot = 200               # number of bootstrap replicates
resids = fit$resid[-1]    # the first resid is NA
x.star = x                # x[1] stays the same throughout
phi.star = rep(NA, nboot)
for (i in 1:nboot) {
 resid.star = sample(resids, replace=TRUE)
  for (t in 1:49) x.star[t+1] =  phi*x.star[t] + resid.star[t]
  phi.star[i] = ar.yw(x.star, order=1)$ar
  }
summary(phi.star);  hist(phi.star);  stem(phi.star)  # and so on ...
```

3.23 Write $w_t(\phi) = x_t - \phi x_{t-1}$ for $t = 1, ..., n$ conditional on $x_0 = 0$. Then $z_t(\phi) = -\partial w_t(\phi)/\partial\phi = x_{t-1}$. Let $\phi_{(0)}$ be an initial guess at ϕ, then

$$\phi_{(1)} = \phi_{(0)} + \frac{\sum_{t=1}^n x_{t-1}(x_t - \phi_{(0)}x_{t-1})}{\sum_{t=1}^n x_{t-1}^2} = \frac{\sum_{t=1}^n x_{t-1}x_t}{\sum_{t=1}^n x_{t-1}^2},$$

and the estimate has converged in one step to the (conditional) SS estimate of ϕ.

3.24 (a) Taking expectation through the model, $\mathrm{E}(x_t) = \alpha + \phi\mathrm{E}(x_{t-1}) + 0 + \theta 0$, we have $\mu = \alpha + \phi\mu$ or $\alpha = \mu(1-\phi)$. Let $y_t = x_t - \mu$, then the model can be written as $y_t = \phi y_{t-1} + w_t + \theta w_{t-1}$. Because $|\phi| < 1$ the process y_t is causal (and hence stationary) and consequently x_t is stationary. The same technique used in Problem 1.17 can be used here to show that y_t, and hence x_t, is strictly stationary.

(b) Because of causality, $y_t = \sum_{j=0}^\infty \psi_j w_{t-j}$ where $\psi_0 = 1$ and $\psi_j = (\phi+\theta)\phi^{j-1}$ for $j = 1, 2, \dots$ (see Example 3.11) and hence $x_t = \mu + \sum_{j=0}^\infty \psi_j w_{t-j}$. Thus, by Theorem A.5, $\bar{x} \sim \mathrm{AN}(\mu, n^{-1}V)$ where $V = \sigma_w^2 \left(\sum_{j=0}^\infty \psi_j\right)^2 = \sigma_w^2 \left(1 + (\phi+\theta)\sum_{j=1}^\infty \phi^{j-1}\right)^2$. Equivalently, $\bar{x} \sim \mathrm{AN}(\alpha/(1-\phi), n^{-1}V)$.

3.25 (a) $\mathrm{E}(x_t) = 0$ and $\gamma_x(h) = \mathrm{E}(s_t + a s_{t-\delta})(s_{t+h} + a s_{t+h-\delta}) = \begin{cases} (1+a^2)\sigma_s^2, & h = 0 \\ a\sigma_s^2, & h = \pm\delta \\ 0, & |h| > 1, \end{cases}$ so the process is stationary.

Also, $x_t - a x_{t-\delta} + a^2 x_{t-2\delta} - \cdots + (-1)^k a^k x_{t-k\delta} = s_t - (-1)^{k+1}a^{k+1}s_{t-k\delta}$, and letting $k \to \infty$ shows $s_t = \sum_{j=0}^\infty (-a)^j x_{t-\delta j}$ is the mean square convergent representation of s_t. Note: If δ is known, the process is an invertible MA(δ) process with $\theta_1 = \cdots = \theta_{\delta-1} = 0$ and $\theta_\delta = a$.

(b) The Gauss-Newton procedure is similar to the MA(1) case in Example 3.30. Write $s_t(a) = x_t - a s_{t-\delta}(a)$ for $t = 1, ..., n$. Then $z_t(a) = -\partial s_t(a)/\partial a = s_{t-\delta}(a) + a\partial s_{t-\delta}/\partial a = s_{t-\delta}(a) - a z_{t-\delta}(a)$. The iterative procedure is

$$a_{(j+1)} = a_{(j)} + \frac{\sum_{t=1}^n z_t(a_{(j)})s_t(a_{(j)})}{\sum_{t=1}^n z_t^2(a_{(j)})} \quad j = 0, 1, 2, \dots$$

where $z_t(\cdot) = 0$ and $s_t(\cdot) = 0$ for $t \le 0$.

(c) If δ is unknown, the ACF of x_t can be used to find a preliminary estimate of δ. Then, a Gauss-Newton procedure can be used to minimize the error sum of squares, say $S_c(a, \delta)$, over a grid of δ values near the preliminary estimate. The values $\hat{a}$ and $\hat{\delta}$ that minimize $S_c(a, \delta)$ are the required estimates.

3.26 (a) By Property 3.9, $\hat{\phi} \sim \mathrm{AN}[\phi, n^{-1}(1-\phi^2)]$ so that $\hat{\phi} = \phi + O_p(n^{-1/2})$.

(b) $x_{n+1}^n = \phi x_n$ whereas $\hat{x}_{n+1}^n = \hat{\phi}x_n$. Thus $x_{n+1}^n - \hat{x}_{n+1}^n = (\phi - \hat{\phi})x_n$. Using Tchebcheff's inequality, it is easy to show $x_n = O_p(1)$. Thus, by the properties of $O_p(\cdot)$,

$$x_{n+1}^n - \hat{x}_{n+1}^n = (\phi - \hat{\phi})x_n = O_p(n^{-1/2})O_p(1) = O_p(n^{-1/2})$$

3.27 Write $\nabla^k x_t = (1-B)^k x_t = \sum_{j=0}^k c_j x_{t-j}$ where c_j is the coefficient of B^j in the binomial expansion of $(1-B)^k$. Because x_t is stationary, $\mathrm{E}(\nabla^k x_t) = \mu_x \sum_{j=0}^k c_j$ independent of t, and (for $h \ge 0$) $\mathrm{cov}(\nabla^k x_{t+h}, \nabla^k x_t) = \mathrm{cov}(\sum_{j=0}^k c_j x_{t+h-j}, \sum_{j=0}^k c_j x_{t-j}) = \sum_{j=0}^{h+k} d_j \gamma_x(j)$, that is, the covariance is a time independent (linear) function of $\gamma_x(0), \dots, \gamma_x(h+k)$. Thus $\nabla^k x_t$ is stationary for any k.

Write $y_t = m_t + x_t$ where m_t is the given q-th order polynomial. Because $\nabla^k x_t$ is stationary for any k, we concentrate on m_t. Note that $\nabla m_t = m_t - m_{t-1} = c_q[t^q - (t-1)^q] + \sum_{j=0}^{q-1} c_j[t^j - (t-1)^j]$; from this it follows that the coefficient of t^q is zero. Now assume the result is true for $\nabla^k m_t$ and show it is true for $\nabla^{k+1} m_t$ [that is, for $k < q$, if $\nabla^k m_t$ is a polynomial of degree $q-k$ then $\nabla^{k+1} m_t$ is a polynomial of degree $q-(k+1)$]. The result holds by induction.

3.28 Write $y_t = x_t - x_{t-1}$, then the model is $y_t = w_t - \theta w_{t-1}$, which is invertible. That is, $w_t = \sum_{j=0}^{\infty} \theta^j y_{t-j} = \sum_{j=0}^{\infty} \theta^j (x_{t-j} - x_{t-1-j})$. Rearranging $w_t = x_t - \theta(1-\theta)x_{t-1} - \theta^2(1-\theta)x_{t-2} - \cdots$, or $x_t = \sum_{j=1}^{\infty} \theta^j (1-\theta) x_{t-j} + w_t$.

3.29 (a) Use induction: $y_{n+1}^n = \delta + \phi y_n$, $y_{n+2}^n = \delta + \phi y_{n+1}^n = \delta(1+\phi) + \phi y_n, \dots$, and

$$y_{n+j}^n = \delta + \phi y_{n+j-1}^n = \cdots = \delta(1 + \phi + \cdots + \phi^{j-1}) + \phi^j y_n = \delta \tfrac{(1-\phi^j)}{(1-\phi)} + \phi^j y_n.$$

(b) Note (1) $\sum_{j=1}^m y_{n+j}^n = \sum_{j=1}^m (x_{n+j}^n - x_{n+j-1}^n) = x_{n+m}^n - x_n$, (2) $\sum_{j=1}^m \delta \frac{(1-\phi^j)}{(1-\phi)} = \frac{\delta}{(1-\phi)}[m - \frac{\phi(1-\phi^m)}{(1-\phi)}]$ and (3) $\sum_{j=1}^m \phi^j y_n = y_n \frac{\phi(1-\phi^m)}{(1-\phi)} = (x_n - x_{n-1}) \frac{\phi(1-\phi^m)}{(1-\phi)}$ do the trick.

(c) First note that $\psi^*(z)(1-\phi z)(1-z) = (1 + \psi_1^* z + \psi_2^* z^2 + \cdots)(1 - [1+\phi]z + z^2) = 1$ yields the given homogeneous equation and solution, that is, $\psi_0^* = 1$ and $\psi_j^* = (1-\phi^{j+1})/(1-\phi)$ for $j \ge 1$. Thus, $P_{n+m}^n = \sigma_w^2 [1 + \frac{1}{(1-\phi)^2} \sum_{j=1}^{m-1} (1-\phi^{j+1})^2]$. (The summation part can be expanded, of course, but it's not clear that it gives any further insights. Note that for large m, the end terms in the summation, $(1-\phi^{j+1})^2 \approx 1$ for large j.)

3.30 See Figure 3.3. The EWMAs are smoother than the data (note the EWMAs are within the extremes of the data). The EWMAs are not extremely different for the different values of λ, the smoothest EWMA being when $\lambda = .75$.

```
x = log(varve[1:100])
x25 = HoltWinters(x, alpha=.75, beta=FALSE, gamma=FALSE)  # alpha = 1 - lambda
x50 = HoltWinters(x, alpha=.5, beta=FALSE, gamma=FALSE)
x75 = HoltWinters(x, alpha=.25, beta=FALSE, gamma=FALSE)
plot(x, type="o", ylab="log(varve)")
lines(x25$fit[,1], col=2)
lines(x50$fit[,1], col=3)
lines(x75$fit[,1], col=4)
legend("topright", c("25%", "50%", "75%"), col=2:4, lty=1)
```

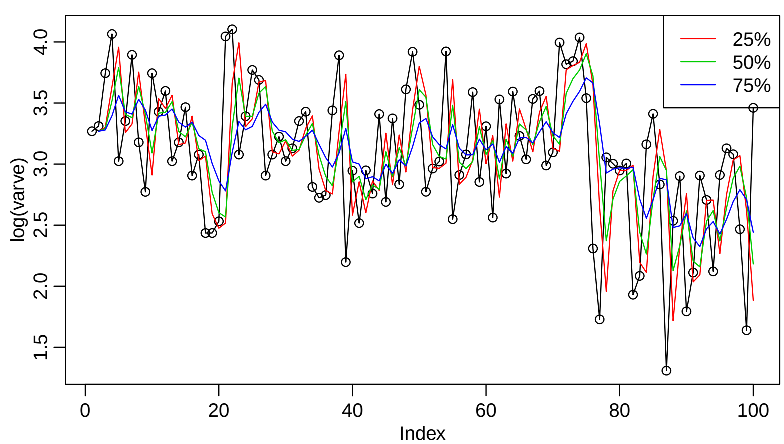

Fig. 3.3. EWMAs for Problem 3.30

3.31 Follow the steps of Examples 3.38 and 3.39, `sarima(gnpgr, 1, 0, 0)` will produce the diagnostics. The results should be similar to those in Example 3.39.

3.32 The ACF of the returns reveals only small amounts of autocorrelation. The most appropriate models seem to be ARMA(1,1) or ARMA(0,3). BIC prefers the ARMA(1,1) whereas AIC prefers the ARMA(0,3). The diagnostics are ok but there are some major outliers that may be affecting the results. R code below:

```
poil = diff(log(oil))
acf2(poil)
sarima(poil, 1, 0, 1)   # BIC favors
sarima(poil, 0, 0, 3)   # AIC favors
```

3.33 An ARIMA(1,1,1) seems to fit the data. Below is R code for the problem:

```
plot(gtemp)
plot(diff(gtemp))
acf2(gtemp)
sarima(gtemp, 1, 1, 1)
sarima.for(gtemp, 10, 1, 1, 1)
```

3.34 There is trend so we consider the (first) differenced series, which looks stationary. Investigation of the ACF and PACF of the differenced suggest an ARMA(0,1) or ARMA(1,1) model. Fitting an ARIMA(0,1,1) and ARIMA(1,1,1) to the original data indicates the ARIMA(0,1,1) model; the AR parameter is not significant in the ARIMA(1,1,1) fit. The residuals appear to be (borderline) white, but not normal – there are numerous outliers. Fitting the ARIMA(0,1,1) to log(so2) removes the outliers, but it doesn't noticeably change anything else.

```
plot(so2);                   plot(log(so2))
plot(diff(so2));             plot(diff(log(so2))
plot(diff(so2));             plot(diff(log(so2))
acf2(diff(so2));             acf2(diff(log(so2))
sarima(so2,0,1,1);           sarima(log(so2),0,1,1)
sarima.for(so2,4,0,1,1);     sarima.for(log(so2),4,0,1,1)
```

3.35 (a) `plot(sales)` - looks like random walk, so difference $d = 1$, (can do this with `log(sales)` but it's not necessary).
`acf2(diff(sales))` - ACF and PACF both tailing, so choose $p = q = 1$
`sarima(sales,1,1,1, no.constant=TRUE)` fits well.

(b) `ccf(diff(sales), diff(lead))` and `lag2.plot(diff(lead), diff(sales),8)` suggest that there is a strong linear relationship between `diff(sales)` and `lag(diff(lead),-3)`. Also, `lag2.plot` shows strong linear relationship.

(c) After fitting the regression, the ACF and PACF indicate an AR(1) for the residuals, which fits well.

```
u = ts.intersect(diff(sales), lag(diff(lead),-3))
ds = u[,1]
dl3 = u[,2]
(fit1 = lm(ds~dl3))          # beta1hat is highly significant
acf2(resid(fit1))             # => an AR(1) for the residuals
(fit2 = sarima(ds, 1,0,0, xreg=dl3))  # reg with ar1 errors
plot(resid(fit2$fit))
acf2(resid(fit2$fit))
```

3.36 (a) The cost is extremely large in the beginning of the series and rapidly decreases over time.

```
plot(cpg)
```

(b) If $c_t \approx \alpha e^{\beta t}$, then $\log c_t$ should be approximately linear. The model fits well.

```
summary(fit <- lm(log(cpg)~time(cpg)))
plot(log(cpg))
abline(fit)
```

(c) The residuals are autocorrelated and behave like an AR(1).

```
plot(resid(fit))
acf2(resid(fit))
```

(d) Fit the regression with AR(1) errors, which fits well.
```
sarima(log(cpg),1,0,0, xreg=time(cpg))
```

3.37 The regression with AR(2) errors is ok:
```
# from problem 2.2
temp = tempr-mean(tempr)
ded = ts.intersect(cmort, trend=time(cmort), temp, temp2=temp^2, part, partL4=lag(part,-4))
summary(fit <- lm(cmort~trend + temp + temp2 + part + partL4, data=ded))
# residuals:
acf2(resid(fit))                          # looks like ar2
sarima(ded[,1], 2,0,0, xreg=ded[,-1])
```

3.38 (a) The model is ARIMA$(0,0,2) \times (0,0,0)_s$ (s can be anything) or ARIMA$(0,0,0) \times (0,0,1)_2$.

(b) The MA polynomial is $\theta(z) = 1 + \Theta z^2$ with roots $z = \pm i/\sqrt{\Theta}$ outside the unit circle (because $|\Theta| < 1$). To find the invertible representation, note that $1/[1-(-\Theta z^2)] = \sum_{j=0}^{\infty}(-\Theta z^2)^j$ from which we conclude that $\pi_{2j} = (-\Theta)^j$ and $\pi_{2j+1} = 0$ for $j = 0, 1, 2, \ldots$. Consequently

$$w_t = \sum_{k=0}^{\infty} (-\Theta)^k x_{t-2k}.$$

(c) Write $x_{n+m} = -\sum_{k=1}^{\infty}(-\Theta)^k x_{n+m-2k} + w_n$ from which we deduce that

$$\tilde{x}_{n+m} = -\sum_{k=1}^{\infty} (-\Theta)^k \tilde{x}_{n+m-2k}$$

where $\tilde{x}_t = x_t$ for $t \le n$. For the prediction error, note that $\psi_0 = 1$, $\psi_2 = \Theta$ and $\psi_j = 0$ otherwise. Thus, $P_{n+m}^n = \sigma_w^2$ for $m = 1, 2$; when $m > 2$ we have $P_{n+m}^n = \sigma_w^2(1 + \Theta^2)$.

3.39 Use the code from Example 3.46 with `ma=.5` instead of `ma=-.5`.

3.40 The first obvious thing to do is to difference the data as in the examples in Chapter 2. Looking at `acf2(diff(chicken))`, the differenced data look like an AR(2) with an annual cycle. Using `sarima(chicken, 2, 1, 0)`, there appears to be some small but significant correlation at the annual lag, so try a seasonal model, `sarima(chicken, 2, 1, 0, 1, 0, 0, 12, no.constant=TRUE)` or `sarima(chicken, 2, 1, 0, 0, 0, 1, 12, no.constant=TRUE)`. Both fit well, but the second model is preferred by the ICs.

3.41 After plotting the unemployment data, say x_t, it is clear that one should fit an ARMA model to $y_t = \nabla_{12}\nabla x_t$. The ACF and PACF of y_t indicate a clear SMA(1) pattern (the seasonal lags in the ACF cut off after lag 12, whereas the seasonal lags in the PACF tail off at lags 12, 24, 36, and so on). Next, fit an SARIMA$(0,1,0) \times (0,1,1)_{12}$ to x_t and look at the ACF and PACF of the residuals. The within season part of the ACF tails off, and the PACF is either cutting off at lag 2 or is tailing off. These facts suggest an AR(2) or and ARMA(1,1) for the within season part of the model. Hence, fit an (i) SARIMA$(2,1,0) \times (0,1,1)_{12}$ or an (ii) SARIMA$(1,1,1) \times (0,1,1)_{12}$ to x_t. Both models have the same number of parameters, so it should be clear that model (i) is better because the MSE is smaller for model (i) and the residuals appear to white (while there may still be some correlation left in the residuals for model (ii)). Below is the R code for fitting model (i), along with diagnostics and forecasting.
```
acf2(diff(diff(unemp),12), 50)
sarima(unemp, 2, 1, 0, 0, 1, 1, 12)
sarima.for(unemp, 12, 2, 1, 0, 0, 1, 1, 12)
```

3.42 This data set is an update to `unemp`, but is in terms of percentage unemployed. The techniques are the same as the previous problem and the model is nearly the same; one addition MA parameter is needed.

```
acf2(diff(diff(unemp),12), 60)
sarima(UnempRate, 2, 1, 1, 0, 1, 1, 12)
sarima.for(UnempRate, 12, 2, 1, 1, 0, 1, 1, 12)
```

3.43 The monthly ($s = 12$) U.S. Live Birth Series can be found in `birth.dat` After plotting the data, say x_t, it is clear that one should fit an ARMA model to $y_t = \nabla_{12}\nabla x_t$. The ACF and PACF of y_t indicate a seasonal MA of order one, that is, fit an ARIMA$(0,0,0) \times (0,0,1)_{12}$ to y_t. Looking at the ACF and PACF of the residuals of that fit suggests fitting a nonseasonal ARMA(1,1) component (both the ACF and PACF appear to be tailing off). After that, the residuals appear to be white. Finally, we settle on fitting an ARIMA$(1,1,1) \times (0,1,1)_{12}$ model to the original data, x_t. The code for this problem is nearly the same as the previous problem.

```
plot(birth)
acf2(diff(diff(birth),12), 50)
sarima(birth, 1, 1, 1, 0, 1, 1, 12)
sarima.for(birth, 12, 1, 1, 0, 0, 1, 1, 12)
```

3.44 Because of the increasing variability, the data, jj_t, should be logged prior to any further analysis. A plot of the logged data, say $y_t = \ln jj_t$, shows trend, and one should notice the differences in the behavior of the series at the beginning, middle, and end of the data (as if there are 3 different regimes). Due to these inconsistencies (nonstationarities), it is difficult to discover an ARMA model and one should expect students to come up with various models. In fact, assigning this problem may decrease your student evaluations substantially.

Next, apply a first difference and seasonal difference to the logged data: $x_t = \nabla_4\nabla y_t$. The PACF of x_t reveals a large correlation at the seasonal lag 4, so an SAR(1) seems appropriate. The ACF and PACF of the residuals reveals an ARMA(1,1) correlation structure for the within the seasons. This seems to be a reasonable fit. Hence, a reasonable model is an SARIMA$(1,1,0) \times (1,1,0)_4$ on the logged data. Below is R code for this problem.

```
plot(diff(diff(log(jj)),4))
acf2(diff(diff(log(jj)),4))
sarima(log(jj),1,1,0,1,1,0,4)
sarima.for(log(jj),4,1,1,0,1,1,0,4)
```

3.45 Clearly $\sum_{j=1}^{p} \phi_j x_{n+1-j} \in \overline{\text{sp}}\{x_k; k \le n\}$, so it suffices to show that $\hat{x}_{n+1}$ satisfies the prediction equations $\text{E}[(x_{n+1} - \hat{x}_{n+1})x_k] = 0$ for $k \le n$. But, by the model assumption, $x_{n+1} - \hat{x}_{n+1} = w_{n+1}$ and $\text{E}(w_{n+1}x_k) = 0$ for all $k \le n$.

3.46 First note that $x_i - x_i^{i-1}$ and $x_j - x_j^{j-1}$ for $j > i = 1, ..., n$ are uncorrelated. This is because $x_i - x_i^{i-1} \in \overline{\text{sp}}\{x_k; k = 1, ..., i\}$ but $x_j - x_j^{j-1}$ is orthogonal (uncorrelated) to $\overline{\text{sp}}\{x_k; k = 1, ..., i\}$ by definition of x_j^{j-1} Thus, by the projection theorem, for $t = 1, 2, \ldots,$

$$x_{t+1}^t = \sum_{k=1}^{t} \theta_{tk}(x_{t+1-k} - x_{t+1-k}^{t-k}) \tag{1}$$

where the θ_{tk} are obtained by the prediction equations. Multiply both sides of (1) by $x_{j+1} - x_{j+1}^j$ for $j = 0, ..., t-1$ and take expectation to obtain

$$\text{E}\left[x_{t+1}^t(x_{j+1} - x_{j+1}^j)\right] = \theta_{t,t-j}P_{j+1}^j.$$

Because of the orthogonality $\text{E}[(x_{t+1} - x_{t+1}^t)(x_{j+1} - x_{j+1}^j)] = 0$ when $j < t$, so equation above can be written as

$$\text{E}\left[x_{t+1}(x_{j+1} - x_{j+1}^j)\right] = \theta_{t,t-j}P_{j+1}^j. \tag{2}$$

Using (1) with t replaced by j, (2) can be written as

$$\theta_{t,t-j} = \text{E}\left[x_{t+1}\left(x_{j+1} - \sum_{k=1}^{j} \theta_{jk}(x_{j+1-k} - x_{j+1-k}^{j-k})\right)\right]\left(P_{j+1}^j\right)^{-1}.$$

Thus

$$\theta_{t,t-j} = \left\{ \gamma(t-j) + \sum_{k=1}^{j} \theta_{jk} \mathrm{E}\left[x_{t+1}(x_{j+1-k} - x_{j+1-k}^{j-k}) \right] \right\} \left(P_{j+1}^{j} \right)^{-1}. \tag{3}$$

Using (2) we can write $\mathrm{E}[x_{t+1}(x_{j+1-k} - x_{j+1-k}^{j-k})] = \theta_{t,t-k} P_{k+1}^{k}$ so (3) can be written in the form of (2.71). To show (2.70), first note that $\mathrm{E}(x_{t+1}x_{t+1}^{t}) = \mathrm{E}[x_{t+1}^{t}\mathrm{E}(x_{t+1}|x_1,...,x_t)] = \mathrm{E}[(x_{t+1}^{t})^2]$. Then, for $t = 1, 2, ...,$

$$P_{t+1}^{t} = \mathrm{E}(x_{t+1} - x_{t+1}^{t})^2 = \gamma(0) - \mathrm{E}[(x_{t+1}^{t})^2] = \gamma(0) - \sum_{j=0}^{t-1} \theta_{t,t-j}^2 P_{j+1}^{j}$$

3.47 (a) From the projection theorem, $x_{n+1}^{n} = \sum_{k=1}^{n} \alpha_k x_k$, where $\alpha = (\alpha_1, ..., \alpha_n)'$ satisfies

$$\begin{bmatrix} 2 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 2 \end{bmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ -1 \end{pmatrix}.$$

Solving recursively we get $\alpha_2 = 2\alpha_1$, $\alpha_3 = 3\alpha_1$, and in general, $\alpha_k = k\alpha_1$ for $k = 1, ..., n$. This fact and the last equation gives $\alpha_1 = -\frac{1}{n+1}$ and the result follows.

(b) MSE$= \gamma(0) - a_1\gamma(-n) - a_2\gamma(-n+1) - \cdots - a_n\gamma(-1) = \sigma_w^2[2 - n/(n+1)] = \frac{(n+2)}{(n+1)}\sigma_w^2$.

3.48 Let $x = (x_1, ..., x_n)'$, then $x \sim \mathrm{N}(0, \Gamma_n)$ with likelihood $L(x) = |\Gamma_n|^{-1/2} \exp\{-\frac{1}{2}x'\Gamma_n^{-1}x\}$ (ignoring a constant). Note that $x_t^{t-1} = P_{\overline{\mathrm{sp}}\{x_1,...,x_{t-1}\}} x_t = \mathrm{E}(x_t|x_1, \ldots, x_{t-1})$ because of the normality assumption. Hence, the innovations, $\epsilon_t = x_t - x_t^{t-1}$, are independent Normal random variables with variance P_t^{t-1} for $t = 1, ..., n$. Because x_t^{t-1} is a linear combination of $x_1, \ldots, x_{t-1}$, the transformation of $x \mapsto \epsilon$, where $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$, is lower triangular with ones along the diagonal; i.e. $x = C\epsilon$, where C is lower triangular. In fact, Problem 3.38 shows that (with θ_{ij} defined there)

$$C = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \theta_{11} & 1 & 0 & \cdots & 0 \\ \theta_{22} & \theta_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & & \cdots & 1 \end{bmatrix}.$$

Thus, $L(x) = L(C\epsilon)$. Noting that $C\epsilon \sim \mathrm{N}(0, CDC')$, where $D = \mathrm{diag}\{P_1^0, P_2^1, \ldots, P_n^{n-1}\}$ we have

$$L(x) = L(C\epsilon) = |CDC'|^{-1/2} \exp\{-\frac{1}{2}\epsilon' C'(CDC')^{-1}C\epsilon\}.$$

This establishes the result noting that $|CDC'| = |C^2|\,|D|$ and $|C^2| = 1$, $|D| = P_1^0 P_2^1 \cdots P_n^{n-1}$, and in the exponential $\epsilon' C'(CDC')^{-1}C\epsilon = \epsilon' D^{-1}\epsilon = \sum_{t=1}^{n}(x_t - x_t^{t-1})^2/P_t^{t-1}$.

3.49 These results are proven in Brockwell and Davis (1991, Proposition 2.3.2).

3.50 The proof of Property 3.2 is virtually identical to the proof of Property 3.1 given in Appendix B.

Chapter 4

Solutions

NOTE: With slight changes, you can replace `mvspec()` with `spectrum()` or `spec.pgram()` in this chapter.

4.1 This part is included mainly to have the information listed somewhere in the text. The hint has enough detail to make it through all the calculations. Also, the proofs can be found in numerous sources, e.g., Fuller (1995), Theorem 3.1.1.

4.2 (a)–(b) The code is basically the same as the example and is given below. The difference is the frequencies in the data (which are .06, .1, .4) are no longer fundamental frequencies (which are of the form $k/128$). Consequently, the periodogram will have non-zero entries near .06, .1, .4 (unlike the example where all other frequencies are zero).

```
t = 1:128
x1 = 2*cos(2*pi*t*6/100) + 3*sin(2*pi*t*6/100)
x2 = 4*cos(2*pi*t*10/100) + 5*sin(2*pi*t*10/100)
x3 = 6*cos(2*pi*t*40/100) + 7*sin(2*pi*t*40/100)
x = x1 + x2 + x3
par(mfrow=c(2,2))
plot.ts(x1, ylim=c(-16,16), main="freq=6/100, amp^2=13")
plot.ts(x2, ylim=c(-16,16), main="freq=10/100, amp^2=41")
plot.ts(x3, ylim=c(-16,16), main="freq=40/100, amp^2=85")
plot.ts(x, ylim=c(-16,16), main="sum")
P = abs(2*fft(x)/128)^2
f = 0:64/128
plot(f, P[1:65], type="o", xlab="frequency",  ylab="periodogram")
```

(c) Use the same code as in the example, but with `x = x1 + x2 + x3 + rnorm(100,0,5).` Now the periodogram will have large peaks at .06, .1, .4, but will also be positive at most other fundamental frequencies.

4.3 (a) Rewrite the transformation as

$$x = \tan^{-1}\frac{z_2}{z_1} \qquad\qquad y = z_1^2 + z_2^2.$$

Note that

$$\frac{\partial}{\partial x}\tan^{-1} u = \frac{1}{1+u^2}\frac{\partial u}{\partial x}.$$

Write the joint density of x and y as

$$g(x,y) = f(z_1,z_2)J,$$

where J denotes the Jacobian, i.e., the determinant of the 2×2 matrix $\{\frac{\partial(z_1,z_2)}{\partial(x,y)}\}$. It is easier to compute

$$\frac{1}{J} = \left|\begin{pmatrix} \frac{\partial x}{\partial z_1} & \frac{\partial x}{\partial z_2} \\ \frac{\partial y}{\partial z_1} & \frac{\partial y}{\partial z_2}\end{pmatrix}\right| = \left|\begin{pmatrix} \frac{-z_2}{z_1^2+z_2^2} & \frac{z_1}{z_1^2+z_2^2} \\ 2z_1 & 2z_2\end{pmatrix}\right| = 2,$$

implying that $J=\frac{1}{2}$. For the joint density of x and y, we obtain

$$g(x,y)=(2\pi)^{-1}\frac{1}{2}\exp\left\{-\frac{1}{2}(z_1^2+z_2^2)\right\}=\frac{1}{2\pi}\times\frac{1}{2}\exp\left\{-\frac{y}{2}\right\}$$

for $-\pi<x<\pi$ and $0<y<\infty$. Integrating over x and y separately shows that the density factors into a product of marginal densities as stated in the problem.

(b) Going the other way, we use

$$h(z_1,z_2)=g(x,y)\frac{1}{J}=\frac{1}{2\pi}2\frac{1}{2}\exp\{-y/2\}=\frac{1}{2\pi}\exp\{-\frac{1}{2}(z_1^2+z_2^2)\},$$

since $z_1^2+z_2^2=(\sqrt{y})^2(\cos^2(x)+\sin^2(x))=y$.

4.4 Most of the work is done in the footnote. Write the terms in the sum (4.4) as $x_{t,k}$ and note that $x_{t,k}$ and $x_{t,\ell}$ are independent for $k\neq\ell$ because the Us are independent (the Us being uncorrelated is sufficient here; this could be pointed out to students - we kept them as independent to go along with the normal and chi-square/uniform theme discussed in (4.2), which requires independence).

$$\begin{aligned}\gamma_k(h)&=\mathrm{E}(x_{t+h,k}x_{t,k})\\&=\mathrm{E}\Big\{\big(U_{1k}\sin[2\pi\omega_k(t+h)]+U_{2k}\cos[2\pi\omega_k(t+h)]\big)\times\big(U_{1k}\sin[2\pi\omega_kt]+U_{2k}\cos[2\pi\omega_kt]\big)\Big\}\\&=\sigma_k^2\Big(\sin[2\pi\omega_k(t+h)]\sin[2\pi\omega_kt]+\cos[2\pi\omega_k(t+h)]\cos[2\pi\omega_kt]\Big)\\&=\sigma_k^2\cos[2\pi\omega_k(t+h)-2\pi\omega_kt]=\sigma_k^2\cos[2\pi\omega_kh]\end{aligned}$$

and $\gamma(h)=\sum_{k=1}^q\gamma_k(h)$ give (4.5).

4.5 (a) $\mathrm{E}w_t=\mathrm{E}x_t=0$ by linearity, $\gamma_w(0)=1$ and zero otherwise; $\gamma_x(0)=(1+\theta^2)$, $\gamma_x(\pm1)=-\theta$, and is zero otherwise. The series are stationary because they are zero mean and the autocovariance does not depend on time but only on the shift.

(b) By Property 4.3, $f_x(\omega)=\sigma_w^2|\theta(e^{-2\pi i\omega})|^2$, thus

$$f_x(\omega)=\sigma_w^2|1-\theta e^{-2\pi i\omega}|^2=\sigma_w^2[1-\theta(e^{-2\pi i\omega}+e^{2\pi i\omega})+\theta^2]=\sigma_w^2[1+\theta^2-2\theta\cos(2\pi\omega)].$$

4.6 (a) This is similar to 4.4(b). By Property 4.3,

$$f_x(\omega)=\sigma_w^2|\phi(e^{-2\pi i\omega})|^{-2}=\sigma_w^2|1-\phi e^{-2\pi i\omega}|^{-2}=\frac{\sigma_w^2}{1+\phi^2-2\phi\cos(2\pi\omega)}.$$

(b) From (4.12), write

$$\begin{aligned}f_x(\omega)&=\sum_{h=-\infty}^{0}\frac{\sigma_w^2\phi^{-h}\mathrm{e}^{-2\pi i\omega h}}{1-\phi^2}+\sum_{h=1}^{\infty}\frac{\sigma_w^2\phi^{h}\mathrm{e}^{-2\pi i\omega h}}{1-\phi^2}=\frac{\sigma_w^2}{1-\phi^2}\left(\sum_{h=0}^{\infty}(\phi\mathrm{e}^{2\pi i\omega})^h+\sum_{h=1}^{\infty}(\phi\mathrm{e}^{-2\pi i\omega})^h\right)\\&=\frac{\sigma_w^2}{1-\phi^2}\left(\frac{1}{1-\phi\mathrm{e}^{2\pi i\omega}}+\frac{\phi\mathrm{e}^{-2\pi i\omega}}{1-\phi\mathrm{e}^{-2\pi i\omega}}\right)=\frac{\sigma_w^2}{1-\phi^2}\frac{1-\phi^2}{|1-\phi\mathrm{e}^{2\pi i\omega}|^2}=\frac{\sigma_w^2}{1+\phi^2-2\phi\cos(2\phi\omega)},\end{aligned}$$

and the spectrum is as claimed, by the uniqueness of the Fourier transform.

4.7 First, note that the autocovariance function is

$$\gamma_x(h)=(1+A^2)\gamma_s(h)+A\gamma_s(h-D)+A\gamma_s(h+D)+\gamma_n(h)$$

Using the spectral representation directly,

$$\gamma_x(h)=\int_{-1/2}^{1/2}\big[\big(1+A^2+A\mathrm{e}^{2\pi i\omega D}+A\mathrm{e}^{-2\pi i\omega D}\big)f_s(\omega)+f_n(\omega)\big]\mathrm{e}^{2\pi i\omega h}d\omega$$

Substituting the exponential representation for $\cos(2\pi\omega D)$ and using the uniqueness gives the required result.

4.8 The product series will have mean $\mathrm{E}(x_t y_t) = \mathrm{E}(x_t)\mathrm{E}(y_t) = 0$ and autocovariance

$$\gamma_z(h) = \mathrm{E}x_{t+h}y_{t+h}x_t y_t = \mathrm{E}(x_{t+h}x_t)\mathrm{E}(y_{t+h}y_t) = \gamma_x(h)\gamma_y(h).$$

Now, by (4.12), (4.13) and (4.14)

$$\begin{aligned} f_z(\omega) &= \sum_{h=-\infty}^{\infty} \gamma_x(h)\gamma_y(h)\exp\{-2\pi i\omega h\} = \int_{-1/2}^{1/2} \sum_{h=-\infty}^{\infty} \gamma_x(h)\mathrm{e}^{-2\pi i\omega h}\mathrm{e}^{-2\pi i\omega h} f_y(\lambda)\, d\lambda \\ &= \int_{-1/2}^{1/2} \sum_{h=-\infty}^{\infty} \gamma_x(h)\mathrm{e}^{-2\pi i(\omega-\lambda)h} f_y(\lambda)\, d\lambda = \int_{-1/2}^{1/2} f_x(\omega-\lambda) f_y(\lambda)\, d\lambda. \end{aligned}$$

4.9 Below is R code that will plot the periodogram on the actual scale and then on a log scale (this produces a generic confidence interval – see Example 4.9 on how to get precise limits). The two major peaks are marked; they are 3 cycles/240 years or the 80 year cycle, and 22 cycles/240 years or about an 11 year cycle.

```
mvspec(sunspotz, taper=0, log="no")   # for CI, remove log="no"
abline(v=3/240, lty="dashed")  #  80 year cycle
abline(v=22/240, lty="dashed") #  11 year cycle
```

4.10 This is like the previous problem; the main component is 1 cycle/16 rows, although there's not enough data to get significance.

```
par(mfrow=c(2,1))      # for CIs, remove log="no" below
mvspec(saltemp, taper=0, log="no")
abline(v=1/16, lty="dashed")
mvspec(salt, taper=0, log="no")
abline(v=1/16, lty="dashed")
```

4.11 (a) Write the model in the notation of Chapter 2 as $x_t = \boldsymbol{\beta}'\boldsymbol{z}_t + w_t$, where $\boldsymbol{z}_t = (\cos(2\pi\omega_k t), \sin(2\pi\omega_k t))'$ and $\boldsymbol{\beta} = (\beta_1, \beta_2)'$. Then

$$\sum_{t=1}^n \boldsymbol{z}_t\boldsymbol{z}_t' = \begin{pmatrix} \sum_{t=1}^n \cos^2(2\pi\omega_k t) & \sum_{t=1}^n \cos(2\pi\omega_k t)\sin(2\pi\omega_k t) \\ \sum_{t=1}^n \cos(2\pi\omega_k t)\sin(2\pi\omega_k t) & \sum_{t=1}^n \sin^2(2\pi\omega_k t) \end{pmatrix} = \begin{pmatrix} n/2 & 0 \\ 0 & n/2 \end{pmatrix}$$

from the orthogonality properties of the sines and cosines. For example,

$$\begin{aligned} \sum_{t=1}^n \cos^2(2\pi\omega_k t) &= \frac{1}{4}\sum_{t=1}^n \left(\mathrm{e}^{2\pi i\omega_k t} + \mathrm{e}^{-2\pi i\omega_k t}\right)\left(\mathrm{e}^{2\pi i\omega_k t} + \mathrm{e}^{-2\pi i\omega_k t}\right) \\ &= \frac{1}{4}\sum_{t=1}^n \left(\mathrm{e}^{4\pi i\omega_k t} + 1 + 1 + \mathrm{e}^{-4\pi i\omega_k t}\right) = \frac{n}{2}, \end{aligned}$$

because, for example,

$$\sum_{t=1}^n \mathrm{e}^{4\pi i\omega_k t} = \frac{\mathrm{e}^{4\pi ik/n}\left(1 - \mathrm{e}^{4\pi ik}\right)}{1 - \mathrm{e}^{4\pi ik/n}} = 0$$

Substituting,

$$\begin{pmatrix} \hat\beta_1 \\ \hat\beta_2 \end{pmatrix} = \frac{2}{n}\begin{pmatrix} \sum_{t=1}^n x_t\cos(2\pi\omega_k t) \\ \sum_{t=1}^n x_t\sin(2\pi\omega_k t) \end{pmatrix} = 2n^{-1/2}\begin{pmatrix} d_c(\omega_k) \\ d_s(\omega_k) \end{pmatrix}.$$

(b) Now,

$$\begin{aligned} SSE &= \boldsymbol{x}'\boldsymbol{x} - 2n^{-1/2}\left(d_c(\omega_k), \quad d_s(\omega_k) \right)\begin{pmatrix} \sum_{t=1}^n x_t\cos(2\pi\omega_k t) \\ \sum_{t=1}^n x_t\sin(2\pi\omega_k t) \end{pmatrix} \\ &= \boldsymbol{x}'\boldsymbol{x} - 2\left[d_c^2(\omega_k) + d_s^2(\omega_k)\right] = \boldsymbol{x}'\boldsymbol{x} - 2I_x(\omega_k). \end{aligned}$$

(c) The reduced model is given by $x_t = w_t$, so that $RSS_1 = \sum_{t=1}^n x_t^2 = \boldsymbol{x}'\boldsymbol{x}$ and RSS is given in part (b). For the F-test we have $q = 2, q_1 = 0$, so that

$$F_{2,n-2} = \frac{2I_x(\omega_k)}{\boldsymbol{x}'\boldsymbol{x} - 2I_x(\omega_k)} \frac{n-2}{2}$$

is monotone in I_x.

4.12 By applying the definition to x_{t-s}, we obtain

$$\begin{aligned}\sum_{s=1}^n a_s x_{t-s} &= n^{-1/2} \sum_{k=0}^{n-1} d_x(\omega_k) \sum_{s=1}^n a_s \mathrm{e}^{2\pi i\omega_k(t-s)} \\ &= \sum_{k=0}^{n-1} d_x(\omega_k) n^{-1/2} \sum_{s=1}^n a_s \mathrm{e}^{-2\pi i\omega_k s} \mathrm{e}^{2\pi i\omega_k t} = \sum_{k=0}^{n-1} d_A(\omega_k) d_x(\omega_k) \mathrm{e}^{2\pi i\omega_k t}.\end{aligned}$$

4.13 Don't forget to difference the data first.

```
mvspec(diff(chicken), spans=c(3,3), log="no", taper=.5)  # looks ok
abline(v=c(1/5,.65,1), lty=2)
```

Aside from the annual cycle (1) there is a 5 year cycle and another at 1/.65, which is about a 1.5 year cycle. We have no particular explanation of these cycles; further details may be obtained by contacting the United Poultry Growers association.

4.14 Continuing from Problem 4.8

```
mvspec(sunspotz, spans=c(7,7), log="no", taper=.5)
# mvspec(sunspotz, spans=c(7,7), taper=.5)  # for log scale with CIs
abline(v=3/240, lty="dashed")  #  80 year cycle
abline(v=22/240, lty="dashed") #  11 year cycle
```

4.15 Continuing from Problem 4.9:

```
par(mfrow=c(2,1))
mvspec(saltemp, spans=c(1,1), log="no", taper=.5)
abline(v=1/16, lty="dashed")
salt.per = mvspec(salt, spans=c(1,1), log="no", taper=.5)
abline(v=1/16, lty="dashed")
```

4.16 R code and discussion below. Also, see Figure 4.1.

```
sp.per = mvspec(speech,  taper=0) # plots the log-periodogram - which is periodic
x = log(sp.per$spec)          # x is the log-periodogram values
par(mfrow=c(2,1))
plot.ts(x)                    # plot x=log-periodogram as a time series
x.sp = mvspec(x, span=5)      # cepstral analysis, x is detrended by default
abline(v=.1035, lty="dashed")
cbind(x.sp$freq, x.sp$spec)   # this lists the quefrencies and cepstra
 [52,] 0.101562500 32.7549412
 [53,] 0.103515625 34.8354468  <- peak is here
 [54,] 0.105468750 30.3669195
```

From Problem 4.6, detrended `x` in the code above is $x_j \sim \cos(2\pi \frac{j}{n} D)$ where n here is `sp.per$n.used = 1024`. Hence, the cepstrum will have a peak around D/n. Thus $\hat{D} = 1024 \times .1035 = 105.984$ points, or $\hat{D} = .0105984$ seconds (recall, 10,000 observations per second). This agrees with Example 1.24.

4.17 For $y_t = h_t x_t$, the DFT is $d_y(\omega_k) = n^{-1/2} \sum_{t=1}^n h_t x_t \mathrm{e}^{-2\pi i\omega_k t}$. Then, the expectation of the modulus squared DFT is

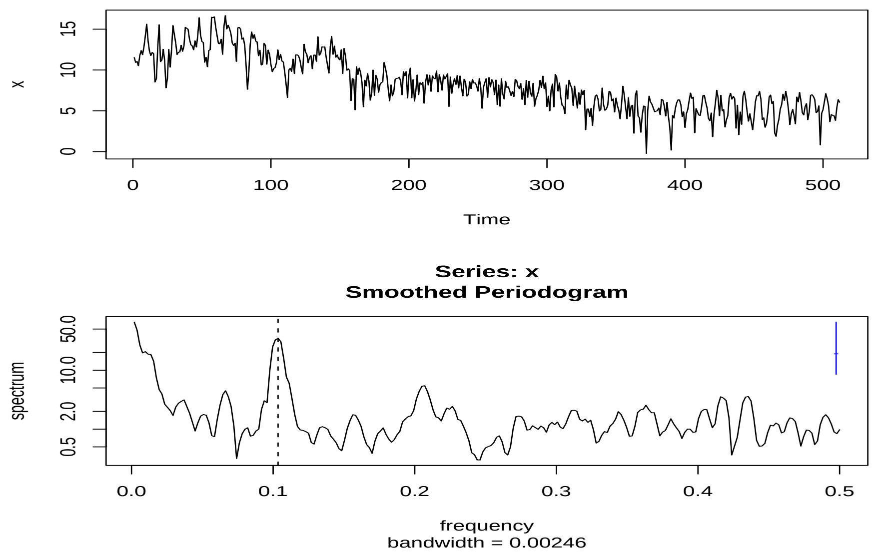

Fig. 4.1. Figure for Problem 4.14

$$
\begin{aligned}
\mathrm{E}|d_y((\omega_k)|^2 &= n^{-1}\sum_{s=1}^{n}\sum_{t=1}^{n} h_s h_t \gamma_x(s-t)\mathrm{e}^{-2\pi i\omega_k(s-t)} \\
&= n^{-1}\int_{-1/2}^{1/2}\sum_{s=1}^{n} h_s \mathrm{e}^{-2\pi i(\omega_k-\omega)s}\sum_{t=1}^{n} h_t \mathrm{e}^{2\pi i(\omega_k-\omega)t} f_x(\omega)\,d\omega = \int_{-1/2}^{1/2}|H_n(\omega_k-\omega)|^2 f_x(\omega)\,d\omega.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\mathrm{E}\Big[L^{-1}\sum_{\ell}|Y(\omega_k+\ell/n)|^2\Big] &= \int_{-1/2}^{1/2}\frac{1}{L}\sum_{\ell}|H_n(\omega_k+\ell/n-\omega)|^2 f_x(\omega)\,d\omega \\
&= \int_{-1/2}^{1/2} W_n(\omega_k-\omega) f_x(\omega)\,d\omega.
\end{aligned}
$$

4.18 (a) Since the means are both zero and the ACF's and CCF's

$$
\gamma_x(h) = \begin{cases} 2 & h=0 \\ -1 & h=\pm 1 \\ 0 & |h|\ge 2 \end{cases} \quad
\gamma_y(h) = \begin{cases} 1/2 & h=0 \\ 1/4 & h=\pm 1 \\ 0 & |h|\ge 2 \end{cases} \quad
\gamma_{xy}(h) = \begin{cases} 0 & h=0 \\ -1/2 & h=1 \\ 1/2 & h=-1 \\ 0 & |h|\ge 2 \end{cases}
$$

do not depend on the time index, the series are jointly stationary.

(b)

$$
f_x(\omega) = |1-\mathrm{e}^{-2\pi i\omega}|^2 = 2(1-\cos(2\pi\omega)) \quad \text{and} \quad f_y(\omega) = \frac{1}{4}|1+e^{-2\pi i\omega}|^2 = \frac{1}{2}(1+\cos(2\pi\omega))
$$

As ω goes from $0 \to \frac{1}{2}$, $f_x(\omega)$ increases, whereas $f_y(\omega)$ decreases. This means x_t has more high frequency behavior and y_t has more low frequency behavior.

(c)

$$
P\Big\{\frac{2La}{f_y(.10)} \le \frac{2L\bar{f}_y(.10)}{f_y(.10)} \le \frac{2Lb}{f_y(.10)}\Big\} = P\Big\{\frac{2La}{f_y(.10)} \le \chi^2_{2L} \le \frac{2Lb}{f_y(.10)}\Big\}
$$

We can make the probability equal to .90 by setting

$$\frac{2La}{f_y(.10)} = \chi^2_{2L}(.95) \quad \text{and} \quad \frac{2Lb}{f_y(.10)} = \chi^2_{2L}(.05)$$

Setting $L = 3, \chi^2_6(.95) = 1.635, \chi^2_6(.05) = 12.592, f_y(.10) = .9045$ and solving for a and b yields $a = .25, b = 1.90$.

4.19 R code for fitting an AR spectrum using AIC is `spec.ar(sunspotz)`. The analysis results in fitting an AR(16) spectrum, which is similar to the nonparametric spectral estimate.

4.21 R code for fitting an AR spectrum using AIC is `spec.ar(rec)`. The analysis results in fitting an AR(13) spectrum, which is similar to the nonparametric spectral estimate.

4.22 We have $2L\bar{f}_x(1/8)/f_x(1/8) \sim \chi^2_{2L}$ where $f_x(1/8) = [1 + .5^2 - 2(.5)\cos(2\pi/8)]^{-1} = 1.842$ from Problem 4.24(a). For $L = 3$, we have $2(3)2.25/1.842 = 7.26$ and does not exceed $\chi^2_6(.05) = 12.59$. For $L = 11$, we have $2(11)2.25/1.842 = 26.87$ and does not exceed $\chi^2_{22}(.05) = 33.92$. Neither sample has evidence for rejecting the hypothesis that the spectrum is as claimed at the $\alpha = .05$ level.

4.23 The conditions imply that under $H_0 : d(\omega_k + \ell/n) \sim CN\{0, f_n(\omega)\}$ and under $H_1 : d(\omega_k + \ell/n) \sim CN\{0, f_s(\omega) + f_n(\omega)\}$. For simplicity in notation, denote $d_\ell = d(\omega_k + \ell/n)$ and $f_s = f_s(\omega), f_n = f_n(\omega)$.

(a) The ratio of likelihoods, under the two hypotheses would be

$$\frac{p_1}{p_0} = \frac{\pi^{-L}(f_s + f_n)^{-L} \prod_\ell \exp\{-|d_\ell|^2/(f_s + f_n)\}}{\pi^{-L}(f_n)^{-\ell} \prod_\ell \exp\{-|d_\ell|^2/f_n\}}$$

and the log likelihood involving the data d_ℓ is proportional to

$$T = \ln \frac{p_1}{p_0} \propto \sum_\ell |d_\ell|^2 \left(\frac{1}{f_n} + \frac{1}{(f_s + f_n)} \right),$$

(b) Write

$$T = \frac{f_s}{f_n(f_s + f_n)} \sum_\ell |d_\ell|^2$$

and note that

$$\frac{2\sum_\ell |d_\ell|^2}{f_n} \sim \chi^2_{2L}$$

under H_0 and

$$\frac{2\sum_\ell |d_\ell|^2}{(f_s + f_n)} \sim \chi^2_{2L}$$

under H_1. Hence

$$T \sim \frac{1}{2} \frac{f_s}{(f_s + f_n)} \chi^2_{2L}$$

under H_0 and

$$T \sim \frac{1}{2} \frac{f_s}{f_n} \chi^2_{2L}$$

under H_1.

(c) Here, we note that

$$P_F = P\{T > K|H_0\} = P\left\{ \chi^2_{2L} > 2K\left(\frac{f_s + f_n}{f_s}\right) \right\} = P\left\{ \chi^2_{2L} > 2K \, \frac{(SNR + 1)}{SNR} \right\},$$

and

$$P_d = P\{T > K|H_1\} = P\left\{\chi^2_{2L} > 2K\left(\frac{f_n}{f_s}\right)\right\} = P\left\{\chi^2_{2L} > \frac{2K}{SNR}\right\},$$

where SNR denotes the signal-to-noise ratio. Note that, as $SNR \to \infty$, $P_F \to P\{T > 2K\}$ and $P_d \to 1$, and the signal detection probability approaches unity for a fixed false alarm rate, as guaranteed by the Neyman-Pearson lemma.

4.24 The analysis is similar to that of Example 4.18. The squared coherency is very large at periods ranging from 16-32 points, or 272-544 feet (1 point = 17 feet). R code below:

```
x = ts(cbind(saltemp,salt))
s = mvspec(x, kernel("daniell",2), taper=0)
s$df                        # = 10
f = qf(.999, 2, s$df-2)     # = 18.49365
C = f/(18+f)                # = 0.5067635
plot(s, plot.type = "coh", ci.lty = 2)
abline(h = C)
cbind(s$freq, s$coh)
            [,1]        [,2]
 [1,] 0.015625 0.598399213
 [2,] 0.031250 0.859492914  <- period 1/.03 = 32
 [3,] 0.046875 0.891469033
 [4,] 0.062500 0.911331648  <- period 1/.06 = 16
 [5,] 0.078125 0.749974642
```

4.25 (a) The CCF is

$$\gamma_{xy}(h) = \text{cov}(x_{t+h}, y_t) = \text{cov}(w_{t+h}, \phi w_t + v_t) = \phi\sigma^2\delta_h^{-D},$$

where $\delta_h^{-D} = 1$ when $h = -D$ and zero otherwise. Thus,

$$f_{xy}(\omega) = \sum_h \gamma_{xy}(h)\exp(-2\pi i\omega h) = \phi\sigma^2\exp(2\pi i\omega D).$$

Also, using Proposition 4.3

$$f_x(\omega) = \sigma^2 \quad \text{and} \quad f_y(\omega) = \sigma^2(1+\phi^2).$$

Finally,

$$\rho^2_{xy}(\omega) = \frac{|\phi\sigma^2\exp(2\pi i\omega D)|^2}{\sigma^2 \times \sigma^2(1+\phi^2)} = \frac{\phi^2}{(1+\phi^2)},$$

which is constant and does not depend on the value of D.

(b) In this case, $\rho^2_{xy}(\omega) = .81/1.81 = .45$. The R code to simulate the data and estimate the coherence is given below. Note that using $L = 1$ gives a value of 1 no matter what the processes are, and increasing L (`span`) gives better estimates.

```
x = rnorm(1024)
y = .9*x + rnorm(1024)
u = ts(cbind(x,y))
s1 = mvspec(u, taper=0, plot=F)$coh           # use this for L = 1
s3 = mvspec(u, span=3, taper=0, plot=F)$coh     # L = 3 and so on
s41 = mvspec(u, span=41, taper=0, plot=F)$coh
s101 = mvspec(u, span=101, taper=0, plot=F)$coh
plot(1:512/1024, s1, col=4, type="l", ylim=c(0,1), ylab="SqCoh", xlab="Frequency")
lines(1:512/1024, s3, col=3)
lines(1:512/1024, s41, col=2, lwd=2)
lines(1:512/1024, s101, col=1, lwd=2)
abline(h = .81/1.81, lty="dashed")
# for a single sq-coherency plot with confidence bands:
plot(mvspec(u, span=101, taper=0, plot=F), plot.type="coh")
```

4.26 (a) It follows from the solution to the previous problem that $\phi_{xy}(\omega) = 2\pi\omega D$. Hence the slope of the phase divided by 2π is the delay D.

(b) Bigger values of L give better estimates. The code is similar to the previous problem.

```
x = ts(rnorm(1025))
y = .9*lag(x,-1)+ rnorm(1025)
u = ts.intersect(x,y)
s1 = mvspec(u, taper=0, plot=F)$phase
s3 = mvspec(u, span=3, taper=0, plot=F)$phase
s41 = mvspec(u, span=41, taper=0, plot=F)$phase
s101 = mvspec(u, span=101, taper=0, plot=F)$phase
plot(1:512/1024, s1, col=4, type="l", ylim=c(-pi,pi), ylab="Phase", xlab="Frequency")
lines(1:512/1024, s3, col=3)
lines(1:512/1024, s41, col=2, lwd=2)
lines(1:512/1024, s101, col=1, lwd=2)
abline(a=0,b=2*pi, lty="dashed")
# for a single phase plot with confidence bands:
plot(mvspec(u, span=101, taper=0, plot=F), plot.type="phase")
```

4.27 (a) The R code for the cross-spectral analysis of the two series is below:

```
par(mfrow=c(3,1))
mvspec(prodn, span=7, taper=.5)
mvspec(unemp, span=7, taper=.5)
plot(mvspec(cbind(prodn,unemp), taper=.5, span=7, plot=FALSE), plot.type="coh")
```

See Figure 4.2 – Figure 4.5. The log spectra with $L = 7$, show substantial peaks at periods of 2.5 months, 3 months, 4 months and 6 months for the production spectrum and significant peaks at those periods plus a 12 month or one year periodicity in the unemployment spectrum. It is natural that the series tend to repeat yearly and quarterly so that 12 month and 3 month periods would be expected. The 6 month period could be winter-summer fluctuations or possibly a harmonic of the yearly cycle. The 4 month period could be a three cycle per year variation due to something less than quarterly variation or possibly a harmonic of the yearly cycle (recall harmonics of 1/12 are of the form $k/12$, for $k = 2, 3, 4, ...$). The squared coherence is large at the seasonal frequencies, as well as a low frequency of about 33 months, or three years, possibly due to a common low frequency business cycle. High coherence at a particular frequency indicates parallel movement between two series at the frequency, but not necessarily causality.

(b) The following code will plot the frequency response functions; see Figure 4.3.

```
w = seq(0,.5, length=1000)
par(mfrow=c(2,1))
FR12 = abs(1-exp(2i*12*pi*w))^2
plot(w, FR12, type="l", main="12th difference")
FR112 = abs(1-exp(2i*pi*w)-exp(2i*12*pi*w)+exp(2i*13*pi*w))^2
plot(w, FR112, type="l", main="1st diff and 12th diff")
```

The frequency response resulting from the application of the standard difference, followed by a seasonal difference shows that the low frequencies and seasonal frequencies should be attenuated, the low frequencies by the difference and the seasonal frequencies by the seasonal difference. The filtered series are plotted in Figure 4.3; the first difference obviously eliminates the trend but there are still regular seasonal peaks, gradually increasing over the length of the series. The final filtered series tends to eliminate the regular seasonal patterns and retain the intervening frequencies.

(c) As mentioned before, the filtered outputs are shown in Figure 4.4. Figure 4.5 shows the spectral analysis of the three series. The first shows the spectrum of the original production series with the low and seasonal frequency components. The second shows the spectrum of the differenced series and we see that the low frequency components have be attenuated while the seasonal component remain. The third shows the spectrum of the seasonally differenced differences, and the power at the seasonal components is essentially notched out. Economists would prefer a flatter response for the seasonally adjusted series and the design of

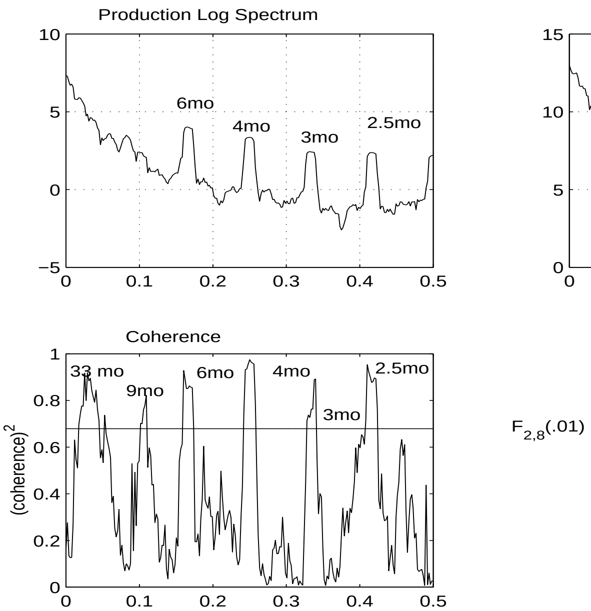

Fig. 4.2. Analysis of Production and Unemployment.

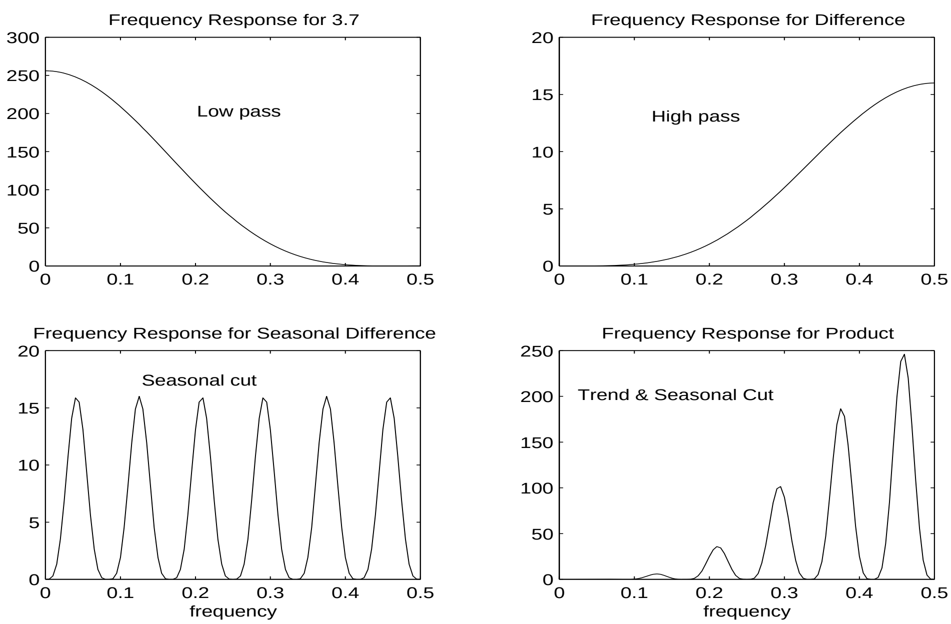

Fig. 4.3. Squared frequency response of various filters.

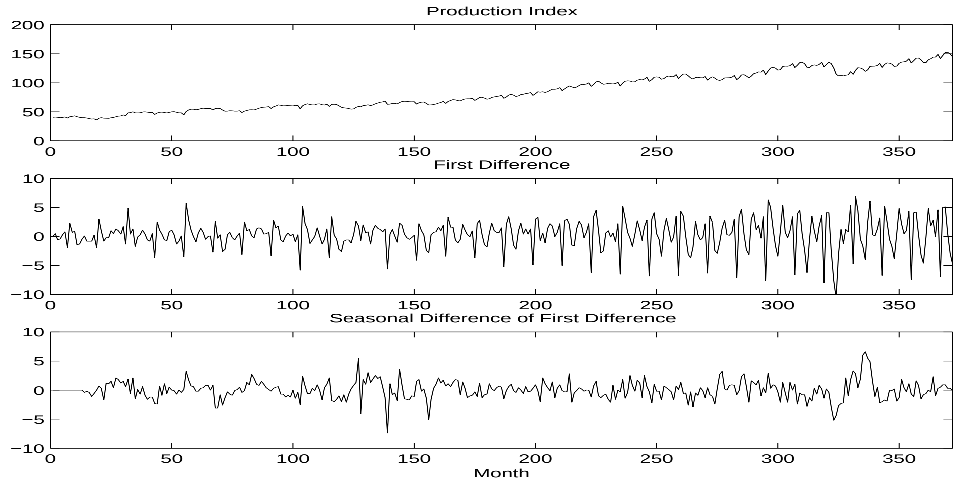

Fig. 4.4. Production and Filtered Series.

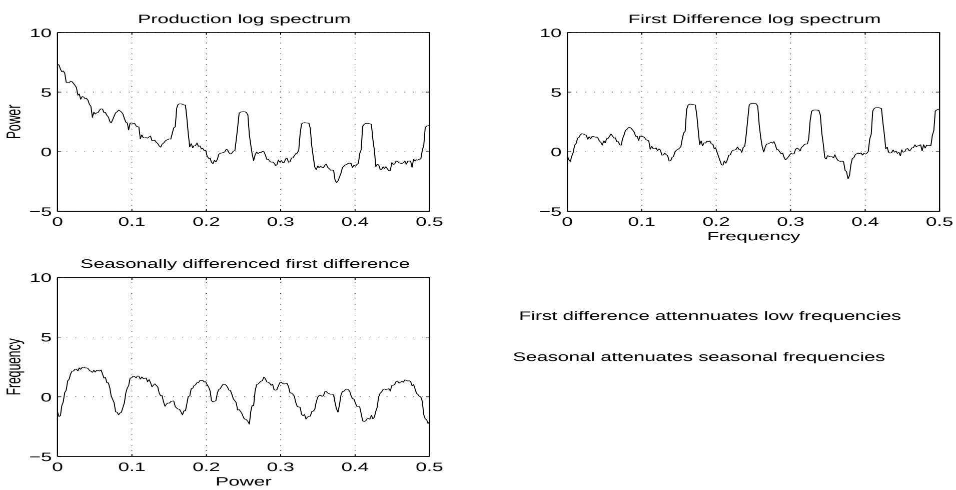

Fig. 4.5. Log spectra of production, differenced production and seasonally differenced production.

seasonal adjustment filters that maintain a flatter response is a continuing saga. Shumway (1988, Section 4.4.3) shows an example.

4.28 Write the filter in the general form (4.99), with $a_2 = a_{-2} = 1, a_1 = a_{-1} = 4, a_0 = 6$. Then

$$
\begin{aligned}
A(\omega) &= \sum_{t=-2}^{2} a_t \mathrm{e}^{-2\pi i\omega t} = (e^{-4\pi i\omega} + e^{-4\pi i\omega}) + 4(e^{-2\pi i\omega} + e^{-2\pi i\omega}) + 6 \\
&= (6 + 2\cos(4\pi\omega) + 8\cos(2\pi\omega))
\end{aligned}
$$

By Property 4.7, the spectrum of the output series is $f_y(\omega) = (6 + 2\cos(4\pi\omega) + 8\cos(2\pi\omega))^2 f_x(\omega)$. The spectrum of the output series will depend on the spectrum of the input series but we can see how frequencies of the input series are modified by plotting the squared frequency response function $|A(\omega)|^2$. After plotting the frequency response function, it will be obvious that the high frequencies are attenuated and the lower frequencies are not. The filter is referred to as a *low pass filter*, because it keeps or passes the low frequencies.

4.29

$$
\begin{aligned}
y_t &= \sum_{k=-\infty}^{\infty} a_k \cos[2\pi\omega(t-k)] = \frac{1}{2}\left[\mathrm{e}^{2\pi i\omega t}A(\omega) + \mathrm{e}^{-2\pi i\omega t}A^*(\omega)\right] = \mathrm{Re}\left\{A(\omega)\mathrm{e}^{2\pi i\omega t}\right\} \\
&= \mathrm{Re}\left\{(A_R(\omega) - iA_I(\omega))(\cos(2\pi\omega t) + i\sin(2\pi\omega t))\right\} = A_R(\omega)\cos(2\pi\omega t) + A_I\sin(2\pi\omega t) \\
&= |A(\omega)|\left[\cos(\phi(\omega))\cos(2\pi\omega t) - \sin(\phi(\omega))\sin(2\pi\omega t)\right] = |A(\omega)|\cos(2\pi\omega t + \phi(\omega)).
\end{aligned}
$$

The periodic cosine input is shifted by $\phi(\omega)$ and multiplied by $|A(\omega)| = \sqrt{A_R^2(\omega) + A_I^2(\omega)}$.

4.30 (a) From Property 4.7, $f_y(\omega) = |A(\omega)|^2 f_x(\omega)$, and $f_z(\omega) = |B(\omega)|^2 f_y(\omega) = |B(\omega)|^2|A(\omega)|^2 f_x(\omega)$.
(b) The frequency response functions of the first difference filter $|A(\omega)|^2$, the seasonal difference filter $|B(\omega)|^2$ and the product of the two are shown in Figure 3. Note that the first difference tends to keep the high frequencies and attenuate the low frequency and is an example of a *high-pass filter*. The seasonal difference tends to attenuate frequencies at the multiples $1/12, 2/12, \ldots 6/12$ which correspond to periods of $12, 6, 4, 3, 2$ months respectively. Frequencies in between are retained and the filter is an example of a *notch filter*, since it attenuates or notches out the seasonal frequencies.
(c) The product of the two filters tends to reject low frequency trends (the high-pass part) and seasonal frequencies (the notch part). Retaining the other frequencies is of interest to economists who seek *seasonal adjustment filters*. Better ones can be designed by specifying a frequency response for the high-pass part that rises more sharply.

4.31 (a) Using Property 4.7, $f_y(\omega) = [1 + a^2 - 2a\cos(2\pi\omega)]^{-1} f_x(\omega)$.
(b) Figure 4.6 plots the frequency response functions of both filters and we note that they are both *low-pass filters* with different rates of decrease. Recursive filters change the phase information and sometimes this can be important. Running the filters twice, once forward and once backward can fix this problem.

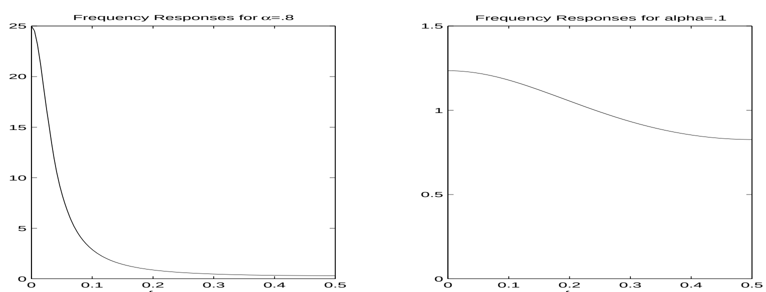

Fig. 4.6. Squared frequency response of recursive filters.

4.32 Note first that

$$
\begin{aligned}
a_k^M &= M^{-1}\sum_{j=0}^{M-1} A(\omega_j)\mathrm{e}^{2\pi i\omega_j k} = M^{-1}\sum_{j=0}^{M-1}\sum_{t=-\infty}^{\infty} a_t \mathrm{e}^{-2\pi i\omega_j t}\mathrm{e}^{2\pi i\omega_j k} \\
&= \sum_{t=-\infty}^{\infty} a_t M^{-1}\sum_{j=0}^{M-1}\mathrm{e}^{-2\pi i\omega_j(t-k)} = \sum_{\ell=-\infty}^{\infty} a_{k+\ell M} = a_k + \sum_{\ell\neq 0} a_{k+\ell M}.
\end{aligned}
$$

Thus

$$
y_t - y_t^M = \sum_{|k|\geq M/2} a_k x_{t-k} - \sum_{\ell\neq 0}\sum_{|k|<M/2} a_{k+\ell M}x_{t-k} \leq \sum_{|k|\geq M/2}|a_k x_{t-k}| + \sum_{\ell\neq 0}\sum_{|k|<M/2}|a_{k+\ell M}x_{t-k}|
$$

$$\le \sum_{|k|\ge M/2} |a_k x_{t-k}| + \sum_{|k|>M/2} |a_k x_{t-k}| \le 2 \sum_{|k|\ge M/2} |a_k x_{t-k}|,$$

where the last steps follow by writing the separate sums for $\ell = \pm 1, \pm 2, \dots$ and simplifying. Then

$$\begin{aligned}\mathrm{E}[(y_t - y_t^M)^2] &\le 4 \sum_{|j|\ge M/2} \sum_{|k|\ge M/2} |a_j||a_k|\mathrm{E}[|x_{t-j}||x_{t-k}|] \\ &\le 4 \sum_{|j|\ge M/2} \sum_{|k|\ge M/2} |a_j||a_k|\mathrm{E}^{1/2}[|x_{t-j}|^2]\mathrm{E}^{1/2}[|x_{t-k}|^2] \le 4\gamma_x(0)\Big(\sum_{|k|\ge M/2}|a_k|\Big)^2,\end{aligned}$$

which goes to zero as M increases as long as the absolute summability condition holds.

4.33 Multiply both sides of the equation by x_{t+h} an use the Fourier representation of the spectra and cross spectra to show that $f_{yx}(\omega) = A(\omega)f_x(\omega)$. Also, $f_y(\omega) = |A(\omega)|^2 f_x(\omega)$. Then, by the definition of squared coherence,

$$\rho^2_{y\cdot x}(\omega) = \frac{|f_{yx}(\omega)|^2}{f_x(\omega)f_y(\omega)} = \frac{|A(\omega)f_x(\omega)|^2}{f_x(\omega)|A(\omega)|^2 f_x(\omega)} = 1$$

4.34 (a) Using the following R code:

```
attach(climhyd)
inflow = log(Inflow) # log inflow
precip = sqrt(Precip) # sqrt precipitation
x = replace(climhyd, 5:6, c(precip,inflow))
u = mvspec(x, span=c(7,7), plot=FALSE)
plot(u, ci=-1, plot.type="coh")
dev.new()
LagReg(precip, inflow, threshold=.009)   # part (b)
detach(climhyd)
```

the squared coherencies of inflow with the other series will appear in the last row. It is clear that the precipitation-inflow coherence is uniformly larger than the others so that precipitation should be considered as the major contributor over the entire frequency range.

(b) Use a two-step procedure as specified in the example of the section, that is, first run `LagReg()` with no threshold, then run it again using a value near .009. Based on the .009 threshold, the predictions are extremely accurate. The estimated model is

```
  The positive lags, at which the coefficients are large
  in absolute value, and the coefficients themselves, are:
     lag s      beta(s)
[1,]     0  0.045197830
[2,]     1  0.022098634
[3,]     2  0.019057057
[4,]     3  0.009269450
[5,]    13 -0.009137882
```

4.35 For the model $y_t = \sum_{r=-\infty}^{\infty} \beta_r x_{t-r} + v_t$, first show that

$$\gamma_y(h) = \sum_{r=-\infty}^{\infty}\sum_{s=-\infty}^{\infty} \beta_r \gamma_x(h - r + s)\beta_s + \gamma_v(h).$$

Substituting the spectral representations for $\gamma_x(h)$ and $\gamma_v(h)$ and identifying the representations for $f_x(\omega)$, $f_v(\omega)$ as well as the Fourier series for β_t, leads to the first required result. Then, note that

$$\gamma_{xy}(h) = \sum_{r=-\infty}^{\infty} \beta_r \gamma_x(h + r) + \gamma_v(h)$$

and substitute again. That is, since x_t and v_t are uncorrelated,

$$\gamma_{xy}(h) = \int_{-1/2}^{1/2} \sum_{r=-\infty}^{\infty} \beta_r e^{2\pi i\omega r} f_x(\omega) e^{2\pi i\omega h}\, d\omega = \int_{-1/2}^{1/2} \overline{B(\omega)} f_x(\omega) e^{2\pi i\omega h}\, d\omega.$$

4.36 Rewrite the first equation as

$$\begin{aligned} y_t - \phi y_{t-1} &= x_t - \phi x_{t-1} + v_t - \phi v_{t-1} \\ &= w_t + v_t - \phi v_{t-1}, \end{aligned}$$

and identify the righthand side as a first-order MA. The required spectrum is the spectrum of an ARMA$(1,1)$ process, which also has an first order MA on the righthand side. Assuming that the process is Gaussian, we can equate the ACF's of the two processes and solve for σ^2 and θ. Letting $u_t = w_t + v_t - \phi v_{t-1}$, we obtain

$$\gamma_u(0) = \sigma_w^2 + (1+\phi^2)\sigma_v^2$$

and $\gamma_u(1) = -\phi\sigma_v^2$. Equating these results to the corresponding results for a first-order MA leads to the equations

$$\sigma^2(1+\theta^2) = \sigma_w^2(1+\phi^2)\sigma_v^2 \quad \text{and} \quad -\theta\sigma^2 = -\phi\sigma_v^2$$

relating the two models. Solving the second and substituting back into the first equation leads to the required results.

4.37 (a) Set up the orthogonality condition

$$E[(x_t - \sum_{s=-\infty}^{\infty} a_s y_{t-s}) y_{t-u}] = 0, u = 0, \pm 1, \ldots,$$

which leads to the normal equations $\sum_{s=-\infty}^{\infty} a_s \gamma_y(u-s) = \gamma_{xy}(u)$. Noting that $\gamma_{xy}(u) = \gamma_x(u)$ and taking Fourier transforms leads to the equation

$$A(\omega) = \frac{f_x(\omega)}{f_y(\omega)} = \frac{\sigma_w^2}{\sigma^2} \frac{1}{|1-\phi e^{-2\pi i\omega}|^2} \frac{|1-\phi e^{-2\pi i\omega}|^2}{|1-\theta e^{-2\pi i\omega}|^2} = \frac{\sigma_w^2}{\sigma^2} \frac{1}{|1-\theta e^{-2\pi i\omega}|^2},$$

which we recognize as the spectrum of a first-order AR process, with variance σ_w^2/σ^2. Hence its Fourier transform will be the autocovariance of a first-order AR process, i.e.

$$a_s = \frac{\sigma_w^2}{\sigma^2} \frac{\theta^{|s|}}{1-\theta^2}$$

for the optimal filter.

(b) The mean squared error will be

$$E[(x - \hat{x}_t)x_t] = E[(x_t - \sum_{s=-\infty}^{\infty} a_s y_{t-s}) x_t] = \gamma_x(0) - \sum_{s=-\infty}^{\infty} a_s \gamma_x(s)$$

$$= \frac{\sigma_w^2}{1-\phi^2} - \frac{\sigma_w^4}{\sigma^2} \frac{1}{1-\theta^2} \frac{1}{1-\phi^2} \sum_{s=-\infty}^{\infty} (\theta\phi)^{|s|} = \frac{\sigma_w^2}{1-\phi^2)} \left[1 - \frac{\sigma_w^2}{\sigma^2} \frac{1}{1-\theta^2} \frac{1+\phi\theta}{1-\phi\theta}\right] = \frac{\sigma_v^2 \sigma_w^2}{\sigma^2(1-\theta^2)}.$$

(c) To get the optimal finite estimator, use the orthogonality principle to get the equation

$$\begin{pmatrix} \gamma_y(0) & \gamma_y(1) \\ \gamma_y(1) & \gamma_y(0) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \gamma_x(1) \\ \gamma_x(2) \end{pmatrix}$$

We obtain $\gamma_y(0), \gamma_y(1)$ from Example 3.11 in Chapter 3, which derives the autocovariance of an $ARMA(1,1)$ process and $\gamma_x(1), \gamma_x(2)$ from Problem 3.5(b). This leads to the equation

$$\begin{pmatrix} .9583 & .8147 \\ .8147 & .9584 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} .8147 \\ .7333 \end{pmatrix}$$

and the solution $\boldsymbol{a} = (.7204, .1527)'$. The mean squared error can be computed from

$$\begin{aligned} \text{MSE} &= \text{E}[(x_t - a_1 y_{t-1} - a_2 y_{t-2}) x_t] \\ &= \gamma_x(0) - a_1\gamma_x(1) - a_2\gamma_x(2) \\ &= .2064. \end{aligned}$$

The optimal mean squared error from the equation in part (b) is .0364.

4.38 (a) Write

$$\begin{aligned} \gamma_y(h_1,h_2) &= \text{E}\Big[\sum_{u_1,u_2,v_1,v_2} a_{u_1,u_2} a_{v_1,v_2} x_{s_1+h_1-u_1, s_2+h_2-u_2} x_{s_1-v_1,s_2-v_2}\Big] \\ &= \sum_{u_1,u_2,v_1,v_2} a_{u_1,u_2} \gamma_x(h_1 - u_1 + v_1, h_2 - u_2 + v_2) a_{v_1,v_2} \end{aligned}$$

(b) Substitute in the autocovariance in (a) the spectral representation, so that

$$\begin{aligned} \gamma_y(h_1,h_2) &= \int_{-1/2}^{1/2} \Big[\sum_{u_1,u_2} a_{u_1,u_2} \text{e}^{-2\pi i(u_1\omega_1 + u_2\omega_2)} \sum_{v_1,v_2} a_{v_1,v_2} \text{e}^{2\pi i(v_1\omega_1+v_2\omega_2)} \\ &\quad \times f_x(\omega_1\omega_2) \text{e}^{2\pi i(\omega_1 h_1 + \omega_2 h_2)}\Big]\, d\omega_1 d\omega_2 \\ &= \int_{-1/2}^{1/2} |A(\omega_1,\omega_2)|^2 f_x(\omega_1,\omega_2) \text{e}^{2\pi i(\omega_1 h_1 + \omega_2 h_2)}\, d\omega_1 d\omega_2. \end{aligned}$$

4.39 We can use the definition (C.13) for $S_n(\omega_k, \omega_\ell)$ in the white noise case [$\gamma(s-t) = \sigma_w^2$, for $s = t$ and 0 otherwise] to write

$$S_n(\omega_k,\omega_\ell) = n^{-1}\sigma_w^2 \sum_{t=1}^{n} \text{e}^{-2\pi i(\omega_k - \omega_\ell)t},$$

which we know to be 1 when $\omega_k - \omega_\ell = 0, \pm n, \pm 2n \dots$ and 0 otherwise. Then, for example,

$$\begin{aligned} \text{E}[d_c(\omega_k) d_c(\omega_\ell)] &= \frac{1}{4}\big[S_n(-\omega_k,\omega_\ell) + S_n(\omega_k,\omega_\ell) \\ &\quad + S_n(\omega_\ell,\omega_k) + S_n(\omega_k,-\omega_\ell)\big] \\ &= \frac{\sigma_w^2}{4}[0+1+1+0] \\ &= \frac{\sigma_w^2}{2}, \end{aligned}$$

for $\omega_k = \omega_\ell$ and is zero otherwise. The other terms are treated similarly.

4.43 Write

$$\begin{aligned} w &= 2\text{Re}[\boldsymbol{a}_c - i\boldsymbol{a}_s)^*(\boldsymbol{x}_c - i x_s)] \\ &= 2(\boldsymbol{a}_c'\boldsymbol{x}_c + \boldsymbol{a}_s'\boldsymbol{x}_s) \\ &= 2\begin{pmatrix} \boldsymbol{a}_c' & \boldsymbol{a}_s' \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_c \\ \boldsymbol{x}_s \end{pmatrix}. \end{aligned}$$

Hence,

$$\begin{aligned}
\operatorname{cov} z &= 4\,(\,\boldsymbol{a}_c' \quad \boldsymbol{a}_s'\,)\,\frac{1}{2}\begin{pmatrix} C & -Q \\ Q & C \end{pmatrix}\begin{pmatrix} \boldsymbol{a}_c \\ \boldsymbol{a}_s \end{pmatrix} \\
&= 2(\boldsymbol{a}_c' + i\boldsymbol{a}_s')(C - iQ)(\boldsymbol{a}_c - i\boldsymbol{a}_s) \\
&= \boldsymbol{a}^* \Sigma \boldsymbol{a},
\end{aligned}$$

where the last step follows because $\Sigma = \Sigma^*$.

Chapter 5

Solutions

5.1 (a) A time plot is shown below. Note the apparent trend in the data.

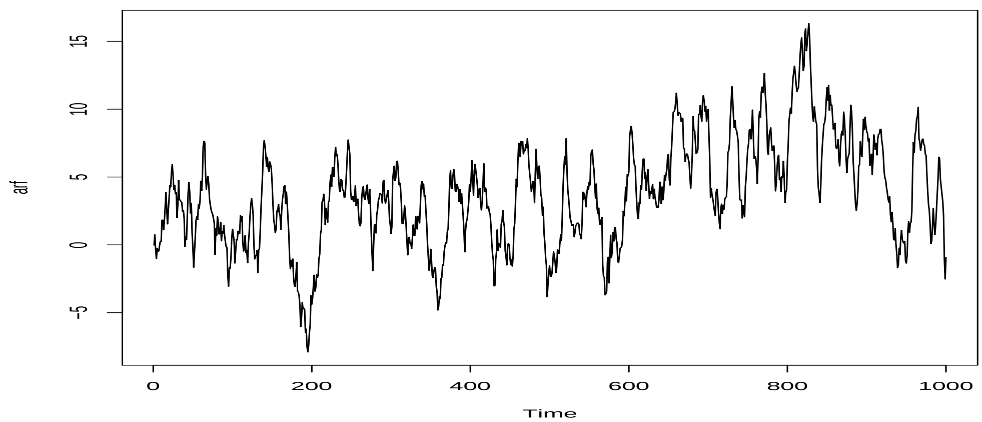

Fig. 5.1. Plot of data for Problem 5.1(a)

(b) Shown below. ACF dampens slowly indicating long memory, but the PACF lag-1 indicates possible random walk.

(c) R code and partial output below (must load `fracdiff` package):

```
plot(arf)     # part (a)
acf2(arf)     # part (b)
library(fracdiff)  # part (c)
summary(fracdiff(arf,nar=1))
 Coefficients:
   Estimate Std. Error z value  p.value
d  0.264631   0.009653   27.41   <2e-16
ar 0.863068   0.016921   51.01   <2e-16
```

(d) The time plot in (a) and the ACF and PACF in (b) suggest nonstationary behavior typically found with a random walk; i.e. slow trend in data, ACF tails slowly, PACF lag-1 near one.

(e) A plot of the ACF and PACF [`acf2(diff(arf))`] suggests stationarity is achieved after differencing. They indicate an AR(1) or an MA(1) with a small coefficient.

(f) Using `sarima(arf,1,1,0)` to fit an ARI(1,1) model fits the data well, with a significant `ar1` coefficient of about .17.

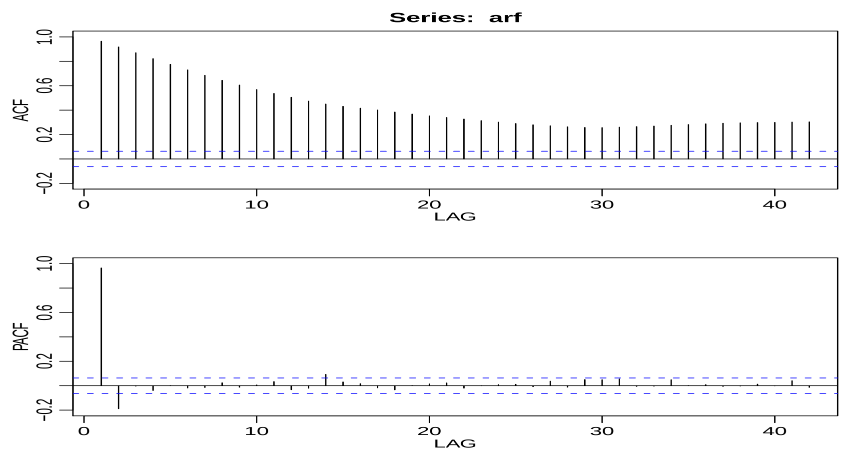

Fig. 5.2. ACF and PACF for Problem 5.1(b)

5.2 R code to fit an ARFIMA(1,1,1) and discussion as comments:

```
acf(abs(nyse), 200)                         # indicates long memory
u = fracdiff(abs(nyse))                     # initial analysis: d = .18
res = diffseries(abs(nyse), u$d)            #  get frac diffed series
acf2(res)                                   # indicates ARMA(1,1) with small coefs
summary(fracdiff(abs(nyse), nar=1, nma=1))  # full estimation
      Estimate Std. Error z value  p.value
  d   0.310079   0.016432  18.870   <2e-16
  ar -0.068736   0.006991  -9.831   <2e-16
  ma  0.166071   0.007021  23.652   <2e-16
```

5.3 The ADF gives a completely different result than DF and PP, which both reject the unit root hypothesis. It's unclear why the difference, although it seems `adf.test` does different things when $k = 0$ and when $k > 0$.

```
plot(globtemp)
library("tseries")
adf.test(globtemp, k=0)
        Augmented Dickey-Fuller Test
data:  globtemp
Dickey-Fuller = -4.0442, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary
Warning message:
In adf.test(globtemp, k = 0) : p-value smaller than printed p-value

adf.test(globtemp)
        Augmented Dickey-Fuller Test
data:  globtemp
Dickey-Fuller = -1.6219, Lag order = 5, p-value = 0.7338
alternative hypothesis: stationary

pp.test(globtemp)
        Phillips-Perron Unit Root Test
data:  globtemp
Dickey-Fuller Z(alpha) = -25.062, Truncation lag parameter = 4, p-value = 0.02003
alternative hypothesis: stationary
```

5.4 All tests reject unit root hyp in favor of explosive.

```
plot(gnp)
```

```
library("tseries")
adf.test(gnp, k=0, alt="e")  # DF  = 0.0696, Lag order = 0, p-value = 0.01  | alt hyp:
    explosive
adf.test(gnp, alt="e")       # ADF = -0.295, Lag order = 6, p-value = 0.01 | ...
pp.test(gnp, alt="e")        # PP  = 0.742, Truncation lag parameter = 4, p-value = 0.01 |
    ...
```

5.5 Follows almost directly from the AR(2) example just below (5.33): Rearrange the model as in the AR(2) example and then subtract x_{t-1} from both sides.

5.6 The time plot of the oil returns indicate GARCH behavior. The [P]ACF indicate an ARMA(1,1) for the mean. The [P]ACF of the squared residuals indicate a GARCH(1,1) with long tailed error distribution. Below is code using a t-distribution for the error.

```
poil=diff(log(oil))
plot(poil)
sarima(gnpgr, 1, 0, 1) # fit an ARMA(1,1)
acf2(innov^2, 24) # get (p)acf of the squared residuals
# innov^2 could be ARMA(1,1) again
library(fGarch)
summary(fit<-garchFit(~arma(1,1)+garch(1,1), cond.dist= "std", data=poil))
```

5.9 Following the advice of many authors [see for example Thanoon (1990), *J. Time Series Anal.*, 75-87] we fit a long AR, to the data with threshold $x_{t-3} > 36.6$. [This information might be related to the students. Anything between 36 and 37 will work.] We found an AR(15) worked well, the residuals appeared to be white, but were somewhat heteroscedastic. The following models were fit:

$$x_t = \alpha^{(1)} + \sum_{j=1}^{15} \phi_j^{(1)} x_{t-j} + w_t^{(1)}, \quad x_{t-3} \le 36.6$$

$$x_t = \alpha^{(2)} + \sum_{j=1}^{15} \phi_j^{(2)} x_{t-j} + w_t^{(2)}, \quad x_{t-3} > 36.6,$$

You can use `tsDyn` but it's a little quirky.

```
library(tsDyn)
vignette("tsDyn")    # for package details
(u = setar(sunspotz, m=15, thDelay=2)) # thDelay=2 is lag 3 (thDelay=0 is lag 1 - I know,
    it's strange)
plot(u)              # for all the plots
```

5.10 (a) We transform inflow and take the seasonal difference $y_t = \ln i_t - \ln i_{t-12}$ which is proportional to the percentage yearly increase in flow. Monthly precipitation has some zero values and we use the square root transformation to stabilize this variable. Fitting to two series separately leads to the two ARIMA models

$$x_t = \nabla_{12} P_t = (1 - .812(.029)B^{12})w_t$$

and

$$y_t = (1 - .764(.033)B^{12})z_t,$$

with $\hat\sigma_w^2 = 32.503$ and $\hat\sigma_z^2 = .225$.

(b) Cross correlating the two transformed series x_t and $(1 - .812B^{12})y_t$ leads to the figure shown below and we note that the inflow series seems to depend on exponentially decreasing lagged values of the precipitation.

5.11 (a) The ACF of the residuals from the ARIMA$(0,0,0)\times(0,1,1)_{12}$ model is well-behaved with all values (except lag 1) well below the significance levels.

(b) The CCF has been computed in Problem 5.11 and is shown in Figure 5.3. The exponential decrease, beginning at lag zero, suggests

$$\alpha(B) = \delta_0(1 + \omega_1 B + \omega_1^2 B^2 + \omega_1^3 B^3 + \ldots)$$

for fitting the exponential decrease.

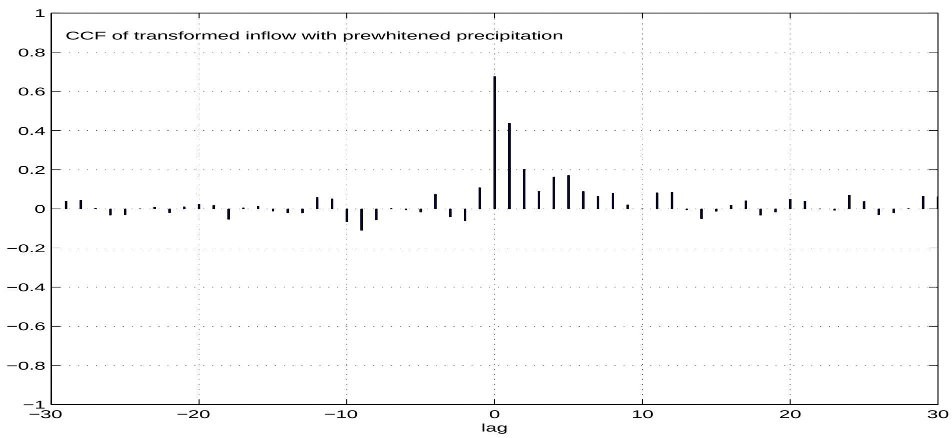

Fig. 5.3. CCF for Problem 5.12

(c) So far, we have

$$y_t = \frac{\delta}{1 - \omega_1 B} x_t + \eta_t$$

which becomes

$$(1 - \omega_1 B) y_t = \delta x_t + (1 - \omega_1 B)\eta_t,$$

or

$$y_t = \omega_1 y_{t-1} + \delta_0 x_t + n_t,$$

where

$$n_t = (1 - \omega_1 B)\eta_t.$$

We can run the regression model above as ordinary least squares, even though the residuals are correlated, obtaining

$$y_t = .526(.026) y_{t-1} + .050(.002) x_t + n_t$$

(d) To model the noise, we take $\hat{n}_t$ from the above model and note that

$$\hat{n}_t = (1 - .526B)\eta_t$$

can be solved for η_t by inverting the first order moving average transformation. These residuals, say $\hat{\eta}_t$ can be modeled by an $ARIMA(1,0,0) \times (0,0,1)_{12}$ model of the form

$$(1 - .384B)\eta_t = (1 - .796B^{12}) z_t,$$

where $\hat{\sigma}_z^2 = .0630$. Hence, the final model is of the form

$$y_t = \frac{.050}{1 - .526B} x_t + \eta_t,$$

where

$$x_t = (1 - .812B^{12}) w_t$$

and the noise η_t is as modeled above.

(e) A possible general procedure would be to forecast x_t and η_t separately and then combine the forecasts using the defining equation above (note that x_t and η_t are assumed to be independent). To forecast x_t, note that

$$x_t = w_t - .812 w_{t-12}$$

and

$$x_{t+m} = w_{t+m} - .812 w_{t+m-12}$$

and the forecast would be

$$\tilde{x}_{t+m} = \begin{cases} -.812 w_{t+m-12} & m \le 12 \\ 0 & m > 12, \end{cases}$$

where the residuals w_t come from applying the model to the data x_t. The forecast variance for $\tilde{x}_{t+m}$ will be σ_w^2 for $m \le 12$ and $(1 + .812^2)\sigma_w^2$ for $m > 12$. To forecast η_t, note that

$$\eta_{t+m} = .384 \eta_{t+m-1} + z_{t+m} - .795 z_{t+m-12},$$

so that

$$\tilde{\eta}_{t+m} = \begin{cases} .384 \tilde{\eta}_{t+m-1} - .795 z_{t+m-12} & m \le 12 \\ .384 \tilde{\eta}_{t+m-1} & m > 12, \end{cases}$$

where $\tilde{\eta}_t = \eta_t$. In this case, solving the defining equation in terms η_t yields

$$\eta_t = \frac{1 - .796 B^{12}}{1 - .384 B} z_t = \psi(B) z_t.$$

The contribution of this term to the forecast variance will then be

$$E[(\eta_{t+m} - \tilde{\eta}_{t+m})^2] = \sigma_z^2 \sum_{j=0}^{m-1} \psi_j^2.$$

5.12 Below is the R code. Note that there is still zero-order correlation among the series, so a model that contains zero-order regressors may have done better (this is entirely reasonable because the data are quarterly, and hence the values evolve over months).

```
x = log(econ5[,1:3])
library(vars)
VARselect(x, lag.max=10, type="both")    # suggests an order 2 or 3 model
summary(fit <- VAR(x, p=2, type="both"))
(fit.pr = predict(fit, n.ahead = 24, ci = 0.95)) # 4 weeks ahead
fanchart(fit.pr) # plot prediction + error
```

Chapter 6

Solutions

For the most part, we stopped using bold face for vectors in the text. However, in the solutions, the bold face remains because we're too lazy to fix it.

6.1 (a)

$$\begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} = \begin{pmatrix} 0 & -.9 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ x_{t-2} \end{pmatrix} + \begin{pmatrix} w_t \\ 0 \end{pmatrix}$$

and

$$y_t = [1 \;\; 0] \begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} + v_t$$

(b) For y_t to be stationary, x_t must be stationary. Note that for $t = 0, 1, 2, ...,$ we may write $x_{2t-1} = \sum_{j=0}^{t-1}(-.9)^j w_{2t-1-2j} + (-.9)^t x_{-1}$ and $x_{2t} = \sum_{j=0}^{t-1}(-.9)^j w_{2t-2j} + (-.9)^t x_0$. From this we see that x_{2t-1} and x_{2t} are independent. Repeating the steps of Problem 3.5, we conclude that setting $x_0 = w_0/\sqrt{1-.9^2}$ and $x_{-1} = w_{-1}/\sqrt{1-.9^2}$ will make x_t stationary. In other words, set $\sigma_0^2 = \sigma_1^2 = \sigma_w^2/(1-.9^2)$.

(c) and (d): The plots are not shown here.

6.2 (i) $s \neq t$: Without loss of generality, let $s < t$, then $\text{cov}(\epsilon_s, \epsilon_t) = \text{E}[\epsilon_s \text{E}(\epsilon_t \mid y_1, ..., y_s)] = 0$.
(ii) $s = t$: Note that $y_t^{t-1} = \text{E}(x_t + v_t \mid y_1, ..., y_{t-1}) = x_t^{t-1}$. Thus $\epsilon_t = y_t - y_t^{t-1} = (x_t - x_t^{t-1}) + v_t$, and it follows that $\text{var}(\epsilon_t) = \text{var}[(x_t - x_t^{t-1}) + v_t] = P_t^{t-1} + \sigma_v^2$.

6.3 This problem is similar to Example 6.3 and Example 6.6 [minus the estimation part].

6.4 (a) Write $\boldsymbol{x} = (x_1, ..., x_p)'$, $\boldsymbol{y} = (y_1, ..., y_q)'$, $\boldsymbol{b} = (b_1, ..., b_p)'$ and $B = \{B_{ij}\}_{i=1,...,p; j=1,...,q}$. The projection equations are

$$\text{E}[(x_i - b_i - B_{i1}y_1 - \cdots - B_{iq}y_q)\; 1] = 0, \qquad i = 1, ..., p; \tag{1}$$

$$\text{E}[(x_i - b_i - B_{i1}y_1 - \cdots - B_{iq}y_q)\; y_j] = 0, \qquad i = 1, ..., p, \quad j = 1, ..., q. \tag{2}$$

In matrix notation, the p equations in (1) are $\text{E}(\boldsymbol{x} - \boldsymbol{b} - B\boldsymbol{y}) = \boldsymbol{0}$ and the pq equations in (2) are $\text{E}(\boldsymbol{xy}' - \boldsymbol{by}' - B\boldsymbol{yy}') = 0$, as was to be shown. Solving (1) leads to the solution for $\boldsymbol{b}$; that is $\boldsymbol{b} = \text{E}(\boldsymbol{x}) - B\text{E}(\boldsymbol{y})$. Inserting this solution into (2) and then solving (2) leads to the solution for B; that is, $\text{E}(\boldsymbol{xy}') - \boldsymbol{\mu}_x\boldsymbol{\mu}_y' = B[\text{E}(\boldsymbol{yy}') - \boldsymbol{\mu}_y\boldsymbol{\mu}_y']$ or $B = \Sigma_{xy}\Sigma_{yy}^{-1}$.

(b) Let $\widehat{\boldsymbol{x}} = P_{\mathcal{M}}\boldsymbol{x} = \boldsymbol{\mu}_x + \Sigma_{xy}\Sigma_{yy}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y)$ as given in (a). The MSE matrix is $\text{E}(\boldsymbol{x} - P_{\mathcal{M}}\boldsymbol{x})(\boldsymbol{x} - P_{\mathcal{M}}\boldsymbol{x})' = \text{E}[(\boldsymbol{x} - P_{\mathcal{M}}\boldsymbol{x})\boldsymbol{x}']$ because $\boldsymbol{x} - P_{\mathcal{M}}\boldsymbol{x} \perp \mathcal{M}$ and $P_{\mathcal{M}}\boldsymbol{x} \in \mathcal{M}$. Thus,

$$MSE = \text{E}[(\boldsymbol{x} - \boldsymbol{\mu}_x)\boldsymbol{x}'] - \Sigma_{xy}\Sigma_{yy}^{-1}\text{E}[(\boldsymbol{y} - \boldsymbol{\mu}_y)\boldsymbol{x}'] = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx},$$

noting, for example, that $\text{E}[(\boldsymbol{x} - \boldsymbol{\mu}_x)(\boldsymbol{x} - \boldsymbol{\mu}_x)'] = \text{E}[(\boldsymbol{x} - \boldsymbol{\mu}_x)\boldsymbol{x}']$.

(c) Consider writing the equation preceding (6.27) in terms of this question. That is,

$$\begin{pmatrix} \boldsymbol{x} = \boldsymbol{x}_t \\ \boldsymbol{y} = \boldsymbol{\epsilon}_t \end{pmatrix} \Big| \, Y_{t-1} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{\mu}_x = \boldsymbol{x}_t^{t-1} \\ \boldsymbol{\mu}_y = \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} = P_t^{t-1} & \Sigma_{xy} = P_t^{t-1} A_t' \\ \Sigma_{yx} = A_t P_t^{t-1} & \Sigma_{yy} = \Sigma_t \end{bmatrix} \right).$$

Then using (a), $\widehat{\boldsymbol{x}} = \boldsymbol{\mu}_x + \Sigma_{xy}\Sigma_{yy}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y)$ corresponds to $\boldsymbol{x}_t^t = \boldsymbol{x}_t^{t-1} + P_t^{t-1}A_t'\Sigma_t^{-1}\boldsymbol{\epsilon}_t$, which is precisely (6.21). Moreover, from (b), the MSE is $\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}$ which corresponds to $P_t^{t-1} - P_t^{t-1}A_t'\Sigma_t^{-1}A_tP_t^{t-1}$, which is precisely P_t^t defined in (6.22). Thus the normal theory and the projection theorem results coincide.

6.5

(a) Because $\boldsymbol{y}_{k+1} - \boldsymbol{y}_{k+1}^k \in \mathcal{L}_{k+1}$, it suffices to show that $\boldsymbol{y}_{k+1} - \boldsymbol{y}_{k+1}^k \perp \mathcal{L}_k$. But

$$\mathrm{E}[\boldsymbol{y}_j(\boldsymbol{y}_{k+1} - \boldsymbol{y}_{k+1}^k)'] = \mathrm{E}\{\boldsymbol{y}_j \mathrm{E}[(\boldsymbol{y}_{k+1} - \boldsymbol{y}_{k+1}^k)' \mid \boldsymbol{y}_1, ..., \boldsymbol{y}_k]\} = 0 \quad j = 1, 2, \ldots, k,$$

as required.

(b) From the problem statement, we have

$$H_{k+1} = \mathrm{E}[\boldsymbol{x}_k(\boldsymbol{y}_{k+1} - \boldsymbol{y}_{k+1}^k)'] \left\{\mathrm{E}[(\boldsymbol{y}_{k+1} - \boldsymbol{y}_{k+1}^k)(\boldsymbol{y}_{k+1} - \boldsymbol{y}_{k+1}^k)]'\right\}^{-1}.$$

Now $\boldsymbol{y}_{k+1} - \boldsymbol{y}_{k+1}^k = A_{k+1}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_{k+1}^k) + \boldsymbol{v}_{k+1}$. From this it follows immediately that

$$\mathrm{E}[(\boldsymbol{y}_{k+1} - \boldsymbol{y}_{k+1}^k)(\boldsymbol{y}_{k+1} - \boldsymbol{y}_{k+1}^k)]' = A_{k+1}P_{k+1}^k A_{k+1} + R.$$

To complete the result write ($\boldsymbol{x}_k \perp \boldsymbol{v}_{k+1}$)

$$\mathrm{E}[\boldsymbol{x}_k(\boldsymbol{y}_{k+1} - \boldsymbol{y}_{k+1}^k)'] = \mathrm{E}[\boldsymbol{x}_k(\boldsymbol{x}_{k+1} - \boldsymbol{x}_{k+1}^k)'A_{k+1}'] = \mathrm{E}[\boldsymbol{x}_k(\boldsymbol{x}_k - \boldsymbol{x}_k^{k-1})'\Phi'A_{k+1}'] = P_k^k\Phi'A_{k+1}'.$$

(c) Using (6.23),

$$A_{k+1}'[A_{k+1}P_{k+1}^kA_{k+1} + R]^{-1} = [P_{k+1}^k]^{-1}K_{k+1}.$$

In addition, using (6.21),

$$K_{k+1}(\boldsymbol{y}_{k+1} - \boldsymbol{y}_{k+1}^k) = \boldsymbol{x}_{k+1}^{k+1} - \boldsymbol{x}_{k+1}^k.$$

From these two facts we find that

$$H_{k+1}(\boldsymbol{y}_{k+1} - \boldsymbol{y}_{k+1}^k) = P_k^k\Phi'[P_{k+1}^k]^{-1}K_{k+1}(\boldsymbol{y}_{k+1} - \boldsymbol{y}_{k+1}^k) = J_k(\boldsymbol{x}_{k+1}^{k+1} - \boldsymbol{x}_{k+1}^k),$$

and the result follows.

(d) and (e) The remainder of the problem follows in a similar manner.

6.6 (a) First write $\nabla y_t = y_t - y_{t-1} = w_t + v_t - v_{t-1}$. Then, $\mathrm{E}(\nabla y_t) = 0$, $\mathrm{var}(\nabla y_t) = \sigma_w^2 + 2\sigma_v^2$, and $\mathrm{cov}(\nabla y_t, \nabla y_{t-h}) = -\sigma_v^2$ for $h = \pm 1$ and 0 for $|h| > 1$. We conclude that ∇y_t is an MA(1) with ACF given by $\rho(h) = \frac{-\sigma_v^2}{\sigma_w^2 + 2\sigma_v^2}\delta_1^h$ for $h = 1, 2, \ldots$. Note, $|\rho(1)| \le .5$ for all values of $\sigma_w^2 \ge 0$ and $\sigma_v^2 > 0$.

(b) The estimates should be about $\widehat{\sigma}_w = .111$ and $\widehat{\sigma}_v = .425$. With these estimates $\widehat{\rho}(1) = -.483$, and hence $\widehat{\theta} = -.77$. These values are close to the values found in Ch. 3. R code:

```
y=log(varve)
num = length(y)
mu0=y[1]
Sigma0=var(y[1:10])
# likelihood
Linn=function(para){
  cQ = para[1]
  cR= para[2]
  kf = Kfilter0(num, y, 1, mu0, Sigma0, 1, cQ, cR)
```

```
  return(kf$like)
  }
init.par=c(.1,.1)
(est = optim(init.par, Linn, NULL, method='BFGS', hessian=TRUE,
    control=list(trace=1,REPORT=1)))
SE = sqrt(diag(solve(est$hessian)))
# Summary of estimation
estimate = c(sig.w=est$par[1], sig.v=est$par[2])
cbind(estimate, SE)
```

6.7 Using (6.64), the essential part of the complete log-likelihood (i.e. dropping any constants) is

$$\ln|\sigma_x^2| + \sum_{t=1}^{n} \frac{x_t^2}{r_t\sigma_x^2} + \ln|\sigma_v^2| + \sum_{t=1}^{n} \frac{(y_t - x_t)^2}{\sigma_v^2}.$$

Following (6.71) and (6.72), the updated estimates will be (using the notation of the problem, ˆ for updates and ˜ for current values)

$$\hat\sigma_x^2 = n^{-1}\sum_{t=1}^{n} \frac{[\tilde x_t^n]^2 + \tilde P_t^n}{r_t} \quad \text{and} \quad \hat\sigma_v^2 = n^{-1}\sum_{t=1}^{n}[(y_t - \tilde x_t^n)^2 + \tilde P_t^n].$$

It remains to determine $\tilde x_t^n$ and $\tilde P_t^n$. These can be obtained from (B.9)–(B.10). Write $X_n = (x_1, ..., x_n)'$ and $Y_n = (y_1, ..., y_n)'$ and drop the ˜ from the notation. Then

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \sim \mathrm{N}\left(\mathbf{0}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}\right),$$

where $\Sigma_{xx} = \mathrm{diag}\,\{r_1\sigma_x^2, \dots, r_n\sigma_x^2\}$, $\Sigma_{xy} = \Sigma_{xx}$ [because $\mathrm{E}(x_t y_t) = \mathrm{E}(x_t^2 + x_t v_t) = \mathrm{E}(x_t^2)$] and $\Sigma_{yy} = \mathrm{diag}\,\{\sigma_v^2, \dots, \sigma_v^2\}$. Using (B.9)–(B.10) it follows that

$$x_t^n = \mathrm{E}(x_t|Y_n) = \frac{r_t\sigma_x^2}{r_t\sigma_x^2 + \sigma_v^2} y_t \quad \text{and} \quad P_t^n = \mathrm{var}(x_t|Y_n) = r_t\sigma_x^2 - \frac{r_t^2\sigma_x^4}{r_t\sigma_x^2 + \sigma_v^2} = \frac{r_t\sigma_x^2\sigma_v^2}{r_t\sigma_x^2 + \sigma_v^2}.$$

The stated results now follow.

6.8 (a) Using Property 6.1, $P_t^{t-1} = \phi^2[1 - K_{t-1}]P_{t-1}^{t-2} + Q = \phi^2([P_{t-1}^{t-2}]^{-1} + R^{-1})^{-1} + Q$. Note $R = \sigma_v^2$ and $Q = \sigma_w^2$

(b) To ease the notation, we write P_t^{t-1} as P_t. Part (a) is then $P_t = \phi^2(P_{t-1}^{-1} + R^{-1})^{-1} + Q$. Using this relationship yields

$$P_t^{-1} - P_{t-1}^{-1} = \phi^2[P_{t-1}^{-1} - P_{t-2}^{-1}]. \tag{1}$$

From (1) we see that $P_t \ge [\le]\, P_{t-1}$ as $P_{t-1} \ge [\le]\, P_{t-2}$, implying that the sequence $\{P_t\}$ is monotonic. In addition, the sequence is bounded below by 0 and above by $\sigma_w^2/(1-\phi^2)$ using the fact that

$$P_t = \mathrm{E}(x_t - x_t^{t-1})^2 \le \mathrm{var}(x_t) \le \sigma_w^2/(1-\phi^2).$$

From these facts we conclude that P_t has a limit, say P, as $t \to \infty$, and from part (a), P must satisfy

$$P = \phi^2(P^{-1} + R^{-1})^{-1} + Q. \tag{2}$$

We are given $R = Q = 1$; solving (2) yields

$$P^2 + (1-\phi^2)P - 1 = 0.$$

(c) Using the notation in (b), $K_t = P_t/(P_t + 1)$ and it follows that $K_t \to K = P/(P+1)$. Also, $0 < (1-K) = 1/(P+1) < 1$ because $P > 0$.

(d) In this problem, $y_{n+1}^n = x_{n+1}^n$, and in steady state

$$\begin{aligned} x_{n+1}^n &= \phi K y_n + \phi(1-K)x_n^{n-1} \\ &= \phi K y_n + \phi^2(1-K)K y_{n-1} + \phi^2(1-K)^2 x_{n-1}^{n-2} \\ &\vdots \\ &= \sum_{j=1}^{\infty} \phi^j K(1-K)^{j-1} y_{n+1-j}. \end{aligned}$$

6.11 Using Property 6.2 and for $m \neq 0, n$:

$$x_m^n = x_m^m + J_m(x_{m+1}^n - x_{m+1}^m). \tag{1}$$

Because x_m is not observed,

$$\bullet\ x_m^m = x_m^{m-1} = \phi x_{m-1}.$$

Moreover,

$$\bullet\ x_{m+1}^n = x_{m+1}$$

and

$$\bullet\ x_{m+1}^m = x_{m+1}^{m-1} = \phi^2 x_{m-1}.$$

Note, $J_m = P_m^m \phi / P_{m+1}^m$. Now from (6.22) with $A_m = 0$,

$$\bullet\ P_m^m = P_m^{m-1} = \sigma_w^2.$$

In addition,

$$\bullet\ P_{m+1}^m = P_{m+1}^{m-1} = \sigma_w^2(1+\phi^2).$$

Thus,

$$\bullet\ J_m = \phi/(1+\phi^2).$$

Inserting the •-ed values in (1) gives the desired result.

Also from Property 6.2 and for $m \neq 0, n$,

$$P_m^n = P_m^m + J_m^2(P_{m+1}^n - P_{m+1}^m).$$

Noting that $P_{m+1}^n = 0$, and using the •-ed values above yields the desired result.

6.12 The estimates are $\widehat{\phi} = 0.786\ (.065)$ and $\widehat{\sigma}_w^2 = 1.143\ (.135)$. The missing value estimates are:

```
 t    x_t   x_t^n       t    x_t   x_t^n       t    x_t  x_t^n
 1   1.01              53  -3.12              93  -1.78
 2   ****   1.00       54   ****  -2.36       94   ****  -1.11
 3   1.05              55  -1.73              95  -0.51

 6   0.76              61  -2.28
 7   ****   1.32       62   ****  -1.38
 8   1.95              63  -0.56

13  -2.81              66  -1.64
```

```
 14   ****  -1.57       67   ****  -1.89
 15  -0.42              68  -2.25

 40  -0.60              79   1.45
 41   ****  -0.88       80   ****   1.03
 42  -1.21              81   0.67
 43   ****   0.04
 44   1.29              85   1.14
 45   0.12              86   ****   1.53
 46   ****   0.02       87   2.00
 47  -0.07              88  -0.59
 48  -0.28              89   ****  -0.58
 49   ****  -1.51       90  -0.60
 50  -2.83
```

6.13 Code is nearly the same (just put `jj = sqrt(jj)`) and everything else is the same. Results are a little better

6.14 This problem is similar to Example 6.11.

```
y = aggregate(unemp, nfrequency=4, FUN = mean)
num = length(y)
A = cbind(1,1,0,0)

# Function to Calculate Likelihood
Linn=function(para){
 Phi = diag(0,4); Phi[1,1] = para[1]
 Phi[2,]=c(0,-1,-1,-1); Phi[3,]=c(0,1,0,0); Phi[4,]=c(0,0,1,0)
 cQ1 = para[2]; cQ2 = para[3]     # sqrt q11 and sqrt q22
 cQ=diag(0,4); cQ[1,1]=cQ1; cQ[2,2]=cQ2
 cR = para[4]       # sqrt r11
 kf = Kfilter0(num,y,A,mu0,Sigma0,Phi,cQ,cR)
 return(kf$like)
 }

# Initial Parameters
mu0 = c(y[1],0,0,0);  Sigma0 = diag(100,4)
init.par = c(1.01,.1,.1,.1)   # Phi[1,1], the 2 Qs and R

# Estimation
est = optim(init.par, Linn, NULL, method='BFGS', hessian=TRUE,
    control=list(trace=1,REPORT=1))
SE = sqrt(diag(solve(est$hessian)))
u = cbind(estimate=est$par,SE)
rownames(u)=c('Phi11','sigw1','sigw2','sigv'); u

# Smooth
Phi = diag(0,4); Phi[1,1] = est$par[1]
Phi[2,]=c(0,-1,-1,-1); Phi[3,]=c(0,1,0,0); Phi[4,]=c(0,0,1,0)
cQ1 = est$par[2]; cQ2 = est$par[3]
cQ = diag(0,4); cQ[1,1]=cQ1; cQ[2,2]=cQ2
cR = est$par[4]
ks = Ksmooth0(num,y,A,mu0,Sigma0,Phi,cQ,cR)

# Plot
Tsm = ts(ks$xs[1,,], start=1948, freq=4)
Ssm = ts(ks$xs[2,,], start=1948, freq=4)
p1 = 3*sqrt(ks$Ps[1,1,]); p2 = 3*sqrt(ks$Ps[2,2,])
windows(height=5)
```

```
par(mfrow=c(2,1), mar=c(3,3,2,1), mgp=c(1.6,.6,0))
plot(Tsm, main='Trend Component', ylab='Trend')
  xx=c(time(y), rev(time(y)))
  yy=c(Tsm-p1, rev(Tsm+p1))
  polygon(xx, yy, border=NA, col=gray(.4, alpha = .2))
plot(Ssm, main='Seasonal Component', ylim=c(-40,50),  ylab='Season')
  xx=c(time(y), rev(time(y)))
  yy=c(Ssm-p2, rev(Ssm+p2))
  polygon(xx, yy, border=NA, col=gray(.4, alpha = .2))

# Forecast
n.ahead=12; y = ts(append(y, rep(0,n.ahead)), start=1948, freq=4)
rmspe = rep(0,n.ahead); x00 = ks$xf[,,num]; P00 = ks$Pf[,,num]
Q=t(cQ)%*%cQ;  R=t(cR)%*%(cR)
for (m in 1:n.ahead){
 xp = Phi%*%x00; Pp = Phi%*%P00%*%t(Phi)+Q
 sig = A%*%Pp%*%t(A)+R; K = Pp%*%t(A)%*%(1/sig)
 x00 = xp; P00 = Pp-K%*%A%*%Pp
 y[num+m] = A%*%xp; rmspe[m] = sqrt(sig)
}

windows(height=4)
par(mar=c(3,3,1,1), mgp=c(1.6,.6,0))
plot(y, type='o', main='', ylab='', ylim=c(200,900), xlim=c(1960,1984))
 upp = ts(y[(num+1):(num+n.ahead)]+3*rmspe, start=1979, freq=4)
 low = ts(y[(num+1):(num+n.ahead)]-3*rmspe, start=1979, freq=4)
  xx=c(time(low), rev(time(upp)))
  yy=c(low, rev(upp))
 polygon(xx, yy, border=NA, col=gray(.4, alpha = .2))
 abline(v=1978.75, lty=3)
```

6.15 (a) That S_{t-5} should be included is evident from the lag-plot.

```
sarima(rec,2,0,0)
rec.r = innov
lag.plot2(soi,rec.r,8)
```

(b) The model is $R_t = \beta S_{t-5} + \phi_1 R_{t-1} + \phi_2 R_{t-2} + v_t$ where R_t is centered (or detrended) Recruitment and S_t is centered (or detrended) SOI. This could easily be fit via regression techniques. In state space form, the model is

$$\begin{pmatrix} x_{t+1} \\ x_t \end{pmatrix} = \begin{pmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{pmatrix} \begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \beta \\ 0 \end{pmatrix} S_{t-5+1} + \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} v_t$$

$$R_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} + v_t.$$

First, we can use regression for initial estimates:

```
R = rec - mean(rec)  # center series
S = soi - mean(soi)
##-- Regression --##
dog = ts.intersect(R, R1=lag(R,-1), R2=lag(R,-2), S5=lag(S,-5), dframe=TRU\E)
summary(lm(dog$R ~ 0 + dog$R1 + dog$R2 + dog$S5))
   Coefficients:
         Estimate Std. Error t value
 dog$R1   1.09076    0.03565  30.594
 dog$R2  -0.25683    0.03448  -7.449
 dog$S5 -17.73344    1.01053 -17.549
```

```
 ---
 Residual standard error: 7.331 on 445 degrees of freedom
 Multiple R-squared:  0.9324,
 F-statistic:  2045 on 3 and 445 DF,  p-value: < 2.2e-16
##-- MLE  --##
fish = ts.intersect(R, S5=lag(S,-5))
y = fish[,1]
input = fish[,2]
num = length(y); A = array(c(1,0), dim = c(1,2,num))
# Function to Calculate Likelihood
 Linn=function(para){
   phi1=para[1]; phi2=para[2]; cR=para[3]; b1=para[4]
   mu0 = matrix(c(0,0), 2); Sigma0 = diag(100, 2)
   Phi = matrix(c(phi1, phi2, 1, 0), 2)
   Theta = matrix(c(phi1, phi2), 2)
   Ups = matrix(c(b1, 0), 2)
   cQ = cR;   S = cR^2
   kf = Kfilter2(num, y, A, mu0, Sigma0, Phi, Ups, Gam=0, Theta, cQ, cR, S, input)
   return(kf$like) }
# Estimation
phi1=1;  phi2=-.25; cR=7; b1=-17
init.par = c(phi1, phi2, cR, b1)  # initial parameters
est = optim(init.par, Linn, NULL, method='L-BFGS-B', hessian=TRUE,
     control=list(trace=1,REPORT=1))
SE = sqrt(diag(solve(est$hessian)))
# Results
u = cbind(estimate=est$par, SE)
rownames(u)=c('phi1','phi2','sigv','S5'); u

        estimate         SE
phi1   1.0896554 0.03552116
phi2  -0.2559608 0.03434807
sigv   7.3063958 0.24426808
S5   -17.7709038 1.00763834
```

6.16 (a) AR(1): $x_{t+1} = \phi x_t + \phi v_t$ and $y_t = x_t + v_t$.
(b) MA(1): $x_{t+1} = 0x_t + \theta v_t$ and $y_t = x_t + v_t$.
(c) IMA(1,1): $x_{t+1} = x_t + (1+\theta)v_t$ and $y_t = x_t + v_t$.

6.17 The proof of Property 6.5 is similar to the proof of Property 6.1. The first step is in noting that in this setup,

$$\begin{aligned}\operatorname{cov}(\boldsymbol{x}_{t+1}, \boldsymbol{\epsilon}_t | y_{:t-1}) &= \Phi P_t^{t-1} A_t' + S \\ \operatorname{cov}(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t) \equiv \Sigma_t &= A_t \Phi P_t^{t-1} A_t' + R \\ \mathrm{E}(\boldsymbol{x}_{t+1} | y_{:t-1}) \equiv \boldsymbol{x}_{t+1}^{t-1} &= \Phi \boldsymbol{x}_t^{t-1} + \Upsilon \boldsymbol{u}_t.\end{aligned}$$

Then we may write

$$\begin{pmatrix} \boldsymbol{x}_{t+1} \\ \boldsymbol{\epsilon}_t \end{pmatrix} \Big| \, y_{:t-1} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{x}_{t+1}^{t-1} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} P_{t+1}^{t-1} & \Phi P_t^{t-1} A_t' + S \\ A_t P_t^{t-1} \Phi' + S' & \Sigma_t \end{bmatrix} \right),$$

and (6.98), (6.100) follow. To show (6.99), write

$$(\boldsymbol{x}_{t+1} - \boldsymbol{x}_{t+1}^t) = (\Phi - K_t A_t)(\boldsymbol{x}_t - \boldsymbol{x}_t^{t-1}) + [I \quad -K_t] \begin{bmatrix} w_t \\ v_t \end{bmatrix}.$$

Then

$$P_{t+1}^t = \mathrm{E}[(\boldsymbol{x}_{t+1} - \boldsymbol{x}_{t+1}^t)(\boldsymbol{x}_{t+1} - \boldsymbol{x}_{t+1}^t)'] = (\Phi - K_tA_t)P_t^{t-1}(\Phi - K_tA_t)' \times [I \quad -K_t] \begin{bmatrix} Q & S \\ S' & R \end{bmatrix} \begin{bmatrix} I \\ -K_t' \end{bmatrix}.$$

and (6.99) follows.

6.19 Repeat Example 6.14 without "windowing' the data in lines 3 and 4 of the example code. We also set `tol = .001` in line 1 and `nboot = 200` in line 2. In contrast to Example 6.14, stochastic regression is appropriate in this case. The estimates using Newton-Raphson with estimated standard errors ("asymptotic' and"bootstrap') are below. Compare Figure 6.1 here with Figure 6.10 in the text. Here, a 95% bootstrap confidence interval for ϕ is: (.76, .97).

```
              MLE     ASYM-SE     BOOT-SE
phi     0.9166138  0.06023241   0.0559607
alpha  -0.7384140  0.42057041   0.2980171
beta    1.0300287  0.16874251   0.1034184
sig.w   0.1111845  0.03496117   0.0238318
sig.v   1.1984293  0.11386013   0.1253356
```

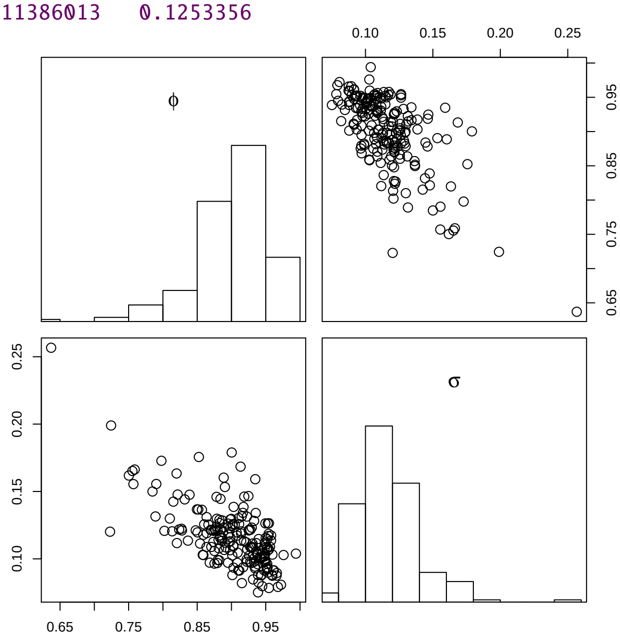

Fig. 6.1. Bootstrap distributions, $B = 200$, of the estimator of ϕ and σ_w for Problem 6.20.

6.20 See Figure 6.2. The R code for each part is below:

```
# (a) spline
plot(gtemp, type='o')
lines(smooth.spline(globtemp), col=4) # default GCV
lines(smooth.spline(globtemp, spar=.7), col=2)

# (b) DLM
num = length(globtemp)
Phi = matrix(c(2,1,-1,0),2)
A = matrix(c(1,0),1)
mu0 = matrix(c(-.28,-.28),2)
Sigma0= diag(c(100,100))
Linn=function(para){
  sigw = para[1]; sigv = para[2]
```

```
  cQ = diag(c(sigw,0))
  kf = Kfilter0(num,globtemp,A,mu0,Sigma0,Phi,cQ,sigv)
  return(kf$like)
  }
#estimation
init.par=c(.1,.1)
(est = optim(init.par, Linn, NULL, method='BFGS', hessian=TRUE,
    control=list(trace=1,REPORT=1)))
SE = sqrt(diag(solve(est$hessian)))
# Summary of estimation
estimate = est$par; u = cbind(estimate, SE)
rownames(u)=c('sigw','sigv'); u
# smooth
sigw=est$par[1]
cQ = diag(c(sigw,0))
sigv=est$par[2]
ks=Ksmooth0(num,globtemp,A,mu0,Sigma0,Phi,cQ,sigv)
xsmoo = ts(ks$xs[1,1,], start=1880)
psmoo = ts(ks$Ps[1,1,], start=1880)
upp=xsmoo+2*sqrt(psmoo)
low=xsmoo-2*sqrt(psmoo)
plot(globtemp, type='o')
lines(xsmoo, ylim=c(-.5,1), ylab='Temp Deviations')
lines(upp, col=4,lty=2)
lines(low, col=4, lty=2)
lines(smooth.spline(globtemp), col=2)  # this is close- within the bnds anyway using GCV

# (c)
plot(globtemp, type='p')
lines(xsmoo, col=2)
lines(upp, col=2, lty=2)
lines(low, col=2, lty=2)
lines(smooth.spline(globtemp,spar=.7), col=4, lty=2, lwd=2)
lines(smooth.spline(globtemp), col=4)
```

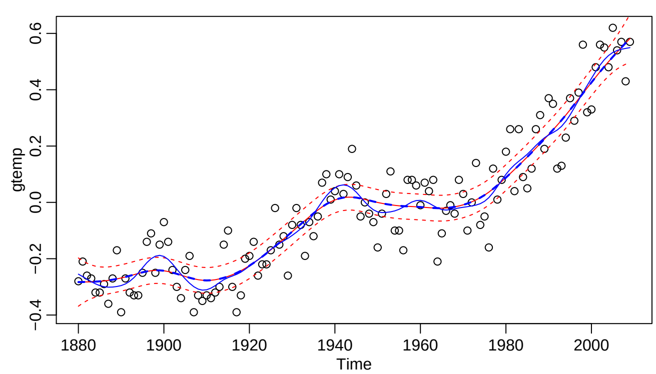

Fig. 6.2. Blue solid is GCV spline; Blue dashed is `spar=.7` spline; Red solid is DLM, (identical to `spar=.7` spline); Red dashed for DLM error bnds;

6.25 Below is the code to get the estimates. It looks like a switching model is not needed because α_1 is not significant. **You might want to give the following suggestions:** Below is the code but without observational noise ($R = 0$) and with α_0 held fixed at the drift of all the data.

```
y = diff(log(gnp))
alpha0 = mean(y)     # this is the drift and it's held fixed
num = length(y); nstate = 4;
M1 = as.matrix(cbind(1,-1,1,0))
M2 = as.matrix(cbind(1,-1,1,1))
prob = matrix(0,num,1)
# to store pi_2(t|t-1)
yp = y
# yp to store y(t|t-1)
xfilter = array(0, dim=c(nstate,1,num))
# -- Function to Calculate Likelihood --#
Linn=function(para){
# model
 phi1=para[1]; phi2=para[2]; sQ=para[3]
 sR=para[4]; alpha1=para[5]
phi=matrix(0,nstate,nstate)
phi[1,1]=phi1; phi[1,2]=phi2;
phi[2,1]=1; phi[3,3]=1; phi[4,4]=1
Q = matrix(0,nstate,nstate)
Q[1,1]=sQ^2;
R=sR^2
# end model
xf=matrix(0, nstate, 1)
# x filter
xp=matrix(0, nstate, 1)
# x pred
Pf=diag(1, nstate)
# filter cov
Pp=diag(1, nstate)
# prec cov
pi11=.75; pi12=.25
# initial switch probs
pi22=.75; pi21=.25
pif1=.5; pif2=.5
# pi_j(t|t) - filter
#-- begin filtering --#
like=0
for(i in 1:num){
xp = phi%*%xf; Pp = phi%*%Pf%*%t(phi) + Q
sig1=as.numeric(M1%*%Pp%*%t(M1) + R)
sig2=as.numeric(M2%*%Pp%*%t(M2) + R)
k1=Pp%*%t(M1)/sig1; k2=Pp%*%t(M2)/sig2
e1=y[i]-M1%*%xp-alpha0; e2=y[i]-M2%*%xp-(alpha0+alpha1)
pip1 = pif1*pi11 + pif2*pi21; pip2 = pif1*pi12 + pif2*pi22;
den1 = (1/sqrt(sig1))*exp(-.5*e1^2/sig1);
den2 = (1/sqrt(sig2))*exp(-.5*e2^2/sig2);
denom = pip1*den1 + pip2*den2;
pif1 = pip1*den1/denom; pif2 = pip2*den2/denom;
pif1=as.numeric(pif1); pif2=as.numeric(pif2)
e1=as.numeric(e1); e2=as.numeric(e2)
xf=xp + pif1*k1*e1 + pif2*k2*e2
eye = diag(1, nstate)
Pf = pif1*(eye-k1%*%M1)%*%Pp + pif2*(eye-k2%*%M2)%*%Pp
like=like-log(pip1*den1 + pip2*den2)
prob[i]<<-pip2
```

```
xfilter[,,i]<<-xf
innov.sig<<-c(sig1,sig2)
yp[i]<<-ifelse(pip1 > pip2, M1%*%xp, M2%*%xp)
}
return(like)
}
# -- Estimation --#
phi1=1; phi2=-.3; sQ=.01; sR=0;  alpha1=0   # R does not move off of 0
init.par = c(phi1,phi2,sQ,sR,alpha1)
(est = optim(init.par,Linn,NULL,method='BFGS',hessian=TRUE,control=list(trace=1,REPORT=1)))
SE = sqrt(diag(solve(est$hessian)))
u = cbind(estimate=est$par, SE)
rownames(u)=c('phi1','phi2','sig.w','sig.r','alpha1');
cat( 'alpha0 =', alpha0, '\n'); u
  alpha0 = 0.008337501
            estimate           SE
 phi1     1.23296864 0.1043216858
 phi2    -0.29143485 0.1014927439
 sig.w    0.00860795 0.0003471278
 sig.r    0.00000000 0.0023890671
 alpha1  -0.42442477 0.5594017787  *Not significant
```

6.26 This problem was done in Gauss many years ago. We're not providing R code, but we give details of the analysis. The number of sunspots rises slowly and decreases rapidly indicating a switch in regime. A threshold model was fit in Problem 5.10. Here we used an AR(2) as the basis of the model for the state, but this could be extended to higher order models.

Consider state equations

$$x_{t1} = \alpha_1 x_{t-1,1} + \alpha_2 x_{t-2,1} + w_{t1}$$

$$x_{t2} = \beta_0 + \beta_1 x_{t-1,1} + \beta_2 x_{t-2,1} + w_{t2}$$

or in vector form

$$\begin{pmatrix} x_{t1} \\ x_{t-1,1} \\ x_{t2} \\ x_{t-1,2} \end{pmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \beta_1 & \beta_2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} x_{t-1,1} \\ x_{t-2,1} \\ x_{t-1,2} \\ x_{t-2,2} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \beta_0 \\ 0 \end{pmatrix} + \begin{pmatrix} w_{t1} \\ 0 \\ w_{t2} \\ 0 \end{pmatrix}$$

with

$$Q = \begin{bmatrix} \sigma_{w1}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{w2}^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and observation equation (with y_t being the centered sunspots data)

$$y_t = A_t \begin{pmatrix} x_{t1} \\ x_{t-1,1} \\ x_{t2} \\ x_{t-1,2} \end{pmatrix} + v_t$$

where

$$A_t = [1\ 0\ 0\ 0] \quad \text{or} \quad [1\ 0\ 1\ 0].$$

Thus there are two regimes and either

$$(a)\ y_t = \mu_y + \alpha_0 + \alpha_1 x_{t-1,1} + \alpha_2 x_{t-2,1} + w_{t1} + v_t \qquad (\text{“}y_t = x_{t1} + \text{noise}')$$

(b) $y_t = \mu_y + \beta_0 + (\alpha_1 + \beta_1)x_{t-1,1} + (\alpha_2 + \beta_2)x_{t-2,1} + w_{t2} + v_t$ ("$y_t = x_{t1} + x_{t2} +$ noise$'$).

The final estimates are (we set $\widehat{\mu}_y = \bar{y}$) shown in the table below. The estimated states, $\widehat{x}_{t1}$ and $\widehat{x}_{t1} + \widehat{x}_{t2}$, are shown in Figure 6.3. Note that the dotted line at 40 indicates the model is selecting $\widehat{x}_{t1} + \widehat{x}_{t2}$; in general, the model selects $\widehat{x}_{t1} + \widehat{x}_{t2}$ during periods when the data is increasing and then peaks.

ESTIMATES FOR PROBLEM 6.22

Parameter	Estimate	SE
α_1	1.700	0.026
α_2	−0.793	0.027
β_1	0.355	0.078
β_2	−0.272	0.090
σ_{w1}	7.495	0.252
σ_{w2}	0.001	1.118
σ_v	0.000	0.371
α_0	−0.218	0.334
β_0	3.675	1.166

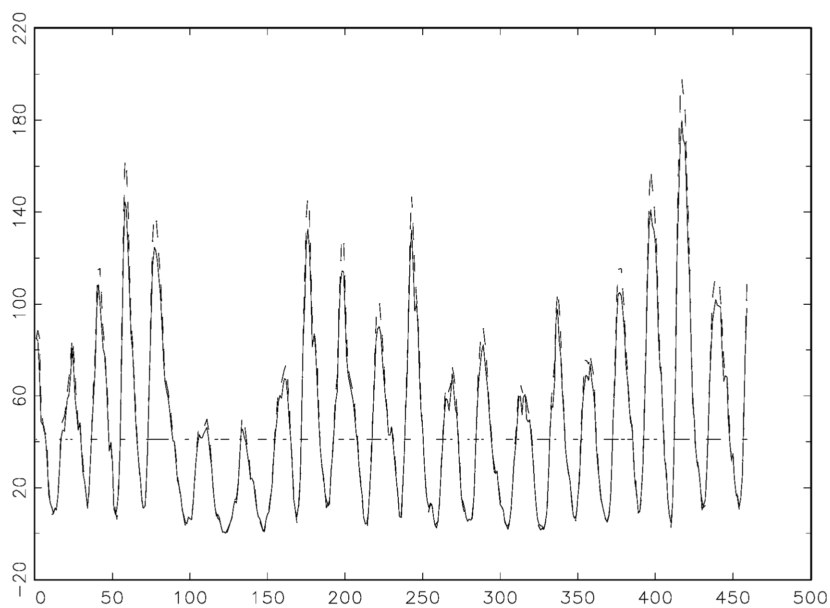

Fig. 6.3. Sunspots analysis, Problem 6.22. $\widehat{x}_{t1}$ (|), $\widehat{x}_{t1} + \widehat{x}_{t2}$ (- - -), and the horizontal dotted line ($\cdots$)at 40 indicates the model is selecting $\widehat{x}_{t1} + \widehat{x}_{t2}$.

6.28 Use the same code as in Example 6.19, but first extract the innovations:

```
ARfit = sarima(diff(log(gnp)),1,0,0)  # or can use the MA(2) fit
y = log(resid(ARfit$fit)^2)
num = length(y)
  .
  .
  .
```

Using the notation of Section 6.10, a stochastic volatility model fit to the GNP residuals is given below.

```
          estimates        SE
phi0   0.001374046 0.1849378
```

```
phi1   0.909416795 0.1043696
sQ     0.333135052 0.2291385
alpha -9.755616207 1.9498158
sigv0  0.889107526 0.2305288
mu1   -2.106589153 0.5140917
sigv1  2.512724902 0.2958002
```

Chapter 7

Solutions

7.1 Note first that

$$(C - iQ)(\boldsymbol{v}_c - i\boldsymbol{v}_s) = \lambda(\boldsymbol{v}_c - i\boldsymbol{v}_s)$$

implies that

$$\begin{pmatrix} C & -Q \\ Q & C \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_c \\ \boldsymbol{v}_s \end{pmatrix} = \lambda \begin{pmatrix} \boldsymbol{v}_c \\ \boldsymbol{v}_s \end{pmatrix}$$

by equating the real and imaginary parts and stacking. Also,

$$\begin{pmatrix} C & -Q \\ Q & C \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_s \\ -\boldsymbol{v}_c \end{pmatrix} = \lambda \begin{pmatrix} \boldsymbol{v}_s \\ -\boldsymbol{v}_c \end{pmatrix}$$

is the same equation, showing that there are two eigenvectors for each eigenvalue. Hence,

$$\left| \frac{1}{2} \begin{pmatrix} C & -Q \\ Q & C \end{pmatrix} \right| = \left(\frac{1}{2}\right)^{2p} |\mathrm{diag}\{\lambda_1, \lambda_1, \ldots, \lambda_p, \lambda_p\}| = \left(\frac{1}{2}\right)^{2p} \prod_{j=1}^{p} \lambda_j^2.$$

But,

$$|f|^2 = \left(\prod_{j=1}^{p} \lambda_j\right)^2 = \prod_{j=1}^{p} \lambda_j^2.$$

Then, it follows that

$$|\Sigma| = \left(\frac{1}{2}\right)^{2p} |f|^2,$$

which verifies the result. To verify the second result, let $\boldsymbol{Y} = \boldsymbol{X} - \boldsymbol{M}$ or

$$\boldsymbol{Y}_c - i\boldsymbol{Y}_s = (\boldsymbol{X}_c - \boldsymbol{M}_c) - i(\boldsymbol{X}_s - \boldsymbol{M}_s)$$

and note that $\boldsymbol{Y}^* f^{-1}\boldsymbol{Y}$ is purely real because $f = f^*$ is Hermitian. Then, let $\boldsymbol{W} = f^{-1}\boldsymbol{Y}$ so that

$$\begin{aligned} \boldsymbol{Y}^* f^{-1}\boldsymbol{Y} &= \boldsymbol{Y}^*\boldsymbol{W} \\ &= (\boldsymbol{Y}_c' + i\boldsymbol{Y}_s')(\boldsymbol{W}_c - i\boldsymbol{W}_s) \\ &= \boldsymbol{Y}_c'\boldsymbol{W}_c + \boldsymbol{Y}_s'\boldsymbol{W}_s, \end{aligned}$$

because the imaginary part is zero. Now, $\boldsymbol{Y} = f\boldsymbol{W}$ implies that

$$\boldsymbol{Y}_c - i\boldsymbol{Y}_s = (C - iQ)(\boldsymbol{W}_c - i\boldsymbol{W}_s),$$

or

$$\begin{pmatrix} C & -Q \\ Q & C \end{pmatrix} \begin{pmatrix} \boldsymbol{W}_c \\ \boldsymbol{W}_s \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y}_c \\ \boldsymbol{Y}_s \end{pmatrix}$$

Then, write the quadratic form

$$\frac{1}{2} \left((\boldsymbol{X}_c - \boldsymbol{M}_c)' \quad (\boldsymbol{X}_s - \boldsymbol{M}_s)' \right) \left[\frac{1}{2} \begin{pmatrix} C & -Q \\ Q & C \end{pmatrix} \right]^{-1} \begin{pmatrix} \boldsymbol{X}_c - \boldsymbol{M}_c \\ \boldsymbol{X}_s - \boldsymbol{M}_s \end{pmatrix}$$

as

$$\begin{aligned} \left(\boldsymbol{Y}_c' \quad \boldsymbol{Y}_s' \right) \begin{pmatrix} C & -Q \\ Q & C \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{Y}_c \\ \boldsymbol{Y}_s \end{pmatrix} &= \left(\boldsymbol{Y}_c' \quad \boldsymbol{Y}_s' \right) \begin{pmatrix} \boldsymbol{W}_c \\ \boldsymbol{W}_s \end{pmatrix} \\ &= (\boldsymbol{Y}_c' \boldsymbol{W}_c + \boldsymbol{Y}_s' \boldsymbol{W}_s) \\ &= \boldsymbol{Y}^* f^{-1} \boldsymbol{Y}. \end{aligned}$$

7.2 Substitute $L\hat{f}$ from (5.6) into (5.5) to rewrite the negative of the log likelihood as

$$\begin{aligned} -\ln L(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_L; f) &\propto L \ln |f| + \operatorname{tr} \left\{ f^{-1} \sum_{\ell=1}^{L} (\boldsymbol{X}_\ell - \boldsymbol{M}_\ell)(\boldsymbol{X}_\ell - \boldsymbol{M}_\ell)^* \right\} \\ &= L \ln |f| + L \operatorname{tr}\{\hat{f} f^{-1}\} \\ &= -L \ln |\hat{f} f^{-1}| + L \operatorname{tr}\{\hat{f} f^{-1}\} + \ln |\hat{f}| \\ &= -L \ln |\hat{f} P P^*| + L \operatorname{tr}\{\hat{f} P P^*\} + \ln |\hat{f}| \\ &= -L \ln |\hat{P}^* \hat{f} P| + L \operatorname{tr}\{P^* \hat{f} P\} + \ln |\hat{f}| \\ &= -L \ln |\Lambda| + L \operatorname{tr}\{\Lambda\} + \ln |\hat{f}| \\ &= -L \sum_{i=1}^{p} \ln \lambda_i + L \sum_{i=1}^{p} \lambda_i - Lp + Lp + \ln |\hat{f}| \\ &= \sum_{i=1}^{p} (\lambda_i - \ln \lambda_i - 1) + Lp + + \ln |\hat{f}| \\ &\geq Lp + \ln |\hat{f}| \end{aligned}$$

with equality when $\Lambda = I$ or $P^* \hat{f} P = I$ so that

$$\hat{f} = (P^*)^{-1} P^{-1} = f$$

7.3

$$\begin{aligned} MSE &= \gamma_y(0) - \sum_{r=-\infty}^{\infty} \boldsymbol{\beta}_r' \boldsymbol{\gamma}_{xy}(-r) \\ &= \int_{-1/2}^{1/2} f_y(\omega)\, d\omega - \int_{-1/2}^{1/2} \left(\sum_{r=-\infty}^{\infty} \boldsymbol{\beta}_r' \mathrm{e}^{-2\pi i \omega r} \right) \boldsymbol{f}_{xy}(\omega)\, d\omega \\ &= \int_{-1/2}^{1/2} [f_y(\omega) - \boldsymbol{B}'(\omega) \boldsymbol{f}_{xy}(\omega)]\, d\omega \\ &= \int_{-1/2}^{1/2} [f_y(\omega) - \boldsymbol{f}_{xy}^*(\omega) f_x^{-1}(\omega) \boldsymbol{f}_{xy}(\omega)]\, d\omega \\ &= \int_{-1/2}^{1/2} f_{y \cdot x}(\omega)\, d\omega. \end{aligned}$$

7.4 Note first that to find the coherence, we must evaluate

$$\rho^2_{\hat{y}y}(\omega) = \frac{|f_{\hat{y}y}(\omega)|^2}{f_{\hat{y}}(\omega) f_y(\omega)}$$

Note that the Fourier representation of the cross-covariance function, $\gamma_{\hat{y}y}(h)$ can be written

$$\begin{aligned}
\gamma_{\hat{y}y} &= E(\hat{y}_{t+h} y_t) \\
&= E\bigg(\sum_{r=-\infty}^{\infty} \boldsymbol{\beta}_r' \boldsymbol{x}_{t+h-r} y_t \bigg) \\
&= \sum_{r=-\infty}^{\infty} \boldsymbol{\beta}_r' \boldsymbol{\gamma}_{xy}(h-r) \\
&= \int_{-1/2}^{1/2} \sum_{r=-\infty}^{\infty} \boldsymbol{\beta}_r' \mathrm{e}^{-2\pi i\omega r} \boldsymbol{f}_{xy}(\omega) \mathrm{e}^{2\pi i \omega h} \, d\omega \\
&= \int_{-1/2}^{1/2} \boldsymbol{B}'(\omega) \boldsymbol{f}_{xy}(\omega) \mathrm{e}^{2\pi i \omega h} \, d\omega \\
&= \int_{-1/2}^{1/2} \boldsymbol{f}_{xy}^*(\omega) f_x^{-1}(\omega) \boldsymbol{f}_{xy}(\omega) \mathrm{e}^{2\pi i \omega h} \, d\omega
\end{aligned}$$

We would also need

$$\begin{aligned}
\gamma_{\hat{y}\hat{y}} &= E(\hat{y}_{t+h} \hat{y}_t) \\
&= E\bigg(\sum_{r=-\infty}^{\infty} \boldsymbol{\beta}_r' \boldsymbol{x}_{t+h-r} \sum_{s=-\infty}^{\infty} \boldsymbol{x}_{t-s}' \boldsymbol{\beta}_s \bigg) \\
&= \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \boldsymbol{\beta}_r' \Gamma_x(h-r+s) \boldsymbol{\beta}_s \\
&= \int_{-1/2}^{1/2} \bigg(\sum_{r=-\infty}^{\infty} \boldsymbol{\beta}_r' \mathrm{e}^{-2\pi i \omega r} \bigg) f_x(\omega) \bigg(\sum_{s=-\infty}^{\infty} \boldsymbol{\beta}_s \mathrm{e}^{2\pi i \omega s} \bigg) \mathrm{e}^{2\pi i \omega h} \, d\omega \\
&= \int_{-1/2}^{1/2} \boldsymbol{B}'(\omega) \boldsymbol{f}_x(\omega) \overline{\boldsymbol{B}(\omega)} \mathrm{e}^{2\pi i \omega h} \, d\omega \\
&= \int_{-1/2}^{1/2} \boldsymbol{f}_{xy}^*(\omega) f_x^{-1}(\omega) f_x(\omega) f_x^{-1}(\omega) \boldsymbol{f}_{xy}(\omega) \mathrm{e}^{2\pi i \omega h} \, d\omega \\
&= \int_{-1/2}^{1/2} \boldsymbol{f}_{xy}^*(\omega) f_x^{-1}(\omega) \boldsymbol{f}_{xy}(\omega) \mathrm{e}^{2\pi i \omega h} \, d\omega
\end{aligned}$$

Substituting into the definition for squared coherence, we obtain

$$\rho^2_{\hat{y}y}(\omega) = \frac{|\boldsymbol{f}_{xy}^*(\omega) f_x^{-1}(\omega) \boldsymbol{f}_{xy}(\omega)|^2}{\boldsymbol{f}_{xy}^*(\omega) f_x^{-1}(\omega) \boldsymbol{f}_{xy}(\omega) f_y(\omega)} = \frac{\boldsymbol{f}_{xy}^*(\omega) f_x^{-1}(\omega) \boldsymbol{f}_{xy}(\omega)}{f_y(\omega)}$$

which is just $\rho^2_{y\cdot x}(\omega)$ as given in (5.21).

7.5 Writing the complex version of the regression model as

$$\boldsymbol{Y}_c - i\boldsymbol{Y}_s = (X_c - iX_s)(\boldsymbol{B}_c - i\boldsymbol{B}_s) + \boldsymbol{V}_c - i\boldsymbol{V}_s$$

shows that

$$\boldsymbol{Y}_c = X_c\boldsymbol{B}_c - X_s\boldsymbol{B}_s + \boldsymbol{V}_c$$

and

$$\boldsymbol{Y}_s = X_s\boldsymbol{B}_c + X_c\boldsymbol{B}_s + \boldsymbol{V}_s$$

which is the matrix equation determining the real regression model. Furthermore, two complex matrices $F = F_c - iF_s, G = G_c - iG_s$ can be multiplied and the real and imaginary components of the product $H = H_c - iH_s = FG$ will appear as components of the real product

$$\begin{pmatrix} F_c & -F_s \\ F_s & F_c \end{pmatrix}\begin{pmatrix} G_c & -G_s \\ G_s & G_c \end{pmatrix} = \begin{pmatrix} H_c & -H_s \\ H_s & H_c \end{pmatrix}$$

This result, along with isomorphism involving the product of a matrix and a vector justifies the last two equations at the end of the problem. Note also that the least squares solution will be

$$\hat{\boldsymbol{B}} = (X^*X)^{-1}X^*Y = \left(\sum_{k=1}^L \overline{X_k}X_k'\right)^{-1}\sum_{k=1}^L \overline{X_k}Y_k$$

It follows that

$$\hat{\boldsymbol{B}}' = \sum_{k=1}^L Y_k\overline{X_k}'\left(\sum_{k=1}^L X_k\overline{X_k}'\right)^{-1} = \hat{\boldsymbol{f}}_{xy}^*\hat{f}_x^{-1},$$

which is the sample version of (5.16). To verify the first part of the last equation, note that

$$\boldsymbol{Y}^*\boldsymbol{Y} = \sum_{k=1}^L |Y_k|^2 = L\hat{f}_y$$

and

$$\begin{aligned}
\boldsymbol{Y}^*X(X^*X)^{-1}X^*\boldsymbol{Y} &= \sum_{k=1}^L \overline{Y_k}\boldsymbol{X}_k'\left(\sum_{k=1}^L \overline{\boldsymbol{X}_k}\boldsymbol{X}_k'\right)^{-1}\sum_{k=1}^L \overline{\boldsymbol{X}_k}Y_k \\
&= \left[\sum_{k=1}^L \overline{Y_k}\boldsymbol{X}_k'\left(\sum_{k=1}^L \overline{\boldsymbol{X}_k}\boldsymbol{X}_k'\right)^{-1}\sum_{k=1}^L \overline{\boldsymbol{X}_k}Y_k\right]' \\
&= \sum_{k=1}^L Y_k\overline{\boldsymbol{X}_k}'\left(\sum_{k=1}^L \boldsymbol{X}_k\overline{\boldsymbol{X}_k}'\right)^{-1}\sum_{k=1}^L \boldsymbol{X}_k\overline{Y_k} \\
&= L\hat{\boldsymbol{f}}_{xy}^*\hat{f}_x^{-1}\hat{\boldsymbol{f}}_{xy}
\end{aligned}$$

and the assertions in the last equations are true.

7.6 Note first that, since

$$\hat{\boldsymbol{\beta}}_t = \sum_{s=-\infty}^{\infty} h_s\boldsymbol{y}_{t-s}$$

$$\begin{aligned}
E\hat{\psi}_t &= E\sum_{r=-\infty}^{\infty}\boldsymbol{a}_r'\boldsymbol{\beta}_{t-r} \\
&= E\sum_{r,s}\boldsymbol{a}_r'h_s\boldsymbol{y}_{t-r-s} \\
&= \sum_{r,s,j}\boldsymbol{a}_r'h_sZ_{t-r-s-j}\boldsymbol{\beta}_j \\
&= \int_{-1/2}^{1/2}\boldsymbol{A}'(\omega)H(\omega)Z(\omega)\boldsymbol{B}(\omega)\mathrm{e}^{2\pi i\omega t}\,d\omega.
\end{aligned}$$

Now,

$$\psi_t = \sum_{r=-\infty}^{\infty} \boldsymbol{a}_r' \boldsymbol{\beta}_{t-r} = \int_{-1/2}^{1/2} \boldsymbol{A}'(\omega)\boldsymbol{B}(\omega) \mathrm{e}^{2\pi i\omega t}\, d\omega$$

and the above two expressions are equal if and only if

$$H(\omega)Z(\omega) = I$$

for all ω. To show that the variance is minimized, subject to the unbiased constraint, note that, for any unbiased estimator

$$\tilde{\psi}_t = \psi_t + \sum_{r,s} \boldsymbol{a}_r' g_s \boldsymbol{v}(t-r-s),$$

we would have

$$E[(\tilde{\psi}_t - \psi_t)^2] = E[(\hat{\psi}_t - \psi_t)^2] + E[(\tilde{\psi}_t - \hat{\psi}_t)^2] + 2E[(\tilde{\psi}_t - \hat{\psi}_t)(\hat{\psi}_t - \psi_t)].$$

The first two terms on the righthand side are positive and the result is shown if the cross product term is zero. We have

$$\begin{aligned}
E[(\tilde{\psi}_t - \hat{\psi}_t)(\hat{\psi}_t - \psi_t)] &= \sum_{r,s}\sum_{j,k} \boldsymbol{a}_r'(g_s - h_s)E[\boldsymbol{v}_{t-r-s}\boldsymbol{v}_{t-j-k}']h_k'\boldsymbol{a}_j \\
&= \int_{-1/2}^{1/2} \boldsymbol{A}'(\omega)\Big[G(\omega) - H(\omega)\Big]H^*(\omega)\, d\omega \\
&= \int_{-1/2}^{1/2} \boldsymbol{A}'(\omega)\Big[G(\omega) - H(\omega)\Big]Z(\omega)\Big(Z^*(\omega)Z(\omega)\Big)^{-1} d\omega \\
&= \int_{-1/2}^{1/2} \boldsymbol{A}'(\omega)\Big[G(\omega)Z(\omega) - H(\omega)Z(\omega)\Big]\Big(Z^*(\omega)Z(\omega)\Big)^{-1} d\omega \\
&= 0
\end{aligned}$$

because $G(\omega)Z(\omega) = H(\omega)Z(\omega) = I$.

7.7 In the model (5.39), make the identifications $\boldsymbol{\beta}_t = (\mu_t, \alpha_t)'$ and

$$z_t = \begin{pmatrix} \delta_t & \delta_{t-\tau_1} \\ \delta_t & \delta_{t-\tau_2} \\ \vdots & \vdots \\ \delta_t & \delta_{t-\tau_N} \end{pmatrix},$$

where $\delta_t = 0, t = 0$ and is zero otherwise. Taking the Fourier transforms gives

$$Z(\omega) = \begin{pmatrix} 1 & \mathrm{e}^{-2\pi i\omega\tau_1} \\ 1 & \mathrm{e}^{-2\pi i\omega\tau_2} \\ \vdots & \vdots \\ 1 & \mathrm{e}^{-2\pi i\omega\tau_N} \end{pmatrix},$$

Then,

$$S_z(\omega) = \begin{pmatrix} N & \sum_{j=1}^N \mathrm{e}^{-2\pi i\omega\tau_j} \\ \sum_{j=1}^N \mathrm{e}^{2\pi i\omega\tau_j} & 1 \end{pmatrix} = N\begin{pmatrix} 1 & \overline{\phi(\omega)} \\ \phi(\omega) & 1 \end{pmatrix}$$

Now, it follows that

$$S_z^{-1}(\omega) = \frac{1}{N}\frac{1}{(1 - |\phi(\omega)|^2)}\begin{pmatrix} 1 & -\overline{\phi(\omega)} \\ -\phi(\omega) & 1 \end{pmatrix}$$

for $\omega \neq 0$. If we apply $Z^*(\omega)$ directly to the transform vector $\boldsymbol{Y}(\omega) = (Y_1,(\omega),\ldots,Y_n(\omega))'$, we obtain a 2×1 vector containing $NY_\cdot(\omega)$ and

$$NB_w(\omega) = \sum_{j=1}^{N} \mathrm{e}^{2\pi i\omega j} Y_j(\omega),$$

which leads to the desired result, on multiplying by the 2×2 matrix $S_z^{-1}(\omega)$.

7.8 For computations, it is convenient to let $u = t - r$ in the model (5.66), so that

$$\boldsymbol{y}_t = \sum_{u=-\infty}^{\infty} z_u \boldsymbol{\beta}_{t-u} + \boldsymbol{v}_t.$$

The estimator is given by (5.68), i.e.,

$$\hat{\boldsymbol{\beta}}_t = \sum_{r=-\infty}^{\infty} h_r \boldsymbol{y}_{t-r}.$$

The orthogonality principle yields the optimal filter as a solution of

$$E[(\boldsymbol{\beta}_t - \hat{\boldsymbol{\beta}}_t)\boldsymbol{y}'_{t-s}] = 0$$

or

$$E[(\boldsymbol{\beta}_t \boldsymbol{y}'_{t-s})] = E[\hat{\boldsymbol{\beta}}_t \boldsymbol{y}'_{t-s})].$$

since $\boldsymbol{\beta}_t$ and $\boldsymbol{v}_t$ are uncorrelated the lefthand side of the above becomes

$$\begin{aligned} E\biggl(\boldsymbol{\beta}_t \sum_{u=-\infty}^{\infty} \boldsymbol{\beta}'_{t-s-u} z'_u\biggr) &= \sum_{u=-\infty}^{\infty} \gamma_\beta(s+u) z'_u \\ &= \int_{-1/2}^{1/2} f_\beta(\omega) Z^*(\omega) \mathrm{e}^{2\pi i\omega s}\, d\omega. \end{aligned}$$

The righthand side becomes

$$\begin{aligned} E\biggl(\hat{\boldsymbol{\beta}}_t \boldsymbol{y}'_{t-s}\biggr) &= E\biggl(\sum_{r=-\infty}^{\infty} h_r \boldsymbol{y}_{t-r} \boldsymbol{y}'_{t-s}\biggr) \\ &= \sum_{r=-\infty}^{\infty} h_r \Gamma_y(s-r) \\ &= \int_{-1/2}^{1/2} H(\omega) f_y(\omega) \mathrm{e}^{2\pi i\omega s}\, d\omega. \end{aligned}$$

Equating the Fourier transforms of the left and right sides gives $Hf_y = f_\beta Z^*$, or

$$H = f_\beta Z^*(Z f_\beta Z^* + f_v I)^{-1},$$

where the frequency arguments are suppressed to save notation. To get the final form, note that (4.56) can be written as

$$AC^*(CAC^* + B)^{-1} = (A^{-1} + C^* B^{-1} C)^{-1} C^* B^{-1}$$

for the complex case, implying the form

$$\begin{aligned} H &= \left(\frac{1}{f_\beta} I + \frac{1}{f_v} Z^* Z\right)^{-1} Z^* \frac{1}{f_v} \\ &= (S_z + \theta I)^{-1} Z^* \end{aligned}$$

for the optimal filters. To derive the mean square error, note that

$$\text{MSE} = E[(\boldsymbol{\beta}_t - \hat{\boldsymbol{\beta}}_t)\boldsymbol{\beta}_t'] = E[(\boldsymbol{\beta}_t\boldsymbol{\beta}_t')] - E[(\hat{\boldsymbol{\beta}}_t\boldsymbol{\beta}_t')].$$

The second term is

$$\begin{aligned} E[(\hat{\boldsymbol{\beta}}_t\boldsymbol{\beta}_t')] &= E\left(\sum_{u,s} h_s z_u \boldsymbol{\beta}_{t-s-u}\boldsymbol{\beta}_t'\right) \\ &= \sum_{s,u} h_s z_u \gamma_\beta(-s-u) \\ &= \int_{-1/2}^{1/2} H(\omega)Z(\omega)f_\beta(\omega)\, d\omega. \end{aligned}$$

Combining the two terms gives

$$\text{MSE} = \int_{-1/2}^{1/2} [f_\beta(\omega) - H(\omega)Z(\omega)f_\beta(\omega)]\, d\omega.$$

Suppressing frequency again, we may write the argument as

$$f_\beta - f_\beta Z^*\bigg(Zf_\beta Z^* + f_v I\bigg)^{-1} Zf_\beta$$

Then, appeal to the complex version of (4.55), i.e.

$$A - AC^*\bigg(CAC^* + B\bigg)^{-1}CA = \bigg(A^{-1} + C^*B^{-1}C)^{-1}$$

to write the argument as

$$\left(\frac{1}{f_\beta}I + Z^*Z\frac{1}{f_v}\right)^{-1}$$

which reduces to (5.72).

7.9 Suppressing the frequency argument, we write

$$\begin{aligned} E[SSR] &= E[\boldsymbol{Y}^*ZS_z^{-1}Z^*\boldsymbol{Y}] \\ &= \text{tr}\left\{ZS_x^{-1}Z^*E[\boldsymbol{YY}^*]\right\} \\ &= \text{tr}\left\{ZS_z^{-1}Z^*(f_\beta ZZ^* + f_v I)\right\} \\ &= f_\beta\ \text{tr}\left\{ZS_z^{-1}Z^*ZZ^*\right\} + f_v\ \text{tr}\left\{ZS_z^{-1}Z^*\right\} \\ &= f_\beta\ \text{tr}\left\{ZZ^*\right\} + f_v\ \text{tr}\left\{S_z^{-1}S_z\right\} \\ &= f_\beta\ \text{tr}\left\{S_z\right\} + qf_v. \end{aligned}$$

When the spectrum is cumulated over L frequencies, the multiplier appears.

7.10 Again, suppressing the frequency subscripts the model $\boldsymbol{Y} = Z\boldsymbol{B} + V$ takes the vector form

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1N} \\ Y_{21} \\ \vdots \\ Y_{2N} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & -1 \\ \vdots & \vdots \\ 1 & -1 \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} + \begin{pmatrix} V_{11} \\ \vdots \\ V_{1N} \\ V_{21} \\ \vdots \\ V_{2N} \end{pmatrix},$$

where B_1 is the DFT of μ_t and B_2 is the DFT of $\alpha_{1t}, \alpha_{2t} = -\alpha_{1t}$. The null hypothesis is that $B_2 = 0$. Now, in (5.52),

$$s_{zy} = \begin{pmatrix} N(Y_{\cdot 1} + Y_{\cdot 2}) \\ N(Y_{\cdot 1} - Y_{\cdot 2}) \end{pmatrix}$$

and

$$\begin{aligned} s_{y\cdot z}^2 &= \sum_{i=1}^{2}\sum_{j=1}^{N} |Y_{ij}|^2 - \left(N(Y_{\cdot 1} + Y_{\cdot 2}) \quad N(Y_{\cdot 1} - Y_{\cdot 2}) \right) \begin{pmatrix} 2N & 0 \\ 0 & 2N \end{pmatrix}^{-1} \begin{pmatrix} N(Y_{\cdot 1} + Y_{\cdot 2}) \\ N(Y_{\cdot 1} - Y_{\cdot 2}) \end{pmatrix} \\ &= \sum_{j=1}^{N} |Y_{1j}|^2 + \sum_{j=1}^{N} |Y_{2j}|^2 - N|Y_{1\cdot}|^2 - N|Y_{2\cdot}|^2 \\ &= \sum_{i=1}^{2}\sum_{j=1}^{N} |Y_{ij} - Y_{i\cdot}|^2, \end{aligned}$$

which is (5.85) Under the reduced model, $s_{1y} = N(Y_{1\cdot} + Y_{2\cdot})$ and $S_{11} = 2N$ so that

$$s_{y\cdot 1}^2 = \sum_{j=1}^{N} |Y_{1j}|^2 + \sum_{j=1}^{N} |Y_{2j}|^2 - \frac{N}{2}|Y_{1\cdot} + Y_{2\cdot}|^2.$$

Then, substitute

$$Y_{\cdot\cdot} = \frac{1}{2}(Y_{1\cdot} + Y_{2\cdot})$$

to obtain

$$s_{y\cdot 1}^2 = \sum_{j=1}^{N} |Y_{1j}|^2 + \sum_{j=1}^{N} |Y_{2j}|^2 - 2N|Y_{\cdot\cdot}|^2.$$

Then,

$$\begin{aligned} RSS &= s_{y\cdot 1}^2 - s_{y\cdot z}^2 \\ &= -2N|Y_{\cdot\cdot}|^2 + N|Y_{1\cdot}|^2 + N|Y_{2\cdot}|^2 \\ &= N\sum_{i=1}^{2} |Y_{i\cdot} - Y_{\cdot\cdot}|^2 \\ &= \sum_{i=1}^{2}\sum_{j=1}^{N} |Y_{i\cdot} - Y_{\cdot\cdot}|^2, \end{aligned}$$

which is (5.84).

7.11 Use the model

$$y_{ijt} = \mu_{it} + v_{ijt},$$

for $i = 1, \ldots, I, j = 1, \ldots, N_i$ and write the frequency domain version in terms of the vector

$$\boldsymbol{Y} = (Y_{11}, \ldots, Y_{1,N_1}, Y_{21}, \ldots, Y_{2N_2}, \ldots, Y_{I1}, \ldots, Y_{IN_I})'$$

The matrix Z has N_1 ones in elements 1 to N_1 of column 1, and zeros elsewhere, N_2 ones in elements $N_1 + 1, \ldots, N_1 + N_2$ of column 2, etc. It follows that

$$S_z = Z^* Z = \text{diag}\,(N_1, N_2, \ldots, N_I)$$

and

$$Z^* \boldsymbol{Y} = (N_1 Y_{1\cdot}, N_2 Y_{2\cdot}, \ldots, N_I Y_{I\cdot})'$$

so that

$$\hat{\boldsymbol{B}} = (Y_{1\cdot}, Y_{2\cdot}, \ldots, Y_{I\cdot})'$$

and

$$\boldsymbol{A}^* \hat{\boldsymbol{B}} = \sum_{i=1}^{I} \overline{A_i} Y_{i\cdot}.$$

Finally,

$$\begin{aligned} Q(A) &= (\overline{A_1}, \overline{A_2}, \ldots, \overline{A_I}) \operatorname{diag}\left(\frac{1}{N_1}, \frac{1}{N_2}, \ldots, \frac{1}{N_I}\right)(A_1, A_2, \ldots, A_I)' \\ &= \sum_{i=1}^{I} \frac{|A_i|^2}{N_i}. \end{aligned}$$

The error variance $s^2_{y \cdot z}$ comes from (5.85).

7.12 Each of the spectra

$$\hat{f}_1 = \sum_{j=1}^{N_1} |Y_{1j} - Y_{1\cdot}|^2 \quad \text{and} \quad \hat{f}_2 = \sum_{j=1}^{N_2} |Y_{2j} - Y_{2\cdot}|^2$$

can be regarded as an error power component, computed from the model

$$y_{ijt} = \mu_{it} + v_{ijt}$$

for a fixed $i = 1, 2$. Hence, from Table 5.2, the error power components will have a chi-squared distribution, say,

$$\frac{2(N_i - 1)\hat{f}_i}{f_i} \sim \chi^2_{2(N_i - 1)}$$

for $i = 1, 2$, and the two samples are assumed to be independent. It follows that the ratio

$$\frac{\hat{f}_1}{\hat{f}_2} \frac{f_2}{f_1} \sim \frac{\chi^2_{[2(N_1-1)]}/2(N_1 - 1)}{\chi^2_{[2(N_2-1)]}/2(N_2 - 1)} \sim F_{[2(N_1-1), 2(N_2-1)]}.$$

It follows that

$$\frac{\hat{f}_1}{\hat{f}_2} \sim \frac{f_1}{f_2} F_{[2(N_1-1), 2(N_2-1)]}.$$

7.13 In the notation of (5.113), $\boldsymbol{\mu}_1 = \boldsymbol{s}, \boldsymbol{\mu}_2 = \boldsymbol{0}, \Sigma_1 = \Sigma_2 = \sigma^2_w I$ and

$$\begin{aligned} d_L(\boldsymbol{x}) &= \frac{1}{\sigma^2_w} \boldsymbol{s}' \boldsymbol{x} - \frac{1}{2} \frac{\boldsymbol{s}' \boldsymbol{s}}{\sigma^2_w} + \ln \frac{\pi_1}{\pi_2} \\ &= \frac{1}{\sigma^2_w} \sum_{t=1}^{n} s_t x_t - \frac{1}{2} \frac{\sum_{t=1}^{n} s_t^2}{\sigma^2_w} + \ln \frac{\pi_1}{\pi_2} \\ &= \frac{1}{\sigma^2_w} \sum_{t=1}^{n} s_t x_t - \frac{1}{2} \left(\frac{S}{N}\right) + \ln \frac{\pi_1}{\pi_2} \end{aligned}$$

When $\pi_1 = \pi_2$, the last term disappears and we may use (5.115) for the two error probabilities with

$$D^2 = \frac{\boldsymbol{s}' \boldsymbol{s}}{\sigma^2_w} = \left(\frac{S}{N}\right).$$

7.14 In this case, $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}$, $\Sigma_1 = (\sigma_s^2 + \sigma_w^2)I$, $\Sigma_2 = \sigma_w^2 I$, so that (5.115) becomes

$$d_q(\boldsymbol{x}) = -\frac{n}{2}\ln\left(\frac{\sigma_s^2 + \sigma_w^2}{\sigma_w^2}\right) - \frac{1}{2}\left(\frac{1}{\sigma_s^2 + \sigma_w^2} - \frac{1}{\sigma_w^2}\right)\boldsymbol{x}'\boldsymbol{x} + \ln\frac{\pi_1}{\pi_2}$$
$$= \frac{1}{2}\frac{\sigma_s^2}{\sigma_w^2(\sigma_s^2 + \sigma_w^2)}\sum_{t=1}^n x_t^2 - \frac{1}{2}\ln\left(\frac{\sigma_s^2 + \sigma_w^2}{\sigma_w^2}\right) + \ln\frac{\pi_1}{\pi_2}$$

In this case, define the signal-to-noise ratio as

$$\left(\frac{S}{N}\right) = \frac{\sigma_s^2}{\sigma_w^2}$$

so that, for the quadratic criterion with $\pi_1 = \pi_2$ we accept Π_1 or Π_2 according to whether the statistic

$$T(\boldsymbol{x}) = \frac{1}{2}\frac{\sigma_s^2}{\sigma_w^2(\sigma_s^2 + \sigma_w^2)}\sum_{t=1}^n x_t^2$$

exceeds or fails to exceed

$$K = \frac{1}{2}\ln\left(\frac{\sigma_s^2 + \sigma_w^2}{\sigma_w^2}\right) = \frac{1}{2}\ln\left[1 + \left(\frac{S}{N}\right)\right].$$

Now, under Π_1, $\sum_t x_t^2 \sim (\sigma_s^2 + \sigma_w^2)\chi_n^2$, whereas, under Π_2, $\sum_t x_t^2 \sim \sigma_w^2\chi_n^2$, so that

$$T(\boldsymbol{x}) \sim \frac{1}{2}\left(\frac{S}{N}\right)\chi_n^2$$

under Π_1 and

$$T(\boldsymbol{x}) \sim \frac{1}{2}\left[1 + \left(\frac{S}{N}\right)\right]^{-1}\chi_n^2$$

under Π_2.

7.15 *Awake-Heat:* Figures 7.1 and 7.2 are the figures corresponding to Figures 7.14 and 7.15 of the text (Example 7.14) except that here, Caudate is included (location number 5). *Awake-Shock:* The corresponding figures are Figures 7.3 and 7.4 below. The listing below is similar to Table 5.8 but for Awake-Heat and Awake-Shock. Note that $\chi_2^2(.95, .975, .99) = 5.99, 7.38, 9.21$.

```
                AWAKE-HEAT                              AWAKE-SHOCK
loc  |e|   chi^2   loc  |e|    chi^2  |  loc  |e|   chi^2   loc  |e|    chi^2
 1  .467  842.01    6  .121     8.09  |   1  .410  233.39    6  0.139  13.952
 2  .450  150.16    7  .012     0.07  |   2  .416  138.42    7  0.161  18.848
 3  .460  623.62    8  .323    64.99  |   3  .387  389.46    8  0.309  252.75
 4  .363  104.45    9  .254    46.76  |   4  .370  352.68    9  0.398  539.46
 5  .230   30.32                      |   5  .269  107.23
```

7.16 (a) See Figure 7.5. The P components have broad power at the midrange frequencies whereas the S components have broad power at the lower frequencies.
(b) See Figure 7.7. There appears to be little or no coherence between the P and S components.
(c) - (d) See Figure 7.6. These figures support the overall conclusion of part (a).
(e) See Figure 7.8. The canonical variate series appear to be strongly coherent (in contrast to the individual series).

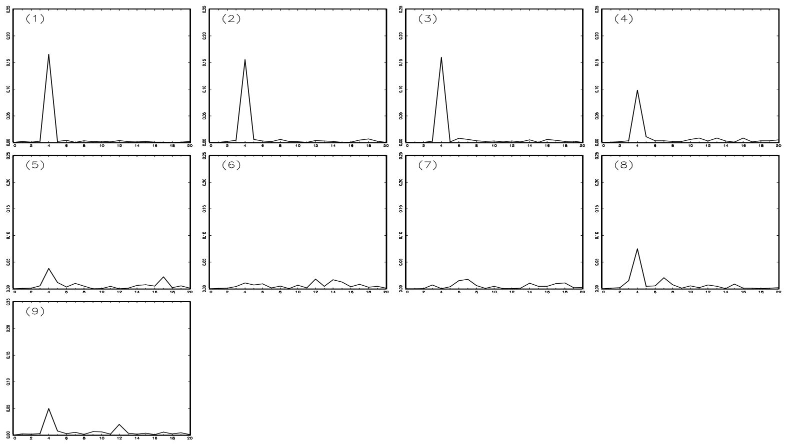

Fig. 7.1. Problem 7.15 (Awake-Heat) individual periodograms

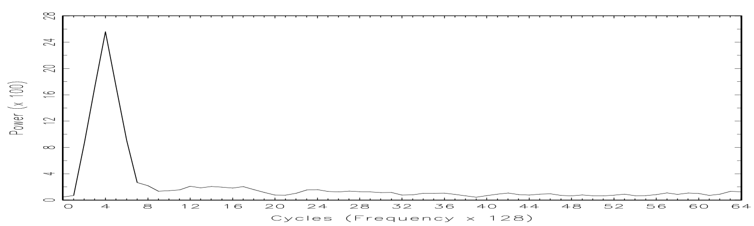

Fig. 7.2. Problem 7.15 (Awake-Heat) spectral density of first PC series.

7.17 For $p = 3$ and $q = 1$, write (5.158) as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} z + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix}.$$

This implies

$$\begin{bmatrix} 1 & .4 & .9 \\ .4 & 1 & .7 \\ .9 & .7 & 1 \end{bmatrix} = \begin{bmatrix} b_1^2 + \delta_1^2 & b_1 b_2 & b_1 b_3 \\ b_1 b_2 & b_2^2 + \delta_2^2 & b_2 b_3 \\ b_1 b_3 & b_2 b_3 & b_3^2 + \delta_3^2 \end{bmatrix}.$$

Now

$$b_1 b_2 = .4 \quad \text{and} \quad b_2 b_3 = .7 \quad \Rightarrow \quad b_1 = \frac{4}{7} b_3.$$

But

$$b_1 b_3 = .9 \quad \Rightarrow \quad \frac{4}{7} b_3^2 = .9.$$

This means that $b_3^2 = .9\frac{7}{4} = 1.575$. But we also have $b_3^2 + \delta_3^2 = 1$ in which case $\delta_3^2 < 0$ which is not a valid variance.

7.19 Note that $f(\omega) = \sum_h \Gamma(h)\cos(2\pi h v) - i\sum_h \Gamma(h)\sin(2\pi h v) = f^{re}(\omega) - i f^{im}(\omega)$. Now $f^{im}(\omega)' = \sum_h \Gamma'(h)\sin(2\pi h v) = \sum_h \Gamma(-h)\sin(2\pi h v) = -\sum_h \Gamma(-h)\sin(-2\pi h v) = -\sum_h \Gamma(h)\sin(2\pi h v) =$

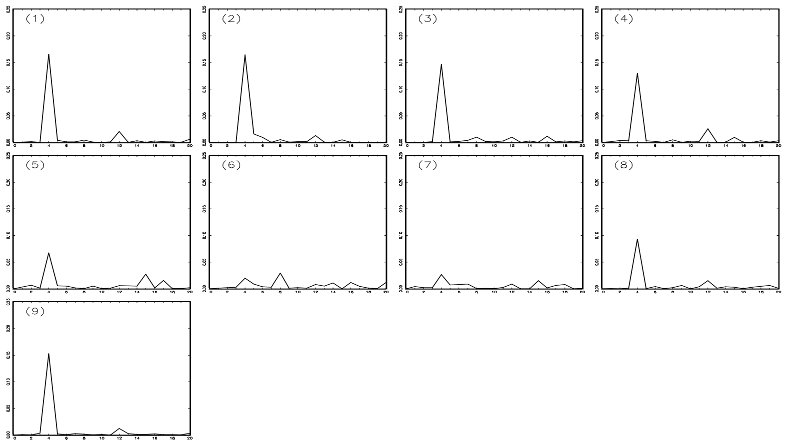

Fig. 7.3. Problem 7.15 (Awake-Shock) individual periodograms

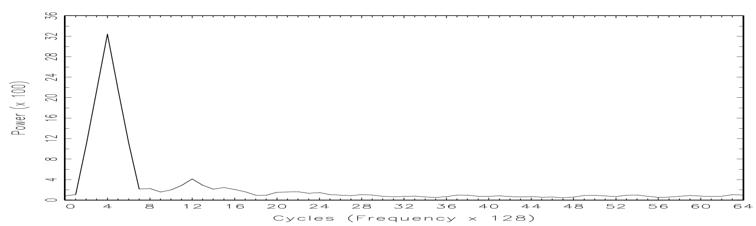

Fig. 7.4. Problem 7.15 (Awake-Shock) spectral density of first PC series.

$-f^{im}(\omega)$; that is, the imaginary part is skew symmetric. Also note that $f^{re}(\omega)' = \sum_h \Gamma(-h)\cos(-2\pi h v) = f^{re}(\omega)$; that is, the real part is symmetric.
Next, because $\boldsymbol{\beta}' f^{im}(\omega)\boldsymbol{\beta}$ is a scalar, $\boldsymbol{\beta}' f^{im}(\omega)\boldsymbol{\beta} = (\boldsymbol{\beta}' f^{im}(\omega)\boldsymbol{\beta})' = \boldsymbol{\beta}' f^{im}(\omega)'\boldsymbol{\beta} = -\boldsymbol{\beta}' f^{im}(\omega)\boldsymbol{\beta}$. This result implies $\boldsymbol{\beta}' f^{im}(\omega)\boldsymbol{\beta} = 0$ for any real-valued vector $\boldsymbol{\beta}$.

7.20 See Figure 7.9. Note the significant peak at $\omega = 1/3$ as opposed to the fourth quarter of EBV (see Figure 7.23).

7.21 (a) The estimated spectral density of r_t is shown in Figure 7.10 and supports the white noise hypothesis.
(b) The spectral envelope with respect to $\mathcal{G} = \{x, |x|, x^2\}$ is shown in Figure 7.11 where substantial power near the zero frequency is noted. The optimal transformation is shown in Figure 7.12 as a solid line and the usual square transformation is shown as a dotted line; the two transformations are similar.

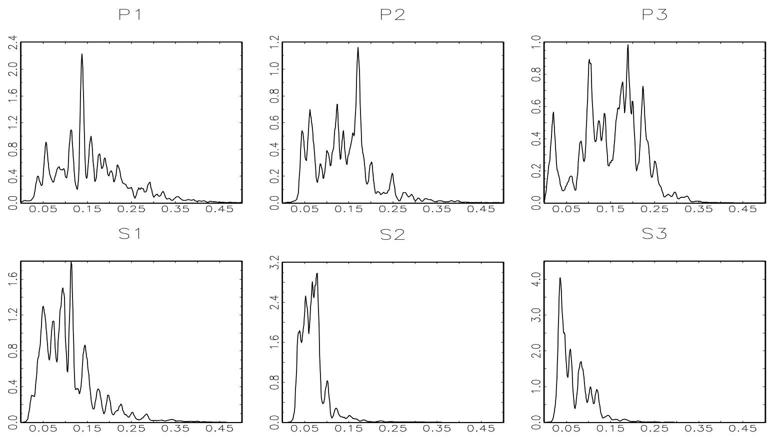

Fig. 7.5. Problem 7.16 (a). Estimated spectral densities.

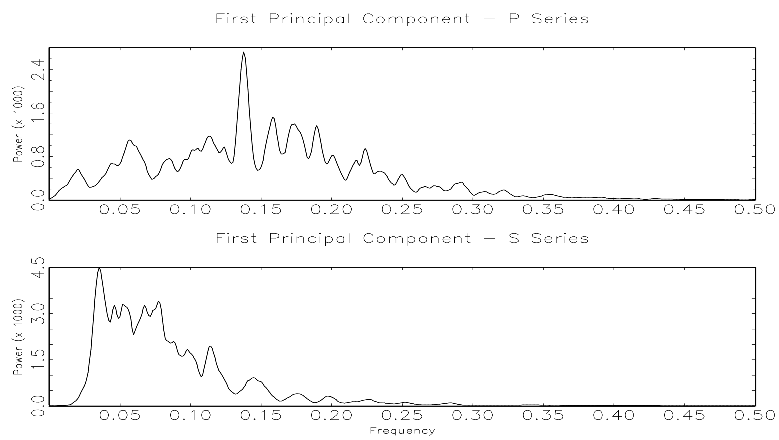

Fig. 7.6. Problem 7.16 (c)-(d). Spectral density of first PC series.

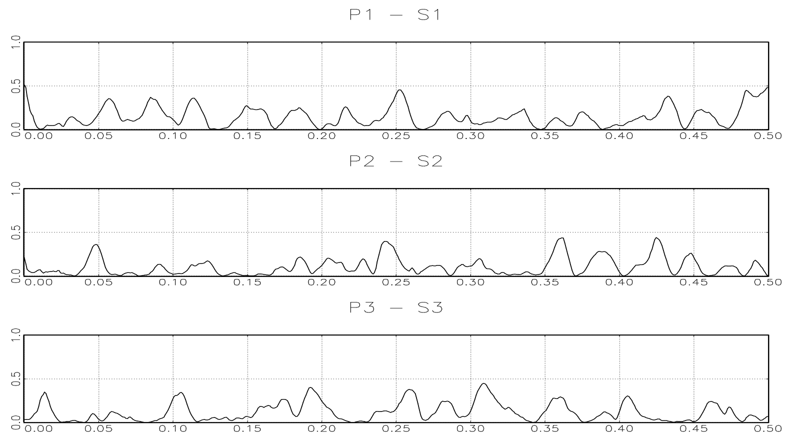

Fig. 7.7. Problem 7.16 (b). Estimated coherencies.

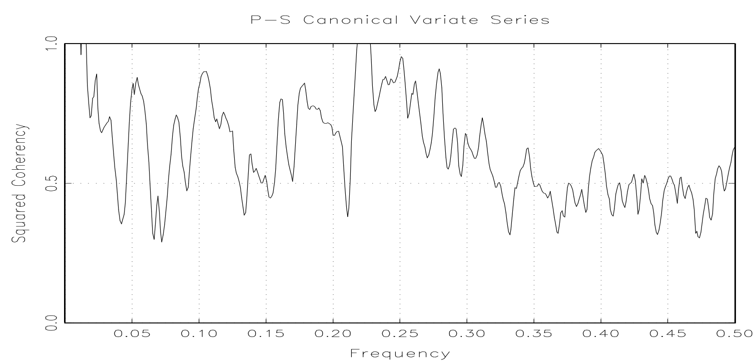

Fig. 7.8. Problem 7.16 (c) (d). Squared coherency between canonical variate series.

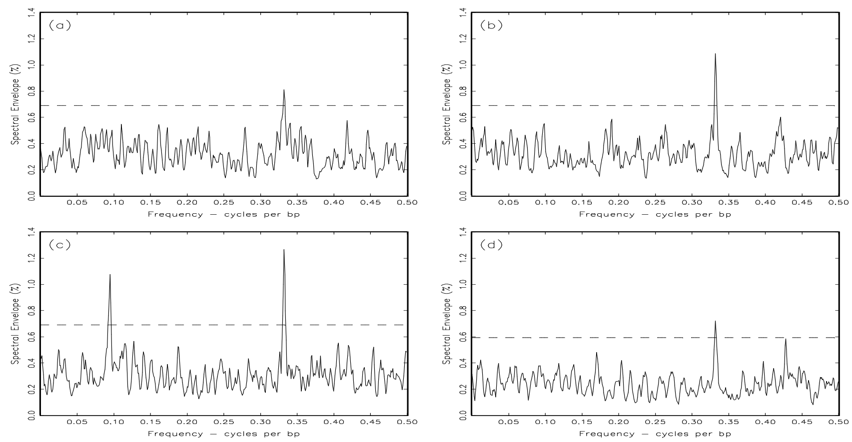

Fig. 7.9. Problem 7.20. This is the equivalent of Figure 7.23 but for the herpesvirus saimiri.

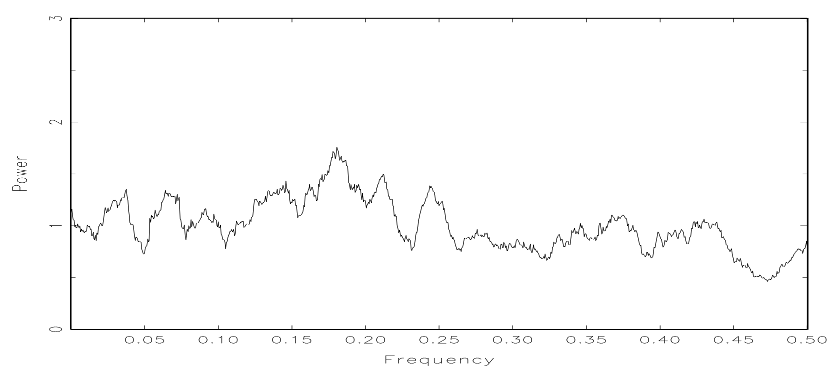

Fig. 7.10. Problem 7.21 (a). Estimated spectral density of NYSE returns.

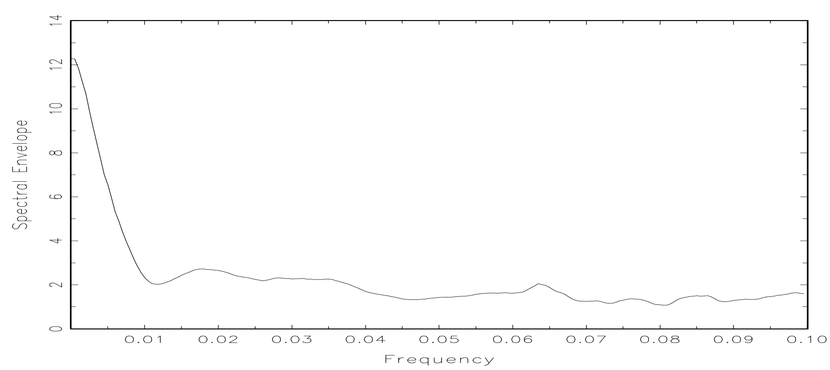

Fig. 7.11. Problem 7.21 (b). Spectral envelope of NYSE returns with respect to $\mathcal{G} = \{x, |x|, x^2\}$.

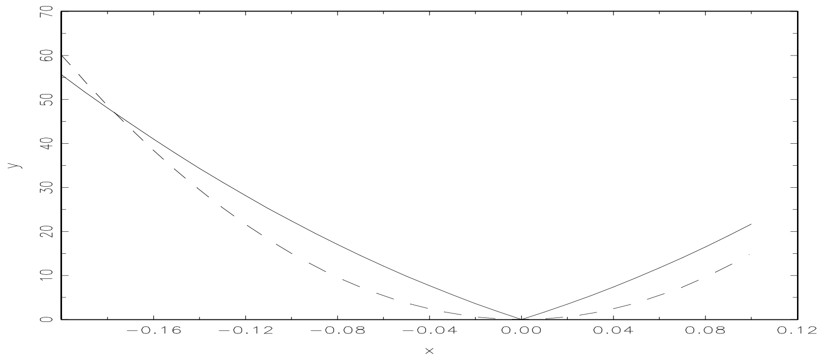

Fig. 7.12. Problem 7.21 (b). The optimal transformation (|) and the usual square transformation (- - -).

http://www.springer.com/978-3-319-52451-1

Time Series Analysis and Its Applications
With R Examples
Shumway, R.H.; Stoffer, D.S.
2017, XIII, 562 p. 148 illus., 70 illus. in color., Softcover
ISBN: 978-3-319-52451-1