STAT 626: Outline of Lecture 5

Transformations to Stationarity (§2.3)

1. **Forms of Nonstatioarity**

$$x_t = \mu_t + \sigma_t y_t,$$

   where $\{y_t\}$ is stationary.

2. **Differencing** (Fixes Mean-Nonstationarity)

3. **Log and Power (Box-Cox) Transforms** (Fixes Variance-Nonstat.)

4. **Regression Methods (§2.2)** (Study this Section for the next Lecture.)

5. **Moving Average and Smoothing (§2.4)**

~~**Example 2.3 Regression With Lagged Variables**~~

~~In Example 1.25, we discovered that the Southern Oscillation Index (SOI) measured at time $t-6$ months is associated with the Recruitment series at time t, indicating that the SOI leads the Recruitment series by six months. Although there is evidence that the relationship is not linear (this is discussed further in Example 2.7), we may consider the following regression,~~

$$\cancel{R_t = \beta_1 + \beta_2 S_{t-6} + w_t,} \tag{2.26}$$

~~where R_t denotes Recruitment for month t and S_{t-6} denotes SOI six months prior. Assuming the w_t sequence is white, the fitted model is~~

$$\cancel{\widehat{R}_t = 65.79 - 44.28_{(2.78)} S_{t-6}} \tag{2.27}$$

~~with $\widehat{\sigma}_w = 22.5$ on 445 degrees of freedom. This result indicates the strong predictive ability of SOI for Recruitment six months in advance. Of course, it is still essential to check the the model assumptions, but again we defer this until later.~~

~~Performing lagged regression in R is a little difficult because the series must be aligned prior to running the regression. The easiest way to do this is to create a data frame that we call `fish` using `ts.intersect`, which aligns the lagged series.~~

```
fish = ts.intersect(rec, soiL6=lag(soi,-6), dframe=TRUE)
summary(lm(rec~soiL6, data=fish, na.action=NULL))
```

2.3 Exploratory Data Analysis

In general, it is necessary for time series data to be stationary, so averaging lagged products over time, as in the previous section, will be a sensible thing to do. With time series data, it is the dependence between the values of the series that is important to measure; we must, at least, be able to estimate autocorrelations with precision. It would be difficult to measure that dependence if the dependence structure is not regular or is changing at every time point. Hence, to achieve any meaningful statistical analysis of time series data, it will be crucial that, if nothing else, the mean and the autocovariance functions satisfy the conditions of stationarity (for at least some reasonable stretch of time) stated in Definition 1.7. Often, this is not the case, and we will mention some methods in this section for playing down the effects of nonstationarity so the stationary properties of the series may be studied.

A number of our examples came from clearly nonstationary series. The Johnson & Johnson series in Figure 1.1 has a mean that increases exponentially over time, and the increase in the magnitude of the fluctuations around this trend causes changes in the covariance function; the variance of the process, for example, clearly increases as one progresses over the length of the series. Also, the global temperature series shown in Figure 1.2 contains some evidence of a trend over time; human-induced global warming advocates seize on this as empirical evidence to advance their hypothesis that temperatures are increasing.

Perhaps the easiest form of nonstationarity to work with is the trend stationary model wherein the process has stationary behavior around a trend. We may write this type of model as

$$x_t = \mu_t + y_t \tag{2.28}$$

where x_t are the observations, μ_t denotes the trend, and y_t is a stationary process. Quite often, strong trend, μ_t, will obscure the behavior of the stationary process, y_t, as we shall see in numerous examples. Hence, there is some advantage to removing the trend as a first step in an exploratory analysis of such time series. The steps involved are to obtain a reasonable estimate of the trend component, say $\widehat{\mu}_t$, and then work with the residuals

$$\widehat{y}_t = x_t - \widehat{\mu}_t. \tag{2.29}$$

Consider the following example.

Example 2.4 Detrending Global Temperature

Here we suppose the model is of the form of (2.28),

$$x_t = \mu_t + y_t,$$

where, as we suggested in the analysis of the global temperature data presented in Example 2.1, a straight line might be a reasonable model for the trend, i.e.,

$$\mu_t = \beta_1 + \beta_2\, t.$$

In that example, we estimated the trend using ordinary least squares[3] and found

$$\widehat{\mu}_t = -11.2 + .006\, t.$$

Figure 2.1 shows the data with the estimated trend line superimposed. To obtain the detrended series we simply subtract $\widehat{\mu}_t$ from the observations, x_t, to obtain the detrended series

$$\widehat{y}_t = x_t + 11.2 - .006\, t.$$

The top graph of Figure 2.4 shows the detrended series. Figure 2.5 shows the ACF of the original data (top panel) as well as the ACF of the detrended data (middle panel).

[3] Because the error term, y_t, is not assumed to be iid, the reader may feel that weighted least squares is called for in this case. The problem is, we do not know the behavior of y_t and that is precisely what we are trying to assess at this stage. A notable result by Grenander and Rosenblatt (1957, Ch 7), however, is that under mild conditions on y_t, for polynomial regression or periodic regression, asymptotically, ordinary least squares is equivalent to weighted least squares.

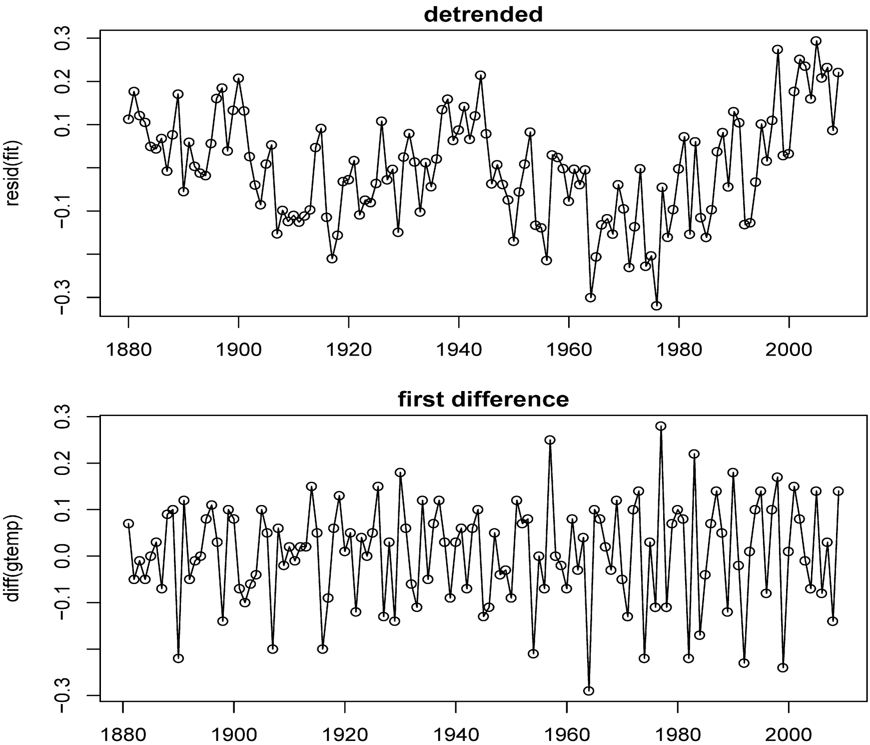

Fig. 2.4. Detrended (top) and differenced (bottom) global temperature series. The original data are shown in Figures 1.2 and 2.1.

To detrend in the series in R, use the following commands. We also show how to difference and plot the differenced data; we discuss differencing after this example. In addition, we show how to generate the sample ACFs displayed in Figure 2.5.

```
fit = lm(gtemp~time(gtemp), na.action=NULL) # regress gtemp on time
par(mfrow=c(2,1))
plot(resid(fit), type="o", main="detrended")
plot(diff(gtemp), type="o", main="first difference")
par(mfrow=c(3,1))  # plot ACFs
acf(gtemp, 48, main="gtemp")
acf(resid(fit), 48, main="detrended")
acf(diff(gtemp), 48, main="first difference")
```

In Example 1.11 and the corresponding Figure 1.10 we saw that a random walk might also be a good model for trend. That is, rather than modeling trend as fixed (as in Example 2.4), we might model trend as a stochastic component using the random walk with drift model,

$$\mu_t = \delta + \mu_{t-1} + w_t, \tag{2.30}$$

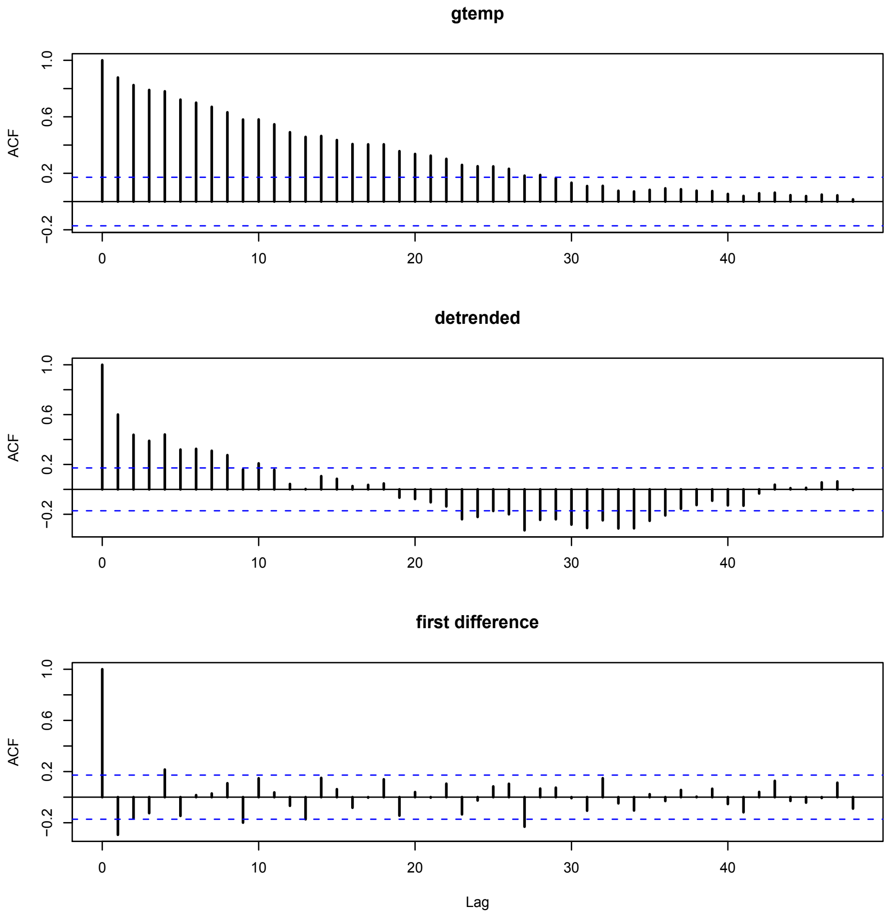

Fig. 2.5. Sample ACFs of the global temperature (top), and of the detrended (middle) and the differenced (bottom) series.

where w_t is white noise and is independent of y_t. If the appropriate model is (2.28), then differencing the data, x_t, yields a stationary process; that is,

$$\begin{aligned} x_t - x_{t-1} &= (\mu_t + y_t) - (\mu_{t-1} + y_{t-1}) \\ &= \delta + w_t + y_t - y_{t-1}. \end{aligned} \tag{2.31}$$

It is easy to show $z_t = y_t - y_{t-1}$ is stationary using footnote 3 of Chapter 1 on page 20. That is, because y_t is stationary,

$$\begin{aligned} \gamma_z(h) = \text{cov}(z_{t+h}, z_t) &= \text{cov}(y_{t+h} - y_{t+h-1}, y_t - y_{t-1}) \\ &= 2\gamma_y(h) - \gamma_y(h+1) - \gamma_y(h-1) \end{aligned}$$

is independent of time; we leave it as an exercise (Problem 2.7) to show that $x_t - x_{t-1}$ in (2.31) is stationary.

One advantage of differencing over detrending to remove trend is that no parameters are estimated in the differencing operation. One disadvantage, however, is that differencing does not yield an estimate of the stationary process y_t as can be seen in (2.31). If an estimate of y_t is essential, then detrending may be more appropriate. If the goal is to coerce the data to stationarity, then differencing may be more appropriate. Differencing is also a viable tool if the trend is fixed, as in Example 2.4. That is, e.g., if $\mu_t = \beta_1 + \beta_2 t$ in the model (2.28), differencing the data produces stationarity (see Problem 2.6):

$$x_t - x_{t-1} = (\mu_t + y_t) - (\mu_{t-1} + y_{t-1}) = \beta_2 + y_t - y_{t-1}.$$

Because differencing plays a central role in time series analysis, it receives its own notation. The first difference is denoted as

$$\nabla x_t = x_t - x_{t-1}. \tag{2.32}$$

As we have seen, the first difference eliminates a linear trend. A second difference, that is, the difference of (2.32), can eliminate a quadratic trend, and so on. In order to define higher differences, we need a variation in notation that we will use often in our discussion of ARIMA models in Chapter 3.

Definition 2.4 *We define the* **backshift operator** *by*

$$Bx_t = x_{t-1}$$

and extend it to powers $B^2 x_t = B(Bx_t) = Bx_{t-1} = x_{t-2}$, *and so on. Thus,*

$$B^k x_t = x_{t-k}. \tag{2.33}$$

It is clear that we may then rewrite (2.32) as

$$\nabla x_t = (1 - B)x_t, \tag{2.34}$$

and we may extend the notion further. For example, the second difference becomes

$$\begin{aligned} \nabla^2 x_t &= (1 - B)^2 x_t = (1 - 2B + B^2)x_t \\ &= x_t - 2x_{t-1} + x_{t-2} \end{aligned}$$

by the linearity of the operator. To check, just take the difference of the first difference $\nabla(\nabla x_t) = \nabla(x_t - x_{t-1}) = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2})$.

Definition 2.5 **Differences of order d** *are defined as*

$$\nabla^d = (1 - B)^d, \tag{2.35}$$

where we may expand the operator $(1-B)^d$ *algebraically to evaluate for higher integer values of* d. *When* $d = 1$, *we drop it from the notation.*

The first difference (2.32) is an example of a linear filter applied to eliminate a trend. Other filters, formed by averaging values near x_t, can produce adjusted series that eliminate other kinds of unwanted fluctuations, as in Chapter 3. The differencing technique is an important component of the ARIMA model of Box and Jenkins (1970) (see also Box et al., 1994), to be discussed in Chapter 3.

Example 2.5 Differencing Global Temperature

The first difference of the global temperature series, also shown in Figure 2.4, produces different results than removing trend by detrending via regression. For example, the differenced series does not contain the long middle cycle we observe in the detrended series. The ACF of this series is also shown in Figure 2.5. In this case it appears that the differenced process shows minimal autocorrelation, which may imply the global temperature series is nearly a random walk with drift. It is interesting to note that if the series is a random walk with drift, the mean of the differenced series, which is an estimate of the drift, is about .0066 (but with a large standard error):

```
mean(diff(gtemp))    # = 0.00659 (drift)
sd(diff(gtemp))/sqrt(length(diff(gtemp))) # = 0.00966 (SE)
```

An alternative to differencing is a less-severe operation that still assumes stationarity of the underlying time series. This alternative, called fractional differencing, extends the notion of the difference operator (2.35) to fractional powers $-.5 < d < .5$, which still define stationary processes. Granger and Joyeux (1980) and Hosking (1981) introduced long memory time series, which corresponds to the case when $0 < d < .5$. This model is often used for environmental time series arising in hydrology. We will discuss long memory processes in more detail in §5.2.

Often, obvious aberrations are present that can contribute nonstationary as well as nonlinear behavior in observed time series. In such cases, transformations may be useful to equalize the variability over the length of a single series. A particularly useful transformation is

$$y_t = \log x_t, \tag{2.36}$$

which tends to suppress larger fluctuations that occur over portions of the series where the underlying values are larger. Other possibilities are power transformations in the Box–Cox family of the form

$$y_t = \begin{cases} (x_t^\lambda - 1)/\lambda & \lambda \neq 0, \\ \log x_t & \lambda = 0. \end{cases} \tag{2.37}$$

Methods for choosing the power λ are available (see Johnson and Wichern, 1992, §4.7) but we do not pursue them here. Often, transformations are also used to improve the approximation to normality or to improve linearity in predicting the value of one series from another.

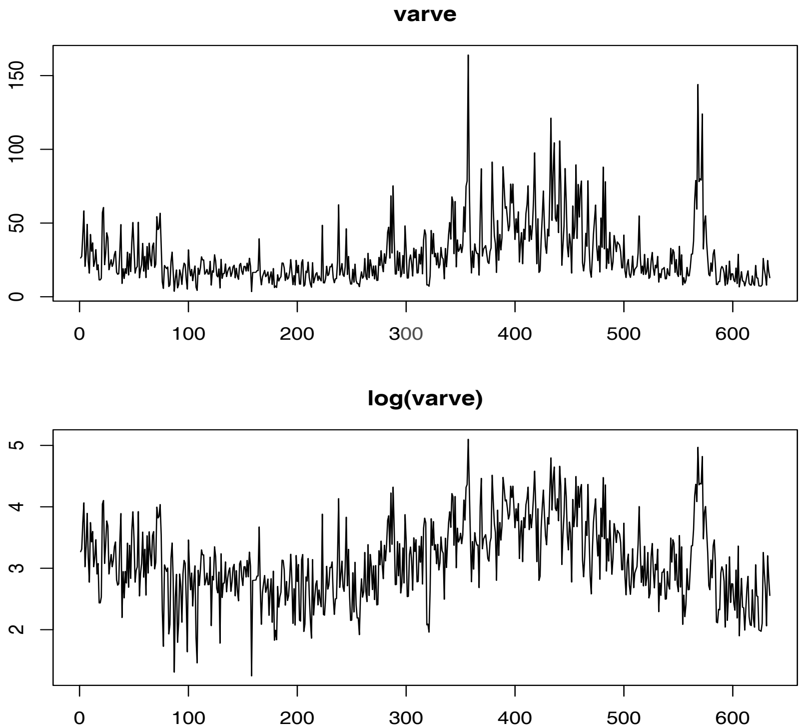

Fig. 2.6. Glacial varve thicknesses (top) from Massachusetts for $n = 634$ years compared with log transformed thicknesses (bottom).

Example 2.6 Paleoclimatic Glacial Varves

Melting glaciers deposit yearly layers of sand and silt during the spring melting seasons, which can be reconstructed yearly over a period ranging from the time deglaciation began in New England (about 12,600 years ago) to the time it ended (about 6,000 years ago). Such sedimentary deposits, called varves, can be used as proxies for paleoclimatic parameters, such as temperature, because, in a warm year, more sand and silt are deposited from the receding glacier. Figure 2.6 shows the thicknesses of the yearly varves collected from one location in Massachusetts for 634 years, beginning 11,834 years ago. For further information, see Shumway and Verosub (1992). Because the variation in thicknesses increases in proportion to the amount deposited, a logarithmic transformation could remove the nonstationarity observable in the variance as a function of time. Figure 2.6 shows the original and transformed varves, and it is clear that this improvement has occurred. We may also plot the histogram of the original and transformed data, as in Problem 2.8, to argue that the approximation to normality is improved. The ordinary first differences (2.34) are also computed in Problem 2.8, and we note that the first differences have a significant negative correlation at lag $h = 1$. Later, in Chapter 5, we will show that perhaps the varve series has long memory and will propose using fractional differencing.

Figure 2.6 was generated in R as follows:

```
par(mfrow=c(2,1))
plot(varve, main="varve", ylab="")
plot(log(varve), main="log(varve)", ylab="" )
```

Next, we consider another preliminary data processing technique that is used for the purpose of visualizing the relations between series at different lags, namely, scatterplot matrices. In the definition of the ACF, we are essentially interested in relations between x_t and x_{t-h}; the autocorrelation function tells us whether a substantial linear relation exists between the series and its own lagged values. The ACF gives a profile of the linear correlation at all possible lags and shows which values of h lead to the best predictability. The restriction of this idea to linear predictability, however, may mask a possible nonlinear relation between current values, x_t, and past values, x_{t-h}. This idea extends to two series where one may be interested in examining scatterplots of y_t versus x_{t-h}

Example 2.7 Scatterplot Matrices, SOI and Recruitment

To check for nonlinear relations of this form, it is convenient to display a lagged scatterplot matrix, as in Figure 2.7, that displays values of the SOI, S_t, on the vertical axis plotted against S_{t-h} on the horizontal axis. The sample autocorrelations are displayed in the upper right-hand corner and superimposed on the scatterplots are locally weighted scatterplot smoothing (lowess) lines that can be used to help discover any nonlinearities. We discuss smoothing in the next section, but for now, think of lowess as a robust method for fitting nonlinear regression.

In Figure 2.7, we notice that the lowess fits are approximately linear, so that the sample autocorrelations are meaningful. Also, we see strong positive linear relations at lags $h = 1, 2, 11, 12$, that is, between S_t and $S_{t-1}, S_{t-2}, S_{t-11}, S_{t-12}$, and a negative linear relation at lags $h = 6, 7$. These results match up well with peaks noticed in the ACF in Figure 1.14.

Similarly, we might want to look at values of one series, say Recruitment, denoted R_t plotted against another series at various lags, say the SOI, S_{t-h}, to look for possible nonlinear relations between the two series. Because, for example, we might wish to predict the Recruitment series, R_t, from current or past values of the SOI series, S_{t-h}, for $h = 0, 1, 2, ...$ it would be worthwhile to examine the scatterplot matrix. Figure 2.8 shows the lagged scatterplot of the Recruitment series R_t on the vertical axis plotted against the SOI index S_{t-h} on the horizontal axis. In addition, the figure exhibits the sample cross-correlations as well as lowess fits.

Figure 2.8 shows a fairly strong nonlinear relationship between Recruitment, R_t, and the SOI series at $S_{t-5}, S_{t-6}, S_{t-7}, S_{t-8}$, indicating the SOI series tends to lead the Recruitment series and the coefficients are negative, implying that increases in the SOI lead to decreases in the Recruitment. The

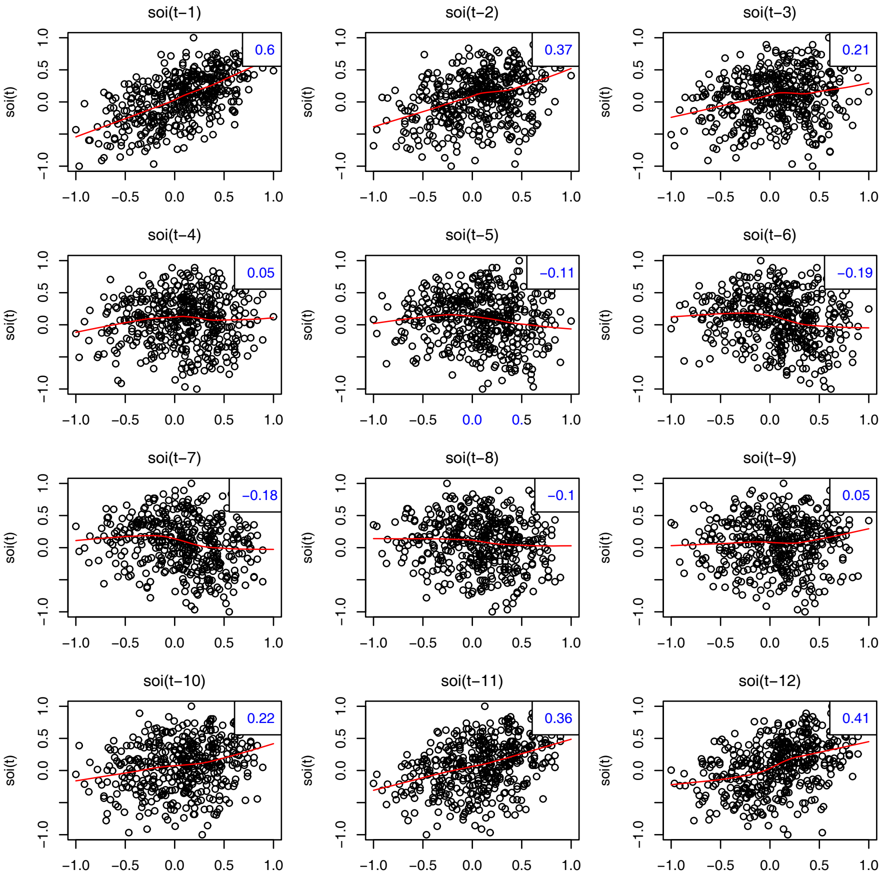

Fig. 2.7. Scatterplot matrix relating current SOI values, S_t, to past SOI values, S_{t-h}, at lags $h = 1, 2, ..., 12$. The values in the upper right corner are the sample autocorrelations and the lines are a lowess fit.

nonlinearity observed in the scatterplots (with the help of the superimposed lowess fits) indicate that the behavior between Recruitment and the SOI is different for positive values of SOI than for negative values of SOI.

Simple scatterplot matrices for one series can be obtained in R using the `lag.plot` command. Figures 2.7 and 2.8 may be reproduced using the following scripts provided with the text (see Appendix R for detials):

```
lag.plot1(soi, 12)       # Fig 2.7
lag.plot2(soi, rec, 8)   # Fig 2.8
```

As a final exploratory tool, we discuss assessing periodic behavior in time series data using regression analysis and the periodogram; this material may be thought of as an introduction to spectral analysis, which we discuss in

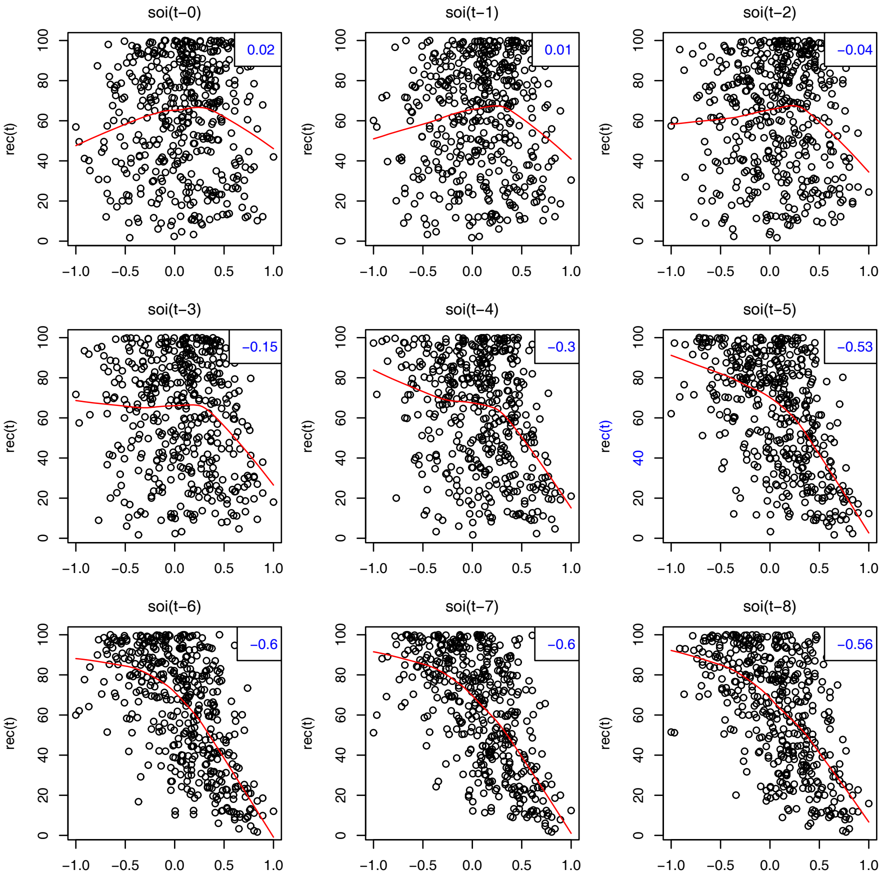

Fig. 2.8. Scatterplot matrix of the Recruitment series, R_t, on the vertical axis plotted against the SOI series, S_{t-h}, on the horizontal axis at lags $h = 0, 1, \dots, 8$. The values in the upper right corner are the sample cross-correlations and the lines are a lowess fit.

detail in Chapter 4. In Example 1.12, we briefly discussed the problem of identifying cyclic or periodic signals in time series. A number of the time series we have seen so far exhibit periodic behavior. For example, the data from the pollution study example shown in Figure 2.2 exhibit strong yearly cycles. Also, the Johnson & Johnson data shown in Figure 1.1 make one cycle every year (four quarters) on top of an increasing trend and the speech data in Figure 1.2 is highly repetitive. The monthly SOI and Recruitment series in Figure 1.6 show strong yearly cycles, but hidden in the series are clues to the El Niño cycle.

Example 2.8 Using Regression to Discover a Signal in Noise

In Example 1.12, we generated $n = 500$ observations from the model

$$x_t = A\cos(2\pi\omega t + \phi) + w_t, \tag{2.38}$$

where $\omega = 1/50$, $A = 2$, $\phi = .6\pi$, and $\sigma_w = 5$; the data are shown on the bottom panel of Figure 1.11 on page 16. At this point we assume the frequency of oscillation $\omega = 1/50$ is known, but A and ϕ are unknown parameters. In this case the parameters appear in (2.38) in a nonlinear way, so we use a trigonometric identity[4] and write

$$A\cos(2\pi\omega t + \phi) = \beta_1 \cos(2\pi\omega t) + \beta_2 \sin(2\pi\omega t),$$

where $\beta_1 = A\cos(\phi)$ and $\beta_2 = -A\sin(\phi)$. Now the model (2.38) can be written in the usual linear regression form given by (no intercept term is needed here)

$$x_t = \beta_1 \cos(2\pi t/50) + \beta_2 \sin(2\pi t/50) + w_t. \tag{2.39}$$

Using linear regression on the generated data, the fitted model is

$$\widehat{x}_t = -.71_{(.30)} \cos(2\pi t/50) - 2.55_{(.30)} \sin(2\pi t/50) \tag{2.40}$$

with $\widehat{\sigma}_w = 4.68$, where the values in parentheses are the standard errors. We note the actual values of the coefficients for this example are $\beta_1 = 2\cos(.6\pi) = -.62$ and $\beta_2 = -2\sin(.6\pi) = -1.90$. Because the parameter estimates are significant and close to the actual values, it is clear that we are able to detect the signal in the noise using regression, even though the signal appears to be obscured by the noise in the bottom panel of Figure 1.11. Figure 2.9 shows data generated by (2.38) with the fitted line, (2.40), superimposed.

To reproduce the analysis and Figure 2.9 in R, use the following commands:

```
set.seed(1000)  # so you can reproduce these results
x = 2*cos(2*pi*1:500/50 + .6*pi) + rnorm(500,0,5)
z1 = cos(2*pi*1:500/50);  z2 = sin(2*pi*1:500/50)
summary(fit <- lm(x~0+z1+z2))  # zero to exclude the intercept
plot.ts(x, lty="dashed")
lines(fitted(fit), lwd=2)
```

Example 2.9 Using the Periodogram to Discover a Signal in Noise

The analysis in Example 2.8 may seem like cheating because we assumed we knew the value of the frequency parameter ω. If we do not know ω, we could try to fit the model (2.38) using nonlinear regression with ω as a parameter. Another method is to try various values of ω in a systematic way. Using the

[4] $\cos(\alpha \pm \beta) = \cos(\alpha)\cos(\beta) \mp \sin(\alpha)\sin(\beta)$.

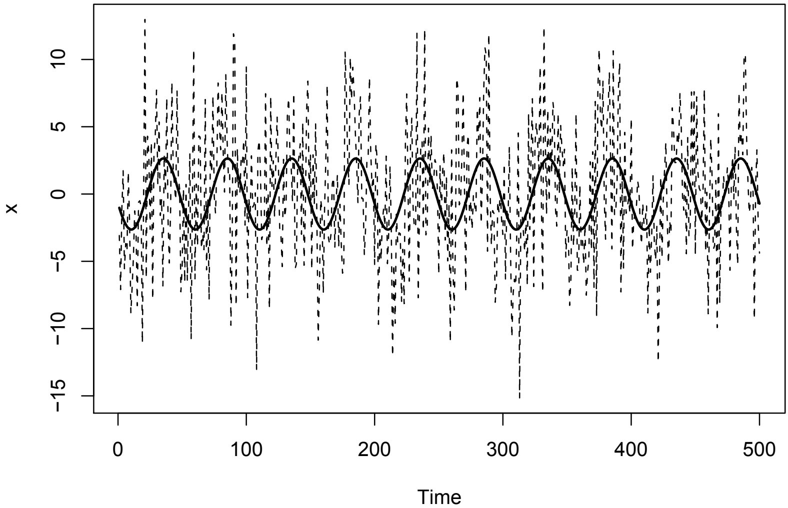

Fig. 2.9. Data generated by (2.38) [dashed line] with the fitted [solid] line, (2.40), superimposed.

regression results of §2.2, we can show the estimated regression coefficients in Example 2.8 take on the special form given by

$$\widehat{\beta}_1 = \frac{\sum_{t=1}^n x_t \cos(2\pi t/50)}{\sum_{t=1}^n \cos^2(2\pi t/50)} = \frac{2}{n}\sum_{t=1}^n x_t \cos(2\pi t/50); \tag{2.41}$$

$$\widehat{\beta}_2 = \frac{\sum_{t=1}^n x_t \sin(2\pi t/50)}{\sum_{t=1}^n \sin^2(2\pi t/50)} = \frac{2}{n}\sum_{t=1}^n x_t \sin(2\pi t/50). \tag{2.42}$$

This suggests looking at all possible regression parameter estimates,[5] say

$$\widehat{\beta}_1(j/n) = \frac{2}{n}\sum_{t=1}^n x_t \cos(2\pi t\, j/n); \tag{2.43}$$

$$\widehat{\beta}_2(j/n) = \frac{2}{n}\sum_{t=1}^n x_t \sin(2\pi t\, j/n), \tag{2.44}$$

where, $n = 500$ and $j = 1, \dots, \frac{n}{2} - 1$, and inspecting the results for large values. For the endpoints, $j = 0$ and $j = n/2$, we have $\widehat{\beta}_1(0) = n^{-1}\sum_{t=1}^n x_t$ and $\widehat{\beta}_1(\frac{1}{2}) = n^{-1}\sum_{t=1}^n (-1)^t x_t$, and $\widehat{\beta}_2(0) = \widehat{\beta}_2(\frac{1}{2}) = 0$.

For this particular example, the values calculated in (2.41) and (2.42) are $\widehat{\beta}_1(10/500)$ and $\widehat{\beta}_2(10/500)$. By doing this, we have regressed a series, x_t, of

[5] In the notation of §2.2, the estimates are of the form $\sum_{t=1}^n x_t z_t \,/\, \sum_{t=1}^n z_t^2$ where $z_t = \cos(2\pi tj/n)$ or $z_t = \sin(2\pi tj/n)$. In this setup, unless $j = 0$ or $j = n/2$ if n is even, $\sum_{t=1}^n z_t^2 = n/2$; see Problem 2.10.

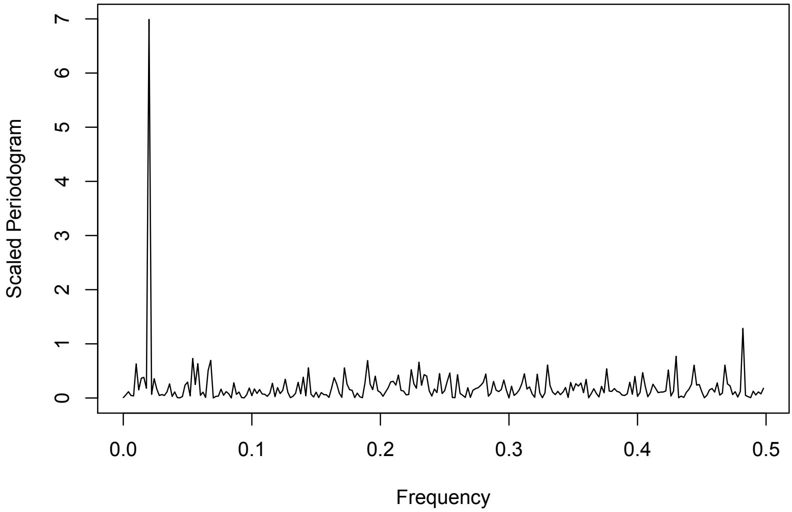

Fig. 2.10. The scaled periodogram, (2.45), of the 500 observations generated by (2.38); the data are displayed in Figures 1.11 and 2.9.

length n using n regression parameters, so that we will have a perfect fit. The point, however, is that if the data contain any cyclic behavior we are likely to catch it by performing these saturated regressions.

Next, note that the regression coefficients $\widehat{\beta}_1(j/n)$ and $\widehat{\beta}_2(j/n)$, for each j, are essentially measuring the correlation of the data with a sinusoid oscillating at j cycles in n time points.[6] Hence, an appropriate measure of the presence of a frequency of oscillation of j cycles in n time points in the data would be

$$P(j/n) = \widehat{\beta}_1^2(j/n) + \widehat{\beta}_2^2(j/n), \tag{2.45}$$

which is basically a measure of squared correlation. The quantity (2.45) is sometimes called the periodogram, but we will call $P(j/n)$ the scaled periodogram and we will investigate its properties in Chapter 4. Figure 2.10 shows the scaled periodogram for the data generated by (2.38), and it easily discovers the periodic component with frequency $\omega = .02 = 10/500$ even though it is difficult to visually notice that component in Figure 1.11 due to the noise.

Finally, we mention that it is not necessary to run a large regression

$$x_t = \sum_{j=0}^{n/2} \beta_1(j/n)\cos(2\pi tj/n) + \beta_2(j/n)\sin(2\pi tj/n) \tag{2.46}$$

to obtain the values of $\beta_1(j/n)$ and $\beta_2(j/n)$ [with $\beta_2(0) = \beta_2(1/2) = 0$] because they can be computed quickly if n (assumed even here) is a highly

[6] Sample correlations are of the form $\sum_t x_t z_t \big/ \left(\sum_t x_t^2 \sum_t z_t^2\right)^{1/2}$.

composite integer. There is no error in (2.46) because there are n observations and n parameters; the regression fit will be perfect. The discrete Fourier transform (DFT) is a complex-valued weighted average of the data given by

$$\begin{aligned} d(j/n) &= n^{-1/2} \sum_{t=1}^{n} x_t \exp(-2\pi i t j/n) \\ &= n^{-1/2} \left(\sum_{t=1}^{n} x_t \cos(2\pi t j/n) - i \sum_{t=1}^{n} x_t \sin(2\pi t j/n) \right) \end{aligned} \tag{2.47}$$

where the frequencies j/n are called the Fourier or fundamental frequencies. Because of a large number of redundancies in the calculation, (2.47) may be computed quickly using the fast Fourier transform (FFT)[7], which is available in many computing packages such as Matlab®, S-PLUS® and R. Note that[8]

$$|d(j/n)|^2 = \frac{1}{n} \left(\sum_{t=1}^{n} x_t \cos(2\pi t j/n) \right)^2 + \frac{1}{n} \left(\sum_{t=1}^{n} x_t \sin(2\pi t j/n) \right)^2 \tag{2.48}$$

and it is this quantity that is called the periodogram; we will write

$$I(j/n) = |d(j/n)|^2.$$

We may calculate the scaled periodogram, (2.45), using the periodogram as

$$P(j/n) = \frac{4}{n} I(j/n). \tag{2.49}$$

We will discuss this approach in more detail and provide examples with data in Chapter 4.

Figure 2.10 can be created in R using the following commands (and the data already generated in x):

```
I = abs(fft(x))^2/500  # the periodogram
P = (4/500)*I[1:250]   # the scaled periodogram
f = 0:249/500          # frequencies
plot(f, P, type="l", xlab="Frequency", ylab="Scaled Periodogram")
```

2.4 Smoothing in the Time Series Context

In §1.4, we introduced the concept of smoothing a time series, and in Example 1.9, we discussed using a moving average to smooth white noise. This method is useful in discovering certain traits in a time series, such as long-term

[7] Different packages scale the FFT differently; consult the documentation. R calculates (2.47) without scaling by $n^{-1/2}$.

[8] If $z = a - ib$ is complex, then $|z|^2 = z\overline{z} = (a - ib)(a + ib) = a^2 + b^2$.