STAT 626: Outline of Lecture 6

Regression in the Time Series Contexts (§2.2,)

1. **Review of Stationarity, Preview of TS Models**

2. **A Quick Review of Multiple Regression**

$$x = Z\beta + w,$$

LSE of β :

$$\widehat{\beta} = (Z'Z)^{-1}Z'x.$$

3. **Tests, CIs and Variable Selection**

4. **AIC (Akaike Information Criterion)**

5. **Example 2.2: Pollution, Temperature and Mortality (PTM) Data**

6. **Example 2.3: Regression with Lagged Variables**

7. **Examples 2.8, 2.9: Trig. Regression and Periodogram**

Review of Stationarity, Preview of TS Models

1. **Linear Processes:** $x_t = \mu + \sum_{j=-\infty}^{+\infty} \psi_j w_{t-j}$ is stationary with the *autocovariance function*

$$\gamma(h) = \sum_{j=-\infty}^{\infty} \psi_{j+h}\psi_j.$$

2. **MA(q) Models**: $x_t = w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}, \quad \theta_q \neq 0$, is stationary, its autocovariance is zero at lags greater than q.

3. **Autoregressive Models of order p or AR(p) Models:**

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + w_t, \quad \phi_p \neq 0.$$

QUESTION: Is a time series $\{x_t\}$ defined via an AR(p) model always stationary? If so, what is its autocovariance function?
To get a feel for the answer consider the AR(1):

$$x_t = \phi x_{t-1} + w_t,$$

what happens when $\phi = 1$?

4. **The Backshift Operator B:** $Bx_t = x_{t-1}$.

5. **MA(q) and B:**

$$x_t = w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q} = (1 + \theta_1 B + \ldots + \theta_q B^q) w_t = \theta(B) w_t.$$

6. **AR(p) and B:**

$$x_t - \phi_1 x_{t-1} - \ldots - \phi_p x_{t-p} = w_t, \quad (1 - \phi_1 B - \ldots - \phi_p B^p) x_t = \phi(B) x_t = w_t.$$

7. **The ROOTS of the polynomial equation**

$$\phi(B) = 0,$$

hold the key to the question of stationarity of the solutions of AR models.

context, and therefore we include some material on transformations and other techniques useful in exploratory data analysis.

2.2 Classical Regression in the Time Series Context

We begin our discussion of linear regression in the time series context by assuming some output or dependent time series, say, x_t, for $t = 1, \ldots, n$, is being influenced by a collection of possible inputs or independent series, say, $z_{t1}, z_{t2}, \ldots, z_{tq}$, where we first regard the inputs as fixed and known. This assumption, necessary for applying conventional linear regression, will be relaxed later on. We express this relation through the linear regression model

$$x_t = \beta_1 z_{t1} + \beta_2 z_{t2} + \cdots + \beta_q z_{tq} + w_t, \tag{2.1}$$

where $\beta_1, \beta_2, \ldots, \beta_q$ are unknown fixed regression coefficients, and $\{w_t\}$ is a random error or noise process consisting of independent and identically distributed (iid) normal variables with mean zero and variance σ_w^2; we will relax the iid assumption later. A more general setting within which to embed mean square estimation and linear regression is given in Appendix B, where we introduce Hilbert spaces and the Projection Theorem.

Example 2.1 Estimating a Linear Trend

Consider the global temperature data, say x_t, shown in Figures 1.2 and 2.1. As discussed in Example 1.2, there is an apparent upward trend in the series that has been used to argue the global warming hypothesis. We might use simple linear regression to estimate that trend by fitting the model

$$x_t = \beta_1 + \beta_2 t + w_t, \quad t = 1880, 1857, \ldots, 2009.$$

This is in the form of the regression model (2.1) when we make the identification $q = 2$, $z_{t1} = 1$ and $z_{t2} = t$. Note that we are making the assumption that the errors, w_t, are an iid normal sequence, which may not be true. We will address this problem further in §2.3; the problem of autocorrelated errors is discussed in detail in §5.5. Also note that we could have used, for example, $t = 1, \ldots, 130$, without affecting the interpretation of the slope coefficient, β_2; only the intercept, β_1, would be affected.

Using simple linear regression, we obtained the estimated coefficients $\widehat{\beta}_1 = -11.2$, and $\widehat{\beta}_2 = .006$ (with a standard error of .0003) yielding a highly significant estimated increase of .6 degrees centigrade per 100 years. We discuss the precise way in which the solution was accomplished after the example. Finally, Figure 2.1 shows the global temperature data, say x_t, with the estimated trend, say $\widehat{x}_t = -11.2 + .006t$, superimposed. It is apparent that the estimated trend line obtained via simple linear regression does not quite capture the trend of the data and better models will be needed.

To perform this analysis in R, use the following commands:

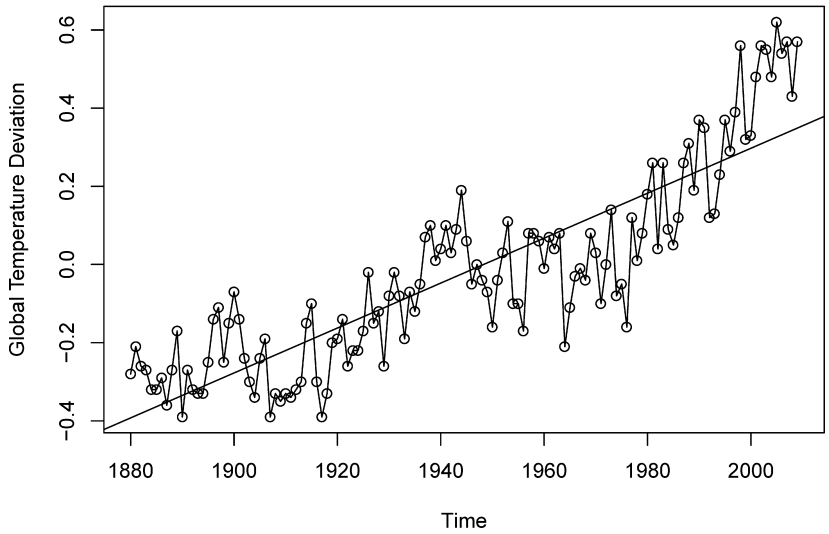

Fig. 2.1. Global temperature deviations shown in Figure 1.2 with fitted linear trend line.

```
summary(fit <- lm(gtemp~time(gtemp)))  # regress gtemp on time
plot(gtemp, type="o", ylab="Global Temperature Deviation")
abline(fit)  # add regression line to the plot
```

The linear model described by (2.1) above can be conveniently written in a more general notation by defining the column vectors $\boldsymbol{z}_t = (z_{t1}, z_{t2}, \ldots, z_{tq})'$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_q)'$, where $'$ denotes transpose, so (2.1) can be written in the alternate form

$$x_t = \boldsymbol{\beta}'\boldsymbol{z}_t + w_t. \tag{2.2}$$

where $w_t \sim$ iid $\mathrm{N}(0, \sigma_w^2)$. It is natural to consider estimating the unknown coefficient vector $\boldsymbol{\beta}$ by minimizing the error sum of squares

$$Q = \sum_{t=1}^{n} w_t^2 = \sum_{t=1}^{n} (x_t - \boldsymbol{\beta}'\boldsymbol{z}_t)^2, \tag{2.3}$$

with respect to $\beta_1, \beta_2, \ldots, \beta_q$. Minimizing Q yields the ordinary least squares estimator of $\boldsymbol{\beta}$. This minimization can be accomplished by differentiating (2.3) with respect to the vector $\boldsymbol{\beta}$ or by using the properties of projections. In the notation above, this procedure gives the normal equations

$$\left(\sum_{t=1}^{n} \boldsymbol{z}_t \boldsymbol{z}_t'\right) \widehat{\boldsymbol{\beta}} = \sum_{t=1}^{n} \boldsymbol{z}_t x_t. \tag{2.4}$$

The notation can be simplified by defining $Z = [\boldsymbol{z}_1 \,|\, \boldsymbol{z}_2 \,|\, \cdots \,|\, \boldsymbol{z}_n]'$ as the $n \times q$ matrix composed of the n samples of the input variables, the observed $n \times 1$ vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)'$ and the $n \times 1$ vector of errors

$\boldsymbol{w} = (w_1, w_2, \ldots, w_n)'$. In this case, model (2.2) may be written as

$$\boldsymbol{x} = Z\boldsymbol{\beta} + \boldsymbol{w}. \tag{2.5}$$

The normal equations, (2.4), can now be written as

$$(Z'Z)\,\widehat{\boldsymbol{\beta}} = Z'\boldsymbol{x} \tag{2.6}$$

and the solution

$$\widehat{\boldsymbol{\beta}} = (Z'Z)^{-1}Z'\boldsymbol{x} \tag{2.7}$$

when the matrix $Z'Z$ is nonsingular. The minimized error sum of squares (2.3), denoted SSE, can be written as

$$\begin{aligned} SSE &= \sum_{t=1}^{n}(x_t - \widehat{\boldsymbol{\beta}}'\boldsymbol{z}_t)^2 \\ &= (\boldsymbol{x} - Z\widehat{\boldsymbol{\beta}})'(\boldsymbol{x} - Z\widehat{\boldsymbol{\beta}}) \\ &= \boldsymbol{x}'\boldsymbol{x} - \widehat{\boldsymbol{\beta}}'Z'\boldsymbol{x} \\ &= \boldsymbol{x}'\boldsymbol{x} - \boldsymbol{x}'Z(Z'Z)^{-1}Z'\boldsymbol{x}, \end{aligned} \tag{2.8}$$

to give some useful versions for later reference. The ordinary least squares estimators are unbiased, i.e., $E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, and have the smallest variance within the class of linear unbiased estimators.

If the errors w_t are normally distributed, $\widehat{\boldsymbol{\beta}}$ is also the maximum likelihood estimator for $\boldsymbol{\beta}$ and is normally distributed with

$$\mathrm{cov}(\widehat{\boldsymbol{\beta}}) = \sigma_w^2\left(\sum_{t=1}^{n}\boldsymbol{z}_t\boldsymbol{z}_t'\right)^{-1} = \sigma_w^2(Z'Z)^{-1} = \sigma_w^2 C, \tag{2.9}$$

where

$$C = (Z'Z)^{-1} \tag{2.10}$$

is a convenient notation for later equations. An unbiased estimator for the variance σ_w^2 is

$$s_w^2 = MSE = \frac{SSE}{n-q}, \tag{2.11}$$

where MSE denotes the *mean squared error*, which is contrasted with the maximum likelihood estimator $\widehat{\sigma}_w^2 = SSE/n$. Under the normal assumption, s_w^2 is distributed proportionally to a chi-squared random variable with $n-q$ degrees of freedom, denoted by χ^2_{n-q}, and independently of $\widehat{\beta}$. It follows that

$$t_{n-q} = \frac{(\widehat{\beta}_i - \beta_i)}{s_w\sqrt{c_{ii}}} \tag{2.12}$$

has the t-distribution with $n-q$ degrees of freedom; c_{ii} denotes the i-th diagonal element of C, as defined in (2.10).

Table 2.1. Analysis of Variance for Regression

Source	df	Sum of Squares	Mean Square
$z_{t,r+1}, \ldots, z_{t,q}$	$q-r$	$SSR = SSE_r - SSE$	$MSR = SSR/(q-r)$
Error	$n-q$	SSE	$MSE = SSE/(n-q)$
Total	$n-r$	SSE_r	

Various competing models are of interest to isolate or select the best subset of independent variables. Suppose a proposed model specifies that only a subset $r < q$ independent variables, say, $\boldsymbol{z}_{t:r} = (z_{t1}, z_{t2}, \ldots, z_{tr})'$ is influencing the dependent variable x_t. The reduced model is

$$\boldsymbol{x} = Z_r \boldsymbol{\beta}_r + \boldsymbol{w} \tag{2.13}$$

where $\boldsymbol{\beta}_r = (\beta_1, \beta_2, \ldots, \beta_r)'$ is a subset of coefficients of the original q variables and $Z_r = [\boldsymbol{z}_{1:r} \mid \cdots \mid \boldsymbol{z}_{n:r}]'$ is the $n \times r$ matrix of inputs. The null hypothesis in this case is H_0: $\beta_{r+1} = \cdots = \beta_q = 0$. We can test the reduced model (2.13) against the full model (2.2) by comparing the error sums of squares under the two models using the F-statistic

$$F_{q-r,n-q} = \frac{(SSE_r - SSE)/(q-r)}{SSE/(n-q)}, \tag{2.14}$$

which has the central F-distribution with $q-r$ and $n-q$ degrees of freedom when (2.13) is the correct model. Note that SSE_r is the error sum of squares under the reduced model (2.13) and it can be computed by replacing Z with Z_r in (2.8). The statistic, which follows from applying the likelihood ratio criterion, has the improvement per number of parameters added in the numerator compared with the error sum of squares under the full model in the denominator. The information involved in the test procedure is often summarized in an Analysis of Variance (ANOVA) table as given in Table 2.1 for this particular case. The difference in the numerator is often called the regression sum of squares

In terms of Table 2.1, it is conventional to write the F-statistic (2.14) as the ratio of the two mean squares, obtaining

$$F_{q-r,n-q} = \frac{MSR}{MSE}, \tag{2.15}$$

where MSR, the *mean squared regression*, is the numerator of (2.14). A special case of interest is $r = 1$ and $z_{t1} \equiv 1$, when the model in (2.13) becomes

$$x_t = \beta_1 + w_t,$$

and we may measure the proportion of variation accounted for by the other variables using

$$R^2 = \frac{SSE_1 - SSE}{SSE_1}, \tag{2.16}$$

where the residual sum of squares under the reduced model

$$SSE_1 = \sum_{t=1}^{n} (x_t - \bar{x})^2, \tag{2.17}$$

in this case is just the sum of squared deviations from the mean $\bar{x}$. The measure R^2 is also the *squared multiple correlation* between x_t and the variables $z_{t2}, z_{t3}, \ldots, z_{tq}$.

The techniques discussed in the previous paragraph can be used to test various models against one another using the F test given in (2.14), (2.15), and the ANOVA table. These tests have been used in the past in a stepwise manner, where variables are added or deleted when the values from the F-test either exceed or fail to exceed some predetermined levels. The procedure, called stepwise multiple regression, is useful in arriving at a set of useful variables. An alternative is to focus on a procedure for model selection that does not proceed sequentially, but simply evaluates each model on its own merits. Suppose we consider a normal regression model with k coefficients and denote the maximum likelihood estimator for the variance as

$$\widehat{\sigma}_k^2 = \frac{SSE_k}{n}, \tag{2.18}$$

where SSE_k denotes the residual sum of squares under the model with k regression coefficients. Then, Akaike (1969, 1973, 1974) suggested measuring the goodness of fit for this particular model by balancing the error of the fit against the number of parameters in the model; we define the following.[1]

Definition 2.1 Akaike's Information Criterion (AIC)

$$\text{AIC} = \log \widehat{\sigma}_k^2 + \frac{n + 2k}{n}, \tag{2.19}$$

where $\widehat{\sigma}_k^2$ is given by (2.18) and k is the number of parameters in the model.

The value of k yielding the minimum AIC specifies the best model. The idea is roughly that minimizing $\widehat{\sigma}_k^2$ would be a reasonable objective, except that it decreases monotonically as k increases. Therefore, we ought to penalize the error variance by a term proportional to the number of parameters. The choice for the penalty term given by (2.19) is not the only one, and a considerable literature is available advocating different penalty terms. A corrected

[1] Formally, AIC is defined as $-2 \log L_k + 2k$ where L_k is the maximized log-likelihood and k is the number of parameters in the model. For the normal regression problem, AIC can be reduced to the form given by (2.19). AIC is an estimate of the Kullback-Leibler discrepency between a true model and a candidate model; see Problems 2.4 and 2.5 for further details.

form, suggested by Sugiura (1978), and expanded by Hurvich and Tsai (1989), can be based on small-sample distributional results for the linear regression model (details are provided in Problems 2.4 and 2.5). The corrected form is defined as follows.

Definition 2.2 AIC, Bias Corrected (AICc)

$$\text{AICc} = \log\ \widehat{\sigma}_k^2 + \frac{n+k}{n-k-2}, \tag{2.20}$$

where $\widehat{\sigma}_k^2$ is given by (2.18), k is the number of parameters in the model, and n is the sample size.

We may also derive a correction term based on Bayesian arguments, as in Schwarz (1978), which leads to the following.

Definition 2.3 Bayesian Information Criterion (BIC)

$$\text{BIC} = \log\ \widehat{\sigma}_k^2 + \frac{k \log n}{n}, \tag{2.21}$$

using the same notation as in Definition 2.2.

BIC is also called the Schwarz Information Criterion (SIC); see also Rissanen (1978) for an approach yielding the same statistic based on a minimum description length argument. Various simulation studies have tended to verify that BIC does well at getting the correct order in large samples, whereas AICc tends to be superior in smaller samples where the relative number of parameters is large; see McQuarrie and Tsai (1998) for detailed comparisons. In fitting regression models, two measures that have been used in the past are adjusted R-squared, which is essentially s_w^2, and Mallows C_p, Mallows (1973), which we do not consider in this context.

Example 2.2 Pollution, Temperature and Mortality

The data shown in Figure 2.2 are extracted series from a study by Shumway et al. (1988) of the possible effects of temperature and pollution on weekly mortality in Los Angeles County. Note the strong seasonal components in all of the series, corresponding to winter-summer variations and the downward trend in the cardiovascular mortality over the 10-year period.

A scatterplot matrix, shown in Figure 2.3, indicates a possible linear relation between mortality and the pollutant particulates and a possible relation to temperature. Note the curvilinear shape of the temperature mortality curve, indicating that higher temperatures as well as lower temperatures are associated with increases in cardiovascular mortality.

Based on the scatterplot matrix, we entertain, tentatively, four models where M_t denotes cardiovascular mortality, T_t denotes temperature and P_t denotes the particulate levels. They are

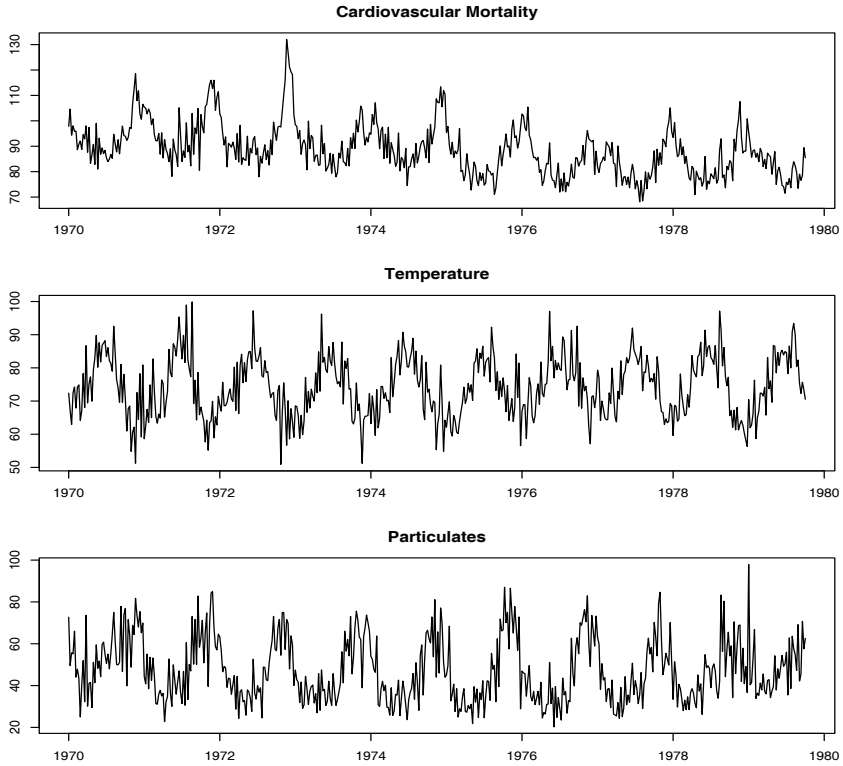

Fig. 2.2. Average weekly cardiovascular mortality (top), temperature (middle) and particulate pollution (bottom) in Los Angeles County. There are 508 six-day smoothed averages obtained by filtering daily values over the 10 year period 1970-1979.

$$M_t = \beta_1 + \beta_2 t + w_t \tag{2.22}$$
$$M_t = \beta_1 + \beta_2 t + \beta_3 (T_t - T.) + w_t \tag{2.23}$$
$$M_t = \beta_1 + \beta_2 t + \beta_3 (T_t - T.) + \beta_4 (T_t - T.)^2 + w_t \tag{2.24}$$
$$M_t = \beta_1 + \beta_2 t + \beta_3 (T_t - T.) + \beta_4 (T_t - T.)^2 + \beta_5 P_t + w_t \tag{2.25}$$

where we adjust temperature for its mean, $T. = 74.6$, to avoid scaling problems. It is clear that (2.22) is a trend only model, (2.23) is linear temperature, (2.24) is curvilinear temperature and (2.25) is curvilinear temperature and pollution. We summarize some of the statistics given for this particular case in Table 2.2. The values of R^2 were computed by noting that $SSE_1 = 50,687$ using (2.17).

We note that each model does substantially better than the one before it and that the model including temperature, temperature squared, and particulates does the best, accounting for some 60% of the variability and with the best value for AIC and BIC (because of the large sample size, AIC

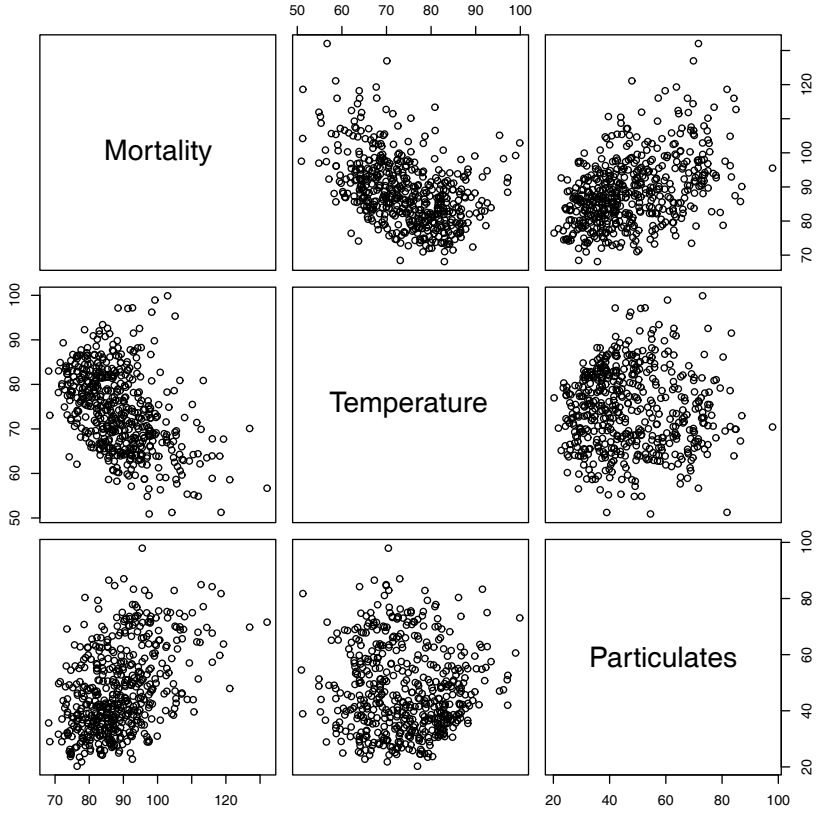

Fig. 2.3. Scatterplot matrix showing plausible relations between mortality, temperature, and pollution.

Table 2.2. Summary Statistics for Mortality Models

Model	k	SSE	df	MSE	R^2	AIC	BIC
(2.22)	2	40,020	506	79.0	.21	5.38	5.40
(2.23)	3	31,413	505	62.2	.38	5.14	5.17
(2.24)	4	27,985	504	55.5	.45	5.03	5.07
(2.25)	5	20,508	503	40.8	.60	4.72	4.77

and AICc are nearly the same). Note that one can compare any two models using the residual sums of squares and (2.14). Hence, a model with only trend could be compared to the full model using $q = 5, r = 2, n = 508$, so

$$F_{3,503} = \frac{(40,020 - 20,508)/3}{20,508/503} = 160,$$

which exceeds $F_{3,503}(.001) = 5.51$. We obtain the best prediction model,

$$\widehat{M}_t = 81.59 - .027_{(.002)}t - .473_{(.032)}(T_t - 74.6) + .023_{(.003)}(T_t - 74.6)^2 + .255_{(.019)}P_t,$$

for mortality, where the standard errors, computed from (2.9)-(2.11), are given in parentheses. As expected, a negative trend is present in time as well as a negative coefficient for adjusted temperature. The quadratic effect of temperature can clearly be seen in the scatterplots of Figure 2.3. Pollution weights positively and can be interpreted as the incremental contribution to daily deaths per unit of particulate pollution. It would still be essential to check the residuals $\widehat{w}_t = M_t - \widehat{M}_t$ for autocorrelation (of which there is a substantial amount), but we defer this question to to §5.6 when we discuss regression with correlated errors.

Below is the R code to plot the series, display the scatterplot matrix, fit the final regression model (2.25), and compute the corresponding values of AIC, AICc and BIC.[2] Finally, the use of `na.action` in `lm()` is to retain the time series attributes for the residuals and fitted values.

```
par(mfrow=c(3,1))
plot(cmort, main="Cardiovascular Mortality", xlab="", ylab="")
plot(tempr, main="Temperature", xlab="", ylab="")
plot(part, main="Particulates", xlab="", ylab="")
dev.new()  # open a new graphic device for the scatterplot matrix
pairs(cbind(Mortality=cmort, Temperature=tempr, Particulates=part))
temp = tempr-mean(tempr)  # center temperature
temp2 = temp^2
trend = time(cmort)  # time
fit = lm(cmort~ trend + temp + temp2 + part, na.action=NULL)
summary(fit)  # regression results
summary(aov(fit))  # ANOVA table   (compare to next line)
summary(aov(lm(cmort~cbind(trend, temp, temp2, part)))) # Table 2.1
num = length(cmort)  # sample size
AIC(fit)/num - log(2*pi)  # AIC
AIC(fit, k=log(num))/num - log(2*pi)  # BIC
(AICc = log(sum(resid(fit)^2)/num) + (num+5)/(num-5-2)) # AICc
```

As previously mentioned, it is possible to include lagged variables in time series regression models and we will continue to discuss this type of problem throughout the text. This concept is explored further in Problems 2.2 and 2.11. The following is a simple example of lagged regression.

[2] The easiest way to extract AIC and BIC from an `lm()` run in R is to use the command `AIC()`. Our definitions differ from R by terms that do not change from model to model. In the example, we show how to obtain (2.19) and (2.21) from the R output. It is more difficult to obtain AICc.

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,

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

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

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

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

STAT 626: Outline of Regrssion Modeling

Simple & Multiple Regressions

Overview of Iterative Regression Modeling

(i) **Model Formulation:** Use Scatterplots, Substantative Knowlegde, Transformation,...,

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim i.i.d. \quad N(0, \sigma^2), \quad i = 1, 2, \ldots, n.$$

(ii) **Model Estimation:** Given data $(x_i, y_i), \quad i = 1, \ldots, n$, minimizing

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2,$$

over the parameters (β_0, β_1), one obtains the least square estimate (LSE):

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad \hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \sum a_i y_i.$$

Fitted Values and Residuals: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \quad e_i = y_i - \hat{y}_i, \quad i = 1, \ldots, n.$

(iii) **Model Checking:** Plot Residuals vs Fitted values, Histogram of Residuals,...

SST=SSR+SSE, Coefficient of Determination r^2,....

Inference: $E(\hat{\beta}_1) = \beta_1, \mathrm{V}(\hat{\beta}_1) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}, S_{\hat{\beta}_1} = \frac{\hat{\sigma}}{\sqrt{\sum (x_i - \bar{x})^2}}, \hat{\sigma} = ?.$

Distribution: $\hat{\beta}_1 \sim N(\beta_1, \mathrm{V}(\hat{\beta}_1))$.

Test Statistic for Testing $H_0 : \beta_1 = 0$, vs $H_a : \beta_1 \neq 0$? $\quad T = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \sim t_{n-2}$.

Confidence Interval (CI) for β_1? $\quad \hat{\beta}_1 \pm t_{\alpha/2, n-2} S_{\hat{\beta}_1}$.

Confidence Interval (CI) for the Mean Response at x_0 or $\mu_{Y.x_0}$:

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\alpha/2, n-2} SE(x_0), \quad SE(x_0) = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}.$$

Prediction Interval (PI) for a Future Value at x_0?:

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\alpha/2, n-2} SE_{\text{pred}}(x_0), \qquad SE_{\text{pred}}(x_0) = \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}.$$

1. Multiple Regression; $Y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon$,

2. Model Assumptions on ε_i's?

3. Full and Reduced Models,

4. Using R, and Data Examples,

5. Model Selection, R^2, AIC,...

6. Standardized Residuals, $e_i^*, \quad i = 1, \ldots, n$.

7. Influential Observations, Cook's Distance,....

Review of Basic PROBAB & STAT

8. If $X_1, \ldots, X_n$ are independent with $E(X_i) = \mu_i, \mathrm{Var}(X_i) = \sigma_i^2$, then
   (a) $E(a_1 X_1 + a_2 X_2) =$??
   (b) $\mathrm{Var}(a_1 X_1 + a_2 X_2) =$??
   (c) Compute $\mathrm{Var}(\bar{X}) =$??

9. A Bonus Problem: Get 5 points for HW; e-mail me typed correct solution by midnight July 4, 2016). Given $x_i, i = 1, \ldots, n$, find c to minimize the function

$$f(c) = \sum_{i=1}^{n} |x_i - c|.$$