Statistical Time Series Analysis

STATISTICS 6555

Summer 2013

Contents

Chapter 1

Introduction

What Constitutes a Time Series

Time Series Analysis: the analysis of experimental data that have been observed at different points in time (usually equally spaced time points).

The obvious correlation introduced by the sampling of adjacent points in time can severely restrict the applicability of conventional statistical methods which are based on independent and identically distributed data.

The first eight weeks or so of this course will cover statistical techniques to forecast future observations based on past observations.

Basic Methodology:

1. a "pattern" is first attained from the data at hand
2. the "pattern" is then extrapolated into the future to prepare a forecast

Assumptions? The "pattern" we've observed will continue...

Some applications:

- predicting/forecasting future gas prices
- global warming? Predicting future global temperatures
- population growth

Components of a Time Series

- Trend: the long-run upward or downward movement of the series
- Cycle: the upward or downward movement of the series around the trend (think of a wave)
- Seasonal Variations: patterns in the data that follow yearly patterns (think of seasonal temperatures)
- Irregular Variations: the remaining erratic movements in the series that cannot be accounted for

Our goal is to estimate the trend, cycle, and seasonal components of a time series so that all that is left is irregular fluctuations (often referred to as white noise).

Chapter 2

Review: Regression Analysis Using R

2.1 Simple Linear Regression

We use ordinary least squares regression to relate a continuous, response variable (y) to a continuous, explanatory variable (x). The basic model:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

- β_0 is the y-intercept coefficient. Interpreted as the average value of y when x=0 although sometimes there is not valid interpretation
- β_1 is the slope coefficient. Interpreted as the change in average value of y for a one-unit increase in x
- ε is the individual error term. Assumed to be IID $N(0, \sigma^2)$

Always start by plotting the data to make sure a linear relationship exists. Then use the `lm()` function in R to analyze the relationship.

Work through Example 3.1 from the text: see corresponding R code

2.2 Multiple Linear Regression

The model depends on the nature of the explanatory variables.

1. If the p predictors $(x_1, \ldots, x_p)$ are all continuous the model is an extension of the one above
$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon$$
2. If you have categorical variables, you will need to introduce a set of indicator variables for each categorical variable. If your categorical variable has L levels, you will need $L - 1$ indicator variables
    - One Continuous and One Categorical Variable:
$$y = \beta_0 + \beta_1 x_1 + \beta_{D_1} \mathbf{I}\{D_1\} + \beta_{D_2} \mathbf{I}\{D_2\} + \cdots + \beta_{D_{L-1}} \mathbf{I}\{D_{L-1}\} + \varepsilon$$

where

$$\mathbf{1}\{D_i\} = \begin{cases} 1 & \text{If } t \text{ is Level } i \\ 0 & \text{Otherwise} \end{cases}$$

Work through Example 4.15 from the text: see corresponding R code

3. The linear model allows functions of your predictor variables as long as the coefficients remain linear.

$$y = \beta_0 + \beta_1 f(x_1) + \cdots + \beta_p f(x_p) + \varepsilon$$

For instance, a k-th degree polynomial:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \cdots + \beta_k x_1^k + \varepsilon$$

Work through Example 4.12 from the text: see corresponding R code

2.3 Residual Analysis

For any particular observed value of y, the corresponding residual is

$$e = y - \hat{y}$$

where the predicted value of y is calculated using the least squares prediction equation

$$\hat{y} = b_0 + b_1 x$$

If the regression assumption hold, the residuals should look like they have been randomly and independently selected from a normal distribution with mean 0 and variance σ^2.

2.3.1 Residual Plots

The calculated residuals should be plotted against

1. The values of each independent variable
2. The fitted values ($\hat{y}$)

When examining the residual plots, they should appear randomly distributed about the zero line with no discernible pattern. Otherwise, this would suggest the assumption are violated.

2.3.2 Normality Assumption

The residuals are assumed to follow a normal distribution, so you should check them for normality by creating a histogram and constructing a normality plot. You should also use some of the many normality tests to check this assumption.

2.3.3 Constant Variance Assumption

To check this assumption, we examine the residual plots. If the plots appear to "fan out" or "funnel in" we would say that variance does not appear constant. In general the residual plots should form relatively horizontal bands within which most of the residuals appear.

2.3.4 Independence Assumption

To check independence, we plot the residuals against indexed order (i.e. the order in which the data was collected), looking for a discernible pattern. This usually occurs with time-dependent data.

You should work through other examples in chapters 3-5 using R to analyze the data and trying to recreate the results produced in the book.

Chapter 3

The Basics

3.1 Stationary Time Series

A stationary time series is one for which the statistical behavior of a set of observations $x_{t_1}, x_{t_2}, \ldots x_{t_k}$ is identical to that of the shifted set of observations $x_{t_{1+h}}, x_{t_{2+h}}, \ldots x_{t_{k+h}}$ for any collection of time points $t_1, t_2, \ldots t_k$ and for any shift h (lag).

This is the definition of *strong stationarity* and is too strong for most applications.

Weak Stationarity:

There is a relaxed definition, referred to as *weak stationarity* which requires only that the first and second moments satisfy the following constraints:

1. $$E(x_t) = \mu \qquad \forall t$$

2. $$\gamma_x(h) = E[(x_{t+h} - \mu)(x_t - \mu)]$$

where E is the usual expectation over the population density, h is the time shift (lag), and $\gamma_x(h)$ is called the autocovariance function and we additionally assert that

$$\gamma_x(h) = \gamma_x(-h)$$

Example:

Let w_t be *white noise* such that $w_t \overset{iid}{\sim} N(0,1)$ and let x_t be a first-order moving averge series (to be discussed in more depth later)

$$x_t = w_t - 0.9w_{t-1}$$

Lets show that:

$$E(x_t) = 0$$

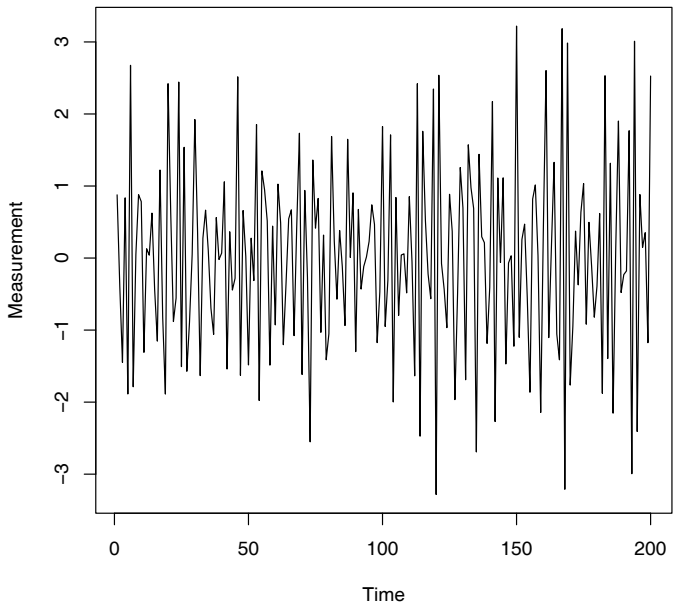
and

$$\gamma_x(h) = \begin{cases} 1 + 0.9^2 & \text{if } h = 0 \\ -0.9 & \text{if } h = \pm 1 \\ 0 & \text{if } |h| \geq 2 \end{cases}$$

$$
\begin{aligned}
E(x_t) &= E(w_t - 0.9w_{t-1}) \\
&= E(w_t) - 0.9E(w_{t-1}) \\
&= 0 - 0.9(0) \\
&= 0 \\
\gamma_x(0) &= E[(x_t - \mu)(x_t - \mu)] \\
&= E[(w_t - 0.9w_{t-1})(w_t - 0.9w_{t-1})] \\
&= E[w_t^2 - 0.9w_t w_t - 0.9w_{t-1}w_t + 0.9^2 w_{t-1}^2] \\
&= E[w_t^2] - 0.9E[w_t]E[w_t] - 0.9E[w_{t-1}]E[w_t] + 0.9^2 E[w_{t-1}^2] \\
&= 1 + 0.9^2 \\
\gamma_x(1) &= E[(x_{t+1} - \mu)(x_t - \mu)] \\
&= E[(w_{t+1} - 0.9w_t)(w_t - 0.9w_{t-1})] \\
&= E[w_{t+1}w_t - 0.9w_{t+1}w_t - 0.9(w_t)^2 + 0.9^2 w_t w_{t-1}] \\
&= E[w_{t+1}]E[w_t] - 0.9E[w_{t+1}]E[w_t] - 0.9E[(w_t)^2] + 0.9^2 E[w_t]E[w_{t-1}] \\
&= -0.9 \\
\gamma_x(2) &= E[(x_{t+2} - \mu)(x_t - \mu)] \\
&= E[(w_{t+2} - 0.9w_{t+1})(w_t - 0.9w_{t-1})] \\
&= E[w_{t+2}w_t - 0.9w_{t+2}w_t - 0.9w_{t+1}w_t + 0.9^2 w_{t+1}w_{t-1}] \\
&= E[w_{t+1}]E[w_t] - 0.9E[w_{t+1}]E[w_t] - 0.9E[w_{t+1}]E[w_t] + 0.9^2 E[w_t]E[w_{t-1}] \\
&= 0
\end{aligned}
$$

R Code:

```
sample.MA1 = arima.sim(list(order=c(0,0,1), ma=-.9),200)
plot(sample.MA1,xlab="Time",ylab="Measurement")
```



Autocorrelation Function:

Closely related to the autocovariance function is the autocorrelation function defined by

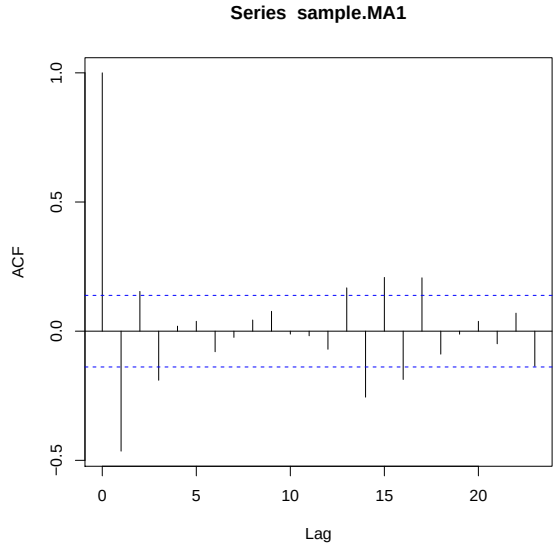
$$\rho_x(h) = \frac{\gamma_x(h)}{\gamma_x(0)} \qquad -1 < \rho_x(h) < 1$$

Much like the relationship between covariance and correlation in standard statistics, autocorrelation is roughly a standardized version of the autocovariance function.

ACF for the above example:

$$\rho_x(h) = \begin{cases} \frac{(1+0.9^2)}{(1+0.9^2)} = 1 & \text{if } h = 0 \\ \frac{-0.9}{(1+0.9^2)} \approx -0.5 & \text{if } h = \pm 1 \\ \frac{0}{(1+0.9^2)} = 0 & \text{if } |h| \geq 2 \end{cases}$$

```
acf(sample.MA1)
```



Practice Problems: Are the following weakly stationary? Find the expected value, the autocovariance function and autocorrelation function for x_t where

1. $x_t = w_{t-1} + 2w_t + w_{t+1}$
2. $x_t = w_t - \theta w_{t-1} + u_t$

Where

$$w_t \overset{iid}{\sim} N(0, \sigma_w^2) \text{ and } u_t \overset{iid}{\sim} N(0, \sigma_u^2)$$

Practice Problems Solutions: We need only focus on the cross product terms that refer to the same time point, I've excluded many of the cross product terms because when they are different time points, they are independent and the expectation is zero based on the distribution of w_t and u_t, these terms would be in the $\cdots$

1. $x_t = w_{t-1} + 2w_t + w_{t+1}$

$$\begin{aligned}
E(x_t) &= E(w_{t-1} + 2w_t + w_{t+1}) \\
&= E(w_{t-1}) + 2E(w_t) + E(w_{t+1}) \\
&= 0 + 2(0) + 0 \\
&= 0 \\
\gamma_x(0) &= E[(x_t - \mu)(x_t - \mu)] \\
&= E[(w_{t-1} + 2w_t + w_{t+1})(w_{t-1} + 2w_t + w_{t+1})] \\
&= E[w_{t-1}^2 + 4w_t^2 + w_{t+1}^2 + \cdots] \\
&= E[w_{t-1}^2] + 4E[w_t^2] + E[w_{t+1}^2] + E[\cdots] \\
&= \sigma_w^2 + 4\sigma_w^2 + \sigma_w^2 + 0 \\
&= 6\sigma_w^2 \\
\gamma_x(1) &= E[(x_{t+1} - \mu)(x_t - \mu)] \\
&= E[(w_t + 2w_{t+1} + w_{t+2})(w_{t-1} + 2w_t + w_{t+1})] \\
&= E[2w_t^2 + 2w_{t+1}^2 + \cdots] \\
&= 2E[w_t^2] + 2E[w_{t+1}^2] + E[\cdots] \\
&= 2\sigma_w^2 + 2\sigma_w^2 + 0 \\
&= 4\sigma_w^2 \\
\gamma_x(2) &= E[(w_{t+1} + 2w_{t+2} + w_{t+3})(w_{t-1} + 2w_t + w_{t+1})] \\
&= E[w_{t+1}^2 + \cdots] \\
&= E[w_{t+1}^2] + E[\cdots] \\
&= \sigma_w^2 + 0 \\
&= \sigma_w^2
\end{aligned}$$

2. $x_t = w_t - \theta w_{t-1} + u_t$

$$\begin{aligned}
E(x_t) &= E(w_t - \theta w_{t-1} + u_t) \\
&= E(w_t) - \theta E(w_{t-1}) + E(u_t) \\
&= 0 + \theta(0) + 0 \\
&= 0 \\
\gamma_x(0) &= E[(x_t - \mu)(x_t - \mu)] \\
&= E[(w_t - \theta w_{t-1} + u_t)(w_t - \theta w_{t-1} + u_t)] \\
&= E[w_t^2 + \theta^2 w_{t-1}^2 + u_t^2 + \cdots] \\
&= E[w_t^2] + \theta^2 E[w_{t-1}^2] + E[u_t^2] + \cdots] \\
&= \sigma_w^2 + \theta^2\sigma_w^2 + \sigma_u^2 + 0 \\
&= (1 + \theta^2)\sigma_w^2 + \sigma_u^2 \\
\gamma_x(1) &= E[(x_{t+1} - \mu)(x_t - \mu)] \\
&= E[(w_{t+1} - \theta w_t + u_{t+1})(w_t - \theta w_{t-1} + u_t)] \\
&= E[-\theta w_t^2 + \cdots] \\
&= -\theta\sigma_w^2 + 0 \\
&= -\theta\sigma_w^2
\end{aligned}$$

3.2 Dealing with Nonstationarity:

Most time series are not stationary to begin with. We must modify the series to improve the approximation of stationarity by

- detrending
- differencing
- transformations
- linear filters

3.2.1 Detrending

The general version of the nonstationary time series given above is to assume a general linear trend of the form

$$x_t = \beta_0 + \beta_1 t + \varepsilon_t$$

The natural thing to do is to consider the residual

$$\hat{e}_t = x_t - \hat{\beta}_0 - \hat{\beta}_1 t$$

as a plausible stationary series where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimated intercept and slope coefficients based on least squares estimation.

This is referred to as **Time Series Regression** and will be covered in more depth.

3.2.2 Differencing

A common first step for achieving stationity is with the first difference

$$\bigtriangledown x_t = x_t - x_{t-1}$$

This method of transforming to stationarity is common also for series with trend.

Higher order differences are defined as successive applications of the operator $\bigtriangledown$. For example, the second difference is

$$\begin{aligned} \bigtriangledown^2 x_t &= \bigtriangledown[\bigtriangledown x_t] \\ &= \bigtriangledown[x_t - x_{t-1}] \\ &= [x_t - x_{t-1}] - [x_{t-1} - x_{t-2}] \\ &= x_t - 2x_{t-1} + x_{t-2} \end{aligned}$$

If the model also contains a quadratic trend term, it is easy to show that taking the second difference usually reduces the model to a stationary form.

Example:

Consider the standard linear regression model

$$x_t = \beta_0 + \beta_1 t + \varepsilon_t$$

where

$$\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$$

The first difference is

$$\begin{aligned} \bigtriangledown x_t &= (\beta_0 + \beta_1 t + e_t) - (\beta_0 + \beta_1(t-1) + e_{t-1}) \\ &= \beta_1 + e_t - e_{t-1} \end{aligned}$$

Is $\bigtriangledown x_t$ stationary? Find $E(\bigtriangledown x_t)$, $\gamma_{\bigtriangledown x_t}(h)$

$$\begin{aligned} E(\bigtriangledown x_t) &= E(\beta_1 + e_t - e_{t-1}) \\ &= E(\beta_1) + E(e_t) - E(e_{t-1}) \\ &= \beta_1 + 0 + 0 \\ &= \beta_1 \\ \gamma_{\bigtriangledown x_t}(0) &= E[(\bigtriangledown x_t - \mu)(\bigtriangledown x_t - \mu)] \\ &= E[(\beta_1 + e_t - e_{t-1} - \beta_1)(\beta_1 + e_t - e_{t-1} - \beta_1) \\ &= E[e_t^2 + e_{t-1}^2 + \cdots] \\ &= E[e_t^2] + E[e_{t-1}^2] + E[\cdots] \\ &= \sigma^2 + \sigma^2 + 0 \\ &= 2\sigma^2 \\ \gamma_{\bigtriangledown x_t}(1) &= E[(\bigtriangledown x_{t+1} - \mu)(\bigtriangledown x_t - \mu)] \\ &= E[(\beta_1 + e_{t+1} - e_t - \beta_1)(\beta_1 + e_t - e_{t-1} - \beta_1) \\ &= E[-e_t^2 + \cdots] \\ &= -E[e_t^2] + E[\cdots] \\ &= -\sigma^2 \end{aligned}$$

Higher Order Differences:

In order to find higher order differences, we must first define the **backshift operator notation (B)**

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

and extend it to powers

$$B^2 x_t = B(Bx_t) = Bx_{t-1} = x_{t-2} \longrightarrow B^k x_t = x_{t-k}$$

then higher order differences are defined as

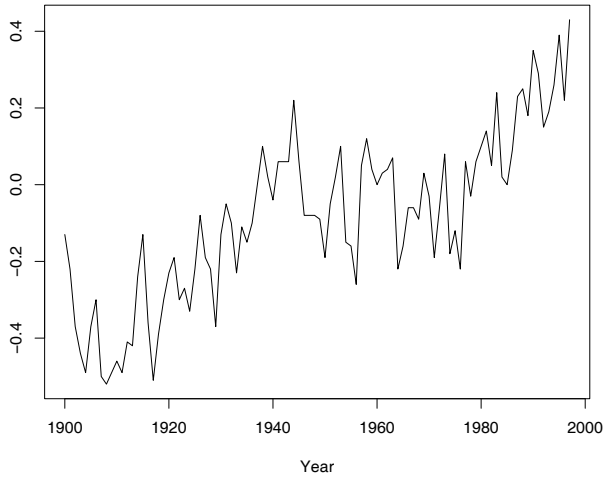
$$\bigtriangledown^d x_t = (1 - B)^d x_t$$

Lets find the fourth difference $\bigtriangledown^4 x_t$, making use of the backshift operator.

$$\begin{aligned} \bigtriangledown^4 x_t &= (1 - B)^4 x_t \\ &= (1 - 4B + 6B^2 - 4B^3 + B^4) x_t \\ &= x_t - 4x_{t-1} + 6x_{t-2} - 4x_{t-3} + x_{t-4} \end{aligned}$$

Example: Global Temperatures

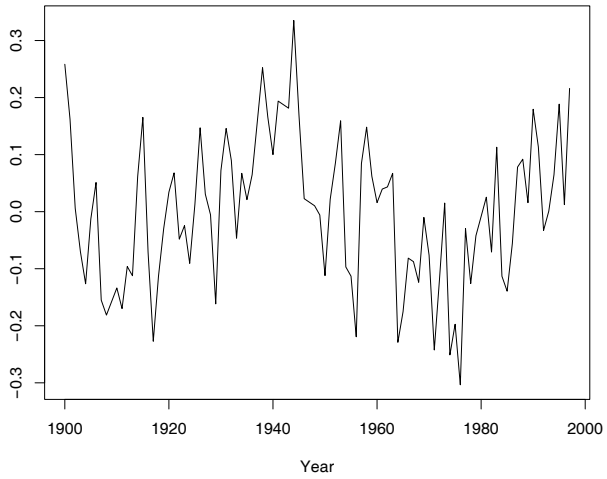
Consider a global temperatures series: the data are a combination of land-air average temperature anomalies for the years 1900-1997.



- The apparent upward trend over time has been used as an argument for the global warming hypothesis
- Note the leveling off at about 1935, then another sharp upward trend at about 1970
- Sometimes hard to check constant variance in presence of trend

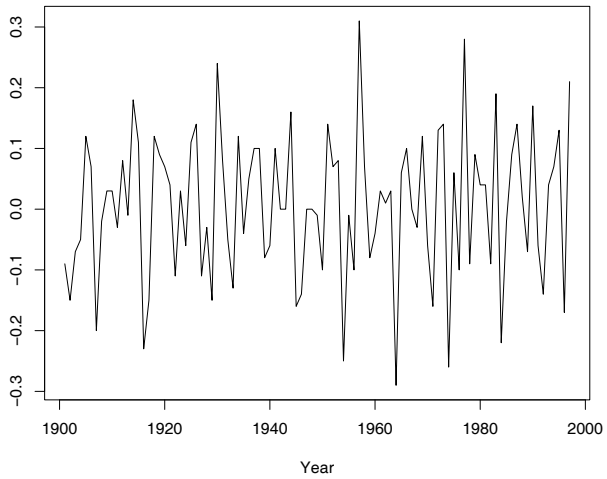
We first look at detrending the time series and examine the residuals from the model

$$\widehat{Temp} = -12.2 + 0.006 * Years$$



- The detrended series improves the behavior of the series, however does not eliminate the nonstationary behavior
- Appears to be a long cyclical trend which needs to be eliminated

Next, look at differencing the time series with the first difference



- The first difference of the series does not contain the long middle cycle observed in the detrended series
- The behavior of this series seems more stationary making it superior to the detrended model

3.2.3 Transformations

A transformation that cuts down the values of larger peaks of a time series and emphasizes the lower values may be effective in reducing nonstationary behavior due to changing variance.

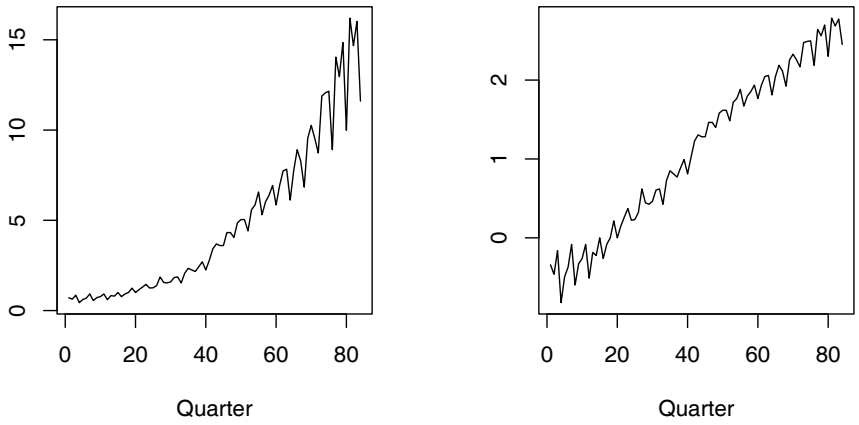
Examples:

- The logarithmic transformation, $y_t = log(x_t)$ usually exponential-base (natural log)
- The square root transformation, $y_t = \sqrt{x_t}$, useful for count data
- More general transformations fall within the Box-Cox family

$$y_t = \begin{cases} \frac{1}{\lambda}(x_t^{\lambda} - 1) & \text{if } \lambda \neq 0 \\ \ln(x_t) & \text{if } \lambda = 0 \end{cases}$$

Example: Johnson & Johnson Data

The following figures shows quarterly earnings per share for Johnson&Johnson from 1960 to 1980 before and after a log transformation.



3.2.4 Linear Filters

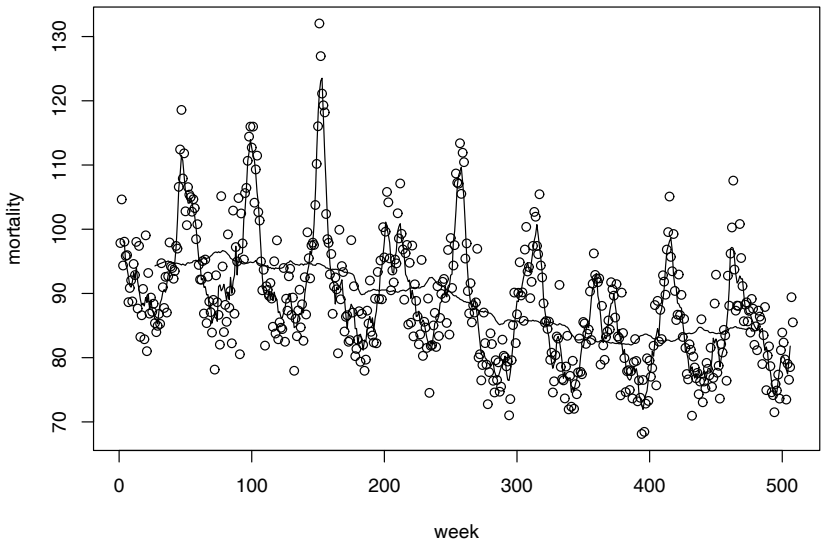
We may defne general linear filters to do kinds of smoothing and roughening of a time series to enhance signals and attenuate noise. Consider the general linear combination of past and future values of a time series given as

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

where $a_j, j = 0, \pm 1, \pm 2, \ldots$ define a set of fixed filter coefficients to be applied to the series of interest. An example of a linear filter is the first difference with $a_0 = 1, a_1 = -1, a_j = 0$ otherwise

Example: Cardiovascular Mortality Rates

The following plot shows the average weekly cardiovascular mortality in Los Angeles County between 1970-1979. Within the data are two moving average filters, a 4 week average (i.e. monthly) and a 52 week average (i.e. yearly)



Chapter 4

Time Series Regression

Most useful when the parameters describing the time series to be forecast remain constant over time.

4.1 Trend Model

$$x_t = TR_t + \epsilon_t \qquad where \qquad \epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$$

1. No Trend
$$TR_t = \beta_0$$
2. Linear Trend
$$TR_t = \beta_0 + \beta_1 t$$
3. Quadratic Trend
$$TR_t = \beta_0 + \beta_1 t + \beta_2 t^2$$
4. Pth Order Polynomial Trend
$$TR_t = \beta_0 + \beta_1 t + \cdots + \beta_P t^p$$

Regardless of the specific trend model, we use standard least squares regression techniques to model the time series. After we've fit our model, the residuals should be weakly stationary, normal distributed. If not, there is a missing element to our model that needs to be captured such as a seasonal or cyclical factor.

4.2 Detecting Autocorrelation

1. All of the above models are assuming that the error terms are independent
2. What if the errors are correlated such that $\rho(h)$ is significant at some lag h
3. First order autocorrelation
$$\varepsilon_t = \phi\varepsilon_{t-1} + a_t \qquad \text{where} \qquad a_t \overset{iid}{\sim} N(0, \sigma^2)$$
   - Can check informally in R using `acf()`

- We can also test for first order autocorrelation using the Durbin-Watson test on the residuals from the fitted model:

  Hypotheses:

  $$\begin{aligned} H_0: &\quad \text{Error terms are not correlated} \\ H_A: &\quad \text{Error terms are correlated} \end{aligned}$$

  Test Statistic:

  $$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=2}^{n} e_t^2}$$

  The value of d is compared with critical values of d provided in Tables A5 and A6 of the textbook.
- If autocorrelation is present we need to estimate ϕ using Box-Jenkins methods.

4.3 Adding a Seasonality Factor

$$x_t = TR_t + SN_t + \epsilon_t \qquad where \qquad \epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$$

1. Main assumption: constant seasonal variation for basic time series regression
2. If there appears to be changing seasonal variation, apply a transformation so that the variation is relatively constant.
   - Power Transformation

     $$x_t^* = (x_t)^\lambda \qquad where \qquad 0 < \lambda < 1$$

   - Natural logarithm transformation:

     $$\lambda \longrightarrow 0 \qquad (x_t)^\lambda \longrightarrow ln(x_t)$$

3. We model the seasonal factor SN_t using indicator variables (aka "dummy" variables)
   - Need to model L seasons in SN_t using $L-1$ indicator variables:

     $$SN_t = \beta_{S_1}\mathbf{I}\{S_1\} + \beta_{S_2}\mathbf{I}\{S_2\} + \cdots + \beta_{S_{L-1}}\mathbf{I}\{S_{L-1}\}$$

     where

     $$\mathbf{I}\{S_i\} = \begin{cases} 1 & \text{If } t \text{ is season } i \\ 0 & \text{Otherwise} \end{cases}$$

4.4 Growth Curve Models

Up to this point, all of our models have been linear in terms of the parameters, although not necessarily in terms of the independent variables. In other words, each term in the

model is the product of a model parameter and a value of the independent variable determined by the observed data. However, what if our paramters are not linear functions of time? For instance, let us consider growth curve models:

$$x_t = \beta_0 \times \beta_1^t \times \varepsilon_t$$

This model is not linear in the parameters since time enters as an exponent and β_0 is multiplied by β_1^t. In addition the error term is multiplicative instead of additive. We can, however, use all of the previous techniques of estimation and prediction if we can linearlize the model

$$\begin{aligned} x_t &= \beta_0 \times \beta_1^t \times \varepsilon_t \\ ln(x_t) &= ln(\beta_0 \times \beta_1^t \times \varepsilon_t) \\ ln(x_t) &= ln(\beta_0) + tln(\beta_1) + ln(\varepsilon_t) \\ x_t^* &= \beta_0^* + \beta_1^* t + \varepsilon_t^* \end{aligned}$$

Point estimates and prediction intervals for the forecasted values are then found for the transformed x_t^* and back transformed using the inverse function used to linearlize the model. In this case, the exponential function.

Chapter 5

Decomposition Models

Two types

- Additive Decomposition
  - The time series exhibits constant seasonal variation
  - The parameters describing the series do not change over time
- Multiplicative Decomposition
  - The time series exhibits increasing or decreasing seasonal variation
  - The parameters describing the series do not change over time

5.1 Additive Decomposition

Model:

$$y_t = TR_t + SN_t + CL_t + IR_t$$

where TR_t, SN_t, CL_t, IR_t are defined to be the trend, seasonal, cyclical, and irregular factors

Our goal is to estimate the above factors with point estimates tr_t, sn_t, cl_t, ir_t and use these estimates for forecasting.

Method:

We start by finding the centered moving average (CMA) based on the expected seasonality, which is an estimate of the trend and cycle factors.

$$CMA_t = tr_t + cl_t$$

You would begin by using a 12-period moving average for monthly data, a 4-period moving average for quarterly data, etc. However this value represents the midpoint value of each month, so we then use a 2-period moving average to obtain values that correspond with our time periods. This is necessary because seasonal terms are usually even, if we

had an odd seasonal term for some reason, the second 2-period moving average would not be necessary.

The model implies that

$$SN_t + IR_t = y_t - (TR_t + CL_t)$$

and it follows that the estimate of the seasonal and irregular factors is

$$sn_t + ir_t = y_t - (tr_t + cl_t) = y_t - CMA_t$$

Estimating SN_T

Next, to obtain sn_t, we group the values of $sn_t + ir_t$ by seasons (months, quarters, years, ...)

For each season we compute the average $\overline{sn}_t$ of the $sn_t + ir_t$ realizing that the ir_t factor will drop out in the average

Then we normalize the seasonal average by subtracting the the mean over the L seasons.

$$sn_t = \overline{sn}_t - \sum_{t=1}^{L} \frac{\overline{sn}_t}{L}$$

Estimating TR_T

Once we have sn_t, we can deseasonalize the observations

$$d_t = y_t - sn_t$$

From d_t, we estimate TR_t by fitting a k^{th} order polynomial regression. Use standard regression techniques to determine the appropriate polynomial.

Estimating CL_T

Go back to the model

$$CL_t + IR_T = y_t - TR_t - SN_t$$

Our estimate of the LHS is then

$$cl_t + ir_t = y_t - tr_t - sn_t$$

In order to average out the ir_t, we perform a three-point moving average to get

$$cl_t = \frac{(cl_{t-1} + ir_{t-1}) + (cl_t + ir_t) + (cl_{t+1} + ir_{t+1})}{3}$$

Estimating IR_T

Go back to the model

$$IR_T = y_t - TR_t - SN_t - CL_t$$

Our estimate is the residual

$$ir_t = y_t - tr_t - sn_t - cl_t$$

Prediction:

Since our goal is prediction, we assume that the pattern continues in the future and that there is no pattern in the irregular component and so we predict IR_t to be zero

Thus the point estimate of the forecast at time t is

$$\hat{y}_t = tr_t + sn_t + cl_t$$

An appropriate prediction interval for y_t is

$$\hat{y}_t \pm B_{t,\alpha}$$

where $B_{t,\alpha}$ is the error bound in a $100(1-\alpha)\%$ prediction interval for d_t

5.2 Multiplicative Decomposition

Model:

$$y_t = TR_t * SN_t * CL_t * IR_t$$

where TR_t, SN_t, CL_t, IR_t are defined to be the trend, seasonal, cyclical, and irregular factors

Our goal is to estimate the above factors with point estimates tr_t, sn_t, cl_t, ir_t and use these estimates for forecasting
Method:

We start by using the centered moving average (CMA) based on the expected seasonality, which is an estimate of the trend and cycle factors

$$CMA_t = tr_t * cl_t$$

The model implies that

$$SN_t * IR_t = \frac{y_t}{(TR_t * CL_t)}$$

and it follows that the estimate of the seasonal and irregular factors is

$$sn_t * ir_t = \frac{y_t}{(tr_t * cl_t)} = \frac{y_t}{CMA_t}$$

Estimating SN_T

Next, to obtain sn_t, we group the values of $sn_t * ir_t$ by seasons (months, quarters, years, ...)

For each season we compute the average $\overline{sn}_t$ of the $sn_t * ir_t$ realizing that the ir_t factor will drop out in the average

Then we normalize the seasonal average by subtracting the the mean over the L seasons.

$$sn_t = \frac{\overline{sn}_t}{\sum_{t=1}^{L} \frac{\overline{sn}_t}{L}}$$

Estimating TR_T

Once we have sn_t, we can deseasonalize the observations

$$d_t = \frac{y_t}{sn_t}$$

From d_t, we estimate TR_t by fitting a k^{th} order polynomial regression. Use standard regression techniques to determine the appropriate polynomial.

Estimating CL_T

Go back to the model

$$CL_t * IR_T = \frac{y_t}{TR_t * SN_t}$$

Our estimate of the LHS is then

$$cl_t * ir_t = \frac{y_t}{tr_t * sn_t}$$

In order to average out the ir_t, we perform a three-point moving average to get

$$cl_t = \frac{(cl_{t-1} * ir_{t-1}) + (cl_t * ir_t) + (cl_{t+1} * ir_{t+1})}{3}$$

Estimating IR_T

Go back to the model

$$IR_T = \frac{y_t}{TR_t * SN_t * CL_t}$$

Our estimate is the residual

$$ir_t = \frac{y_t}{tr_t * sn_t * cl_t}$$

Prediction:

Since our goal is prediction, we assume that the pattern continues in the future and that there is no pattern in the irregular component and so we predict IR_t to be one

Thus the point estimate of the forecast at time t is

$$\hat{y}_t = tr_t * sn_t * cl_t$$

An appropriate prediction interval for y_t is

$$\hat{y}_t \pm B_{t,\alpha}$$

where $B_{t,\alpha}$ is the error bound in a $100(1-\alpha)\%$ prediction interval for d_t.

Chapter 6

Exponential Smoothing

1. Simple Exponential Smoothing: used when the time series has no significant linear trend (i.e. slope) but the mean is changing over time
2. Holt's Trend Corrected Exponential Smoothing: used when a series has a linear trend and a slope that is changing over time
3. Holt-Winter's Method: extension of Holt's trend method for a series that has seasonality
4. Damped Trend Exponential Smoothing: used if the growth rate cannot be sustained into the future and therefore should dampen the effects

6.1 Simple Exponential Smoothing

- If no trend exists and the mean remains constant, then our standard linear model

$$y_t = \beta_0 + \varepsilon \qquad where \qquad \varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$$

can be used giving us the simple estimate:

$$\hat{y} = b_0 = \frac{1}{n} \sum y_t$$

giving equal weight to all data points

- If instead, the mean might be changing slowly over time then it might be better to give more recent observations greater weight than older observations

Simple Exponential Smoothing is used to forecast a time series under these conditions. It gives the most recent observations the greatest weight and allows the forecaster to detect changes in the mean level and incorporate them into the model.

6.1.1 Method

1. Begin with an estimate of the mean at time point zero (before any observations). Average a subset (n_S :the first 12 observations of a large time series, or half of the observations for a small time series) of the observations to get this baseline estimate.

$$l_0 = \frac{1}{n_S} \sum_{t=1}^{n_S} y_t$$

2. Compute updated estimates using the following smoothing equation:

$$l_t = \alpha y_t + (1-\alpha) l_{t-1}$$

where $0 < \alpha < 1$ is chosen to be an arbitrary value to be improved during the optimization process.

3. Next we want to find the best α by minimizing the Sum of Squared Error (SSE)

$$SSE = \sum_{t=1}^{T} (y_t - l_{t-1})^2$$

Note: l_T gives the most recent observation the greatest weight with successively smaller weight to each of the earlier observations.

$$\begin{aligned}
l_T &= \alpha y_T + (1-\alpha) l_{T-1} \\
&= \alpha y_T + (1-\alpha)[\alpha y_{T-1} + (1-\alpha) l_{T-2}] \\
&= \alpha y_T + \alpha(1-\alpha) y_{T-1} + (1-\alpha)^2 l_{T-2} \\
&= \alpha y_T + \alpha(1-\alpha) y_{T-1} + (1-\alpha)^2 [\alpha y_{T-2} + (1-\alpha) l_{T-3}] \\
&= \alpha y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + (1-\alpha)^3 l_{T-4} \\
&\vdots \\
&= \alpha y_T + \alpha(1-\alpha) y_{T-1} + (1-\alpha)^2 y_{T-2} + \cdots + \alpha(1-\alpha)^{T-1} y_1 + (1-\alpha)^T l_0
\end{aligned}$$

and $\quad \alpha > \alpha(1-\alpha) > \alpha(1-\alpha)^2 > \cdots > \alpha(1-\alpha)^{T-1} > (1-\alpha)^T$

So the coefficients of $y_T, \ldots, y_1$ are decreasing exponentially with time, hence the term exponential smoothing.

6.1.2 Forecasting

- We can now forecast into the future, $y_{T+\tau}$. Our estimates will become less accurate as we move further from our last observation, y_t, and the prediction intervals will get wider.
- All point predictions will be based on l_T which is the forecast for y_{T+1}, the first time point after our final observation.
- The prediction interval for any time point, τ, in the future will be:

$$l_T \pm z_{\alpha/2} s \sqrt{1 + (\tau - 1)\alpha^2} \qquad where \qquad s = \sqrt{\frac{SSE}{T-1}}$$

6.1.3 Error Correction Form of Smoothing Equation

$$\begin{aligned}
l_t &= \alpha y_t + (1-\alpha) l_{t-1} \\
&= \alpha y_t + l_{t-1} - \alpha l_{t-1} \\
&= l_{t-1} + \alpha(y_t - l_{t-1})
\end{aligned}$$

- Remember that l_{t-1} is the forecast of y_t so ($y_t - l_{t-1}$ is the forecast error
- Therefore l_t is the estimate of y_t plus a fraction of the forecast error of y_t

6.2 Holt's Trend Corrected Exponential Smoothing

- If a time series is increasing or decreasing at a fixed rate, use linear regression model
$$y_t = \beta_0 + \beta_1 t + \varepsilon \qquad where \qquad \varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$$
where the increase between time,$t-1$, and time, t, is simply the growth rate, β_1
- What if both the mean level and the growth rate are changing with time? We need a model to describe these changing levels

6.2.1 Method

1. Let l_{t-1} estimate the mean at time period, $t-1$, and b_{t-1} estimate the growth rate. Now our estimate of y_t is $l_{t-1} + b_{t-1}$
2. Begin with an estimate of the mean and growht rate at time point zero (before any observations). Use standard linear regression on a small subset of the data such that:
$$l_0 = \hat{\beta}_0 \qquad and \qquad b_0 = \hat{\beta}_1$$
3. Compute updated estimates using the following smoothing equations:
$$\begin{aligned} l_t &= \alpha y_t + (1-\alpha)[l_{t-1} + b_{t-1}] \\ b_t &= \gamma(l_t - l_{t-1}) + (1-\gamma) b_{t-1} \end{aligned}$$
where $0 < \alpha, \gamma < 1$ are chosen to be arbitrary values to be improved during the optimization process.
4. Next we want to find the best α, γ by minimizing the Sum of Squared Error (SSE)
$$SSE = \sum_{t=1}^{T} (y_t - (l_{t-1} + b_{t-1}))^2$$

6.2.2 Forecasting

- We can now forecast into the future, $y_{T+\tau}$. Our estimates will become less accurate as we move further from our last observation, y_t, and the prediction intervals will get wider.
- All point predictions will be based on l_T, b_T and time τ:
$$\hat{y}_{T+\tau} = l_T + \tau b_{T+1}$$
- The prediction interval for any time point, τ, in the future will be:
$$l_T \pm z_{\alpha/2} s \sqrt{1 + \sum_{j=1}^{\tau-1} \alpha^2 (1 + j\gamma)^2} \qquad where \qquad s = \sqrt{\frac{SSE}{T-2}}$$

6.2.3 Error Correction Form of Smoothing Equations

$$\begin{aligned} l_t &= \alpha y_t + (1-\alpha)[l_{t-1} + b_{t-1}] \\ &= \alpha y_t + (l_{t-1} + b_{t-1}) - \alpha[l_{t-1} + b_{t-1}] \\ &= (l_{t-1} + b_{t-1}) + \alpha(y_t - (l_{t-1} + b_{t-1})) \\ b_t &= \gamma(l_t - l_{t-1}) + (1-\gamma)b_{t-1} \\ &= \gamma(l_t - l_{t-1}) + b_{t-1} - \gamma b_{t-1} \\ &= b_{t-1} + \gamma((l_t - l_{t-1}) - b_{t-1}) \\ &= b_{t-1} + \gamma((l_{t-1} + b_{t-1} + \alpha(y_t - (l_{t-1} + b_{t-1})) - l_{t-1} - b_{t-1}) \\ &= b_{t-1} + \alpha\gamma(y_t - (l_{t-1} + b_{t-1})) \end{aligned}$$

6.3 Holt-Winter's Exponential Smoothing

6.3.1 Additive Model

- If we had a fixed trend and fixed seasonal factor, our previous model was:

$$y_t = TR_t + SN_t + \varepsilon \qquad where \qquad \varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$$

- If linear trend is changing with time and has constant seasonal variation we use the additive Holt-Winter's method

Method

1. Let l_{t-1} estimate the mean at time period, $t-1$, and b_{t-1} estimate the growth rate. Let sn_{t-L} denote the most recent estimate of the seasonal factor for the season in time t where L is then number of seasons.

2. To estimate the seasonal component, we need 4 or 5 years of data which will be our subset of observations for baseline.

3. Begin with an estimate of the mean and growth rate at time point zero (before any observations). Use standard linear regression on the subset of the data such that:

$$l_0 = \hat{\beta}_0 \qquad and \qquad b_0 = \hat{\beta}_1$$

4. Our seasonal baseline factors are found as follows:

   (a) Use the regression equation to compute a fitted value ($\hat{y}_t$) for the subset of data

   (b) Detrend the data by computing: $y_t - \hat{y}_t$

   (c) Compute the baseline seasonal factors by averaging the detrended values over the years for each season

5. Compute updated estimates using the following smoothing equations:

$$\begin{aligned} l_t &= \alpha(y_t - sn_{t-L}) + (1-\alpha)[l_{t-1} + b_{t-1}] \\ b_t &= \gamma(l_t - l_{t-1}) + (1-\gamma)b_{t-1} \\ sn_t &= \delta(y_t - l_t) + (1-\delta)sn_{t-L} \end{aligned}$$

where $0 < \alpha, \gamma, \delta < 1$ are chosen to be arbitrary values to be improved during the optimization process.

6. Next we want to find the best α, γ, δ by minimizing the Sum of Squared Error (SSE)

$$SSE = \sum_{t=1}^{T} (y_t - (l_{t-1} + b_{t-1} + sn_{t-L}))^2$$

Forecasting

- We can now forecast into the future, $y_{T+\tau}$. Our estimates will become less accurate as we move further from our last observation, y_t, and the prediction intervals will get wider.
- All point predictions will be based on l_T, b_T, sn_{T-L} and time τ:

$$\hat{y}_{T+\tau} = l_T + \tau b_T + sn_{T+\tau-L}$$

- The prediction interval for any time point, τ, in the future will be:

$$\hat{y}_{T+\tau} \pm z_{\alpha/2} s \sqrt{c_\tau} \qquad \text{where} \qquad s = \sqrt{\frac{SSE}{T-3}}$$

and

$$\begin{aligned} c_\tau &= 1 \qquad \text{if } \tau = 1 \\ &= 1 + \sum_{j=1}^{\tau-1} \alpha^2 (1 + j\gamma)^2 \qquad \text{if } 2 \le \tau \le L \\ &= 1 + \sum_{j=1}^{\tau-1} [\alpha(1 + j\gamma) + d_{j,L}(1 - \alpha)\delta]^2 \qquad \text{if } \tau > L \\ d_{j,L} &= 1 \text{ if } j \text{ is a multiple of L} \\ &= 0 \text{ otherwise} \end{aligned}$$

6.3.2 Multiplicative Model

- If we had a fixed trend and fixed seasonal factor, our previous model was:

$$y_t = TR_t \times SN_t \times \varepsilon \qquad where \qquad \varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$$

- If linear trend is changing with time and has changing seasonal variation we use the multiplicative Holt-Winter's method

Method

1. Let l_{t-1} estimate the mean at time period, $t-1$, and b_{t-1} estimate the growth rate. Let sn_{t-L} denote the most recent estimate of the seasonal factor for the season in time t where L is then number of seasons.
2. To estimate the seasonal component, we need 4 or 5 years of data which will be our subset of observations for baseline.

3. Begin with an estimate of the mean and growth rate at time point zero (before any observations). Use standard linear regression on the subset of the data such that:
$$l_0 = \hat{\beta}_0 \qquad and \qquad b_0 = \hat{\beta}_1$$
4. Our seasonal baseline factors are found as follows:
   (a) Use the regression equation to compute a fitted value ($\hat{y}_t$) for the subset of data
   (b) Detrend the data by computing: $y_t/\hat{y}_t$
   (c) Compute the baseline seasonal factors by averaging the detrended values over the years for each season giving the following values: $\bar{S}_1, \ldots, \bar{S}_L$
   (d) We then need to adjust these estimates so that the average of the seasonal factors is 1:
$$sn_{i-L} = \bar{S}_i \left[\frac{L}{\sum_{i=1}^{L} \bar{S}_i} \right] \qquad \text{where} \qquad i = 1, \ldots, L$$
5. Compute updated estimates using the following smoothing equations:
$$\begin{aligned} l_t &= \alpha(y_t/sn_{t-L}) + (1-\alpha)[l_{t-1} + b_{t-1}] \\ b_t &= \gamma(l_t - l_{t-1}) + (1-\gamma)b_{t-1} \\ sn_t &= \delta(y_t/l_t) + (1-\delta)sn_{t-L} \end{aligned}$$
   where $0 < \alpha, \gamma, \delta < 1$ are chosen to be arbitrary values to be improved during the optimization process.
6. Next we want to find the best α, γ, δ by minimizing the Sum of Squared Error (SSE)
$$SSE = \sum_{t=1}^{T} \left[\frac{y_t - (l_{t-1} + b_{t-1})sn_{t-L}}{(l_{t-1} + b_{t-1})sn_{t-L}} \right]^2$$

Forecasting

- We can now forecast into the future, $y_{T+\tau}$. Our estimates will become less accurate as we move further from our last observation, y_t, and the prediction intervals will get wider.
- All point predictions will be based on l_T, b_T, sn_{T-L} and time τ:
$$\hat{y}_{T+\tau} = (l_T + \tau b_T)sn_{T+\tau-L}$$
- The prediction interval for any time point, τ, in the future will be:
$$\hat{y}_{T+\tau} \pm z_{\alpha/2} s(\sqrt{c_\tau}) sn_{T+\tau-L} \qquad \text{where} \qquad s = \sqrt{\frac{SSE}{T-3}}$$
  and
$$\begin{aligned} c_\tau &= (l_t + b_t)^2 \text{ if } \tau = 1 \\ &= \sum_{j=1}^{\tau-1} \alpha^2(1 + (\tau - j)\gamma)^2(l_T - jb_t)^2 + (l_T + \tau b_T)^2 \qquad \text{if } 2 \le \tau \le L \\ &= \text{even more complicated if} \qquad \tau > L \end{aligned}$$

6.4 Damped Trend Exponential Smoothing

- If it seems that the growth rate cannot continue in a relatively constant manner into the future, the effects should be dampened for the forecasts
- Methods are exactly the same as in the previous models except we add an additional dampening factor the the smoothing equations which must also be optimized while minimizing the SSE

6.5 State Space Models used to find Prediction Intervals

1. These models assume a singel source of error
2. The components of the model are called states

State	Model Parameter	Estimate
Level	L_t	l_t
Growth	B_t	b_t
Season	SN_t	sn_t

3. Each state space model has an observation equation and one or more state equations using the error correction form of the smoothing equations
4. The remaining error after accounting for the states is $\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$

Example: Simple Exponential Smoothing

$$\begin{aligned} \text{Observation Equation} \quad & y_t = L_{t-1} + \varepsilon_t \\ \text{State Equation} \quad & L_t = L_{t-1} + \alpha\varepsilon_t \end{aligned}$$

A general prediction interval for any forecast $y_{T+\tau}$ will have the following form:

$$\hat{y}_{T+\tau} \pm z_{\alpha/2}\sqrt{Var(y_{T+\tau} - E[y_{T+\tau}])}$$

$$\begin{aligned} E[y_{T+\tau}] &= E(L_{T+\tau-1} + \varepsilon_{T+\tau}) \\ &= E(L_{T+\tau-2} + \alpha\varepsilon_{T+\tau-1} + \varepsilon_{T+\tau}) \\ &= E(L_{T+\tau-3} + \alpha\varepsilon_{T+\tau-2} + \alpha\varepsilon_{T+\tau-1} + \varepsilon_{T+\tau}) \\ &\vdots \\ &= E(L_T + \alpha\varepsilon_{T+1} + \cdots + \alpha\varepsilon_{T+\tau-1} + \varepsilon_{T+\tau}) \\ &= E(L_T + \alpha\sum_{j=1}^{\tau-1}\varepsilon_{T+j} + \varepsilon_{T+\tau}) \\ &= E(L_T) + E\left[\alpha\sum_{j=1}^{\tau-1}\varepsilon_{T+j}\right] + E[\varepsilon_{T+\tau}] \\ &= E(L_T) + \alpha\sum_{j=1}^{\tau-1}E(\varepsilon_{T+j}) + E[\varepsilon_{T+\tau}] \\ &= L_T \end{aligned}$$

We expect $y_{T+\tau}$ to be L_T on average, therefore the estimate $\hat{y}_{T+\tau}$ is l_T.

$$\begin{aligned}
Var(y_{T+\tau} - E[y_{T+\tau}]) &= Var(L_{T+\tau-1} + \varepsilon_{T+\tau} - L_T) \\
&= Var(L_T + \alpha \sum_{j=1}^{\tau-1} \varepsilon_{T+j} + \varepsilon_{T+\tau} - L_T) \\
&= Var\left[\alpha \sum_{j=1}^{\tau-1} \varepsilon_{T+j} + \varepsilon_{T+\tau}\right] \\
&= Var\left[\alpha \sum_{j=1}^{\tau-1} \varepsilon_{T+j}\right] + Var(\varepsilon_{T+\tau}) \\
&= \alpha^2 \sum_{j=1}^{\tau-1} Var(\varepsilon_{T+j}) + Var(\varepsilon_{T+\tau}) \\
&= \alpha^2(\tau-1)\sigma^2 + \sigma^2 \\
&= \sigma^2(1 + (\tau-1)\alpha^2)
\end{aligned}$$

Since σ^2 is unknown, we estimate it with s^2 and our prediction interval becomes:

$$l_T \pm z_{\alpha/2} s \sqrt{1 + (\tau-1)\alpha^2}$$

Chapter 7

Box-Jenkins Methodology

7.1 Time Series Relationships

Assume we have two series x_t and y_t that are observed over some set of time points, $t = 1, 2, \ldots, n$

There are three types of correlation functions we will need to consider:

- Autocorrelation Function-measures the relationship between a series with itself at different time points
- Partial Autocorrelation Function-measures the self predictability of a series
- Cross-correlation Function-measures the correlation between several series at different time points

7.1.1 Autocorrelation Function (ACF)

Correlation at adjacent points of the same series is measured by the autocorrelation function (ACF) defined previously. Since the true ACF is unknown, we utilize a sample version of the ACF, say

$$\hat{\rho}_x(h) = \frac{\hat{\gamma}_x(h)}{\hat{\gamma}_x(0)}$$

where

$$\hat{\gamma}_x(h) = \frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x}) \quad with \quad \bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$$

One technique used for the purpose of visualizing the relations between a series at different lags are scatterplot matrices

R: `lag.plot(data, lags=12, diag=F)`

Under the assumption that the underlying process x_t is gaussian white noise (after we have taken the appropriate steps), the approximate standard error of the ACF is

$$\sigma_{\hat{\rho}} = \frac{1}{\sqrt{n}}$$

This implies that

$$\hat{\rho}_x(h) \sim N\left(0, \frac{1}{n}\right)$$

From normal theory, values within $\pm 1.96\sigma_{\hat{\rho}}$ might be reasonable if one is willing to live with each test at $\alpha = 0.05$.

7.1.2 Partial Autocorrelation Function (PACF)

On can think of the PACF as the simple correlation between two points separated by a lag h, say x_t and x_{t-h}, with the effect of the intervening points $x_{t-1}, x_{t-2}, \ldots, x_{t-h+1}$ conditioned out.

This is analogous to finding the partial correlation in regression between the response variable and one of the independent variables conditioning out the other independent variables.
Suppose we want to predict x_t from $x_{t-1}, \ldots, x_{t-h}$ using some linear function of these past values. Consider minimizing the mean square prediction error

$$MSE = E[(x_t - \hat{x}_t)^2]$$

using the predictor

$$\hat{x}_t = a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_h x_{t-h}$$

over the possible values of the weighting coeffcients $a_1, \ldots, a_h$ where we assume, for convenience that x_t has been adjusted to have zero mean.
The partial autocorrelation function (PACF) is defined as the value of the last coeffcient

$$\phi_{hh} = a_h$$

In practice, we minimize the sample error sum of squares

$$SSE = \sum_{t=h+1}^{n} \left[(x_t - \bar{x}) - \sum_{k=1}^{h} a_k (x_{t-k} - \bar{x}) \right]^2$$

with the estimated partial correlation defined as $\hat{\phi}_{hh} = \hat{a}_h$ and

$$\hat{\phi}_{hh} \sim N\left(0, \frac{1}{n}\right)$$

7.1.3 Cross Correlation Function (CCF)

The fact that correlations may occur at some time delay when trying to relate two series to one another at some lag h for purposes of prediction suggests that it would also be useful to plot x_{t+h} against y_t. A measure of the correlation between several series is the cross covariance function and the standardized cross correlation function

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

and

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

Since we do not know the true cross correlation function, we use the sample cross correlation function

$$\hat{\rho}_{xy}(h) = \frac{\hat{\gamma}_{xy}(h)}{\sqrt{\hat{\gamma}_x(0)\hat{\gamma}_y(0)}}$$

where

$$\hat{\gamma}_x y(h) = \frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(y_t - \bar{y})$$

Under the hypothesis that there is no relation at time h and that at least one of the two series is independent and identically distributed, then

$$\hat{\rho}_{xy}(h) \sim N\left(0, \frac{1}{n}\right)$$

7.2 Nonseasonal Methods

7.2.1 Introduction

The Box-Jenkins method consists of a four-step iterative procedure:

1. Use historical data to identify an appropriate model.
2. Estimate parameters of the model using the data.
3. Check the fit of the model using diagnostics and if needed fit an improved model.
4. Use final model to forecast future values.

Box-Jenkins models describe **stationary** time series, so we must determine if the series is stationary and if not, transform it to attain stationarity. We use the ACF to determine if the time series is stationary

- If the ACF dies down quickly (relatively few significant lags) we can consider the series stationary
- If it dies down slowly then we need to try some transformations to either remove the trend or stabilize the variance. We will then use the new transformed series as the "working" time series for the remainder of the analysis.

Once we have achieved stationarity, we can then use the ACF and PACF to identify a potential model:

- If there is a spike ("highly significant") at lag k where all lags afterward are not significant we can focus on the relationship between y_T and y_{T-k}.
- We say the ACF or PACF "cuts off after lag k" if all the lags beyond are not significant.
- Need to realize that both the ACF and PACF might not "cut off" at the same lag. For instance the PACF might not cut off at all while the ACF has.

7.2.2 Autoregressive (AR) Models

The idea behind AR models is that the present value of the series can be explained as a function of p past values

$$AR(p) : y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + w_t$$

We can also use the backshift operator to write the model more succinctly:

$$\Phi(B) y_t = w_t$$

where $\Phi(B) = 1 - \phi_1(B) - \cdots - \phi_p(B^p)$ and recall that $B^k y_t = y_{t-k}$

The general pattern of an AR(p) model will have an ACF that dies down in a steady fashion dominated by damped exponential decay and the PACF will have a significant peak at lag p and then cut off.

Show Examples of AR(1) with +/- ϕ

Lets look at an AR(1):

$$y_t = \phi_1 y_{t-1} + w_t$$

It can be shown that for an AR(1) model the autocorrelation function is:

$$\rho(h) = (\phi_1)^h \qquad for \qquad h \geq 1$$

7.2.3 Moving Average (MA) Models

As an alternative to the AR representation in which y_t is a linear combination of previous observations, the MA model assumes that white noise is combined linearly to form the observed data

$$MA(q) : y_t = w_t \theta_1 w_{t-1} + \theta_2 2_{t-2} + \cdots + \theta_q w_{t-1}$$

We can also use the backshift operator to write the model more succinctly:

$$y_t = \Theta(B) w_t$$

where $\Theta(B) = 1 + \theta_1(B) + \cdots + \theta_q(B^q)$ and recall that $B^k y_t = y_{t-k}$

The general pattern of an MA(q) model will have an PACF that dies down in a steady fashion dominated by damped exponential decay and the ACF will have a significant peak at lag q and then cut off.

Show Examples of MA(1) with +/- θ

Lets look at an MA(1):

$$y_t = w_t + \theta_1 w_{t-1}$$

It can be shown that for an AR(1) model the autocorrelation function is:

$$\begin{aligned} \rho(h) &= \frac{-\theta_1}{1 + \theta_1^2} && \text{for h = 1} \\ &= 0 && \text{for h ¿ 1} \end{aligned}$$

7.3 Seasonal Methods