Module 8: Time Series Analysis

8.1 Serial Correlation

Serial Correlation

- Data points can be either independent from one another or correlated with one another.
- Correlation, or dependence, can occur in either time or space.
- Here we discuss correlation in time sometimes called autocorrelation.

Serial Correlation

- Data points that are independent in time give no information about the next value that will occur. They may give information about the probability distribution that they are samples from but they give no direct information about the next value.
- Alternatively, points correlated in time give information that can be used to predict the next point.

Serial Correlation

- Examples of independent time series:
  - consecutive flips of a fair coin,
  - consecutive rolls of a die,
  - …
- Examples of correlated time series:
  - daily high temperatures
  - Dow Jones Average daily closing values
  - ...

Serial Correlation

- So, if I measured the heights of students in the class in alphabetical order and plotted them in the time order of measuring, the points give information about the average height (mean) and variability (standard deviation).
- However, the fifth height measures doesn't help me at all to predict the sixth.
- They are independent.

Module 8.1

Serial Correlation

- However, if I don't know what day of the year it is but I do know today's high temperature then I can use that to predict tomorrow's high temperature (best guess is today's high)
- If I know it was 10 ^{o}F yesterday that is a good indication of cold weather today. Very different than if it was 90 ^{o}F yesterday.
- Daily high temperatures are positively autocorrelated

Module 8.1

Serial Correlation

- Recall that positive correlation means that high points are associated with other high points and lows with lows
- So, data points positively autocorrelated in time are more like their near neighbors in the series and less like points further away in time.
- Data points negatively autocorrelated are less like their near neighbors, they tend to jump from high to low to high

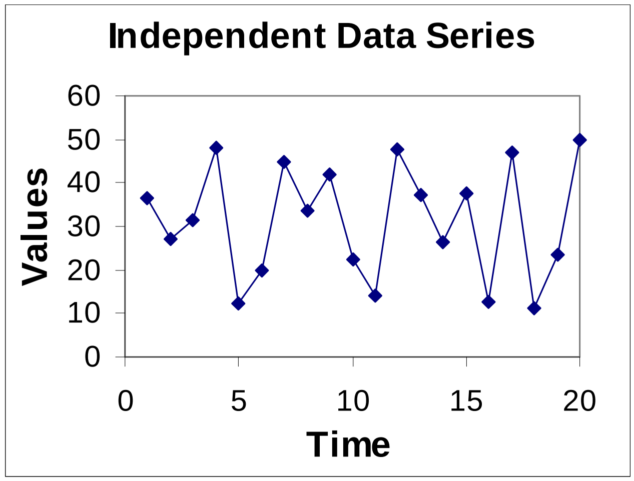
Module 8.1

Serial Correlation

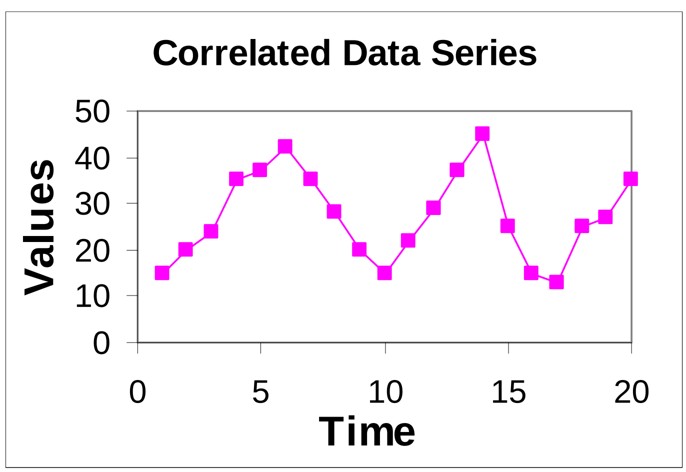
- Independent data points are all random samples from their underlying pdf and give no additional information about their near neighbors in time

Module 8.1

Serial Correlation



Serial Correlation



Serial Correlation

- Autocorrelations can be estimated.
- The correlations are different for
  - r_1 = correlation at lag 1 = points one time step apart
  - r_2 = correlation at lag 2 = points two time steps apart
  - and so on
- Most of the time $r_1 > r_2 > r_3$

Serial Correlation

- However, if there is a seasonal component then that will show up as a large r.
- If data is monthly and there is a seasonal component over the year, then r_{12} will reflect this.
- It shows that Januarys tend to be alike as do Mays, Julys, and so forth.
- Monthly highs or lows would be examples.
- Also monthly retail sales – December for example

Serial Correlation

- So, the autocorrelation at lag k is the correlation between x_i and x_{i+k}

$$r_k = \frac{\sum_{i=1}^{n-k} \frac{(x_i - \bar{x})(x_{i+k} - \bar{x})}{n-k}}{\sum_{i=1}^{n} \frac{\left(x_i - \bar{x}\right)^2}{n}}$$

Serial Correlation

- Once the autocorrelations are calculated, they can be plotted as an autocorrelation function or correlogram. It is a plot of r_k against k.
- Even if significant autocorrelations do not exist, values will exist for each r_i.
- In this case, these will be random values from a normal distribution. Knowing this helps to determine if the r_i's are significantly different from zero.

Serial Correlation

- Most statistical software packages will indicate the value above which the r_i's may be significant.
- If a correlation does appear to be significantly different from zero, then seek to understand why it may exist.
- Keep in mind that marginally significant correlations with no explanation may be reflections of random variability since the test reflects many simultaneous significance tests.

Module 8.1

15

Serial Correlation

- Tests for randomness:
  - Another way to test whether significant autocorrelation exists is to use a test for randomness.
  - There are several simple nonparametric tests:
    - Runs above and below the median
    - Sign test
    - Runs up and down

Module 8.1

16

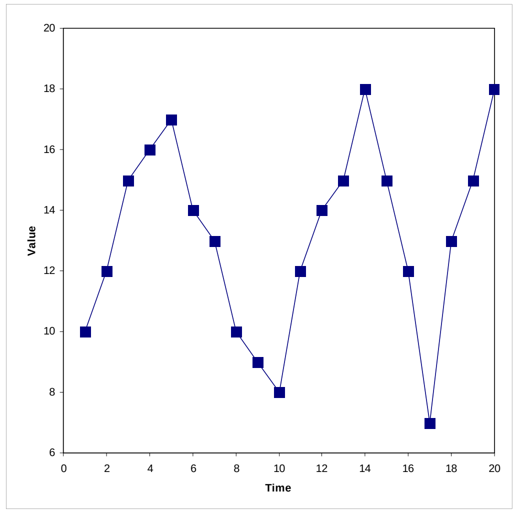
Serial Correlation

- Continuity correction:
  - This correction is needed when using a continuous pdf like the normal distribution to get a critical value for a test based on discrete counts.
  - The correction adjusts for the difference between discrete data and a continuous pdf.
  - To carry out the continuity correction, add or subtract 0.5 from the test statistic as appropriate

Serial Correlation - Example

Time	Data	Runs Above and Below Median	Sign Test	Runs up and Down
1	10	0		
2	12	0	+	+
3	15	1	+	+
4	16	1	+	+
5	17	1	+	+
6	14	1	-	-
7	13	0	-	-
8	10	0	-	-
9	9	0	-	-
10	8	0	-	-
11	12	0	+	+
12	14	1	+	+
13	15	1	+	+
14	18	1	+	+
15	15	1	-	-
16	12	0	-	-
17	7	0	-	-
18	13	0	+	+
19	15	1	+	+
20	18	1	+	+
Median =	**13.5**	**M = 6**	**P = 11**	**R = 5**

Serial Correlation - Example



Serial Correlation - Example

- Tests of significance:
  - Runs Above and Below the Median
    - M = number of runs
    - r = number of zeros
    - $\mu_M = 2r(n-r)/n + 1$
    - $\sigma^2_M=2r(n-r)[2r(n-r)-n]/[n^2(n-1)]$
    - $Z = (M - \mu_M)/ \sigma_M$
  - Z is compared to the table of the standard normal distribution (possibly need continuity correction)

Serial Correlation - Example

- Tests of significance:
  - Runs Above and Below the Median
    - M = 6
    - r = 10
    - μ_M = 2*10*(20-10)/20 + 1 = 11
    - $\sigma^2{}_M$=2*10*(20-10){2*10*(20-10)-20}/{20^2(20-1)} = (200*180)/(400*19) = 36000/7600 = 4.74
    - Z = (M - μ_M)/ σ_M = (6.5 - 11)/2.18 = -2.06
    - P(Z<-2.06 and Z>2.06) = 2*0.02 = 0.04
    - Conclusion: There are significantly fewer runs than expected so data may not be independent

Serial Correlation - Example

- Tests of significance:
  - Runs Above and Below the Median
    - Note: must take square root of $\sigma^2{}_M$ to get σ_M
    - Note: continuity correction was applied as discussed in Manly, bottom of page 187.

Serial Correlation - Example

- Tests of significance:
  - Sign Test
    - P = the number of positive signs
    - m = number of differences (ignoring zeros)
    - $\mu_P = m/2$
    - $\sigma^2_P = m/12$
    - $Z = (P - \mu_P)/\sigma_P$

Serial Correlation - Example

- Tests of significance:
  - Sign Test
    - P = 11
    - m = 19
    - $\mu_P = 19/2 = 9.5$
    - $\sigma^2_P = 19/12 = 1.58$
    - $Z = (P - \mu_P)/\sigma_P = (10.5 - 9.5)/1.26 = 0.79$
    - $P(Z<-0.79 \text{ and } Z>0.79) = 2 * (0.5- 0.285) =$
      $= 2 * 0.215 = 0.43$
    - Conclusion: There are not significantly more positive differences than expected so data may be independent

Serial Correlation - Example

- Tests of significance:
  - Runs Up and Down
    - R = number of runs
    - m = number of differences (ignoring zeros)
    - $\mu_R = (2m + 1)/3$
    - $\sigma^2_R = (16m – 13)/90$
    - $Z = (R - \mu_R)/ \sigma_R$

Module 8.1

Serial Correlation - Example

- Tests of significance:
  - Runs Up and Down
    - R = 5
    - m = 19
    - $\mu_R = (2*19+1)/3 = 13$
    - $\sigma^2_R = (16*19-13)/90 = 3.23$
    - $Z = (R - \mu_R)/ \sigma_R = (5.5 - 13)/1.80 = -4.17$
    - P(Z<-4.17 and Z>4.17) = <0.002
    - Conclusion: There are fewer runs up and down than expected so data may not be independent

Module 8.1

Serial Correlation - Example

- Two of the three tests showed strong evidence of autocorrelation.
- This should lead to serious questioning of the assumption of independence
- Statistical techniques that rely on this assumption (regression analysis for example) should not be used
- Must use time series methods