Math 519 Stochastic Mathematical Modeling

Time Series Homework 1

Homework #4

William Gillespie

**Sheet 1**

I graphed the time vs value columns as a scatter plot. I created two trend lines: an exponential one and a cubic polynomial one. Here are examples of the tables:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| time | value | exp pred | exp resids | poly^3 pred | poly^3 resids |
| 0 | 3.491812519 | 2.6167 | 0.875112519 | 0.45742186 | 3.034390659 |
| 1 | 3.970740063 | 2.728396249 | 1.242343814 | 1.08869849 | 2.882041573 |
| 2 | 3.204296022 | 2.844860355 | 0.359435667 | 1.6830423 | 1.521253722 |

Below is a scatter plot of the data with the trend lines and their equations.

Below is a residual plot of the exponential function with a moving average dotted line (20 units).

For the above residual plot, there does not seem to be a trend in the residuals until Time>80, after which there is a downward trend. The exponential function tends to slightly under predict for Time<80, and over-predict for Time>80.

Below is a residual plot of the polynomial fit trend line (with a moving average fit line).

When one considers the moving average, there is a wavy down-up-down line, indicating that the polynomial has some modest lack of fit. The down-up-down pattern reflects being constrained to be a cubic polynomial, which has a natural down-up-down or up-down-up pattern.

**Sheet 2**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| time | value | exp pred | residuals | poly^2 pred | poly^2 resid |
| 0 | 4.198800676 | 3.056728 | 1.142072676 | 13.671374 | -9.472573324 |
| 1 | 4.435557739 | 3.180266705 | 1.255291034 | 12.695486 | -8.259928261 |
| 2 | 2.453141733 | 3.308798269 | -0.855656536 | 11.7658 | -9.312658267 |

There is a huge pattern here. The values are badly over-estimated between approximately Time 30 to 50.

The points are badly under-estimated between approximately Time 50 and 80.

Although the fit is terrible at the ends, it is not too bad in the middle, from approximately Time 10 to 90.

**Sheet 3**

The quadratic function provides an excellent fit overall. There are some rising and falling patterns between 1790 and 1910. There are two outliers at 1940 and 1950, probably reflecting the drop in population due to deaths during WW2. There is also an outlier in 2000, possibly due to people wanting to have children born in the year 2000.

**Sheet 4**

Questions:

i) What do you notice about the slopes and intercepts of the two lines? Hit F9 to refresh the data a few times.

**The slopes and intercepts stay the same for both plots.**

ii) What do you notice about the R^2 values for the two lines? Hit F9 to refresh the data a few times.

**The R^2 value is always greater in the figure where the points at a given x value are averaged.**

iii) What conclusion can you make about what happens when you average before fitting a trend line?

**Taking the average does not change the slope or intercept, but reduces the variability around the line, thus increasing the R^2.**

iv) Should you average before fitting a trend line? Explain.

**No, one should not average before fitting a trend line. Although the slope and intercept are the same, the R^2 value is artificially inflated. That means the trend line underestimates the variance in the data. There may be some situations where taking the average is appropriate.**

**Sheet 5**

i) Create an artificial data series Y\_t = 3\*t + Normal(0,1)

|  |  |  |
| --- | --- | --- |
| t | Y­\_t | first difference |
| 0 | 0.575716409 | 1.470688269 |
| 1 | 2.046404678 | 2.975618267 |
| 2 | 5.022022945 | 4.714578181 |
| 3 | 9.736601127 | 2.039977654 |

ii) Graph it.

iii) Take the first differences, D\_t = Y\_t - Y\_(t-1)

**Done (shown in table above)**

iv) Plot those differences. Do they look independent from one to the next? We know the original deviations from the line were independent.

**Yes, the differences look quite independent from one to the next. I see no trend in the differences. That is because Y\_t depends only on t and not y(t-1).**

**Sheet 6**

i) Create an artificial data series Y\_t = Y\_(t-1) + 3 + Normal(0,1)

|  |  |  |  |
| --- | --- | --- | --- |
| t | Y\_t | predicted | residuals |
| 0 | 1 | 1.208940465 | -0.208940465 |
| 1 | 4.085336916 | 4.159770269 | -0.074433352 |
| 2 | 8.031954435 | 7.110600072 | 0.921354363 |

ii) Graph it and fit a straight line.

iii) Use the functions =SLOPE and =INTERCEPT to get the slope and intercept of that line in cells,

|  |  |
| --- | --- |
| slope | intercept |
| 3.352938872 | **0.389503297** |

**I used =SLOPE(B15:B35,A15:A35) and =INTERCEPT(B15:B35, A15:A35)**

iv) Create a column of Predicted values, and then a column of Residuals.

**Done. (shown in the table in i).**

v) Plot the Residuals and comment on them as in previous exercises.

**The residuals are not completely random. Although there are occasional jumps in the y value of a given residual, the y values tend to be close to their neighbor. I noticed this after having the sheet reevaluated with new random numbers several times. (The graph on the word document is even updated when I update the excel document). This is because the y value depends on a y-1 value + some random noise. Therefore, the change can only be so far away from the previous y value. The next y value generated will be generated with a normal distribution, plus 3, plus the previous y value.**

vi) Problem 5 and Problem 6 relate to what distinction in de-trending time series?

**The distinction is whether the consecutive points are independent or not. In problem 5 they were independent. In problem 6 they were not. Dependent points lead to autocorrelation in the residuals.**