Assignment 2

The data in the spreadsheet (downloadable by clicking the down-arrow button, 04_cricket_1999to2011.csv) contains data on ODI matches from 1999 to 2011.

1. Using the first innings data alone in the above spreadsheet, find the best fit 'run production functions' in terms of wickets-in-hand w and overs-to-go u. Assume the model $Z(u,w) = Z_0(w)[1 - \exp\{-b(w)u\}]$. Use the sum of squared errors loss function, summed across overs and wickets. You will find one $Z_0(w)$ and $b(w)$ for each w. In your report, you should provide a plot of the ten functions, report the (20) parameters associated with the (10) production functions, and the error per point. Your function should be named:

Z_0, b = DuckworthLewis20Params()

2. Now assume the model $Z(u,w) = Z_0(w)[1 - \exp\{-Lu/Z_0(w)\}]$ and use the sum of squared errors loss function, summed over overs and wickets. Note that this new regression forces all slopes to be equal at $u = 0$. In your report, you should provide a plot of the ten functions, report the (11) parameters associated with the (10) production functions, and the error per point. This function should be named:

Z_0, L = DuckworthLewis11Params()

Feel free to use tools for nonlinear regression available in Python. Some date fields are in different format with an extra comma. Write a short script to clean this up. There's no need to send us the cleaning script or the cleaned csv file.

3. In the summary file, in addition to providing the plots, compare the slopes at u=0 in Problem 1, describe how close they are to each other, and to the slope computed in problem 2. Provide the total error per point.