Coding for Economic Applications

1 Python

You are asked to exploit Python to scrape a database containing National Hockey League statistics for the period 1990-2011.

Data are available from different sources on the web. We suggest the following

https://www.scrapethissite.com/pages/forms/

Your task: write a complete webscraper script (i.e., do not simply download bulk data) allowing to download, for the 21 seasons from 1990 to 2011 the following variables: `team_name`, `year`, `wins` (number of won matches), `losses` (number of lost matches), `win_perc` (percentage of won matches), `goal_for` and `goal_against`. Each line of the script must be described in the pdf document with your comments (point 3 above).

2 Stata/R

Once obtained the data, exploit a statistical software of your choice (R or Stata) to address (report and comment in detail) the following points:

a) Import the data for the analysis and provide some descriptive statistics in formatted tabular form. Discuss the evidence suggested by your tables.

b) Provide a graphical analysis of the average `wins` and `losses` over time. What happened from 1999? Discuss.

c) Run a log-linear regression of the (log) number of won matches (`log(wins)`) on a squared polynomial for `goal_for` and `goal_against`. Interpret the results (magnitude and statistical significance) and comment on the adjusted R^2. Add to the previous model a set of $T - 1$ year dummies. Do the marginal effects of `goal_for` change?