DRAFT PICKS, TEAM WEALTH, TEAM QUALITY

(The attached PDF file has better formatting.)

{This dialogue mentions other team characteristics that affect the regression coefficients.}

Jacob: Suppose the correlations are positive for lags 1-10 and zero for lags 11 and greater. (A *zero* correlation means *not significantly different from zero*.) What do we do next?

Rachel: To optimize the regression equation, form a data base where each row has 11 values: the won-loss record for the forecast year followed by the won-loss records for the ten preceding years. Show columns for the team identifier, the League, and the forecast year, so that we can sort the data base by these columns.

Illustration: In the *F* statistic part of this sample project, we compare National vs American League teams, so we sort by League. If you compare sets of years, or if you compare rich teams vs poor teams or good teams vs bad teams, you might sort them differently.

Jacob: Does the NEAS web site have these data bases?

Rachel: We provide data bases for several sports. Statisticians spend much of their time compiling data. For the student projects, we provide data, so that you spend your time on the statistical techniques.

Publicly available statistics for professional sports is enormous. We have taken data from public web sites and formatted them for the student project on the NEAS web site. This saves you the time spent compiling data.

Jacob: Does the NEAS web site show *rich* teams vs *poor* teams?

Rachel: To compare teams with high player salaries (rich teams) vs teams with low player salaries (poor teams), you compile the data yourself; we don’t have a good data source.

Jacob: How do we identify *good* teams vs *bad* teams?

Rachel: We identify good vs bad teams by their previous won-loss records.

- A good team in year T has a won-loss record in year T-1 above 50%.
- A bad team in year T has a won-loss record in year T-1 below 50%.

Jacob: Do the good and bad teams change each year?

Rachel: That is up to you. You can use a ten year average to identify good and bad teams, or you can have the good and bad teams change each year.

Jacob: Do we expect a difference between good and bad teams?

Rachel: The team with the best won-loss record gets the last draft pick. Its first-round draft pick may not even start on its team roster and has little effect on its won-loss record the next year. The team with the first draft pick, which had the worst won-loss record one year, may have a good won-loss record the next year. A team might be classified as bad one year and good the next year.

Jacob: Why is the draft pick so important? Does the team with the first draft pick have the highest β_1 parameter?

Rachel: We expect the opposite. A high β_1 means that last year's won-loss record is repeated this year. A bad team that lost 70% of its games last year will lose about 70% this year as well (β_1 = 100%). But if this team gets the first draft pick, it may lose only 50% of its games (β_1 = 0%).

The effects are clearest in basketball, where a single draft pick may turn a team around.

~ The worst team, with the first draft pick, may have a *low* β_1 parameter.
~ The next year, with an average draft pick, it may have an *average* β_1 parameter.

Illustration: Suppose we use a two-variable regression model: $WLR_T = \alpha + \beta_1 WLR_{T-1} + \epsilon$.

Player aging, injuries, and retirements cause mean reversion of 10% each year. With a *random* draft, the regression coefficients are α = 5% and β = 90%: Y = 5% + 90% × X.

Each year may bring two or three outstanding new players. The two or three worst teams, which won about 20% of their games and get the first draft picks, may win about 60% the next year. These teams have estimators of α = 65% and β = –25%.

In practice, draft picks don't always turn out as expected, and we have a smoother continuum of player quality. It may turn out that the optimal regression equation for all teams is α = 10% and β = 80%, or α = 20% and β = 60%.

A student project on good vs bad teams tests whether the same regression coefficients should be used for all teams, both good and bad. This is an intriguing project, which you may enjoy.