Appendix

Exemptions on the PBA National Tour

Currently, there are a variety of ways in which an exemption can be earned on the PBA National Tour. These include the following: (1) Win a "major" championship (U.S. Open, USBC Masters, Tournament of Champions, and PBA World Championship) and earn a two or three year exemption (all remaining accomplishments on this list earn one year exemptions) (2) Win any other National Tour event (3) Be one of the highest ranked bowlers on the World Point Ranking list for the previous season (4) Finish in the top seven at the PBA Regional Players Invitational (entrance into this tournament is earned by finishing in the top 25 in points for any of the seven regions) (5) Be the highest ranked non-exempt member on the World Point Ranking list. In previous seasons, it was possible to earn an exemption by being one of the top finishers in the PBA Tour Trials, leading a region in points for a season (along with the Tour Trials, this method of earning an exemption was replaced by the Regional Players Invitational), or earning a Golden Parachute exemption.

Model and Computational Details

The model that was used in this article is known as a Bayesian hierarchical model. It is known as a "hierarchical" model because the values of certain parameters are dependent on the values of other parameters. In particular, every bowler ability parameter is assumed to come from a normal distribution with it's own mean and standard deviation (this is the prior distribution for the bowler ability parameters), where the mean and standard deviation of this distribution are known as "hyper-parameters." These two parameters also have prior

distributions. The tournament difficulty parameters work in a similar fashion. The notation for the prior distributions in the model specification is given below:

Prior distribution for bowler abilities: $\theta_i \sim Normal(\mu_{bowler}, \tau_{bowler})$, i = 1,2,…,3931

Prior distributions for bowler ability hyper-parameters: $\mu_{bowler} \sim Normal(205, 0.000001)$

$\tau_{bowler} \sim Gamma(0.001, 0.001)$

Prior distribution for tournament difficulties: $\gamma_j \sim Normal(0, \tau_{tournament})$, j = 1,2,…,117

Prior distribution for tournament difficulty hyper-parameter: $\tau_{tournament} \sim Gamma(0.001, 0.001)$

Prior distribution for precision of all bowling scores: $\tau \sim Gamma(0.001, 0.001)$

In the model specification above, the prior distributions for the hyper-parameters are highly non-informative due to the lack of prior knowledge about their possible values. These prior distributions allow the data to determine the posterior distributions of the parameters, with very little influence from the priors. A value of 205 is used for the mean of the normal prior distribution on the mean bowler ability parameter because this value is close to the overall mean of all bowler scores of 205.65. The τ's (as well as the value of 0.000001 for the parameter named μ_{bowler}) represent precisions for each of the normal distributions, where the precision is the inverse of the variance. Finally, τ (without a subscript) refers to the precision for all bowlers' average scores, which is transformed to σ (whose value was estimated at 27.47) in the following manner: $\sigma = 1/\sqrt{\tau}$.

The analysis was performed using WinBUGS (Bayesian Inference Using Gibbs Sampling) software. The first 10,000 observations (after setting reasonable initial values for the

parameters of interest) were discarded as the "burn-in" and every 10^{th} observation out of the next 250,000 was kept, for a total sample size of 25,000. Convergence of the posterior distributions for all parameters was verified by various diagnostic checks available in WinBUGS. All parameter values reported in this article are the posterior means of the relevant parameters using the sampling scheme explained in the previous sentences, with one modification. In order for the tournament difficulty estimates to obey the "sum to zero" constraint ($\sum_{j=1}^{117} \gamma_j = 0$), the mean of all tournament difficulty posterior means was subtracted from each tournament difficulty posterior mean and then added to the posterior mean of each bowler ability estimate (these estimates represent the "baseline" average score in an average tournament with a difficulty of 0) and the posterior mean of the mean bowler ability estimate.

Finally, there are a few assumptions that are made by the model that should be addressed. First, the model assumes that each score is independent of all previous scores and that a bowler's estimated average score for a particular tournament is constant for all games bowled in that tournament (where deviations from this average are given by σ = 27.47). Although the article by Dorsey-Palmateer et al (2004) gives evidence that games are not independent, they do not give an indication of the magnitude of dependence from game to game or if this dependence is the same for every bowler. Therefore, the assumption of independence is made both for convenience and because I do not believe that the results presented in the "Probabilities …" section of the article would change dramatically if dependence was introduced in some fashion. It is also likely true that a bowler's true ability can change over time, as is often overtly evident in many sports. Because the primary focus of this article is on estimation of general variation in bowler abilities and tournament difficulties, however, I did not control for changing ability over a bowler's career. For those interested in a model that does attempt to estimate career trajectories

in sports, see Berry et al (1999). One final assumption is that there is no interaction between bowlers and tournaments or between bowlers and oil patterns. This assumption indicates that all bowlers perform equally better or worse in the same tournament when compared with an average tournament. Therefore, if γ = -5 for a particular tournament, then all bowlers' averages are estimated to drop by 5 pins relative to their averages in an average PBA tournament. Again, this is probably not true, as it is likely that certain bowlers' styles are conducive to scoring well in certain tournaments (most likely due to the specific oil pattern that is used by those tournaments). As stated previously, however, the primary focus of this article is on estimation of particular variation in bowler abilities and tournament difficulties. Due to this focus, coupled with the fact that interaction terms can hamper interpretation of main effects and can also be inappropriate in cases where the number of estimated parameters is greater than the number of data points, I decided not to include any interactions in the model.