Web Appendix

Individual Team Incentives and Managing Competitive Balance in Sports Leagues: An Empirical Analysis of Major League Baseball

Michael Lewis

The estimation results reported in Section 5 (Table 3) are used as the basis for the subsequent analyses and policy experiments. The model has an intuitive structure and yields results that possess face validity. Nonetheless, the general modeling approach used in the study requires a significant number of decisions and assumptions. For example, the estimation results are generated under an assumption that team's use a rolling four-year planning horizon. This type of assumption naturally inspires questions. In this appendix we attempt to address these types of concerns by providing additional background material related to the selection of the specification ("main model").

Table A1 provides fit statistics for a variety of alternative model specifications. In terms of alternative specifications the table includes results for several restricted models. The first restricted model removes the effect of divisional finish from the reward function. This restriction yields a log-likelihood of –3793.1 and a Bayesian Information Criteria (BIC) score of 7754.0. In comparison, the "Main Model" yields a log-likelihood of –3783.5 and a BIC score of 7759.9. The second restricted model drops the load factor term from the reward function. The final two restricted models restrict the role of population. In the model labeled (Payroll X Population) the interactions between payroll and payroll squared with the population size are set to zero. The model labeled population adds an additional restriction to the (Payroll × Population) model by eliminating the differential effects of revenue-sharing on market size categories. The main model is preferred over the restricted models in terms of the BIC measure and comparisons of the log-likelihoods of the restricted models to the main model all yield significant chi-square statistics.

Table A1: Fit Statistics

Model	**Description**	**Parms**	**LL**	**BIC**
Main Model	Detailed in equation (3).	27	-3783.5	7748.2
Restricted Models				
Finish	Divisional Finish and Finish squared are dropped from the reward function	25	-3793.1	7754.0
Load Factor	Load Factor is dropped from the reward function.	26	-3792.7	7759.9
Revenue-Sharing	The revenue sharing adjustment terms are set to zero for all market size categories	24	-3797.0	7755.0
Salary X Population	The interactions between payrolls and population are set to zero	25	-3862.9	7893.5
Population	All terms involving population (revenue sharing and payroll interactions) are set to zero.	23	-3894.4	7943.2

To determine the appropriate dynamic structure we sequentially evaluated several different decision making horizons. The model that assumed a four-year decision horizon yielded the best fit with a log-likelihood of –3783.5. In contrast, the log-likelihood for the model that assumed a decision horizon of three years was –3786.1, while for a five-year horizon the log-likelihood was –3784.2.

In addition to a comparison of relative fit we also evaluated model performance via an analysis of out of sample fit. For this analysis the data was randomly divided into an estimation sample of 638 observations and a holdout sample of 184 observations.[1] To assess out of sample fit we begin by computing the implied optimal dynamic salary policies as in Section 5. The dynamic nature of the model makes this a non-trivial task. The key point is that the optimal policies are not just a function of the market size but also a function of the team's previous load factor and previous divisional finish (see Figure 7). The implied optimal payrolls are then compared with actual payroll selections in the holdout sample. Table A2 provides out of sample fit statistics for the "Main Model" and for four of the restricted specifications described in Table A1.

Table A2: Out-of-Sample Fit Results

	Holdout Sample	Main Model	No Finish	No Load Factor	No Rev. Share
Relative Payrolls					
Variance	.116	.099	.0770	.067	.0693
Range	187%	153%	113%	105%	111%
Fit Measures					
MSE		.106	.122	.127	.126
MAD		.232	.279	.285	.285
Within 15%		36.6%	29.2%	30.4%	29.4%
Within 30%		65.3%	55.4%	57.8%	52.9%

The model specification in equation (3) out performed the restricted versions along a variety of criteria such as mean absolute deviation (MAD) and mean square error (MSE). The bottom two rows of the table provide the hit rates for the model's predictions compared to actual payrolls observed in the data. The Main Model predicts 36.6% percent of the relative payrolls within a 15% range and 65% within a 30% range. Of the restricted models the worst performing is the version that drops the revenue sharing terms.

The improvement in fit relative to the restricted models occurs because the focal model is better able to match the variance in the data. In the actual data the relative payrolls range from 20% to 207% of the league average. The Main Model's predictions have a range of 153%. The model that drops the load factor reproduces the least variance with the predicted payrolls ranging from 60% to 165% of the average. The model that restricts the population effects to zero is not included in the holdout sample analysis. The difficulty with this model is that the optimization analysis does not yield solutions within the range of the observed data.

The preceding analyses suggest that the models have some difficulty in accounting for the variance in owners' decisions. This pattern suggests that there may be factors beyond the measures in the model that influence payroll decisions. As such we also evaluated the usefulness of controlling for unobserved heterogeneity across owners. To account for the possibility of unobserved heterogeneity we estimated a finite mixture model. To extend the model to account for variability in preferences a latent class

[1]The estimation results reported in the paper use the full sample of data. Given the number of parameters to be estimated and the limited amount of data we are reluctant to exclude data from the model used to generate policy recommendations.

approach is used (Kamakura and Russell 1989). This approach assumes that the population consists of M types, where π_m is the proportion of type m in the population. For the finite mixture approach, the sample likelihood is as follows:

$$\prod_{i=1}^{N} \sum_{m=1}^{M} \Pr\left(d_{1i}^{m}, d_{2i}^{m}, ..., d_{Ti}^{m} \mid type = m\right) \times \pi_m ,$$

where

$$\Pr(d_i^m(1), d_i^m(2), .. d_n^m(T)) = \prod_{t=1}^{T} \Pr(d(t) \mid S(t), type = m) . \qquad \text{(A1)}$$

The use of a finite mixture model to account for unobserved heterogeneity significantly increases the computational burden because the dynamic optimization problem must be repeatedly solved for each type in the population.

The logic for accounting for unobserved preference heterogeneity across owners is that it is reasonable and likely that owners may differ in their preferences for economic returns and on-field success. While the logic for this type of model is clear, in practice the model was problematic. In terms of the BIC criteria a full two-segment model (all parameters allowed to vary for each segment) did not yield an improvement relative to the homogenous population model (7834.4 versus 7748.2). In addition, this two-segment model yielded many non-significant parameters. Given the policy formulation goals of the research this lack of confidence is a critical shortcoming. From a qualitative perspective, the two-segment results yielded a large segment similar in character to the results reported in Table 3 and a small segment that seems to garner greater utility from higher finishes than most teams. This “on-field success-oriented” segmented was approximately 10% of the population.

In addition, while the likelihood of diverse preferences for winning and revenues increases the appeal of allowing for preference heterogeneity, the policy formulation goals of the research limit the viability of this type of latent class analysis. The obvious implications of segment level results, is that revenue sharing should be customized based not only on market size but also on ownership “type.” While this type of customization might be feasible and advantageous in traditional consumer applications, given that the collective bargaining agreement is something that owners must agree to it is doubtful that discrimination based on a latent trait would be acceptable.

Kamakura, Wagner and Gary J. Russell (1989), “A Probalistic Choice Model for Market Segmentation and Elasticity Structure,” *Journal of Marketing Research*, 26 (November), 379–90.