

Modeling the Evolution of Distributions: An Application to Major League Baseball

GARY KOOP*

*Department of Economics, University of Glasgow,
Glasgow, G12 8RT, U.K.*

E-mail: G.Koop@socsci.gla.ac.uk

May 2001

In this paper, we develop Bayesian techniques for modeling the evolution of entire distributions over time and apply them to the distribution of team performance in Major League baseball for the period 1901-2000. Such models offer insight into many key issues (e.g. competitive balance) in a way that regression-based models cannot. The models involve discretizing the distribution and then modeling the evolution of the bins over time through transition probability matrices. We allow for these matrices to vary over time and across teams. We find that, with one exception, the transition probability matrices (and, hence, competitive balance) have been remarkably constant across time and over teams. The one exception is the Yankees, who have outperformed all other teams.

Keywords: Bayesian, Gibbs sampler, ordered probit, Damn Yankees

*The author is a visiting professor at the University of Edinburgh. I would like to thank John Geweke and Dale Poirier for helpful comments.

1 Introduction

Traditional panel data methods involve the modeling of the conditional mean of a distribution. Inclusion of heteroscedasticity can allow for modeling of the conditional variance. However, there are some cases where the researcher has a panel of data but is interested in modeling the evolution of the entire distribution. The field of economic growth is a prominent example where the influential "Twin Peaks" papers of Danny Quah (e.g. Quah, 1993a,b, 1996, Durlauf and Quah, 2000, see also Paap and van Dijk, 1998) have found that the cross-country distribution of income has become bimodal over time. Such a finding would have been hard to detect by examination of means and variances alone. These Twin Peaks papers have developed (non-Bayesian) statistical methods for modeling the evolution of distributions. One purpose of the present paper is to develop Bayesian analogues to the models used in the Twin Peaks papers and extend them in empirically important directions. However, our primary purpose is to apply these methods to the case of Major League baseball.

From a policy perspective, the large and increasing share of GDP which is devoted to sport motivates an interest in this sector of the economy. From an empirical perspective, sport (and baseball in particular) is a field where extensive and reliable data is readily available. From an economic theory perspective, sports provides an interesting range of markets and institutional structures which require careful theoretical modeling. Furthermore, in many cases theoretical or econometric models developed in a sporting context generalize to other markets or data sets. A large and growing academic literature attests to the interest in sports.

Some of the theoretical work in this area implies an interest in understanding the evolution of the distribution of team performances. Consider, for instance, what Vrooman (2000) calls the "Yankee paradox". The Yankees are a rich team which, over the years, has tended to dominate baseball. This could be bad for baseball in that spectators prefer

competitive balance and, hence, a league where the Yankees are predominant may generate less fan interest (and, hence, less profit) than one where victory is more equally shared among teams.¹ Hence, the Yankees may not simply have an incentive to win as many games as possible, they also want to make sure that their competitors do not lose too much. Economists (see, among many others, Flynn and Gilbert, 2001, Quirk and Fort, 1992, Quirk and El Hodiri, 1974, Rottenberg, 1956, Szymanski, 2001 and Vrooman, 2000), thus, have several interesting issues to model and empirically investigate. For instance, unlike most oligopoly situations (but perhaps relevant for firms like Microsoft seeking to avoid antitrust prosecution) firms (i.e. teams) have an incentive to keep their competitors in business.

Questions relating to competitive balance cannot be empirically addressed by looking at the mean of the cross-team distribution of winning percentage, since this is always .500 by construction. Even an examination of year-by-year standard deviations of team performance provides only limited information since a reduction in standard deviation could be consistent with the same teams winning or losing all the time. It is only through an examination of the entire distribution and how it evolves over time that we can gain an empirical understanding of issues relating to competitive balance in baseball.

The purpose of the present paper is not to develop a economic model of what the distribution of team performance should look like or how it should evolve over time. Our more limited purpose is to develop models and methods for producing the empirical facts which theorists might be interested in. These techniques can be used to address questions such as:

- "Is baseball becoming more unequal in terms of team performance?"
- "How much mobility is there across the performance distribution? Does mobility

¹In contrast, in professional football, structures have been put in place to insure a high degree of competitive balance between teams and some of the sporting press has argued that the result has been too much parity and a lessening of fan interest.

change over time?”

- ”Is there a dominance of rich, large market teams, over poor small market teams? Is it worsening over time?”
- ”How likely are dynasties to occur in baseball?”
- ”How long does it take for new teams to establish themselves?”
- ”How common is it for a historically weak team to break into the upper echelons of baseball?”

We begin with a model similar to the one in Quah (1993a,b). That is, we discretize the distribution of team performance in each year into different bins and model the transition probability matrix (i.e. the probability of moving between bins) using a first order Markov chain. However, such a model assumes that the manner in which team performance distributions evolve is time invariant. Several of the issues we wish to address require time varying transition probabilities (e.g. ”is it harder now for a small market team to build a contender than in the past?”). Furthermore, questions like ”is it difficult for a small market team to build a champion?” relate to the issue of whether transition probabilities vary over teams. Accordingly, we motivate and develop a new model, involving an ordered probit specification, which allows for the possibility that transition probabilities vary over time and over teams. We also develop methods for testing these extensions. These methods are used in a study of team performance in Major League baseball, but we discuss their relevance in other academic fields (e.g. economic growth and the study of wage distributions). Furthermore, we describe many simple extensions which may be relevant in other applications.

Our empirical results indicate that the standard deviation of team performance has been, on average, decreasing over the 1901-2000 period, which indicates that one component of competitive balance is not worsening over time. However, the transition probability

matrix/matrices we estimate do not seem to be changing over time. The latter finding indicates that the degree of persistence and mobility within the distribution has not changed over time in a noticeable way. Hence, competitive balance has at most been minimally affected by the expansion in team numbers, introduction of free agency, growth in team wealth inequality and many other changes which have occurred over the century. Furthermore, the transition probability matrix/matrices we estimate indicates a high degree of mobility by teams within the distribution, which is consistent with baseball being a highly competitive sport where many different teams have done well/poorly over the years. With one exception, we find that all teams are equally competitive (in the sense that they face the same transition probability matrix). The exception is the Yankees who have a transition probability matrix which is significantly different than other teams. In particular, this matrix indicates that, when the Yankees are near the top of the league, they tend to stay there much longer (i.e. the Yankees have had more dynasties) and, when they are near the bottom of the league, they tend to languish there for a shorter period.

The remainder of the paper is organized as follows. The next section discusses the data and presents some summary statistics while the third section presents the model and develops methods for Bayesian estimation and inference. The fourth section presents empirical results and the fifth concludes.

2 The Data

Our data set comprises the winning percentage of every team in every available year in the 1901-2000 period and is taken from The Baseball Archive (www.baseball1.com). There are several issues and characteristics of this data set which deserve comment. Major League baseball is comprised of two leagues, the American (AL) and National (NL) leagues. The choice of the year 1901 is a natural one since this is the year the American League was founded (the NL is older). Although baseball records go back to the mid-19th century,

there were several competing leagues prior to 1901 and the turnover in leagues and teams was substantial. From 1901 to the present, the structure of Major League baseball has been relatively constant and, apart from the Negro Leagues and the short-lived Federal League of 1914-1915, there have been no significant other leagues in competition with Major League baseball in the USA. From 1901 through 1960, the American and National Leagues each consisted of eight teams. In 1961, the American League created two new expansion teams, the Los Angeles Angels and the Washington Senators (the old Washington Senators having moved to Minnesota the previous year). Expansion continued sporadically over the next 40 years and, at present, the Major Leagues consist of 30 teams, 16 in the NL and 14 in the AL. This means that we have 100 observations for 16 teams, but fewer for the remaining 14 teams.

A further characteristic of Major League baseball is that, although no team has ever gone out of business, many teams have moved to different cities. Since such changes involved moving the whole team (including players, management, etc.) to the new city, we do not consider such relocations as implying a new team is created. Nevertheless, formally speaking, it is perhaps better to consider this paper as modeling the evolution of franchises over time. So, for instance, we have 100 years of data on the Athletics franchise even though this franchise played in Philadelphia (1901-1954), Kansas City (1955-1967) and Oakland (1968-2000). Where the franchise name changed, we will use the name of the current team to avoid confusion. So, for instance, what we call the "Orioles franchise" began as the Milwaukee Brewers (1901), before becoming the St. Louis Browns (1902-1953) which in turn became the Baltimore Orioles (1954-2000). Full details of the history of all 30 franchises are given in Table 1.

Our choice of winning percentage as the focus of our study also deserves comment. All baseball games are either won or lost, hence winning percentage is a sensible measure of team performance. Given that the number of games played has changed over time, other choices (e.g. games won) are not appropriate. An alternative measure might be the rank

of the team in the league. For some issues, the rank of a team might be a more suitable variable to model than winning percentage. For instance, the team which did very well in terms of winning percentage, but finished in second place year after year might find their fans arguing that rank is a more important measure of performance than winning percentage. However, modeling of rank is complicated by the expansion of the league and the move to divisional play in 1969 (i.e. the AL and NL were each sub-divided into an East and West division in 1969 and, subsequently, a Central division was added to both leagues). Furthermore, modeling winning percentage allows for an investigation of important issues such as inequality of performance. For instance, if the winning percentage of the top team in each league was very similar to the second ranked team, this likely means that the season was competitive and exciting. Merely looking at the rankings would not reveal this information. Accordingly, we use winning percentage as our variable of interest and focus on the question of how to model the evolution of the distribution of winning percentage over the years.

In this paper we combine teams from the AL and NL together and, hence, for each year we have a distribution involving all Major League teams. Preliminary data analysis indicated that the properties of the AL and NL were quite similar and, hence, combining data the two leagues to obtain larger sample sizes in each year was acceptable.

In summary, our raw data is y_{it} = the winning percentage in year t for franchise i with $i = 1, \dots, 30$ and $t = \underline{t}_i, \dots, 2000$, where \underline{t}_i is the founding year of franchise i . In total, we have 1,956 observations.

Table 1: Name and History of Franchises

1. **Angels**, played as Los Angeles Angels (1961-1964), California Angels (1965-1996), Anaheim Angels (1997-2000).
2. **Diamondbacks**, played as Arizona Diamondbacks (1998-2000).

3. **Braves**, played as Boston Braves (1901-1952), Milwaukee Braves (1953-1965), Atlanta Braves (1966-2000).
4. **Orioles**, played as Milwaukee Brewers (1901), St. Louis Browns (1902-1953), Baltimore Orioles (1954-2000).
5. **Red Sox**, played as Boston Red Sox (1901-2000).
6. **White Sox**, played as Chicago White Sox (1901-2000).
7. **Cubs**, played as Chicago Cubs (1901-2000).
8. **Reds**, played as Cincinnati Reds (1901-2000).
9. **Indians**, played as Cleveland Indians (1901-2000).
10. **Rockies**, played as Colorado Rockies (1993-2000).
11. **Tigers**, played as Detroit Tigers (1901-2000).
12. **Marlins**, played as Florida Marlins (1993-2000).
13. **Astros**, played as Houston Astros (1962-2000).
14. **Royals**, played as Kansas City Royals (1969-2000).
15. **Dodgers**, played as Brooklyn Dodgers (1901-1957), Los Angeles Dodgers (1958-2000).
16. **Brewers**, played as Seattle Pilots (1969), Milwaukee Brewers (1970-2000).
17. **Twins**, played as Washington Senators (1901-1960), Minnesota Twins (1961-2000).
18. **Expos**, played as Montreal Expos (1969-2000).

19. **Mets**, played as New York Mets (1962-2000).
20. **Yankees**, played as Baltimore Orioles (1901-1902), New York Yankees (1903-2000).
21. **Athletics**, played as Philadelphia Athletics (1901-1954), Kansas City Athletics (1955-1967), Oakland Athletics (1968-2000).
22. **Phillies**, played as Philadelphia Phillies (1901-2000).
23. **Pirates**, played as Pittsburgh Pirates (1901-2000).
24. **Padres**, played as San Diego Padres (1969-2000).
25. **Giants**, played as New York Giants (1901-1957), San Francisco Giants (1958-2000).
26. **Cardinals**, played as St. Louis Cardinals (1901-2000).
27. **Mariners**, played as Seattle Mariners (1977-2000).
28. **Devil Rays**, played as Tampa Bay Devil Rays (1998-2000).
29. **Rangers**, played as Washington Senators, expansion² (1961-1971), Texas Rangers (1972-2000).
30. **Blue Jays**, played as Toronto Blue Jays (1977-2000).

It is useful to present a few summary statistics on winning percentage. Pooling all the data together over all years, we obtain a standard deviation of 0.084, a minimum and maximum of 0.230 and 0.760, respectively, the mean being 0.500 by definition. The 25th and 75th percentiles are 0.440 and 0.560. These statistics highlight the well-known fact that baseball is a highly competitive sport on a game-by-game basis. A team which wins 60%

²In 1961 the original Washington Senators moved to Minnesota, Washington was immediately given a new franchise which was also named the Senators.

of its games will usually finish first in its division or league while a team which wins 40% of its games will be considered to have had a very poor year.

Figures 1a)-1c) indicate some other interesting properties of the data. Figure 1a), which is a histogram of all the data, exhibits some non-Normality as it is skewed to the right. Plotting the histogram of winning percentage on a year by year basis is precluded by space considerations. However, it is worth noting that the shape of such histograms often change enormously over time. For instance, in some years the distribution of winning percentage seems roughly Normal, in other years Uniform, in other years U-shaped, in other years continuously increasing to the right, other years decreasing to the right, etc. Such year by year plots indicate that the distribution of winning percentage is evolving over time in a manner which it might be of interest to model formally and highlights the dangers of using a particular parametric family of distributions (e.g. the Normal).

Figure 1b), which plots the standard deviation of y_{it} for each year, demonstrates another interesting property of the evolution of the distribution of winning percentage over time. Despite many irregular fluctuations, overall the standard deviation of winning percentage has tended to drop over time. We do not formally model the relationship between competitive balance and attendance in this paper. But as an interesting tangent, Figure 1c) plots the evolution of the standard deviation of team attendance (obtained from the Baseball Archive) over time. This figure demonstrates a substantial increase in the dispersion of attendance across teams. In the sporting press, one often hears the argument that there is a growing divide between rich and poor teams and, since the rich teams can buy better players, a growing inequality of outcomes. Figures 1b) and 1c) provide evidence which undermines this story. Insofar as team attendance is a proxy for team wealth, Figure 1c) does indicate a substantial increase in wealth inequality over time, however Figure 1b) indicates inequality of wealth does not necessarily imply inequality of team performance.

Insert Figure 1 here

Table 2 gives summary statistics for individual teams, including mean, standard deviation, minimum, maximum, the first order autocorrelation and average attendance. Rows corresponding to the original 16 teams, for which we have a full 100 years of data, have been made bold. Given that attendance has increased substantially over time, the attendance figure is a Z-score (i.e. we subtract off yearly mean attendance and divide by yearly standard deviation). Apart from very recently established teams (e.g. Diamondbacks, Devil Rays, Rockies and Marlins), performance results for each team looks fairly similar. For instance, if we focus on the original 16 teams, they all have roughly similar standard deviations and autocorrelations. The Yankees superior performance reveals itself in having a higher mean winning percentage. Autocorrelations for the original 16 teams tend to be around .4 to .7, indicating a moderately strong degree of persistence in performance. One might expect baseball dynasties to be revealed by high autocorrelations and low standard deviations, but the Yankees have roughly average values for these.

The new teams which were created after 1960 tend to have lower standard deviations and autocorrelations than the original 16. However, these facts could be consistent with the decreasing inequality of Major League baseball noted in Figure 1b). The attendance figures do confirm the common view that there are some small and some large market teams and the latter can, in the long run, sustain a larger fan base. Well-known small market teams in cities such as Cincinnati, Minnesota, Oakland, Pittsburgh and Seattle do tend to have lower average attendance than large market teams in cities like New York, Los Angeles, Boston and Chicago. However, it is worth stressing that the cross-team variation is not enormous (e.g. even the Yankees' attendance has averaged less than a single standard deviation above the mean) and that looking at averages over their entire history obscures interesting changes over time (e.g. the Braves and Indians look like small market teams in Table 2, but both franchises have had extended periods where their attendance figures were among the highest in baseball).

Table 2: Summary Statistics for Franchises

| Franchise | Mean | St.Dev. | Min. | Max. | Auto Corr. | Attendance |
|------------------|-------------|-------------|-------------|-------------|-------------|--------------|
| Angels | .483 | .048 | .406 | .574 | -.164 | .095 |
| D'Backs | .514 | .108 | .401 | .617 | n.a. | 1.068 |
| Braves | .478 | .092 | .248 | .654 | .687 | -.416 |
| Orioles | .478 | .093 | .279 | .673 | .643 | -.432 |
| Red Sox | .512 | .083 | .279 | .691 | .608 | .258 |
| White Sox | .505 | .074 | .325 | .649 | .428 | .110 |
| Cubs | .505 | .083 | .364 | .763 | .637 | .293 |
| Reds | .506 | .076 | .344 | .686 | .492 | -.178 |
| Indians | .512 | .072 | .333 | .721 | .481 | -.197 |
| Rockies | .481 | .042 | .414 | .535 | .267 | 2.017 |
| Tigers | .512 | .070 | .325 | .656 | .398 | .341 |
| Marlins | .448 | .073 | .333 | .568 | -.087 | -.392 |
| Astros | .494 | .059 | .398 | .630 | .460 | -.144 |
| Royals | .505 | .057 | .398 | .630 | .343 | -.046 |
| Dodgers | .522 | .077 | .316 | .682 | .555 | .947 |
| Brewers | .480 | .060 | .395 | .590 | .548 | -.515 |
| Twins | .475 | .081 | .252 | .651 | .506 | -.607 |
| Expos | .488 | .069 | .321 | .649 | .397 | -.466 |
| Mets | .474 | .098 | .250 | .667 | .729 | .507 |
| Yankees | .564 | .082 | .329 | .714 | .604 | .817 |
| Athletics | .480 | .111 | .235 | .704 | .685 | -.357 |
| Phillies | .456 | .087 | .278 | .623 | .620 | -.367 |
| Pirates | .518 | .084 | .273 | .741 | .635 | -.339 |
| Padres | .456 | .069 | .321 | .605 | .391 | -.493 |
| Giants | .538 | .075 | .353 | .693 | .451 | .359 |
| Cardinals | .514 | .082 | .314 | .688 | .533 | .059 |
| Mariners | .455 | .061 | .350 | .562 | .354 | -.656 |
| D'Rays | .414 | .022 | .389 | .419 | n.a. | -.602 |
| Rangers | .467 | .067 | .346 | .586 | .372 | -.423 |
| Blue Jays | .493 | .087 | .327 | .615 | .735 | .821 |

These summary statistics and figures offer a preliminary picture of a league which is becoming slightly more equal over time. However, apart from this there is no strong evidence of major changes in the competitive balance of baseball. Table 2 tells a simple story where big market teams tend to win slightly more often on average than small market teams, but no major differences in standard deviations or autocorrelations exist between teams. However, to gain a deeper understanding of how baseball is evolving over time a formal statistical model is required.

3 The Model

Quah (1993a,b) begins by discretizing the cross-country distribution of GDP per capita into various bins (e.g. bins for rich, middle income and poor countries) and then assumes countries move between bins according to a Markov transition probability matrix which is estimated from the data. The elements of this matrix can be interpreted as, e.g., the probability that a poor country will become rich or stay poor in the following period. In this section, we develop Bayesian methods for a similar model. Our transition probabilities can be interpreted as, e.g., the probability that a team at the bottom of the league one year can turn itself into a contender the next. However, many of the questions we wish to address involve the transition probabilities varying over time (e.g. have free agent rules and increasing wealth inequality made it harder to build a contender?) or over teams (e.g. is it harder for small market teams to build into contenders?). Accordingly, we develop a model which allows the transition probabilities to vary over time and across teams and develop Bayesian methods for estimation and inference.

3.1 An Ordered Probit Model

Our raw data, percentage of games won in a season, is bounded between zero and one. We divide this interval into m bins and let b_0, \dots, b_m be the limits of the bins with $b_0 = 0$ and

$b_m = 1$. Define, for $j = 1, \dots, m$,

$$S_{it} = j \text{ if } b_{j-1} < y_{it} \leq b_j, \quad (3.1)$$

where $i = 1, \dots, N$ indexes teams and $t = \underline{t}_i, \dots, T$ indexes years. Let

$$p_{jit} = \Pr[S_{it} = j | \Omega_t] \quad (3.2)$$

be the probability team i in period t is in bin j where Ω_t is an information set containing information available at time t . There are many possible choices of a model for p_{jit} . In the Twin Peaks literature, it is common assume (3.2) follows a Markov chain with transition probabilities that are constant over time and across individuals. That is, the transition probabilities are given by:

$$p_{jh} = \Pr[S_{it} = j | S_{i,t-1} = h], \quad (3.3)$$

and p_{jh} are unknown parameters estimated from the data for $j = 1, \dots, m-1$ and $h = 1, \dots, m$ (i.e. we have the restriction $p_{mh} = 1 - \sum_{j=1}^{m-1} p_{jh}$). For future reference, note that (3.3) has $m^2 - m$ free parameters. Alternatively, in a macroeconomic time series context with two regimes, Filardo and Gordon (1998) use a probit specification for transition probabilities analogous to p_{jk} . The model developed in this paper can nest these two alternatives.

We use an ordered probit specification (see, e.g. Maddala, 1983, pages 46-49). Since we are modelling a distribution, such a choice is a natural one (i.e. the bins fall in an obvious order). Furthermore, computation is particularly easy with such a model. Of other possible choices, Bayesian analysis of multinomial logit models is computationally difficult in that importance sampling (Koop and Poirier, 1993) or use of a Metropolis-Hastings algorithm (Lahiri and Gao, 2001, Quinn, Martin and Whitford, 1999) is required. Bayesian analysis of multinomial probit is complicated by difficulties associated with identification and prior elicitation of parameters in the error covariance matrix (McCulloch and Rossi, 1994, 2000, McCulloch, Polson and Rossi, 2000 and Nobile, 2000). Relative to simple Markov models of

the Twin Peaks literature, the ordered probit specification allows for transition probabilities to depend on explanatory variables, vary over time or across individuals and include higher order lags of S_{it} in a simple and parsimonious fashion.

If we define the cumulative probability, for $j = 1, \dots, m - 1$:

$$\eta_{jit} = \sum_{h=1}^j p_{hit}, \quad (3.4)$$

then the ordered probit specification is:

$$\eta_{jit} = \Phi(\gamma_j - x'_{it}\beta), \quad (3.5)$$

where $\Phi(a)$ is the c.d.f. of the standard Normal evaluated at point a , x_{it} is a k -vector of explanatory variables, γ_j is a scalar parameter for $j = 1, \dots, m - 1$ with $\gamma_1 < \gamma_2 < \dots < \gamma_{m-1}$ and β is a k -vector of unknown coefficients. As discussed in, e.g., Albert and Chib (1993), we can impose, with no loss of generality, the identifying restriction $\gamma_1 = 0$. This model implies

$$p_{1it} = \Phi(-x'_{it}\beta),$$

$$p_{2it} = \Phi(\gamma_2 - x'_{it}\beta) - \Phi(-x'_{it}\beta),$$

....

$$p_{m-1,it} = \Phi(\gamma_{m-1} - x'_{it}\beta) - \Phi(\gamma_{m-2} - x'_{it}\beta),$$

$$p_{mit} = 1 - \Phi(\gamma_{m-1} - x'_{it}\beta).$$

An alternative way of interpreting this model is in terms of a latent variable. That is, let S_{it}^* be Normally distributed with mean $x'_{it}\beta$ and standard deviation 1. The model in

(3.5) is equivalent to one where we observe $S_{it} = j$ if $\gamma_{j-1} < S_{it}^* \leq \gamma_j$ with $\gamma_0 = -\infty$ and $\gamma_m = \infty$. The latent data formulation provides the motivation for our Gibbs sampling with data augmentation algorithm described below.

To relate this model to the Markov model used in the Twin Peaks literature, consider the case where $x_{it} = (d_{1it}, \dots, d_{mit})'$ where $d_{hit} = 1$ if $S_{i,t-1} = h$ and equals zero otherwise. The coefficient vector then becomes $\beta = (\beta_1, \dots, \beta_m)'$. Define the standard Normal random variable $Z_{ith} = S_{it}^* - \beta_h$. The transition probabilities are (for $h = 1, \dots, m$ and $j = 1, \dots, m-1$):

$$\Pr[S_{it} = j | S_{i,t-1} = h] = \Pr[\gamma_{j-1} < Z_{ith} \leq \gamma_j]. \quad (3.6)$$

In other words, with x_{it} containing only dummies for last period's bins, this model looks identical to a standard Markov model except that the transition probabilities are expressed in terms of integrals over regions of the standard Normal (i.e. compare equations 3.3 to 3.6).

The ordered probit model, however, uses order information to place some (in most cases sensible) restrictions on the transition probability matrix. This is seen most clearly by noting that an unrestricted transition probability matrix has $m^2 - m$ free parameters whereas the transition probability matrix defined by (3.6) has only $2m - 2$ free parameters (i.e. β_1, \dots, β_m and $\gamma_2, \dots, \gamma_{m-1}$). Should these restrictions be deemed objectionable, they can easily be relaxed by taking what Maddala (1983), page 46, refers to as "an obvious generalization" of the ordered probit model and replacing (3.5) by:

$$\eta_{jit} = \Phi(\gamma_j - x'_{it}\beta^{(j)}), \quad (3.7)$$

where $\beta^{(j)}$ is a k -vector which varies across bins. With such a specification, it can easily be verified that, if $x_{it} = (d_{1it}, \dots, d_{mit})'$, there are $m^2 - 2$ free parameters and, after imposing $m - 2$ identifying restrictions, (3.7) yields an unrestricted transition probability matrix.

The model developed here has the advantage that x_{it} can contain be defined in many

ways, allowing for the transition probability to vary over time or across teams. Several different choices are discussed below.

Bayesian inference in the ordered probit model is discussed in Albert and Chib (1993). The likelihood function is determined by (3.5) and is given by:

$$p(S|\beta, \gamma, X) \propto \prod_{i=1}^N \prod_{t=\underline{t}_i}^T \sum_{j=1}^m 1(S_{it} = j) [\Phi(\gamma_j - x'_{it}\beta) - \Phi(\gamma_{j-1} - x'_{it}\beta)], \quad (3.8)$$

where $1(\cdot)$ is the indicator function, $\gamma = (\gamma_2, \dots, \gamma_{m-1})'$ is a vector containing unknown parameters (remember $\gamma_0 = -\infty$, $\gamma_1 = 0$ and $\gamma_m = \infty$), $S = (S_{1\underline{t}_1}, \dots, S_{1T}, S_{2\underline{t}_2}, \dots, S_{2T}, \dots, S_{N\underline{t}_N}, \dots, S_{NT})'$ and

$$X = \begin{bmatrix} x'_{1\underline{t}_1} \\ \cdot \\ \cdot \\ x'_{1T} \\ x'_{2\underline{t}_2} \\ \cdot \\ \cdot \\ x'_{NT} \end{bmatrix}.$$

Bayesian inference is based on the posterior which is proportional to the prior times the likelihood function. In this paper, we use the following prior for β :

$$p(\beta) = f_N(\underline{b}, \underline{B}), \quad (3.9)$$

where $f_N(\underline{b}, \underline{B})$ denotes the multivariate Normal density with mean \underline{b} and covariance \underline{B} . In our empirical section, we carry out a prior sensitivity analysis by computing results for a specific choice for \underline{b} and \underline{B} as well as a noninformative choice where $\underline{B}^{-1} = 0_{k \times k}$. For the elements of γ we use a noninformative prior of the form:

$$p(\gamma_j | \{\gamma_h, h \neq j\}) = f_U(\gamma_{j-1}, \gamma_{j+1}), \quad (3.10)$$

where $f_U(\gamma_{j-1}, \gamma_{j+1})$ is the Uniform density over the interval $(\gamma_{j-1}, \gamma_{j+1})$.

The posterior, which is simply proportional to the product of (3.8), (3.9) and (3.10), does not have a computationally convenient form. However, conditional on the latent data, $S^* = (S_{1\bar{t}_1}^*, \dots, S_{1T}^*, S_{2\bar{t}_2}^*, \dots, S_{2T}^*, \dots, S_{N\bar{t}_N}^*, \dots, S_{NT}^*)'$, things simplify considerably and we obtain:

$$p(\beta|S^*, S, X) = f_N(\bar{b}, \bar{B}), \quad (3.11)$$

where

$$\bar{B} = (\underline{B}^{-1} + X'X)^{-1}$$

and

$$\bar{b} = \bar{B}(\underline{B}^{-1}\underline{b} + X'S^*).$$

The posterior conditionals for γ_j for $j = 2, \dots, m-1$ are given by:

$$p(\gamma_j|S^*, S, X, \{\gamma_h, h \neq j\}) = f_U(\bar{\gamma}_{j-1}, \bar{\gamma}_{j+1}), \quad (3.12)$$

where $h = 1, \dots, m$ and

$$\bar{\gamma}_{j-1} = \max\{\max\{S_{it}^* : S_{it} = j\}, \gamma_{j-1}\},$$

and

$$\bar{\gamma}_{j+1} = \min\{\{S_{it}^* : S_{it} = j+1\}, \gamma_{j+1}\}.$$

Finally, the S_{it}^* 's are conditionally independent over all i and t with

$$p(S_{it}^*|S_{it} = j, X, \beta, \gamma) = f_N(x_{it}'\beta, 1) 1(\gamma_{j-1} < S_{it}^* \leq \gamma_j). \quad (3.13)$$

Hence, a Gibbs sampling algorithm which involves sequentially sampling from (3.11), (3.12) and (3.13) can be set up. Since this algorithm involves only drawing from the Normal,

Uniform and truncated Normal distributions, it is simple to program and computationally efficient.

Several hypotheses of interest relate to subsets of the coefficients. For instance, if X includes time dummies then we are interested in testing whether their coefficients are zero. If they are, then the transition probabilities will be time invariant. Such considerations motivate interest in calculating Bayes factors comparing:

$$M_0 : R\beta = r$$

against

$$M_1 : R\beta \neq r,$$

where R is a known $r \times k$ matrix and r is a known $r \times 1$ vector. It can be shown that (see, e.g., Verdinelli and Wasserman, 1995), for priors of the form used in this paper, the Bayes factor comparing M_0 to M_1 takes the form:

$$B_{01} = \frac{p(R\beta = r|S, X, M_1)}{p(R\beta = r|M_1)}, \quad (3.14)$$

where the numerator and denominator are the posterior and prior, respectively, under M_1 for $R\beta$, evaluated at r . Given that the prior for β is Normal (see 3.9), the denominator can be evaluated directly. Since the posterior of β , conditional on S^* , is Normal (see 3.11) and

$$p(R\beta = r|S, X, M_1) = \int p(R\beta = r|S^*, S, X, M_1) dS^*,$$

we can use output from the Gibbs sampler to evaluate the numerator in a simple fashion.

Results presented below are based on 55,000 Gibbs draws with the initial 5,000 dropped to remove startup effects. The draws pass standard MCMC diagnostic tests.

3.2 A Discussion of Other Models

The purpose of this paper is to develop statistical methods appropriate for modeling the evolution of the winning percentage distribution of Major League baseball. In the following section, we will present evidence that the model outlined above is of sufficient generality to do so. In preliminary empirical work, we considered several more general models. Since these did not add anything to the understanding of Major League baseball, they have not been included here. However, insofar as the statistical methods developed in this paper are intended to be widely applicable in other empirical fields (e.g. in the economic growth or income distribution literatures), it is worthwhile to briefly mention that several potentially useful extensions of the present model can be handled in a simple fashion.

Initially, we had expected that small market teams might have been very different from large market teams and, if so, transition probability matrices could have been very different for these two groups. In our empirical results, we include team dummy variables in X , which effectively allows us to investigate this issue. However, we also investigated a mixture of Normals ordered probit specification. That is, instead of the Normality assumption of (3.13) we had a mixture of more than one Normal. Tools for Bayesian inference in the mixtures of Normals probit model have been developed in Geweke and Keane (1999). Extending these tools to the case of ordered probit was straightforward. We had expected, if there were two distinctly different groups of teams, to find two elements in our mixture. However, we found evidence for only one. Since this was consistent with our finding using the simpler ordered probit model with team dummies, we do not include the mixtures of Normals ordered probit model here.

We investigate whether the transition probability matrix varies over time by including decade or year dummies in X . However, we also considered an alternative approach based on Chib (1998). The latter paper estimates structural breaks in a time series through Markov chain methods. If there are s structural breaks, then the data are divided into

$s + 1$ different regimes or states. Chib (1998) uses a Markov chain to model the movement between states. The structural break nature of the problem imposes restrictions on the transition probability matrix (i.e. in a structural break problem, once you have left a regime the probability of returning to it is zero). Chib (1998) develops MCMC methods to carry out Bayesian estimation and inference in such a model. We developed a model where Chib's state indicator was used to create regime specific dummies which were then entered as latent variables in X . In this manner, structural breaks could occur in the entire transition probability matrix. We developed MCMC methods which simply combined Chib (1998)'s MCMC methods with the ones outlined above. However, our failure to find a break (supported by evidence using the simpler ordered probit model with time dummies) lead us to omit this extension here.

In general, the modular nature of many MCMC algorithms used for Bayesian inference makes it trivial to combine different models in the aid of greater flexibility. For instance, in this section, we have developed methods involving an ordered probit model. These methods can be interpreted as one module in an MCMC algorithm for a mixtures of Normal ordered probit or a ordered qualitative choice model based on a Student-t latent variable or a nonparametric Dirichlet mixtures of Normals ordered probit, etc. Similarly, if components in X are latent data, then the Gibbs sampler given in (3.11)-(3.13) can be treated as one module in an MCMC algorithm, with the other module drawing the latent data conditional on S^* , γ and β . Chib (1998)'s model of structural of breaks provides us with one possible way that elements of X might be latent, but a myriad of others are straightforward to implement (e.g. elements of X could be based on a state space model allowing for gradual change in the transition probability over time or a hierarchical prior introduced which has team-specific effects being treated as latent draws from an underlying distribution, etc.). For our application, none of these extensions are necessary. However, in other applications, they might be of interest. Hence, it is worthwhile stressing the ease of developing MCMC methods for Bayesian inference in a variety of extensions of the present model.

4 Empirical Results

4.1 Preliminaries

Given that the economic success of Major League baseball likely depends to a degree on competitive balance, we are interested in understanding how the distribution of winning percentage evolves over time. The transition probability matrix and the year-by-year standard deviation reflect different aspects of competitive balance. The standard deviation is decreasing over time (see Figure 1), which indicates increasing competitive balance in that the difference between strong and weak teams is decreasing. The transition probability matrix looks at a different aspect of competitive balance: the degree of mobility of teams within the distribution. A league with a great deal of churning, where poor teams evolve into good teams and good teams evolve into poor over a relatively short time span is likely to be more economically successful than one where the same set of teams wins most the time. The year-by-year standard deviation does not necessarily tell us anything about this, but the transition probability matrix will. Hence, a primary focus of this section is on the transition probability matrix and whether it differs over time and across teams.

A point worth noting is that equality of the transition probability matrix across all teams might not be enough to imply perfect competitive balance, at least in the sense that your average baseball fan would define competitive balance. To see why, consider the case where all teams face the same transition probability matrix at all times. If the transition probability matrix exhibits a great deal of persistence, then this may be bad for competitive balance in that the same teams dominate for very long periods of time. For instance, if there are two types of teams, "Winners" ($S_{it} = 1$) and "Losers" ($S_{it} = 2$) and $\Pr[S_{it} = 1|S_{i,t-1} = 1] = 0.99$ and $\Pr[S_{it} = 2|S_{i,t-1} = 2] = 0.99$ then a team which is a winner/loser will, with high probability, stay a winner/loser for a very long time. This may be bad for competitive balance and league profitability in that "Losers" may lose fan support and even supporters of "Winners" may become jaded if the same small subset of

teams always win. Hence, the degree of persistence exhibited by the transition probability matrix/matrices is something we look at below.

Similarly, differences in persistence across teams is often of interest. For instance, a finding that $\Pr[S_{it} = 1|S_{i,t-1} = 1] = 0.99$ for some teams and $\Pr[S_{it} = 1|S_{i,t-1} = 1] = 0.25$ for other teams would be of interest as it would imply that some teams can build dynasties but others can only build up winning teams for short periods of time. In the sporting press, such a belief is commonly discussed. That is, the argument is made that big market teams (e.g. the Yankees) can build up a winning team and maintain it for a long time. Small market teams (e.g. Athletics) may be able to build up a winning team for a short period, but then lose their best players to free agency and quickly sink into mediocrity.

In sum, we would argue that many of the issues surrounding competitive balance in baseball can be revealed by looking at transition probability matrices, examining its persistence properties, and seeing if they evolve over time or differ across teams. We acknowledge that "competitive balance" may involve more issues than the ones discussed here. Nevertheless, we hope that the results presented here provide a big part of the story.

In order to avoid making distributional assumptions, we have followed the tradition of the Twin Peaks literature and work with a discretized distribution into several bins. For most of the time span of our data (i.e. from 1901-1960), the Major Leagues comprised only 16 teams. This limits the number of bins it is sensible to use. The empirical results presented here use four bins with $b_0 = 0, b_1 = .440, b_2 = .500, b_3 = .560, b_4 = 1$. Over the entire century, these numbers are the $0^{th}, 25^{th}, 50^{th}, 75^{th}$ and 100^{th} percentiles (see Section 2). Furthermore, these choices will undoubtedly make sense to students of baseball. For instance, a team with a winning percentage greater than .560 would typically have been contending for first place in a league or division. Fans of teams playing less than .440 baseball would consider their teams to have had very poor years. Hence, we argue that our bin choices loosely classify teams into contenders, good teams, weak teams and very weak teams, respectively. Preliminary data analysis suggested we did not have enough data to

work with 5 bins, but 3 bins was not enough to accurately classify the teams.

4.2 The Ordered Probit Model

Given our interest in the transition probability matrix, all of the specifications below have the first four explanatory variables being $d_{hi,t} = 1$ if $S_{i,t-1} = h$ and equals zero otherwise, for $h = 1, \dots, 4$. The simplest model we will entertain thus has $x_{it} = (d_{1it}, \dots, d_{4it})'$. Since these explanatory variables enter all models, use of a noninformative prior on their coefficients will not cause problems with the interpretation of Bayes factors,³ and, hence, we set the top left-hand 4×4 component of $\underline{B}^{-1} = 0_{4 \times 4}$ in all cases. Before considering more complicated specifications, involving additional explanatory variables, we present some results for this simplest model. We do this since it is most directly comparable to the standard Markov model used in the Twin Peaks literature. In the previous section, we showed that our model with $x_{it} = (d_{1it}, \dots, d_{4it})'$ has fewer free parameters than a model which leaves the transition probability matrix completely unrestricted (apart from the restriction that probabilities sum to one), but that the restrictions implied by our specification were likely harmless. To provide evidence for this contention, which means we do not need to work with the more flexible ordered probit specification given in (3.7), Table 3 presents the posterior mean and standard deviation of the transition probabilities for our model (i.e. the jh^{th} element of this table is the posterior mean and standard deviation of $\Pr[S_{it} = j | S_{i,t-1} = h]$). Table 4 presents maximum likelihood estimates for a standard Markov model. To be precise, if we define $n_{hj} = \sum_{i=1}^N \sum_{t=\underline{L}_i+1}^T 1(S_{it} = j) 1(S_{i,t-1} = h)$ to be the number of switches from bin h to bin j and $n_h = \sum_{i=1}^N \sum_{t=\underline{L}_i}^{T-1} 1(S_{i,t-1} = h)$ then the maximum likelihood estimate of $\Pr[S_{it} = j | S_{i,t-1} = h]$ is $\frac{n_{hj}}{n_h}$.

³Bartlett's paradox is the name given to the problems associated with interpreting Bayes factors when using an improper prior on parameters which are restricted under M_0 (see Poirier, 1995, page 390). It is common (see, e.g., Kass and Raftery, 1995, section 5.3), however, to use non-informative priors over parameters which are common to both M_0 and M_1 and we adopt this practice in this paper.

| | $S_{i,t-1} = 1$ | $S_{i,t-1} = 2$ | $S_{i,t-1} = 3$ | $S_{i,t-1} = 4$ |
|---------------|------------------|------------------|------------------|------------------|
| $S_{i,t} = 1$ | 0.537 (0.023) | 0.247 (0.019) | 0.139 (0.015) | 0.046 (0.008) |
| $S_{i,t} = 2$ | 0.267 (0.017) | 0.289 (0.017) | 0.239 (0.014) | 0.135 (0.010) |
| $S_{i,t} = 3$ | 0.154 (0.011) | 0.301 (0.013) | 0.342 (0.014) | 0.312 (0.015) |
| $S_{i,t} = 4$ | 0.039 (0.007) | 0.163 (0.016) | 0.281 (0.019) | 0.506 (0.022) |

| | $S_{i,t-1} = 1$ | $S_{i,t-1} = 2$ | $S_{i,t-1} = 3$ | $S_{i,t-1} = 4$ |
|---------------|-----------------|-----------------|-----------------|-----------------|
| $S_{i,t} = 1$ | 0.549 | 0.238 | 0.157 | 0.033 |
| $S_{i,t} = 2$ | 0.251 | 0.291 | 0.238 | 0.167 |
| $S_{i,t} = 3$ | 0.137 | 0.306 | 0.353 | 0.289 |
| $S_{i,t} = 4$ | 0.037 | 0.178 | 0.259 | 0.514 |

The point estimates in Tables 3 and 4 are very similar to one another and, in almost every case, the posterior mean is well within one posterior standard deviation of the maximum likelihood estimate. This indicates that the restrictions imposed by our ordered probit specification are not at odds with the data.⁴ We stress that the transition probability matrices in Tables 3 and 4 are both based on the assumption that $\Pr[S_{it} = j | S_{i,t-1} = h]$ is the same for every team and every year, assumptions which we relax below. However, an examination of these tables provides a positive view of competitive balance in baseball. For instance, the upper-left hand corner element of Table 4 says that, of all the teams which were very weak (i.e. winning percentage less than .440), only 54.9% of them were also very weak the following year. 3.7% of these very weak teams actually became contenders

⁴We do not present formal measures of model fit in this paper. However, the similarity between Tables 3 and 4 provides compelling informal evidence that even this, our most simple model, is fitting the data well.

(winning percentage greater than .560) the following year. At the top end of the league there is also considerable churning, with only 51.4% of the contenders in one year remaining contenders the next, and 3.3% of the contenders in one year becoming very weak teams the next. Hence, the probability of persistence in any bin is low.

Of course, the patterns noted in Tables 3 and 4 does not necessarily mean all teams are equally competitive. It is possible that a group of big market teams are dominating the top two categories with occasional and short-lived drops into the bottom two categories and small market teams dominate the lower two categories. Furthermore, it is possible that Tables 3 and 4 hide a worsening competitive balance over time. We can investigate these issues by augmenting our list of explanatory variables. In particular, we set $x_{it} = (d_{1it}, \dots, d_{4it}, td_{1it}, \dots, td_{29,it}, dd_{1it}, \dots, dd_{9it})'$, where td_{hit} are team dummies for 29 teams with the Yankees being the omitted team, and dd_{jit} are decade dummies for 1911-1920, 1921-1930, ..., 1991-2000, with 1901-1910 being the omitted dummy.

Bayesian results were calculated using both a completely noninformative prior with $\underline{B}^{-1} = 0_{k \times k}$ where $k = 42$, and a partially informative prior with $\underline{b} = 0_k$ and the bottom right-hand 38×38 component of $\underline{B} = I_{38}$. Note that this latter prior is completely non-informative for the coefficients on $(d_{1it}, \dots, d_{4it})$. The coefficients on the team and decade dummies are, *a priori*, independent of one another, each Normally distributed with prior mean 0 and prior standard deviation 1.0. To interpret our prior, remember that the ordered probit model expresses choice probabilities in terms of interval probabilities for the Normal with standard deviation 1.0 (see the discussion around equation 3.6) and, since all our explanatory variables are dummies, β_h for $h = 5, \dots, k$ will simply shift the mean of the Normal distribution, leaving the integration regions unchanged (i.e. γ does not depend on the explanatory variables). Our prior allocates approximately 95% its prior weight to the interval $[-2, 2]$. Hence, our prior allows for mean shifts in the underlying Normal of up to two standard deviations. Given that the integration regions are unaffected by the explanatory variables (see equation 3.6), a two standard deviation shift in the mean of a

Normal allows for a quite large change in the choice probabilities. Hence our prior, although informative, is not dogmatic and allows for substantial variation over teams and decades in the transition probability matrix. To aid in interpretation of results, note that negative values of β_h , by shifting the mean of the Normal to the left, lower the probability of being in the good bins relative to the bad ones (e.g. decreases the chance of becoming a contender and increases the chance of becoming a very weak team).

We do not present detailed results for this case with $k = 42$ since, somewhat to our surprise, the coefficients on all of the decade dummies were completely insignificant. The Bayes factor in favor of the model with $\beta_{34} = \dots = \beta_{42} = 0$ imposed was 1.59×10^6 and the largest of the individual t-ratios (i.e. posterior mean divided by posterior standard deviation) was 0.65. Similar findings were obtained when we used 99 year-specific dummies or a single dummy for 1901-1950. Hence, despite the fact that the standard deviation of the team performance distribution has been dropping over time (see Figure 1), the manner in which the distribution evolves over time has not changed in a statistically significant manner.

In light of this strong finding, we drop the decade dummies from the list of explanatory variables and proceed with $x_{it} = (d_{1it}, \dots, d_{4it}, td_{1it}, \dots, td_{29,it})'$. Table 5 presents posterior results for this model. It contains posterior means and standard deviations of all elements of β under both the noninformative and informative priors. For the informative prior we also present the Bayes factor in favor of the hypothesis that $\beta_h = 0$ for $h = 1, \dots, 33$. The last column, labelled "Number of Contenders", requires explanation. Once we include team dummies, each team will have a different transition probability matrix. Space considerations preclude presenting posterior results for 30 such matrices. Accordingly, the last column involves a single function of the transition probability matrix which might be of interest to baseball fans: the number of years in which each team will be a contender in the upcoming century. To be precise, at each pass through the Gibbs sampler, we obtain the transition probability matrix for every team and use this to simulate 100 years of team

performance, counting the number of years that each team is a contender (i.e. in bin 4). By suitably averaging across Gibbs draws, we can obtain the posterior mean and standard deviation of the number of years our model predicts each team will be a contender. We start all teams in bin 1.

| Table 5: Posterior Results for Ordered Probit Model with Team Dummies (st. dev. in parentheses) | | | | |
|--|-------------------------|----------------------|-------------------------------|-------------------------|
| Explanatory Variable | Noninformative Prior | Informative Prior | Bayes Factor $\beta_h = 0$ | Number of Contenders |
| Dummy for: | Posterior Mean | Posterior Mean | | |
| $S_{i,t-1} = 1$ | 0.521 (0.137) | 0.338 (0.117) | 0.148 | n.a. |
| $S_{i,t-1} = 2$ | 1.244 (0.138) | 1.061 (0.118) | 0.000 | n.a. |
| $S_{i,t-1} = 3$ | 1.626 (0.140) | 1.440 (0.120) | 0.000 | n.a. |
| $S_{i,t-1} = 4$ | 2.177 (0.138) | 1.997 (0.122) | 0.000 | n.a. |
| Angels | -0.675 (0.210) | -0.490 (0.196) | 0.093 | 17.80 (8.90) |
| D'Backs | 0.509 (0.784) | 0.399 (0.606) | 1.323 | 22.35 (9.83) |
| Braves | -0.688 (0.167) | -0.517 (0.150) | 0.001 | 14.31 (6.76) |
| Orioles | -0.657 (0.167) | -0.485 (0.151) | 0.002 | 15.00 (6.92) |
| Red Sox | -0.435 (0.163) | -0.273 (0.146) | 0.440 | 18.99 (8.05) |
| White Sox | -0.483 (0.163) | -0.316 (0.146) | 0.166 | 18.34 (7.72) |
| Cubs | -0.546 (0.163) | -0.377 (0.148) | 0.040 | 17.31 (7.45) |
| Reds | -0.496 (0.163) | -0.325 (0.147) | 0.134 | 18.24 (7.72) |
| Indians | -0.406 (0.162) | -0.238 (0.146) | 0.809 | 19.32 (8.10) |
| Rockies | -0.412 (0.409) | -0.214 (0.373) | 2.259 | 17.31 (9.62) |
| Tigers | -0.448 (0.162) | -0.283 (0.146) | 0.326 | 18.86 (7.94) |
| Marlins | -0.717 (0.425) | -0.474 (0.388) | 1.193 | 12.20 (9.79) |

| Table 5 (continued) | | | | |
|----------------------------|-------------------|-------------------|-------------------------------|-------------------------|
| Dummy for: | Posterior Mean | Posterior Mean | Bayes Factor $\beta_h = 0$ | Number of Contenders |
| Astros | -0.545 (0.213) | -0.366 (0.195) | 0.562 | 16.93 (7.87) |
| Royals | -0.472 (0.225) | -0.302 (0.210) | 1.372 | 17.99 (8.10) |
| Dodgers | -0.384 (0.163) | -0.220 (0.147) | 1.140 | 19.51 (8.14) |
| Brewers | -0.661 (0.231) | -0.474 (0.216) | 0.251 | 14.31 (7.87) |
| Twins | -0.761 (0.166) | -0.585 (0.149) | 0.000 | 12.08 (6.39) |
| Expos | -0.609 (0.228) | -0.427 (0.211) | 0.385 | 15.48 (7.87) |
| Mets | -0.640 (0.219) | -0.457 (0.205) | 0.222 | 14.87 (7.73) |
| Athletics | -0.640 (0.166) | -0.467 (0.149) | 0.004 | 15.44 (7.03) |
| Phillies | -0.876 (0.167) | -0.697 (0.151) | 0.000 | 8.51 (5.73) |
| Pirates | -0.415 (0.164) | -0.251 (0.148) | 0.682 | 19.27 (8.06) |
| Padres | -0.805 (0.234) | -0.610 (0.219) | 0.041 | 10.61 (7.49) |
| Giants | -0.221 (0.163) | -0.062 (0.147) | 5.485 | 20.88 (8.63) |
| Cardinals | -0.410 (0.164) | -0.245 (0.147) | 0.741 | 19.28 (8.07) |
| Mariners | -0.818 (0.261) | -0.612 (0.243) | 0.140 | 10.28 (7.87) |
| D'Rays | -62.96 (52.97) | -1.005 (0.729) | 0.550 | 0.09 (1.40) |
| Rangers | -0.818 (0.214) | -0.630 (0.199) | 0.009 | 10.26 (7.13) |
| Blue Jays | -0.461 (0.220) | -0.285 (0.241) | 1.867 | 17.84 (8.35) |

Since the Yankees is the omitted dummy, our prediction for "Number of Contenders" for this team is not included in Table 5. It turns out to have a posterior mean of 53.38 and a posterior standard deviation of 9.59.

The strong message of Table 5 is one of Yankees exceptionalism. With the exception of the Diamondbacks (which only existed for 1998-2000), the point estimates for the coefficients on the team dummies are all negative. Under the noninformative prior all these negative coefficients are at least one posterior standard deviation from zero and 24 out of 29 are more than two standard deviations from zero. Furthermore, with the exception of the Giants, the few teams with negative coefficients only one standard deviation from zero are all recent expansion teams and have relatively few data points. The informative prior tells a similar story (and increases our confidence that our findings are not largely due to prior information). The Bayes factor in favor of the hypothesis that $\beta_5 = \dots = \beta_{33} = 0$ is 0 (to four decimal places) and the individual Bayes factors in favor of $\beta_h = 0$ for $h = 5, \dots, 33$ are virtually all less than one. We stress that this strong evidence for statistical significance of team dummies, both jointly and (in most cases) individually, comes despite our use of a relatively flat priors.⁵

This finding of Yankees exceptionalism, which we will refer to as the "Damn Yankees" finding, is reinforced in Tables 6 and 7 which present the transition probability matrices for the Yankees and a randomly chosen team, the Orioles.

⁵Relatively flat priors under M_1 , by putting considerable prior weight in regions of the parameter space where the likelihood function is negligible, tend to yield support for the restrictions imposed under M_0 . In other words, Bayes factors with relatively flat priors have a strong reward for parsimony built in.

| Table 6: Posterior Mean of Transition Probability Matrix for Yankees (st. dev. in parentheses) | | | | |
|---|------------------|------------------|------------------|------------------|
| | $S_{i,t-1} = 1$ | $S_{i,t-1} = 2$ | $S_{i,t-1} = 3$ | $S_{i,t-1} = 4$ |
| $S_{i,t} = 1$ | 0.368 (0.044) | 0.146 (0.027) | 0.076 (0.017) | 0.024 (0.007) |
| $S_{i,t} = 2$ | 0.306 (0.021) | 0.249 (0.023) | 0.183 (0.022) | 0.092 (0.015) |
| $S_{i,t} = 3$ | 0.234 (0.023) | 0.336 (0.014) | 0.334 (0.015) | 0.261 (0.022) |
| $S_{i,t} = 4$ | 0.091 (0.020) | 0.269 (0.038) | 0.406 (0.043) | 0.624 (0.040) |

| Table 7: Posterior Mean of Transition Probability Matrix for Orioles (st. dev. in parentheses) | | | | |
|---|------------------|------------------|------------------|------------------|
| | $S_{i,t-1} = 1$ | $S_{i,t-1} = 2$ | $S_{i,t-1} = 3$ | $S_{i,t-1} = 4$ |
| $S_{i,t} = 1$ | 0.560 (0.069) | 0.287 (0.060) | 0.175 (0.046) | 0.069 (0.025) |
| $S_{i,t} = 2$ | 0.264 (0.030) | 0.299 (0.022) | 0.264 (0.029) | 0.172 (0.032) |
| $S_{i,t} = 3$ | 0.140 (0.032) | 0.276 (0.030) | 0.326 (0.020) | 0.327 (0.020) |
| $S_{i,t} = 4$ | 0.036 (0.015) | 0.137 (0.039) | 0.236 (0.054) | 0.432 (0.069) |

Clearly, the Yankees' performance has been much better than the Orioles'. The point estimates indicate that, if the Yankees are a contender one year, there is a 62.4% they will be a contender the next. For the Orioles, the relevant figure is 43.2%. When the Yankees have a disastrous year (i.e. winning percentage less than 0.440), there is a 63.1% chance they will move up at least one bin the next and a 32.5% chance they will actually improve to over .500. For the Orioles, the comparable figures are 44.0% and 17.6%, respectively. In other words, relative to other teams, the Yankees have exhibited much more persistence in bin 4 (i.e. maintained dynasties longer) and much less in bin 1 (i.e. their lean periods have been much shorter).

The Damn Yankees finding becomes even clearer if we repeat all the previous empirical work with the Yankees omitted. Brevity precludes presenting all the results for this case. However, the bottom line finding is that, with the Yankees omitted, none of the team or decade dummies are significant and the simplest model with $x_{it} = (d_{1it}, \dots, d_{4it})'$ is preferred by the data. For instance, the Bayes factor in favor of the hypothesis that the team dummies are jointly zero (i.e. $\beta_5 = \dots = \beta_{32} = 0$ where Blue Jays are now the omitted team) is 8.96×10^{11} and the individual Bayes factors in favor of $\beta_h = 0$ for $h = 5, \dots, 32$ are virtually all greater than one. With the noninformative prior, no team dummy has posterior mean more than two standard deviations from zero and only two team dummies have posterior means greater than one standard deviation from zero. The latter two teams (which also have Bayes factors in favor of $\beta_h = 0$ being slightly less than one) are the Giants and Phillies. Thus, the strongest statistical support goes for a model where all teams have the same transition probability matrix and it is constant over time. There is some very weak (and statistically insignificant according to usual standards) evidence that the Giants have a slightly better (in that more probability is associated with higher numbered bins) and the Phillies slightly worse transition probability matrices than average.

With all Yankees data and the insignificant team dummies omitted, the transition probability matrix is given in Table 8.

| Table 8: Posterior Mean of Transition Probability Matrix with Yankees Omitted (st. dev. in parentheses) | | | | |
|--|------------------|------------------|------------------|------------------|
| | $S_{i,t-1} = 1$ | $S_{i,t-1} = 2$ | $S_{i,t-1} = 3$ | $S_{i,t-1} = 4$ |
| $S_{i,t} = 1$ | 0.543 (0.022) | 0.258 (0.019) | 0.145 (0.014) | 0.056 (0.012) |
| $S_{i,t} = 2$ | 0.260 (0.015) | 0.280 (0.016) | 0.233 (0.014) | 0.143 (0.012) |
| $S_{i,t} = 3$ | 0.157 (0.012) | 0.301 (0.014) | 0.344 (0.013) | 0.322 (0.014) |
| $S_{i,t} = 4$ | 0.040 (0.006) | 0.160 (0.014) | 0.278 (0.018) | 0.479 (0.023) |

Table 8 indicates that, once we omit the Yankees, there is equality among teams and a great degree of mobility within the distribution. It paints an idyllic picture of a league with a high degree of competitive balance except, of course, for those Damn Yankees.

5 Conclusions

In this paper, Bayesian techniques for modeling the evolution of distributions over time have been developed and applied to the distribution of team performance in Major League baseball in the twentieth century. We discussed how issues relating to competitive balance are quite important in several theoretical models of sports leagues and our econometric model allows for insights into competitive balance which cannot be obtained using regression-based methods.

Some of the main empirical findings include:

1. There is no evidence that competitive balance has worsened over time despite major changes such as the expansion of number of teams, free agency and growing wealth inequality. Indeed the year-by-year standard deviation in the distribution of winning percentage has tended to fall over time and the transition probability matrices seem to be unchanged over time.
2. The Yankees are an exceptional team with a transition probability matrix which is substantially different from the other teams.
3. With the Yankees data omitted, a model where the transition probability matrix is common to all teams and all years is strongly supported. This transition probability matrix indicates a high degree of team mobility within the distribution.
4. Hence, we have strong evidence that, except for the Yankees, baseball enjoys a very high degree of competitive balance.

6 References

Albert, J. and Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88, 669-679.

Chib, S., 1998. Estimation and comparison of multiple change point models. *Journal of Econometrics* 86, 221-241.

Durlauf, S. and Quah, D. 1999. The new empirics of economic growth. Chapter 4, 231-304 in *Handbook of Macroeconomics*, J. Taylor and M. Woodford (eds.), North Holland Elsevier Science.

Filardo, A. and Gordon, S., 1998. Business cycle durations. *Journal of Econometrics* 85, 99-123.

Flynn, M. and Gilbert, R., 2001. The analysis of professional sports leagues as joint ventures. *The Economic Journal* 111, F27-F46.

Geweke, J. and Keane, M., 1999. Mixtures of Normals probit models. Pages 49-78 in *Analysis of Panels and Limited Dependent Variables*, C. Hsiao, K. Lahiri, L.-F. Lee and M.H. Pesaran (eds.), Cambridge: Cambridge University Press.

Kass, R. and Raftery, A., 1995. Bayes factors. *Journal of the American Statistical Association* 90, 773-795.

Koop, G. and Poirier, D., 1993. Bayesian analysis of logit models using natural conjugate priors. *Journal of Econometrics* 56, 323-340.

Lahiri, K. and Gao, J., 2001. Bayesian analysis of nested logit models by Markov Chain Monte Carlo. SUNY- Albany, manuscript available at www.albany.edu/~klahiri/.

Maddala, G.S., 1983. *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge: Cambridge University Press.

McCulloch R., Polson N. and Rossi, P., 2000. A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal Of Econometrics* 99, 173-193.

McCulloch, R. and Rossi, P., 2000. Reply to Nobile. *Journal Of Econometrics* 99,

347-348.

McCulloch R., Rossi, P., 1994. An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics* 64, 207-240.

Nobile, A., 2000. Comment: Bayesian multinomial probit models with a normalization constraint. *Journal Of Econometrics* 99, 335-345.

Paap, R. and van Dijk, H., 1998. Distribution and mobility of wealth of nations. *European Economic Review* 42, 1269-1293.

Poirier, D., 1995. *Intermediate Statistics and Econometrics: A Comparative Approach*, Cambridge: The MIT Press.

Quah, D., 1993a. Empirical cross section dynamics in economic growth. *European Economic Review* 37, 426-434.

Quah, D., 1993b. Galton's fallacy and tests of the convergence hypothesis. *Scandinavian Journal of Economics* 95, 427-443.

Quah, D., 1996. Empirics for economic growth and convergence. *European Economic Review* 40, 1353-1375.

Quinn, K., Martin, A. and Whitford, A., 1999. Voter choice in multi-party democracies: A test of competing theories and models. *American Journal of Political Science* 43, 1231-1247.

Quirk, J. and El Hodiri, M., 1974. The economic theory of professional sports leagues. In *Government and the Sports Business*, R. Noll (ed.), Washington: Brookings.

Quirk, J. and Fort, R., 1992. *Pay Dirt: The Business of Professional Sports Teams*, Princeton: Princeton University Press.

Rottenberg, S., 1956. The baseball players labor market. *Journal of Political Economy*, 242-258.

Szymanski, S., 2001. Income inequality, competitive balance and the attractiveness of team sports: Some evidence and a natural experiment from English soccer. *The Economic Journal* 111, F69-F84.

Verdinelli, I. and Wasserman, L., 1995. Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *Journal of the American Statistical Association* 90, 614-618.

Vrooman, J., 2000. The economics of American sports leagues. *Scottish Journal of Political Economy* 47, 364-398.

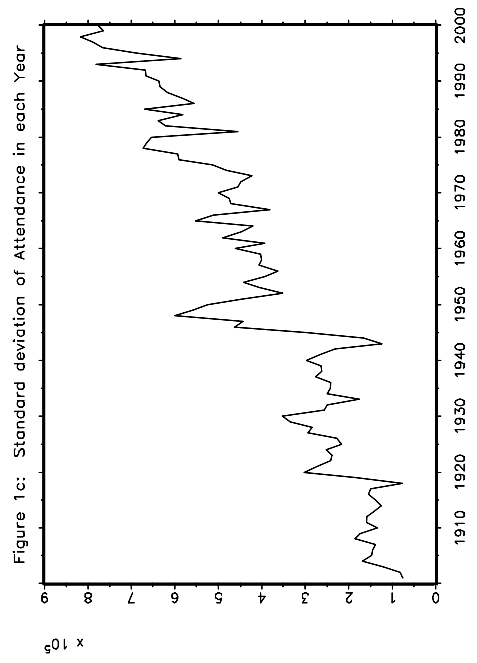
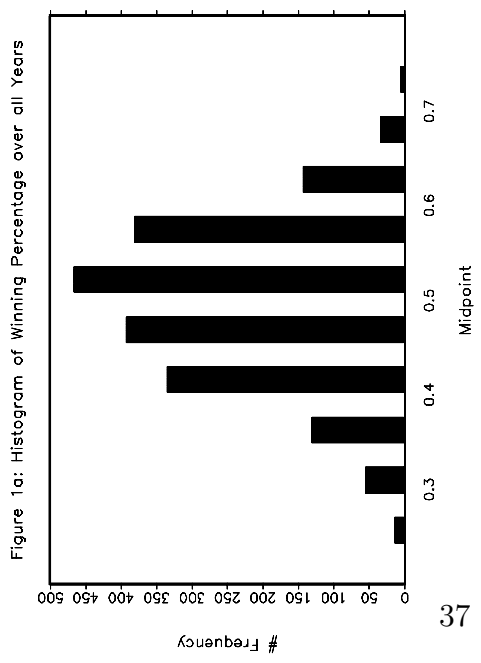
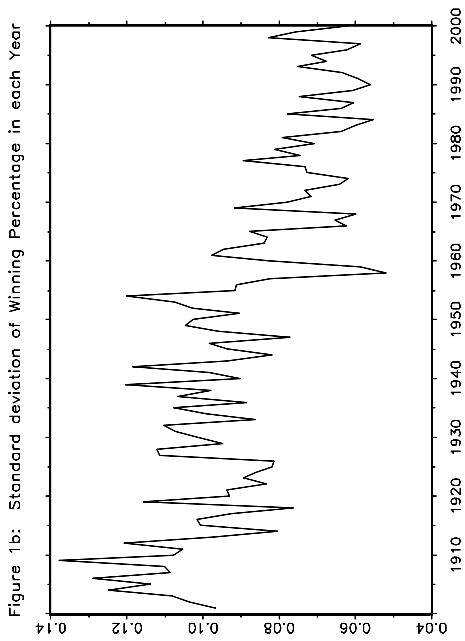


Figure 1: