Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Bridging Different Eras in Sports

Scott M. Berry, C. Shane Reese & Patrick D. Larkey

To cite this article: Scott M. Berry, C. Shane Reese & Patrick D. Larkey (1999) Bridging Different Eras in Sports, Journal of the American Statistical Association, 94:447, 661-676, DOI: 10.1080/01621459.1999.10474163

To link to this article: https://doi.org/10.1080/01621459.1999.10474163

Published online: 17 Feb 2012.

Bridging Different Eras in Sports

Scott M. BERRY, C. Shane REESE, and Patrick D. LARKEY

This article addresses the problem of comparing abilities of players from different eras in professional sports. We study National Hockey League players, professional golfers, and Major League Baseball players from the perspectives of home run hitting and hitting for average. Within each sport, the careers of the players overlap to some extent. This network of overlaps, or bridges, is used to compare players whose careers took place in different eras. The goal is not to judge players relative to their contemporaries, but rather to compare all players directly. Hence the model that we use is a statistical time machine. We use additive models to estimate the innate ability of players, the effects of aging on performance, and the relative difficulty of each year within a sport. We measure each of these effects separated from the others. We use hierarchical models to model the distribution of players and specify separate distributions for each decade, thus allowing the "talent pool" within each sport to change. We study the changing talent pool in each sport and address Gould's conjecture about the way in which populations change. Nonparametric aging functions allow us to estimate the league-wide average aging function. Hierarchical random curves allow for individuals to age differently from the average of athletes in that sport. We characterize players by their career profile rather than a one-number summary of their career.

KEY WORDS: Aging function; Bridge model; Hierarchical model; Population dynamics; Random curve.

1. INTRODUCTION

This article compares the performances of athletes from different eras in three sports: baseball, hockey, and golf. A goal is to construct a statistical time machine in which we estimate how an athlete from one era would perform in another era. For examples, we estimate how many home runs Babe Ruth would hit in modern baseball, how many points Wayne Gretzky would have scored in the tight-checking National Hockey League (NHL) of the 1950s, and how well Ben Hogan would do with the titanium drivers and extra-long golf balls of today's game.

Comparing players from different eras has long been pub fodder. The topic has been debated endlessly, generally to the conclusion that such comparisons are impossible. However, the data available in sports are well suited for such comparisons. In every sport there is a great deal of overlap in players' careers. Although a player that played in the early 1900s never played against contemporary players, they did play against players, who played against players, . . . , who played against contemporary players. This process forms a *bridge* from the early years of sport to the present that allows comparisons across eras.

A complication in making this bridge is that the overlapping of players' careers is confounded with the players' aging process; players in all sports tend to improve, peak, and then decline. To bridge the past to the present, the effects of aging on performance must be modeled. We use a nonparametric function to model these effects in each sport. An additional difficulty in modeling the effects of age on performance is that age does not have the same effect on all players. To handle such heterogeneity, we use random effects for each player's aging function, which allows for modeling players that deviate from the "standard" aging pattern. A desirable effect of using random curves is that each player is characterized by a career profile, rather than by a one-number summary. Player A may be better than player B when they are both 23 years old, and player A may be worse than player B when they are both 33 years old. Section 3.4 discusses the age effect model.

By modeling the effects of age on the performance of each individual, we can simultaneously model the difficulty of each year and the ability of each player. We use hierarchical models (see Draper et al. 1992) to estimate the innate ability of each player. To capture the changing pool of players in each sport, we use separate distributions for each decade. This allows us to study the changing distribution of players in each sport over time. We also model the effect that year (season) has on player performance. We find that, for example, in the last 40 years, improved equipment and course conditions in golf have decreased scoring by approximately 1 shot per 18 holes. This is above and beyond any improvement in the abilities of the players over time. The estimated innate ability of each player, and the changing evolution of each sport is discussed in Section 7.

Gould (1996) has hypothesized that the population of players in sport is continually improving. He claimed there is a limit to human ability—a wall that will never be crossed. There will always be players close to this wall, but as time passes and the population increases, more and more players will be close to this wall. He believes there are great players in all eras, but the mean players and lower end of the tail players in each era are closer to the "wall." By separating out the innate ability of each player, we study the dynamic nature of the population of players. Section

Scott M. Berry is Assistant Professor, Department of Statistics, Texas A&M University, College Station, TX 77843. C. Shane Reese is Technical Staff Member, Statistical Sciences, Los Alamos National Laboratory, Los Alamos, NM, 87545. Patrick D. Larkey is Professor, H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, Pittsburgh, PA 15213. The authors thank Wendy Reese and Tammy Berry for their assistance in collecting and typing in data, Sean Lahman for providing the baseball data, and Marino Parascenzo for his assistance in finding birth years for the golfers. The authors are also grateful for discussions with Jim Calvin, Ed Kambour, Don Berry, Hal Stern, Jay Kadane, Brad Carlin, Jim Albert, and Michael Schell. The authors thank the editor, the associate editor, and three referees for encouraging and helpful comments and suggestions.

Journal of the American Statistical Association
September 1999, Vol. 94, No. 447, Applications and Case Studies

8 describes our results regarding the population dynamics. We provide a discussion of Gould's claims as well.

We have four main goals:

1. To describe the effects of aging on performance in each of the sports, including the degree of heterogeneity among players. Looking at the unadjusted performance of players over their careers is confounded with the changing nature of the players and the changing structure of the sports. We separate out these factors to address the aging effects.
2. To describe the effects of playing in each year in each of the sports. We want to separate out the difficulty of playing in each era from the quality of players in that era. These effects may be due to rule changes, changes in the quality of the opponents, changes in the available (and legal) equipment, and the very nature of the sport.
3. To characterize the talent of each player, independent of the era or age of the player.
4. To characterize the changing structure of the population of players. In a sport involving one player playing against an objective measure with the same equipment that has always been used (e.g. throwing a shot put, lifting weights), it is clear that the quality of players is increasing. We want to know if that is true in these three professional sports.

Addressing any factor that affects performance requires addressing all such factors. If the league-wide performance is used as a measure of the difficulty of a particular year, then this is confounded with the players' ability in that year. In hockey, if an average of 3 goals are scored per game in 1950 and 4 goals scored per game in 1990, it is not clear whether scoring is easier or if the offensive players are more talented. Our aim is to separate out each effect, while accounting for the other effects.

We have found little research in this area. Riccio (1994) examined the aging pattern of golfer Tom Watson in his U.S. Open performances. Berry and Larkey (1998) compared the performance of golfers in major tournaments. Albert (1998) looked at the distribution of home runs by Mike Schmidt over his career. Schell (1998) ranked the greatest baseball players of all time on their ability to hit for average. He used a z-score method to account for the changing distribution of players and estimated the ballpark effects. He ignored the aging effects by requiring a minimum number of at bats to qualify for his method. Both of these effects are estimated separately without accounting for the other changing effects. Our goal is to construct a comprehensive model that makes the necessary adjustments simultaneously, rather than a series of clever adjustments.

The next section examines the measures of performance in each sport and the available data for each. Section 3 describes the models used and the key assumptions of each. Section 4 discusses the Markov chain Monte Carlo (MCMC) algorithms used. The algorithms are standard MCMC successive substitution algorithms. Section 5 looks at the goodness of fit for each of the models. To address the aging effects, Section 6 presents nonparametric aging functions. Random curves are used to allow for variation in aging across individuals. Section 7 discusses the results for each sport, including player aging profiles, top peak performers, and the changes over time for each sport. Section 8 discusses the population dynamics within each sport, and Section 9 discusses the results and possible extensions.

2. SPORTS SPECIFICS AND AVAILABLE DATA

In our study of hockey, we model the ability of NHL players to score points. In hockey, two teams battle continuously to shoot the puck into the other team's goal. Whoever shoots the puck into the goal receives credit for scoring a *goal*. If the puck was passed from teammates to the goal scorer, then up to the last two players to pass the puck on the play receive credit for an *assist*. A player receives credit for a *point* for either a goal or an assist. In hockey there are three categories of players: forwards, defensemen, and goalies. A main task of defensemen and goalies is to prevent the other team from scoring. A main task of forwards is to score goals. Therefore, we consider forwards only. We recorded the number of points in each season for the 1,136 forwards playing at least 100 games between 1948 and 1996. All hockey data are from Hollander (1997). We deleted all seasons played for players age 40 and older. There were very few such seasons, and thus the age function was not well defined greater than 40. Any conclusions in this article about these players is based strictly on their ability to score points, which is not necessarily reflective of their "value" to a hockey team. Some forwards are well known for their defensive abilities; thus their worth is not accurately measured by their point totals.

Considered among the most physically demanding of sports, hockey requires great physical endurance, strength, and coordination. As evidence of this, forwards rotate throughout the game, with three or four lines (sets of three forwards) playing in alternating shifts. In no other major sport do players participate such a small fraction of the time. We do not have data on which players were linemates. The NHL has undergone significant changes over the years. The league has expanded from 6 teams in 1948 to 30 teams in 1996. Recent years have brought a dramatic increase in the numbers of Eastern European and American players, as opposed to almost exclusively Canadians in the early years. Technological developments have made an impact in the NHL. The skates that players use today are vastly superior to those of 25 years ago. The sticks are stronger and curved, helping players control the puck better and shoot more accurately. The style of play has also changed. At different times in NHL history coaches have stressed offense or stressed defense.

In golf, it takes a long time for a player to reach his or her peak. Golf requires a great deal of talent, but it does not take the physical toll that hockey does. It seems reasonable to expect that the skills needed to play golf do not deteriorate as quickly in aging players as do speed and strength in hockey. Therefore, the playing careers of golfers are much longer. Technology is believed to have played an enormous role in golf. Advances in club and ball design have aided modern players. The conditions of courses today are far superior to conditions of 50 years ago: Modern professional golfers experience very few bad lies of the ball on the fairways of today's courses. The speed of the greens has increased over the years, which may increase scores, but this may be offset by a truer roll. The common perception is that technology has made the game easier.

We model the scoring ability of male professional golfers in the four major tournaments, considered the most important events of each golf season. We have individual round scores for every player in the Masters and U.S. Open from 1935–1997 and in the Open Championship (labeled the British Open by Americans) and the PGA of America Championship from 1961–1997. (The Masters and U.S. Open were not played in 1943–1945 because of World War II.) A major tournament comprises four rounds each of 18 holes of play. A "cut" occurs after the second round, and thus playing in a major generally consists of playing either two or four rounds. We found the birth years for 488 players who played at least 10 majors in the tournaments we are considering. We did not find the ages of 38 players who played at least 10 majors. The birth years for current players were found at various web sites (*pgatour.com;www.golfweb.com*). For older players, we consulted golf writer Marino Parascenzo. We had trouble finding the birth years for marginal players from past eras. This bias has consequences in our analysis of the population dynamics in Section 8.

Baseball is rich in data. We have data on every player (nonpitcher) who has batted in Major League Baseball (MLB) in the modern era (1901–1996). We have the year of birth and the home ballpark for each player during each season. The number of at bats, base hits, and home runs are recorded for each season. An official at bat is one in which the player reaches base safely from a base hit or makes an out. An at bat does not include a base on balls, sacrifice, or hit by pitch. (Interestingly, sacrifices were considered at bats before 1950 but not thereafter.) A player's batting average is the proportion of at bats in which he gets a base hit. We also model a player's home run average, which is the proportion of at bats a player hits a home run.

In terms of player aging, baseball is apparently between golf and hockey. Hand-eye coordination is crucial, but the game does not take an onerous physical toll on players. A common perception is that careers in baseball are longer than in hockey, but shorter than in golf. Baseball prides itself on being a traditional game, and there have been relatively few changes in the rules during the twentieth century. Some changes include lowering the mound, reducing the size of the strike zone, and modifications to the ball. The first 20 years of this century were labeled the "deadball era." The most obvious change in the population of players came in the late 1940s, when African-Americans were first allowed to play in the major leagues. MLB has historically been played mainly by U.S. athletes, although Latin Americans have had an increasing influence over the last 40 years.

3. MODELS

In this section we present the bridging model, with details of the model for each sport. To compare players from different eras, we select the most recent season in the dataset as the *benchmark* season. All evaluations of players are relative to the benchmark season. The ability of every player that played during the benchmark season can be estimated by their performance in that season. In the home run example, this includes current sluggers like Mark McGwire (1987–present), Ken Griffey, Jr. (1989–present), and Mike Piazza (1992–present). The ability of players whose careers overlapped with the current players can be estimated by comparing their performances to the current players' performances in common years. In the home run example, this includes comparing players like Reggie Jackson (1967–1987), Mike Schmidt (1972–1989), and Dale Murphy (1976–1992) to McGwire, Griffey, and Piazza. The careers of Jackson, Schmidt, and Murphy overlapped with the careers of players who preceded them, such as Mickey Mantle (1951–1968), Harmon Killebrew (1954–1975), and Hank Aaron (1954–1976). The abilities of Mantle, Killebrew, and Aaron can be estimated from their performances relative to Jackson, Schmidt, and Murphy in their common years. The network of thousands of players with staggered careers extends back to the beginning of baseball. All three sports considered in this article have similar networks.

We estimate a league-wide age effect by comparing each player's performance as they age with their estimated ability. The difficulty of a particular season can be estimated by comparing each player's performance in the season with their estimated ability and estimated age effect during that season. We can estimate other effects, such as ball park and individual rounds in golf, in an analogous fashion. This explanation is an iterative one, but the estimates of these effects are produced simultaneously.

There are two critical assumptions for each model used in this article. The first is that outcomes across events (games, rounds, and at bats) are independent. A success or failure in one trial does not affect the results of other trials. One example of dependence between trials is the "hot-hand" effect: Success breeds success, and failure breeds failure. This topic has received a great deal of attention in the statistics literature. We have found no conclusive evidence of a hot-hand effect. (For interesting studies of the hot-hand, see Albert 1993, Albright 1993, Jackson and Mosurski 1997, Larkey, Smith, and Kadane 1989, Stern 1995, Stern and Morris 1993, and Tversky and Gilovich 1989a,b.) We do not take up this issue here, but we do believe that golf is the most likely sport to have a hot-hand effect (and we have found no analysis of the hot-hand effect in golf).

All of the models used are additive. Therefore, the second critical assumption is that there are no interactions. An interaction in this context would mean that the performances of different players are affected differently by a predictor. For example, if player A is more successful in the modern game than player B, then had they both played 50 years ago player A would still have been better than player B.

We address the question of interactions in the discussion section.

We use the same parameters across sports to represent player and year effects. When necessary, superscripts h, g, a, and r are used to represent hockey, golf, and batting averages and home runs in baseball.

3.1 Hockey

For the hockey data, we have $k = 1{,}136$ players. The number of seasons played by player i is n_i, and the age of player i in his jth season is a_{ij}. The year in which player i played his jth season is y_{ij}, the number of points scored in that season is x_{ij}, and the games played is g_{ij}. Counting the number of points for a player in a game is counting rare events in time, which we model using the Poisson distribution.

Per game scoring for a season is difficult to obtain. To address the appropriateness of the Poisson distribution for one player, we collected data on the number of points scored in each game for Wayne Gretzky in the 1995–1996 season, as shown in Table 1. The Poisson appears to be a reasonable match for the points scored per game (the chi-squared goodness-of-fit test statistic is 5.72, with a p value of .22).

We assume that the points scored in a game are independent of those scored in other games, conditionally on the player and year, and that the points scored by one player are independent of the points scored by other players. The model is

$$x_{ij} \sim \text{Poisson}(\lambda_{ij} g_{ij}), \qquad i = 1, \ldots, k; \qquad j = 1, \ldots, n_i,$$

where the x_{ij} are independent conditional on the λ_{ij}'s. Assume that

$$\log(\lambda_{ij}) = \theta_i + \delta_{y_{ij}} + f_i(a_{ij}).$$

In this log-linear model, θ_i represents the player-specific ability; that is, $\exp(\theta_i)$ is the average number of points per game for player i when he is playing at his peak age ($f_i = 0$) in 1996 ($\delta_{1996} = 0$). There are 49 δ_l's, one for each year in our study. They represent the difficulty of year l relative to 1996. Therefore, we constrain $\delta_{1996} \equiv 0$. We refer to 1996 as the benchmark year. The function f_i represents the aging effects for player i. We use a random curve to model the aging, as discussed in Section 3.4. The function f_i is restricted to be 0 for some age a (player i's peak age).

A conditionally independent hierarchical model is used for the θ's. To allow for the distribution of players to change over time, we model a separate distribution for the θ's for each decade. Let d_i be the decade in which player i was born. In the hockey example, the first decade is 1910–1919, the second decade is 1920–1929, and the last decade, the seventh, is 1970–1979. The model is

$$\theta_i \sim \text{N}(\mu_\theta(d_i), \sigma_\theta^2(d_i)), \qquad i = 1, \ldots, 1{,}136,$$

where $\text{N}(\mu, \sigma^2)$ refers to a normal distribution with a mean of μ and a variance of σ^2. The hyperparameters have the distributions

$$\mu_\theta(d_i) \sim \text{N}(m, s^2), \qquad d_i = 1, \ldots, 7$$

and

$$\sigma_\theta^2(d_i) \sim \text{IG}(a, b), \qquad d_i = 1, \ldots, 7,$$

where $\text{IG}(a, b)$ refers to an inverse gamma distribution with mean $1/b(a-1)$ and variance $1/b^2(a-1)^2(a-2)$. The θ_i are independent conditional on the μ_θ's and σ_θ^2's. The δ_l's are independent with prior distributions

$$\delta_l \sim \text{N}(0, \tau^2), \qquad l = 1948, \ldots, 1995.$$

The average forward scores approximately 40 points in a season, or approximately .5 points per game. Thus we set $m = \log(.5)$, and allow for substantial variability around this number by setting $s = .5$. For the distribution of σ_θ^2, we set $a = 3$ and $b = 3$. This distribution has mean .167 and standard deviation .167. We chose $\tau = 1$. We specified prior distributions that we thought were reasonable and open minded. This prior represents the notion that σ_θ^2 is not huge, but is flexible enough so that the posterior is controlled by the data. We find little difference in the results for the priors that we considered reasonable.

3.2 Golf

The golf study involves $k = 488$ players, with player i playing n_i rounds of golf (a round consisting of 18 holes). For the jth round of player i, the year in which the round is played is y_{ij}, the score of the round is x_{ij}, the age of the player is a_{ij}, and the round number in year y_{ij} is r_{ij}. The round number ranges from 1 to 16 in any particular year, corresponding to the chronological order.

We adopt the following model for golf scores:

$$x_{ij} \sim \text{N}(\beta_{ij}, \sigma_x^2),$$

where the x_{ij} are independent given the β_{ij}'s and σ_x^2. Assume that

$$\beta_{ij} = \theta_i + \delta_{y_{ij}} + \gamma_{y_{ij}, r_{ij}} + f_i(a_{ij}).$$

Parameter θ_i represents the mean score for player i when that player is at his peak ($f_i = 0$), playing a round of average difficulty in 1997 ($\delta = 0$ and $\gamma = 0$). The benchmark year is 1997; thus $\delta_{1997} \equiv 0$, and each δ_l represents the difficulty of that year's major tournaments relative to 1997. There is variation in the difficulty of rounds within a year. Some courses are more difficult than others; the course setup can be relatively difficult or relatively easy, and the weather plays a major role in scoring. The γ's represent the

Table 1. The Points Scored in Each of Wayne Gretzky's 83 Games in the 1995–96 Season

Points	Gretzky	Poisson
0	23/83 = .28	.31
1	34/83 = .41	.36
2	18/83 = .22	.21
3	4/83 = .05	.08
4	4/83 = .05	.02
5	0/83 = 0	.006

NOTE: For each point total the probability of that occurrence, assuming a Poisson distribution with a mean of 1.18, is shown in the second column.

difficulty of rounds within a year. Thus $\gamma_{u,v}$ is the mean difference, in strokes, for round v from the average round in year u. To preserve identifiability, and thus interpretability, we restrict

$$\sum_{v=1}^{16} \gamma_{u,v} \equiv 0.$$

The aging function f_i is discussed in Section 3.4. A decade-specific hierarchical model is used for the θ's. Let d_i be the decade in which a golfer was born. There are seven decades: 1900–1909, 1910–1919, ..., 1960+. Only three players in the dataset were born in the 1970s, so they were combined into the 1960s. Let the θ_i's be independent conditional on the μ_θ's and σ_θ^2's and be distributed as

$$\theta_i \sim \mathrm{N}(\mu_\theta(d_i), \sigma_\theta^2(d_i)), \qquad i = 1, \ldots, 488,$$

where

$$\mu_\theta(d_i) \sim \mathrm{N}(m, s^2), \qquad d_i = 1, \ldots, 7$$

and

$$\sigma_\theta^2(d_i) \sim \mathrm{IG}(a, b), \qquad d_i = 1, \ldots, 7.$$

The δ_l's are independent with prior distributions

$$\delta_l \sim \mathrm{N}(0, \tau^2), \qquad l = 1935, \ldots, 1996$$

and the $\gamma_{u,v}$'s are independent with the priors

$$\gamma_{u,v} \sim \mathrm{N}(0, \phi^2).$$

We specify the hyperparameters as $m = 73, s = 3, a = 3, b = 3, \tau = 3$, and $\phi = 3$. As in the hockey study, here the results from priors similar to this one are virtually identical. The distribution of golf scores has been discussed by Mosteller and Youtz (1993). They modeled golf scores as 63 plus a Poisson random variable. Their resulting distribution looked virtually normal, with a slight right skew. They developed their model based on combining the scores of all professional golfers. Scheid (1990) studied the scores of 3,000 amateur golfers and concluded that the normal fits well, except for a slightly heavier right tail. There are some theoretical reasons why normality is attractive. Each round score is the sum of 18 individual hole scores. The distribution of scores on each hole is somewhat right-skewed, because scores are positive and unlimited from above. A score of 2, 3, or 4 over par on one hole is not all that rare, whereas 2, 3, or 4 under par on a hole is extremely rare, if not impossible. A residual normal probability plot shown in Section 5 demonstrates the slight right-skewed nature of golf scores. We checked models with a slight right skew, and found the results to be virtually identical (not shown). The only resulting difference that we noticed was in predicting individual scores (in which we are not directly interested). Because of its computational ease and reasonable fit, we adopt the normality assumption.

3.3 Baseball

The baseball studies involve $k = 7{,}031$ players, with player i playing in n_i seasons. For player i in his jth season, x_{ij} is the number of hits, h_{ij} is the number of home runs, m_{ij} is the number of at bats, a_{ij} is the player's age, y_{ij} is the year of play, and t_{ij} is the player's home ballpark. (Players play half their games in their home ballpark and the other half at various ballparks of the other teams.)

We model at bats as independent Bernoulli trials, with the probability of success for player i in his jth year equal to π_{ij}. We study both hits and home runs as successes; therefore, we label π_{ij}^a and π_{ij}^r as the probability of getting a hit and of hitting a home run. Thus

$$x_{ij} \sim \text{binomial}(m_{ij}, \pi_{ij}^a),$$

where

$$\log\left(\frac{\pi_{ij}^a}{1 - \pi_{ij}^a}\right) = \theta_i^a + \delta_{y_{ij}}^a + \xi_{t_{ij}}^a + f_i^a(a_{ij}).$$

For the baseball home run study, we use a similar model,

$$h_{ij} \sim \text{binomial}(m_{ij}, \pi_{ij}^r),$$

where

$$\log\left(\frac{\pi_{ij}^r}{1 - \pi_{ij}^r}\right) = \theta_i^r + \delta_{y_{ij}}^r + \xi_{t_{ij}}^r + f_i^r(a_{ij}).$$

The δ parameters are indicator functions for seasons, and the ξ parameters are indicator functions for home ballparks. We include the ξ parameters to account for the possibility that certain stadiums are "hitters'" parks and others are "pitchers'" parks. The aging function f_i is discussed in the following subsection.

Let d_i be the decade in which player i was born. There are 12 decades for the baseball players: 1860–1869, ..., 1970+. A decade-specific conditionally independent hierarchical model is used:

$$\theta_i^a \sim \mathrm{N}(\mu_\theta^a(d_i), (\sigma_\theta^a)^2(d_i)), \qquad i = 1, \ldots, 7{,}031,$$

where the θ_i^a's are independent conditional on the μ_θ^a's and $(\sigma_\theta^a)^2$'s. Assume that

$$\mu_\theta^a(d_i) \sim \mathrm{N}(m^a, (s^a)^2), \qquad d_i = 1, \ldots, 12$$

and

$$(\sigma_\theta^a)^2(d_i) \sim \mathrm{IG}(a^a, b^a), \qquad d_i = 1, \ldots, 12.$$

The δ_l^a's are independent with prior distributions

$$\delta_l^a \sim \mathrm{N}(0, (\tau^a)^2), \qquad l = 1, \ldots, 1995,$$

and the ξ_q^a's are independent with prior distributions

$$\xi_q^a \sim \mathrm{N}(0, (\phi^a)^2).$$

The parameters are selected as $m^a = -1, s^a = 1, a^a = 3, b^a = 3, \tau^a = 1$, and $\phi^a = 1$.

For the home run data, the following decade-specific hierarchical model is used:

$$\theta_i^r \sim \mathrm{N}(\mu^r(d_i), (\sigma_\theta^r)^2(d_i)), \qquad i = 1, \ldots, 7{,}031,$$

where the θ_i^r's are independent conditional on the μ_θ^r's and $(\sigma_\theta^r)^2$'s. Assume that

$$\mu_\theta^r(d_i) \sim \mathrm{N}(m^r, (s^r)^2), \qquad d_i = 1, \ldots, 12$$

$$(\sigma_\theta^r)^2(d_i) \sim \mathrm{IG}(a^r, b^r), \qquad d_i = 1, \ldots, 12.$$

The δ_l^r's are independent with prior distributions

$$\delta_l^r \sim \mathrm{N}(0, (\tau^r)^2), \qquad l = 1901, \ldots, 1995$$

and the ξ_q^r's are independent with prior distributions

$$\xi_q^r \sim \mathrm{N}(0, (\phi^r)^2).$$

We set $m^r = -3.5, s^r = 1, a^r = 3, b^r = 3, \tau^r = 1$, and $\phi^r = 1$. In both the average and the home run studies, the selection of the parameters in the priors have essentially no effect on the final conclusion.

3.4 Aging Functions

A common perception of aging functions is that players improve in ability as they mature, up to a peak level, then slowly decline. It is generally believed that players improve faster while maturing than they decrease in ability while declining. The aging curve is clearly different for different sports with regard to both peak age and the rate of change during maturity and decline. Moreover, some players tend to play at near-peak performance for a long period of time, whereas others have short periods of peak performance. This may be due to conditioning, injuries, or genetics. We assume a *mean aging curve* for each sport. We model the variation in aging for each player using hierarchical models, with the mean aging curve as the standard. In each model, θ_i represents the ability of player i at peak performance in a benchmark year. Thus each player's ability is characterized by a profile rather than one number; it may be that player A is better than player B when they are both 22 years old, but player B is better than player A when they are both 35. For convenience we round off ages, assuming that all players were born on January 1. Lindley and Smith (1972) proposed using random polynomial curves. Shi, Weiss, and Taylor (1996) used random spline curves to model CD4 cell counts in infants over time. Their approach is similar to ours in that it models an effect in longitudinal data with a flexible random curve.

We let $g(a)$ denote the mean aging curve in each sport. We let $\breve{a}$ be the peak age for a player. Without loss, we assume that $g(\breve{a}) = 0$. We use the following model for player i's aging curve:

$$f_i(a) = \begin{cases} g(a)\psi_{1i} & \text{if } a < a_{\mathrm{M}} \\ g(a) & \text{if } a_{\mathrm{M}} \le a \le a_{\mathrm{D}} \\ g(a)\psi_{2i} & \text{if } a > a_{\mathrm{D}}. \end{cases}$$

The parameter $\psi_i = (\psi_{1i}, \psi_{2i})$ represents player i's variation from the mean aging curve. We define the maturing period as any age less than a_{M} and the declining period as any age greater than a_{D}. To preserve the interpretation of ψ_1 and ψ_2 as aging parameters, we select a_{M} and a_{D} where the aging on each side becomes significant. We fit the mean aging function for every player, then select ages (or knots) a_{M} and a_{D}, to represent players after their rise and before their steady decline. For ages a, such that, $a_{\mathrm{M}} \le a \le a_{\mathrm{D}}$, each player ages the same. This range was determined from initial runs of the algorithm. We selected a region in which the players' performance was close to the peak performance. Part of the motivation for a range of values unaffected by individual aging patterns is to ensure stability in the calculations. In each study we use the hierarchical model

$$(\psi_{1i}, \psi_{2i})' \sim \mathrm{N}_2((1,1)^T,\ \mathrm{diag}(\sigma_{\mathrm{M}}^2, \sigma_{\mathrm{D}}^2)),$$

which is a bivariate normal distribution. We use IG(10, 1) priors for σ_{M}^2 and σ_{D}^2 (mean .11 and standard deviation .039). Due to the large number of players in each example, the priors that we considered reasonable had virtually identical results.

In the golf model, $g(a)$ represents the additional number of strokes worse than peak level for the average professional golfer at age a. The maturing and declining parameters for each player have a multiplicative effect on the additional number of strokes. A player with $\psi_1 = 1$ matures the same as the average player. If $\psi_1 > 1$, then the player averages more strokes over his peak value than the average player would at the same age $a < a_{\mathrm{M}}$. If $\psi_1 < 1$, then the player averages fewer strokes over his peak value than the average player would at the same age $a < a_{\mathrm{M}}$. The same interpretation holds for ψ_2, only representing players of age $a > a_{\mathrm{D}}$.

The quantity $\exp(f_i(a))$ has a multiplicative effect on the mean points per game parameter in hockey and on the log-odds of success in baseball. Therefore, $\exp(g(a))$ represents the proportion of peak performance for the average player at age a.

We use a nonparametric form for the mean aging function in each sport:

$$g(a) = \alpha_a, \qquad a = \min(\text{age}), \ldots, \max(\text{age}), \tag{1}$$

where the α's are parameters. The only restriction is that $\alpha_a \equiv 0$ for some value a. We select $\alpha_{\breve{a}} = 0$ by initial runs of the algorithm to find $\breve{a}$. This preserves the interpretation for the θ's as the peak performance values. This model allows the average aging function to be of arbitrary form on both sides of peak age. In particular, the aging function may not be monotone. Although this may be nonintuitive, it allows for complete flexibility. A restriction is that the age of peak performance is the same across a sport. We believe that this is a reasonable assumption. The model is robust against small deviations in the peak age, because the aging function will reflect the fact that players performed well at those ages. By allowing players to age differently during the maturing and declining stages, each player's aging function can better represent good performance away from the league-wide peak. An alternative would be to model the peak age as varying across the population using a hierarchical model.

We tried alternative aging functions that were parametric. We used a quadratic form and an exponential decay (growth) model. Both of these behaved very similar to the nonparametric form close to the peak value. The parametric forms behaved differently for very young and very old players. The parametric form was too rigid in that it predicted far worse performance for older players. A piecewise parametric form may be more reasonable.

4. ALGORITHMS

In this section we describe the Markov chain Monte Carlo algorithms used to calculate the posterior distributions. The structure of the programs is to successively generate values one at a time from the complete conditional distributions (see Gelfand and Smith 1990; Tierney 1994).

In the golf model, all of the complete conditional distributions are available in closed form. In the hockey, batting average, and home run models, a Metropolis–Hastings step is used for most of the complete conditional distributions (see Chib and Greenberg 1995). In all of the models, generating the decade specific means and standard deviations are available in closed form.

Our results are based on runs with burn-in lengths of 5,000. Every third observation from the joint distribution is selected from one chain until 10,000 observations are collected. We used Fortran programs on a 166 MHz Sun Sparc Ultra 1. The golf programs took about 15 minutes, the hockey programs about 30 minutes, and each baseball program took about 80 minutes. With thousands of parameters, monitoring convergence is difficult. We found that most of the parameters depended on the year effects, and so concentrated our diagnostic efforts on the year effects. The algorithm appeared to converge very quickly to a stable set of year effects. Waiting for convergence for 5,000 observations is probably overkill. Monitoring the mixing of the chain is also difficult. Again the year effects were important. We also monitored those effects generated with a Metropolis step. We varied the candidate distributions to assure that the chain was mixing properly.

To validate our approach, we designed simulations and compared the results with the known values. We set up scenarios where players were constant over time, getting gradually better, and getting gradually worse. We crossed this with differing year effects. Some of the aspects of the models were developed using this technique. For example, we adopted different means and standard deviations for each decade based on their increased performance in the simulations. Our models did very well in the simulations. In particular, we found no systematic bias from these models.

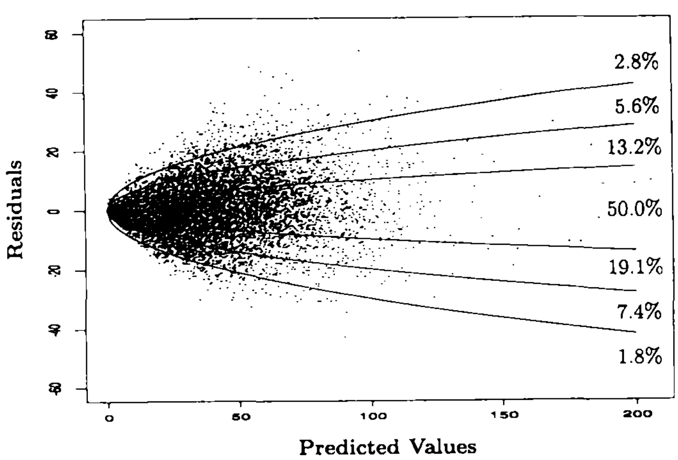

Figure 1. The Residuals in the Hockey Study Plotted Against the Fitted Values. The lines are ± 1, 2, and 3 square root of the predicted values. These are the standard deviations assuming the model and parameters are correct, and the data are truly Poisson. The percentage of observations in each of the regions partitioned by the ± 1, 2, and 3 standard deviations are reported on the graph.

5. GOODNESS OF FIT

In this section we consider the appropriateness of our models and address their overall fit. For each sport we present an analysis of the residuals. We look at the sum of squared errors for our fitted model (referred to as the *full* model) and several alternative models. The *no individual aging* model is a subset of the full model, with the restriction that $\psi_{1i} = 1$ and $\psi_{2i} = 1$, for all i. The *no aging effects* model assumes that $f_i(a) = 0$, for all i and a. The *null* model is a one-parameter model that assumes all players are identical and there are no other effects. Although this one-parameter model is not taken seriously, it does provide some information about the fit of the other models.

The objective sum of squares is the expected sum of squares if the model and parameters are correct. This is an unattainable goal in practice but gives an overall measure of the combined fit of the model and the parameters. We provide an analog to R^2, which is the proportion of sum of squares explained by each model. For each model M, this is defined as $1 - \mathrm{SS}_M/\mathrm{SS}_N$, where SS_M refers to the sum of squared deviations using model M and subscript N indexes the null model. Myers (1990) discussed R^2 in the context of log-linear models.

In calculating the sum of squares, we estimate the parameters with their posterior means.

5.1 Hockey

Figure 1 plots the residual points per player season are plotted against the predicted points per player season. We include curves for $\pm 1, 2$, and 3 times the square root of the predicted values. These curves represent the $\pm 1, 2$, and 3 standard deviations of the residuals, assuming that the parameters and model are correct. The percent of residuals in each region is also plotted. The residual plot demonstrates a lack of fit of the model.

Table 2 presents the sum of squared deviations for each model. The sum of squares for each model is the sum of the squared difference between the model estimated point total and the actual point total, over every player season.

Table 2. The Sum of Squared Deviations (SS) Between the Predicted Point Totals for Each Model and the Actual Point Totals in the Hockey Example

Model	SS	R^2
Objective	346,000	.91
Full	838,000	.79
No individual aging	980,000	.75
No aging effects	1,171,000	.70
Null	3,928,000	

The objective sum of squares is $\sum_{ij} g_{ij}\hat{\lambda}_{ij}$, where $\hat{\lambda}_{ij}$ is the estimate of the points per game parameter from the full model. This represents the expected sum of squares if the model and parameters are exactly correct.

We feel that the model is reasonable but clearly demonstrates a lack of fit. The full model is a huge improvement over the null model, but it still falls well short of the objective. Of the three examples (golf has no objective), hockey represents the biggest gap between the objective and the full model. We believe that this is because strong interactions are likely in hockey. Of the three sports studied, hockey is the most team oriented, in which the individual statistics of a player are the most affected by the quality of his teammates. For example, Bernie Nicholls scored 78 points in the 1987–88 season without Wayne Gretzky as a teammate, and scored 150 points the next season as Gretzky's teammate.

There is also strong evidence that the aging effects and the individual aging effects are important. The R^2 is increased by substantial amounts by adding the age effects and additionally, the individual aging effects. We think that the aging functions have a large effect because hockey is a physically demanding sport in which a slight loss of physical skill and endurance can have a big impact on scoring ability. Two players who have slight differences in aging patterns can exhibit large differences in point totals (relative to their peaks).

5.2 Golf

Figure 2 is a normal probability plot of the standardized residuals in the golf example. The pattern in the residual q-q plot is interesting, showing a deviation from normality. The left tail is "lighter" than that of a normal distribution, and the right tail is "heavier" than that of a normal distribution. As discussed in Section 3.2, this makes intuitive sense. It is very difficult to score low, and it is reasonably likely to score high. We tried various right-skewed distributions but found little difference in the results. The only difference we can see is in predicting individual scores, which is not a goal of this article.

Table 3 presents the sum of squared deviations between the estimated scores from each model and the actual scores.

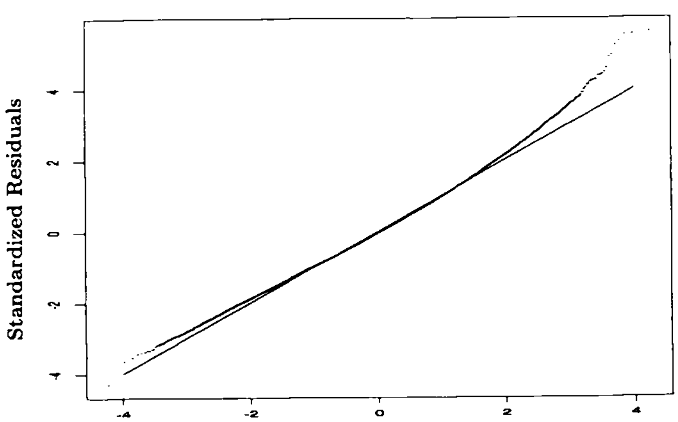

Figure 2. Normal Probability Plot of the Residuals From the Golf Model.

Table 3. The Sum of Squared Deviations (SS) Between the Predicted Score for Each Model and the Actual Score in the Golf Example

Model	SS	R^2
Full	366,300	.30
No individual aging	366,600	.30
No aging effects	372,000	.29
Null	527,000	

Because of the normal model there is no objective sum of squares to compare the fit. The variance in the scores, σ_x^2, is a parameter fitted by the model and thus naturally reflects the fit of the model. Despite the small improvement between the null model and the full model, we feel that this is a very good-fitting model. This conclusion is based on the estimate of σ_x, which is 2.90. The R^2 for this model is only .30, which is small, but we believe there is a large amount of variability in a golf score, which will never be modeled. We were pleased with a standard error of prediction of 2.90. There is little evidence that aging plays an important role in scoring. This is partly due to the fact that most of the scores in the dataset are recorded when players are in their prime. Few players qualified for the majors when they were very old or very young, and for these ages there is an effect. There is also little evidence that individual aging effects are needed, but this suffers from the same problem just mentioned.

5.3 Baseball

The residual plot for each baseball example indicated no serious departures from the model. The normal probability plots showed almost no deviations from normality. Table 4 presents the home run sum of squares; Table 5, the batting average sum of squares. The batting average example presents the sum of squared deviations of the predicted number of base hits from the model from the actual number of base hits. The R^2 is .60 for the full model, very close to the objective sum of squares of .62. We believe the batting average model is a good-fitting model. The home run model does not fit as well as the batting average example. Despite an R^2 of .80, it falls substantially short of the objective sum of squares. The high R^2 is due to the large spread in home run ability across players, for which the null model does not capture.

Aging does not play a substantial role in either measure. This is partly due to the large number of observations close to peak, where aging does not matter, but also can be attributed to the lack of a strong effect due to aging. The contrast between the four examples in the role of aging

Table 4. The Sum of Squared Deviations (SS) Between the Predicted Number of Home Runs for Each Model and the Actual Number of Home Runs in the Home Run Example

Model	SS	R^2
Objective	171,000	.86
Full	238,000	.80
No individual aging	242,500	.80
No aging effects	253,700	.78
Null	1,203,000	

Table 5. The Sum of Squared Deviations (SS) Between the Predicted Number of Hits for Each Model and the Actual Number of Hits in the Batting Average Example

Model	SS	R^2
Objective	1,786,000	.62
Full	1,867,000	.60
No individual aging	1,897,000	.60
No aging effects	1,960,000	.58
Null	4,699,000	

and the individual aging effects is interesting. In the most physically demanding of the sports, hockey, aging plays the greatest role. In the least physically demanding sport, golf, the aging effect plays the smallest role.

6. AGE EFFECT RESULTS

Figures 3–6 illustrate the four mean age effect (g) functions. Figure 3 shows the hockey age function. The y-axis represents the proportion of peak performance for a player of age a. Besides keeping track of the mean of the aging function for each age, we also keep track of the standard deviation of the values of the curve. The dashed lines are the ± 2 standard deviation curves. This graph is very steep on both sides of the peak age, 27. The sharp increase during the maturing years is surprising—20- to 23-year-old players are not very close to their peak. Because of the sharp peak, we specified 29 and older as declining and 25 and younger as maturing.

Figure 4 presents the average aging function for golf. In this model g represents the average number of strokes from the peak. The peak age for golfers is 34, but the range 30–35 is essentially a "peak range." The rate of decline for golfers is more gradual than the rate of maturing. An average player is within .25 shots per round (1 shot per tournament) from peak performance when they are in the 25–40 age range. An average 20-year-old and an average 50-year-old are both 2 shots per round off their peak performance. Because of the peak range from 30–35, we specified the declining stage as 36 and older and the maturing phase as 29 and younger.

Figures 5 and 6 present the aging functions for home runs and batting averages. The home run aging function presents the estimated number of home runs for a player who is a 20–home run hitter at his peak. The peak age for home runs is 29. A 20–home run hitter at peak is within 2 home runs of his peak at 25–35 years old. There is a sharp increase for maturers. Apparently, home run hitting is a talent acquired through experience and learning, rather than being based on brute strength and bat speed. The ability to hit home runs does not decline rapidly after the peak level—even a 40-year-old 20–home run-at-peak player is within 80% of peak performance.

The age effects for batting average are presented for a hitter who is a .300 hitter at his peak. Hitting for average does differ from home run hitting—27 is the peak age, and younger players are relatively better at hitting for average than hitting home runs. An average peak .300 hitter is expected to be a .265 hitter at age 40. For batting average and home runs, the maturing phase is 25 and younger and the declining phase is 31 and older.

7. PLAYER AND SPORT RESULTS

This section presents the results for the individual players and the year effects within each sport. To understand the rankings of the players, it is important to see the relative difficulty within each sport over the years. Each player is characterized by θ, his value at peak performance in a benchmark year, and by his aging profile. We present tables that categorize players by their peak performance, but we stress that their career profiles are a better categorization of the players. For example, in golf Jack Nicklaus is the best player when the players are younger than 43, but Ben Hogan is the best for players over 43. The mean of their maturing and declining parameters are presented for comparison.

7.1 Hockey

The season effect in hockey is strong. Figure 7a shows the estimated multiplicative effects, relative to 1996. From 1948–1968 there were only six teams in the NHL, and the game was defensive in nature. In 1969 the league added six teams. The league continued to expand to the present 30 teams. With this expansion, goal scoring increased. The 1970s and early 1980s were the height of scoring in the NHL. As evidence of the scoring effects over the years, many players who played at their peak age in the 1960s with moderate scoring success played when they were "old" in the 1970s and scored better than ever before (e.g., Gordie Howe, Stan Mikita, and Jean Beliveau). In 1980 the wide-open offensive-minded World Hockey Association, a competitor to the NHL, folded and the NHL absorbed some of the teams and many of the players. This added to the offensive nature and style of the NHL. In the 1980s the NHL began to attract players from the Soviet block, and the United States also began to produce higher-caliber players. This influx again changed the talent pool.

Scoring began to wane beginning in 1983. This is attributed in part to a change in the style of play. Teams went from being offensive in nature to defensively oriented. "Clutching and grabbing" has become a common term to describe the style of play in the 1990s. As evidence of this, in 1998 the NHL made rule changes intended to increase scoring. The seasonal effects are substantial. The model predicts that a player scoring 100 points in 1996 is would have scored 140 points in the mid-1970s or early 1980s.

Table 6 presents the top 25 players, rated on their peak level. Figure 8 presents profiles of some of these best players. It demonstrates the importance of a profile over a one-number summary. Mario Lemieux is rated as the best peak-performance player, but Wayne Gretzky is estimated to be better when they are young, whereas Lemieux is estimated to be the better after peak. The fact that Lemieux is ahead of Gretzky at peak may seem surprising. Gretzky played during the most wide-open era in the NHL, whereas Lemieux played more of his career in a relatively defensive-minded era. Lemieux's career and season totals are a bit misleading, because he rarely played a full season. He missed many games throughout his career, and we rate players on their per game totals.

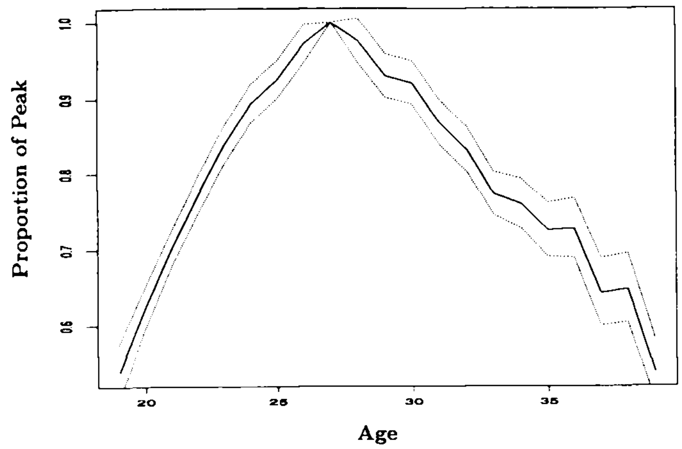

Figure 3. The Estimated Mean Aging Function and Pointwise ±2 Standard Deviation Curves for the Hockey Study. The y-axis is the proportion of peak for a player of age a.

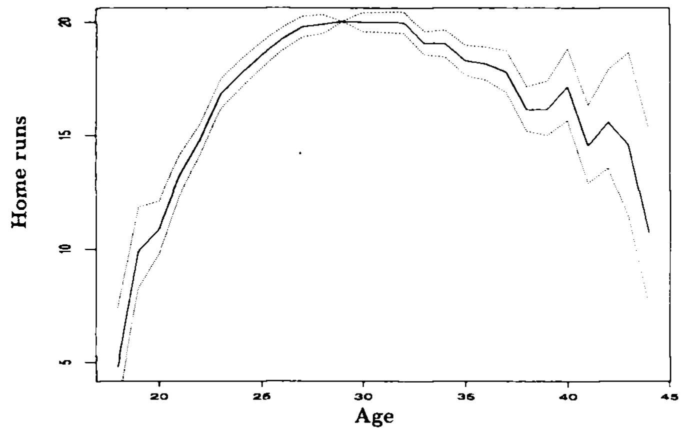

Figure 5. The Estimated Mean Aging Function and Pointwise ±2 Standard Deviation Curves for the Home Run Study. The y-axis is the number of home runs for a player who is a 20-home run hitter at peak performance.

As a cross-validation, we present the model-predicted point totals for the 10 highest-rated peak players who are still active. We estimated the 1997 season effect, $\hat{\delta}_{1997} = -.075$, by the log of the ratio of goals scored in 1997 to goals scored in 1996. Table 7 presents the results. We calculated the variance of each predicted point total using the variance of the Poisson model and the points per game parameter, λ_{ij} (the standard deviation is reported in Table 7). With the exception of Pavel Bure, the predictions are very close.

7.2 Golf

Figure 7b shows the estimate for the difficulty of each round of the Masters tournament. The mean of these years is also plotted. We selected the Masters because it is the one tournament played on the same course (Augusta National) each year and the par has stayed constant at 72. These estimates measure the difficulty of each round, separated from the ability of the players playing those rounds. These estimates may account for weather conditions, course difficulty, and the equipment of the time. Augusta in the 1940s played easier than in the 1950s or 1960s; we are unsure why. There is approximately a 1 shot decrease from the mid-1950s to the present. We attribute the decrease from the 1950s to the present to the improved equipment available to the players. Although it does appear that Augusta National is becoming easier to play, the effects of improved equipment do not appear to be as strong as public perception would have one believe. Augusta is a challenging course in part because of the speed and undulation of the greens. It may be that the greens have become faster, and more difficult, over the years. If this is true, then the golfers are playing a more difficult course and playing it 1 shot better than before. Such a case would imply the equipment has helped more than 1 shot.

Table 8 shows the top 25 players of all time at their peak. Figure 9 shows the profile of six of the more interesting careers. The y-axis is the predicted mean average for each player when they are the respective age. Jack Nicklaus at his peak is nearly .5 shot better than any other player. Nicklaus essentially aged like the average player. Ben Hogan, who is .7 shot worse than Nicklaus at peak, aged very well. He is estimated to be better than Nicklaus at age 43 and older. The beauty of the hierarchical models comes through in the estimation of Tiger Woods' ability. Woods has played very

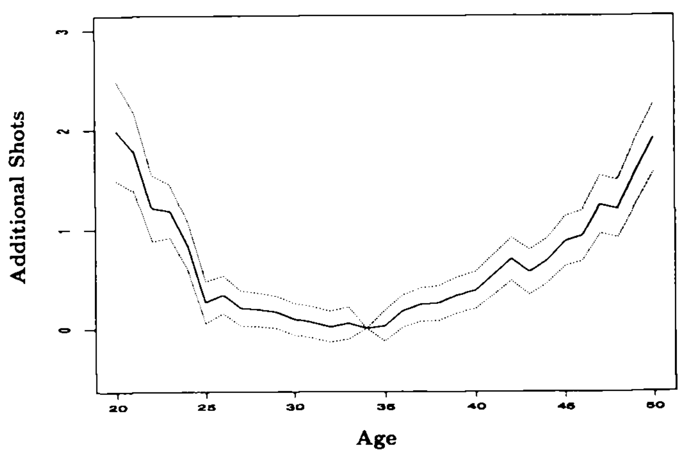

Figure 4. The Estimated Mean Aging Function and Pointwise ±2 Standard Deviation Curves for the Golf Study. The y-axis is the number of shots more than peak value for a player of age a.

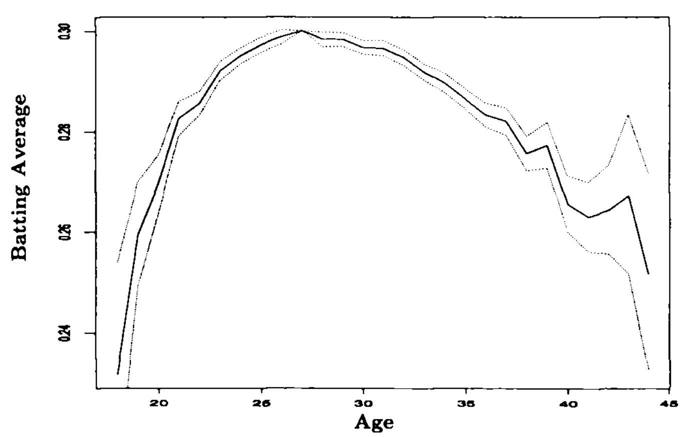

Figure 6. The Estimated Mean Aging Function and Pointwise ±2 Standard Deviation Curves for the Batting Average Study. The y-axis is the batting average for a player who is a .300 hitter at peak performance.

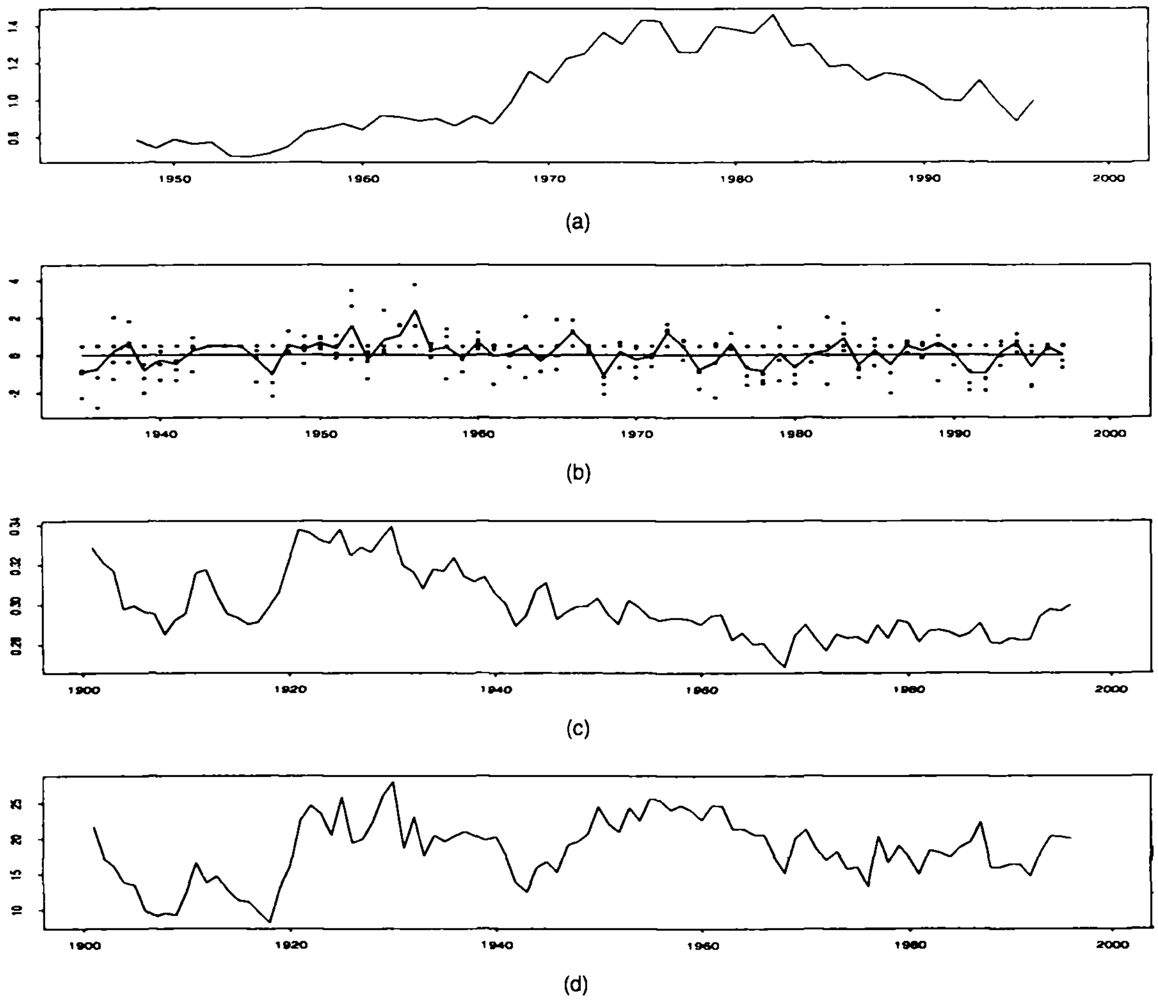

Figure 7. The Yearly Effects for the Hockey (a), Golf (b), Batting Average (c), and Home Run (d) Studies. The hockey plot shows the multiplicative effect on scoring for each year, relative to 1996. The golf plot shows the additional number of strokes for each round in the Masters, relative to the average of 1997. The line is the average for each year. The home run plot shows the estimated number of home runs for a 20–home run hitter in 1996. The batting average plot shows the estimated batting average for a player who is a .300-hitter in 1996.

well as a 21-year-old player (winning the Masters by 12 shots). He averaged 70.4 during the benchmark 1997 year. If he aged like the average player, which means during 1997 he was 1.8 shots per round off his peak performance, then he would have a peak performance of 68.6. If he aged like the average player, then he would be by far the best player of all time. This is considered very unlikely because of the distribution of players. It is more likely that he is a quick maturer and is playing closer to his peak than the average 21 year old. Thus his maturing parameter is estimated to be .52. The same phenomenon is seen for both Ernie Els and Justin Leonard, who are performing very well at young ages.

7.3 Baseball

Figures 7c and 7d illustrate the yearly effects for home runs and batting average. They show that after 1920, when the dead-ball era ended, the difficulty of hitting home runs has not changed a great deal. A 20–home run hitter in 1996 is estimated to have hit about 25 in the mid-1920s. Home run hitting has slowly decreased over the years, perhaps because of the increasing ability of pitchers. The difficulty of getting a base hit for a batter of constant ability has also decreased since the early 1920s. The probability of getting a hit bottomed out in 1968 and has increased slightly since then. The slight increase after 1968 has been attributed to the lowering of the pitcher's mound, thus decreasing the pitchers' ability, and also to expansion in MLB. Most baseball experts believe that umpires are now using a smaller strike zone, which may also play a role. We attribute part of the general decrease over the century in the difficulty of getting a base hit to the increasing depth and ability of pitchers.

Tables 9 and 10 show the top 25 peak batting average and home run hitters. The posterior means for peak performance in the benchmark year of 1996, the maturing parameter (ψ_1), and the declining parameter (ψ_2) are provided. Figures 10 and 11 show the career profiles for the batting average and home run examples. The model selects Mark McGwire as the greatest home run per at bat hitter in history. The model estimates that Babe Ruth's at his prime would hit 5 fewer home runs than McGwire in 1996. Interestingly the all-time career home run king (with 755) Hank Aaron, is only 23rd on the peak performance list. Aaron declined very slowly (the slowest of the top 100). He is higher on the batting average list (13) than on the home run list!

Table 6. The Top 25 Peak Players in the Hockey Study

Rank	Name	Born	Points in 1996	ψ_1	ψ_2
1	M. Lemieux	1965	187 (7)	1.18	.89
2	W. Gretzky	1961	181 (5)	.66	1.66
3	E. Lindros	1973	157 (16)	.93	1
4	J. Jagr	1972	152 (9)	1.37	1
5	P. Kariya	1974	129 (15)	.95	1
6	P. Forsberg	1973	124 (10)	.84	1
7	S. Yzerman	1965	120 (5)	.91	1.43
8	J. Sakic	1969	119 (6)	.95	1
9	G. Howe	1928	119 (7)	1.04	.69
10	T. Selanne	1970	113 (6)	.78	1
11	P. Bure	1971	113 (8)	.81	1
12	J. Beliveau	1931	112 (5)	.67	.90
13	P. Esposito	1942	112 (5)	1.82	1.36
14	A. Mogilny	1969	112 (6)	1.18	1
15	P. Turgeon	1969	110 (6)	.95	1
16	S. Federov	1969	110 (5)	1.05	1
17	M. Messier	1961	110 (4)	1.51	.55
18	P. LaFontaine	1965	109 (5)	1.20	1.32
19	Bo. Hull	1939	108 (4)	.94	1.29
20	M. Bossy	1957	108 (4)	.86	1.02
21	Br. Hull	1964	107 (5)	1.15	1.12
22	M. Sundin	1971	106 (7)	.99	1
23	J. Roenick	1970	106 (6)	.67	1
24	P. Stastny	1956	105 (4)	1.20	1.12
25	J. Kurri	1960	105 (4)	1.11	1.30

NOTE: The means of ψ_1 and ψ_2 are also presented. The Points in 1996 column represents the mean points (with standard deviations given in parentheses) for the player in 1996 if the player was at his peak performance.

Willie Stargell and Darryl Strawberry provide an interesting contrast in profiles. At peak they are both considered 41 home run hitters. Strawberry is estimated to have matured faster than Stargell, whereas Stargell maintained a higher performance during the declining phase.

Ty Cobb, who played in his prime about 80 years ago, is still considered the best batting average hitter of all time. Tony Gwynn is estimated to decline slowly ($\psi_2 = .78$) and is considered a better batting average hitter than Cobb after age 34. Paul Molitor is estimated to be the best decliner of the top 100 peak players. At age 40, in 1996, he recorded a

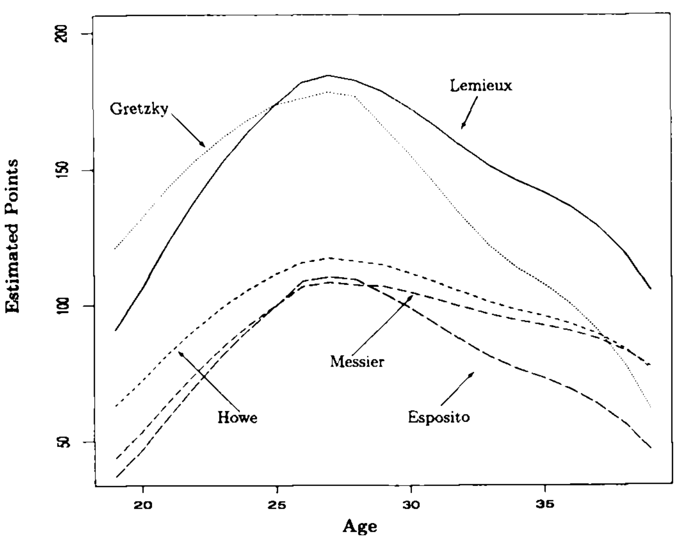

Figure 8. A Profile of Some of the Best Players in the Hockey Study. The estimated mean number of points for each age of the player, if that season were 1996.

batting average of .341. Alex Rodriguez exhibits the same regression to the mean characteristics as Tiger Woods does in golf. In Rodriguez's second year, the benchmark year of 1996, he led the American League in hitting at .358. The model predicts that at his peak in 1996 he would hit .336. Because of the shrinkage factor, as a result of the hierarchical model, it is more likely that Rodriguez is closer to his peak than the average player (i.e., is a rapid maturer) and that 1996 was a "lucky" year for him.

We recorded 78 ballparks in use in MLB beginning in 1901. When a ballpark underwent significant alterations, we included the "before" and "after" parks as different. The constraint for the parks is that $\xi_{\{\text{new Fenway}\}} = 0$ (There is an old Fenway, from 1912–1933, and a new Fenway, 1933–. Significant changes were made in 1933, including moving the fences in substantially.) We report the three easiest and three hardest ballparks for home runs and batting average. (We ignore those with less than 5 years of use unless they are current ballparks.) For a 20–home run hitter in new Fenway, the expected number of home runs in the three easiest home run parks are 30.1 in South End Grounds (Boston Braves, 1901–1914), 28.6 in Coors Field (Colorado, 1995–), and 26.3 in new Oakland Coliseum (Oakland, 1996–). The 20–home run hitter would be expected to hit 14.5 at South Side (Chicago White Sox, 1901–1909), 14.8 at old Fenway (Boston, 1912–1933), and 15.9 at Griffith Stadium (Washington, 1911–1961), which are the three most difficult parks. The average of all ballparks for a 20–home run hitter at new Fenway is 20.75 home runs.

For a .300 hitter in new Fenway, the three easiest parks in which to hit for average are .320 at Coors Field, .306 at Connie Mack Stadium (Philadelphia, 1901–1937), and .305 at Jacobs Field (Cleveland, 1994–). The three hardest parks in which to hit for average are .283 at South Side, .287 at old Oakland Coliseum (Oakland, 1968–1995), and .287 at old Fenway. New Fenway is a good (batting average) hitters' park. A .300 hitter at new Fenway would be a .294 hitter in the average of the other parks. Some of the changes to the ballparks have been dramatic. Old Fenway was a very difficult park in which to hit for average or home runs, but after the fences were moved in, the park became close to average. The Oakland Coliseum went from a very difficult park to a very easy park after the fences were moved in and the outfield bleachers were enclosed in 1996.

Table 7. The Predicted and Actual Points for the Top 10 Model-Estimated Peak Players Who Played in 1997

Rank	Name	Age in 1997	Games played	Predicted points	Actual points
1	M. Lemieux	32	76	135 (14.9)	122
2	W. Gretzky	36	82	103 (13.9)	97
3	E. Lindros	24	52	84 (12.0)	79
4	J. Jagr	25	63	99 (13.1)	97
5	P. Kariya	23	69	86 (12.8)	99
6	P. Forsberg	24	65	83 (12.5)	86
7	S. Yzerman	32	81	84 (13.1)	85
8	J. Sakic	28	65	87 (12.6)	74
10	T. Selanne	27	78	99 (13.6)	109
11	P. Bure	27	63	81 (12.3)	55

NOTE: The model used data only from 1996 and prior to predict the point totals.

Table 8. The Top 25 Peak Players in the Golf Study

Rank	Name	Born	θ	ψ_1	ψ_2
1	J. Nicklaus	1940	70.42 (.29)	1.03	.99
2	T. Watson	1949	70.82 (.23)	.92	1.19
3	B. Hogan	1912	71.12 (.29)	1.13	.27
4	N. Faldo	1957	71.19 (.21)	1.19	1.21
5	A. Palmer	1929	71.33 (.28)	1.19	.95
6	G. Norman	1955	71.39 (.19)	1.21	.64
7	J. Leonard	1972	71.40 (.45)	.68	1
8	E. Els	1969	71.45 (.34)	.78	1
9	G. Player	1935	71.45 (.23)	.87	.62
10	F. Couples	1959	71.50 (.21)	1.00	.97
11	H. Irwin	1945	71.56 (.26)	1.02	.68
12	C. Peete	1943	71.56 (.36)	1	.80
13	J. Boros	1920	71.62 (.37)	1	.61
14	R. Floyd	1942	71.63 (.24)	1.22	.38
15	L. Trevino	1939	71.63 (.29)	1.00	.72
16	S. Snead	1912	71.64 (.27)	1.10	.21
17	J. Olazabal	1966	71.69 (.39)	.74	1
18	T. Kite	1949	71.71 (.23)	.98	.70
19	B. Crenshaw	1952	71.74 (.22)	.43	1.22
20	T. Woods	1975	71.77 (.64)	.52	1
21	B. Casper	1931	71.77 (.26)	1.00	1.09
22	B. Nelson	1912	71.78 (.31)	1.00	1.11
23	P. Mickelson	1970	71.79 (.44)	.79	1
24	L. Wadkins	1949	71.79 (.22)	1.13	.78
25	T. Lehman	1959	71.82 (.30)	1.05	.79

NOTE: The standard deviations are in parentheses. The means of ψ_1 and ψ_2 are also presented.

As cross-validation we present the model predictions for 1997 performance. Recall, the baseball study uses data from 1996 and earlier in the estimation. We estimate the season effect of 1997 by the league-wide performance relative to 1996. For batting average, the estimated year effect for 1997 is −.01. Table 11 presents the model-predicted batting average for the 10 highest-rated peak batting average players of all time who are still active. The estimates are good, except for Piazza and Gwynn, both of whom had batting averages approximately two standard deviations above the predicted values.

Table 12 presents the model-predicted and actual number of home runs, conditional on the number of at bats, for the 10 highest-rated peak home run hitters of all time who are still active. The estimated year effect for 1997 is −.06. Palmer, Belle, and Canseco did worse than their projected values. The model provided a nice fit for Griffey and McGwire, each of whom posted historical years that were not so unexpected by the model. Standard errors of prediction were calculated using the error of the binomial model and the error in the estimates of player abilities, age effects, and ballpark effects.

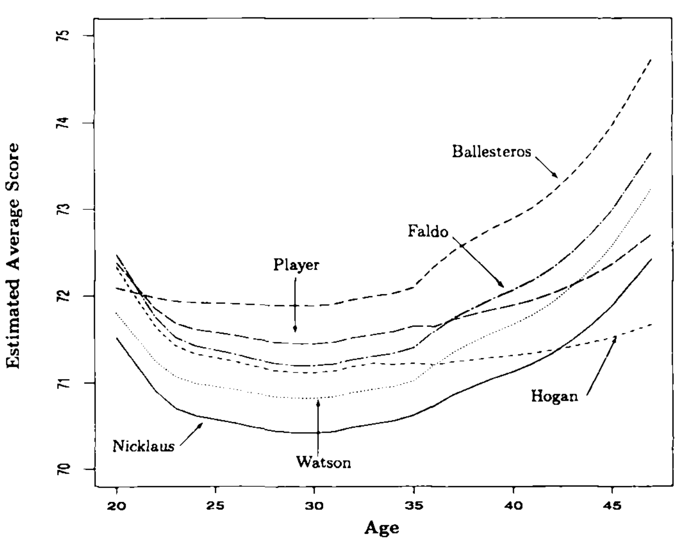

Figure 9. A Profile of Some of the Best Players in the Golf Study. The estimated mean score for each age of the player, if that round were an average 1997 round.

Table 9. The Top 25 Peak Players for the Batting Average Study

Rank	Name	Born	Average	ψ_1	ψ_2
1	T. Cobb	1886	.368 (.005)	1.14	1.31
2	T. Gwynn	1960	.363 (.006)	1.08	.78
3	T. Williams	1918	.353 (.006)	.95	.93
4	W. Boggs	1958	.353 (.005)	1.05	1.17
5	R. Carew	1945	.351 (.005)	1.06	.92
6	J. Jackson	1889	.347 (.007)	.86	1.12
7	N. Lajoie	1874	.345 (.009)	1	1.36
8	S. Musial	1920	.345 (.005)	.98	1.16
9	F. Thomas	1968	.344 (.008)	.99	1
10	E. Delahanty	1867	.340 (.001)	1	1.02
11	T. Speaker	1888	.339 (.006)	.99	1.32
12	R. Hornsby	1896	.338 (.005)	1.02	1.02
13	H. Aaron	1934	.336 (.006)	.89	1.25
14	A. Rodriguez	1975	.336 (.001)	.85	1
15	P. Rose	1941	.335 (.004)	1.25	.89
16	H. Wagner	1874	.333 (.007)	1	1.30
17	R. Clemente	1934	.332 (.005)	1.37	.50
18	G. Brett	1953	.331 (.005)	.92	1.16
19	D. Mattingly	1961	.330 (.006)	.88	1.07
20	K. Puckett	1961	.330 (.006)	1.14	.93
21	M. Piazza	1968	.330 (.009)	1.04	1
22	E. Collins	1887	.329 (.004)	.96	1.01
23	E. Martinez	1963	.328 (.008)	1.22	.79
24	P. Molitor	1956	.328 (.005)	.94	.31
25	W. Mays	1931	.328 (.005)	.99	1.19

NOTE: The standard deviations are in parentheses. The means of ψ_1 and ψ_2 are presented.

Table 10. The Top 25 Peak Players in the Home Run Study

Rank	Name	Born	θ	ψ_1	ψ_2
1	M. McGwire	1963	.104 (.006)	.97	1.12
2	J. Gonzalez	1969	.098 (.008)	1.05	1
3	B. Ruth	1895	.094 (.004)	.72	.93
4	D. Kingman	1948	.093 (.004)	.96	1.05
5	M. Schmidt	1949	.092 (.005)	.99	1.18
6	H. Killebrew	1936	.090 (.005)	.87	1.13
7	F. Thomas	1968	.089 (.007)	.99	1
8	J. Canseco	1964	.088 (.004)	1.05	1.01
9	R. Kittle	1958	.086 (.006)	1.08	.96
10	W. Stargell	1940	.084 (.003)	1.24	.79
11	W. McCovey	1938	.084 (.004)	1.04	1.22
12	D. Strawberry	1962	.084 (.005)	.70	1.10
13	B. Jackson	1962	.083 (.006)	1.06	1.04
14	T. Williams	1918	.083 (.004)	.88	.97
15	R. Kiner	1922	.083 (.004)	1.01	1.05
16	P. Seerey	1923	.081 (.009)	.91	1
17	R. Jackson	1946	.081 (.004)	.83	1.11
18	K. Griffey	1969	.080 (.006)	1.03	1
19	A. Belle	1966	.080 (.006)	1.12	1
20	R. Allen	1942	.080 (.004)	1.16	1.12
21	B. Bonds	1964	.079 (.004)	1.27	1.05
22	D. Palmer	1968	.079 (.007)	1.07	1
23	H. Aaron	1934	.078 (.003)	1.26	.53
24	J. Foxx	1907	.078 (.003)	1.34	1.16
25	M. Piazza	1968	.078 (.006)	.95	1

NOTE: The standard deviations are in parentheses. The means of ψ_1 and ψ_2 are presented.

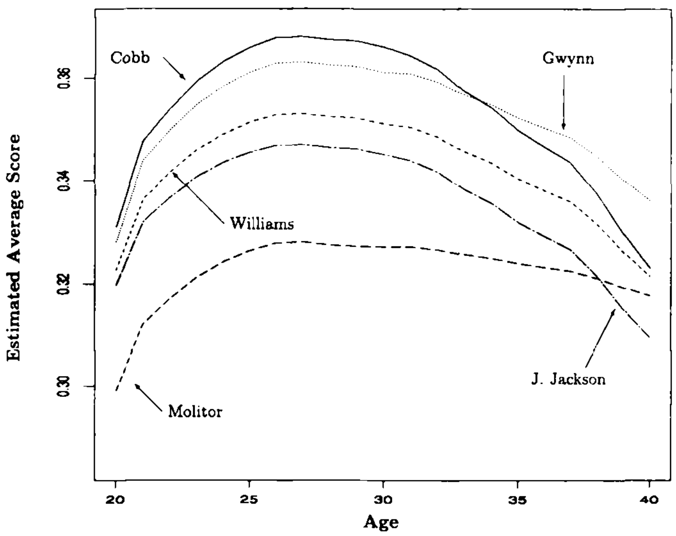

Figure 10. A Profile of Some of the Best Players in the Batting Average Study. The estimated batting average for each age of the player, if that year were 1996.

8. POPULATION DYNAMICS

In this section we address the changing distribution of players within each study. Figures 12–15 present graphs of the peak value estimate for each player, graphed against the year the player was born. These player effects are separated from all the other effects; thus the players can be compared directly.

In hockey there is some slight bias on each end of the population distribution (see Fig. 12). Players born early in the century were fairly old when our data began (1948). They are in the dataset only if they are good players. The restriction that each player plays at least 100 games was harder for a player to reach earlier in this century because a season consisted of 48 games, rather than the current 82 games. Therefore, there is a bias, overestimating the percentiles of the distribution of players for the early years.

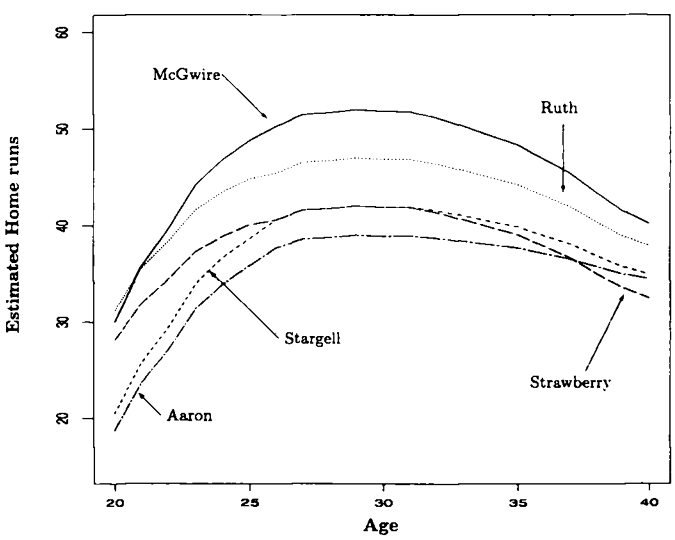

Figure 11. A Profile of Some of the Best Players in the Home Run Study. The estimated number of home runs, conditional on 500 at bats, for each age of the player, if that year were 1996.

Table 11. The Predicted and Actual Batting Averages (BA) for the Top 10 Model-Estimated Peak Players Who Played in 1997

Rank	Name	Age in 1997	At bats	Predicted BA	Actual BA
2	T. Gwynn	37	592	.329 (.021)	.372
4	W. Boggs	39	353	.318 (.027)	.292
9	F. Thomas	29	530	.328 (.023)	.347
14	A. Rodriguez	22	587	.312 (.022)	.300
21	M. Piazza	29	556	.316 (.023)	.362
23	E. Martinez	34	542	.309 (.022)	.330
24	P. Molitor	41	538	.290 (.022)	.305
29	R. Alomar	29	412	.316 (.026)	.333
39	K. Griffey	28	608	.313 (.022)	.304
47	M. Grace	33	555	.308 (.022)	.319

NOTE: The model used 1901–1996 data to predict 1997 totals. Standard deviations are in parentheses.

Of the players born late in this century (after 1970), it is more likely that the good ones are included. Thus the percentiles are probably overestimated slightly.

For hockey players born after 1940, there is a clear increase in ability. Of the top 25 players, 9 are current players who have yet to reach their peak (36%, where only 8% of the players in our data had not reached their peak). It is hard to address Gould's claim with the hockey distribution. This is because not everyone in this dataset is trying to score points. Many hockey players are role players, with the job of to playing defense or even just picking fights with the opposition! The same is true of the distribution of home run hitters in baseball. Many baseball players are not trying to hit home runs; their role may focus more on defense or on hitting for average or otherwise reaching base. The same type of pattern shows up in home run hitting. The top 10% of home run hitters are getting better with time (see Fig. 13). This could be attributed to the increasing size and strength of the population from which players are produced, the inclusion of African-Americans and Latin Americans, or an added emphasis by major league managers on hitting home runs.

It is easier to address Gould's claim with the batting average and golf studies. In baseball, every player is trying to get a base hit. Every player participates in the offense an equal amount, and even the defensive-minded players try to get hits. In golf, every player tries to minimize his

Table 12. The Predicted and Actual Home Runs (HR) for the Top 10 Model-Estimated Peak Players Who Played in 1997

Rank	Name	Age in 1997	At bats	Predicted HR	Actual HR
1	M. McGwire	34	540	55 (7.63)	58
2	J. Gonzalez	28	541	41 (7.08)	42
7	F. Thomas	29	530	42 (7.17)	35
8	J. Canseco	33	388	37 (6.36)	23
12	D. Strawberry	35	29	2 (1.58)	0
18	K. Griffey	28	608	49 (7.74)	56
19	A. Belle	31	634	46 (7.15)	30
21	B. Bonds	33	532	37 (6.48)	40
22	D. Palmer	29	556	34 (6.51)	23
25	M. Piazza	29	542	36 (6.63)	40

NOTE: The model used 1901–1996 data to predict 1997 totals. Standard deviations are in parentheses.

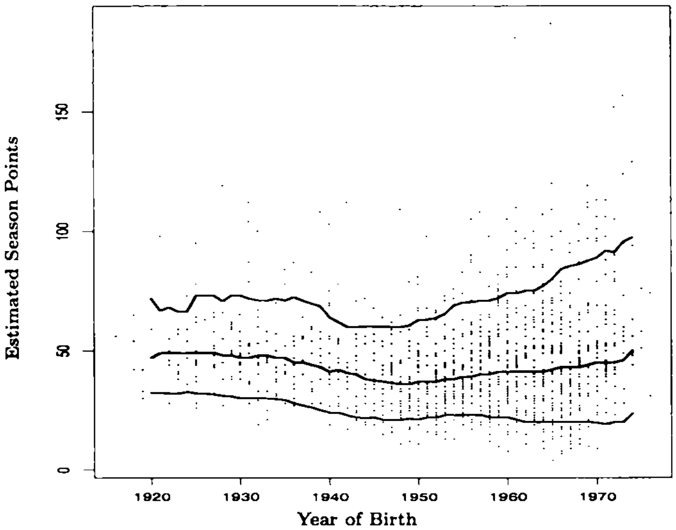

Figure 12. The Estimated Peak Season Scoring Performance of Each Player Plotted Against the Year They Were Born. The y-axis represents the mean number of points scored for each player, at their peak, if the year was 1996. The three curves are the smoothed 10th, 50th, and 90th percentiles.

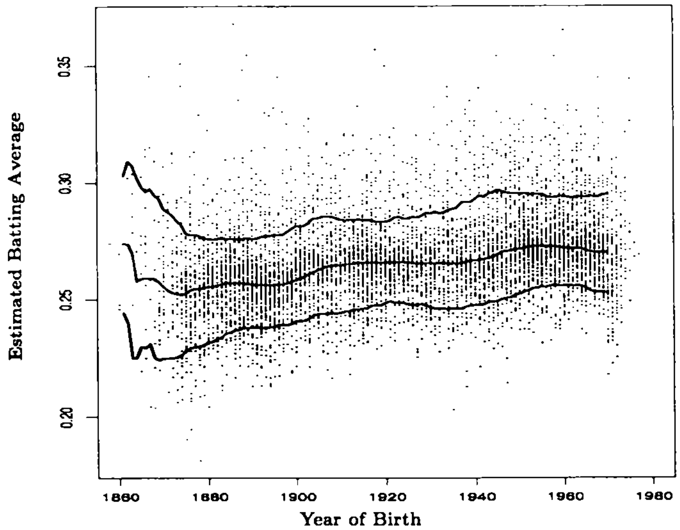

Figure 14. The Estimated Peak Batting Average Performance of Each Player Plotted Against the Year They Were Born. The y-axis represents the mean probability of a hit for each player, at their peak, if the year was 1996. The three curves are the smoothed 10th, 50th, and 90th percentiles.

scores—the only goal for the golfer. In the golf study there is a bias in the players who are in our dataset: Only players with 10 majors are included. It was harder to achieve this in the early years, because we have data on only two majors until 1961. It was also hard to find the birth dates for marginal players from the early years. We believe we have dates for everyone born after 1940, but we are missing dates for about 25% of the players born before then. There is also a slight bias on each end of the batting average graph. Only the great players born in the 1860s were still playing after 1900, and only the best players born in the early 1970s are in the dataset.

Except for the tails of Figures 12 and 13, there is an clear increase in ability. The golf study supports Gould's conjecture. The best players are getting slightly better, but there are great players in every era. The median and 10th percentile are improving rapidly (see Fig. 15). The current 10th percentile player is almost 2 shots better than the 10th percentile fifty years ago. This explains why nobody dominates golf the way Hogan, Snead, and Nelson dominated in the 1940s and 1950s. The median player, and even the marginal player, can have a good tournament and win. Batting average exhibits a similar pattern. The best players are increasing in ability, but the 10th percentile is increasing faster than the 90th percentile (see Fig. 14). It appears as though batting averages have increased steadily, whereas golf is in a period of rapid growth.

These conclusions coincide with the histories of these sports. American sports are experiencing increasing diversity in the regions from which they draw players. The globalization has been less pronounced in MLB, where players

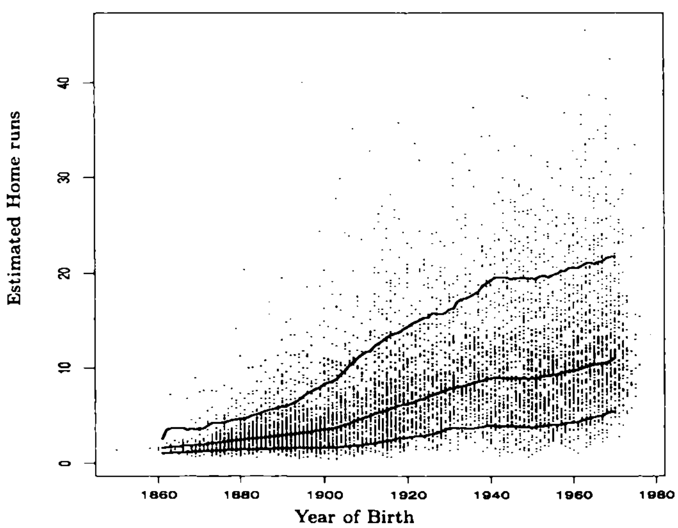

Figure 13. The Estimated Peak Home Run Performance of Each Player Plotted Against the Year They Were Born. The y-axis represents the mean number of home runs for each player, at their peak, if the year was 1996. The three curves are the smoothed 10th, 50th, and 90th percentiles.

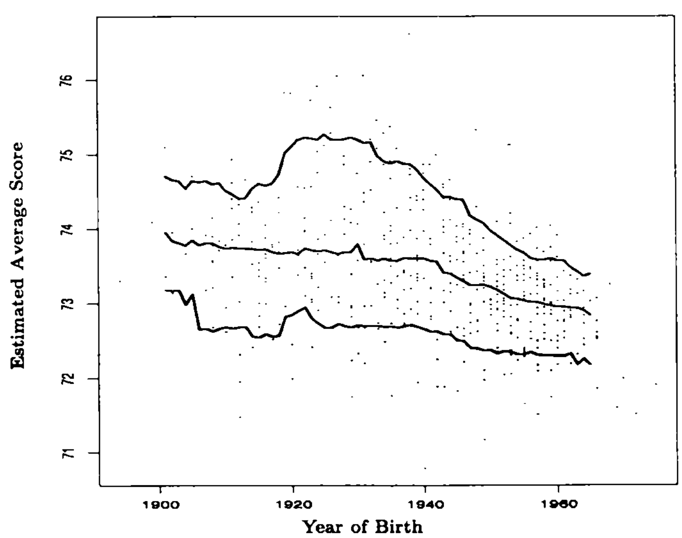

Figure 15. The Estimated Peak Scoring Performance of Each Player Plotted Against the Year They Were Born. The y-axis represents the mean score for each player, at their peak, if the year was 1997. The three curves are the smoothed 10th, 50th, and 90th percentiles.

are drawn mainly from the United States and other countries in the Americas. Baseball has remained fairly stable within the United States, where it has been an important part of the culture for more than a century. On the other hand, golf has experienced a huge recent boom throughout the world.

9. DISCUSSION

In this article we have developed a model for comparing players from different eras in their performance in three different sports. Overlapping careers in each sport provide a network of bridges from the past to the present.

For each sport we constructed additive models to account for various sources of error. The ability of each player, the difficulty of each year, and the effects of aging on performance were modeled. To account for different players aging differently, we used random curves to represent the individual aging effects. The changing population in each sport was modeled with separate hierarchical distributions for each of the decades.

Because of multiple sources of variation not accounted for in scoring, the model for the scoring ability of NHL players did not fit as well as the model in the other three studies. It still provided reasonable estimates, however, and the face validity of the results is very high. The different years in hockey play an important role in scoring. Career totals for individuals are greatly influenced by the era in which they played. Wayne Gretzky holds nearly every scoring record in hockey and yet we estimate him to be the second-best scorer of all time. The optimal age for a hockey player is 27, with a sharp decrease after age 30. A hockey player at age 34, the optimal golf age, is at only 75% of his peak value. Many of the greatest scorers of all time are playing now, NHL hockey has greatly expanded its talent pool in the last 20 years, and the number of great players has increased as well.

The golf model provided a very good fit, with results that are intuitively appealing. Players' abilities have increased substantially over time, and the golf data support Gould's conjecture. The best players in each era are comparable, but the median and below-average players are getting much better over time. The 10th percentile player has gotten about 2 shots better over the last 40 years. The optimal age for a professional golfer is 34, though the range 30–35 is nearly optimal. A golfer at age 20 is approximately equivalent to the same golfer at age 50—both are about 2 shots below their peak level. We found evidence that playing Augusta National now, with the equipment and conditions of today, is about 1 shot easier than playing it with the equipment and conditions of 1950. Evidence was also found that golf scores are not normal. The left tail of scores is slightly shorter than a normal distribution and the right tail slightly heavier than a normal distribution.

The baseball model fit very well. The ability of players to hit home runs has increased dramatically over the century. Many of the greatest home run hitters ever are playing now. Batting average does not have the same increase over the century. There is a gradual increase in the ability of players to hit for average, but the increase is not nearly as dramatic as for home runs. The distribution of batting average players lends good support to Gould's conjecture. The best players are increasing in ability, but the median and 10th percentile players are increasing faster over the century. It has gotten harder for players of a fixed ability to hit for average. This may be due to the increasing ability of pitchers.

Extensions of this work include collecting more complete data in hockey and golf. The aging curve could be extended to allow for different peak ages for the different players. Model selection could be used to address how the populations are changing over time—including continuously indexed hierarchical distributions.

[Received May 1998. Revised January 1999.]

REFERENCES

Albert, J. (1993), Comment on "A Statistical Analysis of Hitting Streaks in Baseball," by S. Albright, *Journal of the American Statistical Association*, 88, 1184–1188.

——— (1998), "The Homerun Hitting of Mike Schmidt," *Chance*, 11, 3–11.

Albright, S. (1993), "A Statistical Analysis of Hitting Streaks in Baseball," *Journal of the American Statistical Association*, 88, 1175–1183.

Berry, S., and Larkey, P. (1998), "The Effects of Age on the Performance of Professional Golfers," *Science and Golf III*, London: E & FN SPON.

Chib, S., and Greenberg, E. (1995), "Understanding the Metropolis–Hastings Algorithm," *The American Statistician*, 49, 327–335.

Draper, D., Gaver, D., Goel, P., Greenhouse, J., Hedges, L., Morris, C., Tucker, J., and Waternaux, C. (1992), *Combining Information: Statistical Issues and Opportunities for Research*, (Vol. 1 of *Contemporary Statistics*), Washington, DC: National Academy Press.

Gelfand, A., and Smith, A. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Gilks, W., Richardson, S., and Spiegelhalter, D. (Eds.) (1996), *Markov Chain Monte Carlo in Practice*, London: Chapman and Hall.

Gould, S. (1996), *Full House: The Spread of Excellence from Plato to Darwin*, New York: Three Rivers Press.

Hollander, Z. (Ed.) (1997), *Inside Sports Hockey*, Detroit: Visible Ink Press.

Jackson, D., and Mosurski, K. (1997), "Heavy Defeats in Tennis: Psychological Momentum or Random Effect?," *Chance*, 10, 27–34.

Larkey, P., Smith, R., and Kadane, J. (1989), "It's Okay to Believe in the Hot Hand," *Chance*, 2, 22–30.

Lindley, D., and Smith, A. (1972), "Bayes Estimates for the Linear Model," *Journal of the Royal Statistical Society*, Ser. B, 34, 1–41.

Mosteller, F., and Youtz, C. (1993), "Where Eagles Fly," *Chance*, 6, 37–42.

Myers, R. (1990), *Classical and Modern Regression With Applications*, Belmont, CA: Duxbury Press.

Riccio, L. (1994), "The Aging of a Great Player; Tom Watson's Play," *Science and Golf II*, London: E & FN SPON.

Scheid, F. (1990), "On the Normality and Independence of Golf Scores," *Science and Golf*, London: E & FN SPON.

Schell, M. (1999), *Baseball's All-Time Best Hitters*, Princeton, NJ: Princeton University Press.

Shi, M., Weiss, R., and Taylor, J. (1996), "An Analysis of Paediatric CD4 Counts for Acquired Immune Deficiency Syndrome Using Flexible Random Curves," *Applied Statistics*, 45, 151–163.

Stern, H. (1995), "Who's Hot and Who's Not," in *Proceedings of the Section on Statistics in Sports, American Statistical Association.*

Stern, H., and Morris, C. (1993), Comment on "A Statistical Analysis of Hitting Streaks in Baseball," by S. Albright, *Journal of the American Statistical Association*, 88, 1189–1194.

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions," *The Annals of Statistics*, 22, 1701–1762.

Tversky, A., and Gilovich, T. (1989a), "The Cold Facts About the "Hot Hand" in Basketball," *Chance*, 2, 16–21.

——— (1989b), "The "Hot Hand": Statistical Reality or Cognitive Illusion," *Chance*, 2, 31–34.