Factors Influencing Demand in Major League Baseball: Steroid Policy, Discrimination, and Uncertainty of Outcome

by

Scott A. Tainsky

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Kinesiology)
in The University of Michigan
2008

Doctoral Committee:

Assistant Professor Jason A. Winfree, Chair
Professor Charles C. Brown
Professor Rodney D. Fort
Assistant Professor Katherine M. Babiak

Table of Contents

List of Figures

Figure

List of Tables

Table

List of Abbreviations

CBA	Collective Bargaining Agreement
CPI	Consumer Price Index
CSU	Consecutive Season Uncertainty
FCI	Fan Cost Index
GLS	Generalized Least Squares
GU	Game Uncertainty
MLB	Major League Baseball
MLS	Major League Soccer
MRP	Marginal Revenue Product
MSA	Metropolitan Statistical Area
MVP	Most Valuable Player
NBA	National Basketball Association
NHL	National Hockey Association
OLS	Ordinary Least Squares
PED	Performance Enhancing Drugs
PU	Playoff Uncertainty
RGRC	Racial and Gender Report Card
RPI	Relative Power Index
SD	Standard Deviation
SOS	Strength of Schedule
SQL	Structured Query Language
THG	Tetrahydrogestrinone
TU	Team Uncertainty
UEFA	Union of European Football Associations
UOH	Uncertainty of Outcome

Abstract

The research presented in this dissertation represents three distinct types of issues influencing demand. Chapter II examines how a policy change affected revenues. The results show that there was a financial incentive to not enact and enforce a policy punishing users of performance enhancing drugs. Chapter III measures the change in demand for teams attributable to foreign-born players on the squad. The results show fans have changed their bias from anti- to pro-foreign players over the period from 1985-2005. Chapter IV questions how fans level of consumption is affected by the relative uncertainty of a game and season's outcome. The findings are that fans prefer to attend games when their team is in a better position to make the playoffs, when the game itself is important to determining playoff qualification, and when their team's standing is similar to that of the previous season. While each study represents a different subset of studies within the literature on sport economics, each is aimed specifically at learning more about fan preferences and measuring their impact on demand.

Chapter I

Introduction

Background

Demand for sporting events becomes an increasingly vital topic as the interest in professional sports continues to grow. In 2006, over 76 million fans attended regular-season Major League Baseball (MLB) games, over three times as many as had attended just 20 years earlier. At an average ticket price of $22 apiece, that comes to roughly $1.67 billion in ticket revenue alone. That figure does not even take into account ancillary local revenues such as parking and concessions. Academic work on demand has similarly flourished during that time. Studies of demand in sports cover game and yearly attendance, the impact of stars, team quality, facility, temporal factors, and work stoppages, uncertainty of outcome, competitive balance, excess demand for tickets, and the secondary ticket market, to name a few. As preferences are prone to change—as evidenced in sports by the recent decline in interest in the NHL, declining NBA television ratings, and emerging popularity of NASCAR—this line of inquiry promises to remain relevant to practitioners and scholars alike.

The work conducted for my dissertation comes in three distinct projects, each under the umbrella of estimating demand in a major North American professional sports league, MLB. Chapter II examines the introduction of a new league policy and estimates the related change in revenues. Specifically, it looks at Major League Baseball's performance-enhancing drug policy and shows how its institution impacted gate attendance. Chapter III applies tried and true methods for measuring consumer discrimination in sports on a previously unexplored group. There is a vast literature on race and gender discrimination in sports. This study's novelty is that it looks at country-of-origin as an input in estimating fan attendance at MLB games and attempts to find the source of that bias. Chapter IV explores game attendance in Major League Baseball and incorporates the most detailed metrics of uncertainty introduced in an academic work. In addition to quantifying all types of uncertainty, it applies the parameters to estimating demand.

Where the three aforementioned studies diverge is in the angle of inquiry taken in estimating owner revenues. The league can be thought of as the collection of its owners. So when the league, by way of the commissioner's office, institutes a new policy, it is the owners who have decided that the benefits of the new policy outweigh the costs. Such is the case of MLB's steroid policy. Although I can, and do, quantify the change in revenue for each owner, the decision-making unit is the collection of owners. In contrast to the work covered in Chapter II, player personnel decisions, the focus of Chapter III, rest in the hands of each individual team. While some leagues have historically restricted the number of international players, MLB has never done so. When considering the marginal change in revenue associated with hiring a foreign-born player, it would be prudent for that team's management to be aware of the implications. The league, as the collection of owners, is only concerned with this hire inasmuch as the collective bottom line is altered via revenue sharing. Once again, however, league policy comes into play in broaching whether there should be regulations in the signing of foreign players. This is somewhat akin to Chapter II in that the scheduling of contests falls under league commission. The subtle difference here is that scheduling, unlike the performance-enhancing drug policy, is one of the very basic duties of the league necessary in order to make play happen. The underlying question, therefore, is not whether or not to implement the policy, rather how to best implement scheduling to maximize revenues while taking into account other considerations such as fairness of play. For example, MLB attendance may be far greater in St. Louis than in many other host cities, but arranging additional home games for the Cardinals when home field provides a competitive advantage would contradict the charge of the league to create a fair regular season schedule. The other component of this chapter, uncertainty, closely resembles the steroid question in that it falls on the league to understand the implications of uncertainty in establishing league policy.

Analytical Techniques

The work that composes this dissertation use a multitude of well-established techniques to investigate the questions posed. Each of the chapters uses regression to predict the attendance at games under a given set of conditions, an accepted convention in the previous literature in setting a proxy for demand. I caution the reader to not overstate the effects found as these can only be confidently applied to the manner in which they influence attendance and not other revenues. Furthermore, this research focuses only on Major League Baseball. Thus, when statements are made regarding fan tendencies, I am referring to the tendencies of baseball fans and additional research would be necessary to draw similar conclusions about other leagues. The subsequent discussion details the additional methodology specific to that section, to be further expounded upon in each respective chapter.

Demand is estimated in Chapter II using runs as a predictor along with other factors known to affect attendance, the dependent variable. The list includes predictors of quality, expectations, population size, and indicator variables for strike years. All of these are placed into the model and regressed using Ordinary Least Squares (OLS) on the log of attendance. The logarithm is employed for ease of comparing percentages. Then, using log of runs as the dependent variable and by controlling for other factors known to affect scoring, such as park factor[1], it is possible to calculate the difference between expected and actual run production in the post-steroid era.

[1] Park factor (PF) indicates the difference between runs scored in a team's home and road games. It is calculated by the formula:

Demand for international players was estimated using a similar design. The notable methodological differences between this and the regression analysis mentioned above are the use of Generalized Least Squares (GLS) regression and a robust covariance matrix. GLS is used to control for heteroskedasticity and autocorrelation. The covariance matrix was used because there was a reasonable expectation that the error terms would be correlated within teams in different seasons. An additional metric was calculated in an effort to study the effect of matching team and market international populations. The matching figure was computed by summing the products of proportion of team and market in each season for each country.

Chapter IV tests the Uncertainty of Outcome Hypothesis (UOH). Broadly based on Simon Rottenberg's original contention, UOH posits that fans want their teams to have a reasonable opportunity to compete to win games and championships. Subsequent research divided the notions of

---

$$PF = \frac{\frac{homeRS + homeRA}{homeG}}{\frac{roadRS + roadRA}{roadG}}$$

where RS is runs scored, RA is runs against and home and road correspond to whether the team is playing at its home park or at a visiting park. Thus, park factors over 1 are said to be hitter friendly and those under 1 pitcher friendly.

game and championship uncertainty, later adding a third category of consecutive season uncertainty. I use game-by-game data to build a probit model to determine the probability of success in a given game in order to evaluate the uncertainty involved in that game and, by extrapolation, season. Game-by-game attendance was then employed as the dependent variable to evaluate the relative merit of several metrics measuring game, playoff, and consecutive season uncertainty controlling for other factors previously demonstrated to influence demand. Once again the log form of attendance is used.

Dissertation Layout

This first chapter provides an overview of each of the three studies that comprise my dissertation. Along with an introduction to each paper and an overview of basic methodology, a summary of the larger findings is presented. Chapter II presents MLB’s performance-enhancing drug policy’s impact on league revenues. Chapter III details the evolution of fan discrimination with respect to international players over the past 20 years. Chapter IV offers a short-run demand analysis of twelve MLB seasons and presents novel findings on the Uncertainty of Outcome Hypothesis. It also uses a probit model to estimate game and playoff uncertainty, and then tests metrics based on the probit results for significance in the model. A short conclusion is presented in Chapter V.

Summary of Findings

In Chapter II the model shows that implementing drug testing with genuine punitive repercussions had an impact on revenues in MLB. The first model demonstrates scoring as a significant predictor of gate attendance while the second quantifies the decreased run production attributable to steroid testing. Estimates of lost revenues for each team are then calculated.

Chapter III shows that fans' preferences vis-à-vis foreign-born players changed during the period studied, 1985-2005. At the outset there was evidence of discrimination against international players. This steadily decreased to the point that, all else equal, there was a preference for foreign players. This peaked right around the year 2000, but has been flattened out since. The lack of significance of the matching statistic suggests that the source of bias is not in the population of residents coming from those countries also represented on the teams.

Chapter IV covers game-by-game demand for MLB from 1996-2007. The research shows that when all aspects of uncertainty are incorporated into the same model, playoff uncertainty is the most influential on demand. The importance of the game in terms of playoff qualification is a significant indicator of attendance, as are the probability a team will make the playoffs and, interestingly, the change in performance from the same point in the previous season.

Chapter II

Financial Incentives and League Policy

Introduction

Sports leagues create policies for a variety of reasons. Some leagues have a salary cap or a luxury tax to increase competitive balance. Other leagues have ownership policies, presumably to ensure that owners' incentives are closely aligned with league incentives. Game rules are typically set for player safety and enjoyment of the fans. Most of these policies are somehow related to league revenues. One way leagues might increase demand is by creating more offense or scoring, since some fans want more scoring (Fort, 2006). Baseball experienced an offensive surge in the 1990s (see Figure 2.1) that could be attributed to a number of factors, among them the use of steroids (Bryant, 2004). It was not until Congress intervened on March 17, 2005 that Major League Baseball (MLB) stringently tested its players and punished the offenders. This work illustrates the financial incentives of not having a meaningful steroid policy in MLB from the time the league acknowledged their presence in the game until finally implementing a policy.

At least the first half of the old football adage “offense sells tickets, defense wins championships” seems to be true in baseball. Because it appeals to fans, increased offense, or run production, is desirable to MLB owners as it has been shown to, ***ceteris paribus***, boost attendance (Winfree, McCluskey, Mittelhammer, and Fort 2004; Domazlicky and Kerr 1990). If there was a link or even a perceived link between steroids and offense, then MLB owners had an incentive to not institute a steroid policy that would decrease steroid use. San Diego Padres general manager Kevin Towers admitted to knowing and feeling guilty about not speaking up about his third baseman and 1996 National League Most Valuable Player (MVP) Ken Caminiti’s steroid use. He further added his belief that general managers all across the league knew about it (steroid use), but did not say anything due in part to the success of the players (Towers in NY Times, 2005). Peter Magowan, managing general partner of the San Francisco Giants, is even more explicit about ownership’s complicity in his admission that steroid problems were overlooked for far too long (Magowan in Jenkins, 2005).

Steroids and other performance enhancing drugs have been a growing issue in many sports over the past two decades. Besides baseball, players have been tested positive for steroids in the National Football League (American football), Tour de France champions have failed blood tests, and a myriad of Olympic champions have failed doping tests. Sports such as football, rugby, and cricket too have also had players test positive for performance enhancing drugs. While I use MLB as a context, this study illustrates the incentive for leagues to not be vigilant in finding athletes who use steroids or other drugs. It may be the case that demand for the National Football League increases when athletes are bigger or that cycling and Olympic fans prefer to see records broken. If this is the case, there are benefits to owners and organizing committees to ignore certain athlete behavior. However, to my knowledge, most other sports league and organizations have been more proactive than MLB on this issue. Regardless, if society wants athletes to stop using performance enhancing drugs, league owners or organizations may not be the most vigilant. This may be especially important since many feel that performance enhancing drug use will continue at least into the near future (Lippi and Guidi 2004).

Although the impact of steroids on baseball performance is one that is widely discussed, to date there has been no conclusive study conducted to determine the relative impact of steroids on batting performance versus pitching performance. De Vany (2006) argues that steroids had no effect on offensive production in MLB. His analysis shows that the distribution of home runs hit by the players is non-normal. Curiously, he extrapolates from this that there is an infinite variance for the distribution of home runs hit by a player in a year. Therefore, even though nearly every offensive statistical category has increased in the steroid era, he argues that "averages signify nothing" (DeVany, 2006, p.23). Other researchers have established that "more players are hitting a higher number of home runs" (Yilmaz et al., 2001, p.181). My analysis shows that offense significantly decreased the same year the policy was put in place.

If steroids indeed had an effect on offensive production, the data will show an increase in a number of metrics. While this may be evident in measuring the increase of home runs, the anecdotal proxy of steroids' impact, there are limitations to measuring this one statistic alone. If batters were able to hit a baseball harder and further because of steroid use, home runs would be just one part of the offensive advantage gained over pitchers. Furthermore, previous literature cites runs, and not simply home runs, as increasing fan attendance (Winfree, et al., 2004, Domazlicky, 1990). Consequently, I have chosen runs rather than home runs as the critical statistic in gauging offensive output.

Gabriel Schechter (2005, as cited in Seeman, 2005), a researcher for the National Baseball Hall of Fame, attributed the rise in offense to several factors, including the relatively small dimensions of modern fields. I therefore control for park effects in predicting offense. Others maintain that expansion is part of the cause of increased offense due to a change in the talent distribution (Bradbury, 2007). However, expansion years prior to the 1990s seem to have no correlation with offensive production.

There exists a widespread perception that the net effect of steroids has been positive for batting and negative for pitching. A recent poll conducted by ***USA Today*** found that only 18% of major league baseball players felt that steroids were not a contributor to record performances in recent years (Jenkins, 2005). ***New York Times*** columnist Buzz Bissinger eloquently sums up the "greedy and feckless" owners' role in the steroid mess as "guilty of cynically jettisoning the game's subtlety and complexity to turn it into a slugfest circus" (Bissinger, 2005). Whether this is the case because more batters have used steroids, the product of usage is greater for batters, or some combination thereof is not the focus of this work. Likewise, a recent article in ***The Washington Post*** cites a number of baseball medical experts who take the position that the physiological benefit of steroids is far greater for batters than pitchers. "While drug experts largely agree that steroids can enhance any hitter's power and likely fueled the unprecedented home run surge in the 1990s, the only evidence to suggest that steroids have significantly affected pitching during the same time period is the injuries that have occurred…Frank Jobe, the longtime Los Angeles Dodgers team physician credited with the invention of ligament replacement surgery (Tommy John surgery), speculated that steroids could earn a pitcher perhaps an extra 2 mph on his fastball" (Shipley, 2006, p.E01). However, even if it were the case that steroids had nothing to do with the rise in offense, there is still a perceived incentive for the owners if they believe such an increase to be so. Further, although the literature on the effect of performance enhancing drugs is sparse due to their illegal nature, certain steroids have been shown to increase muscular strength by 5 to 20 percent (Hartgens and Kuipers 2004).

MLB’s Steroid Policy

On June 17, 1991 the MLB commissioner’s office put forth a memorandum acknowledging the harmful effects of steroids. Also in the document, in the section entitled “Major League Baseball’s Drug Policy,” it is stated that players’ use of steroids “are subject to discipline by the Commissioner and risk permanent expulsion from the game” (Vincent, 1991, p.1). No provision, however, was made concerning testing for steroids or any other banned substance for that matter. The lack of any specific repercussions for a violation of the memorandum is tantamount to baseball’s non-policy, one that would last until the middle of the next decade. Although the 2002 MLB Collective Bargaining Agreement (CBA) did address steroids, the player's union would, after public pressure, only allow steroid testing on a trial basis for 2003 (Staudohar, 2002). It is easy to see why players would choose to use the steroids, especially under pre-policy conditions. As Haugen (2004) shows, it is completely rational for players to cheat and use steroids when the perceived gains are high and the cost is low.

During the 2003 season, provisional testing was conducted in order to gauge the level of usage in baseball. Only anonymous tests were conducted, and only for the garden-variety steroid—designer steroids like Tetrahydrogestrinone (THG) were undetectable by the type of test administered. All players on the 40-man roster were tested in spring with an additional 240 players selected at random to undergo a second test. If less than 2.5% of the tests had come back positive over consecutive seasons, all testing would have ceased under the 2002 CBA. In practice, the usage rate was sufficiently high (between 5 and 7 percent of the 1438 tests administered to major league players came back positive) to warrant future testing under the CBA. This triggered further testing among MLB players for the 2004 season.

Each player was to be tested once over the course of the 2004 campaign. A positive test would result in the player entering treatment, followed by a 15-day suspension or up to $10,000 fine for a second positive test, increasing to up to a year suspension or fine up to $100,000 for a fifth positive test. In 2005, the list of banned substances was expanded to include not only steroids, but steroid precursors, designer steroids, diuretics, and masking agents. Unannounced mandatory testing of each player was conducted over the course of the season. There was further testing of randomly selected players as well as random testing during the off-season. The penalties established for an initial positive result was 10 days, increasing to 30 days, 60 days, and one year suspensions without pay for subsequent positive tests. Under the new policy in 2005, twelve players, including well-known players like Rafael Palmeiro, who testified in the 2005 Congressional hearings that he had never taken steroids, were suspended for ten games apiece for violating MLB's performance enhancing drug policy (Baseball Almanac, n.d.).

Although steroids may have helped the economic viability of MLB, many including West Virginia Senator Jay

Rockefeller suggest this is harmful to society (Rockefeller, 2005). The National Institute on Drug Abuse estimates that as of 2000, between 2.7% and 2.9% of young American adults had taken an anabolic-androgenic steroid at least once in their life. In 2004, United States President George W. Bush stated the following in his State of the Union address:

> To help children make right choices, they need good examples. Athletics play such an important role in our society, but unfortunately, some in professional sports are not setting much of an example. The use of performance-enhancing drugs like steroids in baseball, football, and other sports is dangerous, and it sends the wrong message -- that there are shortcuts to accomplishment, and that performance is more important than character. So tonight I call on team owners, union representatives, coaches, and players to take the lead, to send the right signal, to get tough, and to get rid of steroids now.

Presumably, the U.S. Congressional hearings took place in 2005 for similar reasons. However, I argue that the "team owners, union representatives, coaches, and players" President Bush referred to had a financial incentive to not ban steroids.

Previous Studies

Past studies have used attendance to estimate both short-run (Hill et al., 1982) and long-run demand (Schmidt and Berri, 2004) for MLB. Attendance has also been utilized to estimate player pay (Scully, 1974), the effects of roster turnover (Kahane and Shmanske, 1997), competitive balance (Schmidt and Berri, 2001), uncertainty of outcome (Knowles et al., 1992), and even the designated hitter (Domazlicky and Kerr, 1990).

There is also a vast amount of literature focusing on how increased revenue affects player pay. Scully (1974) showed that before MLB's reserve clause, players would earn roughly 10% of their marginal revenue product (MRP). This implies that if a player can increase a team's revenue by a certain amount, the player would receive about 10% of this increase. However, after the reserve clause, player pay increased dramatically. Krautmann (1997) estimated that journeymen were actually paid more than 100% of their MRP, while Zimbalist (1992) found that they were only given 60% of their MRP. Zimbalist also found that free agents were actually overpaid, while other Scully (1989) work shows that they are only paid about 28% of their MRP. Fort (2006) demonstrates that from 2000-2004, players received an average of 61% of total revenue. These studies illustrate that a percentage of any change in revenue will go to the players, while the rest will go to the owners and management. It follows that if steroids increase runs, and therefore attendance, owners and players both would benefit financially from the increased fan interest.

Model

Using regression analysis it is possible to determine the relationship between attendance and offensive production, namely runs, controlling for factors known to affect attendance, such as win percentage, price, market size, and the impact of the player strikes. Beyond these factors, I employ a trend variable to capture variability that could be attributable to aspects unrelated to offense. I then attempt to determine whether revenue was lost by team owners because of a decrease in offensive production.

To test whether the use of steroids has an effect on attendance, a regression of average yearly attendance by team on an offense-based variable, *Runs*, and a set of control variables is employed. The attendance data for MLB covers the period from 1992-2005. The years were chosen to correspond to those following the commissioner's memorandum. Demand was estimated using the following equation:

$$\ln(attendance) = \beta_0 + \beta_1 \ln(lagattendance) + \beta_2 win\% + \beta_3 aswin\% + \beta_4 \ln(fci) + \beta_5 \ln(pop) + \beta_6 trend + \beta_7 stad + \beta_8 strike94 + \beta_9 strike95 + \beta_{10} NHLstrike05 + \beta_{11} \ln(runs) + \varepsilon \tag{1}$$

where ***Attendance*** is per game season attendance, by team, for each team that competed in the league in a particular season taken from baseballreference.com. The log function was employed to calculate the percentage change in attendance. ***Lagattendance*** is per game attendance in the previous year, by team, for each team that competed in MLB. Once again, the log function was utilized. ***Win%*** is the team's winning percentage at the end of the year. The sign of this variable is expected to be positive because, simply put, fans want to see their home team win. ***Aswin%*** is the team's winning percentage at the all-star break, which is approximately half way through the season. The sign of this variable is also expected to be positive. The use of the ***Aswin%*** variable in addition to the ***Win%*** variable is important for the following reason. Fans like to see their team win or at least contend for championships. It is sometimes the case in MLB that a team makes a strong showing late in the season but never really contends for a championship. Employing the ***Aswin%*** variable captures the effect of a team remaining in the hunt for a playoff berth.

Whitney (1988) had success using a similar design, in that case employing monthly winning percent variables and the number of games behind in the standings. This specification helps capture the overall effect being in contention may have, but uses a less elaborate design because it was not the focus of this analysis. All of the aforementioned variables were collected from baseball-reference.com. ***FCI*** is the fan cost index representing the real price (2005 dollars) of a family of four attending a game. The fan cost index data were collected from teammarketing.com and includes four average-price tickets, four small soft drinks, two small beers, four hot dogs, two game programs, parking, and two adult-size caps. The ***FCI*** is expected to be negative. ***Pop*** is the 2003 estimate of city population (in millions) according to United States Census Bureau. For Canadian cities, the 2001 Census of Population estimates are used. According to fundamental economic principle, larger markets should, above and beyond other factors, draw larger crowds. As such, the sign of this variable is anticipated to be positive. ***Trend*** is a trend variable showing how attendance is changing over time. A cursory glimpse at attendance data across all major American sports shows rising totals. The T***rend*** variable accounts for this development in baseball and is expected to be positive. S*tad* is an indicator variable equaling one if the team had a new stadium that year. New facilities have been shown to increase attendance (Quirk and Fort, 1997). Therefore, ***Stad*** is predicted to be positive. S***trike94*** and ***Strike95*** are indicator variables for strike years. Major League Baseball has a history of protracted labor disputes. Empirical evidence shows that each of these incidents has had a demonstrative effect on short-term demand (Schmidt and Berri, 2002). The dummy variables capture the variability that may stem from the strike independent of the other factors being measured. Accordingly, the ***Strike94*** and ***Strike95*** effects are anticipated to be negative. ***2005season*** is also an indicator variable for the 2005 season, which may have benefited from the absence of National Hockey League (NHL) games being played. The 2005-06 NHL lockout was the first time a major sports league lost an entire season due to a work stoppage and has been shown to affect other sports (Winfree, 2007). While work stoppages do not typically have a statistically significant impact on other sports, Schmidt and Berri (2004) showed that the 1987 NFL strike indeed had a slight impact on MLB attendance. Consequently, the direction of this effect is expected to be positive. ***Runs*** represents the average runs per game for each team during each of the seasons examined and is the major variable of interest in the study. The effect is anticipated to be positive. Although ***Runs, Win%,*** and ***ASWin%*** are all highly correlated ($\rho(Runs, Win\%) = .535$, $\rho(Runs, ASWin\%) = .446$, and $\rho(Win\%, ASWin\%) = .871$), the more runs a team scores the more likely they are to win, I am not interested in the effect that ***Runs*** has through winning percentage. Obviously, only one team can win a given game, therefore the average winning percentage for the league does not change year to year. I am only interested in the effect that ***Runs*** has on attendance apart from winning percentage, so winning percentage is controlled for. However, because of the collinearity of these variables, three regressions are run—one without ***Win%***, one without ***ASWin%***, and the complete model. This is done to check the consistency of the parameter estimate of $\ln(Runs)$.

The change in attendance associated with runs was then utilized to estimate the change in revenue experienced by owners during the 2005 season as a result of the new testing policy. To do so I estimate decreased run production during the 2005 season that might be attributable to steroids. I then used these data to gauge what effect that drop off had on attendance based on actual 2005 attendance figures. Multiplying those figures by the fan cost index yields the revenue lost by the typical MLB owner.

Next I use a second regression analysis to determine the portion of this loss attributable to steroids. In 2005 runs decreased by 4.6%, but this does not control for any new stadiums or trends. Figure 2.1 also shows the dip in runs in 2005, however the drop off certainly does not stand out. The data for the second regression date from 1992-2005 to match the first regression. The specification is as follows,

$$\ln(runs) = \gamma_0 + \gamma_1 trend + \gamma_2 trend^2 + \gamma_{3-47} stadium + \gamma_{48} 2005 + \varepsilon \qquad (2)$$

where ***2005*** is an indicator variable that equals one for 2005 when the steroid policy was put in place. ***Stadium*** represents a matrix of forty-five indicator variables for each stadium to control for different offensive effects in different stadiums.

Once the regressions are estimated, it is then possible to calculate the estimate of lost revenue due to the decrease in offense. If we assume that stadium revenue equals the product of attendance and the FCI, then the estimated loss in revenue can be calculated by the following equation:

$$\Delta \text{Re}venue_i = \frac{81}{4} FCI_i * e^{X_i \beta}\left[e^{\gamma_{48}\beta_{11}} - 1\right] \tag{3}$$

Where X represents the data matrix in equation 1. The left hand side of the equation is multiplied by 81/4 since each team plays 81 home games in a full season and the FCI is the price of a game for 4 people. This equation illustrates the estimated stadium revenue for team i in 2005 subtracted from the estimated stadium revenue that would have been brought in without a policy.

Results

Descriptive statistics for the data are given in Table 2.1. The regression coefficients for both equation (1) and equation (2) were estimated using ordinary least squares (OLS). All explanatory variables were entered into the model simultaneously. Results from the regression estimating the effect of run production on attendance are given in Table 2.2. Most variables—ln*(Lagattendance)* , *Win%*, *Aswin%*, ln*(Pop)*, *Trend*, *Stad*, *Strike95*, and *2005season*—are of the expected sign and statistically significant, with *ln(FCI)*, *Trend* and *Strike94* being exceptions. The main variable of interest ln(*Runs*) was found to significantly increase attendance even after accounting for these other factors. A 1% increase in runs generates a 0.15% increase in attendance with the complete model. The other two demand regressions show a relatively consistent parameter estimate for $\ln(runs)$. Results from the regression measuring the change in runs after the enactment of the league's anti-doping policy are shown in Table 2.3. These results show that runs were 6.8% below the expected run production without the steroid policy. The 6.8% decline in runs scored during the 2005 season, due to the steroid policy, implies a 1.03% decrease in attendance.

Table 2.4 shows what a 1.03% decrease in attendance means for each team in terms of revenue and fans per game. The average club drew 2,504,682 in 2005. A 1.03% decrease represents 25,798 seats per year (318 per game) that went unoccupied because of the decrease in run production. Multiplying those numbers by the fan cost index, the league as a whole lost about $31.2 million in stadium revenue from the steroid policy. It is also important to note that regular season stadium revenue typically accounts for approximately half of all revenue for the team. Therefore, if other revenues behave in a similar fashion, MLB may have lost over $60 million in 2005 due to the steroid policy, money that would have been shared by players and owners. This estimation is conservative given MLB’s own appraisal of the value of the 1998 home run record chase between Mark McGwire and Sammy Sosa. Baseball estimated the financial windfall from the race to break the record at $1.5 billion (Jenkins, 2005). In many ways McGwire has become a symbol of the steroid era, formerly ***Time*** magazine’s Hero of the Year, the ex-slugger’s damaging testimony during the Congressional hearings on steroids certainly played a part in his receiving less than one-third of the votes necessary to be inducted by the Baseball Writers’ Association of America into baseball’s Hall of Fame (Baseball Hall of Fame, 2007) despite sparkling credentials. The league’s appraisal coupled with speculation on McGwire’s involvement with performance enhancing drugs further reinforces the financial disincentive to enact change in doping policy.

Discussion

This illustrates how changes in revenue from league policy can be estimated. Just as in the case of revenue sharing, luxury tax, and salary cap policies, I argue that is also important to look at league and team revenues regarding the steroid policy. The regression estimates indicate that attendance at MLB games was strongly influenced by the increased offensive production experienced during the 1990s and the first half of the current decade. Some other possible explanations for the recent decrease in offense in MLB have been accounted for in the methods. Although the point of this research is not to prove that steroids have increased offense, only that the perception of their use was ignored, the statistics seem to confirm that any perceived increase in offense from steroid use may be correct. Furthermore, it is unlikely that other exogenous factors decreased offense in 2005 given that expected error term in the model is zero for 2005. It also stands to reason that if any other changes coincided with the policy change, baseball officials would have implemented ones to counterbalance the expected decrease in runs. As such, these estimates may, in fact, be underestimates of the true value given reports of MLB's use of means to increase offense such as the use of more tightly wound baseballs in 2005, widely believed to benefit hitters (Ocker, 2006). These findings might help illustrate why both the player's association and owners were unwilling to change their steroid policy until Congress stepped in. MLB may have been gaining over $31 million a year in stadium revenue alone by not controlling steroid use among the players.

The results of this research imply that there are many other questions regarding MLB's steroid policy. Future research might examine how much of this extra revenue went to players and how much went to management or owners. Additionally, as performance enhancing drugs continue to be an issue in other sports like cycling, track and field, horse racing, and football, research gauging the impact of enforcement in these sports too is critical. Representative Tom Davis underscores the importance of vigilant testing beyond the fundamental objective of a level playing field. "We need to understand the dangerous cycle that perception creates. College athletes believe they have to consider steroids if they're going to make the pros; high school athletes, in turn, think steroids are the key to getting a scholarship" (Davis in Jenkins, 2005). This dangerous cycle created by the mere perception of performance enhancing drugs' prevalence in sports is no less applicable to other sports and in other countries. It is also worth looking further into whether college and high school athletes are truly influenced by the perception of performance enhancing drug use in professional sports. It would seem as though future research is warranted in this area as performance enhancing drug use continues to be an important issue in sports.

Table 2.1

Descriptive Statistics for Steroid Study

Parameter	**N**	**Min**	**Max**	**Mean**	**Standard Deviation**
Attendance	401	7935	57570	28657	9210
Lagattendance	401	7935	57570	28570	9246
Win%	401	0.265	0.716	0.501	0.073
Aswin%	401	0.272	0.753	0.501	0.077
Fci	401	83.6	276.2	140.9	28.0
Pop	401	0.317	8.086	1.481	2.008
Trend	401	1	14	7.666	4.000
Stad	401	0	1	0.035	0.184
Strike94	401	0	1	0.070	0.255
Strike95	401	0	1	0.070	0.255
2005season	401	0	1	0.072	0.259
Runs	401	3.380	6.230	4.788	0.536

Table 2.2

Ordinary Least Squares Regression Results

Estimating Log of Attendance per Game (Yearly Data) for MLB

	Model 1			Model 2			Model 3		
Parameter	**Parameter Estimate**	**t-statistic**	**p-value**	**Parameter Estimate**	**t-statistic**	**p-value**	**Parameter Estimate**	**t-statistic**	**p-value**
Constant	1.366***	5.707	<0.001	1.394***	5.741	<0.001	1.274***	5.349	<0.001
Ln(lagattendance)	0.784***	33.184	<0.001	0.787***	32.814	<0.001	0.788***	33.210	<0.001
Win%	0.497**	2.539	0.011	1.077***	9.585	<0.001			
Aswin%	0.612***	3.591	<0.001				0.968***	9.987	<0.001
Ln(fci)	0.017	0.314	0.754	0.014	0.245	0.806	0.023	0.416	0.677
Ln(pop)	0.015**	2.029	0.042	0.014*	1.875	0.061	0.017**	2.224	0.026
Trend	-0.004	-1.483	0.138	-0.003	-1.410	0.159	-0.004*	-1.718	0.086
Stad	0.284***	7.937	<0.001	0.286***	7.875	<0.001	0.281***	7.793	<0.001
Strike94	-0.023	-0.841	0.400	-0.023	-0.819	0.413	-0.028	-0.999	0.318
Strike95	-0.253***	-9.329	<0.001	-0.254***	-9.205	<0.001	-0.256***	-9.379	<0.001
2005season	0.029	1.021	0.307	0.026	0.904	0.366	0.034	1.192	0.233
Ln(runs)	0.151**	2.100	0.036	0.137*	1.877	0.060	0.212***	3.113	0.002
R^2	.863			.859			.861		
N	401			401			401		

* significant at p < .10
** significant at p < .05
*** significant at p < .01

Table 2.3

Ordinary Least Squares Regression Results

Estimating Log of Runs per Game (Yearly Data) for MLB

Parameter	Parameter Estimate	t-statistic	p-value
Constant	1.436	52.886***	<0.001
Trend	0.005	3.268***	0.001
AnaStadium	0.081	2.245**	0.025
BankOneBP	0.075	1.645	0.101
AtlFCStadium	0.072	1.470	0.142
TurnerField	0.116	2.899***	0.004
OriolePark	0.108	3.077***	0.002
FenwayPark	0.165	4.669***	<0.001
Wrigley	0.045	1.271	0.205
NewComiskey	0.163	4.611***	<0.001
RiverfrontStadium	0.079	2.106**	0.036
GreatAmBP	0.058	0.962	0.337
CleMemStadium	0.062	0.879	0.380
JacobsField	0.214	5.817***	<0.001
MileHigh	0.139	1.437	0.152
CoorsField	0.206	5.473***	<0.001
TigersStadium	0.133	3.213***	0.001
ComericaPark	-0.011	-0.245	0.807
ProPlryStadium	0.011	0.299	0.765
Astrodome	0.108	2.604***	0.010
MinuteMdPark	0.118	2.572**	0.011
Kauffman	0.059	1.685*	0.093
MilFCStadium	0.094	2.340**	0.020
MillerPark	-0.034	-0.698	0.486
Metrodome	0.083	2.348**	0.019
OlympicStadium	0.001	0.028	0.978
SheaStadium	0.013	0.373	0.710
YankeeStadium	0.215	6.102***	<0.001
OaklandCol	0.140	3.973***	<0.001
VeteransStadium	0.041	1.104	0.270
CitizensBkPark	0.156	2.194**	0.029
3RiversStadium	0.031	0.775	0.439
PNCPark	-0.042	-0.849	0.396
JackMurphyStadium	0.019	0.505	0.614
PetcoPark	0.032	0.443	0.658
CandlestickPark	0.080	1.918*	0.056

SBCPark	0.104	2.263**	0.024
Kingdome	0.189	4.350***	<0.001
SafecoField	0.129	2.980***	0.003
BuschStadium	0.095	2.687***	0.008
TropicanaField	0.012	0.267	0.789
OldArlington	0.095	1.342	0.180
BPatArlington	0.203	5.538***	<0.001
Skydome	0.117	3.307***	0.001
2005Season	-0.068	-3.264***	0.001

$R^2 = 0.389$

N = 401

* Significant at $p < 0.10$
** Significant at $p < 0.05$
*** Significant at $p < 0.01$

Table 2.4

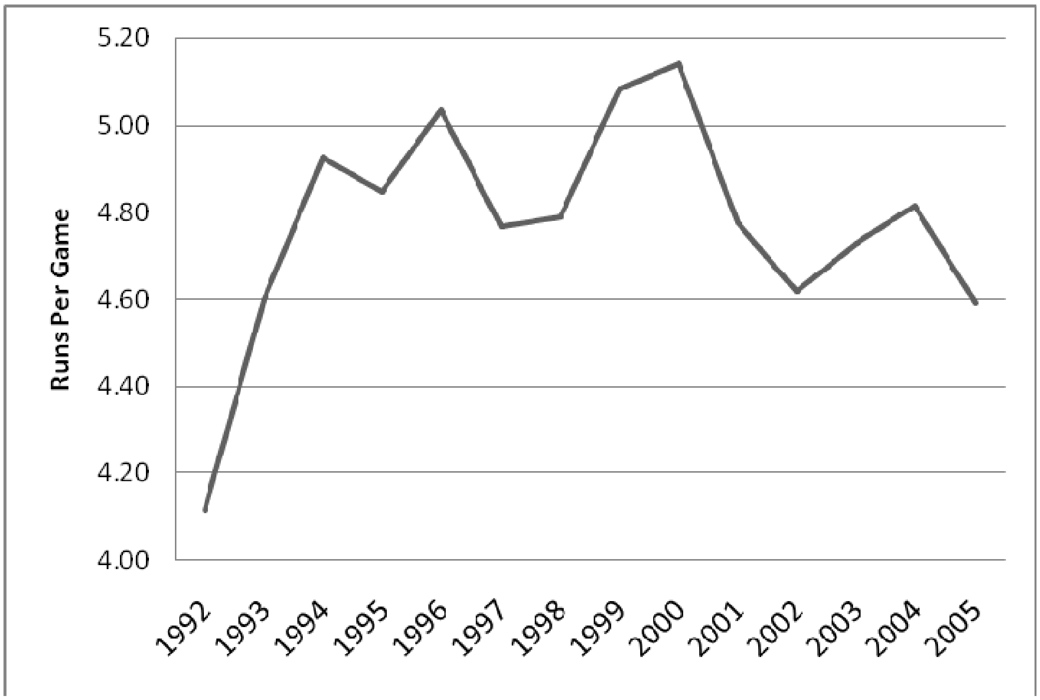
Estimated Change in Attendance and Gate Revenue from Decrease in Offensive Production

Team	Change in Attendance per game	2005 Fan Cost Index	Change in Yearly Stadium Revenue
ANAHEIM ANGELS	431	$125.78	$1,098,767
ARIZONA DIAMONDBACKS	307	$145.97	$906,094
ATLANTA BRAVES	320	$145.55	$943,577
BALTIMORE ORIOLES	339	$158.62	$1,089,275
BOSTON REDSOX	389	$276.24	$2,173,788
CHICAGO CUBS	380	$210.01	$1,615,617
CHICAGO WHITE SOX	311	$188.07	$1,186,295
CINCINNATTI REDS	268	$145.10	$786,092
CLEVELAND INDIANS	259	$156.18	$817,598
COLORADO ROCKIES	264	$141.68	$756,112
DETROIT TIGERS	247	$157.93	$789,303
FLORIDA MARLINS	240	$147.04	$713,959
HOUSTON ASTROS	382	$182.64	$1,413,015
KANSAS CITY ROYALS	191	$119.85	$463,909
LOS ANGELES DODGERS	389	$158.98	$1,250,832
MILWAUKEE BREWERS	263	$130.96	$697,059
MINNESOTA TWINS	257	$146.49	$762,857
NEW TORK METS	310	$185.13	$1,163,410
NEW YORK YANKEES	503	$193.86	$1,974,010
OAKLAND ATHLETICS	289	$152.64	$893,417
PHILADELPHIA PHILLIES	408	$189.31	$1,564,134
PITTSBURGH PIRATES	210	$143.31	$608,912
SAN DIEGO PADRES	374	$176.32	$1,335,172
SAN FRANCISCO GIANTS	355	$191.37	$1,376,966
SEATTLE MARINERS	328	$172.03	$1,143,017
ST. LOUIS CARDINALS	422	$177.66	$1,517,542
TAMPA BAY DEVIL RAYS	159	$143.81	$463,187
TEXAS RANGERS	335	$136.14	$922,946
TORONTO BLUE JAYS	253	$164.53	$842,741
Total			$31,269,603

The Washington Nationals were eliminated from the sample

Figure 2.1

Average Runs per Game for MLB



References

Baseball Almanac (n.d.). Retrieved on 05/13/07 from http://www.baseball-almanac.com/legendary/steroids_baseball.shtml

Baseball Hall of Fame (2007). Retrieved on 05/19/07 from http://www.baseballhalloffame.org/news/2007/election/results.htm

Baseball-Reference (n.d.). Last viewed on 07/24/08 at http://www.baseball-reference.com/

Bissinger, B. Home Runs Wanted. No Questions Asked. Retrieved on 05/28/07 from http://query.nytimes.com/gst/fullpage.html?res=9E00E4DD1730F936A35756C0A9639C8B63&sec=health&spon=&pagewanted=all

Bradbury, J. What Really Ruined Baseball. Retrieved on 04/06/07 from http://www.nytimes.com/2007/04/02/opinion/02bradbury.html

Bryant, H. A Tainted Era? Guilty or Not Players are Feeling Fans' Skepticism as Offensive Numbers Escalate. Retrieved on 04/06/07 from http://www.naturalstrength.com/steroids/detail.asp?ArticleID=1053

Bush, G. State of the Union Address. Retrieved on 04/06/07 from http://www.whitehouse.gov/news/releases/2004/01/20040120-7.html

De Vany, A. Steroids, Home Runs, and the Law of Genius. ***Economic Inquiry*** (forthcoming).

Domazlicky, B.R., and Kerr, P.M. (1990). Baseball Attendance and the Designated Hitter. ***American Economist*** 34(Spring):62-68.

Fort, R.D. (2006). ***Sports Economic,*** 2nd ed. Upper Saddle River, New Jersey: Pearson Prentice Hall.

Hartgens, F. and H. Kuipers. (2004). Effects of androgenic-anabolic steroids in athletes. ***Sports Medicine*** 34 pp. 513-54.

Haugen, Kjetel K. (2004). The performance-enhancing drug game. ***Journal of Sports Economics*** 5(1) pp. 67-86.

Hill, J.R., Madura, J. and Zuber, R.A. (1982). The short run demand for Major League Baseball. ***Atlantic Economic Journal*** 10, 31-35.

Jenkins, C. Players Admit Steroids Changed Baseball. Retrieved on 5/28/07 from http://www.usatoday.com/sports/baseball/2005-03-15-steroids-mlb-cover_x.htm

Jenkins, L. Baseball; Perfect Season to Perfect Storm: Steroids Taint McGwire Legacy. Retrieved on 5/28/07 from http://query.nytimes.com/gst/fullpage.html?sec=health&res=9F05E2DA123CF936A25750C0A9639C8B63

Kahane, L., and Shmanske, S. (1997). Team roster turnover and attendance in Major League Baseball. ***Applied Economics,*** 29, 425-431.

Knowles, G., Sherony, K., and Haupert, M. (1992). The Demand for Major League Baseball: A Test of the Uncertainty of Outcome Hypothesis. ***American Economist*** 36(Fall):72-80.

Krautmann, A. (1999). What's Wrong With Scully-Estimates of a Player's Marginal Revenue Product? ***Economic Inquiry*** 37(2):369-382.

Lippi, G. and Guidi, G.C. (2004) Gene Manipulation and Improvement of Athletic Performances: New Strategies in Blood Doping. ***British Journal of Sports Medicine,*** 38, 641.

National Institute on Drug Abuse. About anabolic steroid abuse. ***NIDA Notes.*** Aug 2000;15:15.

Ocker, S. (2006). Retieved on October 11, 2006 from http://www.philly.com/mld/beaconjournal/sports/baseball/14516394.htm

Quirk, J., and R. D. Fort. (1997). *Pay Dirt: The Business of Professional Team Sports*, 2nd ed. Princeton, NJ: Princeton University Press, 538.

Rockefeller, J (2005). Rockefeller Demands Tougher Steroids Policy for Sake of Young Fans Who See Athletes as Role Models. Retrieved on 04/06/07 from http://www.senate.gov/~rockefeller/news/2005/pr092805.html

Schmidt, M. B., and Berri, D. (2001). Competitive Balance and Attendance: The Case of Major League Baseball. *Journal of Sports Economics* 2(No. 2, May):145-167.

Schmidt, M. B. and Berri, D. (2002). The Impact of the 1981 and 1994-1995 Strikes on Major League Baseball Attendance: A Time-Series Analysis. *Applied Economics,* 34, 471-478.

Schmidt, M. B. and Berri, D. (2004). The Impact of Labor Strikes on Consumer Demand: An Application to Professional Sports. *American Economic Review*, 94(1): 344-357.

Scully, G.W. (1974). Pay and performance in Major League Baseball. *American Economic Review*, 64, 915-930.

Scully, G. W. (1989). *The Business of Major League Baseball*, Chicago: University of Chicago Press.

Seeman, B. Truth About Steroids Unlikely to Show Up in Box Scores. Retrieved on 04/06/07 from http://newhouse.live.advance.net/archive/seeman031005.html

Shipley, A. (2006, April 29). Do Steroids Give a Shot in the Arm: Benefits for Pitchers are Questionable [electronic version]. *Washington* Post, E01. Retrieved on May 11, 2006 from http://www.washingtonpost.com/wp-dyn/content/article/2006/04/29/AR2006042901195_pf.html

Staudohar, P. D. (2002). Baseball Negotiations: A New Agreement. *Monthly Labor Review*, December, 15-22.

Team Marketing (n.d.) Last viewed on 07/24/08 at http://www.teammarketing.com/fancost/mlb/

Towers, K. (2005). Retrieved on October 12, 2006 from http://query.nytimes.com/gst/fullpage.html?sec=health&res=9F04E4DD123DF932A35750C0A9639C8B63

Vincent, F. (1991). *Baseball's Drug Policy and Prevention Program*. Retrieved on 1/17/06 from http://www.businessofbaseball.com/docs/1991Memo_Baseballs_Drug_Policy_And_Prevention_Program.pdf

Whitney, J.D. (1988). Winning Games Versus Winning Championships: The Economics of Fan Interest and Team Performance. *Economic Inquiry*, 26, 703-724.

Winfree, J. (2007). Owners Incentives during the 2004-05 NHL Lockout. Working Paper.

Winfree, J., McCluskey, J., Mittelhammer, R. and Fort, R. (2004). Location and Attendance in Major League Baseball. *Applied Economics* 36(19): 2117-2124.

Yilmaz, M, S Chatterjee and M Habibullah. (2001). Improvement by Spreading the Wealth: The Case of Home Runs in Major League Baseball. *Journal of Sports Economics* 2(2): 181-193.

Zimbalist, A. (1992) *Baseball and Billions*, New York: Basic Books.

Chapter III

Discrimination and Demand: The Effect of International Players on Attendance in Major League Baseball

Introduction

Thirty-three years after Roberto Clemente, MLB's first foreign-born superstar, was inducted into the Hall of Fame, most pundits favored the United States or one of the Latin powerhouses, Puerto Rico, Dominican Republic, or Venezuela, to win the inaugural 2006 World Baseball Classic (WBC). It was an Asian team, Japan, however, that emerged victorious. This makes two points clear—the proficiency of Hispanic players is well-accepted and America's national pastime has gone global. As Figure 3.1 shows, over the last 20 years, there has been a substantial rise in the percentage of international players in MLB. As of 1985, just over ten percent of MLB players were born outside of the United States. In 2005, over one-fourth of players were foreign born. Experts have no choice but to recognize the international presence in American professional sports and even the need for sports to orient themselves toward a global fan base (Liefer, 1995). Even so, while there has been a wealth of research on racial discrimination in professional sports, the influence of nationality has yet to be studied to the degree of race or gender in spite of the fact that foreign participation in MLB is chronologically deep and geographically wide (Osborne, 2006). This study attempts to take one step towards measuring the impact of foreign players on demand for professional baseball in the United States. Specifically, it measures how the presence of international players affects attendance. Further, I evaluate the source of consumer discrimination being expressed.

MLB is different from some sports leagues in that it does not seem to discourage international players. Major League Soccer (MLS) has a quota on foreign players, apparently to promote the development and recognition of American talent. As of 2006, the Canadian Football League required that at least half of the players be Canadian. European football (soccer) leagues also have tried to limit the number of international players. For example, in 1991, the Union of European Football Associations (UEFA) limited teams to playing three international players plus two other international players with domestic playing experience for select games. Similar league policies have caused problems recently because of the European Union's law on the free movement of labor (Leeds and Von Allmen, 2007). Presumably these league policies are put in place to increase the visibility of domestic players at risk of becoming a minority. This, however, is not the case in MLB, as I will point out that in our data sample there are few teams fielding a majority of foreign players.

To begin with, the initial purchasing of playing talent in MLB varies based on a player's background. Players born in the US or Canada first enter the league via a first-year-player draft. Those completing high school or their junior or senior season at a four-year college are draft-eligible. Residents of US territories and players competing in high school baseball in the US also enter the league through the draft. Teams select in a modified reverse order of finish from the previous season whereby the team with the fewest wins from the American and National leagues respectively alternate selecting first. By contrast, MLB does not have any specific provisions regulating the signing of foreign-born players. Consequently, international players are free agents until they choose to sign traditional or developmental contracts with MLB teams. There are no restrictions on the number of foreign-born players signed by an organization or the number competing for the major league club. This chapter raises the question whether MLB's policy with respect to foreign-born players is a boon to league finances or, alternatively, whether policies implemented by other leagues would translate into greater revenues.

Relevant Literature

Becker (1971) contended that discrimination originates from three potential sources—employer, coworker, and customer preferences. Since customer preferences cannot be modified by market forces, teams attuned to their fan base's preferences stand to benefit financially. This is particularly important as, although player productivity is stochastic (Krautmann, 1990), consumer preferences for certain types of players can be ascertained and both play a role in a player's marginal contribution to revenue. Accordingly, when management employs race-neutral hiring practices with respect to players, they may do so at the expense of profits. Burdekin's (1991 and 2005) articles applied Becker to National Basketball Association (NBA) fans and showed that the racial composition of teams and their metropolitan markets to be positively correlated. Subsequent research (Hoang and Rascher, 1999) found that matching team racial composition with the population of the team's market area increases attendance, although McCormick and Tollison (2001) using data from the 1980s found no such relationship.

Customer prejudice in baseball was first examined in the 1970s. Over the years, the impact of race on attendance has been studied extensively, though the results are ambiguous. Gwartney and Haworth (1974) studied the decade immediately following integration and saw that African-American players increased home attendance. They showed that the inclusion of black players on a club increased team quality, measured by games won, and also brought additional customers to the park independent of winning. It is worth noting that later research showed a statistically significant relationship between winning and the presence of black players in the starting lineup in the early 1950s (Hanssen, 1998). By the 1960s, evidence showed the impact of African Americans on attendance had diminished (Scully, 1973). Examining data from the 1970s, there was no evidence that racial composition had any impact on revenues among the first free agent cohort (Sommers and Quinton, 1982). While the aforementioned articles look at customer discrimination as it relates to revenues, other studies introduce alternative sources of discrimination. Bodvarsson and Brastow (1999) argue that employer discrimination was at the source, and discriminatory practices were reduced when monopsonistic power diminished in the late 1980s and early 1990s in the NBA.

Still other studies examine the relationship between race and wage disparity. Kahn (1991) determined that the wage gap between black and white basketball players favored whites by 11 to 25 percent. Referring back to Becker, no entry effect was found to be present (Brown, Spiro, and Keenan, 1991; Kahn and Sherer, 1988), although wage differential was shown to be significant. Kahn (1991) asserted that much turnover (exit) in sports is involuntary given the high incomes earned by professional athletes. Scully (1973) suggested that African-Americans faced retention barriers in MLB. Similarly, Jibou (1988) found that black baseball players had higher exit rates than whites between 1971 and 1985 and Johnson and Marple (1973) found that white reserve players had longer careers than black benchwarmers. Bodvarsson and Partridge (1999) add coworker discrimination as a potential source of variance in salary among divergent groups. Longley (1995 and 2000) showed French Canadians playing hockey in English Canada suffered from salary discrimination and were underrepresented on English Canadian teams relative to US-based teams.

Similarly, country-of-origin has emerged as an important input effect on consumer behavior. Early scholarship on the subject demonstrated that consumers use country-of-origin information to evaluate products (Han, 1989; Hong and Wyer, 1989; Johansson, 1989). Experts and novices both utilize country-of-origin information in evaluations when attribute information is ambiguous (Maheswaran, 1994). Applied to sports, Osborne (2006) tested whether there has been sustained specialization among players from a given foreign country and found Canada’s trend to produce pitching and that of Puerto Rico and Venezuela to produce offense. Pedace’s (2007) paper on English Professional Soccer uses a market test approach to evaluate for the presence of nationality discrimination by estimating the effect of team nationality composition on attendance. He finds owners may benefit from increased attendance with more South American players. No similar studies, however, have been undertaken to assess the presence of MLB fan discrimination relating to nationality.

Hypotheses

These studies have a direct and important application in the world of sports. With the wealth of information freely available it is possible to gauge if there is customer discrimination vis-à-vis international players as well as identifying the source of this bias. I test the following two hypotheses. One, the presence of international players will have an impact on attendance in MLB beyond their contribution to winning. The conclusions to draw with respect to this part of the study are fairly straightforward. If demand decreased when foreign players competed for a team, we can conclude that there is customer discrimination against players born outside the United States. If demand increased we can infer that baseball fans prefer to watch foreign-born players. If there is no positive or negative effect we can say that baseball fans are indifferent to whether players are born in or outside of the United States.

It is important to note that I focus on consumer discrimination as opposed to employer discrimination. Other studies have found evidence of employer discrimination in MLB soon after MLB became racially integrated (Lanning, 2007). However, using this data set, I find no evidence that the number of foreign players affects a team's winning percentage. This gives some indication that owners are not biased for or against international players in their hiring practices.

The second hypothesis is that markets with higher (lower) portions of their populations identifying with foreign demographics will attend in greater force if the team fielded in the corresponding market is similarly high (low). I postulate that individuals identifying themselves as belonging to a non-American nationality on the US Census prefer to watch players born in their country. I therefore predict that the cumulative effect of matching the proportion of an MSA's population from each non-American country to the proportion of players from those countries will increase demand for MLB teams.

Data and Empirical Specification

I chose to study the years 1985-2005, utilizing Structured Query Language (SQL) and the database available at baseball1.com to find player information. Using SQL I identified all the players who competed in the major leagues as well as their country-of-origin. Total counts were tabulated of the number of players from each country who competed for each team and each season. Canadian teams (Toronto Blue Jays and Montreal Expos) were omitted from the study because the application of who is an international player is inconsistent from country to country.

Demand was estimated using the equation:

$$\begin{aligned}\ln(Attendance_{it}) &= \beta_0 + \beta_1(Win\%_{it}) + \beta_2(Lagwin\%_{it}) + \beta_3(Stadium_{it}) + \\ &\beta_4(Stadium5_{it}) + \beta_5\ln(Population_{it}) + \beta_6\ln(Ticketprice_{it}) + \beta_7(TeamsinMSA_{it}) + \\ &\beta_8\ln(Income_{it}) + \beta_9(Trend_t) + \beta_{10}(Foreignplayers_{it}) + \\ &\beta_{11}(Foreignplayers_{it} * trend_t) + \beta_{12}(Foreignplayers_{it} * trend^2{}_t) + \beta_{13}(Matching_{it}) + \varepsilon_{it}\end{aligned} \quad (1)$$

The variables entered in the model are as follows. ***Attendance*** represents a team's average game attendance in year ***t*** for team ***i***. The quality and pricing control factors included ***Win%, Lagwin%, Stadium, Stadium5, Population, Ticketprice, TeamsinMSA, Income,*** and ***Trend***. ***Win%*** corresponds to the winning percentage of the team during that season and was obtained from baseball-reference.com. ***Lagwin%*** is the winning percentage of the team during the previous season. This variable is utilized to control for fan expectations early in the season, but may also pick up any habitual nature of fans (Lee and Smith, forthcoming). ***Stadium*** and ***Stadium5*** are inserted in the model to control for the increased demand attributable to the novelty effect of new stadiums. There is no accepted convention in the sport literature as far as controlling for the duration of this novelty effect. Quirk and Fort found a significant effect using the first five years of a stadium's existence (Quirk and Fort, 1992), thus I have chosen to use five years in this model. ***Stadium*** is an indicator variable for whether a team's home stadium was in its first year of use and ***Stadium5*** an indicator variable for whether a team's home stadium was five or fewer years old. ***Population*** corresponds to the value of the franchise's MSA population. US Census Bureau data is used to record the population for all MSAs where MLB teams are located. Data were recorded for each MSA's 1980, 1990, and 2000 population using American Factfinder. Intervening years were interpolated assuming a constant growth rate. Years after 2000 were extrapolated using the 1990-2000 growth rate identified for each respective franchise in the previous equation. These decadal census figures are the most reliable population data available. In reviewing the differences between the growth rates for each respective population and market population in the 1980s and 1990s (the absolute value of the 1980s growth rate minus the 1990s growth rate), 97% of the figures were within 10% of one another. Consequently, it is fair to conclude that any population movements that may have occurred during this period happened steadily rather than rapidly and are accurately represented in the interpolations. These figures are available in Table 3.2. ***Ticketprice*** is the price of tickets for the team deflated by the Consumer Price Index (2005 dollars). ***Ticketprice*** was retrieved from teammarketing.com. Although baseball tickets are considered normal goods, the price of tickets has been shown to sometimes be set in the inelastic range, thus we do not necessarily want to interpret this effect (Krautmann and Berri, 2007). Consequently, I run two versions of the model—one with the price term, the other without—and allow the reader to interpret one or both of the results. ***TeamsinMSA*** refers to the number of additional teams sharing a given team's market. As MLB teams have been shown to be substitutes for one another (Winfree, McCluskey, Mittelhammer, and Fort, 2004), the influence of close substitutes on demand is expected to be negative. ***Income*** is the mean income for residents of a franchise's MSA. Logs are taken of ***Attendance***, ***Population***, ***Ticketprice***, and ***Income*** because these variables are strictly positive and do not already represent a percentage. ***Trend*** is a variable increasing by one unit for each season to account for changes over time.

The variables of interest are ***Foreignplayers, Foreginplayers*trend, Foreignplayers*trend²,*** and ***Matching***. ***Foreignplayers*** represents the number of foreign players on the team. International players comprised an average of 18.8% of league rosters from 1985-2005. The 2004 Los Angeles Dodgers led all teams in the study with 23 foreign-born players, constituting 53.5% of their roster. Interestingly, twice during the 1985-2005 seasons a franchise did not employ even one foreign-born player—the 1985 Chicago Cubs and the 1992 Detroit Tigers—while the single greatest representation of any one nationality came in 2004, when the Kansas City Royals employed 10 players born in the Dominican Republic. ***Foreignplayers*** is interacted with ***trend*** and ***trend²*** to measure the linear and quadratic pattern of the effect over the whole era. The interaction is measured since we anticipate the marginal impact may have changed over this 20 year period.

Matching was calculated by a simple aggregation of the product of the proportion of each team from a given country and the proportion of the corresponding MSA population. The equation is defined as:

$$\textit{Matching} = \sum_{x=1}^{N} (P_i X_i) \qquad (2)$$

where *P* is the proportion of players, *X* is the proportion of the population, and *N* is the number of nationalities. A value was calculated for each team in each season. Descriptive statistics for the data are provided in 3.1. Effects of players from specific countries on demand were also investigated. However, no meaningful results were found. In addition to the matching of market population and team composition, I also looked into whether there was discrimination against Asian or Hispanic populations (the two primary non-American player populations) without regard to a specific country. Here too no significance was found.

Generalized Least Squares (GLS) regression was run to estimate the equation to correct for any autocorrelation or heteroskedasticity. Also, given the nature of the data, it is reasonable to expect correlated error terms within teams. As such, a robust covariance matrix with clustered standard errors was used.[2]

Finally, I quantified the value of an adding one foreign-born player during each season in our study. The

---

[2] All standard errors were clustered by team to help correct for any correlation pattern within teams over time (Bertrand et al. 2004). The covariance matrix is given by $V = \left(\frac{n-1}{n-k}\right)\left(\frac{G}{G-1}\right)(X'X)^{-1}\left(\sum_{i=1}^{G} X_i' e_i e_i' X_i\right)(X'X)^{-1}$ where X is nxk, there are G teams, and i denotes the observations for one team.

percent change in attendance attributable to the addition of one foreign-born player was evaluated by the equation:

$$\Delta Attendance = \beta_{10} + \beta_{11} * trend + \beta_{12} * trend^2 \quad (3)$$

The change in revenue was then estimated using the formula:

$$\Delta Revenue = \Delta Attendance * \overline{Attendance_t} * \overline{Ticketprice_t} * 81 \quad (4)$$

where $\Delta Attendance$ is the estimation from the previous equation, $\overline{Attendance_t}$ is the average attendance throughout baseball in a given season, and $\overline{Ticketprice_t}$ is the average ticket price in that season deflated by the Consumer Price Index. The equation is multiplied by 81, since MLB teams play 81 home games during the season. Therefore the change in revenue represents the yearly change in ticket revenue for a team.

Results

The regression results are displayed in Table 3.3 both with and without ticket price and the interaction of foreign players and trend in the model. All four models showed statistically significant heteroskedasticity and first-order autocorrelation, which was corrected for by using GLS. The ensuing interpretation discusses the complete model. The majority of the control variables are in the expected direction and significant in the model. Winning percentage, the previous season's winning percentage, first year in a new stadium, and MSA population were all positive and significant. The number of teams in MSA significantly reduced demand. Income and whether or not a team was in its first five years in a new stadium were not found to be significant.

Turning our attention to the variables of interest, all were found to be significant in estimating attendance in the full model with the exception of the matching variable. Considered alone, the number of foreign players on a team increased demand, but along with the interaction terms, the effect is more intricate and quite clearly quadratic. Figure 3.2 depicts the combined effect of the number of foreign players and interaction terms. The overall result of these terms, taken in combination, shows that foreign players had a negative effect on demand at the outset of the period being studied. Although the effect continues to remain negative for almost half the era, it is moving steadily in a positive direction, approaching zero in 1993. The marginal change from season to season continues to move in a positive direction and at roughly similar intervals, peaking around the year 2001, in which attendance increased 1.14% from the presence of an international player. Then, although the net effect remains positive until the end of the sample period, it flattens out. In checking for whether the interaction effects were merely picking up on a quadratic trend effect, trend squared was introduced into the model but was not found to be significant. It is therefore fair to conclude that the foreign players interaction is a significant finding.

Table 3.4 shows the percentage change attributable to adding one foreign player to a team for each season in the study as well as the change in ticket revenue associated with such an addition. The revenue associated with that change was calculated based on the average attendance across the league and ticket price for each season. The largest negative change in revenue from an international player came in 1985 with a loss of $735,528. In 2001, when the effect peaked, teams gained $568,068 from the presence of an international player.

I also ran estimations using year indicator variables, instead of quadratic trend variables, to validate that the findings truly followed a quadratic pattern and were not simply a function of outliers in just a few seasons. Figure 3.2 also depicts the individual year effects along with the estimated quadratic effect.

While not significant, the findings on the matching variable are no less interesting. Were it the case that the matching variable was found to be significant, we could have attributed the increase in attendance associated with international players to the fan population identifying with those countries. However, since the likeness of team and fan population is not significant in the model, the change in demand attributable to foreign-born players cannot be found in the non-international fan population. Rather than ascribing the change in demand to fans who want to see their countrymen compete in MLB, there is no evidence that the increase (decrease) in demand is different in the international and non-international populations.

In light of the findings with respect to the matching variable coupled with the declining presence of African-

American players in MLB, as documented by the Racial and Gender Report Card (RGRC) (Lapchick, 2007), I believed that identifying whether MSAs with higher or lower African-American populations responded differently to the presence of foreign-born players may shed more light on the source of the consumer bias. I looked at whether MSAs with higher proportions of African-Americans responded differently to foreign players, but did not find any such effect.

I further queried the same database to see if there were any confounding factors that may be the true source of the change in demand attributable to foreign players and cite several in the ensuing discussion. The number of foreign pitchers and batters each increase incrementally throughout the sample. As such neither explains the quadratic pattern of demand. I further conjectured that the type of foreign batter in the major leagues may have evolved to the slugger favored by MLB fans, only to decline in recent seasons, but, in fact, the highest ratio of home runs per foreign batter occurred during the recent flattening out, several years after the peak in demand for foreign-born players. Finally, the proportion of batters among foreign players declined gradually and so too does not confound the change in preference for foreign players.

All of these analyses support the idea that there is not a confounding factor in the evolution of fan interest in foreign-born players. These figures are presented in Table 3.

Discussion

This chapter demonstrates how traditional studies of consumer discrimination can be applied to gauge the impact of international players in sports. I show that the net effect of adding an additional international player has evolved during the sample period from a negative to a positive. Furthermore, I demonstrate that the marginal effect from year to year, steadily increasing throughout the first fifteen seasons in the study, recently has begun to level off. Just as in the case of post-integration MLB, as the population of athletes continues to grow internationally heterogeneous, this line of study takes on added importance to owners, management, and policy-makers in all leagues.

The results would imply that leagues may have incentives to make their league more international. The findings are not merely statistically but economically significant. At its peak in 2001, the effect of adding an additional foreign-born player to a major league roster was an increase in over $568,000 in revenue and the average US-based team fielded 10.78 foreign players on its roster. Thus, the additional revenue garnered by the average team was over $6.1 million in that season alone. Furthermore, since teams fielded as many as 14 and as few as 7 foreign players that same season, the difference in revenues between these two teams is almost $4 million. Consequently, franchises identifying the effect stood to gain a considerable advantage in revenues on competing teams in the league. Leagues such as the National Football League and the National Basketball Association have actively tried to become more international by getting international players or scheduling more games outside of the United States. While this study may not have implications for trying to create a more international fan base, it does show that domestic demand may be affected by the presence of international players. Future studies may explore if similar trends exist in other sports and explain why the change in consumer preference occurred when it did in baseball.

Table 3.1

Descriptive Statistics for Foreign Players Study

Parameter	**N**	**Min**	**Max**	**Mean**	**Standard Deviation**
Ln(Attendance)	546	8.998	10.961	10.179	0.333
Win%	546	0.265	0.716	0.500	0.071
Lagwin%	546	0.265	0.716	0.499	0.070
Stadium	546	0	1	0.035	0.183
Stadium5	546	0	1	0.163	0.370
Ln(Population)	546	14.181	16.949	15.355	0.727
Ln(Ticketprice)	546	2.127	3.797	2.680	0.309
Teamsin MSA	546	0	1	0.310	0.463
Ln(Income)	546	10.256	10.756	10.481	0.115
Trend	546	0	20	10.381	6.049
Foreignplayers	546	0	23	8.103	3.962
Matching	546	0.000	0.240	0.073	0.047

Table 3.2

Change in Population Growth

	%change MSA	%change Hisp	%change PR	%change Mex	%change DR	%change Asia	%change Jpn	%change Chi	%change Kor
chgAtl	0.402	8.550	1.521	11.677	0.022	4.401	8.407	6.572	4.440
chgBal	0.016	1.667	2.783	2.010	0.014	1.864	3.052	3.097	4.743
chgBos	0.821	4.031	3.614	1.667	0.000	6.502	5.388	2.937	8.786
chgCha	0.000	1.453	1.211	1.231	0.008	0.833	2.593	0.853	3.508
chgChn	0.000	1.453	1.211	1.231	0.008	0.833	2.593	0.853	3.508
chgCin	0.939	8.170	0.657	12.657	9.934	3.088	4.114	2.791	5.989
chgCle	3.156	2.684	5.181	1.100	0.178	3.119	8.400	2.594	5.169
chgDet	0.885	4.545	1.560	4.871	0.161	0.653	7.867	0.662	4.757
chgHou	0.154	0.080	5.666	0.534	0.019	3.675	2.033	4.281	2.714
chgKca	0.405	3.948	3.898	3.924	0.931	1.827	2.609	2.376	5.679
chgLaa	5.669	5.180	5.705	5.633	0.003	8.936	5.511	9.599	9.468
chgLan	5.669	5.180	5.705	5.633	0.003	8.936	5.511	9.599	9.468
chgMil	0.910	1.671	0.930	1.015	0.094	3.401	1.056	2.746	6.081
chgMin	0.015	6.989	1.580	7.180	1.185	4.191	4.045	0.763	6.796
chgNya	5.486	2.523	2.609	0.838	0.000	6.222	7.686	4.503	11.667
chgNyn	5.486	2.523	2.609	0.838	0.000	6.222	7.686	4.503	11.667
chgOak	5.570	6.504	5.559	8.814	0.051	7.022	8.392	5.270	10.814
chgPhi	1.782	1.266	3.245	7.761	0.007	4.584	5.480	5.335	7.291
chgPit	0.599	4.756	2.200	4.054	0.070	2.097	2.118	0.568	0.630
chgSdn	1.785	1.942	5.076	2.865	0.037	4.972	3.719	4.337	4.704
chgSea	1.419	1.716	7.050	2.123	0.287	3.512	3.732	0.983	9.253
chgSfn	5.570	6.504	5.559	8.814	0.051	7.022	8.392	5.270	10.814
chgStl	0.269	3.543	1.130	4.458	0.695	0.305	6.071	1.146	1.964
chgTex	0.294	0.967	4.498	0.295	0.021	4.754	4.547	5.670	8.511

Table 3.3

Generalized Least Squares Results

Estimating Log of Season Attendance in MLB

Variable	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Constant	6.7351**	2.18	6.4735**	2.03	5.1006*	1.85	4.7426*	1.67
Win%	1.3408***	6.22	1.3414***	6.11	1.4378***	6.54	1.4440***	6.42
Lagwin%	0.8894***	5.98	0.9425***	6.58	1.0547***	7.23	1.1171***	8.40
Stadium	0.1974***	4.55	0.1947***	4.39	0.2240***	5.75	0.2225***	5.68
Stadium5	0.0640	0.93	0.0618	0.86	0.0896	1.42	0.0901	1.36
Ln(Population)	0.1758***	2.88	0.1832***	2.96	0.2185***	3.93	0.2280***	3.98
Ln(Ticketprice)	0.2616*	1.75	0.2695*	1.80				
Teamsin MSA	-0.1803*	-1.76	-0.1935*	-1.78	-0.2473***	-2.47	-0.2642***	-2.46
Ln(Income)	-0.0837	-0.28	-0.0861	-0.29	0.0551	0.21	0.0607	0.23
Trend	-0.0196***	-2.66	-0.0083	-1.44	-0.0116*	-1.74	-0.0009	-0.19
Foreignplayers	-0.0368*	-1.80	0.0057	1.11	-0.0376*	-1.87	0.0093**	2.16
*Foreignplayers*Trend*	0.0061**	2.25			0.0075***	2.68		
Foreignplayers$Trend^2$	-0.0002**	-2.27			-0.0003***	-2.90		
Matching	0.1920	0.33	0.0697	0.13	0.1416	0.25	0.0253	0.05
R^2	0.512		0.509		0.492		0.495	

Table 3.4

Impact of One Additional Foreign Player on Attendance and Ticket Revenue

Season	Change in Attendance	Change in Revenue
1985	-3.68%	-$735,528
1986	-3.09%	-$621,256
1987	-2.54%	-$530,403
1988	-2.02%	-$420,197
1989	-1.55%	-$353,335
1990	-1.11%	-$270,733
1991	-0.71%	-$196,738
1992	-0.35%	-$95,153
1993	-0.03%	-$9,922
1994	0.25%	$87,313
1995	0.49%	$136,460
1996	0.70%	$211,935
1997	0.87%	$297,496
1998	0.99%	$390,943
1999	1.08%	$461,140
2000	1.13%	$537,549
2001	1.14%	$568,068
2002	1.11%	$533,139
2003	1.05%	$501,561
2004	0.94%	$498,271
2005	0.80%	$427,375

Table 3.5

Foreign Batters and Pitchers with Batter Power Production

Year	HR	#Batters	HR/Batter	#Pitchers	PctForeignBatters	PctForeignPitcher
1985	385	111	3.468	33	11.20%	7.57%
1986	390	110	3.545	33	10.88%	7.35%
1987	574	123	4.667	38	11.82%	8.15%
1988	446	138	3.232	48	13.42%	10.50%
1989	417	128	3.258	40	12.05%	8.20%
1990	485	141	3.440	49	12.78%	9.46%
1991	591	156	3.788	58	14.51%	11.74%
1992	462	151	3.060	49	14.34%	10.52%
1993	654	178	3.674	64	15.19%	11.59%
1994	670	163	4.110	65	15.87%	13.24%
1995	777	199	3.905	83	16.14%	13.63%
1996	995	216	4.606	86	17.52%	14.73%
1997	996	247	4.032	104	20.46%	18.06%
1998	1186	256	4.633	107	19.69%	17.54%
1999	1396	277	5.040	121	21.66%	19.24%
2000	1420	293	4.846	132	21.77%	19.97%
2001	1521	308	4.938	131	23.51%	20.44%
2002	1410	317	4.448	146	24.63%	22.60%
2003	1648	323	5.102	148	24.53%	22.39%
2004	1812	342	5.298	165	26.09%	24.81%
2005	1454	347	4.190	151	26.71%	23.59%

Figure 3.1

Percent of International Players in MLB

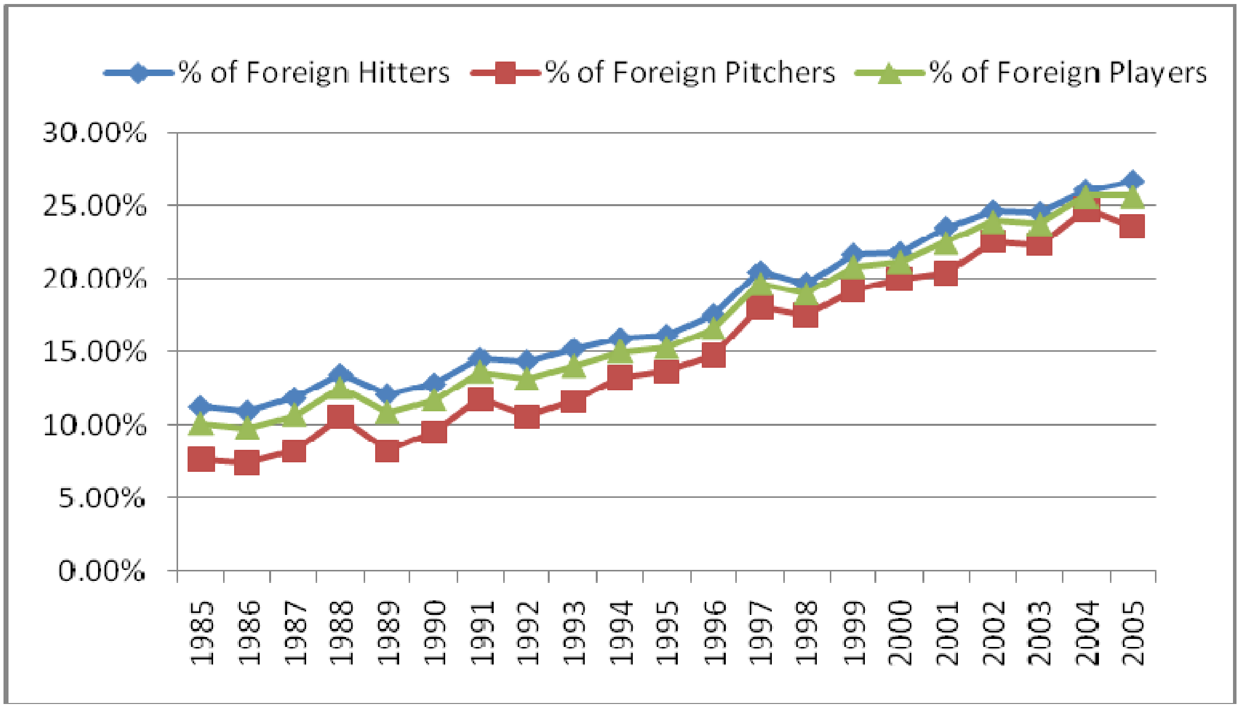
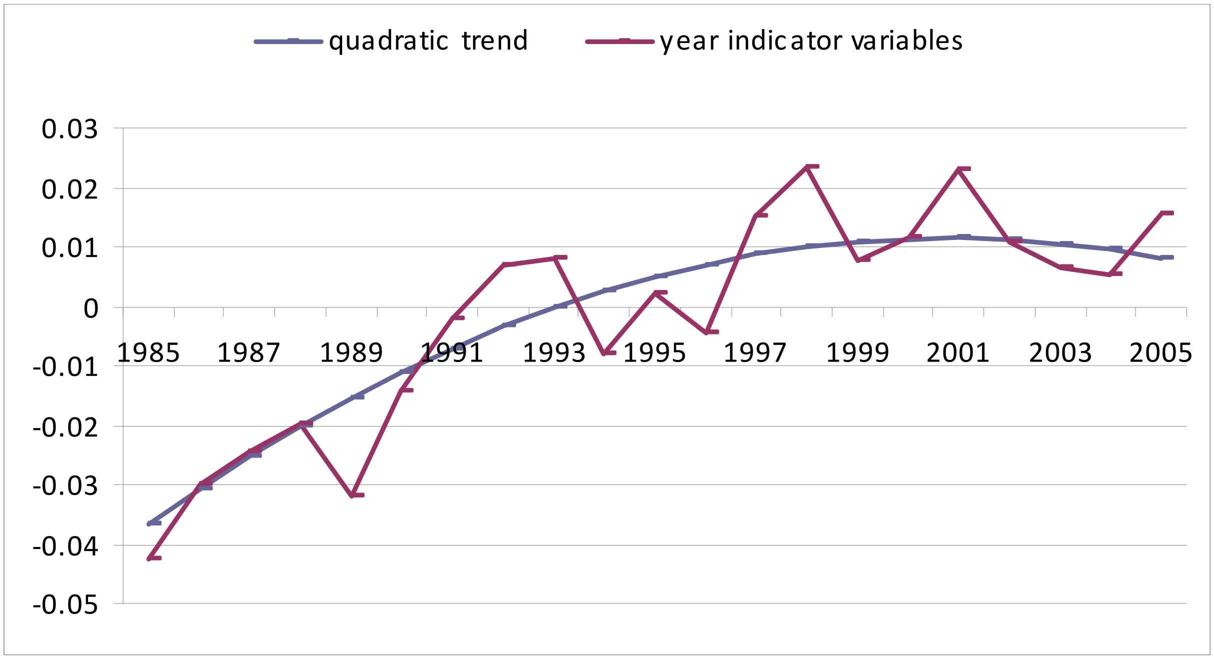


Figure 3.2

Percent Change in Attendance From One Foreign Player



References

American Factfinder (n.d.). Last viewed on 07/24/08 at http://factfinder.census.gov/home/saff/main.html?_lang=en

Baseball-Reference (n.d.). Last viewed on 07/24/08 at http://www.baseball-reference.com/

Becker, G. The Economics of Discrimination, 2nd Edition. Chicago, IL: University of Chicago Press, 1971.

Bertrand, M., Duflo, E. & Mullainathan, S. (2004). How Much Should We Trust Differences-in-differences Estimates? *The Quarterly Journal of Economics*. 1: 249-275.

Bodvarsson, O. & Brastow, R. (1999). A Test of Employer Discrimination in the NBA. *Contemporary Economic Policy*. 17(2), 243-255.

Bodvarsson, O. & Partridge, M. (2001). A Supply and Demand Model of Co-worker, Employer and Customer Discrimination. *Labour Economics*. 8(3), 389-416.

Brown, E., Spiro, R., & Keenan, D. (1991). Wage and Nonwage Discrimination in Professional Basketball: Do Fans Affect It? *American Journal of Economics and Sociology.* 50(3), 333-345.

Burdekin, R., Hossfield, R. & Smith, J. (2005) Are NBA Fans Becoming Indifferent to Race? *Journal of Sports Economics*. 6(2), 144-159.

Burdekin, R. & Idson, T. (1991). Customer Preferences, Attendance and the Racial Structure of Professional Basketball Teams. *Applied Economics*. 23(1), 179-186.

Gwartney, J. & Haworth, C. (1974). Employer Costs and Discrimination: The Case of Baseball. *The Journal of Political Economy*. 82(4), 873-881.

Han, C. (1989). Country Image: Halo or Summary Construct? *Journal of Marketing Research*. 26(2), 222-229.

Hanssen, A. (1998). The Cost of Discrimination: A Study of Major League Baseball. *Southern Economic Journal*. 64(3), 603-627.

Hoang, H. & Rascher, D. (1999). The NBA, Exit Discrimination, and Career Earnings. ***Industrial Relations***. 38(1), 69-91.

Hong, S. & Wyer, R. (1989). Effects of Country-of-Origin and Product-Attribute Information on Product Evaluation: An Information Processing Perspective. ***The Journal of Consumer Research***. 16(2), 175-187.

Jibou, R. (1988). Racial Inequality in a Public Arena: The Case of Professional Baseball. ***Social Forces***. 67, 524-534.

Johansson, J. (1989). Determinants and Effects of the Use of 'Made In' Labels. ***International Marketing Review.*** 6(1).

Johnson, N. & Marple, D. (1973). Racial Discrimination in Professional Basketball: An Empirical Test. ***Sociological Focus***. 6(4), 6-18.

Kahn, L. & Sherer, P. (1988). Is There Discrimination in the Black Man's Game? ***Journal of Labor Economics***. 6(1), 40-61.

Kahn, L. (1991). Discrimination in Professional Sports: A Survey of the Literature. ***Industrial and Labor Relations Review***. 44(3), 395-418.

Krautmann, A. (1990). Shirking or Stochastic Productivity in Major League Baseballl? ***Southern Economic Journal***, 56(4), 961-968.

Krautmann, A. & Berri, D. (2007) Can We Find It at the Concessions? Understanding Price Elastictity in Professional Sports. ***Journal of Sports Economics***, 8(2), 183-191.

Lanning, J. Testing Standard Theories of Economic Discrimination: Productivity, Prejudice, and Lost Profits During Baseball's Integration, Working Paper.

Lapchick, R. (2007).The 2006 Racial and Gender Report Card: Major League Baseball. Retrieved on June 13, 2007 from http://www.bus.ucf.edu/sport/public/downloads/2006_RGRC_MLB.pdf

Lee, Y.H. & Smith T. Why are Americans Addicted to Baseball? An Empirical Analysis of Fandom in Korea and the U.S.? *Contemporary Economic Policy*, forthcoming.

Leeds, M., & von Allmen, P. (2007). *The Economics of Sports* (3rd ed.) Boston: Addison-Wesley.

Liefer, E.(1995). *Making the Majors: The Transformation of Team Sports in America*. Cambride, MA: Harvard University Press.

Longley, N. (1995). Salary Discrimination in the National Hockey League: The Effects of Team Location. *Canadian Public Policy / Analyse de Politiques*, 21(4), 413-422.

Longley, N. (2000). The Underrepresentation of French Canadians on English Canadian NHL Teams: Evidence From 1943-1998. *Journal of Sports Economics*, 1(3), 236-256.

Maheswaran, D. (1994). Country of Origin as a Stereotype: Effects of Consumer Expertise and Attributable Strength on Product Evaluations. *The Journal of Consumer Research*. 21(2), 354-365.

McCormick, R. & Tollison, R. (2001). Why do Black Basketball Players Work More for Less Money? *Journal of Economic Behavior and Organization.* 44(2), 201-219.

Osborne, E. (2006). Baseball’s International Division of Labor. *Journal of Sports Economics*, 7(2), 150-167.

Pedace, R. (forthcoming) Earnings, Performance, and Nationality Discrimination in a Highly Competitive Labor Market: An Analysis of the English Professional Soccer League. *Journal of Sports Economics*.

Quirk, J. & Fort, R.D. (1992). *Pay Dirt: The Business of Professional Team Sports*. Princeton University Press.

Scully, G.W. (1973). Economic Discrimination in Professional Sports. *Law and Contemporary Problems*. 38(1), 67-84.

Sommers, P. & Quinton, N. (1982). Pay and Performance in Major League Baseball: The Case of the First Family of

Free Agents. *The Journal of Human Resources.* 17(3), 426-436.

Swinnerton, K. & Wial, H. (1985). Is Job Stability Declining in the US Economy. *Industrial and Labor Relations Review.* 48(2), 293-305.

Winfree, J., McCluskey, J., Mittelhammer, R. and Fort, R. (2004). Location and Attendance in Major League Baseball. *Applied Economics* 36(19): 2117-2124.

Chapter IV

Short-run Demand and Uncertainty of Outcome in Major League Baseball

Background

Uncertainty of outcome is one of the fundamental differences between sporting events and other forms of entertainment. Specifically, the uncertainty of outcome hypothesis (UOH) states that fans want to see their team win in close games and at least contend for championships. First iterated in Rottenberg's (1956) seminal work, Cairns (1987) neatly separated the notions of game uncertainty (GU), playoff uncertainty (PU), and consecutive season uncertainty (CSU). Ultimately all work on uncertainty of any type, even those not conceived in his original treatise, relates back to Rottenberg's pioneering contention—the more balanced the league, the greater the economic benefit. In the following paragraphs I review some of the previous literature evaluating the impact these aspects of uncertainty have had on demand. My aim in this chapter is to build on the existing literature by incorporating all three types of uncertainty into a model predicting demand for individual games in Major League Baseball (MLB). Like most questions in sport economics, there is a vast literature using MLB as its subject. As it is also the focus of this research, I focus on those studies in our review of the literature.

The earliest work on GU evaluates the relative quality of teams at the time of the contest (Demmert, 1973; Hill, Madura, and Zuber, 1982). Whether the studies use teams' winning percentage or divisional standing, the theory is upheld that demand escalates as teams are more closely matched. By using game attendance, Hill and associates are able to attribute changes in demand for individual games to variables such as the quality of opponent, games contested later in the season, and games played on the weekend. The article is unique, even among later studies, in that it quantifies short run demand and, accordingly, can precisely identify the particular factors influencing the decision to attend each game at any point in the season. The more recent studies (Scully, 1989; Quirk and Fort, 1992; Knowles, Sherony, & Haupert, 1992; Humphreys, 2002) use the dispersion of winning percents, the most commonly utilized measure in competitive balance studies, as a proxy for league balance. Using average annual game attendance as the dependent variable, UOH is tested by the sensitivity of attendance to the ratio of actual standard deviation of winning percent to the standard deviation of a balanced league.

I break briefly from our review of the UOH literature on baseball to recognize other contributions to the research. In his review of competitive balance of Dutch soccer, Koning (1999) develops a simple model isolating the strength of team from that of other factors in team success such as home-field-advantage. He finds that game outcomes (including ties) are a function of the relative contributions of team quality and the impact of home-field advantage as identified by game results of evenly-matched teams.

The idea of playoff uncertainty encompasses both a team's competitiveness as it relates to qualifying for playoffs and Rottenberg's original notion of competing for championships. Studies have used both the straightforward metric of years per championship (Quirk and Fort, 1992) and the more econometrically elegant Gini coefficient of championship concentration normalized on seasons (Quirk and Fort, 1992). Much more attention has been paid to the closeness of the home team to qualification for postseason play (Hill, Madura, and Zuber, 1982; Baade and Tiehen, 1990; Butler, 2002; Schmidt and Berri, 2004). The aforementioned studies aim to capture the importance of being in contention, while Whitney (1988) uses the simple notion of how far the season has progressed and Hill, Madura, and Zuber (1982) use games contested in the final two-thirds of the season as an indication of the importance of games in the minds' of fans. The variable of games behind the divisional leader is found to be significant in several studies (Demmert, 1973; Knowles, et al., 1992), but not in Noll's (1974) analysis of seasonal demand.

Fans of a league want to see "their" team at least compete for championships. Therefore, if there is little variation in league standing at the conclusion of consecutive seasons, the league is said to be more certain and should impact demand adversely. Previous research has employed both correlation of winning percentage in consecutive seasons (Butler, 1995) and Gini coefficient of team winning percentage (Schmidt and Berri, 2001) as well as average standard deviation ratio across teams in multiple seasons (Humphreys, 2002) to measure CSU.

In an attempt to build on the existing research in the area, we call the reader's attention to the following. In their time-series analysis of MLB, Fort and Lee (working paper) point out that while the literature on UOH is extensive, there is a dearth of studies that address all forms of uncertainty. Consequently, little can be learned of the relative importance of these factors. In addition, few studies look at attendance on a game-by-game basis, choosing instead to look solely at average annual attendance in spite of the widespread agreement on the usefulness of conducting empirical testing on individual contests (Szymanski, 2003). Additionally, the informed consumer of sport knows far more about his or her team than has been put into previous models, many of which have demonstrated weak or limited support for the UOH. One way we can advance in this respect is regarding strength of schedule and GU/PU. Consider the example of two teams vying for the last remaining playoff berth, Team A leading Team B by one game as each team has two games remaining on its schedule. Schedule, however, makes a real difference in the probability of playoff qualification in our scenario. Let's further suppose that Team A's final games are against the league juggernaut—an opponent with a winning percent of 1.000, having won every contest in its season. Team B's opponent is completely hapless—a winning percent of 0.000, having won not a single game. The introduction of remaining schedule informs a fan's outlook on each team's chance of qualification. Yet another example of how scheduling comes into play is in past performance. Consider the example of two teams with identical records, but Team C has faced teams with a combined winning percentage of 0.800 while Team D's opponents have a combined winning percentage of 0.200. One can easily deduce that Team C's past performance is more impressive than Team D's because it has come against better competition. If teams C and D have similar remaining schedules, Team C's past performance would tell us that it is more likely to fare better in future contests. Thus, adding the quality of opponents remaining on the schedule and/or previous opponents certainly affects the outlook of an informed fan on the likelihood of playoff qualification.

I feel that this study builds on the existing literature in a number of ways in order to address these issues. I begin by including applying all three notions of uncertainty to MLB. This will allow us to decipher the impact of each one and make more definitive statements about how they impact demand. I also use game attendance rather than annual attendance as the dependent variable in this research and bring strength of schedule into the equation. Lastly, I put forth the most thorough construction of metrics quantifying game and playoff uncertainty to date, employing a probit model of anticipated success in remaining contests based on performance in previous contests to create measures of GU and PU.

The specific purpose of this research is to uncover the extent to which uncertainty of outcome plays a role in fans’ decisions to attend MLB games. The broader subject is that of consumer preferences. Demand for any product, sporting events included, is influenced by a number of factors. Population and income of the consumer base as well as the price of substitutes and expectations about the future are all well-established determinants of demand. Preferences too are an important factor and this research attempts to advance our understanding of consumer preferences regarding uncertainty of outcome.

Methods

If demand is based on all of these factors of uncertainty then:

Attendance=f(X, game uncertainty, playoff uncertainty, within season balance, across season balance)

where *X* are factors other than those related to uncertainty. These may be attributes of game quality, stadium quality, and fan base quality. Game quality includes temporal factors such as whether the game was played during the traditional workweek or on the weekend and day or night, as well as the quality of the home and away teams. Note that including the quality of teams playing in a given contest is necessary to evaluate for uncertainty, or the relative quality of competing teams. Stadium quality too is an important input in measuring demand as new stadiums have been shown to draw additional fans (Coates and Humphreys, 2005). The final consideration is the size and affluence of the fan base. Larger size and wealth of market population where the teams are based should influence the demand for the greater. Thus the overall equation for estimating attendance for team *i* in game *g* of year *y* can be thought of as:

$$Y_{i,g,y} = \beta + \alpha Z + \chi X_1 + \delta X_2 + \gamma X_3 \qquad (1)$$

where Z is UOH, X_1 is game quality, X_2 is stadium quality, and X_3 is fan base quality.

Turning our attention to UOH, a probit model was constructed to find the anticipated result of remaining contests and will describe the model before revisiting the specific metrics of GU, PU, and CSU. Because the ultimate goal is to separate fan preferences with respect to the different types of uncertainty and it is a reasonable conjecture that a team's chances for playoff qualification is negligible or at most a small factor in fans attending early on in a 162 game MLB season, I chose to only use games from July 1 and later, roughly the second half of the year. Thus the sample consists of more than 81 games for each club from 1996 to 2007 collected from baseball-reference.com. This era was selected for its substance, contemporary nature, and the lack of time-series issues. The probit uses games previously played in that season to predict future outcomes. As addressed in the previous section, knowledgeable consumers of sport are familiar with the notion of strength of schedule (SOS). I incorporate this idea into forecasting game outcome. The home team's winning percentage as well the winning percentage of its previous opponents was put into the model along with similar inputs for the away team in order to produce the probability of a home team win in each contest. The next step involved simulating a winner based on the probability established by the model run according to the Monte Carlo method. The model then simulated the remainder of the season one thousand times using a similar design, updating all the probabilities daily. The information generated by this probit was used in creating many of the metrics described in the remainder of this section.

A number of game uncertainty measures were created to separate the various types of expectations.

$$GameUncertainty1 = p(win)_{i,y,g} \quad (2)$$

Our first measure, ***GameUncertainty1***, is the probability of a win in that contest as established by the probit model for team i in year y for game g. Perhaps the best way to think of this measure is the likelihood of a home team win. If fans prefer games that the home team is likely to emerge victorious, demand will increase as ***GameUncertainty1*** increases.

$$GameUncertainty2 = |1-2*p(win)_{i,y,g}| \quad (3)$$

GameUncertainty2 adds the quality of minimizing when the outcome of the game is most uncertain (a value of 0.5). Accordingly, if conventional theory on UOH holds true, as ***GameUncertainty2*** increases, attendance will decrease.

$$TeamUncertainty = |win\%_{i,y} - win\%_{i,y-1}| \quad (4)$$

TeamUncertainty measures the change in team quality for the home club from one year prior. What ***TeamUncertainty*** will elucidate is how fans feel about changes in their team's performance, regardless of whether that change is for the better or for the worse. Therefore, if fans prefer their team to perform consistently from season-to-season, ***TeamUncertainty*** will have a negative impact on demand.

Playoff uncertainty too is based on the probit model. The first two metrics of playoff uncertainty are based on the home team's position in the standings at the outset of the day that game is contested. After the probability of a home team win is defined by the probit, Monte Carlo simulations are run and team records updated for use by the probit in subsequent contests. Thus the probability of playoff qualification can easily be calculated by the model going a step further to predict not just that but all remaining games—we chose to perform 1,000 simulations—and the proportion of the time that the team finishes with a record that would qualify for baseball's postseason establishes our first two playoff uncertainty measures.

$$\textit{PlayoffUncertainty1} = p(\text{playoffs})_{i,\, y,\, g} \qquad (5)$$

and

$$\textit{PlayoffUncertainty2} = |1 * 2p(\text{playoffs})_{i,\, y,\, g}| \qquad (6)$$

Similar to the first two metrics of game uncertainty, employing both of these PU metrics simultaneously allows us to separate whether fans are interested in attending games based on the likelihood of their team qualifying for the playoffs alone and whether there is increased demand for games when the likelihood of playoff qualification is most uncertain.

One additional PU metric is added to this study.

$$MarginalPlayoffUncertainty = p(\text{playoffs}|\text{win}_g)_{i,y,g} - p(\text{playoffs}|\text{loss}_g)_{i,y,g} \quad (7)$$

The model assumes a home team win in game g, then a loss, and then estimates the likelihood of the home team qualifying for the postseason World Series tournament given the two possible outcomes of the game in question. *MarginalPlayoffUncertainty* can be thought of as the likelihood of playoff qualification given a win versus a loss or the game’s importance in terms of playoff qualification.

Finally, I want to evaluate if the unpredictability of the teams qualifying for the playoffs across the league has any bearing on demand for individual games. Thus we introduce

LeagueUncertainty = variance p(playoffs) $_n$ (8)

where n is the number of teams in the league. If fans prefer to attend games that the league's playoff teams are less certain, then demand will increase as *LeagueUncertainty* increases.

CSU measures, within season balance and across season balance, are constructed according to practices employed previously in the UOH literature. *WithinSeasonBalance* is defined as

$$\sigma_{y,g} = \frac{\sqrt{\sum (win\%_i - \overline{win\%_i})^2}}{n-1} \quad (9)$$

and *AcrossSeasonBalance* is the correlation in the performance of all the league's teams at game g in year y and at game g of year y-1.

AcrossSeasonBalance = $\rho(win\%_{n,y,g}\ ,\ win\%_{n,y-1,g})$ (10)

Consequently, if fans prefer a league that is more uncertain, demand will escalate as *AcrossSeasonBalance* diminishes.

Data

Data were collected on every regular season MLB game contested after July 1 from 1996 through 2007 with attendance data collected from baseball-reference.com. *Win%* refers to the home team's winning percentage at the outset of the game being played. *OpponentWin%* is the visiting team's winning percentage entering the game. Additional indicator variables for the year, month, and day of week were also created. 40 stadium dummy variables and 30 team dummy variables were employed to control for stadium and team fixed effects respectively. The other indicator variables we constructed were *DayGame*, representing whether the game was played in the afternoon as opposed to at night, and *LastYearStadium*, *FirstYearStadium*, and *SecondYearStadium*, in order to control for nostalgia and novelty effects of leaving an old ballpark and moving into a new one. *TicketPrice* accounts for the price of attendance. *RunsHome* and *RunsAway* are the average runs scored by the home and visiting teams to that point in the season. Finally, the variables of interest—*GameUncertainty1*, *GameUncertainty2*, *TeamUncertainty*, *PlayoffUncertainty1*, *PlayoffUncertainty2*, *MarginalPlayoffUncertainty*, *LeagueUncertainty*, *WithinSeasonBalance*, and *AcrossSeasonBalance*—were constructed as described earlier in the chapter.

Postulating that there may be relationships between the different types of uncertainty, I constructed a correlation matrix of our uncertainty variables. Due to the nature of the outcome variable—individual games—further testing for autocorrelation and heteroskedasticity was performed. Were there to be correlation in the attendance from one game to the next, I would run our demand estimation as a Generalized Least Squares (GLS) regression with clustered standard errors rather than an Ordinary Least Squares (OLS) model since using OLS under these conditions would overstate the significance of the variables entered. The variables entered into the demand estimation were as described above with the logarithmic form of attendance serving as the dependent variable.

Results

Summary statistics of the probit are found in Table 4.1. With 14,756 observations from the 1996-2007 seasons, the probit used home team winning percentage, home team strength of schedule, road team winning percentage, and road team strength of schedule to successfully predict 56.53% of the game outcomes. A graphical representation of how successful the probit was in its predictions after X games in the season is provided in Figure 4.1. It appears as though the model improves slightly as the season wears on, perhaps as a result of having more information on each competing team. The initial variability can be explained by the smaller sample of games—only occasionally would a home team have played fewer than 80 games by July 1. All of the four variables entered into the probit produced estimates in the expected direction (i.e., the higher the winning percentage of the road team, the lower the probability of a home team win in that contest), but only records of the competing teams themselves were significant in the model. This result indicates that it cannot be stated definitively that a team's strength of schedule contributes to the expected outcome of a Major League Baseball game. Parameter estimates, marginal effects, and t-statistics are provided in Table 4.2.

These results were then utilized in the demand estimation to measure ***GU*** and ***PU***. Summary statistics are available in Table 4.3. The uncertainty variables and winning percentage were then tested for correlation with the results depicted in Table 4.4. A surprising few of the variables were highly correlated. It is worth noting that most correlated variables in the set were ***GU1*** and ***GU2***

(r=0.616), ***GU1*** and ***PU1*** (r=0.568), and ***GU2*** and ***PU1*** (r=0.471).

After conducting tests of autocorrelation and heteroskedasticity, shown in Table 4.5, it was determined that a GLS model with clustered standard errors was appropriate on account of the autocorrelation. In other words, the assumption that the demand for a contest and the game previous are independent could not be made. Thus GLS is used checking only for first order correlation and the results of this regression shown in Table 4.6. Over two-thirds of the variability in attendance can be explained by the model (R^2=0.67). The log form of attendance is used as the dependent variable in the equation in accordance with convention in estimating demand in sports research. The reference game in this model is a Sunday game played in July of 1997[3]. I first cover the control variables. As expected, attendance decreased for games played on weekdays. Interestingly, all else equal, July games are more highly attended than those in August, September, and October, there was a higher demand for Saturday games than Sunday games, and night games were preferable to those played during the day. Additionally, year, team, and

[3] Arizona and Bank One Ballpark are the reference team and stadium respectively.

individual park effects are accounted for and easily interpreted in the model results table. Consistent with previous research on the subject, demand increased for games played in the final year a stadium's use and in new stadiums. Runs per game and opponent's run per game are also included in the model as previous research has demonstrated fan preference for high scoring contests (Winfree, McCluskey, Mittelhammer, and Fort, 2004; Domazlicky and Kerr, 1990). Interestingly, neither home nor road scoring was found to be significant in the GLS model. Home team quality was positive and significant in the model. In other words, the better the home team's winning percentage, the more fans will come to the stadium. By contrast, road team quality too was positive, but nowhere near significant. We are left to conclude that fans care greatly about the quality of their home team, but the same cannot be said about the visiting squad.

We now turn to our variables of interest beginning with those variables related to game uncertainty. Neither ***GU1***, the probability of a home team win in that contest, nor ***GU2***, the uncertainty of the contest, are found to be significant. Accordingly, it cannot be concluded that fans prefer games in which their team is favored. Similarly, we cannot conclude that fans prefer contests of more evenly-matched teams. Both of these findings diverge from conventional thinking and research on UOH. ***TU***, the difference in winning percent for the home team from one year prior, is negative and significant. What this speaks to is fan expectations. Since demand decreases as success changes, it can be inferred that fans prefer that their team's performance remains consistent from year-to-year.

PU1, the probability of playoff qualification, is positive and significant in the model. This means that fans prefer attending games when their team is more likely to garner a postseason berth and are less likely to attend when their team's chance of making the playoff decreases. ***PU2***, the uncertainty of playoff qualification, is not significant in the model. According to this result, it cannot be stated that fans prefer to attend contests as their team is on the playoff "bubble." What is significant in the model and follows traditional thinking on UOH is ***MarginalPU***. The model shows that fans recognize game importance in terms of playoff qualification and prefer contests when more rides on winning and losing that game.

We now will examine the uncertainty variables that apply to not just the team but the league as a whole. ***LeagueUncertainty***, the SD of playoff chances for all teams, is not a significant factor in estimating demand. This challenges the idea that fans prefer when the league champion is less predictable, controlling for their team's success. So too ***WithinSeasonBalance***, estimated by the SD of all team's winning percentage at that point in the season, is not significant in the model. Were this significant, we could have affirmed the longstanding contention that consumers prefer more balanced leagues. Thus this result presents a strong challenge to that assertion. Finally, the variable ***AcrossSeasonBalance***, approximated by the correlation of winning percentages for all teams at the outset of the game and one year prior, was also not significant. I therefore do not maintain that league consistency across seasons plays a role in fans' decisions to attend MLB games.

While steadfast in this approach and its directing the use of GLS because of reasons presented earlier in the paper, I nonetheless address OLS findings on these identical variables. Results for the just the UOH variables are presented in Table 4.7. In contrast to the

GLS results, which showed the significance of ***PU1***, ***MarginalPU***, and ***TeamUncertainty***, using OLS to find what aspects of UOH influence demand for individual MLB games, all of the UOH variables were significant contributors with the exception of ***LeagueUncertainty*** and ***AcrossSeasonBalance***. This is noteworthy principally because the standard OLS approach to questions of UOH would have given undue significance to ***GU1***, ***GU2***, ***PU2***, and ***WithinSeasonBalance*** that is actually attributable to the serial correlation in demand for successive games.

Conclusions and Discussion

This chapter explores how traditional and novel proxies of uncertainty influence demand for MLB contests. It is the first study in many years to use game-by-game attendance as the dependent term and first among all the uncertainty literature to quantify the probability of game success and playoff qualification via a probit model. The probit demonstrates that both the home team and road team's winning percentage are important in predicting game outcomes, while strength-of-schedule is not significant for either team. Furthermore, a team's probability of making the playoffs, the importance of the game in terms of playoff qualification, and the change in winning percentage from one year prior are the significant uncertainty factors in estimating demand for MLB regular-season games. I feel this research adds to the rich canon of work on uncertainty of outcome and provides a springboard for further exploration on a number of related questions.

Fort (2006) emphasizes the need for the examination of UOH principles across more than just MLB. I offer this first analysis of utilizing a probit model to ascertain the actual probability of game and playoff success at the time of any contest and echo the recommendation to expand the use of this model to other leagues.

It is also evident that many of the new metrics introduced in this chapter fall outside of the GU/PU/CSU paradigm. For example, what I call "Team Uncertainty" is connected both to the idea of GU and CSU. So while the metrics are named for their resemblance to the traditional approach to parsing out different types of uncertainty, it may be useful to merely separate the notions quality and uncertainty.

Finally, I reflect back on the utility of this type of analysis. As we learn more about fan preferences, we consider how practitioners in the sport industry can make use of this information. In this case, it is apparent that those who create the MLB schedule can look to this study's findings to, at their leisure, maximize fan welfare or owner profits. In either case, all else equal it may be advisable to construct a schedule that projects to maximize games of great playoff importance. No less important are the implications regarding fans' desire for consistent team performance from season to season.

Table 4.1

Summary Statistics of Winning Percent and Strength of Schedule

Parameter	**Min.**	**Max.**	**Mean**	**SD**
Home Team Win%	0.238	0.754	0.500	0.074
Home Team SOS	0.456	0.540	0.500	0.011
Road Team Win%	0.238	0.765	0.501	0.074
Road Team SOS	0.451	0.539	0.500	0.011
Predicted Probabilities	0.463	0.537		
Actual Probabilities	0.463	0.536		
Correct Predictions	0.565			

N=14,756

Figure 4.1

Success of Probit Model in Predicting Game Outcomes

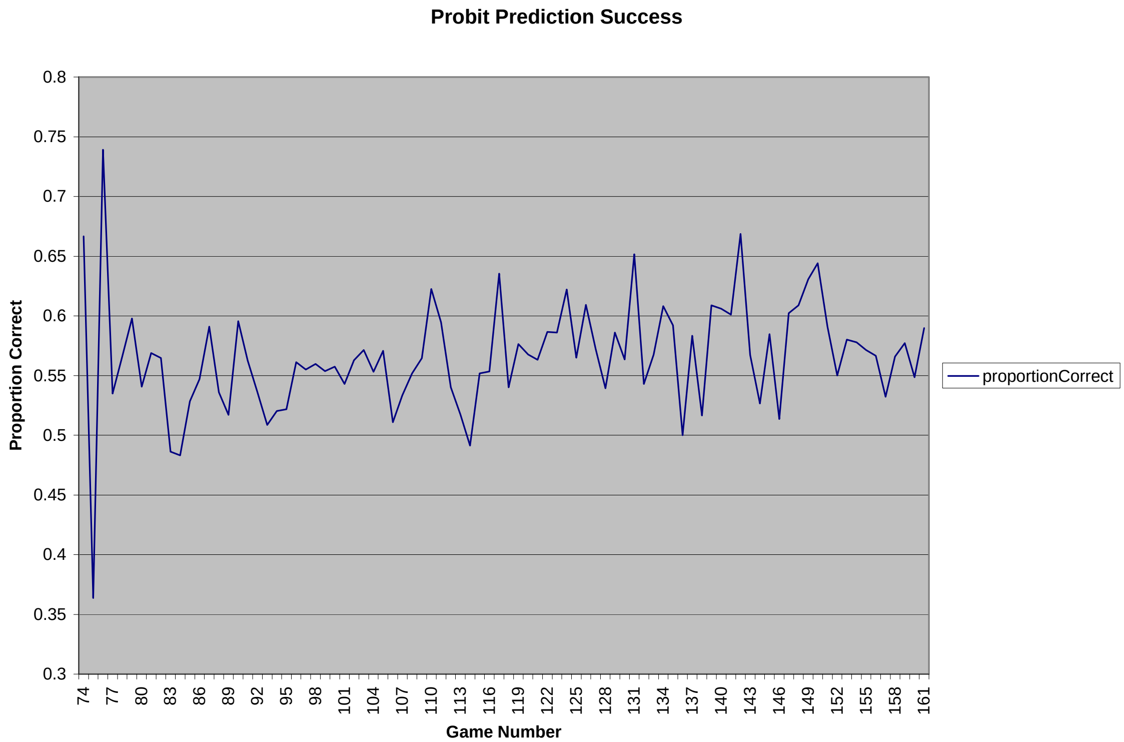


Table 4.2

Probit Model of MLB Contests

Parameter	**Estimate**	**SD**	**T**	**Marginal Effect**
Constant	0.256	0.719	0.357	0.102
Home Team Win%	1.535***	0.162	9.494	0.609
Home Team SOS	0.778	1.083	0.719	0.309
Road Team Win%	-1.737***	0.163	-10.677	-0.690
Road Team SOS	-0.902	1.093	-0.826	-0.358

N=14,756

Table 4.3

Summary Statistics for Factors Affecting Demand in MLB

Parameter	**Min.**	**Max.**	**Mean**	**SD**
Win%	0.238	0.754	0.501	0.074
Opponent's Win%	0.238	0.765	0.500	0.075
August	0	1	0.343	0.475
September	0	1	0.317	0.465
October	0	1	0.018	0.132
Monday	0	1	0.098	0.297
Tuesday	0	1	0.144	0.351
Wednesday	0	1	0.144	0.351
Thursday	0	1	0.123	0.328
Friday	0	1	0.160	0.367
Saturday	0	1	0.166	0.372
Day Game	0	1	0.318	0.466
1998	0	1	0.083	0.275
1999	0	1	0.095	0.293
2000	0	1	0.094	0.293
2001	0	1	0.093	0.290
2002	0	1	0.092	0.289
2003	0	1	0.091	0.288
2004	0	1	0.092	0.289
2005	0	1	0.090	0.286
2006	0	1	0.093	0.291
2007	0	1	0.093	0.290
Turner	0	1	0.034	0.181
Camden Yards	0	1	0.034	0.181
Fenway	0	1	0.035	0.183
Comiskey	0	1	0.034	0.181
Wrigley	0	1	0.034	0.181
Riverfront	0	1	0.019	0.185
Great American BP	0	1	0.016	0.124
Jacobs	0	1	0.035	0.183
Coors	0	1	0.035	0.183
Tiger	0	1	0.009	0.093
Comerica	0	1	0.025	0.158
Dolphins	0	1	0.033	0.178
Astrodome	0	1	0.009	0.096
Minute Maid	0	1	0.025	0.157
Kaufmann	0	1	0.033	0.180
Angels	0	1	0.034	0.182
Dodgers	0	1	0.034	0.182
Milwaukee Cty	0	1	0.012	0.111

Miller	0	1	0.022	0.146
Metrodome	0	1	0.033	0.180
Olympic	0	1	0.021	0.143
Shea	0	1	0.034	0.180
Yankee	0	1	0.035	0.183
McAfee	0	1	0.035	0.183
Veterans'	0	1	0.021	0.144
Citizens Bank	0	1	0.012	0.110
Three Rivers	0	1	0.012	0.109
PNC	0	1	0.021	0.145
Jack Murphy	0	1	0.021	0.144
Petco	0	1	0.012	0.111
Kingdome	0	1	0.006	0.078
Safeco	0	1	0.028	0.164
Candlestick	0	1	0.008	0.091
AT&T	0	1	0.025	0.155
Old Busch	0	1	0.029	0.167
New Busch	0	1	0.006	0.179
Tropicana	0	1	0.028	0.165
Arlington	0	1	0.034	0.182
Skydome	0	1	0.033	0.178
ATL	0	1	0.032	0.175
BAL	0	1	0.034	0.181
BOS	0	1	0.034	0.182
CHA	0	1	0.034	0.182
CHN	0	1	0.033	0.179
CIN	0	1	0.033	0.180
CLE	0	1	0.034	0.181
COL	0	1	0.033	0.179
DET	0	1	0.034	0.182
FLA	0	1	0.033	0.178
HOU	0	1	0.032	0.176
KCA	0	1	0.034	0.182
LAA	0	1	0.034	0.181
LAN	0	1	0.033	0.179
MIL	0	1	0.033	0.180
MIN	0	1	0.035	0.183
MON	0	1	0.025	0.157
NYA	0	1	0.034	0.181
NYN	0	1	0.032	0.176
OAK	0	1	0.033	0.178
PHI	0	1	0.033	0.179
PIT	0	1	0.033	0.179
SD	0	1	0.034	0.180
SEA	0	1	0.034	0.181
SF	0	1	0.034	0.182

STL	0	1	0.033	0.178
TB	0	1	0.031	0.173
TEX	0	1	0.034	0.180
TOR	0	1	0.034	0.182
Last Year Stadium	0	1	0.044	0.205
First Year Stadium	0	1	0.036	0.185
Second Year Stadium	0	1	0.030	0.170
Ticket Price	8.22	47.71	18.213	6.272
Runs	3.163	6.625	4.843	0.513
Opponent's Runs	3.157	6.654	4.834	0.521
Game Uncertainty1	0.307	0.755	0.537	0.068
Game Uncertainty2	0	0.510	0.125	0.091
Team Uncertainty	-0.268	0.249	0.000	0.076
Playoff Uncertainty1	0	1	0.270	0.362
Playoff Uncertainty2	0	1	0.808	0.289
Marginal PU	-0.012	0.515	0.035	0.051
League Uncertainty	0.257	0.460	0.366	0.041
Within Season Balance	0.251	0.262	0.257	0.002
Across Season Balance	-0.048	0.882	0.523	0.223

N=13,272

Table 4.4

Correlation Matrix of Uncertainty Variables

	GU1	**GU2**	**TU**	**PU1**	**PU2**	**MPU**	**LU**	**WSB**	**ASB**	**Win%**
GU1	1	0.616	0.343	0.568	-0.287	0.260	0.019	-0.008	-0.012	0.694
GU2	0.616	1	0.200	0.471	-0.126	0.079	0.035	0.056	-0.007	0.420
TU	0.343	0.200	1	0.377	-0.231	0.203	0.013	-0.014	-0.023	0.504
PU1	0.568	0.471	0.377	1	-0.347	0.338	0.013	-0.004	-0.008	0.810
PU2	-0.287	-0.126	-0.231	-0.347	1	-0.825	0.297	0.009	0.026	-0.415
MPU	0.260	0.079	0.203	0.338	-0.825	1	-0.163	-0.042	0.017	0.404
LU	0.019	0.035	0.013	0.013	0.297	-0.163	1	0.019	0.023	0.015
WSB	-0.008	0.056	-0.014	-0.004	0.009	-0.042	0.019	1	0.258	-0.011
ASB	-0.012	-0.007	-0.023	-0.008	0.026	0.017	0.023	0.258	1	-0.021
Win%	0.694	0.420	0.504	0.810	-0.415	0.404	0.015	-0.011	-0.021	1

Table 4.5

Tests of Autocorrelation and Heteroskedasticity

	Parameter	SE	T
Heteroskedasticity			
Constant	5.162	0.593	8.704
Linear Error	-0.928	0.117	-7.947
Squared Error	0.042	0.006	7.304
Autocorrelation			
Constant	0.001	0.002	0.738
Previous Game	0.502	0.008	65.547

Table 4.6

Generalized Least Squares Results with Clustered Standard Errors

Demand Estimation for UOH and MLB Games

Parameter	**Estimate**	**SD**	**T**
Constant	8.413***	1.098	7.660
Win%	1.525**	0.714	2.136
Opponent's Win%	0.282	0.750	0.375
August	-0.017**	0.008	-2.033
September	-0.078***	0.017	-4.511
October	-0.079*	0.044	-1.814
Monday	-0.189***	0.012	-15.260
Tuesday	-0.194***	0.011	-17.453
Wednesday	-0.172***	0.010	-16.722
Thursday	-0.155***	0.010	-16.292
Friday	-0.043***	0.010	-4.437
Saturday	0.072***	0.008	9.461
Day Game	-0.026***	0.007	-3.684
1998	-0.004	0.048	-0.083
1999	-0.019	0.049	-0.396
2000	-0.046	0.052	-0.886
2001	-0.054	0.049	-1.101
2002	-0.134**	0.053	-2.521
2003	-0.123**	0.052	-2.371
2004	-0.066	0.054	-1.208
2005	-0.092*	0.056	-1.664
2006	-0.057	0.061	-0.941
2007	-0.016	0.058	-0.279
Turner	-0.088**	0.044	-1.990
Camden Yards	0.106	0.070	1.525
Fenway	-0.359***	0.092	-3.905
Comiskey	-0.309***	0.057	-5.446
Wrigley	0.098*	0.051	1.942
Riverfront	-0.184***	0.053	-3.484
Great American BP	-0.167***	0.043	-3.844
Jacobs	-0.064	0.066	-0.969
Coors	0.172**	0.070	2.464
Tiger	-0.214*	0.116	-1.837
Comerica	-0.120**	0.056	-2.162
Dolphins	-0.483***	0.087	-5.530
Astrodome	-0.165***	0.051	-3.249
Minute Maid	0.008	0.042	0.190
Kaufmann	-0.352***	0.047	-7.494

Angels	0.003	0.069	0.039
Dodgers	0.247***	0.039	6.331
Milwaukee Cty	-0.246***	0.069	-3.574
Miller	0.005	0.043	0.110
Metrodome	-0.349***	0.075	-4.675
Olympic	-0.874***	0.095	-9.153
Shea	-0.013	0.056	-0.213
Yankee	0.029	0.070	0.415
McAfee	-0.371***	0.058	-6.447
Veterans	-0.239***	0.058	-4.093
Citizens Bank	-0.067	0.050	-1.349
Three Rivers	-0.321***	0.049	-6.515
PNC	-0.193***	0.050	-3.865
Jack Murphy	-0.041	0.047	-0.876
Petco	-0.028	0.051	-0.544
Kingdome	-0.035	0.111	-0.316
Safeco	0.034	0.050	0.686
Candlestick	-0.247***	0.062	-4.008
AT&T	0.097**	0.048	2.036
Old Busch	0.067	0.058	1.158
New Busch	-0.071	0.106	-0.670
Tropicana	-0.448***	0.059	-7.604
Arlington	-0.051	0.049	-1.037
Skydome	-0.160***	0.054	-2.980
ATL	0.068***	0.024	2.871
BAL	0.032	0.030	1.069
BOS	0.147***	0.031	4.694
CHA	-0.010	0.030	-0.340
CHN	0.192***	0.024	8.011
CIN	0.049**	0.021	2.314
CLE	0.056*	0.029	1.926
COL	0.001	0.025	0.046
DET	-0.005	0.029	-0.168
FLA	-0.021	0.022	-0.956
HOU	0.007	0.021	0.325
KCA	0.002	0.030	0.073
LAA	0.000	0.029	-0.004
LAN	0.073***	0.021	3.489
MIL	-0.024	0.021	-1.160
MIN	-0.015	0.027	-0.541
MON	-0.058**	0.027	-2.110
NYA	0.251***	0.033	7.555
NYN	0.068***	0.019	3.638
OAK	0.017	0.027	0.631
PHI	-0.021	0.026	-0.815
PIT	-0.004	0.023	-0.191

SD	0.015	0.019	0.784
SEA	0.059**	0.027	2.165
SF	0.082***	0.021	3.962
STL	0.109***	0.026	4.219
TB	-0.009	0.030	-0.281
TEX	-0.022	0.030	-0.739
TOR	-0.005	0.028	-0.180
Last Year Stadium	0.058**	0.027	2.137
First Year Stadium	0.138***	0.036	3.779
Second Year Stadium	0.105**	0.045	2.358
Ticket Price	0.013***	0.003	4.381
Runs	0.019	0.024	0.802
Opponent's Runs	0.016	0.010	1.589
Game Certainty1	0.070	1.154	0.061
Game Certainty2	0.027	0.062	0.442
Team Uncertainty	-0.751***	0.118	-6.376
Playoff Uncertainty1	0.058*	0.035	1.669
Playoff Uncertainty2	0.016	0.028	0.554
Marginal PU	0.383***	0.123	3.124
League Uncertainty	0.142	0.198	0.720
Within Season Balance	2.556	3.114	0.821
Across Season Balance	-0.002	0.052	-0.031

N=13,272

R^2=0.671

Table 4.7

Ordinary Least Squares Results

Demand Estimation for UOH and MLB Games

Parameter	**Estimate**	**SD**	**T**
GameUncertainty1	1.482**	0.571	2.593
GameUncertainty2	0.062*	0.034	1.804
TeamUncertainty	-0.907***	0.039	23.007
PlayoffUncertainty1	0.071***	0.012	6.012
PlayoffUncertainty2	0.060***	0.015	4.022
MarginalPU	0.745***	0.080	9.311
LeagueUncertainty	0.159	0.120	1.324
WithinSeasonBalance	4.301**	1.706	2.520
AcrossSeasonBalance	-0.048	0.034	1.417

N=13,272

References

Baade, R.A. & Tiehen, L.J. (1990). An Analysis of Major League Baseball Attendance, 1969-1987. *Journal of Sport & Social Issues.* 14(1), 14-32.

Baseball-Reference (n.d.). Last viewed on 07/24/08 at http://www.baseball-reference.com/

Butler, M.R. 1995. Competitive Balance in Major League Baseball. *American Economist* 39, 46-52.

Cairns, J.A. 1987. Evaluating Changes in League Structure: The Reorganization of the Scottish Football League. *Applied Economics* 19, 259-275.

Coates, D & Humphreys, B.R. (2005). Baseball Strikes and the Demand for Attendance. *Journal of Sports Economics.* 6(3), 282-302.

Demmert, H.G. (1973). *The Economics of Professional Team Sports*. Lexington,MA:Lexington Books.

Domazlicky, B.R., and Kerr, P.M. (1990). Baseball Attendance and the Designated Hitter. *American Economist* 34(Spring), 62-68.

Fort, R.D. (2006). Competitive balance in North American professional sports. In J. Fizel (Ed.), Handbook of sports economics research (pp. 190-208). Armonk, NY: M.E. Sharpe.

Fort, R.D., and Lee, Y.H. (2006). The Uncertainty of Outcome Hypothesis: Major League Baseball Time Series. Working Paper, October, 2006.

Hill, J.R., Madura, J. & Zuber, R.A. (1982) The Short Run Demand for Major League Baseball. *Atlantic Economic Journal.* 10, 31-35.

Humphreys, B.R. (2002). Alternative Measures of Competitive Balance in Sports Leagues. *Journal of Sports Economics*. 3(2), 133-148.

Knowles, G, Sherony, K, & Haupert. (1992) The Demand for Major League Baseball: A Test of the Uncertainty of Outcome Hypothesis. *American Economist*. 36, 72-80.

Koning, R.H. (1999), Competitive Balance in Dutch Soccer, SOM Research Report 99B04, Faculty of Economics, University of Groningen.

Noll, R. (1974). Attendance and Price Setting. In *Government and the Sport Business*. Washington, DC: Brookings Institution.

Quirk, J., and R. D. Fort. (1997). *Pay Dirt: The Business of Professional Team Sports*, 2nd ed. Princeton, NJ: Princeton University Press.

Rottenberg, S. 1956. The Baseball Player's Labor Market. *Journal of Political Economy*. 64, 242-258.

Schmidt, M.B. & Berri, D.J. (2001). Competitive Balance and Attendance: The Case of Major League Baseball. *Journal of Sports Economics*. 2, 145-167

Schmidt, M. B. and Berri, D. (2004). The Impact of Labor Strikes on Consumer Demand: An Application to Professional Sports. *American Economic Review*, 94(1): 344-357.

Scully, G. W. (1989). *The Business of Major League Baseball*, Chicago: University of Chicago Press.

Szymanski, S. (2003). The Economic Design of Sporting Contests. *Journal of Economic Literature*. 41(4), 1137-1187.

Whitney, J.D. (1988). Winning Games Versus Winning Championships: The Economics of Fan Interest and Team Performance. *Economic Inquiry*, 26, 703-724.

Winfree, J., McCluskey, J., Mittelhammer, R. and Fort, R. (2004). Location and Attendance in Major League Baseball. *Applied Economics* 36(19): 2117-2124.

Chapter V

Conclusions

Summary of Findings

This dissertation takes a look at three types of issues in MLB related to demand. Chapter II focuses on the revenue change associated with the major policy change MLB undertook in testing for performance-enhancing drugs and implementing penalties on the offending players. The impact of the findings extend to most if not all sports as Major League Baseball is not the sole organization whose athletes have used PEDs to gain the upper hand on their competition. Chapter III looks at consumer discrimination with respect to international players in MLB. I find an evolution of the effect of foreign-born players on consumer demand. Over the course of two decades, fans went from preferring American-born players to favoring those born abroad. A follow-up study is warranted since the degree of preference flattens out in the final few seasons included in the study. Chapter IV tests nine types of uncertainty in MLB contests. Using probit analysis to establish probabilities of winning a given game and, along with the probit, Monte Carlo simulations to determine the probability of playoff qualification, the nine uncertainty metrics were placed in a regression analysis with game attendance as the dependent variable. The analysis shows how fans turn out in greater numbers for games in which the outcome is of great importance in terms of playoff qualification and when their team is more likely to qualify for the playoffs as well as that fans react adversely to changes in team performance from the previous season. These findings are consistent with some of the recent work on UOH. Just as interesting are the metrics not found to significantly impact demand, contradicting some of the established theories of fan preferences for close games and playoff uncertainty.

The totality of these studies representing three different areas of research on demand in MLB is a well-rounded view of the types of studies that aim to gauge fan preferences. The decision to implement drug testing is one of many policies the league has enacted. Any of these policies can be looked at in a manner similar to that in the Chapter II to measure the impact on revenues. So too the Chapter III represents a type of inquiry undertaken by sport researchers. The presence of foreign players is notable, and having increased over time, warrants the same attention of integration in estimating demand. Finally, the longstanding notion that fans desire uncertainty is tested in the Chapter IV. This is something that is unique to sport, as in most forms of entertainment the outcome has been decided in advance. Their place within different subsets of the research on sport economics notwithstanding, each of these chapters reveals to us something about fan preferences and their impact on demand.