The State of Research on Markets for Sports Betting and Suggested Future Directions

Raymond D. Sauer
Clemson University
September 2005

I thank Jahn Hakes and Chad Turner for their help in estimating models on which much of this analysis depends, and John Gandar and Rick Zuber for kindly supplying data used in this paper.

Introduction: The State of the Literature

This paper is written from the perspective of a writer who, about a decade ago, began absorbing the economic literature on wagering markets and developing an interpretation of its findings (Sauer, 1998). Although any literature is a mixed bag with both insight and error, the early papers on wagering markets were a remarkable lot. They combined data from unique and valuable institutional settings with appropriate theoretical and econometric tools. Clever exploitation of this "natural laboratory" created a rich set of findings on the nature of prices in these markets.

The central theme of this literature is the efficient markets hypothesis. In this context, efficiency implies that odds or prices summarize all that is known about a event, and more generally, implies the absence of profitable betting opportunities.

The early literature established that efficiency was a very useful benchmark for actual prices observed in betting markets. Although there are many published exceptions which reject efficient pricing, it is the benchmark result that prices are approximately efficient which yields insights which economics should be proud to lay claim to. For example, odds at the racetrack imbed subjective estimates of the probability of winning that are close to those obtained using sophisticated statistical methods, and point spreads are unbiased predictors of score differences in football and basketball games. These may be surprising findings to those tied to the principle that "all bettors are fools," but not to scholars with an understanding of economics.

That is an important, and underappreciated body of work. Yet the most interesting papers were those whose results depart in interesting ways from canonical representations of the efficiency hypothesis. Papers documenting price adjustments, differential rates of return across betting time and bettor type, and optimizing responses to inside information by price-makers, together make it clear that the formation of anything approaching an efficient price does not take place in a vacuum. Prices, at least at this point in history, are not formed in a frictionless world with complete and fully dispersed information. Rather, the process of pricing itself, as Hayek (1945) emphasized, aggregates, creates, and communicates information.

It is this aspect of pricing, on full display in many early papers (Asch et. al, 1982, Crafts, 1985, Brown and Sauer, 1993, Gandar et. Al, 1998), that is both compelling, yet puzzlingly absent in much of the recent literature. In fact, the literature on sports betting has grown somewhat stale, as evidenced by the lack of published articles in the leading journals in recent years. The typical paper fine-tunes a specification from an existing model, or even more common, conducts efficiency/profitability analysis in a narrow sense.

More creative approaches are available within the efficient markets paradigm. In particular, creative application of the efficient markets hypothesis can reveal information in ways that are otherwise quite difficult. This notion has been readily adopted in the emerging event or information markets literature (Hanson, 2005, and Wolfers and

Zitzewitz, 2004). Papers on event markets tend to skip over the literature on sports for various reasons, the principal one being the nature of the questions addressed by the typical paper. And while sports papers continue to look for profit opportunities in the back alleys of betting markets, the big picture on pricing of events - sporting or otherwise - has already been absorbed. Hanson (2005), a pioneer in the design and study of information markets, puts it this way:

> "So far, speculative markets have done well in every known head to head field comparison with other forecasting institutions. Orange Juice futures improve on National Weather Service forecasts, horse race markets beat horse race experts, Oscar markets beat columnist forecasts, gas demand markets beat gas demand experts, stock markets beat the official NASA panel at fingering the guilty company in the Challenger accident, election markets beat national opinion polls, and corporate sales markets beat official corporate forecasts."

Meanwhile, papers focused on sports betting continue to report rejections of simple efficiency and fragile evidence of profitable betting opportunities. But scholars outside of the club are not listening, so it is time to rethink our approach to this field.

Constructive Applications of the Efficiency Hypothesis

The narrow question of whether a market is efficient or not is no longer interesting. The benchmark result referred to in the introduction is widely known. It is equally obvious from the published literature that models rejecting efficient pricing are easily produced, although perhaps less easy to interpret.

What does one make of this apparent contradiction? The resolution in Camerer's (1989) hot hand analysis makes a good deal of sense. A class of bettors may be subject to occasional biases - in Camerer's case, over-estimation of the relevance of being on a temporary hot streak. Their influence on betting prices is limited, however, by the existence of informed bettors, who capitalize on pricing discrepancies. There is ample evidence of differentially informed pools of bettors in the literature, and adjustments that result when the relatively informed move prices towards the efficient benchmark. But since transactions costs inhibit complete adjustment, final prices can leave traces of bettor misperceptions that are not completely washed away by informed trading.

Some deviations of prices from the efficiency benchmark are persistent, and evidence of cognitive biases. Yet these findings, thought they make use of efficiency as a foil, do not replace the concept with a more useful, general tool for the study of prices. Statistical rejection of efficiency does not imply a large gap between prices in the data and the benchmark. Nor does documenting an apparently profitable betting rule imply that prices depart sufficiently from the benchmark to render the discrepancy exploitable (Brailsford et. al, 1995). Simply put, most nominal rejections of efficient pricing in this literature are accurately viewed as minor deviations from the benchmark.

The stale nature of the current literature is due to the manner in which the efficiency hypothesis is employed. The questions addressed come predominately from a behavioralist perspective - what bias are we testing for today? The results of this enterprise may bring occasional cheer to behavioral partisans, but by design, they are incapable of much more. And a failure to reject the efficiency hypothesis these days brings nothing but yawns from everyone.

The literature on sports betting minimizes its influence on economics because of the nature of the analyses pursued. While these may be relevant to aficionados, we can and should expand our sights by rethinking the questions we ask of the data. The narrow focus on questions of profitability and efficiency should be abandoned, and fresh, creative approaches should be explored. Two suggestions explored in the following sections are (1) applications in which efficient prices are assumed, with the aim being to discover information revealed by the pricing mechanism; and (2) studies of variation in real time pricing in betting networks to better understand the process by which efficient prices are determined.

These might be called constructive applications of the efficiency hypothesis to wagering markets. The purpose is not to "resurrect" efficiency per se, but to suggest ways in which the hypothesis can contribute to a literature that gets more recognition in the profession. What *is* essential is a creative infusion of ideas into the field. There are signs in recent working papers that this may be imminent. Strumpf (2003), for example, studies the records of illegal bookmakers seized in New York City. Among other findings, Strumpf reports clear evidence of price discrimination against known Yankee partisans. Weinbach (2005) studies pari-mutuel betting on greyhound races, focusing on liquidity constrained agents. Weinbach documents a pattern in volume which implies that relatively poor recipients of transfer payments account for spikes coincident with the arrival of checks in the mail from the federal government. Neither of these papers has an axe to grind on the efficiency question, but each contains the sort of fresh ideas required to bring relevance to the literature.

Two Examples

This section presents two constructive applications of the efficiency hypothesis from the betting market for baseball games. Each example makes use of a probability model of a baseball game developed by Hakes and Sauer (2004), and presents an analysis of the pricing mechanism in a way which has yet to appear in the literature.

The first application uses pre-game betting odds to extract the betting market's estimate of the impact of starting pitching on the probability of winning. The market's estimate is then compared to the actual performance of pitchers. This comparison is used to test a conjecture by Bill James on the value of pitching relative to other skills in baseball.

The second application compares real-time betting prices within a ballgame to corresponding probabilities derived from the fundamentals in Hakes and Sauer (2004).

The ballgame chose for study is one associated with a famous "crash" - an event in which trading is commonly alleged to go haywire. Somewhat crude comparisons between the real-time prices and the model's probabilities (this research is at an early stage) are nevertheless quite illustrative.

Baseball Betting Markets and the Bill James Conjecture on the Value of Pitching

Extracting the value of a starting pitcher from betting market data is a straightforward estimation problem. Betting odds imply probabilities of winning under the efficiency hypothesis, hence variation in the odds and the identity of the starting pitchers should be sufficient to identify the impact of a particular pitcher on his teams' chances of winning. Although perhaps of limited relevance on its own, this information from the betting market can be combined with the probability model to address an age-old question: "how important are the relative skills of hitting, pitching, and fielding in major league baseball."

Hakes and Sauer first confronted this question when we attempted to construct a probability-based measure of player performance. The assumption of a stationary distribution of scoring across teams and innings generates powerful implications. In particular, it allows a probability of winning to be calculated for either team, given any situation (summarized by the team, inning, outs, & runners on base). Plays, whether they be a strikeout, a hit, or anything else, alter the state of the game and thus the probability of winning. (See Hakes and Sauer, 2004, for a complete description of the method).

Players involved in a given play are responsible for the change in probability associated with the outcome. But how should that change be apportioned across players? Some sense of the representative impact of batters, pitchers, and fielders on the "typical" play is necessary if one wishes to apportion probability across players if one wishes to compare value across different players at different positions.

Answers to this question have been posed, but many have been dogmatic, and none of the answers has been tested. Connie Mack, the manager holding the record for lifetime wins, is alleged to have claimed that pitching was "eighty per cent of the game.[1]" Earl Weaver, Mack's modern-day counterpart, claimed that "the reason you win or lose is darn near always the same - pitching." [2] While there is certainly some truth in this, when a great pitcher has an off day, good hitters will make more of the situation than poor hitters, which suggests that these managers may overstate the case.

Bill James, the modern dean of baseball statistical analysts, makes a such conjecture in his recent book *Win Shares.* The object of win shares is to apportion the credit for victories across players at different positions, but there is no "pure approach" for doing so that is available. James thus resorts to conjecture (without any empirical or theoretical support): "48% of Win Shares are assigned to hitters/baserunners, 35% are assigned to

[1] http://www.baseballlibrary.com/baseballlibrary/ballplayers/M/Mack_Connie.stm
[2] http://www.baseball-almanac.com/quotes/quoweav.shtml

pitchers, and 17% are assigned to fielders (James and Henzler, 2002, p. 10)." While one must respect the past accomplishments of James, the same must be said of the winningest manager in baseball, and their estimates of the proportion of the game attributed to pitching - 80 percent vs. 35% - are quite far apart. It turns out that the betting market, combined with the probability can provide a resolution of this dispute.

The first part of the resolution is to employ the efficient markets hypothesis: we assume that the betting market correctly estimates the effect of starting pitchers on the probability of winning a game. The data used are pitching matchups and betting odds from Computer Sports World for the 1999 season.[3] An example set of odds is presented in Table 1 for a game between where the Reds visited the Houston Astros.

A three step procedure is used to convert the odds to probabilities. First, adjust the "Money Line" representation of the odds into "true" odds per $1 bet. Since"+140" indicates that a winning bettor nets $140 for every $100 bet, the odds Equivalent is 1.4/1. Similarly, "-150" indicates that a winning bet of $150 on the Astros would net $100. Hence the true odds per $1 are /1.5 or 0.67.

Second, if the bets were fair, they would yield a zero expected return. These defines the implied fair probability of winning, $p^{FAIR} = 1/(1+\text{Odds})$. But since bookmakers are not in business for the fun of it, they don't offer fair odds, and the sum of the implied fair probabilities exceeds 1.0. Hence the third step: obtain normalized fair probabilities, p^{FAIR*} by dividing each probability by the associated sum (1.0167 in this case). Probabilities for all games in the 1999 season were obtained in this fashion.

The empirical model includes individual dummy variables for each pitcher with at least 10 starts in the season (other starters are lumped into the omitted variable.) Each teams' offensive ability is measured by its OPS, a powerful statistic (in the sense of predicting runs scored) that simply adds the teams' on base and slugging percentages.[4] Fielding ability and the ability of a team's relief pitchers are proxied in the model by incorporating a 0/1 dummy variable for each team.

The model uses the home team's probability of winning as the dependent variable. The model estimated is thus $p_h^{FAIR*} = g(ops_h, ops_v, d_h, d_v, d_p)$ where d_h and d_v are 0/1 dummies for each team, and d_p is a composite pitcher dummy variable. The latter variable is constructed in the following way. Let d_{ph} and d_{pv} be 0/1 dummies for the home team's and visiting teams' pitchers, respectively. $d_{ph} = 1$ uniquely for each pitcher when pitching at his home park, and $d_{pv} = 1$ uniquely for each pitcher when pitching on the road. For parsimony, we assume that a pitcher's ability is independent of the park he is pitching in, which implies that a pitcher's impact on the home team winning when he's pitching on the road is the negative of his impact when he's at home. This is not the assumption made on the offensive side - we allow for the possibility that offense is tailored to the home teams'

---

[3] The data were graciously provided by John Gandar and Rick Zuber.

[4] A moving average of team OPS over the prior 40 games is used in estimation. For games early in the season, a weighted average of the prior year's OPS and current OPS is used, with weights declining on the prior year from 40 to 1 as each game is played.

park, at the cost of estimating one additional parameter. In contrast, the assumption of identical home/away impacts for pitchers reduces the number of estimated parameters by about 150, or a factor of 2.

We adopt a simple approach to estimation and assume a linear probability model. The cost in efficiency should not be great since the mass of data for most baseball games is near .5, indeed there are no extreme probabilities near the 0 or 1 limit. The model is thus:

$$p_h^{FAIR*} = \beta_0 + \beta_1 * ops_h + \beta_2 * ops_v + \beta_3 * d_h + \beta_4 * d_v + \beta_5 * d_p + \varepsilon$$

β_0 is thus a coefficient which measures the home field advantage, independent of an offense's greater ability in its own park. Table 2a presents summary statistics and coefficient estimates for β_0, β_1, and β_2. The model is reasonably precise, with a value of Rbar-squared of .85 and a root mean square error of .036 in probability. The estimate for β_1 exceeds β_2, suggesting that part of the advantage of playing at home derives from an offense's ability "fitting" the contours of the home team's ballpark.

The parameters of interest are in the vector of coefficients for pitchers, β_5. Table 2b reports the largest and smallest coefficient estimates for the pitcher dummy variables. The cast of characters with the largest positive impact on the probability of winning are familiar names to baseball fans. Pedro Martinez tops the list at .15, followed by Curt Schilling (.15), Kevin Brown (.14), Mike Mussina (.14) and Randy Johnson (.13). These estimates imply that each time one of these pitchers stepped to the mound in 1999, they added an incremental probability of winning of between .13 and .15 over and above run-of-the-mill (the omitted) pitchers.

The pitchers with the lowest estimates are essentially indistinguishable from "no-name" starters, the pitchers who did not start a sufficient number of times to obtain their own coefficient estimate. As a long-suffering Houston Astro fan, I certainly understand the minus.03 estimate for Chris Holt, arguably the worst pitcher in baseball that year. The one puzzling estimate is for Andy Pettitte (-.005), which suggests the betting market that year viewed him as indistinguishable from a no-name pitcher.[5]

Table 2 utilizes the efficient markets hypothesis to produce estimates of the relative value of pitchers among themselves. In addition, we understand from the Table that the best pitchers in 1999 yielded an increment in winning probability of .1 to .15. For a healthy pitcher that started 30 games, a .15 incremental probability of winning each time he stepped on the mound results in an expected return of 4.5 additional wins for his ballclub, over and above the return expected by using a run-of-the-mill pitcher. That is, a Pedro Martinez turns an average 81-81 club into and 85 or 86 win team, on the verge of playoff contention.

---

[5] Pettite is now recognized as one of baseball's best pitchers. The problem may be that he was viewed in 1999 as a minor member or perhaps indistinguishable from the Yankees staff, with the staff effect being captured in the coefficient for the Yankee dummy variable. Either possibility suggests that adjustment of the specification to capture differences in the quality of the starting rotations would be useful.

It may be that back-of-the-envelope calculations such as this is what led James to posit his conjecture that pitching was "35% of the game," but he does not provide a basis for his conjecture.[6] In any event, we can test this conjecture using play-by-play data for the 1999 season and the a model of the probability of winning the game.[7] Each play changes the state of the game, and hence the probability that a given team is the victor. Table 3 presents results obtained in Hakes and Sauer (2004) for the typical walk, single, double, etc. One can see in the table that the average single is about 1.5 times as valuable as the average walk, with a similar ratio for doubles to singles. The typical out, of course, reduces a team's chances of winning.

For each play in the 1999 season, the average value of that play is recorded, and a portion of that value (the change in probability) is assigned to the pitcher responsible for the batter. The apportionment is determined by the Bill James conjecture: 35% of the values in Table 3 are assigned for each play to the pitcher. These changes are then summed over all plays in which the pitcher participated, and divided by the number of games started. Define this statistic as DP_sum.

We are now in position to conduct a joint test of market efficiency and the Bill James conjecture. If the market prices games efficiently, the implicit impact of each pitcher on the probability of winning a game is given by β_5. If James is approximately correct, then the net actual impact of the pitcher on a game is DP_sum. The hypothesis that James is correct, under the maintained assumption of efficiency, is that across pitchers, β_5 = DP_sum. This is easily tested.
Table 5 presents the results of the following regression:

$$\text{DP\_sum} = \gamma_0 + \gamma_1 * d_p + \nu$$

Under the hypothesis that $\gamma_0 = 0$ and $\gamma_1 = 1$, the regression implies that the betting market's evaluation of the player's impact on the probability of winning equals the Jamesian apportionment of that player's actual impact on the game. The model's explanatory power (.36) is quite good, and the estimated coefficients (-.02, 1.069) accord well with this hypothesis, which cannot be rejected at the .05 level. Bill James' conjecture is given support by the betting market. In contrast to the hyperbole of Christy Mathewson, this analysis confirms that the contribution of pitching skill to game outcomes is roughly 35% of the value of plays made during the game.

---

[6] James was not the first to come out with a such a figure. The classic board game and baseball simulation "Strat-O-Matic," anticipated James and Henzler's percentage about 40 years before the fact. Strat-O-Matic was designed so that half of the plays were determined on the hitter's card, and half on the pitcher's card, with "X-chances" for fielding ability distributed on the pitcher's card such that the weights were 3/6 hitting + 2/6 pitching + 1/6 fielding (50-33-17).
[7] The play-by-play data we use were purchased by Stats Inc, the supplier of official statistics to such sites as ESPN.com.

Betting on Baseball in Real Time

Recent innovations in internet-based betting markets allow trading to occur on a continuous basis while a game is being played. In a typical market, trading is based on a contract which pre-specifies a return R, where = $1. Offers to buy and sell various quantities of these contracts are posted and revised on a continuous basis while the game is being played.

Consider a game between the Chicago Cubs and an opponent, with a contract that pays $1 if the Cubs win the game, and $0 otherwise. Trading in the market determines the price of this contract, i.e. c, the cost of purchasing a contingent claim to $1 should the Cubs win.

Let there be a probability p that the Cubs win some point during the game. At that moment, any wager $c < pR$ that would realize R at the game's conclusion would be a profitable ex ante. Obviously, such profit opportunities are ruled out in an efficient betting market. Market efficiency in the presence of transaction costs implies that $c >= pR$ for all bets at all times during a game.

For simplicity, assume zero transaction costs in the betting. The efficiency condition then becomes an equality: $c = pR$. Further, for the case of a simple $1 contract, the price of the contract would equal the probability that the Cubs win.

In what follows, we examine this relation for a famous playoff game, Game 6 of the National League Championship Series between the Cubs and Marlins in 2003. Data from Tradesports.com was obtained for all of the trades made during this game. This data includes the time of each bet, the size, and the price c. Figure 1 contains the series of prices, a series which appears to contain a spectacular crash.

If one were analyzing asset markets, the temptation to refer to this as the bursting of a speculative bubble might arise. But in this case, we have the opportunity to refer to a continuous stream of fundamentals-based prices - i.e. the estimate of the probability that the Cubs would win, which is synonymous with the efficient price. Actual plays, their timing, and their impact on the state of the game were recorded by listening to the recorded broadcast stored at MLB.com on "GameDay Audio." After each play, the Hakes and Sauer (2004) model was be used to calculate the probability p that the Cubs would win.

Putting the prices together with the probabilities confirms a remarkable match, particularly around the time of the crash. This research is at an early stage, so this visual impression sums up all that can be said at this point.

Several data and modeling issues inhibit the relevance of formal statistical tests. First, since the time stamps for prices and plays in the two series are relative and not absolute, the two series were synchronized in an ad hoc manner by matching the largest change in probability with the largest change in price. This achieves no better than a close approximation, leaving some trades and plays potentially mismatched in time.[8]

In addition, the probability model is somewhat naïve, and does not incorporate differences in team ability or the quality of the starting pitchers. Although not a serious issue when using it to measure differences in probability, as in the prior example. But it does affect conclusions based on levels. In particular, the naïve assumption affects the starting value for p, which is 0.5 for the model. The starting value for the price is much higher, and remains so for several innings. This gap is likely accounted for by the presence of the Cub's ace pitcher, Mark Prior, on the mound. The gap between the two series shrinks over time, as the relevance of the initial conditions fades as actions takes place. In this game, Prior shut out the Marlins for the first seven innings, while the Cubs built a seemingly safe 3-0 lead. Both prices and probability reached a maximum near 0.95.

Then came the fateful eighth inning. With one out and a man on first, a Mr. Bartman interfered with Moises Alou's catch of a foul ball. From there, things went from bad to worse, as the Marlins scored an improbable eight unanswered runs to win the game, 8-3. In the Figure 2, the dramatic crash in both prices and probabilities is clearly evident. What is particularly noteworthy is how close these series match each other *as the crash is taking place*. In this case, the model supplies a set of fundamentals which matches the crash in prices, play-by-play.

Conclusion

This paper argued that the literature on betting markets has become stale, and that changes in the nature of the questions asked in the literature are required for it to recapture relevance among economists. Much work continues to be focused on applications of the efficient markets hypothesis. Rejections of this hypothesis have been a vehicle for expanding the domain of behavioralist principles (e.g. Camerer, 1989). But the impression that results from these studies, combined with rather boring failures to reject in others, is misleading. It is hoped that the two examples presented here provide evidence that much remains to be learned through creative application of this paradigm.

Betting market prices are packed with information that is worth extracting. The literature can make real progress by constructing creative models which make sense of these prices, models which uncover the information produced in these markets but hidden elsewhere. The hypothesis of efficient pricing is an essential ingredient in this enterprise.

---

[8] We are working on an algorithm to rid the data of this problem.

References

Asch, Peter; Burton G. Malkiel, and Richard E. Quandt. 1982. “Racetrack Betting and Informed Behavior,” J. Finan. Econ., 10, pp. 187–94.

Berg, Joyce, Forrest Nelson, and Thomas Rietz, “Accuracy and Forecast Standard Error of Prediction Markets,” Technical Report, University of Iowa, College of Business Administration, 2001.

Brailsford, Timothy J., Stephen A. Easton, Philip K. Gray, and Stephen F. Gray, "The Efficiency of Australian Football Betting Markets," Australian Journal of Management, 20, 2, December 1995.

Brown, William O. and Sauer, Raymond D. "Fundamentals or Noise? Evidence from the Point Spread Betting Market," Journal of Finance, 48:4, 1993.

Camerer, Colin F. "Does the Basketball Market Believe in the Hot Hand?” American Economic Review, 79, 1257.

Crafts, N.F.R. 1985. “Some Evidence of Insider Knowledge in Horse Race Betting in Britain,” Economica, 52, pp. 295–304.

Figlewski, Stephen “Subjective Information and Market Efficiency in a Betting Market,” Journal of Political Economy 87, no. 1 (1979): 75–88.

Gandar, John; W. Dare, Craig Brown and R. Zuber. 1998. “Informed Traders in the Betting Market for Professional Basketball Games,” J. Finance, 53, pp. 385–401.

Hakes, Jahn K. and Raymond D. Sauer, "A Probability-Based Measure of Productivity in Major League Baseball With Application to the Questions of Clutch Performance and the Value of Pitching," manuscript, http://hubcap.clemson.edu/~sauerr/papers.html, 2004.

Hakes, Jahn K., Raymond D. Sauer and Kerry Waller, " The Progress of the Betting in a Baseball Game," paper presented at the 2005 Meeting of the Western Economic Association, http://hubcap.clemson.edu/~sauerr/papers.html.

Hanson, Robin, "Foul Play in Information Markets," in *Information Markets: A New Way of Making Decisions in the Public and Private Sectors*, ed. Bob Hahn and Paul Tetlock, AEI-Brookings Press, 2005.

James, Bill and Jim Henzler. *Win Shares,* Stats Inc. Publishing, 2002.

Maloney, Michael T. and J. Harold Mulherin, “The Complexity of Price Discovery in an Efficient Market: The Stock Market Reaction to the Challenger Crash,” Journal of Corporate Finance 9, no. 4 (2003): 453–79.

Pennock, David M. C. Lee Giles, and Finn A. Nielsen, “The Real Power of Artificial Markets,” Science 291 (2001): 987–88.

Roll, Richard “Orange Juice and Weather,” The American Economic Review 74 no. 5, (1984): 861–80.

Sauer, Raymond. "The Economics of Wagering Markets," Journal of Economic Literature, December 1998.

Spencer, Jakab “New ICAP-Nymex Derivatives Have U.S. Gas Markets Number,” Wall Street Journal, August 4, 2004.

Strumpf, Koleman, "Illegal Sports Bookmakers," working paper, http://unc.edu/cigar/papers.

Weinbach, Andrew. "Gambling and the Poor," PhD Dissertation, The John E. Walker Department of Economics, Clemson University, 2005.

Wolfers, Justin, and Zitzewitz, Eric, "Prediction Markets," Journal of Economic Perspectives, Volume 18, Number 2, Spring 2004, pp. 107–126.

Table 1

Odds and Pitchers for 9/29/99: Reds Vs Astros

Team	Pitcher	Money Line	Odds Equivalent	Odds to $1	p^{FAIR}	p^{FAIR*}
Reds	Parris	+140	1.4/1	1.4	.4167	.4103
Astros	Hampton	-150	1/1.5	0.67	.5988	.5897

Source: Computer Sports World

Table 2a

The Linear Probability Model of Winning

Dependent Variable: p_h

Variable	Coef.	Std. Err.	t-ratio
constant	0.504	0.020	25.8
ops_h	0.076	0.016	4.74
ops_v	-0.055	0.017	-3.2
Team dummies	yes		
Pitcher dummies	yes		
Observations	2385		
R-bar sq	.851		
Root MSE	.036		

Table 2b

Coefficient Estimates for Pitchers

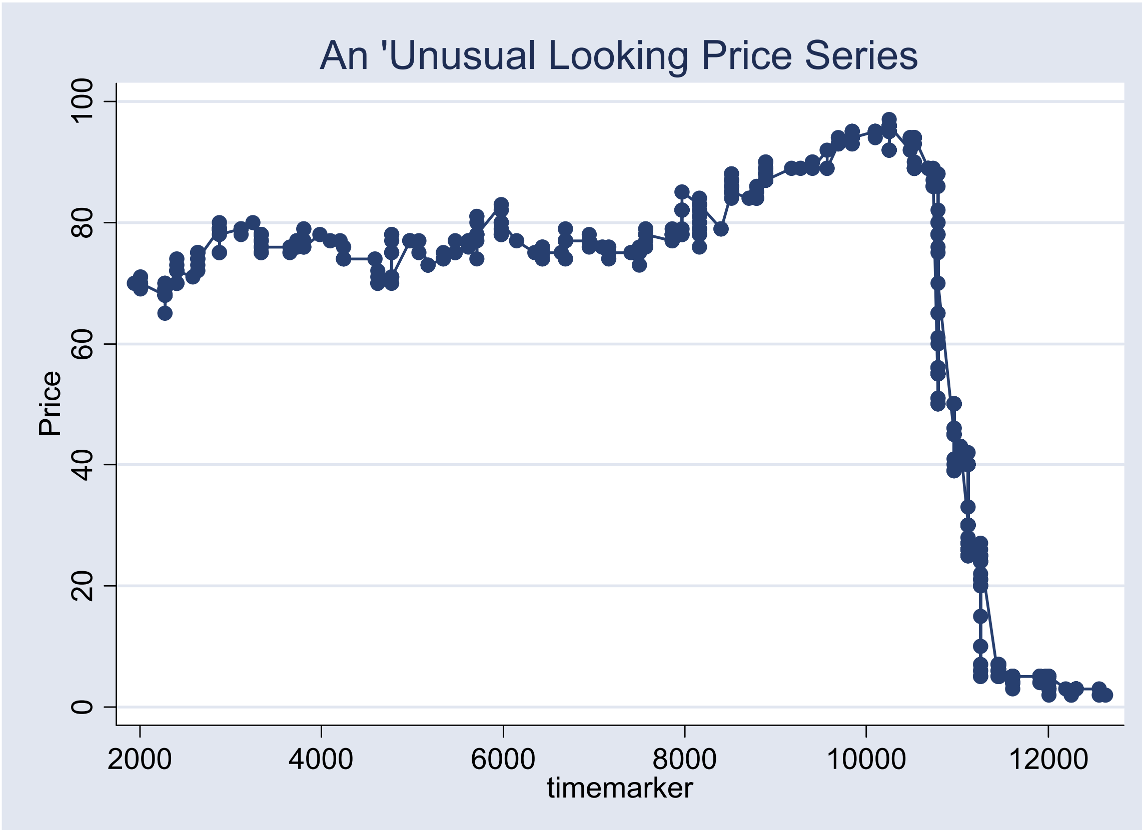
Largest:		Smallest:	
P.Martinez	0.153852	M.Yoshii	-0.00298
C.Schilling	0.148916	L.Hawkins	-0.00319
K.Brown	0.141394	K.Mercker	-0.00469
M.Mussina	0.135116	A.Pettitte	-0.00582
R.Johnson	0.130619	J.Witasick	-0.00685
G.Maddux	0.104611	M.Gardner	-0.00719
K.Tapani	0.079895	M.Portugal	-0.01183
B.Radke	0.079373	P.Rapp	-0.01408
C.Finley	0.074121	M.Morgan	-0.02192
J.Lieber	0.073567	C. Holt	-0.02949

Table 3: Average Effect of Selected Events Upon Probability of Winning

Event	Frequency	Mean Change in P(win)	Std. Error
Walk	17,028	0.0281	0.0002
Hit by pitch	1,572	0.0284	0.0006
Single	29,686	0.0418	0.0003
Double	8,902	0.0646	0.0007
Triple	952	0.0948	0.0026
Home run	5,693	0.1217	0.0013
Strikeout	31,254	-0.0276	0.0001
Ground out	35,191	-0.0220	0.0001
Fly out	25,279	-0.0248	0.0001
Ground into double play	3,833	-0.0753	0.0010

Table 4
Testing the Bill James Conjecture
Dependent Variable: DP_sum

Variable	Coeff	Std err	t-ratio
constant	-0.023	0.006	-3.77
d_p	1.069	0.126	8.41
Rbar2	0.361		



An 'Unusual Looking Price Series
Price
0
20
40
60
80
100
2000
4000
6000
8000
10000
12000
timemarker

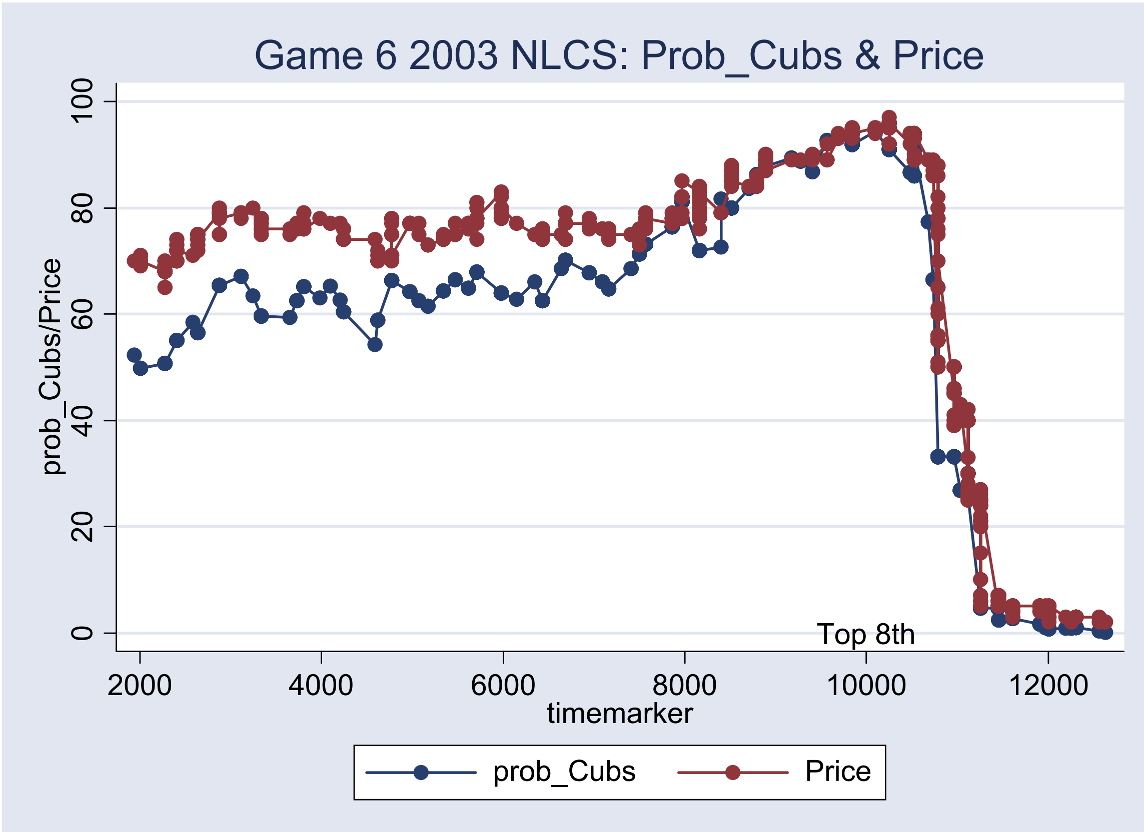


Figure 2: Betting Market Prices & the Probability of Winning, 2003 NLCS Game 6