Give Freakonomics a Chance

IAN AYRES

David Stern has a humongous problem. As preseason opens, the scandal of a guilty plea by NBA referee, Tim Donaghy, calls into question the integrity of the entire league. Donaghy admitted in court that he sold "inside information on NBA players, referees and coaches to James ('Sheep') Battista." The NBA is big business whose product is, at least in part, fair competition. The customers may stay away if they start worrying that the games are rigged.

Stern claims that Donaghy was an isolated rogue employee. But how are fans supposed to know?

Ian Ayres is Professor of Law and Business, Yale University. He is the author of numerous academic articles and books, as well as a commentator on public radio's Marketplace *and a columnist at* Forbes. *His latest book is* Super Crunchers: Why Thinking-By-Numbers Is the New Way to Be Smart.

Stern has hired former federal prosecutor Lawrence B. Pedowitz to study the NBA's anti-gambling efforts, including how it monitors officials.

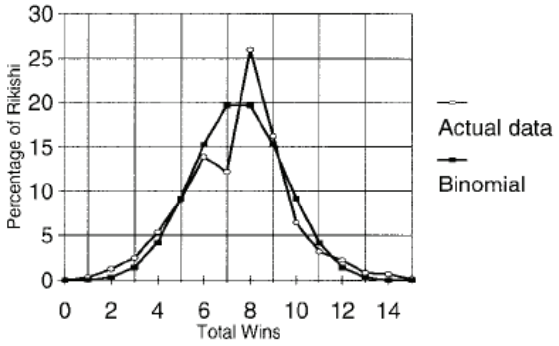
Stern has promised the NBA will work "to be transparent in the sense that our fans know how the system works." But procedural transparency isn't sufficient. Merely telling the fans more about how the current system operates isn't enough. The NBA should be more substantively transparent.

Stern should give Freakonomics a chance. In the last decade, a cadre of empirical economists led by Steve Levitt has crunched numbers to test for cheating. They've created a new field of forensic econometrics that can provide statistical evidence of criminality.

In 2002, Levitt (together with Mark Dugan) used statistics to identify corruption in sumo wrestling. There are 15 bouts in sumo tournaments and you are only promoted if you win at least 8. Levitt found that some sumo wrestlers who were on the cusp had an uncanny tendency to win their eight match. The following graph (Figure 1, next page) shows that the distribution of wins follows the expected bell curve except that some of the expected probability mass is shifted from 7 win to 8 win results.

The authors have to worry of course that this disparity is caused by additional effort that wrestlers make when they have more at stake. And several ancillary pieces of evidence suggest that the success rate of wrestlers on the bubble is correlated with the costs of corruptions. When public attention is focused on match rigging (the costs of corruption increases), the bubble bump goes away. When both opponents are from major stables with ample opportunities for

Figure 1



Wins in a Sumo Tournament (Actual vs. Binomial)

[Source: Dugan & Levitt]

reciprocal accommodations (the cost of corruption decreases), the bubble bump increases.

In 2004, Levitt (together with Brian Jacob) provided striking evidence that several Chicago grade school teachers were helping their classes cheat on standardized tests. Once again corruption left behind a digital trail. For example, take a look at the following table of test answers (Figure 2). Each row reports the answers of a different student. The letters A, B, C, and D represent the correct answers to questions while the numbers represent wrong answers. 1 is a student who wrongly answered A.

Jacob and Levitt looked for unusual strings of consecutive identical answers within a class. The red rectangle below is an extreme example of strings that do not seem to have occurred by chance. Because of this study, the city fired several teachers.

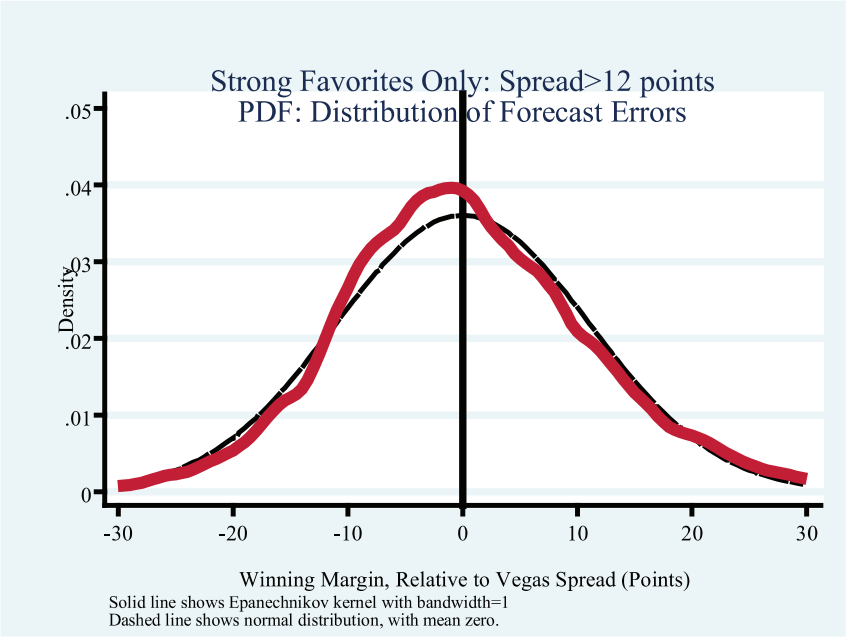
Gambling cheats often leave a statistical trail. Indeed, Justin Wolfers has already used forensic econometrics to uncover malfeasance in basketball games. Wolfers compared the outcome of thousands of college basketball games to the Las Vegas point spread. We should expect that

Figure 2

Suspected Cheating Classroom			
112A4A342CB214D0001ACD24A3A12DADBCB4A0000000	1.9	5.3	4.4
1B2A34D4AC42D23B141ACD24A3A12DADBCB4A2134141	4.3	5.6	4.3
DB2ABAD1ACBDDA212B1ACD24A3A12DADBCB400000000	3.0	6.5	5.1
1142340C2CBDDADB4B1ACD24A3A12DADBCB43D133BC4	3.6	6.3	4.9
D43A3A24ACB1D32B413ACD24A3A12DADBCB422143BC0	5.2	5.9	4.9
D43AB4D1AC3DD434212340D24A3A12DADBCB400000000	4.8	5.3	3.6
DBA2BA21AC3D2AD3C4C4CD40A3A12DADBCB400000000	1.9	6.1	3.6
DBAA4ADC4CBD24DBCB2A1110A3A12DADBCB400000000	3.3	6.3	6.2
144A3ADC4CBDDADBCBC2C2CC43A12DADBCB4211AB343	3.0	6.8	4.9
D43ABA3CACBDDADBCBCA42C2A3212DADBCB42344B3CB	4.8	7.1	6.6
214AB4DC4CBDD31B1B2213C4AD412DADBCB4ADB00000	3.6	6.1	4.3
313A3AD1AC3D2A23431223C000012DADBCB400000000	3.8	4.7	5.1
D4AAB2124CBDDADBCB1A42CCA3412DADBCB423134BC1	5.5	6.6	7.7
3B3AB4D14C3D2AD4CBCAC1C003A12DADBCB4ADB40000	3.0	6.5	6.6
DBAAB3DCACB1DADBC42AC2CC31012DADBCB4ADB40000	3.8	7.1	5.6
DB223A24ACB11A3B24CACD12A241CDADBCB4ADB4B300	4.9	6.5	5.8
D122BA2CACBD1A13211A2D02A2412D0DBCB4ADB4B3C0	3.6	6.1	6.2
1423B4D4A23D241314132341 23A243A2413A21441343	4.9	2.5	5.6
DB4ABADCACB1DAD3141AC212A3A1C3A144BA2DB41B43	5.9	6.5	7.7
DB2A33DCACBD32D313C21142323CC300000000000000	3.8	4.4	5.6
1B33B4D4A2B1DADBC3CA22C0000000000000000000	5.0	4.4	7.2
D12443D43232D32323C213C22D2C23234C332DB4B300	3.3	3.8	3.6
D4A2341CACBDDAD3142A2344A2AC23421C00ADB4B3CB	6.4	5.9	6.2

[Source: Jacob & Levitt]

Figure 3



[Source: Wolfers]

the favored team beats the spread almost exactly half the time. And indeed that's just what Wolfers found when the Las Vegas spread was small. But when the point spread was greater than 12, Wolfers found that the favored team was likely to just miss covering the spread. Indeed, Wolfers produced a graph (Figure 3) that was quite analogous to Levitt's sumo bell curve.

Instead of having a bump just to the right of the mean, now we see an unexpected bump just to the left of the mean. In some of these lopsided games, players on the favored team were holding back just slightly at the end of the game. They could get paid for point shaving without really increasing the chance that their team would lose.

This evidence from college basketball and Sumo wrestling makes me worry about the NBA.

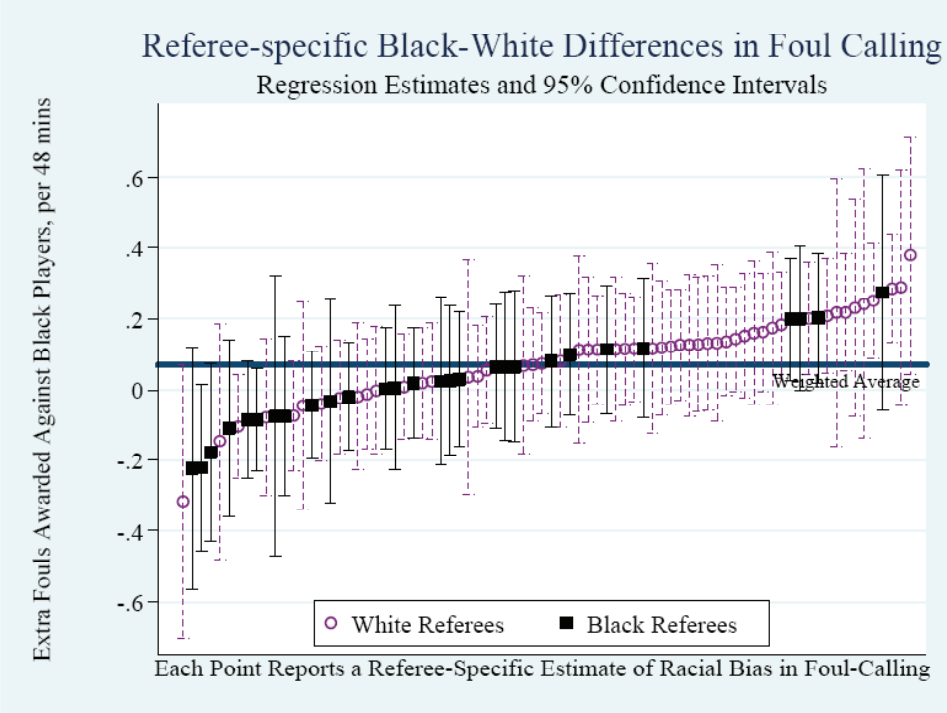
The application of Freakonomics to the NBA in a limited way has already been done. Earlier this year, Wolfers (together with Joseph Price) analyzed data on NBA games to suggest that white refs were more likely to call fouls on black players than black refs. For example, they found after controlling for a ton of other variables that a player "earns 0.18 fewer fouls per 48 minutes played when facing three referees of his own race than when facing three opposite-race referees." You can see the impact of ref and player race in the following diagram (Figure 4, next page).

Dots in this graph represent racial disparity estimates for individual referees. Referees who called an above-average number of fouls against Black players (relative to White players) lie above the horizontal line, while those who were estimated to have called a below-average number of fouls against Black players are the dots lying below the horizontal line. Notice how many of these above-average refs were white (indicated by the hollow circles) and how many of the below-average refs were black (indicated by the filled squares).

Price and Wolfers have suggested that this "own race" bias was large enough to impact the outcome of games. For example, they estimate that if one team has one fewer black starter than its opponent, its chance of winning the game increases by approximately 3.4 percentage points if the game is refereed by an all white crew than by an all black crew.

The Price/Wolfers data had one big limitation. It didn't include information on the foul

Figure 4



[Source: Price & Wolfers]

calls by individual referees. At the prompting of Mark Cuban, the NBA started collecting this data a few years back. But the league refuses to release it.

When Wolfers' study was publicized, David Stern arrogantly rejected Wolfers' conclusions in the most vehement terms, calling them "wrong" and "disingenuous." The NBA had done its own study, using the more detailed data with information on foul calls by individual refs. Stern claimed that it was "more powerful, more robust, and demonstrates that there is no bias." After several players (including Charles Barkley) sided with the Commissioner, the Wolfers controversy subsided. But imagine how it might have played out if the evidence of NBA referee bias had broken now, when we're not as sure about the league's oversight of its referees.

I actually reviewed the NBA's statistical study and disagree strongly with Stern's characterization that there is no evidence of a racial disparity. I'm limited in what I can say about the NBA's materials—but you can read more about its weaknesses here. The NBA study (which is actually based on fewer observations that the Price/Wolfers study) seems to have either been done by less competent empiricists and/or that to have been overly "lawyered." There was no attempt to start with the Price/Wolfers specification and show that the results no longer held with the more detailed data. In the end, I believe even the NBA's analysis tends to support a conclusion of own-race bias. The real point, though, is that this question should not be decided by unsupported claims by any one person. I am a fan of David Stern, but he should do more than say "trust me."

Especially now, the NBA should release its refereeing data and let forensic econometricians search for impropriety. This is not just about confirming Donaghy's malfeasance, or racial disparities. A Levitt or a Wolfers might find malfeasance or unaccountable disparities by other

officials. They might find, for example that in games refereed by other officials, the combined team scores were systematically to cover the over/under or the spread. Sure, the NBA should hire its own in-house forensic statistician. But it should also let outsiders test for malfeasance.

Fans want to know if they can trust NBA refereeing. Instead of business as usual, Stern should be willing to put his refs to the test. Let facts be submitted to a candid world.

Letters commenting on this piece or others may be submitted at submit.cgi?context=ev.