CONTEMPORARY ECONOMIC POLICY

BIASED REFEREES?: RECONCILING RESULTS WITH THE NBA'S ANALYSIS

JOSEPH PRICE and JUSTIN WOLFERS*

In a recent study, we used National Basketball Association (NBA) box-score data over a 13-year period including nearly 600,000 foul calls to show that NBA referees call relatively more fouls on players of another race. The NBA commissioned its own study using internal data with information on calls made by individual referees and claimed the results show there is no bias among the referees in the NBA. This paper is an attempt to reconcile these competing claims. (*JEL* J15, J71)

I. INTRODUCTION

In a recent study, we found that National Basketball Association (NBA) referees call relatively more fouls on players of the other race (Price and Wolfers 2010). This analysis was based on box-score data from the 1991 to 2003 seasons encompassing over a quarter of a million observations of players in games in which more than 600,000 fouls were called. These results were robust to the inclusion player, referee, team, and game-specific characteristics, allowing us to rule out many alternative explanations for the bias we observe. The impact of the bias extends to many other aspects of the game including minutes played, points scored, and ultimately who wins the game.

In response to our study, the NBA commissioned its own study using internal data based on the decisions of individual referees.[1] David Stern, the NBA commissioner, claimed quite emphatically that, "We think our cut at the data is more powerful, more robust, and demonstrated that there is no bias." In addition, Joel Litvin, the NBA president stated, "The paper, the study, is completely wrong, as far as we're concerned. We've proven it through our own studies. We believe their studies are inferior. Their methodology is inferior. And we don't have any problems publicizing our findings of the more than 148,000 calls over a 2½-year period we looked into studying ourselves." The goal of this paper is to reconcile these two competing claims.

This paper makes three claims. First, the NBA fundamentally misunderstood our research question. Of the 15 tables in the NBA's study, 12 examine racial differences in the fouls called on players and fouls called by referees. Our study was not about whether black players receive more fouls (they do not) or whether black referees call more fouls, but rather whether a black player will receive relatively more fouls when there are more black referees on the court.

Second, the NBA fundamentally mistreated their data. They dropped all players who received zero fouls and weighted their observations based on fouls called. This leads to a very mistaken view of the average number of fouls called on players in the NBA. In addition, the referee's race is recorded at the call level (a major advantage of the NBA's data) but the dependent variable is the total fouls called on the player for the game (and not for the particular referee). This removes any advantage they might have had over the box-score data.

*We are grateful to the research assistance of Matt Gibb and Tim Larsen. This paper has benefited from the comments of seminar participants at Chicago Law School, Dartmouth, Delaware, Missouri, Northwestern Law School, UCLA, Wharton, the Society of Labor Economists, the American Law and Economics Association, and the Western Economic Association annual meetings.

Price: Department of Economics, Brigham Young University, Provo, UT 84602. Phone (801) 422-5296, Fax (801) 422-0194, E-mail joe_price@byu.edu

Wolfers: The Wharton School, University of Pennsylvania, Philadelphia, PA 19104. Phone (215) 898-3013, Fax (215) 898-7635, E-mail jwolfers@wharton.upenn.edu

1. Fluhr (2007), the CEO of the consulting firm that the NBA commissioned for its study, provides a brief description of the analysis that his firm did for the NBA. Throughout our paper, we will refer to this study as "the NBA's analysis" although it was actually conducted by the Segal Company. The Segal Company is an actuarial and consulting firm that provides clients with HR consulting services.

ABBREVIATION

IAT: Implicit Association Test
NBA: National Basketball Association

Contemporary Economic Policy (ISSN 1465-7287)
Vol. 30, No. 3, July 2012, 320–328
Online Early publication June 27, 2011

doi:10.1111/j.1465-7287.2011.00268.x

Third, the NBA's analysis provides enough information for us to back out the information needed to replicate the type of analysis we did in our original study. Once we adjust the NBA's results appropriately, we find that their data support our original findings, though it appears the own-race bias is largely confined to players who spend fewer minutes on the court.

The approach we use in this paper also relates to a larger issue of replicating results in sports economics, which is discussed in two recent papers by Winfree (2010) and Fort (2010). Winfree notes three reasons that can explain the inability to replicate past results include changes in computation power over time, differences across statistical programs, and slight differences across available data. The NBA's analysis, however, does not follow the normal pattern of past replication studies, but is rather an attempt to refute our study using a different data set (one we do not have access to) and using a different methodology. This paper provides an example of the ability to reconcile results across different studies even when the two studies differ along many important dimensions.

The motivation for this paper also relates to the ongoing debate about the economic impact of sports facilities. In their review of this literature, Siegfried and Zimbalist (2000) note that this is one of the few topics in empirical economic research where there is "virtual unanimity of findings." In another review of the literature, Coates and Humphries (2008) note that this consensus view only exists within the body of academic research (and primarily among that research done by economists) with an opposite consensus view occurring within the "promotional literature," which consists of reports written by consulting firms on behalf of the team or other organization that is vying for public subsidies for a sports facility. The competing claims that we describe in this paper can be seen as yet another example of academic research being challenged by the claims of a consulting firm working on behalf of a group with private interests at stake.

Hudson (2001) notes that one of the primary challenges in these debates is that the reports produced by the consulting firms are not published in reputed journals or listed in easily accessible databases. Instead, they are shared with the local media and used to sway public opinion and then disappear. Several studies have preserved the information from these past debates by confronting the specific claims and describing the methodological flaws in the promotional literature. Notable examples include Crompton (1995), Siegfried and Zimbalist (2000), and Hudson (2001), all of which reference specific studies conducted by consulting firms. This study follows in that tradition by documenting the methodological approach of the NBA's analysis, documenting some of the empirical mistakes that were made, and providing a framework in which outside parties can consider the strength of the two competing claims.

II. THE NBA'S ANALYSIS

The NBA's data contain information on over 155,000 individual referee calls in 3,482 games from November 1, 2004 to March 25, 2007. For each foul called, they record the referee who called the foul, the player on whom the foul was called, and the race of both player and referee. They also measure the average number of fouls called by each referee, fouls received by each player and team, and average fouls called on home and away teams for each season. Their analysis uses these measures as additional controls in a way similar to our use of player, referee, and team fixed effects. In some of the specifications, they also include controls for the player's position where the different positions include center, center-forward, forward-center, forward-guard, guard-forward, guard, and forward.

Since the NBA has information on who blew the whistle for each call, they use the foul call as the unit of analysis. However, the dependent variable they use is the number of fouls that the player received during the game (rather than fouls called on a player by a particular referee). Thus, a player who receives four fouls in a game will show up in the data set as four observations, each assigned to the race combination measure for that call. All four observations would be assigned the same dependent variable value of four fouls. In addition, any players who did not receive any fouls during the game (regardless of how many minutes they played) would not show up in the data set.

The NBA's analysis consists of 15 regression models, each run separately in four groups based on the number of minutes the player spent on the court. There are 6,234 players with 0–9 minutes, 35,265 with 10–19 minutes, 51,439 with 20–29 minutes, and 55,263 with 30+ minutes. Table 1 provides a summary of

TABLE 1
Summary of Results from the NBA's Analysis

	0–10 min	10–20 min	20–30 min	30+ min
Panel A				
Models 1 and 3				
Black player	−0.167**	0.047**	−0.123**	−0.030*
	(0.036)	(0.017)	(0.014)	(0.016)
Constant	2.728	3.044	3.483	3.486
	(0.029)	(0.015)	(0.012)	(0.014)
Models 2 and 4				
Black official	0.051	0.019	0.020	0.018
	(0.034)	(0.015)	(0.012)	(0.011)
Constant	2.593	3.071	3.382	3.453
	(0.023)	(0.010)	(0.008)	(0.007)
Models 5–8				
Black player	−0.167**	0.047**	−0.122**	−0.030*
	(0.036)	(0.017)	(0.014)	(0.016)
Black official	0.051	0.019	0.019	0.018
	(0.034)	(0.015)	(0.012)	(0.011)
Models 10 and 12[a]				
White player	0.051	−0.005	−0.023	0.005
	(0.033)	(0.016)	(0.013)	(0.015)
White official	−0.023	0.008	0.004	0.007
	(0.033)	(0.015)	(0.012)	(0.011)
Model 14[b]				
White player	0.086*	−0.146**	−0.001	−0.084**
	(0.038)	(0.017)	(0.014)	(0.016)
White official	−0.049	0.015	0.005	0.006
	(0.036)	(0.015)	(0.012)	(0.012)
Panel B				
Model 9				
Player/official same race	−0.006	−0.016	0.032**	0.016
	(0.034)	(0.015)	(0.012)	(0.011)
Models 11 and 13[a]				
Player/official same race	−0.043	−0.039**	0.003	0.000
	(0.031)	(0.014)	(0.011)	(0.000)
Model 15[b]				
Player/official same race	−0.018	−0.049**	0.004	−0.007
	(0.034)	(0.015)	(0.012)	(0.011)

Notes: * and **indicate significance at the 5% and 1% levels, respectively. All regressions include a constant term. We only report the constant term for models 1–4.

[a]Includes controls for the player, team, referee, and home-away average number of fouls during the season.

[b]Includes the same controls as models 10 and 12 but also controls for the player's position.

the regression coefficients from the 15 tables included in the NBA's report. The tables in the NBA's report also include the t-statistic, p-value, and 95% confidence interval for each coefficient. We have omitted this additional information from the results we report in order to fit all of the results into a single table. Moreover, all of the relevant information is contained in the coefficient and standard error.

We have separated the table into two panels based on whether the results examine the role of own-race bias or not. Panel A looks at differences in fouls called based on either the player's race or the referee's race. Panel B looks at whether players receive fewer fouls when they are the same race as the referee. All of the coefficients in Table 1 are based on a linear regression, most of which (models 1–9) include only the racial measures of the player and referee without any additional controls. Four of the models (10–13) include controls for the average calls against the player, average calls against the team, average calls made by the official, and average calls made against home or visiting teams overall (assigned based on whether the player is on the home or away team), all in a given season. Models 14 and 15 include an additional control for the player's position.

The first point that emerges from these tables is that many of the regression models the NBA runs in their analysis are redundant. For example, models 5–8 are literally identical, with the only difference being which group is used as the omitted category. The same is true for models 1 and 3 and models 2 and 4. In each case, the coefficients on the player or referee race dummy variable are the same but with the opposite sign. Models 10, 12, and 14, while not completely redundant, are simply the same as the basic tables with additional controls for the player, team, referee, home-away average foul calls for the season, and the player's position. Note that these controls used by the NBA are crude proxies of our use of player, team, referee, year, stadium, and game fixed effects.

A second point that emerges from these tables is that most of the results are not relevant for the specific question addressed in our original research. Our question was whether a player would receive relatively more fouls from referees of the other race. None of the regressions reported in panel A address this question. For example, models 1 and 3 show that black players generally get fewer fouls. In our paper, we found that black players generally do have fewer fouls per minute played until one controls for the player's weight, height, and position. This is simply because of the fact that white players are (on average) taller, heavier, and more likely to play center (and are thus more likely to be in locations on the court where they will commit fouls). None of this speaks to racial bias on the part of NBA referees.

Models 2 and 4 show there is no difference in the fouls called based on the race of the officials. Models 5–8 show that when we include the player's race and the referee's race in the same regression, the coefficients are roughly the same as when they are run in separate regressions. This is exactly what we would expect if the two variables are orthogonal. This agrees with the fact that the league does not consider the racial mix of the team when assigning a referee crew to a game. In our study, we tested this claim by showing the fraction of a team's starters that were white was uncorrelated with the racial mix of the refereeing crew.

Thus, all of the models in panel A of Table 1 (which accounts for 12 of the NBA's 15 tables) are irrelevant to our research question. In panel B of Table 1, we provide the estimates for the three models which do test for an own-race bias. At first glance, it might appear that the results of panel B provide little evidence of own-race bias. Of the 12 coefficients, only 3 are statistically significant, and 1 of these has the wrong sign (or provides evidence of an "anti-own-race bias"). However, the first row of results in panel B (model 9) includes no controls for the player of referee, which is particularly problematic as these regressions include an interaction term without including the main effects for the player and referee's race (an issue we will return to later).

If we focus on the eight coefficients included in the two models that actually include the additional controls, we find that two are significant at the 1% level, and the four coefficients for players playing 20 minutes or less are all negative.[2] Given this pattern, the NBA's claim that there is no evidence of racial bias is a bit tenuous, even when using the results upon which their conclusions were based. However, our original results were not focused on the statistical significance but rather the magnitude of the coefficients (or the amount of racial bias that is occurring).[3]

Of the three tables that do test for an own-race bias, model 11 provides a natural starting point for examining the magnitude of the coefficients in the NBA's analysis. The specification in model 11 includes a measure of the average number of fouls for the player, team, and referee. There is also an indicator variable for whether the player and referee were of the same race. Of the four subgroups of players, the one group that has a significant coefficient (p-value $= .005$) are players with 10–20 minutes. The coefficient for this group indicates that these players will receive 0.039 fewer fouls when the referee is of the same race.

Thus, moving from an all-black crew to an all-white crew would lead a white player to receive 0.117 fewer fouls per game. To rescale this estimate to match our measure based on playing 48 minutes, we would multiply 0.117 by 3.2 (48 minutes divided by the midpoint of the group of 15 minutes), leading to a comparable estimate of about 0.374 fewer fouls

2. The coefficients for the players who played 0–10 minutes are very similar to those of players who played 10–20 minutes, at least for models 11 and 13. However, there are about six times fewer observations of players with 0–10 minutes, making the estimated coefficients for this group much less precise. It is important to note that having estimates with large coefficients and large standard errors does not provide evidence that there is no racial bias. In one sense, it simply means that you are unable to rule out even larger amounts of racial bias.

3. One concern with our original study was that we had so many observations that we would certainly find a statistically significant effect. The discussion of our results was always focused on the magnitude of the effect. The large sample size allows us to estimate this magnitude with a very high level of precision.

TABLE 2
Price-Wolfers Estimates of Own-Race Bias for Original Sample and Sample Covering the Same Period as NBA Analysis

	Black Players	White Players	Difference: Black−White Foul Rate	Slope: Δ(Black−White)/ Δ% White Referees
Panel A: original sample 1991/92–2003/04 (N = 266,984 player-game observations)				
0% White referees	4.418	5.245	−0.827	—
	(0.043)	(0.094)	(0.106)	
33% White referees	4.317	4.992	−0.675	0.455
	(0.016)	(0.035)	(0.038)	(0.331)
67% White referees	4.335	4.989	−0.654	0.064
	(0.010)	(0.023)	(0.025)	(0.137)
100% White referees	4.322	4.897	−0.574	0.240**
	(0.013)	(0.029)	(0.032)	(0.121)
Average slope:				Diff-in-diff
ΔFouls/Δ%White referees	−0.022	−0.204***		0.182***
	(0.027)	(0.066)		(0.066)
				(p = .006)
Panel B: NBA sample period 2004/05–2006/07 (N = 69,047 player-game observations)				
0% White referees	4.439	5.058	−0.619	—
	(0.063)	(0.127)	(0.145)	
33% White referees	4.415	4.831	−0.416	0.607
	(0.026)	(0.052)	(0.059)	(0.465)
67% White referees	4.377	4.692	−0.314	0.306
	(0.021)	(0.042)	(0.047)	(0.225)
100% White referees	4.275	4.619	−0.344	−0.088
	(0.034)	(0.068)	(0.075)	(0.268)
Average slope:				Diff-in-diff
ΔFouls/Δ%White referees	−0.173***	−0.387***		0.214*
	(0.053)	(0.112)		(0.121)
				(p = .077)

Note: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

per 48 minutes played. Table 2 provides the estimates from our original sample (1991–2003) as well as results using box-score data over the same period as the NBA's analysis. For this comparable period, the estimated own-race bias using the box-score data is 0.214 fouls per 48 minutes played.

The other group that approaches a significant estimate (p-value = .167) is the players with fewer than 10 minutes played in a game. Again, the estimate for this set of players points in the direction of an own-race bias (although not statistically significant at conventional levels) with players receiving 0.043 fewer fouls from referees of their same race. A similar rescaling leads to a black player receiving 1.238 more fouls per 48 minutes played when we switch from an all-black crew to an all-white crew.

The analysis for the two groups of players with 20–30 minutes and 30+ minutes of playing time have point estimates of 0.003 and 0.000, respectively, with p-values of .783 and .977. These results would seem to suggest that the bias we are detecting in our study is mainly driven by the low-profile players who receive fewer minutes each game. This is consistent with the work by Govan and Williams (2004) who find the negative associations detected with the Implicit Association Test (IAT) disappeared when the images shown included notable black people such as Bill Cosby, Michael Jordan, or Eddie Murphy.[4]

The specification in model 15 simply adds the player's position to the variables already included in model 13 and finds roughly similar results. At face value, the results in model 9 provide the one piece of evidence against

4. The Implicit Association Test (IAT) is a psychological tool that measures biases by recording the timing of the response to images of black and white people and words that have a good or bad connotation. Harvard hosts an IAT test that takes about 15 minutes and is freely available for anyone to take. In his commentary on our research, Ian Ayres mentioned that he would pay $100 per referee to take the test so that we could compare IAT measures of bias with those that we detect using foul calls. We attempted to do the same thing using a $50 reward and were unable to get any NBA referees to take the IAT test.

our study. This specification includes only one covariate: whether the player is of the same race as the referee. For the group of players in the 20–30-minute group, model 9 finds these players receive 0.032 more fouls when the referee is of the same race (p-value = .007). However, this is not the evidence of the kind of "no bias" that David Stern spoke about, but rather some form of anti-own-race bias (something which quickly goes away when even the most basic controls are included).

Thus, even without further investigation, it seems clear that the NBA has misrepresented their own findings. If anything, the results in models 9, 11, and 15 (the only tables that are relevant to our original research question) would provide evidence that supports the hypothesis that referees exhibit an own-race bias (with some offsetting evidence in the analysis that includes no additional controls).

There are still some important concerns about the way in which the NBA conducted their analysis. Our primary concerns relate to the interpretation of an interaction effect when there are no main effects included in the analysis, the removal of the players with zero fouls from the analysis, and the fact that their analysis did not explicitly control for the number of minutes played (other than simply separating the analysis into four groups based on playing time). By addressing each of these issues, we will show that the results presented by the NBA indicate that the data underlying the NBA's analysis (which we do not have access to) support our original finding and provide surprisingly similar results using individual foul calls to those we found using crew-level box-score data.

III. RECONCILING THE TWO DATA SETS

While the NBA sent us the raw output from their regression tables, they did not clarify the exact nature of their data set. The first step in comparing our results is to understand how their data set is constructed and what they are measuring. The NBA's analysis is based on over 155,000 individual referee calls in 3,482 games from November 1, 2004 to March 25, 2007.[5] While our original analysis was only based on data through the 2003/2004 season, we were able to collect additional box-score data from nba.com to cover the entire period of the NBA's data. When we replicate the same period of time using the box-score data, we find that there were 157,631 foul calls in 3,496 games.[6]

We should note that we will not be able to fully reconcile our numbers because we only code whether a player is black or non-black ("white"), while the NBA analysis explicitly drops all players who are neither black nor white. Again, the NBA did not tell us which players are excluded from their analysis though they explicitly stated that Asians were excluded, but they made no reference to what they did with Hispanic players (since there are black-Hispanic players). We went back to our original race data on all of the players and included additional markers for whether the player was Asian (there are 3 in the sample) or Hispanic (there are 17 in the sample) and then excluded Asian and non-black Hispanic players.[7]

The NBA notes that their racial exclusion drops the number of foul calls in the sample from over 155,000 to about 148,000. When we exclude the Asian and Hispanic players, the total fouls called in our sample drops from over 155,000 to about 153,000. In addition, there is also one Hispanic referee in the NBA (Tommy Nunez) but since our data is based on referee crews and not individual referee decisions, throwing out games in which he officiated would overcorrect our sample.[8] As the NBA's analysis aggregates the fouls called on each player to the game level, it is not clear how they would treat games in which Tommy Nunez participated.

A good starting point in reconciling the two data sets is to try to simply recreate the numbers from the NBA's model 1. Because this regression does not rely on the race of the referee who blew the whistle, the NBA's own data should yield the same answer as an analysis based on box-score data. Model 1 simply estimates the average number of fouls by black and white players. We use our own data to recreate the same four groups as the NBA study. We

5. The NBA's own document suggests that their sample ended in January 2007, but a column by Gwen Knapp reported that their sample ends on March 25, 2007, a date more consistent with the number of observations they report.

6. We exclude one game because we were unable to find the race of Gary Forest (an NBA official) and another game in which only two referees were present due to weather conditions. The NBA notes that its study excludes one game from the 2004 season and two from the 2005 season.

7. Information on Latino players in the NBA comes from an article on nbadraft.net by Joshua Motenko.

8. Tommy Nunez was the first Hispanic referee in the NBA. In a 2003 article at tommynunezfoundation.com he is quoted as saying, "I'm proud to be the only 'Ain't' in the NBA. I ain't white and I ain't black."

TABLE 3
Reconciling the Two Data Sets (Fouls per Game)

	0–10 min	10–20 min	20–30 min	30+ min
Panel A: NBA data				
White players	2.728	3.044	3.483	3.486
Black players	2.562	3.091	3.361	3.456
N	6,235	35,266	51,440	55,264
Panel B: Price-Wolfers box-score data				
White players	1.153	1.827	2.524	2.670
Black players	1.077	1.887	2.415	2.628
N	6,027	19,418	21,219	22,383
Panel C: Price-Wolfers attempted reconstruction of NBA data set				
White players	2.678	3.043	3.469	3.519
Black players	2.549	3.111	3.365	3.457
N	6,618	36,354	51,758	58,977

Notes: The number in each cell is the average fouls per game for each group. The changes we make to our data between panels B and C is to drop all of the players with zero fouls and weight the outcomes by the number of fouls called.

then use the regression coefficient and constant from model 1 to get the measure of the average number of fouls called on black and white players. These numbers are shown for each group of players in panel A of Table 3. We recreate the same output using information from the box scores and display these in panel B. The numbers in these two panels simply do not match at all.

It should be obvious to any NBA fan that players playing 0–10 minutes per game are not likely to earn 2.5–2.7 fouls, on average. Indeed, it appears that the NBA analysis is not even internally consistent. For instance, the NBA claims to have over 155,000 fouls from 3,482 games. This implies that there are around 44 fouls per game. Inspection of any box score will show that this figure is roughly accurate. This suggests that given there are two teams and, on average, ten players on each team get playing time, the average player earns 2.2 fouls per game. This is consistent with the Price-Wolfers data, but not with the NBA analysis. The NBA analysis finds *all* groups of players earn *at least* 2.5 fouls per game. Indeed, given the proportions of players in each group, the NBA study implies that the average player earns around 3.3 fouls per game. This would mean that an average of 66 fouls take place per game.

Our best attempt at reconstructing their data set involves two key adjustments. First, we suspect that they omit all players from their data who earn zero fouls, even if those players were on the court for much of the game. Second, we suspect that they include a player-game observation for each foul the player commits. That is, a player who commits five fouls will be in their data five times, while a player who commits two fouls will be in their data twice.

We use these two adjustments to reconstruct their data using our box-score data and report these results in panel C of Table 3. The similarity here is quite striking, and it seems that any remaining differences are often in the third decimal place. That is, the algorithm described above allows us to match the NBA's estimates of both the number of observations and the average number of fouls earned by black and white players in each group.

The NBA analysis—by weighting by fouls and eliminating players earning zero fouls—estimates the average number of fouls incorrectly. More importantly, this mismeasurement is itself a function of the number of fouls a player earns, and thus it is not simply random noise, but a form of measurement error likely correlated with the variables we are interested in analyzing.

IV. RECONCILING THE RESULTS

As mentioned earlier, a large fraction of the regression models estimated by the NBA were irrelevant to the question of own-race bias. However, these additional regressions provide a way to use the NBA's data to roughly recreate our results. Model 1 gives the average number of fouls called on white and black players. Model 2 gives the average number of fouls called by white and black referees. Model 9 gives the average number of calls called in opposite-race and own-race interactions. Finally, model 5 gives the average number of fouls a player receives controlling for both player and referee race (but no interaction of the two).

Using the information from these models, we construct the proportion of the sample that involves each of the possible player/referee race combinations that can occur as well as the average number of fouls that occur in each type of interaction. The four models above provide us seven facts. We combine this with an eighth fact which is that the sum of the proportions of each type of interaction is one. This provides us eight equations for which we are finding eight unknowns, a problem we can address using a system of simultaneous equations. The NBA

TABLE 4
Difference in Difference Estimates Based on the NBA's Analysis

	0–10 min (N = 6,235)			**10–20 min (N = 35,266)**		
	Black Players	**White Players**	**Difference: Black– White Foul Rate**	**Black Players**	**White Players**	**Difference: Black– White Foul Rate**
White referee	2.533	2.656	−0.123	3.089	2.994	0.095
Black referee	2.582	3.187	−0.605	3.094	3.081	0.013
Difference:			Diff-in-diff			Diff-in-diff
white referee – black referee	−0.050	−0.532	+0.482	−0.005	−0.087	+0.083
	20–30 min (N = 51,440)			**30+ min (N = 55,264)**		
	Black Players	**White Players**	**Difference: Black– White Foul Rate**	**Black Players**	**White Players**	**Difference: Black– White Foul Rate**
White referee	3.353	3.479	−0.126	3.433	3.479	−0.046
Black referee	3.372	3.488	−0.116	3.466	3.499	−0.033
Difference:			Diff-in-diff			Diff-in-diff
white referee – black referee	−0.019	−0.009	−0.010	−0.033	−0.021	−0.012

Note: These tables were recreated using a system of linear equations based on the stylized facts created by the NBA's analysis.

data are reported to three decimal points, which will slightly limit our accuracy, but not much. The Mathematica code used to reconstruct the NBA data is available from the authors.

Panel A in Table 4 recreates our original table using the NBA's data for players with 0–10 minutes. The estimated bias for the set of players with little playing time is more than twice as large as our original estimates using the full sample. When we reconstruct the data in a similar way for the other cohorts of players, we find that the estimated bias for players with 10–20 minutes is 0.083, for players with 20–30 minutes it is −0.010, and for players with 30+ minutes it is −0.012.

Panel A in Table 4 also illustrates why the NBA's analysis in model 9 comes to the conclusion that there is no racial bias. The NBA's analysis is comparing the average fouls from the own-race cells (2.582 and 2.656) with the opposite-race cells (2.533 and 3.187). Using the proportions of the sample from each cell, we can recreate their estimates for the own-race interaction as:

$$(1) \qquad (39\% \times 2.582 + 28\% \times 2.656) / (39\% + 28\%) = 2.613.$$

We can do the same for their estimate the opposite-race interaction using:

$$(2) \qquad (29\% \times 2.533 + 4\% \times 3.187) / (29\% + 4\%) = 2.612.$$

This explains why the NBA did not find an effect. White players are more than seven times more likely to face an own-race referee. This confounds two facts. First, white players earn fewer fouls under own-race referees, which should lead to a negative own-race effect. Second, white players earn more fouls than black players. Because white players are more likely to have own-race referees, this leads to an offsetting positive bias to the own-race effect, thus masking the own-race bias that is occurring. Had the NBA included a control for the player's race in their specification with the player-referee race interactions, the own-race bias in their report would have been even more pronounced.

V. LEVEL OF AGGREGATION

Throughout this paper, we have been working to reconcile our results with the analysis of the NBA based on individual-level decisions. This is primarily in response to the NBA's claim (and the claims of many sports reporters) that our analysis could not test for bias because we did not know who blew the whistle. However, in many ways the crew is the relevant unit of analysis in terms of implementing possible changes in policy. Clearly identifying individual referees who have pronounced racial biases would provide a relatively easy solution. However, our earlier results showed that the bias was quite pervasive across all of the referees rather than a small set of referees driving the results. Our results speak to the larger issue of what would happen to the treatment of black players if the league were to alter the racial mix of the referees.

There are a number of other settings in which this type of policy is the most relevant and feasible. Donohue and Levitt (2001) find that as the fraction of officers that are black in a police force increases, the number of blacks arrested is unaffected and the arrest rate for whites goes up. McCrary (2007) examines court-ordered racial hiring quotas imposed on municipal police departments and finds that the court orders increase the fraction of police officers who are black and reduce the fraction of offenders who are black.

VI. CONCLUSION

The goal of this paper was to reconcile the competing claims of our original study with the study commissioned by the NBA to refute our results. The NBA fundamentally misunderstood our research question and as a result spent the bulk of their analysis on regression models that were largely irrelevant to our research question. These additional regressions allowed us to use the NBA's data (with information on individual calls) to reconstruct the type of analysis we did with box-score data and show that the data underlying their analysis support the conclusions of our original research.

Since our original study, there has been confirming evidence from studies using data from baseball. Parsons et al. (forthcoming) find that a strike is more likely to be called when the pitcher and umpire are of the same race. Chen (2007) finds that white umpires provide a larger strike zone to white pitchers and a smaller strike zone to white batters. One of the most encouraging aspects of this additional research is that the own-race bias completely disappears in stadiums with a QuesTec system (devices that provide nearly perfect monitoring of umpires' decisions about whether a pitch was a strike) and that the own-race bias of a white home-plate umpire is reduced when working with a racially diverse crew of officials.

In May 2007, David Stern described the results of our research on racial bias among referees by saying, "... my major concern about it is that it's wrong.... We ran the data and came up with something that said quite starkly that there is no bias amongst NBA officials." It is true that the NBA has been one of the leaders in promoting racial equality. While earlier research suggested that black NBA players suffered substantial wage discrimination (Kahn and Sherer 1988; Koch and Vander Hill 1988), over recent decades these racial gaps appear to have receded or even disappeared (Bodvarsson and Brastow 1999; Hamilton 1997). We hope one day the same will be said about racial bias in officiating.

REFERENCES

Bodvarsson, O., and R. Brastow. "A Test of Employer Discrimination in the NBA." *Contemporary Economic Policy*, 17, 1999, 243–55.

Chen, J. *What Does Baseball Teach Us about Reducing Racial Discrimination? Evidence from Two Natural Experiments.* Mimeo, Stanford University, 2007.

Coates, D., and B. Humphries. "Do Economists Reach a Conclusion on Subsidies for Sports Franchises, Stadiums, and Mega-Events?" *Econ Journal Watch*, 5, 2008, 294–315.

Crompton, J. "Economic Impact Analysis of Sports Facilities and Events: Eleven Sources of Misapplication." *Journal of Sport Management*, 9, 1995, 14–35.

Donohue, J., and S. Levitt. "The Impact of Race on Policing and Arrests." *Journal of Law and Economics*, 44, 2001, 367–94.

Fluhr, H. "Superior Data, Analysis Shows No Race Bias among NBA Referees." *Sports Business Journal*, July 23, 2007, 44.

Fort, R. "Observation, Replication, and Measurements in Sports Economics." *Journal of Sports Economics*, 11, 2010, 3–16.

Govan, C., and K. Williams. "Changing the Affective Valence of the Stimulus Items Influences the IAT by Re-defining the Category Labels." *Journal of Experimental Social Psychology*, 40, 2004, 357–65.

Hamilton, B. "Racial Discrimination and Professional Basketball Salaries in the 1990s." *Applied Economics*, 29, 1997, 287–96.

Hudson, I. "The Use and Misuse of Economic Impact Analysis: The Case of Professional Sports." *Journal of Sport and Social Issues*, 25(1), 2001, 20–39.

Kahn, L., and P. Sherer. "Racial Differences in Professional Basketball Players' Compensation." *Journal of Labor Economics*, 6, 1988, 40–61.

Koch, J., and C. W. Vander Hill. "Is There Discrimination in the 'Black Man's Game'?" *Social Science Quarterly*, 69, 1988, 83–94.

McCrary, J. "The Effect of Court-Ordered Hiring Quotas on the Composition and Quality of Police." *American Economic Review*, 97, 2007, 318–53.

Parsons, C., J. Sulaeman, M. Yates, and D. Hamermesh. Forthcoming. "Strike Three: Umpires' Demand for Discrimination." *American Economic Review.*

Price, J., and J. Wolfers. "Racial Discrimination Among NBA Referees." *Quarterly Journal of Economics*, 125(4), 2010, 1859–77.

Siegfried, J., and A. Zimbalist. "The Economics of Sports Facilities and Their Communities." *Journal of Economic Perspectives*, 14(3), 2000, 95–114.

Winfree, J. "Issues with Replicating Results in Sports Economics." *Journal of Sports Economics*, 11, 2010, 48–59.