SEGAL COMPANY SUMMARY OF KEY FINDINGS MADE IN ITS STUDY OF ALLEGED BIAS IN FOUL CALLING BY NBA REFS

The summary below is a compilation of the key findings made by the Segal Company in its study of alleged bias in foul calling by NBA refs. We raised each of these points with the Times in discussions and meetings over the past two weeks. Page numbers refer to the paper provided to us by the New York Times on April 18.

This document is divided into the four topics shown below.

1. A review of our methodology
2. The material findings from our analyses
3. Other relevant findings
4. Additional review of the Wolfers/Price paper

1. A review of our methodology

We reviewed over 155,000 individual referee calls in 3,482 games from November 2004 through January 2007. Within this group, there were about 148,000 relevant calls, since they involved a player and an Official who were either black or white. Calls involving Officials and/or players who are neither black nor white are excluded (e.g., Asian players). (Note: there is one game excluded from the 2004 season and two games excluded from the 2005 season.)

The call-level data fields, by season, used in this analysis are:

- Game Number (the NBA internal game number code, which allows us to group call level data by season)
- Home Team
- Visiting Team
- Player Whistled for a Foul
- Player's Team
- Referee Calling the Foul
- Race of Player
- Race of Referee

We used regression models as our primary tool to test if race has an impact on calling. By using a regression model, we are measuring the predictive power that the variable of race has with respect to calling. The dependent variable in all of the tests referenced below is the number of calls made against an individual player in a game. Therefore, we are testing the question, "Does knowing the race of a player and Official have any predictive value in calculating the number of calls made against a player?"

First, we developed a series of simple linear regression models that mapped the number of calls, by player, in a game onto each of the separate race variables (i.e., black player, white player, black Official, white Official). Results from these separate regressions failed to account for either a statistical or a material amount of variance in calls made against a single player (R-squared ~ 0%). Further, the beta coefficients from these regressions were not material (e.g., the largest beta was 0.016). Next we combined the race variables (e.g., black Official/white player, and separately, same/different race pairing) into a multiple regression model. Similar results were found for the multiple regression models in that the R-squared statistic and beta coefficients did not indicate a statistically significant or meaningful relationship between Official or player race and call volume.

Finally, a more sophisticated form of regression was used to determine the effects of the race variables in the presence of the other variables. Specifically, we developed a hierarchical regression model to determine the predictive value of a number of independent variables:

- Average calls against the player, in a given season
- Average calls against the team, in a given season
- Average calls made by the Official, in a given season
- Average calls made against the home team or visiting team, in a given season
- Race of Official / Race of player
- "Same Race" or "Different Race" (conducted separately)

Hierarchical regression involves the comparison of two or more regression models (via the incremental change in R-squared) to determine the incremental variance in calls that can be accounted for by adding additional predictor variables into the base regression model. This technique is commonly used to test the predictive power of variables of interest in subsequent models after controls are accounted for in preceding models. In this case, we used race as the additional predictor variables and the historical average call variables listed above as the base variables.

We used the seasonal average for each player, team, Official, and home/away combination for each call reviewed. For example, if a call was made by Mark Wunderlich against Tracy McGrady during the 2004-2005 season when Houston was at home, the seasonal dependent variables associated with that call were:

- Average calls made against McGrady in any game in 2004-2005
- Average calls made against Houston in any game in 2004-2005
- Average calls made by Mark Wunderlich in any game in 2004-2005
- Average calls made against home teams in any game in 2004-2005

We did this to account for seasonal changing play styles, team strategies and rule interpretations and applications.

To conduct the analysis, we ran comparative hierarchical regression models at the call level to test, 1) the predictive power of the season-based histories of the players, teams, Officials and venue, and 2) the incremental predictive value of knowing the race of the Officials and the players. We then replicated this, using time cohort groups, described below.

2. The material findings from our analyses

With respect to our first two hierarchical regression models, the base model produced an R-squared of 0.14, meaning that knowing the histories of the teams, players, Officials and venues involved with each call could only account for about 14% of the variance in calling. R-squared is a measure of the relative predictive power of a model, and it can be between 0 and 1. The closer it is to one, the better the model. *The second model, which included the additional predictor variable (race), produced no measurable change in the R-squared value (0.14), meaning that knowing race does not add any predictive value to the number of calls made against a player.*

We also conducted this analysis using “same race / different race” variables and found similar results. Tests on individual independent variables concerning race (e.g., black referee) and race combinations (e.g., black referee/white player) did not reveal any statistically significant results.

Further, we replicated the above analyses based on average game-time cohorts:

- Players with less than 10 minutes of average play time
- Players with 10 or more minutes and less than 20 minutes of average play time
- Players with 20 or more minutes and less than30 minutes of average play time
- Players with more than 30 minutes of average play time

We conducted this additional analysis to address the fact that players who play more minutes per game tend to have different roles than those who play for a short period of time, and as such, have different confounding factors that affect their interaction with Officials. We think this method of addressing play time exposure is more sound than the "normalization" method (48-minute call rate) applied in the Wolfers/Price paper since it better captures the nature of play within each group, rather than effectively considering all players the same.

Similar to the previous findings, the cohort analyses show that race variables do not account for a statistical or material amount of predictive power in Official call volume. When evaluating the R-square values across the cohort models, the addition of race variables to the base model that already accounts for player, team, referee and venue variables, adds virtually no incremental change in R-Squared values (and in some cases, none). Two interesting findings from this

analysis are that (a) as one would expect, the R-squared values for the cohort models are higher than those for the model that analyzes the whole population, likely indicating different play-styles (and thus foul call patterns) amongst the player cohorts, and (b) the more minutes played, the less predictive power our models have in terms of predicting call volume.

Based on these analyses, we conclude that there is no material relationship between race and calling.

3. Other relevant findings

One of the shortcomings of these analyses and the Wolfers analysis is that even if there were a measurable and statistically significant difference in the volume of calls made based on race (which our analysis found does not exist and we believe Wolfers failed to convincingly demonstrate), it fails to answer the real question, which is, "Are the calls correct?"

To that end, we reviewed the accuracy of calls made for one and one-half seasons of data from November 2005 to January 2007 with respect to race. We found no material difference in the accuracy of calling in any Official/player race combination. Furthermore, we also analyzed the accuracy of calls based on the player who had been fouled.

4. Additional review of the Wolfers/Price paper

Below are some of the most significant methodological and inferential issues raised by the Wolfers/Price paper. Unless otherwise noted, we are referring to the April 18th paper:

- Using box-scores, the authors review only crew-level data, that is, the total number of calls made by the crew, without knowledge of individual calls made by black or white Officials. The NBA analyses were conducted at the individual level, which is as close to the phenomenon of racial bias as data can be. In their section on Behavior Interpretations, the authors make a number of invalid conclusions about individual behavior based on crew-level data.

- When applying their regression analyses, Wolfers/Price consistently show that race alone is not a statistically valid measure to predict calling or game outcome. This is evident in their measure of validity concerning race (as measured by R-squared, in columns 1 in Tables 4, 5 and 7). These columns show that race does not account for calling. It is only when the authors add additional factors known to be associated with foul volume (position, etc.) that the tests show some statistical validity (Columns 2 through 9). When looking at a number of factors (e.g., race, position, teams calling, etc.), the authors do not test for or isolate the incremental

usefulness of knowing race. The authors never test (in their paper) to see if a model that includes race has more usefulness than a model that does not include race.

» Page 35 shows that with respect to the number of personal fouls called against a black player when the makeup of the officiating crew changes, there is a large standard deviation (3.34) relative to the mean (4.44) and a slim coefficient for the effects of white officials (0.122 to 0.215), indicating no meaningful implication due to race (even if the coefficient is statistically significant).

» In answering the question, "Are these effects large enough to affect game outcome?" on page 12, the authors use a "Win Score" formula (page 13) consisting of nine variables (points, field goal attempts, free throw attempts, turnovers, rebounds, steals, blocks, assists and fouls) to calculate the effects of the regression models on winning margin (Table 5). Without commenting on the validity of the "Win Score" measure itself, note that page 35 shows that of these nine variables, only two variables (fouls and points) show consistently, statistical significance among the coefficients.

» Note substantial changes in material assumptions made between the March 22 and April 18th paper, including the author's repeat statements concerning the importance of the NBA's "random" and "arbitrary" scheduling methodology in the validity of results in the earlier paper, and the change in tone and substance in the later paper.