When Superstars Compete: New Evidence They Are Not So Super After All[1]

Robert A. Connolly and Richard J. Rendleman, Jr.

March 22, 2016

[1]Robert A. Connolly is Associate Professor, Kenan-Flagler Business School, University of North Carolina, Chapel Hill. Richard J. Rendleman, Jr. is Professor Emeritus, Kenan-Flagler Business School, University of North Carolina, Chapel Hill and Visiting Professor at the Tuck School of Business at Dartmouth. The authors thank the PGA Tour for providing the ShotLink data and a portion of the Official World Golf Rankings data used in connection with this study and Mark Broadie, Edwin Burmeister, David Dicks, Robert Hansen, Connie Helfat, Shijie Lu, and Kent Womack for providing very helpful comments. We thank Jennifer Brown for providing us with a condensed version of her STATA code, providing her weather and Official World Golf Ranking data, and providing lists of events employed in her study. We also thank her for responding to many of our questions in correspondence and providing comments on earlier drafts of this paper. Please address comments to Robert Connolly (email: robert_connolly@unc.edu; phone: (919) 962-0053) or to Richard J. Rendleman, Jr. (e-mail: richard_rendleman@unc.edu; phone: (919) 962-3188).

When Superstars Compete: New Evidence They Are Not So Super After All

Abstract

Recent research finds that the presence of an exceptionally skillful competitor, a "superstar" such as Tiger Woods, can lead to lower effort from other competitors. We dispute this conclusion, showing that it reflects a flawed empirical design that does not capture potential temporal changes in the terms of competition, allows a portion of any potential superstar effect to be absorbed by one or more explanatory variables, does not fully account for variation in scoring common to all players in a given round, and if uncorrected, leads to an exceptionally high superstar false discovery rate. Counterfactually, we show that when Woods was unexpectedly absent from PGA Tour competition, players may have performed worse. In addition, we find that in tournaments in which Woods participated some but not all of the time, field quality was higher when Woods actually did compete compared with when he did not.

KEY WORDS: Tournaments, Incentives, Peer Effects, False Discovery, Superstar, Golf.

1. Introduction

For several decades, economists have studied the properties of tournaments as mechanisms for allocating resources, selecting job candidates, auctioning assets, and rewarding skillful performance in athletic and other contests.[1] The underlying assumption is that competitors vary effort in response to incentives. The broad conclusion of the empirical research is that tournament competition produces improved performance. It is less clear whether this is due to self-selection or reflects the 'heat of competition' when high-skill competitors interact.[2]

Brown (2011) presents a model in which the presence of an exceptionally skillful competitor, a "superstar," actually leads to *lower* effort from other competitors. In her paper, she studies the potential adverse incentive effects associated with a superstar's participation in professional golf competition and concludes that in his prime, Tiger Woods' presence in a PGA Tour event resulted in reduced performance among his fellow competitors. This contrasts with typical tournament models, where smaller prizes, rather than the presence of a high-skill competitor, tend to lead to lower effort.

Since these findings pose such a fundamental challenge to existing theory and evidence, it is important to establish their accuracy. Our general conclusion is that Brown's findings are not reliable; in general, we find no support for her superstar hypothesis. Moreover, our analysis suggests that any empirical framework similar to that of Brown, designed to isolate the effect of a superstar in tournament competition, is fundamentally flawed. Indeed, one of the contributions of this paper is to show how complex it can be empirically to isolate a 'superstar effect.'

In Section 2 we provide a general framework of statistical design for addressing the issue of whether a superstar's presence is associated with the adverse performance of other economic "players," in this case professional golfers. We show that the estimation of a potential superstar effect, particularly as it relates to professional golf, is inherently problematic. In essence, the empirical model must control for round- (or event-) specific variation in scoring without allowing such controls to absorb any scoring effects that might otherwise be due to the superstar's participation. As

[1] The seminal paper on tournaments and incentives is Lazear and Rosen (1981). Prendergast (1999) provides a review of tournament theory applications in the context of firms. Ehrenberg and Bognanno (1990) are the first to study tournament incentives in golf. Szymanski (2003) surveys issues in the economic design of sports contests and tournaments.

[2] See Coffey and Maloney (2010) offer an additional perspectives on the source of the improved performance.

we show, this is an onerous task.

In Section 3, we reexamine Brown's main tests as reported in her Table 2.[3] We show that two important regression controls, TV viewership and tournament field quality, appear to absorb a *favorable* portion of any Woods-related superstar effect, therefore causing the remaining portion, *if viewed as the full effect*, to appear (too) unfavorable. We also show that Brown's evidence of a Woods-related superstar effect is highly sensitive to assumptions regarding regression error structure and the implicit assumption that individual player skill and relative course difficulty are fixed for the entire 1999-2006 regression estimation period. Although (generally) statistically insignificant, we show that estimated Woods-related regression coefficients tend to be positive (implying an unfavorable effect) for the first half of the 1999-2006 estimation period and negative (implying a favorable effect) for the second half.

In Section 4, we address two additional tests that Brown conducts to identify conditional competition disincentives. Brown finds that during her main 1999-2006 estimation period, players tended to perform better during periods when Woods performance was "cool" compared to periods when his performance was "hot." We find no support for such an effect. Brown also analyzes Woods' surprise absences from competition during portions of 2008-2010, finding that players performed relatively well in tournaments in which Woods was expected to play but did not participate due to injury or personal reasons. We find (if anything) the opposite: players performed worse when Woods was absent from events in which he would have normally been expected to compete.

Finally, in Section 5 we show that tournament field quality was higher in events in which Woods participated compared with events in which he did not compete. This creates a potential statistical estimation problem if Woods participation *caused* field quality to be higher. But perhaps more importantly, the finding is at odds with Brown's hypothesis that players, especially the best players, would give less effort when Woods is in a tournament field. At the margin, why would the best players go out of their way to compete with Woods if they did not plan to give their best efforts?

[3] In the course of our work, we discovered numerous errors and event omissions in Brown's data, described in detail in Online Appendices A and D. Although we would reach our final conclusions, whether our tests were based on data that approximates that of Brown (data set 1) or data corrected for errors and omissions (data set 2), we believe it is important to report results using the best data available to estimate any Woods-related superstar effect, whether it existed or not, as accurately as possible.

2. Modelling Scoring in Golf

A number of golf-related studies have employed the specification described in Equation (1) below, sometimes with slight variation, to model 18-hole scoring in golf. Typically, the purpose is to estimate mean individual player scores (skill) while simultaneously controlling for the relative difficulty of each 18-hole round. Such estimates are often referred to as estimates of "neutral" scores, since variation in scoring due to differences in relative round difficulty is removed.

Following Broadie and Rendleman (2013), who estimate mean neutral individual player scores in two-year windows to examine potential bias in the Official World Golf Rankings (OWGR), assume that each player, i, is endowed with a fixed skill level that generates an expected neutral score of μ_i per 18-hole round. Each round (j)-course (k) combination, $\{j, k\}$, has a fixed level of difficulty that causes the expected score of each player to be $\delta_{j,k}$ strokes higher than in a round that is otherwise neutral.[4] Therefore, the actual score of player i in round j on course k, $s_{i,j,k}$, is modelled as follows:

$$s_{i,j,k} = \mu_i + \delta_{j,k} + \epsilon_{i,j,k}, \tag{1}$$

where $\epsilon_{i,j,k}$ is the usual error term. One can then employ Ordinary Least Squares (OLS) regression, as specified in Equation (2) below, to estimate each individual player's mean neutral score while simultaneously controlling for the relative difficulty of the rounds in which the scores are recorded.

$$s_{i,j,k} = \mu_i P_i + \delta_{j,k} R_{j,k} + \epsilon_{i,j,k} \tag{2}$$

In (2), P_i is a matrix of dummy variables that identifies players, and $R_{j,k}$ is a matrix of dummies that identifies the interaction of round and course. The μ_i and $\delta_{j,k}$ are estimated simultaneously as fixed effects using one-zero indicator variables for players and round-course combinations. As specified, dummy indicators for one player or one round-course combination must be omitted. For ease of interpretation, assume that dummy indicators associated with the first round-course

[4] On the PGA Tour, most events are conducted on a single course, but there are several events each year conducted on up to four courses. For example, in a typical two-course event, half the players would play on the primary course, course 1, in round 1 and the other half would play on course 2. In round two, each player would play the course he did not play in round 1. A cut would then occur after round 2, with approximately half the field continuing play for two more rounds on the primary course. Since each course would not necessarily play with the same level of difficulty in each round, it is important to condition scoring on the interaction of the round and course on which the score is recorded.

combination are omitted; therefore, μ_i can be interpreted as the estimated score of player i when playing round-course combination $\{j = 1,\ k = 1\}$.

Although (2) does not take specific information about course setup, weather conditions, etc. into account, their mean effects on scoring should be reflected in the $\delta_{j,k}$ estimates. Moreover, if such effects are non-linear or otherwise mathematically complex, their total effect on scoring should still be reflected in the $\delta_{j,k}$ estimates, and the resulting estimates of round-course effects should be more accurate than if round-specific variation in scoring were estimated using a combination of specific round-course-related explanatory variables. Slight variations of (2) include Berry (2001) and Broadie (2012), who estimate the $\delta_{j,k}$ as random effects, and Connolly and Rendleman (2008, 2009, 2012a, and 2012b), who estimate time-varying player skill, while simultaneously estimating random round-course and player-course effects and first-order autocorrelation in individual player residual scores

It is important to note that the 18-hole score generating process represented by (1) reflects that there are two important sources of variation in scoring, one due to variation in individual player skills (expected neutral scores) and another due to the intrinsic difficulty of the round-course being played. Slight modifications, as in the Connolly-Rendleman studies, could also reflect a dependence between individual player scores and the course being played: that is, a player might perform better on some courses than on others.[5] In any event, most studies of 18-hole scoring in golf that have required player skill to be estimated have recognized that a significant portion of variation in scoring is due to differences in relative round-course difficulty; players in general tend to record their highest (worst) scores on difficult courses under the most difficult conditions and their lowest scores on the easiest courses under the most benign conditions. Failure to recognize or to fully account for this important source of variation in scoring can lead to incorrect statistical inferences.

In contrast to the above studies, where the purpose is to estimate individual player skill while controlling for differences in relative round-course difficulty, the focus of Brown's study is the opposite – to determine whether scoring differences by round (or event) are related to the presence of a superstar, while controlling for differences in player skill.

Assume that the presence of a superstar in round j causes the scores of all players to be β strokes higher than they would otherwise be. Building on the scoring model above, the actual score

[5] Random player-course effects as estimated by Connolly and Rendleman (2008) tend to be very small.

of player i in connection with round-course combination $\{j, k\}$ becomes:

$$s_{i,j,k} = \mu_i + \delta_{j,k} + \beta \times tif_j + \epsilon_{i,j,k}, \tag{3}$$

where tif_j (Tiger in field) is a one-zero dummy variable that takes a value of 1 if the superstar (Tiger Woods) participates the tournament associated with round j.

Let $\theta_{j,k} = \delta_{j,k} + \beta \times tif_j$. Equation (3) can then be re-expressed as:

$$s_{i,j,k} = \mu_i + \theta_{j,k} + \epsilon_{i,j,k}. \tag{4}$$

In this form, each μ_i and $\theta_{j,k}$ in a regression such as (2) can be estimated, but without further structure, the round effect, $\delta_{j,k}$, cannot be estimated separately from the superstar effect, β.[6] This is the essence of the estimation problem – how to determine the extent to which variation in scoring by round is due to differences in intrinsic round difficulty or due to Woods' participation in the tournament in which the round is played.

Although not stated explicitly in these terms, Brown's solution to the estimation problem is to use a matrix of round- and event-specific controls, Y_j, to control for round-specific variation in scoring instead of including direct estimates of fixed (or random) round-course effects.[7] She also estimates fixed player effects, $\mu_{i,k}$, conditional upon the course, k, being played. Thus, Brown's estimation model, shown below and expanded further in Equation (6), can be expressed as follows:

$$s_{i,j,k} = \mu_{i,k} P_{i,k} + \gamma Y_j + \beta \times tif_j + \epsilon_{i,j,k}, \tag{5}$$

where $P_{i,k}$ is a matrix of dummy variables that identifies the interaction of players and courses.

[6]In an effort to estimate a Woods-related superstar effect, Connolly and Rendleman (2009) make the critical error of estimating the effect by regressing residuals from a regression much like (2) on Woods-in-field dummies, not recognizing that any Woods-related effect, if present, would have been absorbed by the initial $\delta_{j,k}$ estimates.

[7]We do not include a k subscript with Y_j, because none of the controls included among Brown's Y_j distinguish among courses in multiple-course events.

3. Brown's Approach to Estimating the Woods Effect

Brown's main estimation model, which can be viewed as an expansion of Equation (5), can be stated as follows:

$$
\begin{aligned}
net_{i,j,k} = \alpha_0 + \mu_{i,k} P_{i,k} + \gamma Y_j + \alpha_1 HRanked_{i,j} + \alpha_2 LRanked_{i,j} \\
+ \beta_1 tifH_{i,j} + \beta_2 tifL_{i,j} + \beta_3 tifU_{i,j} + \epsilon_{i,j,k} \qquad (6)
\end{aligned}
$$

In (6), $net_{i,j,k}$ is the score of player i net of par in connection with round-course combination $\{j,k\}$.[8] $HRanked_{i,j}$ is a dummy variable indicating that player i is ranked in the top 20 of the OWGR at the time he plays round j, $LRanked_{i,j}$ is a dummy indicating that player i is ranked between 21-200 at the time of round j, and $URanked_{i,j}$ is a dummy variable indicating that the player's OWGR at the time of round j exceeds 200. $tifH_{i,j}$ is the interaction of the tif dummy variable and the dummy variable $HRanked_{i,j}$. $tifL_{i,j}$ and $tifU_{i,j}$ are defined in similar fashion. Y_j is a matrix of event- (or round-) specific controls, some of which are described in greater detail below, including inflation-adjusted tournament purse, inflation-adjusted tournament purse squared, weather-related variables, field quality, TV viewership, and a dummy variable indicating whether the tournament associated with round j is a "major." $\epsilon_{i,j,k}$ is the error term. The assumption behind this specification is that there is no portion of round-course-related scoring that is not captured by γ. Otherwise, $cov(Y_j^*, \epsilon_{i,j,k}) \neq 0$ (where Y_j^* denotes the properly measured 'true' determinants of round-specific variation in scoring), and the tif-related β estimates will be biased.

Brown estimates (6) two ways, using first-round scores only and again using four-round (event-level) tournament scores, but in the latter case, only for players who played the final of four rounds, therefore excluding those who missed cuts. In first-round regressions, round-course combination $\{j,k\}$ refers to a first round played on course k, but in event-level regressions, round-course combination $\{j,k\}$ refers to the primary course used in connection with the event.

Brown's superstar hypothesis implies that the β coefficients in (6) are positive. However, she emphasizes that β_1 and β_2 should be positive more so than β_3, since "players who are low in the

[8] In golf, each hole is assigned a "par" value, almost always 3, 4 or 5, which represents an estimate of the score for a high-skill player, assuming the player takes two putts to complete the hole. The sum of par values for each of the 18 holes is the course par, the value Brown subtracts from each respective 18-hole score to obtain the net score.

distribution of relative skill or who expect to finish in the nearly flat portion of the tournament prize distribution may not be adversely affected by a top competitor" (p. 999). The inclusion of α_0, $\alpha_1 HRanked_{i,j}$, and $\alpha_2 LRanked_{i,j}$ allows for β_1, β_2, and β_3 to be estimated as incremental Tiger-in-field effects, conditional upon OWGR categories, after controlling for the potential scoring effect associated with each of the three OWGR categories not already captured by estimated player-course effects.

3.1. Data

Our analysis of Brown's work relies primarily on two data sets, both of which are described in detail in Online Appendix A. Data set 1 closely approximates the data set actually used by Brown, including all errors and event exclusions of which we are aware. Data set 2 uses corrected data and events consistent with Brown's selection criteria as implied in her paper and in correspondence. In estimating Equation (6), we employ data from 1999-2006, as in Brown.

We obtained 18-hole scoring data, tournament purse values, course yardages, and identifiers for players, tournaments, courses and tournament rounds from the PGA Tour's ShotLink files. We obtained weekly OWGR data for 2003-2006 from the PGA Tour and collected 1999-2002 data ourselves from the OWGR archives.[9,10] We collected weather-related data for 1999-2006 from the National Climatic Data Center of the National Oceanic and Atmospheric Administration (NOAA) (details are provided in Online Appendix B) and also obtained weather-related data from Brown. Finally, we purchased proprietary television viewership data from The Nielsen Company.

3.2. Brown's Main Regression Results and Our Replication Results in Brown-type and Corrected Data

Table 1 summarizes Brown's main regression results (as shown in her Table 2) and our comparable results using data set 1 (designed to approximate that used by Brown) and data set 2 (corrected for known data errors and event omissions). To maintain consistency with Brown, we compute robust standard errors clustered by player-year, but later, we employ a different method of clustering,

[9] Data for 2004-2006 include individual player rankings and average ranking points for players ranked 1-999, whereas data for 1999-2003 just include rankings and points for players ranked 1-200, since during that period the OWGR did not report rankings and points beyond position 200.

[10] Brown also provided us with the OWGR data she used in connection with her study, which is updated monthly. Due to difficulties matching her monthly OWGR data to our ShotLink scoring data, as described in Online Appendix A, Section E, we do not attempt to replicate any of Brown's results using her actual OWGR data.

which we believe to be more appropriate.

In regression specification (6), if a player-course interaction is observed only once, data observations associated with the interaction can only contribute to estimating specific player-course intercepts but, otherwise, cannot provide any explanatory power in the regression; therefore, we omit such observations, resulting in "effective" samples with a smaller number of observations (denoted by N^*) than full samples (with sample sizes denoted by N). This will be the case for all observations on courses played only once during the 1999-2006 period (primarily "majors" other than the Masters) and for all observations involving players who otherwise played a specific course only one time.

In both data sets, approximately 1/3 of the observations are associated with player-course interactions observed only once and, therefore, are implicitly excluded. In data set 2, a total of 30 PGA Tour events are implicitly excluded, and Tiger Woods was in the field in 24 of these events. The 30 omissions include 18 of a possible 32 major championships and three highly-prestigious events in the World Golf Championships series.[11] In an effort to reduce the time required to run the regressions and to avoid over-stating adjusted R^2 values and robust standard errors, we run all of our regressions with "effective" samples, rather than full samples. (See footnote 12, below, regarding the potential overstatement of robust standard errors.) In addition, we not only run each regression as specified in (6), where the Tiger-in-field indicator variable, tif, is interacted with $HRanked$, $LRanked$, and $URanked$, we also run each regression with an unconditional version of the tif indicator.

The entries in Table 1 show a close correspondence between regression estimates reported by Brown in first-round regressions and those that we obtain using data set 1, designed to approximate her data. If anything, evidence of a Woods-related superstar effect in our Brown-type first-round data is even stronger than that presented by Brown; our tif-related coefficient estimates are slightly higher, and our standard error estimates are lower, resulting in statistically significant coefficient estimates for $tifH$, $tifL$, and $tifU$, with all p-values less than 0.03. We suspect that our lower

[11] We note that in her Table 2, Brown uses two different types of first-round and total-score data sets – data sets that include both regular tournaments and majors and "regulars only" data sets that exclude majors. Given the implied omission of majors in the more inclusive "regular and majors" data sets, only the eight 1999-2006 Masters events and two years each of the U.S. Open, British Open and PGA Championship are excluded incrementally in the data sets that include regular tournaments only. To conserve space, and in recognition that Brown's statistical design implicitly leaves out over half the majors to begin with, we do not analyze "regulars only" here.

standard errors reflect that we employed effective samples, rather than full samples, when estimating regression (6).[12] In addition, the coefficient estimate of 0.257, associated with unconditional tif is highly significant (p-$value = 0.002$).[13]

Using corrected first-round data, coefficient estimates for $tifH$, $tifL$, and $tifU$ are all lower than those using data set 1, but two estimates are still statistically significant at the .05 level (those associated with $tifH$ and $tifU$). Although the 0.169 coefficient estimate associated with unconditional tif is lower than that estimated in connection with data set 1, the estimate is still statistically significant at the .05 level (p-$value = 0.028$).

In event-level regressions, coefficient estimates using data set 1 are lower than those reported by Brown, and all p-values for $tifH$, $tifL$, and $tifU$ are 0.092 or higher. With corrected data, coefficient estimates are lower still, with estimates for $tifH$, $tifL$, and $tifU$ of 0.685, 0.018 and -0.154, respectively, bearing little relationship to those estimated by Brown except in relative order of magnitude. Moreover, none of the three estimates is close to being statistically significant at normally accepted levels.[14] Also, using corrected data, the unconditional tif estimate is much lower than when estimated in connection with data set 1 (0.051 vs. 0.407, with p-values of 0.807 and 0.079, respectively).

Overall, using regression specification (6), we find almost no evidence of a superstar effect in event-level regressions using Brown-like and corrected data but do find evidence in first-round regressions. However, as discussed below, potential problems with the dependent and explanatory variables employed in (6) as well as problems in computing clustered robust standard errors could cause the regression results reported in Table 1 to misrepresent the magnitude of any Woods-related

[12] Although we ran all of our regressions in R, we have confirmed that when running regressions in STATA, the statistical package used by Brown, estimated clustered robust standard errors are higher in connection with full samples than in comparable effective samples. In the STATA code that Brown shared with us, it appears that she estimated her regressions using full samples.

[13] The adjusted R^2 values of 0.188 and 0.375 that we report for data set 1 in connection with first-round and event-level regressions are both substantially lower than corresponding estimates of 0.29 and 0.48, respectively, reported by Brown. Our adjusted R^2 values were computed using effective samples, but we believe that Brown's were computed using full samples. When we run the same regressions using full samples, our adjusted R^2 values are 0.256 for first-round regresions (compared with Brown's 0.29) and 0.485 for event-level regression (compared with Brown's 0.48), suggesting that we are running essentially the same regressions as Brown. We note that the full event-level sample size for data set 1 is almost identical to that of Brown, 18,822 vs. 18,805 observations, but the full sample sizes for first-round regressions, 36,379 and 34,966, in connection with data set 1 and Brown's actual data, respectively, are quite a bit different, and we have been unable to reconcile the difference. This likely contributes to explaining the larger difference in adjusted R^2 values in connection with the two first-round regressions.

[14] Although statistically insignificant, using event-level data, a $tifH$ value of 0.685 indicates that four-round total scores are 0.685 strokes higher for players in the OWGR top-20 when Woods is in the field. By contrast, *tif*-related coefficient estimates using first-round data represent single-round scoring estimates.

superstar effect and its statistical significance.

3.3. Potential Problems with Brown's Model Specification

3.3.1. Dependent Variable

The dependent variable in (6) is $net_{i,j,k}$, the score of player i net of par in connection with round-course combination $\{j, k\}$, rather than the actual 18-hole score, $s_{i,j,k}$, as in (5). If $par_{j,k}$, the "par" assigned to course k in round j, were the same in every round over the 1999-2006 period, such that $par_{j,k} = par_k$ for all rounds, j, it would make no difference in estimating a Woods effect if $net_{i,j,k}$ or the actual score $s_{i,j,k}$ served as the dependent variable, since for any course k, estimated player-course effects while using $s_{i,j,k}$ as the dependent variable would be exactly par_k strokes higher than if the dependent variable were $net_{i,j,k}$. As such, no other coefficient estimates would be affected. In fact, however, over the 1999-2006 estimation period, par was reduced in data set 2 by 1 stroke for four of 87 courses, increased by 1 stroke for one course, and for another course took on values of 70, 71, and 72 in various orders. Similarly, in data set 1, par changed for five of 78 courses. As such, using the score net of par as the dependent variable without conditioning player-course identifiers and associated $\mu_{i,k}$ estimates on par creates a potential estimation problem, but a problem that can be avoided altogether if the dependent variable is the actual 18-hole score rather than the net score.[15] To be consistent with Brown, for the regressions summarized in Table 1, we employ the net score as the dependent variable when using data set 1, but when using corrected data, we employ the actual 18-hole score, $s_{i,j,k}$. Generally, in most regressions that we estimate, including those summarized in Table 1, using $s_{i,j,k}$ as the dependent variable rather than $net_{i,j,k}$ causes all tif-related coefficient estimates to be lower, and in some cases substantially lower, depending upon the regression form. Moreover, in smaller-scale regressions designed to estimate the effect of Woods' surprise absences from golf competition (see Section 4.2), using $s_{i,j,k}$ as the dependent variable rather than $net_{i,j,k}$ has a very large impact on some tif-related coefficient estimates.

[15]It should be noted that the term "par" is not included in the rules of stroke-play competition, the method of play represented in all scoring data referenced in this study. If in a simple single 18-hole competition, player A scores 72 and player B scores 68, player A wins by four strokes. It makes no difference whether the par assigned to the course were 70, in which case A would score 2-over par and B would score 2-under par, resulting, again, in a four-stroke difference, or if the par were 72, resulting in an even-par score for A and a 4-under par score for B and, again, a four-stroke difference.

3.3.2. Miss-measured Round-related Explanatory Variables

There are also potential problems with some round-related explanatory variables in regression specification (6). First, the γY_j component of (6), which represents the estimation of round- (event-) specific variation in scoring, must reflect how scoring would have been affected *in the absence of Woods* and, therefore, not include potential scoring effects due to Woods' tournament participation. Otherwise, a portion of any estimated Woods effect will be absorbed by the γ estimate, resulting in β estimates that reflect only the remaining portion of the effect not already captured by γ. Depending upon whether absorbtion within γ is positive or negative, estimates of the Woods effect as reflected solely in the β estimates will be biased low or high, respectively.

Brown includes TV viewership, as measured by Nielsen ratings, among the proxies for round-specific variation in scoring, recognizing that "[i]f intense media attention [as measured by TV viewership] improves (or hurts) players' performances, then the television viewer controls should capture these effects" (p. 998). But, as is widely recognized,[16] if viewership is higher due to Wood's tournament participation, the TV portion of any γ estimate would reflect both a general media attention effect and a Woods effect, resulting in β estimates that are too high (low) if players perform better (worse) when media attention is intense. (Note, in golf, "better" scores are lower scores.)

Using the event-level version of data set 2, we find that mean TV viewership is 2.21 million in events with Woods and 0.69 million without Woods (*p-value* < 0.0001). In the same data set, if we include only events in which Woods participated in some years but not all years, estimated TV viewership is 1.25 and 0.86 million in events with and without Woods, respectively (*p-value* < 0.0001). We note that when we estimate (6) directly and with variation, the coefficient estimates associated with TV viewership are negative in all 20 regressions that we summarize in Tables 2 and 3, indicating that scores are lower (better) when TV viewership is high. Since Woods appears to cause TV viewership to be high, any estimate of the effect of Woods on player performance should not only include the Tiger-in-field coefficient estimates, β_1, β_2, and β_3, but should also include Woods' favorable effect on scoring that is absorbed in the estimated TV viewership coefficient. As such, the Woods-related β estimates will tend to over-state any adverse effect associated with

[16]See for example, Kay (2015) and Myers (2015).

Woods' presence in a tournament field.

Brown also includes an estimate of field quality among the Y_j, measured as average OWGR points of the field, excluding Tiger Woods and the player associated with each respective scoring observation.[17] If Woods' expected participation in a PGA Tour event has an effect on the participation choices of other players, the estimates of the β coefficients in a regression like (6) could again be affected.

Using the first-round version of data set 2, we find that mean field quality as measured by average OWGR points is 1.98 in events with Woods and 1.16 without Woods ($p\text{-}value < 0.0001$).[18] In the same data set, if we include only events in which Woods participated in some years but not all years, estimated field quality is 1.67 and 1.34 in events with and without Woods, respectively ($p\text{-}value = 0.001$).[19] When we estimate (6) directly and with variation, the coefficient estimate associated with field quality is negative in 18 of 20 regressions that we summarize in Tables 2 and 3, resulting in β estimates that will tend to over-state any adverse Woods-related effect. (The tendency for the field quality coefficient to be negative is consistent with what Brown reports in her footnote 20, "The controls for the quality of the field are negative for the regressions in Table 2, suggesting that stronger fields may lead to lower scores" (page 1000)). In all remaining results that we report involving regression specification (6) and its variations, we report coefficient estimates, standard errors, and p-values for TV viewership and field quality to illuminate the potential problems associated with the inclusion of these two explanatory variables in the regressions.

3.3.3. Subsample Stability

A third potential problem is the implicit assumption that individual player and course effects are fixed for the entire 1999-2006 estimation period or, alternatively, that interacted player-course effects are fixed. As Connolly and Rendleman (2008, 2012b) have shown, individual player skill

[17] Inasmuch as the field quality measure associated with the score of player i excludes player i's own OWGR points, we should actually notate Y_j as $Y_{i,j}$. Since no other proxy variables associated with round-specific variation in scoring are player-specific, for expositional simplicity we continue to notate round-related proxy variables as Y_j.

[18] Here, we employ a single field quality measure for each event, excluding Woods, but not excluding other individual players.

[19] Using the event-level version of data set 2, we find that mean field quality is 2.37 in events with Woods and 1.44 without Woods ($p\text{-}value < 0.0001$). In the same data set, if we include only events in which Woods participated in some years but not all years, estimated field quality is 1.98 and 1.69 in events with and without Woods, respectively ($p\text{-}value = 0.003$). These values are higher than corresponding values in first-round data, since they are based on players who made tournament cuts, who on average, should be stronger players than those who missed cuts.

can vary quite substantially over time.

To illustrate how player skill varied over the 1999-2006 estimation period, using corrected first-round data, we estimate a version of regression (6) in which we estimate fixed player effects and fixed course effects separately, while allowing fixed player effects for each player with 20 or more first-round scores to be divided into first- and second-half estimates, where first- and second-half refer to the first and second halves of the players scores, not the first- and second halves of the 1999-2006 sample period. For players with first- and second-half skill (fixed player effect) estimates, the mean of the difference in first- and second-half skill is 0.065 strokes, indicating little change, on average, but the standard deviation of first- and second-half estimates of skill differences is 1.021, suggesting substantial time-dependency in mean player scoring.

It is also important to note that relative course difficulty may not have been constant over the entire 1999-2006 estimation period. During this time, improvements in golf technology and player conditioning enabled players to achieve greater distances with their shots, and many courses were lengthened to help compensate for these improvements.[20] Among courses that were played in all years, 1999-2006, the mean yardage of courses included in the corrected first-round data set increased by $7,116 - 6,977 = 139$ yards. Among courses that were not played in each of the eight years, the mean yardage in 2006 was 7,277 compared with 7,072 in 1999, a difference of 205 yards. Clearly, the "competitive playing field" was not fixed, and, therefore, estimating a potential Woods effect under the assumption that player skills and relative course difficulty were constant could produce misleading results.

Along similar lines, it may not be appropriate to estimate a Woods-related superstar effect as if the effect itself were constant over the entire 1999-2006 period. In fact, we find considerable differences in estimated superstar effects in the first- and second-halves of the estimation period.

3.3.4. Robust Standard Errors

Finally, although not evident in Brown's regression specification per se, the standard errors that she reports are robust standard errors clustered by player-year, and to maintain consistency with Brown, the standard errors that we report in all Table 1 regressions are calculated in similar fashion. But such standard errors do not capture round-course-specific variation in scoring beyond that reflected

[20] See, for example, Herity (2009).

in her γ estimates. As we show in Section 3.4 below, in regressions like (6), robust standard errors clustered by round-course tend to be two or more times those clustered by player-year (essentially the same as OLS standard errors).[21,22] This not only affects inferences of statistical significance, but also suggests potential model miss-specification issues. (See King and Roberts (2014).) In an effort to estimate a potential Woods-related superstar effect as accurately as possible, unless stated otherwise, all regression results that we report in the remainder of this paper are based on corrected data and employ robust standard errors clustered by round-course.[23] In this respect, our use of robust standard errors clustered by round-course is similar to the statistical methodology of Ehrenberg and Bognanno (1990) in their seminal work on tournament incentives in which they model event-specific variation in 72-hole (full tournament) scoring as a random effect. (Later, as described in Section 3.5, we also employ a random effects specification.)

3.4. Further Insight Into Brown's Main Regression Results

Regression (6) includes the estimation of fixed interacted player-course effects. As noted above, the inclusion of interacted player-course effects has the effect of implicitly eliminating roughly one-third of data observations. Moreover, the observations that are implicitly eliminated may not represent random omissions. This implicit loss of observations and possible bias in connection with non-random omissions can be avoided if fixed player and course effects are estimated separately. The cost of doing so is the inability to capture the possibility that players might perform better on some

[21]Each year, several events are conducted on more than one course. In first-round regressions, we define the course played as the course played in the first round and cluster by the first round course, which properly reflects that players could be assigned to different courses in the first round of some events. In event-level regressions, and consistent with Brown (per correspondence), we define the course as the course played in the final tournament round and cluster by the final-round course. Even in multiple-course events, all players who participate in the final round are assigned to the same course. We note that defining event-level courses in this fashion does not properly reflect differences in course rotation assignments leading to the final round. Although not reported, we have also defined the "course" and corresponding cluster for a multiple-course event as the course played in the first round. In any single multiple-course event, the first-round course assignment defines the course rotation. However, there is no guarantee that the first-round assignment, which defines the rotation in year t, would also define the assignment in other years. Although not reported, there are essentially no practical differences in any of our event-level results when we define event-level courses as first-round courses rather than final-round courses. See footnote 4 for further detail about course rotations in multiple-course events.

[22]Inasmuch as only one first-round or event-level scoring observation per event is reflected in regressions like (6), clustering by round-course is equivalent to clustering by year-course.

[23]We note that in both first-round and event-level regressions, Brown employs robust standard errors clustered by event (approximately the same as clustering by round-course), but does not report the results. In her footnote 17, Brown states "I also consider clustering by event. The standard errors on all coefficients are higher than when clustering by player-year but lead to a similar pattern of statistical significance. The exception is regression 3 in [Brown's] table 2 ..., where the coefficient on the effect for players ranked twenty-first to 200th is not statistically significant. These results are not reported."

courses than on others.

Panel A of Table 2 summarizes regression results using corrected data for the same regression specifications as in Table 1 but with two variations. The first variation is that we compute robust standard errors clustered by round-course rather than player-year (see footnote 21 for further details about round-course clustering), and the second is that we also estimate fixed player and course effects separately, thereby producing much larger effective sample sizes compared with regressions in which fixed player and course effects are estimated as interacted effects.

The two regressions summarized in panel A in which player and course effects are interacted produce coefficient estimates identical to those for corresponding regressions summarized in Table 2, but standard errors in Table 2 regressions are much higher, in many cases over twice those reported in Table 1, since they reflect round-course clustering rather than player-year clustering. The net effect is that with robust standard errors clustered by round-course, no *tif*-related coefficient estimate in player-course-interacted regressions summarized in panel A is statistically significant at the .05 level, and only one is close (the estimate associated with *tifH* in first-round data, with $p\text{-}value = 0.079$).

In the same panel we also show *tif*-related coefficient estimates for first-round and event-level regressions in which fixed player and course effects are estimated separately. This has the effect of increasing the effective sample size by including the events that are excluded implicitly when player and course are interacted.[24]

In first-round regressions where fixed player and course effects are estimated separately, all *tif*-related coefficient estimates are lower than those shown in connection with player-course interacted regressions, and the most important estimate for Brown's superstar hypothesis, that associated with *tifH*, falls from 0.505 to 0.184. In event-level regressions, the same estimate falls from 0.685 to 0.023, although estimates for *tifL* and *tifU* are somewhat higher in event-level regressions when player and course effects are estimated separately. Overall, in regressions for which player and course effects are estimated separately, the p-value most favorable to Brown's superstar hypothesis is 0.388 (that associated with *tifH* in first-round regressions), suggesting that there is no evidence

[24] As shown in Table 2, even when we estimate fixed player and course effects separately, the effective sample size is less than the full sample size, since we eliminate observations associated with players whose scores are observed only once. The exclusion of such players cannot affect coefficient estimates but will result in more accurate estimates of robust standard errors and adjusted R^2 values.

of a Woods-related superstar effect using that particular regression specification.

Panel B summarizes results for the same types of regressions but with time-dependent player and course effects rather than effects that are assumed to be fixed for the entire 1999-2006 estimation period. More specifically, for all players who recorded 20 or more scores, we interact individual player dummies with a first- and second-half dummy, where first- and second-half is defined in terms the first and second halves of the player's scores rather than calendar time. In addition, we interact course dummies with two dummies associated with calendar time, one indicating that the scoring observation occurred in the 1999-2002 period and another indicating that the observation occurred at some point during 2003 to 2006. These time-dependent player and course dummies come into play, not only in regressions in which we estimate player and course effects separately, but also in regressions with interacted player-course effects.

For each panel-B regression, the adjusted R^2 is higher than that reported in panel A, suggesting that conditioning player and course dummies on first- and second-half dummies improves the regression specifications. Note that in regressions in which player and course are interacted, the effective sample size is approximately half the size of the full sample, indicating that such regressions may not be as informative as those in which player and course effects are estimated separately.

In first-round regressions summarized in panel B, coefficient estimates tend to be lower than those for corresponding regressions summarized in panel A. In fact, in first-round panel B regressions in which player and course effects are estimated separately, coefficient estimates associated with *tifH*, *tifL*, and *tifH* are essentially zero (0.011, -0.076 and 0.003 , respectively). In panel B event-level regressions, *tif*-related coefficient estimates tend to be a little higher than estimates for corresponding regressions in panel A, but none is close to being statistically significant (minimum $p\text{-}value = 0.340$).

In Table 2, we also report coefficient estimates for TV viewership and field quality. In each regression, the coefficient estimate associated with each of these explanatory variables is negative, and in panel B, each of the four TV-related coefficient estimates is statistically significant at the .05 level. As mentioned previously, if Tiger Woods' presence in a PGA Tour event causes TV viewership and field quality to be higher than in an event without Woods, as appears to be the case, negative coefficient estimates associated with these two explanatory variables are an indication that any Woods-related effect on player scoring, if present, would actually be lower than that estimated in connection with the various *tif*-related coefficients. Together, the relatively low and statisitically insignificant *tif*-related coefficient estimates in combination with negative coefficient estimates for TV viewership and field quality suggests that there is little, if any, evidence that Woods' presence in a PGA Tour event had an adverse affect on the performance of other players.

Panel A of Table 3 summarizes coefficient estimates for regressions identical in form to those in Panel B of Table 2 except all *tif*-related dummies are interacted with dummies indicating whether the scoring observation occurred in 1999-2002 (denoted as ":1") or 2003-2006 (denoted as ":2"). With only one exception (*tifH*:2, in connection with first-round player-course-interacted regressions), all *tif*-related coefficient estimates associated with the 1999-2002 period are positive while those associated with the 2002-2006 period are negative. Moreover, only one of 24 *tif*-related coefficients interacted with *HRanked*, *LRanked*, and *URanked* and further interacted with 1999-2002 and 2003-2006 dummies is statistically significant at the .05 level (*tifH:*1 in player-course interacted event-level regressions), approximately what one would expect if 24 p-values were evaluated when there was no Woods effect at all. As in Table 2 panel B, all coefficient estimates related to TV viewership and field quality are negative, indicating that any adverse Woods effect as revealed in *tif*-related coefficient estimates may be over-stated.

In panel B of Table 3, we show coefficient estimates in connection with the same type of regressions summarized in panel A of Table 2 (that is, regressions in which player and course dummies are not interacted with first- and second-half dummies), but here, we estimate regressions separately for the 1999-2002 and 2003-2006 periods. In all 1999-2002 regressions (denoted as "1:"), each *tif*-related coefficient estimate is positive but only two are statistically significant at the .10 level. In all 2003-2006 regressions (denoted as ":2"), each tif-related coefficient estimate is negative and three of 12 are statistically significant at the .07 level. As in previous regressions, all coefficient estimates related to TV viewership are negative, and six of eight coefficient estimates related to field quality are negative, again suggesting that any adverse Woods effect as revealed in *tif*-related coefficient estimates may be over-stated.[25]

[25] If we re-estimate the same regressions summarized in Tables 2 and 3 using data set 1 rather than data set 2, the general tendencies that we have noted in this section are still evident, although *tif*-related coefficient estimates are somewhat higher, 2003-2006 estimates tend to be negative but not as negative as with corrected data, only 13 of 20 TV-related coefficient estimates are negative, but all 20 coefficient estimates associated with field quality are negative. We believe that the problems associated with Brown's TV viewership data, described in detail in Online Appendix A, most likely contribute to the smaller number of negative TV-related coefficient estimates.

3.5. Random Effects Specification

In addition to OLS with robust standard errors clustered by round-course, we also consider a random effects specification to account for round-specific variation in scoring. As described below, there is very little practical difference between the robust round-course results previously reported, and those which we obtain via the random effects model.

Referring back to Equations (3) and (4), although it would seem natural to estimate round-to-round variation in scoring in terms of fixed round-course effects, any *tif*-related effect would be absorbed by estimated fixed round-course effects; therefore, one cannot estimate fixed round-course and Tiger-in-field effects simultaneously. However, it may be possible to estimate Tiger-in-field effects while simultaneously estimating *random* round-course effects, provided the random effects model proves to be statistically consistent via the Hausman (1978) test relative to the corresponding OLS specification.

When considering a possible random effects specification, one often thinks of the Hausman test as a test of a random effects model versus a corresponding fixed effects specification. However, since there is no fixed effects model that corresponds directly to our random effects specification, we conduct the Hausman test, instead, as a test of a model *with* random effects model versus the corresponding OLS model *without* random effects. In general, the Hausman test evaluates the consistency of an estimator (in this case, random effects) relative to a less efficient alternative (in this case, OLS) that is known (or, at least, assumed) to be be consistent.

We estimate the random effects model using corrected data for all versions of the regression models summarized in Table 2 and in Panel A of Table 3. When we estimate fixed player and course effects separately, five of six p-values associated with the Hausman tests are approximately 1.0, with the remaining $p\text{-}value = 0.347$, providing strong support for the six random effects specifications. (Note, a p-value less than 0.05 would be required to reject the random effects model in favor of OLS at the .05 level.) Inasmuch as the six Hausman tests do not reject the random effects specifications, one would expect *tif*-related coefficients estimated in connection with the random effects models to be approximately the same as those estimated in connection with the corresponding OLS specifications. This is, in fact, the case, with the maximum difference across 24 tif-related estimates (interacted with *Hranked*, *Lranked*, and *Uranked*) of 0.092. The standard errors associated with the 24 coefficient estimates are all slightly higher than those associated with the alternative OLS-based model (using robust standard errors clustered by round-course), with a maximum difference of 29%. As such, whether we employ the OLS model with fixed player and course effects estimated separately (with robust standard errors clustered by round-course) or the random effects model, we generally come to the same conclusions regarding the magnitude of any tif-related effects. However, with slightly higher standard errors, the evidence in favor of a Tiger-in-field effect, if it exists at all, is even less compelling.

By contrast, when we estimate interacted fixed player-course effects rather than separate effects, it is not clear that random effects is the preferred model, since the Hausman statistic is negative in five of six regression specifications. Notwithstanding the general lack of evidence in favor of the random effects model, the tif-related coefficient estimates are similar in magnitude to those estimated via OLS. Moreover, standard errors are generally of the same order of magnitude as robust standard errors clustered by round-course, but there is no clear pattern of higher random effects-based standard errors relative to robust standard errors. Inasmuch as there is no compelling evidence in favor of the random effects model when player and course effects are estimated as interacted effects, we do not employ the random effects model in connection with interacted fixed player-course effects in any of our remaining work.

3.6. False (or Additional) Discoveries

In their seminal paper, Bertrand, Duflo, and Mullainathan (2004), show that with an improperly specified residual error structure, one can falsely discover far more statistically significant differences-in-differences "effects" than implied by the test level of statistical significance when, in fact, nothing at all is going on. Although our analysis does not involve differences-in-differences, we address the same general problem by asking how many players would be identified as superstars if each were allowed to take on the role of the superstar in individual regressions.[26] Even if no individual player affected the performance of others within the same tournament, with a properly specified empirical design, one would expect approximately 5 percent of players to be identified as

[26]For each designated player we create the dummy variable $star_j$, denoting whether the player is in the field and then remove all observations from the underlying data set involving the designated player. In these regressions, Tiger Woods is treated just like any other player, unless he happens to be the player identified as the superstar in a particular regression.

having statistically significant effects on the performance of other players at a .05 test significance level. If substantially more than 5 percent of players appear to affect the performance of other players, one can conclude that either 1) the underlying empirical design is flawed or 2) there are many players who affect the performance of others. In the latter case, if Woods happens to be in this group, the story of Woods as a superstar is not particularly interesting. (Of course, there is a practical limit to the number of players whose presence in a PGA Tour event can appear to affect the performance of other players.)

We re-run regressions for the same data and regression model specifications underlying our Tables 1 and 2 for all players who recorded at least 100 18-hole scores in PGA Tour stroke-play competition over the 1999-2006 period (368 players total).[27] For each specification, we set the dummy variable *star* equal to 1 in each event in which the designated "star" participated and estimate the coefficients for $star \times HRanked$, $star \times LRanked$, and $star \times URanked$ (corresponding to $tifH$, $tifL$ and $tifU$ when Woods, himself, is the potential superstar) and the statistical significance level for each, computing robust standard errors clustered by both player-year (as in Brown) and by round-course. We also estimate random effects versions of each regression specification, but only for regressions models in which player and course effects are estimated separately. After running all 368 regressions, we note the proportion of players in connection with each regression specification for whom each *star*-related coefficient estimate is statistically significant at the .05 level. These proportions, or "discovery rates," are reported in our Table 4 along with indications as to whether Woods is among the players so identified.

The first section of Table 4 summaries discovery rates for regressions using Brown's original player-course interacted regression specification and data that we believe closely approximates that which she employed. As such, the values that we report in the first section represent (approximately) what we believe Brown would have reported if she had undertaken the same "discovery" exercise. When computing robust standard errors clustered by player-year, as in Brown, between 13% and 55% of the 368 star-related coefficient estimates are identified as statistically significant at the .05 level. Woods is among those producing statistically significant coefficient estimates for four of

[27] Over half of the players who competed in PGA Tour events during the 1999-2006 period played 4 or fewer stroke-play rounds. (Woods played 561.) These players were mainly one- or two-time qualifiers for the U.S. Open, British Open and PGA Championship, who, otherwise, had no opportunity to compete in PGA Tour events. Including these players in the analysis along with others who played in just a small number of rounds would cause a disproportionate number of designated "superstars" to be playing in many of the same events.

six estimates. However, when we compute robust standard errors clustered by round-course, the discovery rates fall to the range of 4% to 15%, and no coefficient estimate is statistically significant at the .05 level when Woods is the designated superstar.

The second section of Table 4 summarizes discovery rates for the same regression specifications using corrected data rather than data that approximates that of Brown. Here we observe similar discovery rate patterns, although in most cases discovery rates are slightly lower.

The third section summarizes discovery rates for the same regression specifications, except fixed player and course effects are estimated separately rather than as interacted effects. As mentioned earlier, this has the effect of increasing effective sample sizes for each regression. Again, we observe similar discovery rate patterns when computing robust standard errors clustered by player-year and by round-course, although in most cases, the rates are slightly higher than those reported in the second section. However, with the random effects specification, discovery rates are much lower, falling between 5% and 8%.[28]

The fourth and fifth sections of Table 4 summarize discovery rates for the same regression specifications as the second and third sections, respectively, except player and course IDs are split into first- and second-half IDs as in panel B of Table 2. Again, we observe similar discovery rate patterns, with the lowest discovery rates occurring in connection with the random effects model when fixed player and course effects are estimated separately.

In summary, it is clear that Brown's original regression specification would have produced statistically significant star-related coefficient estimates for many players, not just Woods. But with regression error structures that account for round-course variation in scoring not otherwise captured by round-related proxy variables (i.e., round-course clustered standard errors and the random (round-course) effects model), .05-level discovery rates fall much closer to 5%, and Woods is never among the "discovered" players.

[28]We do not conduct the Hausman test in connection with each of the 368 regressions but, instead, only for the single regression in which Tiger Woods is the designated superstar. See Section 3.5 for further detail.

4. Additional Tests

4.1. Hot/Cool Analysis

Brown interprets her model as predicting an important relationship between player effort and the extent to which Woods might have been playing well or poorly over the same 1999-2006 period that underlies her main tests. More specifically, Brown states "When the superstar was relatively 'hot' ... , effort would be low and the superstar effect should be large. When the superstar was relatively 'cool' ..., effort would be high and the superstar effect should be small" (p. 1005). She then presents evidence that players performed differently when Woods was in a tournament field depending upon whether he had been "hot", "cool," or playing more typically during the month preceding the tournament.

Brown identifies hot, cool, and typical periods as follows (p. 1005):

> "I identify hot and cool periods by calculating the difference between Woods' average score and other ranked players' average score in the previous month. When Woods' performance is not remarkably better than other golfers' performances – score differences in the bottom quintile – he is in a cool period. When Woods' scores are remarkably lower than his competitors' scores – score differences in the top quintile – he is in a hot period. Score differences in the second to fourth quintiles represent Woods' typical performance."

In principle, the hot/cool approach should represent a more powerful test of Brown's incentives model than the basic superstar tests based on Equation (6), since the essence of her story is that when Tiger Woods was 'killing it' on the course, the rational response (according to her incentives model) was for players to throttle back their own efforts. Consistent with her model, Brown shows that players performed relatively poorly when Woods was 'hot' during the 1999-2006 period but played well during periods in which Woods' performance had 'cooled.'

Details of our analysis of player performance conditional upon Woods being "hot" or "cool" are presented in Online Appendix C. Unfortunately, we do not obtain results consistent with those of Brown when using data set 1, designed to closely approximate the data set that she used (see the B panels of Tables C1 and C2 in the appendix). Moreover, we find no evidence of the performance patterns Brown reports, whether we attempt to replicate Brown's results using data set 1 (B panels) or data set 2 (C panels), both with interacted player-course effects, or employ corrected data (data set 2) in connection with a regression model in which fixed player and course effects are estimated separately (D panels). Our inability to replicate Brown's pattern of results is independent of the error structure we assume (robust standard errors clustered by player-year or by round-course). These observations hold in connection with first-round and event-level regressions using both data sets 1 and 2. Differences in our results and those reported by Brown may reflect differences between our 'hot,' 'typical,' and 'cool' classifications and those of Brown when Woods did not participate in PGA Tour competition during the prior month. Further details associated with this possible source of discrepancy are provided in the appendix.

4.2. Surprise Absences from Tour

Brown points out that "a simple comparison of scores with and without Woods could be misleading" if "Woods plays only the most difficult courses or if particularly talented players avoid tournaments in which Woods is in the field" (p. 1000). (We note that the estimation of fixed player-course effects in Brown's regressions should control for relative course difficulty. Also, as we show in Section 5, field quality is higher, not lower, in events that include Woods.) An ideal approach to settling this question would require golfers other than Woods to commit in advance to a playing schedule with Woods participating randomly. While this is not practical, Brown exploits an alternative approach based on two periods when Wood's unexpectedly withdrew from PGA Tour competition. With respect to these two periods, Brown states:

> "In these scenarios, players prepared similarly in the months leading up to the competitive season but faced a different competitive environment at the start of the events. Given the timing of announcements about Woods' absences, it is reasonable to expect that players could not adjust the intensity of their long-term training programs but could adjust the intensity of the preparation activities for any given tournament." (p. 1001)

In this section we revisit the design and execution of Brown's tests in connection with Woods' surprise absences from Tour. Unfortunately, the analysis is tedious, due to problems in event selection and differences between the regression specification in Brown's paper and that in her STATA code. We address these issues in detail in Online Appendix D.

4.2.1. Evidence Presented in Brown's Table 3

In her Table 3, Brown compares player performance in tournaments in which Woods was expected to, but did not actually participate, with performance in adjacent years in which Woods was expected to, and did participate. Her tests reflect two unexpected absences from competition: absences due to Woods' knee surgery, announced in June 2008, which caused him to miss the remainder of the 2008 PGA Tour season, and absences due to well-publicized personal difficulties, which became evident in November 2009 and caused Woods to miss the January-March portion of the 2010 season.

In her Table 3, which focuses on mean event-level scores net of par with no controls for differences in player skill or relative course difficulty, Brown examines mean net scores per OWGR ranking category for events in which Woods normally competed but missed due to knee surgery and personal difficulties. As it turns out, the events that Brown actually includes in her Table 3 analysis are not consistent with the description of events in her paper. We provide more detail in Section A of Online Appendix D and describe "our" version of the data, which conforms, as best we can tell, with Brown's stated event selection criteria.

(Our) Table 5 summarizes our analysis of Brown's Table 3, which shows mean total scores net of par among players who made cuts for the three OWGR ranking categories in tournaments with and without Woods. Panel A shows results as presented by Brown. Panel B shows our results using Brown's events and OWGR categories. Panel C shows our results after making the event selection corrections noted in Section A of Online Appendix D and correcting OWGR categories as defined in data set 2 (see Online Appendix A, Section J.3).

Values in panel B correspond closely to Brown's actual values as presented in panel A, but most likely differ due to differences in OWGR ranking classifications.[29] Notwithstanding the slight differences in mean scores net of par as shown in panels A and B, as in Brown (p. 1003), "[the] summary statistics ... suggest a large adverse superstar effect in these periods of interest." However, once the data are corrected to include events consistent with Woods' periods of surprise absence, the

[29]In both Brown's Table 3 and panel A of our Table 5, there are 753 total observations. The mean score net of par is 1.12 in Brown's Table 3 and 1.14 in our panel B. This difference of 0.02 most likely reflects that the mean scores per category reported by Brown, from which we computed the 1.12 value, are rounded to two decimal places of precision. As we describe in Online Appendix A, Section E, we are unable to match Brown's monthly OWGR data to our ShotLink scoring records, and, therefore, we cannot replicate Brown's results using her actual OWGR ranking classifications.

mean overall score net of par is 4.33 in tournaments with Woods and 4.66 without Woods (values drawn from panel C but not shown in the table); i.e., on average, when Woods was unexpectedly absent from competition, players performed a little worse. Therefore, using events consistent with Brown's event selection criteria, there is no evidence in tests intended to mimic those that Brown reports in her Table 3 that players performed better when Woods was unexpectedly absent from competition. Of course, without controls for player skill and relative course difficulty, there is little information value in these results, even when corrected to include events that meet Brown's selection criteria.

4.2.2. Evidence Presented in Brown's Table 4

Brown's Table 4 provides a summary of a regression-based analysis of the same empirical strategy underlying her Table 3. In her Table 4 analysis, Brown includes a number of additional events during Woods' surgery and personal issues periods, not just those in which Woods' normally participated but missed unexpectedly, upon which her Table 3 (and our Table 5) analysis is based. As with her selection of events for her Table 3, the events she includes in her Table 4 analysis are not consistent with the description of events in her paper or in correspondence. We provide more detail in Section B of Online Appendix D along with descriptions of our corrected first-round and event-level data sets.

Brown's apparent aim is to extend her Table 3 analysis by controlling for player skill and relative course difficulty. We use the word "apparent," because, as described in Section C of this same appendix, we believe that Brown might not have run the regression that controls for both player skill and course difficulty as described in her paper and, instead, controlled only for player skill, at least in some regressions.[30] Notwithstanding this particular problem, we note that there is nothing in the regression specification reported in Brown's paper nor that which we believe she may have run that reflects the estimation of the *incremental* impact of Woods' unexpected absences. Neither regression (D-1) in Online Appendix D (the regression shown in Brown's paper) nor (D-2) (the regression shown in her STATA code) makes a distinction between events that Woods missed by choice and those he missed by surprise. This is a critical issue in properly interpreting Brown's

[30]Unlike the regressions underlying her main analysis, the regressions in this section do not interact player with course. As such, they are not subject to the problem of implicitly omitting events in which player-course interactions are observed only once.

results.

Rather than focus here on the effects of Brown's regression specification discrepancies and event choices, we address these issues in Online Appendix D. Instead, we focus on the results we believe Brown would have obtained if she had included events consistent with her selection criteria and adjusted the regression reported in her paper to include the incremental effect of Woods' surprise absences.

We base our analysis on the following regression specification, the last two lines of which represent the form of Brown's regression as stated in her paper.

$$\begin{aligned} s_{i,j} = {} & \phi_1 surprise_j \times HRanked_{i,j} + \phi_2 surprise_j \times LRanked_{i,j} + \phi_3 surprise_j \times URanked_{i,j} \\ & + \beta_1 tif_j \times HRanked_{i,j} + \beta_2 tif_j \times LRanked_{i,j} + \beta_3 tif_j \times URanked_{i,j} \\ & + \gamma_0 + \gamma_1 P_i + \gamma_2 C_j + \gamma_3 Z_j + \lambda_1 purse_j + \lambda_2 purse_j^2 + \epsilon_{i,j} \end{aligned} \tag{7}$$

In (7), P_i is a matrix of dummy variables that identifies players, C_j is a dummy variable matrix that identifies courses, and Z_j is a vector of event-specific controls for pre-event rainfall, wind, and temperature. (Except as noted below, all other variables in the second and third lines are as defined earlier.)

In this specification, we set $tif_j = 1$ for all observations associated with events that Woods played, or otherwise would have been expected to play, if he had not been absent from golf. We create a new dummy variable, $surprise_j$, which we set to zero for all observations except those associated with events that Woods missed by surprise. Thus $tif_j = 0$ for all events Woods missed by choice, $tif_j = 1$ for all events Woods played or, otherwise, would have been expected to play, and $surprise_j = 1$ for all events Woods missed by surprise. The estimated coefficients for $surprise_j$ interacted with $HRanked_{i,j}$, $LRanked_{i,j}$, and $URanked_{i,j}$, provide estimates by OWGR category of the incremental effect of Woods' surprise absences on player scoring. To be consistent with Brown's hypothesis, one would expect estimated $surprise_j$ coefficients to be negative, since players should have performed better (recorded lower scores) when Woods was unexpectedly absent from competition. Table 6 summarizes our regression results.

In the table, only three of 12 estimated *surprise*-related coefficients are negative, and none of the three is statistically significant. By contrast, six of the 12 coefficient estimates are positive and significant at the 0.001 level or better, all of which are associated with Woods' surprise absences for personal reasons. We doubt that there is a story here. But if there is, the story would be that when Woods was unexpectedly absent from Tour for personal reasons (but not due to knee surgery), players performed worse. (We note that all standard errors reported in Table 6 are robust standard errors clustered by player-year, as in Brown. Given these results, we see no need to report additional results using robust standard errors clustered by round-course.)

5. Field Quality with and without Woods

As we mention in Section 3.3, in corrected first-round data, we find that mean field quality is 1.98 in events with Woods and 1.16 without Woods, where field quality is defined as the mean OWGR points of all tournament participants, excluding Woods. In the same data set, if we include only events in which Woods participated in some years but not all years, estimated field quality is 1.67 and 1.34 in events with and without Woods, respectively.

As is well-known among those who follow professional golf, Woods highest priority was and continues to be winning "majors," the four events each year considered to be the most prestigious on the PGA Tour. Also, golf's major championships have always tended to attract the best golfers; in fact, eligibility requirements limit tournament fields of majors to golf's most successful players. Thus, the observation that tournament fields are the strongest in events with Woods could simply reflect that all of golf's best players, including Woods himself, are attracted to golf's major championships. The same could be said for events that offer the largest purses.[31]

In an effort to estimate the effect of Woods' tournament participation on field quality, while controlling for purse size and whether the event in which field quality is measured is a major championship, we estimate the following OLS regression for the 306 events represented in our corrected first-round data set:

$$Q_j = \underset{(0.178)}{0.893} + \underset{(0.050)}{0.078}\ purse_j + \underset{(0.121)}{0.366}\ major_j + \underset{(0.077)}{0.695}\ tif_j, \tag{8}$$

where Q_j denotes the field quality of tournament j, $purse_j$ denotes the inflation-adjusted purse of

[31]In general, whether a PGA Tour event is a major or non-major, priority is given to players who have been the most successful.

tournament j (in \$ millions), $major_j$ is a dummy variable that takes a value of 1 if tournament j is a major, the adjusted $R^2 = 0.331$, and robust standard errors are in parentheses.

If we include all events, but condition tif_j on Ψ_j and $(1 - \Psi_j)$, where Ψ_j is a dummy variable that takes a value of 1 if tournament j is among the events that Woods played in some years but not all years, we obtain:

$$Q_j = \underset{(0.173)}{1.019} + \underset{(0.049)}{0.042}\ purse_j - \underset{(0.139)}{0.024}\ major_j + \underset{(0.084)}{0.494}\ tif_j \times \Psi_j + \underset{(0.111)}{1.121}\ (1 - tif_j \times \Psi_j), \tag{9}$$

with N adjusted $R^2 = 0.383$. (We obtain similar results for both regressions when we estimate each in connection with data set 1.)

These estimates indicates a strong positive association between Woods' presence in a tournament field and the quality of the field, after controlling for purse size and whether the tournament is a major championship. Whether Woods' presence *causes* field quality to be higher is, perhaps, open to further testing. Nevertheless, it is clear that when Woods participates, tournament fields tend to be strong. This evidence, together with Brown's superstar hypothesis, suggests that golf's best players seek to participate in events that also include Woods and then give substandard effort once competition begins. We do not believe this is the story Brown is trying to tell. It seems inconceivable that the top golfers in the professional tour elect to play with Tiger Woods, but then throttle back their effort because he is in the tournament.

6. Summary and Conclusions

The superstar story, as Brown has studied it in golf, has important implications beyond professional sports. Firms that rely on high-skill employees often make use of explicit and implicit tournaments to establish the identify of their best performers. Taken at face value, Brown's evidence suggests performance in PGA Tour events may be adversely affected by the presence of a superstar competitor, in her case Tiger Woods, during the prime of his professional career and raises the question as to whether employment-related tournaments in other settings might suffer from the same effect. In this paper we show that Brown's evidence for an adverse superstar effect in professional golf simply doesn't survive.

In our study, we replicate and expand on Brown's main tests, using two data sets, one that approximates data employed by Brown and another, which reflects extensive corrections of known errors and includes additional PGA Tour events consistent with the event selection criteria as stated in Brown's paper and in correspondence. Our conclusion that Woods' presence in PGA Tour competition had no adverse effect of the performance of his fellow competitors is based only in small part on data corrections and the inclusion of additional events. Despite the relatively small impact, we believe it is important to report results using the best data available to estimate any Woods-related superstar effect as accurately as possible, whether it existed or not.

To obtain an estimate of the superstar effect, the challenge is to estimate the effect of the superstar's presence on player performance while controlling for other determinants of player scoring, i.e., individual player skill and across-the-board round-specific variation in scoring, that would otherwise be unaffected by the superstar's participation in the event. This is a non-trivial task, since it is difficult to separate the round- and/or event-specific effect of the superstar's presence on scoring from round- or event-specific variation in scoring that would have occurred naturally in the absence of the superstar. Moreover, it is nearly impossible to control for all possible sources of round- or event-specific variation in scoring without such controls absorbing all or part of a potential Woods-related effect, if present.

Brown's solution to the estimation problem is to employ a small set of round- and event-specific controls to proxy for the variation in scoring that would have otherwise occurred in Woods' absence. However, her estimates of standard errors in connection with Woods-related coefficient estimates do not reflect any potential round- or-event-specific variation in scoring beyond that reflected in her round- and event-specific proxies. If one estimates robust standard errors clustered by round (or event) and/or treats round- or event-specific variation beyond that captured by Brown's proxy variables as random effects, standard errors of Woods-related coefficient estimates can be two or more times those reported by Brown.

A typical, but implicit, assumption in tournament analysis is that the nature of the competition is stable over time. We show that this was not the case in professional golf over Brown's 1999-2006 estimation period and that it is essential to control for changes in the nature of the competition over time. For example, if we allow estimates of fixed player and course effects (and also interacted fixed-player and course effects) to be time-dependent, the effect of Wood's presence in PGA Tour competition effectively diminishes to zero. Moreover, if we allow the estimated superstar effect itself to be time-dependent, we find remarkable differences in the sizes and signs of estimated effects from the first half of the sample period to the second half. During the first half, estimates of Woods' effect on the scoring of others is positive (i.e., scores tend to be worse) but generally statistically insignificant while second-half estimates tend to be negative (i.e., scores tend to be better) but, again, statistically insignificant.

We also conclude that the inclusion of two round- or event-specific explanatory variables may cause any estimated adverse Woods-related superstar effect to be overstated. In particular, we find that higher TV viewership and stronger tournament fields both lead to *lower* scores. If either is caused by Wood's presence, the estimated superstar effect as reflected in direct Woods-related coefficient estimates will be overstated. We provide evidence that both TV viewership and field quality are higher in events in which Woods participates. If Woods is the cause, and there appears to be ample anecdotal evidence that he is the cause, at least when it comes to TV ratings, Brown-type estimates of a Woods-related adverse superstar effect will be too high.

We also address the issue of false discovery rates, which if high, could suggest a problematic empirical design. The premise behind Brown's superstar hypothesis is that there is a single dominant competitor, Woods, whose presence in a tournament has an adverse effect on the performance of all other players, especially the best players on Tour. Our false discovery analysis shows that Brown's main regression model, if applied to 368 active players taking on the role of the superstar, including Woods himself, would produce *far* more superstars than one might expect by chance alone. However, after allowing for time-dependent variation in fixed player and course effects and also accounting for round-specific variation in scoring when computing standard errors associated with superstar-related coefficient estimates, we find that the number of players who would be classified as 'superstars' is much closer to what one would expect by chance, and Woods is not among those so identified.

Although not part of what we consider to be Brown's main tests, we address two additional sets of tests that Brown presents in her paper. The first concerns the performance of players over the same 1999-2006 PGA Tour seasons during periods in which Woods performance during the previous month was 'hot' relative to that of other players compared with performance during his prior-month periods of 'typical and 'cool' performance. We find no evidence, even using data with known errors and event omissions, that players performed differentially worse when Woods was 'hot.' Differences in our results and those reported by Brown may reflect differences between our 'hot,' 'typical,' and 'cool' classifications and those of Brown when Woods did not participate in PGA Tour competition during the prior month.

Brown also provides evidence that when Woods was unexpectedly absent from PGA Tour competition during two separate periods, one in 2008 and the other in 2010, that players performed better in events in which Woods would have otherwise been expected to compete compared with their performance in the same events when Woods was able to participate. We find numerous discrepancies between included events, as stated in her paper, and events actually employed in her analysis. Moreover, we note that Brown's regression methodology does not account for the *incremental* impact of Woods' surprise absences from competition. Once corrected, we find no evidence that players performed better when Woods was unexpectedly absent from competition. If anything, they actually performed worse.

Although a byproduct of our work, we find that tournament field quality was higher over the 1999-2006 period in events with Woods compared with events without Woods. Moreover, when we restrict the analysis to events in which Woods participated in some years but not all years, we find the same – higher field quality when Woods participated. If a Woods-related superstar effect actually existed, as Brown suggests, we find it ironic that top players would be attracted to events that included Woods and then throttle back their efforts once competition began.

References

Berry, Scott M.. 2001. "How Ferocious is Tiger?" *Chance* 14:51-56.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics* 119: 249-275.

Broadie, M. 2012. "Assessing Golfer Performance on the PGA TOUR." Interfaces 42(2):146-165.

Broadie, Mark and Richard Rendleman, Jr. 2013. "Are the Official World Golf Rankings Biased?" *Journal of Quantitative Analysis in Sports*, 9:127-140

Brown, Jennifer. 2008. "Quitters Never Win: The (Adverse) Incentive Effects of Competing with Superstars." Working paper dated September 2008.

Brown, Jennifer. 2011. "Quitters Never Win: The (Adverse) Incentive Effects of Competing with Superstars." *Journal of Political Economy* 119:982-1013.

Coffey, Bentley and M.T. Maloney. 2010. "The Thrill of Victory: Measuring the Incentive to Win." *Journal of Labor Economics.* 28:87-112.

Connolly, Robert A., and Richard J. Rendleman, Jr. 2008."Skill, Luck, and Streaky Play on the PGA Tour." *Journal of the American Statistical Association.* 103:74-88.

Connolly, Robert A., and Richard J. Rendleman, Jr. 2009. "Dominance, Intimidation, and 'Choking' on the PGA Tour." *Journal of Quantitative Analysis in Sports.* 5 (May, article 6).

Connolly, Robert A., and Richard J. Rendleman, Jr. 2012a, "Tournament Selection Efficiency: An Analysis of the PGA TOUR's FedExCup." *Journal of Quantitative Analysis in Sports.* 8 (November, article 1).

Connolly, Robert A., and Richard J. Rendleman, Jr. 2012b. "What it Takes to Win on the PGA Tour (If Your Name is "Tiger" or If It Isn't)." *Interfaces.* 42:554-576.

Ehrenberg, Ronald G., and Michael L. Bognanno. 1990. "Do Tournaments Have Incentive Effects?" *Journal of Political Economy.* 98:1307-1324.

Hausman, Jerry. A. 1978. "Specification Tests in Econometrics." *Econometrica.* 46: 2511271.

Herity, Jim. 2009. "The Shapes: Evolution of Improvement." *golfdigest.com.* http://www.golfdigest.com/story/hotlistevolution-0902. (January 3).

Kay, Emily. 2015. "Tiger Woods' Return to Golf Boosts the 2015 Masters." http://www.sbnation.com/golf/2015/4/13/8400427/tiger-woods-masters-2015-tv-ratings. (April 13)

King, Gary and Margaret E. Roberts. 2015. "How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About It." *Political Analysis.* 23:159179.

Lazear, Edward P., and Sherwin Rosen. 1981. "Rank-order Tournaments as Optimum Labor Contracts." *Journal of Political Economy.* 89: 841-64.

Myers, Alex. 2015. "CBS Owes Tiger a Big Thank-you for these Ridiculous TV Ratings." http://www.golfdigest.com/story/cbs-owes-tiger-a-big-thank-you-for-these-ridiculous-tv-ratings.

Napoli, Philip M. 1997. "Rethinking Program Diversity Assessment: An Audience-Centered Approach." *Journal of Media Economics.* 10:59-74.

Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature.* 37: 7-63.

Szymanski, Stefan. 2003. "The Economic Design of Sporting Contests." *Journal of Economic Literature*. 61: 1137-1187.

Table 1: Analysis of Brown's Table 2 Using Data that Include Majors (Robust Standard Errors Clustered by Player-Year)

Analysis Type	Data Set	Stat.	tif	tifH	tifL	tifU	N / N*	Adj R^2
First-Round	Brown Actual	Est.	n/r	0.596	0.161	0.202	34,966	0.29
		S.E.	n/r	(0.281)	(0.113)	(0.126)	24,066 (est.)	
		p-value	n/r	0.034	0.154	0.109		
	Data Set 1	Est.	0.257	0.715	0.207	0.247	36,379	0.188
	(Brown-Type)	S.E.	(0.081)	(0.228)	(0.095)	(0.111)	25,039	
		p-value	0.002	0.002	0.029	0.026		
	Data Set 2	Est.	0.169	0.505	0.089	0.223	43,256	0.204
	(Corrected)	S.E.	(0.077)	(0.205)	(0.090)	(0.103)	30,684	
		p-value	0.028	0.014	0.322	0.031		
Event-Level	Brown Actual	Est.	n/r	1.358	0.804	0.596	18,805	0.48
		S.E.	n/r	(0.726)	(0.318)	(0.396)	12,155 (est.)	
		p-value	n/r	0.061	0.011	0.132		
	Data Set 1	Est.	0.407	0.964	0.388	0.174	18,822	0.375
	(Brown-Type)	S.E.	(0.231)	(0.572)	(0.254)	(0.336)	12,166	
		p-value	0.079	0.092	0.126	0.603		
	Data Set 2	Est.	0.051	0.685	0.018	-0.154	22,312	0.452
	(Corrected)	S.E.	(0.208)	(0.485)	(0.228)	(0.290)	14,789	
		p-value	0.807	0.158	0.936	0.595		

n/r denotes not reported. As in Brown, standard errors are robust standard errors clustered by player-year. Brown does not report p-values, but those shown in the "Brown Actual" sections are based on the coefficient estimates and standard errors that Brown reports. As described in Section J of Online Appendix A, data set 1 closely matches that used by Brown, and data set 2 corrects known errors and includes all events that meet Brown's selection criteria as implied in her paper and in correspondence. N is the actual (or full) sample size and N^* is the "effective" sample size, after omitting observations in connection with player-course interactions observed only once. We estimate effective sample sizes for Brown's actual regressions by multiplying her actual sample sizes by the ratio of the effective sample size in data set 1 to the actual sample size in the same data set. Adjusted R^2 values for data set 1 and data set 2 are based on regressions using effective sample sizes; using actual sample sizes, R^2 values are higher. R^2 values reported by Brown appear to have been based based on actual samples.

Table 2: Re-analysis of Brown's Main Results Using Corrected Data, Robust Standard Errors Clustered by Round-Course, and Player and Course Effects Divided into First- and Second-Half Components

Panel A: Player and Course Effects (and Player-Course) Effects are Fixed for the Entire 1999-2006 Period

Analysis Type	Plr & Course	Stat.	tif	tifH	tifL	tifU	TV	Quality	N / N*	Adj R^2
First-round	Interacted	Est.	0.169	0.505	0.089	0.224	-1.081	-0.730	43,256	0.204
		S.E.	(0.197)	(0.288)	(0.210)	(0.199)	(0.945)	(0.293)	30,684	
		p-value	0.390	0.079	0.671	0.261	0.253	0.013		
First-round	Separate	Est.	0.113	0.184	0.079	0.140	-0.460	-0.618	43,256	0.262
		S.E.	(0.181)	(0.213)	(0.185)	(0.184)	(0.796)	(0.251)	42,402	
		p-value	0.534	0.388	0.669	0.448	0.564	0.014		
Event-level	Interacted	Est.	0.051	0.685	0.018	-0.154	-0.387	-0.767	22,312	0.452
		S.E.	(0.556)	(0.782)	(0.564)	(0.596)	(0.406)	(0.553)	14,789	
		p-value	0.927	0.382	0.974	0.796	0.341	0.165		
Event-level	Separate	Est.	0.076	0.023	0.075	0.093	-0.484	-0.353	22,312	0.498
		S.E.	(0.535)	(0.577)	(0.538)	(0.554)	(0.402)	(0.498)	22,110	
		p-value	0.887	0.968	0.889	0.867	0.228	0.477		

Panel B: Player and Course Effects (and Player-Course) Effects Split into First and Second Halves

Analysis Type	Plr & Course	Stat.	tif	tifH	tifL	tifU	TV	Quality	N / N*	Adj R^2
First-round	Interacted	Est.	0.072	0.290	-0.076	0.316	-2.946	-0.689	43,256	0.233
		S.E.	(0.257)	(0.339)	(0.271)	(0.255)	(1.249)	(0.395)	22,370	
		p-value	0.779	0.391	0.779	0.214	0.018	0.081		
First-round	Separate	Est.	-0.037	0.011	-0.076	0.003	-2.211	-0.394	43,256	0.279
		S.E.	(0.200)	(0.229)	(0.201)	(0.204)	(1.026)	(0.347)	42,402	
		p-value	0.852	0.963	0.704	0.988	0.031	0.256		
Event-level	Interacted	Est.	0.209	0.813	0.216	-0.276	-0.796	-0.442	22,312	0.471
		S.E.	(0.659)	(0.852)	(0.674)	(0.749)	(0.407)	(0.694)	10,037	
		p-value	0.752	0.340	0.749	0.713	0.050	0.524		
Event-level	Separate	Est.	0.217	0.113	0.220	0.243	-0.745	-0.446	22,312	0.519
		S.E.	(0.556)	(0.587)	(0.556)	(0.581)	(0.369)	(0.630)	22,110	
		p-value	0.696	0.847	0.693	0.676	0.043	0.479		

All regression results based on data set 2 (corrected data). "TV" denotes TV viewership. "Quality" denotes field quality. Standard errors are robust standard errors clustered by round-course. N is the actual (or full) sample size. When interacted player-course effects are estimated, N^* is the sample size after omitting observations in connection with player-course interactions observed only once. When player and course effects are estimated separately, N^* is the sample size after omitting observations in connection with players observed only once. In Panel B, fixed player effects are divided into first- and second-half components for players with 20 or more scoring observations. (Here, first- and second-halves are defined in terms of individual player observations, not in terms of calendar time, 1999-2002 and 2003-2006.) First- and second-half course effects are defined in terms of calendar time, 1999-2002 and 2003-2006. Adjusted R^2 values are based on regressions using effective samples of size N^*.

Table 3: Re-analysis of Brown's Main Results Using Corrected Data, Robust Standard Errors Clustered by Round-Course, Player and Course Effects Divided into First- and Second-Half Components (Panel A Only), and *tif*-Related Effects Conditional on First- and Second-Half of Sample

Panel A: Single 1999-2006 Regression with tif-Related Coefficients Estimated for 1999-2002 (1) and 2003-2006 (2)

Analysis Type	Plr & Course	Stat.	tif:1	tif:2	tifH:1	tifL:1	tifU:1	tifH:2	tifL:2	tifU:2	TV	Quality	N / N*	Adj R^2
First-round	Interacted	Est.	0.369	-0.342	0.384	0.224	0.639	0.134	-0.487	-0.166	-2.902	-0.713	43,256	0.234
		S.E.	(0.321)	(0.365)	(0.432)	(0.346)	(0.302)	(0.476)	(0.373)	(0.386)	(1.232)	(0.399)	22,370	
		p-value	0.250	0.349	0.374	0.519	0.034	0.778	0.192	0.667	0.018	0.074		
First-round	Separate	Est.	0.234	-0.347	0.197	0.209	0.266	-0.210	-0.393	-0.307	-2.181	-0.407	43,256	0.280
		S.E.	(0.263)	(0.282)	(0.308)	(0.266)	(0.265)	(0.311)	(0.283)	(0.291)	(1.012)	(0.348)	42,402	
		p-value	0.372	0.219	0.522	0.432	0.315	0.500	0.166	0.290	0.031	0.242		
Event-level	Interacted	Est.	1.242	-1.146	2.163	1.180	0.894	-0.677	-1.072	-1.982	-0.729	-0.392	22,312	0.473
		S.E.	(0.955)	(0.695)	(1.099)	(0.946)	(1.090)	(1.076)	(0.751)	(0.718)	(0.394)	(0.697)	10,037	
		p-value	0.194	0.099	0.049	0.212	0.412	0.529	0.153	0.006	0.064	0.574		
Event-level	Separate	Est.	1.230	-1.026	1.055	1.260	1.225	-1.018	-1.037	-1.003	-0.708	-0.381	22,312	0.520
		S.E.	(0.806)	(0.643)	(0.847)	(0.797)	(0.850)	(0.704)	(0.652)	(0.659)	(0.362)	(0.624)	22,110	
		p-value	0.127	0.111	0.213	0.114	0.150	0.148	0.112	0.128	0.051	0.541		

Panel B: Regressions Estimated Separately, 1999-2002 (1) and 2003-2006 (2)

Analysis Type	Plr & Course	Stat.	tif:1	tifH:1	tifL:1	tifU:1	TV	Quality	N / N*	Adj R^2
First-round Sample Split	Interacted	Est.	0.420	0.485	0.299	0.590	-6.714	-0.780	21,424	0.212
		S.E.	(0.312)	(0.428)	(0.331)	(0.316)	(2.356)	(0.384)	13,654	
		p-value	0.179	0.257	0.368	0.062	0.004	0.042		
First-round Sample Split	Separate	Est.	0.298	0.300	0.286	0.310	-5.840	-0.581	21,424	0.275
		S.E.	(0.271)	(0.320)	(0.275)	(0.276)	(2.121)	(0.340)	20,807	
		p-value	0.273	0.350	0.297	0.262	0.006	0.088		
Event-level Sample Split	Interacted	Est.	1.217	2.159	1.388	0.134	-0.544	-1.769	10,958	0.464
		S.E.	(1.137)	(1.295)	(1.106)	(1.405)	(0.669)	(0.927)	3,661	
		p-value	0.285	0.096	0.210	0.924	0.416	0.056		
Event-level Sample Split	Separate	Est.	1.042	0.775	1.196	0.851	-0.920	-1.244	10,958	0.537
		S.E.	(0.815)	(0.860)	(0.810)	(0.857)	(0.526)	(0.699)	10,835	
		p-value	0.201	0.368	0.140	0.321	0.080	0.075		

Analysis Type	Plr & Course		tif:2	tifH:2	tifL:2	tifU:2	TV	Quality	N / N*	Adj R^2
First-round Sample Split	Interacted	Est.	-0.332	-0.044	-0.416	-0.233	-0.837	-0.747	21,832	0.222
		S.E.	(0.349)	(0.470)	(0.365)	(0.346)	(1.198)	(1.092)	13,436	
		p-value	0.342	0.926	0.254	0.501	0.485	0.494		
First-round Sample Split	Separate	Est.	-0.334	-0.246	-0.378	-0.290	-0.412	-0.390	21,832	0.271
		S.E.	(0.285)	(0.311)	(0.287)	(0.296)	(0.854)	(0.958)	21,279	
		p-value	0.241	0.429	0.187	0.327	0.629	0.684		
Event-level Sample Split	Interacted	Est.	-1.417	-0.077	-1.605	-1.788	-0.048	0.257	11,354	0.492
		S.E.	(0.657)	(1.423)	(0.685)	(0.919)	(0.681)	(1.551)	3,559	
		p-value	0.031	0.957	0.019	0.052	0.944	0.869		
Event-level Sample Split	Separate	Est.	-1.097	-1.044	-1.181	-0.941	-0.142	1.122	11,354	0.513
		S.E.	(0.629)	(0.679)	(0.640)	(0.652)	(0.508)	(1.276)	11,206	
		p-value	0.081	0.124	0.065	0.149	0.780	0.379		

All regression results based on data set 2 (corrected data). "TV" denotes TV viewership. "Quality" denotes field quality. Standard errors are robust standard errors clustered by round-course. N is the actual (or full) sample size. When interacted player-course effects are estimated, N^* is the sample size after omitting observations in connection with player-course interactions observed only once. When player and course effects are estimated separately, N^* is the sample size after omitting observations in connection with players observed only once. In Panel A, fixed player effects are divided into first- and second-half components for players with 20 or more scoring observations. (Here, first- and second-halves are defined in terms of individual player observations, not in terms of calendar time, 1999-2002 and 2003-2006.) First- and second-half course effects are defined in terms of calendar time, 1999-2002 and 2003-2006. Adjusted R^2 values are based on regressions using effective samples of size N^*. In Panel B, two separate regressions are run, the first using 10999-2002 data and the second using 2003-2006 data. In both cases, player and course effects are assumed to be fixed for the entire 4-year estimation period.

Table 4: Discovery Rate (.05 Level) Among Payers who Recorded at Least 100 18-hole Scores over the 1999-2006 Period, 368 Players Total

Original Table Reference	Player and Course IDs Split in Halves	Player and Course	Data Set	Coefficient	Player-Year Clustering		Round-Course Clustering		Random Effects Model	
					First Round	Event-Level	First Round	Event-Level	First Round	Event-Level
1	No	Interacted	Brown-Type	$star \times HRanked$	0.130†‡	0.231	0.035	0.109	n/a	n/a
				$star \times LRanked$	0.443†	0.554†‡	0.084	0.149	n/a	n/a
				$star \times URanked$	0.364†	0.424	0.071	0.141	n/a	n/a
2A	No	Interacted	Corrected	$star \times HRanked$	0.114†	0.193	0.035	0.076	n/a	n/a
				$star \times LRanked$	0.459	0.457	0.111	0.120	n/a	n/a
				$star \times URanked$	0.326†	0.353	0.073	0.087	n/a	n/a
2A	No	Separate	Corrected	$star \times HRanked$	0.182	0.274	0.060	0.120	0.052	0.079
				$star \times LRanked$	0.524	0.535	0.122	0.087	0.082	0.063
				$star \times URanked$	0.448	0.484	0.092	0.076	0.084	0.052
2B	Yes	Interacted	Corrected	$star \times HRanked$	0.098	0.253	0.046	0.141	n/a	n/a
				$star \times LRanked$	0.402	0.470	0.084	0.182	n/a	n/a
				$star \times URanked$	0.299	0.269	0.098	0.079	n/a	n/a
2B	Yes	Separate	Corrected	$star \times HRanked$	0.171	0.299	0.076	0.144	0.057	0.076
				$star \times LRanked$	0.500	0.576	0.147	0.147	0.101	0.106
				$star \times URanked$	0.451	0.508	0.122	0.103	0.092	0.071

"Original Table Reference" indicates the table and panel (if applicable) where the corresponding regressions involving Woods only are summarized. $star \times HRanked$, etc. denotes the interaction of the player identified as the superstar and the OWGR dummies, *HRanked*, *LRanked*, and *URanked*. Each table entry denotes the proportion of regression coefficient estimates identified as statistically significant at the .05 level in 368 regressions (per regression group) in which one of 368 players who recorded at least 100 18-hole scores over the 1999-2002 period takes on the role of the superstar. † indicates that the coefficient estimate is statistically significant when Tiger Woods takes on the role of the superstar. ‡ indicates that the coefficient estimate using the standard error reported by Brown is statistically significant at the .05 level (when Woods is the superstar). n/a indicates that we did not run the discovery rate analysis for these regressions.

Table 5: Mean Tournament Scores Relative to Par in Selected Events With and Without Tiger Woods

	Panel A: As Brown Reports			Panel B: Replication of Brown			Panel C: With Event Corrections		
	Knee Surgery 2007-9		Personal Issues 2009-10	Knee Surgery 2007-9		Personal Issues 2009-10	Knee Surgery 2007-9		Personal Issues 2009-10
	July	August	March	July	August	March	July	August	March
Ranked 1-20									
With Woods	-4.00	4.51	4.60	-3.60	4.50	4.60	1.55	4.71	-3.75
	(3.61)	(7.29)	(4.16)	(4.38)	(7.18)	(4.16)	(6.18)	(5.94)	(7.65)
Observations	5	35	5	10	36	5	38	62	24
Without Woods	-8.00	-0.17	-1.00	-8.00	-0.42	-1.00	8.41	2.33	-4.00
	(5.24)	(5.06)	(6.04)	(5.24)	(5.04)	(6.04)	(9.81)	(6.52)	(5.76)
Observations	9	18	9	5	19	9	17	30	27
Ranked 21-200									
With Woods	2.01	3.84	5.84	1.87	4.15	5.71	3.41	6.93	0.68
	(4.70)	(10.05)	(5.20)	(4.76)	(10.33)	(5.02)	(5.19)	(6.57)	(7.38)
Observations	86	136	44	90	149	49	182	228	106
Without Woods	-5.97	2.63	1.87	-6.24	2.59	1.78	7.18	7.54	-0.22
	(4.93)	(6.44)	(5.07)	(4.64)	(6.39)	(4.95)	(10.78)	(7.71)	(5.49)
Observations	73	57	47	80	58	49	88	117	100
Unranked (over 200)									
With Woods	4.13	-2.68	6.48	4.67	-5.76	7.00	5.66	11.65	5.55
	(4.69)	(11.43)	(4.14)	(4.24)	(9.27)	(4.33)	(5.45)	(6.33)	(6.23)
Observations	53	59	23	48	45	18	70	17	20
Without Woods	-4.09	5.80	3.21	-3.54	11.33	3.65	7.59	10.67	3.88
	(4.58)	(7.82)	(5.75)	(4.69)	(1.15)	(6.03)	(9.63)	(7.79)	(6.41)
Observations	70	5	19	63	3	17	49	6	16

Note.–Values in parentheses are standard errors. Months and years of interest reflect the timing of Woods' unexpected absences from competition. Only scores of players who made cuts are included, and these scores are net of par. "With Woods" indicates that Woods participated in the tournament and "Without Woods" includes the same tournaments in which Woods did not participate in the years of interest. Scores for Woods are excluded. Values in Panel B employ OWGR categories as defined for data set 1. Values in Panel C employ OWGR categories as defined in data set 2.

Table 6: Analysis of the Incremental Impact of Woods' Surprise Absences from Tour

Regression	Estimate	surprise × Hranked	surprise × Lranked	surprise × Uranked	Obs.
First Round	Coef.	0.257	-0.178	0.198	2,565
Surgery	Std Error	0.885	0.617	0.723	
	p-value	0.772	0.774	0.784	
Event-Level	Coef.	-0.787	-1.520	0.439	1,484
Surgery	Std Error	1.870	1.201	1.494	
	p-value	0.674	0.206	0.769	
First Round	Coef.	3.580	3.416	5.499	2,447
Personal	Std Error	0.941	0.627	0.916	
	p-value	0.000	0.000	0.000	
Event-Level	Coef.	9.641	9.519	10.779	1,245
Personal	Std Error	2.130	1.349	2.236	
	p-value	0.000	0.000	0.000	

Note.– Regressions reflect data that meet Brown's event selection criteria as described in Online Appendix D with correction for data errors. Standard errors are computed as robust standard errors clustered by player-year as in Brown and in Table D3 of Online Appendix D.

Online Appendix A
Programming Issues and Data Corrections

A. Information Sources

Brown provided us with the following:

1. A condensed listing of STATA code used in connection with her study.
2. Her weather-related data.
3. Her Official World Golf Ranking (OWGR) data.
4. Lists of events included in regressions summarized in her Tables 2 and 5.
5. A list of events included in her Table 4 regressions.

We were able to infer the events included in Brown's Table 3 analysis from results shown in the paper. We also obtained other information from email correspondence with Brown.

B. Programming Issues

Brown provided us with a condensed listing of the STATA code she used in connection with her study. Some of the regression code varies slightly from the regression specifications shown in her paper. Also, some of the code is in error, which causes us to believe that the condensed code that she provided us might not have been representative of the code she actually used.[32] Although we were able to resolve some programming issues through correspondence, other issues continue to be unresolved.

Within the STATA code, in all regressions in which inflation-adjusted tournament purse and the square of the inflation-adjusted tournament purse are used as explanatory variables, each is interacted with the dummy variables *Hranked*, *Lranked*, and *Uranked*, which indicate whether the player's OWGR is 1-20, 21-200, or greater than 200, respectively. Within the paper, however, no indication is provided that the purse-related explanatory variables are interacted with the three OWGR categories. In fact, in footnote 20, Brown indicates that "higher purses actually appear to

[32] For example, the code for computing field strength associated with first round scores is as follows.

1. egen n_players_beforecut=count(score), by(course_location year)
2. egen fieldrank_beforecut=mean(OWGRrank), by(tournament year)
3. gen fieldstrength_beforecut=(n_players_beforecut*fieldrank_beforecut- OWGRrank)/(n_players_beforecut-1)

In statement (2), the function mean(OWGRrank) has the effect of computing the mean OWGR rank of only those players with non-missing values of OWGRrank, i.e., players in ranking positions 1-200 only. Thus, Brown's unranked players, associated with approximately 47% of the observations in her first-round data set, would not factor into the calculation of mean(OWGRrank). The third statement has the effect of assigning missing values to fieldstrength_beforecut for the unranked players which, in turn, causes observations associated with these missing values to be excluded from subsequent regressions. Thus, if these STATA statements are taken literally, approximately 47% of Brown's first-round data observations would be excluded from her regressions, since STATA regressions omit records with missing data. We note that such exclusions would prohibit the estimation of regression coefficients for any explanatory variables related to her dummy variable *URanked*. As such, we believe that this portion of the STATA code Brown sent to us must be in error; otherwise she could not have estimated the regression coefficients associated with *URanked* that she reports.

induce slightly higher scores: raising the purse by $100,000 is associated with a 0.04 increase in the first-round score." This would suggest that she employed unconditional versions of the purse and purse squared variables rather than versions in which the two were interacted with OWGR categories. To maintain consistency with what Brown actually reports in her paper, we have chosen not to interact purse-related explanatory variables with $Hranked$, $Lranked$, and $Uranked$.

C. Event Selection

C.1. Events Included in Brown's Table 2 and Table 5 Analyses

In the 1999-2006 first-round and event-level (total score) "Regulars and Majors" data sets used in her test summarized in her Tables 2 and 5, Brown excludes the following events, which we believe should have been included based on the selection criteria stated in her paper and in correspondence.

1. 12 events, originally scheduled for 5 rounds, are excluded from both the first-round and event-level data sets.[33]

2. Three events, originally scheduled for 4 rounds, but cut short to 2 or 3 rounds due to adverse weather, are excluded from the event-level data set only.[34]

3. All 1999-2006 occurrences of THE PLAYERS Championship, Greater Milwaukee Open (US Bank Championship), Worldcom (MCI, Verizon) Heritage Classic and Kemper Insurance (FBR, Booze Allen) Open are excluded from both data sets. Also, all occurrences of the Air Canada Championship, conducted in years 1999-2002 only, are excluded.

4. 11 additional events are excluded from both data sets but not in each year.[35]

5. We believe there may be other omissions, especially in Brown's first-round data set, since we cannot reconcile the number of observations she should have obtained with the observations she reports in her Table 2.[36]

C.2. Events Included in Brown's Analysis of Woods' Surprise Absences (Tables 2 and 3)

In Online Appendix D, we address issues related to Brown's event selection and differences between her stated regression equation and that shown in her STATA code.

[33]These include the 1999-2006 Bob Hope Chrysler Classic and the 1999-2002 Las Vegas Invitational. After 2002, the Las Vegas event was shortened to four rounds.

[34]These events include the 2000 and 2005 BellSouth Classic and 2005 Nissan Open. The 1999 AT&T Pebble Beach National Pro-Am, cut short to three rounds, is excluded altogether and included among the events listed in item 4, below.

[35]These events include the 1999 and 2002 AT&T Pebble Beach National Pro-Am, 2002 and 2005 Bay Hill Invitational, 2004 Honda Classic, 2002 Compaq Classic of New Orleans, 1999 B.C. Open, 1999 and 2000 Buick Challenge, 1999 British Open, and 2006 WGC-American Express.

[36]Brown reports 34,986 first-round and 18,805 total-score observations in Table 2. These exclude observations involving Tiger Woods. Using the 259 first-round events that Brown told us that she used in correspondence, there are a total of 36,494 observations, including those of Woods. Excluding 115 observations involving Woods, gives 36,379 net of Woods. Thus, Brown's reported 34,986 first-round observations fall $36,379 - 34,986 = 1,393$ observations short of what we calculate for the same 259 events. This would suggest that she omitted approximately 10 additional events in her first-round data set. Using similar methodology, the 18,805 observations that Brown reports for the 256 events in her total-score data set fall 17 observations short of our count for the same events. An alternative explanation for these omissions is that Brown may not have properly matched her OWGR data to the PGA Tour ShotLink scoring data for some players. See Section E, below, for further explanation.

D. Errors in Weather Data and Application in Regressions

We discovered a significant error in Brown's pre-event total rainfall calculation for 60 of 127 events for which we could find detailed rainfall amounts from the NOAA website for weather stations that Brown identifies in her data. Rather than compute total rainfall in the three days prior to each event, as described in the paper, Brown computes total rainfall over the four rounds of each event. Moreover, for all events, Brown uses average temperature and wind speed over the entire event to predict first-round scores.[37] As in her use of after-the-fact temperature and wind speed measures, the error in her pre-event rainfall calculation also results in the use of weather-related data that has not yet occurred to predict first-round scoring.

In the documentation of her weather-related data that she sent to us, Brown indicates that for some events (mainly international), rainfall amounts are monthly averages, and for one event, the wind speed amount is the average for the year.

In correspondence, we asked Brown to confirm the errors we discovered in pre-event rainfall. Her response was "I collected the rainfall data by hand from the NOAA website in the summer of 2007 using individual searches of tournament locations. My notes reflect the following calculation: I calculated the sum of the rainfall for the three days [prior] to the day in question."

Given the problems in the weather-related data that Brown provided us, we collected new weather data ourselves. We describe our weather data and collection process in Online Appendix B.

E. Official World Golf Rankings

In an attempt to replicate Brown's work, we tried to merge the monthly OWGR data that she provided us with our PGA Tour ShotLink scoring data. Brown's OWGR data is for the top 200 players in the OWGR, updated monthly, and contains no ShotLink player ID numbers. Therefore, we attempted to match her OWGR data to the ShotLink data by player name, the way she matches the two data sets in her STATA code.[38]

We were unable to associate approximately half the ShotLink scoring records with Brown's OWGR data. Many of the unmatched records were associated with players legitimately outside the OWGR top 200 and, therefore, could be coded properly as such. However, other records could have been unmatched due to differences in player name spellings. For example, we were unable to match ShotLink records for Davis Love III, a very prominent PGA Tour player, to any of Brown's 1999-2006 OWGR records, because his name was coded as "DAVIS LOVE", without the "III", in the 1999-2006 portion of Brown's OWGR data and "DAVIS LOVE III" (after converting both to upper case) in the ShotLink file.[39] Since we were unable to determine whether an unmatched ShotLink record was due to a player being ranked outside the OWGR top 200 or due to a difference in name spelling, we abandoned our attempt to replicate Brown's work using her monthly OWGR data.

Examining Brown's monthly OWGR data and STATA code, OWGR rankings associated with

[37] Per correspondence with Brown.

[38] We cannot tell from Brown's STATA code whether she changed the player names in the ShotLink scoring data to match those in her OWGR data. If she had not, she would have run into the same name matching problems we describe here.

[39] In the 1999-2006 portion of her OWGR data, Brown codes the names of each of the following players two different ways: Frank Lickliter II, Gonzalo Fernández-Castaño, Choi Kyung-Ju, Jeev Milkha Singh, and Rodney Pampling. This would have resulted in her not properly matching these players to her ShotLink scoring data in her 1999-2006 data set, unless she used two different name spellings for each of these players in her scoring data to compensate.

players who participated in the first tournament ending in a given month are based on rankings that were published after the event had been completed, which, in turn, would reflect the outcomes of the tournament itself. By contrast, all of our weekly OWGR rankings are associated with players who participated in tournaments the week after the OWGRs were published.

F. Field Quality

Brown describes her field quality variable as follows: "The competitiveness of the field of players is proxied by the average OWGR rank points of the participants (excluding Woods). For each player, I calculate this average excluding his own contribution to the strength of the field" (pp. 997-998). We interpret this definition as an average of OWGR points rather than ranking positions. This interpretation is reinforced by Brown's statement, later in the paper (footnote 20, page 1000), "The controls for the quality of the field are negative for the regressions in Table 2, suggesting that stronger fields may lead to lower scores." (If Brown had averaged OWGR ranking positions, rather than points, one would expect the relationship between field quality and scoring to be positive.) Our interpretation is reinforced further in her definition of field quality in an earlier version of her paper (Brown (2008), p. 18), "Note [that] I include an index of the quality of the field, the average OWGR points of the participants, in all reported specifications."

With this definition in mind, we base our field quality variable on average OWGR points. Our OWGR data include player rankings and points for positions 1-200 only from 1999 through September 12, 2004, but starting on September 19, 2004 the data include rankings and points for positions 1-999. Typically, OWGR points for position 999 are 0.04 to 0.05 and approximately 0.8 for position 200. On September 19, 2004, the average OWGR points for players in positions 201-500 and positions 201-999, respectively were 0.4 and 0.2. Under the assumption that most players in our data set without OWGR rankings prior to September 19, 2004 would have been ranked between positions 201 and 500, we assign OWGR points equal to 0.4 for all such unranked players. Starting September 19, 2004, we assign OWGR points equal to zero for players in positions 1000 and greater. The OWGR data that Brown provided us includes rankings and points for positions 1-200 only.

G. Tournament Purses

In correspondence, Brown indicated "I was missing purse data for several events." We obtained purse data for all PGA Tour events by summing individual prize winnings for each tournament from the ShotLink event files.

H. TV Viewership

In her paper, and as further clarified in correspondence, Brown indicates that she calculates TV viewership from proprietary Nielsen data as the sum of viewing households per minute for the first round of each event and tournament-level TV viewership as the sum of viewing households per minute over the four rounds of each event. Upon further inquiry, Brown indicated that although she had intended to calculate tournament-level TV viewership as the sum of viewing households per minute over the four rounds of each event, she actually used first-round TV viewership data when she analyzed tournament-level scoring. In correspondence, Brown states "I can report that the results remain very similar both in terms of magnitudes and statistical significance. ... In summary, this inadvertent oversight in the code has no impact on the findings in my paper."

Notwithstanding the oversight described above, we do not believe that Brown's TV viewership calculation is appropriate. Suppose, that event A is broadcast for four hours and has an average of 3 million viewing households per minute. Assume also that event B is broadcast sequentially for two hours each on two networks (for example, Golf Channel followed by CBS) with an average of 3 million viewing households per minute on each network. Clearly, the viewership for the two events is the same, 3 million viewers per minute for four hours. However, Brown would compute the viewership for B as twice that of A. Following Napoli (1997), we believe that the appropriate way to estimate TV viewership is to weight each viewership amount by the length of the broadcast and then sum over all first-round broadcasts for the analysis of first-round scores and over all four rounds for the analysis of event-level scores.

Like Brown, we also purchased proprietary TV viewership data from Nielsen. We believe that the data Nielsen initially sent to us and that which they sent to Brown are the same, since our Nielsen sales rep told us that the data we had originally received was the same that they had sent to Brown. Unfortunately, we found the data to be essentially unusable with many missing broadcasts.

In her STATA code, Brown uses an input file named RAW_NielsenData.dta. After the data are read into STATA, there is no further modification of the data in her code except the setting of missing values to zero. If she is using the same data Nielsen initially sent to us, some broadcast days are contaminated with irrelevant broadcasts of golf-related events that have nothing to do with PGA Tour competition, for example Senior (Champions) Tour events broadcast on the same day. In correspondence with Brown, we stated "Our file contains numerous records that clearly are not associated with a given PGA Tour broadcast. For example, our data has records for April 3 and April 4 1999 broadcasts of The Bell South Classic on NBC. Our data also includes records on the same dates for The Tradition, a Senior Tour event broadcast on ABC. If you were using the same raw data that we have, it would appear from your code that you would have included viewership for both events for April 3 and April 4. Can we assume that you included only PGA Tour broadcast data in your RAW_NielsenData.dta file and, as such, it would not have included data for Senior Tour and other non-PGA Tour broadcasts such as highlight shows?" Unfortunately, our correspondence with Brown did not resolve this issue.

Inasmuch as we found the TV viewership data that Nielsen originally sent to us to be unusable, except in our attempt to replicate Brown's work, we requested and received a much more comprehensive and complete set of viewership data from Nielsen. We use these data and the Napoli-based viewership calculation described above in our data set that we believe corrects the errors in Brown's data.

Although there are few events represented in both our original and updated data sets, such as the Bob Hope Chrysler Classic, that are not included in our actual analyses, our original and updated Nielsen data sets include 292 and 437 non-zero first-round event observations, respectively, for the 1999-2006 period and 331 and 493 non-zero event-level observations. As noted above, Brown inadvertently used first-round viewership records in her analysis of tournament-level scoring. Thus, ignoring the few events such as the Bob Hope, and assuming her TV data is the same as our original Nielsen data, we estimate that Brown would have used approximately 292 non-zero viewership observations when analyzing event-level scoring compared to our 493. We acknowledge that even our updated data set appears to have a few missing records. However, when we asked our Nielsen representative about these missing records, we were told "the data was run based on the official schedules for the PGA tour, so if [an event] is missing from the latest data, I would assume it did not air for one reason or another."

I. Net Total Score Calculation

Each year several PGA Tour events are played on more than one course. In a 2-course event, players rotate on courses 1 and 2 for the first two rounds, after which a cut is made. Then, the final two rounds are played on course 1, which we call the "primary course." In a 3-course event, players rotate on courses 1, 2 and 3 for the first three rounds, after which a cut is made. Then, a final round is played on course 1.

There is no requirement that each course in a multiple-course event have the same par. Per correspondence, Brown computes total score net of par per four-round event as the sum of actual scores minus 4 times the par of the primary course, rather than the sum of each of the four net scores. The two methods produce the same total net score for single-course events, but Brown's method will likely produce the incorrect total net score in a multiple-course event if the par for each course is not the same.[40] In our version of Brown's 256-event total-score data set, the difference between the correct net total score and Brown's net total score is 1 in 240 of 18,932 observations (140 of 12,255 effective observations, where player-course interactions observed only once are omitted).

J. Data Sets

In each subsection below, item 1 describes the data we use when we attempt to replicate Brown's work, and item 2 describes corrected data, and in some cases, more accurate data (weather, OWGR and TV viewership data), which we use when we attempt to correct Brown's results.

J.1. Event Selection

1. We include the same events that Brown has told us in correspondence that she actually included.

2. We include the events that meet Brown's selection criteria as stated in her paper and in correspondence.

J.2. Weather Data

1. We use the weather data that Brown sent to us and, consistent with Brown's use of the data as indicated in correspondence, use event-level weather data for the analysis of both event-level and first-round scores.

2. We use the weather data we collected ourselves. We employ the same pre-event rainfall totals for the analysis of both first-round and event-level scores. We employ first-round temperature (and corresponding dummy variables) and wind speed data for the analysis of first-round scores and average temperature (and corresponding dummy variables) and wind speed over all four rounds for the analysis of event-level scores.

J.3. Official World Golf Rankings

1. Using our own OWGR data, we update rankings and points monthly, and as in Brown, incorrectly associate the first set of OWGR data in each month with the first PGA Tour event that was completed in that month.

[40] Brown's method would compute the total net score correctly in a 3-course event if, for example, par for the primary course were 71, and par for the other two courses were 70 and 72.

2. Using our own OWGR data, we update rankings and points weekly and associate the rankings established at the end of each week t with the tournament completed in week $t+1$.

J.4. Field Quality

1. We base our field quality on the average OWGR points of the field, excluding Tiger Woods (or the player who takes on the role of the superstar) and the player associated with each respective scoring observation. For pre-September 19, 2004 data, we assign 0.4 OWGR points to each unranked player (players ranked worse than 200). Starting on September 19, 2004, we assign zero OWGR points to each unranked player (players ranked worse than 1,000).

2. We compute field quality the same as described above.

J.5. Tournament Purses

1. Since we do not have access to Brown's purse data, and it is not clear from correspondence which events were missing such data (see Section G above), we compute tournament purses by summing individual prize winnings for each tournament from the ShotLink event files.

2. We compute field quality the same as described above.

J.6. TV Viewership

1. We use the original TV viewership data sent to us by Nielsen, which we believe is the same as that used by Brown. We exclude all broadcasts that do not appear to be associated with an actual PGA Tour event. For each day, we compute TV viewership as the sum of household viewership rating values (in 1,000s) over all relevant broadcasts for the day.

2. We use the expanded TV viewership data sent to us by Nielsen. We exclude all broadcasts that do not appear to be associated with an actual PGA Tour event. For each day, we compute TV viewership as the sum of household viewership rating values (in millions) multiplied by broadcast duration (in minutes) over all relevant broadcasts for the day.

J.7. Net Scores

1. In single-course events, we subtract par from each gross score. In multiple-course events, per correspondence with Brown, we subtract the par of the primary course from each gross score.

2. Per discussion in Section 3.3, we do not use net scores as the explanatory variable but, instead, use actual 18-hole scores.

Online Appendix B: Weather Data Collection Methodology

Given the problems we discovered with Brown's weather data, as described in Online Appendix A, we went ahead and collected weather-related data ourselves. We began with a process, similar to that described by Brown, where we attempted to match each PGA Tour event site to the closest NOAA site, while trying to maintain geographic similarities. However, we found this process to be very subjective and, therefore, abandoned it in favor of a more scientific and replicable approach.

In correspondence, prior to providing us with her weather-related data, Brown indicated that we could collect weather data ourselves from the NOAA website:
http://www7.ncdc.noaa.gov/CDO/cdoselect.cmd?datasetabbv=GSOD&countryabbv=&georegionabbv=. The data set provided at the website is known as the Global Summary of the Day (GSOD). That is where we started in our initial efforts. However, we were told by user support people at NOAA to use the Global Historical Climate Network-Daily (GHCN-D) data, accessible at http://www.ncdc.noaa.gov/cdo-web/#t=secondTabLink, since this data set tends to be more accurate.[41] As it turns out, we were able to obtain the required data items for domestic weather stations only from the GHCN-D data. Therefore, we used GHCN-D data for domestic weather stations and GSOD data for stations outside the U.S.

We collected complete 1999-2010 state-wide GHCN-D data for all states where PGA Tour events were held in the U.S. along with complete state-wide data for several adjacent states. Using a process similar to that described by Brown, we attempted to find the closest individual GSOD weather stations to each international event and, where possible, also included all GSOD weather stations shown in Brown's weather file, under the assumption that her stations might be closer. The GHCN-D data includes GPS latitude and longitude coordinates for each weather station, but the GSOD data does not. Using Google Maps, we found GPS coordinates for each GSOD international weather station from which we collected data, including those of Brown. We also found GPS coordinates for each golf course. When an event was conducted on more than one course, we used the mean of the coordinates as a single location for the event. Finally, we assembled all statewide GHCN-D and international GSOD data into one file.

For each daily weather-related data item (total rainfall, maximum and minimum temperature (GHCN-D), average daily temperature (GSOD) and average wind speed) we recorded the item from the closest weather station with available data, regardless of source.[42] (As such, it is possible that weather data associated with a Canadian golf course could have been collected from a U.S. weather station, if the U.S. station happened to be the closest.) In some cases, individual data items for a given event were collected from different weather stations on the same day, and in a few cases, the closest weather station was in a state different from that of the actual event. Moreover, it is possible that data items of the same type for the same event were collected from different stations on different days.

The following table provides summary statistics for weather items used in our (CR) data set and that of Brown for the 256 events included in Brown's total-score data. As is evident from the table, there is a very high correlation between our average daily temperature and that of Brown. This is not surprising, since one would expect temperature to be relatively uniform across weather stations within the same general vicinity. Average wind speed is less correlated. According to a

[41] In some cases, we found that weather measures reported from the same location were identical, or nearly identical, in the two data sets but off by one day. For example, weather reported in the GSOD data set on a Wednesday might be the same as that reported in the GHCN-D data set on Tuesday.

[42] More precisely, we computed the chord distance between each golf course and weather station and took the data for a given item from the station with the shortest chord distance.

NOAA representative, wind speed in the U.S. is always collected at airports, with two exceptions, Central Park and The University of Southern California. If the airport from which one of our wind speed items was collected were different from that of Brown, it would not be surprising if wind speeds were different, since airports tend not to be close, and wind speed is not as uniform as temperature. Also, as noted earlier, GHCN-D and GSOD wind speed data collected from the same airport could be the same, but differ by one day in the way they are reported. Initially we were surprised by the very low correlation between our pre-event rainfall measure and that of Brown. This led us to examine both data sets more closely, and in the process, we discovered that for 60 of 127 events, Brown's pre-event rainfall appeared to be computed as the total rainfall over the four days of each tournament rather than the total rainfall over the three days prior to the event. (See Online Appendix A for more detail.)

Table B1: Weather Summary Using 256 Events in Brown's Total Score Data Set

Weather Item	Min	1st Qu.	Median	Mean	3rd Qu.	Max
CR Rainfall (in.)	0.00	0.00	0.01	0.27	0.26	4.41
Brown Rainfall (in.)	0.00	0.00	0.11	0.40	0.49	3.55
CR Mean Temp (F)	47.26	62.21	69.24	68.07	74.38	87.01
Brown Mean Temp (F)	48.62	63.09	69.38	68.45	74.43	87.12
CR Mean Wind (mph)	1.73	5.52	6.91	7.32	8.85	14.71
Brown Mean Wind (mph)	1.18	4.45	5.77	6.10	7.83	16.50
CR Rain Distance (mi.)*	0.27	2.99	5.53	6.44	8.69	53.77
CR Temp Distance (mi.)*	0.83	2.99	5.75	6.82	8.91	53.77
CR Wind Distance (mi.)*	1.09	5.67	8.91	9.72	11.87	53.77

CR vs. Brown	Correlation
Pre-Event Rainfall	0.167
Mean Temperature	0.929
Mean Wind Speed	0.685

Note.–*Computed over seven days starting three days prior to each event and going through the fourth day of each event. (Only pre-event total rainfall and round 1-4 mean temperature and wind speed are used in our regressions and reflected in the correlations shown in the table.) CR denotes Connolly-Rendleman.

There is insufficient detail within Brown's weather data to determine mean distances from tournament sites to weather stations for each data item. Therefore, we only report mean distances for our own data items. Generally, rainfall and temperature are collected from weather stations within three to nine miles of each tournament site, with wind speed collection stations being a little more distant, reflecting that, with two exceptions, wind speed in the U.S. is collected at airports only.

Online Appendix C: Effect of Woods Being 'Hot' or 'Cool'

When estimating the hot/cool effect, we follow the same procedure as Brown for identifying Woods' hot, cool and typical periods. We define scoring quintiles relative to the observations of Woods' scoring in the preceding month. However, there are some observations for which there are no prior-month scoring records for Woods. In these cases, we record the observations as "typical." As in Brown, we interact *hot*, *cool* and *typical* dummies with $tifH$, $tifL$, and $tifU$.[43] Tables C1 and C2 summarize our results.

We note that for this set of tests we employ the same programming code and data that we used in connection with our previously reported results, except we interact *hot*, *cool* and *typical* dummies with our tif-related dummies. Perhaps the large differences in our results and those of Brown reflect the way we code *hot*, *cool*, and *typical* dummy variables when we have no prior-month scoring records, but without full access to Brown's data and code, we have no way of knowing.

[43]We are unable to determine how Brown dealt with observations for which there were no prior-month scoring records for Woods.

Table C1: Analysis of Browns Table 5, First Round Hot/Cool Analysis

Panel A: Results as Reported by Brown

Coefficient	Est.	Player-Yr Cluster SE	Player-Yr Cluster p-Value	R-C Cluster SE	R-C Cluster p-Value
$tifH \times typical$	0.358	0.302	0.236	n/r	n/r
$tifH \times hot$	0.982	0.321	0.002	n/r	n/r
$tifH \times cool$	0.279	0.439	0.525	n/r	n/r
$tifL \times typical$	-0.088	0.127	0.488	n/r	n/r
$tifL \times hot$	0.611	0.139	0.000	n/r	n/r
$tifL \times cool$	-0.220	0.171	0.198	n/r	n/r
$tifU \times typical$	0.129	0.152	0.396	n/r	n/r
$tifU \times hot$	0.523	0.149	0.000	n/r	n/r
$tifU \times cool$	-0.294	0.216	0.173	n/r	n/r

Panel B: Data Set 1, Interacted Player-Course Effects

Coefficient	Est.	Player-Yr Cluster SE	Player-Yr Cluster p-Value	R-C Cluster SE	R-C Cluster p-Value
$tifH \times typical$	0.801	0.249	0.001	0.328	0.015
$tifH \times hot$	0.503	0.340	0.139	0.491	0.306
$tifH \times cool$	0.608	0.336	0.071	0.489	0.214
$tifL \times typical$	0.169	0.101	0.095	0.242	0.487
$tifL \times hot$	0.236	0.138	0.087	0.225	0.294
$tifL \times cool$	0.338	0.151	0.025	0.362	0.350
$tifU \times typical$	0.141	0.126	0.261	0.234	0.545
$tifU \times hot$	0.259	0.186	0.164	0.260	0.319
$tifU \times cool$	0.692	0.196	0.000	0.309	0.025

Panel C: Data Set 2, Interacted Player-Course Effects

Coefficient	Est.	Player-Yr Cluster SE	Player-Yr Cluster p-Value	R-C Cluster SE	R-C Cluster p-Value
$tifH \times typical$	0.811	0.227	0.000	0.297	0.006
$tifH \times hot$	-0.158	0.310	0.611	0.404	0.696
$tifH \times cool$	0.031	0.308	0.919	0.414	0.939
$tifL \times typical$	0.259	0.096	0.007	0.227	0.255
$tifL \times hot$	-0.082	0.127	0.518	0.310	0.792
$tifL \times cool$	-0.277	0.130	0.033	0.299	0.354
$tifU \times typical$	0.299	0.114	0.009	0.213	0.161
$tifU \times hot$	0.122	0.170	0.473	0.248	0.623
$tifU \times cool$	0.115	0.170	0.499	0.289	0.691

Panel D: Data Set 2, Separate Player and Course Effects

Coefficient	Est.	Player-Yr Cluster SE	Player-Yr Cluster p-Value	R-C Cluster SE	R-C Cluster p-Value
$tifH \times typical$	0.385	0.140	0.006	0.234	0.100
$tifH \times hot$	-0.192	0.219	0.381	0.283	0.496
$tifH \times cool$	-0.086	0.208	0.680	0.305	0.779
$tifL \times typical$	0.279	0.073	0.000	0.196	0.154
$tifL \times hot$	-0.202	0.100	0.043	0.249	0.418
$tifL \times cool$	-0.299	0.101	0.003	0.263	0.255
$tifU \times typical$	0.254	0.076	0.001	0.195	0.192
$tifU \times hot$	-0.081	0.115	0.482	0.250	0.746
$tifU \times cool$	0.055	0.115	0.631	0.260	0.832

"R-C Cluster" denotes robust standard errors clustered by round-course. p-values in Panel A are based on coefficient estimates and standard errors reported by Brown. As described in Section J of Online Appendix A, data set 1 closely approximates that used by Brown, and data set 2 corrects errors and includes all events that meet Brown's selection criteria as stated in her paper and implied in correspondence. n/r denotes not reported.

Table C2: Analysis of Brown's Table 5, Event-Level Hot/Cool Analysis

Panel A: Results as Reported by Brown

		Player-Yr Cluster		R-C Cluster	
Coefficient	Coef	SE	p-Value	SE	p-Value
$tifH \times typical$	1.016	0.833	0.223	n/r	n/r
$tifH \times hot$	2.074	0.775	0.007	n/r	n/r
$tifH \times cool$	0.500	1.049	0.634	n/r	n/r
$tifL \times typical$	0.669	0.365	0.067	n/r	n/r
$tifL \times hot$	1.377	0.369	0.000	n/r	n/r
$tifL \times cool$	-0.332	0.486	0.495	n/r	n/r
$tifU \times typical$	0.433	0.460	0.347	n/r	n/r
$tifU \times hot$	1.387	0.504	0.006	n/r	n/r
$tifU \times cool$	-1.151	0.663	0.083	n/r	n/r

Panel B: Data Set 1, Interacted Player-Course Effects

		Player-Yr Cluster		R-C Cluster	
Coefficient	Coef	SE	p-Value	SE	p-Value
$tifH \times typical$	0.786	0.590	0.183	1.059	0.458
$tifH \times hot$	0.849	0.962	0.378	1.348	0.529
$tifH \times cool$	1.983	0.740	0.007	1.069	0.064
$tifL \times typical$	0.784	0.270	0.004	0.830	0.345
$tifL \times hot$	-0.046	0.342	0.892	0.889	0.958
$tifL \times cool$	-0.463	0.382	0.226	0.831	0.578
$tifU \times typical$	0.622	0.377	0.099	0.926	0.502
$tifU \times hot$	-0.770	0.491	0.117	0.922	0.404
$tifU \times cool$	-0.249	0.640	0.697	0.996	0.802

Panel C: Data Set 2, Interacted Player-Course Effects

		Player-Yr Cluster		R-C Cluster	
Coefficient	Coef	SE	p-Value	SE	p-Value
$tifH \times typical$	0.744	0.512	0.146	0.846	0.379
$tifH \times hot$	-0.008	0.758	0.991	1.095	0.994
$tifH \times cool$	1.229	0.662	0.063	0.893	0.168
$tifL \times typical$	0.306	0.249	0.220	0.645	0.636
$tifL \times hot$	-0.385	0.312	0.217	0.702	0.584
$tifL \times cool$	-0.541	0.362	0.135	0.608	0.373
$tifU \times typical$	-0.092	0.333	0.782	0.698	0.895
$tifU \times hot$	-0.376	0.444	0.398	0.805	0.641
$tifU \times cool$	0.109	0.609	0.858	0.866	0.900

Panel D: Data Set 2, Separate Player and Course Effects

		Player-Yr Cluster		R-C Cluster	
Coefficient	Coef	SE	p-Value	SE	p-Value
$tifH \times typical$	-0.036	0.316	0.909	0.630	0.954
$tifH \times hot$	0.019	0.465	0.967	0.879	0.982
$tifH \times cool$	0.241	0.417	0.563	0.651	0.711
$tifL \times typical$	0.294	0.182	0.106	0.606	0.627
$tifL \times hot$	-0.331	0.231	0.153	0.671	0.622
$tifL \times cool$	-0.258	0.264	0.329	0.594	0.664
$tifU \times typical$	0.238	0.210	0.257	0.645	0.712
$tifU \times hot$	-0.211	0.305	0.488	0.713	0.767
$tifU \times cool$	-0.021	0.371	0.955	0.800	0.979

"R-C Cluster" denotes robust standard errors based on round-course assignment in first round. p-values in Panel A are based on coefficient estimates and standard errors reported by Brown. As described in Section J of Online Appendix A, data set 1 closely matches that used by Brown, and data set 2 corrects errors and includes all events that meet Brown's selection criteria as stated in her paper and implied in correspondence. n/r denotes not reported.

Online Appendix D
Data and Regression Specification Problems in Connection with Brown's Surprise Absence Analysis

A. Events Included in Brown's Table 3 Analysis

We were able to infer the events included in Brown's Table 3 analysis from results shown in the table.[44] Although not mentioned or implied, Brown leaves out the two majors that were conducted during each of the three years of the July-September 2007-9 "surgery period." In the two majors that Woods missed in 2008, participating players performed an average of 4.06 strokes worse net of par than in 2007 and 2009 when Woods did compete.[45] During the same period, she includes the 2007 Buick Open as an event that Woods missed due to knee surgery, but the surgery actually occurred a year later.

Similarly, Brown omits the WGC-CA Championship, conducted in March, an event in which Woods competed in 2009 but missed in 2010 due to personal issues. But on page 1001, she mentions the WGC-CA, along with the Arnold Palmer Invitational, as examples of events that <u>should</u> be included in the analysis. As it turns out, the January-March 2009-2010 personal issues analysis, summarized in Brown's Table 3, includes only the 2009 (with Woods) and 2010 (without Woods) versions of the Arnold Palmer event, hardly sufficient to draw inferences about Woods' surprise absence from Tour during the January-March 2010 period.[46]

Our "corrected" version of Brown's data, which we believe conforms with her event selection criteria, includes the following events: the 2007 and 2009 Bridgestone Invitational, the 2007 and 2009 AT&T National, the 2007 and 2009 British Open, and the 2007 and 2009 PGA Championship, in which Woods participated, and the 2008 Bridgestone, 2008 AT&T National, 2008 Brisish Open and 2008 PGA Championship, in which Woods did not participate. Events we include in Wood's personal issues period are the 2009 Arnold Palmer Invitational and 2009 WGC-CA, in which Woods participated, and the 2010 Arnold Palmer and 2010 WGC-CA, in which Woods did not participate.

B. Events Included in Brown's Table 4 Analysis

Brown states (implicitly) that her Table 4 analysis excludes majors but provides no explanation for these exclusions. (If Brown had controlled for differences in relative course difficulty in her regressions, as indicated in the paper, there should have been no reason to have excluded majors.) Otherwise, the events included do not satisfy her inclusion criteria implied in her paper and in correspondence. For example, Brown indicates that the personal issues analysis covers events in January-March 2009 and 2010. But she actually includes all non-major events in April, but only for 2009, and during the stated January-March period, she includes four events for just one of the two years.

[44] Events which Brown included in the analysis of Woods surgery period were the 2007 and 2009 Bridgestone Invitational and the 2007 and 2009 AT&T National, in which Woods participated, and the 2008 Bridgestone, 2008 AT&T National and 2007 Buick Open, in which Woods did not participate. Events included in Wood's personal issues period include the 2009 Arnold Palmer Invitational, in which Woods participated, and the 2010 Arnold Palmer, in which Woods did not participate.

[45] Much of this difference could be attributable to differences in difficulty among the six different courses on which the majors were conducted. This, in fact, may be the reason Brown omitted the two majors from her analysis. But it is not mentioned in the paper, and the issue as to why the two majors were excluded remains unresolved.

[46] Unfortunately, our correspondence with Brown concerning the WGC-CA omission did not resolve this issue.

Several events that meet Brown's inclusion criteria for the July-September 2007-9 surgery period are omitted altogether from her Table 4 analysis.[47] In the paper, Brown states that she specifically excludes FedExCup Playoffs events (p. 1003), because incentives for these events are different, but she actually includes the 2007 and 2009 versions of the BMW Championship, the third event in the Playoffs, but not the 2008 version of the same event. Within the surgery period, she includes the US Bank Championship, an "alternate" event held at the same time as the British Open, but excludes the Reno-Tahoe Open, an alternate event held opposite the WGC-Bridgestone Invitational during the same July-September 2007-9 period. Although she does not specifically state that alternate events should be excluded in her Table 4 analysis, she does specifically exclude alternate events in her main analysis (her Tables 2 and 5), and she excludes two alternate events that would otherwise meet her inclusion criteria in the January-March 2009-2010 personal issues portion of her Table 4 analysis.

Given the many differences between the events Brown actually included in her Table 4 analysis and the events that meet her selection criteria, we have put together Tables D1 and D2, which summarize the differences. These tables include all stroke-play events conducted during the July-September 2007-9 surgery and January-March 2009-10 personal issues periods. They also include all stroke-play events conducted in April 2009 and 2010, since Brown included all non-major events during April 2009, despite the fact that they fell outside her stated "personal issues" time frame. In each table, we provide an indication of whether each event was included in Brown's first-round data set (Brown 1) and her event-level data set (Brown 4). We also provide an indication for each event of whether it should have been included (CR 1 for first-round scores and CR 4 for event-level scores, respectively) based on Brown's selection criteria. For each event, we provide an indication of whether the event is among those that Woods missed by surprise due to knee surgery or personal issues. Finally, for each event that is excluded from one or more of the data sets, we provide an explanation or comment on its exclusion. All events for which Brown's inclusion or omission is inconsistent with her event selection criteria are marked with an asterisk.

C. Surprise Absence Regression Specification (Table 4)

According to the footnote in Table 4 of her paper, Brown estimates regression equation (D-1) below for Woods' knee surgery period (July-September 2007-9) and separately for his personal issues period (January-March 2009-10).

$$net_{i,j} = \beta_1 tifH_{i,j} + \beta_2 tifL_{i,j} + \beta_3 tifU_{i,j} + \gamma_0 + \gamma_1 P_i + \gamma_2 C_j + \gamma_3 Z_j \tag{D-1}$$

In (D-1), P_i is a matrix of dummy variables that identifies players, C_j is a dummy variable matrix that identifies courses, and Z_j is a vector of event-specific controls for pre-event rainfall, wind, temperature, inflation-adjusted purse and the square of inflation-adjusted purse. (All other variables are as defined earlier.) In contrast to Brown's regression specification (6), which underlies her Table 2 and Table 5 analyses, fixed player and course effects are estimated separately, and a smaller set of event-specific controls is employed.

[47] The Turning Point Resort Championship, conducted September 20-23, 2007 and the Viking Classic, conducted September 27-30, 2007 and September 18-21, 2008 appear to meet Brown's criteria for inclusion but are left out. We note that the Viking Classic was held opposite the President Cup in 2007 and the Ryder Cup in 2008. However, the Valero Texas Open, held opposite these two events in 2004-2006, is included in Brown's main 1999-2006 data set. Therefore, we see no reason why the Viking Classic should be excluded.

By contrast, Brown's STATA code indicates that the following regression was actually run:[48]

$$
\begin{aligned}
net_{i,j} = {} & \beta_1 tifH_{i,j} + \beta_2 tifL_{i,j} + \beta_3 tifU_{i,j} \\
& + \alpha_1 HRanked_{i,j} + \alpha_2 LRanked_{i,j} + \gamma_0 + \gamma_1 P_i + \gamma_3 Z_j \\
& + \lambda_1 Q_j + \lambda_2 TV_j + \lambda_3 major_j + \epsilon_{i,j}
\end{aligned} \tag{D-2}
$$

In (D-2), Q_j is field quality, TV_j is TV viewership, and $major_j$ is a dummy variable that indicates whether the tournament is a "major."[49] Despite including more explanatory variables (Q_j, TV_j and $major_j$), the most significant difference between (D-1) and (D-2) is that (D-1) reflects the estimation of both fixed player and course effects, while (D-2) controls for fixed player effects only. In this application, a regression run without controls for relative course difficulty would be largely uninformative, since the scoring average of one golfer can be lower (i.e., better) than another, not because of superior skill, but because of playing on easier courses. It is well-known that Woods and other elite players tend to play in tournaments conducted on the most challenging courses.

In correspondence, Brown indicates "It appears that I had neglected to include the course fixed effects in the specification for column (2) [event-level regressions during Woods' surgery period]; however, the estimates are still positive and statistically significant when I include those fixed effects." In a subsequent followup message, we asked Brown "So just to be perfectly clear, are you saying that the regressions for columns 1, 3 and 4 in Table 4 did include fixed course effects and that the column 2 regression is the only regression that did not include fixed course effects?" As of this writing, our correspondence with Brown has not resolved this issue; therefore, we do not know which regression form she actually used.

D. Brown's Table 4 Results

Table D3 summarizes our analysis of the results reported in Brown's Table 4. In general, Brown interprets positive and statistically significant coefficient estimates of $tifH$, $tifL$, and $tifU$, as being consistent with her superstar hypothesis; in essence, scores tend to be higher (worse) when Woods plays.

Panel A presents Brown's results as stated in her paper. Panel B shows our own results when we attempt to replicate those of Brown, using the events she actually used, explanatory variables from data set 1, which closely match those of Brown, and the regression form shown in Brown's STATA code (regression form (D-2)), which omits the estimation of fixed course effects. In all panels, we compute robust standard errors clustered by player-year. (To conserve space, we do not report results for robust standard errors clustered by round-course, but as it turns out, it does not matter.) Note that in all panel-B regressions except the last, our observation counts are the same as Brown's.[50] Also, coefficient estimates and p-values are generally of the same order of magnitude that Brown reports, and the "story" that might arise from panel B is essentially the same as that

[48] As discussed in Online Appendix A, Section B, Brown's STATA code indicates that inflation-adjusted purse and the square of inflation-adjusted purse were interacted with *Hranked*, *Lranked*, and *Uranked*. However, we do not employ these interactions in the regressions that we run.

[49] Including the "major" dummy is redundant, since majors are explicitly excluded from the regression.

[50] The 5-player difference reflects that Brown's data include 78 observations for the 2009 Quail Hollow Championship, whereas our data include 73. Inasmuch as five players withdrew from competition after the start of play, we suspect that Brown may have included their scores. If included, the total scores prior to withdrawal of 35, 38, 42, 78 and 111 (source, Yahoo Sports) for the five players would have the potential to distort Brown's total-score regression results. The total four-round winning score for the Quail Hollow event was 277. Thus, these scores, if used by Brown, are much lower than typical four-round scores.

associated with panel A. Inasmuch as the panel B regressions omit controls for relative course difficulty, we believe there is nothing informative about the panel B results other than the fact that they correspond closely to Brown's results as shown in panel A and, as such, suggest that Brown may have omitted the estimation of fixed course effects in her regressions.

In correspondence, Brown seems to imply that she inadvertently omitted controls for relative course difficulty in just a single regression, her event-level regression for Woods' surgery period, not all regressions. In any event, the regressions summarized in panel C are identical to our panel B regressions, except we do include the estimation of fixed course effects. Therefore, if Brown did include the estimation of fixed course effects in her other three regressions, we would expect our panel C coefficient estimates for these three regressions to be even closer to Brown's results than the results we report in panel B. A comparison of coefficient estimates in connection with the first three regressions in panels B and C versus those in panel A is inconclusive.

In each of the panel-D and E regressions, we include events consistent with Brown's selection criteria, as shown in the CR1 and CR4 columns of our Tables D1 and D2. In panel D, we summarize the same regressions as in panel C but use explanatory variables corrected for data errors (data set 2, as described in Online Appendix A) and events consistent with Brown's selection criteria. Here, no coefficient estimates are statistically significant except those associated with event-level scores during Woods' personal issues period. But these coefficient estimates are huge, in absolute magnitude, approximately -17 to -16 strokes per four-round event, and of the wrong sign to be consistent with Brown's hypothesis.

The regressions summarized in panel E are identical to panel D regressions, except we use regression (D-1), the regression form indicated in Brown's paper. We view the panel E results as corresponding closely to the results Brown would have reported if she had used corrected data, had run the regressions shown in the paper and had included the set of events consistent with her event selection criteria. Note that these results are inconsistent with Brown's hypothesis; six of seven coefficient estimates that are statistically significant at the 0.05 level are of the wrong sign.

Notwithstanding our foregoing analysis, none of the regressions summarized in panels B-E nor the regressions Brown summarizes in her paper, enable one to estimate the incremental effect of Woods' surprise absences. In these regressions, $tif_j = 1$ when Woods is in the field, and $tif_j = 0$ when he is not. As such, one cannot distinguish the effect of Woods being absent by surprise from the effect of Woods being absent by choice. We address this issue in Section 4.2.2 of the main text.

Table D1: Summary of All Stroke-Play Events During Brown's Table 4 July-September 2007-9 Surgery Period

Event		Start	End	Was Event Included? Brown 1	Brown 4	CR 1	CR 4	Surprise		Comments
2007	Buick Open	28-Jun	1-Jul	No	Yes	No	Yes	No	*	Woods missed, but not due to 2008 knee surgery. CR include total score, but not first round, to be consistent with Brown. Otherwise, inclusion questionable.
2007	AT&T National	5-Jul	8-Jul	Yes	Yes	Yes	Yes	No		
2007	John Deere	12-Jul	15-Jul	Yes	Yes	Yes	Yes	No		
2007	U.S. Bank	19-Jul	22-Jul	Yes	Yes	No	No	No	*	Alternate event, Brown should not have included
2007	British Open	19-Jul	22-Jul	No	No	No	No	No		Major championship
2007	Canadian Open	26-Jul	29-Jul	Yes	Yes	Yes	Yes	No		
2007	Reno-Tahoe	2-Aug	5-Aug	No	No	No	No	No		Alternate event
2007	WGC-Bridgestone	2-Aug	5-Aug	Yes	Yes	Yes	Yes	No		
2007	PGA Champ	9-Aug	12-Aug	No	No	No	No	No		Major championship
2007	Wyndham	16-Aug	19-Aug	Yes	Yes	Yes	Yes	No		
2007	The Barclays	23-Aug	26-Aug	No	No	No	No	No		FedExCup Playoffs event
2007	Deutsche Bank	31-Aug	3-Sep	No	No	No	No	No		FedExCup Playoffs event
2007	BMW	6-Sep	9-Sep	Yes	Yes	No	No	No	*	FedExCup Playoffs, Brown should not have included
2007	TOUR Championship	13-Sep	16-Sep	No	No	No	No	No		FedExCup Playoffs event
2007	Turning Stone	20-Sep	23-Sep	No	No	Yes	Yes	No	*	Brown excluded, meets inclusion criteria
2007	Viking Classic	27-Sep	30-Sep	No	No	Yes	Yes	No	*	Brown excluded, meets inclusion criteria
2008	AT&T National	3-Jul	6-Jul	Yes	Yes	Yes	Yes	Yes		
2008	John Deere	10-Jul	13-Jul	Yes	Yes	Yes	Yes	No		
2008	U.S. Bank	17-Jul	20-Jul	Yes	Yes	No	No	No	*	Alternate event, Brown should not have included
2008	British Open	17-Jul	20-Jul	No	No	No	No	Yes		Major championship
2008	Canadian Open	24-Jul	27-Jul	Yes	Yes	Yes	Yes	No		
2008	Reno-Tahoe	31-Jul	3-Aug	No	No	No	No	No		Alternate event
2008	WGC-Bridgestone	31-Jul	3-Aug	Yes	Yes	Yes	Yes	Yes		
2008	PGA	7-Aug	10-Aug	No	No	No	No	Yes		Major championship
2008	Wyndham	14-Aug	17-Aug	Yes	Yes	Yes	Yes	No		
2008	The Barclays	21-Aug	24-Aug	No	No	No	No	Yes		FedExCup Playoffs event
2008	Deutsche Bank	29-Aug	1-Sep	No	No	No	No	Yes		FedExCup Playoffs event
2008	BMW	4-Sep	7-Sep	No	No	No	No	Yes		FedExCup Playoffs event
2008	Viking Classic	18-Sep	21-Sep	No	No	Yes	Yes	No	*	Brown excluded, meets inclusion criteria
2008	TOUR Championship	25-Sep	28-Sep	No	No	No	No	Yes		FedExCup Playoffs event
2009	AT&T National	2-Jul	5-Jul	Yes	Yes	Yes	Yes	No		
2009	John Deere	9-Jul	12-Jul	Yes	Yes	Yes	Yes	No		
2009	U.S. Bank	16-Jul	19-Jul	Yes	Yes	No	No	No	*	Alternate event, Brown should not have included
2009	British Open	16-Jul	19-Jul	No	No	No	No	No		Major championship
2009	Canadian Open	23-Jul	26-Jul	Yes	Yes	Yes	Yes	No		
2009	Buick Open	30-Jul	2-Aug	Yes	Yes	Yes	Yes	No		
2009	Reno-Tahoe	6-Aug	9-Aug	No	No	No	No	No		Alternate event
2009	WGC-Bridgestone	6-Aug	9-Aug	Yes	Yes	Yes	Yes	No		
2009	PGA	13-Aug	16-Aug	No	No	No	No	No		Major championship
2009	Wyndham	20-Aug	23-Aug	Yes	Yes	Yes	Yes	No		
2009	The Barclays	27-Aug	30-Aug	No	No	No	No	No		FedExCup Playoffs event
2009	Deutsche Bank	4-Sep	7-Sep	No	No	No	No	No		FedExCup Playoffs event
2009	BMW	10-Sep	13-Sep	Yes	Yes	No	No	No	*	FedExCup Playoffs, Brown should not have included
2009	TOUR Championship	24-Sep	27-Sep	No	No	No	No	No		FedExCup Playoffs event

Note.–Includes all events conducted during July-September, 2007-9. Brown 1 = Brown's first-round events; Brown 4 = Brown's total-score events; CR 1 = Connolly-Rendleman first-round events; CR 4 = Connolly-Rendleman total-score events. Surprise = "Yes" denotes that Woods did not participate due to knee surgery but otherwise would have been expected to participate. * denotes an event choice by Brown inconsistent with her implied event selection criteria.

Table D2: Summary of All Stroke-Play Events During Brown's Table 4 January-March 2009-10 Personal Issues Period

				Was Event Included?						
Event		Start	End	Brown 1	Brown 4	CR 1	CR 4	Surprise		Comments
2009	Mercedes-Benz	8-Jan	11-Jan	No	No	No	No	No		Small-field event
2009	Sony Open	15-Jan	18-Jan	Yes	Yes	Yes	Yes	No		
2009	Bob Hope	21-Jan	25-Jan	No	No	No	No	No		5-round event
2009	FBR	29-Jan	1-Feb	Yes	Yes	Yes	Yes	No		
2009	Buick Inv	5-Feb	8-Feb	Yes	Yes	Yes	Yes	No		
2009	AT&T Pebble Beach	12-Feb	15-Feb	No	No	Yes	No	No	*	3-round event. Brown excluded first round but meets first-round inclusion criteria.
2009	Northern Trust Open	19-Feb	22-Feb	Yes	Yes	Yes	Yes	No		
2009	Mayakoba	26-Feb	1-Mar	No	No	No	No	No		Alternate event
2009	Honda Classic	5-Mar	8-Mar	Yes	Yes	Yes	Yes	No		
2009	WGC-CA	12-Mar	15-Mar	No	No	Yes	Yes	No	*	Brown excluded, meets inclusion criteria
2009	Puerto Rico	12-Mar	15-Mar	No	No	No	No	No		Alternate event
2009	Transitions	19-Mar	22-Mar	Yes	Yes	Yes	Yes	No		
2009	Arnold Palmer	26-Mar	29-Mar	Yes	Yes	Yes	Yes	No		
2009	Shell Houston Open	2-Apr	5-Apr	Yes	No	No	No	No	*	April event, Brown should not have included
2009	Masters	9-Apr	12-Apr	No	No	No	No	No		April event and major, should not be included
2009	Verizon Heritage	16-Apr	19-Apr	Yes	Yes	No	No	No	*	April event, Brown should not have included
2009	Zurich Classic	23-Apr	26-Apr	Yes	Yes	No	No	No	*	April event, Brown should not have included
2009	Quail Hollow	30-Apr	3-May	Yes	Yes	No	No	No	*	April/May event, Brown should not have included
2010	SBS (Mercedes)	7-Jan	10-Jan	No	No	No	No	No		Small-field event
2010	Sony Open	14-Jan	17-Jan	Yes	Yes	Yes	Yes	No		
2010	Bob Hope	20-Jan	24-Jan	No	No	No	No	No		5-round event
2010	Farmers Ins (Buick Inv)	28-Jan	31-Jan	Yes	Yes	Yes	Yes	No		
2010	Northern Trust	4-Feb	7-Feb	Yes	Yes	Yes	Yes	No		
2010	AT&T Pebble Beach	11-Feb	14-Feb	Yes	Yes	Yes	Yes	No		
2010	Mayakoba	18-Feb	21-Feb	No	No	No	No	No		Alternate event
2010	Waste Mgt (FBR)	25-Feb	28-Feb	No	No	Yes	Yes	No	*	Brown excluded, meets inclusion criteria
2010	Honda Classic	4-Mar	7-Mar	No	No	Yes	Yes	No	*	Brown excluded, meets inclusion criteria
2010	WGC-CA	11-Mar	14-Mar	Yes	Yes	Yes	Yes	Yes		
2010	Puerto Rico	11-Mar	14-Mar	No	No	No	No	No		Alternate event
2010	Transitions	18-Mar	21-Mar	No	No	Yes	Yes	No	*	Brown excluded, meets inclusion criteria
2010	Arnold Palmer	25-Mar	28-Mar	Yes	Yes	Yes	Yes	Yes		
2010	Shell Houston Open	1-Apr	4-Apr	No	No	No	No	No		April event, should not be included
2010	Masters	8-Apr	11-Apr	No	No	No	No	No		April event and major, should not be included
2010	Verizon Heritage	15-Apr	18-Apr	No	No	No	No	No		April event, should not be included
2010	Zurich Classic	22-Apr	25-Apr	No	No	No	No	No		April event, should not be included
2010	Quail Hollow	29-Apr	2-May	No	No	No	No	No		April and May event, should not be included

Note.–Includes all events conducted during January-March, 2009-10, and all conducted in April 2009-10. Brown 1 = Brown's first-round events; Brown 4 = Brown's total-score events; CR 1 = Connolly-Rendleman first-round events; CR 4 = Connolly-Rendleman total-score events. Surprise = "Yes" denotes that Woods did not participate due to personal issues but otherwise would have been expected to participate. * denotes an event choice by Brown inconsistent with her implied event selection criteria.

Table D3: Analysis of Brown's Table 4

Regression	Estimate	Panel A: As Brown Reports tifH	tifL	tifU	Obs.	Panel B: Using STATA Regressions and Data Set 1 tifH	tifL	tifU	Obs.
First Round Surgery	Coef	0.244	0.317	0.119	2,676	0.447	0.604	0.211	2,676
	Std Error	0.614	0.309	0.377		0.507	0.236	0.268	
	p-value	0.691	0.305	0.752		0.378	0.011	0.431	
Event-Level Surgery	Coef	3.919	5.647	5.755	1,616	2.966	4.409	6.630	1,616
	Std Error	1.462	0.956	1.094		1.135	0.530	0.775	
	p-value	0.007	0.000	0.000		0.009	0.000	0.000	
First Round Personal	Coef	1.125	1.510	1.210	2,339	0.765	0.931	0.514	2,339
	Std Error	1.125	0.447	0.484		0.882	0.307	0.354	
	p-value	0.317	0.001	0.012		0.386	0.002	0.146	
Event-Level Personal	Coef	2.754	3.519	3.554	1,172	1.688	3.240	2.494	1,167
	Std Error	2.038	1.050	1.351		1.392	0.797	1.183	
	p-value	0.177	0.001	0.009		0.226	0.000	0.035	

Regression	Estimate	Panel C: Using STATA Regressions + Course Effects and Data Set 1 tifH	tifL	tifU	Obs.	Panel D: Using STATA Regressions + Course Effects and Data Set 2 tifH	tifL	tifU	Obs.
First Round Surgery	Coef	0.245	0.214	-0.174	2,676	-0.401	0.401	0.142	2,565
	Std Error	0.625	0.366	0.396		1.231	1.048	1.068	
	p-value	0.695	0.559	0.661		0.745	0.702	0.895	
Event-Level Surgery	Coef	5.389	4.836	6.229	1,616	1.836	1.608	2.502	1,484
	Std Error	1.292	0.657	0.833		1.900	1.281	1.363	
	p-value	0.000	0.000	0.000		0.334	0.210	0.067	
First Round Personal	Coef	n/a	0.092	-0.191	2,339	-1.265	-1.132	-2.032	2,447
	Std Error	n/a	0.846	0.854		1.415	1.234	1.341	
	p-value	n/a	0.913	0.823		0.371	0.359	0.130	
Event-Level Personal	Coef	n/a	1.442	1.524	1,167	-17.465	-16.178	-16.156	1,245
	Std Error	n/a	1.167	1.396		2.957	2.758	3.225	
	p-value	n/a	0.217	0.275		0.000	0.000	0.000	

Regression	Estimate	Panel E: Using Regressions from Paper and Data Set 2 tifH	tifL	tifU	Obs.
First Round Surgery	Coef	-0.404	0.195	-0.211	2,565
	Std Error	0.748	0.593	0.597	
	p-value	0.589	0.743	0.724	
Event-Level Surgery	Coef	0.748	0.452	1.564	1,484
	Std Error	1.326	0.647	0.754	
	p-value	0.573	0.484	0.038	
First Round Personal	Coef	-3.630	-3.622	-4.302	2,447
	Std Error	0.803	0.610	0.845	
	p-value	0.000	0.000	0.000	
Event-Level Personal	Coef	-10.740	-9.412	-9.195	1,245
	Std Error	1.810	1.306	1.889	
	p-value	0.000	0.000	0.000	

Note.– Values in Panel A are as Brown reports in her Table 4, except for p-values, which we compute. n/a denotes that variable is collinear with other explanatory variables, and the estimated regression coefficient cannot be computed. Regressions in Panels B and C employ the set of events that Brown actually used. Those in Panels D and E use events consistent with Brown's selection criteria, shown in the CR1 and CR4 columns of Table D1. Data set 1 closely matches data employed by Brown. Data set 2 uses data that has been corrected for errors.