NBER WORKING PAPER SERIES

THE BENEFITS OF COLLEGE ATHLETIC SUCCESS:
AN APPLICATION OF THE PROPENSITY SCORE DESIGN WITH INSTRUMENTAL VARIABLES

Michael L. Anderson

Working Paper 18196
http://www.nber.org/papers/w18196

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2012

I thank David Card and Jeremy Magruder for insightful comments and am grateful to Tammie Vu and Yammy Kung for excellent research assistance. Funding for this project was provided by the California Agricultural Experiment Station. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

The Benefits of College Athletic Success: An Application of the Propensity Score Design with Instrumental Variables
Michael L. Anderson
NBER Working Paper No. 18196
June 2012
JEL No. C23,C26,I20,I23,J24

ABSTRACT

Spending on big-time college athletics is often justified on the grounds that athletic success attracts students and raises donations. Testing this claim has proven difficult because success is not randomly assigned. We exploit data on bookmaker spreads to estimate the probability of winning each game for college football teams. We then condition on these probabilities using a propensity score design to estimate the effects of winning on donations, applications, and enrollment. The resulting estimates represent causal effects under the assumption that, conditional on bookmaker spreads, winning is uncorrelated with potential outcomes. Two complications arise in our design. First, team wins evolve dynamically throughout the season. Second, winning a game early in the season reveals that a team is better than anticipated and thus increases expected season wins by more than one-for-one. We address these complications by combining an instrumental variables-type estimator with the propensity score design. We find that winning reduces acceptance rates and increases donations, applications, academic reputation, in-state enrollment, and incoming SAT scores.

Michael L. Anderson
Department of Agricultural and Resource Economics
207 Giannini Hall, MC 3310
University of California, Berkeley
Berkeley, CA 94720
and NBER
mlanderson@berkeley.edu

1 Introduction

College athletic spending at National Collegiate Athletic Association (NCAA) Division I schools exceeded \$7.9 billion in 2010 (Fulks 2011). This scale of expenditures is internationally unique and is partly justified on the basis that big-time athletic success, particularly in football and basketball, attracts students and generates donations. An extensive literature examines these claims but reaches inconsistent conclusions. A series of papers find positive effects of big-time athletic success on applications and contributions (Brooker and Klastorin 1981; Sigelman and Bookheimer 1983; Grimes and Chressanthis 1994; Murphy and Trandel 1994; Tucker 2004; Humphreys and Mondello 2007; Pope and Pope 2009), but a number of other studies find no impact of big-time athletic success on either measure (Sigelman and Carter 1979; Baade and Sundberg 1996; Turner et al. 2001; Meer and Rosen 2009; Orszag and Israel 2009). A central issue confronting all studies is the non-random assignment of athletic success. Schools with skilled administrators may attract donations, applicants, and coaching talent (selection bias), and surges in donations or applications may have a direct impact on athletic success (reverse causality). It is thus challenging to estimate causal effects of athletic success using observational data.

This article estimates the causal effects of college football success using a propensity score design. Propensity score methods are difficult to apply because researchers seldom observe all of the important determinants of treatment assignment. Treatment assignment is thus rarely ignorable given the data at the researcher's disposal (Rosenbaum and Rubin 1983; Dehejia and Wahba 1999). We overcome this challenge by exploiting data on bookmaker spreads (i.e., the expected score differential between the two teams) to estimate the probability of winning each game for NCAA "Division I-A" football teams. We then condition on these probabilities to estimate the effect of football success on donations and applications. If potential outcomes are independent of winning games after conditioning on bookmaker expectations, then our estimates represent causal effects.

We face two complications when estimating these effects. First, the treatment – team wins – evolves dynamically throughout the season, and the propensity score for each win depends on the outcomes of previous games. We address this issue by independently estimating the effect of wins in each week of the season. However, this introduces the second complication: a win early in the season is associated with a greater than one-for-one increase in total season wins because the winning team has (on average) revealed itself to be better than expected. We address this issue by combining an instrumental variables-type estimator with the propensity score estimator.

Applying this framework we find robust evidence that football success increases athletic donations, increases the number of applicants, lowers a school's acceptance rate, increases enrollment of in-state students, increases the average SAT score of incoming classes, and enhances a school's academic reputation. The estimates are up to twice as large as comparable estimates from the previous literature. There is less evidence that football success affects donations outside of athletic programs or enrollment of out-of-state students. The effects appear concentrated among teams in the six elite conferences classified as "Bowl Championship Series" (BCS) conferences, with less evidence of effects for teams in other conferences.

The paper is organized as follows. Section 2 describes the data, and Section 3 summarizes the cross-sectional and longitudinal relationships between football success, donations, and student body measures. Section 4 discusses the propensity score framework and estimation strategy. Section 5 presents estimates of the causal relationships between football success, donations, and student body measures. Section 6 concludes.

2 Data

Approximately 350 schools participate in NCAA Division I sports (the highest division of intercollegiate athletics). Of these schools, 120 field football teams in the Football Bowl Subdivision (FBS, formerly known as "Division I-A"). Teams in this subdivision play 10 to 13 games per season and are potentially eligible for post-season bowl games. Games between teams in this subdivision are high-profile events that are widely televised. We gathered data on games played by all FBS teams from 1986 to 2009 from the website *Covers.com*. Data include information on the game's date, the opponent, the score of each team, and the expected score differential between the two teams (known as the spread).

We combined these data with data on alumni donations, university academic reputations, applicants, acceptance rates, enrollment figures, and average SAT scores. Donations data come from the Voluntary Support of Education survey (VSE), acceptance rate and academic reputation data come from a survey of college administrators and high school counselors conducted annually by *US News and World Report*, and application, enrollment, and SAT data come from the Integrated Postsecondary Education Data System (IPEDS). Reporting dates for these measures range from 1986 to 2008.

The first column of Table 1 presents summary statistics for key variables. Each observation represents a single season for a single team. Actual season wins and expected season wins are both 5.4 games per season (out of an average of 10.7 games played per season). We exclude post-season games (bowl games) when calculating wins as participation in these games is endogenously determined by regular season wins. Alumni donations to athletic programs average $2.4 million per year, and total alumni donations (including both operating and capital support) average $18.8 million per year. The average school receives 13,748 applicants every year and accepts 70% of them. A typical incoming class contains 3,343 students and has a 25th percentile SAT score of 1,054 (IPEDS reports 25th and 75th percentile SAT scores; using the 75th percentile instead of the 25th percentile does not affect our conclusions).

The next two sets of columns in Table 1 present summary statistics for BCS and non-BCS conferences respectively. The six BCS conferences are the ACC, Big East, SEC, Big Ten, Big Twelve, and Pac-10 (now Pac-12). Winners of these conferences are automatically eligible for one of ten slots in the five prestigious BCS bowl games, and through 2012 only three non-BCS conference teams had ever played in a BCS bowl game. Teams in BCS conferences have more wins (note that inter-conference play is common), more alumni donations, better academic reputations, lower acceptance rates, and more applicants and enrolled students than teams in non-BCS conferences. Since BCS conference football teams have higher profiles, we expect that team success may have a larger impact for these schools (particularly for alumni donations), and we estimate results separately for BCS conferences.

3 Cross-Sectional and Longitudinal Results

3.1 Cross-sectional Results

We first estimate the cross-sectional relationship between lagged win percentage, alumni donations, and class characteristics. We estimate linear regressions of the form

$$y_{it+1} = \beta_0 + \beta_1 season\ wins_{it} + \beta_2 season\ games_{it} + \phi_{t+1} + \varepsilon_{it+1} \tag{1}$$

where y_{it+1} represents an outcome for school i in year $t+1$ (e.g., alumni donations, applicants, or acceptance rate), $season\ wins_{it}$ represents school i's football wins in year t, $season\ games_{it}$ represents school i's football games played in year t, and ϕ_{t+1} represents a year fixed effect that controls for aggregate time trends. The coefficient of interest is β_1. We lag the win measure by one year because the college football season runs from September to December, so the full effects of a winning season on donations or applications are unlikely to materialize until the following year.

The first set of columns in Table 2 reports results from estimating equation (1). One extra win is associated with a \$340,400 increase in alumni athletic donations and a \$960,100 increase in total alumni donations. Out-of-state and in-state enrollment increase by 34 and 91 students respectively. However, there is no significant relationship between wins and non-athletic operating donations, the average donation rate, academic reputation, applications, the acceptance rate, or the 25th percentile SAT score.

The large number of hypothesis tests raises the issue of multiple hypothesis testing – throughout the paper we test for effects on 10 different outcomes using multiple specifications and two subgroups (BCS and non-BCS schools). We address this issue by reporting "*q*-values" that control the False Discovery Rate (FDR) across all tables, along with standard per-comparison *p*-values. The False Discovery Rate is the proportion of rejections that are false discoveries (type I errors). Controlling FDR at 0.1, for example, implies that less than 10% of rejections will represent false discoveries. To calculate FDR *q*-values we use the "sharpened" FDR control algorithm from Benjamini et al. (2006), implemented in Anderson (2008). In most cases statistical significance remains even after controlling FDR.

3.2 Longitudinal Results

It is unlikely that the cross-sectional estimates in Table 2 represent causal effects; some unobserved factors that affect donations or applications are probably correlated with long-term athletic success. To remove unobserved factors that vary across schools but are fixed over time, we estimate linear regressions of the form

$$\Delta y_{it+1} \quad = \quad \beta_0 + \beta_1 \Delta season\ wins_{it} + \beta_2 \Delta season\ games_{it} + \phi_{t+1} + \varepsilon_{it+1} \qquad (2)$$

where

$$\Delta y_{it+1} = y_{it+1} - y_{it-1}$$
$$\Delta season\ wins_{it} = season\ wins_{it} - season\ wins_{it-2}$$
$$\Delta season\ games_{it} = season\ games_{it} - season\ games_{it-2}$$

Other variables are as defined in equation (1), and the coefficient of interest is again β_1. We difference over two years rather than one year because $season\ wins_{it}$ may affect y_{it}, so estimating a regression using one-year differences may attenuate our estimates of β_1. Indeed, differencing over one year rather than two years generates estimates that are 46% smaller in magnitude on average.[1]

The second set of columns in Table 2 reports results from estimating equation (2). A one win increase is associated with a $74,000 increase in alumni athletic donations but no statistically significant increases in total alumni donations or the alumni giving rate. Academic reputation increases by 0.002 points (0.003 standard deviations). Applications increase by 104 per year, acceptance rates drop 0.2 percentage points, and in-state enrollment increases by 17 students. There is no significant relationship between increases in wins and out-of-state enrollment or the 25th percentile SAT score.

Two patterns appear when comparing the cross-sectional and longitudinal estimates. First, the longitudinal estimates are much more precise, with standard errors that are ap-

[1] The divergence between the two-year differences results and the one-year differences results is largest for the acceptance rate estimate, which falls by 71%, and smallest for the SAT score estimate, which falls by 4%

proximately 3 to 15 times smaller. This is because much of the unexplained variation in y_{it} occurs across schools rather than within schools, and the differencing transformation in equation (2) removes most of the cross-school variation in y_{it}. Second, with the exception of acceptance rates, the longitudinal estimates are all smaller in magnitude than their cross-sectional counterparts. In several cases (alumni athletic donations, out-of-state enrollment, and in-state enrollment) it is possible to reject the hypothesis that the cross-sectional and longitudinal estimates converge to the same value. The divergence in the two sets of estimates suggests the presence of selection bias or reverse causality, though it could also indicate persistence in effects (i.e., several winning seasons may have a larger impact than a single winning season).

3.3 BCS Results

Teams in BCS conferences have higher profiles and thus may experience greater impacts from football success. Table 3 reports cross-sectional and longitudinal results for BCS teams. The first set of columns estimates equation (1). There is a significant relationship between wins and alumni athletic donations, but no other coefficients are significant at the $p = 0.05$ level. In most cases the cross-sectional estimates from the BCS sample are smaller than the cross-sectional estimates from the pooled sample. Estimates from the non-BCS sample are smaller still (see Appendix Table A1). Since conditioning on BCS status is equivalent to flexibly controlling for BCS status, this pattern suggests that unobserved differences between BCS and non-BCS schools may affect the pooled cross-sectional estimates.

The second set of columns in Table 3 estimates equation (2). Significant longitudinal relationships arise between wins and alumni athletic donations, academic reputation, applications, acceptance rates, and in-state and out-of-state enrollment. Longitudinal estimates from the BCS sample are of the same sign and typically larger magnitude when compared to longitudinal estimates from the pooled sample. This suggests that effects may be concentrated among BCS schools. Indeed, longitudinal estimates from the non-BCS sample are generally smaller in magnitude, and all but one are statistically insignificant (see Appendix

Table A1).

4 The Propensity Score Design

4.1 Theoretical Framework

Longitudinal estimates control for unobserved factors that differ across schools but are constant over time. Nevertheless, they should be interpreted with caution (LaLonde 1986); changes in donations or admissions could be related to changes in wins through reverse causality, and trends in other factors (e.g., coaching talent) might be related to both sets of variables. One way to improve the research design is to condition on observable factors that determine football wins. However, two problems arise in this context. First, we do not have data on a wide range of factors that plausibly determine whether a team wins. Second, even if such data were available, conditioning on a large number of factors introduces dimensionality problems and makes estimation via matching or subclassification difficult.

In cases with binary treatments, conditioning on the propensity score – the probability of treatment given the observable characteristics – is equivalent to conditioning on the observables themselves (Rosenbaum and Rubin 1983; Dehejia and Wahba 1999). Conditioning on the propensity score, or the probability of a win, is attractive in this case for two reasons. First, it is readily estimable using bookmaker spreads. Second, it is of low dimension.

However, the treatment $season\ wins_{it}$ is not binary but can instead realize integer values from 0 to 12. Furthermore, it is dynamically determined – each game occurs at a different point in the season. Recent research has extended propensity score methods to cases with categorical and continuous treatments (Hirano and Imbens 2004; Imai and van Dyk 2004). Since the distribution of a bounded random variable is defined by its moments, we could in principle calculate the conditional expectation, variance, and skewness of $season\ wins_{it}$ and condition on these quantities. In practice, however, we cannot calculate these conditional moments because bookmaker spreads are updated throughout the season. Importantly, bookmaker spreads in week s are a function of the team's performance in weeks 1 through $s-1$. Bookmaker spreads are thus endogenously determined by the treatment itself.

To overcome this problem we separately estimate the effect of winning for each week of the season. This ensures that the set of conditioning variables is always predetermined with respect to the treatment. Formally, let the treatment W_{ist} equal one if school i wins in week s of year t and zero otherwise. The $1 \times S$ row vector $\mathbf{w_{it}}$ contains the outcome of each game for the entire season, with the sth column containing the outcome for week s. The potential outcome $Y_{it+1}(w_{i1t}, ..., w_{iSt})$ is the value of the outcome for school i in year $t+1$ as a function of $\mathbf{w_{it}}$. Conditional on wins and losses in other weeks, the causal effect of a win in week s for school i in year t is $\beta_{ist} = Y_{it+1}(w_{i1t}, ..., w_{is-1t}, 1, w_{is+1t}, ..., w_{iSt}) - Y_{it+1}(w_{i1t}, ..., w_{is-1t}, 0, w_{is+1t}, ..., w_{iSt})$. For notational convenience we write $Y_{ist}(1) = Y_{it+1}(w_{i1t}, ..., w_{is-1t}, 1, w_{is+1t}, ..., w_{iSt})$ and $Y_{ist}(0) = Y_{it+1}(w_{i1t}, ..., w_{is-1t}, 0, w_{is+1t}, ..., w_{iSt})$.

While there are many potential outcomes for any combination of i, s, and t, we can only observe $Y_{it+1}(\mathbf{w_{it}})$, where $\mathbf{w_{it}}$ is the realized series of wins and losses for school i in year t. If wins were randomly assigned, then $Y_{it+1}(\mathbf{w_{it}})$ would be independent of W_{ist} for all values of $\mathbf{w_{it}}$, and we could estimate $E[Y_{ist}(1)]$ and $E[Y_{ist}(0)]$ using the sample analogs of $E[Y_{it+1}|W_{ist} = 1]$ and $E[Y_{it+1}|W_{ist} = 0]$. However, in all likelihood the potential outcomes $Y_{it+1}(\mathbf{w_{it}})$ are correlated with W_{ist}.

To estimate the causal effects of football wins we rely on two assumptions. First, we make the standard ignorability assumption:

Assumption 1 *Strong Ignorability of Treatment Assignment (Rosenbaum and Rubin 1983):* $W_{ist} \perp \{Y_{it+1}(\mathbf{w_{it}}) : \mathbf{w_{it}} \in \mathcal{W}\} \mid \mathbf{X_{ist}}$. *Furthermore,* $0 < P(W_{ist} = 1|\mathbf{X_{ist}}) < 1 \; \forall \; \mathbf{X_{ist}} \in \mathcal{X}$.

The ignorability assumption is crucial. It implies that, conditional on the observables $\mathbf{X_{ist}}$, W_{ist} is independent of the potential outcomes $Y_{it+1}(\mathbf{w_{it}})$. The ignorability assumption has two unique features in our study. First, $\mathbf{X_{ist}}$ represents the set of covariates observed by bookmakers in week s of year t rather than set of covariates available to us the researchers. Second, there is a strong economic reason to believe that $\mathbf{X_{ist}}$ contains all of the important observables. If there were an observable characteristic x^*_{ist} that predicted

W_{ist} and were not included in $\mathbf{X_{ist}}$, then professional bettors could use x^*_{ist} to form superior predictions of $P(W_{ist} = 1)$ than those formed by the bookmakers. This discrepancy would represent an arbitrage opportunity, and over time bookmakers would go bankrupt if they did not condition their spreads on x^*_{ist}. Thus, unlike in many data sets, we have a compelling reason to believe that the ignorability assumption holds. Studies of betting markets support this conjecture in that it has proven difficult to build models that outperform bookmakers' spreads by an economically meaningful margin (Glickman and Stern 1998; Levitt 2004; Stern 2008). Of course, it is theoretically possible that some characteristic that is unobservable to everyone and affects the probability of winning is correlated with the potential outcomes. As a robustness check, we reestimate the effects later in the season when team quality becomes well known, and we find similar results (see Section 5.2).

Proposition 1 *Let* $p(\mathbf{X_{ist}}) = P(W_{ist} = 1 \mid \mathbf{X_{ist}})$. *Under strong ignorability,* $W_{ist} \perp \{Y_{it+1}(\mathbf{w_{it}}) : \mathbf{w_{it}} \in \mathcal{W}\} \mid p(\mathbf{X_{ist}})$.

Proof: See Appendix A.

Corollary: Under strong ignorability, if $Cov(Y_{it+1}, W_{ist} \mid p(\mathbf{X_{ist}})) \neq 0$, then W_{ist} must have a causal effect on Y_{it+1} for some values of i, s, t.

Proof: Suppose that W_{ist} has no casual effect on Y_{it+1} for any values of i, s, t. Then $Y_{it+1}(\mathbf{w_{it}}) = Y_{it+1}$ for any $\mathbf{w_{it}} \in \mathcal{W}$. Thus, by Proposition 1, $W_{ist} \perp Y_{it+1} \mid p(\mathbf{X_{ist}})$.

Proposition 1 implies that we can test for a causal effect of W_{ist} on Y_{it+1} by conditioning on $p(\mathbf{X_{ist}})$ and estimating a regression of Y_{it+1} on W_{ist}. A significant regression coefficient in this case allows us to reject the sharp null hypothesis of no effect, and we use this proposition to estimate *p*-values in Tables 4–7. These tests are similar in spirit to propensity score based tests in recent research by Angrist and Kuersteiner on the effects of monetary policy shocks. However, to estimate the magnitude of the causal effect we make two additional assumptions.

Assumption 2 *Ignorability of Future Treatment Assignments:* $(W_{is+1t}, \ldots, W_{iSt}) \perp \{Y_{it+1}(\mathbf{w_{it}}) : \mathbf{w_{it}} \in \mathcal{W}\} \mid \mathbf{X_{ist}}$.

Assumption 3 *Homogenous Effects Across Weeks:* $Y_{it+1}(\mathbf{w_{it}}) = \mathbf{w_{it}}\boldsymbol{\beta_{it}} + Y_{it+1}(\mathbf{0})$*, where* $\boldsymbol{\beta_{it}}$ *is a* $S \times 1$ *column vector containing* β_{it} *in each entry, and* $\mathbf{0}$ *is a* $S \times 1$ *row vector of zeros.*

Assumption 2 extends the standard ignorability assumption to include weeks in year t following week s. There is minimal cost to making this assumption because it is difficult to imagine a scenario in which Assumption 1 holds but Assumption 2 does not. For example, Assumption 2 would be violated if there were some unobserved factor x^*_{ist} that affected team performance and were correlated with the potential outcome $Y_{it+1}(\mathbf{w_{it}})$. In that case, however, Assumption 1 would likely be violated as well.

Assumption 3 is a simplifying assumption that implies the effect of a win in week s for team i in year t is always β_{it}. This assumption consists of two components. First, it imposes a form of the Stable Unit Treatment Value Assumption (Rubin 1980) in that a team's performance in week s' is assumed not to impact the causal effect of performance in week s. Second, it imposes a homogeneity assumption that wins have equal effects in any week and that the margin of victory does not matter. We still allow effect heterogeneity by school and year. While Assumption 3 is not guaranteed to hold, it is no more restrictive than those imposed when interpreting the linear regression coefficients in Sections 3.1 and 3.2. It is also not needed to test the sharp null hypothesis of no effect – it is only necessary for estimating the magnitude of the effect.

Assumptions 2 and 3 are useful because they allow us to rescale the observed effect of winning in week s by the average increase in season wins associated with winning in week s. This is significant because winning in week s is associated with a higher probability of winning in later weeks, even after conditioning on $p(\mathbf{X_{ist}})$. This occurs because a team that wins reveals itself to be of unexpectedly higher quality. If effects are homogenous across weeks we can simply rescale the effect of a win in week s by the average increase in season wins associated that win. For example, if a win in week 3 is associated with an average of 1.5 extra wins during the course of a season (after conditioning on $p(\mathbf{X_{ist}})$), then we would rescale the estimated effect of a win in week 3 by $1/1.5 = 0.67$.

Proposition 2 *Under Assumptions 1, 2, and 3,*

$$\begin{aligned}
&E[Y_{it+1} \mid W_{ist}=1, p(\mathbf{X_{ist}})] - E[Y_{it+1} \mid W_{ist}=0, p(\mathbf{X_{ist}})] \\
&\quad = \quad E[\beta_{it} \mid p(\mathbf{X_{ist}})] \\
&\qquad \times (1 + E[\sum_{j=s+1}^{S} W_{ijt} \mid W_{ist}=1, p(\mathbf{X_{ist}})] - E[\sum_{j=s+1}^{S} W_{ijt} \mid W_{ist}=0, p(\mathbf{X_{ist}})]).
\end{aligned} \tag{3}$$

Thus,

$$\frac{E[Y_{it+1} \mid W_{ist}=1, p(\mathbf{X_{ist}})] - E[Y_{it+1} \mid W_{ist}=0, p(\mathbf{X_{ist}})]}{1 + E[\sum_{j=s+1}^{S} W_{ijt} \mid W_{ist}=1, p(\mathbf{X_{ist}})] - E[\sum_{j=s+1}^{S} W_{ijt} \mid W_{ist}=0, p(\mathbf{X_{ist}})]}$$

$$= \frac{\pi_{p(\mathbf{X_{ist}})}}{1+\gamma_{p(\mathbf{X_{ist}})}} = E[\beta_{it} \mid p(\mathbf{X_{ist}})].$$

Proof: See Appendix A.

Proposition 2 suggests a two-step procedure to estimate $E[\beta_{it} \mid p(\mathbf{X_{ist}})]$. First estimate $\pi_{p(\mathbf{X_{ist}})} = E[Y_{it+1} \mid W_{ist} = 1, p(\mathbf{X_{ist}})] - E[Y_{it+1} \mid W_{ist} = 0, p(\mathbf{X_{ist}})]$, or the change in Y_{it+1} associated with winning in week s. Then estimate $\gamma_{p(\mathbf{X_{ist}})} = E[\sum_{j=s+1}^{S} W_{ijt} \mid W_{ist} = 1, p(\mathbf{X_{ist}})] - E[\sum_{j=s+1}^{S} W_{ijt} \mid W_{ist} = 0, p(\mathbf{X_{ist}})]$ or the change in season wins associated with winning in week s. Finally combine the two estimates to generate $\beta_{p(\mathbf{X_{ist}})} = \pi_{p(\mathbf{X_{ist}})}/(1+\gamma_{p(\mathbf{X_{ist}})})$. In practice this estimator is similar to an instrumental variables (IV) estimator in which W_{ist} is the instrument (which is exogenous after conditioning on $p(\mathbf{X_{ist}})$) and $\sum_{j=s}^{S} W_{ijt}$ is the endogenous treatment to be instrumented.

4.2 Estimation

Estimation proceeds in three steps. First, for each game we estimate the propensity score $p(\mathbf{X_{ist}})$ using the published bookmakers' spread. Next, for each week s we stratify the sample into 12 bins based on the estimated propensity score (Dehejia and Wahba 1999).[2] Within each bin we estimate two linear regressions: a regression of the outcome y_{it+1}

[2] We experimented with greater numbers of bins (e.g., 20) and found generally similar results. A larger number of bins is attractive in that it allows a more flexible relationship between the estimated propensity score and the outcome. However, the statistical precision of our "first stage" regressions relating total season wins to winning in week s becomes poor as the number of bins increases (see discussion at the end of this subsection). Thus we limited the number of bins to 12.

on w_{ist} and a regression of the expected number of wins for the remainder of the season ($p^{+}_{ist} = \sum_{j=s+1}^{S} p(\mathbf{X_{ijt}})$) on w_{ist}. The first regression estimates $\pi_{p(\mathbf{X_{ist}})}$, and the second regression estimates $\gamma_{p(\mathbf{X_{ist}})}$. We include the estimated propensity score as a control in both regressions to eliminate any remaining imbalance within bins. Finally, we combine the estimates to form $\beta_{p(\mathbf{X_{ist}})} = \pi_{p(\mathbf{X_{ist}})}/(1+\gamma_{p(\mathbf{X_{ist}})})$. We compute the average of $\beta_{p(\mathbf{X_{ist}})}$ across all 144 bin-by-week combinations (12 bins per week by 12 weeks per season), weighting each estimate by the relevant sample size.[3]

To estimate the propensity score for each observation, we must translate the bookmakers' "point spread" into the probability of a win. The point spread represents a team's expected margin of victory. Previous research has shown that the margin of victory in National Football League (NFL) games is approximately normal, allowing a simple translation from point spreads to win probabilities (Stern 1991). However, these results may not apply directly to NCAA football games. To allow flexibility in the relationship, we estimate a logistic regression of w_{ist} on a fifth-order polynomial of the point spread. The results are highly significant; a likelihood ratio test of the hypothesis that all five coefficients equal zero produces a χ^2_5 statistic of 11,333. We use the fitted values from the logistic regression, $\hat{p}_{ist}$, as the propensity score. The minimum and maximum estimated propensity scores are 0.005 and 0.994 respectively.

For each week s we stratify the sample into 12 bins based on $\hat{p}_{ist}$. Bin 1 contains observations with lowest values of $\hat{p}_{ist}$, and bin 12 contains observations with the highest values of $\hat{p}_{ist}$. We construct the bins so that approximately the same number of observations fall into each bin. Within each bin b we estimate two regressions. Defining $\hat{p}^{+}_{ist} = \sum_{j=s+1}^{S} \hat{p}_{ijt}$, our "first stage" regression is:

$$\hat{p}^{+}_{ist} = \alpha_{0sb} + \gamma_{sb} \cdot w_{ist} + \alpha_{1sb} \cdot \hat{p}_{ist} + u_{ist} \tag{4}$$

[3] An alternative way to condition on the estimated propensity score is to weight treated observations by $\sqrt{1/\hat{p}}$ and untreated observations by $\sqrt{1/(1-\hat{p})}$. Hirano et al. (2003) show that weighting leads to efficient estimation of the average treatment effect. However, the weighting estimator does not apply directly to our context because we need to estimate and combine two separate coefficients, both of which may vary with the propensity score. If we simply estimate the "reduced form" effect of w_{ist} on y_{it+1} (i.e., we do not adjust for the fact that winning in week s is associated with more than one additional win over the course of a season), then we get qualitatively similar results if we stratify on the propensity score as described above or if we apply the weighting estimator.

The regression coefficient $\hat{\gamma}_{sb}$ estimates the relationship between winning in week s and winning during the remainder of the season, after conditioning on the propensity score in week s. We can replace the dependent variable p_{ist}^{+} with the sum of observed wins, $\sum_{j=s+1}^{S} w_{ijt}$, without changing the results (though precision is reduced).

Our second regression is equivalent to a "reduced form" regression in an IV setting. Our reduced form regression is:

$$y_{it+1} = \delta_{0sb} + \pi_{sb} \cdot w_{ist} + \delta_{1sb} \cdot \hat{p}_{ist} + v_{ist} \tag{5}$$

This regression estimates the relationship between winning in week s and the outcome y_{it+1}. Within each bin we then construct $\hat{\beta}_{sb} = \hat{\pi}_{sb}/(1 + \hat{\gamma}_{sb})$. After estimating equations (4) and (5) for each bin b in each week s, we compute

$$\hat{\beta} = \sum_{s=1}^{S} \sum_{b=1}^{12} \hat{\beta}_{sb} \cdot r_{sb} \tag{6}$$

where the weights r_{sb} sum to one and are proportional to the number of observations (games) in bin b in week s.

We make two additional modifications to improve the procedure described above. First, to ensure that the overlap assumption holds (i.e., $0 < p(\mathbf{X_{ist}}) < 1 \ \forall \ \mathbf{X_{ist}} \in \mathcal{X}$) we trim the sample for each week s to eliminate all observations with propensity scores less than the minimum score among winning observations for week s or greater than the maximum score among losing observations for week s. To be conservative, we further restrict the sample by dropping any observations with propensity scores less than 0.05 or greater than 0.95, and we eliminate any bins with less than two winning or two losing observations. Our main conclusions are robust to relaxing these restrictions.[4]

Our second modification focuses on the "first stage". Our raw first stage estimates of γ_{sb} are relatively noisy since a typical bin contains less than 150 observations. If $\hat{\gamma}_{sb}$ is near

[4] For example, adding the restriction dropping any observations with propensity scores less than 0.05 or greater than 0.95 does not change any of the estimated effects by more than 10%, except for the effect on SAT scores (which changes 13%). All of the restrictions combined do not change the statistical significance of any of the effects, though they do impact the sizes of some estimates (in particular, the effects on alumni donations, acceptance rates, and SAT scores change by 21%, 18%, and 30% respectively when all of the restrictions are added).

-1 in even a small number of cases, the confidence interval for $\hat{\beta}$ becomes very wide, since $\hat{\beta}_{sb} = \hat{\pi}_{sb}/(1 + \hat{\gamma}_{sb})$. This problem is well known in the instrumental variables literature (Bound et al. 1995). To improve the precision of our first stage estimates, we assume that the relationship between w_{ist} and $\sum_{j=s+1}^{S} w_{ijt}$ is reasonably smooth in the propensity score. For each week s we thus regress $\hat{\gamma}_{sb}$ on a quadratic in the bin number (b and b^2) and use the fitted values from this regression as our estimates of γ_{sb}. This improves precision by using information in bins near bin b to form the estimate of γ_{sb}. Alternatively, we could assume that the first stage coefficient γ_{sb} does not vary with the covariates $\mathbf{X_{ist}}$ and estimate a first stage regression that pools observations across all 12 bins for week s. This is standard in the instrumental variables literature (Angrist and Imbens 1995) and generates qualitatively similar results.

5 Propensity Score Results

5.1 Baseline Results

Table 4 reports the results of estimating equation (6) using all FBS ("Division I-A") schools. Each row reports the results for a different outcome. In the first set of columns we specify the outcome in levels (y_{it+1}), analogous to equation (1). In the second set of columns we specify the outcome in two-year differences (Δy_{it+1}), analogous to equation (2). Both sets of estimates should be consistent under Assumptions 1–3, but estimates in the second column are more precise because differencing removes much of the unexplained cross-school variation in y_{it+1}. Furthermore, differencing can help eliminate any bias that remains after conditioning on the propensity score (Smith and Todd 2005). In this section we focus on the differenced results (the second set of columns) because precision is limited in the first set of columns.

Estimates from the propensity score design imply that an extra win increases alumni athletic donations by $136,400. There are no statistically significant effects on non-athletic donations, total donations, or the alumni giving rate, though we lack sufficient precision to rule out economically significant effects. An extra win increases a school's academic reputation by 0.004 points (0.006 standard deviations) and increases the number of applicants by 135 (1%). Acceptance rates decrease by 0.3 percentage points (0.4%), in-state enrollment increases by 15 students (0.6%), and the 25th percentile SAT score increases 1.8 points (0.02 standard deviations). All of these results remain significant when controlling FDR at the $q = 0.10$ level, and all but the effects on academic reputation and acceptance rate remain significant when controlling FDR at the $q = 0.05$ level.

A natural question is how the propensity score results compare to the longitudinal results from Section 3.2. In general, the propensity score design produces larger estimates than the differenced regression. For example, the estimated effect on athletic donations is 84% larger, on academic reputation is 71% larger, on applicants is 31% larger, on acceptance rate is 45% larger, and on SAT scores is 205% larger. The only statistically significant effect that is smaller with the propensity score design than the differenced regression is in-state enrollment (which is 8% smaller). This pattern implies that changes in expected wins are associated with smaller outcome gains than changes in unexpected wins. On the one hand, this may occur because changes in expected wins are correlated with confounding factors that attenuate the true effects. On the other hand, it is possible that alumni and applicants react more strongly to unexpected wins than to expected wins. Previous research has found that negative reactions are more pronounced following unexpected NFL losses than following expected NFL losses (Card and Dahl 2011).

Table 5 reports propensity score results estimated separately for BCS and non-BCS teams. In all regressions the dependent variable is specified in two-year differences (Δy_{it+1}). For most outcomes the estimated effect for BCS teams (the first column) is larger (or more positive) than the estimated effect for non-BCS teams (the second column). However, for in-state enrollment and SAT scores, there is a larger effect among non-BCS schools than among BCS schools. Both these measures pertain to attracting students rather than satisfying alumni, and it is possible that winning seasons have a larger effect on visibility for lower-profile non-BCS schools than for high-profile BCS schools.

5.2 Effects Excluding Early Season Games

The propensity score design is particularly credible for games occurring later in the season. At that point bookmakers have better knowledge of each team's skill, and wins and losses are closer to truly random after conditioning on bookmaker odds. As a robustness check we estimate the results while excluding the first month of the season. The second set of columns in Table 6 reports these results (the first set of columns reports the baseline results from Table 4 for comparison).

In most cases the estimates change little when we exclude the first four games of the season. Statistically significant effects remain for alumni athletic donations, academic reputation, applicants, in-state enrollment, and SAT scores (the one exception is the acceptance rate, which becomes statistically insignificant). Precision falls since there are now fewer games used in the estimation sample, and many effects are only marginally significant after controlling FDR. Nevertheless, the similarity in the point estimates between the two columns for outcomes that are statistically significant in the baseline results is reassuring.

5.3 Persistent Effects

The baseline results demonstrate that wins in year t affect outcomes in year $t+1$. If wins have persistent effects, then wins in year t may also affect outcomes in year $t+2$ or beyond. Let w_{it} be wins in year t (i.e., $w_{it} = \sum_{s=1}^{S} w_{ist}$). It is tempting to estimate the effect of w_{it} on y_{it+2} by simply replacing y_{it+1} with y_{it+2} in the "reduced form" equation, equation (5). However, doing so overlooks the fact that winning in year t is correlated with winning in year $t+1$, even after conditioning on the propensity score. This occurs for the same reason that winning in week s is correlated with winning in week $s+1$ even after conditioning on the propensity score – a win in week s can reveal that a team has more talent than expected. Some of the estimated effect of winning in year t on y_{it+2} may thus result from increased wins in year $t+1$.

We use the following procedure to estimate the effect of w_{it} on y_{it+2} while holding w_{it+1} constant. First, we replace y_{it+1} with y_{it+2} in equation (5) and estimate equation (6). Denote this estimate as $\hat{\psi}$; $\hat{\psi}$ estimates the "reduced form" effect of w_{it} on y_{it+2} without controlling for changes in w_{it+1}. Next, we estimate the relationship between w_{it} and w_{it+1}. To do this we replace y_{it+1} with w_{it+1} in the "reduced form" equation (5) and estimate equation (6). Denote this estimate as $\hat{\lambda}$; $\hat{\lambda}$ estimates the relationship between w_{it} and w_{it+1} (after conditioning on the propensity score). Finally, we calculate $\hat{\theta} = \hat{\psi} - \hat{\lambda}\hat{\beta}$, where $\hat{\beta}$ is the causal effect of w_{it} on y_{it+1} (estimated in Section 5.1). In short, we adjust the reduced form effect of w_{it} on y_{it+2} to account for the fact that w_{it+1} is increasing in w_{it} (i.e., $\hat{\lambda}$) and that w_{it+1} affects y_{it+2} (i.e., $\hat{\beta}$).

The second set of columns in Table 7 reports estimates of $\hat{\theta}$, the effect of winning in year t on outcomes in year $t + 2$. There is little evidence that winning has effects that persist for two years. Most of the estimates are statistically insignificant and smaller than comparable estimates from Table 4 (reproduced in the first set of columns). None are statistically significant after controlling FDR. The effects of winning in year t appear to be concentrated in the following year.

6 Conclusions

For FBS schools, winning football games increases alumni athletic donations, enhances a school's academic reputation, increases the number of applicants and in-state students, reduces acceptance rates, and raises average incoming SAT scores. The estimates imply that large increases in team performance can have economically significant effects, particularly in the area of athletic donations. Consider a school that improves its season wins by 5 games (the approximate difference between a 25th percentile season and a 75th percentile season). Changes of this magnitude occur approximately 8% of the time over a one-year period and 13% of the time over a two-year period. This school may expect alumni athletic donations to increase by \$682,000 (28%), applications to increase by 677 (5%), the acceptance rate to drop by 1.5 percentage points (2%), in-state enrollment to increase by 76 students (3%), and incoming 25th percentile SAT scores to increase by 9 points (1%). These estimates are equal to or larger than comparable estimates from the existing literature. For example, among studies that found significant effects, a 5-win increase in team performance was associated with a 2.1% to 2.5% increase in applications (Murphy and Trandel 1994; Pope and Pope 2009) and a \$687,000 increase in restricted donations (Humphreys and Mondello 2007).

Do these effects imply that investing in team quality generates positive net benefits for an FBS school? Answering this question is difficult because we do not know the causal relationship between team investments and team wins. Nevertheless, we consider a simple back-of-the-envelope calculation to establish the potential return on team investments.

Orszag and Israel (2009) report that a \$1 million increase in "football team expenditures" is associated with a 6.7 percentage point increase in football winning percentage (0.8 games). If we interpret this relationship as causal, it implies that a \$1 million investment in football team expenditures increases alumni athletic donations by \$109,000, increases annual applications by 108, and increases the average incoming SAT score by 1.4 points. These effects seem too modest by themselves to offset the additional expenditures. However, if increases in team expenditures generate commensurate increases in athletic revenue (another finding in Orszag and Israel (2009), though a portion of this relationship is presumably due to reverse causality), then the effects estimated here represent a "bonus" that the school gets on top of the increased athletic revenue.

Two additional caveats apply when interpreting our results. First, we estimate the effects of unexpected wins. If a school invests in its football program and improves its record, alumni and applicant expectations will eventually change. The effects of a persistent increase in season wins may therefore differ from the effects we estimate here.

Second, the effects we observe likely operate through two channels. One channel is team quality – a team that plays well is more enjoyable to watch than a team that plays poorly, even holding constant the game's outcome. This is in part why the NFL can charge much higher ticket prices than competing leagues that employ less skilled players (e.g., the Arena Football League). The second channel is winning itself – fans and alumni enjoy winning games regardless of how well the team plays. Team records, however, are by definition a zero-sum game; one team's win is another team's loss. The effects demonstrated here thus do not change the "arms race" nature of team investment, as each team purchases its wins at the cost of other teams. While improving the overall level of play in the NCAA may attract more fans and alumni support through the first channel, it cannot have any effect on the second channel. A simultaneous investment of $1 million in each BCS football team will likely generate smaller effects on donations and applications than the estimates presented in this paper.

These caveats notwithstanding, we demonstrate that big-time athletic success can attract donations and students. We do so by extending the propensity score design to a dynamic setting in which multiple treatments occur at different points in time. In this setting the propensity score for any given treatment depends on the realized values of previous treatments. We apply this framework in a context in which the ignorability assumption is likely to hold and in which previous research has generated inconsistent conclusions. While the ignorability assumption does not apply in many circumstances, in those that it does these tools may facilitate estimation of causal effects.

Appendix A: Verification Of Propositions 1 and 2

Verification of Proposition 1 (Rosenbaum and Rubin 1983). Under Assumption 1

$$
\begin{aligned}
& P(W_{ist} = 1 \mid \{Y_{it+1}(\mathbf{w_{it}})\}_{\mathbf{w_{it}} \in \mathcal{W}}, p(\mathbf{X_{ist}})) \\
&= E[W_{ist} \mid \{Y_{it+1}(\mathbf{w_{it}})\}_{\mathbf{w_{it}} \in \mathcal{W}}, p(\mathbf{X_{ist}})] \\
&= E[E[W_{ist} \mid \{Y_{it+1}(\mathbf{w_{it}})\}_{\mathbf{w_{it}} \in \mathcal{W}}, p(\mathbf{X_{ist}}), \mathbf{X_{ist}}] \mid \{Y_{it+1}(\mathbf{w_{it}})\}_{\mathbf{w_{it}} \in \mathcal{W}}, p(\mathbf{X_{ist}})] \\
&= E[E[W_{ist} \mid \{Y_{it+1}(\mathbf{w_{it}})\}_{\mathbf{w_{it}} \in \mathcal{W}}, \mathbf{X_{ist}}] \mid \{Y_{it+1}(\mathbf{w_{it}})\}_{\mathbf{w_{it}} \in \mathcal{W}}, p(\mathbf{X_{ist}})] \\
&= E[E[W_{ist} \mid \mathbf{X_{ist}}] \mid \{Y_{it+1}(\mathbf{w_{it}})\}_{\mathbf{w_{it}} \in \mathcal{W}}, p(\mathbf{X_{ist}})] \\
&= E[p(\mathbf{X_{ist}}) \mid \{Y_{it+1}(\mathbf{w_{it}})\}_{\mathbf{w_{it}} \in \mathcal{W}}, p(\mathbf{X_{ist}})] \\
&= p(\mathbf{X_{ist}}) \\
&= P(W_{ist} = 1 \mid p(\mathbf{X_{ist}}))
\end{aligned}
$$

Verification of Proposition 2. Let $\mathbf{W_{is^-t}} = [W_{i1t}, ..., W_{is-1t}]$ and $\mathbf{W_{is^+t}} = [W_{is+1t}, ..., W_{iSt}]$. Let $\boldsymbol{\beta}^-_{\boldsymbol{it}}$ be a $(s-1) \times 1$ column vector containing β_{it} in each row and $\boldsymbol{\beta}^+_{\boldsymbol{it}}$ be a $(S-s) \times 1$ column vector containing β_{it} in each row. Note that under Assumptions 1, 2, and 3, combined with Proposition 1, β_{it} is independent of $\mathbf{W_{it}}$ after conditioning on $p(\mathbf{X_{ist}})$, as $\beta_{it} = \frac{1}{S} \cdot (Y_{it+1}(\mathbf{1}) - Y_{it+1}(\mathbf{0}))$. Thus

$$
\begin{aligned}
& E[Y_{it+1} \mid W_{ist} = 1, p(\mathbf{X_{ist}})] - E[Y_{it+1} \mid W_{ist} = 0, p(\mathbf{X_{ist}})] \\
&= E[Y_{it+1}(\mathbf{W_{is^-t}}, 1, \mathbf{W_{is^+t}}) \mid W_{ist} = 1, p(\mathbf{X_{ist}})] \\
&\quad - E[Y_{it+1}(\mathbf{W_{is^-t}}, 0, \mathbf{W_{is^+t}}) \mid W_{ist} = 0, p(\mathbf{X_{ist}})] \\
&= E[\mathbf{W_{is^-t}} \boldsymbol{\beta}^-_{\boldsymbol{it}} + \beta_{it} + \mathbf{W_{is^+t}} \boldsymbol{\beta}^+_{\boldsymbol{it}} \mid W_{ist} = 1, p(\mathbf{X_{ist}})] \\
&\quad - E[\mathbf{W_{is^-t}} \boldsymbol{\beta}^-_{\boldsymbol{it}} + \mathbf{W_{is^+t}} \boldsymbol{\beta}^+_{\boldsymbol{it}} \mid W_{ist} = 0, p(\mathbf{X_{ist}})] = \\
&= E[\mathbf{W_{is^-t}} \boldsymbol{\beta}^-_{\boldsymbol{it}} \mid W_{ist} = 1, p(\mathbf{X_{ist}})] - E[\mathbf{W_{is^-t}} \boldsymbol{\beta}^-_{\boldsymbol{it}} \mid W_{ist} = 0, p(\mathbf{X_{ist}})] + E[\beta_{it} \mid W_{ist} = 1, p(\mathbf{X_{ist}})] \\
&\quad + E[\mathbf{W_{is^+t}} \boldsymbol{\beta}^+_{\boldsymbol{it}} \mid W_{ist} = 1, p(\mathbf{X_{ist}})] - E[\mathbf{W_{is^+t}} \boldsymbol{\beta}^+_{\boldsymbol{it}} \mid W_{ist} = 0, p(\mathbf{X_{ist}})] \\
&= E[\mathbf{W_{is^-t}} \boldsymbol{\beta}^-_{\boldsymbol{it}} \mid p(\mathbf{X_{ist}})] - E[\mathbf{W_{is^-t}} \boldsymbol{\beta}^-_{\boldsymbol{it}} \mid p(\mathbf{X_{ist}})] + E[\beta_{it} \mid p(\mathbf{X_{ist}})] \\
&\quad + E[\mathbf{W_{is^+t}} \mid W_{ist} = 1, p(\mathbf{X_{ist}})] \cdot E[\boldsymbol{\beta}^+_{\boldsymbol{it}} \mid W_{ist} = 1, p(\mathbf{X_{ist}})] \\
&\quad - E[\mathbf{W_{is^+t}} \mid W_{ist} = 0, p(\mathbf{X_{ist}})] \cdot E[\boldsymbol{\beta}^+_{\boldsymbol{it}} \mid W_{ist} = 0, p(\mathbf{X_{ist}})]
\end{aligned}
$$

$$
\begin{aligned}
&= E[\beta_{it} \mid p(\mathbf{X_{ist}})] \\
&\quad + E[\mathbf{W_{is^+t}} \mid W_{ist} = 1, p(\mathbf{X_{ist}})] E[\boldsymbol{\beta}^{+}_{\boldsymbol{it}} \mid p(\mathbf{X_{ist}})] - E[\mathbf{W_{is^+t}} \mid W_{ist} = 0, p(\mathbf{X_{ist}})] E[\boldsymbol{\beta}^{+}_{\boldsymbol{it}} \mid p(\mathbf{X_{ist}})] \\
&= E[\beta_{it} \mid p(\mathbf{X_{ist}})] + (E[\mathbf{W_{is^+t}} \mid W_{ist} = 1, p(\mathbf{X_{ist}})] - E[\mathbf{W_{is^+t}} \mid W_{ist} = 0, p(\mathbf{X_{ist}})]) \cdot E[\boldsymbol{\beta}^{+}_{\boldsymbol{it}} \mid p(\mathbf{X_{ist}})] \\
&= E[\beta_{it} \mid p(\mathbf{X_{ist}})] + (E[\sum_{j=s+1}^{S} W_{ijt} \mid W_{ist} = 1, p(\mathbf{X_{ist}})] \cdot E[\beta_{it} \mid p(\mathbf{X_{ist}})] \\
&\quad - E[\sum_{j=s+1}^{S} W_{ijt} \mid W_{ist} = 0, p(\mathbf{X_{ist}})] \cdot E[\beta_{it} \mid p(\mathbf{X_{ist}})]) \\
&= E[\beta_{it} \mid p(\mathbf{X_{ist}})] \cdot (1 + E[\sum_{j=s+1}^{S} W_{ijt} \mid W_{ist} = 1, p(\mathbf{X_{ist}})] - E[\sum_{j=s+1}^{S} W_{ijt} \mid W_{ist} = 0, p(\mathbf{X_{ist}})])
\end{aligned}
$$

References

M. L. Anderson. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American Statistical Association*, 103(484):1481–1495, Dec. 2008.

J. D. Angrist and G. W. Imbens. Two-Stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90(430):431–442, 1995.

J. D. Angrist and G. M. Kuersteiner. Causal effects of monetary shocks: Semiparametric conditional independence tests with a multinomial propensity score. *Review of Economics and Statistics*, 93(3):725–747, 2011.

R. A. Baade and J. O. Sundberg. Fourth down and gold to go? Assessing the link between athletics and alumni giving. *Social Science Quarterly*, 77(4):789–803, 1996.

Y. Benjamini, A. M. Krieger, and D. Yekutieli. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507, Sept. 2006.

J. Bound, D. Jaeger, and R. Baker. Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430):443–450, 1995.

G. W. Brooker and T. D. Klastorin. To the victors belong the spoils? College athletics and alumni giving. *Social Science Quarterly*, 62(4):744–50, Dec. 1981.

D. Card and G. B. Dahl. Family violence and football: The effect of unexpected emotional cues on violent behavior. *The Quarterly Journal of Economics*, 126(1):103–143, Feb. 2011.

R. H. Dehejia and S. Wahba. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94 (448):1053–1062, 1999.

D. L. Fulks. Revenues and expenses 2004–2010: NCAA Division I intercollegiate athletics programs report. Technical report, National Collegiate Athletic Association, Indianapolis, IN, 2011.

M. E. Glickman and H. S. Stern. A State-Space model for national football league scores. *Journal of the American Statistical Association*, 93(441):25–35, 1998.

P. W. Grimes and G. A. Chressanthis. Alumni contributions to academics. *American Journal of Economics and Sociology*, 53(1):27–40, 1994.

K. Hirano and G. Imbens. The propensity score with continuous treatments. In A. Gelman and X.-L. Meng, editors, *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family*, chapter 7, pages 73–84. John Wiley and Sons, Ltd, 2004.

K. Hirano, G. Imbens, and G. Ridder. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4):1161–1189, 2003.

B. R. Humphreys and M. Mondello. Intercollegiate athletic success and donations at NCAA division i institutions. *Journal of Sport Management*, 21(2):265–280, 2007.

K. Imai and D. A. van Dyk. Causal inference with general treatment regimes. *Journal of the American Statistical Association*, 99(467):854–866, Sept. 2004.

R. J. LaLonde. Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review*, 76(4):604–20, 1986.

S. Levitt. Why are gambling markets organised so differently from financial markets? *Economic Jounal*, 114(495):223–246, April 2004.

J. Meer and H. S. Rosen. The impact of athletic performance on alumni giving: An analysis of microdata. *Economics of Education Review*, 28(3):287–294, June 2009.

R. G. Murphy and G. A. Trandel. The relation between a university's football record and the size of its applicant pool. *Economics of Education Review*, 13(3):265–270, Sept. 1994.

J. Orszag and M. Israel. The empirical effects of collegiate athletics: An update based on 2004-2007 data. Technical report, Compass Lexecon, 2009.

D. G. Pope and J. C. Pope. The impact of college sports success on the quantity and quality of student applications. *Southern Economic Journal*, 75(3), 2009.

P. R. Rosenbaum and D. B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41 –55, Apr. 1983.

D. Rubin. Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association*, 75(371):591–593, 1980.

L. Sigelman and S. Bookheimer. Is it whether you win or lose? Monetary contributions to Big-Time college athletic programs. *Social Science Quarterly*, 64(2):347–59, June 1983.

L. Sigelman and R. Carter. Win one for the giver? Alumni giving and Big-Time college sports. *Social Science Quarterly*, 60(2):284–94, 1979.

J. Smith and P. Todd. Does matching overcome lalonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125(1-2):305–353, 2005.

H. Stern. On the probability of winning a football game. *The American Statistician*, 45(3): 179–183, 1991.

H. S. Stern. Point spread and odds betting: Baseball, basketball, and american football. In D. B. Hausch and W. T. Ziemba, editors, *Handbook of Sports and Lottery Markets*, pages 223–237. Elsevier, Oxford, UK, 2008.

I. B. Tucker. A reexamination of the effect of big-time football and basketball success on graduation rates and alumni giving rates. *Economics of Education Review*, 23(6): 655–661, Dec. 2004.

S. E. Turner, L. A. Meserve, and W. G. Bowen. Winning and giving: Football results and alumni giving at selective private colleges and universities. *Social Science Quarterly*, 82 (4):812–826, 2001.

Table 1: Summary Statistics

	All Teams		BCS		Non-BCS		Dates	
Variable:	Mean	N	Mean	N	Mean	N	Available	Source
Season Wins	5.4	2,360	5.9	1,477	4.5	883	1987–2009	Covers.com
	(2.7)		(2.6)		(2.5)			
Season Games	10.7	2,360	10.7	1,477	10.5	883	1987–2009	Covers.com
	(0.9)		(0.8)		(1.0)			
Expected Wins	5.4	2,360	5.9	1,477	4.5	883	1987–2009	Covers.com
	(2.1)		(2.1)		(1.8)			
Alumni Athletic Operating Donations	$2,439	925	$3,829	534	$540	391	1999–2009	VSE
	(3,272)		(3,705)		(0,593)			
Alumni Nonathletic Operating Donations	$7,136	925	$11,000	534	$1,912	391	1999–2009	VSE
	(12,100)		(14,500)		(3,684)			
Total Alumni Donations	$18,800	1,793	$27,000	1,129	$4,908	664	1990–2009	VSE
	(26,700)		(30,400)		(6,520)			
Alumni Giving Rate	0.141	1,837	0.165	1,141	0.102	696	1990–2009	VSE
	(0.078)		(0.077)		(0.061)			
Academic Reputation	3.22	1,044	3.47	701	2.71	343	1997–2008	US News
	(0.63)		(0.55)		(0.44)			
Applicants	13,748	840	16,623	502	9,479	338	1999–2007	IPEDS
	(8,187)		(8,058)		(6,305)			
Acceptance Rate	0.698	1,591	0.669	1,064	0.756	527	1988–2008	US News
	(0.181)		(0.184)		(0.159)			
First-Time Out-of-State Enrollment	802	1,498	1,015	913	469	585	1986–2008	IPEDS
	(632)		(596)		(536)			
First-Time In-State Enrollment	2,541	1,498	2,814	913	2,115	585	1986–2008	IPEDS
	(1,404)		(1,522)		(1,067)			
25th Percentile SAT	1,054	718	1,097	451	981	267	1999–2008	IPEDS
	(119)		(103)		(108)			

Notes: Donation measures are in thousands of dollars. Parentheses contain standard deviations. One observation is a single school for a single year.

Table 2: OLS and Longitudinal Relationships Between Football Wins and Outcomes

	OLS Relationships				Longitudinal Relationships			
Outcome:	Coefficient	*p*-val	FDR *q*-val	N	Coefficient	*p*-val	FDR *q*-val	N
Alumni Athletic Operating Donations	340.4	0.001	0.017	914	74.0	0.002	0.017	698
	(99.8)				(23.2)			
Alumni Nonathletic Operating Donations	161.6	0.527	0.604	914	117.2	0.246	0.381	698
	(254.7)				(100.3)			
Total Alumni Donations	960.1	0.039	0.095	1,771	207.7	0.150	0.281	1,403
	(460.6)				(143.2)			
Alumni Giving Rate	0.0025	0.150	0.281	1,814	0.0003	0.358	0.497	1,437
	(0.0017)				(0.0003)			
Academic Reputation	0.015	0.363	0.497	1,031	0.002	0.029	0.082	727
	(0.016)				(0.001)			
Applicants	314.8	0.104	0.204	826	103.6	0.000	0.001	585
	(192.0)				(26.8)			
Acceptance Rate	-0.001	0.841	0.845	1,562	-0.002	0.009	0.041	1,084
	(0.004)				(0.001)			
First-Time Out-of-State Enrollment	34.4	0.002	0.017	1,397	2.9	0.262	0.401	1,076
	(11.0)				(2.6)			
First-Time In-State Enrollment	91.3	0.004	0.027	1,397	16.6	0.001	0.017	1,076
	(31.3)				(4.9)			
25th Percentile SAT	3.1	0.355	0.497	707	0.6	0.079	0.165	474
	(3.3)				(0.3)			

Notes: Donation measures are in thousands of dollars. Parentheses contain standard errors clustered at the school level. FDR *q*-values control the False Discovery Rate across all tables. Each coefficient is from a separate regression in which the outcome is regressed on season wins. All regressions include season games and year fixed effects as controls, and all variables in longitudinal regressions are differenced over two years (see Section 3.2).

Table 3: OLS and Longitudinal Relationships Between Football Wins and Outcomes for BCS Schools

	OLS Relationships				Longitudinal Relationships			
Outcome:	Coefficient	p-val	FDR q-val	N	Coefficient	p-val	FDR q-val	N
Alumni Athletic Operating Donations	355.6	0.008	0.038	532	121.9	0.002	0.017	417
	(129.5)				(37.3)			
Alumni Nonathletic Operating Donations	-427.4	0.349	0.497	532	191.5	0.281	0.405	417
	(452.6)				(175.7)			
Total Alumni Donations	37.1	0.955	0.850	1,126	328.2	0.131	0.256	920
	(659.1)				(214.7)			
Alumni Giving Rate	-0.0010	0.686	0.743	1,138	0.0004	0.410	0.526	926
	(0.0024)				(0.0004)			
Academic Reputation	-0.016	0.378	0.517	699	0.002	0.028	0.082	519
	(0.018)				(0.001)			
Applicants	108.0	0.633	0.720	500	121.3	0.002	0.017	360
	(225.2)				(37.4)			
Acceptance Rate	0.000	0.940	0.850	1,061	-0.003	0.001	0.017	742
	(0.005)				(0.001)			
First-Time Out-of-State Enrollment	3.0	0.824	0.845	851	7.2	0.002	0.017	654
	(13.3)				(2.2)			
First-Time In-State Enrollment	88.4	0.060	0.131	851	21.4	0.002	0.017	654
	(46.2)				(6.6)			
25th Percentile SAT	-2.0	0.623	0.711	449	0.3	0.458	0.545	311
	(4.0)				(0.4)			

Notes: Donation measures are in thousands of dollars. Parentheses contain standard errors clustered at the school level. FDR q-values control the False Discovery Rate across all tables. Each coefficient is from a separate regression in which the dependent variable is regressed on season wins. All regressions include season games and year fixed effects as controls, and all variables in longitudinal regressions are differenced over two years (see Section 3.2). The sample is restricted to BCS schools.

Table 4: Effects of Football Wins on Outcomes

	Dependent Variable: y_{it+1}				Dependent Variable: Δy_{it+1}			
Outcome:	Coefficient	p-val	FDR q-val	N	Coefficient	p-val	FDR q-val	N
Alumni Athletic Operating Donations	175.1	0.029	0.082	818	136.4	0.001	0.017	637
	(67.6)				(41.1)			
Alumni Nonathletic Operating Donations	64.3	0.894	0.845	818	227.9	0.210	0.355	637
	(283.6)				(171.4)			
Total Alumni Donations	527.1	0.228	0.377	1,565	311.9	0.450	0.539	1,264
	(429.4)				(295.7)			
Alumni Giving Rate	0.0017	0.155	0.286	1,605	-0.0001	0.698	0.743	1,293
	(0.0012)				(0.0005)			
Academic Reputation	-0.003	0.717	0.743	919	0.004	0.015	0.056	669
	(0.013)				(0.002)			
Applicants	4.9	0.942	0.850	740	135.3	0.005	0.029	533
	(184.3)				(49.9)			
Acceptance Rate	0.000	0.955	0.850	1,379	-0.003	0.022	0.069	967
	(0.003)				(0.001)			
First-Time Out-of-State Enrollment	11.9	0.236	0.379	1,237	-0.4	0.983	0.851	1,038
	(11.1)				(3.4)			
First-Time In-State Enrollment	10.5	0.783	0.838	1,237	15.2	0.007	0.035	1,038
	(25.0)				(5.8)			
25th Percentile SAT	2.6	0.387	0.525	640	1.8	0.002	0.017	431
	(2.9)				(0.6)			

Notes: Donation measures are in thousands of dollars. Parentheses contain standard errors clustered at the school level, and p-values test the sharp null hypothesis of no effect. FDR q-values control the False Discovery Rate across all tables. Each coefficient represents the effect of an additional football win on the outcome, computed from equation (6) as described in Section 4.2.

Table 5: Effects of Football Wins on Outcomes for BCS and Non-BCS Schools

	Sample: BCS Schools				Sample: Non-BCS Schools			
Outcome:	Coefficient	p-val	FDR q-val	N	Coefficient	p-val	FDR q-val	N
Alumni Athletic Operating Donations	175.0	0.004	0.027	381	17.7	0.157	0.286	258
	(62.7)				(11.7)			
Alumni Nonathletic Operating Donations	387.4	0.146	0.281	381	39.9	0.420	0.536	258
	(274.1)				(46.8)			
Total Alumni Donations	566.0	0.242	0.379	829	-248.3	0.038	0.095	441
	(434.4)				(131.9)			
Alumni Giving Rate	0.0000	0.990	0.851	832	-0.0004	0.506	0.604	462
	(0.0007)				(0.0006)			
Academic Reputation	0.004	0.022	0.069	472	0.002	0.242	0.379	197
	(0.002)				(0.003)			
Applicants	129.9	0.013	0.054	325	126.0	0.277	0.404	212
	(58.2)				(79.3)			
Acceptance Rate	-0.003	0.030	0.083	657	0.001	0.853	0.845	316
	(0.002)				(0.002)			
First-Time Out-of-State Enrollment	2.3	0.473	0.564	637	-5.1	0.266	0.402	404
	(4.8)				(5.1)			
First-Time In-State Enrollment	12.3	0.086	0.178	637	22.7	0.020	0.069	404
	(7.8)				(9.2)			
25th Percentile SAT	1.0	0.098	0.198	282	2.9	0.092	0.188	154
	(0.7)				(1.4)			

Notes: In all specifications the dependent variable is Δy_{it+1} (two year differences). Donation measures are in thousands of dollars. Parentheses contain standard errors clustered at the school level, and p-values test the sharp null hypothesis of no effect. FDR q-values control the False Discovery Rate across all tables. Each coefficient represents the effect of an additional football win on the outcome, computed from equation (6) as described in Section 4.2.

Table 6: Effects of Football Wins in Later Weeks on Outcomes

	Sample: All Weeks				Sample: Weeks 5–12			
Outcome:	Coefficient	*p*-val	FDR *q*-val	N	Coefficient	*p*-val	FDR *q*-val	N
Alumni Athletic Operating Donations	136.4	0.001	0.017	637	141.6	0.012	0.051	637
	(41.1)				(53.8)			
Alumni Nonathletic Operating Donations	227.9	0.210	0.355	637	194.9	0.289	0.414	637
	(171.4)				(187.3)			
Total Alumni Donations	311.9	0.450	0.539	1,264	579.8	0.185	0.336	1,264
	(295.7)				(376.6)			
Alumni Giving Rate	-0.0001	0.698	0.743	1,293	0.0000	0.871	0.845	1,293
	(0.0005)				(0.0006)			
Academic Reputation	0.004	0.015	0.056	669	0.004	0.035	0.089	669
	(0.001)				(0.002)			
Applicants	135.3	0.005	0.029	533	119.9	0.049	0.110	533
	(49.9)				(62.1)			
Acceptance Rate	-0.003	0.022	0.069	967	-0.002	0.200	0.355	967
	(0.001)				(0.002)			
First-Time Out-of-State Enrollment	-0.4	0.983	0.851	1,038	-2.8	0.525	0.604	1,038
	(3.4)				(4.5)			
First-Time In-State Enrollment	15.2	0.007	0.035	1,038	15.4	0.034	0.089	1,038
	(5.8)				(7.5)			
25th Percentile SAT	1.8	0.002	0.017	431	1.9	0.017	0.062	431
	(0.6)				(0.8)			

Notes: In all specifications the dependent variable is Δy_{it+1} (two year differences). Donation measures are in thousands of dollars. Parentheses contain standard errors clustered at the school level, and *p*-values test the sharp null hypothesis of no effect. FDR *q*-values control the False Discovery Rate across all tables. Each coefficient represents the effect of an additional football win on the outcome, computed from equation (6) as described in Section 4.2.

Table 7: Effects of Football Wins on Outcomes Two Years Later

	Dependent Variable: Δy_{it+1}				Dependent Variable: Δy_{it+2}			
Outcome:	Coefficient	*p*-val	FDR *q*-val	N	Coefficient	*p*-val	FDR *q*-val	N
Alumni Athletic Operating Donations	136.4	0.001	0.017	891	78.6	0.233	0.379	612
	(41.1)				(65.4)			
Alumni Nonathletic Operating Donations	227.9	0.210	0.355	891	-207.8	0.433	0.536	612
	(171.4)				(264.2)			
Total Alumni Donations	311.9	0.450	0.539	1,712	86.5	0.865	0.845	1,294
	(295.7)				(507.6)			
Alumni Giving Rate	-0.0001	0.698	0.743	1,755	-0.0004	0.589	0.680	1,327
	(0.0005)				(0.0007)			
Academic Reputation	0.004	0.015	0.056	1,010	0.000	0.956	0.850	707
	(0.002)				(0.002)			
Applicants	135.3	0.005	0.029	805	-88.3	0.326	0.463	496
	(49.9)				(89.5)			
Acceptance Rate	-0.003	0.022	0.069	1,515	-0.004	0.044	0.104	1,141
	(0.001)				(0.002)			
First-Time Out-of-State Enrollment	-0.4	0.983	0.851	1,359	-13.5	0.040	0.096	622
	(3.4)				(6.5)			
First-Time In-State Enrollment	15.2	0.007	0.035	1,359	-2.7	0.814	0.845	622
	(5.8)				(11.5)			
25th Percentile SAT	1.8	0.002	0.017	693	1.0	0.273	0.404	398
	(0.6)				(0.9)			

Notes: Donation measures are in thousands of dollars. Parentheses contain standard errors clustered at the school level, and *p*-values test the sharp null hypothesis of no effect. FDR *q*-values control the False Discovery Rate across all tables. Each coefficient represents the effect of an additional football win on the outcome, computed as described in Section 5.3.

Table A1: OLS and Longitudinal Relationships Between Football Wins and Outcomes for Non-BCS Schools

	OLS Relationships				Longitudinal Relationships			
Outcome:	Coefficient	p-val	FDR q-val	N	Coefficient	p-val	FDR q-val	N
Alumni Athletic Operating Donations	19.6	0.402	0.525	382	7.7	0.299	0.427	281
	(23.2)				(7.3)			
Alumni Nonathletic Operating Donations	59.8	0.598	0.683	382	36.5	0.275	0.404	281
	(112.5)				(33.0)			
Total Alumni Donations	-22.7	0.896	0.845	645	-46.4	0.591	0.680	483
	(172.7)				(85.7)			
Alumni Giving Rate	0.0011	0.509	0.604	676	0.0002	0.665	0.743	511
	(0.0017)				(0.0004)			
Academic Reputation	-0.002	0.888	0.845	332	0.001	0.568	0.671	208
	(0.013)				(0.002)			
Applicants	-204.7	0.307	0.436	326	76.5	0.002	0.017	225
	(198.2)				(23.3)			
Acceptance Rate	0.007	0.115	0.222	501	0.001	0.520	0.604	342
	(0.004)				(0.001)			
First-Time Out-of-State Enrollment	19.8	0.364	0.497	546	-3.8	0.486	0.580	422
	(21.6)				(5.4)			
First-Time In-State Enrollment	14.0	0.614	0.704	546	9.2	0.201	0.355	422
	(27.5)				(7.1)			
25th Percentile SAT	-2.6	0.528	0.604	258	1.1	0.105	0.204	163
	(4.1)				(0.6)			

Notes: Donation measures are in thousands of dollars. Parentheses contain standard errors clustered at the school level. FDR q-values control the False Discovery Rate across all tables. Each coefficient is from a separate regression in which the dependent variable is regressed on season wins. All regressions include season games and year fixed effects as controls, and all variables in longitudinal regressions are differenced over two years (see Section 3.2). The sample is restricted to non-BCS schools.