The Benefits of College Athletic Success: An Application of the Propensity Score Design *

Michael L. Anderson

UC Berkeley and NBER

January 21, 2016

Abstract

Spending on big-time college athletics is often justified on the grounds that athletic success attracts students and raises donations. 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. We find that winning reduces acceptance rates and increases donations, applications, academic reputation, in-state enrollment, and incoming SAT scores.

JEL Codes: C22, C26, I23, Z20 Keywords: Selection on observables; ignorable treatment assignment; big-time football; instrumental variables; sequential treatment effects

*Michael L. Anderson is Associate Professor, Department of Agricultural and Resource Economics, University of California, Berkeley, CA 94720 (E-mail: mlanderson@berkeley.edu). He thanks David Card, John Siegfried, Jeremy Magruder, and Josh Angrist for insightful comments and suggestions and is grateful to Tammie Vu and Yammy Kung for excellent research assistance. All mistakes are the author's.

1 Introduction

Athletic spending at National Collegiate Athletic Association (NCAA) Division I schools exceeded $7.9 billion in 2010, and only 18% of athletic programs at the 120 Football Bowl Subdivision (FBS) schools covered their operating costs (Fulks 2011). At the same time, college football attendance reached 49.7 million spectators (Johnson 2012). This scale of expenditures, subsidy, and attendance is internationally unique and has generated a spirited debate within and across schools about the appropriate level of athletic spending (Thomas 2009a,b; Drape and Thomas 2010).

High spending is justified partly 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; Tucker and Amato 1993; Grimes and Chressanthis 1994; Murphy and Trandel 1994; Mixon Jr et al. 2004; Tucker 2004, 2005; Humphreys and Mondello 2007; Pope and Pope 2009), but a number of other studies find mixed evidence or no impact of big-time athletic success on either measure (Sigelman and Carter 1979; McCormick and Tinsley 1987; Bremmer and Kesselring 1993; Baade and Sundberg 1996; Rhoads and Gerking 2000; Turner et al. 2001; Litan et al. 2003; Meer and Rosen 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" (now "FBS") 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 in two manners. First, we combine an instrumental variables-type estimator with the propensity score estimator. Under an assumption of additively separable treatment effects this estimates a weighted average of team-specific treatment effects. Second, we estimate the effects of an entire season of wins and losses using a sequential treatment effects model.

Applying these estimators 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 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 discusses the propensity score framework and estimation strategies. Section 4 presents estimates of the causal relationships between football success, donations, and student body measures. Section 5 concludes.

2 Data

Approximately 350 schools participate in NCAA Division I sports (the highest division of intercollegiate athletics). Enrollment at these schools totals 4.5 million students, or 65% of total enrollment at all NCAA schools (NCAA 2014; most public or nonprofit 4-year degree-granting institutions are part of the NCAA). Within Division I schools, 120 field football teams in the Football Bowl Subdivision (FBS, formerly known as "Division I-A"). Participation in Division I sports in general, and the FBS in particular, requires substantial financial resources; average athletic spending in 2010 was $46.7 million at FBS schools and $13.1 million at other Division I schools (Fulks 2011). Since participation in Division I requires scale, most schools are not in Division I, but the majority of 4-year undergraduate students attend Division I schools.

Teams in the FBS 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, and the spread, or expected score differential between the two teams.

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.

Within the FBS there is a subset of six conferences known informally as "Bowl Championship Series" (BCS) conferences. The six BCS conferences are the Atlantic Coast Conference (ACC), Big East (now American Athletic Conference), Southeastern Conference (SEC), Big Ten, Big Twelve, and Pac-10 (now Pac-12).[1] Until 2014 winners of these conferences were automatically

[1]There is spatial clustering in conference membership. ACC and Big East teams are on the East Coast, SEC teams are in the Southeast, Big Ten teams are in the northern Midwest, Big

eligible for one of ten slots in five prestigious BCS bowl games, and only five non-BCS conference teams had ever played in a BCS bowl game.[2] Membership in a BCS conference is a signal of prestige for a football program, and we expect that success may have larger effects for BCS teams than for non-BCS teams. We thus estimate separate effects for BCS and non-BCS teams, and we code a team as BCS if it was in a BCS conference at the beginning of our sample.[3]

Table 1 presents summary statistics for key variables by BCS status. Each observation represents a single season for a single team. For BCS teams, actual (expected) season wins are 5.9 (5.8) games per season out of an average of 10.8 games played. Non-BCS schools win (expect to win) only 4.6 (4.7) games per season since the two types of teams regularly play each other. In both cases we exclude post-season games (bowl games) when calculating wins as participation in these games is endogenously determined by regular season wins, and we do not observe the propensity score of post-season participation.[4] Alumni donations to athletic programs average $4.0 million per year at BCS schools and $0.7 million per year at non-BCS schools, and total alumni donations (including both operating and capital support) average $27.6 million per year at BCS schools and $5.4 million per year at non-BCS schools. The average BCS (non-BCS) school receives 16,815

Twelve teams are in the southern Midwest and Texas, and Pac-10 teams are on the West Coast. Nevertheless, there is geographic overlap between different conferences; the eastern most Big Ten school – Penn State – lies east of several ACC and Big East schools.

[2]Starting in 2014, the NCAA has switched to a system in which performance in playoff games – which previously did not exist – determines participation in major bowl games.

[3]In only one case during our sample period did a BCS team move from a BCS conference to a non-BCS conference; in 2004 Temple University transitioned from the Big East to independent status (and later to the Mid-American Conference) due to poor attendance and non-competitiveness. In several cases, however, non-BCS teams joined BCS conferences. Cincinnati, Louisville, and South Florida joined the Big East in 2005, and UConn joined the Big East in 2002.

[4]When interpreting our regular season results, post-season participation is a potential channel through which the effects may operate. Thus, while our results correctly estimate the average effect of a regular season win, it is possible that the effect of winning may be larger when a regular season win induces a team to participate in a post-season bowl game, and smaller when it does not.

(9,660) applicants every year and accepts 67% (76%) of them. A typical incoming BCS (non-BCS) class contains 3,849 (2,611) students and has a 25th percentile SAT score of 1,101 (984).[5] In general, BCS schools have more resources, larger student bodies, and more qualified students than non-BCS schools; these differences corroborate our choice to estimate results separately for BCS and non-BCS teams.

3 The Propensity Score Design

Consider linear regressions of the form

$$\Delta Y_{i(t+1)} = \beta_0 + \beta_1 W_{it} + \beta_2 W_{i(t-2)} + \beta_3 S_{it} + \beta_4 S_{i(t-2)} + \phi_{t+1} + \varepsilon_{i(t+1)} \quad (1)$$

where $Y_{i(t+1)}$ represents an outcome for school *i* in year $t+1$ (e.g. alumni donations, applicants, or acceptance rate), W_{it} represents school *i*'s football wins in year t, S_{it} represents school *i*'s football games played in year t, Δ is a two-period differencing operator (i.e. $\Delta Y_{it} = Y_{it} - Y_{i(t-2)}$), 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 relative to the outcome measure 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.[6] Nevertheless, W_{it} may affect Y_{it} towards the end of the year, so we difference over two years rather than one year to avoid attenuating the estimates of β_1.

Differencing models control for unobserved factors that vary across units but are constant over time. Nevertheless, time-varying unobservables correlated with treatment assignment may con-

[5]IPEDS reports 25th and 75th percentile SAT scores; using the 75th percentile instead of the 25th percentile does not affect our conclusions.

[6]In the case of outcomes measured on a fiscal year basis (e.g. donations), there could be no causal effect from contemporaneous wins; for most schools the 2012 fiscal year ends before the 2012 football season begins. Failing to lag the win measure severely attenuates many of the results, which supports our hypothesis that effects do not – and in many cases cannot – materialize until the following calendar year.

found fixed effects or differenced estimates (LaLonde 1986). In our context 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, but 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 our case for two reasons. First, it is readily estimable using bookmaker spreads. Second, it is of low dimension.

The treatment W_{it}, however, 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, and the outcome of a game in week s may affect expectations about the outcome of a game in week $s+1$.[7] We address these issues in three ways. First, we exploit the conditional independence

[7]Hirano and Imbens (2004) and Imai and van Dyk (2004) extend propensity score methods to cases with categorical and continuous treatments. 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 W_{it} and condition on these quantities if we could observe them at the beginning of each season. 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. One manifestation of this issue is that a regression of a team's total wins during a season on the sum of its weekly propensity scores generates a regression coefficient that is significantly greater than one. If the weekly propensity scores were calculated only using information determined at the beginning of the season – such

assumption to conduct nonparametric tests of the sharp null hypothesis that wins in year t do not affect outcomes in subsequent years. Second, we add an assumption of additively separable treatment effects that, when combined with the conditional independence assumption, allows us to estimate the causal effects of wins in a framework that combines instrumental variables and propensity score estimators. Third, we adopt an inverse probability weighting approach exploiting sequential conditional independence assumptions from the literature on sequential treatment effects (STE); this approach allows us to relax the additively separable treatment effects assumption in exchange for a stronger common support assumption.

3.1 Conditional Independence Tests

Angrist and Kuersteiner (2011) develop causality tests in the context of monetary policy shocks that rely on a propensity score model and have the correct size regardless of the model determining the outcomes. We conduct tests of the sharp null hypothesis of no causal effect using the same identifying assumptions as Angrist and Kuersteiner (2011), albeit in a very different context. The key insight in both papers is that the conditional independence assumption generates a testable prediction under the sharp null: outcomes in periods $t+1$ and beyond should be independent of treatment assignment in period t after conditioning on the probability of treatment in period t.

To set notation, let team i play S games in season (year) t. The $1 \times S$ row vector $\mathbf{W_{it}}$ contains the outcome of each game for the entire season, with the sth element, W_{ist}, equal to unity if the team wins in week s and zero otherwise. Let W_{it} denote the sum of elements in $\mathbf{W_{it}}$, or total wins for team i in season t. Denote a potential sequence of wins and losses as the $1 \times S$ row vector $\mathbf{w} \in \mathcal{W}$, with w representing the sum of the elements of $\mathbf{w}$. The potential outcome $Y_{i(t+1)}(\mathbf{w})$ is the value of outcome Y for school i in year $t+1$ as a function of the entire series of wins and losses in season t. There are 2^S potential outcomes for team i in season t; in general 2^S ranges from 1,024 to 4,096. A causal effect compares two potential outcomes, $Y_{i(t+1)}(\mathbf{w})$ and $Y_{i(t+1)}(\mathbf{w}')$, so there are $\binom{2^S}{2}$ causal effects that we might consider. Note that $Y_{i(t+1)}(\mathbf{w})$ need

as the previous season's wins – then this regression would generate a coefficient equal to one in expectation.

not equal $Y_{i(t+1)}(\mathbf{w}')$ even if $w = w'$; different sequences with the same number of wins may generate different outcomes. For generality we express our assumptions, propositions, and proofs in terms of a generic potential outcome $Y_{i(t+1)}(\mathbf{w})$, but in our empirical results we use $\Delta Y_{i(t+1)}$ as the outcome to increase precision.[8]

Our test relies on the standard ignorability assumption. Let $\mathbf{X_{ist}}$ represent the set of observable characteristics, measured in week s of year t, that determine W_{ist}. Denote the history of observable characteristics from weeks 1 through s as $\underline{\mathbf{X}}_{\mathbf{ist}} = [\mathbf{X_{i1t}} \ ... \ \mathbf{X_{ist}}]$. These observables include the history of wins and losses through week $s-1$. In our context the ignorability assumption requires the following:

Assumption 1 *Strong Ignorability of Treatment Assignment:*

1. $Y_{i(t+1)}(\mathbf{w}) \perp W_{ist} \mid \underline{\mathbf{X}}_{\mathbf{ist}} = \underline{\mathbf{x}}_{\mathbf{ist}}, \ \forall \ \mathbf{w} \in \mathcal{W}, \ \forall \ s : s \in \{1, ..., S\}, \ \forall \ t : t \in \{1, ..., T-1\}$

2. $0 < P(W_{ist} = 1 \mid \underline{\mathbf{X}}_{\mathbf{ist}} = \underline{\mathbf{x}}_{\mathbf{ist}}) < 1, \ \forall \ \underline{\mathbf{x}}_{\mathbf{ist}} \in \underline{\mathcal{X}}_{ist}, \ \forall \ s : s \in \{1, ..., S\},$
   $\forall \ t : t \in \{1, ..., T-1\}$

The first part of the assumption is the standard conditional independence assumption; it implies that in each week s of the season, winning in week s is independent of the potential outcomes $Y_{i(t+1)}(\mathbf{w})$ when conditioning on the history of observables and wins and losses through week s. The second part of the assumption is the standard common support assumption.

The conditional independence assumption has two unique features in our study. First, $\underline{\mathbf{X}}_{\mathbf{ist}}$ 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 $\underline{\mathbf{X}}_{\mathbf{ist}}$ contains all of the important observables. If there were an observable characteristic x^*_{ist} that predicted W_{ist} and were not included in $\underline{\mathbf{X}}_{\mathbf{ist}}$, then professional bettors could use x^*_{ist} to form predictions of $P(W_{ist} = 1)$ that were superior to those formed by the bookmakers. This discrepancy

[8]Define $\Delta Y_{i(t+1)}(\mathbf{w})$ to be $Y_{i(t+1)}(\mathbf{w}) - Y_{i(t-1)}$. It is trivial to see that if $Y_{i(t+1)}(\mathbf{w})$ is conditionally independent of $\mathbf{W_{it}}$, then $\Delta Y_{i(t+1)}(\mathbf{w})$ is conditionally independent of $\mathbf{W_{it}}$; $Y_{i(t-1)}$ is an observable, so it must be conditionally independent of $\mathbf{W_{it}}$. Thus conditional independence of $Y_{i(t+1)}(\mathbf{w})$ and $\mathbf{W_{it}}$ implies conditional independence of $\Delta Y_{i(t+1)}(\mathbf{w})$ and $\mathbf{W_{it}}$.

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 conditional independence 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). In our setting, the conditional independence assumption would fail if school alumni or applicants conditioned their decisions on some variable x^*_{ist} that affected team performance and was observable to them but not to bookmakers. This seems unlikely. Nevertheless, as a robustness check we reestimate the effects later in the season when team quality becomes well known to all observers, and we find similar results (see Section 4.2).

Denote the conditional probability of winning as $P(W_{ist} = 1 \mid \underline{\mathbf{X}}_{\mathbf{ist}} = \underline{\mathbf{x}}_{\mathbf{ist}}) = p(\underline{\mathbf{x}}_{\mathbf{ist}})$. For binary treatments the ignorability assumption implies that conditioning on the probability of treatment, or propensity score, is sufficient to guarantee independence between treatment assignment and potential outcomes (Rosenbaum and Rubin 1983):

$$Y_{i(t+1)}(\mathbf{w}) \perp W_{ist} \mid p(\underline{\mathbf{x}}_{\mathbf{ist}}), \ \forall \ \mathbf{w}, s, t \tag{2}$$

Under the sharp null hypothesis the treatment has no effect and all potential outcomes are identical. Thus under the null $Y_{i(t+1)}(\mathbf{w}) = Y_{i(t+1)} \, \forall \, \mathbf{w}$. This fact generates the testable implication that $Y_{i(t+1)} \perp W_{ist} \mid p(\underline{\mathbf{x}}_{\mathbf{ist}})$ under the null. Tests of the sharp null based on this implication should have the correct size regardless of the model generating $Y_{i(t+1)}$.[9]

In practice we test whether $\text{Cov}(Y_{i(t+1)}, W_{ist} \mid p(\underline{\mathbf{x}}_{\mathbf{ist}})) = 0$. If we reject this hypothesis, then W_{ist} must have a causal effect on $Y_{i(t+1)}$. To implement the test we first estimate the propensity score. This requires translating the bookmakers' "point spread" – or predicted margin of victory – into the probability of a win. To allow flexibility in the relationship we estimate a logistic regression of W_{ist} on a fifth-order polynomial of the point spread.[10] The results are highly significant; a

[9]The power will of course depend on the alternative hypothesis and the specific test statistic.

[10]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.

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.[11]

After forming $\hat{p}_{ist}$, in each week $s \in \{1, ..., 12\}$ we condition on $\hat{p}_{ist}$ via subclassification and estimate regressions of $\Delta Y_{i(t+1)}$ on W_{ist}. We then test whether the average of those 12 regression coefficients is significantly different than zero. These tests, the *p*-values from which we report in Tables 2 through 7, rely only on strong ignorability (Assumption 1). They do not depend on the data generating process for $Y_{i(t+1)}$ or the potentially dynamic process generating W_{ist}. However, to estimate the magnitude of any causal effects we must impose an additional assumption.

3.2 Instrumental Variables Estimator

Since we separately estimate the effect of a win in each week s for $s = 1, ..., 12$, it is tempting to average the 12 effects to compute the average effect of a win. However, the outcome of a game in week s may reveal information about team quality and affect expectations about the outcomes of games in weeks $s + 1$ and beyond. A win in week s thus correlates with a greater than one unit change in total season wins, w_{it}, even when conditioning on $p(\underline{\mathbf{x}}_{\mathbf{ist}})$. The coefficient in a regression of $Y_{i(t+1)}$ on W_{ist} (while conditioning on $p(\underline{\mathbf{x}}_{\mathbf{ist}})$) thus overstates the causal effect of winning in week s because it includes the effects of an increased likelihood of winning in subsequent weeks as well.

We employ two strategies to estimate the causal effects of wins. The first strategy estimates the effects of a win during each week of the season separately and applies an instrumental variables type estimator to account for the fact that a win early in the season reveals additional information about the quality of the team. The intuition underlying this estimator is straightforward: a win in

[11] Recent papers by Graham et al. (2012) and Imai and Ratkovic (2014) suggest using inverse probability tilting or covariate balancing propensity scores to ensure more exact covariate balance than may be achieved via propensity score matching or weighting. In our case, however, these methods are less relevant, because we do not observe the actual covariates being used to form bookmaker expectations and the corresponding propensity scores.

week s is uncorrelated with potential outcomes (after conditioning on $p(\underline{\mathbf{x}}_{\mathbf{ist}})$) and correlated with total wins, thus satisfying the two conditions of a valid instrument. The strength of this approach is that it is straightforward to ensure common support across observations, but the downside is that we must assume that winning in week s does not modify the causal effect of winning in week $s+1$ (and vice versa). An alternative strategy, presented in Section 3.3, models the entire sequence of wins during a season.

The causal effect on $Y_{i(t+1)}$ of a sequence of wins $\mathbf{w}$ relative to an alternative sequence $\mathbf{w}'$ is $Y_{i(t+1)}(\mathbf{w}) - Y_{i(t+1)}(\mathbf{w}')$. Conditional on wins and losses in other weeks, the causal effect of a win in week s for school i in year t is the difference between the two potential outcomes:

$$\begin{aligned}\beta_{ist}(\mathbf{w}_{\mathbf{i(-s)t}}) = Y_{i(t+1)}(w_{i1t}, ..., w_{i(s-1)t}, 1, w_{i(s+1)t}, ..., w_{iSt}) \\ -Y_{i(t+1)}(w_{i1t}, ..., w_{i(s-1)t}, 0, w_{i(s+1)t}, ..., w_{iSt}).\end{aligned}$$

The vector $\mathbf{w}_{\mathbf{i(-s)t}}$ contains team i's wins in weeks prior to and following week s. There is an asymmetry between wins in weeks prior to s and wins in weeks subsequent to s. Wins in weeks prior to s are predetermined, and we can treat them as we would any other predetermined covariate. Wins in weeks subsequent to s are potentially endogenous and may be affected by W_{ist}. In particular, W_{ist} could have a causal effect on $W_{i(s+1)t}$, or the realization of W_{ist} could reveal information that affects expectations of $W_{i(s+1)t}$. The data imply that one or both of these factors is present; winning in week s is associated with a higher probability of winning in later weeks, even after conditioning on $p(\underline{\mathbf{x}}_{\mathbf{ist}})$.

To estimate the causal effects of wins we assume that the effects of wins across different weeks of the same season are additively separable. This reduces the dimensionality of the parameter space that we consider to a manageable degree, as β_{ist} is no longer a function of $\mathbf{w}_{\mathbf{i(-s)t}}$. Let $\mathbf{0}$ be a $1 \times S$ row vector containing zeroes and $\boldsymbol{\beta_{it}}$ be a $S \times 1$ column vector containing β_{ist} in row s.

Assumption 2 *Additively Separable Treatment Effects:* $Y_{i(t+1)}(\mathbf{w}) = Y_{i(t+1)}(\mathbf{0}) + \mathbf{w}\boldsymbol{\beta_{it}}$.

Assumption 2 allows the effect of winning to vary arbitrarily by team, season, and week, but it does not allow for interactions of these effects across different weeks. This assumption is a form of the Stable Unit Treatment Value Assumption (SUTVA; Rubin 1980) in that a team's performance

in week s' is assumed not to modify the causal effect of its performance in week s. It would be violated if, for example, the effect of winning in week 2 depended on whether a team won in week 1. While Assumption 2 is not guaranteed to hold, it is less restrictive than the linear functional form assumptions imposed when interpreting regression coefficients, as is standard in this literature. If Assumption 2 applies, then we can define an instrumental variables (IV) estimator that converges to a weighted average of treatment effects in different games.

We define our IV estimator as follows. First estimate the "reduced form" effect of winning in week s on $Y_{i(t+1)}$:

$$\pi_{p(\underline{\mathbf{x}}_{\mathbf{ist}})} = E[Y_{i(t+1)} \mid W_{ist} = 1, p(\underline{\mathbf{x}}_{\mathbf{ist}})] - E[Y_{i(t+1)} \mid W_{ist} = 0, p(\underline{\mathbf{x}}_{\mathbf{ist}})] \tag{3}$$

If there were no relationship between winning in week s and winning in future weeks, then $\pi_{p(\underline{\mathbf{x}}_{\mathbf{ist}})}$ would represent the causal effect of winning in week s under Assumption 1. However, we must account for the fact that winning in week s is associated with a higher probability of winning in future weeks. Thus we estimate the "first stage" relationship between winning in week s and winning in future weeks:

$$\gamma_{p(\underline{\mathbf{x}}_{\mathbf{ist}})} = \sum_{j=s+1}^{S} [P(W_{ijt} = 1 \mid W_{ist} = 1, p(\underline{\mathbf{x}}_{\mathbf{ist}})) - P(W_{ijt} = 1 \mid W_{ist} = 0, p(\underline{\mathbf{x}}_{\mathbf{ist}}))] \tag{4}$$

Finally, combine the two estimates to generate:

$$\beta_{p(\underline{\mathbf{x}}_{\mathbf{ist}})} = \frac{\pi_{p(\underline{\mathbf{x}}_{\mathbf{ist}})}}{1 + \gamma_{p(\underline{\mathbf{x}}_{\mathbf{ist}})}} \tag{5}$$

This estimator is identical to an IV estimator in which W_{ist} is the instrument (which is exogenous after conditioning on $p(\underline{\mathbf{x}}_{\mathbf{ist}})$), and $\sum_{j=s}^{S} W_{ijt}$ is the endogenous treatment to be instrumented.

The IV estimator is valid regardless of whether the correlation between current wins and future wins arises because W_{ist} has a causal effect on W_{ijt} (for $j > s$) or because the realization of W_{ist} reveals information that affects expectations of W_{ijt}. Nevertheless, when interpreting the IV estimator it is helpful to draw a formal distinction between these two possibilities. Using the language of Angrist et al. (1996) we define "compliers" as teams for whom winning in week s has a (positive) causal effect on winning in week j, changing W_{ijt} from zero to one. Let ρ^{c}_{sj} denote

the share of teams that are week j compliers. All other teams are "noncompliers" in the sense that winning in week s has no causal effect on winning in week j. As is standard, we rule out "defiers" (i.e. losing in week s cannot cause a team to win in week j).

Proposition 1 *Let* C *(*NC*) denote the event that team i in year t is (is not) a complier for weeks s and j ($j > s$). Assume that the causal effect of W_{ist} on W_{ijt} is weakly positive. Under Assumptions 1 and 2,*

$$\begin{aligned} & E[Y_{i(t+1)} \mid W_{ist} = 1, p(\underline{\mathbf{x}}_{\mathbf{ist}})] - E[Y_{i(t+1)} \mid W_{ist} = 0, p(\underline{\mathbf{x}}_{\mathbf{ist}})] \\ = \;& E[\beta_{ist} \mid p(\underline{\mathbf{x}}_{\mathbf{ist}})] + \sum_{j=s+1}^{S} (\rho_{sj}^{c} E[\beta_{ijt} \mid p(\underline{\mathbf{x}}_{\mathbf{ist}}), \mathrm{C}] + (1 - \rho_{sj}^{c}) E[\beta_{ijt} \mid p(\underline{\mathbf{x}}_{\mathbf{ist}}), \mathrm{NC}] \\ & \cdot \{P(W_{ijt} = 1 \mid W_{ist} = 1, p(\underline{\mathbf{x}}_{\mathbf{ist}}), \mathrm{NC}) - P(W_{ijt} = 1 \mid W_{ist} = 0, p(\underline{\mathbf{x}}_{\mathbf{ist}}), \mathrm{NC})\}) \end{aligned}$$

Thus,

$$\frac{\pi_{p(\underline{\mathbf{x}}_{\mathbf{ist}})}}{1 + \gamma_{p(\underline{\mathbf{x}}_{\mathbf{ist}})}} = \frac{E[\beta_{ist} \mid p(\underline{\mathbf{x}}_{\mathbf{ist}})] + \sum_{j=s+1}^{S} (\rho_{sj}^{c} E[\beta_{ijt} \mid p(\underline{\mathbf{x}}_{\mathbf{ist}}), \mathrm{C}] + (1 - \rho_{sj}^{c}) \mu_{isjt} E[\beta_{ijt} \mid p(\underline{\mathbf{x}}_{\mathbf{ist}}), \mathrm{NC}])}{1 + \sum_{j=s+1}^{S} (\rho_{sj}^{c} + (1 - \rho_{sj}^{c}) \mu_{isjt})},$$

where

$$\mu_{isjt} = P(W_{ijt} = 1 \mid W_{ist} = 1, p(\underline{\mathbf{x}}_{\mathbf{ist}}), \mathrm{NC}) - P(W_{ijt} = 1 \mid W_{ist} = 0, p(\underline{\mathbf{x}}_{\mathbf{ist}}), \mathrm{NC}).$$

Proof: See Appendix A.

In this weighted average of treatment effects, the weight μ_{isjt} represents the increase in the probability of winning in week j that is associated with a win in week s (conditional on $p(\underline{\mathbf{x}}_{\mathbf{ist}})$). These weights are higher for two types of teams: compliers and teams with greater uncertainty about their quality. For compliers μ_{isjt} is set to one by definition because winning in week s causes W_{ijt} to change from zero to one. For noncompliers μ_{isjt} will be greater if a win reveals more information about team quality, and it typically decreases with s because wins late in the season reveal less new information about team quality than wins early in the season. Compliers receive a higher weight than noncompliers, but unlike in a typical IV setting the noncompliers do not receive zero weight. If winning in prior weeks never has a causal effect on winning in subsequent weeks, then there are no compliers, $\rho_{sj}^{c} = 0$, and the sum of weights simplifies to $1 + \sum_{j=s+1}^{S} \mu_{isjt}$.

The mechanics of the IV estimator are straightforward. For simplicity suppose that wins in all weeks have an equal effect, β. In that case we may rescale the observed effect of winning in week s by the average increase in total season wins associated with winning in week s. For example, if a win in week 3 is associated with an average of 1.5 additional wins during the entire course of a season (including the week 3 win itself, and after conditioning on $p(\underline{\mathbf{x}}_{\mathbf{ist}})$), then we would rescale the estimated effect of a win in week 3 by $1/1.5 = 0.67$. In reality effects may vary by school, season, and week, and in that case the IV estimator converges to a weighted average of heterogeneous treatment effects.

Estimation proceeds in three steps. First, for each game we estimate the propensity score $p(\underline{\mathbf{x}}_{\mathbf{ist}})$ using the published bookmakers' spread (see Section 3.1). To ensure common support we trim all observations with $\hat{p}_{ist}$ less than 0.05 or greater than 0.95. Next, for each week s we stratify the sample into 12 bins based on the estimated propensity score (Dehejia and Wahba 1999).[12] Within each bin we estimate two linear regressions: a regression of the outcome $Y_{i(t+1)}$ on W_{ist} and a regression of the expected number of wins for the remainder of the season ($p_{ist}^{post} = \sum_{j=s+1}^{S} p(\underline{\mathbf{x}}_{\mathbf{ijt}})$) on W_{ist}. The first regression estimates the "reduced form" coefficient $\pi_{p(\underline{\mathbf{x}}_{\mathbf{ist}})}$ (Equation 3), and the second regression estimates the "first stage" coefficient $\gamma_{p(\underline{\mathbf{x}}_{\mathbf{ist}})}$ (Equation 4). 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 $\hat{\beta}_{p(\underline{\mathbf{x}}_{\mathbf{ist}})} = \hat{\pi}_{p(\underline{\mathbf{x}}_{\mathbf{ist}})}/(1+\hat{\gamma}_{p(\underline{\mathbf{x}}_{\mathbf{ist}})})$ (Equation 5). We then compute $\hat{\beta}$ as the weighted average of $\hat{\beta}_{p(\underline{\mathbf{x}}_{\mathbf{ist}})}$ across all 144 bin-by-week combinations (12 bins per week by 12 weeks per season), weighting each estimate by the relevant sample size.[13] Appendix B provides implementation details.

[12] 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 in Appendix B). Thus we limited the number of bins to 12.

[13] 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

3.3 Sequential Treatment Effects Model

The additively separable treatment effects assumption (Assumption 2) accommodates arbitrary treatment effect heterogeneity by school, season, and week. Nevertheless, it does not accommodate the possibility that the effects of winning in week s may depend on whether or not a team wins in week j ($j \neq s$). To accommodate this possibility we consider a dynamic model of sequential treatment effects (Lechner 2009; Lechner and Miquel 2010). This model allows us to relax Assumption 2; its downside is that given the long series of treatments within a season, it is impractical to guarantee common support across observations.

Recall that the row vector $\mathbf{w}$ represents a potential sequence of wins and losses in a season, and the quantity w represents the sum of the elements of $\mathbf{w}$ (i.e. total season wins). There are 2^S potential outcomes for team i in season t, $Y_{i(t+1)}(\mathbf{w})$, and $\binom{2^S}{2}$ causal effects that we might consider. For tractability we focus on the average potential outcome for the set of sequences that contain $w = k$ wins,

$$\frac{1}{r_k} \sum_{\mathbf{w} \ni w=k} E[Y_{i(t+1)}(\mathbf{w})]. \tag{6}$$

In this expression r_k is the number of unique possible sequences with exactly k wins. Equation (6) allows for heterogeneity in the average outcome for different sequences with the same number of wins, but it does not attempt to separately estimate the average outcome for each sequence. Estimating the average outcome for each sequence would be infeasible with our data, as the number

weighting leads to efficient estimation of the average treatment effect. We apply an inverse probability weighting estimator in our sequential treatment effects model in Section 3.3. However, the weighting estimator does not apply directly to our IV estimator because we need to estimate and combine two separate coefficients, both of which may vary with the propensity score. In particular, the weighting estimator does not allow us to estimate different coefficients for different values of the propensity score. If we simply estimate the "reduced-form" effect of W_{ist} on $Y_{i(t+1)}$ (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.

of potential outcomes exceeds the number of observations in our data set.

Estimating Equation (6) for $k = 0, ..., 12$ requires the weak dynamic conditional independence assumption (W-DCIA). W-DCIA includes a conditional independence assumption, applied to the case of a dynamically determined sequence of treatments, and a common support assumption. If we assume that the history of observable characteristics through week s, $\underline{\mathbf{X}}_{\mathbf{ist}}$, includes previous wins and losses (which it almost surely must, as these quantities are readily observable and are clearly informative about team quality), then W-DCIA is equivalent to the ignorability assumption (Assumption 1).[14]

Under random assignment of wins we could estimate Equation (6) using the simple average

$$\bar{y}_{i(t+1)k} = \frac{\sum_{i,t} Y_{i(t+1)} \cdot 1\{W_{it} = k\}}{\sum_{i,t} 1\{W_{it} = k\}}.$$

To address selection we apply inverse probability weights (IPW) to construct estimators of the form

$$\bar{y}_{i(t+1)k} = \frac{1}{q_{it}(k)} \sum_{i,t} \frac{Y_{i(t+1)} \cdot 1\{W_{it} = k\}}{\hat{p}(\mathbf{w_{it}})};\ \ q_{it}(k) = \sum_{i,t} \frac{1\{W_{it} = k\}}{\hat{p}(\mathbf{w_{it}})}. \tag{7}$$

The weights $\hat{p}(\mathbf{w_{it}})$ represent the estimated probability of observing sequence $\mathbf{w_{it}}$, and if W-DCIA holds then the IPW solve the selection problem (Lechner 2009). The $q_{it}(k)$ terms normalize the weights to sum to one. Following Robins et al. (2000) and Lechner (2009) we express the joint

[14]Following Lechner (2009), W-DCIA in our context requires:

1. $Y_{i(t+1)}(\mathbf{w}) \perp W_{i1t} \mid \mathbf{X_{i1t}} = \mathbf{x_{i1t}}\ \forall\ \mathbf{w} \in \mathcal{W}$;
   $Y_{i(t+1)}(\mathbf{w}) \perp W_{ist} \mid \underline{\mathbf{X}}_{\mathbf{ist}} = \underline{\mathbf{x}}_{\mathbf{ist}}, W_{i1t} = w_{i1t}, ..., W_{i(s-1)t} = w_{i(s-1)t}\ \forall\ \mathbf{w} \in \mathcal{W}$
   $\forall\ s : s \in \{2, ..., S\}$

2. $0 < P(W_{i1t} = 1 \mid \mathbf{X_{i1t}} = \mathbf{x_{i1t}}) < 1\ \forall\ \mathbf{x_{i1t}} \in \mathcal{X}_{i1t}$;
   $0 < P(W_{ist} = 1 \mid \underline{\mathbf{X}}_{\mathbf{ist}} = \underline{\mathbf{x}}_{\mathbf{ist}}, W_{i1t} = w_{i1t}, ..., W_{i(s-1)t} = w_{i(s-1)t}) < 1\ \forall\ \underline{\mathbf{x}}_{\mathbf{ist}} \in \underline{\mathcal{X}}_{ist}$
   $\forall\ s : s \in \{2, ..., S\}$

These conditions follow from Assumption 1 as long as $\underline{\mathbf{X}}_{\mathbf{ist}}$ includes $W_{i1t}, ..., W_{i(s-1)t}$.

probability that $\mathbf{W_{it}} = \mathbf{w_{it}}$ as the product of the sequential transition probabilities:

$$p(\mathbf{w_{it}}) = P(W_{i1t} = w_{i1t} \mid \mathbf{X_{i1t}} = \mathbf{x_{i1t}}) \cdot P(W_{i2t} = w_{i2t} \mid \underline{\mathbf{X}}_{\mathbf{i2t}} = \underline{\mathbf{x}}_{\mathbf{i2t}}, W_{i1t} = w_{i1t})$$
$$\cdot \ldots \cdot P(W_{iSt} = w_{iSt} \mid \underline{\mathbf{X}}_{\mathbf{iSt}} = \underline{\mathbf{x}}_{\mathbf{ist}}, W_{i1t} = w_{i1t}, ..., W_{i(S-1)t} = w_{i(S-1)t})$$

This formulation concurs with our data in that for each week of the season we observe the estimated probability of winning conditional on the events that occur up until that week. It is thus straightforward to construct $1/\hat{p}(\mathbf{w_{it}})$, our estimated IPW, and estimate Equation (7) for $k = 0, ..., 12$.

We exploit Equation (7) to generate estimates in two ways. First, we estimate nonparametric local polynomial regressions of $\Delta Y_{i(t+1)}$ on W_{it}, weighting each observation by $1/\hat{p}(\mathbf{w_{it}})$. We present these estimates graphically. Then, to summarize these local polynomial regressions, we estimate linear regressions of $\Delta Y_{i(t+1)}$ on W_{it}, again weighting each observation by $1/\hat{p}(\mathbf{w_{it}})$:

$$\Delta Y_{i(t+1)} = \beta_0 + \beta_1 W_{it} + \beta_2 W_{i(t-2)} + \beta_3 S_{it} + \beta_4 S_{i(t-2)} + \phi_{t+1} + \varepsilon_{i(t+1)}, \quad (8)$$
$$\text{with weights } a_{it} = \frac{1}{\sqrt{\hat{p}(\mathbf{w_{it}})}}.$$

We present the results of these regressions in our tables. Although the IPW solve the selection problem, the regression results represent minimum mean squared error linear approximations to the potentially nonlinear dose-response function relating $\Delta Y_{i(t+1)}$ and W_{it} (represented in the figures).

The key issue in applying our sequential treatment effects model is that for many potential outcomes we observe little or no data. There are 2^S possible sequences – in most seasons 2^S lies between 1,024 and 4,096 – and less than 2,000 observations. For the majority of sequences $\mathbf{w} \in \mathcal{W}$, the set of observations with $\mathbf{w_{it}} = \mathbf{w}$ is empty or a singleton. Furthermore, with so few observations per sequence it is unlikely that the common support assumption holds when, for example, comparing an observation with a sequence with one win to an observation with a sequence with ten wins. Thus, although we trim observations with extreme weights to help induce common support, our estimates are best interpreted as the average effect of a marginal change in win percentage.[15] The estimated dose-response relationships do not necessarily indicate what would happen to a team that exogenously changes from one win to ten wins.

[15]Specifically, we trim all observations with IPW that are above the 90th percentile. Trimming at the 95th percentile instead generates qualitatively similar results.

4 Propensity Score Results

Table 2 presents results demonstrating balance in pre-treatment outcomes between winning and losing teams when conditioning on the propensity score. Each row reports the results for a different dependent variable. The first set of columns in Table 2 present results from our IV model (Equation 5). To match the main specification we use in our results we specify the outcome in differences ($\Delta Y_{i(t+1)}$). However, instead of estimating the relationship between wins in year t and outcomes in year $t+1$, we estimate the relationship between wins in year $t+2$ and outcomes in year t.[16] Significant results in Table 2 would imply that current trends in outcomes predict future wins, even after conditioning on the propensity score. This would represent a violation of Assumption 1. The results reveal no significant relationship between any outcome and future wins, nor a consistent trend in the signs of the coefficients (seven of the ten coefficients are in the "negative" direction, implying that an unfavorable trend in the outcome weakly correlates with future wins).

The second set of columns in Table 2 presents results from estimating the sequential treatment effects models (Equation 8). Again we estimate the relationship between wins in year $t+2$ and outcomes in year t (while conditioning on the propensity score via inverse probability weights). There is still no significant relationship between any outcome and future wins nor a consistent trend in the signs of the coefficients. Overall the results in Table 2 suggest that winning and losing schools are comparable in pre-trends after conditioning on the propensity score.

4.1 Baseline Results

Table 3 reports the results of estimating the effects of winning using all FBS ("Division I-A") schools. In the first set of columns we estimate the IV model (Equation 5). We specify the outcome in differences ($\Delta Y_{i(t+1)}$) instead of levels because differencing removes much of the unexplained cross-school variation in $Y_{i(t+1)}$, generating more precise estimates. Furthermore, differencing can

[16] Recall that we use two year differences because some variables are measured on the academic or fiscal year, while others are measured on the calendar year. Thus we set a two year gap between wins and outcomes for the placebo tests as well.

help eliminate any bias that remains after conditioning on the propensity score (Smith and Todd 2005). We report p-values that test the sharp null hypothesis of no relationship between wins in year t and outcomes in year $t+1$. As noted in Section 3.1 the validity of these tests relies *only* on strong ignorability (Assumption 1) and does not depend on the data generating process for $Y_{i(t+1)}$ or the potential dynamics in the process generating W_{ist}.

Estimates from the IV model 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. The point estimate on non-athletic donations is positive, however, suggesting that the increase in athletic donations does not "crowd out" non-athletic donations. 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).

The large number of outcomes tested raises the issue of multiple hypothesis testing. We address this issue by reporting "q-values" that control the False Discovery Rate (FDR) across all results tables (Tables 3 through 7) alongside the 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 should represent false discoveries. To calculate FDR q-values we use the "sharpened" FDR control algorithm from Benjamini et al. (2006), implemented in Anderson (2008). All of our significant results in Table 3 remain significant when controlling FDR at the $q = 0.10$ level.

The second set of columns in Table 3 reports estimates from the sequential treatment effects model (Equation 8). There is a reasonable degree of concordance between the IV model and STE model results, which increases our confidence in both. The STE model estimates that an additional win increases alumni athletic donations by $191,200 ($t = 2.9$). Focusing on other results for which we can reject the sharp null hypothesis based on our conditional independence tests, we find that academic reputation increases by 0.003 points, applications increase by 81 applicants, acceptance

rates decrease by 0.3 percentage points, in-state enrollment increases by 13 students, and the 25th percentile SAT score increases by 0.8 points. Most of these estimates are of similar magnitude to our IV estimates; the two exceptions are effects on applicants (135 versus 81) and effects on the 25th percentile SAT score (1.8 versus 0.8). The confidence intervals for these results are wider, and we would not be able to reject the null hypothesis if we only examined the STE model *t*-statistics.

An advantage of the STE model over the IV model is that we can consider the potential non-linearity of the functions relating wins to outcomes. Figure 1 plots local polynomial regressions relating outcomes to changes in wins. Each panel presents a different outcome, and we weight each observation by its IPW to address selection. Panel 1 reveals a strong, generally linear relationship between wins and alumni athletic operating donations. The relationships for applicants (Panel 6) and in-state enrollment (Panel 9) also appear approximately linear. The relationships for the acceptance rate (Panel 7) and 25th percentile SAT score (Panel 10), however, display concavity, although the confidence intervals are too wide to reach strong conclusions; in both cases a straight line would fit within the regions defined by the confidence intervals. Nevertheless, it is notable that SAT scores exhibit both the largest divergence between the IV and STE estimates and the strongest evidence of concavity in the nonparametric regressions. This pattern suggests that for SAT scores, the additive separability assumption (Assumption 2) may not be appropriate.

Tables 4 and 5 report IV and STE model results estimated separately for BCS teams and non-BCS teams respectively. For most outcomes the estimated effect for BCS teams (Table 4) is larger (or more positive) than the estimated effect for non-BCS teams (Table 5). However, for applications, in-state enrollment, and SAT scores, there is a larger effect among non-BCS schools than among BCS schools in one or both models. All three of 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.[17] Nevertheless, the effects for non-BCS schools generally fail to achieve statistical significance, so it is difficult to

[17]Another interesting stratification would be to compare results for public and private institutions. However, there are too few private institutions playing FBS football (17) to reliably estimate the propensity score design.

draw any firm conclusions.

4.2 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. It is also the case that early season games tend to be against out-of-conference opponents that are specifically selected because they are weak (Thamel 2006). As a robustness check we estimate the IV and STE models while excluding the first month of the season. Table 6 reports these results.

In most cases the estimates change little relative to Table 3 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 several effects fail to achieve significance at the $q = 0.10$ level after controlling FDR. Nevertheless, the similarity in the IV point estimates between Tables 3 and 6 for outcomes that are statistically significant in the baseline results is reassuring. There is somewhat more variability in the STE point estimates; the effects for athletic operating donations, academic reputation, and applicants fall by an average of 40%, while the effects for the acceptance rate, in-state enrollment, and 25th percentile SAT score rise by an average of 84%. Nevertheless, the overall average effect for these outcomes in the STE model is almost identical between Tables 3 and 6 (increasing 5%).

4.3 Unexpected Wins vs. Unexpected Losses

The effects that we estimate by definition represent the effects of unexpected wins and losses. Prospect theory and previous research suggest that individuals may respond differently to unexpected losses and unexpected wins. For example, Card and Dahl (2011) find that domestic violence increases following unexpected NFL losses but does not decrease following unexpected NFL wins.

To explore this possibility we separately estimate Equation (5) for games in which the propensity score is greater than 0.5 and games in which the propensity score is less than 0.5.[18] In the former case oddsmakers expect the team to win, while in the latter case they expect the team to lose.

Appendix Table A1 presents results for these two separate samples. The first set of columns estimates the effect of avoiding unexpected losses (the sample is limited to games in which the propensity score exceeds 0.5), while the second set of columns estimates the effect of achieving unexpected wins (the sample is limited to games in which the propensity score is less than 0.5). There is no consistent pattern when comparing the two sets of estimates. Unexpected wins appear to have a larger impact on alumni measures, while avoiding unexpected losses may have a larger impact on incoming student measures. The differences, however, are not statistically significant. Overall there is little evidence suggesting that the effects of winning vary strongly with expectations.

4.4 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 $Y_{i(t+2)}$ by simply replacing $Y_{i(t+1)}$ with $Y_{i(t+2)}$ in the "reduced-form" equation, Equation (3). 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 $Y_{i(t+2)}$ may thus result from increased wins in year $t+1$, $W_{i(t+1)}$.

We use the following procedure to estimate the effect of W_{it} on $Y_{i(t+2)}$ while controlling for $W_{i(t+1)}$. First, we replace $Y_{i(t+1)}$ with $Y_{i(t+2)}$ in Equation (3) and estimate our IV model. Denote

[18] It makes less sense to estimate the sequential treatment effects model in this context because the STE model estimates the effect of an entire sequence of wins and losses, rather than a single win or loss.

this estimate as $\hat{\psi}$; $\hat{\psi}$ estimates the "reduced-form" relationship between W_{it} and $Y_{i(t+2)}$ without controlling for changes in $W_{i(t+1)}$. Next, we estimate the relationship between W_{it} and $W_{i(t+1)}$. To do this we replace $Y_{i(t+1)}$ with $W_{i(t+1)}$ in Equation (3) and estimate our IV model. Denote this estimate as $\hat{\lambda}$; $\hat{\lambda}$ estimates the "reduced-form" relationship between W_{it} and $W_{i(t+1)}$ (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 $Y_{i(t+1)}$ (estimated in Section 4.1). In short, we adjust the reduced-form effect of W_{it} on $Y_{i(t+2)}$ to account for the fact that $W_{i(t+1)}$ is increasing in W_{it} (i.e. $\hat{\lambda}$) and that $W_{i(t+1)}$ affects $Y_{i(t+2)}$ (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 3 (reproduced in the first set of columns). None are statistically significant at any level after controlling FDR. The effects of winning in year t appear to be concentrated in the following year.

5 Conclusions

For FBS schools, our results reveal that 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 three games (the approximate difference between the median team and an 85th percentile team). Changes of this magnitude occur approximately 20% of the time over a one-year period and 27% of the time over a two-year period. This school may expect alumni athletic donations to increase by $409,000 (17%), applications to increase by 406 (3%), the acceptance rate to drop by 0.9 percentage points (1.3%), in-state enrollment to increase by 46 students (1.8%), and incoming 25th percentile SAT scores to

increase by 2.4 points (0.2%).[19] These estimates are approximately twice as large as comparable estimates from the existing literature. For example, among studies that used fixed effects with panel data, a 3-win increase in team performance was associated with a 1.5% to 1.7% increase in applications (Murphy and Trandel 1994; Pope and Pope 2009), while our estimates suggest a 3% increase.

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 0.6 points. These effects seem too modest by themselves to justify 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 surely 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. In particular, increased expectations may diminish the return to achieving any given level of success. 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

[19]For the SAT score calculation we use the STE coefficient rather than the IV coefficient, as there is suggestive evidence that the additive separability assumption (Assumption 2) fails for SAT scores (see Section 4.1).

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 to the detriment 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.

References

Michael 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, December 2008.

Joshua D. Angrist and Guido 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, August 2011.

Joshua D. Angrist, Guido W. Imbens, and Donald B. Rubin. Identification of causal effects using

instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455, June 1996.

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

Yoav Benjamini, Abba M Krieger, and Daniel Yekutieli. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507, September 2006.

Dale Bremmer and Randall Kesselring. The advertising effect of university athletic success: A reappraisal of the evidence. *Quarterly Review of Economics and Finance*, 33(4):409–421, Winter 1993.

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

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

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

Joe Drape and Katie Thomas. As colleges compete, major money flows to minor sports. *The New York Times*, September 2, 2010.

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

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

Bryan S. Graham, Cristine Campos De Xavier Pinto, and Daniel Egel. Inverse Probability Tilting for Moment Condition Models with Missing Data. *The Review of Economic Studies*, 79(3): 1053–1079, July 2012.

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

Keisuke Hirano and Guido Imbens. The propensity score with continuous treatments. In Andrew Gelman and Xiao-Li 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.

Keisuke Hirano, Guido Imbens, and Geert Ridder. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4):1161–1189, July 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, April 2007.

Kosuke Imai and Marc Ratkovic. Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):243–263, January 2014.

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

Gary Johnson. NCAA football attendance reaches new heights again, January 2012. URL `http://www.ncaa.org/wps/wcm/connect/public/NCAA/Resources/Latest+News/2012/January/NCAA+football+attendance+reaches+new+heights+again`.

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

Michael Lechner. Sequential causal models for the evaluation of labor market programs. *Journal of Business and Economic Statistics*, 27(1):71–83, January 2009.

Michael Lechner and Ruth Miquel. Identification of the effects of dynamic treatments by sequential conditional independence assumptions. *Empirical Economics*, 39(1):111–37, August 2010.

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

Robert Litan, Jonathan Orszag, and Peter Orszag. The empirical effects of collegiate athletics: An interim report. Technical report, Sebago Associates, 2003.

Robert E. McCormick and Maurice Tinsley. Athletics versus academics? Evidence from SAT Scores. *Journal of Political Economy*, 95(5):1103–1116, October 1987.

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

Frank G. Mixon Jr, Len J. Treviño, and Taisa C. Minto. Touchdowns and test scores: exploring the relationship between athletics and academics. *Applied Economics Letters*, 11(7):421–424, 2004.

Robert G. Murphy and Gregory 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, September 1994.

NCAA. NCAA recruiting facts, 2014. URL `http://www.ncaa.org/sites/default/files/Recruiting%20Fact%20Sheet%20WEB.pdf`.

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

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

Thomas A. Rhoads and Shelby Gerking. Educational contributions, academic quality, and athletic success. *Contemporary Economic Policy*, 18(2):248–258, April 2000. ISSN 1465-7287.

James M. Robins, Miguel Angel Hernan, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–60, September 2000.

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

Donald Rubin. Randomization analysis of experimental data: The Fisher Randomization Test comment. *Journal of the American Statistical Association*, 75(371):591–593, September 1980.

Lee Sigelman and Samual 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.

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

Jeff Smith and Petra Todd. Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125(1-2):305–353, March–April 2005.

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

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

Pete Thamel. In college football, big paydays for humiliation. *The New York Times*, 23 August 2006.

Katie Thomas. Call to curb athletic spending strikes some as unrealistic. *The New York Times*, October 27, 2009a.

Katie Thomas. SUNY weighs the value of Division I sports. *The New York Times*, November 16, 2009b.

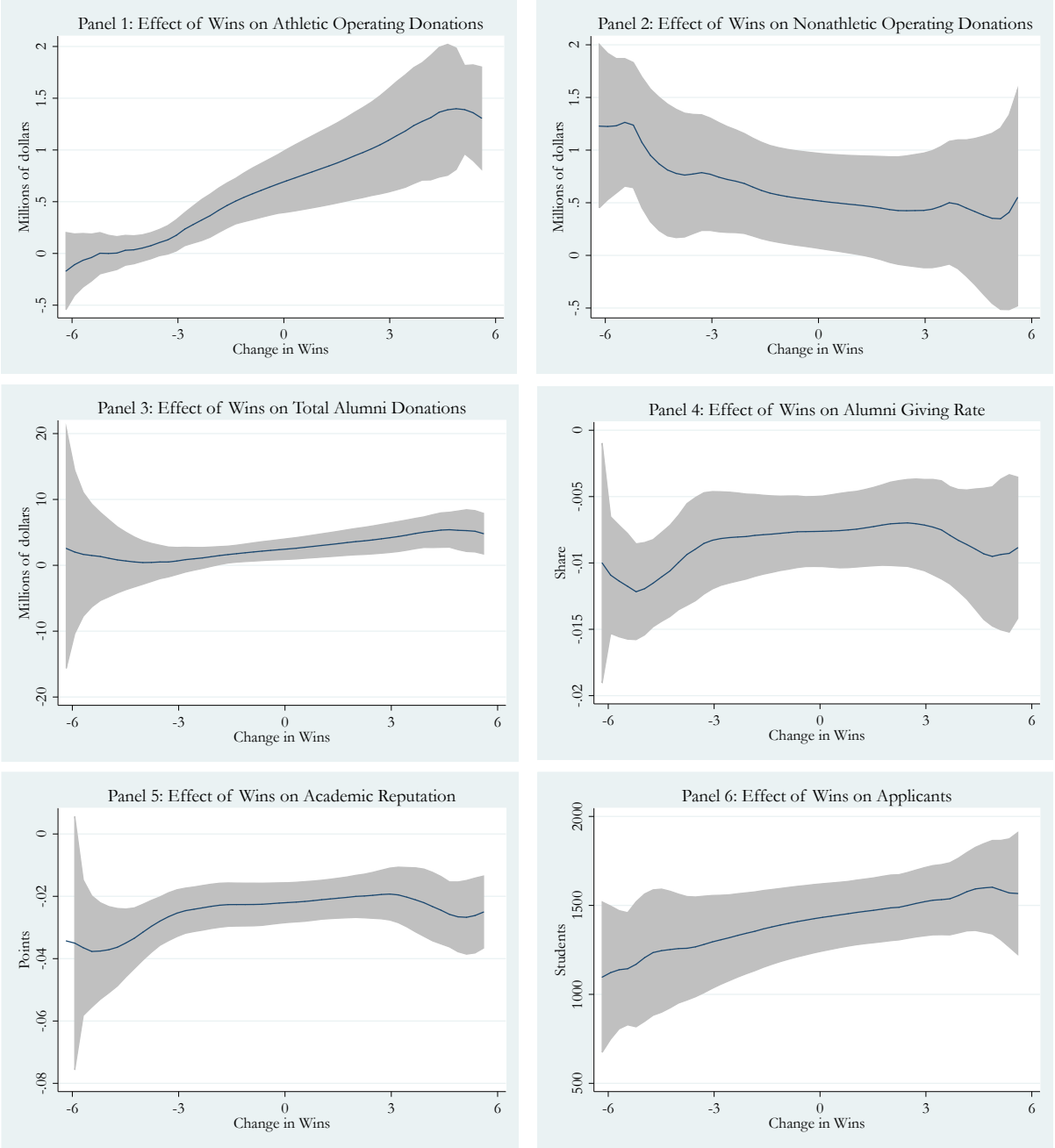
Irvin B. Tucker. A reexamination of the effect of big-time football and basketball success on graduation rates and alumni giving rates. *Economics of Education Review*, 23(6):655–661, December 2004.

Irvin B. Tucker. Big-time pigskin success: Is there an advertising effect? *Journal of Sports Economics*, 6(2):222–229, May 2005.

Irvin B. Tucker and Louis Amato. Does big-time success in football or basketball affect SAT scores? *Economics of Education Review*, 12(2):177–181, June 1993.

Sarah Elizabeth Turner, Lauren A Meserve, and William G Bowen. Winning and giving: Football results and alumni giving at selective private colleges and universities. *Social Science Quarterly*, 82(4):812–826, December 2001.

Figure 1: Effects of Wins on Outcomes



Notes: Figures plot local polynomial regressions of $\Delta Y_{i(t+1)}$ on ΔW_{it}, weighting by $1/\hat{p}(\mathbf{w_{it}})$. Both $\Delta Y_{i(t+1)}$ and games won are residualized with respect to lagged wins, season games, lagged season games, and time effects.

Figure 1 Continued: Effects of Wins on Outcomes

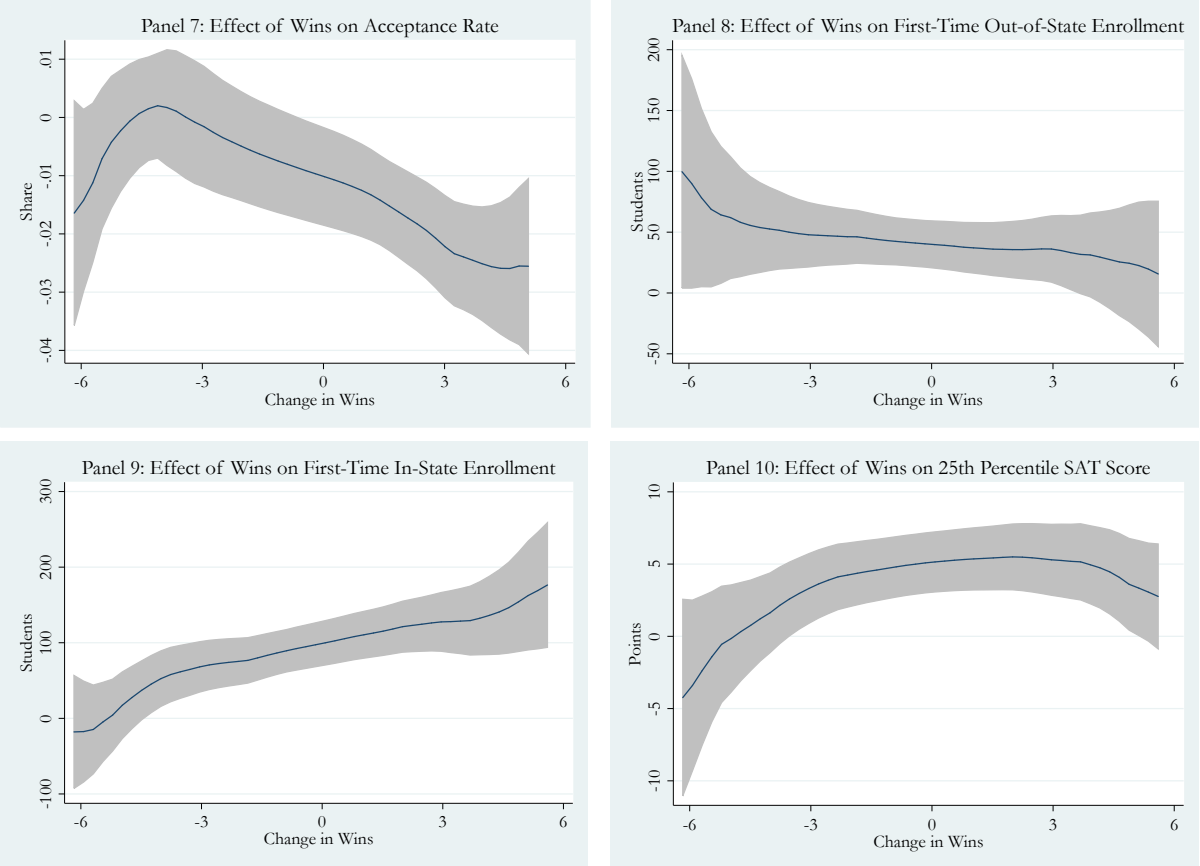


Table 1: Summary Statistics

	BCS			Non-BCS			Dates	
Variable:	Mean	N	Teams	Mean	N	Teams	Available	Source
Season Wins	5.9	1,437	63	4.6	923	53	1987–2009	Covers.com
	(2.6)			(2.5)				
Season Games	10.8	1,437	63	10.5	923	53	1987–2009	Covers.com
	(0.8)			(1.0)				
Expected Wins	5.8	1,437	63	4.7	923	53	1987–2009	Covers.com
	(1.9)			(1.7)				
Alumni Athletic Operating Donations	$3,953	495	54	$695	430	48	1999–2009	VSE
	(3,805)			(836)				
Alumni Nonathletic Operating Donations	$11,600	495	54	$1,992	430	48	1999–2009	VSE
	(14,800)			(3,569)				
Total Alumni Donations	$27,600	1,084	63	$5,359	709	50	1990–2009	VSE
	(30,900)			(6,751)				
Alumni Giving Rate	0.167	1,104	63	0.103	733	51	1990–2009	VSE
	(0.077)			(0.061)				
Academic Reputation	3.499	679	62	2.712	365	43	1997–2008	US News
	(0.54)			(0.43)				
Applicants	16,815	480	63	9,660	360	51	1999–2007	IPEDS
	(8,043)			(6,403)				
Acceptance Rate	0.667	1,036	62	0.755	555	44	1988–2008	US News
	(0.185)			(0.158)				
First-Time Out-of-State Enrollment	1,038	886	63	461	612	53	1986–2008	IPEDS
	(591)			(525)				
First-Time In-State Enrollment	2,811	886	63	2,150	612	53	1986–2008	IPEDS
	(1,537)			(1,073)				
25th Percentile SAT	1,101	431	62	984	287	48	1999–2008	IPEDS
	(103)			(106)				

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

Table 2: Placebo Effects of Future Wins on Current Outcomes

	IV Model		STE Model		Sharp null
Outcome:	Coefficient	N	Coefficient	N	p-val
Alumni Athletic Operating Donations	95.0	485	-44.8	477	0.193
	(72.5)		(59.2)		
Alumni Nonathletic Operating Donations	-16.2	485	16.1	477	0.934
	(196.9)		(89.6)		
Total Alumni Donations	261.1	1,085	291.3	1,094	0.402
	(310.1)		(341.8)		
Alumni Giving Rate	-0.0002	1,095	0.0000	1,106	0.676
	(0.0006)		(0.0007)		
Academic Reputation	-0.002	615	0.003	600	0.240
	(0.002)		(0.002)		
Applicants	-26.2	532	-16.0	526	0.640
	(55.9)		(60.1)		
Acceptance Rate	0.001	911	0.003	926	0.578
	(0.002)		(0.002)		
First-Time Out-of-State Enrollment	0.2	917	2.9	941	0.957
	(3.7)		(3.8)		
First-Time In-State Enrollment	-1.8	917	-4.1	941	0.791
	(6.7)		(8.1)		
25th Percentile SAT	-0.2	432	1.4	427	0.797
	(0.8)		(0.8)		

Notes: Coefficients in the first set of columns are from IV regressions of the outcome ΔY_{it} on winning game s in year $t+2$, after stratifying on $\hat{p}(\underline{\mathbf{x}}_{\mathbf{is(t+2)}})$ (analogous to Equation 5). Coefficients in the second set of columns are from STE regressions of the outcome ΔY_{it} on the number of games won in year $t+2$, weighting by $1/\hat{p}(\mathbf{w}_{\mathbf{i(t+2)}})$ and controlling for lagged wins, season games, lagged season games, and time effects (analogous to Equation 8). 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 relationship.

Table 3: Effects of Football Wins on Outcomes

	IV Model		STE Model		Sharp null	FDR
Outcome:	Coefficient	N	Coefficient	N	p-val	q-val
Alumni Athletic Operating Donations	136.4	637	191.2	616	0.001	0.053
	(41.1)		(65.0)			
Alumni Nonathletic Operating Donations	227.9	637	-137.4	616	0.210	0.289
	(171.4)		(96.1)			
Total Alumni Donations	311.9	1,264	267.4	1,258	0.450	0.539
	(295.7)		(266.9)			
Alumni Giving Rate	-0.0001	1,293	0.0002	1,287	0.698	0.612
	(0.0005)		(0.0007)			
Academic Reputation	0.004	669	0.003	650	0.015	0.087
	(0.001)		(0.002)			
Applicants	135.3	533	81.1	528	0.005	0.063
	(49.9)		(60.4)			
Acceptance Rate	-0.003	967	-0.003	979	0.022	0.100
	(0.001)		(0.002)			
First-Time Out-of-State Enrollment	-0.4	1,038	1.6	962	0.983	0.704
	(3.4)		(5.0)			
First-Time In-State Enrollment	15.2	1,038	12.6	962	0.007	0.069
	(5.8)		(6.4)			
25th Percentile SAT	1.8	431	0.8	426	0.002	0.053
	(0.6)		(0.7)			

Notes: Each coefficient represents the effect of an additional football win on an outcome. Coefficients in the first set of columns are from IV regressions of the outcome $\Delta Y_{i(t+1)}$ on winning game s in year t, after stratifying on $\hat{p}(\underline{\mathbf{x}}_{\mathbf{ist}})$ (Equation 5). Coefficients in the second set of columns are from STE regressions of the outcome $\Delta Y_{i(t+1)}$ on the number of games won in year t, weighting by $1/\hat{p}(\mathbf{w}_{\mathbf{it}})$ and controlling for lagged wins, season games, lagged season games, and time effects (Equation 8). 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.

Table 4: Effects of Football Wins on Outcomes for BCS Schools

	IV Model		STE Model		Sharp null	FDR
Outcome:	Coefficient	N	Coefficient	N	p-val	q-val
Alumni Athletic Operating Donations	190.7	353	245.8	341	0.003	0.053
	(68.5)		(100.0)			
Alumni Nonathletic Operating Donations	537.1	353	-183.9	341	0.094	0.170
	(332.0)		(151.6)			
Total Alumni Donations	670.9	796	398.9	795	0.183	0.289
	(461.8)		(404.1)			
Alumni Giving Rate	0.0000	805	0.0007	807	0.935	0.669
	(0.0007)		(0.0009)			
Academic Reputation	0.004	457	0.003	453	0.016	0.087
	(0.002)		(0.002)			
Applicants	79.4	312	56.9	312	0.110	0.184
	(58.9)		(73.7)			
Acceptance Rate	-0.003	642	-0.005	658	0.040	0.130
	(0.002)		(0.002)			
First-Time Out-of-State Enrollment	2.1	623	10.8	571	0.491	0.567
	(5.0)		(5.2)			
First-Time In-State Enrollment	13.8	623	14.5	571	0.053	0.131
	(8.0)		(6.4)			
25th Percentile SAT	0.7	268	0.5	270	0.220	0.289
	(0.6)		(0.6)			

Notes: Each coefficient represents the effect of an additional football win on an outcome. Coefficients in the first set of columns are from IV regressions of the outcome $\Delta Y_{i(t+1)}$ on winning game s in year t, after stratifying on $\hat{p}(\underline{\mathbf{x}}_{\mathbf{ist}})$ (Equation 5). Coefficients in the second set of columns are from STE regressions of the outcome $\Delta Y_{i(t+1)}$ on the number of games won in year t, weighting by $1/\hat{p}(\mathbf{w_{it}})$ and controlling for lagged wins, season games, lagged season games, and time effects (Equation 8). 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.

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

	IV Model		STE Model		Sharp null	FDR
Outcome:	Coefficient	N	Coefficient	N	p-val	q-val
Alumni Athletic Operating Donations	8.6	288	40.1	275	0.886	0.669
	(15.5)		(17.1)			
Alumni Nonathletic Operating Donations	49.4	288	-24.1	275	0.255	0.328
	(41.3)		(49.7)			
Total Alumni Donations	-260.1	478	-26.2	463	0.031	0.124
	(137.8)		(100.2)			
Alumni Giving Rate	-0.0003	495	-0.0006	480	0.674	0.612
	(0.0006)		(0.0010)			
Academic Reputation	0.004	214	0.002	197	0.091	0.170
	(0.003)		(0.003)			
Applicants	96.3	226	48.2	216	0.379	0.454
	(74.4)		(113.2)			
Acceptance Rate	-0.001	335	-0.001	321	0.642	0.612
	(0.002)		(0.002)			
First-Time Out-of-State Enrollment	-3.2	428	-14.1	391	0.400	0.471
	(4.7)		(7.9)			
First-Time In-State Enrollment	17.8	428	-2.4	391	0.054	0.131
	(9.0)		(11.7)			
25th Percentile SAT	1.9	166	1.3	156	0.201	0.289
	(1.4)		(1.4)			

Notes: Each coefficient represents the effect of an additional football win on an outcome. Coefficients in the first set of columns are from IV regressions of the outcome $\Delta Y_{i(t+1)}$ on winning game s in year t, after stratifying on $\hat{p}(\underline{\mathbf{x}}_{\mathbf{ist}})$ (Equation 5). Coefficients in the second set of columns are from STE regressions of the outcome $\Delta Y_{i(t+1)}$ on the number of games won in year t, weighting by $1/\hat{p}(\mathbf{w_{it}})$ and controlling for lagged wins, season games, lagged season games, and time effects (Equation 8). 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.

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

	IV Model		STE Model		Sharp null	FDR
Outcome:	Coefficient	N	Coefficient	N	p-val	q-val
Alumni Athletic Operating Donations	141.6	637	110.4	374	0.012	0.087
	(53.8)		(67.6)			
Alumni Nonathletic Operating Donations	194.9	637	216.6	374	0.289	0.353
	(187.3)		(104.8)			
Total Alumni Donations	579.8	1,264	266.7	787	0.185	0.289
	(376.6)		(331.7)			
Alumni Giving Rate	0.0000	1,293	0.0016	810	0.871	0.669
	(0.0006)		(0.0007)			
Academic Reputation	0.004	669	0.003	390	0.035	0.124
	(0.002)		(0.003)			
Applicants	119.9	533	40.0	320	0.049	0.131
	(62.1)		(93.3)			
Acceptance Rate	-0.002	967	-0.005	630	0.200	0.289
	(0.002)		(0.002)			
First-Time Out-of-State Enrollment	-2.8	1,038	7.4	604	0.525	0.582
	(4.5)		(5.0)			
First-Time In-State Enrollment	15.4	1,038	28.3	604	0.034	0.124
	(7.5)		(8.7)			
25th Percentile SAT	1.9	431	1.7	252	0.016	0.087
	(0.8)		(1.2)			

Notes: Each coefficient represents the effect of an additional football win on an outcome. Coefficients in the first set of columns are from IV regressions of the outcome $\Delta Y_{i(t+1)}$ on winning game s in year t (for $s > 4$), after stratifying on $\hat{p}(\underline{\mathbf{x}}_{\mathbf{ist}})$ (Equation 5). Coefficients in the second set of columns are from STE regressions of the outcome $\Delta Y_{i(t+1)}$ on the number of games won after week 4 in year t, weighting by $1/\hat{p}(\mathbf{w_{it}})$ and controlling for lagged wins, season games, lagged season games, and time effects (Equation 8). 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.

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

	Dependent variable: $\Delta Y_{i(t+1)}$		Dependent variable: $\Delta Y_{i(t+2)}$		FDR
Outcome:	Coefficient	N	Coefficient	N	q-val
Alumni Athletic Operating Donations	136.4	637	82.0	612	0.289
	(41.1)		(66.1)		
Alumni Nonathletic Operating Donations	227.9	637	-191.7	612	0.551
	(171.4)		(263.8)		
Total Alumni Donations	311.9	1,264	45.0	1,294	0.669
	(295.7)		(507.6)		
Alumni Giving Rate	-0.0001	1,293	-0.0005	1,327	0.582
	(0.0005)		(0.0007)		
Academic Reputation	0.004	669	0.000	707	0.669
	(0.001)		(0.002)		
Applicants	135.3	533	-86.6	496	0.386
	(49.9)		(88.1)		
Acceptance Rate	-0.003	967	-0.004	1,141	0.131
	(0.001)		(0.002)		
First-Time Out-of-State Enrollment	-0.4	1,038	-13.2	622	0.131
	(3.4)		(6.5)		
First-Time In-State Enrollment	15.2	1,038	-2.8	622	0.632
	(5.8)		(11.5)		
25th Percentile SAT	1.8	431	1.0	398	0.353
	(0.6)		(0.9)		

Notes: Each coefficient represents the effect of an additional football win on an outcome. Coefficients in the first set of columns are from IV regressions of the outcome $\Delta Y_{i(t+1)}$ on winning game s in year t, after stratifying on $\hat{p}(\underline{\mathbf{x}}_{\mathbf{ist}})$ (Equation 5). Coefficients in the second set of columns are from IV regressions of the outcome $\Delta Y_{i(t+2)}$ on winning game s in year t, after stratifying on $\hat{p}(\underline{\mathbf{x}}_{\mathbf{ist}})$ and controlling for the relationship between winning game s in year t and winning games in year $t+1$. 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.