Vassar College
Digital Window @ Vassar

Senior Capstone Projects

2017

Marginal Revenue Products of Collegiate Basketball Players: What's March Madness Worth to Your Team?

Nicholas K.S Johnson
Vassar College

Follow this and additional works at: http://digitalwindow.vassar.edu/senior_capstone

Part of the Econometrics Commons, Economic Theory Commons, and the Labor Economics Commons

Recommended Citation
Johnson, Nicholas K.S, "Marginal Revenue Products of Collegiate Basketball Players: What's March Madness Worth to Your Team?" (2017). *Senior Capstone Projects*. 624.
http://digitalwindow.vassar.edu/senior_capstone/624

This Open Access is brought to you for free and open access by Digital Window @ Vassar. It has been accepted for inclusion in Senior Capstone Projects by an authorized administrator of Digital Window @ Vassar. For more information, please contact library_thesis@vassar.edu.

Marginal Revenue Products of Collegiate Basketball Players: What's March Madness Worth To Your Team?

Nicholas K.S. Johnson

Vassar College

April 14, 2017

Abstract

Basketball is considered a revenue sport within the context of the National Collegiate Athletic Association (NCAA). The annual *March Madness* tournament brings in over a billion dollars in advertising revenue alone. This is just one illustration of the influx of money lining the coffers of the NCAA and Division 1 teams each year, leading many to question why the student-athletes don't receive a bigger piece of the pie. Some pundits argue that the athletic scholarship caps and restriction of player movement enforced by the NCAA lead to players contributing more in revenue to their schools than they receive in scholarship value. We test this claim by using a modified approach to estimate marginal revenue products (MRPs) across collegiate basketball teams. We find that for 37% of the teams in our panel, the players are collectively contributing more to revenues than the amount they cost their institutions. We argue that, based on how a MRP is defined within this context, we cannot say with certainty that the current on-court revenue contributions of the remaining teams do not exceed their costs.

1 Introduction

It could be the nonstop television coverage, from one of the ubiquitous sports commentary programs and even your nightly news update, or the sudden resurgence of "bracketology", a word that seems almost dormant for 9 months of the year, but has become so entrenched within the American lexicon that you can actually find it in the Oxford dictionary. These are just a few of the harbingers each year, letting us know that the NCAA tournment season, or March Madness as it is affectionately dubbed, is about to begin. From Selection Sunday all the way through to the championship game, millions of americans tune in each year as the event draws the attention of a broad gamut of viewers: from casual sports fans to die-hard enthusiasts, from petty gamblers to the most speculative Vegas bettors.

Viewership was down in 2016 due to the broadcasting of the Final Four and Championship games on Turner Sports channels in lieu of a free traditional commercial broadcast television network. In 2015, however, when CBS took over broadcasting duties, 28.3 million people tuned in to watch the Duke Blue Devils defeat the Wisconsin Badgers in the title game, the

most in about two decades.[1] Additionally, an average of 11.3 million viewers tuned in each day throughout the entirety of the three-week event.[2]

Within the context of the National Collegiate Athletics Association (NCAA), basketball is considered a "revenue sport", in that it often generates net revenues for the schools. As fans of different races, creeds and ideals tune in to Division 1 (D1) collegiate basketball during the regular season and throughout the tournament, the elephant in the room always seems to find its way into debates across the the country, from living rooms and dining tables, bars and restaurants, to television and radio sports studios: *Why aren't these athletes being paid?* This question has become a polemic issue, exacerbated by the tumultuous and undoubtedly corrupt history of the NCAA as an institution, that has generated robust debate with valid arguments on both sides. This paper is meant to be a contribution to this debate. It is important to note that these are non-trivial issues with billions of dollars at stake, and more importantly, far-reaching implications for the economic well-being of the 99% of collegiate basketball players who never make it to the NBA, and more poignantly, the 81% who never play basketball in any kind of professional capacity,[3] many of whom are from low-income, disadvantaged backgrounds. Although the scope scope of this paper is limited to the context of D1 men's collegiate basketball, we hope that it will also be a contribution to the broader discussion on the NCAA's policies regarding collegiate athlete movement and compensation across all sports.

Now, of course, the majority of these athletes receive full-ride athletic scholarships for a tertiary education, and many other perks, some quantifiable and some not. At the same time however, as individuals who help generate billions of dollars in ticket sales, broadcast revenue and licensing fees year after year for their institutions and the NCAA, it is certainly not ludicrous to think that these "student-athletes" should be receiving a bigger piece of the pie. The NCAA, many believe, exploits the notion of "amateurism" and the term "student-athlete" to substantially reduce the pecuniary benefits that should be accruing to players in revenue sports like basketball. As a matter of policy, the NCAA limits the total amount of financial aid that a student athlete can receive to what is known as a full grant-in-aid (GIA), which covers tuition, room and board, and necessary textbooks. In 2011, the NCAA increased this scholarship limit by $2000 to account for "incidental expenses such as school supplies other than required textbooks and travel between the school and the students home" (Lane, 2014).

The NCAA has been involved in myriad lawsuits, some still ongoing, relating to their purported exploitation of student-athletes. In July 2009, for example, Ed O'Bannon, a former star player at UCLA, filed an antitrust class action lawsuit against the NCAA alleging that their use of players' names, images and likenesses for commercial purposes violated the Sherman Antitrust Act. In July 2014, the presiding judge, Judge Claudia Wilken, ruled, among other things, that the NCAA "unreasonably restrains trade" and ordered the organization to allow student-athletes to receive deferred compensation beyond the cost of attendance.[4] In 2015, a federal appelate court, in a 2-1 decision, overturned major parts of Judge Wilken's ruling, deeming compensation beyond the cost of attendance "a quantum leap" from the current business model.[5] This litigation and others like it, combined with ongoing debate within the public discourse, beg the following question: *What is a collegiate athlete really worth?*. This paper uses an economic approach to tackle that question within the context of D1 men's collegiate basketball.

[1] http://time.com/money/4283816/ncaa-basketball-tv-ratings-streaming/
[2] http://money.cnn.com/2015/04/07/media/march-madness-tv-ratings/
[3] http://www.ncaa.org/about/resources/research/estimated-probability-competing-professional-athletics
[4] http://www.cbssports.com/college-football/news/obannon-judge-rules-ncaa-violates-antitrust-law/
[5] http://www.huffingtonpost.com/entry/ncaa-violates-antitrust-law_us_560bee3fe4b0768126fff02a

In order to address this question, one needs to first think about the marginal benefits and costs of student-athletes. For team sports like basketball, the most immediately obvious benefit is that they enhance their team's performance, which should translate to a better product on the court/field, through greater competitiveness and winning percentages, and presumably higher revenues for the schools through various mediums. It is important to note, however, that the benefits student-athletes bring go way beyond their physical ability. Players with a certain cachet help the coaching staff in the recruiting process, which ultimately leads to more talent on future teams and the sustained success of the athletic program. The players also attract a non-trivial amount of financial donations from alumni and other special interest groups, as well as sponorship money from blue-chip companies like Nike and Adidas. These realities, among others, result in pecuniary and non-pecuniary benefits that do not directly arise from an athlete's performance on the field/court of play. In recognition of the complexity associated with attempting to quantify and isolate these additional benefits, this paper will focus exclusively on value specifically derived from the student-athletes' on-court performance. As such, any estimates of value presented can be thought of as a lower-bound on the total value contributed by these individuals.

The most direct cost to a school/athletic department of taking on an additional student is the value of his/her athletic scholarship (Lane, 2014). However, Martin (2004)[9] found that "...increasing returns to scale and significant fixed costs in the intermediate term imply that marginal cost [of college education] is less than average cost". This means that the marginal cost of educating and housing an additional student may be considerably less than the tuition and room and board charges. Moreover, "The marginal cost of a student-athlete also depends on whether the student-athlete is an additional student or is replacing another student, and if replacing a student, whether that student would pay full price or receive financial aid from the school." (Lane, 2014). Because of a lack of reliable data and complications stemming from costs shared with other sports teams such as the maintenance of training facilities, among other reasons, this paper uses a player's athletic scholarship as a proxy for his marginal cost to the school. Given the preceding arguments, the reader may view this as an upper-bound on the true marginal cost to the school.

This paper applies an innovative approach to an established methodology within the sports labor market context in order to quantify the on-court value of D1 collegiate basketball players, and then compares them to their costs as measured by scholarship value. Two sets of marginal revenue products are calculated, and both use maximum likelihood methods to estimate the production function. It is important to note that the results of this paper fall within a partial equilibrium framework. More will be said about this in the concluding remarks.

We will proceed as follows. A review of relevant literature on labor markets within the context of athletics and sports is presented, which provides a broad context for the methodologies and arguments that follow. We will then discuss our approach, present the estimation model we have developed to address the issue at hand, and highlight its limitations. This is followed by a review of our data sources, after which we present our results, discuss them, and finally provide concluding remarks.

2 Literature Review

The literature on labor markets within the context of sports, both professional and collegiate, is quite extensive and addresses a slew of issues from different angles. The motives of the studies within this research space span a broad gamut of perspectives, but they generally are seeking to do one, or both, of the following two things: test an economic theory, or analyse some idiosyncratic characteristic of the specific labor market in question, such as the extent of monopsony rents that are generated. For my purposes, the main theoretical question is whether competition in these markets is inhibited in any way. There are inherent market imperfections that exist, whether they be compensation restrictions by the NCAA in collegiate sports, or salary cap restrictions in professional sports. If these markets were unencumbered and perfectly functioning, an athletes salary/compensation would be equated to his marginal revenue product (MRP).

A central question that arises throughout the literature relates to how these MRPs can be measured. In addition to a nuanced understanding of the nature of the sport and the drivers of performance, this issue requires one to postulate a specific framework. The first issue that needs to be considered is how we should think about production functions within an athletic context. How do we measure output? The convention, pioneered by Scully (1974)[12] in his seminal paper, is to use games, as this is what fans pay to see. Specifically, the output is considered to be the number of wins in a season, or the team win/loss record, as one can argue that fans do not just patronize sports neutrally, but in the hope of seeing their team succeed. In sports where draws are common, such as soccer, points or goals scored may be a more appropriate measure of team output (Schofield, 1988)[11]. Other measures of team output include attendance (Gustafson, 1999)[4] and TV broadcast revenue (Hausman and Leonard, 1997)[5], and some studies such as Carmichael (2000)[3] have used individual games as the time unit, in lieu of the conventional single season.

The next question to be considered is what inputs should be included in the production function. Clearly this varies by sport, but the most consistent input vector has been a vector of performance variables, given that most studies are trying to estimate an athlete's value strictly from his contribution on the field of play. Other common input variables include some measure of coach quality, such as Kahn (1993)[6] which shows positive effects from managerial quality on player performance, and division/geographical area. One would surmise, particularly within the collegiate context, that an important input/control vector should to include season and school-fixed variables. This was sorely lacking early in the literature. Regarding college basketball, Lane, Nagel and Netz (2014)[8] first included these effects.

Functional form specification is the next major issue. How are the inputs combined to produce this output? The Scully method, which is the methodology that a preponderance of the literature builds on, implicitly assumes additive separability of the production function. In sports such as cricket and baseball, which is the sport Scully originally focused on, experience suggests this may be a reasonable assumption. Otherwise though, sports performance is highly interdependent. One would deduce that interaction of the inputs needs to be accounted for, and this is another area where the literature leaves much to be desired. There are some log-linear models that account for interactions more generally (Gustafson, 1999), and others (Atkinson, 1988)[1] that employ structural approaches, leveraging in-depth knowledge of the sport to account for interactions between different positions on the roster.

Within the Scully framework, the marginal product is the regression coefficient vector on the performance variables in the production function, and the marginal revenue comes from a revenue equation regressing team revenue on the team performance variables and other controls. A player's MRP is the product of these coefficients, multiplied by his "contribution"

to team performance. One of the more polemic issues within this space, which we will address later, is how an individual's contribution to overall team performance is measured.

There are other methods that have been used to estimate MRPs. Lane (2014) uses the distribution of professional basketball player salaries within the collegiate basketball context. Brown and Jewell (2006)[2] tried to estimate the MRP of star players on the collegiate level, essentially through a regression of team revenue on the number of players drafted professionally during that season, plus other controls.

Results within this field vary substantially. For my purposes, some of the more salient points include the fact that MRP estimates within the professional context are quite ambiguous, with some studies finding that athletes are grossly underpaid and others finding that athletes are slightly, or even substantially, overpaid. These findings have implications for future collective bargaining agreement negotiations, league policy, and public sentiment, just to name a few. Within the collegiate context, while estimates do vary significantly, the general conclusions are fairly consistent. The majority of papers find that the NCAA, through their restrictive policies on compensation and player transfers, generate monopsony rents in conjunction with D1 institutions. McKenzie and Sullivan (1987)[10] present coherent economic arguments, commonly cited by the NCAA in litigation hearings, in support of the NCAAs policies, but their arguments have not been empirically validated in any way.

Though comprehensive in some ways, the literature is lacking on a variety of fronts. The overwhelming use of ordinary least squares (OLS) estimation, for me, is problematic for a variety of reasons. If your output variable is the team's win-loss record, OLS does not bound your results between 0 and 1. Also, the interdependence that lies at the core of all sporting environments makes for a multiplicity of endogeneity problems that I am not convinced the literature spends enough time considering. It is no surprise that baseball is the most studied North American sport, because the assumptions of additive separability and the independence of players marginal products make the estimation problems much more tractable; that is, it becomes easier to isolate individual contributions to team success. However, this leaves a large gap in other areas, particularly at the collegiate level, as basketball and football are the sports that generate the most revenue.

In addition to the lack of attention paid to the interaction of inputs within the production function, many papers don't provide adequate rationale justifying the inputs they use. Also, many studies don't explicitly distinguish between control variables, and variables for which they are trying to tease out causality. For example, many studies include some measure of managerial quality. Using this variable as a control is fine, but trying to interpret the coefficient causally, which many papers do, gives rise to further endogeneity issues.

In the same vein, many studies include too many variables in their performance vector, and this is problematic within the context of a linear regression. It is particularly dangerous as it relates to basketball, where many of these performance statistics are highly correlated. This leads to inefficient estimates and less powerful tests. I think that in the era of advanced analytics in which we live, one can comprehensively measure performance across all facets of the game with just a couple statistics, which is much more efficient. Another flaw in the literature is that most studies assume that every team in their sample has the same production function. However, different teams may choose to combine inputs differently. For example, a team with a comparative advantage in a particular style of play probably exhibits a different production function.

Finally, the Scully method upon which much of the literature builds has well-documented flaws. While some studies have acknowledged these short-comings, and other have tried to address them, I think it's important to point out some of the more salient ones that remain. Firstly, the method is purely predicated on performance. Consequently, if one is trying to estimate individual MRPs, players that do not get much or any playing time will have MRPs

equal to zero, which is not plausible. Secondly, the Scully method is extremely sensitive to the way in which individual contributions to team performance are measured. Krautmann (1999)[7] presents a comprehensive illustration of this flaw, among other issues. Thirdly, as I have alluded to before, this literature has been primarily concerned with the performance-related revenue contribution of players. Not many studies have empirically tried to measure what the off-the-field value may be to a team. This is not as germane within the professional context, where athletes can independently sign endorsement deals that value their names and likeness. It is particularly relevant within the collegiate context, though ultimately impossible to measure because of the structure of the NCAA and the rules that prevail.

3 Our Approach

The overarching question that this paper addresses is whether the marginal revenue products (MRPs) of D1 college basketball teams, as measured strictly by on-court performance, exceed their cost to the schools, as measured by the combined maximum possible scholarship value allocated to the team. Previous literature, both within the professional and collegiate contexts, has sought to estimate the MRPs of individual players. For the purposes of this paper, there are two main reasons why we believe it is more prudent to try and estimate collective MRPs of teams, which is essentially just the sum of individual player MRPs. Firstly, due to the preponderance of interdependence that exists particularly within the sport of basketball, we think it is highly unlikely that one will be able to account for interactions accurately enough to estimate the specific revenue contribution of a particular player. Secondly, and more importantly, from a pragmatic point of view we should not particularly care about individual MRPs. Amateurism has and always will be what the NCAA purports to be its bedrock principle. As such, there will never be a time, at least in the foreseeable future, where collegiate athletes are each compensated differently according to their "value" to the team, as in professional markets. Any policy change/concession on the part of the NCAA as it relates to collegiate athlete compensation will apply across the board, to every team and every collegiate athlete. Indeed, Title IX rules would dictate as much. Testing whether or not an individual's MRP exceeds his scholarship value thus has very few policy implications. A more compelling endeavour, therefore, would be to try and figure out whether D1 collegiate basketball teams, as a collective, contribute more in revenue to their institutions than they receive in scholarship value, as any policy change or lack thereof engendered by such a project would call for team-wide shifts in NCAA policy.

Theory suggests that, due to scholarship caps and restriction of player transfers within the NCAA, students' athletic scholarship values may be depressed below their actual revenue contributions to the school. In order to measure a MRP, one needs a production function, to derive the marginal product, and a revenue equation, to derive the marginal revenue. As stated in the literature review, most of the empirical literature within this field builds, in one way or another, on the Scully method. Our proposed method for estimating MRPs of teams deviates from the Scully method, in that we estimate the production function differently. While most studies have used the team win-loss record as the output variable in the production function, we are proposing to use a binary variable that indicates whether or not the team appeared in the NCAA tournament that year. In our view, there is a compelling argument that what matters most to teams, schools and fans within the collegiate basketball context is not necessarily regular season wins, but whether or not they earn a chance to play for the ultimate prize: a championship. While each institution will have its own nuances, we believe it reigns true that, if given the choice between a great regular season record and a tournament appearance, both teams and fans would overwhelmingly choose a tournament appearance. Moreover, it is usually the case in the NBA and other professional leagues that a team's regular season record determines home-court advantage in the postseason. This is not the case within the collegiate basketball context; NCAA tournament games are played at neutral sites, and so no team has home-court advantage. The fact that there is no home-court advantage at stake in the regular season lends even more credence to the argument that teams are trying to maximize the likelihood of appearing in the tournament, and not necessarily trying to maximize their win-loss record during the regular season.

Now obviously, the two output variables have some relation. A team with more wins, ceteris paribus, has a greater likelihood of making the tournament. We prefer the tournament appearance output variable because it is novel to the literature. In both methods, we are concerned with the marginal effect of on-court performance, that is a vector of team-performance

variables, on the output variable.

The hypothesis of the Scully method is that, each season, teams are trying to maximize their win-loss record. Our hypothesis is that, each season, teams are trying to maximize the probability that they receive a tournament bid. There are well-known issues with OLS when the dependent variable is binary. These include the fact that your estimates are not constrained between 0 and 1, and the error term is highly non-normal as well as heteroskedastic. Therefore, we propose to estimate the production function using maximum likelihood estimation.

Consider our dependent variable,y_i, for whether or not a team makes the NCAA tournament in a particular season. It is a Bernoulli random variable, where the probability density function (PDF) is:

$$f(y_i) = \begin{cases} \theta^{y_i}(1-\theta)^{1-y_i}, & y_i = 0, 1 \\ 0 \end{cases}$$

where $\theta = Pr(y_i = 1)$. Clearly, within this framework, one should not expect that $Pr(y_i = 1)$ is constant, but that it depends on i, that is, contingent upon school i's characteristics during that season. Probit, a popular estimation approach for such issues, models θ_i as a cumulative distribution function (CDF). Specifically, $\theta_i = \Phi(x_i'\beta)$, where x_i is a vector of appropriate characteristics that help determine y_i. The maximum likelihood estimation procedure then estimates β parameters, in lieu of θ. The log-likelihood function for this problem is the natural logarithm of the product of the marginal probabilities:

$$\sum_{n=1}^{\infty} y_i \ \ln \ \Phi(x_i'\beta) + (1 - y_i) \ \ln \ (1 - \Phi(x_i'\beta))$$

For our model, we take the probit approach and modify it to account for the way my data was actually generated, that is, to account for the actual dynamics at play in the determination of whether or not a school makes the tournament. There are 32 D1 conferences in the NCAA. At the end of the regular season, conference champions are determined via conference tournaments, and each of these teams is awarded an automatic bid to the March Madness tournament. Subsequent to this, on what is dubbed Selection Sunday, the NCAA selection committee awards at-large bids, based on weeks of deliberation, to 36 other teams, for a total of 68. With this in mind, consider the following model for θ_i, the probability that a team makes the tournament that year:

$$\theta_i = \Phi(x_i'\beta) + \Phi(z_i'\delta)\left(1 - \Phi(x_i'\beta)\right) \text{ where}$$

- $\Phi(x_i'\beta)$ is the CDF of a standard normal random variable indicating the probability that team i won their conference.
- $\Phi(z_i'\delta)$ is the CDF of a standard normal random variable indicating the probability that team i was granted an at-large bid by the selection committee.
- x_i is is a vector of characteristics that determine whether or not team i wins their conference. This vector will contain the on-court performance variables whose marginal effect will be our marginal product in the MRP estimation.
- z_i is a vector of characteristics that determine whether or not team i is selected by the committee.

The maximum likelihood procedure then estimates our β and δ parameters. It is clear that even though θ_i is now a function of two CDF probabilities instead of one, it is still constrained between 0 and 1, as desired, and with this framework we have accounted for the

way the data is actually being generated. Another way to think of the production function presented above is that we have two latent equations jointly determining the probability of a tournament appearance by each school:

1. $y^*_{1_i} = x_i'\beta + \epsilon_{1_i}$ where $y_{1_i} = 1$ if $y^*_{1_i} > 0$, and
2. $y^*_{2_i} = z_i'\delta + \epsilon_{2_i}$ where $y_{2_i} = 1$ if $y^*_{2_i} > 0$

$\epsilon_{1_i}, \epsilon_{2_i} \sim i.i.d.N(0,1)$ given x_i and z_i respectively.

It is important to note that our marginal product with this model, that is the marginal effect of our performance variables on the probability θ_i, will vary for each team, as it will depend on that teams specific characteristics. This is a distinct advantage over OLS models, as most of the previous literature has either assumed constant returns or the same effect across all teams.

The revenue equation will be a standard panel regression controlling for fixed effects. Specifically:

$$rev_{it} = \alpha + \beta Tourn_{it} + \delta D_{it} + \gamma Conf_i + \rho Year_t + \lambda F_i + \epsilon_{it}, \text{ where}$$

- i and t index the team and season respectively.
- *Tourn* is the tournament appearance dummy variable.
- D is a vector containing other determinants of revenue.
- *Conf* is a vector of conference controls.
- *Year* represents the season.
- F represents team fixed effects.

4 Limitations

As with any model, there are assumptions that undergird our approach and various limitations that arise. Some of the more salient ones are outlined below (please note that this is in no way exhaustive):

- As mentioned previously, the marginal effects of our performance variables on the probability of making the tournament, θ_i, will depend on each team's characteristics and thus are not constant, but different for each team in each year. However, there seems to be a mathematical drawback in the way that we have modelled the production function. The marginal effect of a single performance variable tells us how the probability of making the tournament changes due to this performance variable, holding everything else constant. That is, it is the derivative of θ_i with respect to that performance variable p:

  $$\frac{\partial Pr(y_i = 1|x_i, z_i)}{\partial p} = \beta_p \phi(x_i'\beta)[1 - \Phi(z_i'\delta)].$$

  It is clear that the marginal effect will be essentially 0 if $\Phi(z_i'\delta)$, the probability that team i is selected by the committee, is arbitrarily close to 1. Years of historical precedent allow us to hypothesize, a priori, that teams with a high probability of being selected by the committee, and thus whose MRPs will be 0 within this framework, are high quality teams from top conferences. This is problematic as it stretches credulity to believe that players on these teams do not contribute to their school's revenue stream through their on-court performance. In taking account of how the data is actually generated and ensuring that marginal effects depend on the team in question, it seems that we have given rise to a shortcoming in that we are unable to value some of the top teams.

  In order to address this shortcoming, we will estimate a second set of MRPs. The methodology will be the same, except that the production function will be modelled using the conventional probit framework. This does not take into account the dynamics of the data generation and thus can be thought of as a more general approach. However it is still theoretically sound in that we are estimating the probability of making the NCAA tournament using on-court performance variables and other factors that are assumed to be exogenous.

- The likelihood function assumes independence across observations. Because our data is a panel, this implies two things:

  1. There is no cross-sectional correlation. That is, one team making the tournament does not influence whether or not another team makes the tournament.
  2. There is no serial correlation. That is, team i making the tournament one year does not influence whether or not they make the tournament in another year.

  The first of these assumptions, while seemingly precarious at first glance, is not all that problematic. Different studies such as Lane's (2014) have shown that factors such as opponents' rank index do not significantly impact a team's performance. Also, the selection committee, which selects more than half of the tournament participants, has no set rubric that they must adhere to in terms of conference/geographical representation. This means that, technically, having 6 teams from your conference selected for the tournament does not make it less likely your team is selected, and conversely, having no teams from your conference in the tournament does not make it more likely that your team will be selected.

  The second of these assumptions is much stronger. It doesn't seem plausible that a team making the tournament one year does not influence their competitiveness and

ability to make the tournament the next year, particularly if most of the players are returning. The experience and maturity are likely to translate into a similarly strong team the following year. That being said, our estimates will still be consistent, as are all maximum likelihood estimators. What will be affected is the asymptotic efficiency of the estimator, in that it will no longer achieve the Cramer-Rao lower bound, and thus our estimates will be less precise and our tests less powerful. We ran our models correcting the standard errors for the type of clustering autocorrelation that concerned us, and the results were essentially identical.

- If we had data specifying conference winners and at-large bid selections in each year, we could estimate the latent equations and parameters directly as a robustness check on our maximum likelihood model. However, Monte-Carlo simulations have been used to show that the production function does indeed produce the expected coefficients, and so this is not an issue.
- One of the critiques in our literature review was that most of the models in this space assume that the production function for each team is the same. Our model makes this same asumption. However, it is much more appropriate within the context of collegiate basketball. Unlike sports such as football, there is no broad distinction of styles geographically. An illustration of this is the West Coast offense versus the East Coast offense that is still relevant in both collegiate and professional football. More importantly, in basketball, and particularly collegiate basketball, the rules more or less dictate what styles of play are successful and feasible for a coach to implement. Consequently, the styles of play of each team, while not facsimiles, are broadly similar. Different production functions representing different styles of play would be more appropriate in other sports, or if we were comparing teams across eras, in which case the rules of the game, and thus the style of play, would be appreciably different.

5 Data Sources

The basketball team revenue, expense and participation data is from the U.S. Department of Educations Office of Postsecondary Educations Equity in Athletics Data Analysis Cutting Tool website. The performance statistics and coaching statistics are from Basketball Reference, an online repository of all things basketball. Arena capacity data is from NCAA online archives. Grant-in-aid and Cost of Attendance data is from the Department of Education, National Center for Education Statistics, IPEDS (Integrated Postsecondary Education Data System). While we do not have data on individual scholarships, most players receive the maximum allowable amount (Lane, 2014), and so that is what is used in cost calculations.

Descriptive statistics are presented in the appendices.

6 Results and Discussion

The Production Function

As discussed, we postulate that the output teams produce is not their win/loss record, but an appearance in the NCAA tournament at the end of the regular season. The production function is estimated using a maximum likelihood approach as follows:

$$\theta_i = \Phi(x_i'\beta) + \Phi(z_i'\delta)\left(1 - \Phi(x_i'\beta)\right)$$

where θ_i is the probability of a tournament appearance, $\Phi(x_i'\beta)$ is the probability a team wins their conference, and $\Phi(z_i'\delta)$ is the probability that a team is awarded an at-large bid by the selection committee.

In the vector x_i we include a selection of performance variables, whose marginal effect on the probability of a tournament appearance we are ultimately interested in, and two measures of managerial quality and experience. These variables are assumed to be exogenous in the maximum likelihood approach. The performance variables included are:

1. True Shooting (TS) Percentage: This is a popular statistic within basketball circles of pundits and casual fans alike. It is a measure of shooting efficiency, on a scale of 0-100, that takes into account different types of attempts at the basket including 2-point field goals and 3-point field goals. The higher a player or team's TS percentage, the more efficient they are at scoring the basketball, and thus the more likely they are to win games.
2. Free-throw (FT) percentage: When a player is fouled on a shot attempt, or if the opponents are over the foul limit, they are sent to the free throw line to shoot either 1, 2 or 3 free throws, depending on the context. This is a critical part of the game, particularly within the collegiate context where games are being called tightly and more fouls are being called. Many games are decided based on made or missed free-throws, and the greater percentage a player or team shoots from the line, the better.
3. Free Throws per Field Goal Attempt (FT/FGA): We included this measure specifically because of the context of collegiate basketball and recent rule changes. In an attempt to make the product, that is the games, more appealing, the NCAA has made a conscious effort to call games more tightly in order to encourage more player movement on the court and foster a more free-flowing style of play.[6] As mentioned above, this has resulted in more fouls are being called, and more free-throws being shot. Free throws are considered the easiest way to accumulate points, given that you are at the line by yourself and there is no one guarding you. This shift in officiating has led to more teams trying to get to the free-throw line more frequently. Success at this strategy can make a significant difference, particularly against an opponenent that plays great defense, where open shots and good offense are much more difficult to generate. Consequently, greater FT/FGA should improve a team's winning prospects throughout the season.
4. Total Rebounds (TRB): Rebounding is an integral part of the game, both on the offensive and defensive ends of the court. Not only does it limit the number of field goal attempts your opponent has when your team is on defense, and increase the number of opportunities that you have to accumulate points when your team is on offense, but it is also commonly thought of as a barometer on the level of effort the players are exerting, as rebounding is considered a "hustle" statistic. The more rebounds a team accumulates, the better.

[6] For more on this, see here: http://herosports.com/news/college-basketball-officiating-all-time-low

5. Total Steals (STL): Steals are another significant part of the sport that enhances your play on the defensive end of the floor. They limit your opponents' opportunities to accumulate points and allow you to score easy baskets in transition before your opponent has a chance to set up their defense. A steal is also considered a "hustle" statistic and thus a measure of the level of overall effort being put forth. The more steals, the better.

We expect the coeffcients on the aforementioned performance variables to be positive. Coaches, particularly in collegiate basketball, are considered important because the players are still kids between the ages of 18 and 22, and as such many have not fully developed their games and basketball acumen. Consequently, we include 2 measures of the quality of a team's coach: their overall win/loss record across the various teams they have coached, and the number of times they have led their team to the NCAA tournament. We expect the coefficient on these variables to be positive as well.

In the vector z_i, we include a selection of variables that, based on either official statements by members of the selection committee or conventional beliefs espoused by experts within the media, we believe are central to a team being awarded an at-large bid. There are 32 conferences in NCAA division 1 competition, however there are 6 conferences, commonly referred to as the BCS Power Conferences, that stand above the rest in terms of revenue from broadcasting, recruiting, performance and overall athletic success. These are: the Southeastern Conference (SEC), the Pac-12 Conference, the Big-10 Conference, the Big-12 Conference, the Atlantic Coast Conference (ACC), and the Big East Conference. We include a dummy variable for whether or not a team is in a top conference, as it is commonly cited amongst pundits that top teams within the top conferences are virtually guaranteed a tournament spot, as the selection committee will vote them in even if they do not win their conference. We include a variable known as the simple rating system (SRS) to account for the general quality of a team that the committee would consider in their appraisals. This variable is denominated in points above/below average, where zero is average.[7] We also include a variable known as the strength of schedule (SOS), that is denominated in the same way as the SRS, which accounts for the quality of the opposition a team has faced throughout the regular season. It is well known, as many committee members have stated publicly, that a team's strength of schedule is a crucial part of the committee's deliberations.

We include the number of away wins that a team accumulated that season, as committee members have stressed that teams need to play well away from home, particularly since the tournament games are all played at neutral sites. Additionally, we include 2 interaction terms: the number of conference wins that season interacted with the the top conference dummy variable, and the number of wins outside your conference interacted with a non-top conference dummy variable. The rationale for this is that if you're in a top conference, the committee will more likely be concerned with your wins within the conference, as the competition is stiff and the quality of opponent is higher. If you are in a weaker conference however, the committee will more likely be concerned with your wins outside of the conference, given that the quality of opposition within your conference is not as strong. Of course, not every conference outside of the power conferences is considered weak, and performance varies each season, so the measure is somewhat imperfect in that sense. Yet and still, it should be fairly accurate. We expect the coefficients on all the variables in the vector z_i to be positive.

The results from the maximum likelihood estimation are presented below in Table 1. Most of the coefficients are of the expected sign, and are significant at the 5% level, except for the coefficient on FT percentage, which is significant at the 10% level.[8] One notable exception is the coefficient on the top conference dummy variable in the z_i vector. It is

[7] http://www.sports-reference.com/cbb/seasons/2016-school-stats.html

[8] p-value = 0.087

negative, and significant at the 5% level. This contravenes our hypothesis that being in a top conference should increase the probability that you are awarded an at-large bid by the selection committee. One possible explanation for this is that these 6 conferences are top-heavy, meaning that there are 2 or 3 great teams in each conference each year, with the others perhaps much weaker on a relative basis when compared with teams from other conferences.

VARIABLES	(Conference Win) $x_i'\beta$	(At-Large Bid) $z_i'\delta$
True Shooting Percentage	14.26***	
	(3.235)	
Free Throw Percentage	3.959*	
	(2.311)	
Free Throws per FGA	4.063**	
	(1.970)	
Total Rebounds	0.00308***	
	(0.000636)	
Total Steals	0.00477***	
	(0.00164)	
Coach Win/Loss	1.415	
	(0.942)	
Coach # of Tournament Appearances	0.0539**	
	(0.0220)	
Top Conference		-3.112*
		(1.588)
Simple Rating System		0.332***
		(0.0661)
Strength of Schedule		0.289***
		(0.109)
Away Wins		0.309***
		(0.0942)
Top Conference · Conference Wins		0.293***
		(0.109)
Non-Top Conference · Non-Conference Wins		0.0953
		(0.0864)
Constant	-18.43***	-7.577***
	(2.081)	(1.344)
Observations	1,385	1,385

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1: Results from Maximum Likelihood estimation of Production Function. Dependent variable is tournament appearance.

In order to calculate the collective players' MRPs, we need marginal products. We want to know how an increase in rebounds, steals or the like, by a specfic collection of players, impacts

the probability that that team makes the NCAA tournament. These marginal products, of course, are not the coefficients presented in Table 1 from the maximum likelihood estimation. As discussed previously, the marginal effect of one of the performance variables, call it p, is:

$$\frac{\partial Pr(y_i = 1|x_i, z_i)}{\partial p} = \beta_p \phi(x_i'\beta)[1 - \Phi(z_i'\delta)].$$

We derive linear predictions based on the coefficients from the maximum likelihood estimation, meaning predictions of $x_i'\beta$ and $z_i'\delta$, and use them to calculate the marginal effect of the 5 performance variables included. Each team's characteristics differ, so the predictions of $x_i'\beta$ and $z_i'\delta$ differ, leading to different marginal effects for each team and each year in the panel. The averages of these different marginal effects for the 5 variables in question are presented in Table 2. These averages are what Stata reports using the *margins* command in a conventional probit, logit, or any other maximum likelihood model that Stata has programmed automatically.

Performance Variables	(1) Obs.	(2) Average Marginal Effect	(3) S.D
True Shooting Percentage	1,385	1.259	1.445
Total Rebounds	1,385	0.000272	0.000313
Free-Throw Percentage	1,385	0.350	0.401
Free-Throws per FGA	1,385	0.359	0.412
Total Steals	1,385	0.000421	0.000484
Number of Schools	345	345	345

Table 2: Averages and Standard Deviations of performance variable marginal effects.

As it relates to interpretation, the information above tells us that an additional rebound, on average, increases the probability of a tournament appearance by 0.00027%, a non-trivial effect. Note that one additional rebound is negligible when thinking in terms of an entire season, given that the mean of our TRB variable is ~ 1113 and the standard deviation is ~ 135. Since this is not a linear specification, we cannot exactly say that a 1 standard deviation increase in total rebounds for the season increases the probability of a tournament appearance by $.00027 \cdot 135 \sim 3.67\%$ on average.

Variables like TS and FT percentage are bounded between 0 and 1, and so their interpretation is a bit different. For example, we can glean from the table above that making 10% more of your free throws increases your chances of being in the tournament by $\sim 3.5\%$ on average, again an appreciable impact.

The engaged reader will be curious as to why we don't include these performance statistics as explanatory variables for whether or not a team is awarded an at-large bid. We decided a priori that the performance variables probably have a much stronger impact on the probability that a team wins their conference than on the probability that a team is selected by the committee. The committee is much more concerned with the landscape from a relative perspective, which is why they consider a team's rating and strength of schedule rather than their metrics in isolation.[9] For example, the committee will place less of a premium on

[9] As an illustration of the committee's comparative approach, see the following article discussing why Syracuse was not awarded an at-large bid to the 2017 tournament: http://www.syracuse.com/orangebasketball/index.ssf/2017/03/ncaa_selection_committee_chairman_on_why_syracuse_was_left_out_of_tournament.html

the good performance statistics of a school in a bad conference, on the premise that those statistics are inflated due to inferior competition.

If we were to model the production function with the performance measures as explanatory variables in both vectors, the marginal effect of a performance variable p becomes:

$$\frac{\partial Pr(y_i = 1|x_i, z_i)}{\partial p} = \beta_p \phi(x_i'\beta)[1 - \Phi(z_i'\delta)] + \delta_p \phi(z_i'\beta)[1 - \Phi(x_i'\beta)].$$

We estimated the production function in this way and found the coefficients on the performance variables in the z_i vector to be mostly insignificant, thus confirming our a priori hypothesis.

The Revenue Equation

Team revenue is modelled as follows:

$$rev_{it} = \alpha + \beta Tourn_{it} + \delta D_{it} + \gamma Conf_i + \rho Year_t + \lambda F_i + \epsilon_{it}$$

In the vector D_{it} we include the team's home arena capacity, their strength of schedule, the top conference dummy variable, and a variable *Pace* that measures an up-tempo style of play. The rationale for including arena capacity is straightforward: the more seats in the stadium, the more paying fans that each team can accomodate at their home games. Strength of schedule, top conference and pace are included to capture some aspects of fan demand. A stronger strength of schedule means you are playing better teams, which probably means that more fans will want to see that matchup. It probably also means that networks will be interested in more of your games, bringing in more broadcasting revenue. The same applies to the inclusion of the top conference dummy variable. We include the pace variable as we believe that a more up-tempo, free-flowing style of play is more exhilarating to watch and thus more appealing to fans.

Different conferences have different revenue-sharing agreements. For example, "some conferences split gate revenues evenly; some split a minimum, with revenues above the minimum accruing to the home team; and some guarantee a minimum to the visiting team" (Lane, 2014). We therefore include $Conf_i$, a vector of conference dummies, to account for these different revenue sharing agreements.

We include year-fixed effects to account for general macro-economic factors, and we estimate the panel regression with and without school-fixed effects. The results are presented below in Table 3. The estimates with fixed effects differ immensely from those without fixed effects. For example, the effect of a tournament appearance on revenue drops from almost \$2 million when not accounting for school-fixed effects, to $\sim$ \$206,000 when these effects are accounted for. This disparity is consistent with what Lane (2014) found, even though that paper uses the win/loss record as its measure of output, and finds a much smaller drop-off. While we did not expect the disparity to be this large, the drop should make sense to anyone familiar with the landscape of college sports in America. The sports culture at some institutions is much more palpable than at others, and those fans support their teams much more fervently than fans of other teams, through both good seasons and bad seasons. Some teams also receive consistently higher institutional support from their teams, in the form of better facilities, more resources like physical trainers, team planes and other amenities. These are just a few reasons that may help explain the disparity. This finding lends even more

credence to the belief that past studies that have excluded fixed effects have dramatically over-estimated marginal revenue products.

It goes without saying that the school-fixed effects are jointly significant. The conference dummies are jointly significant as well. Arena capacity is omitted when we add fixed effects to the regression,[10] however it is significant in the regression without fixed effects. SOS and the top conference dummy both have the expected signs, and SOS is significant at all relevant levels of significance. We get a negative coefficient for the pace variable, which goes against what we expected. However, this coefficient is not significant, and thus the impact on revenue is uncertain.

VARIABLES	(**Without School-Fixed Effects**) Estimated Coefficients	(**With School-Fixed Effects**) Estimated Coefficients
Tournament	1,819,271***	206,531***
	(266,440)	(75,347)
Arena Capacity	339.7***	
	(42.28)	
Strength of Schedule	242,624***	48,135***
	(50,620)	(17,724)
Top Conference	4,061,599***	1,459,706
	(573,586)	(947,799)
Pace	24,900	-18,408
	(24,570)	(11,890)
Constant	-1.133e+06	5.460e+06***
	(1.733e+06)	(829,293)
Observations	1,366	1,366
Conference Dummies	YES	YES
Year-Fixed Effects	YES	YES
School-Fixed Effects	NO	YES
Number of Schools		345

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Revenue Regression Results. Dependent variable is team's annual revenues.

In order to calculate a team's MRP, we take the sum of the marginal effects of the performance variables for each team, which tell us how an increase in performance by the collection of players impacts the probability of a tournament appearance, and multiply by the tournament coefficient in the revenue equation, which tells us the increase in revenue that arises when a team makes the tournament. Papers that try to calculate MRPs for an individual player go a step further, by multiplying the previous product by some measure of that player's proportional contribution to team performance.

The estimated MRPs are positively skewed, and we summarize them in the Table 4. Across the panel, the MRPs range from 0 to around $1.8 million, and on average, NCAA D1 basketball players, when aggregated by team, collectively contribute around $406,000 in

[10]Each team's home arena capacity is fixed each year.

revenue to their schools. Recall that we are isolating on-court contributions, and are not accounting for off-the-court value through name, likeness, branding, recruiting, etc.

(1) Mean	(2) SD	(3) Min	(4) Max	(5) Skewness	(6) P25	(7) P50	(8) P75	(9) P95
406,471	466,593	0	1,834,670	1.298	41,035	209,921	643,639	1,478,000

Table 4: Summary statistics for marginal revenue products across panel.

We want to compare each team's MRP in each year to our upper-bound on what these athletes cost the school. We do this by first assuming that each player receives the maximum allowable scholarship amount, which, as indicated previously, is equal to full grant-in-aid (GIA) plus an additional $2,000 for incidental expenses. Full GIA covers tuition and fees, room and board and required textbooks, and so this maximum allowable scholarship amount will clearly vary by institution. We then take this maximum allowable amount and multiply it by the number of participants on the team. This then becomes our upper-bound on the marginal cost of these athletes to the school/athletic department. We also repeat the same process with the estimated true cost of attendance, which is typically higher than the maximum allowable scholarship amount. We do this because many have asserted that student athletes should at least receive scholarships equal to the estimated cost of attendance, and so we want to compare the on-court value of the athletes to these cost of attendance estimates.

We find that, across the panel, collective MRPs exceed the collective maximum possible scholarship amount allocated to the team and collective cost-of-attendance estimates for $\sim$ 37% and $\sim$ 36% of schools respectively. The numbers are remarkably consistent across the years included in the panel, except for 2014. Collective MRPs exceeded collective scholarship values for 35-37% of schools in the years 2012, 2013 and 2015. In 2014, however, the proportion jumped to 44% (See Figure 1).

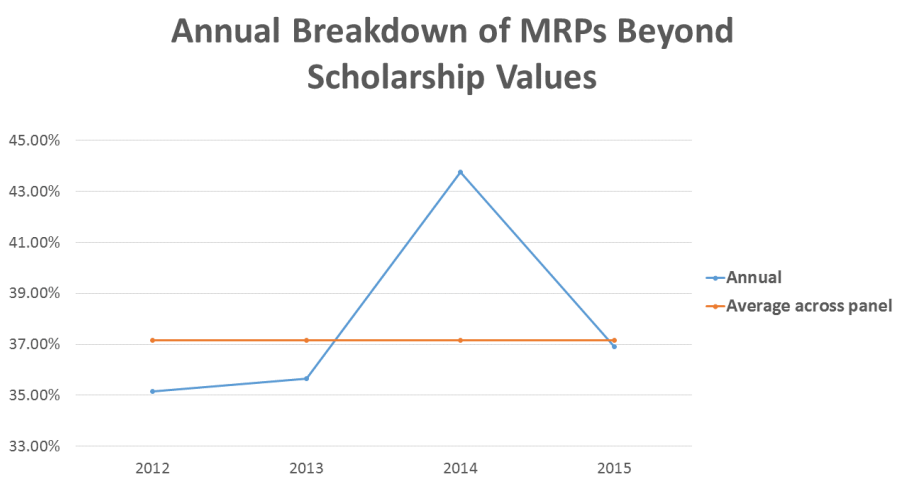

Figure 1: Percentages of teams each year with MRPs greater than combined maximum possible scholarship value.

These numbers may seem surprisingly low, particularly if one is reading this having internalized the preconceived notion that these college athletes are being egregiously exploited.

The results certainly surprised us. Lane, Nagel and Netz (2014), using their method, found that individual MRPs exceeded individual scholarship limits for around 66% of men's collegiate basketball players. Even though they do not estimate collective MRPs, it seems fair to surmise that if this was true for 66% of individual players, then aggregating across teams would lead to a somewhat similar number. If we make that assumption, we see that our MRP estimates are much lower.

As indicated in the limtations section, in modelling the production function to account for the way the data is actually generated, the mathematical equation for the marginal effects is such that if teams, based on their characteristics, have a high probability of an at-large bid (a high $\Phi(z_i'\delta)$), then their marginal products, and thus their MRPs, will be close to zero. There is a high correlation between teams with large revenues[11] and teams with high predicted probabilities of receiving at-large bids, as one would expect. If we assume that, for these schools, collective MRPs actually do exceed their scholarship values, then the overall proportion of school's for which this is true jumps to 51%. This is an appreciable increase, but still below what we think Lane (2014) would've found using a different methodology.

A Slight Modification

As we indicated in the limitations section, we want to compare the MRP estimates above with the estimates that arise when we model the production function a bit differently. In the above analysis, some of the schools with MRPs equal to zero because of their quality on the court, as reflected in a high $\Phi(z_i'\delta)$, include Duke, Kentucky, Gonzaga, and Villanova. Anyone even remotely familiar with NCAA basketball is cognizant of the fact that these are quality programs that produce perennial contenders year after year. In fact, Villanova was the 2016 NCAA champion, and Gonzaga was the 2017 runner-up. One would expect that the players on these teams contribute an appreciable amount to revenue.

Modelling the production function using the conventional probit framework ostensibly solves the issue, as the marginal effect of a performance variable p now becomes:

$$\frac{\partial Pr(y_i = 1|x_i, z_i)}{\partial p} = \beta_p \phi(x_i'\beta).$$

The exact same variables are used to estimate the production function in this setting, and the variables from the z_i vector are included in the new x_i vector above. The theoretical grounding is the same; this approach is simply less direct in that it does not account for the way the data is actually generated.

Following the same procedure, we find that estimated MRPs are much lower than the previous specification, ranging from essentially 0 to around $962,000, with a much lower mean of around $275,000. Across the panel, estimated collective MRPs exceed the combined maximum possible scholarship amount allocated to the team around 31% of the time, as opposed to the 37% we found previously. The results from this specification are remarkably consistent in one sense: many of the powerhouse programs have marginal products, and thus MRPs, that are essentially zero. Previously, it was easy to discern from the mathematical formulation which teams would have very low marginal products. With this specification, those same teams (those with high $\Phi(z_i'\delta)$) continue to display very low marginal products; the only difference is that this time they arise due to very low values for $\phi(x_i'\beta)$.

These results, particularly the consistently low MRPs for the players at these powerhouse schools across both specifications, were initially quite surprising to us and difficult to put into

[11] We somewhat arbitrarily define large-revenue schools as those reporting revenues greater than 10 million.

perspective. A bit of reflection, however, was helpful on that front. One of our immediate thoughts was the possibility that, perhaps for these renowned and storied programs, the players are really just cogs within an overarching system and sports culture on which they have little direct influence. That is to say, it really does not matter what players these schools put on the court, the die-hard fans would still patronize the product and support the team. Perhaps its the institution, the supposed innocence of these young kids out there giving it their all, and the notion of a "student-athlete" that really entices these fans to support their teams year after year, either through ticket purchases, cable subscriptions, team paraphernalia or the like. Anecdotally and historically, these programs certainly don't ever seem to lack support for their teams, whether they're filled with future NBA lottery picks, or just a collection of solid players.

A closer look at what we're measuring, and the law of diminishing marginal returns, gives rise to a more logical interpretation of our findings. The teams in question are at or near full-capacity in terms of the inputs they are combining to produce their output, that is, the performance statistics that have been included. The law of diminishing marginal returns dictates that, if the collection of players were to increase their level of performance, this increase will have a much smaller impact on the output, which in this case is the probability of a tournament appearance, than it otherwise would have if capacity were at a lower level. In fact, as we've seen, it may have no effect at all. In other words, adding performance or ability to an already great team will have a negligible effect on output, resulting in a negligible MRP. This explanation has a certain intuitive appeal.

Mathematically, conventional probit models bound the dependent variable between 0 and 1. Looking at Figure 2, a graph of the standard normal CDF, it is clear that as the probability gets closer to 0 or 1, the rate of change of the function gets closer to 0. Even though our approach differs from the conventional probit approach in that θ_i is modelled as a function of two distinct probabilties, the mathematical features of the model are consistent. Therefore, for schools with predicted probabilities close to 0 or 1, their marginal products will be close to 0, as we have found.

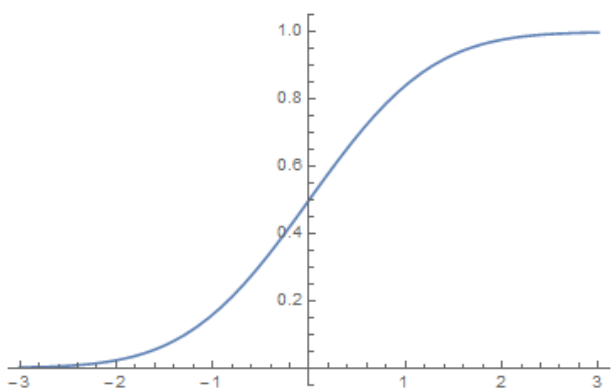

Figure 2: Standard Normal CDF with probability on the vertical axis. The function gets flater as the probability approaches both 0 and 1.

Other studies, such as Scully's (1974) original paper and the fairly recent Lane (2014) paper which also focuses on collegiate basketball, have most likely overstated MRPs. This is so because the Scully method assumes that the marginal effect is the same across teams. In doing so, and working with a linear regression, their results can be thought of in terms of averages. As seen earlier, the average collective MRP from our original method was over $400,000, a significant amount. What Scully, Lane and others have done is taken a number

akin to this one (by multiplying coefficients from 2 linear specifications, which are essentially averages) and then multiplied by some measure of a player's individual contribution to his team in order to obtain the player's MRP. This gives rise to a kind of incongruence, as they have taken an average that is based upon all the teams in their sample, and then multiplied by an individual contribution measure that is completely unique to that player's particular team. To treat this value as a player's MRP, that is the increase in revenue that would arise if he contributed to an increase in performance by his team, is incorrect, particularly if his team is operating at or near full capacity.

We believe the MRPs we have presented here are accurate in the sense that they are measuring what a MRP is defined to be. However, we cannot help but ask ourselves whether a MRP is actually telling us what we'd like to know, particularly as it relates to the aforementioned powerhouse schools. The MRPs in this paper tell us how an additional, rebound, steal, percentage of free throws made etc, contributed by a collection of players, impacts revenue through its effect on the probability of a tournament appearance. A conventional profit-maximizing firm will employ additional units of input/labor until the MRP of the last worker is equal to their marginal cost. Finding that a MRP, within our framework, is below our proxy for marginal cost tells us that an increase in performance statistics would not generate more in revenue than what the athletes cost the school. Does this mean that their current on-court performance does not exceed this marginal cost proxy? Some would argue that if it didn't, the school would do what a conventional profit-maximizing firm would do and cut down on the players, and by extension inputs, that they are employing. Others would argue that this is not necessarily the case, as the schools are not conventional firms, and they have an obligation to their student-body and their fans to field a team regardless of the economics. Additionally, schools make scholarship decisions based on a player's expected contribution, not their performance that ends up materializing over time. This further complicates the matter.

There is also the question of whether we should be comparing the MRPs we have calculated here to the combined scholarship value allocated to the team. The player's are already on the team, and so any increase in performance that they contribute should not technically cost the school anything. Our analysis would probably be more straightforward if we could isolate the direct causal impact of a specfic player/group of players on the revenues that the schools report. Scully (1974) also acknowledges this in his seminal paper. He says, "It would be better if the relationship between marginal revenue product and player performance could be estimated using the player as the unit of measurement". This, however, is much more difficult than it sounds.

Still and all, we can safely say that for a non-trivial 37% of teams across the panel, the players are collectively contributing more to revenues than the amount they cost their institutions. We make this assertion based on a consideration of the contrapositive. If they were not, the additional revenue generated from an increase in performance, their MRP, could not be greater than our marginal cost proxy. This however would contradict our finding that MRPs for these teams actually do exceed our marginal cost proxy.

There are two final observations that the reader may find useful when thinking about the value of the on-court performance of collegiate basketball players. Firstly, even though we have highlighted the average MRP of the teams in our panel, it is important to realize that this differs from the MRP of an average team. The latter,which the reader may find more interesting, tells us the additional revenue that would result if we improved the performance of an average team in our sample. Evaluating the marginal effects of our performance variables at the means of our vector-products $x_i'\beta$ and $z_i'\delta$, and proceeding in the same manner, we find that the MRP of a team with average characteristics is $\sim 395,000$, a significant amount.

Secondly, as discussed at length, throughout this paper we have been considering how

small increases in player contributions impact revenue. The reader may find it more informative to consider what the additional on-court production of the players on a great team would mean, in terms of additional revenue, to a team of lesser quality. We can use the predicted probabilities of a tournament appearance from our model to analyze such scenarios.

Recall our original definition of θ_i, the probability that our tournament appearance dummy variable is equal to 1:

$$\theta_i = \Phi(x_i'\beta) + \Phi(z_i'\delta)\left(1 - \Phi(x_i'\beta)\right).$$

We obtain predicted probabilities using linear predictions of $x_i'\beta$ and $z_i'\delta$ after estimation of the production function. These predicted probabilities are heavily skewed to the right, with a skewness measure of 1.75. Even up to the 25th percentile, the predicted probability of a tournament appearance for that team is less than 1%.

(1) Mean	(2) SD	(3) Min	(4) Max	(5) Skewness	(6) P25	(7) P50	(8) P75	(9) P95
0.193	0.310	7.16e-10	1	1.750	0.00534	0.0369	0.202	0.998

Table 5: Summary statistics for predicted probabilities across panel.

For a school around the 50th percentile, such as Wake Forest's team in 2014, the probability of making the NCAA tournament comes in at around 3.6%. We are interested in what would happen to this probability if we were to replace the performance statistics of Wake Forest with statistics around the 95th percentile of their respective categories, such as those of Duke's team in the same year. It is paramount that we only change the performance variables, as we are only interested in the impact of improvements on the court. Performing this exercise, we find that Wake Forest's predicted probability of making the tournament jumps from 3.6% to approximately 62%, which based on our estimates would result in additional revenue to the program of approximately $121,000. Therefore, all other things being equal, the on-court production of the Duke players, or the players on any other top-tier team, would have been worth $121,000 plus whatever the direct revenue contribution was of the Wake Forest players at the time.

For a school around the 75th percentile, such as Villanova's team in 2012, the probability of a tournament appearance was around 20%. Villanova's program improved tremendously over the next 4 years, as they went on to win the entire NCAA tournament in 2016. However, replacing their on-court performance with that of the North Carolina Tar Heels, a great team in 2012 and the overall champions in 2017, reveals that this increase in on-court performance results in a predicted probability of around 83%. This probability jump would generate additional revenue of over $131,000.

Looking at these distinct changes in probabilities does not yield infallible analysis. As we have seen, each school has different institutional traits, and so what applies to Wake Forest and Villanova in particular years may not apply in the same way to other schools with similar on-court performance. Also, there is no guarantee that the players from a top team would replicate the same on-court performance under different circumstances, including a different coach, different fans, different practice facilities, among other factors. That being said, it is fair to say that ceteris paribus, the on-court production of players on upper echelon teams would be worth approximately $121,000 and $131,000 plus the ex-post revenue contributions of the players on Wake Forest and Villanova respectively.

7 Conclusion

The contentious debate regarding collegiate athlete compensation in the United States is not likely to be resolved any time soon. This paper contributes to the field of empirical research surrounding this debate by trying to assess the on-court revenue contribution of D1 collegiate basketball players, through the estimation of marginal revenue products. As we discussed, Title IX rules and the structure of the NCAA as an institution render the prospect of isolating individual player MRPs somewhat irrelevant. Moreover, we argue in this paper that individual MRPs estimated in the work of Scully and others do not represent what the authors purport, because of computational flaws. We focus on collective MRPs of teams during the years 2012-2015 inclusive.

We find that for 37% of the teams in our panel, the players are collectively contributing more to revenues than the amount they cost their institutions, given that their collective MRPs exceed the maximum possible scholarship amounts that could be allocated to the teams, the latter being upper bounds on the cost of these athletes to their schools. We argue that, based on how a MRP is defined within this context, we cannot say with certainty that the current on-court revenue contributions of the remaining teams do not exceed their costs.

The average MRP across the panel is around $406,000, while the MRP of an average team is approximately $395,000. We also find that if we replace the players on teams like Wake Forest in 2015 and Villanova in 2012 with the players on top teams like Duke or North Carolina in the respective years, ceteris paribus, the probability of an appearance in the NCAA tournament for the weaker team jumps significantly. For Wake Forest, it jumps from around 3% to approximately 61%, while for Villanova, it jumps from 20% to 83%. These increases in probability would be associated with additional revenue of approximately $121,000 and $131,000 respectively.

There are many other issues that are germane to the pay-for-play debate. For example, there are other revenue issues, such as the fact that schools receive pecuniary benefits not only from players' on-court performance, but from their names and likenesses. It is also important to remember that many of these players come from poor, disadvantaged backgrounds, and aside from the handful that are able to pursue sports as a vocation, many find themselves in an quandary. They are not good enough to play professionally, and at the same time, the committment of time and energy demanded by their coaches prevents them from paying adequate attention to the academic responsibilities necessary to complete their degree.[12] These realities give rise to moral and ethical issues.

There are also issues that arise from the myriad of onerous rules enforced by the NCAA. Transfer students have to sit out a year before they are allowed to compete, while the same restriction of movement does not apply to coaches. Scholarships are not guaranteed for the entirety of the 4 year period in which a student-athlete is expected to be enrolled, which means that an injured player could be denied the opportunity to complete his/her education without recourse. There is also the rule that NBA prospects have to play a year in college before being eligible for the draft, an agreement between the NCAA and NBA that many view as collusion.

Thinking about some of the aforementioned problems, and the feeling of mistrust that they foster among the NCAA, the schools and the student-athletes, it is easy to convince yourself that something is not right. One cannot help but get the feeling that some of the basic tenets on which the United States is supposedly built, such as fairness and liberty, are

[12] There have been a multiplicity of academic fraud cases in NCAA history, see most recently with the 2017 NCAA champion UNC: http://www.cbssports.com/college-football/news/north-carolina-ad-lays-out-academic-fraud-defense-ncaa-overcharged-tar-heels/

being violated. In trying to incubate different ways to address these issues, however, the job gets very complex, very quickly. As we have seen in this paper, the task of trying to ascertain the on-court revenue contributions of players brings with it a ream of complications that one needs to account for or acknowledge. If we then add to that the multiplicity of off-court factors that one needs to consider, and the additional layers of complexity that accompany them, it is easy to see why these issues have persisted for so long.

Having said that, there are many areas ripe for additional research. One could, for example, seek to come up with a model to value monetarily the off-court contribution of student athletes, whether it be in revenue sports such as basketball, or less profitable sports. One could also try and find a way to use counterfactuals in order to measure the causal impact of players on revenue, which would allow us to estimate MRPs using the players as the unit of measurement in lieu of performance statistics. Additionally, as highlighted in the introduction, the results in this paper should be viewed from a partial equilibrium perspective. One may want to explore the implications for general equilibrium analysis, for example not only considering changes in performance statistics for a single school in isolation, but within the context of similar changes for every other D1 collegiate basketball team.

8 Appendix

- Descriptive statistics for variables included in production function:

VARIABLES	(1) Obs.	(2) Mean	(3) S.D	(4) Min	(5) Max
Simple Rating System	1,393	-0.618	9.824	-35.85	28.72
Strength of Schedule	1,393	-0.257	5.106	-12.71	12.13
Away Wins	1,393	4.813	2.680	0	14
Free-Throw Percentage	1,393	0.693	0.0373	0.541	0.814
Total Rebounds	1,393	1,123	134.9	717	1,711
Steals	1,393	208.9	44.23	101	430
True Shooting Percentage	1,393	0.529	0.0297	0.433	0.616
Free-Throws per FGA	1,393	0.260	0.0387	0.146	0.415
Coach Win/Loss Record	1,385	0.515	0.128	0	0.939
Coach # of Tournament Appearances	1,393	2.636	4.481	0	31
Top Conference Dummy	1,393	0.213	0.410	0	1
Top Conference · Conference Wins	1,393	1.902	4.032	0	18
Non-top Conference · Non-Conference Wins	1,393	5.857	4.424	0	21

- Descriptive statistics for variables included in revenue regression (excluding conference dummies that accounted for different revenue sharing agreements):

VARIABLES	(1) Obs.	(2) Mean	(3) S.D	(4) Min	(5) Max
Strength of Schedule	1,393	-0.257	5.106	-12.71	12.13
Pace	1,393	66.10	2.897	57.10	77.60
Team Revenue	1,373	4.097e+06	5.147e+06	347,223	4.584e+07
Tournament Appearance Dummy	1,393	0.194	0.395	0	1
Top Conference Dummy	1,393	0.213	0.410	0	1
Arena Capacity	1,377	8,158	4,945	878	33,000

- Descriptive statistics for combined scholarship values and combined cost of attendance values. These were calculated by taking the maximum allowable individual scholarship and the estimated cost of attendance for each school, and multiplying by the number of participants on each team.

VARIABLES	(1) Obs.	(2) Mean	(3) S.D	(4) Min	(5) Max
Combined Full GIA	1,394	428,054	147,873	0	719,712
Combined COA	1,394	439,429	153,478	0	738,480

References

[1] Atkinson, S. E., Stanley, L. R., and Tschirhart, J. (1988). Revenue sharing as an incentive in an agency problem: An example from the national football league. *The Rand journal of economics*, pages 27–43.

[2] Brown, R. W. and Jewell, R. (2006). The marginal revenue product of a women's college basketball player. *Industrial Relations: A Journal of Economy and Society*, 45(1):96–101.

[3] Carmichael, F., Thomas, D., and Ward, R. (2000). Team performance: the case of english premiership football. *Managerial and decision Economics*, pages 31–45.

[4] Gustafson, E., Hadley, L., Ruggiero, J., et al. (1999). Alternative econometric models of production in major league baseball. *Sports economics: current research.*, pages 95–107.

[5] Hausman, J. A. and Leonard, G. K. (1997). Superstars in the national basketball association: Economic value and policy. *Journal of Labor Economics*, 15(4):586–624.

[6] Kahn, L. M. (1993). Managerial quality, team success, and individual player performance in major league baseball. *Industrial & Labor Relations Review*, 46(3):531–547.

[7] Krautmann, A. C. (1999). What's wrong with scully-estimates of a player's marginal revenue product. *Economic Inquiry*, 37(2):369–381.

[8] Lane, E., Nagel, J., and Netz, J. S. (2014). Alternative approaches to measuring mrp: Are all mens college basketball players exploited? *Journal of Sports Economics*, 15(3):237–262.

[9] Martin, R. E. (2004). Tuition discounting without tears. *Economics of Education Review*, 23(2):177–189.

[10] McKenzie, R. B. and Sullivan, E. T. (1987). Does the ncaa exploit college athletes-an economics and legal reinterpretation. *Antitrust Bull.*, 32:373.

[11] Schofield, J. A. (1988). Production functions in the sports industry: an empirical analysis of professional cricket. *Applied Economics*, 20(2):177–193.

[12] Scully, G. W. (1974). Pay and performance in major league baseball. *The American Economic Review*, 64(6):915–930.