Published in Yale Journal of Economics, Spring 2014

Returns To Schooling: A College Athlete's Perspective

Alex Dombrowski [0]

Abstract

Two decades ago 90% of the college draftees into the NBA each season had just finished their senior year of college. Today that number is only 30%. Collegiate basketball players are making the jump to the NBA earlier on, after their freshman, sophomore, or junior season. This paper studies a player's decision of when to go pro through both an empirical and theoretical framework. I estimate returns to schooling by comparing the earnings of two groups of NBA players: those who went pro out of high school or after freshman year of college from 1989 to 2005 versus those who went pro after freshman year from 2006 to 2012. A new rule made 2005 the last year players could move directly from high school to the NBA. The rule forced players to wait at least one year after their high school graduation before going pro. Hence the latter group contains players who were forced to take an additional year of school. I find no significant difference in earnings between the two groups.

1 Introduction

In economics, returns to schooling is commonly thought of as the expected change in earnings due to an additional year of school. So for example, an undergraduate considering whether or not to pursue graduate study may like to know the expected salary difference from getting those additional years of education.

This paper approaches returns to schooling from the perspective of a collegiate athlete. In particular, I consider collegiate basketball players, but as a motivating example and illustration of how this phenomenon is widespread, consider the football player Matt Barkley. Barkley, class of 2013, was a quarterback at USC who had a phenomenal junior season. Speculators thought he would forego his senior year at USC to go right into the NFL. Instead, Barkley opted to stay at USC for his senior year, intending to go pro right after. Unfortunately, his senior year performance wasn't nearly as compelling and he was drafted much lower and with a contract estimated to be worth millions of dollars less than if he had went pro after junior year. In Barkley's case, one would say that his returns to schooling for that year were *negative*.[1]

Figure 1 shows the composition of the NBA draft from 1989 to 2012 and illustrates how players have been going pro earlier. In each draft, the 30 teams select two players a piece for a total of 60 players moving from college into the NBA. Throughout the early 1990s, around 90% of those drafted were seniors. This percentage has fallen dramatically over the last two decades. In 2012 only 35% of those drafted were seniors.

---

[0]University of California, Berkeley Class of 2014. Triple major: Economics, Mathematics, and Statistics. a.dombrowski@berkeley.edu Written Fall 2013. Advised by Professor David Card.
I sincerely thank Professor Card for his time and support. As always, thanks Mom and Dad.

[1]There are of course other benefits to finishing senior year (e.g. earning the degree) but here I focus solely on maximizing the player's earnings as a professional athlete.

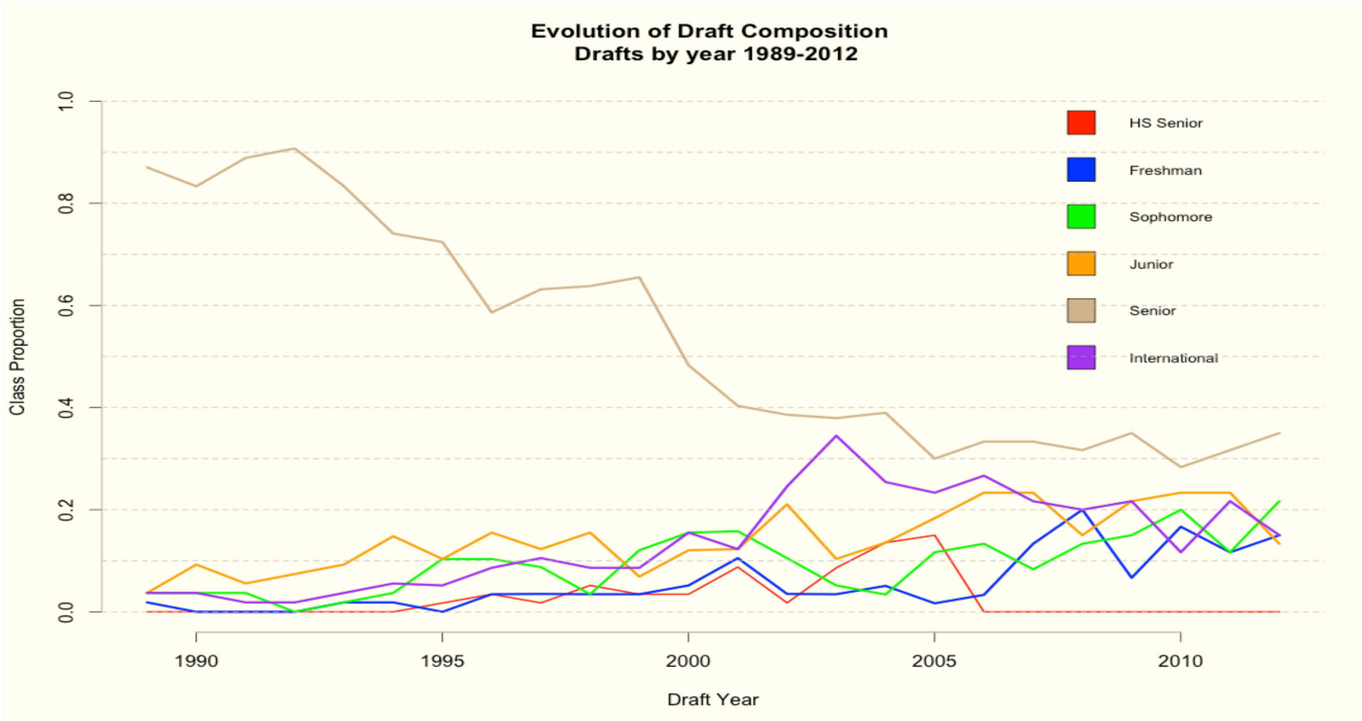

Figure 1: Of the players drafted into the NBA each year, the percentage that are seniors has been declining. Underclassmen and international players are making up a larger share of the draft class.

So why strategize about when to leave college and go pro? *If a collegiate basketball player enters the NBA draft, he foregoes his remaining college eligibility regardless of whether or not he is drafted.*[2] For example, if a freshman enters the draft then he can no longer play collegiate basketball even if he doesn't get drafted.

The paper is laid out as follows: Section 2 gives a brief literature review. Section 3 is an overview of the NBA. Section 4 is a study testing the effect of playing an additional year of collegiate basketball on earnings. Sections 5, 6, and 7 contain the theoretical framework. Section 8 concludes.[3]

2 Literature Review

There are a few notable papers which discuss contracts and early entry. Li and Rosen (1998) analyze when and why contracts are made early. Winfree and Molitor (2007) analyze returns to schooling for baseball players. In particular, they focus on a recent high school graduate's decision of whether to go to college or go directly into the Major League. Arel and Tomas (2012) view declaring for the NBA draft as exercising an American style put option early.

[2] If the player did not sign with an agent, has never applied for a previous NBA draft, and withdraws his name from the draft by the deadline, then he can retain his eligibility. However, the NCAA rules only allow players to enter the draft once without losing eligibility.

[3] All figures throughout this paper are original.

3 Overview of NBA

The NBA can be viewed as a labor market where each year the 30 teams (firms) hire 60 players (employees) from a pool of players nationwide. The NBA players have a labor union, the National Basketball Players Association (NBPA) which negotiates rules with the League, which is comprised of the commissioner and team owners. Every several years the League and NBPA form a new Collective Bargaining Agreement (CBA) which lays the foundation for how the NBA is governed. The CBA contains information about how the draft works, how basketball revenue is allocated among the players and the League, and the minimum and maximum amount a team can pay its players. For example, the 1995 CBA introduced the rookie pay scale, which structured how incoming players would be payed. Players drafted in the first round are guaranteed a two-year contract, followed by two one-year team options. Players drafted in the second round are not guaranteed contracts. Also, being picked later in the draft results in lower pay.

The NBA draft is held annually at the end of June. Starting in 1989, the draft instated a two round system, in which each team selects one player in each round. This new drafting system is the main reason I focus on draft data beginning in 1989. I constructed a data set of 1382 players who were drafted from 1989 to 2012.[4] Data was collected from www.basketball-reference.com and verified using basketball.realgm.com. During the 24 draft years 1989-2012, 53.11% of the total 1382 players drafted were seniors, 14.40% were juniors, 14.26% were international, 9.62% were sophomores, 5.79% were freshmen, and 2.82% were high school seniors.

4 Returns to Schooling: An Additional Year

The 2005 NBA Collective Bargaining Agreement made 2005 the last season a high school player could go directly into the NBA. Beginning in 2006, a player had to be one year removed from high school before going into the NBA.[5] Figure 2 shows how the number of high school seniors and college freshmen has evolved in drafts dating back to 1989. The blue line falls to zero in 2006 as a result of the new CBA. In 2007, the red line spikes from 2 to 8. This spike represents the 2006 high school class that was forced to play a year in college, along with a couple of 2006 high school graduates who would've chosen to play their freshman year even without the new rule. Hence the red line post 2005 can be thought of as a merging of the pre 2006 blue line and the pre 2006 red line. In this study, I explore differences in earnings between the following two groups of players: Freshmen and High School Seniors drafted 1989-2005 versus Freshmen drafted 2006-2012. The first group, which I call the "Pre-law" group, has 67 players. The second group, the "Post-law" group, has 52 players. Of these 119 players, 3 never played a game in the NBA and were removed from the data.[6] Table 1 provides a detailed comparison of these two groups.

[4] Although in 2012 there were 30 total teams, at the beginning of this time period there were only 27. Consequently, earlier drafts saw less than 60 players drafted.

[5] The player doesn't *have* to attend college for that year, however, attending college for that year is standard.

[6] Ousmane Cisse (2001 draft), Ricky Sanchez (2005 draft), and Keith "Tiny" Gallon (2010 draft) never played in the NBA.

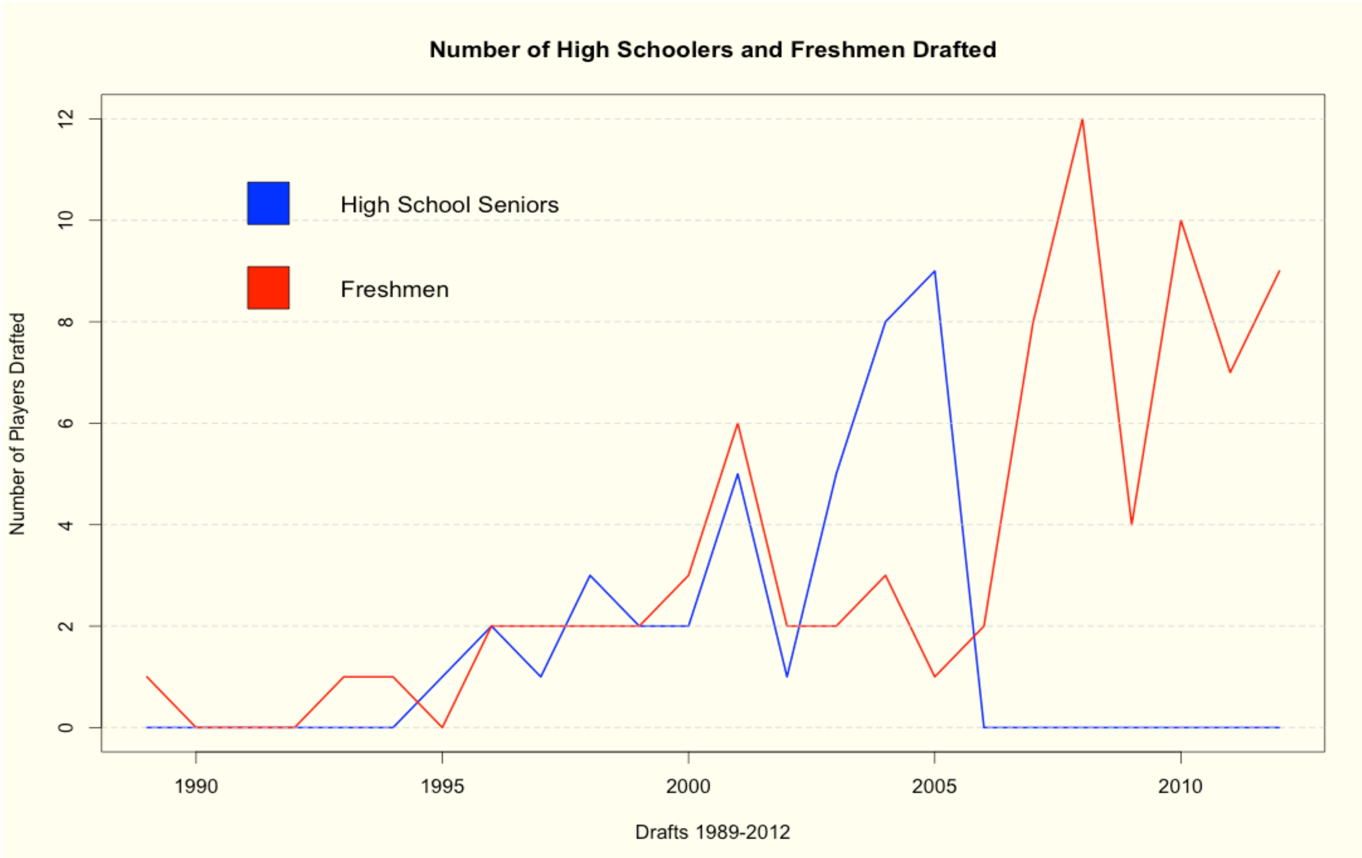

Figure 2: The two groups merge in 2006.

Table 1: Comparison of Pre-law and Post-law group performance.

	1989-2005			2006-2012		
Career Statistics	Mean	Median	SD	Mean	Median	SD
Games Played	548.9	560.0	323.9	193.4	170.5	129.9
Minutes Played	15740	14250	11557.1	4883	3568	4176.8
Points	7556	5709	6750.9	2201.0	1400.0	2281.2
Total Rebounds	3129.0	2516.0	2785.2	884.0	601.0	846.6
Total Assists	1302.0	840.0	1464.4	425.9	156.0	549.6
FG Shooting	0.4439	0.4540	0.092	0.4454	0.4455	0.0883
3 PT Shooting	0.2626	0.3130	0.1541	0.2471	0.2985	0.1458
FT Shooting	0.7007	0.7410	0.15571	0.6858	0.7330	0.1506
Minutes Per Game	23.80	24.90	9.9822	21.95	21.45	9.3834
PPG	10.91	9.90	6.4508	9.425	8.250	5.7259
Rebound PG	4.679	4.200	2.73399	4.015	3.650	2.5015
Assists PG	1.858	1.500	1.5396	1.862	1.200	1.8257
Career Length (years)	9.269	9.000	4.269	3.481	3.000	1.862

From Table 1, we can informally compare the statistics of the two groups. Since the Pre-law group has players with longer careers than those in the Post-law group, it may be more difficult to compare across statistics like points and total rebounds. To compare across these, note that the last row shows that the median career length of the Pre-law group is three times that of the Post-law group. Hence multiplying the Post-law group's total points by three may give a reasonable number to compare to the Pre-law group's total points. For a finer comparison use minutes played instead of career length. The median minutes played of the Pre-law group is four times that of the Post-law group. Hence numbers could be scaled appropriately by three to four to make more accurate comparisons.

4.1 Analysis of Earnings

Salary data was collected from www.basketball-reference.com and checked against http://www.eskimo.com/∼pbender/.[7] Salary was put into real 2013 dollars using CPI numbers from FRED. Figure 3 shows how earnings have evolved throughout this time period. All 116 players are plotted, with each player represented by a line. The red lines are the Pre-law players and the blue lines are the Post-law players. The figure is meant to give a general sense of how earnings progress as the players gain more years of experience. Notable players Kevin Garnett (1995 draft), Kobe Bryant (1996 draft), and Kevin Durant (2007 draft) are highlighted.

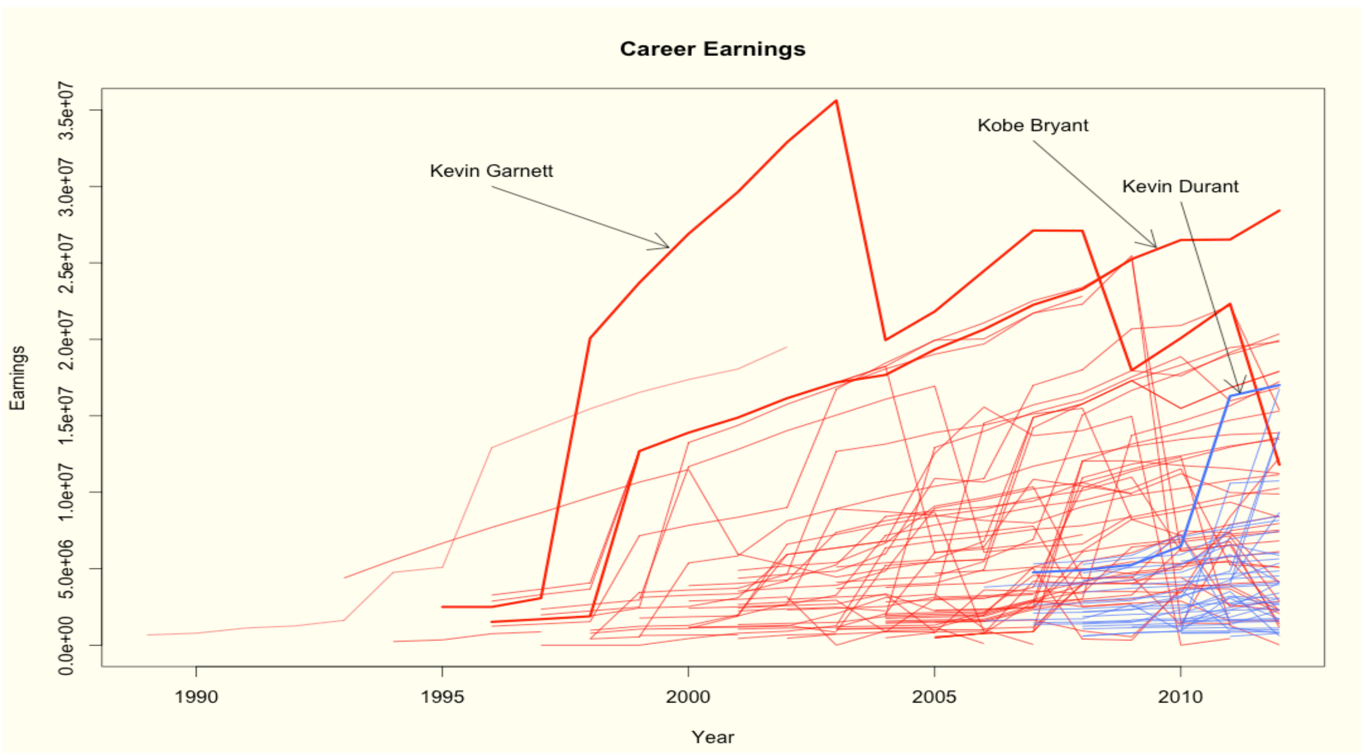

Figure 3: Career earnings in real 2013 dollars.

[7]For the NBA lockouts in 1998 and 2011, I used the full years' earnings, not the prorated salary.

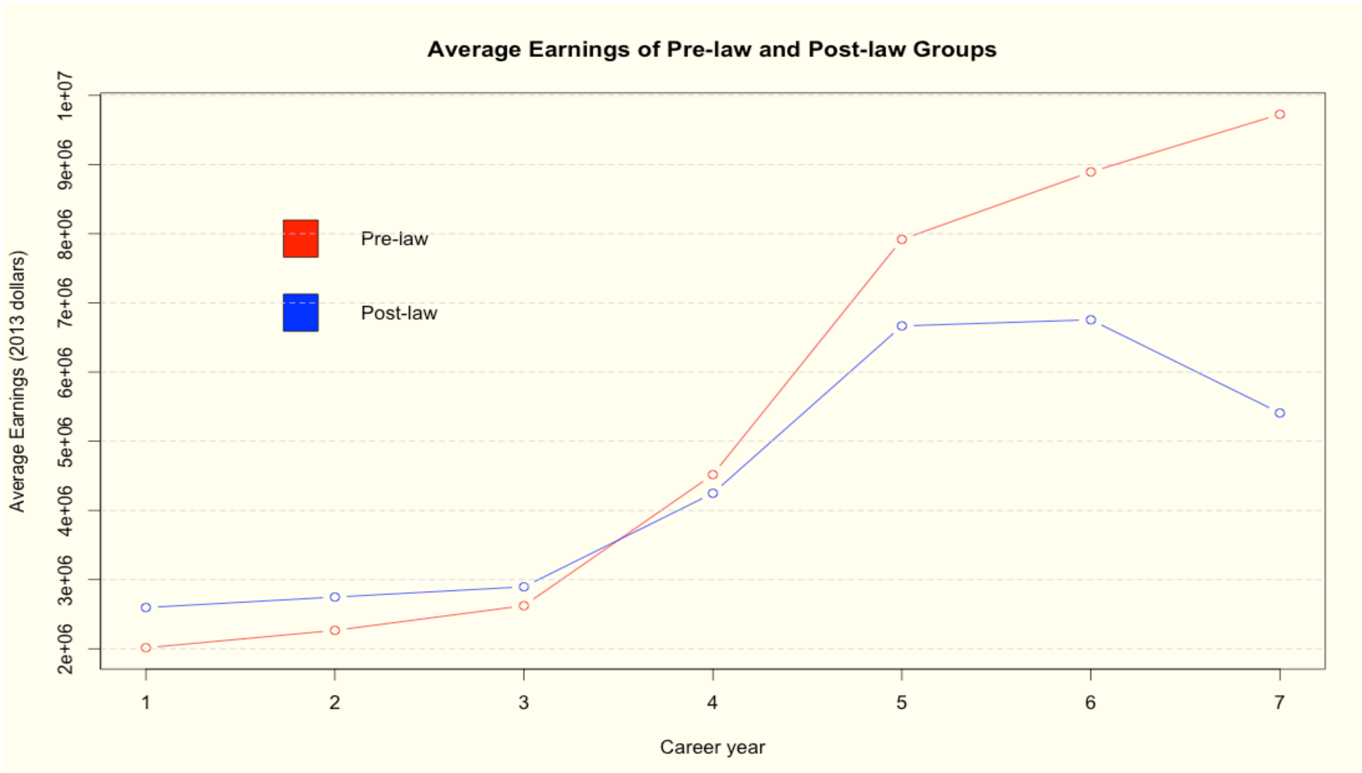

Figure 4: The Post-law group earned more in each of the first 3 years, then was worse off.

Figure 4 is a condensed version of Figure 3. Figure 4 illustrates the difference in earnings between the Pre-law and Post-law groups. Average earnings in real 2013 dollars are plotted against career year. So for example, the Post-law group (blue line) for career year one corresponds to the average earnings of players in that group during their rookie year.

In career year one, the Post-law group earned on average \$580,000 more than the Pre-law group. In career year two, the Post-law group earned on average \$480,000 more than the Pre-law group. In career year three, the Post-law group earned on average \$270,000 more than the Pre-law group. In career years 4 and 5, the Pre-law group out-earned the Post-law group by \$270,000 and \$1,250,000 respectively.[8]

Draftees often hire agents to negotiate their rookie contract. The rookie pay scale is not completely rigid: The team can pay between 80% and 120% of the salary specified by the rookie pay scale. However, nearly all contracts end up at 120%. Arel and Tomas (2012) find that of the players drafted in the first round between 2006 and 2012, 98% had contracts for 120% of the amount specified by the rookie pay scale. Therefore I didn't control for the quality of the agent in the analysis.

The following two subsections use a Two Sample Z Test and Mann-Whitney Test to determine if the difference in earnings between the two groups is significant.

[8] I don't analyze career years 6 and 7 in detail because there are less than ten players from the Post-law group who played six or more years, and only two who played all seven.

4.1.1 Parametric: Two Sample Z Test

The test assumes $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{N}(\mu_X, \sigma^2)$ and $Y_1, \ldots, Y_m \overset{iid}{\sim} \mathrm{N}(\mu_Y, \sigma^2)$ where σ^2 is estimated by a pooled variance

$$s_p^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{m+n-2} \qquad \text{where} \qquad s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{1}$$

In this case, the observations X_i, Y_i are the annual earnings of each player for some specified year. The observations are reasonably independent and identical. The histograms in Figure 5 show the distribution of earnings for each group in the first two years. Though the data are not convincingly normal, the next section's analysis uses the nonparametric Mann-Whitney test and gives very similar results to this two sample z test.

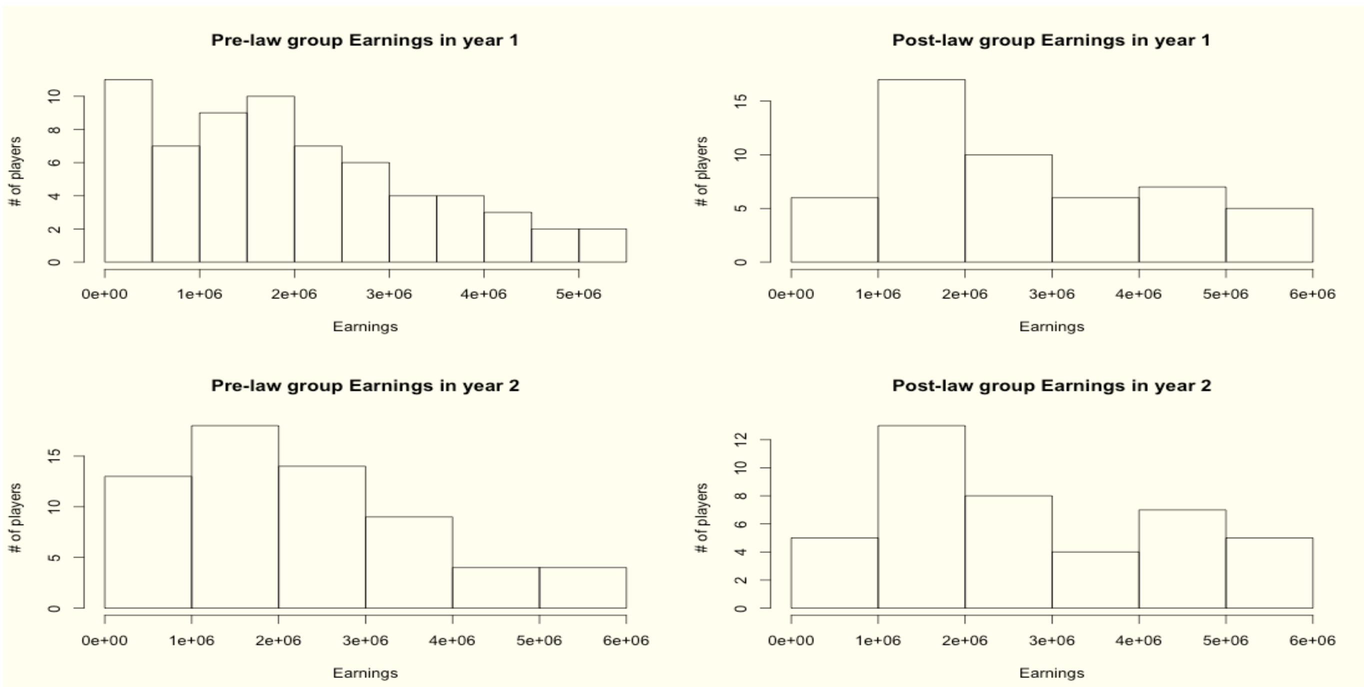

Figure 5: Earnings in years 1 and 2 for the Pre-law group and Post-law group.

Table 2 gives the results from this two sample z test (two sided) for career years one through five. In each test, the null hypothesis is

$$H_0 : \mu_{Xj} = \mu_{Yj} \qquad j = 1, 2, 3, 4, 5 \tag{2}$$

where for example μ_{Xj} is the mean of the distribution of earnings in career year j for the Pre-law group.

The difference in earnings between the two groups during career year one is significant at the 5% level. The difference in earnings between the two groups during career year two is just shy of being significant at the 10% level. The other tests do not yield significant results.

Table 2: Results of Two Sample Z Test.

Career year	t-statistic	p-value
1	2.12749991	0.03337857
2	1.6017363	0.1092139
3	0.8169752	0.4139426
4	-0.3234246	0.7463737
5	-0.9189319	0.3581312

Table 3: Results of Mann-Whitney Test.

Career year	t-statistic	p-value
1	-2.06083935	0.03931837
2	-1.6164613	0.1059946
3	-0.7660718	0.4436336
4	-0.7233642	0.4694561
5	-0.7660718	0.4436336

4.1.2 Nonparametric: Mann-Whitney Test

The Mann-Whitney test is nonparametric, which means it makes no assumptions about the underlying distribution of the observations. The test instead ranks the observations from smallest to largest and compares the sum of ranks. The null hypothesis is that the treatment has no effect, where in this case the treatment is the additional year of school the Post-law group experienced. Table 3 has the results, which are similar to those in Table 2. Again, the Post-law group's average earnings in career year one are significantly higher than the Pre-law group's average earnings in career year one.

4.2 Linear Regression Model

The following model is the form for the regressions:

$$\ln \text{Earnings}_i = \beta_0 + \beta_1 S_i + \beta_2 \text{Ability}_i + \beta_3 \text{Draft Pick}_i + \beta_4 \text{Experience}_i + \epsilon_i \tag{3}$$

The dependent variable *Earnings* varies from regression to regression. S is an indicator variable taking on zero for the Pre-law group and one for the Post-law group. *Ability* is measured by rookie year statistics: minutes played, points, assists, and rebounds. *Experience* is measured by a player's total career points. *Experience* can be thought of as long term ability, controlling for whatever the ability regressor doesn't. Table 4 gives the results.

Table 4 has four regressions. The choice of regressions is unconventional in the sense that all regressions use the same set of regressors, but have different dependent variables. Instead of settling on one measure for earnings, it seemed more appropriate to give several measures. The dependent variables in regressions (1) and (2) are *Career Average Yearly Earnings* and *Year 1 Earnings*. The dependent variables in regressions (3) and (4) are *Earnings in first 2 years* and *Earnings in first 3 years*.

For all regressions, the coefficient of *Year of School Dummy* is not significant. Hence in these tests, there is no evidence that supports that one group had significantly different earnings than the other group.

The results of regressions (2), (3), and (4) are very similar, mostly because the of the rookie pay scale, which was introduced in 1995.[9] Under the new contract system, earnings in years one and two are highly correlated.[10]

[9] Although this data set of 116 players encompasses the time period 1989-2012, only 3 players were drafted before 1995: Shawn Kemp 1989, Shawn Bradley 1993, and Dontonio Wingfield 1994.

[10] Of the 116 players, 104 played at least 2 years. The correlation between these players' earnings in years one and two of their career is 0.996

Table 4: The effect of an additional year of school on earnings.

	Dependent variable:			
	(1)	(2)	(3)	(4)
Year of School Dummy	-0.111	0.091	0.088	0.027
	(0.104)	(0.284)	(0.319)	(0.360)
Minutes Played (Rookie)	-0.00002	-0.001^{*}	-0.001^{*}	-0.001
	(0.0002)	(0.001)	(0.001)	(0.001)
Rebounds (Rookie)	0.001^{*}	0.001	0.002	0.002
	(0.0005)	(0.001)	(0.002)	(0.002)
Assists (Rookie)	0.0002	0.001	0.001	0.001
	(0.001)	(0.002)	(0.002)	(0.002)
Points (Rookie)	-0.0001	0.001	0.001	0.001
	(0.0004)	(0.001)	(0.001)	(0.001)
Draft Pick	-0.024^{***}	-0.074^{***}	-0.073^{***}	-0.068^{***}
	(0.004)	(0.011)	(0.012)	(0.013)
Total Career Points	0.0001^{***}	-0.00004	-0.00004	-0.00004
	(0.00001)	(0.00003)	(0.00003)	(0.00003)
Constant	14.895^{***}	15.919^{***}	16.650^{***}	17.055^{***}
	(0.140)	(0.383)	(0.414)	(0.460)
Observations	116	116	104	96
R^2	0.770	0.411	0.372	0.298
Adjusted R^2	0.755	0.372	0.326	0.243
Residual Std. Error (df=108)	0.445	1.217	1.297 (df=96)	1.400 (df=88)
F Statistic (df=7; 108)	51.585^{***}	10.747^{***}	8.116^{***} (df=7; 96)	5.348^{***} (df=7; 88)

*p<0.1; **p<0.05; ***p<0.01

(1) Career Average Yearly Earnings (log)
(2) Year 1 Earnings (log)
(3) Earnings in first 2 years (log)
(4) Earnings in first 3 years (log)

In each regression, *Draft Pick* is significant at 1%. This is not surprising for regressions (2) and (3) since as a player is drafted lower his pay will decline according to the rookie pay scale. It's not as obvious that *Draft Pick* is significant in regression (1) where *Career Average Yearly Earnings* is the dependent variable. The coefficient of *Draft Pick* in regression (1) is -0.024, which is smaller in absolute value than the corresponding coefficient in the other three regressions. This is consistent with the rookie pay scale's influence on average career yearly earnings becoming diluted because of the expiration of the contract and room for more variability in earnings in later years.

The regressions highlight the importance of draft pick on average yearly earnings throughout a player's career. Regression (1) estimates that being drafted one position later leads to a 2.4% decrease in average career earnings, on average. Table 5 summarizes how each group was drafted. A Mann-Whitney test on the draft number for the Pre-law and Post-law groups yielded a p-value of 0.16. Hence there is no evidence that suggests draft positions are significantly different between the two groups.

Table 5: Summary of Draft Pick Number

Group	Min	1st Qu.	Median	Mean	3rd Qu.	Max
Pre-law	1.00	6.00	13.00	17.98	26.00	56.00
Post-law	1.00	4.00	11.00	14.75	22.00	49.00

References

[1] Arel, B. and Tomas, M. (2012). The NBA Draft: A Put Option Analogy. *Journal of Sports Economics*, Vol. 13, No. 3 : pp. 223-249

[2] FRED, Federal Reserve Economic Data, Federal Reserve Bank of St. Louis: Consumer Price Index for All Urban Consumers: All Items (CPIAUCSL); http://research.stlouisfed.org/fred2/series/CPIAUCSL/ ; accessed September 2013

[3] Hlavac, Marek (2013). stargazer: LaTeX code and ASCII text for well-formatted regression and summary statistics tables. R package version 4.5.3. http://CRAN.R-project.org/package=stargazer

[4] Presman, E. and Sonin, I. (1973). The Best Choice Problem for a Random Number of Objects. *Theory of Probability and Its Applications*, Vol. 17, No. 4 : pp. 657-668

[5] 2005 Collective Bargaining Agreement, http://www.nbpa.org/cba/2005

[6] http://basketball.realgm.com/

[7] http://www.eskimo.com/∼pbender/

[8] http://www.nytimes.com/2013/04/28/sports/football/matt-barkley-stayed-at-usc-and-paid-for-it-with-late-draft-pick.html (if copy and pasting this link, delete the space before 'with')

[9] www.basketball-reference.com/draft/

5 Optimal Entry

A major concern for collegiate players is when to go pro. Players may play all four years of college ball, or may opt to enter the draft early. In the NBA, once a player declares for the draft, he can never play college basketball again regardless of whether he's drafted to an NBA team or not.[11] A very talented underclassman may want to enter the draft early for many reasons. He may have had an outstanding season or his team may have won the national championship. He could've received a prestigious award like MVP or may be concerned his performance will be worse during his later years of college. He also could get injured in a later season. There are many cases where collegiate athletes went pro at the "wrong time". Hence we would like to analyze optimal entry.

Consider a collegiate player wanting to go pro. He wants to be drafted high, not low (i.e. in the first round rather than the second or third round).
Let

$$P_H := \text{probability of being drafted high}$$

$$P_L := \text{probability of being drafted low}$$

$$E_H := \text{high career earnings (when drafted high)}$$

$$E_L := \text{low career earnings (when drafted low)}$$

where $E_H > E_L > 0$.
Consider a player seeking to go pro and assume this player won't go undrafted.[12] That is,

$$P_H + P_L = 1 \tag{4}$$

Suppose that P_H and P_L depend only on that player's skill level, S. Hence

$$P_H = P_H(S) \qquad \text{and} \qquad P_L = P_L(S) \tag{5}$$

where these two functions have the property

$$\frac{dP_H}{dS} > 0 \qquad \text{and} \qquad \frac{dP_L}{dS} < 0 \tag{6}$$

since a player's chance of being drafted high should be monotonically increasing in his skill level. Likewise, a player's chance of being drafted low should decline as the player's skill increases. A player's skill level will vary over time as he progresses through college. So letting y denote years of college experience we have

$$S = S(y) \qquad\qquad y \in [0, 4] \tag{7}$$

where $y = 2$ for example corresponds to 2 years of experience. Let the player have a utility function, $u(e)$, where e is career earnings as a professional with $u'(e) > 0$ and $u''(e) < 0$. In other words, the player's utility is increasing and concave in his career earnings. Hence to maximize utility, it's sufficient to maximize expected career earnings, $\mathbb{E}(e)$. Thus we have the optimization problem,

$$\max_y \ \mathbb{E}(e) \ = \max_y \Big[P_H(S) \cdot E_H + P_L(S) \cdot E_L \Big] \tag{8}$$

From (4) and (5), (8) becomes

$$\max_y \Big[P_H\big(S(y)\big) E_H \ + \ \Big(1 - P_H\big(S(y)\big)\Big) E_L \Big] \tag{9}$$

[11] In the NBA, if the player did not sign with an agent, has never applied for a previous NBA draft, and withdraws his name from the draft by the deadline, then he can retain his eligibility. However, the NCAA rules only allow players to enter the draft once without losing eligibility

[12] Relaxing this assumption leads to the same conclusion. See Appendix for proof.

The solution to (9) involves making $P_H\big(S(y)\big)$ as large as possible, which occurs when $S(y)$ is as large as possible. Let's verify that this is indeed the case. The first order condition says,

$$\begin{aligned}\frac{d}{dy}\Big[P_H\big(S(y)\big)E_H \;+\; \Big(1 - P_H\big(S(y)\big)\Big)E_L\Big] &= E_H P_H'\big(S(y)\big)S'(y) - E_L P_H'\big(S(y)\big)S'(y)\\ &= (E_H - E_L)P_H'\big(S(y)\big)S'(y) = 0 \qquad (10)\end{aligned}$$

By assumption, $E_H > E_L$ and so $E_H \neq E_L$. Also by assumption, $P_H(S)$ is monotonically increasing, and hence $P_H'(S) \neq 0$ for all S.
Therefore,

$$(E_H - E_L)P_H'\big(S(y)\big)S'(y) = 0 \qquad \Longleftrightarrow \qquad S'(y) = 0 \qquad (11)$$

Let y^* be such that

$$S'(y^*) = 0 \qquad \text{and} \qquad S''(y^*) < 0 \qquad (12)$$

Then y^* denotes the number of years of college experience that maximizes skill level.

<u>*Claim*</u>:

$$\max_{y_L \leq y \leq y_R} \mathbb{E}(e) = \mathbb{E}(e)\big|_{y=y^*}$$

provided that y^* is the absolute maximum of $S(y)$.

<u>*Proof*</u>: We already know $S'(y^*) = 0$ which makes y^* a critical point of $\mathbb{E}(e)$. So we must check

$$\frac{d^2y}{dy^2}\left[\mathbb{E}(e)|_{y=y^*}\right] < 0$$

From (10),

$$\frac{d}{dy}\Big[(E_H - E_L)P_H'\big(S(y)\big)S'(y)\Big] = (E_H - E_L)\Big[P_H'\big(S(y)\big)S''(y) + S'(y)P_H''\big(S(y)\big)S'(y)\Big]$$

Letting $y = y^*$,

$$= (E_H - E_L)\Big[P_H'\big(S(y^*)\big)S''(y^*) + \big(S'(y^*)\big)^2 P_H''\big(S(y^*)\big)\Big]$$

Since by assumption $S'(y^*) = 0$ this becomes

$$= (E_H - E_L)P_H'\big(S(y^*)\big)S''(y^*)$$

By assumption, $E_H > E_L$ and $S''(y^*) < 0$. Also, $P_H'(S) > 0$ for all S. So in particular, $P_H'\big(S(y^*)\big) > 0$.
Therefore,

$$(E_H - E_L)P_H'\big(S(y^*)\big)S''(y^*) < 0 \qquad (13)$$

From (13) we conclude

$$\max_{y} \mathbb{E}(e) = \mathbb{E}(e)\big|_{y=y^*} \qquad \text{for} \qquad y_L < y < y_R$$

where y_L and y_R and the leftmost and rightmost y values of the interval.[13] Now evaluate expected earnings at the boundary and compare to $\mathbb{E}(e)\big|_{y=y^*}$:

Let y_B generically denote either y_L or y_R. Then

$$\begin{aligned}\mathbb{E}(e)\big|_{y=y^*} > \mathbb{E}(e)\big|_{y=y_B} &\Longleftrightarrow P_H\big(S(y^*)\big)E_H + \Big[1 - P_H\big(S(y^*)\big)\Big]E_L > P_H\big(S(y_B)\big)E_H + \Big[1 - P_H\big(S(y_B)\big)\Big]E_L\\ &\Longleftrightarrow P_H\big(S(y^*)\big)E_H - P_H\big(S(y^*)\big)E_L > P_H\big(S(y_B)\big)E_H - P_H\big(S(y_B)\big)E_L\\ &\Longleftrightarrow E_L\Big[P_H\big(S(y_B)\big) - P_H\big(S(y^*)\big)\Big] > E_H\Big[P_H\big(S(y_B)\big) - P_H\big(S(y^*)\big)\Big]\end{aligned}$$

[13] For this model, take $y_L = 0$ and $y_R = 4$.

Since $E_L < E_H$,

$$\iff P_H\big(S(y_B)\big) - P_H\big(S(y^*)\big) < 0$$

Since $P_H'(S) > 0$,

$$\iff S(y^*) > S(y_B)$$

$\blacksquare$

Therefore to maximize expected career earnings, and hence utility, a player should go pro when his college skills are best. This conclusion is simple. The complication is that a player doesn't know when his skills will be best. He may have a great freshman year and expect to get better, but actually have worse years later in college. Players don't know when $S(y)$ is at its maximum, just like stock market agents don't know when the price of a stock is at its maximum. This naturally leads to random walks and optimal stopping rules, which are considered next.

6 Stochastic Model

In the previous section, we concluded that a player can optimize expected career earnings by strategically entering the draft when his skill level is highest. Thus we should analyze how skill level moves from year to year and find when it's likely to be highest.

Let S_n be skill level after n years of experience and let $\{S_n, n = 1, 2, 3, 4\}$ be a random walk defined by

$$S_{n+1} = S_n + \mu + \epsilon \tag{14}$$

where $\epsilon \sim N(0, \sigma^2)$ and the jumps from year to year are independent. Epsilon allows for variation in the amount of skill gained. Mu represents a drift and is thought of as the baseline amount of skill gained from year to year. It's more often than not that $\mu > 0$ and so we make this assumption in the algebraic solution, though the simulations in section 6.2.3 allow for $\mu \leq 0$. Note that $S(y)$ is more naturally thought of as being continuous since a player's skill evolves perhaps continuously throughout his career. However, we can consider our discrete time analysis as using only the values $S(1), S(2), S(3)$, and $S(4)$ of the continuous time $S(y)$. The probability skill increases over a year is given by

$$\begin{aligned} P(S_{n+1} > S_n) &= P(S_n + \mu + \epsilon > S_n) \\ &= P(\epsilon > -\mu) \\ &= 1 - \Phi(-\mu) \\ &= \Phi(\mu) \end{aligned}$$

where

$$\Phi(\mu) = \int_{-\infty}^{\mu} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} \, dx \tag{15}$$

The conditional expectation and variance are

$$E(S_{n+1}|S_n) = E(S_n + \mu + \epsilon \mid S_n) = S_n + \mu \tag{16}$$

$$Var(S_{n+1}|S_n) = Var(S_n + \mu + \epsilon \mid S_n) = \sigma^2 \tag{17}$$

Next, consider the maximum value of the random walk. Define "probability functions" $\mathbb{P}_i$ to be

$$\mathbb{P}_i := P\Big(\max\{S_1, S_2, S_3, S_4\} = S_i\Big) \qquad i = 1, 2, 3, 4$$

We would like to find expressions for $\mathbb{P}_i$. For $\mu = 0$, the expressions are simple. However, for $\mu > 0$, the expressions are messier. The following solves for $\mathbb{P}_i$ when $\mu = 0$, gives an outline of the solution for $\mathbb{P}_i$ when $\mu > 0$, and provides a simulation which illustrates the estimated solution for $\mathbb{P}_i$ for all μ.

6.1 No drift ($\mu = 0$)

Let's find expressions for $\mathbb{P}_i,\ i = 1, 2, 3, 4$ when $\mu = 0$. To simplify notation, let $p := \Phi(\mu)$, the probability the random walk goes up. Since $\mu = 0$, we could use $p = \Phi(\mu) = \Phi(0) = 1/2$, however, the next subsection generalizes these expressions, so we don't explicitly use $1/2$ here. The following lemma illustrates the logic used in this section.

Lemma 9.1) For the walk $S_{n+1} = S_n + \epsilon$, P(max$\{S_0, S_1, S_2\} = S_1$) = p(1-p).

Verification:

$$\begin{aligned} P(\max\{S_0, S_1, S_2\} = S_1) &= P(S_1 > S_0,\ S_1 > S_2) \\ &= P(S_1 > S_2 | S_1 > S_0) P(S_1 > S_0) \\ &= P(S_1 > S_1 + \mu + \epsilon)\Phi(\mu) \\ &= P(\epsilon < -\mu)\Phi(\mu) \\ &= (1 - \Phi(\mu))\Phi(\mu) \\ &= (1 - p)p \end{aligned}$$

■

Now let's find $\mathbb{P}_1$, the probability that S_1 is the maximum (i.e. the player's skill is highest after freshman year). S_1 is the maximum in four cases: The walk goes

1. down down down
2. down down up, with the two down steps being larger than the up step
3. down up down, with the first down step being larger than the up step
4. down up up, with the down step being larger than the two up steps

Hence

$$\mathbb{P}_1 = (1-p)^3 \ + \ p(1-p)^2 \cdot P\Big(N(0, 2\sigma^2) > N(0, \sigma^2)\Big) \ + \ p(1-p)^2 \cdot P\Big(N(0, \sigma^2) > N(0, \sigma^2)\Big) \tag{18}$$

$$+ p^2(1-p) \cdot P\Big(N(0, \sigma^2) > N(0, 2\sigma^2)\Big)$$

where we've used the fact that $N(0, \sigma^2) + N(0, \sigma^2) \stackrel{d}{=} N(0, 2\sigma^2)$. Also, by symmetry, if $X \sim N(0, \sigma_X^2)$ and $Y \sim N(0, \sigma_Y^2)$, then

$$P(X < Y) = P(X > Y) = 1/2 \qquad \forall\ \sigma_X, \sigma_Y \tag{19}$$

From (18), using (19) and simplifying gives

$$\mathbb{P}_1 = (1-p)^3 + \frac{p}{2}(1-p)^2 + \frac{p}{2}(1-p)^2 + \frac{p^2}{2}(1-p)$$

$$\rightarrow \mathbb{P}_1 = 1 - 2p + \frac{3}{2}p^2 - \frac{1}{2}p^3 \tag{20}$$

Continuing in this manner, let's find $\mathbb{P}_2, \mathbb{P}_3, \mathbb{P}_4$. S_2 is the maximum in two cases: The walk goes

1. up down up, with the down step begin larger than the last up step
2. up down down

Hence

$$\mathbb{P}_2 = p^2(1-p) \cdot P\Big(N(0, \sigma^2) > N(0, \sigma^2)\Big) + p(1-p)^2$$

$$\rightarrow \mathbb{P}_2 = p - \frac{3}{2}p^2 + \frac{1}{2}p^3 \tag{21}$$

S_3 is the maximum in two cases: The walk goes

1. down up down, with the up step greater than the first down step
2. up up down

Hence

$$\mathbb{P}_3 = p(1-p)^2 \cdot P\Big(N(0,\sigma^2) > N(0,\sigma^2)\Big) + p^2(1-p)$$
$$\rightarrow \mathbb{P}_3 = \frac{1}{2}p - \frac{1}{2}p^3 \tag{22}$$

S_4 is the maximum in four cases: The walk goes

1. up up up
2. down down up, with the up step larger than the sum of the two down steps
3. down up up, with the sum of the two up steps larger than the down step
4. up down up, with the last up step larger than the down step

So

$$\mathbb{P}_4 = p^3 + p(1-p)^2 \cdot P\Big(N(0,\sigma^2) > N(0,2\sigma^2)\Big) + p^2(1-p) \cdot P\Big(N(0,2\sigma^2) > N(0,\sigma^2)\Big)$$
$$+p^2(1-p) \cdot P\Big(N(0,\sigma^2) > N(0,\sigma^2)\Big)$$
$$\rightarrow \mathbb{P}_4 = \frac{1}{2}p + \frac{1}{2}p^3 \tag{23}$$

As a sanity check, let's verify that

$$\mathbb{P}_1 + \mathbb{P}_2 + \mathbb{P}_3 + \mathbb{P}_4 = 1 \tag{24}$$

Summing (20)-(23),

$$\begin{aligned}
&= \Big(1 - 2p + \frac{3}{2}p^2 - \frac{1}{2}p^3\Big) + \Big(p - \frac{3}{2}p^2 + \frac{1}{2}p^3\Big) + \Big(\frac{1}{2}p - \frac{1}{2}p^3\Big) + \Big(\frac{1}{2}p + \frac{1}{2}p^3\Big) \\
&= 1 + \Big(-2 + 1 + \frac{1}{2} + \frac{1}{2}\Big)p + \Big(\frac{3}{2} - \frac{3}{2}\Big)p^2 + \Big(-\frac{1}{2} + \frac{1}{2} + -\frac{1}{2} + \frac{1}{2}\Big)p^3 \\
&= 1
\end{aligned}$$

Equations (20)-(23) are only valid for $p = 1/2$, as assumed in this section. Evaluating these four expressions when $p = 1/2$ gives

$$\mathbb{P}_1 = 5/16 \qquad \mathbb{P}_2 = 3/16 \qquad \mathbb{P}_3 = 3/16 \qquad \mathbb{P}_4 = 5/16$$

Hence when a player has no drift in skill, the probability his skill will be highest after freshman year is 5/16, after sophomore year is 3/16, after junior year is 3/16, and after senior year is 5/16. The assumption of no drift may hold for some players, but it's more revealing to incorporate drift and build the full model (14).

6.2 Probability functions with positive drift ($\mu > 0$)

We want to find expressions similar to (20)-(23) which relax the assumption of zero drift. The process for constructing the analogue to (18) is the same. That is, there are still the same four ways S_1 could be the maximum. However, the second, third, and fourth ways will have a different form.
Three problems need to be solved: cases 2, 3, and 4 from (18).

I) The probability the walk goes down down up, with the two down steps being larger than the up step.
II) The probability the walk goes down up down, with the first down step being larger than the up step.
III) The probability the walk goes down up up, with the down step being larger than the two up steps.

After finding I, II, and III, these values can be substituted in (18) to get the new $\mathbb{P}_1$. Similarly, substituting these values (or their complements) into the old expressions for the other $\mathbb{P}_i$ will give the new probability functions. We begin with case II.

6.2.1 Case II

The task here is to find the probability of the walk going down, up, down, with the first down step being larger than the up step. Keeping with notation, let the probability of an up jump be $p := \Phi(\mu)$. Let A be the event the walk goes down, then up, then down. Let B be the event that the down step is larger than the up step. Then

$$P(AB) = P(A) \cdot P(B|A) = p(1-p)^2 P(B|A) \tag{25}$$

So we must solve $P(B|A)$, the probability the down jump is larger then the up jump, given the walk went down then up on those first two steps. Let

$$X := \text{size of the up jump} \qquad \text{and} \qquad Y := \text{size of the down jump}$$

Figure 6 shows how the sampling is done.

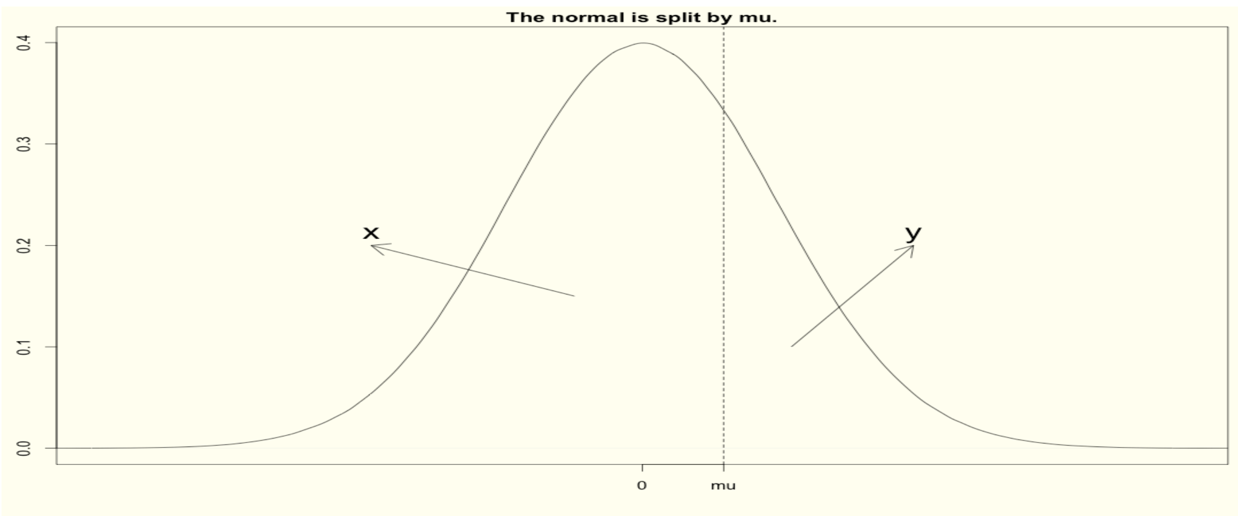

Figure 6: An up jump is drawn from the left side of μ. A down jump is drawn from the right side of μ.

First, make X and Y into densities by scaling the original $N(0, \sigma^2)$ by the appropriate factor. The density of X is given by

$$\widetilde{f}_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}x^2} \frac{1}{\Phi(\mu)} \qquad -\mu \leq x < \infty \tag{26}$$

The density of Y is given by

$$\widetilde{f}_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}y^2} \frac{1}{1-\Phi(\mu)} \qquad \mu \leq y < \infty \tag{27}$$

Since we're concerned with $|X - \mu|$ and $|Y - \mu|$, shift the densities so the support is $[0, \infty)$. So the density of X becomes

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \frac{1}{\Phi(\mu)} \qquad x \geq 0 \tag{28}$$

The density of Y is now given by

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{y+\mu}{\sigma})^2} \frac{1}{1-\Phi(\mu)} \qquad y \geq 0 \tag{29}$$

Using (28) and (29), compute $P(X < Y)$:

$$\begin{aligned}
P(X<Y) &= \iint_R f_{X,Y}(x,y)\ dA \qquad\qquad R = \{(x,y) : x \geq 0,\ y > x\} \\
&= \iint_R f_X(x) f_Y(y)\ dA \\
&= \int_{x=0}^{x\to\infty} \int_{y=x}^{y\to\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \frac{1}{\Phi(\mu)} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{y+\mu}{\sigma})^2} \frac{1}{1-\Phi(\mu)}\ dy\ dx \\
&= \int_{x=0}^{x\to\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \frac{1}{\Phi(\mu)} \left[\int_{y=x}^{y\to\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{y+\mu}{\sigma})^2} \frac{1}{1-\Phi(\mu)}\ dy\right] dx
\end{aligned}$$

The inner integral involves a $N(-\mu, \sigma^2)$, which can be shifted and scaled to get a $N(0,1)$.

$$\begin{aligned}
&= \int_{x=0}^{x\to\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \frac{1}{\Phi(\mu)} \cdot \frac{1}{1-\Phi(\mu)} \left[1 - \Phi\left(\frac{x+\mu}{\sigma}\right)\right]\ dx \\
&= \frac{1}{\Phi(\mu)(1-\Phi(\mu))} \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \cdot \left[1 - \Phi\left(\frac{x+\mu}{\sigma}\right)\right]\ dx \\
&= \frac{1}{\Phi(\mu)(1-\Phi(\mu))} \left[\int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\ dx - \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \cdot \Phi\left(\frac{x+\mu}{\sigma}\right)\right]
\end{aligned}$$

The first integral can be shifted and scaled. The second we do not solve explicitly:

$$= \frac{1}{\Phi(\mu)(1-\Phi(\mu))} \left[\Phi(\mu/\sigma) - \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \cdot \Phi\left(\frac{x+\mu}{\sigma}\right)\right] \tag{30}$$

The expression in (30) can be substituted into (25) for $P(B|A)$.
As a sanity check on (30), plug in $\mu = 0$ and $\sigma = 1$. Then (30) reduces to

$$\frac{1}{(1/2)(1-1/2)} \left[\Phi(0) - \int_0^\infty \phi(x)\Phi(x)\ dx\right] = 4(1/2 - 3/8) = 1/2$$

This is exactly what is expected if $\mu = 0$ since the up jump and down jump come from the same distribution.

6.2.2 Cases I and III

I outline the solution for cases I and III, but do not solve explicitly for them. This is because the next section gives the full approximate probability functions through simulation, which are a lot more enlightening than the algebraic derivations. For case I, we want to solve $P(Y+Y > X)$. First, use convolution to find the density of $Y+Y$, then set up an integral as in the previous case. For case III, we want to solve $P(Y > X+X)$. Again, use convolution to find the density of $X+X$, then set up an integral.

6.2.3 Simulated Probability Functions

Figure 7 plots the probability functions $\mathbb{P}_1, \mathbb{P}_2, \mathbb{P}_3$, and $\mathbb{P}_4$ for various levels of σ. All plots have μ along the x-axis.

Note that for each plot in Figure 7, when $\mu = 0$, the probability functions coincide with the numbers from section 6.1: $\mathbb{P}_1 = 5/16 = \mathbb{P}_4$ and $\mathbb{P}_2 = 3/16 = \mathbb{P}_3$. The plots include values of negative drift, just to include players who may be best at the start of college, then steadily decline throughout their college career. The top left plot is for a player with $\sigma = 0.5$. This player has low variation above and beyond his usual drift from season to season. The orange line in the top left plot shows that if this player has a $\mu > 1$, it's very likely his skills will be highest after senior year. However, if the player's drift is between 0 and 1/2, the orange line is much lower and so the probability his skill level is highest before senior year is more substantial (sum of the heights of the blue, green, and red lines). The bottom right plot is for a player with

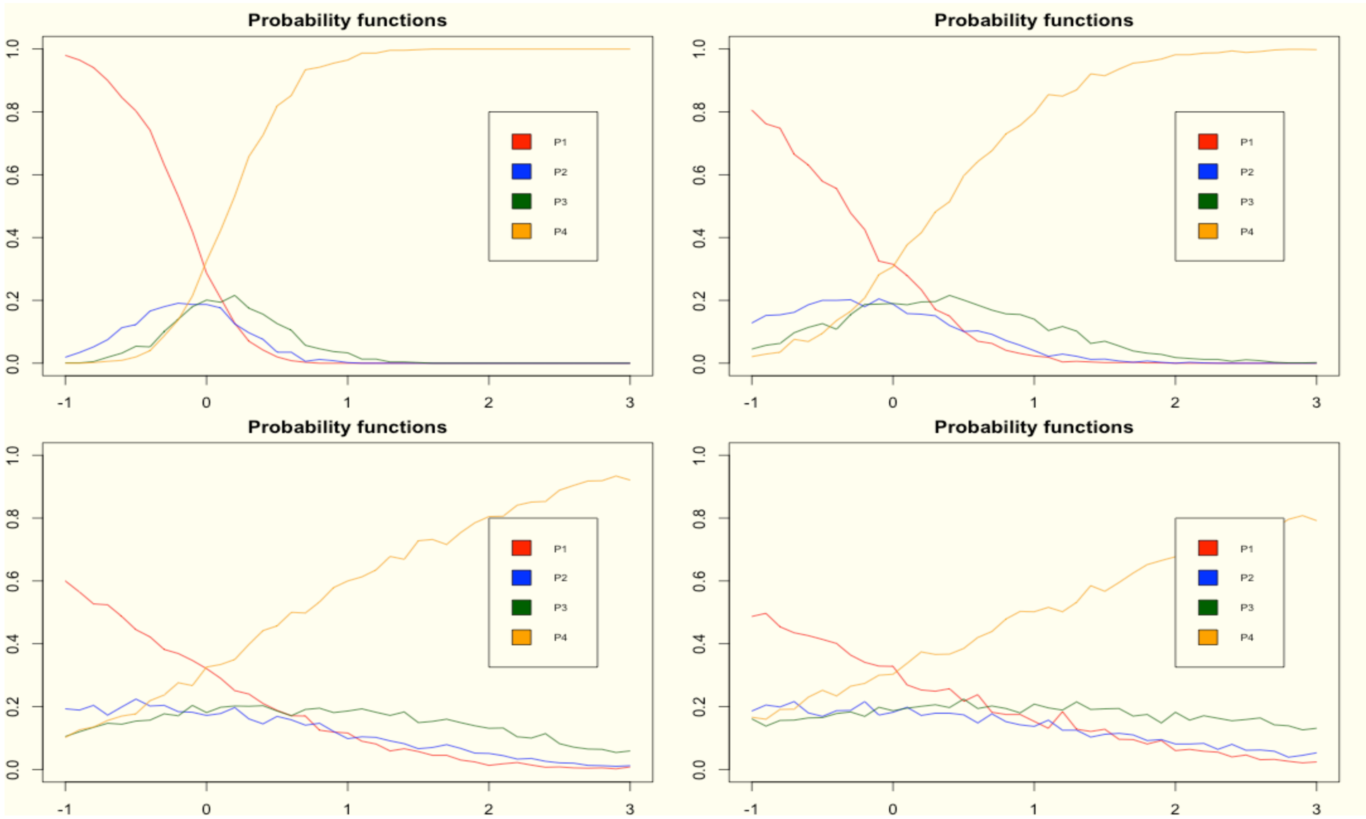

Figure 7: Top left is $\sigma = 0.5$. Top right is $\sigma = 1$. Bottom left is $\sigma = 2$. Bottom right is $\sigma = 3$. R Code is given in the appendix.

high volatility in his skills above and beyond the season to season drift ($\sigma = 3$). The orange line is still monotonic, yet increases much slower. The probability a high volatility player is best after senior year is less than the probability a low volatility is best after senior year for all $\mu > 0$.

6.3 The Stochastic model as a predictive model

From Figure 7, we could give an estimate as to when the player is most likely to have the highest skills, based on his σ and μ. To use Figure 7 as a predictor, we'd like to measure a player's μ and σ. I don't pursue this idea here, rather suggest it as potential future work.

One way is to look at the player's high school statistics or the player's game by game statistics in his first year of college. For each year in high school say, use statistics like the player's points, rebounds, etc to construct a proxy for skill level that year. Then for each of the four years, we'd have a number which corresponds to skill level. Plot these four skill levels versus years 1, 2, 3, and 4. The best fit line through these points could give an estimate for μ and σ: The slope of the best fit line would be $\hat{\mu}$ and the sum of squared residuals would be $\hat{\sigma}$. For each player, we could construct a $\hat{\mu}$ and $\hat{\sigma}$ and compare across players. So for example a $\hat{\sigma}$ in the first quartile of all players' $\hat{\sigma}$'s would classify that player as a low volatility player, whereas a $\hat{\sigma}$ in the fourth quartile of all players' $\hat{\sigma}$'s would classify that player as a high volatility player. The same could be done for the drift values. This would allow us to compare players and predict when a player's skill level would most likely be highest, relative to other players.

6.4 Threshold Idea

Section 5 determined a player can maximize expected career earnings by entering the draft when his skills are the best (i.e. by maximizing $S(y)$). Section 6 showed that if the player has a positive drift, then he is most likely to be best after his senior year, regardless of the actual value of the drift and regardless of the volatility in ϵ. Hence if a player has positive drift, which I suspect almost all do, we should not expect players to enter the draft before senior. So why is this occurring in practice?

To reconcile this, I suggest that players do indeed want to maximize their success by strategically entering the draft when their skills are best, however, if a player is very talented he may already be good enough to go pro. That is, perhaps there is some "skill threshold" players want to reach before going pro. If they attain this threshold before senior year, then they enter the draft before senior year. For example, after a player's team wins the national championship or after the player receives a prestigious award or after the player averages more than a certain number of points in a season. A player coming off a big year receives widespread attention which puts him in the spotlight and may make him more likely to go pro. The player may feel like his chances of being drafted are especially good, despite the fact that his skills may indeed improve if he stays in college for an additional year.

7 The Optimal Stopping Problem

The decision to stop playing college ball and go pro is a stopping problem. Since a player wants to optimize earnings with this decision, it is an optimal stopping problem. Figure 8 shows how after completing each year in college a player must choose whether to stop (S) and try to go pro or continue (C) playing in college. Stopping after year i leads to career earnings of e_i.

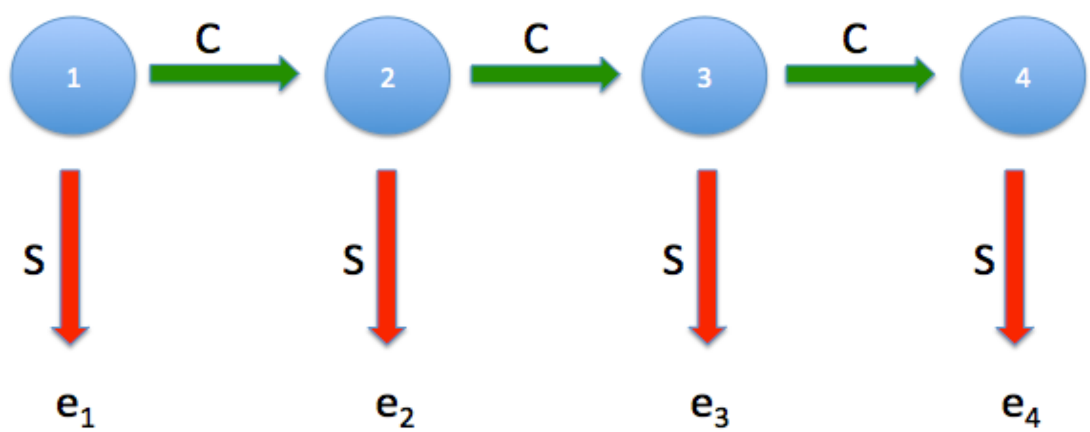

Figure 8: At each node a player must choose whether to continue playing college ball or stop and go pro.

Figure 8 captures the player's dilemma, but it simplifies the problem because it doesn't consider the probability a player is drafted to the NBA. Not all early entrants are drafted[14] and so this approach should incorporate the chance of being drafted.

Consider the finite state markov chain in Figure 9 with state space

$$S = \{0, 1, 2, 3, 4, d1, d2, d3, d4, NBA, ND\}$$

[14]Using my data set of 856 players from 1989 to 2012 who remained early entry, 35.5% went undrafted. Arel and Tomas (2012) find that from 2006 to 2010, 38% went undrafted.

where

(i) State 0 means the player has 0 years of college experience, 1 means the player has 1 year of college experience, etc.

(ii) d_j := declare for draft with j years of experience, $j = 1, 2, 3, 4$

(iii) NBA := player was drafted to the NBA. ND := player was not drafted to the NBA (both absorbing)

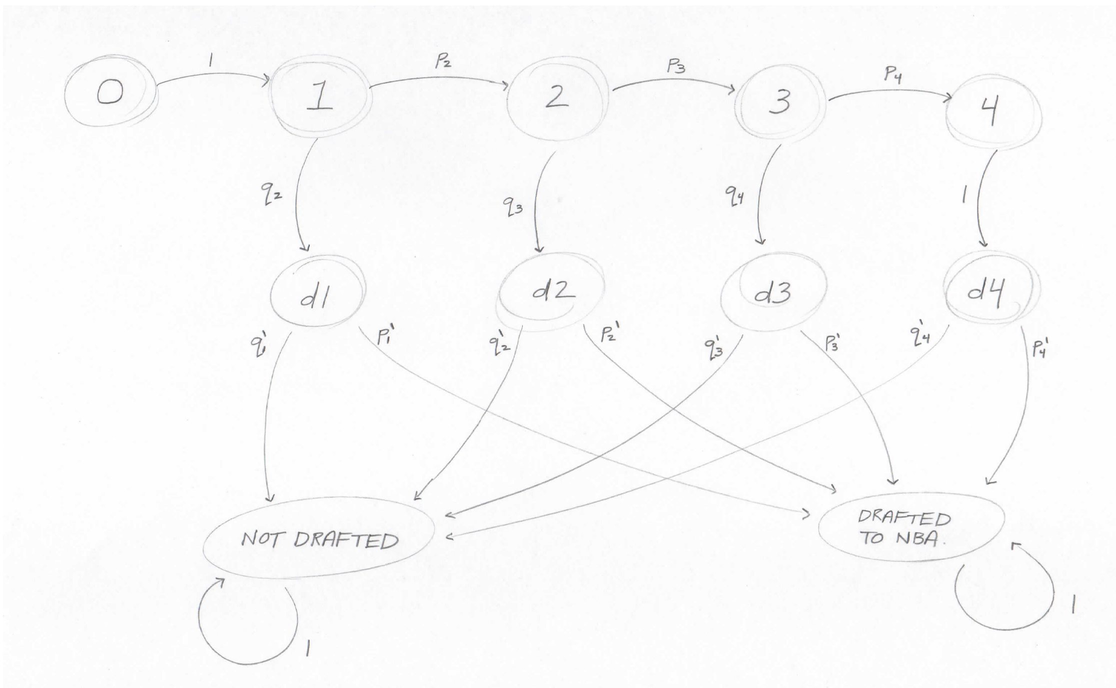

Figure 9: A player may or may not be drafted.

The stochastic matrix is

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_2 & 0 & 0 & q_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_3 & 0 & 0 & q_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & p_4 & 0 & 0 & q_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p'_1 & q'_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p'_2 & q'_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p'_3 & q'_3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p'_4 & q'_4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

P has dimension 11 by 11, where the states run across the top and down the side in the order 0, 1, 2, 3, 4, d1, d2, d3, d4, NBA, ND. So for example, the $(1, 2)$ entry of P gives the probability of moving from state 0 to state 1, which is 1 in Figure 9 since a player must be at least a freshman to declare for the draft.

The k^{th} step stochastic matrix ($k \geq 6$) is

$$P^k = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & q_2p_1' + p_2q_3p_2' + p_2p_3q_4p_3' + p_2p_3p_4p_4' & q_2q_1' + p_2q_3q_2' + p_2p_3q_4q_3' + p_2p_3p_4q_4' \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & q_2p_1' + p_2q_3p_2' + p_2p_3q_4p_3' + p_2p_3p_4p_4' & q_2q_1' + p_2q_3q_2' + p_2p_3q_4q_3' + p_2p_3p_4q_4' \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & q_3p_2' + p_3q_4p_3' + p_3p_4p_4' & q_3q_2' + p_3q_4q_3' + p_3p_4q_4' \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & q_4p_3' + p_4p_4' & q_4q_3' + p_4q_4' \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_4' & q_4' \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_1' & q_1' \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_2' & q_2' \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_3' & q_3' \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_4' & q_4' \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The entry of concern in this matrix is the $(1, 10)$ entry, which represents the probability of going from state 1 to state NBA:

$$P^k(1, NBA) = q_2p_1' + p_2q_3p_2' + p_2p_3q_4p_3' + p_2p_3p_4p_4' \qquad (k \geq 6) \tag{31}$$

That is, (31) is the probability a player who just finished freshman year will be in the NBA by the time he graduates. $P^6(1, NBA)$ is the probability he makes it to the NBA after freshman year (q_2p_1') plus the probability he makes it to the NBA after sophomore year ($p_2q_3p_2'$) plus the probability he makes it after junior year ($p_2p_3q_4p_3'$) plus the probability he makes it after senior year ($p_2p_3p_4p_4'$). Hence $P^6(1, NBA)$ is the probability he eventually makes it to the NBA. Clearly, a player would like this probability to be as high as possible. Since $p_i + q_i = 1$ for $i = 2, 3, 4$, substitute $q_i = 1 - p_i$ in (31) to get

$$P^6(1, NBA) = (1 - p_2)p_1' + p_2(1 - p_3)p_2' + p_2p_3(1 - p_4)p_3' + p_2p_3p_4p_4' \tag{32}$$

We can use data from the empirical section to estimate $P^6(1, NBA)$. To get estimates for p_i' is simple: Consider all players who applied to the NBA draft after i years of college and look at how may were drafted into the NBA.[15] This is a good first approximation for p_i'. The p_i are more difficult to estimate. This would involve looking at all players who finished i years of college ball and seeing how many declared for the draft versus how many stayed for another year. Trying to estimate all the college players who didn't apply for the draft is challenging since there are many colleges in the nation with many college players on each team.

To solve this issue, consider p_2 as being a function of p_1'. That is, $p_2 = p_2(p_1')$. Recall, p_2 is the probability a rising sophomore remains in college for a second year and p_1' is the probability a rising sophomore makes it to the NBA after applying for the draft. It's reasonable to assume

$$\frac{dp_2}{dp_1'} < 0$$

That is, as the probability of being drafted increases, the player will be less and less likely to want to remain in college for a second year. We can imagine that $p_2(0) \approx 1$ since the player has no chance of being drafted and so it's likely he'll remain in college. Also, $p_2(1) \approx 0$ since the player will for sure be drafted and so he forgoes his second year to apply for the draft. This reasoning makes the analysis of this optimal stopping section more appropriate for an average collegiate player. A very talented player may still be reluctant to try for the draft just because he thinks he'll be drafted. The talented player would probably strive for a high draft position, whereas the average player would be happy to be drafted at all.
The last thing to notice about $p_2(p_1')$ is it's second derivative:

$$\text{risk neutral} \Rightarrow \frac{d^2p_2}{dp_1'^2} = 0 \qquad \text{risk averse} \Rightarrow \frac{d^2p_2}{dp_1'^2} < 0 \qquad \text{risk loving} \Rightarrow \frac{d^2p_2}{dp_1'^2} > 0$$

These three properties are summarized by Figure 10.

[15]Using my data set of 856 domestic players from 1989-2012 who remained early entry, 78% of freshmen were drafted, 66% of sophomores were drafted, and 54% of juniors were drafted. Also, 81% of high school seniors who applied were drafted. p_4' is difficult to estimate, but is likely to be well less than 50%.

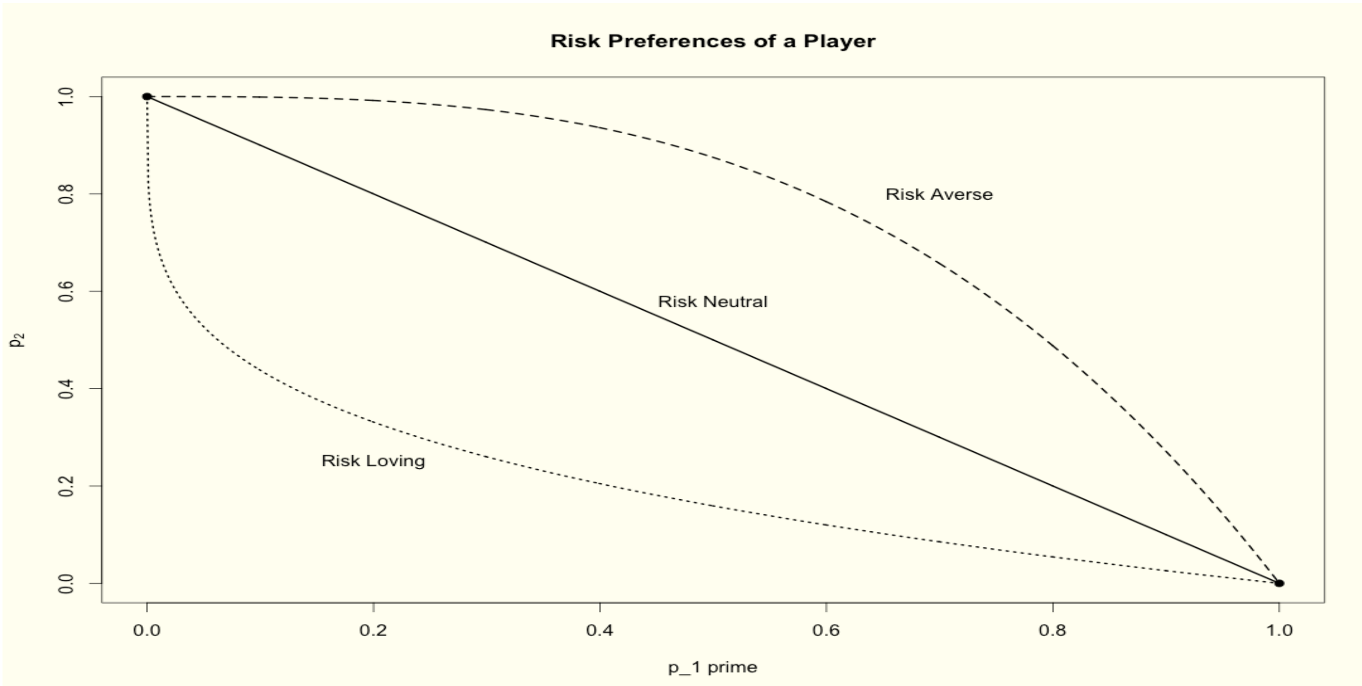

Figure 10: The shape of $p_2 = p_2(p_1')$ will vary depending on the player's risk preferences.

To work with explicit equations, I assume the following functional form of $p_2(p_1')$:

$$p_2(p_1') = 1 - (p_1')^k \qquad\qquad k > 0$$

If the player is risk loving, $k < 1$. If the player is risk neutral, $k = 1$. If the player is risk averse, $k > 1$. The argument above for reasoning that p_2 is a function of p_1' can be used to conclude that p_3 is a function of p_2' and p_4 is a function of p_3'. These other two functions behave exactly the same.
Hence (32) becomes

$$f(k) = \Big[1-\big(1-p_1'^k\big)\Big]p_1'+\big(1-p_1'^k\big)\Big[1-\big(1-p_2'^k\big)\Big]p_2'+\big(1-p_1'^k\big)\big(1-p_2'^k\big)\Big[1-\big(1-p_3'^k\big)\Big]p_3'+\big(1-p_1'^k\big)\big(1-p_2'^k\big)\big(1-p_3'^k\big)p_4'$$

which, after some algebra, simplifies to

$$f(k) = p_4' + (p_3' - p_4')p_3'^k + (p_1' - p_4')p_1'^k + (p_2' - p_4')p_2'^k + (p_4' - p_2')(p_1'p_2')^k \tag{33}$$
$$+(p_4' - p_3')(p_2'p_3')^k + (p_4' - p_3')(p_1'p_3')^k + (p_3' - p_4')(p_1'p_2'p_3')^k$$

$f(k)$ in (33) can be thought of as the chance of making it to the NBA eventually, where k is a measure of the player's risk averseness. Regardless of the values of p_1', p_2', p_3', and p_4', $f(0) = p_1'$ and $\lim_{k\to\infty} f(k) = p_4'$. Data from the previous page in footnote 15 showed $p_1' \approx 0.78 > p_2' \approx 0.66 > p_3' \approx 0.54 > p_4'$. Using these values in (33) and taking $p_4' = 0.3$ (without loss of generality), Figure 11 depicts $f(k)$.

Unlike the first optimal entry model where we assumed the player was talented (i.e. would be drafted if he were to apply to the draft), a not so talented player cannot be as picky about when he applies to the draft. The average player may have only one shot throughout his college career to make a feasible attempt to go pro. Therefore, if an average player wants to optimize his chances of making it to the NBA he should be as risk loving as possible. If he thinks he has any shot of making it to the NBA after completing any year of college, then he should drastically reduce the probability he stays in college for another year and strongly consider applying for the NBA draft.

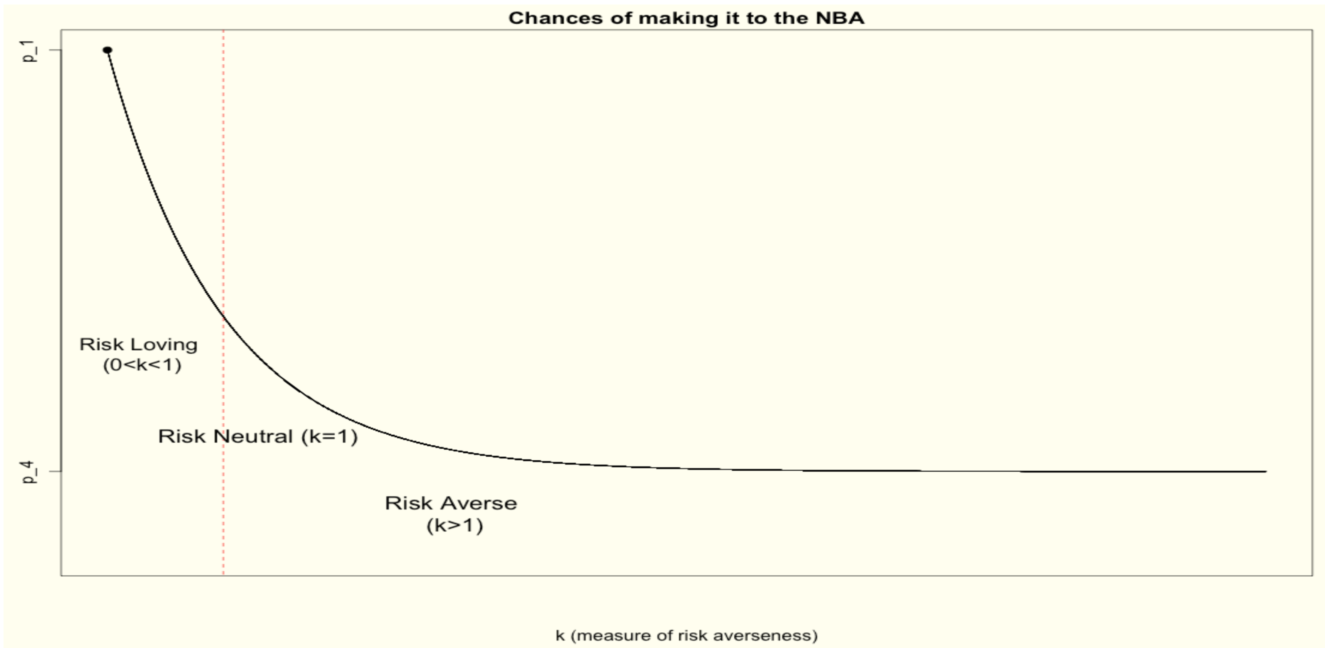

Figure 11: An average player's chances of making it to the NBA are highest when he's risk loving, lower when he's risk neutral, and lowest when he's risk averse.

7.1 Secretary Problem Approach

The classic Secretary Problem from optimal stopping theory can provide another template for when a collegiate basketball player should go pro. The secretary problem is as follows: An employer is looking to hire a single secretary from an applicant pool of n applicants. The applicants can be ranked from best to worst (i.e. there is a best applicant, a second best, etc). The employer does not know the talent of an applicant until that applicant is interviewed. The employer interviews applicants one by one hoping to find the best applicant. After an interview, the employer can hire that applicant and never get to see the remaining applicants, or *reject that applicant forever.*

The optimal solution involves automatically rejecting the first n/e applicants immediately after the interview, then hiring the first applicant who is better than all of those interviewed so far. If no such applicant exists, the employer hires the last applicant interviewed. This strategy can be shown to be successful in hiring the best applicant in the pool with probability $1/e$, as $n \to \infty$.

This solution method can be applied to a collegiate basketball player's decision about when to make the jump to the NBA. A player can opt to go pro after any one of four years. But after electing to go pro, he cannot try to go pro in later years because he has relinquished his college eligibility. At the end of each year of college ball the player assesses his "offer", which can be thought of as his chance of going pro, his potential earnings, and projected overall initial success in the NBA. This is analogous to interviewing an applicant. The player doesn't necessarily know if his skill level will be higher or lower the following year, or if his offer will be better or worse. The secretary problem solution implies the player should reject the first n/e applicants automatically. In the case of the player, $n/e = 4/e \approx 1.47$ offers. Therefore, the player should not go pro after freshman year. Instead, he should go pro in the subsequent year that gives him a better offer than the offer he received his freshman year. In the case his offers after sophomore year and junior year are worse than after freshman year, the player should go pro after senior year.

Using this approach it can be shown that the player successfully accepts the best offer 46% of the time. To see why, rank the four offers as 1, 2, 3, 4 where 1 corresponds to the best offer and 4 the worst. Since there are twenty-four permutations, all equally likely, just directly count the number of cases for which this secretary solution successfully has the player select 1. For example consider the order 2, 3, 1, 4. This means the player will receive the best offer following his junior year, second best offer after freshman year, third best offer after sophomore year, and worst offer after senior year. The solution says to reject the offer after freshman year, then select the first offer that's better than the offer freshman year. Since the offer after sophomore year is worse, it is rejected. Since the offer after junior year is better than the offer after freshman year, the player accepts this offer. He never plays a senior year of basketball and never sees what his offer after senior year would've been. Therefore this method was successful in picking out the best offer. In the twenty-four cases, the best offer is picked in 11 of them ($\approx 46\%$). In 7 cases the second best offer is picked, in 4 cases the third best offer is picked, and in only 2 cases (1, 2, 3, 4 and 1, 3, 2, 4) the worst offer is picked.

The above example is encouraging since the player would have been rather unlucky if he remained for his senior year. The player of course doesn't know if his offer after senior year will be better than after his junior year. Coming off a great junior year he may have been tempted to play senior year thinking that he would get an even better offer after senior year. This method results in the player stopping at the optimal time.

7.1.1 Nondeterministic Secretary Problem

A final thought introduces more randomness into the player's decision about when to go pro. In the previous section it was assumed that the number of applicants n was known. However, this does not have to be the case. Letting N be a random variable denoting the number of applications received, the employer would have to find a new optimal stopping rule to optimize the probability of selecting the best applicant (see Presman and Sonin 1973). Likewise for the player, $n = 4$ is not always the case. The player for example may suffer a long term injury junior year which ends his collegiate career. Hence N took on the value 2, since he only saw two offers, both of which he rejected. Using N also makes sense in the following way: Suppose in the previous section we defined offer as being the event where the probability of being drafted once a player enters the draft is nonzero. Then the player may not have offers every year. That is, there may be some years where he simply may not be talented enough, in which case if he applied to the draft he would almost certainly not be drafted. Thus N would denote the number of years in which he has a nonzero probability of being drafted.

8 Conclusion

This paper analyzes when a player should go pro through both an empirical and theoretical framework. The empirical section estimates returns to schooling by comparing two groups of players: high school seniors and freshmen drafted from 1989 to 2005 versus freshmen drafted from 2006 to 2012. A new law made 2005 the last year players could go directly into the NBA out of high school and so the latter group contains individuals who were forced to take an additional year of school. Comparing the earnings of the two groups shows that the "Post-law" group earned $580,000 more during their first year in the NBA. However, the regressions, which controlled for ability, experience, and draft position, found no evidence to support a significant difference in earnings.

The theoretical section showed players can optimize their career earnings by going pro when their skills are highest. Players who are especially talented should go pro as soon as they attain some minimum skill threshold, despite the possibility that their skills may be higher in later years. Players who are average should go pro if they receive any positive signals that indicate they have a chance of being drafted. That is, they should have risk loving preferences.

9 Appendix

Here I generalize section 5 to include the probability of the player going undrafted. This is mainly to show the robustness of the analysis in sections 5-7 and also allow for "average players" to be defined more broadly.

Let P_H, P_L be defined as before and let P_U be the probability the player is undrafted. If the player is undrafted, his earnings are zero: $E_H > E_L > 0 = E_U$. Assume $dP_U/dS < 0$ and that

$$P_H + P_L + P_U = 1$$

The optimization problem is

$$\max_y \mathbb{E}(e) = \max_y \Big[E_H P_H\big(S(y)\big) + E_L P_L\big(S(y)\big)\Big]$$

which becomes

$$\max_y \Big[E_H P_H\big(S(y)\big) + E_L\Big(1 - P_H\big(S(y)\big) - P_U\big(S(y)\big)\Big)\Big]$$

Differentiating with respect to y gives

$$\frac{d\mathbb{E}(e)}{dy} = E_H P_H'\big(S(y)\big)S'(y) - E_L P_H'\big(S(y)\big)S'(y) - E_L P_U'\big(S(y)\big)S'(y) = 0$$

$$\Rightarrow\ S'(y)\Big[E_H P_H'\big(S(y)\big) - E_L\Big(P_H'\big(S(y)\big) + P_U'\big(S(y)\big)\Big)\Big] = 0 \tag{34}$$

which holds when $S'(y) = 0$. To see why the term in the bracket cannot be zero, rearrange to get

$$\frac{E_H}{E_L} = \frac{P_H'\big(S(y)\big) + P_U'\big(S(y)\big)}{P_H'\big(S(y)\big)} = 1 + \frac{P_U'\big(S(y)\big)}{P_H'\big(S(y)\big)} \tag{35}$$

Since $E_H > E_L$, $E_H/E_L > 1$. But the term on the right in (35) is less than one since $P_U'(S) < 0$ and $P_H'(S) > 0$. Therefore, the maximum of $S(y)$ is the only candidate to optimize earnings. Letting y^* be such that $S'(y^*) = 0$ and $S''(y^*) < 0$, let's verify this is indeed the maximizer.
Differentiating (34),

$$\begin{aligned}\frac{d^2\mathbb{E}(e)}{dy^2}\Big|_{y=y^*} &= S'(y^*)\Big[E_H P_H''\big(S(y^*)\big)S'(y^*) - E_L\Big(P_H''\big(S(y^*)\big)S'(y^*) + P_U''\big(S(y^*)\big)S'(y^*)\Big)\Big] \\ &+S''(y^*)\Big[E_H P_H'\big(S(y^*)\big) - E_L\Big(P_H'\big(S(y^*)\big) + P_U'\big(S(y^*)\big)\Big)\Big] \\ &= S''(y^*)\Big[\big(E_H - E_L\big)P_H'\big(S(y^*)\big) - E_L P_U'\big(S(y^*)\big)\Big] < 0\end{aligned}$$

since $S''(y^*) < 0$ and the term in brackets is positive.
Lastly, check the boundary:

$$\mathbb{E}(e)\Big|_{y=y^*} > \mathbb{E}(e)\Big|_{y=y_B}$$

$$\iff E_H P_H\big(S(y^*)\big) + E_L\Big(1 - P_H\big(S(y^*)\big) - P_U\big(S(y^*)\big)\Big) > E_H P_H\big(S(y_B)\big) + E_L\Big(1 - P_H\big(S(y_B)\big) - P_U\big(S(y_B)\big)\Big)$$

$$\iff E_L\Big[P_H\big(S(y_B)\big) - P_H\big(S(y^*)\big) + P_U\big(S(y_B)\big) - P_U\big(S(y^*)\big)\Big] > E_H\Big[P_H\big(S(y_B)\big) - P_H\big(S(y^*)\big)\Big] \tag{36}$$

For (36) to hold, each term in brackets must be negative.
To make the term in the right bracket negative, it's necessary that $S(y^*) > S(y_B)$.
The term in the left bracket of (36) is $P_H\big(S(y_B)\big) - P_H\big(S(y^*)\big) < 0$ and $P_U\big(S(y_B)\big) - P_U\big(S(y^*)\big) > 0$. To see how the first difference dominates, recall

$$P_H + P_L + P_U = 1 \quad\Longrightarrow\quad \frac{dP_H}{dS} + \frac{dP_L}{dS} + \frac{dP_U}{dS} = 0 \quad\Longrightarrow\quad \frac{dP_H}{dS} > \left|\frac{dP_U}{dS}\right|$$

since we assumed the derivative of P_L is nonzero. Therefore $S(y^*)$ optimizes expected earnings.
The player should go pro when his skills are highest.