

Road to the Hall of Fame:
An Analysis of Time to Enshrinement for NBA Players

Kekona Sorenson
Statistics M213 Final Project
December 9, 2008

Abstract

Every fall, around the time the new NBA season starts, pundits begin analyzing the year's Hall of Fame class. One frequent discussion that comes up is how for certain players it was "long-overdue" while for others it was surprising "how quick they made it in." The conversation inevitably steers towards an argument on how influential the player was and citing the many accolades they received. They go on to attempt to rank each player relative to one another and make their cases for why one is better than the other. It is this process that this paper attempts to formalize. The conclusion is that the various league accolades, such as the MVP award, can be used in conjunction with a cohort rank variable in a semi parametric event history model to explain time-to-enshrinement for NBA players.

It is safe to say that when a player receives an award, they have garnered the respect of the league for that year. Does this, however, translate into a larger hazard for NBA Hall of Fame (HOF) induction? One important thing to note about HOF induction is that a player's ability to get enshrined does not exist in a vacuum. Due to a yearly quota of HOF inductees, a player's performance is measured against all others that are eligible that year. So does this mean that the "higher" the player is ranked in a given class, the higher the hazard of induction? What about when player awards are considered? These are the main questions this paper aims to address. First, the process of how a player gets enshrined is introduced, and then there is a discussion on how this is an event history type question, then other research in the area is introduced, then the data is presented, the model and choice of covariates are justified and finally the results are presented.

The HOF enshrinement process is vote based. A player is nominated and completes an application. The application consists of "a completed nomination form procured from the NBA HOF, and news clippings, magazine articles or other informative, factual data about the candidate."¹ Once the application is complete, it is presented to a committee where they vote. If the player secures enough votes, then they are enshrined². One unique characteristic of the NBA HOF is that they not only take in NBA players, but also coaches, referees, contributors, players from abroad and women players. While this mix of types of basketball contributors makes for an interesting analysis, they will not be considered here. The focus of this paper will be on male NBA players and the length of time until they become enshrined in the HOF.

¹Guidelines For Nomination and Election Into the Naismith Memorial Basketball Hall of Fame. From <http://www.hoophall.com/ot/bhof-enshrinement.html> (accessed 12/1/2009)

² IBID

This process fits event history well. The event in question is induction to the HOF for male NBA players. According to the HOF, a player becomes eligible for HOF consideration five years after they retire³. All retirement and induction years dates are well defined in the data so from an analysis standpoint, this is fairly easy to identify. The clock is started five years after retirement and stopped when the player gets inducted to the HOF, or becomes censored (data gathered in 2007). Since data accounts for all years and since the NBA officially started in 1946 and the first HOF class was in 1959 there is no problem with left censoring. However there is right censoring and it is defined as not yet inducted to the HOF at the time of data collection. Also since a player can only become a HOF member once in their lifetime, there are no repeated events. The time to event is the number of years from HOF eligibility to induction or censoring. Since NBA HOF inductions only occur once per year, discrete-time methods for analysis are appropriate.

Other research in this area has been restricted to baseball HOF or to NBA career length. First, Todd Jewell wrote a paper titled “Voting for the Baseball Hall of Fame: The Effects of Race on Election Date”⁴ which analyzes the effect of race on hall of fame enshrinement in baseball. He concludes that there was no racial effect on baseball hall of fame enshrinement, and cites that this was consistent with previous studies. The main motivations for these studies were to isolate any racial bias in hall of fame induction. While this topic is interesting and has immense sociological value, it is not the aim of this paper. There has also been some event history work done to determine

³ IBID

⁴ Jewell, Todd (2003), “Voting for the Baseball Hall of Fame: The Effects of Race on Election Date,” Jan 2003, *Industrial Relations Journal*, pg 87-100

how draft order affects playing time and survival for rookies in the NBA⁵ and a duration analysis of career length⁶. These both attempt to isolate patterns in NBA career longevity. For this analysis there is less interest in what determines length of a player's career and more in the post-career length of time to basketball hall of fame induction.

One of the advantages to analyzing sports related topics is the availability and usability of the data. The data was gathered and assembled from www.databaseBasketball.com, www.nba.com, and www.hoopshall.com. All sources contained quality-checked comma separated files that were easily imported and analyzed in STATA. The data consists of all NBA players from 1946 until 2007 and the available variables include career averages for points, assists, blocks, steals and rebounds, flag for NBA championship, flag for MVP award, flag for selection to the All NBA first team, flag for selection to All Defensive team, HOF monitor score and of course the retirement year and year of HOF induction. HOF prediction score was created by www.DatabaseBasketball.com and is used as an indicator of likelihood to be admitted to the HOF. For the purposes of this paper, this value is used as a crude way to generate yearly time-varying cohort ranks that vary with time and across players.

The Cox proportional hazards model is appropriate for this analysis because it allows analyzing the effect of several covariates on surviving. Also, there is no theoretical justification for choosing a baseline hazard for this analysis. Since the interest of this paper is in analyzing the affects awards on time to event, all awards that are offered by the NBA were researched and screened. Now this is not to say that every award made it in the model, but rather a subset was selected that indicated the kind of

⁵ Staw, B. M., & Hoang, H. (1995). Sunk costs in the NBA: Why draft order affects playing time and survival in professional basketball. *Administrative Science Quarterly*, 40, 474-493.

⁶ Groothuis and Hill (2007), "Exit Discrimination in the NBA: A Duration Analysis of Career Length" *Economic Inquiry*, Vol 42 Issue 2 pg. 341-349 26 Mar 2007

player that should be in the HOF. Ultimately he had to be a respected, all-around type player. More precisely, he had to have the ability to make players around him better (or be a team player), play good offense and good defense all while garnering the respect from fellow players, coaches and the league. Below are the covariates that capture this type of a player.

The pinnacle individual achievement that a player can attain while playing is the NBA MVP award. This variable is included because this is the closest measure of all-around player performance as viewed by the league. It does not just depend on statistics, but rather overall achievement, and the ability to play well with. An exploratory analysis of the data shows that every HOF eligible player who has won an NBA MVP award has gone on to be admitted to the HOF. Since only one is awarded per year, a flag for selection the All NBA First Team is also included. This award is essentially accounts for the “runner-ups” in the voting.

To further capture team playing ability, a variable that indicates whether or not the player has ever won an NBA championship in their career is included. The goal of this indicator is to identify good team players. Since basketball is a team sport, it is essential to be recognized as a contributing member of a winning team. For individual defensive success a flag for election to the All Defense Team is included. As an additional proxy for individual offensive success career scoring average is included. This is a quantitative variable that measures the player’s regular season career scoring average.

Finally, cohort rank was included because it is an important indicator of strength of class. It is a time-varying covariate that ranks a player (with a rank of 1 representing

the highest HOF monitor score for that year) each year relative to the other at risk players. This rank was derived from the HOF Monitor Score and is updated from year to year depending on who is eligible for HOF induction that year. Due to the yearly quota one can expect a player's probability of enshrinement to depend on the quality of players available at the time. This variable aims to model this phenomenon.

HOF Monitor Score was used to generate the cohort rank. This score was generated by www.databasebasketball.com and was intended to be used as a rough guide for a player's chances for induction to the HOF. It uses a weighted linear combination of various factors in determining the score, including career statistics, efficiency scores, accolades and player value. This was the most feasible and objective measure available to capture cohort rank. This score was also used as a systematic way to trim down the data by removing players that were obviously not HOF contenders. The majority of players that were removed were players who played only one year then got cut, got injured or made no significant contribution to their teams or the league. There may be a slight bias towards more current players involved with this, but due to the sheer number of players in the data and the scarcity of HOF inductees, something had to be done to reduce the data.

The descriptive summary of the awards flags and HOF induction are listed in table 1. Within this table there is also a chi-square test for variable independence. All of the covariates that are used were statistically significant indicating that there is some evidence of variable dependence between each award and HOF status. Since the chi-square test could have misleading results if the cell sizes are too small, a 1 and 2 tailed fisher exact test was conducted. These tests hold better when a table has small cell

values. A look at these results further confirms the chi-square results. These were just brief tests to further confirm the evidence of dependence between the award flags and HOF induction.

Table 1: Independence tests between awards and HOF induction

Award	Result	HOF Flag		Chi-Square Test for Independence		Fishers Exact Test	
		No	Yes	Score	P-Value	One-Sided	Two-Sided
MVP Flag	No	1,865	0				
	Yes	65	17	390.06	0.000	0.000	0.000
All-NBA First Team Flag	No	1,844	21				
	Yes	35	47	735.80	0.000	0.000	0.000
All Defensive Team	No	1,838	27				
	Yes	67	15	105.60	0.000	0.000	0.000
NBA Championship	No	1,628	237				
	Yes	29	53	167.08	0.000	0.000	0.000

To confirm that the Cox model and the covariates are appropriate, the assumptions were tested. Note that to account for the any possible endogeneity in the time-dependent variable cohort rank, a lagged version of cohort rank was used.

The proportional-hazard assumption was the first thing tested. According to Cleves, “under the proportional-hazards assumption, the plotted curves should thus be parallel.”⁷ Looking at figures 3a-3d in the appendix, it can be seen that all plots are roughly parallel indicating evidence in favor of the proportional hazard assumption for each covariate. One exception is the All Defensive team flag that shows some slight non-parallel behavior towards the end of the time frame. Since it’s mostly parallel, we

⁷ Mario Cleves (2008) , An Introduction fo Survival Analysis Using Stata, (College Station: Stata Press) 204

will still include it in our analysis⁸. To test the assumption for the cohort-rank variable, a likelihood ratio test was conducted.

“By using the `lrtest` commands it is possible to tests all the time dependent covariates together by comparing the smaller model without any time dependent covariates to the larger model that includes all the time dependent covariates.”⁹

By using this method we generate a coefficient for cohort lag that is insignificant. Since the coefficient for cohort lag in the smaller model is not significant, it can be concluded that there is evidence in favor of the proportional-hazards assumption for the effect of cohort lag¹⁰. A link test was run to further test the proportional-hazards assumption. The results of this test were that the coefficient on the squared linear predictor was insignificant. Thus there is sufficient evidence to claim that the proportional-hazards assumption holds.

Now that the proportional-hazards assumption has been satisfied, overall model fit can be assessed by running a likelihood ratio test against the null model. After confirming that the number of observations between the null and full model were correct, each log-likelihood was calculates. The null model generated a log-likelihood of -495.21 compared to -280.04 from the full model. The difference between these two times negative two is 430.35. This has an approximate chi-square distribution with six degrees of freedom. This produces statistically significant results indicating that there

⁸ IBID

⁹ SAS FAQ: Tests of proportionality in SAS, Stata, R and SPLUS. UCLA: Academic Technology Services, Statistical Consulting Group. From http://www.ats.ucla.edu/stat/sas/faq/test_proportionality.htm (accessed December 4, 2008).

¹⁰ IBID

is evidence to believe that the full model is better than the null model. There is now sufficient reason to believe that the full Cox regression model is appropriate.

When running the Cox model, the exact partial method for ties was used. The results of the final model show that not only are each of the hypothesized covariates statistically significant at the 5% level, but they also act in accordance with what we expect theoretically. To further confirm the results, a cloglog model with indicators for time was conducted and it was confirmed that all signs, significance levels and coefficient values were similar. As can be seen in table 2, all hazard ratios, except for cohort rank, are greater than one indicating that the hazard rate for players with these awards exceed that of players who have not won the awards.

For variables MVP, All NBA First team, Championship and All NBA defensive team it is expected that if a player has these items on their application they would become more attractive to the HOF selection committee and thus have a higher hazard rate compared to those who do not have the award. Specifically, the interpretations of these ratios is as follows: if a player has won an MVP award, the hazard for HOF induction increases by 389% relative to those who have not won an MVP, for a player who has won an NBA championship the hazard for HOF induction increases 153% relative to those who have not won one, for a player who was a member of the All-NBA Defensive Team the hazard for HOF induction increases 108% relative to those who were not members and for a player who was a member of the All-NBA First Team the hazard for HOF induction increases 234% relative to those who have never made that team. These dramatically large hazard rates indicate that winning these awards are a dependable way for a player to be put at a higher hazard of HOF induction.

For Career Average in Points we see that for each unit increase in career average, the hazard for HOF induction increases by 8%. This was largely expected because higher offensive output generally indicates a better NBA player. For the time-varying covariate cohort rank, we have a hazard ratio that is 0.650. This means that for every unit increase (lower rank) in rank, the risk of HOF enshrinement decreases by 35%. This inverse relationship was indeed expected. If a player is ranked lower (or as rank increases from, say, 2 to 3) relative to other at risk player, it is expected that they would have a lower hazard rate. Each of these interpretations are, of course, *ceteris paribus*.

Table 2: Results from Cox Regression – Final Model

Number of Observations	43,253		
Number of Subjects	1,897	LR Chi2(6)	430.35
Number of Failures	79	Prob > Chi2	0.0000
Covariate	Hazard Ratio	Z-Score	P-Value
MVP Flag	4.260	3.33	0.001
NBA Championship Flag	2.531	3.46	0.001
All-Defensive Team Flag	2.081	1.98	0.048
All-NBA First Team Flag	3.342	4.14	0.000
Career Average for Points	1.080	3.09	0.002
Cohort Rank	0.650	-5.06	0.000

The hazard plots for the overall model, the baseline model and the effect of each covariate are listed in the appendix. In the overall hazard plot one can see a noticeable increase from zero to about seven years, and then a decrease from year seven to year seventeen then it flattens out until around thirty-seven years then finally descends. The initial increase can be interpreted to mean that from initial eligibility to about seven years after eligibility the intensity at which HOF enshrinement is occurring is increasing, while the decrease can be thought of as the intensity at which HOF enshrinement is occurring is decreasing. For a NBA player this makes sense. It is

expected that within the first few years after eligibility, player's accomplishments are fresh in the minds of the voting panel, and as a player gets farther from initial eligibility, it is expected that they become more forgotten. Also, if a player truly is great in the minds of the panel, then they want to get them in as soon as possible.

The analysis here provide estimates of the effects of various NBA awards, career scoring average and cohort rank on time to HOF induction. A common-sense approach was taken and in the end came out with common-sense results. It was confirmed that if a player received awards from the league during their career, then they have a dramatically higher hazard rates relative to those who have not won one. Further, when accounting for a player's strength relative to the other eligible players, it was discovered that a player who is higher ranked in the election year, has a higher hazard.

Appendix:

Figure 1a-1f. Estimated hazard funtions

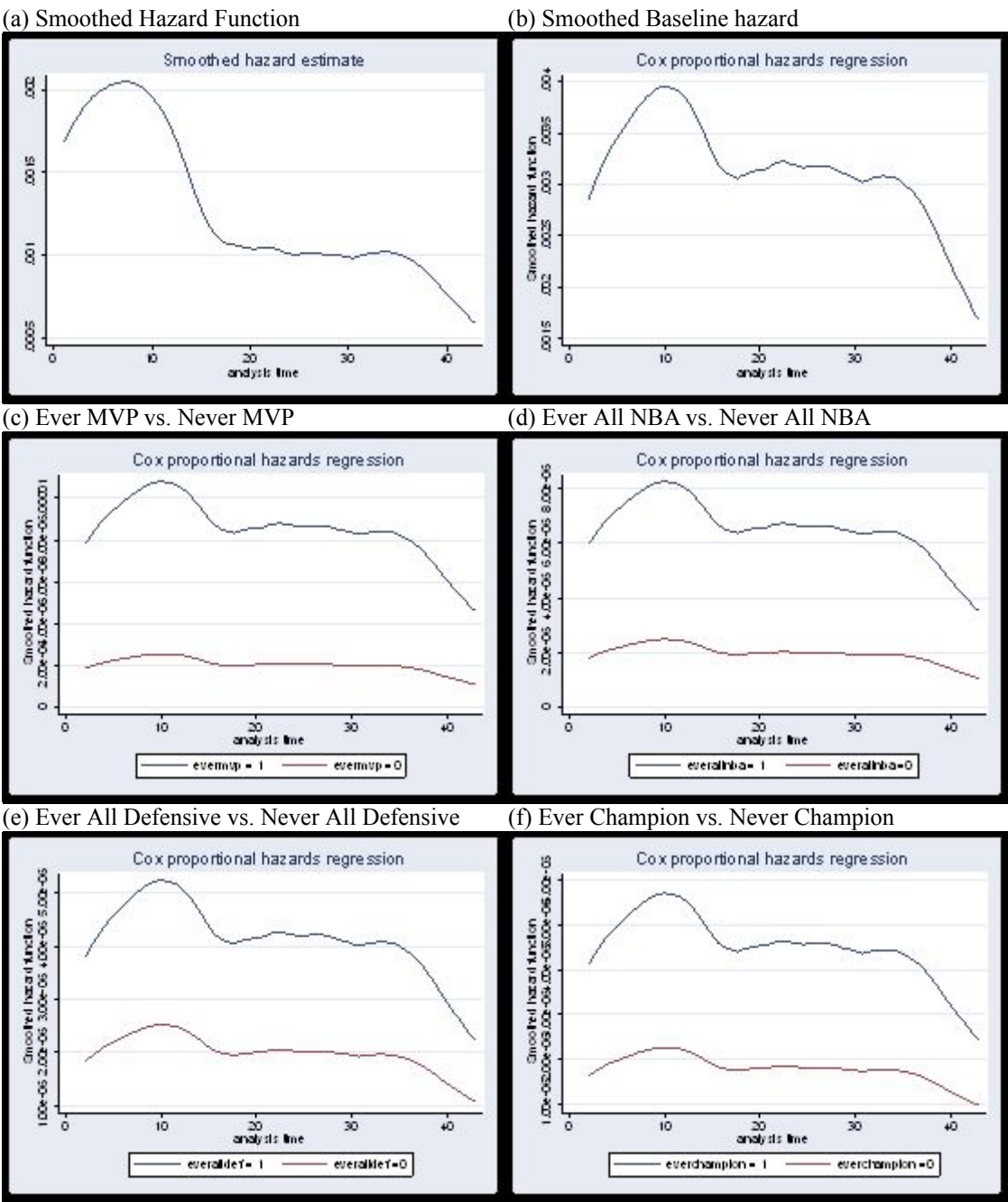


Figure 2. Estimated baseline cumulative hazard

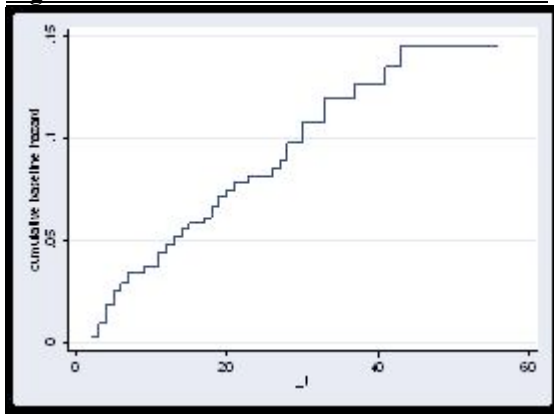
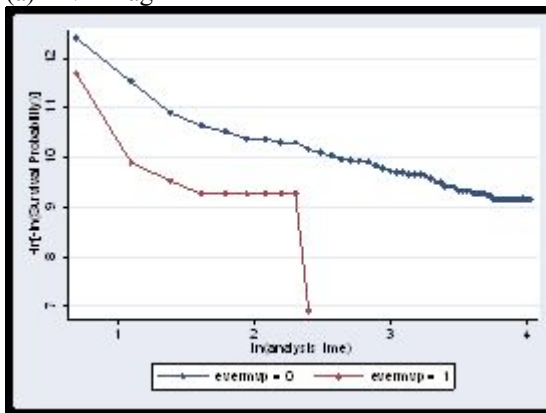
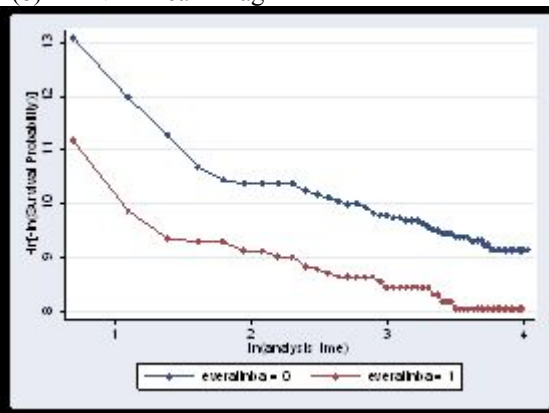


Figure 3a-3d. Test of the proportional-hazards assumptions

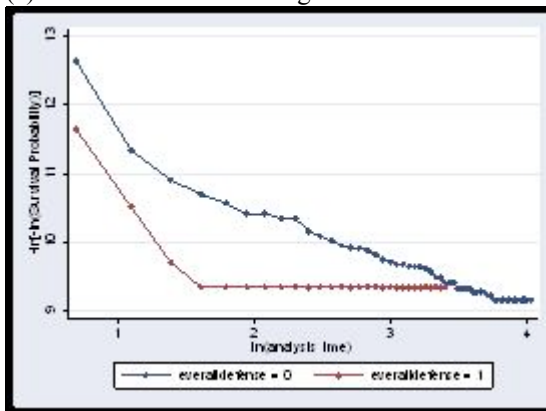
(a) MVP Flag



(b) All NBA Team Flag



(c) All Defensive Team Flag



(d) NBA Championship Flag

